Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome

B. A. Blount, G.-O. F. Gowers, J. C. H. Ho, R. Ledesma-Amaro, D. Jovicic, R. M. McKiernan, Z. X. Xie, B. Z. Li, Y. J. Yuan & T. Ellis

Synthetic biology tools, such as modular parts and combinatorial DNA assembly, are routinely used to optimise the productivity of heterologous metabolic pathways for biosynthesis or substrate utilisation, yet it is well established that host strain background is just as important for determining productivity. Here we report that in vivo combinatorial genomic rearrangement of *Saccharomyces cerevisiae* yeast with a synthetic chromosome V can rapidly generate new, improved host strains with genetic backgrounds favourable to diverse heterologous pathways, including those for violacein and penicillin biosynthesis and for xylose utilisation. We show how the modular rearrangement of synthetic chromosomes by SCRaMbLE can be easily determined using long-read nanopore sequencing and we explore experimental conditions that optimise diversification and screening. This synthetic genome approach to metabolic engineering provides productivity improvements in a fast, simple and accessible way, making it a valuable addition to existing strain improvement techniques.
The transplantation of pathways encoding phenotypes of interest into a more easily engineered host organism is a common process in metabolic engineering, biotechnology and synthetic biology. Optimisation of these heterologous pathways is typically achieved by determining the combination of pathway enzymes and promoter sequences that ensure their ideal expression levels. Alternatively, pathway productivity can also be enhanced by changing the host cell within which the pathway is active, either by switching to an industrial strain or by deleting or overexpressing native genes, such as those involved in primary metabolism. Strain improvement via changes to the host cell genome can be either rational or model-guided or can use adaptive or directed evolution to promote an improved phenotype. In commonly used industrial organisms like yeast, one can also make use of the many host diversification approaches currently available such as mutagenesis and large-scale sexual mating.

The recent arrival of Saccharomyces cerevisiae yeast strains where natural chromosomes are replaced by designed, synthetic chromosomes now offers a radically new form of genome diversification of genes being deleted or altered in their expression. This is afforded by the Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution (SCRaMBLE) system that has been designed into the synthetic sequence of chromosomes produced by the Sc2.0 project. SCRaMBLE is a combinatorial rearrangement and gene deletion system that utilises the placement of symmetrical loxP recombination sites in the 3′ untranslated regions (3′UTR) of all non-essential genes. When Sc2.0 chromosomes are exposed to Cre recombinase in vivo, rearrangements occur that quickly result in major topological changes to the synthetic chromosomes within that culture.

A previous study on SCRaMBLE used genome sequencing to reveal the typical outcomes when a 100 kb circular region of synthetic chromosome DNA is rearranged in yeast grown in standard growth conditions. Owing to the complexity of some recombination events (e.g., tandem duplications), the post-SCRaMBLE DNA topology could not be determined for all sequenced strains using only short-read technology. However, in those strains where sequencing did resolve the new layout, extensive rearrangements were observed, both in cells that maintained normal growth and those exhibiting impaired growth.

The diversity of outcomes seen just with this short synthetic DNA region highlights how millions of unique genotypes could be produced by SCRaMBLE of synthetic chromosomes within a population of cells. We reasoned that this would offer a huge potential phenotype space that could be exploited for strain improvement, and specifically for generating strains with altered genetic backgrounds that provide a benefit for heterologous pathways of industrial relevance. Thus we set out to develop an approach that combines the chromosomal diversification seen previously with SCRaMBLE, with the expression and screening of heterologous pathways hosted in synthetic yeast. To go beyond previous work and scale-up the potential phenotypic space, we used yeast now containing a fully synthetic 536 kb chromosome, and to ensure that the rearrangements could be easily mapped, we sequenced resulting strains of interest using long-read nanopore sequencing. Our work describes SCRaMBLE as a powerful method for host strain diversification. When haploid synthetic yeast strains expressing heterologous pathways are subject to a short SCRaMBLE process with no applied selective pressure, diverse genetic backgrounds are quickly produced that significantly improve the productivity of multiple heterologous metabolic pathways of industrial interest.

**Results**

SCRaMBLE generates a strain with improved biosynthesis yield. To investigate whether induced SCRaMBLE of a synthetic chromosome in vivo can enhance a host for biosynthesis, we assessed production of a secondary metabolite using the pigment-producing violacein biosynthesis pathway from Chromobacterium violaceum. An exemplar synthetic biology study on heterologous expression optimisation was recently able to achieve a two-fold increase in yield from this pathway in S. cerevisiae using model-guided iterative design. A 2µ plasmid, pJCH017, was constructed to express the five pathway genes (Supplementary Fig. 1). To minimise the effects of any criteria other than host genome SCRaMBLE, no loxP sites were included in the plasmid and all genes were assembled with strong promoters. pJCH017 and the Cre recombinase expression plasmid, pSCW11-creEBD, were transformed into synV, a haploid yeast strain in which the natural chromosome V has been replaced with a synthetic version containing 174 loxPsym sites. SCRaMBLE was then induced in a growing culture of this strain for 4 h and cells plated onto SDO agar medium to maintain selection for the violacein plasmid. Selection for the Cre recombinase plasmid was not maintained post-SCRaMBLE.

To screen for violacein production, 87 of the resultant colonies, along with un-SCRaMBLEd controls, were picked and grown to an optical density between 0.5 and 0.7 before spotting 10 µl of each culture onto SDO-URA− agar medium (Fig. 1a, b). Of these spotted cultures, the strain in position B2 (named VB2) was visibly darker than controls and thus likely to be producing more of the violacein pigment molecule. Viola
cin production by VB2 was compared to that of the synV-pJCH017 strain by colourimetric assay and was found to be 2.3 times higher (Fig. 1c). VB2 cells cured of the plasmid and retransformed with pJCH017 from the original DNA stock retained the high production, ruling out plasmid-based mutations as an explanation for the enhanced phenotype (VB2c-pJCH017, Fig. 1c). Microscopic analysis of VB2 cells revealed these to be typically larger than synV equivalents, with this increased size more pronounced in rich medium than in synthetic medium (Supplementary Figs. 2, 3).

To determine whether the enhanced violacein production is related to the pathway metabolism or to increased expression of the pathway enzymes, cured synV and VB2 strains were next transformed with a 2µ vector expressing superfolder green fluorescent protein (sfGFP) and a URA3 auxotrophic marker (pBAB012). VB2-pBAB012 showed enhanced sfGFP production, compared to synV-pBAB012, supporting the case for VB2 providing a general boost in the expression of genes from the 2µ vectors (Fig. 1d, Supplementary Fig. 4–6). To determine whether this effect is 2µ replicon specific, fluorescence from sfGFP expressed from a CEN6/ARS replicon (pBAB011) was also assessed in both VB2 and synV hosts (Fig. 1e) revealing no increase in expression. To further rule out URA3-specific factors, sfGFP expression was measured with LEU2 auxotrophic marker vectors giving a similar pattern of results (Fig. 1e). Lower fluorescence output was seen from plasmids with a LEU2 marker, consistent with previous findings.

Quantitative PCR (qPCR) measurement of the kanR, viob and viod genes on pJCH017 next determined the relative copy numbers of violacein pathway plasmid in VB2-pJCH017 and synV-pJCH017 (Fig. 1f). This revealed a 1.89-fold increase in 2µ DNA copies per genomic DNA in VB2 cells compared to synV, and this increased plasmid copy number likely explains the observed increase in violacein production in this strain.

To demonstrate that the increase in plasmid copy number in VB2 can also enhance biosynthesis from other 2µ constructs, a pathway for penicillin G production in S. cerevisiae was next assessed. The best-performing pathway constructs from a
previous study, encoded on a 2μ and a CEN6 vector, were co-transformed into both synV and VB2 hosts. These strains, along with an existing BY4741 strain containing the pathway plasmids, were then assayed for secreted penicillin yields by liquid chromatography–mass spectrometry (LC-MS). Penicillin production was indeed enhanced by 2.1-fold in the VB2 strain compared to that produced by synV or BY4741 hosts (Fig. 1g, Supplementary Fig. 7). The penicillin G pathway has proved difficult to express in yeast, requiring subcellular localisation of pathway enzymes and library-based screening strategies to achieve bioactive levels of the antibiotic molecule. The yield achieved with the VB2 strain, 14.9 ng ml$^{-1}$, is significantly higher than the previous highest recorded yields between 5 and 6 ng ml$^{-1}$.

Thus VB2, generated by a short SCRaMbLE protocol with no selective pressure, now provides a new host capable of increasing the expression of pathways and genes on 2μ plasmids.

SCRaMbLE to generate a strain with enhanced growth on xylose. In addition to engineering S. cerevisiae with heterologous genes to produce biomolecules of interest or value, altering cellular metabolism to allow or enhance growth on alternative energy sources is often desirable. To demonstrate that SCRaMbLE
can also improve the growth of yeast in an attractive alternative carbon source, the heterologous pathway of xylose utilisation was introduced (Fig. 2a). Xylose is a much-targeted alternative carbon source for yeast growth, due to its high abundance in lignocellulosic biomass. As S. cerevisiae is unable to grow using xylose as the sole carbon source, plasmid pJCH006 containing the XYL1 and XYL2 genes from Scheffersomyces stipitis and an additional copy of the XKS1 xylulokinase from S. cerevisiae (Supplementary Fig. 8), was constructed and transformed into the synV strain to provide the oxidoreductase pathway for xylose utilisation. SCRaMbLE was induced, as before, for 4 h with synV-pJCH006/pSCW11-creEBD cells and these were plated onto SDO URA− agar medium. Eighty seven colonies were picked at random and grown at 30°C in a microwell plate in selective media with xylose as the sole carbon source (SCX URA−). Optical densities of cultures were monitored over 5 days to identify colonies showing improved growth in xylose compared to synV-pJCH006 controls. Strain XD4 was selected from the SCRaMbLE colonies as it exhibited dramatically enhanced growth in xylose in this preliminary screen (Fig. 2b). To further characterise XD4, plate-based growth assays were performed comparing XD4-pJCH006 to synV-pJCH006 in xylose- and glucose-containing media (Fig. 2c, d). XD4-pJCH006 shows on average 5.85 times the growth rate of synV-pJCH006 in SCX URA− xylose medium, whilst having only 0.79 times the growth rate in SDO URA− glucose medium (Fig. 2e). A similar increase in growth rate in xylose medium was also observed when the strains were grown in 60 ml cultures in baffled flasks (Supplementary Fig. 9). Of the 19 XD4-pJCH006 cultures assayed in 96-well plate format, the average growth rate was 0.2 h−1, with some cultures having a growth rate over 5 h measuring as high as 0.238 h−1. This compares favourably to the 0.18 h−1 maximum growth rate reported in similar conditions after extensive optimisation of this pathway in the diploid CEN.PK 113-5D strain and the 0.21 h−1 recently achieved in CEN.PK 113-7D via targeted overexpression and mutation strategies. 

Mutations in pJCH006 were discounted as being responsible for the phenotype by curing XD4 of this plasmid and retransforming it with original pJCH006 DNA stock. After 48 h growth in xylose media in 96-well format, the endpoint optical density of this retransformed strain (XD4c-pJCH006) was equivalent to that of XD4-pJCH006 (Supplementary Fig. 10). Microscopy of XD4 cells also revealed no noticeable morphological differences to synV in SX URA− medium (Supplementary Figs. 11, 12).

To further investigate the XD4 phenotype, 25 ml cultures of XD4-pJCH006 and synV-pJCH006 with 4% xylose, 4% glucose or 2% xylose plus 2% glucose were grown in liquid chromatography (HPLC) (Fig. 2f, g, Supplementary Fig. 13). These analyses show that, in the 4% xylose cultures, XD4-pJCH006 removes around 13.5 g of xylose from the medium, whereas synV-pJCH006 does not remove a detectable amount. In the 2% xylose plus 2% glucose medium, the synV-pJCH006 strain accumulates xylitol, whereas this is not the case with XD4-pJCH006. The oxidoreductase pathway used here is known to generate a reduct imbalance in S. cerevisiae that results in xylitol accumulation and it seems that XD4 is able to bypass this.

Long-read sequencing to determine SCRaMbLE rearrangements. In order to confirm that SCRaMbLE events have occurred and determine what type of chromosome rearrangements have generated the phenotypes of interest, it is desirable to sequence the genomes of chosen strains. While short-read technologies are
The line at 20 kbp indicates the DNA shear length during library preparation. Inset is the same analysis of the subset of the data that aligned by LAST to the synthetic chromosome V sequence. The alignment of XD4 nanopore sequencing reads to parental synV sequence around the identified inversion region. An alignment of XD4 nanopore sequencing reads to parental synV sequence around the identified MXR1 deletion region. Full alignment of the VB2 contig, and the alignment of the XD4 contig to the parental synV. The inversion and deletion regions are expanded to show diagrammatic interpretations of the SCRaMbLe events, showing pre- (top) and post-SCRaMbLe (bottom) configurations, with locations of promoters (arrows), coding sequences (polygons), terminators (T), autonomous replicating sequences (ARS, triangle) and loxPsym sites (red dotted lines) indicated. Red coding sequences are deleted by SCRaMbLe events and pink coding sequences have their 3′ UTR regions altered.

To demonstrate the use of nanopore sequencing to determine post-SCRaMbLe chromosome rearrangements, genomic DNA preparations from XD4 and VB2 were sequenced, each on a single R9 flow cell using the portable ONT MinION device. For VB2, the single 48 h run yielded 2.83 Gbp of base-called sequence data with a read-length distribution with a mean average of 11 kb and an n50 of 16.3 kb. This resulted in a ~233× average genome coverage when reads below 1 kb were excluded. For XD4, the single 48 h run yielded 1.72 Gbp of base-called sequence data with a read-length distribution with a mean average of 12.5 kb and an n50 of 24 kb, corresponding to an average genome coverage of ~133× when reads below 1 kb were excluded (Fig. 3a). Reads were corrected using Canu v1.5 and filtered to yield reads between 1 kb and 60 kb in length (Fig. 3b, c).

Deletions of UBP3 and MXR1 were recreated in BY4741 using CRISPR-mediated recombination, to determine whether these events lead to the VB2 and XD4 phenotypes, respectively. The BY4741 ubp3Δ-pBAB012 strain showed no increase in sGFP

Fig. 3  Nanopore sequencing to identify the SCRaMbLe events in enhanced strains. a  Base-called read length distribution of the XD4 nanopore sequencing run. The line at 20 kbp indicates the DNA shear length during library preparation. Insert is the same analysis of the subset of the data that aligned by LAST to the synthetic chromosome V sequence. b  An alignment of XD4 nanopore sequencing reads to parental synV sequence around the identified inversion region. c  An alignment of XD4 nanopore sequencing reads to parental synV sequence around the identified MXR1 deletion region. d  Full alignment of the VB2 contig, and e  the alignment of the XD4 contig to the parental synV. The inversion and deletion regions are expanded to show diagrammatic interpretations of the SCRaMbLe events, showing pre- (top) and post-SCRaMbLe (bottom) configurations, with locations of promoters (arrows), coding sequences (polygons), terminators (T), autonomous replicating sequences (ARS, triangle) and loxPsym sites (red dotted lines) indicated. Red coding sequences are deleted by SCRaMbLe events and pink coding sequences have their 3′ UTR regions altered.

The sequence of VB2 shows that two SCRaMbLe events have occurred in synthetic chromosome V, deleting the 2971 bp region encoding UBP3 and inverting the 5337 bp region containing the SWI4 and LSM4 coding sequences (Fig. 3d). In this inversion event, LSM4 retains its 3′ UTR but SWI4 does not. The gene TMN3 also loses its 3′ UTR. The recombination sites all map to loxPsym sites, confirming that SCRaMbLe was responsible for the chromosomal changes. These SCRaMbLe events were further confirmed by PCR screening (Supplementary Fig. 14). Sequencing of XD4 also revealed just two SCRaMbLe events in synthetic chromosome V, an inversion of a 7 kb locus encoding GCN4, YEL008W, MIT1 and YEA6 genes (plus ARS510), and deletion of a 785 bp region containing the short MXR1 coding sequence (Fig. 3b, c, e). As with the VB2 sequencing, these recombination events mapped to loxPsym sites and were confirmed by PCR amplification with region-specific primers (Supplementary Fig. 15).

Deletions of UBP3 and MXR1 were recreated in BY4741 using CRISPR-mediated recombination, to determine whether these events lead to the VB2 and XD4 phenotypes, respectively. The BY4741 ubp3Δ-pBAB012 strain showed no increase in sGFP
OD600 achieved after 48 h growth in synthetic xylose medium of XD4-pJCH006, synV-pJCH006 and the 13 fastest growing re-SCRaMbLEd XD4-SCRaMbLE. For ease of analysis, we chose to screen for increased output compared to BY4741-pBAB012 (Supplementary Fig. 16), and while BY4741 mxr1Δ-pJCH006 did show a significant increase in xylose growth compared to BY4741-pJCH006, the improvement was not on the scale seen in XD4 (Supplementary Fig. 16), as varying the inducer concentration has a limited effect on post-SCRaMbLE cell viability levels (Supplementary Fig. 17).

Comparing conditions for SCRaMbLE induction and screening. Both VB2 and XD4 strains were identified from 4 h SCRaMbLE reactions followed by screening of randomly picked colonies. We avoided applying selective pressure during SCRaMbLE and subsequent plating as selective pressures are not always available for a desired phenotype, and <100 colonies were screened each time, simply for the convenience of using 96-well plate format. The fact that improved strains were identified by this method for two quite different cases led us to further investigate the likelihood of seeing an improved strain by SCRaMbLE. For ease of analysis, we chose to screen for increased green fluorescence when sfGFP is expressed from a 2 μm plasmid, as seen previously with VB2, and we induced synV SCRaMbLE for varying times, and also screened both with and without a prior selection. β-estradiol concentration at induction was kept constant at 1 μM, as varying the inducer concentration has a limited effect on post-SCRaMbLE cell viability levels (Supplementary Fig. 17).

pSCW11-creEBD and pBAB016 (2 μm plasmid with sfGFP) were transformed into synV. In quadruplicate, cultures of this strain underwent SCRaMbLE with a β-estradiol induction time of 2, 4, 8 or 24 h. Control cultures also underwent the process but without addition of the inducer. To represent no selective pressure, 83 SCRaMbLE-induced colonies and 6 uninduced colonies were randomly selected for each condition and sfGFP fluorescence per cell was quantified after 24 h (Fig. 4a). To represent selective pressure, this process was repeated but this time only taking colonies post-SCRaMbLE that, by eye, appeared to fluoresce strongly under blue light (Fig. 4b).

Colony counts from the plated SCRaMbLE cultures indicate that viability reduces with longer induction times, ranging from 17.3% with 2 h induction to 5.5% with 8 h induction (Fig. 4c), presumably as increased SCRaMbLE events lead to the deletion or inactivation of more essential genes. This trend does not continue to the 24 h induction culture, which has an 8% viability,
presumably due to a combination of the inducer degrading over time within the culture and previous viability losses being mitigated by cell growth from survivors.

With 96-well throughput, cells with similar sfGFP output to that seen with VB2 were identified by both random and selective screening (Fig. 4a, b). Selective screening markedly increased the proportion with an increased output, with 94.7% of selected 8 h induction colonies achieving this compared to only 23.5% of the randomly selected equivalents. With either approach, an 8 h induction of SCRaMbLE resulted in the highest proportion of cells with mean fluorescence above all un-SCRaMbLEd controls. As with the cell viability data, the percentage of improved strains arising rose as SCRaMbLE induction increased to 8 h but was reduced by a 24 h SCRaMbLE (Fig. 4c).

Analysis by PCR of the genomic DNA from the highest-performing sfGFP strains from the random and selective screens (GB2 and GH3, respectively) showed that neither had replicated either of the two SCRaMbLE events identified by sequence analysis of VB2, and so each had achieved increased fluorescence via different chromosome rearrangements (Supplementary Fig. 14). It is not unexpected that each SCRaMbLE will generate a unique outcome, because the total number of possible rearrangement combinations with hundreds of loxPsym sites is astronomical.

Finally, to investigate whether additional SCRaMbLE of a selected strain can further improve a phenotype, a second round of SCRaMbLE was performed on the xylose-utilising XD4 strain. XD4-pChO06/pSCW11-creEBD cells were SCRaMbLEd with a 4 h induction, but this time plated directly onto SCX URA- agar to accelerate selection. Visible colonies were not evident for 7 days and viable colony numbers were far lower than previously seen with just a primary SCRaMbLE (7 \times 10^5 cfu ml\(^{-1}\) vs. 2.6 \times 10^5 cfu ml\(^{-1}\), Supplementary Fig. 18). The 13 largest colonies were assayed for growth in SCX URA\(^{-}\) medium but all displayed slower growth than XD4 (Fig. 4d). This indicates that the initial SCRaMbLE provided strain improvement, but further SCRaMbLE of an already SCRaMbLEd strain proved detrimental. This may be due to the reduction in possible genetic and phenotypic space that results from the isolation of a particular SCRaMbLEd strain, with further events reducing fitness, causing inviability or simply not affecting the desired phenotype. These results indicate that a longer SCRaMbLE induction would typically be a more effective strategy for strain improvement, rather than rounds of SCRaMbLE punctuated by screening and isolation of individual strains.

**Discussion**

The in vivo rearrangement of synthetic yeast chromosomes offers a powerful approach for optimising the host for improved production of molecules or for enhanced growth on alternative substrates. As demonstrated here, SCRaMbLE effectively offers an extreme form of inducible mutation capable of quickly generating host strains with significantly enhanced performance for different tasks of relevance to biotechnology. The process compares favourably in several ways to existing strain improvement methods, such as transcription machinery engineering, gene deletion library screens and genome shuffling (Table 1), and can be used as an alternative to or to complement these established techniques. It is rapid, with diversification taking only a few hours, and it does not require application of a selective pressure throughout the process. Optimisation by SCRaMbLE is inexpensive and relatively non-technical, requiring minimal equipment or reagents. Additionally, when using SCRaMbLE to improve heterologous pathway performance, such as in VB2 and XD4, results are also achieved with minimal knowledge of the pathway. Indeed, the pathway genes used in the cases described here were all paired with strong promoters with no attempts to balance their expression levels beforehand.

As with other mutation methods used to improve strains, SCRaMbLE can simply be treated as a black box process where enhanced hosts are generated without the genetic events causing the improvements being fully understood. However, as shown here, desktop long-read sequencing now makes it possible to quickly determine the layout and content of rearranged fully synthetic chromosomes. While knowledge of the genomic changes is only part of the puzzle towards explaining global outcomes such as altered metabolism, it allows us to speculate on mechanisms for the phenotypic changes and provides the foundations for future work that can fully uncover how SCRaMbLE leads to a given strain improvement.

For the two strains sequenced here, VB2 and XD4, the identified inversions and deletions point towards possible routes for increased 2\(\mu\) copy number and improved xylose utilisation, respectively. For example, in the inversion event in VB2, the SWI4 gene loses its 3'UTR and thus may have subsequently have altered expression or regulation. SWI4 encodes the DNA-binding component of the SBF transcription complex, which has a major role in regulating several G1-specific genes, including DNA synthesis genes\(^{27}\). As 2\(\mu\) vector replication occurs in S phase, a change in the regulation of DNA synthesis genes in G1, due to changes to SWI4 expression, could be responsible for increased 2\(\mu\) copy number if levels of DNA synthesis machinery are altered going into S phase\(^{28}\). For XD4, the inversion event leaves GCN4 without its standard 3'UTR sequence and so also may alter its expression or regulation. This is of particular interest given that Gcn4 is a global metabolic transcription factor and has previously been identified to be of importance for xylose utilisation\(^{29,30}\). In addition, the deletion of MXR1 may improve xylose utilisation by indirectly correcting the redox imbalance known to lead to xylitol accumulation in un-SCRaMbLEd yeast\(^{24}\). MXR1 encodes a cytosolic methionine-S-sulfoxide reductase involved in oxidative stress that is indirectly linked to NADPH via its use of the redox carrier thioredoxin\(^{31,32}\).

In future work, RNAseq, proteomics and metabolomics studies of these SCRaMbLEd strains or others will help us better understand how various deletions and rearrangements on synthetic chromosomes lead to different phenotypic changes. The full effects of deletions and inversions on local gene expression may well be highly complex, as the altered topology at SCRaMbLEd loci is also likely to affect the expression of neighbouring genes through multiple mechanisms, such as changes in local nucleosome positioning, disruption of nearby promoter regulation and repositioning of genes relative to one another. Each phenotype may well be due to one or a combination of many of these various genomic changes and will be an interesting challenge to explore. It is important to also consider that these changes may have other unanticipated effects, such as reducing the fitness of the host in conditions not being screened for. As with hosts generated by other strain diversification methods, further testing of performance would be needed before using SCRaMbLEd-generated strains for production at commercial scale.

The unique power of the SCRaMbLE method is in being able to generate so many different events randomly and combinatorially with ease, and offering a rapid approach for yeast host strain improvement that we show here can significantly benefit multiple heterologous pathways. Given that the synthetic yeast strains being generated by the Sc2.0 project are all freely available and have no associated intellectual property restrictions\(^{33}\), host optimisation by SCRaMbLE represents a powerful and accessible strain enhancement technology with broad applications.
### Table 1 Comparison of SCRaMbLE to common S. cerevisiae strain improvement strategies

| Technique                        | SCRaMbLE          | Mass mating/ genome shuffling | Transcription machinery engineering/ rewiring | Deletion library screening | Targeted deletion/ overexpression/ underexpression | Random mutagenesis/ directed evolution |
|----------------------------------|-------------------|-------------------------------|----------------------------------------------|-----------------------------|-----------------------------------------------------|---------------------------------------|
| Strain prerequisites             | Synthetic chromosomes and recombinase vectors | Natural isolates with desired properties | Efficient transformation into strain | Deletion library, efficient transformation into strain | —                                      | —                                    |
| Prior knowledge                  | —                 | —                             | Identification of transcriptional regulators/ regulated promoters | —                          | Prior gene characterisation studies, knowledge of metabolic processes | —                                    |
| Preparation work                 | Transform pathway into one strain | Transform pathway into half of mating strains | Large-scale library construction | Transform pathway into many strains | Construction of TAR cassettes | Transform pathway into one strain |
| Technical requirements           | Basic strain cultivation | DNA assembly, library scale DNA transformation | Library scale DNA transformation | DNA assembly/gene editing tools | DNA assembly/gene editing tools | DNA assembly/gene editing tools |
| Timescale to improved strain     | 1-2 weeks         | Months to years | Weeks to months | Months | Months | Months |
| Strain diversity generated       | Coding sequence deletions, inversions, duplications, translocations | Novel regulation | Coding sequence deletions | Specific targeted changes | Point mutations | — |
| Other factors                    | —                 | High prevalence of deleterious alleles, mitotic instability | High throughput screening often necessary | Higher numbers of combined knockouts exponentially increases transformation requirements | Understanding of biological mechanisms often not sufficient for effective improvement | High number of off-target effects, selective pressure often required |

Technical requirements do not include screening methods, which are common to all methods. Timescale to improved strain estimates length of time from introduction of a heterologous pathway to isolation and characterisation of a strain with an improved phenotype.

### Methods

#### Strains and media.

Yeast strains BY4741 and YZXX846 (BY4741 with synthetic chromosome V, referred to as synV in the text) were generated in previous studies. All other yeast strains were generated during this study. Escherichia coli DH10B (Life Technologies) was used as a bacterial cloning and plasmid propagation host. Luria Bertani medium was used for bacterial growth, YPD (10 g L⁻¹ glucose, 20 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, 20 g L⁻¹ yeast nitrogen base, 14 g L⁻¹ tryptophan, 10 g L⁻¹ histidine, 10 g L⁻¹ leucine, 10 g L⁻¹ uracil) was used for bacterial growth, YPD (10 g L⁻¹ glucose, 20 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, 20 g L⁻¹ yeast nitrogen base, 14 g L⁻¹ tryptophan, 10 g L⁻¹ histidine, 10 g L⁻¹ leucine, 10 g L⁻¹ uracil) was used for selective growth.

#### Plasmid construction.

All plasmids generated in this study were assembled using the MoClo-Yeast Toolkit (YTK)³⁴. pCH017 was constructed by amplifying vioA, vioB, vioC, vioD and vioE from pNC104 and assembling into five intermediate plasmids, encoding pTDH3-vioA-TDH1, pPGK1-vioB-ENO2, pHHF2-vioC-pPGK1, pCCW12-vioD-IADH1 and pTEF1-vioE-ISSA1, respectively. The expression cassettes were then assembled together with a URA3 auxotrophic marker, a 2 μm yeast replicon and ColE1 kanR bacterial replicon components to yield the final plasmid. pCH017 was constructed by amplifying vioA, vioB, vioC, vioD and vioE from pNC104 and assembling into five intermediate plasmids, encoding pTDH3-vioA-TDH1, pPGK1-vioB-ENO2, pHHF2-vioC-pPGK1, pCCW12-vioD-IADH1 and pTEF1-vioE-ISSA1, respectively. These expression cassettes were then assembled together with a URA3 auxotrophic marker, a 2 μm yeast replicon and ColE1 kanR bacterial replicon components to yield the final plasmid. pCH017 was constructed by amplifying vioA, vioB, vioC, vioD and vioE from pNC104 and assembling into five intermediate plasmids, encoding pTEF2-XYL1-IADH1, pTEF1-XYL1-TDH1 and pTDH3-XKS1-ENO2, respectively. The expression cassettes were then assembled together with a URA3 auxotrophic marker, a 2 μm yeast replicon and ColE1 kanR bacterial replicon components to yield the final plasmid. pCH017 was constructed by amplifying vioA, vioB, vioC, vioD and vioE from pNC104 and assembling into five intermediate plasmids, encoding pTEF2-XYL1-IADH1, pTEF1-XYL1-TDH1 and pTDH3-XKS1-ENO2, respectively. These expression cassettes were then assembled together with a URA3 auxotrophic marker, a 2 μm yeast replicon and ColE1 kanR bacterial replicon components to yield the final plasmid.

#### Yeast transformations.

All yeast transformations were performed using the lithium acetate method with a 14 min 42 °C heat shock followed by a 10 min recovery in 5 mM calcium chloride prior to plating on appropriate selective media.

DNA was isolated for PCR analysis using the Qiagen spin Miniprep Kits (Qiagen). Plasmids were isolated from bacterial hosts using the Qiagen spin Miniprep Kits (Qiagen).

SCRaMbLE. Single colonies were picked and used to inoculate 5 ml of appropriate selective glucose medium and cultured overnight at 30 °C. The overnight culture was used to inoculate a 5 ml of appropriate selective medium 1/100 and grown shaking at 30 °C. After 2 h, β-estradiol (Sigma Aldrich) was added to a final concentration of 1 μM and cultures were grown shaking at 30 °C for a further 4 h (unless otherwise stated) before being washed in Dulbecco’s phosphate-buffered saline (DPBS, Life Technologies), serially diluted and plated onto appropriate solid media without selection for pJC104 with added 1 μM β-estradiol. To determine the extent of cell death or inviability due to SCRaMbLE in a particular experiment, an equivalent culture was inoculated from the same overnight culture as the SCRaMbLE-induced culture and went through the same process but without the addition of β-estradiol. Viability is the colony count from the induced culture as a proportion of the colony count from the uninduced culture.

Post-SCRaMbLE selection for violacein production. Following SCRaMbLE of synV-pCH017, colonies were picked and used to inoculate 600 μl SDO URA- cultures in a 96 deep-well plate and grown shaking at 700 r.p.m. at 30 °C. After the plates were grown to an OD₆0₀ between 0.5 and 0.7, 10 μl of culture from each well was spotted onto SDO URA- agar and grown at 30 °C for 3 days. The darkest spotted culture was selected for further characterisation.

#### 96-well plate assays.

Single colonies were picked and used to inoculate 5 ml of appropriate selective glucose medium and cultured overnight at 30 °C. The OD₆0₀ of cultures was then determined by spectrophotometry using an Eppendorf BioSpectrometer. Cultures were normalised to an OD₆0₀ of 0.2, washed in DPBS and then resuspended in the appropriate growth medium. Normalised cultures were then used to inoculate wells in a 96-well plate to an OD₆0₀ of 0.02, up to a final volume of 100 μl. Plates were incubated in a Biotek Synergy HT plate reader for the duration of the growth assay at 30 °C, shaking at the medium setting between readings. For growth curves, all values were normalised to the average OD₆0₀ value of blank media wells measured at the same time as that time point prior to statistical analysis. For fluorescence values, wells were measured with an excitation wavelength of 485 nm and an emission wavelength of 528 nm. All values were normalised to the average OD₆0₀ value of blank media wells and then to negative control cells without a plasmid (with media blank subtracted) grown in the same assay plate prior to statistical analysis. Endpoint sfGFP quantification for comparison of SCRaMbLE induction times was performed in the same manner.
except with plates inoculated using 1 μl of saturated overnight 96-well 100 μl SC LEU cultures up to a final volume of 100 μl.

**Statistical analysis.** Unless otherwise stated, values given are mean average values. Significance was determined via two-sample t-tests with assumption of variance determined by F-test. Two-tailed p-values are given with *p* ≤ 0.05, **p** ≤ 0.01, ***p*** ≤ 0.001 and ****p*** ≤ 0.0001. Replicate numbers given indicate the number of biological replicates, except for in the case of qPCR sample analysis where in each case two technical replicates were performed of each of three biological replicates.

**Violacein extraction and quantification.** Violacein extract collection was performed by adapting a previously described method. Briefly, strains were grown for 2 days in 5 ml SC URA, with 1 ml of culture then removed, pelleted by centrifugation and resuspended in 500 μl methanol. This suspension was heated to 95 °C for 4 min, allowed to cool down for 2 min, and then cell debris was removed by centrifugation. Extracts were quantified for violacein content by measurement of absorbance at 577 nm using an Eppendorf BioSpectrometer. Concentration of violacein was determined using the Beer-Lambert Law, assuming the extinction coefficient at 577 nm of violacein in methanol to be 1.7 × 10⁶ dm⁻³ mol⁻¹ cm⁻¹

**Quantitative PCR.** All qPCR reactions were performed in an MasterCycler ep RealPlex 4 (Eppendorf) using SYBR FAST Universal qPCR Master Mix (Kapa Biosystems) according to the manufacturer’s instructions. DNA quantification reactions all contained 1 ng template DNA, as determined by Qubit. A calibration curve for each amplicon was constructed by quantifying known purified plasmid pCH017 amounts at 100, 50, 20, 10, 5, 2.5, 1.25 pg, as determined by Qubit. To prepare a calibration curve, the threshold cycle (Ct) values of calibration samples were used to generate calibration curves, each with an R² value of >0.97, with which sample Ct values were converted to mass values. ACT1 control amplicons were included to ensure that starting material was equal between samples. The qPCR cycle used was an initial 15 min 95 °C denaturation step followed by 40 cycles of 95 °C for 30 s and then 60 °C for 30 s. All qPCR primers were designed using the web-based Integrated DNA Technologies RealTime PCR Design Tool (http://idtdna.com/scitools/applications/realtimepcr). To assess relative plasmid copy number, kanR DNA was amplified using primers BB_kanR_F (ATTCCGGTACCAGTTAGTCTT) and BB_kanR_R (ATGTCGGGCAATCAGGTG), respectively. This template DNA was co-transformed into pJCH006, colonies were picked and used to inoculate 600 μl 1X YPDA (sugar), with 4% glucose, SX URA, BB632 (GCCCTAGGTTTGGCTGG)/BB633 (GGTGAGGACTCGTTCGG) and GB007 (GG031 (TGGGCGAATAGCAGAGCTG), GG032 (GATGACGAATCGAGACCTG), GG034 (CTAGTGCGAGCATCACTGAC), GG035 (GCCACACC TTGATCGC) and GG038 (TCTCGAATAAGGCTAAATGC). The XD4 MXR1 deletion was confirmed by amplification with primers GG031 (TGGGCGAATAGCAGAGCTG), GG032 (TCAGTGCGAGCATCACTGAC), GG035 (GCCACACC TTGATCGC) and GG038 (TCTCGAATAAGGCTAAATGC) using primers BB_kanR_F (ATTCCGGTACCAGTTAGTCTT) and BB_kanR_R (ATGTCGGGCAATCAGGTG), respectively. This template DNA was co-transformed into YPDA along with linear fragments encoding Cas9 and a gRNA targeted to either UBPS (CTAAGGCGAATCGAGGCTAAATG) and GB007 (GCCACACC TTGATCGC) as appropriate, following the protocol described at http://benchling.com/pub/ellis-crispr-tools.

**Data availability.** The authors declare that the data supporting the findings of this study are available from the corresponding author on request.

Received: 21 July 2017 Accepted: 23 January 2018
Published online: 22 May 2018

**References**

1. Na, D., Kim, T. Y. & Lee, S. Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* 13, 363–370 (2010).
2. Smanski, M. J. et al. Synthetic biology to access and expand nature’s chemical diversity. *Nat. Rev. Microbiol.* 14, 135–149 (2016).
3. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195 (2012).

**Microscopy.** Five millilitres of cultures were grown overnight and visualised on an Eclipse Ti inverted microscope (Nikon) at 90x magnification. Images were captured using the NIS-Elements Microscope Imaging Software (Nikon). The area of a minimum of 24 cells was determined for each strain and condition using ImageJ (http://imagej.net).

**Nanopore sequencing.** Genomic DNA was sheared to 20 kb using a g-TUBE (Covaris) and then underwent library preparation using a Ligation Sequecing Kit. The library was then sequenced on a Nanopore Technologies machine. A standard 48 h sequencing run was performed using the MinKnow 1.5.5 software using local calling.

**Sequencing analysis.** Raw fast5 files were converted into fastq and fasta using Poretools. Raw reads were corrected using Canu (v1.5) (www.canu.readthedocs.io). Owing to erroneous ultra-long reads for XD4, reads >60 kb and <1 kb were discarded from this corrected pool using a custom python script in combination with samtools (www.sambtools.sourceforge.net), bedtools (www.bedtools.readthedocs.io), last (maf-convert function, www.last.bcbi.jp) and the fasta-kitool (www.hannonlab.cshl.edu). For VB2, all raw reads were used for assembly. Smartdenovo (www.github.com/ruanjue/smartdenovo) was run on corrected reads to de novo assemble a contiguous sequence (contig) using default flags. This contig was compared to synV using lastal (-1100 flag) (www.last.bcbi.jp) and viewed using ACT (www.sanger.ac.uk). Corrected reads were also aligned to synV using lastal (flag --1100) and viewed on integrative genome viewer (IGV) (www.software.broadinstitute.org) to verify SCRaMbLE events seen in the assembled contig.

**PCR verification of SCRaMbLE events.** Genomic loci identified as potentially containing SCRaMbLE events were amplified from genomic DNA by PCR using Phusion Hi-Fidelity DNA Polymerase (New England BioLabs). The VB2 deletion of UBPS was confirmed by amplification with primers GG047 (TCTCTGAGAATAA GTATGGTGCTGTC), GG048 (CTGGCGAGACGATCTAAGAG), GG049 (GAATTAGAATGTTTCTGGCC) and GG050 (GTGACAGCCTGCTGC ACTCC). The VB2 inversion event was confirmed with amplification with primers GG051 (GCGCTGGAGATCGTTCCAAGTG), GG052 (TCAAACGCGCTT GCCAGCAGATT), GG054 (CAGAATTCAAGCGGCTTTGTTACTCG) and GG057 (CGAGAATTCAAGCGGCTTTGTTACTCG) (www.sanger.ac.uk). The XD4 MXR1 deletion was confirmed by amplification with primers GG035 (TGGGCGAATAGCAGAGCTG), GG032 (TCAGTGCGAGCATCACTGAC), GG035 (GCCACACC TTGATCGC) and GG038 (TCTCGAATAAGGCTAAATGC) using primers BB_kanR_F (ATTCCGGTACCAGTTAGTCTT) and BB_kanR_R (ATGTCGGGCAATCAGGTG), respectively. This template DNA was co-transformed into BY4741 along with linear fragments encoding Cas9 and a gRNA targeted to either UBPS (CTAAGGCGAATCGAGGCTAAATG) and MXR1 (CCACAAACTCTG TTAAGA) as appropriate, following the protocol described at http://benchling.com/pub/ellis-crispr-tools.

**CRISPR-mediated recombination.** SCRaMbLE deletion regions of VB2 and XD4 were amplified from gDNA prepared from those strains by PCR using primer pairs (for VB2) BB632 (GCCCTAGGTTTGGCTGG)/BB633 (GGTGAGGACTCGTTCGG) and BB548 (ACCACAGAAGAATACGTTGATG)/BB549 (ACGTTACTGGT GCCTTTCTGGTAG), respectively. This template DNA was co-transformed into BY4741 along with linear fragments encoding Cas9 and a gRNA targeted to either UBPS (CTAAGGCGAATCGAGGCTAAATG) and MXR1 (CCACAAACTCTG TTAAGA) as appropriate, following the protocol described at http://benchling.com/pub/ellis-crispr-tools.

**HPLC analysis of glucose and xylose cultures.** XD4-pCh006 and synV-pCh006 5 ml overnight SD URA cultures were inoculated, grown overnight, washed and normalised for OD₆₀₀ as described for the baffled flask growth curves. In duplicate, cultures were used to inoculate 250 ml conical flasks containing 20% SC URA with 4% glucose, SX URA with 4% xylose or SCX URA with 2% glucose and 2% xylose. Cultures were incubated at 30 °C by shaking at 250 rpm. At 0, 3, 5, 21.33, 25, 30, 45.33, 70.33 and 118.33 h time points, 1 ml samples were taken from each culture with 100 μl analysed by spectrophotometry to determine the concentration of glucose, and the remaining 900 μl being frozen at –20 °C for downstream analysis. Samples were filtered through 0.45 μm membranes prior to quantification of glucose, xylose and xylitol. Quantification was performed on samples and standards by HPLC using an UltiMate 3000 (Dionex-Thermo Fisher Scientific) with an Anion HFX+7H column eluted with 0.01 N H₂SO₄ at room temperature with a 0.6 ml min⁻¹ flow rate.
4. Tyo, K. E., Alper, H. S. & Stephanopoulos, G. N. Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol.* **25**, 132–137 (2007).
5. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355–1358 (2010).
6. Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* **33**, 1061–1072 (2015).
7. Chatterjee, R. & Yuan, L. Directed evolution of metabolic pathways. *Trends Biotechnol.* **24**, 28–38 (2006).
8. Steenels, J. et al. Improving industrial yeast strains: exploiting natural and artificial diversity. *FEBS Microbiol. Rev.* **38**, 947–995 (2014).
9. Annañuru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
10. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
11. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
12. Jovicic, D., Blount, B. A. & Ellis, T. Total synthesis of a eukaryotic chromosome: redesigning and SCarMbLe-ing yeast. *Bioessays* **36**, 855–860 (2014).
13. Shen, Y. et al. SCarMbLe generates designed combinatorial stochastic diversity in synthetic chromosomess. *Genome Res.* **26**, 36–49 (2016).
14. Mercer, G., et al. 3D organization of synthetic and scrambled chromosomes. *Science* **355**, eaaf597 (2017).
15. Mitchell, L. A. et al. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res.* **43**, 6620–6630 (2015).
16. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Duerer, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **41**, 10668–10678 (2013).
17. Xie, Z. X., et al. “Perfect” designer chromosome V and behavior of a ring derivative. *Science* **355**, eaaf704 (2017).
18. Karim, A. S., Curran, K. A. & Alper, H. S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEBS Yeast Res.* **13**, 107–116 (2013).
19. Awan, A. R. et al. Biosynthesis of the antibiotic nonribosomal peptide penicillin in *S. cerevisiae*. *Nat. Commun.* **8**, 15202 (2017).
20. Hahn-Hagerdal, B., Galbe, M., Gorwa-Grauslund, M. F., Liden, G. & Zacchi, G. Bio-ethanol—the fuel of tomorrow from the residues of today. *Trends Biotechnol.* **24**, 549–556 (2006).
21. Walfridsson, M., Anderlund, M., Bao, X. & Hahn-Hagerdal, B. Expression of different levels of enzymes from the *Pichia stipitis* XYL1 and XYL2 genes in *Saccharomyces cerevisiae* and its effects on product formation during xylose utilisation. *Appl. Microbiol. Biotechnol.* **48**, 218–224 (1997).
22. Pereira, F. et al. Yeast pathway kit: a method for metabolic pathway assembly with automatically simulated executable documentation. *ACS Synth. Biol.* **5**, 386–394 (2016).
23. Verheoven, M. D. et al. Mutations in PMR1 stimulate xylose isomerase activity and anaerobic growth on xylose of engineered *Saccharomyces cerevisiae* by influencing manganese homeostasis. *Sci. Rep.* **7**, 46155 (2017).
24. Moyes, D. N., Reis, V. C., de Almeida, J. F., Costa, L. M. & Torres, F. A. Xylose fermentation by *Saccharomyces cerevisiae* and its effects on product formation during xylose fermentation. *Microb. Biotechnol.* **9**, 1067–1086 (2016).
25. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
26. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
27. Horak, C. E. et al. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16**, 3017–3032 (2002).
28. Zakian, V. A., Brewer, B. J. & Fangman, W. L. Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase. *Cell* **17**, 923–934 (1979).
29. Feng, X. & Zhao, H. Investigating host dependence of xylose utilization in recombinant *Saccharomyces cerevisiae* strains using RNA-seq analysis. *Biotechnol. Biofuels* **6**, 96 (2013).
30. Moles, J. F. et al. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc. Natl. Acad. Sci. USA* **106**, 6477–6482 (2009).
31. Moskovitz, J., Berlett, B. S., Poston, J. M. & Stadtman, E. R. The yeast peptide-methionine sulfone reductase functions as an antioxidant in vivo. *Proc. Natl. Acad. Sci. USA* **94**, 9885–9889 (1997).
32. Toledano, M. B., Delaunay-Moisant, A., Oulton, C. E. & Igbarni, A. Functions and cellular compartmentation of the thioredoxin and glutathione pathways in yeast. *Antioxid. Redox. Signal.* **18**, 1699–1711 (2013).
33. Silva, A., Yang, H., Boeke, J. D. & Mathews, D. J. Freedom and responsibility in synthetic genomics: the Synthetic Yeast Project. *Genetics* **200**, 1021–1028 (2015).
34. Lee, M. E., DeLoache, W. C., Cervantes, R. & Duerer, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
35. Blount, B. A., Driessen, M. R. & Ellis, T. GC prep: fast and easy extraction of stable yeast genomic DNA. *Sci. Rep.* **6**, 26863 (2016).
36. Antonio, R. V. & Creczynski-Pasa, T. B. Genetic analysis of vaccine biosynthesis by *Chromobacterium violaceum*. *Genet. Mol. Res.* **3**, 85–91 (2004).
37. Ledesma-Amaro, R. et al. Metabolic engineering of *Yarrowia lipolytica* to produce chemicals and fuels from xylose. *Metab. Eng.* **18**, 115–124 (2016).
38. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30**, 3399–3401 (2014).

**Acknowledgements**
We thank Leslie Mitchell and Jef Boeke for plasmid pC104, David Bell and Ali Awan for assistance with penicillin yield measurements, Young-Kyoung Park and Soo Me Chee for assistance with HPLC measurements, John Heap for advice on pathway redesign and Suaiji Kumar and Henry Lee for discussions regarding nanopore sequencing. This work was funded in the UK by BBSRC awards BB/K019791/1 and BB/1P504579/1 and funded in China by the Ministry of Science and Technology of China (2014CB745100 and 2015DFB00960) and the National Natural Science Foundation of China (21390203 and 21621004).

**Author contributions**
B.A.B. and T.E. conceived and designed the experiments, B.A.B., G.O.F.G., J.C.H.H., R.L.-A. and D.I. performed the experiments, B.A.B., J.C.H.H., G.O.F.G. and R.L.-A. analysed the experimental data, G.O.F.G., R.M.McK and T.E. analysed the sequencing data, Z.X.X., B.Z.L. and Y.J.Y. generated the synV strain, B.A.B., G.O.F.G. and T.E. prepared the manuscript.

**Additional information**
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03143-w.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/.

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