Received Date : 24-Jul-2013

Revised Date : 29-Oct-2013

Accepted Date : 30-Oct-2013

Article type : Short Communication

Editor : Jens Nielsen

An Autonomously Replicating Sequence for use in a wide range of budding yeasts.

Ivan Liachko[1] and Maitreya J. Dunham[1]*

1. Department of Genome Sciences, University of Washington, Seattle, WA, United States of America

* Corresponding author.

Maitreya J. Dunham

Foege Building S403B, Box 355065

3720 15th Ave NE

Seattle, WA 98195

206-543-2338

maitreya@uw.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1567-1364.12123

Abstract

The initiation of DNA replication at replication origins is essential for the duplication of genomes. In yeast, the autonomously replicating sequence (ARS) property of replication origins is necessary for the stable maintenance of episomal plasmids. However, because the sequence determinants of ARS function differ among yeast species, current ARS modules are limited for use to a subset of yeasts. Here we describe a short ARS sequence that functions in at least 10 diverse species of budding yeast. These include, but are not limited to members of the *Saccharomyces*, *Lachancea*, *Kluyveromyces*, and *Pichia* (*Komagataella)* genera spanning over 500 million years of evolution. In addition to its wide species range, this ARS and an optimized derivative confer improved plasmid stability relative to other currently used ARS modules.

DNA replication is an essential function of cellular biology. It is highly regulated at the initiation stage which occurs at loci termed replication origins. Yeast replication origins retain their initiation activity in a plasmid context allowing autonomous episomal plasmid maintenance (Stinchcomb et al. 1980). This *cis*-acting autonomously replicating sequence (ARS) function has been useful for both understanding the basic science of DNA replication (Nieduszynski et al. 2007; Liachko et al. 2013) and for industrial applications (Böer et al. 2007).

The well-studied ARSs of the baker's yeast, *Saccharomyces cerevisiae,* are short (<100 bp) modular DNA sequences that require an 11-17bp core sequence element called the ARS Consensus Sequence (ACS) as well as less well defined flanking sequences

(Méchali et al. 2013). The ACS serves as a binding site for the Origin Recognition Complex (ORC), a six-member protein complex that serves as the landing pad for downstream replication initiation machinery.

Large-scale studies have elucidated a diversity of ARS sequence determinants among the budding yeasts. Pre-Whole Genome Duplication (WGD) yeast *Kluyveromyces lactis* uses a 50 bp ACS motif that is very dissimilar from the canonical *S. cerevisiae* ACS (Liachko et al. 2010). Another pre-WGD species, *Lachancea waltii,* uses a motif that resembles a chimeric fusion between the *S. cerevisiae* and *K. lactis* ACS motifs (Di Rienzi et al. 2012) whereas its relative *L. kluyveri* has more relaxed sequence requirements (Liachko et al. 2011). While ARSs have also been described in other yeast species (Iwakiri et al. 2005; Iborra & Ball 1994; Vernis et al. 1997; Wright & Philippsen 1991; Cregg et al. 1985; Yang et al. 1994), the low-throughput nature of the relevant studies has precluded drawing any overarching conclusions about their origin structure.

Due to the diversity of sequences required for origin function in different yeast species, ARSs are usually restricted to function in only a few yeast species. For example, *K. lactis* ARSs rarely work in non-*Kluyveromyces* yeasts and ARSs from other species rarely function in *K. lactis* host cells (Liachko et al. 2010; 2011). On the other hand, *L. kluyveri* is a permissive host species and can utilize most ARSs from *S. cerevisiae* and *K. lactis* (Liachko et al. 2011). The methylotrophic budding yeast *Pichia pastoris* uses at least two different kinds of ARS sequences, neither of which function in *S. cerevisiae*

(Liachko *et. al.*, submitted). Since ARSs are required for plasmid maintenance, an ARS that functions across all yeasts would be a useful genetic tool to develop shuttle vectors for cross-species studies, but to date such a module does not exist.

We have identified a 452 bp *K. lactis* genomic fragment that retains ARS function in at least 10 budding yeast species with diverse ARS sequence requirements. This sequence (which we have named "panARS") maps to coordinates 781040-781491 bp on chromosome F of the *K. lactis* genome (strain NRRL Y-1140 (Dujon et al. 2004)). The DNA fragment was originally identified as an ARS in *K. lactis* using a predict-and-verify approach used to generate a comprehensive *K. lactis* ARS map (Liachko et al. 2010). This ARS was subsequently cloned into a commonly used ARS-less *URA3* vector, pRS406. The resulting plasmid (named pIL20) as well as the original plasmid from the *K. lactis* experiment were used to transform *ura3-* strains of *S. cerevisiae, S. paradoxus, S. bayanus var uvarum, L. waltii, L. kluyveri, K. lactis, K. wickerhamii, and P. pastoris.* ARS activity is exhibited by high-transformation efficiency and robust colony formation on selective media. We detected ARS activity (>500 colonies per microgram of transforming plasmid DNA) in all species tested (Fig. 1a). Additionally, for each species several colonies were re-streaked on selective medium agar plates and inoculated into selective liquid medium where they grew robustly at 30□. We were able to recover plasmids from re-streaked colonies and cultures of all species using standard techniques. Sequencing and restriction digestion analysis confirmed the identity of the recovered plasmids to be the same as the input ARS plasmid. Recovered plasmids were used to transform the host species and displayed robust colony formation on

selective media in all cases. These results suggest that panARS allows episomal plasmid maintenance in the yeast species listed above. We also detected ARS activity in *Naumovozyma castellii* and *Hansenula polymorpha* when panARS was cloned into vectors bearing antibiotic resistance markers (Chee & Haase 2012) (data not shown).

To delineate the region of panARS required for function in each of the different species we sheared the 452 bp ARS fragment and cloned a library containing ARS sub-fragments. This library was used to transform the different yeast species in order to identify sub-fragments of the ARS that retain function. Short ARS fragments isolated from this screen were also tested for function across multiple species. In this manner we were able to isolate the minimal region of the ARS that confers function across all species to a region near one end of the ARS (Fig. 1b). All species listed except *P. pastoris* were able to initiate replication with ARS sub-fragments in a region between relative positions 188-316. For ARS function, *P. pastoris* required ARS DNA fragments within relative coordinates 256-371 (Fig. 1b).

We modified the sequence of panARS in an attempt to simultaneously improve its function across multiple species. The sequence determinants of ARS function are not yet understood in most yeasts, precluding targeted optimization across the entire species panel. We introduced mutations into the best match to the *S. cerevisiae* and *K. lactis* ACS sequences within the functional panARS region and one strong match to the *K. lactis* ACS outside the minimal region (since this may be a dimeric *K. lactis* ARS) to improve the sequence matches to these known motifs (Supplementary Figure 1). The

resulting mutations improved all motif matches as assayed by the FIMO motif-alignment program (Grant et al. 2011): the q-value of the *S. cerevisiae* ACS match decreased from 0.003 to 3.11e-05, and the q-value of the two *K. lactis* ACS matches decreased from 1.6e-08 to 7.25e-11 and from 1.89e-07 to 3.32e-12.

We cloned the full length (452bp) optimized ARS mutant sequence into vector pRS406 and tested ARS function in different yeast species. The mutant ARS fragment retained robust ARS activity in all species listed above (data not shown). We also performed plasmid loss assays as described (Donato et al. 2006) to measure relative plasmid stability in the eight aforementioned species (Fig. 1c). The plasmid loss assay (also known as the minichromosome maintenance assay) measures the retention of the plasmid-borne selectable marker during growth in non-selective media. YPD media was inoculated with cells transformed with relevant plasmids and grown for 10-20 generations. Proportions of Ura+ cells within each culture were measured at the start and end of the non-selective growth by plating on YPD and selective agar plates and counting colonies. Plasmids with low or absent ARS activity are quickly lost from the population whereas plasmids with increased ARS activity are more readily retained during non-selective growth. The mutant ARS sequence showed a slightly improved stability (indicated by lower plasmid loss/generation) in the *S. cerevisiae, S.bayanus,* and *L. waltii* hosts relative to the original ARS sequence (one-tailed two-sample t-test p-values = 0.0007, 0.0403, and 0.0086 respectively). In *K. lactis* and *K. wickerhamii*, we did not detect a significant change in plasmid stability between the two alleles (p-values = 0.3872 and 0.1678 respectively). This may be due to the fact that this ARS originates

from *K. lactis* and is already maximally efficient. In *P. pastoris* the optimized ARS showed improved efficiency relative to the wild type sequence (p-value = 0.0115).

We also tested the plasmid loss rate of the same vector backbone bearing the previously described *P. pastoris* ARS, PARS1. This 167 bp sequence is currently the most commonly used ARS module in *P. pastoris* (Lee et al. 2005; Cregg et al. 1985). Additionally, we tested the efficiency of pRS316, a *S. cerevisiae* ARS/CEN plasmid which replicates in *P. pastoris*. Plasmids carrying both the wild type and optimized ARS alleles were more stable than both the PARS1 plasmid and pRS316 (Supplementary Figure 2).

In summary, we have identified a 452bp ARS element which originates from *K. lactis*, but also retains ARS function in a number of other species with diverse sequence requirements for initiating DNA replication. The synthetically optimized mutant version of this sequence performs either equivalently to or better than the wildtype sequence. Additionally, this module performs significantly better than other characterized ARS plasmids in *P. pastoris*, with a stability that resembles ARS/CEN plasmids in better studied models. These results suggest that panARS may be an efficient ARS module in other related yeast species and may be a superior construct even when cross-species performance is not required.

Acknowledgements

The authors thank James Cregg for strain JC308, Duncan Greig for strain YDG613, Mark Johnston for strain FM628, Carol Newlon for strain MW98-8C, Bonita Brewer for

the *L. waltii ura3* strain, David Bartel for strain 4310, USDA Agricultural Research Service for strains Y-8286 and Y-5445, as well as Rachel Youngblood for technical assistance. This work was supported by NSF grant 1243710. IL was supported by NIH grant F32 GM090561. MJD is a Rita Allen Foundation Scholar and a CIFAR Fellow. The authors have filed a patent based on this work.

References

Böer E, Steinborn G, Kunze G & Gellissen G (2007) Yeast expression platforms. *Appl. Microbiol. Biotechnol.* 77: 513–523.

Broach JR, Li YY, Feldman J, Jayaram M, Abraham J, Nasmyth KA & Hicks JB (1983) Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harb. Symp. Quant. Biol.* 47 Pt 2: 1165–1173.

Chee MK & Haase SB (2012) New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 2: 515–526.

Cregg JM, Barringer KJ, Hessler AY & Madden KR (1985) *Pichia pastoris* as a host system for transformations. *Mol. Cell. Biol.* 5: 3376–3385.

Di Rienzi SC, Lindstrom KC, Mann T, Noble WS, Raghuraman MK & Brewer BJ (2012) Maintaining replication origins in the face of genomic change. *Genome Res.* 22: 1940–1952.

Donato JJ, Chung SCC & Tye BK (2006) Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. *PLoS Genet.* 2: e141.

Dujon B et al. (2004) Genome evolution in yeasts. *Nature* 430: 35–44.

Grant CE, Bailey TL & Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018.

Iborra F & Ball MM (1994) *Kluyveromyces marxianus* small DNA fragments contain both autonomous replicative and centromeric elements that also function in *Kluyveromyces lactis*. *Yeast* 10: 1621–1629.

Iwakiri R, Eguchi S, Noda Y, Adachi H & Yoda K (2005) Isolation and structural analysis of efficient autonomously replicating sequences (ARSs) of the yeast *Candida utilis*. *Yeast* 22: 1049–1060.

Lee CC, Williams TG, Wong DWS & Robertson GH (2005) An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. *Plasmid* 54: 80–85.

Liachko I et al. (2011) Novel features of ARS selection in budding yeast *Lachancea kluyveri*. *BMC Genomics* 12: 633.

Liachko I, Bhaskar A, Lee C, Chung SCC, 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, Youngblood RA, Keich U & Dunham MJ (2013) High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. *Genome Res.* 23: 698–704.

Liachko I et al. (submitted) GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*.

Méchali 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 CA, Hiraga S-I, Ak P, Benham CJ & Donaldson AD (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res.* 35: D40–D46.

Stinchcomb DT, Thomas M, Kelly J, Selker E & Davis RW (1980) Eukaryotic DNA segments capable of autonomous replication in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 77: 4559–4563.

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

Wright MC & Philippsen P (1991) Replicative transformation of the filamentous fungus *Ashbya gossypii* with plasmids containing *Saccharomyces cerevisiae* ARS elements. *Gene* 109: 99–105.

Yang VW, Marks JA, Davis BP & Jeffries TW (1994) High-efficiency transformation of *Pichia stipitis* based on its *URA3* gene and a homologous autonomous replication sequence, *ARS2*. *Appl. Environ. Microbiol.* 60: 4245–4254.

Figure Legends.

Figure 1. The function of panARS in different budding yeast species. (a) The ARS-less *URA3* vector pRS406 and its counterpart bearing the panARS sequence (pIL20) were used to transform *ura3* strains of different budding yeast species. Transformations were plated on agar plates lacking uracil. The growth of *URA3* colonies indicates ARS activity conferred by the panARS plasmid (right side of each plate) and not by the empty vector (left side of each plate). (b) Relative coordinates of recovered functional subfragments of panARS in different species. (c) Plasmid loss assays were performed on indicated yeast species transformed with plasmids bearing the wildtype panARS sequence (ARS) or the optimized mutant allele of panARS (OPT). Plasmid loss per generation of growth in non-selective medium is shown with error bars representing standard deviations between at least four replicate experiments. Increased plasmid loss is indicative of weaker ARS function and decreased plasmid stability.

Supplementary Figure 1 - Comparison of ARS Consensus motifs and panARS sequences. (a) Phylogenetic relationships and previously published ACS motifs are shown for *S. cerevisiae* (Broach et al. 1983; Liachko et al. 2013), *L. waltii* (Di Rienzi et al. 2012), *L. kluyveri* (Liachko et al. 2011), and *K. lactis* (Liachko et al. 2010). (b) The sequences of the native (ARS) and optimized (OPT) panARS elements are shown. The region highlighted in orange represents the only significant match to the *K. lactis* ACS within the minimal functional region of the panARS. The region highlighted in green represents a strong match to the *K. lactis* ACS outside of the main functional region. The region highlighted in red represents the best match to the *S. cerevisiae*/*L. waltii*/*L. kluyveri* ACS motifs. Functional ARS sequence determinants in other species are not

yet known. The mutations introduced into the optimized version of ARS are indicated by lowercase letters.

Supplementary Figure 2 - Plasmid loss rates in *Pichia pastoris*. (a) The *P. pastoris* strain was transformed with plasmid pRS316 and a derivative of pRS406 bearing PARS1. Plasmid loss rates are shown for these, as well as wt and optimized panARS plasmids for comparison. (b) One-tailed two-sample T-tests were performed on data from plasmid loss assays. The resulting P-values are listed as a table.

Supplementary Table 1 - Strains and plasmids used in this study

Supplementary File 1 - a fasta file with the full-length sequences of wildtype and optimized panARS.