Improved betulinic acid biosynthesis using synthetic yeast chromosome recombination and semi-automated rapid LC-MS screening

G.-O.F. Gowers, S.M. Chee, D. Bell, L. Suckling, M. Kern, D. Tew, D.W. McClymont & T. Ellis

Synthetic biology, genome engineering and directed evolution offer innumerable tools to expedite engineering of strains for optimising biosynthetic pathways. One of the most radical is SCRaMbLE, a system of inducible in vivo deletion and rearrangement of synthetic yeast chromosomes, diversifying the genotype of millions of Saccharomyces cerevisiae cells in hours. SCRaMbLE can yield strains with improved biosynthetic phenotypes but is limited by screening capabilities. To address this bottleneck, we combine automated sample preparation, an ultra-fast 84-second LC-MS method, and barcoded nanopore sequencing to rapidly isolate and characterise the best performing strains. Here, we use SCRaMbLE to optimise yeast strains engineered to produce the triterpenoid betulinic acid. Our semi-automated workflow screens 1,000 colonies, identifying and sequencing 12 strains with between 2- to 7-fold improvement in betulinic acid titre. The broad applicability of this workflow to rapidly isolate improved strains from a variant library makes this a valuable tool for biotechnology.
Modifying cells to improve production of a desired endogenous or heterologous metabolite is a broad aim for many areas of academic and industrial biotechnology and biosciences. Classic approaches to metabolic engineering use tools such as knockout libraries and flux-balance analysis, among others, to predict rational changes to both the biosynthetic pathway and the native metabolism of the host organisms that will improve metabolite titre. Synthetic biology now offers many tools to rapidly modify pathways and genomes to optimise metabolite titres. For example, modular DNA pathway construction methods are ubiquitous and now available for a variety of microorganisms and biosciences. The nascent field of synthetic genomics now allows entirely new chromosome designs that introduce new tools for genome optimisation for metabolic engineering purposes. The synthetic yeast project (Sc2.0) aims to synthesise strains in biosynthesis of heterologous metabolites, including strains engineered with pathways for violacen and for carotene and lycopene.

These past studies all improved phenotypes that either affect growth or elicit a colourful phenotype allowing strains to be selected or screened with relative ease. However, the vast majority of metabolites of interest to industry exhibit no easily selectable or screenable phenotype. Instead, mass spectrometry coupled to chromatography is the gold standard for quantifying production of industrially relevant metabolites like these, but the time required for each sample is significant, limiting its use for screening large strain libraries that are quickly generated by methods like SCRaMbLE. In this study we aim to overcome this bottleneck by employing rapid and semi-automated technologies into a streamlined workflow that is broadly applicable to many metabolic engineering applications, using strain diversification by SCRaMbLE as an ideal test case.

Notably, our previous work has indicated the SCRaMbLE approach is ultra-fast LC-MS, a modified version of a recent technique developed for therapeutic drug monitoring in patient blood plasma. This uses only a guard column in place of the standard liquid chromatography analytical column and allows each sample to be processed sevenfold faster than with standard LC-MS. Following LC-MS screening we utilise multiplex nanopore sequencing to quickly resolve selected strain genotypes and use this information to determine genetic changes that contribute to improved performance. This workflow is designed to complement the speed of SCRaMbLE by quickly isolating top performing strains and characterising these genetically as well as for the product of interest.

**Results**

**Ultra-fast LC-MS screening of 1000 strains.** Our post-SCRaMbLE screening workflow consists of semi-automated colony picking and sample preparation, ultra-fast LC-MS measurement, and multiplexed nanopore sequencing. To assess our workflow, we first constructed and verified a heterologous pathway for BA biosynthesis in *S. cerevisiae*. Biosynthesis of BA at a detectable titre was achieved by introducing three heterologous genes (*AILUSI*, *AATRI1*, and *BPL0*) and overexpressing *ERG9* and a truncated version of *HMG1* (*hHMG1*). This was constructed in a yeast strain containing a synthetic chromosome V (*synV*), with 176 LoxPsym recombination sites throughout the chromosome. The resulting strain, yGG066, also containing synV, produced BA to 0.28 µg/gCDW when grown under the conditions used in this study. To trigger SCRaMbLE, yGG066 was transformed with a Cre recombinase expression plasmid (pSCW11-creEBD, *LEU*<sup>+</sup>). SCRaMbLE was then induced for 4 h by addition of 1 µM beta-estradiol, before cells were washed and plated. After 3 days of growth on URA-agar, 70 uninduced control colonies and 964 induced colonies were picked by an automated PIXL robot (Singer Instruments) taking under 15 min per 96-well plate. Colonies were picked into 11-deep well 96-well plates containing 500 µl SDO URA-*media* and grown overnight with shaking. This was used to reinoculate 11 plates containing fresh media (1:100 dilution) with the remaining culture stock at −80 °C in glycerol. The inoculation was grown for 2 days prior to solvent extraction of BA. An automated protocol for solvent extraction of BA was developed using a CyBio FeliX robot, which reduced the time of extraction to under 30 min per plate and removed variation due to human error.

The gold standard for BA detection has previously been LC-MS, but as this takes ~5 min per sample, screening 1000 strains would take more than two standard working weeks (~83 h). To tackle 1000 samples we adapted an ultra-fast LC-MS method that only uses a guard column as opposed to a standard chromatography column. The guard column provides sufficient chromatographic separation for measurement with a retention time of only 40 seconds for BA, compared to ~4.8 min for conventional LC-MS. With a runtime of 72 s with 12 s of injection time, the guard column reduces the time for screening from ~5 min per well (~8 h per 96-plate) to 84 s per well (<2.3 h per 96-plate) offering a throughput advantage of ~4-fold (Fig. 1b). This rapid method maintains comparable ion separation, even for complex samples without filtering (Supplementary Fig. 5), and has sufficient assay robustness with minimal sample drift even over 11 plates (Supplementary Fig. 1).

Total cell growth was measured prior to BA extraction to normalise for any growth differences that may be caused by SCRaMbLE. BA titres were calculated using BA standards spiked into negative control biological samples (10 µg/ml – 1 ng/ml) and normalising by cell dry weight (inferred from OD<sub>600</sub>). A threshold for significant titre improvement was set 4 standard deviations above the mean control values (Fig. 1c). A total of 12 screened strains exceeded this threshold. These strains all grew slower than the pre-SCRaMbLE controls (even when cured of the
BA-production plasmid) but produced more total BA even when not normalised by OD600 (Supplementary Figs. 2 and 3).

**Multiplexed nanopore sequencing of top SCRaMBLE strains.** The top 12 BA-producing strains (BC01-12) identified by the screen were selected from the glycerol stock plates. All strains were checked for any remaining recombinase plasmid, and in two strains where this was found (BC01 and BC03), the plasmid was removed by curing before further work. The top 12 strains were then re-analysed with ultra-fast LC-MS (n = 7) exactly as before to better characterise their improved BA titre phenotypes. All 12 strains did indeed show significantly higher BA titres of between two- and sevenfold above that of the yGG066 control (Fig. 2a). Despite the use of plasmid based-pathway expression in these strains, BA biosynthesis was still maintained over 6 days of continuous culturing (BC01, 02, 03, 07, and 11 tested), with only 1 in 30 isolates losing BA biosynthesis (Supplementary Fig. 7). Although the auxotrophic selection of the plasmid used in this study requires no addition of chemicals, genomic integration of the pathway genes would be more desirable for future industrial use.

To determine the genotypic cause of the improved phenotypes it is necessary to map the chromosomal rearrangements of the synthetic chromosome in each strain. Oxford nanopore sequencing yields long reads ideal to bridge chromosomal rearrangements that could be missed or difficult to infer by short read sequencing28. Therefore, we next used a barcoding approach to multiplex all 12 strains and sequenced these together on a single MinION nanopore flow cell. In total, we achieved 11.5 Gbp of reads which, after barcode deconvolution, resulted in an average of 886 Mbp of read data per strain, representing 73.2-fold coverage of the yeast genome (Fig. 2b). Raw reads were aligned to the reference synV chromosome using a LASTAL algorithm in order to quickly identify SCRaMBLE events (Supplementary Fig. 4). Many strains exhibited complex rearrangements involving combinations of deletions, inversions, and duplications, in some cases with regions over 15 kb in length rearranged (Supplementary Data 4). Notably, no obvious rearrangement was detected by our sequencing approach for three of the strains. For these we
speculate that natural mutations elsewhere in the genome may also have yielded beneficial improvements. Our highest performing strain (BC03) is capable of producing ~7-fold more BA per cell than the pre-SCRaMbLE control strain in 500 µl shaking cultures (Fig. 2a) and ~3-fold more in 50 ml shaking flasks over 48 h (Supplementary Fig. 8). Sequencing of this strain revealed a complex combination of four SCRaMbLE rearrangements on synthetic chromosome V. Two non-coding deletions of 551 and 391 bp, respectively, were seen alongside a 10192 bp inversion and 8724 bp duplication (Fig. 2c, left). In contrast, BC11 (right) exhibits only a single non-coding deletion of 723 bp. Source data are provided as a Source Data file.

A genotype-to-phenotype relationship is established. To verify a genotypic cause for increased BA titres we sought to artificially recreate a SCRaMbLE event in a control strain. While BC03 was our top producer, our previous efforts have found it extremely challenging to use genome engineering to recreate large inversions and duplications. Such work may also be of limited value when looking to establish a specific genotype-to-phenotype relationship, as it is likely that only a subset of the affected genes is responsible for the improved titre. Instead, for simplicity, we chose to re-create the single small non-coding deletion in BC11 that only removes the 3′ untranslated region of TIR1 (Fig. 3a; confirmed by PCR in Supplementary Fig. 6).

Quantitative PCR (qPCR) analysis was first used to determine the effect of this deletion on the expression of surrounding genes. At mid-exponential phase TIR1 exhibits 42% lower expression, while PRE1 exhibits 67% higher expression compared to the pre-SCRaMbLE control strain (Fig. 3b). Next, we used CRISPR-based genome engineering to artificially re-create this deletion in the control yGG066 strain, removing the TIR1 3′ UTR region in synV as seen in the BC11. The re-created deletion was confirmed by PCR (Supplementary Fig. 6) and this strain, ‘yGG066 Δ(TIR1)’, was subject to LC-MS characterisation exactly as before along with the pre-SCRaMbLE yGG066 strain (negative control) and BC11 (positive control) (Fig. 3c). We observe the same BA titre as in BC11, confirming that this single SCRaMbLE event is responsible for the threefold increase in BA biosynthesis. Revealing this non-intuitive strain improvement was only made possible due to the workflow devised here that links post-SCRaMbLE high throughput LC-MS screening with multiplexed long read sequencing (Fig. 3d).

Discussion

Rational strain engineering approaches focus on systematically up- or down-regulating endogenous genes, typically informed by mathematical modelling. Inherent in this approach is bias based on prior knowledge of metabolic networks. An alternative, and complementary, approach is to employ whole-genome mutation strategies and other ‘black-box’ methods that generate unpredictable alterations to the strain genotype. SCRaMbLE is one of the most drastic of these, offering a quick and simple way to introduce a range of genomically diverse deletions to a strain without compromising robustness and accuracy. This workflow offers improved throughput and greater specificity, allowing for a much more efficient screening process. Furthermore, this workflow does not suffer from the bias of genomic regions already containing or lacking the barcode sequences, thus allowing for higher throughput. The use of ultra-fast LC-MS and multiplexed nanopore sequencing demonstrates the superior sensitivity of this workflow and provides an opportunity for high-throughput genotyping of strains. The simplicity of the genome engineering strategy also allows for the generation of a diverse range of deletions with improved or desired properties.
was designed to accelerate strain characterisation post-SCRaMbLE, where typically only 1% of colonies show improvement. However, it could easily be applied to other black-box methods that diversify strain genotypes, such as the mating-based methods commonly used with budding yeast.

Interestingly, black-box methods like SCRaMbLE often reveal unconventional strategies to improve biosynthesis performance, and these can greatly help inform further rational design. In BC11, we identified that a small non-coding DNA deletion that lowers expression of TIR1 and increases PRE1 expression can boost BA titres by threefold, a non-obvious finding. Tir1p is a cell wall mannoprotein and previous studies have shown that lowered expression of TIR1 is a stress response to both cold- and heat-shock. We suspect that BA accumulates at the cell wall and therefore any changes to the cell wall structure may affect the accumulation of BA. Pre1p is the beta subunit of the 20S proteasome in yeast and is associated with the endoplasmic reticulum. Further larger scale changes were observed in other SCRaMbLE strains leading to increased BA production (Supplementary Data 4). These genes are unlikely to have been identified in a rational search for candidate genes for systematic modification, but the increased BA titre in BC11 persisted when we recreated this deletion in the control strain, demonstrating it to be solely responsible for the phenotype change.

Central to the success of the workflow devised here is the ultra-fast LC-MS guard column method, which allowed us to characterise samples every 84 s, fourfold faster than when using a standard chromatographic column. Unlike other rapid MS techniques such as Rapid-Fire (Agilent), the method presented here requires no additional equipment or infrastructure. Furthermore, unlike the method presented here, other MS techniques such as REIMS or MALDI-TOF-MS lack any chromatographic separation and so may struggle to resolve some metabolites. We anticipate that the ultra-fast LC-MS method used in this study could replace many standard LC-MS workflows seeking to quantify a broad range of metabolites.

The use of automation with this method allowed us to pick, grow, extract, and analyse 1000 strains in 26 h with minimal input from the user, ensuring the screen was robust, rapid, and bias-free. Scaling this to cover larger library sizes would simply require scaling the timescale of this screening or increasing multiplexing through automation capabilities. Crucially, our workflow requires no significant new instrumentation and is therefore likely to be immediately implementable at the increasing number of Bio-foundries found around the world.

**Methods**

**Strains and media.** *Escherichia coli* Turbo Comp cells (NEB) were used for standard bacterial cloning and plasmid propagation with growth in Luria Bertani media (VWR). Yeast strain yXZX846 (BY4741 with synthetic chromosome V, here referred to as synV) was generated previously. synV yeast was grown in YPD (10 g l\(^{-1}\) yeast extract, 20 g l\(^{-1}\) peptone, 20 g l\(^{-1}\) glucose) or SDO (6.7 g l\(^{-1}\) yeast nitrogen base, 1.4 g l\(^{-1}\) yeast synthetic drop out medium supplement without histidine, leucine, tryptophan and uracil, 20 g l\(^{-1}\) glucose) supplemented with histidine (20 mg l\(^{-1}\)) and leucine (120 mg l\(^{-1}\)), tryptophan (20 mg l\(^{-1}\)), and uracil (20 mg l\(^{-1}\)) as necessary. Two per cent bacteriological agar (VWR) was added to YPD or SDO media when necessary. Unless otherwise stated, all media components were supplied by Sigma Aldrich. SCRaMbLEd yeast strains were stocked in glycerol to a final concentration of 25% (v/v) at −80 °C. OD\(_{600}\) measurements were performed using a Synergy HT (BioTek) microplate reader in clear bottom 96-well plates.

**Plasmids and yeast transformation.** Yeast strain yGG037 (synV with AtATR1 integrated at the HO locus with HIS marker) was constructed by transforming synV with plasmid pGG127 (Supplementary Fig. 1, Supplementary Data 2).
linearised by NotI digestion (NEB), followed by selection on HIS– SDO agar. yGg066 (BA-producing control strain) was transformed by transforming yGg037 with pGg052 (Supplementary Fig. 1; Supplementary Data 2) followed by selection on URA– SDO agar. pGg127 and pGg052 were constructed using the MoClo Yeast ToolKit (YTK). DNA parts not already in the original YTK collection were synthesised by GeneArt (listed in Supplementary Data 3). pScW11-creEBD was generated in a previous study.21

SCRAMBLE yGg066-pSCW11-creEBD was grown overnight in SDO URA– LEU– media (30 °C, 250 r.p.m. shaking). Culture was diluted to OD600 0.2 in 5 ml SDO URA– LEU– media and grown for 4 h (30 °C, 250 r.p.m. shaking). Beta-estradiol was added to the experimental culture (SCRAMBLE) (final concentration 1 µM) and 1 µl ethanol to the control culture (pSCW11) (Supplementary Fig. 1, Supplementary Data 2) followed by centrifugation and resuspension in URA– selective media. Cells were diluted in SDO URA– media (10−4, SCRAMBLE culture and 10−5 pre-SCRAMBLE culture) and spread onto SDO URA–agar media. Plates were left at 30 °C for 3 days.

Automated colony picking. Cells were picked from agar dishes into 2 ml 96-deep well plates (Grenier Bio-One) containing 500 µl SDO URA– media using the PIXL automated colony-picking robot (Singer Instruments). 40 pre-SCRAMBLE control strains and 48 SCRAMBLE strains were picked into the first 96-well plate. Subsequent 96-well plates contained three pre-SCRAMBLE control strains each. Wells were reserved for a spiked standard curve on the first plate. Colonies were automatically identified by the PIXL imaging software (version 2.18.0920.01.12) prioritising minimum proximity to minimise the risk of picking contamination.

Automated extraction of BA. Prior to extraction colonies were picked into SDO URA– media and grown overnight (30 °C, 250 r.p.m. before being diluted 1:100 into fresh media and grown for 2 days (30 °C, 250 r.p.m.). OD600 measurements were taken and negative control wells (synV) were spiked with 10-fold dilutions of a BA standard (Sigma Aldrich) prepared in isopropanol (10 µg/ml–1 ng/ml).

Liquid handling steps were developed on CyBio® Composer (Analytik Jena) for execution on a FeliX robot (R 96/250 µl pipetting head, Analytik Jena). Cells grown in a fume hood and resuspended in 70 µl fresh IPA (2000 r.p.m., 2 min, Bioshake QInstruments) and centrifuged (5 min, 2250 r.c.f.). Supernatant was aspirated and transferred to a 96-well round-bottom plate (Grenier Bio-One) containing 500 µl SDO URA– media and grown for 4 h (30 °C, 250 r.p.m. shaking). Cells were washed twice by centrifugation (3220 r.c.f., 5 min) and 1 µl ethanol to the control culture (pre-SCRaMbLE) and both grown for 4 h (30 °C, 180 r.p.m.). Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Quantitative PCR. qPCR was performed in a MasterCycler ep Realplex 4 (Eppendorf) using a SYBR Green JumpStart qPCR kit. High Capacity cDNA RT kit (Thermo Scientific) was used to generate cDNA. Primers were designed using the IDT Primer Design Tool (http://www.idtdna.com/scitools/applications/design). Pre: Gg116 (CTTCAAGAGGAGTCAAAGG), Gg117 (TGGCTGAGAATGATCAGCCG); TIR1: Gg118 (AGCTGGTGTTTTGGATAATC), Gg119 (ACACGCAAAGTCAAACCTCAG); and housekeeping gene ACT1, Gg114 (TGCACAAGAAATGTGAAACGG), Gg115 (GGCGAGTTCACACCCCAAC).

References

1. Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. Cell 164, 1185–1197 (2016).
2. Liu, W. & Jiang, R. Combinatorial and high-throughput screening approaches for strain engineering. Appl. Microbiol. Biotechnol. 99, 2093–2104 (2015).
3. Jones, J. A. & Koffas, M. A. G. Optimizing metabolic pathways for the improved production of natural products. Methods Enzymol. 575, 179–193 (2016).
4. Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biotechnol. 79, 563–590 (2010).
5. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth. Biol. 4, 975–986 (2015).
6. 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).
7. Andreou, A. I. & Nakayama, N. Mobius assembly: a versatile golden-gate framework towards universal DNA assembly. PLoS ONE 13, 1–18 (2018).
8. Fonseca, J. P. et al. A toolkit for rapid modular construction of biological circuits in mammalian cells. ACS Synth. Biol. 8, 2593–2606 (2019).
9. Storch, M. et al. BASIC: a new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. ACS Synth. Biol. 4, 781–787 (2015).
10. Taylor, G. M., Mordaka, P. M. & Heap, J. T. Start-stop assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. Nucleic Acids Res. 47, e17 (2019).
11. Strickland, S. M. et al. Design of a synthetic yeast genome. Science 355, 1040–1044 (2017).
12. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. Science 344, 55–59 (2014).
13. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477, 471–476 (2011).
14. Dymond, J. & Boeke, J. The Saccharomyces cerevisiae SCRAMBLE system and genome minimization. Bioeng. Bugs 3, 168–171 (2012).
15. Ma, L. et al. SCRAMBLE generates evolved yeasts with increased alkali tolerance. Microb. Cell Fact. 18, 52 (2019).
30. Auxillos, J. Y. et al. Multiplex genome engineering for optimizing bioproduction. Nat. Commun. 9, 1–12 (2018).
31. Luo, Z. et al. Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. Nat. Commun. 9, 1–10 (2018).
32. Blount, B. A. et al. Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. Nat. Commun. 9, 1992 (2018).
33. Kempa, E. E., Hollywood, K. A., Smith, C. A. & Barran, P. E. High throughput screening of complex biological samples with mass spectrometry—from bulk measurements to single cell analysis. Analyst 144, 872–891 (2019).
34. Bolt, F. et al. Automated high-throughput identification and characterization of clinically important bacteria and fungi using rapid evaporative ionization mass spectrometry. Anal. Chem. 88, 9419–9426 (2016).
35. Zhang, L., Smart, S. & Sandrin, T. R. Biomarker- and similarity coefficient-based approaches to bacterial mixture characterization using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). Sci. Rep. 5, 1–10 (2015).
36. Milne, I. et al. Using tablet for visual exploration of second-generation sequencing data. Brief. Bioinformatics 14, 193–202 (2013).

Acknowledgements
We thank Christopher Benton (Agilent) for assistance in setting up ultra-fast LC-MS capabilities at SynbiCITE. We also thank Dr. Rochelle Aw for assistance in planning and running the qPCR experiment. We thank Prof. John Duerber for distribution of the MoClo YTKit and Dr. William Shaw for construction of pre-assembled YTKit vectors. This work was funded in the UK by BBSRC awards BB/P005791/1, BB/L027852/1, and EPSRC awards EP/S001859/1 and EP/L011573/1.

Author contributions
G.-O.F.G. conceived and designed the experiments. G.-O.F.G., L.S., D.B. and S.M.C. designed and performed sample preparation for, and execution of, LC-MS analysis. LC-MS data analysis was performed by G.-O.F.G., S.M.C. and D.B. G.-O.F.G. constructed strains and performed nanopore sequencing with subsequent analysis. M.K., D.T., T.E. and D.M. assisted in interpreting results. G.-O.F.G. and T.E. prepared the manuscript, while all other authors contributed to manuscript improvement.

Competing interests
GSK support G.-O.F.G. through a BBSRC industrial CASE studentship. Singer Instruments provided SynbiCITE with access to a PIXL robot for a limited time. No industrial contributors had input into initial study design nor on decision of when and where to publish this manuscript. All other authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-10708-z.

Correspondence and requests for materials should be addressed to T.E.

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

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.