

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Current Research in Microbial Sciences

journal homepage: www.sciencedirect.com/journal/current-research-in-microbial-sciences

Various short autonomously replicating sequences from the yeast *Kluyveromyces marxianus* seemingly without canonical consensus

Babiker M.A. Abdel-Banat^{a,b,*}, Hisashi Hoshida^{c,*}, Rinji Akada^c

^a Date Palm Research Center of Excellence, King Faisal University, Al-Ahsa 31982, Saudi Arabia

^b Department of Crop Protection, University of Khartoum, Shambat 13314, Sudan

^c Department of Applied Molecular Bioscience, Yamaguchi University Graduate School of Medicine, Tokiwadai, Ube, Japan

ARTICLE INFO

Keywords:

Functional validation
Interchanged sequences
KmARS
NHEJ
Transformability

ABSTRACT

Eukaryotic autonomously replicating sequences (ARSs) are composed of three domains, A, B, and C. Domain A is comprised of an ARS consensus sequence (ACS), while the B domain has the DNA unwinding element and the C domain is important for DNA-protein interactions. In *Saccharomyces cerevisiae* and *Kluyveromyces lactis* ARS101, the ACS is commonly composed of 11 bp, 5'-(A/T)AAA(C/T)ATAAA(A/T)-3'. This core sequence is essential for *S. cerevisiae* and *K. lactis* ARS activity. In this study, we identified ARS-containing sequences from genomic libraries of the yeast *Kluyveromyces marxianus* DMKU3-1042 and validated their replication activities. The identified *K. marxianus* DMKU3-1042 ARSs (*KmARSs*) have very effective replication ability but their sequences are divergent and share no common consensus. We have carried out point mutations, deletions, and base pairs substitutions within the sequences of some of the *KmARSs* to identify the sequence(s) that influence the replication activity. Consensus sequences same as the 11 bp ACS of *S. cerevisiae* and *K. lactis* were not found in all minimum functional *KmARSs* reported here except *KmARS7*. Moreover, partial sequences from different *KmARSs* are interchangeable among each other to retain the ARS activity. We have also specifically identified the essential nucleotides, which are indispensable for replication, within some of the *KmARSs*. Our deletions analysis revealed that only 21 bp in *KmARS18* could retain the ARS activity. The identified *KmARSs* in this study are unique compared to other yeasts' ARSs, do not share common ACS, and are interchangeable.

1. Introduction

Duplication of genomes requires precise initiation of DNA replication at replication origins. Eukaryotic replication origins are divergent but generally encompassed binding sites for origin recognition complex (ORC), regulatory sequences, and transcription units (Gilbert, 2001). An essential component of the replication origins is the *cis*-acting autonomously replicating sequence (ARS). ARS has been shown to allow stable maintenance of episomal plasmids within the yeast cell (Liachko and Dunham, 2014). Generally, intergenic sequences that contain more than 75% A-T are potential initiation sites for DNA replication in yeasts (Liachko et al., 2010). In *Saccharomyces cerevisiae*, short sequences less than 100 bp are defined as ARSs that contain 11-17 bp ARS consensus sequence (ACS) in addition to fairly defined flanking sequences (Liachko and Dunham, 2014; Méchali et al., 2013). However, Méchali (2010) reported that the presence of an ACS is not sufficient to predict a functional DNA replication origin because, among the 12,000 ACS

sequences discovered in *S. cerevisiae* genomes, only 400 are active replicators (Nieduszynski et al., 2006). On the other hand, different groups within the genus *Saccharomyces* have varying ARS elements as components of the replication origin (Dhar et al., 2012). Most of *Kluyveromyces lactis* ARSs utilize 50 bp as an ACS motif, which is completely divergent from the canonical *S. cerevisiae* ACS (Liachko et al., 2010) except the ARS101 of *K. lactis* that shares the common ACS of *S. cerevisiae* (Irene et al., 2004). The yeast *Lachancea kluyveri* ARSs require a sequence that is similar but much longer than the ARS consensus sequence well defined in *S. cerevisiae* (Liachko et al., 2011). ARS elements in *Schizosaccharomyces pombe* are more than 1 kb in size, rich in AT residues, but lacking a common sequence motif. High-affinity binding of *S. pombe* ARS to SpORC requires no specific sequence (Clyne and Kelly, 1995; Kelly and Callegari, 2019; Reeves and Nissen, 1990). An ARS of 60 bp was reported as indispensable and adequate to confer ARS function to shuttle plasmids and linear DNAs in the yeast *Candida guilliermondii* (Foureau et al., 2013).

* Corresponding authors at: Date Palm Research Center of Excellence, King Faisal University, Al-Ahsa 31982, Saudi Arabia
E-mail addresses: bahmed@kfu.edu.sa, babikera@hotmail.com (B.M.A. Abdel-Banat), hoshida@yamaguchi-u.ac.jp (H. Hoshida).

<https://doi.org/10.1016/j.crmicr.2021.100053>

Received 25 May 2021; Received in revised form 8 July 2021; Accepted 27 July 2021

Available online 31 July 2021

2666-5174/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The yeast *Kluyveromyces marxianus* DMKU3-1042 is thermotolerant, fast growing on various carbon biomass, cost-effective, and high-temperature ethanol fermenting yeast (Abdel-Banat et al., 2010a; Limtong et al., 2007). It tends to effectively integrate linear DNA fragments randomly into its chromosomes (Nonklang et al., 2008) via its highly active non-homologous end-joining (NHEJ) pathway (Abdel-Banat et al., 2010b) and it does not need homology sequences at the fragment ends for effective recombination unless otherwise its NHEJ pathway is disrupted. To utilize the advantages of the strain, we developed a simple one-step method for NHEJ-based cloning and constructed several *K. marxianus* circular plasmids with different selection markers for recombinant DNA (Hoshida et al., 2014). Using this method, 36 promoters were cloned to express RFP, and promoters' activities and expression profiles were analyzed in a real-time manner (Suzuki et al., 2015). The outstanding notice is that transformation of a mixture of two PCR-amplified DNA fragments could generate correct recombinant DNA in *K. marxianus* and the replication of plasmids within the yeast cells was driven by the 60-bp sequence of *KmARS7* (Hoshida et al., 2014).

In this study, we demonstrate isolation and analysis of more *KmARSs* from the yeast *K. marxianus* DMKU3-1042. Following a simple functional validation approach and post-transformation cellular events, we identified several robust *KmARSs*. In addition, the impact of site-specific mutations and deletions on the activity of some *KmARSs* were determined. We also demonstrate the influence of short interchanged sequences of *KmARSs* on the replication activity. The *KmARSs* reported here indicate that the strain DMKU3-1042 uses various autonomously replicating sequences that have no obvious canonical consensus.

2. Materials and methods

2.1. Strains, media, and transformation procedures

Yeast strains (Table 1) were regularly maintained at 28°C in YPD medium [1% yeast extract, 2% peptone, 2% glucose] or SD medium [0.17% yeast nitrogen base without amino acids and ammonium sulphate (US Biological, MA, USA), 0.5% ammonium sulphate, 2% glucose and required nutrients]. SD(-U) was an SD medium with necessary nutrients but lacking uracil (Ausubel et al., 1999). 5-Fluoroorotic acid (5-FOA) medium was prepared according to the protocol described by Akada et al. (2006). Luria-Bertani (LB) medium containing 100 µg/ml ampicillin (Sigma-Aldrich, MO, USA) was used for the selection of *E. coli* strain DH5α cells that transformed with plasmids bearing the Amp^R marker gene. Solid media contained 2% agar. Yeast strains were grown in fresh YPD plates at 28°C for 1~2 days before being used for transformation experiments. Yeast competent cells were prepared as previously described (Abdel-Banat et al., 2010b). Briefly, a mixture containing final concentrations of 40% w/v polyethylene glycol 3350 (PEG), 200 mM lithium acetate (LiAc), and 100 mM dithiothreitol (DTT) was dissolved in sterilized distilled water. This mixture was referred to

Table 1.
Yeast strains used in this study.

Strain	Genotype	Parental strain	Reference
RAK3596	<i>K. marxianus</i> DMKU3-1042 wild type	-	Limtong et al. (2007)
RAK3605	<i>K. marxianus ura3-1</i>	RAK3596	Nonklang et al. (2008)
RAK3908	<i>K. marxianus ura3-1 ade2-1</i>	RAK3596	Hoshida et al. (2014)
RAK4174	<i>K. marxianus ura3 leu2</i>	RAK3605	Abdel-Banat et al. (2010b)
RAK4736	<i>K. marxianus ura3-1 leu2 ku70::ScLEU2</i>	RAK4174	Abdel-Banat et al. (2010b)
BY4704	<i>S. cerevisiae MATa ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63</i>	<i>S. cerevisiae MATa</i>	Brachmann et al. (1998)

as the transformation mixture (TM). Aliquots of auxotrophic mutant *K. marxianus* cell suspension prepared in the TM retain their competence for up to 14 months when stored at -80°C (Abdel-Banat et al., 2010b). The transformation was accomplished by thawing the yeast competent cells at room temperature, followed by the addition of PCR-amplified linear or plasmid DNA, heat shock for 15 min at 47°C, and then plating on SD(-U) medium for selection.

2.2. Screening and isolation of autonomously replicating sequences from *K. marxianus* (*KmARSs*)

The yeast *K. marxianus* DMKU3-1042 chromosomal DNA and the yeast *S. cerevisiae* shuttle vector pRS316 (Sikorski and Hieter, 1989) were digested with *EcoRI* and *XhoI* restriction enzymes as instructed by the manufacturer (New England Biolabs, MA, USA). The recovered *K. marxianus* DNA was ligated into the digested vector using the T4 DNA ligase kit (New England Biolabs, MA, USA) and the reaction was terminated by heating for 10 min at 65°C. The ligation product was transformed into competent cells of *E. coli*. Approximately 14,959 *E. coli* colonies carrying plasmids with *K. marxianus* chromosomal DNA fragments were pooled from the LB selection plates, cultured overnight in liquid LB medium at 37°C and the recombinant plasmids were extracted and purified from *E. coli* cells using QIAprep® spin miniprep kit (Qiagen). The purified plasmids were transformed again into the *K. marxianus* strain RAK3605 (*ura3-1*) as described previously (Abdel-Banat et al., 2010b). RAK3605 cells that were transformed with the genomic library were cultured in MM(-U) medium to identify the cells that harbor recombinant pRS316 with potential autonomously replicating sequences of *K. marxianus* (*KmARSs*). The recovered cells were spread on YPD plates to produce colonies and subsequently, at least six transformants from each construct were inoculated on 5-FOA plates (Boeke et al., 1987) to detect whether these plasmids can replicate autonomously.

2.3. Sequence identification of *KmARSs*

To identify the sequence of *KmARS*-containing plasmids that confirmed replicating autonomously within *K. marxianus* cells, yeast transformants were cultured individually on MM(-U) liquid media and grown overnight at 28°C. Then plasmids were extracted using a Zymo-prep™ Yeast Plasmid Miniprep Kit II (Zymo Research, Orange, CA, USA) and Zymolyase 100 T (Seikagaku Biobusiness, Tokyo, Japan), as previously reported (Nonklang et al., 2008). Again, the isolated plasmids were cloned in *E. coli* DH5α competent cells and purified as stated in section 2.2. Throughout the empirical work in this study, the concentration of all kinds of DNA was quantified by Qubit® fluorometer (Thermo Fisher Scientific Inc.) using Quant-iT™ dsDNA assay kit. The sequences of *KmARSs* were determined by the cycle sequencing protocols used for the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems™) according to the supplier's instructions. Recombinant pRS316 plasmids with inserted *KmARSs* are listed in Table 2.

2.4. DNA manipulation

PCR was performed using KOD plus DNA polymerase (Toyobo, Osaka, Japan) according to the manufacturer's instructions. The primers used are listed in Table 3. The *S. cerevisiae URA3* gene (*ScURA3*), including its promoter and terminator, was amplified by PCR from BY4704 chromosomal DNA with the following primer pairs: URA3-223 and URA3-300c; URA3-300 and URA3-300c; 9C-URA3-223 and URA3-300c; and 9C-URA3-223 and 3CG9-URA3+880c. The 9C and 3CG9 sequences flanking the *URA3* gene were utilized subsequently in two discrete PCR reactions (Cha-aim et al., 2009, 2012; Hoshida et al., 2014) to anneal the *KmARSs* at either or both ends for further analysis. The minimum active sequences of *KmARSs* (Table 2) were determined empirically by PCR-directed deletion of the *KmARS* sequences from both

Table 2
Recombinant plasmids with *K. marxianus* genomic sequence and *KmARS* characteristics.

Plasmid*	<i>KmARS</i> size (bp)	Minimum functional ARS (bp)	Chromosomal site**	ACS***	GenBank® accession number
pRS316+ <i>KmARS7</i>	789	Nucleotides 201~247: (47)	Chr 3: 1026687~1027475	143-TTTTATATTTT-153	AB861609.1
pRS316+ <i>KmARS11</i>	154	Nucleotides 46~95: (50)	Chr 1: 1488492~1488645	-	MZ514902
pRS316+ <i>KmARS16</i>	1187	Nucleotides 721~770 (50)	Chr 3: 1557209~1556023	-	MZ514892
pRS316+ <i>KmARS18</i>	1349	Nucleotides 116~136: (21)	Chr 3: 861517~860169	-	MZ514893
pRS316+ <i>KmARS22</i>	1175	Nucleotides 1001~1050: (50)	Chr 6: 728944~730118	-	MZ514894
pRS316+ <i>KmARS36</i>	1183	Nucleotides 291~328: (38)	Chr 2: 770606~769424	-	MZ514895
pRS316+ <i>KmARS51</i>	795	Nucleotides 491~550: (60)	Chr 3: 358583~357789	-	MZ514896
pRS316+ <i>KmARS14</i>	1038	NI	Chr 7: 752365~753402	-	MZ514897
pRS316+ <i>KmARS45</i>	568	NI	Chr 4: 1411986~1412553	-	MZ514898
pRS316+ <i>KmARS3</i>	2590	NI	Chr 5: 839614~842203	-	MZ514899
pRS316+ <i>KmARS20F</i>	906	NI	Chr 6: 265137~266042	-	MZ514900
pRS316+ <i>KmARS20R</i>	905	NI	Chr 2:949414~950318	-	MZ514901

*pRS316 is a *S. cerevisiae* CEN6/ARSH4 shuttle vector (Sikorski and Hieter, 1989).

**Sequence coordinates represent the chromosomes of *K. marxianus* DMKU3-1042 (Lertwattanasakul et al., 2015).

***ACS, ARS consensus sequence commonly found in *S. cerevisiae* and *K. lactis* [(A/T)TTTAT(A/G)TTT(A/T)].

NI, Not Investigated.

sides and rejoining the amplified fragments together with the *URA3* gene as described before (Hoshida et al., 2014).

2.5. Functional validation of *K. marxianus* ARSs (*KmARS*s) by linear *KmARS* transformation

To determine the minimum active sequences of *KmARS*s, three steps were followed (Fig. S1A). First, the *ScURA3* gene was amplified by PCR with the primers 9C-*URA3*-223 and 3CG9-*URA3*+880c. Second, a linker of 9Cs (5'-cccccccc-3') or 3CG9 (5'-cccgggccc-3') was designed at the 3' end of *KmARS* primers to anneal the truncated *KmARS*s sequences to the *ScURA3* gene prepared in the first step. Third, short truncated sequences of some *KmARS*s were divided into two parts to design primers. One part was flanked with 9C and the other with 3CG9 in an intention to leave the central joining sequence of the *KmARS* free after running the PCR with both primers (Fig. S1A). These steps were used to identify the minimum active sequences for *KmARS7*, *KmARS11*, *KmARS16*, *KmARS18*, *KmARS22*, *KmARS36*, and *KmARS51* by transforming the *ScURA3*+*KmARS* into *K. marxianus* strain RAK3606 and selection on MM-U and replica-plating on 5-FOA. To examine whether segments of minimum *KmARS*s can be exchanged with each other while retaining the ARS activity, a combination of primer pairs representing discrete *KmARS*s were used to anneal them by PCR at the ends of the *ScURA3* gene as described in the third step above then followed by routine selection and replica-plating procedures (Fig. S1A).

2.6. Analysis of *K. marxianus* ARS consensus sequence (ACS)

To detect the ACS within *KmARS*s, deletions and/or substitutions experiments were performed on the minimum active sequences of *KmARS7*, *KmARS11*, *KmARS18*, *KmARS22*, and *KmARS36*. Deletion primers were designed from the minimum active sequences of *KmARS7* (201-250) and *KmARS36* (291-340) by deleting triple nucleotides at a time, while for *KmARS18* (111-138) primers, deletion of a single base was carried out in addition to single base substitution for all bases. In the case of *KmARS11* (46-105), five nucleotides were deleted at a time from the 3' end and ten nucleotides were deleted at a time from the 5' end. For *KmARS22* (991-1060), ten nucleotides were deleted at a time from either the 5' or 3' side.

3. Results

3.1. Autonomously replicating sequences from *K. marxianus* DMKU3-1042 (*KmARS*s)

In this study, more than twenty-eight plasmids harboring *K. marxianus* DMKU3-1042 autonomously replicating sequences

(*KmARS*s) were isolated from the genomic libraries. Sequencing of the DNA inserts revealed that many of these plasmids with identical insert sequences, and finally, twelve plasmids were identified as having unique *KmARS*s (Table 2). Plasmids found with identical sequence include two pRS316+*KmARS7*, two pRS316+*KmARS16*, two pRS316+*KmARS18*, two pRS316+*KmARS36*, two pRS316+*KmARS45*, five pRS316+*KmARS3*, two pRS316+*KmARS20F*, and two pRS316+*KmARS20R*. The sequences (Fig. S2) were deposited at the GenBank™ database with the accession numbers (MZ514892 through MZ514902). The size of the insert DNAs with *KmARS*s ranged from 154 to 2,590 base pairs. These insert DNAs are distributed in the seven chromosomes of the yeast *K. marxianus* DMKU3-1042 (Lertwattanasakul et al., 2015). Four ARSs (*KmARS7*, *KmARS16*, *KmARS18*, and *KmARS51*) belong to chromosome 3, two ARSs (*KmARS36* and *KmARS20R*) belong to chromosome 2, two ARSs (*KmARS22* and *KmARS20F*) belong to chromosome 6, while a single ARS was identified from chromosome 1 (*KmARS11*), chromosome 4 (*KmARS45*), chromosome 5 (*KmARS3*), and chromosome 7 (*KmARS14*) (Table 2).

3.2. Functional validation of *KmARS*s

We have previously shown that the circular plasmid pRS316 did not replicate in *K. marxianus* DMKU3-1042 but its linear DNA efficiently integrated into the chromosomes of this strain (Abdel-Banat et al., 2010b; Nonklang et al., 2008; Hoshida et al., 2014). In this study, a simple approach based on a linear transformation protocol was adopted to concept-proof the activities of *KmARS*s (Fig. S1A). After series of sequence alignments (Fig. S3) with known ARSs from *S. cerevisiae* (Deshpande and Newlon, 1992) and *Kluyveromyces lactis* (Iborra and Ball, 1994), *KmARS*s sequences ranging from 21 to 70 bp were identified for replication in *K. marxianus* DMKU3-1042. To analyze the sequences more precisely, these *KmARS*s were fused to the *ScURA3* marker gene and subjected to transformation. Upon transformation, the yeast *K. marxianus* uses its NHEJ pathway to attach the ends of these linear constructs to form circular DNA and transformants. However, some transformants may have produced by chromosomal integration of the DNA introduced. To confirm plasmid formation, transformants were inoculated on 5-FOA plates. Yeast cells with autonomously replicating DNA successfully grow on 5-FOA, while cells with chromosomally integrated *ScURA3* gene fail to grow on 5-FOA (Fig. S1A & B). Using this easy functional validation and post-transformation cellular events, truncated but functional sequences of seven *KmARS*s were verified (Fig. 1A). The functional sequences of *KmARS7* (50 bp), *KmARS11* (60 bp), *KmARS16* (70 bp), *KmARS18* (49), *KmARS22* (50 bp), *KmARS36* (50), and *KmARS51* (60 bp) were shown in Fig. 1A. These *KmARS*s replicate very effectively in *K. marxianus* giving at least 40×10^5 CFU μg^{-1} transforming DNA. Replica plating of the transformants regularly

Table 3
Primers used in this study.

Primer name	Sequences (5'→3')
URA3-223	AAGCTTTTCAATTCATCTTTTTTTTTTTG
9C-URA3-223	cccccccccAAGCTTTTCAATTCATCTTTTTTTTTTTG
URA3-300c	TGTTGTGAAGTCATGACACAG
3CG9-URA3+880c	ccccgggccGTAATAACTGATATAATTAATGA
KmARS7 (201-250) 9c	CAAGACTTCTTGAAGTGAAAACCAACTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (216-250) 9c	GAAAACCAACTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-250) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (225-250) 9C	ACTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (219-250) 9c	AACCAACTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (222-250) 9c	CCAACCTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (225-250) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (230-250) 9C	CAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-247) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-244) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-241) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-238) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (226-235) 9c	CTTTCAGTCTTCAAACATAAAATGAccccccccc
KmARS7 (201-218) c-3CG9	TCACTTCAAGAAGTCTTGCccggggccc
KmARS7 (201-221) c-3CG9	TTTTCACTTCAAGAAGTCTTGCccggggccc
KmARS7 (201-224) c-3CG9	TGGTTTTCACTTCAAGAAGTCTTGCccggggccc
KmARS7 (201-225) c-3CG9	TTGGTTTTCACTTCAAGAAGTCTTGCccggggccc
KmARS7 (201-229) c-3CG9	AAAGTTGGTTTTCACTTCAAGAAGTCTTGCccggggccc
KmARS11 (46-105) 10c	TCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (46-60) c-3CG9	AAAGTTAATTTTGAACccggggccc
KmARS11 (61-100) 9C	CTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATcccccccc
KmARS11 (46-100) 9c	TCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATcccccccc
KmARS11 (46-95) 9c	TCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATcccccccc
KmARS11 (76-105) 9C	TATTTTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (46-75) c3CG9	TGACATTTAGCTTGAAGAAGTAAATTTGGAcccggggccc
KmARS11 (26-105) 10c	CACCTTTTACACTGTGACGTTCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (11-90) 10c	AATCAATGATTCATACACTTTTACACTGTGACGTTCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (36-105) 10c	ACTGTGACGTTCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (46-105) 10c	TCCAAAATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (56-105) 10c	ACTTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGAccccccccc
KmARS11 (51-120) 10c	AATTAACCTTCTAAGCTAAATGTCATATTCGCAATAAAAATAAAGAATATAGATATCAAAGTCTGTGcccccccc
KmARS16 (721-780) 9c	TTTTATTTTTTTTAACTCAATTTCCAGTTTAAACACCAAAATACGTTTCCCCcccc
KmARS16 (721-770) 9c	TTTTATTTTTTTTAACTCAATTTCCAGTTTAAACACCAAAATACGTTTCCCCcccc
KmARS16 (731-790) 10c	TTTTAACTCAATTTCCAGTTTAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (741-790) 10c	ATTTCCAGTTTAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (751-790) 10c	TAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (721-752) c-3CG9	AAAATACGTTTCCATATAATTGAAAAGGAAAGccccggccc
KmARS16 (753-790) 9C	GGTGTTTAACTGGAAATGAGTTAAAAAAAATAAAAccccccccc
KmARS16 (731-790) 10c	TTTTAACTCAATTTCCAGTTTAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (741-790) 10c	ATTTCCAGTTTAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (751-790) 10c	TAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS16 (721-790) 10c	TTTTATTTTTTTTAACTCAATTTCCAGTTTAAACACCAAAATACGTTTCCATATAATTGAAAAGGAAAGcccccccc
KmARS18 (111-159) 10C	TCCATAAATTGGAATGAAAAGTCACTTTAGGTTCACTATATAATGAAAACcccccccc
KmARS18 (111-149) c-3cG9	ATAGTGAACTTAAAGTACCTTCAATTTCCAAATTTATGGAccccggggccc
KmARS18 (121-149) 9C	GGAATTTGAAAAGTCACTTTAGGTTCACTATATAATGAAAACcccccccc
KmARS18 (111-138) c-3CG9	AAAGTGACTTCAATTTCAAAATTTATGGAccccggggccc
KmARS18 (139-159) 9C	AGGTTCACTATATAATGAAAACcccccccc
KmARS18 (111-134) c-3CG9	TGACTTTCAATTTCAAAATTTATGGAccccggggccc
KmARS18 (135-160) 9C	CTTTAGGTTCACTATATAATGAAAAGcccccccc
KmARS18 (135-156) 9C	CTTTAGGTTCACTATATAATGAccccccccc
KmARS18 (135-153) 9C	CTTTAGGTTCACTATATAAccccccccc
KmARS18 (135-150) 9C	CTTTAGGTTCACTATAcccccccc
KmARS18 (135-147) 9C	CTTTAGGTTCACTTcccccccc
KmARS18 (135-144) 9C	CTTTAGGTTcccccccc
KmARS18 (135-141) 9C	CTTTAGGcccccccc
KmARS18 (135-138) 9C	CTTTcccccccc
KmARS22 (991-1060) 9c	TGTTATCTTTTTTCGCTTCAAAGTTACTTTGGATTCTAAATAAAGAAAAAATAAAACCAAACCAATCcccccccc
KmARS22 (1001-1060) 9c	TTCCGCTTCAAAGTTACTTTGGATTCTAAATAAAGAAAAAATAAAACCAAACCAATCcccccccc
KmARS22 (1001-1050) 9c	TTCCGCTTCAAAGTTACTTTGGATTCTAAATAAAGAAAAAATAAAACcccccccc
KmARS22 (1001-1020) c-3CG9	AAAGTAACTTTTGAAGCGAACCccggggccc
KmARS22 (1021-1050) 9C	GGATTCAAATATAAAGAAAAAATAAAACcccccccc
KmARS36 (291-340) 10c	TCTTTAATATATTTTTCACTTTCAAAGAGTGTGAAAATAAAATTAATGcccccccc
KmARS36 (291-306) c-3CG9	AAAATAATATTAAGAGAcccggggccc
KmARS36 (307-340) 9C	TCATTTCAAAGAGTGTGAAAATAAAATTAATGcccccccc
KmARS36 (307-337) 9C	TCATTTCAAAGAGTGTGAAAATAAAATTAACcccccccc
KmARS36 (307-334) 9C	TCATTTCAAAGAGTGTGAAAATAAAATTTcccccccc
KmARS36 (307-331) 9C	TCATTTCAAAGAGTGTGAAAATAAAACcccccccc
KmARS36 (307-328) 9C	TCATTTCAAAGAGTGTGAAAATcccccccc
KmARS36 (307-325) 9C	TCATTTCAAAGAGTGTGAAcccccccc
KmARS36 (307-322) 9C	TCATTTCAAAGAGTGTcccccccc

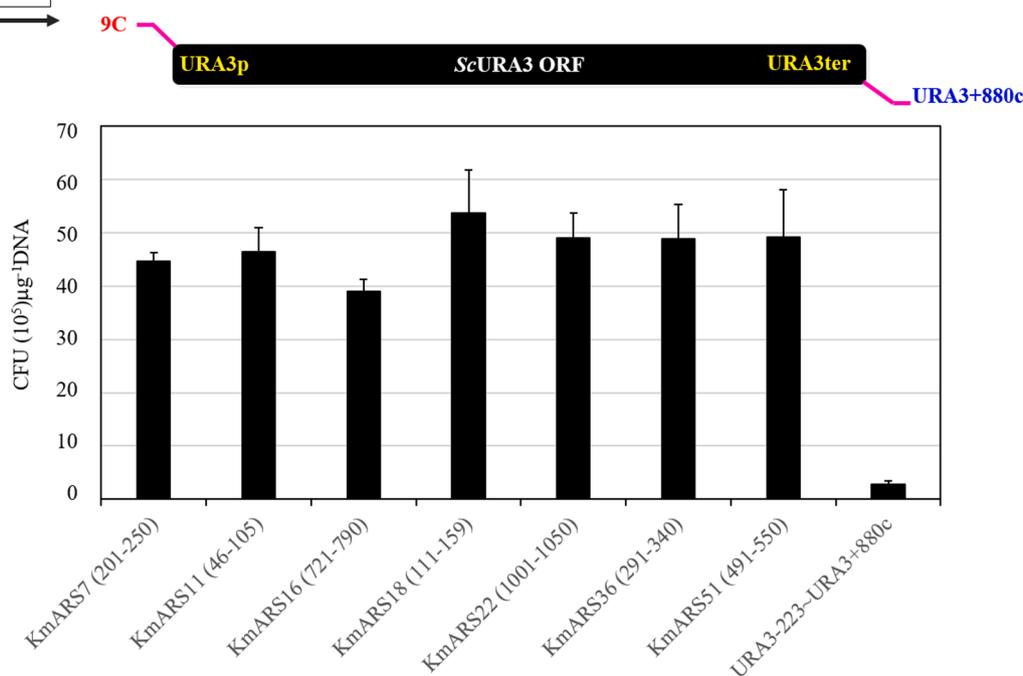
(continued on next page)

Table 3 (continued)

Primer name	Sequences (5'→3')
KmARS36 (316-340) 9C	AAAGTGTGAAATAAAAATTTAAATGccccccccc
KmARS36 (291-315) c3CG9	TTGAAATGAAAAATAATTTAAAGAcceggggccc
KmARS51 (491-550) 9c	AATATTTATGAATAAAAGTAACTTTTGTAGTTTCAAATACAAAAATATTAATTACAAAGccccccccc
KmARS51 (491-515) c-3CG9	AAAGTTACTTTTATTCATAAAATATTccccggggccc
KmARS51 (516-550) 9C	TTAGTTTCAAATACAAAAATATTAATTACAAAGccccccccc

(A)

ARS name	Sequence	(bp)
KmARS7 (201-250):	CAAGACTTCT TGAAGTAAA ACCAACTTTC AGTCTTCAA CTA AAAATGA	(50)
KmARS11 (46-105):	TCCAAAATTA ACTTTCTAAG CTA AATGTCA TATTTGCAA TAAAATAATA AGAATATAGA	(60)
KmARS16 (721-790):	CTTCCTTTTT CAATTATATG GAAACGTATT TTGGTGTTTA AACTGGAAAT TGAGTTAAAA AAAAATAAAA	(70)
KmARS18 (111-159):	TCCATAATTT GGAATTGAAA GTCAC TTTAG GTTCACTATA TAATGAAAA	(49)
KmARS22 (1001-1050):	TTCGCTTCAA AAGTTACTTT GGATTCTAAT ATAAGAAAA AAATAAAAAAC	(50)
KmARS36 (291-340):	TCTTTAATAT TATTTTTCAT TTCAAAAAGT GTGAAATAAA AATTAAAATG	(50)
KmARS51 (491-550):	AATATTTATG AATAAAAAGTA ACTTTTGTAGT TTCAAATACT AAAAAATATT AATTACAAAG	(60)



(B)

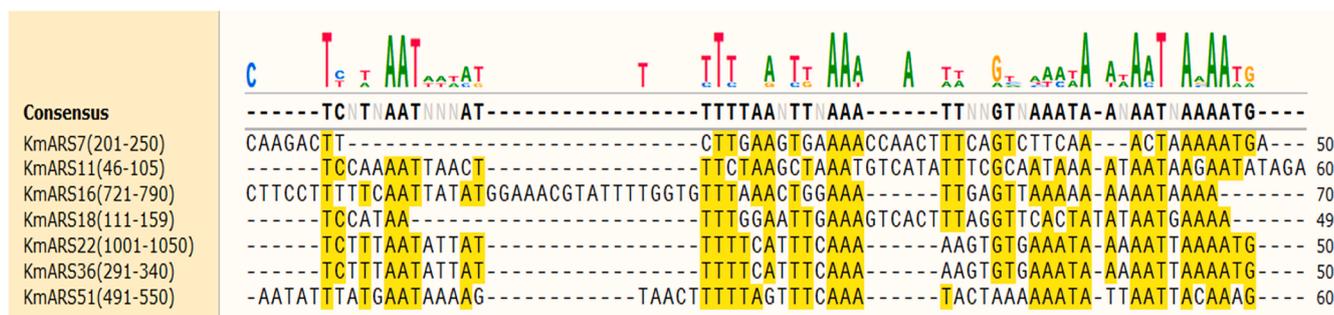


Fig. 1. The activity of short sequences of KmARs. (A) Short sequences of different KmARs ranging from 49 to 70 bp are shown (top panel). Seven KmARS sequences are fused to the ScURA3 gene at its 5' end and the transformation efficacies of these constructs are depicted (bottom panel). Sequence alignment and logos of the short KmARs are depicted in panel (B).

gives more than 80% rescued colonies from 5-FOA toxicity, an indication of intracellular replication as plasmids. It is noteworthy that, the alignment of these short functional sequences showed no prominent common consensus but the AT stretches prevail the sequences (Fig. 1B).

3.3. Impact of truncations and triple nucleotide deletions on the activity of the region 201-250 of KmARS7

We have previously demonstrated that 60 nucleotides of KmARS7 (201-260) effectively drove the replication of the ScURA3 gene

(Hoshida et al., 2014). However, the *KmARS7* retains its potent activity even after further truncations of this region. The region 201-250 gave an average of 40×10^5 colony-forming units (CFU) μg^{-1} transforming DNA, but the number of transformants was dropped drastically when the regions 216-250 or 226-250 were transformed in conjunction with the *ScURA3* gene (Fig. 2A). Truncations of the region *KmARS7* (201-250) were also investigated by triple nucleotide deletions. Two separate primers for each construct were used to amplify the *ScURA3* marker. One primer was *KmARS7* (201-225)c-3CG9 and the other set of primers were *KmARS7* (226-250)9C and its triple nucleotide truncations (Fig. 2B). Deletion of three nucleotides from the 3' end of *KmARS7* (201-250) resulted in the reduction of the transformation efficiency on average by about 25%, while further deletions reduced the transformability of *KmARS7* to a level comparable to the transformation of *ScURA3* gene only. Colonies that appeared on the plates transformed with *KmARS7* (201-244), (201-241), (201-238), and (201-235) are mostly the result of integration activity rather than autonomous replication as judged by replica plating on 5-FOA (data not shown).

3.4. Functional characteristics of *KmARS11*

The whole insert sequence of *KmARS11* is 154 bp. The regions *KmARS11* (46-105) and (46-100) gave comparable high transformability, while the transformability of the regions *KmARS11* (46-95) and (56-105) was declined but produced significant levels of transformants compared with transformation without any ARS. The transformability of the regions *KmARS11* (66-105) was not distinct from that of the *ScURA3* gene (Fig. 3). As a result, 50 bp of *KmARS11* (56-105) retains the transformability.

3.5. Functional characteristics of *KmARS18*

Deletion of seven nucleotides from the 3' end of *KmARS18* (111-160) slightly decreased the transformation efficiency of *KmARS18*. Further triple-nucleotide deletions resulted in the reduction of the transformation efficiency of *KmARS18* on average to levels as low as 34%. Surprisingly, the region *KmARS18* (111-138), which is 28 bp-long, showed elevated transformation efficiency (Fig. 4A). This encouraged us to look for extra ARS active sequences from *KmARS18*. The regions *KmARS18* (111-138) [28 bp], *KmARS18* (139-159) [21 bp], and *KmARS18* (121-149) [29 bp] were tested for transformability. Both the region *KmARS18* (121-149) and the longer *KmARS18* (139-159) showed reduced transformation efficiency relative to *KmARS18* (111-138) (data not shown). This region, *KmARS18* (111-138), was thoroughly investigated by single nucleotide deletion from both sides (Fig. 4B). The deletion of seven nucleotides from the 5' end (TCCATAA) resulted in the generation of fewer transformants. Moreover, an additional single nucleotide deletion from this region completely abolished its transformability. On the other hand, the deletion of four nucleotides from the 3' end (135-CTTT-138) resulted in the elimination of transformability. The region as short as 21 bp-long of *KmARS18* that covers the nucleotides (116-136) was capable to drive efficient transformation (Fig. 4B). Replacement of three nucleotides 131-GTC-133 with CCA, the addition of A at the position 131, deletion of G at position 122, and replacement of the region *KmARS18* (111-TCCATAATT-119) by the introduction of nine nucleotides of *KmARS7* (201-CAAGACTTC-209) at the same site negatively affect the transformation efficiency of the region *KmARS18* (111-138) (Fig. 4B). Furthermore, as shown in Fig. 5A, a single nucleotide substitution in the region *KmARS18* (111-138) induces moderate to weak effect or complete loss of transformability. However, the substitutions at some sites did not affect the transformability and the mutants gave transformants similar to the original sequence. Substitution at the sites T118G, T118C, T119A, T119C, G121C, A128C, A129G, A129T, or A129C made the *KmARS18* (111-138) lose the ability to develop transformants (Fig. 5A). In other cases, very few but small transformants were developed upon base substitution at the sites T111C, A117G,

A117C, T118A, G121T, G122T, T125G, T126A, T126G, G127A, A128T, A130C, G131T, G131C, or T132A (Fig. 5A). Additionally, the region of the 21 nucleotides [*KmARS18* (116-136)] that showed highly efficient transformation (Fig. 4B) was capped by adding five nucleotides, "CGCGC", at its free end after joining it to the marker gene. Transformation of this construct and a similarly capped region *KmARS18* (111-159) as a control, revealed that the region *KmARS18* (116-136) is very sensitive to additional bases at its 3'-end (Fig. 5B) but 5'-capping by the "CGCGC" did not interfere with the efficient transformability of the region *KmARS18* (111-159) [49 bp].

3.6. Functional characteristics of *KmARS22* and *KmARS36*

The region of *KmARS22* (991-1060) [70 bp] with high replication propensity was truncated (Fig. 6A). Regions of *KmARS22* (1001-1060) [60 bp] and *KmARS22* (1001-1050) [50 bp] behave similarly as effective replicators, while the regions *KmARS22* (1021-1050) [30 bp] and *KmARS22* (1001-1020) [20 bp] separately failed to drive the replication process (Fig. 6A). The region (1001-1020) [20 bp] complements the region (1021-1050) [30 bp] to retain the function of *KmARS22* (Fig. 6A).

The region (291-340) of *KmARS36* [50 bp] gave an average of 3.9×10^5 CFU μg^{-1} DNA. Contrary to the other *KmARSs*, the transformability was increased gradually upon deletion of triple nucleotides at a time and reached up to 7.73×10^5 CFU μg^{-1} DNA when nine nucleotides were deleted from the 3' end leaving a region of 41 bp [*KmARS36* (291-331)]. When twelve nucleotides were deleted leaving a region of 38 bp [*KmARS36* (291-328)], the transformability was slightly declined compared with *KmARS36* (291-331) but showed higher transformability than *KmARS36* (291-340), indicating that the 38 bp-long region is still capable to drive the autonomous replication. Further deletions from the 3' end, leaving the regions 291-328 or 291-325, caused the loss of transformability (Fig. 6B). The nucleotides that covering the region 326-ATAAAA-331 are indispensable for the activity of *KmARS36*.

3.7. Impact of *KmARSs* interchanged sequences on transformability

There is clear variation in the sequences among the identified core sequences of *KmARSs* (Fig. 1B) and these *KmARSs* have no sequence identity with the optimized *KIARS* (Liachko and Dunham, 2014) (Fig. S4). Although sequences of the regions *KmARS18* (1181-1240) and *KmARS7* (123-182) have sites with fairly high identity to other yeast ARS consensus sequences (ACS), these regions did not drive efficient transformability relative to their corresponding regions of the *KmARS18* (111-159) and the *KmARS7* (201-250) (Fig. S5). Due to the disparities in the consensus and lengths of the identified *KmARSs*, short sequences of these ARSs were interchanged with each other to judge whether or not they could induce efficient transformability. As shown in Table 4, the majority of various regions of the *KmARSs* when interchanged, they generate in some instances even more transformants than do the corresponding regions of individual *KmARSs*. The most prominent results were the highly efficient transformability of *KmARA18* (111-138) when interchanged with six other *KmARSs* namely *KmARS7* (230-250), *KmARS11* (61-100), *KmARS16* (753-790), *KmARS22* (1021-1050), *KmARS36* (307-340), and *KmARS51* (516-550). Interchanged regions of *KmARS11* (61-100), *KmARS16* (753-790), and *KmARS36* (516-550) respectively with *KmARS7* (201-229), *KmARS11* (46-60), *KmARS16* (721-752), *KmARS18* (111-138), *KmARS22* (1001-1020), *KmARS36* (291-306), and *KmARS51* (491-515) also induced highly effective transformability (Table 4). Meanwhile, these interchanged sequences showed fewer consensus identities and the similarities mainly skewed towards the 3' and 5' ends without clear consensus in the middle (Fig. S6). It is noticeable that transformants from the interchanged constructs gave between 81 to 100% colony growth on 5-FOA.

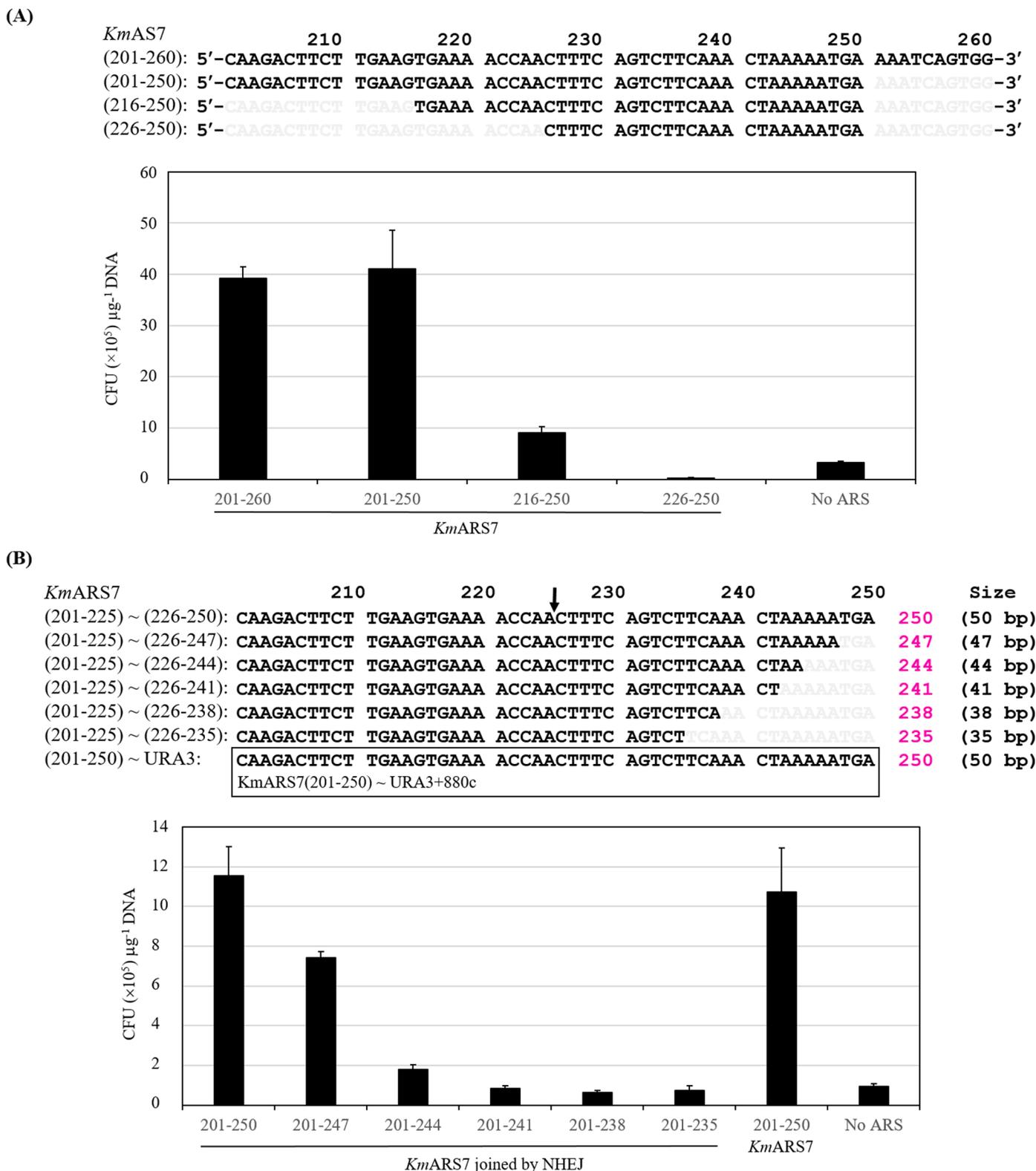


Fig. 2. Impact of truncations and nucleotides deletions on *KmARS7*. (A) *KmARS7* (201-250) and its truncated fragments were fused at the 5' end of the marker gene (top panel). A chart for the transformation efficacy of the regions *KmARS7* (201-260) and *KmARS7* (201-250) and its truncated fragments is depicted (bottom panel). (B) Sequences of *KmARS7* (201-205) and its triple nucleotides deletion fragments are shown on the top and their corresponding transformation efficacies are depicted on the bottom. The values of transformation efficiencies (CFU μg^{-1} DNA) in (A) and (B) are due to the use of different preparations of RAK3605 competent cells. Therefore, the charts in (A) and (B) represent the general patterns of transformability of *KmARS7* (201-250) and its truncation and deletion products.

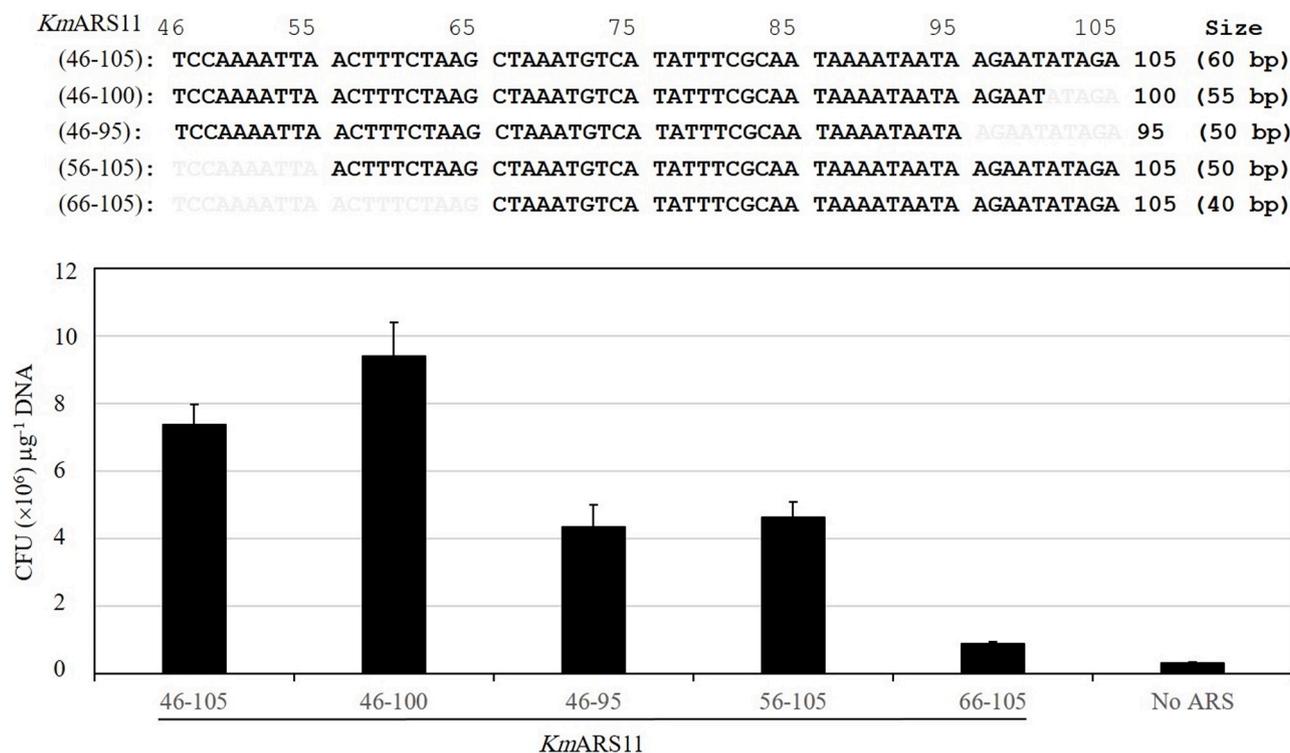


Fig. 3. Effect of 5' and 3' deletions on the activity of the region (46-105) of *KmARS11*. Nucleotide sequences of *KmARS11* (46-105) and its deletion fragments are depicted on the top. A chart of transformation efficiencies of ARSs into the RAK3605 competent cells is shown on the bottom.

4. Discussion

Autonomously replicating sequences (ARS) are the replicator elements to which bind the initiator protein that unwind the DNA double helix and recruits additional factors to initiate the process of DNA replication. The proteins that regulate replication are highly conserved, including the origin recognition complex (ORC), which binds directly to replication origin sequences, but Gilbert (2001) stated, "In several eukaryotic replication systems, it appears that any DNA sequence can function as a replicator". However, many studies on yeast ARS helped to define specific sequences that function as origin replicators in *S. cerevisiae*, *S. pombe*, *K. lactis*, and *C. guilliermondii* (Stinchcomb et al., 1980; Clyne and Kelly, 1995; Irene et al., 2004; Liachko et al., 2010; Liachko and Dunham, 2014; Foureau et al., 2013). Here we report the identification of twelve functional *KmARS* from the strain DMKU3-1042 capable to replicate plasmid DNA but have no common consensus sequences.

Previously, Iborra and Ball (1994) reported the isolation of three small DNA fragments from *K. marxianus* strain ATCC12424 [ARS1 (1267 bp), ARS2 (1206 bp), and ARS3 (1200 bp)]. ARS1 and ARS2 contain both ARS and centromeric elements, while ARS3 contains ARS core sequence only and all function in *K. lactis*. Only two of our *KmARS*s identified in the current study share identity with ARS1 and ARS2 from the strain ATCC12424. One of them is *KmARS3* (Table 2), which shares 89.53% identity to ARS1 and the other is a portion consists of 128 nucleotides from *KmARS20F* that shares 100% identity to ARS2. However, none of the *KmARS*s reported here share significant identity to the ARS3 from the strain ATCC12424, which indicates that in this study ten *KmARS*s are identified for the first time from *K. marxianus*. This might be either the ARS3 replicator is not functional in *K. marxianus* DMKU3-1042 or its rival was missed during our libraries' screening. It has been reported that very similar ACS of nonanucleotide (5'-TTTATGTGTT-3') is common between *K. marxianus* and *K. lactis* (Iborra and Ball, 1994). However, this same ACS is not found in any of the currently investigated *KmARS*s.

In this study, we also identified minimal sequences that function as ARS. These sequences indicated again that ACS found in *S. cerevisiae* and *K. lactis* does not exist in the *K. marxianus* ARSs. In addition, generally within 50-bp *KmARS* sequences, at least 21-bp are functioning as ARS for plasmid replication. Among the identified minimal functional sequences, any clear consensus sequence was not found, indicating that the essential sequence of ARSs in *K. marxianus* are divergent.

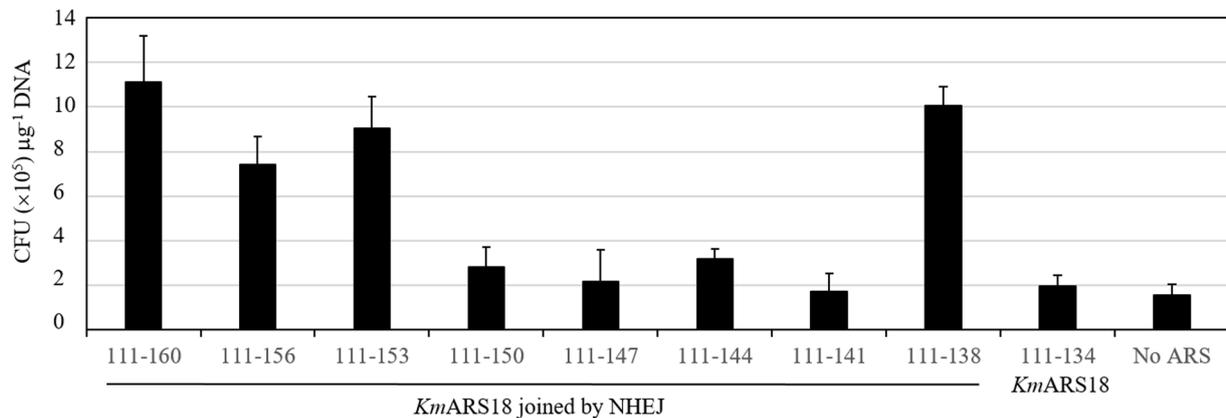
With some exceptions, the majority of the interchanged sequences for *KmARS*s replicate effectively. The sequences 5'-AAA(G/A)T($\times\times\times$) (T/A)TT-3' and 5'-AAAA(T/A)AAAAAT-3' are likely the common consensus in the interchanged *KmARS*s of the strain DMKU3-1042 (Fig. S6) and their sequences position weight matrix logo (Crooks et al., 2004; Liachko et al., 2010) suggest the prevalence of poly-A at the 3' termini of the interchanged sequences. It is noteworthy that, the transformability of the regions *KmARS7* (123-182) and *KmARS18* (1181-1240) is completely different from its counterpart regions *KmARS7* (201-250) and *KmARS18* (111-159) (Fig. S5). Because the regions *KmARS7* (123-182) and *KmARS18* (1181-1240) did not function as effective replicators though they contain sites that share remarkably high similarity to the ACS of *K. marxianus* strain ATCC12424 (Iborra and Ball, 1994) and ACS of *K. lactis* (Liachko et al., 2010).

5. Conclusion

Identification of the short sequences that function as *K. marxianus* autonomous replication origins using a novel and simple approach for the validation of the ARS function. ACSs of *K. marxianus* DMKU3-1042 are diverse among the *KmARS*s as well as from those of *K. lactis*, indicating that eukaryotic replication systems are not necessarily having common ACS. That is evidenced by the fact that no site-specificity was detected in early embryos of frogs, flies, and fish (DePamphilis, 2003). However, mammals contain genetically required sequences that convey origin activity when translocated to other chromosomal sites, but they lack identifiable, genetically required consensus sequences such as ACS in budding yeast replicators (Prioleau et al., 2003). A single nucleotide

(A)

<i>KmARS18</i>	111	120	130	↓	140	150	160	Size
(111-134) ~ (135-160):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	160	(50 bp)
(111-134) ~ (135-156):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	156	(46 bp)
(111-134) ~ (135-153):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAAAGAAA	153	(43 bp)
(111-134) ~ (135-150):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	150	(40 bp)
(111-134) ~ (135-147):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	147	(37 bp)
(111-134) ~ (135-144):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	144	(34 bp)
(111-134) ~ (135-141):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	141	(31 bp)
(111-134) ~ (135-138):	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	138	(28 bp)
(111-134) ~ URA3:	TCCATAATTT	GGAATTGAAA	GTCACTTTAG		GTTCACTATA	TAATGAAAAG	134	(24 bp)



(B)

111	120	130	138	size	HET
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(28 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(27 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(26 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(25 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(24 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(23 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(22 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(21 bp)	++
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(20 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(19 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	137	(27 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	136	(21 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	136	(26 bp)	+++
TCCATAATTT	GGAATTGAAA	GTCACTTT	135	(25 bp)	++
TCCATAATTT	GGAATTGAAA	GTCACTTT	134	(24 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	133	(23 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	132	(22 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	131	(21 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	130	(20 bp)	-
TCCATAATTT	GGAATTGAAA	CCA ACTTT	138	(28 bp)	-
TCCATAATTT	GGAATTGAAA	AGTCACTTT	139	(29 bp)	-
TCCATAATTT	GGAATTGAAA	GTCACTTT	138	(27 bp)	-
CAAGACTTCT	TGGAATTGAA	AGTCACTTT	138	(29 bp)	-

Fig. 4. Impact of nucleotides deletion on the activity of the regions *KmARS18* (111-160) and *KmARS18* (111-138). (A) Sequences of *KmARS18* (111-160) and its successive deletion fragments are shown on the top. The downward arrow indicates the position that separates the two primers for each construct. The transformation efficacy of *KmARS18* (111-160) and its deletion fragments are shown on the bottom. (B) Impact of single nucleotide deletion on the activity of *KmARS18* (111-138). The sequences that produce highly efficient transformation (HET) are indicated by (+++), active regions after the deletion are indicated by (++) and the regions lost the activity are indicated by (-). The minimum sequence of *KmARS18* with highly efficient transformability is indicated in yellow background. Other modifications presented are the replacement of GTC with CCA, the addition of A at position 131 of *KmARS18* (111-138), deletion of G at position 122 of *KmARS18* (111-138), and replacement of the nucleotides from 111-119 of *KmARS18* (111-138) with sequences from *KmARS7*. These modifications are shown in sequences # 20, 21, 22, and 23, respectively.

(A)

Mutation	111	138	HET ⁺	Mutation	111	138	HET ⁺
T111A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++	T125A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
T111G-	GCCATAATTTGGAATTGAAAGTCAC	TTC	+++	T125G-	TCCATAATTTGGAAGTGAAAGTCAC	TTC	+
T111C-	CCCATAATTTGGAATTGAAAGTCAC	TTC	+/-	T125C-	TCCATAATTTGGAAC TGAAAGTCAC	TTC	+++
C112A-	TACATAATTTGGAATTGAAAGTCAC	TTC	+++	T126A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
C112G-	TGCATAATTTGGAATTGAAAGTCAC	TTC	++	T126G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
C112T-	TTCCATAATTTGGAATTGAAAGTCAC	TTC	+++	T126C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
C113A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++	G127A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
C113G-	TCGATAATTTGGAATTGAAAGTCAC	TTC	++	G127C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
C113T-	TCCTATAATTTGGAATTGAAAGTCAC	TTC	+++	G127T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
A114G-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	+++	A128G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A114T-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	+++	A128T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
A114C-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	++	A128C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	-
T115A-	TCCAGAATTTGGAATTGAAAGTCAC	TTC	++	A129G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	-
T115G-	TCCAGAATTTGGAATTGAAAGTCAC	TTC	+++	A129T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	-
T115C-	TCCAGAATTTGGAATTGAAAGTCAC	TTC	+++	A129C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	-
A116G-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	+++	A130C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
A116T-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	++	A130T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
A116C-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	++	A130C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
A117G-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	+	G131A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A117T-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	++	G131T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
A117C-	TCCGTAATTTGGAATTGAAAGTCAC	TTC	+	G131C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
T118A-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+	T132A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+
T118G-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	T132G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
T118C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	T132C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
T119A-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	C133A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
T119G-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	C133G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
T119C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	C133T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
T120A-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+++	A134G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
T120G-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	A134T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
T120C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	A134C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
G121A-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+++	C135A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
G121C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	-	C135G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
G121T-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+	C135T-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
G122A-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T136A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
G122C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T136G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
G122T-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+	T136C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
A123G-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+++	T137A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++
A123T-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T137G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A123C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T137C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A124G-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T138A-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A124T-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	+++	T138G-	TCCATAATTTGGAATTGAAAGTCAC	TTC	++
A124C-	TCCATAAATTTGGAATTGAAAGTCAC	TTC	++	T138C-	TCCATAATTTGGAATTGAAAGTCAC	TTC	+++

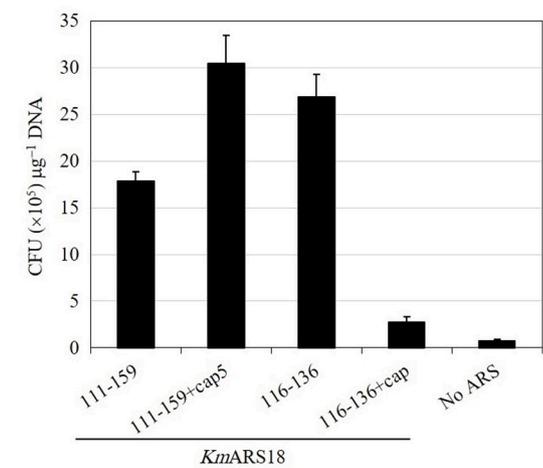
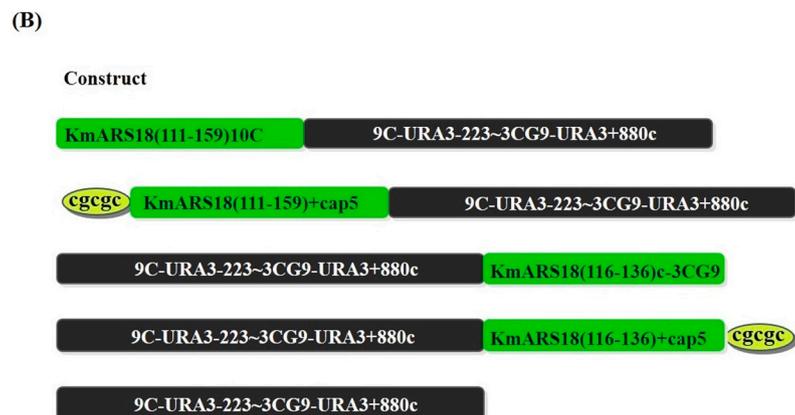


Fig. 5. Effect of nucleotide substitutions and addition of cap sequences on the function of *KmARS18* regions. (A) Influence of single nucleotide substitutions on the transformability of *KmARS18* (111-138). Sequences with the symbol (+++) give a highly efficient transformation, those with the symbol (++) give moderate transformation, those with the symbol (+) give weak transformation, and those with the symbol (-) completely lost the activity. The sequence with (+/-) give variant transformability (mainly small colonies). (B) Sensitivity of *KmARS18* (116-136) to cap. The addition of cap sequences at the end of *KmARS18* (116-136) adversely affects the ARS function of this region. The addition of cap "cgcgc" to the region *KmARS18* (111-159) positively enhanced the transformability, while the transformability of the capped *KmARS18* (116-136) is greatly declined relative to the uncapped same region.

mutagenesis approach helps to identify specifically the essential nucleotides within the span of the active *KmARs*s. The *KmARS18* ACS termini are very sensitive to nucleotides substitution. All defined minimum active *KmARs*s, except *KmARS22* and *KmARS16*, are located at the intergenic sequences of the genome. Overall, the minimum *KmARs*s reported here are capable to induce the formation of circular DNA and effectively replicate within the yeast cells. The *KmARs*s described in this study will provide additional options that are versatile and more

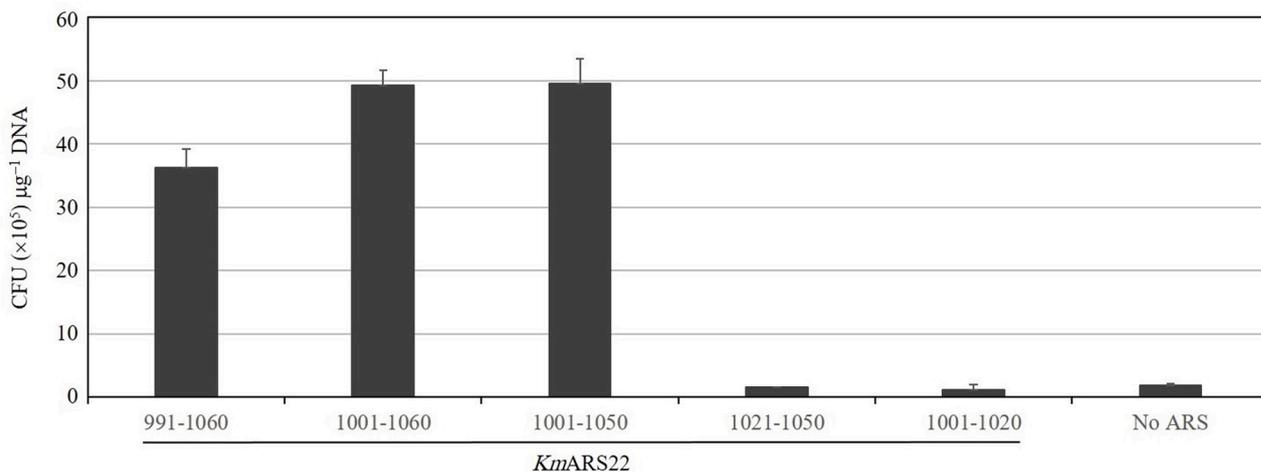
effective to develop large sets of molecular tools for better engineering of this strain.

Authors' contributions

BMAA and RA: Conceptualization. **BMAA:** Methodology, Investigation, and Validation. **BMAA and HH:** Writing- Original draft preparation. **BMAA, HH, and RA:** Writing- Reviewing and Editing. **HH and**

(A) *KmARS22*

	991	1000	1010	1020	1030	1040	1050	1060	Size
(991-1060):	TGTTATCTTT	TTCGCTTCAA	AAGTTACTTT	GGATTCTAAT	ATAAGAAAAA	AAATAAAAAAC	AAACCAAATC		(70 bp)
(1001-1060):	TGTTATCTTT	TTCGCTTCAA	AAGTTACTTT	GGATTCTAAT	ATAAGAAAAA	AAATAAAAAAC	AAACCAAATC		(60 bp)
(1001-1050):	TGTTATCTTT	TTCGCTTCAA	AAGTTACTTT	GGATTCTAAT	ATAAGAAAAA	AAATAAAAAAC	AAACCAAATC		(50 bp)
(1021-1050):	TGTTATCTTT	TTCGCTTCAA	AAGTTACTTT	GGATTCTAAT	ATAAGAAAAA	AAATAAAAAAC	AAACCAAATC		(30 bp)
(1001-1020):	TGTTATCTTT	TTCGCTTCAA	AAGTTACTTT	GGATTCTAAT	ATAAGAAAAA	AAATAAAAAAC	AAACCAAATC		(20 bp)



(B) *KmARS36*

	300	310	320	330	340	size
(291-306) ~ (307-340):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAAATAAA	AATTAAAATG	(50 bp)
(291-306) ~ (307-337):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAAATAAA	AATTAAAATG	337 (47 bp)
(291-306) ~ (307-334):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAAATAAA	AATTAAAATG	334 (44 bp)
(291-306) ~ (307-331):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAAATAAA	AATTAAAATG	331 (41 bp)
(291-306) ~ (307-328):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAAATA	AATTAAAATG	328 (38 bp)
(291-306) ~ (307-325):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GTGAA	AATTAAAATG	325 (35 bp)
(291-306) ~ (307-322):	TCTTTAATAT	TATTTTTTCAT	TTCAAAAAGT	GT	AATTAAAATG	322 (32 bp)

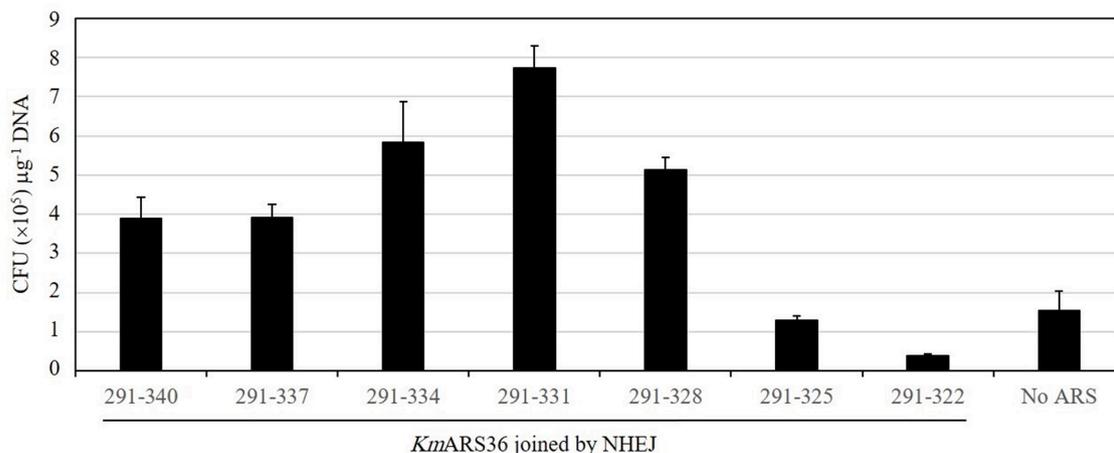


Fig. 6. Impact of truncations and nucleotides deletions on *KmARS22* and *KmARS36*. **(A)** Effect of 5' and 3' ends truncations on *KmARS22*. Sequences of *KmARS22* (991-1060) and its truncated fragments are shown on the top. These sequences are attached to the *ScURA3* marker gene and the transformation efficiencies are shown on the bottom. **(B)** Influence of nucleotides deletions on *KmARS36*. The region *KmARS36* (291-340) was divided into two primers as indicated by the downward arrow. Each primer was attached to one end of the *ScURA3* marker gene. Sequences after successive deletions are shown (top panel). A chart for the transformation efficiency of the *KmARS36* (291-340) and its deletion variants is shown (bottom panel).

Table 4
Influence of interchanged sequences of *KmARSs* on the ARS activity.

ARS fused at the 3' end of <i>ScURA3</i>	ARS fused at the 5' end of <i>ScURA3</i>						
	<i>KmARS7</i> (230-250)	<i>KmARS11</i> (61-100)	<i>KmARS16</i> (753-790)	<i>KmARS18</i> (139-159)	<i>KmARS22</i> (1021-1050)	<i>KmARS36</i> (307-340)	<i>KmARS51</i> (516-550)
<i>KmARS7</i> (201-229)	49*	44.2	53.2	1.55	41.45	32.5	22
<i>KmARS11</i> (46-60)	41.45	44.35	60.25	2.7	3.0	55.1	26.95
<i>KmARS16</i> (721-752)	3.0	63.85	41.55	2.0	1.7	10.9	1.85
<i>KmARS18</i> (111-138)**	67.85	57.4	63.05	45.8	61.95	61.1	66.4
<i>KmARS22</i> (1001-1020)	4.05	38.75	41.65	1.95	28.65	30.8	45.4
<i>KmARS36</i> (291-306)	1.9	55.1	59.15	2.0	2.7	11.85	2.3
<i>KmARS51</i> (491-515)	8.2	51.9	40.35	36.75	38.75	38.45	51.3

*Transformation efficiencies of the interchanged ARS sequences are tabulated as CFU ($\times 10^5$) μg^{-1} DNA. Using the same lot of yeast competent cells (RAK3605), the marker gene alone gave approximately 1.26×10^5 CFU μg^{-1} DNA. **Tested colonies from transformants of the *KmARS18* (111-138) in combination with all other regions of *KmARSs* that shown in this table gave 81 to 100 percent growth on 5-FOA.

RA: Resources. RA: Acquisition of the financial support for the project leading to this publication.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of Competing Interest

None

Acknowledgments

We thank Jun Asakawa and Yuki Misumi for their competent technical assistance. This work was supported in part by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN), Japan.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.crmicr.2021.100053](https://doi.org/10.1016/j.crmicr.2021.100053).

References

- Abdel-Banat, B.M.A., Hoshida, H., Ano, A., Nonklang, S., Akada, R., 2010a. High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? *Appl. Microbiol. Biotechnol.* 85, 861–867.
- Abdel-Banat, B.M.A., Nonklang, S., Hoshida, H., Akada, R., 2010b. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast* 27, 29–39.
- Akada, R., Kitagawa, T., Kaneko, S., Toyonaga, D., Ito, S., Kakihara, Y., Hoshida, H., Morimura, S., Kondo, A., Kida, K., 2006. PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*. *Yeast* 23, 399–405.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., 1999. *Short Protocols in Molecular Biology*. Wiley, New York.
- Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D., 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115–132.
- Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R., 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* 154, 164–175.
- Cha-aim, K., Fukunaga, T., Hoshida, H., Akada, R., 2009. Reliable fusion PCR mediated by GC-rich overlap sequences. *Gene* 434, 43–49.
- Cha-aim, K., Hoshida, H., Fukunaga, T., Akada, R., 2012. Fusion PCR via novel overlap sequences. *Methods Mol. Biol.* 852, 97–110.
- Clyne, R.K., Kelly, T.J., 1995. Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J* 14, 6348–6357.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res* 14, 1188–1190.
- DePamphilis, M.L., 2003. Eukaryotic DNA replication Origins: Reconciling disparate data. *Cell* 114, 274–275.
- Dhar, M.K., Sehgal, S., Kaul, S., 2012. Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.* 163, 243–253.
- Deshpande, A.M., Newlon, C.S., 1992. The ARS consensus sequence is required for chromosomal origin function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4305–4313.
- Foureau, E., Courdavault, V., Gallón, S.M.N., Besseau, S., Simkin, A.J., Crèche, J., Atehortúa, L., Giglioli-Guivarc'h, N., Clastre, M., Papon, N., 2013. Characterization of an autonomously replicating sequence in *Candida guilliermondii*. *Microbiol. Res.* 168, 580–588.
- Gilbert, D.M., 2001. Making sense of eukaryotic DNA replication origins. *Science* 294 (5540), 96–100.
- Hoshida, H., Murakami, N., Suzuki, A., Tamura, R., Asakawa, J., Abdel-Banat, B.M.A., Nonklang, S., Nakamura, M., Akada, R., 2014. Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*. *Yeast* 31, 29–46.
- Iborra, F., Ball, M.M., 1994. *Kluyveromyces marxianus* small DNA fragments contain both autonomous replicative and centromeric elements that also function in *Kluyveromyces lactis*. *Yeast* 10, 1621–1629.
- Irene, C., Maciariello, C., Cioci, F., Camilloni, G., Newlon, C.S., Fabiani, L., 2004. Identification of the sequences required for chromosomal replicator function in *Kluyveromyces lactis*. *Mol. Microbiol.* 51, 1413–1423.
- Kelly, T., Callegari, A.J., 2019. Dynamics of DNA replication in a eukaryotic cell. *Proc. Natl. Acad. Sci. U. S. A.* 116 (11), 4973–4982.
- Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotechnol. Biofuels* 8, 47.
- Liachko, I., Bhaskar, A., Lee, C., Chung, S.C.C., Tye, B.-K., Keich, U., 2010. A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet* 6, e1000946.
- Liachko, I., Dunham, M.J., 2014. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res* 14, 364–367.
- Liachko, I., Tanaka, E., Cox, K., Chung, S.C.C., Yang, L., Seher, A., Lindsay Hallas, L., Cha, E., Kang, G., Pace, H., Jasmine Barrow, J., Inada, M., Tye, B.-K., Keich, U., 2011. Novel features of ARS selection in budding yeast *Lachancea kluyveri*. *BMC Genomics* 12, 633.
- Limtong, S., Sringiew, C., Yongmanitchai, W., 2007. Production of fuel ethanol at high temperature from sugar cane juice by a newly isolated *Kluyveromyces marxianus*. *Bioresour. Technol.* 98, 3367–3374.
- Méchal, M., 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* 11 (10), 728–738.
- Méchal, M., Yoshida, K., Coulombe, P., Pasero, P., 2013. Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr. Opin. Genet. Dev.* 23, 124–131.
- Nieduszynski, C.A., Knox, Y., Donaldson, A.D., 2006. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* 20, 1874–1879.
- Nonklang, S., Abdel-Banat, B.M.A., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M., Rinji Akada, R., 2008. High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042. *Appl. Environ. Microbiol.* 74, 7514–7521.
- Prioleau, M.N., Gendron, M.C., Hyrien, O., 2003. Replication of the chicken β -globin locus: Early-firing origins at the 5' HS4 insulator and the ρ - and β^A -globin genes show opposite epigenetic modifications. *Mol. Cell. Biol.* 23, 3536–3549.
- Reeves, R., Nissen, M.S., 1990. The A-T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.* 265, 8573–8582.
- Sikorski, R.S., Hieter, P., 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19–27.
- Stinchcomb, D.T., Thomas, M., Kelly, J., Selker, E., Davis, R.W., 1980. Eukaryotic DNA segments capable of autonomous replication in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 77, 4559–4563.
- Suzuki, A., Fujii, H., Hoshida, H., Akada, R., 2015. Gene expression analysis using strains constructed by NHEJ-mediated one-step promoter cloning in the yeast *Kluyveromyces marxianus*. *FEMS Yeast Res* 15, fov059.