UC Berkeley
UC Berkeley Previously Published Works

Title

A FAIR-compliant parts catalogue for genome engineering and expression control in Saccharomyces cerevisiae

Permalink

https://escholarship.org/uc/item/4061z0j4

Journal

Synthetic and Systems Biotechnology, 7(2)

ISSN

2405-805X

Authors

D'Ambrosio, Vasil
Hansen, Lea G
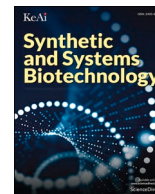Zhang, Jie
et al.

Publication Date

2022-06-01

DOI

10.1016/j.synbio.2022.02.001

Copyright Information

Peer reviewed

A FAIR-compliant parts catalogue for genome engineering and expression control in *Saccharomyces cerevisiae*

Vasil D'Ambrosio [a], Lea G. Hansen [a], Jie Zhang [a], Emil D. Jensen [a], Dushica Arsovska [a], Marcos Laloux [a], Tadas Jakočiūnas [a], Pernille Hjort [a], Davide De Lucrezia [b], Serena Marletta [b], Jay D. Keasling [a,c,d,e,f], Michael K. Jensen [a,*]

[a] *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark*
[b] *Explora Biotech Srl, Italy*
[c] *Joint BioEnergy Institute, Emeryville, CA, USA*
[d] *Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[e] *Department of Chemical and Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, CA, USA*
[f] *Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes of Advanced Technologies, Shenzhen, China*

ARTICLE INFO

ABSTRACT

The synthetic biology toolkit for baker's yeast, *Saccharomyces cerevisiae*, includes extensive genome engineering toolkits and parts repositories. However, with the increasing complexity of engineering tasks and versatile applications of this model eukaryote, there is a continued interest to expand and diversify the rational engineering capabilities in this chassis by FAIR (findable, accessible, interoperable, and reproducible) compliance. In this study, we designed and characterised 41 synthetic guide RNA sequences to expand the CRISPR-based genome engineering capabilities for easy and efficient replacement of genomically encoded elements. Moreover, we characterize in high temporal resolution 20 native promoters and 18 terminators using fluorescein and LUDOX CL-X as references for GFP expression and OD600 measurements, respectively. Additionally, all data and reported analysis is provided in a publicly accessible jupyter notebook providing a tool for researchers with low-coding skills to further explore the generated data as well as a template for researchers to write their own scripts. We expect the data, parts, and databases associated with this study to support a FAIR-compliant resource for further advancing the engineering of yeasts.

1. Introduction

One of the primary goals of synthetic biology is to establish a robust and scalable methodological framework for the forward engineering of biological systems. Standardization plays an instrumental role in fulfilling this vision as standards provide an effective way to build upon previous results by accelerating knowledge transfer and facilitating innovation. Data standardization allows practitioners to leverage data generated by others to shorten the design-build-test-learn (DBTL) cycle. Although the benefits of adopting standards are clear, the scientific community recognizes the lack of widely accepted standards as one of the major challenges in the field, with many initiatives (e.g. BioRoboost, Joint Institute for Metrology in Biology, SynbioLEAP, and COMBINE) promoting the adoption of specific standards in different areas across the DBTL cycle [1–5].

A great effort was required to establish fundamental standards in synthetic biology, and more specifically to introduce harmonized measurement units, by rescaling output data against reference material to normalize values [6–8], or to convert arbitrary units to absolute values in for instance flow cytometry data analysis [9]. Despite the wealth of data, there is a clear shortage of methodological standards to such an extent that 77% of biologists stated that they had tried and failed to

reproduce someone else's result in the lab [10]. Indeed, experimental protocols are often specific to a laboratory or even to an individual researcher, and are made available to the community mostly through condensed sections covering methods and materials in scientific publications. In this regard, public protocol repositories (e.g. GitHub, Protocol.io, Benchling, JoVE) and large scale community-validated protocols [4] are valuable resources to provide a standardized and traceable framework for methodological standards. Finally, efforts in establishing reproducible protocols and robust normalization procedures are seriously jeopardized if different researchers and research groups cannot access reliable specification standards for the execution of the procedures. Reference materials are specifically designed to tackle this problem as they are manufactured to have specified properties, such as homogeneity and stability, to be suitable for use in a measurement process. Reference material can be sourced from non-certified vendors able to ensure material availability and continuity, or from certified bodies such as The National Institute for Biological Standards and Control (NIBSC) or the National Institute of Standards and Technology (NIST). A pioneering example of the use of reference material to standardize measurement across labs was the design of "Universal RNA Standards", a pool of synthetic sequences not expressed in any known genome to be used to control for gene expression [11].

The work presented in this paper specifically addresses the development and validation of a standardization framework for the characterization of regulatory DNA parts and guide-RNA for genome engineering in yeast. We have established fundamental, methodological, and specification standards for the *in vivo* characterization of regulatory DNA parts in *S. cerevisiae* CEN.PK2–1C. The framework comprises a definition of fundamental standards for the normalization of measurements and related protocols supplied with a list of reference materials widely available worldwide to execute the protocols. Whenever possible, reference materials for the construction of calibration curves and execution of quality control are provided in the SOP's description. Likewise, instructions for data collection and processing methods are provided whenever possible. In addition, the data generated are made available through the FAIR-compliant [12] online repository. In particular, public DNA parts and constructs are "findable" through both the webapp (https://doulix.com/public-biomodules/and https://doulix.com/public-constructs/, respectively) using Doulix's query tool by both logged and unregistered use and restful API (https://doulix.com/api/docs/#tag/Biomodule and https://doulix.com/api/docs/#tag/Full-Length-Construct). DNA parts are also accessible via Identifiers.org, just as DNA parts are "accessible" as the repository is free of charge and does not require account creation to query the database. Lastly, DNA parts are also "interoperable" as data can be exported as.fasta,.gb and support the SBOL format, and "reusable" as users can freely make a personal copy of the digital record for further editing.

2. Materials and methods

2.1. Cultivation media and conditions

Chemically competent *Escherichia coli* DH5α strain was used as a host for cloning and plasmid propagation. The cells were cultivated at 37 °C in 2xYT supplemented with 100 μg/mL ampicillin. The *Saccharomyces cerevisiae* strains used in this study were grown at 30 °C and 250 rpm in synthetic complete (SC) medium (6.7 g/L yeast nitrogen base without amino acids with appropriate drop-out medium supplement) supplemented with glucose 20 g/L as carbon source.

2.2. Plasmids and strains construction

All plasmids used in this study were assembled by USER™ (uracil-specific excision reagent) cloning (New England Biolabs). Biobricks used to assemble the plasmids were amplified by PCR using PhusionU polymerase (Thermo Fisher Scientific). *S. cerevisiae* strains were transformed by the lithium acetate/single-stranded carrier DNA/PEG method previously described [13]. The complete list of strains and plasmids used in this study is provided in Supplementary Tables S1 and S2.

2.3. gRNA selection and integration efficiency

Six-hundred random DNA sequences were generated with the sample function in R. Following recommendation from Takera Bio (Takerabio.com) the base at position 1 was changed to G and the base in position 17 was changed to T. Next sequences were filtered based on CG content, removing sequences where GC % < 45% or >60%. The online version of Cas-OFFinder was used to search for potential genomic target sites (Bae et al., 2014). All sequences predicted to target anywhere in the genome of *S. cerevisiae* (S288C) with 2 or fewer mismatches were removed. Of the remaining sequences 41 were randomly selected for testing with AGG as PAM site by introducing them into a pESC based vector under the control of SNR52 promoter and SUP4 terminator. Next, a gRNA cassette containing any one of the 41 target sequences was fused to an mKate2 expression cassette and introduced in a CEN.PK2–1C background strain.

A replacement cassette containing yeGFP controlled by pTDH3 was amplified to have 35bp homology for the terminators used to assemble the mKate2 expression cassette. To transform the cells 1000 ng of gRNA plasmid in combination with 500 ng of replacement cassette were used. The efficiency, expressed as the ratio between green colonies (successful integration) and red colonies (failed/no integration) was calculated (Fig. 3C).

2.4. Fluorescence measurements

All the tested promoters and terminators were integrated in a CEN.PK2–1C background strain in the EasyClone site X-2 [14] together with an expression cassette for super-folder GFP (sfGFP). The constructed strains were grown in a 96-deep well format for 24 h in SC medium. The samples were then diluted 1:5 in fresh medium and grown for 24 h. Finally each sample was diluted 1:100 and cultivated at 30 °C and 1000 rpm in a Bioshake 3000-T elm shaker (Qinstruments) on a custom built Tecan Evo platform. Every 2 h 5 μL of the samples were transferred to a 384-well plate and diluted to a final volume of 25 μL. The OD and fluorescence intensity were then measured on a Tecan Infinite 200 Pro plate reader (Tecan). The parameters used to measure the fluorescence intensity and OD can be found in the Supplementary Tables S3 and S4. Fluorescein and LUDOX CL-X (45% colloidal silica suspension) used in this study were obtained by Sigma-Aldrich. Fluorescein intensities, calibration curves and sequences of characterised promoters and terminators are provided as separate supplementary files.

2.5. Data analysis and data access

Raw data generated during this experiment was processed using in-house R and python scripts. The output for each time point measurement consists of a single raw data file. The Tidyverse R package was used to unify all the files into a single dataframe and to remove background fluorescence and absorbance values [15]. Next, the following python packages were used to perform data analysis and data visualisation described in the jupyter notebooks: Numpy, pandas, matplotlib, seaborn, SequenceMatcher, Biopython and Scikit-Learn [16–22].

3. Results

3.1. Promoter characterisation

Saccharomyces cerevisiae is engineered for widely different health and biotech applications, including metabolic engineering to support the biosynthesis of increasingly complex molecules through the expression of elaborate heterologous biosynthetic pathways [23–25]. These

pathways can cause additional metabolic burdening to the host by competing for resources with the native yeast metabolism or by the production of toxic intermediates [23,26]. One method to mitigate this is to temporally separate the expression of the heterologous pathway from growth. However, many promoters commonly used to construct complex pathways are glycolytic or constitutive [23,27–29]. Thus, there is a need for the characterisation of more promoters that are active in different growth stages. With this in mind, 20 native yeast promoters were selected based on transcriptome data acquired during the lag, mid-exponential and post-exponential phases [23]. Among these, 9 promoters have previously been characterized in different growth conditions [28–31], and chosen herein as references. Next, we measured promoter strengths using reporter assays in a 24-h interval in Synthetic Complete (SC) medium (Fig. 1A). SC medium was preferred to other commonly used media such as YPD because of lower background fluorescence, and because of its defined composition (Supplementary Fig. S1). To enable direct comparison between different experiments and operators, and to improve reproducibility, measured fluorescence intensities were normalized to known amounts of fluorescein, a reference

fluorophore for fluorescence measurements. Furthermore, provided a single point measurement for LUDOX CL-X (45% colloidal silica suspension) which provides a conversion factor for absorbance that can be used to compare the absorbance values with other instruments (Supplementary Table S4) [32].

To provide expression data for different growth stages and to allow comparison with previously published parts characterisation and transcriptomics data we initially focused on measurements at 4, 8 and 24 h [23,28,29,31] (Fig. 1B–D, Supplementary Fig. S2). In line with previous findings, we observed that pTDH3 was the strongest promoter across the majority of tested time-points, reaching a maximum fluorescence intensity of 108.2 ng of fluorescein per unit of OD600 (FvOD) following 8 h (Fig. 1C) [28,31]. Similarly, promoters often used in metabolic engineering efforts (e.g. pPGK1, pTEF1, pENO2, pTEF2) were also highly expressed at 4 and 8 h, and their normalized expression strongly correlated with previous findings (Fig. 1B–C, Supplementary Fig. S3). Furthermore, at the 8 h measurement, pTDH2 is the only promoter that doubles its intensity compared to 4 h, ultimately reaching comparable expression values to pTEF1. Moreover, the YPR036W-A promoter,
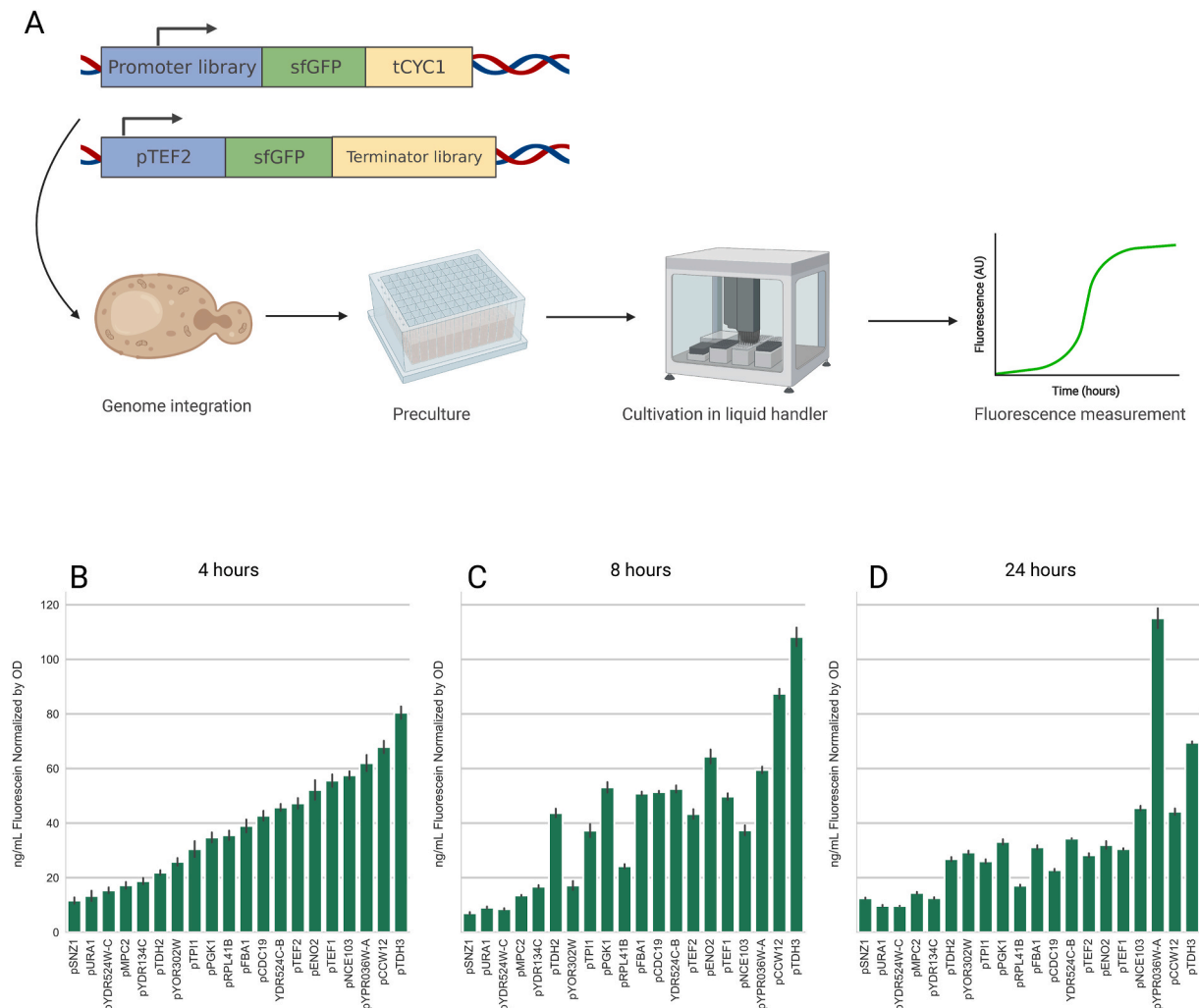

Fig. 1. Experimental designs and fluorescence intensity of the characterised promoters.
(A) The tested designs were integrated in *S. cerevisiae*
and grown in a 96 deep-well plate format. Next they were transferred to a liquid handler and the fluorescence was evaluated every 2 h
(B) Fluorescence intensity values after 4 h
of cultivation. (C) Fluorescence intensity values after 8 h
of cultivation. (D) Fluorescence intensity values after 24 h
of cultivation.The values are expressed as ng/mL of fluorescein per OD600 absorbance value. Each bar represents the average of four (n = 4) technical replicates with error bars representing mean ± standard deviation.

involved in sporulation in *S. cerevisiae*, showed a strong increase in sfGFP fluorescence intensity in the 24-h measurement compared to 4 h and 8 h measurements, reaching a fluorescence value of 115 ng/ml of fluorescein per unit of OD600, higher than the maximum pTDH3 measurement (Fig. 1A and D) [33]. Vice-versa, the five weakest promoters at 4 h measurement, remained the weakest promoters throughout the 24 h sampling window (Fig. 1B–D).

Finally, we compared the fluorescence per OD600 data to the transcriptomics dataset from Rajkumar et al. [23], and observed a correlation coefficient >0.9 for promoters pCDC19, pTEF2, pCCW12, pYDR524C–B, pTEF1, pYPR036W-A, pYOR302W, and pTPI1 (Supplementary Table S5), spanning both promoters with medium and strong activity.

In summary, the tested promoters spanned 6.3–115 ng of fluorescein per unit of OD600 across a 24 h sampling window, with pYPR036W-A and pSNZ1 observed to yield the highest and lowest fluorescein counts (Fig. 1C). Furthermore, fluorescein data for several medium and high strength promoters correlated well with mRNA levels independently obtained in yeast cultivated in different media.

3.2. Terminator characterisation

Terminators are genetic parts capable of altering the level of gene expression by affecting mRNA half-life, polyadenylation and translational efficiency, and it has been previously reported that they can alter gene expression [34,35]. Furthermore, terminators have not been characterised to the same extent as promoters, and often only few terminators, such as tCYC1, tADH1 or tPGK1, are selected per default [34, 35].

Here, we selected and characterised 18 terminator sequences, either native or synthetic, and assessed their effect on fluorescence intensity when coupled to sfGFP expressed under the control of pTEF2 (Fig. 1A). The different terminators were selected based on existing literature to systematically investigate their impact on sfGFP expression controlled by pTEF2 [36–39]. From this screen we observed a 5.6-fold difference in fluorescence after 4 h of cultivation between t3 and tVPS13, the weakest and strongest terminator, respectively (Fig. 2A). Furthermore, after 8 h of cultivation, terminators mostly maintained the same relative strength as in the 4 h measurement (Fig. 2B). This is, in contrast to the temporal expression profile investigated for several of the promoters (Fig. 1B–D).

Similarly to the 8 h measurements, the terminators' relative strength remained mainly unchanged for the 24 h measurement (Fig. 2C).

In summary, to increase the available toolkit of terminators we temporally characterized 18 terminator sequences that support regulation of expression in a range of 5.6-fold. This is in line with previous reports stating that terminators have a smaller effect on protein expression compared to promoters. However, the strength of terminators also depends on the promoter they are coupled to. Previous results show that terminators can alter the expression of weak promoters to a greater extent compared to strong or inducible promoters [34–36,38].

3.3. Characterization of synthetic gRNAs

In the last decade CRISPR/Cas has become an indispensable tool for molecular biologists [40–42]. Several Cas9-based tools have been established to speed up yeast strain construction by providing gRNA sequences for native regions or by developing synthetic landing pads [43–45]. Existing systems, however, rely on the previous integration of synthetic landing pads and do not allow for an easy replacement of previously integrated constructs.

To allow for easy replacement of complete or partial integration cassettes we designed and tested 41 synthetic gRNA sequences (see Materials and Methods) which can be fused to integrated genes or promoters of interest, thus allowing for easy replacement of a previously integrated DNA part (Fig. 3A–C). To assess the efficiency of the synthetic gRNA sequences we integrated a landing pad containing the target sequence for all the synthetic gRNAs fused to a mKate2 expression cassette in the genome of *S. cerevisiae* (Fig. 3C). Next, we tested their efficiency in a reporter-replacement assay. Here the mKate2 gene and its promoter were sought to be replaced with a GFP expression cassette upon transformation of a plasmid expressing the synthetic gRNA into a strain expressing Cas9. By scoring the numbers of colonies expressing mKate2 and GFP we calculated the genome engineering efficiency (Fig. 3D, Supplementary Table S6). In addition to the 41 synthetic gRNAs, 9 gRNA sequences targeting native genomic loci were also assessed by simple readings of integration efficiencies of a yeGFP cassette. Next, after comparing the similarity score with the Sequence-Matcher package to ensure high variability between the sequences, we compared the characteristics of the best and worst performing gRNA sequences to uncover their effect on gRNA efficiency (Figure 3B, 3D-E,
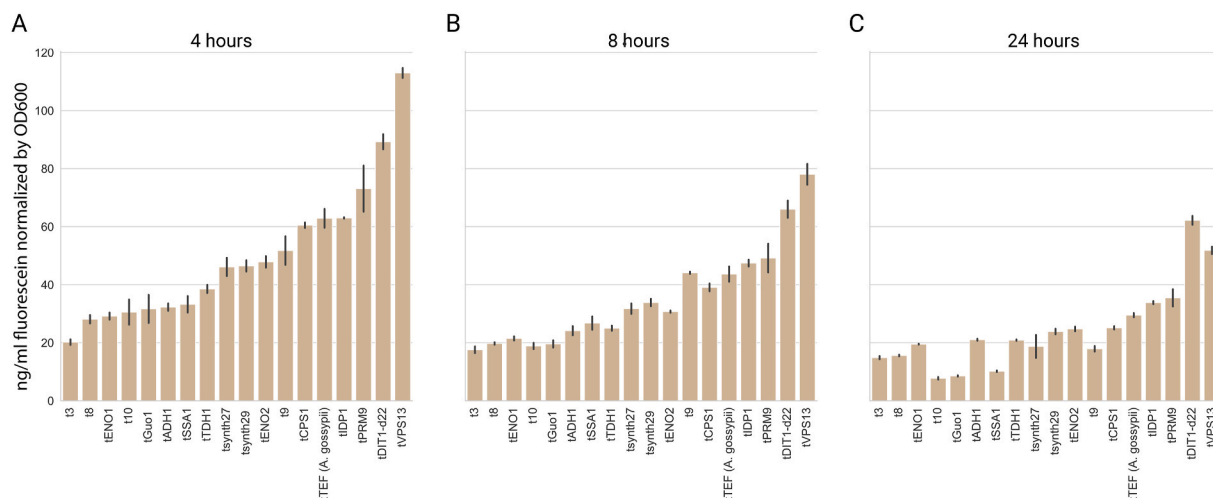

Fig. 2. Fluorescence intensity of the characterised terminators.
(A) Fluorescence intensity values after 4 h
of cultivation. (B) Fluorescence intensity values after 8 h
of cultivation. (C) Fluorescence intensity values after 24 h
of cultivation.The values are expressed as ng/mL of fluorescein per OD600 absorbance value. Each bar represents the average of four (n = 4) technical replicates with error bars representing mean ± standard deviation.
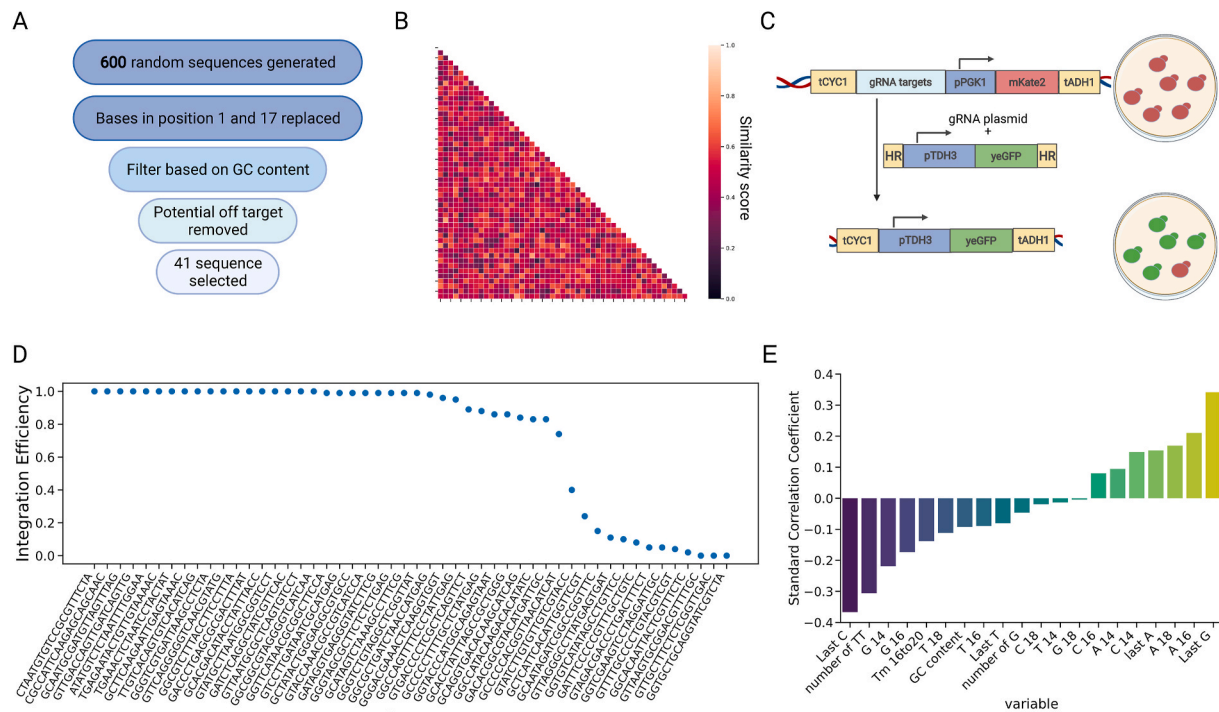
Fig. 3. Experimental design and results for synthetic gRNA sequences.
(A) Selection criteria for the 41 randomly generated gRNA sequences tested in this study (see Methods 2.3). (B) Sequence similarity scores between the tested gRNA sequences.Replacement strategy used to evaluate efficiency. (C) Yeast colonies containing mKate2 (red) are transformed with the tested gRNA plasmid in combination with the replacement cassette. Successfully transformed colonies are green while unsuccessful transformants are red. (D) gRNA efficiency, expressed as the ratio between successful and unsuccessful integrations, for all the tested sequences. (E) Standard correlation coefficient plot for the tested variables compared to genome engineering efficiency. . (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Supplementary Fig. S4, Supplementary Figs. S5A–E).

By doing so, we observed that the last base is strongly correlated to gRNA efficiency (Supplementary Fig. S4). More precisely, guanine (G) is the most common base in this position for the best performing gRNA sequences (34%). Also, none of the worst performing gRNA sequences ends with a G, which indicates that a G in position 20 favours efficiency. On the other hand, cytosine (C) is the most abundant base in the worst performing sequences (Supplementary Fig. S4). Specifically, 64% of poorly performing gRNA sequences end with a cytosine (C) (Supplementary Fig. S4). Another parameter that is strongly correlated to efficiency is the number of TT dimers. Sequences with zero or one TT dimer are over-represented among the best performing sequences with 21% and 55% of all the best performing sequences, respectively (Supplementary Fig. S5B). Other parameters, such as the base at position 14, 16 or 18 also have an impact on the gRNA sequence efficiency. However our data suggests that their contribution is less significant (Fig. 3F, Supplementary Figs. S5C–E).

Interestingly, when looking at the distribution of bases at positions 14, 16, 18 and 20 we observed that there is not a clearly preferred base for each position but rather that specific bases at specific locations have a negative effect. For example, G as the last base is positively correlated with efficiency not because it is over-represented in the best performing sequence group but rather because none of the worst performing sequences had a G at this position (Supplementary Fig. S4). The same is true for position 18 where adenine (A) is not found among the worst performing sequences while the other bases are comparably present (Supplementary Fig. S5).

In summary, in this study we characterized the efficiency of 50 gRNA sequences, 9 native and 41 synthetic. We also provided a temporal characterisation of the effect on the expression of sfGFP of both promoters and terminators. By providing 41 synthetic gRNA sequences we developed a system that can be easily used to replace previously integrated constructs. In line with previous results we observed that the most important parameters when designing efficient gRNA sequences are the G in position 20 and the number of TT dimers.

4. Discussion

The increasingly complex engineering efforts performed in *S. cerevisiae* require new highly characterised toolkits that are findable, accessible, interoperable and reproducible (FAIR). However, even though parts characterisation efforts are becoming more and more ambitious with a higher number of designs being assembled and tested, efforts to standardize acquired data and SOPs remains limited. Similarly, the availability of generated data is often restricted to manuscript figures and condensed methods sections, which limits auxiliary data analysis and hypothesis testing. In an effort to mitigate these challenges we took several approaches. Firstly, we made use of fluorescein, a reference standard for fluorescence intensity, as well as providing a LUDOX CL-X (45% colloidal silica suspension) measurement to include an absorbance reference value which can be used to convert absorbance values between instruments. Secondly, to encourage data re-usability and data analysis transparency, we provided jupyter notebooks containing the python scripts used to perform data analysis and visualisation. Lastly, in compliance with FAIR principles [12], DNA parts are "findable" and "accessible" via a publicly available repository, "interoperable" as data can be exported in SBOL-supported formats (e.g. as. fasta,.gb), and "reusable" as any user can freely make a personal copy of the digital records for personalized use.

Having said this, our data is but a snapshot of how regulatory DNA elements perform when expressed at one well-characterized landing pad in the genome of *S. cerevisiae* strain CEN.PK2–1C cultivated under defined medium conditions. Still, the data obtained in this study correlates well with the fluorescence-based characterizations conducted in other laboratories (Supplementary Fig. S3) [31], while at the same time recapitulate time-resolved transcriptomics data, albeit only for selected

medium and strong promoters (Supplementary Table S5), with some of the discrepancies between promoter data sets from this study and the previously reported transcriptomics data [23], observed for especially weak promoters, could be based on differences in cultivation conditions as well as medium acidification and limited oxygen availability at later time points affecting sfGFP maturation and fluorescence [46,47].

5. Conclusions

In this study we characterized multiple gRNA sequences that can be used as a toolkit to easily swap components of genetic cassettes. Next we characterised 38 genetic components, between promoters and terminators and we were able to identify a promoter (pYPR036W-A), that after 24 h induced a fluorescence intensity of 115 ng/ml of Fluorescein per unit of OD600, higher than the maximum pTDH3. Finally, in an effort to increase usability we provided formatted data and python scripts that allow us to easily extract the values or visualise the data at any available time-point or to perform any additional data analysis.

Data availability

The datasets analysed in this study are available from https://github.com/vasdam/Parts_characterisation.

Declaration of competing interest

D.D.L. has financial interests in Explora Biotech. J.D.K. has a financial interest in Amyris, Lygos, Demetrix, Napigen, Apertor Pharmaceuticals, Maple Bio, Ansa Biotechnologies, Berkeley Yeast, and Zero Acre Farms.

CRediT authorship contribution statement

Vasil D'Ambrosio: Project administration, lead, Writing – original draft, lead, Formal analysis, equal, Writing – review & editing, equal. **Lea G. Hansen:** Investigation, lead, Writing – review & editing, equal. **Jie Zhang:** Conceptualization, equal, Investigation, equal. **Emil D. Jensen:** Investigation, supporting, Writing – review & editing, supporting. **Dushica Arsovska:** Investigation, supporting. **Marcos Laloux:** Investigation, supporting. **Tadas Jakočiūnas:** Investigation, Supporting. **Pernille Hjort:** Investigation, supporting. **Davide De Lucrezia:** Conceptualization, equal, Writing – original draft, supporting, Writing – review & editing, equal. **Jay D. Keasling:** Funding acquisition, lead. Writing – review & editing, equal. **Michael K. Jensen:** Conceptualization, equal, Writing – original draft, supporting, Formal analysis, equal, Writing – review & editing, equal, Funding acquisition, lead.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2022.02.001.

References

[1] Arkin A. Setting the standard in synthetic biology. Nat Biotechnol 2008;26:771–4. https://doi.org/10.1038/nbt0708-771.

[2] Müller KM, Arndt KM. Standardization in synthetic biology. Methods Mol Biol 2012;813:23–43. https://doi.org/10.1007/978-1-61779-412-4_2.

[3] Hillson NJ, Plahar HA, Beal J, Prithviraj R. Improving synthetic biology communication: recommended practices for visual depiction and digital submission of genetic designs. ACS Synth Biol 2016;5:449–51. https://doi.org/10.1021/acssynbio.6b00146.

[4] Beal J, Goñi-Moreno A, Myers C, Hecht A, de Vicente MDC, Parco M, et al. The long journey towards standards for engineering biosystems: are the Molecular Biology and the Biotech communities ready to standardise? EMBO Rep 2020;21:e50521. https://doi.org/10.15252/embr.202050521.

[5] Hecht A, Filliben J, Munro SA, Salit M. A minimum information standard for reproducing bench-scale bacterial cell growth and productivity. Commun Biol 2018;1:219. https://doi.org/10.1038/s42003-018-0220-6.

[6] Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, et al. Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng 2009;3:4. https://doi.org/10.1186/1754-1611-3-4.

[7] Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. Science 2016;352:aac7341. https://doi.org/10.1126/science.aac7341.

[8] Canton B, Labno A, Endy D. Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol 2008;26:787–93. https://doi.org/10.1038/nbt1413.

[9] Castillo-Hair SM, Sexton JT, Landry BP, Olson EJ, Igoshin OA, Tabor JJ. FlowCal: a user-friendly, open source software tool for automatically converting flow cytometry data from arbitrary to calibrated units. ACS Synth Biol 2016;5:774–80. https://doi.org/10.1021/acssynbio.5b00284.

[10] Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452–4. https://doi.org/10.1038/533452a.

[11] Cronin M, Ghosh K, Sistare F, Quackenbush J, Vilker V, O'Connell C. Universal RNA reference materials for gene expression. Clin Chem 2004;50:1464–71. https://doi.org/10.1373/clinchem.2004.035675.

[12] Wilkinson MD, Dumontier M, Ijj Aalbersberg, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.

[13] Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2007;2:31–4. https://doi.org/10.1038/nprot.2007.13.

[14] Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, et al. EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae. FEMS Yeast Res 2014;14:238–48. https://doi.org/10.1111/1567-1364.12118.

[15] Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1686. https://doi.org/10.21105/joss.01686.

[16] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585:357–62. https://doi.org/10.1038/s41586-020-2649-2.

[17] Pedregosa F. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12: 2825–30.

[18] Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9:90–5. https://doi.org/10.1109/MCSE.2007.55.

[19] McKinney W. Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference. SciPy; 2010. https://doi.org/10.25080/majora-92bf1922-00a.

[20] Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422–3. https://doi.org/10.1093/bioinformatics/btp163.

[21] Waskom M. seaborn: statistical data visualization. J Open Source Softw 2021;6: 3021. https://doi.org/10.21105/joss.03021.

[22] Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks - a publishing format for reproducible computational workflows. ELPUB 2016. https://doi.org/10.3233/978-1-61499-649-1-87.

[23] Rajkumar AS, Özdemir E, Lis AV, Schneider K, Qin J, Jensen MK, et al. Engineered reversal of function in glycolytic yeast promoters. ACS Synth Biol 2019;8:1462–8. https://doi.org/10.1021/acssynbio.9b00027.

[24] Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. Science 2015;349:1095–100. https://doi.org/10.1126/science.aac9373.

[25] Rahmat E, Kang Y. Yeast metabolic engineering for the production of pharmaceutically important secondary metabolites. Appl Microbiol Biotechnol 2020;104:4659–74. https://doi.org/10.1007/s00253-020-10587-y.

[26] Xiong D, Lu S, Wu J, Liang C, Wang W, Wang W, et al. Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. Metab Eng 2017; 40:115–23. https://doi.org/10.1016/j.ymben.2017.01.006.

[27] Peng B, Williams TC, Henry M, Nielsen LK, Vickers CE. Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. Microb Cell Factories 2015;14:91. https://doi.org/10.1186/s12934-015-0278-5.

[28] Partow S, Siewers V, Bjørn S, Nielsen J, Maury J. Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae. Yeast 2010;27:955–64. https://doi.org/10.1002/yea.1806.

[29] Xiong L, Zeng Y, Tang R-Q, Alper HS, Bai F-W, Zhao X-Q. Condition-specific promoter activities in Saccharomyces cerevisiae. Microb Cell Factories 2018;17:58. https://doi.org/10.1186/s12934-018-0899-6.

[30] Sun J, Shao Z, Zhao H, Nair N, Wen F, Xu J-H, et al. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. Biotechnol Bioeng 2012;109:2082–92. https://doi.org/10.1002/bit.24481.

[31] Reider Apel A, d'Espaux L, Wehrs M, Sachs D, Li RA, Tong GJ, et al. A Cas9-based toolkit to program gene expression in Saccharomyces cerevisiae. Nucleic Acids Res 2017;45:496–508. https://doi.org/10.1093/nar/gkw1023.

[32] Beal J, Baldwin GS, Farny NG, Gershater M, Haddock-Angelli T, Buckley-Taylor R, et al. Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs. PLoS One 2021;16:e0252263. https://doi.org/10.1371/journal.pone.0252263.

[33] Hurtado S, Kim Guisbert KS, Sontheimer EJ. SPO24 is a transcriptionally dynamic, small ORF-encoding locus required for efficient sporulation in Saccharomyces cerevisiae. PLoS One 2014;9:e105058. https://doi.org/10.1371/journal.pone.0105058.

[34] Curran KA, Karim AS, Gupta A, Alper HS. Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab Eng 2013;19: 88–97. https://doi.org/10.1016/j.ymben.2013.07.001.

[35] Wei L, Wang Z, Zhang G, Ye B. Characterization of terminators in Saccharomyces cerevisiae and an exploration of factors affecting their strength. Chembiochem 2017;18:2422–7. https://doi.org/10.1002/cbic.201700516.

[36] Curran KA, Morse NJ, Markham KA, Wagman AM, Gupta A, Alper HS. Short synthetic terminators for improved heterologous gene expression in yeast. ACS Synth Biol 2015;4:824–32. https://doi.org/10.1021/sb5003357.

[37] MacPherson M, Saka Y. Short synthetic terminators for assembly of transcription units in vitro and stable chromosomal integration in yeast S. cerevisiae. ACS Synth Biol 2017;6:130–8. https://doi.org/10.1021/acssynbio.6b00165.

[38] Lee ME, DeLoache WC, Cervantes B, Dueber JE. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol 2015;4:975–86. https://doi.org/10.1021/sb500366v.

[39] Ito Y, Kitagawa T, Yamanishi M, Katahira S, Izawa S, Irie K, et al. Enhancement of protein production via the strong DIT1 terminator and two RNA-binding proteins in Saccharomyces cerevisiae. Sci Rep 2016;6:36997. https://doi.org/10.1038/srep36997.

[40] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339. https://doi.org/10.1126/science.1231143.

[41] Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 2013;154: 442–51. https://doi.org/10.1016/j.cell.2013.06.044.

[42] Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE. RNA-guided human genome engineering via Cas9. Science 2013;339. https://doi.org/10.1126/science.1232033.

[43] Baek S, Utomo JC, Lee JY, Dalal K, Yoon YJ, Ro D-K. The yeast platform engineered for synthetic gRNA-landing pads enables multiple gene integrations by a single gRNA/Cas9 system. Metab Eng 2021;64:111–21. https://doi.org/10.1016/j.ymben.2021.01.011.

[44] Bourgeois L, Pyne ME, Martin VJJ. A highly characterized synthetic landing pad system for precise multicopy gene integration in yeast. ACS Synth Biol 2018;7: 2675–85. https://doi.org/10.1021/acssynbio.8b00339.

[45] Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, et al. EasyClone-MarkerFree: a vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. Biotechnol J 2016;11:1110–7. https://doi.org/10.1002/biot.201600147.

[46] Hansen MC, Palmer RJ, Udsen C, White DC, Molin S. Assessment of GFP fluorescence in cells of Streptococcus gordonii under conditions of low pH and low oxygen concentration. Microbiology 2001;147:1383–91. https://doi.org/10.1099/00221287-147-5-1383.

[47] Zhang J, Petersen SD, Radivojevic T, Ramirez A, Pérez-Manríquez A, Abeliuk E, et al. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nat Commun 2020;11: 4880. https://doi.org/10.1038/s41467-020-17910-1.