Construction of a synthetic *Saccharomyces cerevisiae* pan-genome neo-chromosome

Dariusz R. Kutyna1, Cristobal A. Onetto1, Thomas C. Williams2, Hugh D. Goold2,3, Ian T. Paulsen2, Isak S. Pretorius2,4, Daniel L. Johnson1,6 & Anthony R. Borneman1,5

The Synthetic Yeast Genome Project (Sc2.0) represents the first foray into eukaryotic genome engineering and a framework for designing and building the next generation of industrial microbes. However, the laboratory strain S288c used lacks many of the genes that provide phenotypic diversity to industrial and environmental isolates. To address this shortcoming, we have designed and constructed a neo-chromosome that contains many of these diverse pan-genomic elements and which is compatible with the Sc2.0 design and test framework. The presence of this neo-chromosome provides phenotypic plasticity to the Sc2.0 parent strain, including expanding the range of utilisable carbon sources. We also demonstrate that the induction of programmable structural variation (SCRaMbLE) provides genetic diversity on which further adaptive gains could be selected. The presence of this neo-chromosome within the Sc2.0 backbone may therefore provide the means to adapt synthetic strains to a wider variety of environments, a process which will be vital to transitioning Sc2.0 from the laboratory into industrial applications.

1The Australian Wine Research Institute, PO Box 197 Glen Osmond, SA 5064, Australia. 2ARC Centre of Excellence in Synthetic Biology and Department of Molecular Sciences, Macquarie University, Sydney, NSW 2019, Australia. 3New South Wales Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Woodbridge Road, Menangle, NSW 2568, Australia. 4The Chancellery, Macquarie University, Sydney, NSW 2109, Australia. 5School of Wine, Food and Agriculture, The University of Adelaide, Adelaide, SA 5005, Australia. 6Present address: The Chancellery, Macquarie University, Sydney, NSW 2109, Australia. ◄Email: anthony.borneman@awri.com.au

https://doi.org/10.1038/s41467-022-31305-4
By way of its ease of propagation and well-defined genetics, the yeast *Saccharomyces cerevisiae* represents one of the most intensively studied eukaryotic model organisms and was the first for which a fully characterized genome sequence was available. The International Synthetic Yeast Genome Project (Sc2.0) has now also positioned *S. cerevisiae* at the forefront of genome engineering, with this species being the first eukaryote to be synthetically engineered at the whole-chromosome scale.

Most studies regarding the biology of *S. cerevisiae*, including the initial genome sequencing and Sc2.0 efforts, have focused on the laboratory strain S288c (or derivatives thereof). However, there are hundreds of diverse strains of *S. cerevisiae* and many display distinctive phenotypes that provide selective advantage within specific environmental niches or industries. E.g. fermenting wine, leavening bread or brewing beer. These phenotypic differences are the direct result of intraspecific genetic variation, often in the form of strain-specific genes or gene clusters. The differential presence of these genes between strains can impart striking phenotypic consequences, including the ability to synthesize vitamins or to survive specific environmental stresses or inhibitory compounds. The common theme across these strain comparisons relative to S288c, is that this laboratory strain appears to represent an almost minimal core of common genes, displaying few open reading frames (ORFs) that are absent in most other strains (except for a large number of transposon integrations), which likely reflects genetic streamlining afforded by selection under ideal laboratory growth conditions. Studies that focus solely on this strain do not therefore consider these pan-genomic ORFs and their phenotypic impacts.

To address this missing genetic variation and provide the potential for additional phenotypic plasticity in the Sc2.0 parental strain, we have sought to assemble an array of pan-genomic elements, normally associated with industrial or environmental isolates of *Saccharomyces cerevisiae* into a seventeenth chromosome for inclusion within the Sc2.0 background.

### Results and discussion

**Design and de novo synthesis of a pan-genome neo-chromosome.** As building blocks for this pan-genome neo-chromosome (PGNC), seventeen unique pan-genome sequences (1.1–60.3 kb), were identified from across whole-genome sequences of more than 200 diverse strains of *S. cerevisiae*. The final collection comprised a non-degenerate set of sequences from eight wine, sake, biofuel, human pathogen and natural isolates (Supplemental Dataset 1). These fragments were concatenated in silico into a single DNA molecule, to which global systematic changes were introduced in accordance with the Sc2.0 project. This included the substitution of TAA for TAG stop-codons, the introduction of oligonucleotide watermarks within 36 ORFs (Supplementary Table 1, Supplementary Fig. 1) and 63 bi-directional Cre-recombinase recognition (loxPsym) sites. The total length of the final synthetic PGNC was 211,409 bp and contained 75 predicted ORFs (Fig. 1a). The sequences of each ORF, along with strain origins and functional annotation are provided in Supplementary Dataset 1, with a full annotated sequence of the PGNC presented in Supplementary Dataset 2.

To allow for DNA synthesis, the final PGNC design was divided into 21 fragments (chunks), of ~10 kb in length (Supplementary Fig. 1). Each chunk was flanked by 200 bp of overhanging sequences at both termini, which were designed to allow for in vivo assembly in *S. cerevisiae* via homologous recombination. Two auxotrophic marker genes, *URA3* and *LEU2*, were also synthesized with specific flanking sequences, allowing for them to be alternatively integrated during the progressive steps of the assembly (Fig. 1b).

A yeast centromeric vector (p416-natR) was used as the backbone for the neo-chromosome assembly, which provided a functional centromere fused to an autonomously replicating sequence and a nourseothricin (clonNAT) resistance marker (Fig. 1b). Assembly was initiated with linearized p416-natR, two pan-genome chunks (Fig_01, Fig_02) and the URA3 marker. For the second round of assembly three chunks (Fig_03, Fig_04, Fig_05) were introduced into a strain containing the completed first round assembly, along with the alternative auxotrophic marker (*LEU2*). Eight rounds of assembly were ultimately conducted, with between one and four pan-genome chunks incorporated per cycle. A ninth round was then used to replace the remaining *LEU2* marker with a BFP-expression cassette, producing the final, circular neo-chromosome (PGNCcirc). The integrity of the PGNCcirc strain was then confirmed by both PCR across each inter-chunk junction (Fig. 1c, Supplementary Fig. 2) and by whole-genome sequencing. *S. cerevisiae* has been shown previously to be able to host large heterologous episomes, such as whole bacterial genomes, in yeast-bacterial shuttle vectors. Circular variants of native *S. cerevisiae* chromosomes have also been engineered as part of the Sc2.0 consortium, where they behave normally, except during meiosis.

To compare the behaviour of circular and linear chromosomal variants of the PGNC, linearized versions were engineered using the telomerase at three different loci within PGNCcirc (Fig. 1d, Supplementary Fig. 2). This resulted in three linear chromosomal variants (PGNClin1, PGNClin2 and PGNClin3), which differed only in the arrangement of genes relative to the newly introduced telomeric sequences. Growth curves were performed to assess the effect of these PGNC chromosomal variants on overall strain fitness in rich media (Fig. 1e). While PGNCcirc displayed a growth curve that was comparable to the wildtype strain, the linear variants all displayed slightly extended lag periods and reduced total cell densities and maximum specific growth rates (WT, 0.53 h⁻¹; PGNCcirc, 0.53 h⁻¹; PGNClin1, 0.51 h⁻¹; PGNClin2, 0.5 h⁻¹ and PGNClin3, 0.46 h⁻¹). All the PGNC variants displayed a lower final optical density than the parental strain, indicating that PGNC elements were impacting strain fitness under standard laboratory conditions.

**PGNC stability.** PGNCcirc is only 20 kb smaller than the native chromosome I of *S. cerevisiae* (smallest native chromosome) and growth curves suggested that it may impart a selective disadvantage to the PGNC-carrying strains in the absence of clonNAT-induced selection. To assess the mitotic stability of the circular and linear variants of the PGNC, representative strains were serially passaged under non-selective (media without clonNAT) conditions. Isolates from each population were assessed for the presence of the PGNC after 25 and 50 generations (Fig. 2, Supplementary Figs. 4 and 5). PGNCcirc displayed the highest stability, averaging 61.7 ± 4.5% and 40 ± 6.9% retention over 25 and 50 generations, respectively. Of the linear versions, PGNClin1 displayed the highest retention (25 gen, 54.0 ± 1.0%; 50 gen, 20.0 ± 2.6%), while PGNClin2 was very unstable, with only 21.3 ± 2.1% (25) and 4.0 ± 2.6% (50 generations).

To attempt to address the stability issues of the PGNC, additional ARS sequences were inserted into the PGNC at two different sites. Two variants were made to the PGNCcirc, whereby either a single copy of ARS305 (efficient, early firing origin from chromosome III) was inserted between ORF58 and ORF59 (ARS_305_01), or a dual-variant with a second ARS305 inserted between ORF19 and ORF20 (ARS_305_02) (Supplementary Fig. 1). The inclusion of the additional ARS sequences were shown to modestly improve the stability of the PGNCcirc element, although not to levels that would preclude the use of selective...
media for the long-term stability of the PGNC (Fig. 2d). Complicating these results, the dual-variant did not provide increased stability relative to the ARS_305_01 alone, suggesting that there is a multi-factored interplay between total ARS number and overall stability.

In future studies, improvement in the stability of the PGNC element could be investigated through screening of additional combinations of positions of ARS elements or through the addition of an essential gene or fusion to another chromosome, as reported for synI22, to drive the maintenance of this element without requiring drug-based selection.

PGNC imparts distinct phenotypes in the Sc2.0 strain background. Given the coding potential of the PGNC, in addition to existing reports of phenotypic outcomes of some of the genes known to be present within the neo-chromosome12,23, the phenotypic consequence of the presence of PGNC in BY4742 (WT) strain. Data are presented as mean values +/− SD, based upon values obtained from three independent biological replicates. Source data are provided as a Source Data file.

**Fig. 1 Construction of a yeast pan-genome neo-chromosome.** a Schematic representation of the circular PGNC construct (PGNC\textsuperscript{circ}). b Iterative assembly of the PGNC via homologous recombination and using alternating selectable markers. c Confirmation of the PGNC assembly using PCR reactions that span each inter-chunk junction. Identical PCR results were obtained across at least two repetitions and during sub-assembly d Production of three linear PGNC variants using the telomerator (PGNC\textsuperscript{lin1}, PGNC\textsuperscript{lin2} and PGNC\textsuperscript{lin3}). Yellow boxes define the position of the centromeric-plasmid backbone. Red boxes denote positions of synthetic telomeric repeat sequences added by the telomerator process. e Microplate growth kinetics of the four PGNC variants (PGNC\textsuperscript{circ}, PGNC\textsuperscript{lin1}, PGNC\textsuperscript{lin2} and PGNC\textsuperscript{lin3}) compared to the BY4742 (WT) strain. Data are presented as mean values +/− SD, based upon values obtained from three independent biological replicates. Source data are provided as a Source Data file.
strain, with melibiose, palatinose and butyric acid being utilized only in the presence of PGNC (Fig. 3a). As the utilization of the carbon sources, palatinose and melibiose were expected to be due to the presence of specific glycosidases (with potential candidate enzymes annotated in the PGNC), these compounds were selected for additional, larger-scale fermentation to confirm the BioLog results and map the regions of the PGNC that were responsible.

Palatinose (Isomaltulose) is a disaccharide composed of glucose and fructose linked via an alpha-1,6-glycosidic bond. While *S. cerevisiae* BY4742 encodes several isomaltases (with potential candidate enzymes annotated in the PGNC), this strain shows very slow utilization of palatinose (Fig. 3b). In contrast, PGNC strains show efficient palatinose utilization, reaching a stationary phase within 24 h (Fig. 3b). To define the pan-genome ORF(s) that might be responsible for this phenotype, intermediate PGNC strains (produced during the stepwise assembly process) were tested for the palatinose-utilization phenotype (Fig. 3c). While strains containing the chunks Frag_01 through Frag_14 displayed a non-utilizing phenotype, intermediate strains containing Frag_15 to Frag_17 displayed a phenotype that was indistinguishable from the full PGNC. Functional annotation of ORFs within Frag_15 provided a candidate cluster of three ORFs, predicted to encode an alpha-1,6-glycosidase family enzyme (ORF48), a putative zinc-finger transcription factor (ORF49) and a sugar transporter (ORF50) (Supplementary Fig. 1). The presence of this cluster of ORFs alone (ORF48–50) was shown to also provide robust growth on palatinose (Fig. 3b). When the function of this cluster was further refined by the expression of individual ORFs, ORF49 was shown to provide the same levels of growth as the three ORF clusters. The putative transcription factor encoded by ORF49 is therefore responsible for the palatinose-utilisation phenotype and the ability for this ORF to stimulate utilisation of palatinose in the S288c background, presumably occurs through activation of MAL-family glucosidases that are present in the S288c genome.

Melibiose is a disaccharide composed of galactose and glucose, which are linked via an alpha-1,6-glycosidic bond. Unlike palatinose, *S. cerevisiae* BY4742 is unable to utilize melibiose as a sole carbon source, even under extended periods of growth, while all of the PGNC variants displayed robust growth on this sugar (Fig. 3d). Analysis of the PGNC intermediates located the region responsible for this phenotype within Frag_06 to Frag_08, which contained 11 predicted ORFs (Fig. 3e). From this group, ORF21, predicted to encode an alpha-galactosidase, was the clear candidate for this phenotype. Expression of ORF21 in isolation was subsequently shown to be sufficient for the melibiose-utilization phenotype (Fig. 3d).

Fig. 2 Mitotic stability of the PGNC in the absence of selection. a Populations were established for each four PGNC variants (PGNC<sub>circ</sub>, PGNC<sub>lin1</sub>, PGNC<sub>lin2</sub> and PGNC<sub>lin3</sub>) and serially passaged eight times for ~50 generations of total growth. b Differential plating of 96 colonies for phenotypic assessment of loss of the PGNC element. c Results of passaging experiments (96 colonies per replicate) for each of the PGNC strains after both 25 and 50 generations. d Results of passaging experiments (96 colonies per replicate) for each of the PGNC strains containing additional ARS sequences after both 25 and 50 generations. All graphs are presented as mean values ±/− SD, based upon values obtained from three independent biological replicate passaging experiments. Source data are provided as a Source Data file.
utilizing phenotype and the over-expression strain (ORF21FBA1p) displayed significant enhancement in its utilization of melibiose (Fig. 3d), confirming the role of this ORF.

**SCRaMbLE-induced phenotypic diversity.** One of the key attributes of the Sc2.0 design is the ability to stimulate genetic diversity through recombination-mediated rearrangement of the loxPsym sites (termed SCRaMbLE) that were inserted throughout the Sc2.0 synthetic chromosomes.29 As the PGNC design included 63 loxPsym sites, the effect of SCRaMbLE on the structure of the PGNC was investigated. Given the clear melibiose-utilization phenotype provided by the PGNC, combined with the evidence for further improvement in growth (provided by the ORF21FBA1p results), this phenotype was chosen as a test for SCRaMbLE-induced adaptive improvement. As the PGNC circ element displayed growth kinetics closest to the parental strains (Fig. 1e), the strain containing this element was chosen as the basis for the adaptive experiments. SCRaMbLE was performed on the strain containing the PGNC circ element, with the resulting mixed population subjected to competitive growth on melibiose (Fig. 4a). After serial passaging, nine single colonies from the Cre-expressing population (Cre+), and three colonies from the control population (Cre-) were assessed for growth of melibiose relative to the PGNC circ and ORF21FBA1p strains (Fig. 4b). Of the twelve isolates, all nine from the Cre+ population displayed
growth rates on melibiose that were significantly improved relative to PGNC\textsuperscript{circ}, although none were able to match the very high isolates from the Cre-expressing populations; green, isolates from the control populations). Source data are provided as a Source Data file.

At this stage, it is not known how these combined structural variants influenced the ability of these strains to utilise melibiose, especially given that strains contained multiple individual mutations that may be synergistic, additive or neutral. While it is relatively straightforward to reconcile increased PGNC copy-number with increased expression of ORF21 (and melibiose utilisation), hypotheses pertaining to other variants are harder to postulate, although the alterations to the 3’ untranslated region of ORF21 could suggest that altered transcript stability may be responsible for some increases that were observed. The processes underpinning the increased growth rate of the two SCRaMbLEd strains with no detectable structural variation or increased copy number (isolates 4 and 10 in Fig. 4) may be due to alterations in the interplay between the natural genome and the PGNC or to smaller-scale mutations such as single-nucleotide polymorphisms.

In summary, the Sc2.0 chassis provides a framework for engineering the next era of industrial microbes. The ability to introduce neo-chromosomes, such as the PGNC, has been demonstrated to greatly expand the genetic and phenotypic diversity that can be achieved within the Sc2.0 background. This provides the means to adapt this, and other, synthetic strains to a variety of environments, a process which will be vital to transitioning Sc2.0 from the laboratory into industrial applications.

**Methods**

**PGNC design.** A total of 17 unique pan-genome sequences (1.1–60.3 kb), were selected from whole-genome sequences of more than 200 diverse strains of S. cerevisiae\textsuperscript{a}. These fragments were concatenated in silico in descending size order into a single DNA molecule, to which global systematic changes were introduced in accordance with the Sc2.0 project\textsuperscript{b}. In short, these changes included: the substitution of TAA stop-codons for TAG stop codons, the introduction of oligonucleotide watermarks in 36 ORFs using the principles of codon redundancy (Supplementary Table 1) and the introduction of 63 bi-directional Cre-
recombinase recognition sequences (loxP/ym), located 3 bp after the stop codon of selected high-confidence ORFs.

Functional annotation for the 75 predicted ORFs (Supplemental Dataset 2) was performed using the Interproscan 5 pipeline (v. 5.32–660). Annotation of Carbohydrate-active enzyme (CAZYme) classes was performed through an HMImerc v.3.3.2.21 of the dbCAN HMIMdb v.9.0 database.22 KEGG orthology assignments were obtained using BlastKOALA v.2.2.23 and prediction of signal peptides was performed using SignalP v.4.1.24

For in vivo assembly, the PGNC was divided into 22 fragments (chunks) of ~10 kb in length (Supplementary Fig. 1). Each chunk was flanked with Pmel and/or NotI restriction sites to allow for release from the plasmid vector backbone (pUG57), in addition to 200 bp of overhanging sequences at the 5′- and 3′-ends, which were homologous to their neighbouring fragments. Two auxotrophic markers, URA3 and LEU2, were also designed with specific flanking sequences, allowing them to be alternatively integrated during the processive steps of the assembly.

PGNC synthesis and assembly. All 21 chunks and the selectable markers cassettes were synthesized and cloned into a plasmid vector (GenScript). For the neo-chromosome assembly, the yeast centromeric vector p416-natR created by replacing URA3 auxotrophic marker with clonNAT resistance marker in p416-GFP vector38 (Supplemental Fig. 1). Assembly was initiated with linearization of the p416-natR vector with CaiI endonuclease and release of Frag 01, Frag 02 and URA3 from pUG57 using Pmel and NotI. Fragments were pooled in even ratios and transformed into S. cerevisiae BY4742 (Supplementary Table 1). After transformation, colonies were selected on a solid yeast nitrogen base (YNB) medium lacking uracil and supplemented with 100 µg/mL of clonNAT and Ura- and NatR-resistant colonies were incubated at 30°C for 2 days. Colonies were confirmed using PCR, with confirmation primers designed to amplify across the junctions between each pair of adjacent chunks (Supplementary Table 3).

For the second round of assembly, the chunks Frag 03, Frag 04, Frag 05 were introduced into a strain containing the confirmed first-round assembly product, along with the alternative auxotrophic marker (LEU2). In total, 7 rounds of assembly were conducted, with a varying number of assembled synthetic pan-genome DNA molecules (1–4 per assembly cycle), while alternating the auxotrophic markers. The complete set of diagnostic PCRs, utilizing the primer sequences directly up- and down-stream of the inserted CORE7 cassette36.

The remaining LEU2 auxotrophic marker, which was present in the sequence of PGNC after the last round of the assembly, was removed using selection/counter selection approach using the CORE7 cassette36.

For the selection step, the CORE7 cassette was amplified from the plasmid using primers equipped with 50 bp flanking regions that were homologous to sequences directly flanking the LEU2 gene (Supplementary Table 4). The CORE7 cassette was transformed into yeast and selected on solid YPD medium supplemented with 100 µg/mL of Hygromycin B (Sigma-Aldrich). Transformant colonies were tested for successful CORE7 cassette insertion using PCR with the primers FR21-F and FR21-V-R (Supplemental Table 3). A single transformant that displayed the expected PCR pattern was then used for the counter-selection step. Here, the FR21-F/FR21-V-R PCR product was used to replace each mutated locus with functional sequences that were PCR amplified using genomic DNA of a prototrophic strain. HIS3, LEU2 and LYS2 PCR products were pooled in equal amounts and transformed into BY4742 containing PGNC6 cassette element. Transformants were selected on solid YNB-clonNAT medium lacking histidine, leucine and lysine. Transformant colonies were tested for correct genomic integration using PCR (Supplementary Table 5). The URA3 auxotrophic mutation was not addressed at this stage as URA3 auxotrophy was needed as a marker for the introduction of the telomerator(see below).

Construction of a SceI expression vector. A SceI expression vector (pTL85-SceI), was constructed from two yeast shuttle vectors, pTL85-ARSA3 (Dr. Tiziana Lodi, University of Parma, Italy) and pUDCD37,39 (Euroscarf). Both vectors were digested with PsvII, in case of pTL85, this resulted in the isolation of the plasmid backbone, which carried the kanamycin resistance cassette (KanMX), and in the case of pUDCD37, a partial PsvII digestion resulted in the isolation of the GAL1::SceI::CycC cassette (Supplementary Fig. 9). All restriction fragments were purified from the agarose gel (Wizard SV Gel and PCR Clean-Up System, Promega) prior to ligation (Blunt/TA Ligation Master Mix, New England BioLabs). Ligation reactions were transformed into high efficiency NEB 10 β Competent E. coli (New England BioLabs) following the manufacturer’s instruction and confirmed by restriction digest.

Linearization of the PGNC. The PGNC6 cassette element was linearized using the telomerator20 (Supplementary Fig. 2). The telomerator cassette was synthesized (GeneScript) and PCR amplified from the vector using primers paired with 50 bp flanking regions, homologous to one of three genomic locations within the PGNC (Supplementary Fig. 2a, Supplementary Table 6).

The PGNC strain was transformed with each of the three separate telomerator PCR products, transformed colonies in three distinct locations (PGNC641, PGNC642 and PGNC643) (Supplementary Fig. 2a). Transformed strains were selected on solid YNB - Ura + clonNAT medium. Insertion at the expected location was tested by PCR (Supplementary Table 7). Strains with correct telomerator insertions were then tested for lack of growth on agar plates containing 1 µg/mL of 5-fluoroorotic acid (5-FOA), which selects for loss of the LEU2 marker.

To confirm the purity of the selected telomerator variants, strains were plated out in serial dilutions onto YPD rich medium containing 100 µg/mL clonNAT and incubated for 48 h at 30°C along with the control strain (PGNC6). Isolates of each variant were then transferred onto YNB, YNB - URA, and YNB + 1 mg/mL of 5-FOA, all containing 100 µg/mL clonNAT.

To induce linearization of the telomerator, confirmed strains were transformed with the pTL85-SceI vector. Transformed strains were plated onto solid YNB medium supplemented with 100 µg/mL of clonNAT (selecting for the PGNC) and 200 µg/L G418 (selecting for pTL85-SceI). Transformant colonies were tested for the presence of pTL85-SceI vector using PCR with M13 primers. Three transformants (each carrying the telomerator in distinct location), as well as the control strain carrying only the PGNC were then inoculated into separate YPD cultures supplemented with 100 µg/mL of clonNAT, 200 µg/L G418, and incubated overnight, with shaking, at 30°C. Cells from these cultures were harvested by centrifugation and washed twice with dH2O and incubated (OD600 0.1) into YPGal media to allow for galactose-mediated expression of the S. cerevisiae -1, 1, 1 expression vector (pTL85-SceI), which carried the kanamycin resistance cassette (KanMX), and in the case of pTL85-SceI, this resulted in the isolation of the plasmid backbone, allowing them to be alternatively integrated during the processive steps of the assembly. The linearization process was performed using the Interproscan 5 pipeline v. 5.52

Mitotic stability of the PGNC variants. The mitotic stability of the circular and three linearized variants of the PGNC were tested using replicative colony picking. Strains were grown overnight in YPD supplemented with 100 µg/mL of clonNAT.YPD cultures (100 mL) were then inoculated in triplicate (OD600 0.1) and incubated at 30°C for 24 h before being diluted and passaged into fresh YPD medium (OD600 0.1).Passaging was repeated ten times (5–10 generations of non-selective mitotic divisions). Colonies were scored for growth. Single colonies were plated onto solid YNB -URA, and YNB + clonNAT medium with a PIXL robotic system (Singer Instruments) and incubated at 30°C for ~72 h, with the proportion of clonNAT resistant and sensitive colonies used to infer stability.

Autonomously Replicating Sequence (ARS 365) was introduced to the PGNC using the pCR365 plasmid (ATUM) expression vector. The ARS 365 cassette was linearized by the telomerator using PCR primers specific to each of the three distinct regions where the telomerator was to be introduced (Supplementary Fig. 2, Supplementary Table 7).

Following linearization, the pTL85-SceI vector was removed by growth under non-selective conditions for the plasmid (5–6 generations). To phenotypically test successful linearization, ten colonies of each strain were then plated onto YNB, YNB -URA, and YNB + 5-FOA (all containing 100 µg/mL clonNAT), (Supplementary Fig. 3). After the completion of the linearization process all strains were cured of the URA3Δ0 mutagen using transformation-based method described above.

BioLog phenotyping. Analysis of growth in the presence of an array of different nitrogen sources, carbon sources, and potentially toxic compounds was assessed using a Biolog Phenotype Microarray24 with the growth of PGNC6 compared with BY4742 containing the pFA-TagRFP-T-GDH51 plasmid. Plates PM1 and PM2 were supplemented with dye mix D. Plates PM3B, PM4A, PM5, PM6, PM7, and PM8 were supplemented with 100 mM D-Glucose and dye mix D. Plates PM9, PM10, PM20B, PM21D, PM22D, PM23A, PM24C, and PM25D were supplemented with 100 mM D-Glucose and dye mix E. Plates were incubated at 30°C and sampled every 15 min for 24 h. References of the location on the Microplate to the compounds tested could be found on the manufacturer's website (www.biolog.com/products-platform-overview/phenotype-microarrays-for-microbial-cells). Plates coated with 6x concentrated dye mixes, turbidimeter and media were sourced from BioLog (Hayward, CA 94545 USA).

Biolog raw data were analyzed using the R package open23 to extract values for maximum curve height (A). Data were then exported and log, ratios calculated for pairs of values from the PGNC WT and control plates. Ratios of PGNC-WT maximum curve heights ≥1 were classified as displaying an increased growth rate for that compound. Wells were excluded from analysis if at
least one strain did not exceed the negative control value by at least 50 units or if the positive control for the plate failed to reach 100 units.

**Constructing α-galactosidase expressing vectors.** p416-natR-Gala was constructed utilizing the p416-natR backbone. The Gala gene sequence (ORF with ER-LBD) and promoter, Cre recombinase fused to the oestrogen nuclear receptor alpha ligand-binding domain (ER-LBD) and CYC1 terminator (GAL1p::ER-LBD::CYC1t). Both fragments were gel-purified (Wizard SV Gel and PCR Clean-Up System, Promega) and ligated (Blunt/TA Ligase Master Mix, New England BioLabs). Ligations were transformed into competent *E. coli* cells (NEB 10-beta, New England Biolabs). The control vector, which lacks the GAL1p::ER-LBD::CYC1t element, was created by digesting and purified pTL85 vector backbone.

**SCRaMbLE for population-based improvement of melibiose utilisation.** Yeast cells carrying the circular version PGNC were transformed separately with the pTL85-Cre-EBD and pTL85-Ctrl vectors. Transformed cells were plated on YPD medium containing clotNaT (to select for the presence of the PGNC), and G418 (to select for the presence of the plasmid). After incubation at 30 °C for 48 h, isolated colonies were inoculated into 50 mL Falcon tubes containing 10 mL of medium containing clonNAT (to select for the presence of the PGNC), and G418. Inoculated cultures were then harvested by centrifugation, washed twice in the inoculum medium and then diluted to OD600 ~ 0.025. 200 µL aliquots were added to bottomed microtiter plates. Microtiter plates were sealed using gas-permeable membranes (Breathe-Easy). Incubations were carried out at 30 °C with shaking for 48 h and then passed back into new media (OD600 0.1). This cycle was repeated four times (~ 20 generations).

**Genome sequencing and structural variation analyses of the PGNC.** Yeast DNA was isolated by lysis of protoplasts formed through zymolyase digestion and potassium acetate precipitation 142,143. Sequencing libraries for nanopore whole-genome sequencing were prepared using the Native Barcoding Kit 1D (EXP-NBD104) in combination with the Ligation Sequencing Kit (SQK-LSK109) and loaded into a FLO-MIN106 R9 flow cell. Sequencing was performed using the MinKNOW (v19.10.1) on the MinION platform (Oxford Nanopore Technologies, UK). Fast5 files were base called and demultiplexed using Guppy v3.2.1 (Oxford Nanopore Technologies, UK). Reads with a minimum qscore of 7 were retained for genome assembly using Canu v1.7.12,13, with assemblies polished using Nanopolish v0.11.1.2. Contigs that contained the completely resolved PGNC were located by mapping the p416-natR backbone region to the genome assemblies. For structural variation analyses, the PGNC contig was replaced by the original sequence of the PGNC and reads were mapped back to each genome assembly using Minimap2 v.2.17.74. Structural variants were identified using Sniffles v1.0.11.15 and confirmed by manual inspection. The relative copy number of the PGNC was calculated by mapping reads back to each genome assembly using Minimap2 v.2.17.74 and the ratio between the average coverage of all contigs larger than 200 kb and the PGNC was obtained using CoverM v.0.4.0 (https://github.com/wwood/CoverM).

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

**Data availability** The strains and plasmids generated in this study have been deposited and are available from the Australian Wine Research Institute Culture Collection. The raw DNA sequence data generated in this study have been deposited under BioProject accession code PRJNA615683. All other data generated in this study are provided in the Supplementary Information/Source Data files. Source data are provided with this paper.

Received: 26 August 2021; Accepted: 14 June 2022; Published online: 24 June 2022

**References**

1. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–563 (1996).
2. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
3. Annaluri, N. et al. Total DNA content is a functional determinant of eukaryotic chromosome. *Science* **344**, 55–58 (2014).
4. Warringer, J. et al. Trait variation in yeast is defined by population history. *PLoS Genet.* **7**, e1002111 (2011).
5. Borneman, A. R. et al. Whole-genome comparison reveals novel genetic mechanisms that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* **7**, e1001287 (2011).
6. Peter, J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
7. Novo, M. et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl Acad. Sci. USA* **106**, 16333–16338 (2009).
8. Hall, C. & Dietrich, F. S. The reacquisition of biontin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* **177**, 2293–2307 (2007).
9. Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. & Schmidt, S. A. Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3 (Bethesda)* https://doi.org/10.1534/g3.115.025692 (2016).
10. Liti, G. et al. Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
11. Akao, T. et al. Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res.* **18**, 423–434 (2011).
12. Marsit, S., Sanchez, L., Galeote, V. & Dequin, S. Horizontally acquired oligopeptide transporters favour adaptation of *Saccharomyces cerevisiae* wine yeast to oenological environment. *Environ. Microbiol.* **18**, 1148–1161 (2016).
13. Nishimura, A., Kotani, T., Sasano, Y. & Takagi, H. An antioxidative mechanism mediated by the yeast N-acetyltransferase Mpr1: oxidative stress-induced arginine synthesis and its physiological role. *FEMS Yeast Res.* **10**, 687–698 (2010).
14. Sasano, Y., Takahashi, S., Shimaa, J. & Takagi, H. Antioxidant N-acetyltransferase Mpr1/2 of industrial baker’s yeast enhances fermentation ability after air-drying stress in bread dough. *Int J. Food Microbiol.* **138**, 181–185 (2010).
15. Ness, F. & Aigle, M. RTM1: a member of a new family of telomeric repeated elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae* s. *Genetics* **153**, 1381–1387 (2009).
16. Mitchell, L. A. & Boeke, J. D. Circular permutation of a synthetic eukaryotic chromosome. *Proc. Natl Acad. Sci. USA* **106**, 1695–1698 (2009).
17. Ness, F. & Aigle, M. RTM1: a member of a new family of telomeric repeated elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae* s. *Genetics* **153**, 1381–1387 (2009).
18. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **344**, 1381–1387 (2012).
19. Xie, Z.-X. et al. Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Cold Spring Harb. Symp. Quant. Biol.* **78**, 1517–1518 (2013).
20. Mitchell, L. A. & Boeke, J. D. Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Proc. Natl Acad. Sci. USA* **106**, 1695–1698 (2009).
21. Gibson, D. G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
22. Hutchinson, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
23. Xie, Z.-X. et al. Perfect designer chromosome V and behavior of a ring derivative. *Science* **355**, eaaf4704 (2017).
24. Mitchell, L. A. & Boeke, J. D. Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Proc. Natl Acad. Sci. USA* **111**, 17003–17010 (2014).
25. Newlon, C. S. et al. Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. *Cold Spring Harb. Symp. Quant. Biol.** 58**, 415–423 (1993).
26. Luo, J. et al. Synthetic chromosome fusion: effects on genome structure and function. Preprint at bioRxiv https://doi.org/10.1101/381137 (2018).
27. Marsit, S. et al. Evolutionary advantage conferred by an Eukaryote-to-Eukaryote gene transfer event in wine yeasts. *Mol. Biol. Evol.* **32**, 1695–1707 (2015).
ACKNOWLEDGEMENTS

Thank you to Toni Cordente, Jef Boeke, Markus Herderich and Ella Robinson for valuable comments and John Dilley for advice on suitable ARS sequences. The AWRI, a member of the Wine Innovation Cluster in Adelaide. Research at the AWRI is supported by Australia’s grape growers and winemakers through their investment body Wine Australia with matching funds from the Australian Government. Macquarie University’s synthetic biology research is supported by Bioplatforms Australia, the New South Wales (NSW) Chief Scientist and Engineer, and the NSW Government’s Department of Primary Industries. The Macquarie-led ARC Centre of Excellence in Synthetic Biology is funded by the Australian Government through its investment agency, the Australian Research Council.

AUTHOR CONTRIBUTIONS

D.R.K. constructed the PGNC, designed and performed the characterization, phenotyping and SCRaMBLE experiments and wrote the manuscript; T.C.W., H.D.G. and I.T.P. undertook the BioLog screening and assisted with SCRaMBLE methodologies; G.A.D. performed the nanopore sequencing, assembly and comparative genomics; I.S.P. and D.L.J. organized funding and provided guidance for the PGNC research; A.R.B. conceived the PGNC research project, designed the PGNC element and experimental procedures and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31305-4.

Correspondence and requests for materials should be addressed to Anthony R. Borneman.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022