Applied Microbiology and Biotechnology (2021) 105:5959–5972
https://doi.org/10.1007/s00253-021-11399-4

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

Revisiting the unique structure of autonomously replicating sequences in *Yarrowia lipolytica* and its role in pathway engineering

Carmen Lopez[1,2] · Mingfeng Cao[3,4] · Zhanyi Yao[2,5] · Zengyi Shao[1,2,4,5,6]

Received: 17 December 2020 / Revised: 19 May 2021 / Accepted: 24 May 2021 / Published online: 6 August 2021

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 2.2-fold increase in gene expression and 1.7-fold 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.*

Key points

- *In Y. lipolytica, an ARS comprises an origin of replication, a spacer, and a centromere.*
- *The unannotated spacer present in the wild-type ARS stabilizes episomal expression.*
- *Plasmid stability and intramolecular rearrangements exhibit a clear cargo dependency.*

Keywords Plasmid stability · *Yarrowia lipolytica* · Autonomously replicating sequence · Plasmid loss and intramolecular rearrangements · 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

✉ Mingfeng Cao
mifcao@gmail.com

✉ Zengyi Shao
zyshao@iastate.edu

[1] Interdepartmental Microbiology Program, Iowa State University, Ames, IA 50011, USA

[2] NSF Engineering Research Center for Biorenewable Chemicals, Iowa State University, Ames, IA 50011, USA

[3] Department of Chemical and Biological Engineering, University of Illinois, Urbana, IL 60801, USA

[4] DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL 61801, USA

[5] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA 50011, USA

[6] The Ames Laboratory, Ames, IA 50011, USA

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 et al. (2017); Thorwall et al. (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 et al. 2018; Miller and Alper 2019; Spagnuolo et al. 2019; Xu et al. 2016). Manipulation of the *Y. lipolytica* genome has been demonstrated by various 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 et al. 2019) or the presence of numerous ribosome DNA clusters that serve as reusable "landing pads" for integration of heterologous DNA (Lv et al. 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 et al. 2015; Wong et al. 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).

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 et al. 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 et al. 2020; Ma et al. 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 et al. 2020). 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 et al. 2012). In some species such as *S. cerevisiae*, *Pichia pastoris*, and *Scheffersomyces stipitis*, autonomously replicating sequences (ARS) and CEN elements do not need to be associated in the same chromosome when being isolated (Cao et al. 2017; Clarke and Carbon 1980; Nakamura et al. 2018). On the contrary, it was demonstrated in an earlier report that successful plasmid replication in *Y. lipolytica* requires the obligatory presence of both origin of replication (ORI) and centromere (CEN), joined by a few hundred base pairs (Vernis et al. 1997). Isolation of ARSs that enable plasmid construction in *Y. lipolytica* can be traced back to the early 1990s. In the counted examples of ARS isolation from the *Y. lipolytica* genome, it has been identified that ARS sequences are composed of an ORI, a CEN, and a linker sequence that varies in length (Fournier et al. 1991; Matsuoka et al. 1993). To date, four ARSs have been discovered and utilized at various frequencies (Fig. 1 and Supplemental Table S1). ARS1, ARS2, ARS18, and ARS68 were isolated from chromosomes 1, 5, 3, and 1, respectively. Compared to these four ARSs, there is also an ORI-CEN sequence whose usage has become more popular in recent years. In fact, ORI-CEN is a minimized version of ARS1 lacking the linker sequence between ORI and CEN (Matsuoka et al. 1993; Yamane et al. 2008a, 2008b) (often labeled as ori1001-CEN1-1 in the literature).

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. 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, which uncovered a clear cargo-dependency of plasmids with the engineered ARSs and suggested an apparent non-modularity feature of the ARS sub-elements. In practice, plasmid loss occurred in terms of stability and integrity. This work underscores the importance of ARS selection for achieving an optimal behavior of a plasmid platform and the obligated characterization of compatibility of the target pathway with the plasmid backbone.

Materials and methods

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

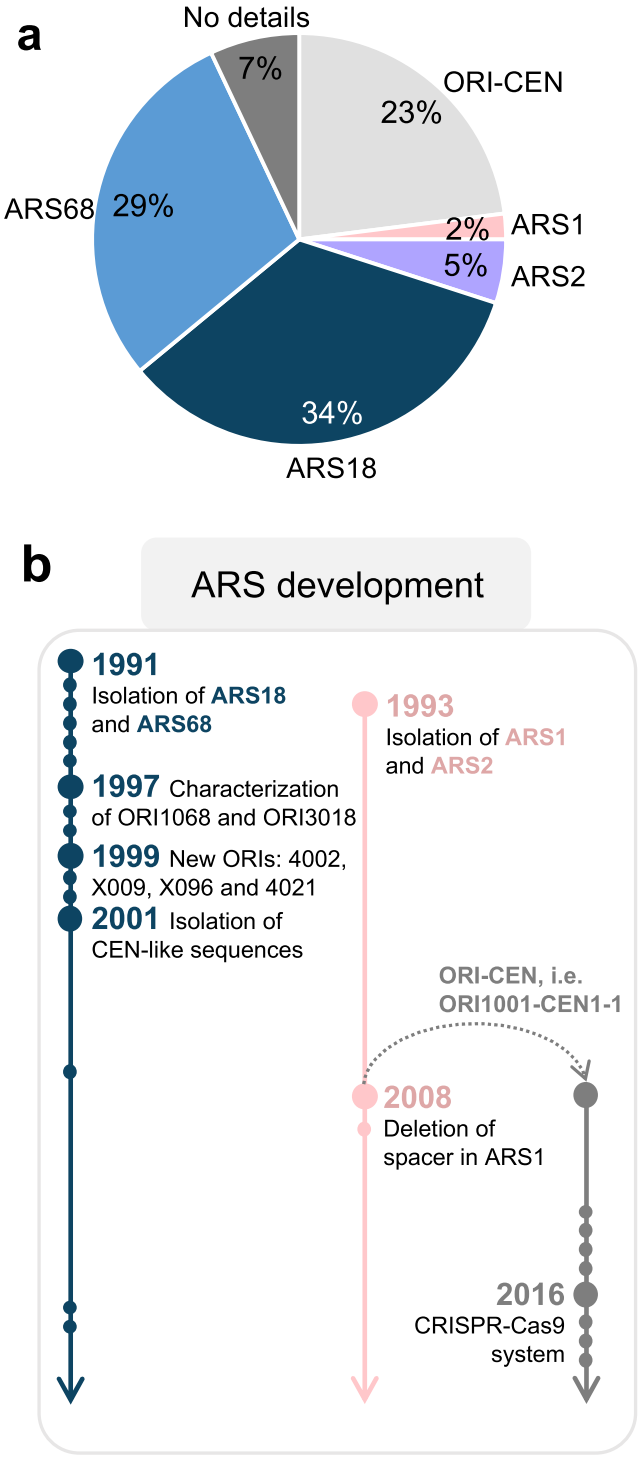


Fig. 1 Autonomously replicating sequences (ARSs) employed in plasmid systems for *Y. lipolytica*. **a** Distribution of ARS usage in literature. The sample consists of 46 research articles. Description of methodology is provided in the Supplementary Information (Supplemental Table S1). **b** Timeline for ARS development. The schematic shows the usage of ARS elements in different research eras, with ORI-CEN (labeled as ORI1001-CEN1-1 in literature) being the preferred choice in recent years. Large dots indicate the hallmarks in ARS characterization, and small dots represent subsequent publications reporting the use of the ARS shown at the top of each branch

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 ($lacI^q$ $rrnB_{T14}$ $\Delta lacZ_{WJ16}$ $\Delta phoBR580$ $hsdR514$ $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$ $galU95$ $endA_{BT333}$ $uidA(\Delta MluI)::pir^+$ *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 due to their high transformation efficiency. The strains and corresponding genotypes are summarized in Supplemental Table S2.

Plasmid construction and transformation

We employed DNA assembler (Shao et al. 2009, 2012; Shao and Zhao 2014) to construct all our plasmids, which are listed in Supplemental Table S2. In the construction of the green fluorescent protein (GFP) plasmid series, plasmid p3887 (a gift from Dr. Suzanne Sandmeyer at University of California, Irvine) was used as a template to amplify ori1001 and CEN1-1, pCRISPRyl (Schwartz et al. 2015) (Addgene #70007) was used to obtain ARS18, and genomic DNA was used for ARS1 and ARS2. 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) and whole ARS replacement were achieved by restriction of pSCARS1 by *Cla*I and *Bam*HI (New England Biolabs, Ipswich, MA) and coupling with the desired ARS sequence amplified from gBlocks® (IDT, Coralville, IA) (Supplemental 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 et al. 2011).

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 FACSCanto flow cytometer (BD Biosciences, San Jose, CA, USA), and the distribution and percentage of GFP^+ cells were calculated with the accompanying software. Wild-type *Y. lipolytica* was included as a negative control in each of the flow cytometry runs for careful threshold gating.

qPCR for transcription assay and plasmid copy number determination

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 N4-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 et al. 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 as an internal standard. Gene quantification was performed using the corresponding standard curve and was reported as relative gene expression. Data was first normalized with respect to *ACT1* and then normalized to the percentage of GFP^+ cells collected at 24-h growth.

Plasmid copy number was determined for the ARS variants from the above GFP-expressing constructs and for the β-carotene producing variants carrying ORI-CEN, ARS1, ORI-foreign sp-CEN, and ORI-EXT-foreign sp N4-CEN. Pellets of OD_{600}=2 were collected at 24 h for total DNA extraction following a reported protocol (Moriya et al. 2006), and 8 μL of the extracts were used as templates. Copy number calculation was performed using the Pfaffl method for GFP plasmids; primer efficiencies were accounted, and a strain with a single copy of the GFP expression cassette integrated into the *Y. lipolytica* genome was used as a calibrator sample (Pfaffl 2001). To calculate the copy numbers of the β-carotene plasmids, ΔCt was employed. Data was first normalized with respect to *ACT1* and then normalized to the percentage of GFP^+ cells collected at 24-h growth. Both assays were performed in triplicate.

β-carotene quantification

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. 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).

Mitotic stability determination

Mitotic stability of plasmids containing assorted ARS sequences was assessed using a classical assay that is widely reported (Dani and Zakian 1983). Cells were streaked from frozen stocks to SC-URA plates and incubated at 30 °C for 2 days. For each construct, three colonies were inoculated to 3.5 mL SC-URA and allowed to grow overnight. Generally, OD_{600} was close to 1; for more concentrated cultures, cells were diluted by PBS buffer. Serial dilutions

were prepared up to 10^{-5} in 96-well plates, and 5 μL were spotted in both SC-URA and YPAD plates. Plates were incubated at 30 °C for 24 h, and colonies were counted. The percentage of mitotic stability was calculated as the ratio of colony forming units (CFU) grown on selective plates and YPAD plates × 100.

Estimation of fragment loss in plasmids by colony PCR

From the freshly streaked plates used for the mitotic stability determination, three additional colonies were picked for each construct to provide DNA template for colony PCR. To facilitate membrane disruption, colonies were dissolved in 10 μL of 0.02 M NaOH and heated at 100 °C for 10 min. One microliter solution was used in each PCR reaction. To evaluate potential fragment loss occurring in plasmids, the presence of relevant fragments of interest was inspected for the six GFP constructs and the four β-carotene constructs varying in ARS. For the GFP plasmids, the selected fragments of interest included the GFP expression cassette (expected phenotype), ARS (essential for *Y. lipolytica* growth), and a segment with the ampicillin cassette joined to the ORI of *E. coli* (labeled as AMP; dispensable for *Y. lipolytica* growth). For β-carotene constructs, the carRP expression cassette was inspected in lieu of the GFP cassette. The data is reported as a color-coded score that ranges from 0 to 100%, where 100% indicates that all three inspected colonies have lost the corresponding region on the plasmid.

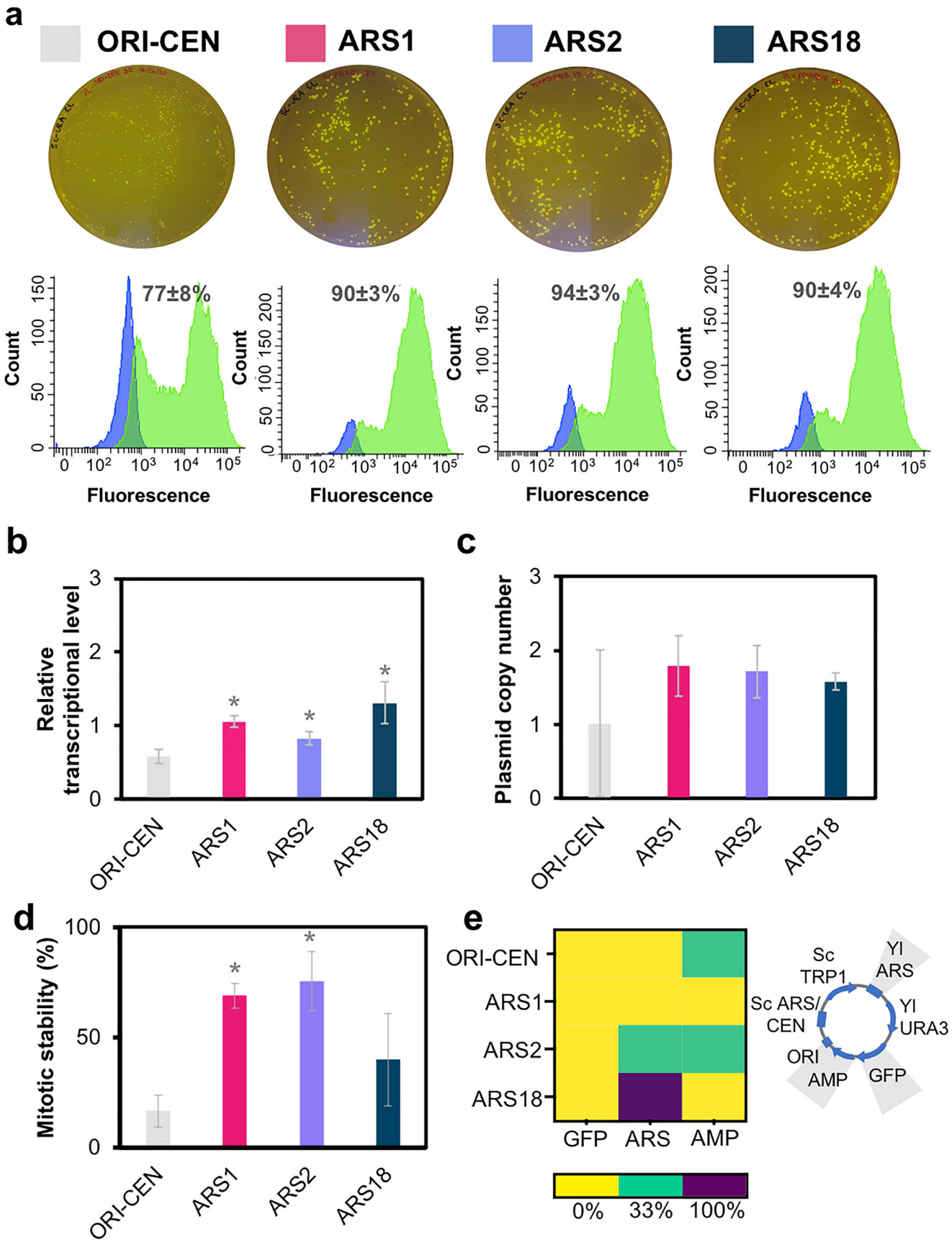
Results

The effect of ARS selection on plasmid behavior

One of the longstanding issues with *Y. lipolytica* plasmid platforms is 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 species like *S. cerevisiae*, *S. stipitis*, *P. pastoris*, and *Issatchenkia orientalis* (Cao et al. 2017; Fitzgerald-Hayes et al. 1982; Nakamura et al. 2018; 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 constitute 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 (Vernis et al. 1997). This observation led us to hypothesize that the spacer (i.e., the linker) between CEN and ORI in a 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 ARS18 and ORI-CEN and the less popular ones ARS2 and ARS1 (Fig. 1a). ORI-CEN was generated from ARS1 by removing the spacer (Yamane et al. 2008b). We did not include ARS68 because it was just labeled differently for being sourced from a different strain and in fact shares a high similarity with ARS1 (Fournier et al. 1991; Matsuoka et al. 1993). A schematic to illustrate the similarity of ARS1 and ARS68 is presented in Supplemental Fig. S1. Supplemental Table S3 lists the sequences of all the ARSs in this study, and their relations are shown in Fig. 1b.

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 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 (Fig. 2a). Flow cytometry revealed that the population bearing a wild-type ARS had more than 90% GFP^+ cells compared to the cells containing ORI-CEN, in which 77% of the cells were GFP^+. Histograms also showed that GFP expression exhibited a trimodal 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 blue peak in Fig. 2a; determined by comparing to the profile presented by wild-type *Y. lipolytica*). ARS1, ARS2, and ARS18 had comparable percentages of GFP^+ cells that were significantly different from that from ORI-CEN ($p<0.05$). Taking the percentage of GFP^+ cells and the fluorescence profiles 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 yielded 15–20% more GFP^+ cells than the minimal ORI-CEN sequence, and this impact appeared to be sustained over a long culture period (Supplemental Fig. S2).

We sought to investigate the potential roots for the desirable effects that the wild-type ARSs conferred to plasmid-borne expression. Therefore, we asked whether choosing a spacer-containing ARS affected the transcriptional level of the reporter gene. qPCR results demonstrated that simply swapping the minimal ORI-CEN by a wild-type ARS yielded a significant increase in GFP expression ranging from 1.4 to 2.2-fold, with the latter obtained when using ARS18 (Fig. 2b). In spite of the observed upregulation, plasmid copy number did not differ significantly (Fig. 2c) and about 1–2 copies were kept per cell, which was in


a
ORI-CEN
ARS1
ARS2
ARS18
77±8%
90±3%
94±3%
90±4%
Count
Fluorescence
b
Relative transcriptional level
c
Plasmid copy number
d
Mitotic stability (%)
e
GFP
ARS
AMP
0%
33%
100%
Sc TRP1
YI ARS
Sc ARS/ CEN
YI URA3
ORI
AMP
GFP

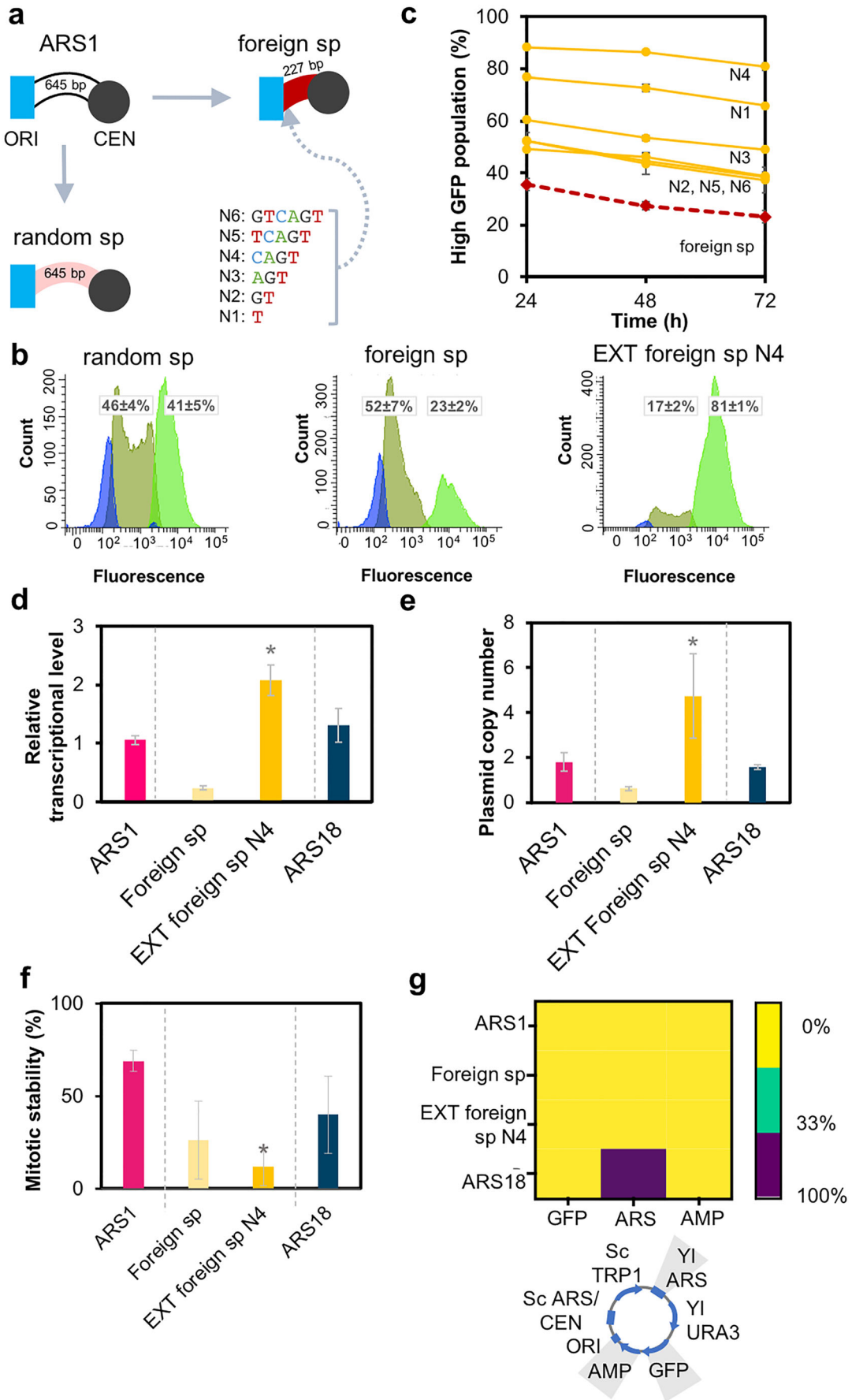
◀ **Fig. 2** Evaluation of plasmid stability with the selected wild-type ARSs. **a** Transformation plates after incubation at 30 °C for 2 days and representative fluorescence profiles generated by flow cytometry after 24-h cultivation (GFP^+ shown in green; plasmid-less cells shown in blue). A percentage for GFP^+ population is presented with a standard deviation of three biological replicates. **b** Determination of the *GFP* transcriptional level with respect to *ACT1* and normalized to the percentage of GFP^+ cells collected at 24-h growth. **c** Determination of plasmid copy number normalized to the percentage of GFP^+ cells collected at 24-h growth. **d** Comparison of mitotic stabilities for different ARS variants. **e** Inspection of plasmid fragment loss by colony PCR. Three target regions are presented in gray shades in the simplified plasmid map, including the GFP expression cassette, ARS, and the ampicillin cassette joined to the *E. coli*'s ORI. Three colonies were randomly picked for each construct, and the absence of the target region is represented with a percentage. Yellow corresponds to 0% and indicates no loss of the fragment of interest in all three inspected colonies. Purple corresponds to 100% and indicates the absence of the fragment of interest in all three colonies. For (b), (c), and (d), significant differences (*) with respect to ORI-CEN were determined by a two-sample *t*-test ($p < 0.05$). Error bars represent error propagation in (b) and (c) and standard deviation of three biological replicates in (d).

accordance with a previous report (Matsuoka et al. 1993). Attention needs to be drawn to a large variation in plasmid copy number of the ORI-CEN construct: since compared to the wild-type ARSs, ORI-CEN causes a more pronounced heterogeneity of cells, with appropriately 23% of the population losing the corresponding plasmid (the blue peak in Fig. 2a). Owing to this observation and considering that a stable plasmid should be maintained in cells for a longer period, we assessed plasmid stability by a classical replica-plating assay (Dani and Zakian 1983). For this study, cells were grown selectively overnight in the SC-URA medium and then spotted on both selective and rich plates (SC-URA and YPAD) in parallel to yield a ratio reported as mitotic stability. Figure 2d suggests that inclusion of either ARS1 or ARS2 increased stability from 17% obtained with ORI-CEN to up to 69–75%. Considering that unstable plasmids might also undergo intramolecular rearrangements and lose some regions, another parameter that we investigated was plasmid integrity. Assessment of plasmid integrity usually involves enrichment of the shuttle plasmids by *E. coli* after their isolation from yeast; however, such a procedure is biased towards the molecules that unequivocally maintain the elements responsible for functioning in bacteria. To avoid this bias and obtain a more accurate snapshot of any potential fragment loss occurring already in *Y. lipolytica*, colony PCR was performed for three selected regions including the GFP expression cassette, the ARS element, and the ampicillin cassette together with *E. coli*'s ORI (labeled AMP). To this end, we used a percentage of fragment loss that represents the absence of the targeted fragment after inspection of three colonies. As expected, the GFP expression cassette was present regardless of the ARS utilized (Fig. 2e). Selection of an ARS impacted the maintenance of both the ARS and the AMP fragment. Analysis of the data showed that only ARS1 allowed the recovery of all three inspected fragments which is indicative of a higher likelihood of plasmid integrity.

Overall, these results suggest that among the few available ARSs, the 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 exhibited slightly lower transcriptional levels compared to ARS18, they exhibited high mitotic stability, but only ARS1 could more likely support the maintenance of an intact plasmid.

Characterization of plasmids with synthetic ARSs

After confirming that the spacer could potentially modulate plasmid replication and segregation, we were curious about whether an ARS demands any sequence-specificity of the spacer for a desirable performance. We hypothesized that the length and the actual sequence of the spacer represent two parameters implicated in an 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 to as random sp hereafter) (Fig. 3a and Supplemental Table S3 for sequences). Comparison of the histograms in Fig. 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 72 h as opposed to 41% for random sp). Interestingly, the replacement of the native spacers in ARS2 and ARS18 by the spacer in ARS1 hinted that the benefit associated with the ARS1 spacer could be maintained, as these swaps did not drastically impact the percentage of GFP^+ cells as compared with their wild-type counterparts (Supplemental Fig. S3). These observations shaped a new hypothesis that consists of the adjustment of the distance between ORI and CEN to potentially modulate the coordination of these two functioning units. 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 Fig. 3a. The performance of plasmids containing these synthetic ARSs was evaluated in a fashion like that applied to the wild-type ARSs. The resulting GFP profiles provided an 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 where 37–81% of the populations could be classified as high GFP at 72 h (Fig. 3c and Supplemental Fig. S4). Among them, extension of the foreign spacer by four nucleotides (referred


a
ARS1
645 bp
ORI
CEN
foreign sp
227 bp
random sp
645 bp
N6: GTCAGT
N5: TCAGT
N4: CAGT
N3: AGT
N2: GT
N1: T
c
High GFP population (%)
Time (h)
N4
N1
N3
N2, N5, N6
foreign sp
b
random sp
foreign sp
EXT foreign sp N4
46±4%
41±5%
52±7%
23±2%
17±2%
81±1%
Count
Fluorescence
d
Relative transcriptional level
ARS1
Foreign sp
EXT foreign sp N4
ARS18
e
Plasmid copy number
ARS1
Foreign sp
EXT Foreign sp N4
ARS18
f
Mitotic stability (%)
ARS1
Foreign sp
EXT foreign sp N4
ARS18
g
ARS1
Foreign sp
EXT foreign sp N4
ARS18
GFP
ARS
AMP
0%
33%
100%
Sc TRP1
YI ARS
Sc ARS/ CEN
YI URA3
ORI
AMP
GFP

◀ **Fig. 3** Engineering the spacer to improve plasmid stability. **a** Schematic depicting the spacer swap in ARS1. A random sequence or a 227-bp spacer from ARS18 replaces the spacer in ARS1. Further modifications include extension by a few nucleotides at the 5′-end of the ARS18 spacer (red). **b** Representative fluorescence profiles obtained by flow cytometry after 72-h cultivation. Profiles for additional extensions (N1-N6) are presented in Supplemental Fig. S4. High GFP^+ cells are shown in green, low GFP^+ in olive, and plasmid-less cells in blue. Each percentage for GFP^+ population is presented with a standard deviation of three biological replicates. **c** Effect of spacer extension on the fraction of cells showing high GFP^+ signals. Red dotted line represents the wild-type ARS18 spacer (foreign sp) used as the negative control. **d** Determination of *GFP* transcriptional level with respect to *ACT1* and normalized to the percentage of GFP^+ cells collected at 24-h growth. **e** Determination of plasmid copy number normalized to the percentage of GFP^+ cells collected at 24-h growth. **f** Comparison of mitotic stabilities for synthetic ARS variants. **g** Inspection of plasmid fragment loss by colony PCR. Three target regions are presented in gray shades in the simplified plasmid map, including the GFP expression cassette, ARS, and the ampicillin cassette joined to the *E. coli*'s ORI. Three colonies were randomly picked for each construct, and the absence of the target region is represented with a percentage. Yellow corresponds to 0% and indicates no loss of the fragment of interest in all three inspected colonies. Purple corresponds to 100% and indicates the absence of the fragment of interest in all the inspected colonies. Statistical analysis was performed using a two-sample *t*-test ($p < 0.05$). For d), e) and f), significant differences only for the extended foreign spacer N4 are shown (*); these were determined with respect to ARS1, ARS18, and the foreign spacer. Error bars represent error propagation in (d) and (e) and standard deviation of three biological replicates in (c) and (f)

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 with the use of any wild-type ARS (Fig. 3b, c). However, appending the same four extra nucleotides to the 3′-end of the foreign spacer did not lead to a desirable GFP distribution as most of the cells depicted lower levels of fluorescence, suggesting a directionality in the interaction of the N4 sequence with the surrounding genetic context (Fig. 3a and Supplemental Fig. S3a).

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 showed that the sequence harboring the extended foreign spacer N4 was the only synthetic ARS that presented an increased *GFP* transcription, which was about 2-fold compared to ARS1 and ARS18 (Fig. 3d). The low level of *GFP* transcription generated by the sequence with the foreign spacer appeared to be consistent with its low copy number (Fig. 3d) and the fact that the population was composed of a large fraction of the GFP^+ cells with low fluorescence (52%) (Fig. 3b). However, the ORI-EXT foreign sp N4-CEN plasmid fared poorly in mitotic stability, reaching only 12% compared to 69% exhibited by ARS1 (Fig. 3f). In terms of evaluating potential fragment loss occurring in the plasmids, the use of synthetic ARSs appeared to be favorable since the three inspected fragments were kept as good as when ARS1 was used (Fig. 3g). These results suggested that while most of the cells exhibited improved GFP expression with the extension of the spacer by the four nucleotides at the 5′-end, plasmid stability was still interrupted compared to the wild-type ARS1. 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.

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 mitotic stability. 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 N4 could maintain the plasmid copy number and 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 and 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 in SC-URA medium for 96 h. Analysis of different variants involved the implementation of an absorbance-based assay to quantify β-carotene production enabled by plasmids with assorted ARSs including the minimal ARS sequence ORI-CEN, the wild-type ARS1, ORI-foreign sp-CEN, and ORI-EXT foreign sp N4-CEN. Carotenogenic cells bearing the wild-type ARS1 were the best producers, and their titer doubled that of the cells expressing the pathway by the ORI-CEN plasmid (Fig. 4 a and b). The observed trend was consistent with the GFP case illustrated in Fig. 2a when ORI-CEN was compared with ARS1. Importantly, cells with the extended foreign spacer N4 were among the lowest producers in this experiment, and the titer was even lower than that of the cells

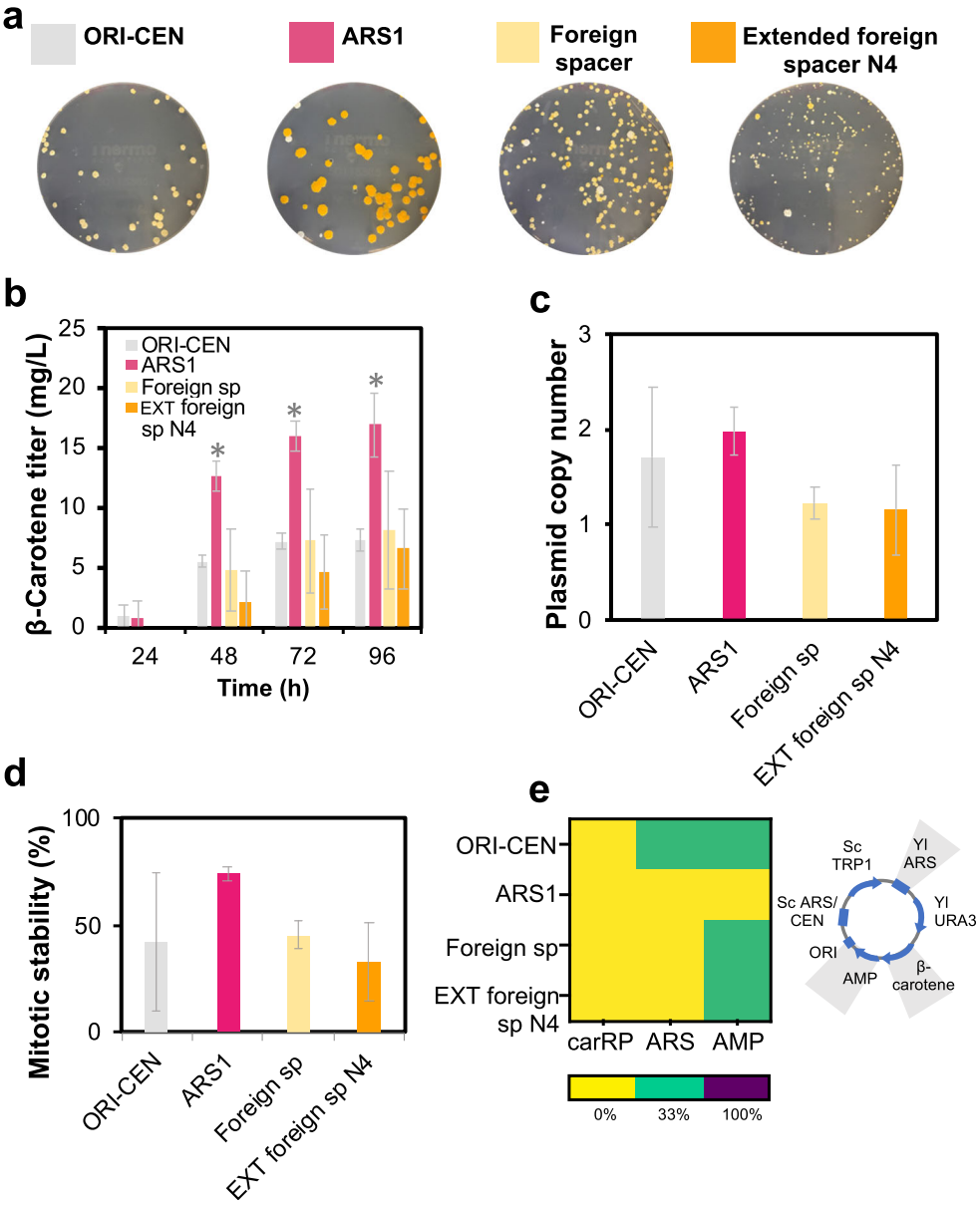


Fig. 4 Assessment of cargo-dependent plasmid stability. **a** The β-carotene pathway was cloned into the plasmids carrying different replicative elements. Transformation plates suggest pronounced differences in production. **b** β-carotene producers were cultured in 12-mL medium, and titers were analyzed by absorbance at 455 nm. **c** Determination of plasmid copy number for the samples collected at 24-h growth. **d** Comparison of mitotic stabilities for synthetic ARS variants. **e** Inspection of plasmid fragment loss by colony PCR. Three target regions are presented in gray shades in the simplified plasmid map, including the carRP expression cassette, ARS, and the ampicillin cassette joined to *E. coli*'s ORI. Three colonies were randomly picked for each construct, and the absence of the target region is represented with a percentage. Yellow corresponds to 0% and indicates no loss of the fragment of interest in all three inspected colonies. Purple corresponds to 100% and indicates the absence of the fragment of interest in all three inspected colonies. Statistical analysis was performed using a two-sample *t*-test ($p < 0.05$), and significant difference in β-carotene production was found for ARS1 with respect to ORI-CEN (*). For (b), (c), and (d), error bars represent the standard deviation of three biological replicates.

with the minimal ORI-CEN sequence. The data presented in Fig. 4c suggests that production was not dependent on plasmid copy number. Taking into consideration that the extended foreign spacer N4 benefited GFP expression but was associated with low mitotic stability, we speculated that the results of low β-carotene production arose from a more pronounced instability issue in the plasmids containing the synthetic ARSs (Fig. 4d). Based on the analysis of fragment loss occurring in the β-carotene plasmids (Fig. 4e), inclusion of a particular ARS impacted plasmid integrity. ARS1 still ranked the highest as it allowed the maintenance of the three inspected fragments, followed by both synthetic ARS variants for which some cells seemed to have lost the AMP fragment while the minimal ARS, i.e., ORI-CEN, supported the maintenance of carRP but not AMP nor the ARS element. The expression cassette for the β-carotene synthetizing gene was confirmed to be present in all the assessed orange colonies regardless of the ARS variant being used. 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* using a synthetic ARS. The increase in copy number and the improved GFP expression observed with the extended spacer N4 were not present for the analogous β-carotene construct.

Considerations for plasmid intactness analysis

The view that *Y. lipolytica* plasmids could undergo intracellular rearrangements during cultivation is shared in the field (Wong et al. (2017); Gu et al. (2020); and our personal

communications with other groups working with *Y. lipolytica*). We attempted to amplify regions of interest via yeast colony PCR, based on which to assess fragment loss occurring in the plasmids during *Y. lipolytica* cultivation (Figs. 2e, 3g, and 4e). The drawback associated with this colony PCR strategy is that a positive amplification cannot guarantee that all the cells belonging to a single colony contain identical plasmids because rearrangements could also occur during colony formation. As mentioned earlier, a more common method to validate plasmid integrity entails isolation of yeast plasmids, transformation to *E. coli* for enrichment, examination of restriction digestion, and plasmid sequencing to identify the changes occurring in plasmids. We attempted such a strategy to compare the β-carotene constructs either with ARS1 or ORI-EXT foreign sp N4-CEN. When the plasmids isolated from *Y. lipolytica* were transformed to *E. coli*, the ARS1-β-carotene construct yielded 80-fold higher number of *E. coli* transformants than the one obtained with ORI-EXT foreign sp N4-CEN (Supplemental Fig. S5a). After five carotenogenic colonies were randomly picked for each construct, all the ARS1-β-carotene construct retrieved from *E. coli* were found to be intact (i.e., I = 1 in Supplemental Fig. S5a). In contrast, only one out of five plasmid isolations on average yielded an intact plasmid with ORI-EXT foreign sp N4-CEN in the backbone (i.e., I=0.2). When incorrect plasmids, as identified per a wrong restriction digestion pattern were sequenced, several sequencing reactions were unsuccessful, indicating that the corresponding primer-annealing regions were absent in the plasmid. Some of the changes that plasmids underwent are summarized in Supplemental Fig. S5b. 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 important to state that such a method has a high degree of bias towards 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 ORI-EXT foreign sp N4-CEN construct suggested that the disruption could likely have occurred on the *E. coli* elements.

Although neither of the two above strategies is ideal, they led to an agreeable conclusion—the wild-type ARS1 stands out as the best replicative element. The benefit of maintaining the natural spacer in ARS1 was sustained for two plasmid-borne cargoes (i.e., the *GFP* gene and the β-carotene pathway). Still, we highly recommend rigorous plasmid stability examination whenever a new cargo is loaded.

Discussion

Considering that ORI-CEN has been generally employed in recent engineering endeavors, the demonstration of the wild-type ARS1 in plasmid design that resulted in enhanced performance allows its prominent usage for *Y. lipolytica* in the future. To the best of our knowledge, this study constitutes the first effort to optimize *Y. lipolytica* as a host using ARS as a target. Another interesting phenomenon is the trimodal GFP expression distribution (Fig. 3b). An on/off bimodal distribution is common in yeasts such as *S. cerevisiae* and *S. stipitis* (Cao et al. 2017). The trimodal expression distribution occurring in *Y. lipolytica* suggests a higher-level heterogeneity, which was most likely caused by the more severe instability issue of the plasmid. As shown in this study, not only the stability heavily depended on the cloned cargo, but the plasmid also experienced various fragment loss or rearrangements, which in turn altered the copy number and ultimately the expression distribution in flow cytometry.

Previously, the Alper group demonstrated a successful promoter engineering approach showing that incorporation of tandem UASs enabled significant increases in promoter strength (Blazeck et al. 2011). Notably, these sequences laid between the ORI and CEN elements. Combined with the results in our study, whether the improved promoter strength was also contributed by altering the distance between ORI and CEN remains as an interesting question. In a separate study (Liu et al. 2014), the same group placed a series of promoter right in front of CEN and obtained a nearly 2.7-fold dynamic range of GFP expression (i.e., from −30% to +80%). Interestingly, such a strategy completely abolished the CEN activity and yielded unstable plasmids in studies conducted in our group. The project involved CEN identification for *S. stipitis*, in which recruiting transcriptional machinery by a strong promoter presumably interrupted kinetochore formation, a key step during plasmid segregation (Cao et al. 2017). Nonetheless, all these studies suggested that the neighboring sequences impact the functioning of the ORI and CEN elements and encouraged us to examine the influence of lengthening the spacer on plasmid stability and gene expression.

The stringent mode of segregation in *Y. lipolytica* hinders the isolation of ARSs. There have been reports of only four ARSs in the last 3 decades, and among those, ARS1 and ARS68 share remarkable similarity despite being sourced from different strains (Fournier et al. 1991; Matsuoka et al. 1993). In an early study that enabled plasmid construction in *Y. lipolytica* in the 1990s, the authors explored the potential role of the spacer in ARS activity by shortening its length from 1002 to 410 bp in an ORI1068-spacer-CEN3 configuration and concluded that spacer length is not critical for plasmid replication (Vernis et al. 1997). This view was strengthened when a minimal ARS lacking the spacer yielded a transformation efficiency of 4×10^3 colonies/μg DNA, albeit with a mitotic stability less than expected (Yamane et al. 2008b). The dispensability of the linker length is contrary to our findings, since we

observed drastic changes when comparison was made between ARS1 and its minimal version, i.e., ORI-CEN. We attribute this discrepancy to the methods employed to survey the impact of altering the distance between ORI and CEN. They relied on the dichotomous question of whether yeast colonies appeared on plates after transformation, which masked the subtle differences in plasmid stability and integrity, not to mention the variation rendered by the cargo.

More exhaustive future studies aimed at understanding the uniqueness of the replication model in *Y. lipolytica* are important, especially those 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 the CENs attached to the spindles) (Amati and Gasser 1988). ORIs in *S. cerevisiae* and higher eukaryotes (e.g., humans) were found to be able to bind to nuclear scaffolds (Amati and Gasser 1988; Wang et al. 2019). Considering that SAR activity was not observed in the ORI 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 et al. 1993). In *S. cerevisiae*, physical tethering between plasmids and chromosomes was observed (Ghosh et al. 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*, more comprehensive characterization of relevant features at the sequence level, including base composition, motif, bendability, and roll angle, might constitute a plausible strategy to ascertain the rules of design that drive rational engineering efforts.

Lastly, to overcome the plasmid instability 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 a 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 concern.

Altogether, we were able to establish that spacers are important elements within ARSs that significantly impact plasmid behavior. This work promotes ARS1, which outperforms other more commonly used replicative elements in *Y. lipolytica*. Furthermore, it highlights the criticality of selecting and designing the genetic parts required to sustain plasmid-borne gene expression that is not only suitable for the target pathway, but also aid in proper plasmid maintenance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-021-11399-4.

Acknowledgements Flow cytometry analysis was performed by Dr. Shawn Rigby from the Office of Biotechnology at Iowa State University.

Author contribution C. L., M. C., and Z. S. conceived the idea; C. L. and M. C. performed the initial experiments to evaluate the effect of ARS selection on plasmid behavior; C. L. and Z. Y. constructed the plasmids with the extended foreign spacer N1 to N6; C. L. performed all the rest of the experiments and analyzed the data; C. L. and Z. S. wrote the manuscript. All the authors read and approved the final manuscript.

Funding This work was supported by the National Science Foundation Grants (1716837 and 1749782) and the Iowa State University bio-based product seed grant. ZS was also partically funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Amati BB, Gasser SM (1988) Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. Cell 54:967–978. 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. Appl Environ Microbiol 77:7905–7914. https://doi.org/10.1128/AEM.05763-11

Bogacz-Radomska L, Harasym J (2018) β-Carotene—properties and production methods. Food Qual Saf 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. Exp Cell Res 208:104–114. https://doi.org/10.1006/excr.1993.1227

Cao M, Gao M, Lopez-Garcia CL, Wu Y, Seetharam AS, Severin AJ, Shao Z (2017) Centromeric DNA facilitates nonconventional yeast genetic engineering. ACS Synth Biol 6:1545–1553. 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 287:504–509. 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*. Biotechnol Bioeng 116:354–363. https://doi.org/10.1002/bit.26863

Dani GM, Zakian VA (1983) Mitotic and meiotic stability of linear plasmids in yeast. Proc Natl Acad Sci U S A 80:3406–3410. https://doi.org/10.1073/pnas.80.11.3406

Darvishi F, Ariana M, Marella ER, Borodina I (2018) Advances in synthetic biology of oleaginous yeast *Yarrowia lipolytica* for producing non-native chemicals. Appl Microbiol Biotechnol 102:5925–5938. https://doi.org/10.1007/s00253-018-9099-x

Dhar MK, Sehgal S, Kaul S (2012) Structure, replication efficiency and fragility of yeast ARS elements. Res Microbiol 163:243–253. 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 29:235–244. https://doi.org/10.1016/0092-8674(82)90108-8

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

Gao M, Cao M, Suástegui MS, Walker J, Rodriguez Quiroz N, Wu Y, Tribby D, Okerlund A, Stanley L, Shanks JV, Shao Z (2016) Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. ACS Synth Biol 6:29–38. https://doi.org/10.1021/acssynbio.6b00132

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

Ghosh SK, Hajra S, Paek A, Jayaram M (2006) Mechanisms for chromosome and plasmid segregation. Annu Rev Biochem 75:211–241. 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 Synth Biol 9:2096–2106. https://doi.org/10.1021/acssynbio.0c00185

Guo Y, Dong J, Zhou T, Auxillos J, Li T, Zhang W, Wang L, Shen Y, Luo Y, Zheng Y, Lin J, Chen G-Q, Wu Q, Cai Y, Dai J (2015) YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. Nucleic Acids Res 43:e88–e88. https://doi.org/10.1093/nar/gkv464

Guo ZP, 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. Commun Biol 3:199. https://doi.org/10.1038/s42003-020-0936-y

Liu L, Otoupal P, Pan A, Alper HS (2014) Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. FEMS Yeast Res 14:1124–1127. https://doi.org/10.1111/1567-1364.12201

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

Löbs A-K, Schwartz C, Wheeldon I (2017) Genome and metabolic engineering in non-conventional yeasts: current advances and applications. Synth Syst Biotechnol 2:198–207. 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 Synth Biol 9:706–717. 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 Synth Biol 8:568–576. https://doi.org/10.1021/acssynbio.8b00535

Ma J, Gu Y, Marsafari M, Xu P (2020) Synthetic biology, systems biology, and metabolic engineering of *Yarrowia lipolytica* toward a sustainable biorefinery platform. J Ind Microbiol Biotechnol 47:845–862. 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/10.1007/BF00279435

Miller KK, Alper HS (2019) *Yarrowia lipolytica*: more than an oleaginous workhorse. Appl Microbiol Biotechnol 103:9251–9262. https://doi.org/10.1007/s00253-019-10200-x

Moriya H, Shimizu-Yoshida Y, Kitano H (2006) In vivo robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. PLoS Genet 2:e111. https://doi.org/10.1371/journal.pgen.0020111

Nakamura Y, Nishi T, Noguchi R, Ito Y, Watanabe T, Nishiyama T, Aikawa S, Hasunuma T, Ishii J, Okubo Y, Kondo A (2018) A stable, autonomously replicating plasmid vector containing *Pichia pastoris* centromeric DNA. Appl Environ Microbiol 84:e02882–e02817. https://doi.org/10.1128/AEM.02882-17

Park HJ, Bae JH, Ko HJ, Lee SH, Sung BH, Han JI, Sohn JH (2018) Low-pH production of d-lactic acid using newly isolated acid tolerant yeast *Pichia kudriavzevii* NG7. Biotechnol Bioeng 115:2232–2242. https://doi.org/10.1002/bit.26745

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45

Rigouin C, Gueroult M, Croux C, Dubois G, Borsenberger V, Barbe S, Marty A, Daboussi F, André I, Bordes F (2017) Production of medium chain fatty acids by *Yarrowia lipolytica*: combining molecular design and TALEN to engineer the fatty acid synthase. ACS Synth Biol 6:1870–1879. https://doi.org/10.1021/acssynbio.7b00034

Schwartz CM, Hussain MS, Blenner M, Wheeldon I (2015) Synthetic RNA polymerase III promoters facilitate high efficiency CRISPR-Cas9 mediated genome editing in *Yarrowia lipolytica*. ACS Synth Biol acssynbio.5b00162 5:356–359. https://doi.org/10.1021/acssynbio.5b00162

Shao Z, Zhao H (2014) Manipulating natural product biosynthetic pathways via DNA assembler. In: Current Protocols in Chemical Biology. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 65–100

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

Shao Z, Luo Y, Zhao H (2012) Chapter 17 DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*. In: Barredo JL (ed) Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 251–262

Sorokina KN, Taran OP, Medvedeva TB, Samoylova YV, Piligaev AV, Parmon VN (2017) Cellulose biorefinery based on a combined catalytic and biotechnological approach for production of 5-HMF and ethanol. ChemSusChem 10:562–574. https://doi.org/10.1002/cssc.201601244

Spagnuolo M, Yaguchi A, Blenner M (2019) Oleaginous yeast for biofuel and oleochemical production. Curr Opin Biotechnol 57:73–81 https://10.1016/j.copbio.2019.02.011

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

Thorwall S, Schwartz C, Chartron JW, Wheeldon I (2020) Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. Nat Chem Biol 16:113–121. https://doi.org/10.1038/s41589-019-0452-x

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

Vernis L, Poljak L, Chasles M, Uchida K, Casare S, Gola Â, Ka E, Matsuoka M, Gaillardin C, Fournier P (2000) Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. J Mol Biol 305:203–217. https://doi.org/10.1006/jmbi.2000.4300

Wang JH, Hung W, Tsai SH (2011) High efficiency transformation by electroporation of *Yarrowia lipolytica*. J Microbiol 49:469–472. https://doi.org/10.1007/s12275-011-0433-6

Wang XY, Zhang X, Wang TY, Jia YL, Xu DH, Yi DD (2019) Shortened nuclear matrix attachment regions are sufficient for replication and maintenance of episomes in mammalian cells. Mol Biol Cell 30:2761–2770. 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*. Metab Eng Commun 5:68–77. https://doi.org/10.1016/J.METENO.2017.09.001

Xu P, Qiao K, Ahn WS, Stephanopoulos G (2016) Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. Proc Natl Acad Sci U S A 113:10848–10853. https://doi.org/10.1073/pnas.1607295113

Yamane T, Ogawa T, Matsuoka M (2008a) Derivation of consensus sequence for protein binding site in *Yarrowia lipolytica* centromere. J Biosci Bioeng 105:671–674. https://doi.org/10.1263/JBB.105.671

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