https://doi.org/10.1038/s41587-025-02913-4

Building synthetic chromosomes one yeast at a time: insights from Sc2.0

Paige E. Erpf, Felix Meier, Roy S. K. Walker, Hugh D. Goold, Jef D. Boeke, Ian T. Paulsen & Isak S. Pretorius

The Synthetic Yeast Genome Project (Sc2.0) set out to redesign and chemically synthesize an entire eukaryotic genome. This Comment summarizes the design- and construction-related defects revealed during the construction of 16 synthetic chromosomes, and the solutions applied, drawing out the key biological and technical insights that will inform future genome-scale engineering.

Sc2.0 emerged as a global collaborative effort to redesign and synthesize the entire *Saccharomyces cerevisiae* genome from scratch[1]. It presented a unique set of opportunities and challenges that prompted the use of rational design principles to enhance genome stability, flexibility and engineering potential. Key features and changes to the native genome included the removal of transposable elements, the introduction of symmetrical *loxP* (loxPsym) sites for genome rearrangement (via the SCRaMbLE inducible evolution system[2]), the reassignment of stop codons and an efficient design to support future genome minimization and reprogramming (Supplementary Fig. 1a). Each synthetic chromosome followed the same design principles used in the assembly of the first semi-synthetic chromosome, *synIXR*[2], and was subsequently constructed by different international teams[1]. An overview of the Sc2.0 project is provided in Supplementary Table 1.

In 2025, the final synthetic chromosome was completed, marking the culmination of more than a decade of iterative design and international collaboration. The timeline of completion reflects not only the distributed nature of the consortium, but also the ongoing evolution and optimization of synthetic genomics as a discipline (Supplementary Fig. 1b). Several strategies and techniques were used to overcome the many challenges faced by the consortium. Here, we discuss how these findings will shape the future of synthetic genomics.

How to build a synthetic yeast genome with SwAP-In

New approaches were required to construct the synthetic yeast genome. A key design feature of Sc2.0 was the development and use of switching auxotrophies progressively for integration (SwAP-In), a stepwise method for assembling synthetic chromosomes in *S. cerevisiae*[1,2]. In this approach, chromosomes were divided into large segments known as megachunks (30–60 kb), each composed of smaller chunks (about 10 kb) or in some cases minichunks (about 3.0 kb) that were first assembled in vitro using restriction enzyme-based ligation. Megachunks were sequentially integrated into the yeast genome using the cell's native homologous recombination machinery. To facilitate the selection and tracking of integration events, the alternating selectable markers *URA3* and *LEU2* were overwritten with each incoming segment. The presence of unique primer binding sites, termed PCRTags, also provided an efficient and low-cost means to validate the integration of synthetic DNA.

Although SwAP-In served as the standard methodology across the consortium, teams iterated on the original method to suit their individual needs and experimental challenges. These variations arose in both major phases of construction: in vitro assembly of chunks and in vivo integration into yeast.

Most groups adopted the canonical SwAP-In strategy for in vitro assembly, but several teams implemented strategic deviations to address chromosome-specific challenges. For example, the *synIV* and *synVIII* teams developed an approach to directly clone and assemble megachunks from pre-synthesized DNA fragments, eliminating the intermediate chunk-assembly phase. This modification reduced assembly complexity and minimized potential recombination artifacts. Meanwhile, the *synIX* team identified residual wild-type DNA scraps in synthetic strains following conventional enzyme-based minichunk ligation. To circumvent this contamination, they adopted the megachunk–bacterial artificial chromosome (BAC) strategy developed by the *synIV* team. In this method, megachunks are assembled as extrachromosomal BACs in yeast, transferred to *Escherichia coli* for purification and verification, and then reintroduced into yeast after restriction-based excision from the plasmid backbone. This strategy offered cleaner integration and higher sequence fidelity in final constructs.

Variations were also common in the in vivo phase of SwAP-In. Several teams (*synII*, *synVII*, *synXIII*, *synXIV* and *synXVI*) adopted a parallel integration strategy, in which megachunks were assembled into two separate parental strains of opposite mating types, each carrying synthetic segments from one end of the native chromosome. These strains were then crossed to generate a heterozygous diploid carrying both synthetic chromosome halves. To assemble these chromosome halves into a complete, markerless synthetic chromosome, targeted double-strand breaks were introduced at the junction sites using the I-SceI endonuclease, facilitating homology-directed repair between the overlapping synthetic DNA sequences. This approach enhanced assembly efficiency and enabled precise spatial control over homologous recombination sites. Another strategy involved the creation of multiple semi-synthetic strains, each carrying one or more synthetic megachunks, which were then crossed and sporulated to assemble the complete synthetic chromosome using meiotic recombination-mediated assembly (applied in *synIV*, *synVIII*, *synXII* and *synXVI*).

To address some of the limitations inherent to SwAP-IN, the *synXV* team developed the CRISPR–Cas9-mediated mitotic recombination with endoreduplication (CRIMiRE) system as an alternative approach for precise chromosome-scale engineering. In this method, CRISPR–Cas9 was used to rationally recombine synthetic and wild-type chromosome segments into yeast. Here, megachunks were first combined

Table 1 | Design- and construction-related defects reported across synthetic yeast chromosomes

	Synthetic chromosomes																
	Syn I	**Syn II**	**Syn III**	**Syn IV**	**Syn V**	**Syn VI**	**Syn VII**	**Syn VIII**	**Syn IX**	**Syn X**	**Syn XI**	**Syn XII**	**Syn XIII**	**Syn XIV**	**Syn XV**	**Syn XVI**	**Total**
Size (kb)	181	770	272	1,454	536	242	1,028	506	405	707	659	999	883	753	1,048	902	
Design defects																	
PCRTags	✓	✓		✓		✓				✓		✓	✓		✓	✓	9
REPEATSMASHER	✓			✓				✓								✓	4
LoxPsym sites						✓	✓	✓		✓	✓			✓	✓	✓	8
SwAP-In		✓									✓	✓				✓	4
Removal of tRNAs		✓					✓		✓			✓					4
Removal of introns										✓		✓		✓			3
Construction defects																	
Patchwork DNA		✓	✓	✓	✓	✓			✓	✓			✓				8
Missing stop codons and LoxPsym sites	✓	✓	✓	✓	✓			✓		✓	✓	✓	✓	✓	✓	✓	13
Megachunk duplications		✓		✓				✓		✓	✓			✓			6
Retention of vector backbones		✓									✓	✓				✓	4
Mitochondrial loss and petite colonies						✓				✓	✓	✓		✓		✓	6
Karyotyping instability and genome duplication		✓	✓						✓		✓		✓				5
Total	3	8	3	5	2	4	2	4	3	7	7	7	4	5	3	7	

into four hyperchunks (HC1, megachunks A–M; HC2, N–AA; HC3, AA–GG; and HC4, GG–PP), which were then fused through targeted mitotic recombination. CRIMiRE used uniquely recoded sequences, selectable markers and a conditional centromere to drive recombination and selectively eliminate undesired chromosome versions, allowing the isolation of fully synthetic, homozygous diploid strains. This strategy enabled the complete and accurate assembly of *synXV*, offering a method for customizable chromosome reconfiguration in genome-scale engineering.

Although the SwAP-In protocol was designed with a unified framework, the evolution of methods across teams offered valuable insights into the strengths and limitations of chromosome-scale assembly methods. Importantly, these variations represent refinements that may guide more efficient and accurate genome assembly methods in the future. As synthetic genomics continues to scale, the lessons learned from SwAP-In and its derivatives provide a blueprint for more robust construction strategies for use in yeast and beyond.

Shared challenges in genome-scale engineering

Some of the most valuable lessons from the Sc2.0 consortium arose from the identification of 'bugs' following the integration of synthetic DNA. Eleven laboratories constructed one or more designated chromosome(s) according to standardized consortium design principles. Despite the methodological differences, a striking pattern emerged: many of the same sequence-level issues and assembly 'bugs' appeared across the consortium. Recurring errors were traced to two main sources: design decisions that later proved problematic and spontaneous, complex mutations that arose during construction (Table 1).

Design-related defects

Saccharomyces cerevisiae remains one of the most thoroughly characterized eukaryotic organisms, and systematic understanding of yeast genetics and genome structure enabled the rational design of Sc2.0 without compromising organismal viability. Nonetheless, the scale and novelty of the synthetic genome introduced design-level challenges that were not fully anticipated.

PCRTags. The development of a watermarking system was a critical feature of Sc2.0, enabling efficient screening for successful integration of synthetic sequences while replacing native DNA[2]. To implement this, short (28-bp) recoded sequences known as PCRTags were systematically introduced into open reading frames (ORFs) using BioStudio. PCRTags were designed to be silent, incorporating only synonymous codon changes unique to either the wild-type or synthetic genome (Fig. 1a). Although the vast majority of PCRTags introduced into Sc2.0 had no detectable effects on fitness, multiple groups across the consortium reported that they occasionally introduced unexpected functional defects, including disruptive mRNA secondary structures (*RRP8* in *synIV* and *PRE4* in *synVI*), interference with transcription factor-binding sites (*FIP1* in *synX*), formation of divergent promoter regions (*IRA2* in *synXV*) and nonsynonymous mutations (*PPQ1* in *synXVI*). In other cases, genes such as *PRM9* (*synI*), *TSC10* (*synII*) and *MMM1* (*synXII*) were identified as related to PCRTag-related growth defects and simply reverted to wild type without investigation. These findings suggest that although watermarking is a valuable feature, minimizing the number of PCRTags in future designs may reduce the burden of downstream debugging, as proposed by the *synXVI* team.

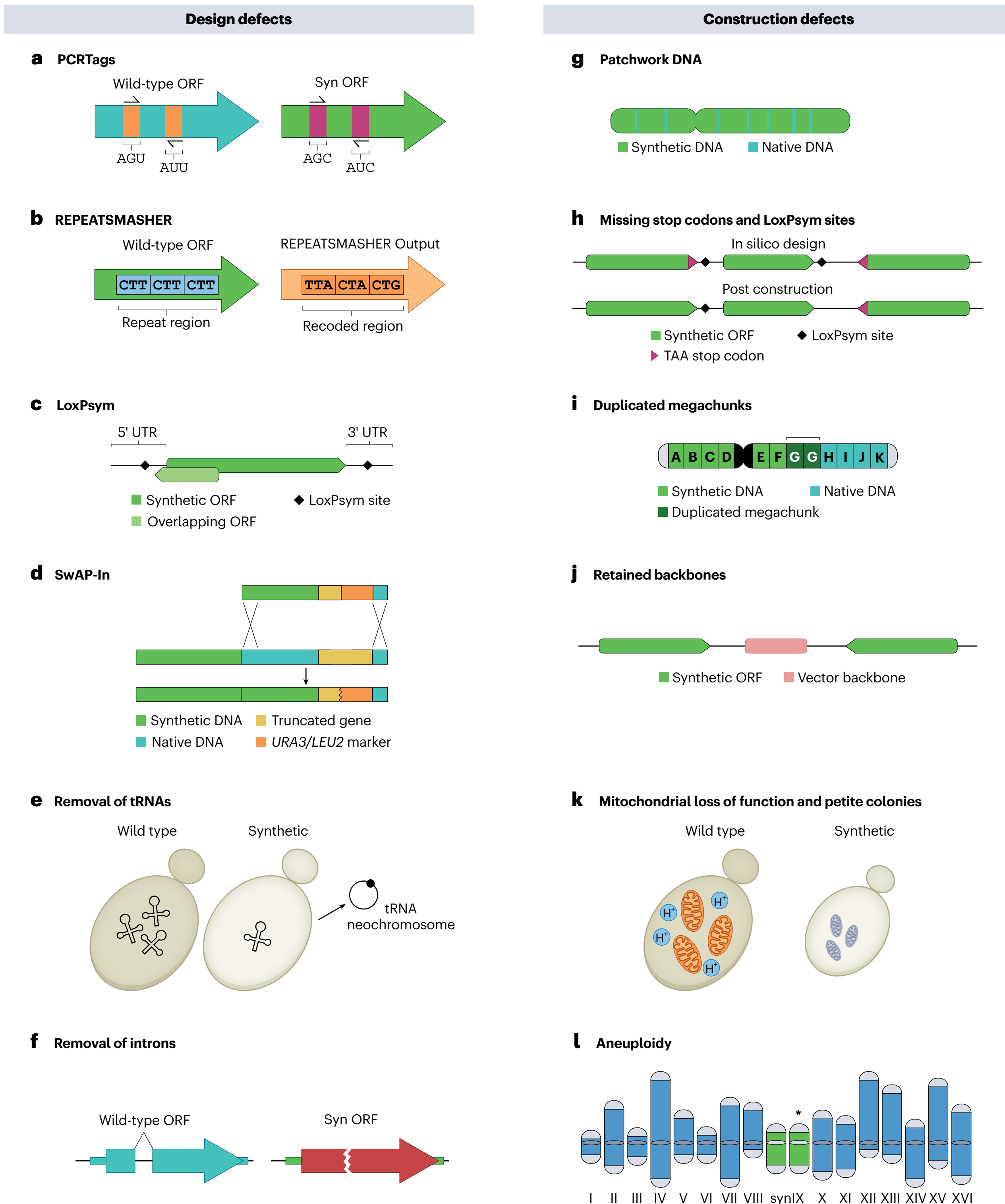

Fig. 1 | Schematic overview of design- and construction-related defects observed across Sc2.0. a–f, The left column illustrates design-related defects, including unintended sequence changes introduced by PCRTags (**a**); gene recoding errors from REPEATSMASHER (**b**); loxPsym insertions that interfered with regulatory regions or gene function (**c**); growth defects linked to SwAP-In integration sites (**d**); the removal and relocation of tRNA genes (**e**); and loss of essential introns (**f**). **g–l**, The right column illustrates construction-related defects, including patchwork retention of native DNA (**g**); missing stop codons and loxPsym sites (**h**); duplicated megachunks (**i**); retained plasmid backbone sequences (**j**); mitochondrial loss of function and petite colonies (**k**); and aneuploidy arising during chromosome assembly (with * highlighting the duplicated synthetic chromosome) (**l**).

Even though most teams encountered PCRTags as a source of unintended defects, one group observed the opposite. In *synXIII*, a synonymous codon change introduced by a PCRTag in the *RRN9* gene led to increased expression and a corresponding extension of cellular lifespan. This finding highlights the potential for subtle design modifications to uncover previously unrecognized aspects of gene regulation and cell physiology.

REPEATSMASHER. A key design feature of Sc2.0 was the removal of tandem repeat sequences within coding DNA sequences, with the goal of reducing repetitiveness and improving DNA synthesis efficiency. To facilitate this, the REPEATSMASHER module from GeneDesign was used to modify sequences flagged as repetitive[1,3]. However, in at least five cases, REPEATSMASHER recoded entire ORFs rather than just the repetitive segments, including essential genes such as *NOP1* and *PCF11* (*synIV*; Fig. 1b). These recoding events highlight the critical need for careful in silico validation of synthetic designs before synthesis, particularly for essential genes, as changes can have detrimental effects on gene function and overall strain viability.

LoxPsym sites. A defining feature of the Sc2.0 genome design was the systematic integration of palindromic loxPsym sites, positioned 3 bp downstream of all nonessential ORFs. These 34-bp sequences support the synthetic chromosome rearrangement and modification by *loxP*-mediated evolution (SCRaMbLE) system, which enables large-scale in vivo genome rearrangement to create a combinatorially rich diversity of genotypes and corresponding phenotypes. Across the consortium, a total of 3,932 loxPsym sites were designed for insertion[1]. Although the vast majority were tolerated, a subset of these sites introduced functional defects.

Eight synthetic chromosome groups reported loxPsym-associated phenotypes, affecting 13 genes. The underlying mechanisms of disruption varied but could be broadly grouped into two categories. The first involved the introduction of loxPsym sites near dubious ORFs, which inadvertently disrupted the overlapping promoter regions of nearby verified genes (Fig. 1c; *NSR1* in s*ynVII*, *ATP* in *synX*, *CPA1* in *synXV* and *CTR1* in s*ynXVI*). The second common interference was the loss of 3′ UTR-mediated mitochondrial mRNA targeting, leading to compromised respiratory function. This was observed for the nuclear genes encoding the mitochondrially located proteins *OMA1*, *MRS4* and *MRPL20* (*synXI*), and *MRPL19* (*synXIV*). In these instances, it is presumed that the loxPsym insertion disrupted elements required for proper mRNA localization.

Additional loxPsym-related anomalies included a 34-bp deletion in *CEN11*, in which the *synXI* centromere was inadvertently truncated and the loxPsym site was misplaced 66 bp downstream. A related instance occurred in *synVI*, in which cellular His2 levels were traced to a loxPsym site that had replaced a tRNA gene located upstream of *HIS2*, probably creating a cryptic transcriptional start site or causing transcriptional interference with the native promoter. Restoration of the upstream sequence resolved this defect. A further issue occurred in *synXI*, in which loxPsym elements had been incorrectly assigned to the 3′ ends of *HXT4* and *MTG2* owing to annotation errors in gene boundaries. To repair the defects, these sites were manually removed to prevent disruption of coding regions.

Although the inclusion of loxPsym sites is central to enabling post-construction genome editing and evolution, the Sc2.0 experience underscores the importance of carefully curating insertion sites. Based on these findings, it is recommended that loxPsym sites be excluded from nuclear encoded mitochondrial genes, particularly those requiring mRNA localization signals, and not assigned within or near regulatory regions such as promoters or untranslated regions of verified ORFs.

SwAP-In. Although SwAP-In was designed to target homology arms to nonessential genes, this strategy occasionally led to functional defects. Some disrupted genes, nonessential in terms of viability, were found to cause substantial growth defects, an undesirable outcome for the project (Fig. 1d). In four cases, SwAP-In integration led to observable growth defects (Table 1). In *synII*, disruption of *NCL1*, encoding a tRNA:m^{5}C-methyltransferase, by the *URA3* cassette resulted in temperature sensitivity and slow growth phenotypes that were mitigated upon reintegration of an intact *NCL1*. Similar defects were observed in *synXI*, where subsequent rounds of SwAP-In restored normal growth.

Phenotypic disruptions observed in *synXII* and *synXVI* showed irreversible characteristics, persisting despite restoration of native gene sequences. These synthetic chromosomes harbored modifications affecting genes encoding mitochondrial function-related proteins. In *synXII*, growth was substantially slowed, delaying the integration cycle. In *synXVI*, mitochondrial loss of function occurred, an outcome that could not be resolved through standard repair, as *S. cerevisiae* lacks the ability to regenerate mitochondrial DNA de novo. In these cases, the underlying genomic modifications responsible for mitochondrial genome instability first had to be identified and corrected; only then could strains be backcrossed into wild-type backgrounds carrying intact mitochondrial DNA, allowing synthetic chromosome construction to resume.

Removal of tRNAs. The relocation of all tRNA genes onto a dedicated neochromosome was a key design feature of the project, aimed at reducing genomic instability[1]. This strategy was adopted because tRNA loci are known hotspots for replication fork stalling events, presumed to be caused by collisions between components associated with the replication fork and heavily transcribed tRNA genes[4], and for Ty element insertions, both of which can compromise genome integrity[5].

The removal of tRNA genes presented a recurring challenge across multiple synthetic chromosomes (Fig. 1e, Table 1). Although they are not individually essential, the loss of some low-copy tRNA genes can create issues with translational supply and demand. In *synII*, the deletion of 13 tRNA genes led to a compensatory upregulation of translation machinery, a response that subsided once the tRNA array was introduced. In *synVII*, removal of the tRNA *tN(GUU)G* gene led to a fitness defect that was subsequently resolved through restoration of the gene. A more complex case was observed in *synIX*, where deletion of a *tRNA^{Asp}* gene and associated noncoding elements upstream of *EST3* resulted in reduced Est3p expression and shortened telomeres, which impaired growth at 37 °C. Reintroduction of the deleted region, including the tRNA and an adjacent Ty1 long terminal repeat element, restored fitness and telomere length. Similarly, *synXII* showed a growth defect traced to the removal of two leucine tRNA genes (*tL(UAG)* and *tL(UAG)*) during megachunk assembly. Restoration of either gene, or inclusion of a synthetic array containing all tRNA loci from chromosome XII, fully rescued growth and cell cycle progression. Together, these cases highlight the context-dependent roles of tRNA genes beyond their standard function in translation, with implications for genome design, copy number thresholds and chromosomal organization.

Removal of introns. Prior to the synthetic yeast genome project, several studies on yeast introns reported that only about 5% of genes in *S. cerevisiae* contain introns, with roughly one-third of those not

essential for growth[6]. This work also noted that the most abundantly spliced pre-mRNAs correspond to ribosomal protein genes[7]. Moreover, deletion of ribosomal protein gene introns frequently caused growth defects, even though most of these genes are present in duplicate[7]. Consequently, the decision was made mid-project to retain this intron class in the larger synthetic chromosomes, although several chromosomes entirely lack them. Even for essential gene introns whose removal impaired fitness, fitness could often be restored by adjusting gene expression levels, leaving open the long-term question of whether all pre-mRNA introns can ultimately be eliminated from Sc2.0. Based on these findings, the design of the synthetic yeast genome called for introns to be systematically removed from all genes except those known to cause a fitness defect[1] (Fig. 1f). However, in three separate cases across the consortium, intron removal proved problematic. In *synX*, the intron of the uncharacterized gene *YJR079W* was restored, although no explanation was provided; as *YJR079W* overlaps the mitochondrial architecture gene *AIM24*, it remains unclear whether the defect arose from disruption of one or both genes. In *synXII*, deletion of a hypothetical intron within the 5′ UTR of *COQ9* led to a transcriptional block and defects in mitochondrial function, probably due to partial deletion of the promoter region. In *synXIV*, removal of introns from *NOG2* and *SUN4* caused fitness defects; *NOG2* encodes the small nucleolar RNA snR191 and its intron deletion produced a marked growth defect, whereas the *SUN4* intron caused only a minor stress-linked phenotype. Reintroduction of these introns restored or stabilized fitness, and they were retained in the final synthetic strains.

These cases underscore the fact that introns, though often considered dispensable, might have critical regulatory roles. Careful evaluation of intron function and consultation of earlier large-scale studies remain essential when designing synthetic chromosomes to avoid unintended effects on gene expression and cellular fitness.

Spontaneous construction-related defects

The process of constructing synthetic chromosomes also introduced unforeseen variation. The complexity of multistep integration and propagation occasionally gave rise to spontaneous mutations and structural anomalies. These events emerged independently of intentional edits, reflecting the inherent plasticity and occasional instability of the yeast genome under intensive engineering.

Patchwork DNA. As synthetic chromosomes were assembled and sequenced across the Sc2.0 consortium, many groups encountered 'patchwork' DNA; unexpected regions of native DNA embedded within synthetic chromosomes (Fig. 1g, Table 1). These mosaic segments often arose from short stretches of wild-type DNA that went undetected during PCRTag-based validation, particularly if they laid entirely between tag sites. To address this, several teams adopted refined assembly approaches to improve fidelity. For example, the megachunk–BAC strategy applied by the *synIV* and *synIX* teams enabled assembly of megachunks as circular plasmids, reducing reliance on multiple homologous overlaps. The *synVIII* team circumvented this issue by sequence-verifying chunks, minimizing integration-induced errors. These adaptations helped to reduce patchwork DNA and improved the overall quality and consistency of synthetic chromosome assemblies, and they should be noted for future synthetic genome designs.

Missing stop codons and loxPsym sites. Over half of the synthetic chromosomes assembled during the Sc2.0 project were initially found to be missing key design features, such as reassigned stop codons and loxPsym sites (Fig. 1h, Table 1). These problems may have arisen from patchwork DNA or from biological selection pressures, whereby cells preferentially retained or reverted to wild-type sequences during integration. In each case, the missing synthetic elements were corrected over successive assembly iterations until a fully synthetic chromosome was achieved. These findings underscore the importance of sequence verification in synthetic genome construction to ensure that all intended design features are incorporated.

Megachunk duplications. Tandem duplications and structural amplifications were identified across multiple synthetic chromosomes during the Sc2.0 project (Fig. 1i, Table 1). These events were mostly detected through whole-genome sequencing and pulsed-field gel electrophoresis (PFGE), often revealing unintended duplications of individual chunks or entire megachunks. In most cases, duplications were attributed to recombination between loxPsym sites or cohesive-end termini, particularly during SwAP-In. Repair strategies varied case by case but typically involved CRISPR–Cas9-mediated editing or donor-directed recombination, followed by PCR screening and sequencing to confirm resolution. Though often phenotypically silent, these rearrangements focus on the need for precise integration control and post-assembly validation when constructing synthetic genomes.

Retention of vector backbones. Throughout Sc2.0, several synthetic chromosomes were found to have retained vector backbone sequences, often carried over during minichunk or megachunk integration (Fig. 1j, Table 1). These extraneous sequences frequently contributed to complex structural variations and assembly defects that required repair.

In *synII*, a structural variant in megachunk T included more than 30 kb of complex DNA, comprising multiple copies of chunks T4 and T5 alongside a partial sequence from the pSBGAK vector backbone. *SynXI* also showed backbone contamination, whereby repeats of chunk J1 were interspersed with vector-derived sequences. Long-read sequencing revealed foreign DNA elements, which were previously missed in short-read assemblies, exposing the addition of an *E. coli* transposase sequence, presumably integrated during plasmid propagation in bacterial hosts. In *synXVI*, genome sequencing uncovered a range of assembly anomalies, including pUC vector insertions. These defects were later resolved through CRISPR-directed gap repair. In *synXII*, a retained vector sequence in chunk P was also identified and removed, although the specific strategy used for correction was not reported.

With the advancement of sequencing technologies, the adoption of long-read platforms has proven critical for detecting hidden backbone insertions and structural anomalies that elude short-read analysis. Such tools reduce dependence on PCRTag-based validation and allow earlier intervention during strain construction. Moving forward, stringent quality control during plasmid propagation and genome integration will be essential, as undetected vector sequences have the potential to undermine genome integrity and complicate downstream characterization.

Loss of mitochondrial function and petite colonies. A recurring phenotype observed in many strains developed in the Sc2.0 project was a growth defect on glycerol, a nonfermentable carbon source (observed in *synVI*, *synX*, *synXI*, *synXII*, *synXIV* and *synXVI*). This defect, indicative of respiratory dysfunction, was frequently associated with mitochondrial loss of function and the emergence of petite colonies, another indicator of compromised mitochondrial integrity (Fig. 1k). Documented on six separate occasions (Table 1), this phenotype was often traced to

disruptions in nuclear encoded mitochondrial genes or those linked to mitochondrial processes. These disruptions arose through a variety of design- or assembly-related events, including the introduction of PCRTags (*synVI*), REPEATSMASHER (*synXVI*), the insertion of loxPsym sites within 5′ UTRs and 3′ UTRs (*synX*, *synXI*, *synXIV* and *synXVI*), removal of tRNAs (*synX*) and errors during megachunk integration (*synXII*).

The consequences of mitochondrial loss of function are problematic, as yeast cells cannot regenerate mitochondrial genomes de novo. Even after the causative genes were identified and repaired, mitochondrial function could not be restored without additional intervention, such as replacing the damaged mitochondrial DNA. Several groups used CRISPR–Cas9-mediated genome editing and backcrossing strategies to reintroduce the synthetic chromosome into a wild-type background with an intact mitochondrial genome. The repeated isolation of strains with respiratory defects underscores a critical consideration in synthetic genome design: maintaining mitochondrial function during assembly is essential, particularly in organisms that lack yeast's amenability to mitochondrial restoration through transformation or backcrossing, such as plant or mammalian cells. Continuous evaluation of mitochondrial function during assembly is therefore crucial, with growth on nonfermentable carbon sources at 37 °C serving as a stringent and comprehensive assay of strain fitness.

Karyotype instability and genome duplications. Throughout the Sc2.0 project, several synthetic strains showed unexpected karyotypic abnormalities (*synII* and *synXIII*), including chromosomal rearrangements (*synIII*), disomy (*synIX*) and diploidy (*synXI*). Although these anomalies were often unrelated to the synthetic chromosome, they introduced challenges in strain validation and necessitated additional genetic interventions.

In *synII* and *synIII*, abnormal karyotyping involving non-target chromosomes were detected via PFGE and genome sequencing, probably having arisen spontaneously during assembly. These anomalies were corrected through backcrossing strategies to restore the expected genome architecture. In *synIX*, a case of whole-chromosome disomy was identified, with the synthetic chromosome itself being duplicated (Fig. 1l). It was hypothesized that the duplication was a result of CRISPR-induced editing instead of selection for a second copy of *synIX* owing to haploinsufficiency, and it was resolved using chromosome destabilization techniques.

Similarly, *synXI* strains were discovered to be homozygous diploids that had remained undetected by sequencing, requiring sporulation and tetrad dissection to recover haploid isolates. Homozygous diploidy was also observed in a cell housing the tRNA neochromosome. In *synXIII*, aberrations on chromosomes V and VIII were also resolved through backcrossing.

These observations emphasize the value of routine karyotyping and phenotypic validation during synthetic genome construction. Even when synthetic sequences are correctly assembled, genome integrity can be compromised by secondary chromosomal events, reinforcing the need for comprehensive quality control at each stage of strain development.

Developing novel approaches to debug synthetic chromosomes

A central goal of the Sc2.0 project was that each synthetic chromosome should function indistinguishably from its wild-type counterpart, with no significant growth or phenotypic defects. Achieving this required an iterative debugging process throughout chromosome assembly. As megachunks were sequentially integrated, many teams monitored for growth abnormalities or other phenotypic changes after each step. Although this stepwise phenotypic screening was the standard approach, several teams implemented more systematic debugging methodologies. These ranged from established genetic techniques such as backcrossing and transcriptomic analysis via RNA-seq to the development of novel high-throughput tools specifically designed for scalable synthetic genome debugging.

One strategy developed by the *synX* team was pooled PCRTag mapping (PoPM), a high-throughput method for linking genotype to phenotype after each megachunk integration (Fig. 2a). PoPM uses a simple PCR-based assay to distinguish synthetic DNA from native sequences, allowing researchers to quickly identify which segments were associated with fitness defects. By pooling colonies into phenotypic groups such as normal versus slow-growing and comparing the presence or absence of specific DNA tags, PoPM localized bugs to small genomic regions for analysis and correction. This approach accelerated the correction of design errors and was scalable in complex synthetic strains, including multi-synthetic or watermarked chromosomes, and adaptable to both mitotic and meiotic mapping contexts. As synthetic genomics progresses and the scale of projects increases, validation tools such as PoPM are vital for ensuring accuracy, functionality and design feedback in genome-scale engineering.

The methods used to correct defects also evolved over the course of the project. Every synthetic chromosome reported included synonymous and nonsynonymous single-nucleotide polymorphisms as well as frameshift mutations. Each mutation was either neutral and left in place or caused observable defects and required repair. Early in the project, cases that required correction relied on classical counterselection approaches such as 5-fluoroorotic acid selection and the *URA3* marker to correct the synthetic sequence. Teams later adopted additional genome editing systems such as Cas9 to achieve more precise replacements or targeted corrections in the construction of *synXI*, *synXIV* and *synXVI*.

Another key tool developed during the project was CRISPR D-BUGS, which combines targeted CRISPR-based genome editing with selectable markers to rapidly identify and fix problematic synthetic regions[8] (Fig. 2b). This approach allowed teams to troubleshoot and resolve assembly errors more precisely and efficiently than through phenotype screening alone, and its flexibility makes it broadly useful for synthetic genome engineering beyond yeast.

In addition, CRIMiRE, incorporating a design–build–test–learn cycle, was particularly effective for diagnosing subtle or cumulative defects (Fig. 2c). For example, when *synXV* showed slow growth and poor sporulation, CRIMiRE was used to identify and correct a single faulty PCRTag within the *IRA2* gene, restoring normal function in the updated *synXV* strain.

Although traditional phenotype screening remains valuable, techniques such as PoPM, CRISPR D-BUGS, and CRIMiRE offer scalable, targeted approaches to debugging at genome scale. Importantly, they not only ensured the success of Sc2.0 but also lay the groundwork for debugging increasingly complex synthetic genomes in future projects.

From yeast to multicellular life

The completion of the Sc2.0 project establishes a platform for expanded genome engineering and functional studies. From the shift between assembly to application, the power of a fully synthetic genome will lie not only in its construction but also in its capacity to be iteratively reconfigured, redesigned and reimagined. With a complete genome

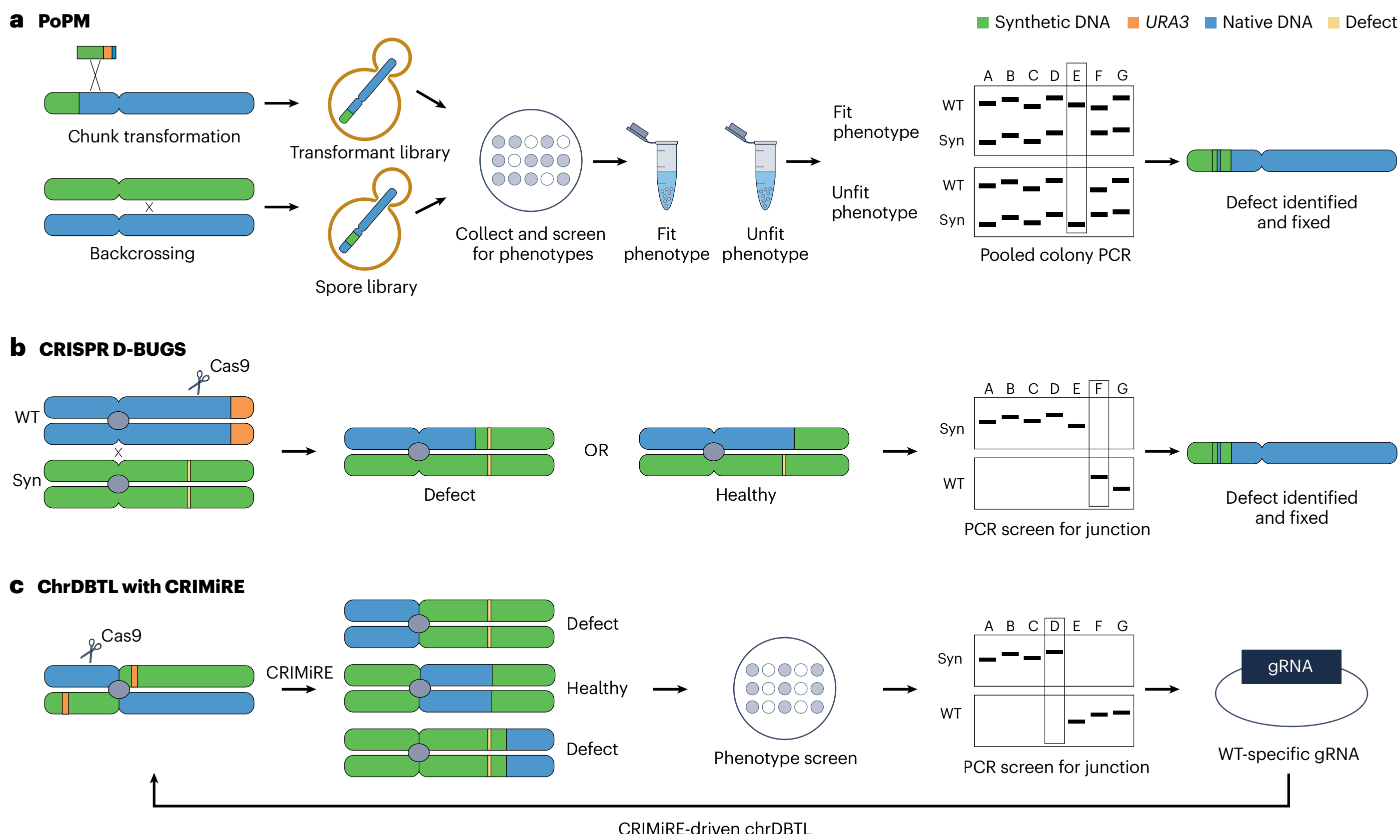

Fig. 2 | Schematic overview of novel debugging approaches developed during Sc2.0. a, Pooled PCRTag mapping (PoPM), a high-throughput strategy that links genotype to phenotype by screening pooled colonies and localizing defects to specific genomic regions. **b**, CRISPR D-BUGS, a targeted editing approach that combines CRISPR–Cas9 with selectable markers to identify and correct problematic sequences using heterozygous diploid strains and mitotic recombination triggered by CRISPR cutting. **c**, CRIMiRE, a CRISPR–Cas9-mediated recombination system integrated with a design–build–test–learn (DBTL) cycle, enabling iterative refinement of synthetic chromosomes through systematically isolation and repair of subtle or cumulative defects.

and thousands of nonessential genes retained, the Sc2.0 project has created a resource for exploring gene function and testing design principles at scale. The incorporation of features including loxPsym sites and PCRTags has enabled scalable rearrangement via SCRaMbLE, allowing researchers to reverse-engineer genome function by linking genotype to phenotype in real time. This resource is already being used to work backward to define essential gene networks, as demonstrated in recent efforts to map fitness landscapes across SCRaMbLEd libraries[9].

Concepts from Sc2.0 have begun to inspire synthetic genome initiatives in organisms with distinct evolutionary histories and biological architectures, including *E. coli*[10], the moss *Physcomitrium patens*[11], algae[12,13] and mammalian systems[14].

As synthetic genome design matures, efforts such as Yeast 3.0 propose to not simply write the genome but reorganize it through minimization and functional streamlining[9,15,16]. One proposed trajectory envisions a modular genome architecture in which genes are grouped by function into discrete chromosomal units[12]. Such an arrangement would enable plug-and-play engineering of metabolic and regulatory pathways, or neochromosomes, supporting rapid reconfiguration for diverse applications. In this context, the conceptual 'Yeast 4.0' framework envisions an industry-ready platform: a robust, customizable chassis optimized for biomanufacturing, synthetic circuit testing or the production of high-value compounds.

An important aspect of this vision is the deliberate design of neochromosomes tailored to specific purposes, carrying entire pathways or functional modules on a single synthetic element[17]. For example, a dedicated neochromosome could be built to host a complete biosynthetic pathway, enabling rapid iteration and optimization without perturbing core genomic regions. Functionally organized modules could also be leveraged for discovery; assembling all cell wall-related genes into a single synthetic cassette would allow targeted studies of *Fusarium* pathogenesis within a yeast context, accelerating insights into plant–pathogen interactions. Similarly, modular synthetic chromosomes may enable orthogonal system development, using *S. cerevisiae* as a stable, genetically tractable vessel to host and iterate on non-native systems, such as engineered plant plastomes or synthetic chloroplast genomes.

Although *S. cerevisiae* has shown considerable plasticity, tolerating chromosome fusions, splits and even circularized chromosomes, questions remain around the long-term stability of engineered karyotypes, centromere compatibility and the limits of genomic reorganization under industrial conditions. One emerging constraint for highly modular genome architectures is the biological burden associated with centromere number and compatibility. Although *S. cerevisiae* can tolerate some degree of karyotype engineering, the introduction of multiple synthetic chromosomes, each carrying the

same centromere, can lead to instability, aneuploidy or segregation defects. This presents a barrier to scaling chromosome number beyond a certain threshold, particularly in designs that rely on small, functionally grouped modules. Additionally, high centromere copy number may complicate cell cycle regulation and increase metabolic load, limiting the efficiency and scalability of synthetic genome production. Future design strategies will need to consider centromere redundancy, suppression or alternative segregation mechanisms to enable the creation of large chromosome complements without compromising genomic stability.

Meeting these challenges will require continued innovation in genome engineering technologies. The toolkit has expanded since the inception of Sc2.0, from CRISPR-based debugging systems to long-read sequencing, and precision editing tools such as REXER[18], BASIS[19] and CRMAGE[20] now enable more subtle construction and iterative refinement. Advances in DNA synthesis from longer, high-fidelity fragments to novel delivery methods are beginning to close the gap between in silico design and in vivo implementation. Projects such as GP-write[21] signal a global, coordinated push to scale synthetic genomics to the level of mammalian chromosomes and entire functional systems.

As we domesticate yeast further, optimizing it for rapid growth on minimal medium, enhancing its tolerance for industrial conditions or layering in biosynthetic circuits, the synthetic genome becomes not just a scientific tool but a foundation for programmable, evolvable cell factories.

Paige E. Erpf[1], Felix Meier[1], Roy S. K. Walker[1], Hugh D. Goold[1,2], Jef D. Boeke[3,4,5], Ian T. Paulsen[1,6] ✉ & Isak S. Pretorius[1] ✉

[1]School of Natural Sciences and ARC Centre of Excellence in Synthetic Biology, Macquarie University, Sydney, New South Wales, Australia. [2]New South Wales Department of Primary Industries and Regional Development, Elizabeth Macarthur Agriculture Institute, Advanced Gene Technology Centre, Menangle, New South Wales, Australia. [3]Institute for Systems Genetics, NYU Langone Health, New York, NY, USA. [4]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA. [5]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY, USA. [6]The Australian Genome Foundry, Sydney, New South Wales, Australia.
✉e-mail: ian.paulsen@mq.edu.au; sakkie.pretorius@mq.edu.au

Published online: 11 December 2025

References

1. Richardson, S. M. et al. *Science* **355**, 1040–1044 (2017).
2. Dymond, J. S. et al. *Nature* **477**, 471–476 (2011).
3. Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D. & Bader, J. S. *Nucleic Acids Res.* **38**, 2603–2606 (2010).
4. Admire, A. et al. *Genes Dev.* **20**, 159–173 (2006).
5. Ji, H. et al. *Cell* **73**, 1007–1018 (1993).
6. Parenteau, J. et al. *Mol. Biol. Cell* **19**, 1932–1941 (2008).
7. Parenteau, J. et al. *Cell* **147**, 320–331 (2011).
8. Zhao, Y. et al. *Cell* **186**, 5220–5236 (2023).
9. Xu, X. et al. *Nat. Commun.* **14**, 1984 (2023).
10. Fredens, J. et al. *Nature* **569**, 514–518 (2019).
11. Chen, L. G. et al. *Nat. Plants* **10**, 228–239 (2024).
12. Goold, H. D., Moseley, J. L. & Lauersen, K. J. *Cell Genomics* **4**, 100505 (2024).
13. Karas, B. J. et al. *J. Biol. Eng.* **7**, 30 (2013).
14. Gibson, D. G., Smith, H. O., Hutchison, C. A. III, Venter, J. C. & Merryman, C. *Nat. Methods* **7**, 901–903 (2010).
15. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. *Genome Biol.* **21**, 205 (2020).
16. Schindler, D. *Bioeng.* **7**, 137 (2020).
17. Schindler, D., Walker, R. S. K. & Cai, Y. *Cell Rep. Methods* **4**, 100761 (2024).
18. Robertson, W. E. et al. *Nat. Protoc.* **16**, 2345–2380 (2024).
19. Zürcher, J. F. et al. *Nature* **619**, 555–562 (2023).
20. Ronda, C., Pedersen, L. E., Sommer, M. O. & Nielsen, A. T. *Sci. Rep.* **6**, 19452 (2016).
21. Boeke, J. D. et al. *Science* **353**, 126–127 (2016).

Acknowledgements

Our Synthetic Yeast Genome Project (Sc2.0) work was supported by the Australian Research Council Centre of Excellence in Synthetic Biology and by external grants from Bioplatforms Australia, the New South Wales (NSW) Chief Scientist and Engineer, and the NSW Government's Department of Primary Industries. Figures and tables were edited and refined by Bronte Turner of Serpentine Studio.

Author contributions

P.E.E., H.D.G., R.S.K.W., I.T.P. and I.S.P. conceived, developed and edited the manuscript, and K.M. and J.D.B. contributed to sections of the article and significantly helped with the refinement of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-025-02913-4.