nature plants

Article

https://doi.org/10.1038/s41477-023-01595-7

A designer synthetic chromosome fragment functions in moss

Received: 19 September 2023

Accepted: 22 November 2023

Published online: 26 January 2024

Check for updates

Lian-Ge Chen[1,13], Tianlong Lan[2,3,4,13], Shuo Zhang[1,5,13], Mengkai Zhao[6,13], Guangyu Luo[7], Yi Gao[2], Yuliang Zhang[1,5], Qingwei Du[1,5,8], Houze Lu[9], Bimeng Li[2], Bingke Jiao[1,5], Zhangli Hu[6], Yingxin Ma[7], Qiao Zhao[7], Ying Wang[2] ✉, Wenfeng Qian[1,2] ✉, Junbiao Dai[6,7,10] ✉ & Yuling Jiao[1,4,5,11,12] ✉

Rapid advances in DNA synthesis techniques have enabled the assembly and engineering of viral and microbial genomes, presenting new opportunities for synthetic genomics in multicellular eukaryotic organisms. These organisms, characterized by larger genomes, abundant transposons and extensive epigenetic regulation, pose unique challenges. Here we report the in vivo assembly of chromosomal fragments in the moss *Physcomitrium patens*, producing phenotypically virtually wild-type lines in which one-third of the coding region of a chromosomal arm is replaced by redesigned, chemically synthesized fragments. By eliminating 55.8% of a 155 kb endogenous chromosomal region, we substantially simplified the genome without discernible phenotypic effects, implying that many transposable elements may minimally impact growth. We also introduced other sequence modifications, such as PCRTag incorporation, gene locus swapping and stop codon substitution. Despite these substantial changes, the complex epigenetic landscape was normally established, albeit with some three-dimensional conformation alterations. The synthesis of a partial multicellular eukaryotic chromosome arm lays the foundation for the synthetic moss genome project (SynMoss) and paves the way for genome synthesis in multicellular organisms.

Following the completion of the sequencing of many entire genomes, advances in DNA synthesis techniques have propelled genome synthesis to the forefront of scientific exploration. Viral and bacterial genomes as well as yeast chromosomes have been engineered and reassembled to facilitate synthesis-based techniques, such as accelerated evolution, multiplex gene deletions and the introduction or elimination of genetic codons[1–7]. However, genome synthesis in multicellular organisms remains uncharted territory.

The advent of multicellularity was accompanied by a substantial increase in gene number and genome size[8]. Consequently, genome synthesis in multicellular organisms poses a formidable challenge due to the experimental difficulties associated with the transformation and assembly of large DNA fragments as well as regeneration from transformed cells. The rapid expansion of epigenetic diversity accompanying multicellularity[9] introduced yet another challenge to genome synthesis. Thus, the implications of artificial genome simplification in the epigenetic landscape remain uncertain.

The amplification of genome size in multicellular organisms is predominantly attributed to the disparate and often copious quantities of repetitive DNA derived from transposable elements (TEs). Two fundamentally divergent perspectives exist regarding the ubiquity of TEs in multicellular eukaryotic genomes[10,11]. It has been postulated

A full list of affiliations appears at the end of the paper. ✉e-mail: yingwang@ucas.ac.cn; wfqian@genetics.ac.cn; daijunbiao@caas.cn; yuling.jiao@pku.edu.cn

https://doi.org/10.1038/s41477-023-01595-7

that TEs are slightly deleterious but not sufficiently so to warrant effective removal by natural selection, especially in small populations[12,13]. Conversely, it has been speculated that TEs are indispensable for chromosomal integrity and the survival of the organism[14,15]. The chromosomal-level elimination of TEs should help solve this conundrum. If the comprehensive removal of TEs does not affect the wild-type (WT) phenotype, the optimization of multicellular organisms can potentially be achieved through synthetic genomics approaches, even if it may not be attainable through natural selection alone.

The early terrestrial plant *Physcomitrium* (*Physcomitrella*) *patens* is a well-established model organism among non-seed plants owing to its short growth cycle, ease of regeneration and dominance of the haploid generation[16]; it represents a suitable testing ground for genome synthesis in multicellular organisms. As *P. patens* is a broadly used model in evolutionary developmental and cell biological studies, its 481.75 megabase (Mb) genome has been fully sequenced, with 26 chromosomes[17–19]. Notably, the *P. patens* genome shows a relatively high TE content of ~60% and a sophisticated seed plant-like epigenetic landscape[20], rendering it suitable for testing genome simplification strategies. Importantly, *P. patens* possesses efficient homologous recombination[21] and a remarkable protoplast regeneration ability, both of which are essential for genome manipulations. Moreover, *P. patens* has been used for decades as a versatile synthetic biology chassis for expressing recombinant therapeutic proteins and small natural products of commercial value[22]. Therefore, we selected *P. patens* as a platform for exploring synthetic genomics in multicellular organisms[23].

Results

Genome design

In the initial phase of a synthetic moss genome project (SynMoss), we established design principles for the moss genome (Supplementary Box 1) with the objective of substantially reducing repetitive sequences while preserving the WT phenotype[1,3]. We further implemented an inhouse computational pipeline to standardize genome design. In brief, we retained all coding sequences as well as upstream and downstream regulatory regions but replaced all TAG/TGA stop codons with TAA codons, liberating the first two codons for prospective genetic code expansion[6]. Furthermore, we devised PCRTags, which use primers positioned within or across deleted regions, to expedite a polymerase chain reaction (PCR)-based assay for rapidly distinguishing WT and synthetic sequences.

The short arm of chromosome 18 (chr. 18L) was chosen as it is the shortest chromosome arm in the *P. patens* v.3.3 genome assembly. However, the new assembly, which finished as this work was ongoing, identified an extra 400 kb region at the telomere end of chr. 18L (ref. 19). We designed a simplified, condensed 197,892 basepair (bp) configuration of the updated 1,143,148 bp chr. 18L, removing >80% of the DNA content. This design procedure was conducted in two distinct phases (I and II), with minor variations in sequence design concerning the elimination of intergenic regions, the development of PCRTags and the insertion of LoxPsym sites (Supplementary Box 1 and Methods). During phase I, we also altered the position of Pp3c18_90 to facilitate the assembly of mid-chunks (Extended Data Fig. 1) and the design of PCRTags and removed Pp3c18_140, a gene deemed non-essential for growth, from the designed genome. The number of genes and genomic elements in the WT and synthetic sequence fragments are delineated in Supplementary Table 1. The gene functions and expression information[24] related to phase I are shown in Supplementary Table 2.

Homologous recombination of large fragments

The actual in vivo replacement of WT sequences with redesigned large synthetic fragments presented a formidable technical challenge. The in vivo assembly of multiple DNA fragments is possible in *P. patens* but has been tested only for fragments with sizes up to 5 kb in total[25]. For the replacement of larger chromosomal regions, we assumed and subsequently demonstrated that longer homologous arms would

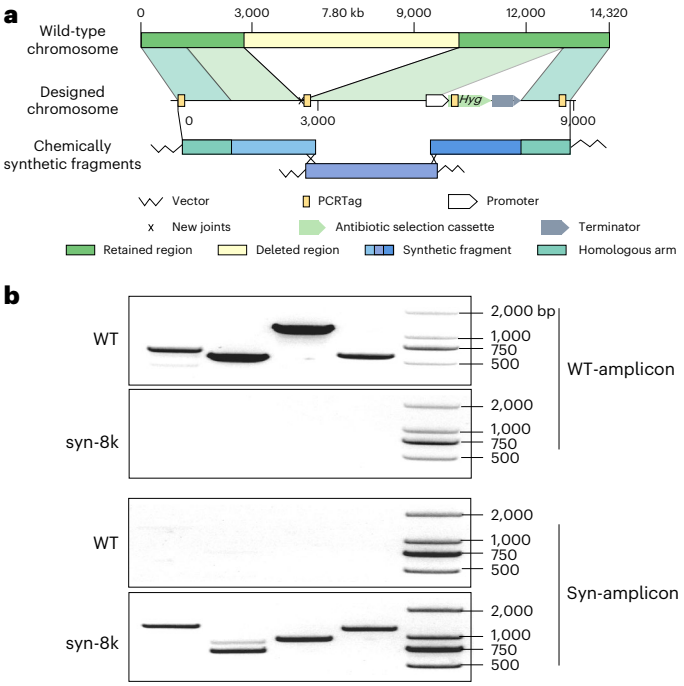


Fig. 1 | The design and characterization of the 8.7 kb pilot experiment. **a**, Schematic diagram of the design and assembly. Top, a 14,322 bp WT sequence in the short arm of chromosome 18 was modified by deleting a 7,798 bp non-gene sequence and adding a hygromycin gene (*Hyg*). Bottom, the designed sequence was divided into three chemosynthetic fragments of ~3 kb each, which were then linearized and transferred into *P. patens* to replace the corresponding WT sequence. **b**, PCRTag analysis. A total of four recombination replacement sites were detected, including two sites between two chemosynthetic fragments and two sites between chemosynthetic fragments and WT sequences. Synthetic sequences could only be detected in the obtained synthetic lines, while WT sequences could not be detected. Three independent experiments were conducted on each line using three independent samples and similar results were obtained.

facilitate the replacement of larger fragments. We first targeted an 8,718 bp region of chr. 18L with three overlapping 3 kb mini-chunk fragments (Fig. 1a). Each mini-chunk overlapped with its adjacent neighbour by 150 bp. The two free ends contained 1 kb regions homologous to the endogenous sequences in *P. patens*. A hygromycin-resistance cassette was placed next to one of the homologous arms. The ~3 kb mini-chunks were obtained through chemical synthesis and amplified in *Escherichia coli*. We transformed *P. patens* with an equimolar number of linearized mini-chunks through polyethylene glycol (PEG)-mediated transformation. Antibiotic-resistant lines were recovered and screened for the presence of synthetic sequences by PCR. We were able to identify lines with endogenous sequences replaced by in vivo-assembled mini-chunks, at a frequency of 30% (Fig. 1b), suggesting that replacing large chromosomal regions is possible. We further characterized one of these lines, named syn-8k, and found that it showed a WT-like phenotype at various developmental stages (Fig. 2a) and it had the same ploidy as the WT with most of its cells containing a haploid genome (Fig. 2b).

Strategy for synthesis and assembly

We next aimed to achieve sequence replacement at a larger scale and chose a 155,181 bp region for replacement with a 68,530 bp redesigned sequence spanning about one-third of the length of the redesigned chr. 18L (phase I design). We added a kanamycin-resistance cassette to the centromeric end (Fig. 3b). To ensure the success of the project, we used two strategies: in vivo mid-chunk assembly and single mega-chunk replacement.

https://doi.org/10.1038/s41477-023-01595-7

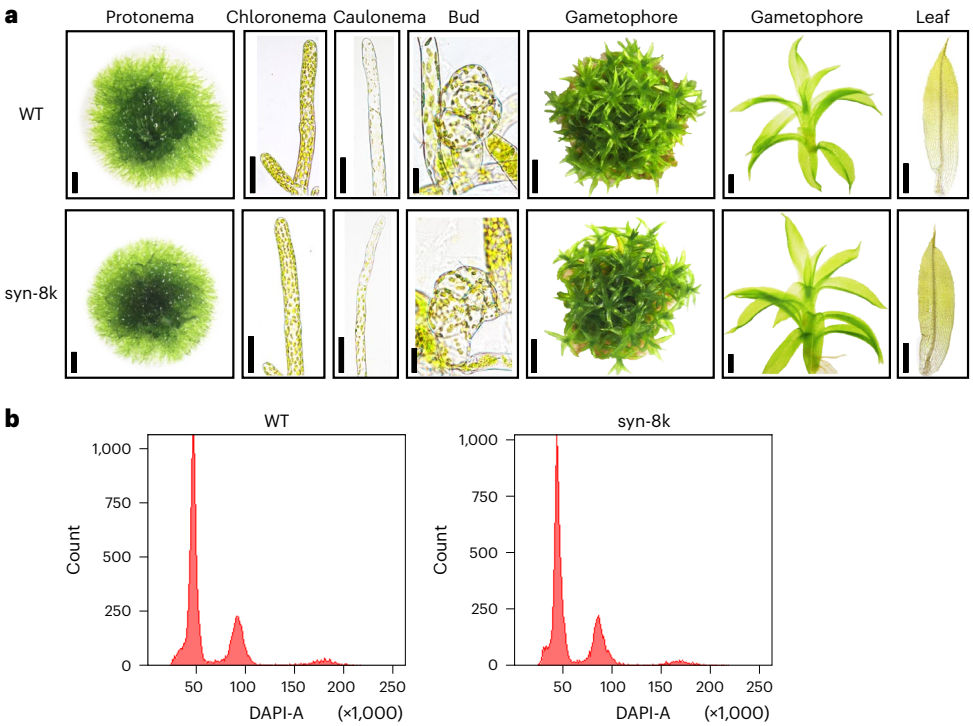


Fig. 2 | Characterization of syn-8k. a, Phenotypes of WT and syn-8k at various developmental stages. The figure shows the phenotypes of representative samples (more than three independent samples were observed for each stage and each line). Scale bars from left to right, 500 µm, 50 µm, 50 µm, 20 µm, 2 mm, 1 mm and 500 µm. **b**, The ploidy of WT and syn-8k was identical, as determined by flow cytometry.

We first designed three 20–30 kb mid-chunks (Extended Data Fig. 1), each of which overlapped with the next by an overhang of 1 kb. Homologous arms of 1 kb were included at the ends of the entire region. Mid-chunks were obtained by assembling 3 kb mini-chunks in yeast, followed by propagation in *E. coli*. Through PEG-mediated transformation into *P. patens*, antibiotic selection and PCR-based screening, we assembled two mid-chunks in vivo, which together replaced 83.6% of the entire 155,181 bp targeted region and the obtained synthetic line was named syn-50k (Fig. 3a and Extended Data Fig. 2a–c). The obtained lines were phenotypically indistinguishable from the WT (Extended Data Fig. 2d). Although ~1,000 antibiotic-resistant lines were screened, we were not able to obtain a line with all three mid-chunks simultaneously integrated into the genome.

We next assembled a single mega-chunk of 68.53 kb in yeast that contained a 1 kb homologous arm at either end (Fig. 3b). For this purpose, we applied a recently introduced pipeline for the construction of scar-free large DNA fragments[26]. In brief, mid-chunks were amplified in *E. coli* and assembled in yeast using a yeast centromere-containing bacteria artificial chromosome (YCp/BAC) vector. The assembled contig was again propagated in *E. coli* before PCR and restriction digestion verification. The linearized 68 kb mega-chunk was released from the YCp/BAC vector for PEG-mediated transformation (Fig. 3c).

After the PCR-based screening of ~1,000 lines carrying the selection cassette, we were able to identify potential lines in which the entire 155,181 bp targeted WT region was substituted by the synthetic sequences (Fig. 4a and Supplementary Table 1). PCRTag-based genotyping confirmed complete substitution in one line, which was named semi-syn18L (Fig. 4a). We applied flow cytometric analysis to ensure that the somatic cells of semi-syn18L were haploid, similar to those of the WT (Fig. 4c). In addition, we recovered several lines in which only partial replacement was achieved (Extended Data Fig. 3).

The whole-genome sequencing (WGS) of semi-syn18L reconfirmed the insertion of the synthetic sequence (Fig. 4b). The WGS data also showed that the synthetic sequence was not present in other genomic regions. When we scrutinized the WGS data, we indeed identified two single nucleotide polymorphisms (SNPs) in the synthetic region that differed from our design (Supplementary Table 3). Debugging is often necessary when a designed or unplanned substitution causes lethality or severe phenotypic changes[1]. Since the SNPs here are not located on genes and do not cause phenotypic changes, we did not correct them. Cotransforming protoplasts with CRISPR/Cas9 plasmids and linear oligonucleotide templates can efficiently (>30%) correct undesired SNPs in further synthesis projects[27].

Phenotypic characterization

We thoroughly compared the phenotypes of semi-syn18L with those of the WT under normal growth conditions or various stress challenges. *P. patens* germinates from a haploid spore, producing filamentous protonema cells through tip growth. Protonema cells can form branches to establish a two-dimensional protonema network. There are two types of protonema cells, chloronema and caulonema cells. The initial chloronemata transform into fast-growing caulonema cells. As the plant matures, gametophores appear on caulonemata-like side branches and form leaflets in a chiral pattern[28]. At the top of each gametophore, both male and female sexual organs form and produce sperm and eggs, which fuse upon fertilization to form zygotes and the zygotes further develop into sporophytes and produce spores. We found that all of these structures were normally formed in semi-syn18L, similar to the WT (Fig. 4d). Notably, spores were produced, suggesting that the life cycle was unaffected in semi-syn18L. We also challenged plants with stress conditions, including salt and osmotic stress. Semi-syn18L showed phenotypes comparable to those of the WT (Extended Data Fig. 4).

Epigenome profiling

Histone and DNA modifications are prevalent in *P. patens*, as in other multicellular organisms and are actively involved in cell fate

https://doi.org/10.1038/s41477-023-01595-7

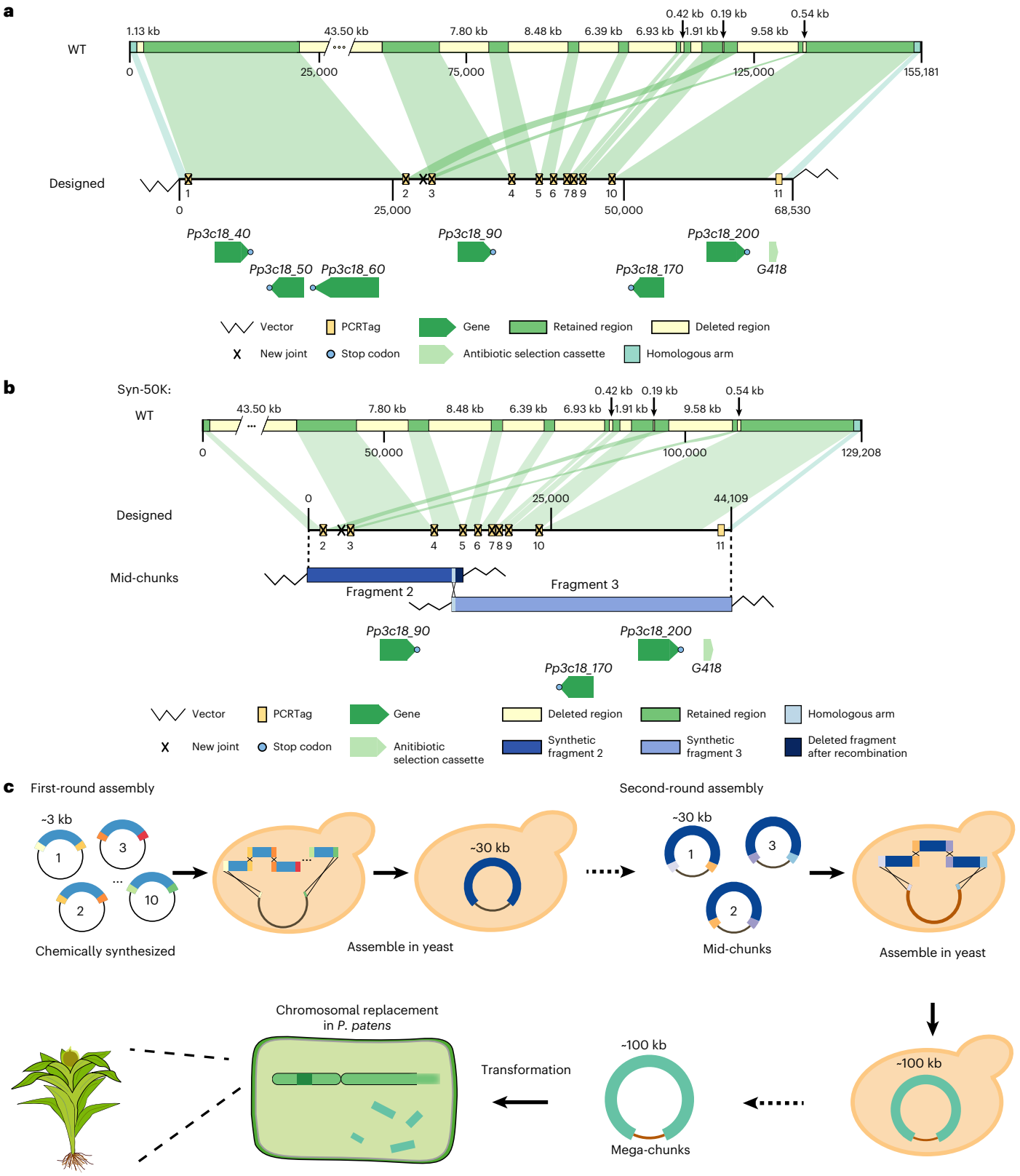


Fig. 3 | Genome design, synthesis and assembly. a, Schematic sequence diagram corresponding to semi-syn18L. The top section shows the 155,181 bp WT genome sequence, from which 86,651 bp was deleted and 68,530 bp was retained. The middle section shows the synthetic sequence and the corresponding gene distribution is shown below. **b**, Assembly replacement of two mid-chunks in syn-50k. The schematic diagram in the middle shows the design sequence of two 20–30 kb mid-chunks. The top section displays the corresponding WT genome sequences, while the bottom section illustrates the gene distribution pattern. **c**, Flow diagram for assembly and replacement. In yeast, we assembled ten 3 kb mini-chunks into an ~30 kb mid-chunk and subsequently assembled these three mid-chunks into a mega-chunk of ~100 kb. Finally, we transformed this ~100 kb mega-chunk into *P. patens* to replace the corresponding WT genome sequence.

https://doi.org/10.1038/s41477-023-01595-7

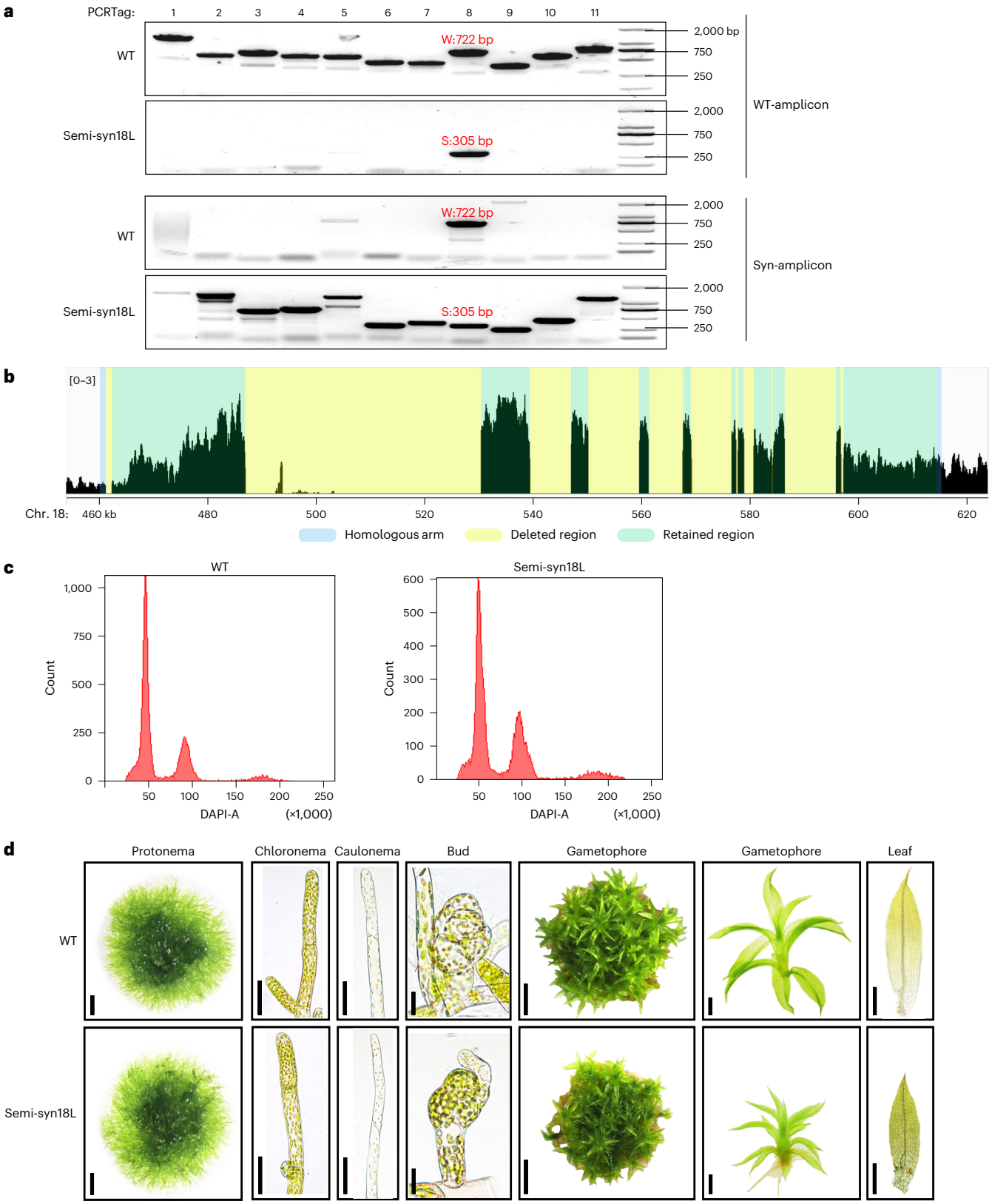


Fig. 4 | Characterization of semi-syn18L. a, PCRTag analysis. A total of 11 recombination replacement sites were detected, including nine sites in deleted regions and two in homologous arms. The red fonts indicate the different lengths of PCR products amplified from WT and semi-syn18L at PCRTag-8. For other PCRTags, only synthetic sequences, but not WT sequences, can be amplified from semi-syn18L. Three independent experiments were conducted on each line using three independent samples and similar results were obtained. **b**, Whole-genome sequencing analysis of semi-syn18L. No reads containing sequences from the deleted regions were found and reads were identified covering all new junctions. **c**, Flow cytometric analysis indicated that semi-syn18L is haploid, similar to the WT. **d**, The morphologies of WT and semi-syn18L at various developmental stages. The figure shows the phenotypes of representative samples. (More than three independent samples were observed for each stage and each line.) Scale bars from left to right, 500 μm, 50 μm, 50 μm, 20 μm, 2 mm, 1 mm and 500 μm.

https://doi.org/10.1038/s41477-023-01595-7

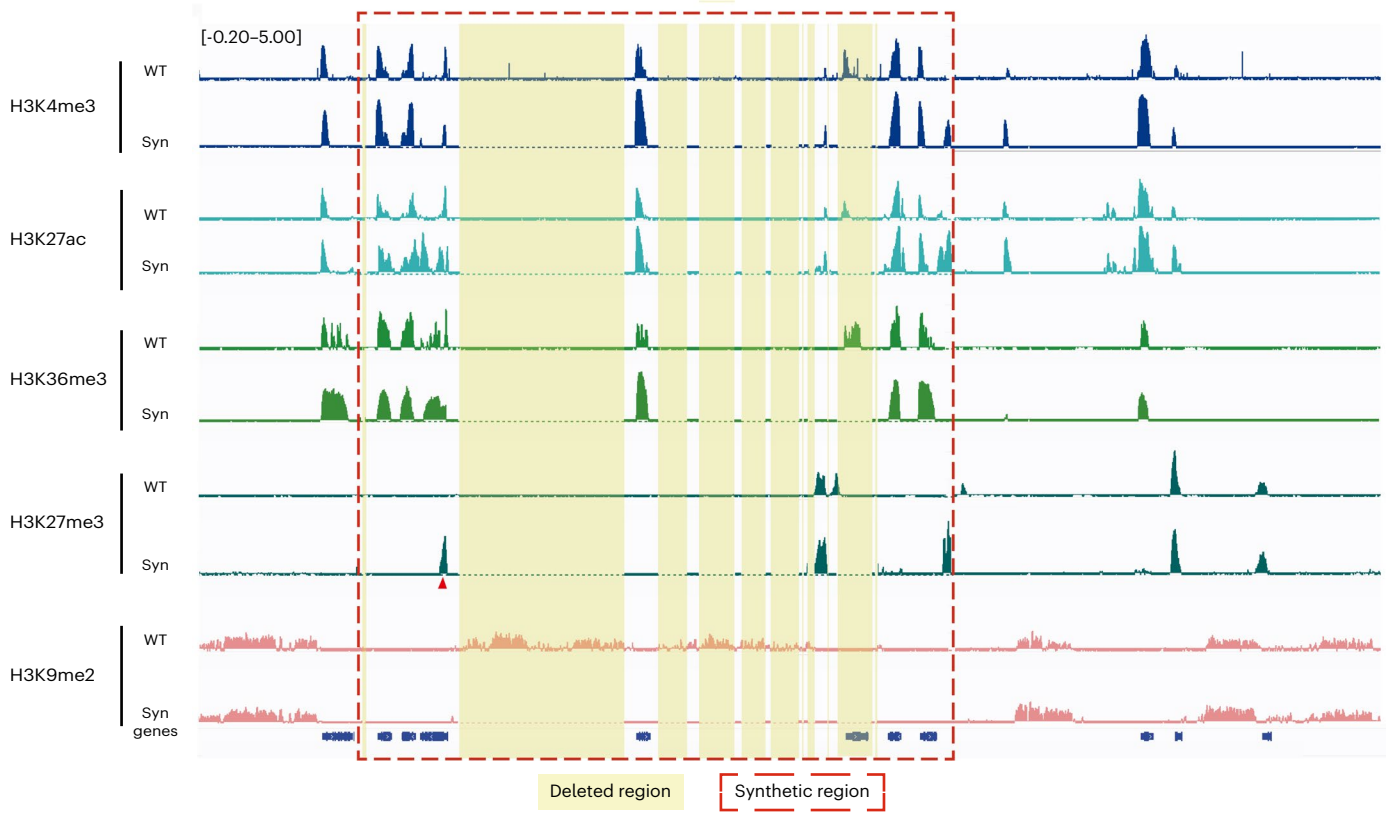


Fig. 5 | Distribution of five histone modifications (H3K4me3, H3K9me2, H3K27me3, H3K36me3 and H3K27ac) spanning the synthetic region. The signals shown were normalized using the BPM method (bins per million mapped reads, same as transcripts per million (TPM) in RNA-seq) after deducting the input. The deleted area is indicated by a dotted line to align the peaks. Note that the additional peaks in semi-syn18L at the right edge of the replacement region are associated by the insertion of the selection cassette. The red arrow indicates newly emerged H3K27me3 peak in semi-syn18L.

specification[20]. Using chromatin immunoprecipitation followed by deep sequencing (ChIP–seq), we obtained the genome-wide chromatin landscape of five important histone modifications (H3K4me3, H3K9me2, H3K27me3, H3K36me3 and H3K27ac) for both semi-syn18L and the WT (Fig. 5 and Extended Data Fig. 5). Pairwise comparison indicated that the synthetic line showed highly similar epigenetic marks to the WT on the whole genome (Extended Data Figs. 5 and 6b), with only a few differences (Supplementary Table 7). Epigenetic marks were established completely de novo on the synthetic fragments, which were amplified in *E. coli* and presumably lack epigenetic marks. Within the synthetic region, we found that H3K9me2 marks were almost completely removed (Fig. 5 and Extended Data Fig. 6c). Because repetitive elements are characterized by strong enrichment of H3K9me2 marks, the lack of H3K9me2 in the synthetic region was probably attributable to the removal of repetitive elements. Likewise, the levels of DNA methylation, which is preferentially associated with TEs[29], were substantially reduced in the synthetic region (Fig. 6b). Notably, preserved intergenic regions in the synthetic segment also showed greatly reduced levels of repressive epigenetic marks (Extended Data Fig. 6c), more prominently for H3K9me2 marks than for DNA methylation. Histone marks that are correlated with active transcription, including H3K4me3, H3K36me3 and H3K27ac, showed a distribution pattern consistent with the WT in the synthetic region but had slightly increased peak values. We also performed the assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and observed a moderate increase in chromatin accessibility for genes located in the synthetic region and in neighbouring region (Fig. 6a), which may be related to changes in chromosome structure caused by replacement. Consistently, RNA-seq analysis indicated that several genes within the region presented higher expression in semi-syn18L (Supplementary Table 8). Nevertheless, in the synthetic region, we also detected slightly higher levels of the H3K27me3 modification, which normally accumulates at silenced gene loci. In addition, a new H3K27me3 peak was located upstream of Pp3c18_60, whose expression, on the contrary, slightly increased.

Three-dimensional chromatin organization

Next, we profiled three-dimensional chromatin organization patterns using Hi-C in the WT and semi-syn18L. We identified many pairwise-interacting chromatin domains resembling chromatin loops, which modulate high-order chromatin organization and promoter–enhancer interactions[30,31]. We found that the regions connected by chromatin loops were enriched with genes (Fisher's exact test $P < 0.05$).

Genome design can alter chromosome conformation[32] and we found clear changes in chromatin loops within the redesigned region as well as in neighbouring chromosomal sequences (Fig. 6c). Two chromatin loops within the redesigned region were lost in the synthetic version. In addition, interactions between the redesigned region and the distal sequences were lost in semi-syn18L. On the other hand, new connections were established between the synthetic region and the proximal neighbouring region, including one connection close to Pp3c18_310, which encodes a putative kinase (Supplementary Table 2). Notably, we found that Pp3c18_310 expression in semi-syn18L was more than three times higher than that in the WT (Supplementary Table 8). Nevertheless, genes showing similar new connections but located further away from the synthetic region, did not show expression changes. Together, the results showed that redesigning

https://doi.org/10.1038/s41477-023-01595-7

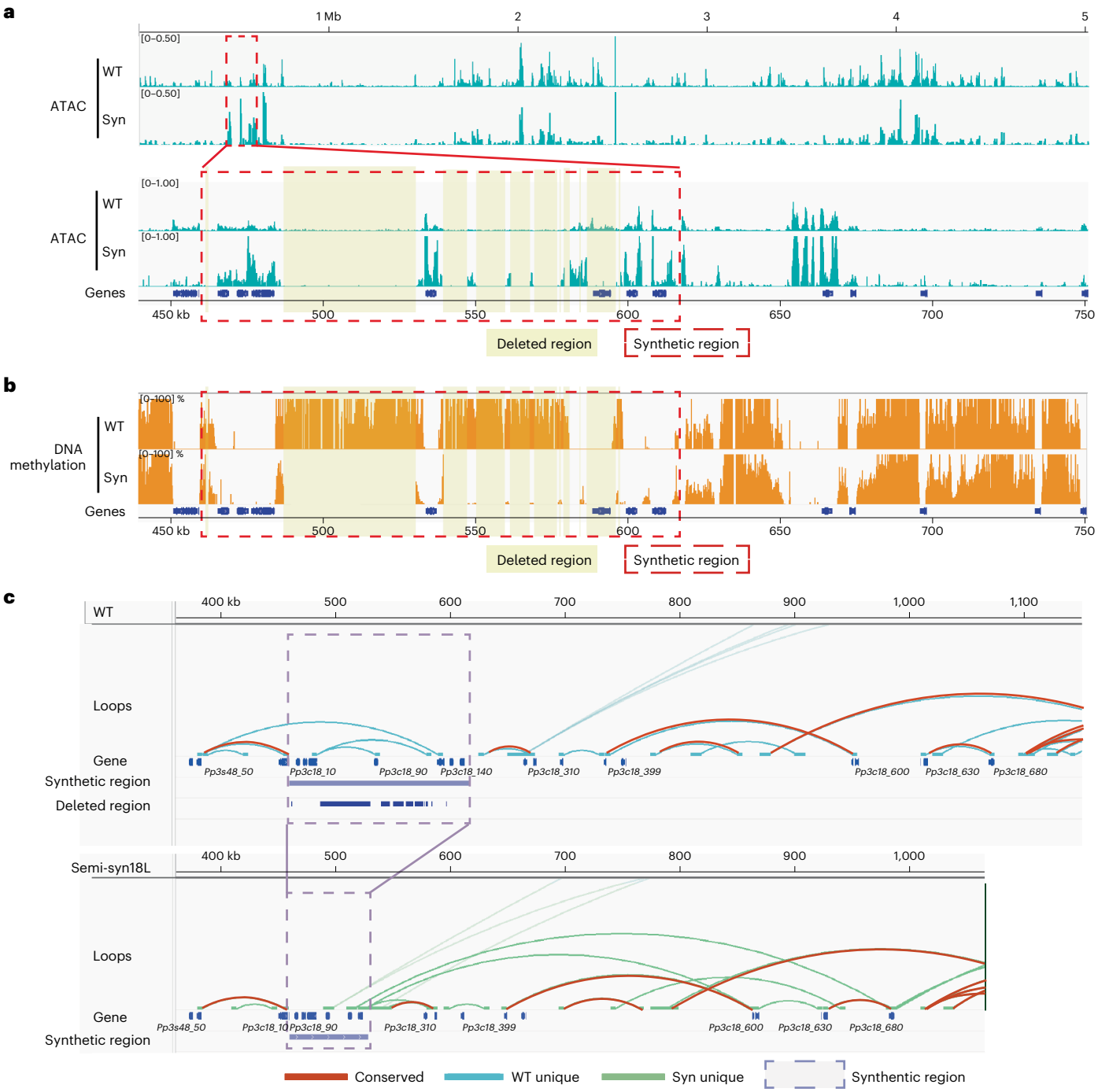


Fig. 6 | The epigenetic characteristics of *P. patens* and the epigenetic changes in semi-syn18L. a, Comparison of chromatin accessibility levels between WT and semi-syn18L around the synthetic region. The light-yellow region represents the deleted region in semi-syn18L. **b**, Comparison of DNA methylation levels between WT and semi-syn18L around the synthetic region. The light-yellow region represents the deleted region in semi-syn18L. **c**, The distribution of chromatin loops of WT and semi-syn18L spanning the synthetic region. The conserved chromatin loops are marked in dark red. The loops unique to WT are shown in blue. New chromatin loops that are unique to semi-syn18L are shown in green.

the genome can alter chromosome conformation and may alter the expression of retained genes.

Discussion

The complete synthesis and replacement of a 155 kb endogenous chromosomal region resulting in an ~55.8% size reduction (phase I) represents a substantial stride toward the design and synthesis of the *P. patens* genome, as well as those of other multicellular plants and animals. By comparing different assembly strategies, we demonstrate the feasibility of large-scale genome replacement in *P. patens*. Although the simultaneous assembly of multiple short fragments is favoured in budding yeast, our findings indicate that single mega-chunk replacement is more feasible in *P. patens*, as we did not obtain a complete assembly using three ~30 kb mid-chunks in planta. In this study, we took advantage of the efficient homologous recombination of *P. patens*[21]. For plant and animal species lacking efficient homologous recombination,

genome-editing tools may facilitate chromosome engineering[33]. In future studies, we expect applying CRISPR/Cas9 and its variants to slice homologous arm would improve assembly efficiency.

The genomes of multicellular organisms frequently show a profusion of repetitive sequences. However, the essentiality of TEs in multicellular organisms remains untested. Natural selection may be incapable of eliminating TEs because of their potentially insufficient deleterious effects[10–13]. Conversely, evidence suggests that TEs may play a role in shaping the epigenetic landscape and function as enhancers to regulate the expression of adjacent genes[14,15]. Our findings demonstrate that a relatively aggressive removal of repetitive sequences is viable in genome design, laying the foundation for the SynMoss. This study also implies that multicellular organisms are robust to radical alterations in genome architecture. Moreover, the modifications incorporated in the genome design, including the removal of repetitive sequences and TAG/TAA stop codons, appear to exert a minimal impact on plant morphology, development and stress resistance. While epigenetic regulation is pervasive in multicellular organisms, our results indicate that the epigenetic landscape can establish normally in a large-scale synthetic region, resulting in largely normal gene expression, thereby ensuring the feasibility of genome synthesis in multicellular organisms.

The obtained lines harbouring synthetic chromosomal fragments show morphological similarity to the WT, include all cell types examined in our analysis and show successful transmission across generations. However, we observed overall increase in gene expression within the synthetic region and some neighbouring genes. Although the epigenetic landscape was re-established, substantial deletion of repetitive sequences led to reductions in H3K9me2 modifications and DNA methylation in the retained sequences. We also observed alterations in three-dimensional chromosome organization, which is closely associated with gene expression, within the simplified sequences. Techniques such as CRISPR-based precise epigenome editing[34] and the tethering approach[35] can be used to fine-tune the epigenetic landscape as necessary.

Having concluded the experiments for phase I of the SynMoss, the achieved outcomes instill us with confidence in our ability to undertake the subsequent phase (phase II) of the project. Upon the completion of phase II, the entire short arm of chromosome 18 (chr. 18L) will be replaced with the simplified and synthesized sequences. The findings from these experiments will further pave the way for the artificial design and synthesis of the entire moss genome. Genome synthesis has emerged as an efficacious approach for comprehending genome organization and function and serves as a foundation for new technologies[5,36,37]. Beyond viruses and microbes, synthetic genomics offers immense potential for application in plants and animals, bolstered by the continuous decline in DNA synthesis costs and advancements in large fragment assembly. This study will serve as a cornerstone for genome synthesis in other multicellular species, including seed plants and animals.

Methods

Design of a simplified synthetic chromosome fragment

To design our simplified genome fragment, we used a list of genes annotated from the *P. patens* genome based on *Physcomitrium patens* v.3.3 (https://phytozome-next.jgi.doe.gov/info/Ppatens_v3_3), with additional inhouse annotation. Supplementary Data 1–4 provide the genome sequence and corresponding annotations for the phase I and II designs. To simplify the chromosomal fragment of chromosome 18, we used the principles described in Supplementary Box 1, which involved several specific steps described below. The inhouse computational pipeline for standardizing genome design is available on Zenodo (https://doi.org/10.5281/zenodo.7894207).

First, we removed intergenic regions to condense the moss genome using slightly different criteria between phases I and II. In phase I, we marked coding sequences, untranslated regions, putative promoters (3 kb sequence upstream) and putative terminators (2 kb sequence downstream) for each gene. In the unmarked genomic regions, we deleted repetitive sequences predicted by RepeatMasker (Open-4.0, http://www.repeatmasker.org) as well as any remaining sequence fragments shorter than 1 kb. In phase II, we removed all sequences beyond the 3 kb upstream or 2 kb downstream sequences of annotated genes to further simplify the synthetic genome fragment.

Second, we replaced all TGA and TAG stop codons with the TAA codon. In phase I, we replaced a total of five TAG codons and one TGA codon, while in phase II, we replaced 13 TAG codons and 12 TGA codons. In phase II, we also inserted LoxPsym sites 3 bp downstream of the stop codon for all 37 genes, following ref. 3. This LoxP sequence is palindromic and capable of recombining in either direction[38].

Third, in phase I, we designed a set of three primers for amplifying DNA fragments that covered each 'retaining–removing' junction site to enable the detection of successful replacements of synthetic fragments (serving as a PCRTag). Two primers were designed upstream (UP) and downstream (DP) of each DNA region that we aimed to remove. These primers could amplify the synthetic fragment but not the original WT fragment because of an insufficient elongation time in PCR. Another primer (RP) designed in the removed region, located near the retaining–removing junction, together with UP (or DP), was able to amplify the WT fragment but not the synthetic fragment. We designed a total of 11 sets of primers for phase I, which are listed in Supplementary Table 2.

In phase II, we introduced altered nucleotides as PCRTags in the designed sequences in introns or coding sequences. *P. patens* contains many introns, which provided us with an opportunity to design PCRTags without affecting synonymous codon usage[39]. Specific design information can be obtained in Supplementary Box 1.

Fourth, we removed Pp3c18_140 from the synthetic sequences and adjusted the position of the Pp3c18_90 gene to ensure a mid-chunk length of 20–30 kb in phase I. The simplified genomes and their corresponding annotations are provided in Supplementary Data 5–8.

Plant materials and growth conditions

In this study, the background of all strains was the Gransden WT strain of *P. patens*. All strains were grown on BCDAT medium at 25 °C under long-day conditions (16 h of light and 8 h of dark). The BCDAT medium was composed of 1.84 mM KH_2PO_4 (pH 6.5), 10 mM KNO_3, 1 mM $MgSO_4$, 45 μM $FeSO_4$, trace element solution (10 μM H_3BO_3, 0.22 μM $CuSO_4$, 2 μM $MnCl_2$, 0.1 μM Na_2MoO_4, 0.19 μM $ZnSO_4$, 0.23 μM $CoCl_2$ and 0.17 μM KI), 1 mM $CaCl_2$, 5 mM ammonium tartrate and 0.7% agar.

The protonemal tissues for protoplast transformation were cultured on cellophane-overlaid BCDAT medium for 1 week. We used 7-day-old protonemal and 28-day-old gametophores in phenotypic analyses. In the sporophyte induction stage, protonemal tissues were cultured on seedling blocks (composed of coconut fibre crumbs) at 25 °C under long-day conditions for 1 month and then transferred to 16 °C under short-day conditions (8 h of light and 16 h of dark) for an extra 6 weeks. The cultures were irrigated from week 3 to week 6 to facilitate fertilization.

P. patens genetic transformation

The transformation of *P. patens* was conducted by PEG-mediated protoplast transformation.

Preparation of transformed DNA

The target plasmid was extracted from *E. coli* with a high-purity plasmid extraction kit (Tiangen, catalogue no. DP116), then the plasmids were linearized by overnight digestion with the restriction enzyme FastDigest NotI (Thermo Fisher, catalogue no. FD0596). The linearized DNA was purified and concentrated with ammonium acetate. The extraction was performed using an equal volume of phenol chloroform isoamyl alcohol (25:24:1). Then, DNA was precipitated with 0.4 times the volume of ammonium acetate and 2.5 times the volume of anhydrous ethanol at 4 °C for 30 min. Finally, the DNA was washed twice with 70% ethanol and dissolved in water. The obtained linearized DNA was used for protoplast transformation.

Preparation of protoplasts
The protonemal tissues were placed in the enzymatic solution (8% mannitol, 0.5% cellulase (Yakult, catalogue no. L0012) and 0.15% pectinase (Yakult, catalogue no. L0021)) and gently shaken for 1 h. Protoplast filtrate was collected by a 70 μm cell strainer (BD Falcon, catalogue no. 352350) filtration and then the protoplasts were collected by centrifugation at 4 °C 200*g*. Protoplasts were washed twice with 8% mannitol before transformation.

Protoplast transformation
Approximately 1.5×10^6 protoplasts were resuspended in 600 μl of MMM solution (15 mM $MgCl_2$, 9.1% mannitol and 1% MES (pH 5.6)). A total of 80 μg of purified linearized DNA (up to 60 μl) was mixed with 600 μl of protoplast MMM solution suspension and 700 μl of PEG solution (40% PEG 6000, 0.1 M $Ca(NO_3)_2$, 10 mM Tris-HCl (pH 8.0) and 8% mannitol) was added to the mixture and incorporated by gently tilting the tubes. The mixture was incubated at room temperature for 30 min (mixed once or twice halfway through) and then diluted with 3 ml of W5 solution (5 mM KCl, 125 mM $CaCl_2$, 154 mM NaCl and 2 mM MES (pH 5.6)) on ice. Protoplasts collected by centrifugation were suspended in 12 ml of PRM/T medium (BCDAT medium with 6% mannitol, 10 mM $CaCl_2$ and 0.4% agar) and 2 ml of protoplasm suspension was coated onto a cellophane-overlaid plate containing PRM/B medium (BCDAT medium with 6% mannitol, 10 mM $CaCl_2$ and 0.8% agar).

Culturing and antibiotic screening of regenerating plants
Protoplasts were cultured at 25 °C under long-day conditions, first on PRM/B medium for 5 to 7 d and were then transferred onto BCDAT medium supplemented with antibiotics for selection (hygromycin (30 μg ml^{-1}) or G418 (25 μg ml^{-1})). The regenerated plants were grown on a screened medium containing antibiotics for 14 d and then transferred to a non-resistant BCDAT medium for recovery culture for 14 d, after which antibiotic levels were increased for a second round of screening (hygromycin (50 μg ml^{-1}) and G418 (50 μg ml^{-1})).

Mid-chunks and mega-chunks in yeast
The mid-chunks and mega-chunks were obtained as described[26] with minor modifications.

PCRTag analysis of synthetic strains
The PCRTag design principles are provided in section 'Design of a simplified synthetic chromosome fragment'. If only WT genome sequences could be amplified and no synthetic sequences could be amplified, no homologous recombination substitution occurred at this site. If both the WT genome sequence and synthetic sequence could be amplified, there was no homologous recombination at the site but the synthetic sequence might show homologous recombination at other sites or the linear synthetic fragment might show self-cyclization and be unstable in vivo or it might be a chimaera. If the WT genome sequence could not be amplified, only the synthetic sequence could be amplified and homologous recombination of the synthetic sequence and the WT genome sequence could then occur at the site. A strain was considered a true positive synthetic strain if all loci could only be amplified with synthetic sequences and no WT genome sequences could be amplified (Supplementary Tables 2 and 3).

P. patens genomic DNA preparation for PCRTag analysis
A small amount of protonemal tissue was picked with a needle and placed into 10× PCR buffer (500 mM KCl, 15 mM $MgCl_2$, 100 mM Tris-HCl (pH 8.3)) and the tissues were disrupted in a grinder ((20 Hz, 3 min) (Retsch, catalogue no. MM400)), incubated at 68 °C for 1 h and centrifuged at 4,500*g* for 5 min. The supernatant contained DNA that could be used as a template for PCRTag analysis.

Flow cytometry
We first collected two gametophore colonies, added 600 μl of nucleus extraction buffer (15 mM Tris-HCl (pH 8.0), 2 mM Na_2EDTA, 0.5 mM spermine, 80 mM KCl, 15 mM β-mercaptoethanol and 0.1% Triton X-100) and chopped the tissue with a blade. After allowing the samples to stand for 5–20 min, we filtered and collected the liquid with a 30 μm sieve (Miltenyi, catalogue no. 130-098-458). The whole process was performed on ice. Chromosome ploidy was analysed by flow cytometry after DAPI (10 μg ml^{-1}) was added to the collection fluid.

Microscopy
Microscopy images of protonema and gametophore colonies and the single tip of the gametophore were taken by using a stereomicroscope (Nikon SMZ25) with materials placed on BCDAT medium. Microscopy images of the chloronema, caulonema, bud, leafy and reproductive organs were taken by using an optical microscope (Nikon Ni-U) with materials placed on a slide.

Stress treatments
In these experiments, all the plant materials were gametophores grown on BCDAT medium for 28 d.

Salt stress treatment
Gametophore colonies were cultured in BCDAT medium with salt concentrations of 0, 200, 300, 400 and 500 mM for 3 d on BCDAT. Then, the materials were collected.

Osmotic stress treatment
Gametophore colonies were cultured in BCDAT medium with sorbitol concentrations of 0, 400, 500, 600 and 750 mM for 3 d on BCDAT. Then, the materials were collected.

Determination of chlorophyll content
The collected plant materials were ground into a fine powder in liquid nitrogen using a mortar. Approximately 400 mg of plant material was placed in a 10 ml tube and 5 ml 80% (v/v) acetone was added to extract chlorophyll. The sample was mixed vigorously for 5 min so that chlorophyll was extracted from the plant. The samples were centrifuged at 14,000*g* for 5 min to remove cell debris. The chlorophyll content of the supernatant was measured with a spectrophotometer at 645 and 663 nm. After measurement, the supernatant was completely redelivered to the initial tube and the samples were dried in a rapid vacuum centrifuge at room temperature until a stable dry weight was obtained for each sample. The total chlorophyll calculation was as follows: chlorophyl per dry weight (mg) = ((A663) (0.00802) + (A645) (0.0202)) × 1.5 per dry weight (mg).

Whole-genome sequencing
Whole-genome DNA was extracted from 50 mg of protonema materials grown on BCDAT medium for 7 d using a DNA extraction kit (Tiangen, catalogue no. DP360). Three independent biological replicates were performed for each genotype. The whole-genome DNA was broken into fragments of ~500 bp with an ultrasonicator (Diagenode, Bioruptor Plus). DNA-seq libraries were prepared according to the VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme, catalogue no. ND607) and sequenced using an Illumina NovaSeq system in 150 nucleotide (nt) paired-end mode.

The raw reads were filtered using Trimmomatic v.0.39 and aligned to the *P. patens* genome (available at https://figshare.com/articles/dataset/ChIP_track_rar/23648046) with BWA v.0.7.17. PCR duplications were removed with Picard v.2.27.5. The depth of the synthetic area was normalized using the BPM method of the bamCoverage tool of deepTools v.3.5.7. Visualization was performed using IGV v.2.14.0.

Transcript expression analyses by RNA sequencing

Total RNA was extracted using the AxyPrep Multisource Total RNA Miniprep Kit (Axygen, catalogue no. AP-MN-MS-RNA). Three independent biological replicates were performed for each genotype. RNA-seq libraries were prepared using the NEBNext ultra II DNA library prep kit (NEB, catalogue no. E7645) and sequenced using an Illumina NovaSeq system in 150 nt paired-end mode.

The reads were aligned to the reference genome of *P. patens* v.3.3 from Phytozome (https://phytozome-next.jgi.doe.gov) using STAR v.2.7.10b. Transcript expression was counted by featureCounts v.2.0.3. Differentially expressed genes were determined using DESeq2 (R package), with a cutoff value of >1 $\log_2$ fold change and Benjamini–Hochberg false discovery rate of <0.05. Transcripts per million (TPM) values were calculated using the standard formula with R.

ChIP–seq analysis

For ChIP, ~1 g of protonema material grown on BCDAT medium for 7 d was collected with cross-linking and then stored at −80 °C before use. Plant materials were ground into a fine powder in liquid nitrogen. DNA was extracted and fragmented into ~500 bp fragments with an ultrasonicator (Diagenode, Bioruptor Plus). A total of 5 µg of the corresponding antibody (Abcam, H3K4me3 (catalogue no. ab8580), H3K9me2 (catalogue no. ab1220), H3K27me3 (catalogue no. ab6002), H3K27ac (catalogue no. ab4729), H3K36me3 (catalogue no. ab9050)) was added to each sample (diluted to 1/200, the final concentration is 5 µg ml^{-1}), which was then incubated at 4 °C for at least 4 h and up to overnight with rotation. Dynabeads Protein A/Protein G were washed and used to capture DNA associated with H3K27me3, H3K9me2 H3K4me3, H3K27me3, H3K36me3 or H3K27ac. The chromatin was eluted and decross-linked overnight at 65 °C. DNA was extracted from immunoprecipitated chromatin by the phenol–chloroform–alcohol method and precipitated with ethanol. For ChIP–seq, three independent biological replicates were performed for sequencing library preparation. DNA-seq libraries were prepared according to the VAHTS Universal DNA Library Prep Kit for Illumina V3 kit (Vazyme, catalogue no. ND607) and sequenced using an Illumina NovaSeq system in 150 nt paired-end mode.

The raw reads were trimmed using Trimmomatic v.0.39 and aligned to the *P. patens* genome using BWA v.0.7.17. In addition, reads of semi-syn18L were mapped to the reference genome modified according to our design. PCR duplicates were removed through Picard v.2.27.5. Reads with a MAPQ value <30 were removed with SAMtools v.1.16.1. Narrow peaks were called using MACS2 v.2.25 with an extsize value of 165. The depth of ChIP–seq reads was normalized using the BPM method of the bamCoverage tool of deepTools v.3.5.7. DiffBind v.3.8.4 was used to explore the peaks of ChIP–seq that differed between semi-syn18L and the WT. The BigWig files of ChIP–seq are available at Figshare (https://figshare.com/articles/dataset/ChIP_track_rar/23648046).

DNA methylation analysis

Whole-genome DNA was extracted from 50 mg of protonema material grown on BCDAT medium for 7 d using a DNA extraction kit (Tiangen, catalogue no. DP360). The whole-genome DNA was broken into fragments of ~500 bp with an ultrasonicator (Diagenode, Bioruptor Plus). DNA-seq libraries were prepared using the NEBNext ultra II DNA library prep kit (NEB, catalogue no. E7645) and methylation was treated by the EpiTect Fast DNA Bisulfite Kit (Qiagen, catalogue no. 59824). For methylation analysis, three independent biological replicates were performed for library preparation. The libraries were sequenced using an Illumina NovaSeq system in 150 nt paired-end mode. The reads were mapped to the *P. patens* genome and all methylation information was extracted using Bismark v.2.5.0.

ATAC-seq analysis

A total of 5,000 protoplast cells ('Preparation of protoplasts' section) were collected in a 200 µl centrifuge tube and centrifuged at 14 °C 200*g* for 5 min. The supernatant was discarded and the cells were resuspended in 6 µl of lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40). The cells were then incubated on ice for 10 min. The TruePrep DNA Library Prep Kit v.2 for Illumina (Vazyme, catalogue no. TD502) was used for transposable enzyme digestion and DNA library construction. Three independent biological replicates were performed for library preparation. The libraries were sequenced using an Illumina NovaSeq system in 150 nt paired-end mode.

The raw reads were filtered by Trimmomatic v.0.39 and aligned to the *P. patens* genome using BWA v.0.7.17. The reads of semi-syn18L were also mapped to the reference genome modified according to our design. PCR duplicates and reads with a MAPQ <30 were removed for ChIP–seq analyses. Narrow peaks were called using MACS2 v.2.25 with a shift value of −100 and an extsize of 200. The ATAC-seq peaks that differed between semi-syn18L and the WT were explored using DiffBind v.3.8.4.

In situ Hi-C analysis

In situ Hi-C analysis was performed as previously described[40]. Briefly, ~1 g of protonema materials were collected in MC buffer (10 mM K_3PO_4, 50 mM NaCl, 100 mM sucrose) with cross-linking (fixed by 1% formaldehyde for 15 min under vacuum and then stopped by 150 mM glycine for 10 min under vacuum). The samples were stored at −80 °C until use. The plant materials were ground into a fine powder in liquid nitrogen and resuspended in nuclear isolation buffer (20 mM HEPES, 250 mM sucrose, 1 mM $MgCl_2$, 5 mM KCl, 40% glycerol, 0.25% Triton X-100, 0.1 mM PMSF, 0.1% 2-mercaptoethanol). The solution was filtered with four layers of Miracloth (Merck Millipore, 475855) and the nuclei were washed repeatedly until no green colour remained. The precipitate was gently resuspended in 150 µl of 0.5% SDS and incubated at 62 °C for 5 min. Then, 50 U of DpnII was added and the mixture was incubated overnight at 37 °C. On the second day, the digested DNA was passivated using the Klenow Fragment (Vazyme, catalogue no. N104-01), with the addition of biotin-14-dCTP (Invitrogen, catalogue no. 19518018). After ligation with T4 DNA ligase, the DNA was purified using phenol–chloroform–isoamyl alcohol extraction. Finally, the DNA was broken into fragments of ~500 bp using an ultrasonic wave (Diagenode, Bioruptor Plus). The sheared DNA was selected with AMPure XP beads (Beckman, A63880) and then subjected to biotin enrichment using Dynabeads MyOne Streptavidin C1 beads (Invitrogen, 65001). Following biotin enrichment, bead end repair and adaptor ligation were carried out. After washing, the beads were resuspended in 15 µl of 10 mM Tris-HCl buffer (pH 8.0) and DNA was isolated from Dynabeads MyOne Streptavidin C1 beads by incubation at 98 °C for 10 min. Library amplification was performed with 12 cycles of PCR and the resulting PCR products were purified using a conventional PCR product purification kit. The library was sequenced on an Illumina NovaSeq system in 150 nt paired-end mode.

The reads generated in the Hi-C experiment were mapped to the reference genome (the reads of WT were mapped to the *P. patens* genome; the reads of semi-syn18L were mapped to the reference genome modified according to our design.) using juicer v.1.6 and the effective interaction pairs were calculated with the default parameters. TAD domains were detected using the 'Arrowhead' module of juicer-tools at a resolution of 5 kb; chromatin loops were detected using the 'CPU HiCCUPS' module of juicer-tools at resolutions of 5 and 10 kb. The relative distances between regions linked by chromatin loops and genes were calculated with bedtools v.2.30.0.

Reporting summary

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

Data availability

All high-throughput sequencing data (ATAC-seq, Hi-C, ChIP–seq, RNA-seq, whole-genome sequencing and whole-genome bisulfite sequencing) in this paper are contained in the Sequence Read Archive (SRA) (PRJNA970280). The raw gel image of PCR, the raw data of flow cytometry and the data of stress treatments are provided in the Supplementary Information. The BigWig files of ChIP–seq are available at figshare (https://figshare.com/articles/dataset/ChIP_track_rar/23648046). *P. patens* genome v.3.3 is available at Phytozome (https://phytozome-next.jgi.doe.gov). Source data are provided with this paper.

Code availability

All original codes used in high-throughput sequencing analysis have been deposited at Github (https://github.com/lanntianlong/SynMoss) and Zenodo (https://doi.org/10.5281/zenodo.8000393). The inhouse computational pipeline for standardizing genome design is available on Zenodo (https://doi.org/10.5281/zenodo.7894207).

References

1. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
2. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
3. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
4. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
5. Zurcher, J. F. et al. Refactored genetic codes enable bidirectional genetic isolation. *Science* **378**, 516–523 (2022).
6. Isaacs, F. J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
7. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
8. Blommaert, J. Genome size evolution: towards new model systems for old questions. *Proc. Biol. Sci.* **287**, 20201441 (2020).
9. Wang, S. Y. et al. Role of epigenetics in unicellular to multicellular transition in *Dictyostelium*. *Genome Biol.* **22**, 134 (2021).
10. Arkhipova, I. R. Neutral theory, transposable elements and eukaryotic genome evolution. *Mol. Biol. Evol.* **35**, 1332–1337 (2018).
11. Bourgeois, Y. & Boissinot, S. On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes* **10**, 419 (2019).
12. Elliott, T. A., Linquist, S. & Gregory, T. R. Conceptual and empirical challenges of ascribing functions to transposable elements. *Am. Nat.* **184**, 14–24 (2014).
13. Doolittle, W. F., Brunet, T. D., Linquist, S. & Gregory, T. R. Distinguishing between 'function' and 'effect' in genome biology. *Genome Biol. Evol.* **6**, 1234–1237 (2014).
14. Gemmell, N. J. Repetitive DNA: genomic dark matter matters. *Nat. Rev. Genet.* **22**, 342 (2021).
15. Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev. Genom. Hum. Genet.* **8**, 241–259 (2007).
16. 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).
17. Lang, D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
18. Rensing, S. A. et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64–69 (2008).
19. Bi, G. et al. Telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nat. Plants* (in the press).
20. 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).
21. Schaefer, D. G. & Zrÿd, J. P. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **11**, 1195–1206 (1997).
22. Reski, R., Bae, H. & Simonsen, H. T. *Physcomitrella patens*, a versatile synthetic biology chassis. *Plant Cell Rep.* **37**, 1409–1417 (2018).
23. Jiao, Y. & Wang, Y. Towards plant synthetic genomics. *BioDes. Res* **5**, 0020 (2023).
24. Perroud, P.-F. et al. The *Physcomitrella patens* gene atlas project: large-scale RNA-seq based expression data. *Plant J.* **95**, 168–182 (2018).
25. King, B. C. et al. *In vivo* assembly of DNA-fragments in the moss, *Physcomitrella patens*. *Sci. Rep.* **6**, 25030 (2016).
26. Jiang, S. et al. Efficient *de novo* assembly and modification of large DNA fragments. *Sci. China Life Sci.* **65**, 1445–1455 (2022).
27. Yi, P. & Goshima, G. Transient cotransformation of CRISPR/Cas9 and oligonucleotide templates enables efficient editing of target loci in *Physcomitrella patens*. *Plant Biotechnol. J.* **18**, 599–601 (2020).
28. Zagórska-Marek, B., Sokolowska, K. & Turzańska, M. Chiral events in developing gametophores of *Physcomitrella patens* and other moss species are driven by an unknown, universal direction-sensing mechanism. *Am. J. Bot.* **105**, 1986–1994 (2018).
29. Domb, K. et al. DNA methylation mutants in *Physcomitrella patens* elucidate individual roles of CG and non-CG methylation in genome regulation. *Proc. Natl Acad. Sci. USA* **117**, 33700–33710 (2020).
30. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
31. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
32. Mercy, G. et al. 3D organization of synthetic and scrambled chromosomes. *Science* **355**, aaf4597 (2017).
33. Ronspies, M., Dorn, A., Schindele, P. & Puchta, H. CRISPR–Cas-mediated chromosome engineering for crop improvement and synthetic biology. *Nat. Plants* **7**, 566–573 (2021).
34. Nakamura, M., Gao, Y. C., Dominguez, A. A. & Qi, L. S. CRISPR technologies for precise epigenome editing. *Nat. Cell Biol.* **23**, 11–22 (2021).
35. Dawe, K. R. et al. Synthetic maize centromeres transmit chromosomes across generations. *Nat. Plants* **9**, 433–441 (2023).
36. Brooks, A. N. et al. Transcriptional neighborhoods regulate transcript isoform lengths and expression levels. *Science* **375**, 1000–1005 (2022).
37. Yim, S. S. et al. Robust direct digital-to-biological data storage in living cells. *Nat. Chem. Biol.* **17**, 246–253 (2021).
38. Steensels, J., Gorkovskiy, A. & Verstrepen, K. J. SCRaMbLEing to understand and exploit structural variation in genomes. *Nat. Commun.* **9**, 1937 (2018).
39. Kong, J., Zhang, S., Qian, W. & Li, K. Synonymous somatic mutations that alter proximal out-of-frame downstream ATGs are associated with aberrant gene expression levels in cancer cells. *J. Genet. Genomics* **50**, 447–449 (2023).
40. Liu, C. *In situ* Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. *Methods Mol. Biol.* **1629**, 155–166 (2017).

Acknowledgements

This work was funded by the National Key R&D Program of China grant no. 2019YFA0903900 (Q.Z., W.Y., W.Q. and Y.J.), the National

Natural Science Foundation of China grant nos. 31825002 (Y.J.), 31800069 (J.D.), 31800082 (J.D.) and 32270345 (Y.W.), Shenzhen Science and Technology Program grant no. KQTD20180413181837372 (J.D.), Guangdong Provincial Key Laboratory of Synthetic Genomics grant no. 2019B030301006 (J.D.), Guangdong Basic and Applied Basic Research Foundation grant no. 2023A1515030285 (J.D.), Bureau of International Cooperation, Chinese Academy of Sciences grant no. 172644KYSB20180022 (J.D.), CAS Strategic Priority Research Program grant nos. XDA24020203 (Y.J.), XDA24020103 (W.Q.) and XDA28030402-5 (W.Q.), Shenzhen Outstanding Talents Training Fund (J.D.) and the China Postdoctoral Science Foundation no. 2020M680740 (L.-G.C.). We thank the National Center for Protein Sciences at Peking University for assistance with flow cytometry analysis.

Author contributions

Y.J., J.D., W.Q., Y.W., L.-G.C. and Q.Z. designed the experiments and analysed the data. S.Z., Y.Z., Q.D., B.J. and W.Q. performed genome design. M.Z., G.L., Z.H and Y.M. performed genome assembly in yeast. L.-G.C. and T.L. performed plant genome assembly, genome sequencing, transcriptome, epigenome and Hi-C analyses. L.-G.C., T.L., H.L. and B.L. performed genotyping and sequence analysis. Y.G. performed phenotype analysis. Y.J., J.D., W.Q., Y.W., L.-G.C., T.L. and S.Z. wrote the paper.

Competing interests

L.-G.C., Y.W. and Y.J. are inventors on patent applications covering the results described in this study. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-023-01595-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-023-01595-7.

Correspondence and requests for materials should be addressed to Ying Wang, Wenfeng Qian, Junbiao Dai or Yuling Jiao.

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

Reprints and permissions information is available at www.nature.com/reprints.

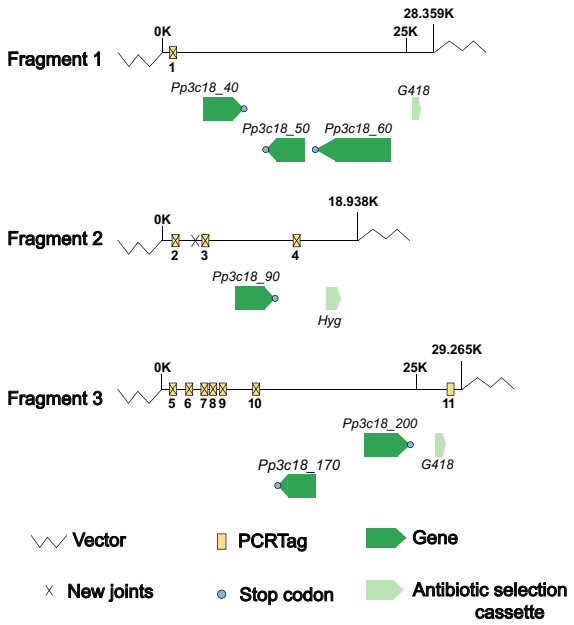
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

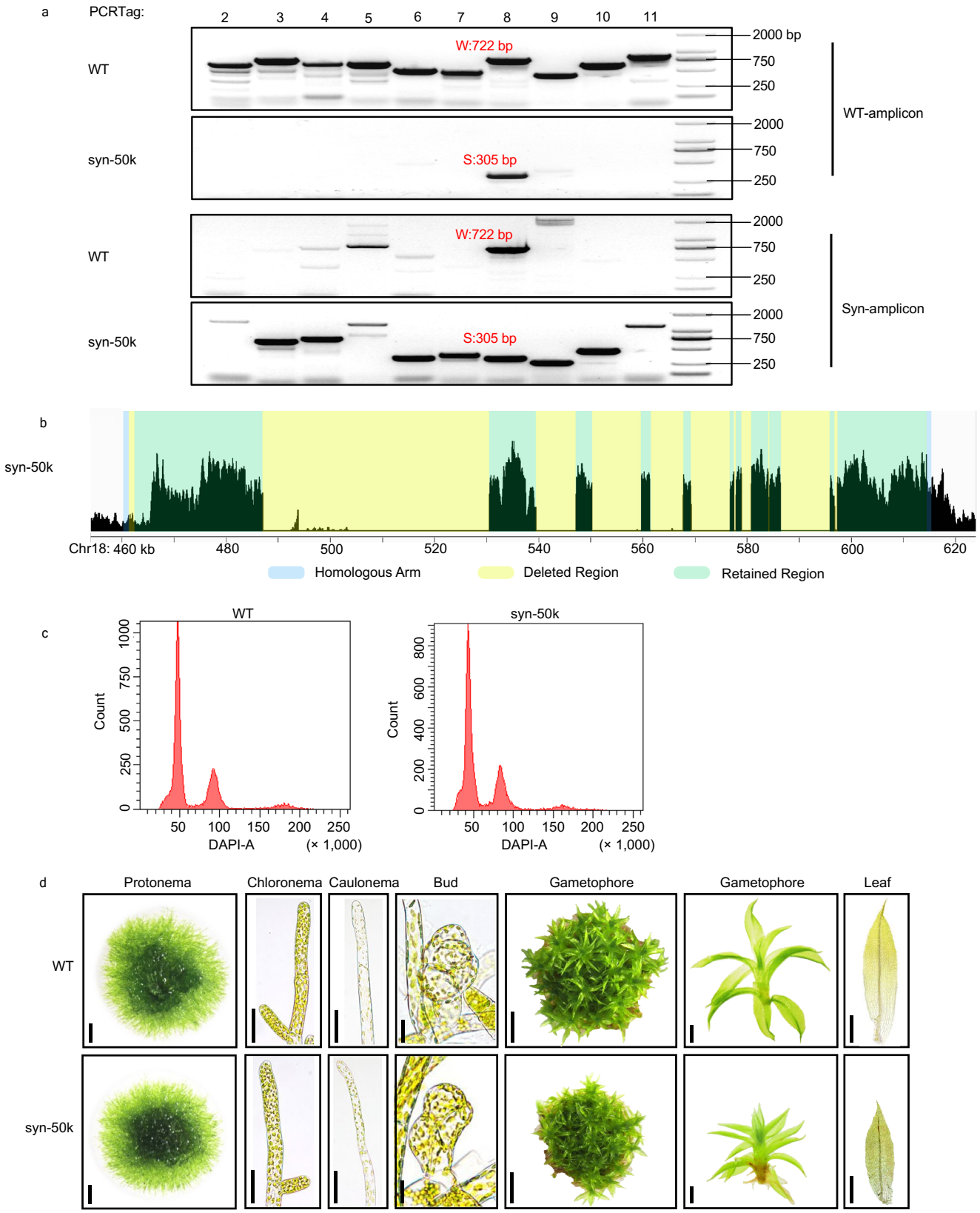
[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China. [2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [4]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [5]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. [6]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China. [7]CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [8]Beijing Agro-Biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. [9]School of Earth and Space Sciences, Peking University, Beijing, China. [10]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [11]Peking-Tsinghua Center for Life Sciences, Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [12]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, China. [13]These authors contributed equally: Lian-Ge Chen, Tianlong Lan, Shuo Zhang, Mengkai Zhao.
✉e-mail: yingwang@ucas.ac.cn; wfqian@genetics.ac.cn; daijunbiao@caas.cn; yuling.jiao@pku.edu.cn

https://doi.org/10.1038/s41477-023-01595-7



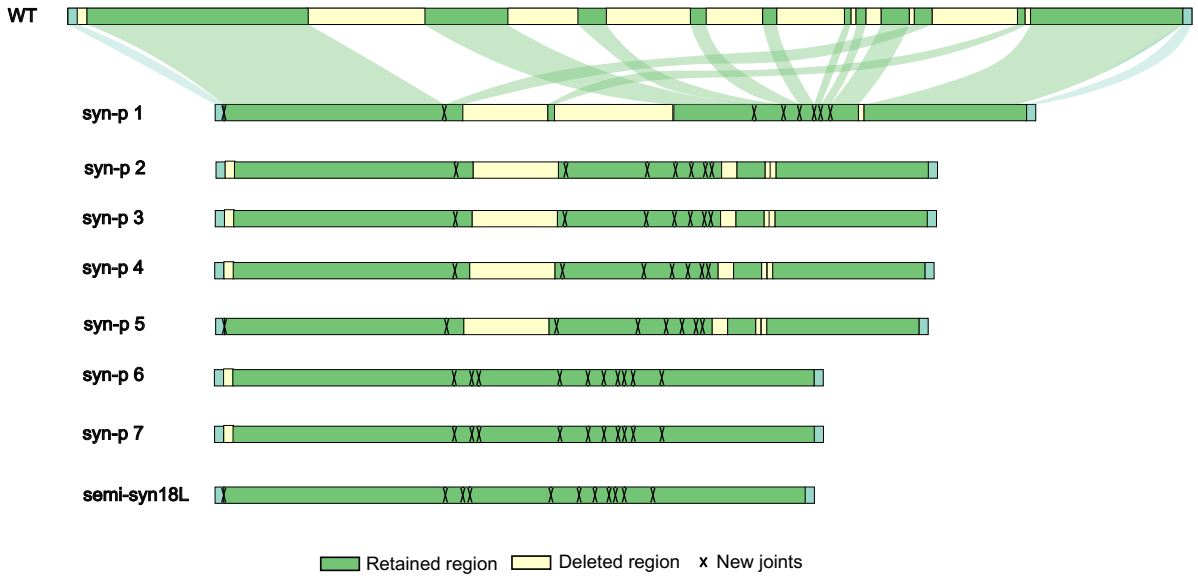
Extended Data Fig. 1 | The design and assembly of three 20–30 kb mid-chunks. We split the 68,530 bp mega-chunk into three mid-chunk fragments and added a resistant cassette (kanamycin or hygromycin) to the centromeric end of each mid-chunk. If it is not possible to achieve the goal of replacing three 30 kb mid-chunks in the assembly, the replacement of two or one mid-chunks should be implemented. Each mid-chunk overlapped with the next one by a 1 kb overhang and 1 kb homologous arms were included at either end of the entire region.

https://doi.org/10.1038/s41477-023-01595-7



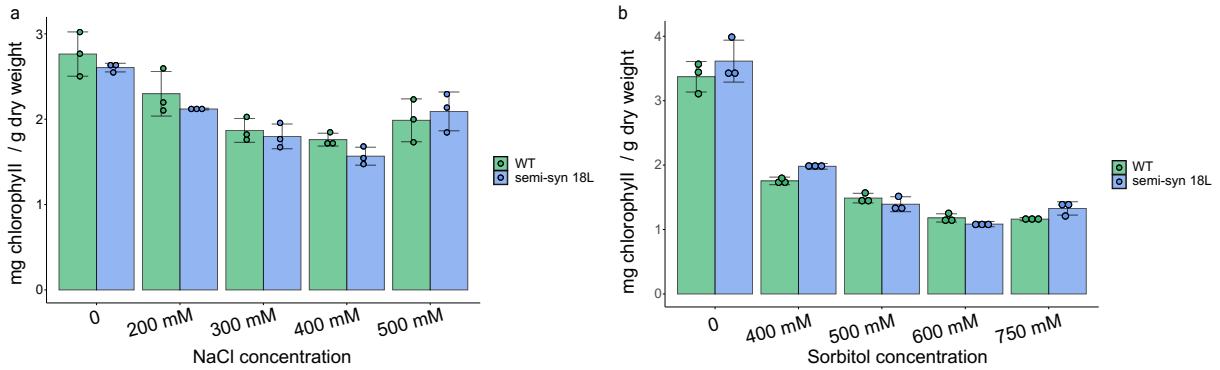
Extended Data Fig. 2 | Characterization of syn-50k. (**a**) PCRTag analysis. A total of ten recombination replacement sites were detected, including eight sites in deleted regions and two in homologous arms. Only synthetic sequences and not wild-type sequences can be amplified from semi-syn18L. Three independent experiments were conducted on each line using three independent samples and similar results were obtained. (**b**) Whole-genome Illumina resequencing analysis of syn-50k. No reads containing sequences from the deleted regions were found and reads were identified covering all new junctions. (**c**) Flow cytometry analysis indicated that semi-syn18L-2 was as haploid as the wild type. (**d**) The phenotypes of the wild type and syn-50k at various developmental stages. The figure shows the phenotypes of representative samples. (More than 3 independent samples were observed for each stage and each line). Bars from left to right, 500 μm, 50 μm, 50 μm, 20 μm, 2 mm, 1 mm and 500 μm.

https://doi.org/10.1038/s41477-023-01595-7



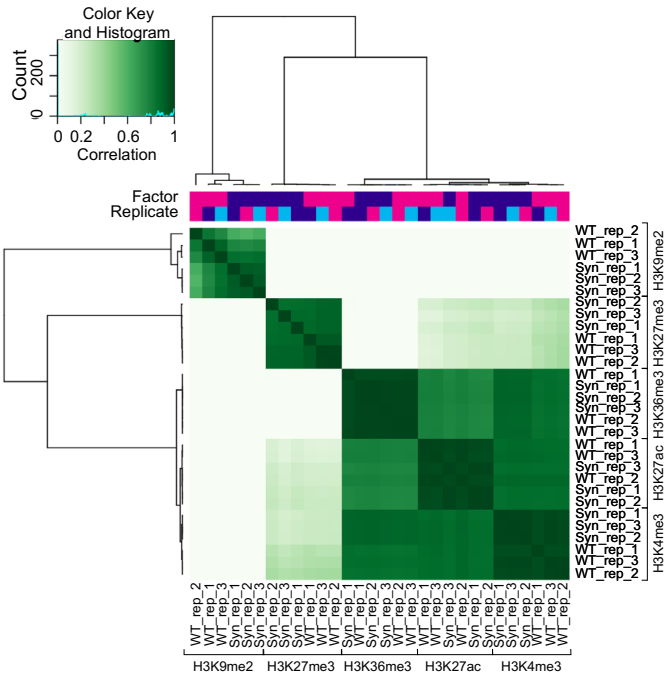
Extended Data Fig. 3 | Schematic diagram of the replacement of the synthetic region in 7 partial replacement lines. The light yellow segment represents the region that should have been deleted according to the design and the green segment represents the preserved region.

https://doi.org/10.1038/s41477-023-01595-7



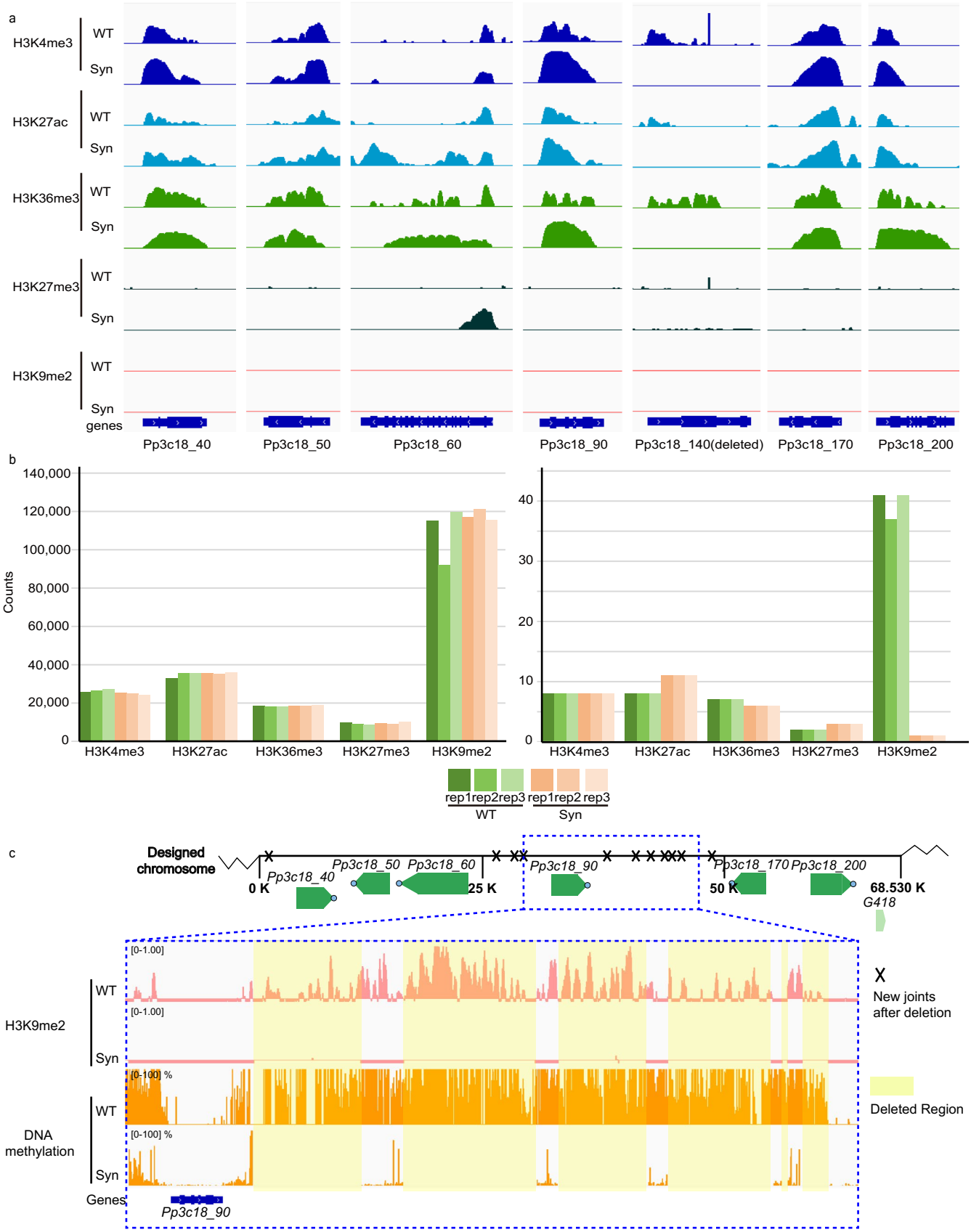
Extended Data Fig. 4 | Stress treatment of wild type and semi-syn18L. (**a**) NaCl treatment. The chlorophyll content of wild type and semi-syn18L was determined after 0 mM, 200 mM, 300 mM, 400 mM and 500 mM NaCl stress treatments (3 d). The bars shown in the histogram are means ± SDs. The points in the figure show the specific values in each repetition, n= 3 biological replicates. (**b**) Sorbitol treatment. The chlorophyll contents of wild type and semi-syn18L were determined after 0 mM, 400 mM, 500 mM, 600 mM and 750 mM sorbitol stress treatments (3 d). The bars shown in the histogram are means ± SDs. The points in the figure show the specific values in each repetition, n = 3 biological replicates.

https://doi.org/10.1038/s41477-023-01595-7



Extended Data Fig. 5 | Heatmap showing correlations among histone marks in the wild type and semi-syn18L. The correlation coefficient was calculated using the normalized read counts in each peak region. Colours represent the correlation coefficients, with darker green indicating higher similarity. Hierarchical clustering is shown at the left and top of the heatmap. Note that active histone modifications cluster together, while H3K9me2, which often represents heterochromatin, differs the most from other modifications. Overall, there is minimal difference in the distribution of histone modifications between the wild type and semi-syn18L from a genome-wide perspective. (also see Supplementary Table 7).

https://doi.org/10.1038/s41477-023-01595-7



Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41477-023-01595-7

Extended Data Fig. 6 | Other epigenetic changes in semi-syn18L. (**a**) Distribution of five histone modifications (H3K4me3, H3K9me2, H3K27me3, H3K36me3 and H3K27ac) on genes in the synthetic region. The signals shown were normalized using the BPM method (Bins Per Million mapped reads, same as TPM in RNA-seq) after deducting the input. (**b**) The number of peaks for histone modifications in wild-type and semi-syn18L. The left figure shows the number of histone modification peaks across the entire genome, while the right figure shows the number of histone modification peaks in the replacement region. The counts of three replicates were separately calculated. (**c**) Changes in H3K9me2 modification and DNA methylation levels at the remaining intergenenic region in semi-syn18L. The top panel shows designed chromosome fragment. The Xs on the top panel refer to new joints formed by deleting duplicate sequences, which correspond to the yellow delete regions below.

nature portfolio

Corresponding author(s): Yuling Jiao, Junbiao Dai, Wenfeng Qian, Ying Wang

Last updated by author(s): Nov 8, 2023

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	There were no specific codes used during the data collection process.
Data analysis	software/packages used: bedtools v2.30.0 bismark v2.5.0 BWA v0.7.17 Deeptools v3.5.7. featureCounts v2.0.3 IGV v2.14.0 juicer v1.6 macs2 v2.25 picard v2.27.5 samtools v1.16.1 Trimmomatic v0.39 BD FACSDiva Software v9.0.1 DESeq2 DiffBind v3.8.4 Primer3 RepeatMasker v4.1.5 Codes uesd in high-throughput sequencing analysis have been deposited at Github (https://github.com/lanntianlong/SynMoss) and Zenodo

(DOI:10.5281/zenodo.8000393). The in-house computational pipeline for standardizing genome design is available on Zenodo (DOI: 10.5281/zenodo.7894207).

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

All high-throughput sequencing data (ATAC-seq, Hi-C, ChIP-seq, RNA-seq, Whole-genome Sequencing and Whole Genome Bisulfite Sequencing) in this paper are contained in SRA (PRJNA970280), The raw gel image of PCR, the raw data of flow cytometry and the data of stress treatments are provided in the Supplementary Information file. The BigWig files of ChIP-seq are available at Figshare (https://figshare.com/articles/dataset/ChIP_track_rar/23648046). P. patens genome v3.3 are available at Phytozome (https://phytozome-next.jgi.doe.gov).

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	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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	In the experiments comparing semi-syn 18L and WT (ATAC-seq, ChIP-seq, RNA-seq, WGBS, phenotype observation, stress treatments), at least three samples were selected from each line for independent replication experiments to improve the reliability of the results.
Data exclusions	The low-quality reads from sequencing were removed using Trimmomatic, and after alignment, duplicate reads were removed using Picard.
Replication	We conducted three biological replicates for each individual experiment, and all of them were successful.
Randomization	The allocation of samples was random, and the materials we obtained were all obtained in the same culture environment.
Blinding	Because the analysis and evaluation in this study are non-subjective, the identification of synthetic lines (PCRTag analysis and resequencing), and the epigenetic analysis have clear and strict definition standards. Therefore, blinding is not necessary in this study.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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

Antibodies

Antibodies used	Rabbit polyclonal to Histone H3 (tri methyl K4) - ChIP Grade (Cat. # ab8580), Mouse monoclonal [mAbcam 1220] to Histone H3 (di methyl K9) - ChIP Grade(Cat. # ab1220), Mouse monoclonal [mAbcam 6002] to Histone H3 (tri methyl K27) - ChIP Grade (Cat. # ab6002), Rabbit polyclonal to Histone H3 (acetyl K27) - ChIP Grade (Cat. # ab4729), Rabbit polyclonal to Histone H3 (tri methyl K36) - ChIP Grade (Cat. # ab9050)
Validation	Rabbit polyclonal to Histone H3 (tri methyl K4) - ChIP Grade (Cat. # ab8580) IgG; from rabbit; suitable for: PepArr, ChIP, WB, IHC-P, ICC/IF ; reacts with: Cow, Human; be predicted to be used for Mouse, Rat, Rabbit, Pig, Saccharomyces cerevisiae, Tetrahymena, Xenopus laevis, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Indian muntjac, Oikopleura, Plants, Zebrafish, Mammals, Trypanosoma cruzi, Common marmoset, Rice, Xenopus tropicalis Mouse monoclonal [mAbcam 1220] to Histone H3 (di methyl K9) - ChIP Grade(Cat. # ab1220) IgG2a; from mouse; suitable for: ICC/IF, WB, ELISA, IHC-P, ChIP; reacts with: Cow, Human, Arabidopsis thaliana, Drosophila melanogaster, Rice; be predicted to be used for Mouse, Rat, Sheep, Chicken, Saccharomyces cerevisiae, Xenopus laevis, Caenorhabditis elegans, Schizosaccharomyces pombe, Corn, Common marmoset, Other species Mouse monoclonal [mAbcam 6002] to Histone H3 (tri methyl K27) - ChIP Grade (Cat. # ab6002) IgG3; from mouse; suitable for: ChIP, ELISA, WB, IHC - Wholemount, ICC/IF; reacts with: Mouse, Cow, Human, Recombinant fragment; be predicted to be used for Rat, Rabbit, Chicken, Xenopus laevis, Arabidopsis thaliana, Drosophila melanogaster, Plants, Zebrafish, Rhesus monkey, Chinese hamster, Rice Rabbit polyclonal to Histone H3 (acetyl K27) - ChIP Grade (Cat. # ab4729) IgG; from rabbit; suitable for: ICC/IF, WB, IHC-P, ChIP, PepArr; reacts with: Mouse, Rat, Cow, Human, Recombinant fragment; Reacts with: Mouse, Rat, Cow, Human, Recombinant fragment; be predicted to be used for Chicken, Xenopus laevis, Arabidopsis thaliana, Drosophila melanogaster, Monkey, Zebrafish, Plasmodium falciparum, Rice, Cyanidioschyzon merolae Rabbit polyclonal to Histone H3 (tri methyl K36) - ChIP Grade (Cat. # ab9050) IgG; from rabbit; suitable for: ICC/IF, WB, ChIP; suitable for: ICC/IF, WB, ChIP; reacts with: Cow, Human; be predicted to be used for Mouse, Rat, Saccharomyces cerevisiae, Xenopus laevis, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Plants, Schizosaccharomyces pombe, Zebrafish, Silk worm, Rice, Xenopus tropicalis, Trypanosoma brucei

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	*State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	*For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Reporting on sex	*Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

Outcomes | *Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

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	In this study, the background of all strains was the Gransden wild-type strain of P. patens.
Novel plant genotypes	Semi-syn 18L was generated through homologous recombination. We recombined the target sequence with homologous arms into genome through protoplast transformation.
Authentication	Different genotypes can be distinguished through PCR Tag we designed

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Raw sequence data generated during this study are available in SRA (PRJNA970280), BigWig files are available in figshare (https://figshare.com/articles/dataset/ChIP_track_rar/23648046)
Files in database submission	WT: B27ac-WT-1_L4_701D04.R1.fastq.gz In-27ac-WT-2_L2_702D04.R1.fastq.gz Input-H3K4-WT-1_L4_704D01.R1.fastq.gz B27ac-WT-1_L4_701D04.R2.fastq.gz In-27ac-WT-2_L2_702D04.R2.fastq.gz Input-H3K4-WT-1_L4_704D01.R2.fastq.gz B27ac-WT-2_L4_702D04.R1.fastq.gz In-27ac-WT-3_L4_703D04.R1.fastq.gz Input-H3K9me2-WT-1_L3_704D02.R1.fastq.gz B27ac-WT-2_L4_702D04.R2.fastq.gz In-27ac-WT-3_L4_703D04.R2.fastq.gz Input-H3K9me2-WT-1_L3_704D02.R2.fastq.gz B27ac-WT-3_L4_703D04.R1.fastq.gz In-27me-WT-2_L4_710D05.R1.fastq.gz Input-H3K9me2-WT-2_L3_705D02.R1.fastq.gz B27ac-WT-3_L4_703D04.R2.fastq.gz In-27me-WT-2_L4_710D05.R2.fastq.gz Input-H3K9me2-WT-2_L3_705D02.R2.fastq.gz B27me-WT-2_L4_712D05.R1.fastq.gz In-36-WT-3_L4_704D04.R1.fastq.gz Input-H3K9me2-WT-3_L3_706D02.R1.fastq.gz B27me-WT-2_L4_712D05.R2.fastq.gz In-36-WT-3_L4_704D04.R2.fastq.gz Input-H3K9me2-WT-3_L3_706D02.R2.fastq.gz B36-WT-3_L4_704D04.R1.fastq.gz In-K27me-WT-2_L2_703D06.R1.fastq.gz K27me-WT-2_L2_709D06.R1.fastq.gz B36-WT-3_L4_704D04.R2.fastq.gz In-K27me-WT-2_L2_703D06.R2.fastq.gz K27me-WT-2_L2_709D06.R2.fastq.gz H3K4-WT-1_L2_709D01.R1.fastq.gz In-K27me-WT-3_L2_704D06.R1.fastq.gz K27me-WT-3_L2_710D06.R1.fastq.gz

H3K4-WT-1_L2_709D01.R2.fastq.gz In-K27me-WT-3_L2_704D06.R2.fastq.gz K27me-WT-3_L2_710D06.R2.fastq.gz
H3K9me2-WT-1_L3_710D02.R1.fastq.gz In-K36-WT-2_L2_705D06.R1.fastq.gz K36-WT-2_L2_711D06.R1.fastq.gz
H3K9me2-WT-1_L3_710D02.R2.fastq.gz In-K36-WT-2_L2_705D06.R2.fastq.gz K36-WT-2_L2_711D06.R2.fastq.gz
H3K9me2-WT-2_L3_711D02.R1.fastq.gz In-K36-WT-3_L2_706D06.R1.fastq.gz K36-WT-3_L2_712D06.R1.fastq.gz
H3K9me2-WT-2_L3_711D02.R2.fastq.gz In-K36-WT-3_L2_706D06.R2.fastq.gz K36-WT-3_L2_712D06.R2.fastq.gz
H3K9me2-WT-3_L3_712D02.R1.fastq.gz In-K4-WT-2_L1_701D06.R1.fastq.gz K4-WT-2_L2_707D06.R1.fastq.gz
H3K9me2-WT-3_L3_712D02.R2.fastq.gz In-K4-WT-2_L1_701D06.R2.fastq.gz K4-WT-2_L2_707D06.R2.fastq.gz
In-27ac-WT-1_L3_701D04.R1.fastq.gz In-K4-WT-3_L2_702D06.R1.fastq.gz K4-WT-3_L2_708D06.R1.fastq.gz

semi-syn 18L:
In-H3K4-14-1_L4_D503D701.R1.fastq.gz In-K36me3-2-14_L1_Q0014W0072.R1.fastq.gz IP-K27ac-14-3_L1_Q0012W0073.R1.fastq.gz
In-H3K4-14-1_L4_D503D701.R2.fastq.gz In-K36me3-2-14_L1_Q0014W0072.R2.fastq.gz IP-K27ac-14-3_L1_Q0012W0073.R2.fastq.gz
In-H3K4-14-2_L4_Q0013W0066.R1.fastq.gz In-K36me3-3-14_L1_Q0021W0072.R1.fastq.gz IP-K27me3-14-1_L1_Q0009W0073.R1.fastq.gz
In-H3K4-14-2_L4_Q0013W0066.R2.fastq.gz In-K36me3-3-14_L1_Q0021W0072.R2.fastq.gz IP-K27me3-14-1_L1_Q0009W0073.R2.fastq.gz
In-H3K4-14-3_L1_Q0028W0066.R1.fastq.gz In-K9me2-1-14_L1_Q0023W0072.R1.fastq.gz IP-K27me3-14-2_L2_Q0248W0073.R1.fastq.gz
In-H3K4-14-3_L1_Q0028W0066.R2.fastq.gz In-K9me2-1-14_L1_Q0023W0072.R2.fastq.gz IP-K27me3-14-2_L2_Q0248W0073.R2.fastq.gz
In-K27ac-14-1_L1_Q0016W0073.R1.fastq.gz In-K9me2-2-14_L1_Q0013W0072.R1.fastq.gz IP-K27me3-14-3_L1_Q0007W0073.R1.fastq.gz
In-K27ac-14-1_L1_Q0016W0073.R2.fastq.gz In-K9me2-2-14_L1_Q0013W0072.R2.fastq.gz IP-K27me3-14-3_L1_Q0007W0073.R2.fastq.gz
In-K27ac-14-2_L1_Q0014W0073.R1.fastq.gz In-K9me2-3-14_L1_Q0028W0072.R1.fastq.gz IP-K36me3-1-14_L1_Q0006W0072.R1.fastq.gz
In-K27ac-14-2_L1_Q0014W0073.R2.fastq.gz In-K9me2-3-14_L1_Q0028W0072.R2.fastq.gz IP-K36me3-1-14_L1_Q0006W0072.R2.fastq.gz
In-K27ac-14-3_L1_Q0021W0073.R1.fastq.gz IP-H3K4-14-1_L4_Q0016W0066.R1.fastq.gz IP-K36me3-2-14_L1_Q0010W0072.R1.fastq.gz
In-K27ac-14-3_L1_Q0021W0073.R2.fastq.gz IP-H3K4-14-1_L4_Q0016W0066.R2.fastq.gz IP-K36me3-2-14_L1_Q0010W0072.R2.fastq.gz
In-K27me3-14-1_L1_Q0023W0073.R1.fastq.gz IP-H3K4-14-2_L4_Q0014W0066.R1.fastq.gz IP-K36me3-3-14_L1_Q0012W0072.R1.fastq.gz
In-K27me3-14-1_L1_Q0023W0073.R2.fastq.gz IP-H3K4-14-2_L4_Q0014W0066.R2.fastq.gz IP-K36me3-3-14_L1_Q0012W0072.R2.fastq.gz
In-K27me3-14-2_L1_Q0013W0073.R1.fastq.gz IP-H3K4-14-3_L4_Q0021W0066.R1.fastq.gz IP-K9me2-1-14_L1_Q0009W0072.R1.fastq.gz
In-K27me3-14-2_L1_Q0013W0073.R2.fastq.gz IP-H3K4-14-3_L4_Q0021W0066.R2.fastq.gz IP-K9me2-1-14_L1_Q0009W0072.R2.fastq.gz
In-K27me3-14-3_L1_Q0028W0073.R1.fastq.gz IP-K27ac-14-1_L2_Q0006W0073.R1.fastq.gz IP-K9me2-2-14_L1_Q0248W0072.R1.fastq.gz
In-K27me3-14-3_L1_Q0028W0073.R2.fastq.gz IP-K27ac-14-1_L2_Q0006W0073.R2.fastq.gz IP-K9me2-2-14_L1_Q0248W0072.R2.fastq.gz
In-K36me3-1-14_L1_Q0016W0072.R1.fastq.gz IP-K27ac-14-2_L2_Q0010W0073.R1.fastq.gz IP-K9me2-3-14_L1_Q0007W0072.R1.fastq.gz
In-K36me3-1-14_L1_Q0016W0072.R2.fastq.gz IP-K27ac-14-2_L2_Q0010W0073.R2.fastq.gz IP-K9me2-3-14_L1_Q0007W0072.R2.fastq.gz

Genome browser session
(e.g. UCSC)

The IGV session can be downloaded from (https://figshare.com/articles/dataset/ChIP_track_rar/23648046)

Methodology

Replicates

Three replicate experiments were performed for each ChIP-seq of H3K9me2, H3K27ac, H3K27me3, H3K4me3, and H3K36me3 in both wild-type and semi-syn 18L

Sequencing depth

All the reads are paied-end and length of all reads is 150bp. The specific number of reads is as follows:
WT:
sample total-read mapped-read properly-paired
K27ac_1 79689402 79448460 76165864
K27ac_2 97725532 97515351 96460870
K27ac_3 78692249 78482710 75084042
K27ac_1_input 63263696 62557633 61367058
K27ac_2_input 63144650 62528096 61581578
K27ac_3_input 67027290 66317939 65419392
K27me3_1 43766779 30889441 30520286
K27me3_2 92814413 92291425 91207946
K27me3_3 113901290 113192616 111527464
K27me3_1_input 120667137 117719493 116885856
K27me3_2_input 82475569 78945334 76418650
K27me3_3_input 109830020 107434667 106752766
K36me3_1 106202245 105585543 104689518
K36me3_2 98125596 97379936 96603472

K36me3_3 68179232 67565961 62801430
K36me3_1_input 97108831 96555432 96037752
K36me3_2_input 111508738 111044432 110436402
K36me3_3_input 50263973 48001136 46955478
K4me3_1 49804909 31724125 29173048
K4me3_2 101985111 100435527 100047114
K4me3_3 123385975 122685369 121276338
K4me3_1_input 51402187 50995958 49812408
K4me3_2_input 75875784 75200587 74607406
K4me3_3_input 104706958 104344840 103781244
K9me2_1 51168260 51048895 45920798
K9me2_2 61238734 61008124 53071170
K9me2_3 62462344 62293464 58550118
K9me2_1_input 51033526 50371281 49694424
K9me2_2_input 46225271 45400624 44553354
K9me2_3_input 43029138 42431573 41655540

Semi-syn 18L:
In-H3K4-14-1_L4_D503D701_raw.bam 122876569 122667544 121200918
In-H3K4-14-2_L4_Q0013W0066_raw.bam 57229535 57126186 56033038
In-H3K4-14-3_L1_Q0028W0066_raw.bam 54582993 54504032 53421062
In-K27ac-14-1_L1_Q0016W0073_raw.bam 49741606 49655923 48742736
In-K27ac-14-2_L1_Q0014W0073_raw.bam 49261504 49179279 47948348
In-K27ac-14-3_L1_Q0021W0073_raw.bam 61699589 61621941 60449670
In-K27me3-14-1_L1_Q0023W0073_raw.bam 52304532 52178532 51320860
In-K27me3-14-2_L1_Q0013W0073_raw.bam 54697716 54559009 53822992
In-K27me3-14-3_L1_Q0028W0073_raw.bam 46110948 46051516 45087146
In-K36me3-1-14_L1_Q0016W0072_raw.bam 60881570 60779726 60156530
In-K36me3-2-14_L1_Q0014W0072_raw.bam 50113209 50038646 49640638
In-K36me3-3-14_L1_Q0021W0072_raw.bam 61607087 61510559 60799026
In-K9me2-1-14_L1_Q0023W0072_raw.bam 51147730 51041111 50407868
In-K9me2-2-14_L1_Q0013W0072_raw.bam 66070353 65915822 65128914
In-K9me2-3-14_L1_Q0028W0072_raw.bam 60797866 60691901 60127766
IP-H3K4-14-1_L4_Q0016W0066_raw.bam 41039772 40927094 38482870
IP-H3K4-14-2_L4_Q0014W0066_raw.bam 58153987 57936548 54845802
IP-H3K4-14-3_L4_Q0021W0066_raw.bam 67090357 66845770 62450824
IP-K27ac-14-1_L2_Q0006W0073_raw.bam 100251045 100047153 94421282
IP-K27ac-14-2_L2_Q0010W0073_raw.bam 109371518 109036765 98290884
IP-K27ac-14-3_L1_Q0012W0073_raw.bam 71956710 71768209 65785562
IP-K27me3-14-1_L1_Q0009W0073_raw.bam 92570848 88471493 79020986
IP-K27me3-14-2_L2_Q0248W0073_raw.bam 88606718 85098033 76746588
IP-K27me3-14-3_L1_Q0007W0073_raw.bam 56499876 54234823 49978236
IP-K36me3-1-14_L1_Q0006W0072_raw.bam 85017886 84860543 81342520
IP-K36me3-2-14_L1_Q0010W0072_raw.bam 76959673 76838700 71637720
IP-K36me3-3-14_L1_Q0012W0072_raw.bam 60442763 60364302 57509364
IP-K9me2-1-14_L1_Q0009W0072_raw.bam 42832715 42809269 40413730
IP-K9me2-2-14_L1_Q0248W0072_raw.bam 55235200 55206014 52942518
IP-K9me2-3-14_L1_Q0007W0072_raw.bam 42891635 42869033 41158606

Antibodies

Rabbit polyclonal to Histone H3 (tri methyl K4) - ChIP Grade (Cat. # ab8580),
Mouse monoclonal [mAbcam 1220] to Histone H3 (di methyl K9) - ChIP Grade(Cat. # ab1220),
Mouse monoclonal [mAbcam 6002] to Histone H3 (tri methyl K27) - ChIP Grade (Cat. # ab6002),
Rabbit polyclonal to Histone H3 (acetyl K27) - ChIP Grade (Cat. # ab4729),
Rabbit polyclonal to Histone H3 (tri methyl K36) - ChIP Grade (Cat. # ab9050)

Peak calling parameters

```
bwa mem -M -t 50 -R '@RG\tID:804D$ID\tPL:illumina\tLB:library\tSM:SYN_WT' $ref ${base}_1_paired.fq.gz ${base}_2_paired.fq.gz |
samtools view -S -b -> ${base}_raw.bam
java -jar /data/ltl/jar/picard.jar MarkDuplicates REMOVE_DUPLICATES=true I=${base}_sorted.bam O=${base}_sorted_rm.bam M=
${base}_sorted_rm_log
samtools view -h -f 2 -q 30 ${base}_sorted_rm.bam | samtools sort -O bam -@ 10 -o - > ${base}_last.bam
macs2 callpeak -t ${base}_last.bam -c ${base}_input_last.bam -f BAM --outdir ${base}_Chipseq -n ${base}_Chipseq -B --nomodel --
extsize 165 --keep-dup all
```

Data quality

ID Factor Replicate FRiP
H3K27me3_syn_1 syn 1 0.19
H3K27me3_syn_2 syn 2 0.21
H3K27me3_syn_3 syn 3 0.2
H3K27me3_wt_1 wt 1 0.25
H3K27me3_wt_2 wt 2 0.29
H3K27me3_wt_3 wt 3 0.31
H3K36me3_syn_1 syn 1 0.2
H3K36me3_syn_2 syn 2 0.2
H3K36me3_syn_3 syn 3 0.2
H3K36me3_wt_1 wt 1 0.31
H3K36me3_wt_2 wt 2 0.27

H3K36me3_wt_3 wt 3 0.26
H3K4me3_syn_1 syn 1 0.47
H3K4me3_syn_2 syn 2 0.47
H3K4me3_syn_3 syn 3 0.48
H3K4me3_wt_1 wt 1 0.39
H3K4me3_wt_2 wt 2 0.48
H3K4me3_wt_3 wt 3 0.48
H3K9me2_syn_1 syn 1 0.24
H3K9me2_syn_2 syn 2 0.25
H3K9me2_syn_3 syn 3 0.25
H3K9me2_wt_1 wt 1 0.24
H3K9me2_wt_2 wt 2 0.24
H3K9me2_wt_3 wt 3 0.24
H3K27ac_syn_1 syn 1 0.41
H3K27ac_syn_2 syn 2 0.4
H3K27ac_syn_3 syn 3 0.41
H3K27ac_wt_1 wt 1 0.4
H3K27ac_wt_2 wt 2 0.44
H3K27ac_wt_3 wt 3 0.42

Software	Trimmomatic v0.39, BWA v0.7.17, Picard v2.27.5, SAMtools v1.16.1, MACS2 v2.25, deepTools v3.5.7

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	The gametophore stage materials were cut into pieces by blades, added to the nucleus extraction buffer, filtered out the fragments, collected the filtrate and added DAPI to stain the nucleus, and then put into flow cytometry for analysis.
Instrument	BD FACSAria II Flow Cytometer 2 Laser Base Configuration
Software	BD FACSDiva Software v9.0.1
Cell population abundance	According to the strength of DAPI signal, the cells are mainly divided into haploid and dividing cells. About 80% of them are haploid
Gating strategy	The cell division indexes were as follows: FSC was 1 ~ 1×105, SSC was 500 ~ 3×104. Then the target cells were divided according to the DAPI intensity of more than 20,000.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*