Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Investigating the role of noncoding regulatory DNA in plasmid development for Yarrowia lipolytica

Carmen Lopez[1], Mingfeng Cao[2], Zhanyi Yao[1], and Zengyi Shao[1]

[1]Iowa State University
[2]University of Illinois at Urbana-Champaign

December 2, 2020

Abstract

Production of industrially relevant compounds in microbial cell factories can employ either genomes or plasmids as an expression platform. Selection of plasmids as pathway carriers is advantageous for rapid demonstration but poses a challenge of stability. Yarrowia lipolytica has attracted great attention in the past decade for the biosynthesis of chemicals related to fatty acids at titers attractive to industry, and many genetic tools have been developed to explore its oleaginous potential. Our recent studies on the autonomously replicating sequences (ARSs) of nonconventional yeasts revealed that the ARSs from Y. lipolytica showcase a unique structure that includes a previously unannotated sequence (spacer) linking the origin of replication (ORI) and the centromeric (CEN) element and plays a critical role in modulating plasmid behavior. Maintaining a native 645-bp spacer yielded a 4.5-fold increase in gene expression and higher plasmid stability compared to a more universally employed minimized ARS. Testing the modularity of the ARS sub-elements indicated that plasmid stability exhibits a pronounced cargo dependency. Instability caused both plasmid loss and intramolecular rearrangements. Altogether, our work clarifies the appropriate application of various ARSs for the scientific community and sheds light on a previously unexplored DNA element as a potential target for engineering Y. lipolytica.

Submitted to Biotechnology and Bioengineering on November 30th, 2020

Investigating the role of noncoding regulatory DNA in plasmid development for *Yarrowia lipolytica*

Carmen Lopez[1,3], Mingfeng Cao[2,*], Zhanyi Yao[2,3], and Zengyi Shao[1,2, 3,4,*]

[1]Interdepartmental Microbiology Program, Iowa State University, Ames, Iowa, USA

[2]Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, USA

[3]NSF Engineering Research Center for Biorenewable Chemicals, Iowa State University, Ames, Iowa, USA

[4]The Ames Laboratory, Ames, Iowa, USA

[*]Correspondence to:

Dr. Zengyi Shao,

4140 Biorenewables Research Laboratory,

Iowa State University, Ames, IA 50011, USA

Phone: 515-294-1132

Email: zyshao@iastate.edu

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Dr. Mingfeng Cao,

1206 E. Gregory Street, M/C 412,

University of Illinois, Urbana, IL 60801, USA

Email: mfcao@illinois.edu

Grant numbers: National Science Foundation Grants (1716837 and 1749782) and the Iowa State University bio-based product seed grant

Running title: Plasmid development for *Yarrowia lipolytica*

Abstract

Production of industrially relevant compounds in microbial cell factories can employ either genomes or plasmids as an expression platform. Selection of plasmids as pathway carriers is advantageous for rapid demonstration but poses a challenge of stability. *Yarrowia lipolytica* has attracted great attention in the past decade for the biosynthesis of chemicals related to fatty acids at titers attractive to industry, and many genetic tools have been developed to explore its oleaginous potential. Our recent studies on the autonomously replicating sequences (ARSs) of nonconventional yeasts revealed that the ARSs from *Y. lipolytica* showcase a unique structure that includes a previously unannotated sequence (spacer) linking the origin of replication (ORI) and the centromeric (CEN) element and plays a critical role in modulating plasmid behavior. Maintaining a native 645-bp spacer yielded a 4.5-fold increase in gene expression and higher plasmid stability compared to a more universally employed minimized ARS. Testing the modularity of the ARS sub-elements indicated that plasmid stability exhibits a pronounced cargo dependency. Instability caused both plasmid loss and intramolecular rearrangements. Altogether, our work clarifies the appropriate application of various ARSs for the scientific community and sheds light on a previously unexplored DNA element as a potential target for engineering *Y. lipolytica.*

Keywords: plasmid stability, *Yarrowia lipolytica,*autonomously replicating sequence, gene expression, genetic manipulation tools

Introduction

The baker's yeast, *Saccharomyces cerevisiae* , is considered a eukaryotic workhorse in the synthetic biology community, attributed to its extensive characterization and capacity to produce a wide array of valuable chemicals, even at industry-relevant scales. However, nature has exerted selective pressure on many other lesser-known microorganisms, forcing them to adapt to their dynamic environments. These so-called "non-conventional" microorganisms possess unique metabolic and physiological capabilities, such as superior xylose fermentation ability (Gao et al., 2016), high thermotolerance (Sorokina et al., 2017), and acid tolerance (Park et al., 2018) [see (Löbs, Schwartz, & Wheeldon, 2017; Thorwall, Schwartz, Chartron, & Wheeldon, 2020) for review of nonconventional yeasts], that can be exploited in industrial settings. In particular, *Yarrowia lipolytica* has garnered much attention in recent years owing to its powerful oleaginous ability to produce lipid-relevant products and other non-native chemicals at titers attractive to industry (Darvishi, Ariana, Marella, & Borodina, 2018; Miller & Alper, 2019; Spagnuolo, Yaguchi, & Blenner, 2019; Xu, Qiao, Ahn, & Stephanopoulos, 2016). Manipulation of the *Y. lipolytica* genome has been demonstrated by random approaches that exploit inherent features of this yeast, such as its preference for non-homologous end joining for repair of chromosomal double-strand breaks (Cui, Jiang, Zheng, Qi, & Hou, 2019) or the presence of numerous ribosome DNA clusters that serve as reusable "landing pads" for integration of heterologous DNA (Lv, Edwards, Zhou, & Xu, 2019). Genome editing has become amenable due to the development of more precise technologies, such as transcription activator-like effector nucleases (TALEN) (Rigouin et al., 2017) and CRISPR-Cas9 (Schwartz, Hussain, Blenner, & Wheeldon, 2015; Wong, Engel, Jin, Holdridge, & Xu, 2017). Nonetheless, extrachromosomal expression*via* plasmids remains a desirable objective, since it enables engineering in a more rapid fashion, especially at the initial stage of strain development, when genetic elements need to be 'mixed and matched' for testing (Guo et al., 2015).

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Progress on this front for *Y. lipolytica* is considerable. For example, the Alper group created a plasmid series bearing an array of hybrid tunable promoters with tandem upstream activating sequences (UAS) and attained a large dynamic range of transcriptional levels (Blazeck, Liu, Redden, & Alper, 2011). More recently, YaliBricks, an innovative platform that relies on compatible restriction enzymes for the facile assembly of long pathways, has been used for the production of squalene and aromatic compounds, including 2-phenylethanol, *p* -coumaric, and violacein (Liu, Wang, Deng, & Xu, 2020; Ma, Gu, Marsafari, & Xu, 2020; Wong et al., 2017). Although these episomal expression platforms have enabled higher gene expression and metabolite production, plasmid instability presents a persistent issue in *Y. lipolytica* (Blazeck et al., 2011; Wong et al., 2017). For example, during an engineering endeavor to produce an array of chemicals derived from the shikimate pathway, it was demonstrated that an increase in the rate-limiting precursor, erythrose 4-phosphate, required the phosphoketolase gene to be integrated into the genome to address the potential issue of biased segregation of plasmids (Gu, Ma, Zhu, Ding, & Xu, 2020). Nonetheless, further inspection of alternative genetic elements and their potential to impact plasmid stability has not been explored. Isolation of autonomously replicating sequences (ARSs) that enable plasmid construction can be traced back to the early 1990s. To date, four ARSs have been discovered and utilized at various frequencies (**Figure 1 and Table S1**). ARS1, ARS2, ARS18, and ARS68 are isolated from chromosomes 1, 5, 3, and 1, respectively. ORI-CEN is a minimized version of ARS1 lacking a the linking sequence between the origin of replication (ORI) and the centromere (CEN) (Matsuoka et al., 1993; Yamane, Sakai, Nagahama, Ogawa, & Matsuoka, 2008) (often labeled as ori1001-CEN1-1 in the literature), whose usage has become more popular in recent years compared to the other four ARSs.

The purpose of this study was to evaluate the plasmid stability driven by different ARSs and to clarify plasmid selection for engineering *Y. lipolytica* . More specifically, a noncoding DNA sequence linking ORI and CEN in ARS was proven to play a critical role in plasmid stability. This linker, referred to as a *spacer* hereafter, is not interchangeable between ARS sequences without compromising plasmid stability. Assays with plasmids bearing either a fluorescence gene reporter or the β-carotene pathway demonstrated that use of an ARS containing a spacer generates phenotypically enhanced strains with more active transcription, and improved plasmid stability. We also attempted a synthetic biology approach that consisted of the exchange of a native spacer varying in either size or sequence; this uncovered a clear cargo-dependency of plasmids with engineered ARSs and suggested an apparent non-modularity feature of the ARS sub-elements. In practice, plasmid loss occurred in terms of both copy number and integrity. This work underscores the importance of ARS selection for achieving optimal behavior of plasmid platforms and the obligated characterization of compatibility of the target pathway with the plasmid backbone.

Materials and Methods

2.1 Strains and culture media

Y. lipolytica PO1f (MATa *ura* 3–302, *leu* 2–270,*xpr* 2–322, *axp* 2–deltaNU49, and XPR2:SUC2) was purchased from ATCC (Manassas, VA, USA) and served as the host. *S. cerevisiae* YSG50 (MATα, *ade* 2-1, *ade* 3Δ22, *ura* 3-1,*his* 3-11, 15, *trp* 1-1, *leu* 2-3, 112, *can* 1-100) was used to construct the plasmids. YPAD consists of 1% (or 10 g/L) yeast extract, 2% peptone, 0.01% adenine hemisulfate, and 2% dextrose. Synthetic complete media were prepared using 0.083% complete supplement mixture lacking the relevant amino acid, 0.16% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, and 2% dextrose. *Escherichia coli* BW25141 (*lacI* ^{q} *rrnB* $_{T14}$Δ*lacZ* $_{WJ16}$ Δ*phoBR580 hsdR514*Δ*araBAD* $_{AH33}$ Δ*rhaBAD* $_{LD78}$*galU95 endA* $_{BT333}$*uidA* (Δ*MluI*)[?]pir^{+} *recA1*) was cultured in Luria Bertani medium supplemented with 100 μg/mL ampicillin for plasmid enrichment. The commercial *E. coli*NEB® 10-beta (C3019I) competent cells were used for plasmid rescue analysis for the recovered yeast plasmids to their high transformation efficiency. The strains and corresponding genotypes are summarized in **Table S2** .

2.2 Plasmid construction and transformation

We employed DNA assembler (Shao, Luo, & Zhao, 2012; Shao, Zhao, & Zhao, 2009; Shao & Zhao, 2014) to construct all our plasmids, which are listed in **Table S2** . For construction of the green fluorescent protein

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

(GFP) plasmid series, plasmids p3887 (a gift from Dr. Suzanne Sandmeyer at University of California, Irvine) and pCRISPRYL (Schwartz et al., 2015) (Addgene #70007) were used as templates to amplify the ori1001, CEN1-1, and the spacer from ARS18. Q5 high-fidelity DNA polymerase (New England BioLabs, Ipswich, MA, USA) was used for PCR. ARS sub-elements were incorporated into a design containing the replicative elements and selection markers for both *E. coli and S. cerevisiae*to facilitate DNA assembly, yielding pSCARS1 and pYL-ORI1001-CEN1-1. Swapping of the ARS1 spacer by either the ARS18 spacer or a random spacer (i.e., 645 bp amplified from the backbone of a plasmid belonging to *S. stipitis* , another nonconventional yeast species being studied by our group), was achieved by restriction of pSCARS1 by*Cla* I and *BamH* I (New England BioLabs, Ipswich, MA) and coupling with the desired ARS sequence amplified from gBlocks® (IDT, Coralville, IA) (**Table S3, Replicative elements**). Subsequent insertions at the 5'-end of the spacers were incorporated in primers whose typical structures comprise a 40-bp homology arm to the plasmid backbone, 1-6 nucleotides that constitute the extension and a 20-bp priming region to the target spacer for PCR amplification. These primers amplified a newly extended spacer, which was then combined with ori1001, CEN1-1, and the digested pSCARS1 backbone for a new round of assembly. Briefly, the assembly entailed co-transformation of the desired fragments into *S. cerevisiae* by electroporation. Cells were then plated on selective medium for 2-3 days followed by inoculation into liquid selection medium. The plasmid was isolated using Zymoprep Yeast Plasmid Miniprep II (Zymo Research, Irvine, CA) and transformed into *E. coli* for enrichment. Intactness score was determined by restriction digestion pattern and Sanger sequencing (ISU DNA Facility, Ames, IA). Genomic DNA of strains CIBTS1606 and CIBTS1608 (Gao et al., 2017) (a gift from Dr. Sheng Yang at Shanghai Institutes for Biological Sciences, Chinese Academy of Science) was used as the template for amplifying the required elements for the β-carotene biosynthetic pathway. An intermediate helper plasmid, pYL-precursor, was first assembled to incorporate the precursor genes for β-carotene synthesis (i.e., *ggs1* and *tHmg*). Next, pYL-precursor was digested with *Mre* I and *Pac* I, and the expression cassettes for *carB* and *carRP* were incorporated by a second round of assembly, yielding pYL-βC which served as a template for pathway amplification. To construct the β-carotene plasmid series, the set of selected GFP plasmids was restricted by *Nde* I and *Mlu* I to serve as backbones, to which the β-carotene pathway was incorporated by another round of assembly. The verified plasmids were transformed to *Y. lipolytica* by electroporation at 12.4 kV/cm (Wang, Hung, & Tsai, 2011).

2.3 Fluorescence-based analysis by flow cytometry

We used a DR46B transilluminator (Clare Chemical Research, Dolores, CO, USA) to screen colonies transformed with the GFP plasmid series by fluorescence. The screened colonies were cultured in 2 mL SC-URA medium at 30 °C and 250 rpm for 48 h. Cells were diluted in 10 mM phosphate-buffered saline (pH 7.4) to reach an optical density at 600 nm (OD_{600}) of 0.1–0.2. We analyzed the fluorescence intensity of the cells at 488 nm with a BD Biosciences FACSCanto flow cytometer, and the distribution and percentage of GFP^+cells were calculated with the accompanying software.

2.4 Transcriptional level assay by qPCR

GFP expression was measured for the following ARS variants: ORI-CEN, ARS1, ARS2, ARS18, ORI-random sp-CEN, ORI-foreign sp-CEN, and ORI-EXT-foreign sp-CEN. Cells were collected from the same cultures used for fluorescence-based analysis. A pellet of at least $OD_{600} = 10$ was collected at 24 h for RNA extraction. The workflow for qPCR analysis was similar to that of a previous report (Lopez, Zhao, Masonbrink, & Shao, 2020). The primer pair YL-GFP (f-CAACCTGATCGAGGAGATGTT; r-CCGGTGATGGTCTTCTTCAT) quantified the fluorescent reporter gene, and the pair YL-ACT1 (f- CAAGTCCAACCGAGAGAAGATG; r-GGCCTGGATAGAGACGTAGAA) quantified*ACT1* , a housekeeping gene. Gene quantification was performed using the corresponding standard curve and was reported as relative gene expression with respect to *ACT1* . The assay was performed in triplicate.

2.5 β-ςαροτενε χυαντιφιςατιον

To evaluate the generalizability of the plasmid platform with varying ARS sequences, β-carotene was quantified. Colonies carrying plasmids from the β-carotene series were subjected to examination by color intensity.

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Three random colonies from the orange group on the transformation plates were inoculated into 3 mL of SC-URA medium at 30 ˚C and 250 rpm for 48 h as seed cultures, and subsequently transferred into 10 mL of SC-URA medium at an initial OD_{600} of 0.2. The cells were cultivated at 30 ˚C and 250 rpm for 96 h. Samples collected at 24, 48, 72, and 96 h were prepared following a previously described protocol (Gao et al., 2017). The supernatant was diluted with acetone at a ratio of 1:10 to prevent saturation of the signal measured by the Synergy Eon Microplate Spectrophotometer in 96-well polypropylene plates at 455 nm (Greiner Bio-One, Solingen, Germany) containing 200 μL of each diluted sample. β-carotene (Sigma-Aldrich, Milwaukee, WI, USA) was used to prepare a standard curve (0-100 mg/L).

2.6 Plasmid rescue analysis

Stability of plasmids containing assorted ARS sequences was assessed. The analysis started with the inoculation of five yeast colonies (per evaluated construct) from a transformation plate to 3 mL of SC-URA medium and cultivation for 48 h at 30 ˚C and 250 rpm. Plasmids were isolated from yeast using Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research, Irvine, CA) and transformed to *E. coli*NEB® 10-beta. Although these commercial cells are highly competent, some transformations failed to produce colonies or yielded only one colony. Regardless, the observed *E. coli*colonies were recorded as the number of rescued colonies, i.e., total colony count. Whenever possible, two *E. coli* colonies from each transformation plate were grown under standard conditions, and plasmids were isolated using the QIAprep Spin Plasmid Mini-prep kit (QIAGEN, Valencia, CA) to verify whether plasmids had been kept intact. In case that only one *E. coli* colony appeared on the plate, the single colony was inoculated for plasmid isolation. The isolated plasmids were inspected by restriction digestion using *Xba* I and *Mlu* I. The number of plasmids showing the expected digestion pattern was divided by 10 (i.e., ideally 10 intact *E. coli* plasmids isolated from 5 yeast colonies) was reported as the intactness score for each construct.

Results

3.1 The effect of ARS selection on plasmid behavior

One of the longstanding issues with plasmid platforms is the potential for instability, which has been documented in recent reports on *Y. lipolytica* (Blazeck et al., 2011; Wong et al., 2017). Yeast plasmids, in general, rely on ARS for replication and CEN for stable segregation, which is a common feature shared among some species, including *S. cerevisiae* , *Scheffersomyces stipitis* , and *Issatchenkia orientalis* (Cao et al., 2017; Fitzgerald-Hayes, Clarke, & Carbon, 1982; Sun et al., 2020). In these species, ARS can independently direct plasmid replication in the absence of CEN. However, in *Y. lipolytica* , a CEN and an ORI arranged at the ends of an unannotated DNA sequence constitutes a functional ARS. An ORI cannot support an autonomously replicating function solely, and it has to be coordinated with the CEN end to direct plasmid replication. This observation led us to hypothesize that the spacer (i.e., the linker) between CEN and ORI in the natural ARS may exert a coordinative impact to facilitate plasmid replication and segregation. In order to test this hypothesis, we chose to characterize the most prevalent replicative elements: the popularly used ORI-CEN and ARS18, and the less popular ones ARS2 and ARS1 (**Figure 1a**). ORI-CEN was generated from ARS1 by removing the spacer (Yamane, Ogawa, & Matsuoka, 2008). We did not include ARS68 because of its high similarity with ARS1; it was labeled differently for being sourced from a different strain (Fournier, Guyaneux, Chasles, & Gaillardin, 1991; Matsuoka et al., 1993). A schematic to illustrate the similarity of ARS1 and ARS68 is presented in Supplementary**Figure S1** . **Table S3** lists the sequences of the ARSs in this study and their relations are shown in Figure 1.

We constructed a set of plasmids bearing an expression cassette for green fluorescent protein GFP and incorporated a different ARS within each construct. After verification, plasmids varying solely in the ARS were transformed to *Y. lipolytica* and a fluorescence assay was implemented to benchmark plasmid-borne gene expression. Inspection of plates during standard incubation qualitatively supported our initial hypothesis, as colonies containing a native ARS, namely the spacer-containing ARS, were brighter and grew faster than those expressing GFP by the ORI-CEN plasmid (**Figure 2a**). Flow cytometry revealed that populations bearing wild type ARSs had more than 90% GFP^+ cells compared to the ones containing ORI-CEN, in

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

which 77% of the cells were GFP$^+$. Histograms also showed that GFP$^+$ cells exhibited a bimodal fluorescence distribution with two evident subpopulations observed in the form of a tall peak and a left shoulder next to a plasmid-less cell fraction. The distribution pattern was consistent with other reports for this yeast (Blazeck et al., 2011; Liu, Otoupal, Pan, & Alper, 2014). Furthermore, similar fluorescence profiles observed in*S. stipitis* , another non-conventional yeast being studied in our group, have shown that distinction between subpopulations is due to the difference in plasmid copy number (Cao et al., 2017). ARS1, ARS2, and ARS18 had comparable percentages of GFP$^+$ cells that were significantly different from ORI-CEN ($p < 0.05$). Taking the percentage of GFP$^+$ cells and the fluorescence profile together, it seems that spacer-containing ARSs are beneficial for plasmid segregation, and consequently, more stable compared to the minimal ORI-CEN sequence that has been in prevalent usage recently. Selection of a wild type ARS yields 15–20% more GFP$^+$ cells than the minimal ORI-CEN sequence, and this impact appears to be sustained over long culture periods (**Figure S2**).

We sought to investigate the potential roots for the desirable effects that wild-type ARSs conferred to plasmid-borne expression. Therefore, we asked whether the positive effect of choosing a spacer-containing ARS affected the transcriptional level of the reporter gene as well. qPCR results demonstrated that simply swapping the minimal ORI-CEN by a wild-type ARS could yield a significant increase in GFP expression ranging from 2.8 to 4.5-fold, with the latter obtained when using ARS18 (**Figure 2b**). In addition, considering that a truly stable plasmid should be maintained in cells for longer periods, we assessed plasmid stability by a plasmid rescue experiment that was modified on previous protocols (Fournier et al., 1993; Matsuoka et al., 1993; Vernis et al., 1997). Yeast plasmids were purified from 48-h cultures grown to similar densities and subsequently transformed into *E. coli* , which enabled us to estimate the plasmid load of GFP$^+$cells from the number of *E. coli* colonies retrieved.**Figure 2c** suggests that inclusion of the native spacer within ARS leads to much higher numbers of rescued colonies compared to the minimal ORI-CEN sequence (281–972 vs. 44 colonies, respectively), which was consistent with the analysis at the transcriptional level. Another parameter that we investigated was plasmid intactness, considering that unstable plasmids might also undergo intramolecular rearrangements. To this end, we used an intactness score (I) that represents the percentage of plasmids that were verified to maintain the same sequence after recovery from *Y. lipolytica* as the initially transformed plasmids. ARS1, ARS2, and ORI-CEN were effective in maintaining plasmid integrity (i.e., I = 1) compared to ARS18, for which 1 out of 5 analyzed clones showed an unexpected digestion pattern (i.e., I = 0.8). Overall, these results suggest that among the few available ARSs, wild-type ARSs that contain an intact, native spacer sequence outperform the more commonly used spacer-less ORI-CEN in terms of both stability and bulk transcriptional level. Furthermore, we found that while ARS1 and ARS2 exhibit a slightly lower transcriptional level than ARS18, they exhibit a higher stability as indicated by their intactness scores.

3.2. Construction of synthetic ARSs and characterization of plasmid behavior

After confirming that the spacer could potentially modulate plasmid replication and segregation, we were curious to see whether an ARS demands any sequence-specificity of the spacer for desirable performance. We hypothesized that the length and the actual sequence of the spacer represent two parameters implicated in ARS performance and, consequently, influence plasmid behavior. Therefore, we implemented a synthetic biology approach in which the 645 bp native spacer of ARS1 was changed by either the 227-bp spacer from ARS18 (Fournier et al., 1991) (referred to as *foreign sp*) or a 645-bp random sequence (referred as *random sp* hereinafter) (**Figure 3a** and **Table S3** for sequences). Comparison of the histograms in **Figure 3b**indicates that the introduction of the shorter *foreign sp*disrupted the desirable distribution of GFP$^+$ cells, yielding a much smaller fraction of high GFP$^+$ cells (23% at 48 h as opposed to 41% for *random sp*). This observation shaped a new hypothesis that consisted of the adjustment of the distance between ORI and CEN to potentially modulate the coordination of these two functioning units epigenetically. We constructed a new set of plasmids by incorporating extensions of up to six randomly chosen nucleotides at the 5'-end of the foreign spacer, as shown in**Figure 3a** . The performance of plasmids containing these synthetic ARSs was evaluated in a fashion similar to that applied to the wild-type ARSs. The resulting GFP profiles provided evidence that regardless of the number of added nucleotides, these modifications helped to shift cells from an undesirable distribution where most of the GFP$^+$ cells had a low level of fluorescence to improved profiles

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

where 37–81% of the populations could be classified as high-GFP (**Figure 3c and Figure S3**). Among them, extension of the foreign spacer by four nucleotides (referred to as *EXT-foreign sp N4* hereinafter) was the most beneficial modification, as it generated a distribution with only 2% of the cells being plasmid-less and 81% being high GFP^+, a profile even better than that obtained from the use of any wild-type ARS (**Figure 3b, c**).

To explore whether the improved synthetic ARS impacts plasmid behavior in a manner similar to that of the wild-type sequences, further characterization was performed for the synthetic ARSs in the following configurations: ORI-random sp-CEN, ORI-foreign sp-CEN, and ORI-EXT-foreign sp N4-CEN. The transcriptional level assay based on qPCR shows that the sequence harboring the extended foreign spacer is the only synthetic ARS that presented an increased GFP expression, which was about 2-fold compared to ARS1 and ARS18 from which the individual replicative elements or the spacer were sourced (**Figure 3d**). The low level of GFP expression generated by sequences with either random or foreign spacers appears to be consistent with the fact that the population was composed of a large fraction of the GFP^+ cells with low fluorescence (46% and 52%, respectively) (**Figure 3b**). Unexpectedly, the highly promising ORI-EXT-foreign sp N4-CEN plasmid fared poorly in the intactness test. A low number of rescued colonies and low intactness scores were obtained for the extended foreign spacer (**Figure 3e**), suggesting the possibility that although most of the cells exhibited improved GFP expression, many of the plasmids might have undergone rearrangements that led to a loss of the elements required for plasmid propagation in*E. coli* . Nevertheless, these results corroborate that the modulation effect of spacer length on plasmid performance takes place at a single base-pair resolution. Future efforts could involve library creation and examination of synthetic ARSs at both length and sequence diversities.

3.3. Assessment of cargo-dependent plasmid stability

The outstanding effect of spacer extension on the performance of the synthetic foreign spacer to improve GFP expression was contradictory to its deleterious impact on plasmid integrity. This drew our attention to the possibility that other regions on the plasmid backbone, not obviously related to plasmid stability, might have passively interfered with plasmid replication and stable segregation. To test this hypothesis, we sought to investigate whether the extended foreign spacer could not keep its beneficial impact on upregulating the expression level after the replacement of the GFP cassette by a long heterologous pathway. To that end, we decided to use the β-carotene biosynthetic pathway as a case study. β-carotene is an orange-colored terpenoid often used as a food additive and a supplement for cosmetics and nutraceuticals (Bogacz-Radomska & Harasym, 2018). *Y. lipolytica*has been demonstrated as a well-suited host for β-carotene production because of its unique ability to store highly hydrophobic compounds in lipid droplets (Gao et al., 2017). The minimal β-carotene biosynthetic pathway comprises the heterologous genes *carB* and *carRP* , and the endogenous genes, *tHmgR* and *ggs1* . A new series of plasmids was constructed to swap GFP by the minimal β-carotene pathway, followed by transformation to *Y. lipolytica* . After incubation for 2–3 days, carotenogenic cells were screened and cultured for 96 h. Analysis of different variants was enabled by the implementation of an absorbance-based assay to quantify β-carotene production facilitated by plasmids with assorted ARS sequences that included the minimal ARS sequence ORI-CEN, wild type ARS1, ORI-foreign sp-CEN, and ORI-EXT-foreign sp N4-CEN. We found that carotenogenic cells bearing wild-type ARS1 were the best producers, and their titer doubled that of cells expressing the pathway by the ORI-CEN plasmid (**Figure 4a and 4b**). The observed trend was consistent with the GFP case illustrated in **Figure 2a** . Importantly, cells that had the extended foreign spacer N4 were among the lowest producers in this experiment, and the titer was even lower than that of the cells with the minimal ORI-CEN sequence. Taking into consideration that the extended foreign spacer N4 benefited GFP expression but was associated with low*E. coli* transformants, we speculated that the results of low β-carotene production arose from more pronounced instability issues in plasmids containing the synthetic ARSs and the instability exhibited a cargo dependency.

To verify this hypothesis, we continued with the characterization of this set of plasmids. Considering that the colorless colonies unequivocally underwent plasmid loss or rearrangement, we randomly picked five carotenogenic colonies for the subsequent plasmid rescue analysis. The results in **Figure 4c** are in line

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

with the idea that the predictable behavior of ARS1 is likely caused by a relatively high plasmid stability, which is supported by a significantly higher number of rescued colonies and an intactness score of 1. On the contrary, recovery from β-carotene plasmids lacking a spacer (i.e., ORI-CEN), containing a foreign spacer, or carrying the extended foreign sp N4, hardly yielded colonies. Specifically, less than ten *E. coli* transformants could be obtained from any of these three plasmid contexts, and a third of the transformation plates did not show any colonies at all, which strongly indicated that the cells were more prone to plasmid loss or rearrangement. Carotenogenesis on agar plates hinted the occurrence of rearrangement, as the presence of white, light yellow, and orange colonies was prevalent for the foreign spacer and the extended foreign spacer N4 variants, which was not typically observed in the other two contexts (ORI-CEN and ARS1, **Figure 4a**). Moreover, upon focusing on the carotenogenic cells that were supposed to sustain an intact β-carotene pathway, we found that the plasmids bearing the extended foreign spacer N4 had the lowest intactness score (I = 0.2,**Figure 4c**). As for ORI-CEN and ORI-foreign sp-CEN, they scored comparably at 0.4 and 0.5 for intactness, respectively, and had consistently low number of rescued *E. coli* clones. Considering that all the clones were selected from the carotenogenic groups, these results suggested that some rearrangements could have caused the loss of the elements responsible for plasmid replication in *E. coli* .

To find whether plasmid rearrangement occurred with a systematic trend, a few rescued plasmids containing the extended foreign spacer N4 were sequenced. In addition to unexpected restriction patterns yielding low intactness scores, several sequencing reactions were unsuccessful, indicating that the corresponding primer-annealing regions were absent in the plasmid. **Figure 4d** illustrates some of the changes that plasmids underwent. Among all the retrieved plasmids, we could not identify any consistent pattern of rearrangement, implying that the changes might have been generated randomly. It is worth pointing out that the current procedure could biasedly lead to the recovery of the plasmids with intact *E. coli* ori and Amp^+selection marker. The encounter of no *E. coli* transformants for some of the colonies picked for the three engineered ARSs (i.e., ORI-CEN, ORI-foreign sp-CEN, and ORI-EXT foreign sp N4-CEN) suggested that the disruption could likely have occurred on *E. coli*elements. Compared to the results obtained with GFP as a cargo, it appears that the orthogonality required for a plug-and-play plasmid system to express any pathway has not been established for *Y. lipolytica* . Although wild-type ARS1 appears to be the best replicative element to date, we highly recommend rigorous plasmid stability examination whenever a new cargo is loaded.

Discussion

In this study, we present the critical modulating role played by a noncoding DNA sequence, linking the previously designated ORI and CEN elements of an ARS, on the performance of an episomal plasmid in the oleaginous yeast *Y. lipolytica.* A major finding is that the use of wild-type ARS sequences, i.e., those containing an intact spacer, consistently leads to reliable and improved plasmid behavior evaluated as per phenotypic variation, changes in gene expression, and stability. Considering that ORI-CEN, a minimal ARS sequence lacking the spacer, has been generally preferred in recent engineering endeavors, the demonstration of wild-type ARS1 in plasmid design that resulted in enhanced performance allows its potentially prominent usage for this yeast in the future. To the best of our knowledge, this study constitutes the first effort to optimize *Y. lipolytica* as an expression host using ARS as a target.

The paradigm of yeast plasmid replication is established with respect to the model yeast *S. cerevisiae.* It entails the use of an ARS sequence with 'well-conserved DNA domains' that sustains plasmid maintenance during cell growth (Dhar, Sehgal, & Kaul, 2012). However, it was observed that in some yeasts (e.g., *S. cerevisiae* ,*P. pastoris* , and *S. stipitis*), plasmids borne by cells display marked instability when only an ARS was included in the design. This stability issue can be mitigated by incorporating a self-standing CEN with the native role of directing chromosome stable segregation during cell division (Cao et al., 2017; Clarke & Carbon, 1980; Nakamura et al., 2018). In these yeasts, the ARS and CEN elements do not need to be associated in the same chromosome when being isolated. In contrast, *Y. lipolytica* presents an ARS configuration that deviates from this canonical structure. In all of the counted examples of ARS isolation from the *Y. lipolytica* genome, it has been identified that sequences are composed of an ORI, a CEN, and a linker sequence that varies in length (Fournier et al., 1991; Matsuoka et al., 1993). It was also

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

demonstrated in an early report that successful plasmid replication in *Y. lipolytica* requires the obligatory presence of both an ORI and a CEN, joined by a few hundred base pairs (Vernis et al., 1997). In that study, the authors explored the potential role of the spacer in ARS activity by shortening its length from 1002 bp to 410 bp in an ORI1068-spacer-CEN3 configuration and concluded that spacer length is not critical for plasmid replication. This view was strengthened when a minimal ARS lacking the spacer yielded a transformation efficiency of 2.6×10^3 colonies/μg DNA, albeit with a mitotic stability less than expected (Yamane, Sakai, et al., 2008). The dispensability of the linker length is contrary to our findings, since we observed drastic changes when comparisons were drawn between ARS1 and its minimal version, i.e., ORI-CEN, which constitutes the corresponding analogous experiment. Furthermore, the benefit of maintaining the natural spacer sequence was sustained for two plasmid-borne cargoes, the GFP gene and the β-carotene pathway, and was also confirmed at the mRNA and protein levels in the case of GFP. We attribute this discrepancy to the conclusion made in the earlier study (Vernis et al., 1997) to the methods employed to survey the impact of altering the distance between ORI and CEN. They relied on a simple dichotomous question of whether yeast colonies appeared on plates after transformation by electroporation, which masked the subtle differences in plasmid stability and intactness, not to mention the variation rendered by the cargo.

The stringent mode of segregation in *Y. lipolytica* hinders the isolation of ARSs. There have been reports of only four ARSs in the last three decades, and among those, ARS1 and ARS68 share remarkable similarity despite being sourced from different strains (Fournier et al., 1991; Matsuoka et al., 1993). Our efforts intended to explore the effects of mixing and matching of ARS sub-elements (i.e., ORI, CEN, and the spacer) point to the fact that spacer swapping among ARSs jeopardizes the plasmid stability. This illustrates the importance of more exhaustive future studies aimed at understanding the uniqueness of the replication model in *Y. lipolytica,* especially pertaining to the mechanism by which a spacer elicits better plasmid performance. Earlier *in vivo* studies demonstrated that CEN1-1, CEN3-1, and, to a lesser degree, the spacers, depict nuclear scaffold binding activity (Vernis et al., 2000). Scaffold-associated regions (SAR) are genomic DNA sequences that interact with nuclear scaffolds, looped nuclear structures devoid of histones during mitotic interphase (DNA replication) or metaphase (chromosomes align with CENs attached to the spindles) (Amati & Gasser, 1988). ORIs in *S. cerevisiae* and higher eukaryotes (e.g., humans) were found to be able to bind to nuclear scaffolds (Amati & Gasser, 1988; Wang et al., 2019). Considering that SAR activity was not observed in the ORIs from *Y. lipolytica* (Vernis et al., 2000), it is possible that some regions in the CEN and/or the spacer could fulfill this role. More interestingly, as SAR are determined by structure, rather than sequence, we speculate that the native spacer could facilitate the establishment of a helical structure optimal for the loading of replication machinery. Analogously, in the *Drosophila* genome, SAR are localized in regions with low helical stability, which implicates a potential role of the structural element in the replication process (Brun, Surdej, & Miassod, 1993). In*S. cerevisiae* , physical tethering between plasmids and chromosomes was observed (Ghosh, Hajra, Paek, & Jayaram, 2006). Given that unwinding of DNA is the first hurdle in the initiation of replication, it is possible that a spacer might aid in opening the double helical structure of a *Yarrowia* plasmid. A more open helix could facilitate the interaction of plasmids *via* CEN and spacer with the scaffold proteins responsible for chromosome replication and segregation. In the future, to fully elucidate the mechanism implemented by *Y. lipolytica* , comprehensive characterization of relevant features at the sequence level, including base composition, motif, bendability, roll angle, and others, might constitute a plausible strategy to ascertain the rules of design that drive future rational engineering efforts.

An unexpected finding arose from the implementation of the rescue analysis for plasmids bearing the engineered ARS, which revealed that in addition to complete loss, a plasmid could undergo critical intramolecular changes during cell growth. The view that plasmids are often modified through their passage by *Y. lipolytica* is shared in the field (Wong *et al.* 2017; Gu *et al.* 2020; and our personal communications with other groups working with *Y. lipolytica*). The improved fluorescence profile and higher GFP transcription promoted by the extended foreign spacer N4 were not in agreement with the fact that plasmids were retrieved at reduced sizes. This is most likely caused by the focus of a functional cargo during flow cytometry and qPCR-mediated characterizations, but biased recovery of the plasmids guaranteed to maintain the fully functional *E. coli* replicable elements. It is possible that the seemingly improved phenotype is presented by a rearranged plas-

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

mid with changes occurring in the backbone region related to plasmid replication in *E. coli* . To overcome this issue, intuitive strategies could arise from further attempts to make a plasmid resemble a chromosome. For example, a recent study reported that adding terminal telomeres facilitated the assembly of an artificial chromosome-like linear molecule containing metabolic pathways, an ARS, and selection markers with a total length longer than 20 kb in *Y. lipolytica* with high efficiency, and the resulting molecule can be maintained under selective or nonselective conditions (Guo et al., 2020). Collectively, these findings emphasize that the development of a robust plasmid system for a less studied microorganism demands innovative solutions that address the longstanding stability concerns.

Conclusions

Altogether, we were able to establish that spacers are important elements within ARSs that significantly impact plasmid behavior. This work positions a promising ARS rarely employed by the community, which outperforms other more commonly used replicative elements in *Y. lipolytica* . Furthermore, this work highlights the criticality of selecting and designing the genetic parts required to sustain plasmid-borne gene expression that are not only suitable for the target pathways, but also aid in proper plasmid maintenance.

Acknowledgments

Flow cytometry analysis was performed by Dr. Shawn Rigby from the Office of Biotechnology at Iowa State University. This work was supported by the National Science Foundation Grants (1716837 and 1749782) and the Iowa State University bio-based product seed grant.

Declaration of interest

The authors declare that they have no conflict of interest.

References

Amati, B. B., & Gasser, S. M. (1988). Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. *Cell* ,*54* (7), 967–978. https://doi.org/10.1016/0092-8674(88)90111-0

Blazeck, J., Liu, L., Redden, H., & Alper, H. (2011). Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach.*Applied and Environmental Microbiology* , *77* (22), 7905–7914. https://doi.org/10.1128/AEM.05763-11

Bogacz-Radomska, L., & Harasym, J. (2018). β-Carotene—properties and production methods. *Food Quality and Safety* , *2* (2), 69–74. https://doi.org/10.1093/fqsafe/fyy004

Brun, C., Surdej, P., & Miassod, R. (1993). Relationship between scaffold-attached regions, sequences replicating autonomously in yeast, and a chromosomal replication origin in the *Drosophila* rDNA.*Experimental Cell Research* , *208* (1), 104–114. https://doi.org/10.1006/excr.1993.1227

Cao, M., Gao, M., Lopez-Garcia, C. L., Wu, Y., Seetharam, A. S., Severin, A. J., & Shao, Z. (2017). Centromeric DNA Facilitates Nonconventional Yeast Genetic Engineering. *ACS Synthetic Biology* ,*6* (8), 1545–1553. https://doi.org/10.1021/acssynbio.7b00046

Clarke, L., & Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* ,*287* (5782), 504–509. https://doi.org/10.1038/287504a0

Cui, Z., Jiang, X., Zheng, H., Qi, Q., & Hou, J. (2019). Homology-independent genome integration enables rapid library construction for enzyme expression and pathway optimization in*Yarrowia lipolytica* . *Biotechnology and Bioengineering* , *116* (2), 354–363. https://doi.org/10.1002/bit.26863

Darvishi, F., Ariana, M., Marella, E. R., & Borodina, I. (2018). Advances in synthetic biology of oleaginous yeast *Yarrowia lipolytica* for producing non-native chemicals. *Applied Microbiology and Biotechnology* , *102* (14), 5925–5938. https://doi.org/10.1007/s00253-018-9099-x

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Dhar, M. K., Sehgal, S., & Kaul, S. (2012). Structure, replication efficiency and fragility of yeast ARS elements. *Research in Microbiology* , *163* (4), 243–253. https://doi.org/10.1016/j.resmic.2012.03.003

Fitzgerald-Hayes, M., Clarke, L., & Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell* , *29* (1), 235–244. https://doi.org/10.1016/0092-8674(82)90108-8

Fournier, A. P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D. M., Yaver, D., ... Gaillardin, C. (1993). Colocalization of Centromeric and Replicative Functions on Autonomously Replicating Sequences Isolated from the Yeast *Yarrowia lipolytica* . *Proceedings of the National Academy of Sciences of the United States of America* ,*90* (11), 4912–4916. https://doi.org/10.1073/pnas.90.11.4912

Fournier, P., Guyaneux, L., Chasles, M., & Gaillardin, C. (1991). Scarcity of ars sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica* . *Yeast* , *7* (1), 25–36. https://doi.org/10.1002/yea.320070104

Gao, M., Cao, M., Suastegui, M. S., Walker, J., Rodriguez Quiroz, N., Wu, Y., ... Shao, Z. (2016). Innovating a Nonconventional Yeast Platform for Producing Shikimate as the Building Block of High-Value Aromatics. *ACS Synthetic Biology* , *6* (1), 29–38. https://doi.org/10.1021/acssynbio.6b00132

Gao, S., Tong, Y., Zhu, L., Ge, M., Zhang, Y., Chen, D., ... Yang, S. (2017). Iterative integration of multiple-copy pathway genes in *Yarrowia lipolytica* for heterologous β-carotene production. *Metabolic Engineering* , *41* (April), 192–201. https://doi.org/10.1016/j.ymben.2017.04.004

Ghosh, S. K., Hajra, S., Paek, A., & Jayaram, M. (2006). Mechanisms for Chromosome and Plasmid Segregation. *Annual Review of Biochemistry* , *75* (1), 211–241. https://doi.org/10.1146/annurev.biochem.75.101304.124037

Gu, Y., Ma, J., Zhu, Y., Ding, X., & Xu, P. (2020). Engineering *Yarrowia lipolytica* as a Chassis for de Novo Synthesis of Five Aromatic-Derived Natural Products and Chemicals. *ACS Synthetic Biology* , *9* (8), 2096–2106. https://doi.org/10.1021/acssynbio.0c00185

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., ... Dai, J. (2015). YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae* . *Nucleic Acids Research* , *43* (13), e88–e88. https://doi.org/10.1093/nar/gkv464

Guo, Z. P., Borsenberger, V., Croux, C., Duquesne, S., Truan, G., Marty, A., & Bordes, F. (2020). An artificial chromosome ylAC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing. *Communications Biology* , *3* (1). https://doi.org/10.1038/s42003-020-0936-y

Liu, H., Wang, F., Deng, L., & Xu, P. (2020). Genetic and bioprocess engineering to improve squalene production in *Yarrowia lipolytica* . *Bioresource Technology* , *317* , 123991. https://doi.org/10.1016/j.biortech.2020.123991

Liu, L., Otoupal, P., Pan, A., & Alper, H. S. (2014). Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. *FEMS Yeast Research* ,*14* (7), n/a-n/a. https://doi.org/10.1111/1567-1364.12201

Löbs, A.-K., Schwartz, C., & Wheeldon, I. (2017). Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synthetic and Systems Biotechnology* , *2* (3), 198–207. https://doi.org/10.1016/J.SYNBIO.2017.08.002

Lopez, C., Zhao, Y., Masonbrink, R., & Shao, Z. (2020). Modulating Pathway Performance by Perturbing Local Genetic Context. *ACS Synthetic Biology* , *9* (4), 706–717. https://doi.org/10.1021/acssynbio.9b00445

Lv, Y., Edwards, H., Zhou, J., & Xu, P. (2019). Combining 26s rDNA and the Cre-loxP System for Iterative Gene Integration and Efficient Marker Curation in *Yarrowia lipolytica* . *ACS Synthetic Biology* ,*8* (3), 568–576. https://doi.org/10.1021/acssynbio.8b00535

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Ma, J., Gu, Y., Marsafari, M., & Xu, P. (2020). Synthetic biology, systems biology, and metabolic engineering of *Yarrowia lipolytica*toward a sustainable biorefinery platform. *Journal of Industrial Microbiology and Biotechnology* , 1–18. https://doi.org/10.1007/s10295-020-02290-8

Matsuoka, M., Matsubara, M., Daidoh, H., Imanaka, T., Uchida, K., & Aiba, S. (1993). Analysis of regions essential for the function of chromosomal replicator sequences from *Yarrowia lipolytica* ,*237* , 327–333. https://doi.org/https://doi.org/10.1007/BF00279435

Miller, K. K., & Alper, H. S. (2019). *Yarrowia lipolytica* : more than an oleaginous workhorse. *Applied Microbiology and Biotechnology* , *103* , 9251–9262. https://doi.org/10.1007/s00253-019-10200-x

Nakamura, Y., Nishi, T., Noguchi, R., Ito, Y., Watanabe, T., Nishiyama, T., ... Kondo, A. (2018). A Stable, Autonomously Replicating Plasmid Vector Containing *Pichia pastori* s Centromeric DNA.*Applied and Environmental Microbiology* , *84* (15), e02882-17. https://doi.org/10.1128/AEM.02882-17

Park, H. J., Bae, J. H., Ko, H. J., Lee, S. H., Sung, B. H., Han, J. I., & Sohn, J. H. (2018). Low-pH production of d-lactic acid using newly isolated acid tolerant yeast *Pichia kudriavzevii* NG7.*Biotechnology and Bioengineering* , *115* (9), 2232–2242. https://doi.org/10.1002/bit.26745

Rigouin, C., Gueroult, M., Croux, C., Dubois, G., Borsenberger, V., Barbe, S., ... Bordes, F. (2017). Production of Medium Chain Fatty Acids by *Yarrowia lipolytica* : Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase. *ACS Synthetic Biology* ,*6* (10), 1870–1879. https://doi.org/10.1021/acssynbio.7b00034

Schwartz, C. M., Hussain, M. S., Blenner, M., & Wheeldon, I. (2015). Synthetic RNA polymerase III promoters facilitate high efficiency CRISPR-Cas9 mediated genome editing in *Yarrowia lipolytica* .*ACS Synthetic Biology* , acssynbio.5b00162. https://doi.org/10.1021/acssynbio.5b00162

Shao, Z., Luo, Y., & Zhao, H. (2012). Chapter 17 DNA Assembler Method for Construction of Zeaxanthin-Producing Strains of *Saccharomyces cerevisiae* . In J. L. Barredo (Ed.), *Methods in Molecular Biology*(Vol. 898, pp. 251–262). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-61779-918-1_17

Shao, Z., & Zhao, H. (2014). Manipulating Natural Product Biosynthetic Pathways via DNA Assembler. In *Current Protocols in Chemical Biology* (Vol. 6, pp. 65–100). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9780470559277.ch130191

Shao, Z., Zhao, H., & Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways.*Nucleic Acids Research* , *37* (2), 1–10. https://doi.org/10.1093/nar/gkn991

Sorokina, K. N., Taran, O. P., Medvedeva, T. B., Samoylova, Y. V., Piligaev, A. V., & Parmon, V. N. (2017). Cellulose Biorefinery Based on a Combined Catalytic and Biotechnological Approach for Production of 5-HMF and Ethanol. *ChemSusChem* , *10* (3), 562–574. https://doi.org/10.1002/cssc.201601244

Spagnuolo, M., Yaguchi, A., & Blenner, M. (2019). Oleaginous yeast for biofuel and oleochemical production. *Current Opinion in Biotechnology* , *57,* 73-81. https://doi.org/10.1016/j.copbio.2019.02.011

Sun, W., Vila-Santa, A., Liu, N., Prozorov, T., Xie, D., Faria, N. T., ... Shao, Z. (2020). Metabolic engineering of an acid-tolerant yeast strain *Pichia kudriavzevii* for itaconic acid production.*Metabolic Engineering Communications* , *10* , e00124. https://doi.org/10.1016/j.mec.2020.e00124

Thorwall, S., Schwartz, C., Chartron, J. W., & Wheeldon, I. (2020). Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. *Nature Chemical Biology* , *16* (2), 113–121. https://doi.org/10.1038/s41589-019-0452-x

Vernis, L., Abbas, A., Chasles, M., Gaillardin, C. M., Brun, C., Huberman, J. A., & Fournier, P. (1997). An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica* . *Molecular and Cellular Biology* ,*17* (4), 1995–2004. https://doi.org/10.1128/MCB.17.4.1995

Posted on Authorea 2 Dec 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160691063.38058320/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Vernis, L., Poljak, L., Chasles, M., Uchida, K., Casare, S., Gola, Â., ... Fournier, P. (2000). Only Centromeres Can Supply the Partition System Required for ARS Function in the Yeast *Yarrowialipolytica* . *J. Mol. Biol* , *305* , 203–217. https://doi.org/10.1006/jmbi.2000.4300

Wang, J. H., Hung, W., & Tsai, S. H. (2011). High efficiency transformation by electroporation of *Yarrowia lipolytica* .*Journal of Microbiology* , *49* (3), 469–472. https://doi.org/10.1007/s12275-011-0433-6

Wang, X. Y., Zhang, X., Wang, T. Y., Jia, Y. L., Xu, D. H., & Yi, D. D. (2019). Shortened nuclear matrix attachment regions are sufficient for replication and maintenance of episomes in mammalian cells.*Molecular Biology of the Cell* , *30* (22), 2761–2770. https://doi.org/10.1091/mbc.E19-02-0108

Wong, L., Engel, J., Jin, E., Holdridge, B., & Xu, P. (2017). YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica* . *Metabolic Engineering Communications* , *5* , 68–77. https://doi.org/10.1016/J.METENO.2017.09.001

Xu, P., Qiao, K., Ahn, W. S., & Stephanopoulos, G. (2016). Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proceedings of the National Academy of Sciences of the United States of America* ,*113* (39), 10848–10853. https://doi.org/10.1073/pnas.1607295113

Yamane, T., Ogawa, T., & Matsuoka, M. (2008). Derivation of consensus sequence for protein binding site in *Yarrowia lipolytica*centromere. *Journal of Bioscience and Bioengineering* ,*105* (6), 671–674. https://doi.org/10.1263/JBB.105.671

Yamane, T., Sakai, H., Nagahama, K., Ogawa, T., & Matsuoka, M. (2008). Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. *Journal of Bioscience and Bioengineering* ,*105* (6), 571–578. https://doi.org/10.1263/JBB.105.571

Hosted file

`Figures-final.pdf` available at https://authorea.com/users/380787/articles/496621-investigating-the-role-of-noncoding-regulatory-dna-in-plasmid-development-for-yarrowia-lipolytica