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

**A R T I C L E   I N F O**

**Keywords:**  
Functional validation  
Interchanged sequences  
*Km*ARS  
NHEJ  
Transformability

**A B S T R A C T**

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 (*Km*ARSs) have very effective replication ability but their sequences are divergent and share no common consensus. We have carried out point mutations, deletions, and base pair substitutions within the sequences of some of the *Km*ARSs 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 *Km*ARSs reported here except *Km*ARS7. Moreover, partial sequences from different *Km*ARSs 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 *Km*ARSs. Our deletions analysis revealed that only 21 bp in *Km*ARS18 could retain the ARS activity. The identified *Km*ARSs 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 (Clayne 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., 2015).

**Email addresses:** bahmed@kfuf.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; Lim tong 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 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 sections of ARSs 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)-containing final concentrations of 40% w/v polyethylene glycol 3350 (PEG), 200 mM lithium acetate (LiAc), and 100 mM dithiothreitol (DTT) were added to the liquid LB medium at 37°C, and the recombinant plasmids were extracted and purified from *E. coli* cells using Qiagen spin miniprep kit (Qiagen). The purified plasmids were transformed again into the *K. marxianus* strain RAK3605s (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.

2.2. 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 RAK3605s (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 ZymoPrep 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; URA3-223 and URA3-300c; 9C-URA3-223 and URA3-300c; and 9C-URA3-223 and 3C9G-URA3+880c. The 9C and 3C9G 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
3. Results

3.1. Autonomously replicating sequences from K. marxianus DMKU3-1042 (KmARSs)

In this study, more than twenty-eight plasmids harboring K. marxianus DMKU3-1042 autonomously replicating sequences (KmARSs) 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 KmARSs (Table 2). Plasmids found with identical sequences included 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 KmARSs ranged from 154 to 2,590 base pairs. These insert DNAs are distributed in the seven chromosomes of the yeast K. marxianus DMKU3-1042 (Lertwattanaskul 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, and 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 KmARSs

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 KmARSs (Fig. S1A). After series of sequence alignments (Fig. S3) with known ARSs from S. cerevisiae (Deshpande and Newton, 1992) and Kluyveromyces lactis (Iborra and Ball, 1994), KmARSs sequences ranging from 21 to 70 bp were identified for replication in K. marxianus DMKU3-1042. To analyze the sequences more precisely, these KmARSs were fused to the ScURA3 strain RAK3606 and selection on MM-U and replica-plating on 5-FOA. To examine whether segments of minimum KmARSs can be exchanged with each other while retaining the ACS activity, a combination of primer pairs representing discrete KmARSs 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 KmARSs, 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.
Table 3

| Primer name | Sequences (5’→3’) |
|-------------|-------------------|
| URA3-223    | AAAGCTTTCTCATCCCTTACCTTTTTTTTTTTTTTTTG |
| 9C-URA3-223 | ccccccccAAAGCTTTCTCATCCCTTACCTTTTTTTTTTTTTTTTG |
| URA3-300c   | TGGTTGAAAGTCTTGGACAG |
| 4C-URA3-300c| ccccccccTGGTTGAAAGTCTTGGACAG |
| URA3-223 c  | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-223 | ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |
| URA3-300c   | CTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| 9C-URA3-300c| ccccccccCTTCAGCTTCCTACATCCATTTTTTTTTTTTTTTTG |
| URA3-300c   | CAGTTTGTGCGTTTGGACAG |
| 3CG9-URA3   | ccccccccCAGTTTGTGCGTTTGGACAG |

(continued on next page)
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 \( \text{KmARS7} \)

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

![Fig. 1. The activity of short sequences of \( \text{KmARSs} \). (A) Short sequences of different \( \text{KmARSs} \) ranging from 49 to 70 bp are shown (top panel). Seven \( \text{KmARS} \) sequences are fused to the \( \text{ScURA3} \) gene at its 5’ end and the transformation efficacies of these constructs are depicted (bottom panel). Sequence alignment and logos of the short \( \text{KmARSs} \) are depicted in panel (B).](image-url)
retains the transformability of the regions (56-105) was declined but produced significant levels of transformation. 

KmARS11 (46-105) and (46-95) gave comparable high transformation rates as judged by replica plating on 5-FOA (data not shown). This region, mostly the result of integration activity rather than autonomous 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.

3.6. 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 KiARSs (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 transmitters 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 transmitters from the interchanged constructs gave between 81 to 100% colony growth on 5-FOA.

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 (TCCATAAA) resulted in the generation of fewer transmitters. 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 transformants. 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-TCCATAAT-119) by the introduction of nine nucleotides of KmARS7 (201-CAGACTTC-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 transmitters similar to the original sequence. Substitutions 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 transmitters were developed upon base substitution at the sites T111C, A117G, A117C, T118A, G122T, T125G, T126A, 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).
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.
89.53% identity to ARS1 and the other is a portion consists of 128 nucleotide sequence only and all function in both ARS and centromeric elements, while ARS3 contains ARS core (bp), ARS2 (1206 bp), and ARS3 (1200 bp). ARS1 and ARS2 contain sequences.

Identification of the short sequences that function as ARS3 (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 KmARSs reported here share significant identity to the ARS3 from the strain ATCC12424, which indicates that in this study ten KmARSs 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′-TTATTGTGTT-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 KmARSs.

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 KmARSs 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 KmARSs reported here share significant identity to the ARS3 from the strain ATCC12424, which indicates that in this study ten KmARSs 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′-TTATTGTGTT-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 KmARSs.

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 KmARSs replicate effectively. The sequences 5′-AAA(G/A)T(×××) (T/A)T-T-3′ and 5′-AAA(A/T)AAAAT-3′ are likely the common consensus in the interchanged KmARSs of the strain DMKU3-1042 (Fig. 56) 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).

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.

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 KmARSs replicate effectively. The sequences 5′-AAA(G/A)T(×××) (T/A)T-T-3′ and 5′-AAA(A/T)AAAAT-3′ are likely the common consensus in the interchanged KmARSs 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 KmARSs as well as from those of K. lactis, indicating that euakaryotic 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 (Prieleau et al., 2003). A single nucleotide
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.
The mutagenesis approach helps to identify specifically the essential nucleotides within the span of the active $K_m$ ARSs. The $K_m$ ARS18 ACS termini are very sensitive to nucleotide substitution. All defined minimum active $K_m$ ARSs, except $K_m$ ARS22 and $K_m$ ARS16, are located at the intergenic sequences of the genome. Overall, the minimum $K_m$ ARSs reported here are capable to induce the formation of circular DNA and effectively replicate within the yeast cells. The $K_m$ ARSs 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

Fig. 5. Effect of nucleotide substitutions and addition of cap sequences on the function of $K_m$ ARS18 regions. (A) Influence of single nucleotide substitutions on the transformability of $K_m$ ARS18 (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 $K_m$ ARS18 (116-136) to cap. The addition of cap sequences at the end of $K_m$ ARS18 (116-136) adversely affects the ARS function of this region. The addition of cap “cgcgc” to the region $K_m$ ARS18 (111-159) positively enhanced the transformability, while the transformability of the capped $K_m$ ARS18 (116-136) is greatly declined relative to the uncapped same region.
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 ScuRA3 | ARS fused at the 5’ end of ScuRA3 | KmARS7 (230-250) | KmARS11 (61-100) | KmARS16 (753-790) | KmARS18 (139-159) | KmARS22 (1021-1050) | KmARS36 (307-340) | KmARS51 (516-550) |
|---------------------------------|---------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| KmARS7 (201-229)                | 49*                            | 44.2           | 53.2           | 1.55           | 41.45          | 32.5           | 22             |                 |
| KmARS11 (46-60)                | 41.45                          | 44.35          | 60.25          | 2.7            | 3.0            | 55.1           | 26.95          |                 |
| KmARS16 (721-752)              | 3.0                            | 63.85          | 41.55          | 2.0            | 1.7            | 10.9           | 1.85           |                 |
| KmARS18 (111-130)**            | 67.85                          | 57.4           | 63.05          | 45.8           | 61.95          | 61.1           | 66.4           |                 |
| KmARS22 (1001-1020)            | 4.05                           | 38.75          | 41.65          | 1.95           | 28.65          | 30.8           | 45.4           |                 |
| KmARS36 (291-306)              | 1.9                            | 55.1           | 59.15          | 2.0            | 2.7            | 11.85          | 2.3            |                  |
| KmARS51 (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 (×10^9) µg^-1 DNA. Using the same lot of yeast competent cells (RAK3605), the marker gene alone gave approximately 1.26×10^9 CFU µg^-1 DNA. **Tested transforms from the variations of the KmARS7 (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 Yukie 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.crmicro.2021.100053.

References
Abdel-Banat, B.M.A., Hoshida, H., Aro, A., Nonklamg, 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 gene integrations through the control of non-homologous end joining in the yeast Kluyveromyces marxianus. Yeast 23, 399–405.

Aunuel, F.M. Brent, R., Kingston, R.E., Moore, D.D., 1999. Short Protocols in Molecular Biology. Wiley, New York.

Brambilla, C.A., Davies, A., Cost, C., 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.

Chakrabarti, D., Davie, H., Cost, H., 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.

Chakrabarti, D., Davie, H., Cost, H., 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.

Deshpande, A.M., Newton, C.S., 1992. The ARS consensus sequence is required for chromosomal origin function in Saccharomyces cerevisiae. Mol. Cell. Biol. 12, 4305–4315.

Dhar, M.K., Sehgal, S., Kaul, S., 2012. Structure, replication efficiency and fragility of yeast ARS elements. Res. Microbiol. 163, 249–255.