# Genetic analysis of the *Komagataella phaffii* centromeres by a color-based plasmid stability assay

Luiza Cesca Piva[1], Janice Lisboa De Marco[1], Lidia Maria Pepe de Moraes[1], Viviane Castelo Branco Reis[1], Fernando Araripe Gonçalves Torres[1]*

[1]Laboratório de Biologia Molecular, Universidade de Brasília, Brasília, DF, Brazil, 70910-900.

*Corresponding author

E-mail: ftorres@unb.br

## Abstract

The yeast *Komagataella phaffii* is widely used as a microbial host for heterologous protein production. However, molecular tools for this yeast are basically restricted to a few integrative and replicative plasmids. Four sequences that have recently been proposed as the *K. phaffii* centromeres could be used to develop a new class of mitotically stable vectors. In this work we designed a color-based genetic assay to investigate genetic stability in *K. phaffii*. Plasmids bearing each centromere and the *ADE3* marker were evaluated in terms of mitotic stability in an *ade2/ade3* auxotrophic strain which allows plasmid screening through colony color. Plasmid copy number was verified through qPCR. Our results confirmed that the centromeric plasmids were maintained at low copy number as a result of typical chromosome-like segregation during cell division. These features, combined with high transformation efficiency and *in vivo* assembly possibilities, prompt these plasmids as a new addition to the *K. phaffii* genetic toolbox.


## Author summary

The methylotrophic yeast *Komagataella phaffii* is considered as one of the most important platforms for the production of proteins and metabolites. We sought in this study to develop a color-based genetic system widely used in other yeasts to assess mitotically stability of vectors carrying the proposed *K. phaffii* centromeres. First, we constructed a *K. phaffii* strain (LA3) mutant for *ADE2* and *ADE3*; this resulted in a strain that forms white colonies and when transformed with a vector (pPICH-ADE3) carrying *ADE3* turns red. Next, the four *K. phaffii* centromeres were cloned into pPICH-

ADE3 and tested in LA3 for copy number and plasmid stability. Centromeres are responsible for proper chromosome segregation during cell division, hence guaranteeing that both daughter cells receive one copy of the duplicated DNA. Our results show that three *K. phaffii* centromeres behaved as expected conferring extra stability to the replicative plasmids and maintaining them at low copy number. Once characterized, centromeres can be used as parts in the construction of advanced genetic manipulation tools, thus allowing the construction of strains capable of expressing large metabolic pathways for the production of complex biochemicals.

## Introduction

*Komagataella phaffii* is a methylotrophic yeast of great industrial importance which has been used for more than 30 years as a heterologous protein production platform [1]. Its genome was first published in 2009 and has since then been refined and thoroughly studied [2,3]. As a result, in addition to a protein factory, *K. phaffii* has also been widely considered as a platform for the production of chemicals, biopharmaceuticals, vitamins and other molecules. However, the construction and regulation of new pathways demand complex molecular biology tools which are not readily available for this yeast [4].

*K. phaffii* genetic manipulation traditionally involves the use of shuttle vectors assembled in *Escherichia coli* and subsequently integrated into the yeast´s genome [5]. Recent studies have described the development of a wide range of genetic parts for use in this yeast, as well as new methods of plasmid assembly and transformation [6]. An alternative to integrative strategies is the use of replicative plasmids, which are usually

62 based on the well-known ARS1 sequence [1]. These plasmids may overcome some
63 drawbacks such as genetic instability in multi-copy strains and non-specific integration
64 [7,8]. In addition, they present higher transformation efficiency when compared to
65 integrative vectors and can be assembled by *in vivo* recombination, which eliminates the
66 need for bacterial transformation [9,10]. However, replicative plasmids show low
67 mitotic stability when compared to integrative vectors and few vector options are
68 available for use [11]. Stability problems can be circumvented by the creation of
69 centromeric plasmids, which may provide proper segregation during mitosis. A greater
70 mitotic stability as well as low copy number allow stable and constant protein
71 expression [12]. Centromeric plasmids can be constructed *in vivo,* allowing the
72 assembly and cloning of large sequences including whole metabolic pathways and
73 regulatory regions [13]. Therefore, the construction of such vectors would be of great
74 value for *K. phaffii* strain development in the context of synthetic biology.

75 Centromeres are typically surrounded by large heterochromatin sections in most
76 organisms [14]. Their structure ranges from simple "point" centromeres of only ~125 bp
77 in *Saccharomyces cerevisiae* to epigenetic, sequence-independent centromeres, such as
78 those present in plants and animals. The reason for this phenomenon is that, for most
79 eukaryotes, centromeres are maintained epigenetically and not genetically. Sequence
80 homologies are rare in and between species, hampering the definition of a consensus
81 sequence. In addition, some DNA regions can be centromeric or not depending on its
82 function in previous cell cycles, which highlights the epigenetic nature of the
83 centromere [15].

84 As for non-conventional yeasts there are wide variations in centromere size and
85 structure. *Candida glabrata* has centromeres that show some homology to the CDEI and

86 CDEIII regions of *S. cerevisiae* while *Kuraishia capsulata* centromeres have 200-bp
87 conserved sequences [16,17]. On the other hand, *Candida tropicalis,*
88 *Schizosaccharomyces pombe* and *Candida albicans* have regional centromeres named
89 after their sizes which range from 3 to 110 kb [18–20].

90 *K. phaffii* centromeres have recently been identified, bearing no sequence
91 similarities to those of any other yeast [3]. Since centromere function relies strongly on its
92 structure rather than on its sequence, a centromere-specific histone H3 variant (CSE4)
93 was used in the search for centromeric regions in *K. phaffii*. A CSE4 homolog was
94 identified in chromosome 2 and tagged with a fluorescence marker. The corresponding
95 nuclear localization of the histone-DNA complex indicated a centromere pattern typical
96 of budding yeasts [3]. Tridimensional conformation analysis followed the centromere
97 clustering pattern observed in yeasts and narrowed down all four *K. phaffii* centromere
98 locations to 20 kb windows [21].

99 Considering that a low transcription rate is typical of centromeric regions, RNA-
100 seq analysis allowed to pinpoint the putative centromeric locations for all four *K. phaffii*
101 centromeres [3]. Similarly to *C. tropicalis* and *S.* pombe, *K. phaffii* centromeres are
102 formed by inverted repeats. All four sequences have two inverted repeats of ~2,5 kb,
103 separated by a central segment of 800 to 1300 bp. Chromatin immunoprecipitation
104 sequencing analysis showed that the CSE4 histone binds preferably to the central
105 region, but also along the inverted repeats [22].

106 *K. phaffii* centromeric sequences contain early replication peaks with
107 autonomously replicating sequences, characteristics that are also observed in
108 centromeres of other yeasts [23,24]. According to recently published studies, there are

109 native ARS sequences contained within centromeres 2 and 4. These comprise regions
110 within the inverted repeats, as well as unique adjacent sequences [22,25].
111 In order to expand the functional analysis of the *K. phaffii* centromeres we
112 sought in this study to develop a genetic system based on an *ade2/ade3* auxotrophic
113 strain and a replicative vector carrying the wild-type *ADE3*. Vectors carrying each
114 individual centromere were used to assess plasmid copy number and mitotic stability.
115

116 **Results and Discussion**

117

118 In yeasts, adenine synthesis pathway is used as a tool for auxotrophic selection,
119 gene copy number indicator and for plasmid stability analysis [26]. Many genes from
120 this pathway have been deleted in *S. cerevisiae* in order to create auxotrophic strains,
121 while in *K. phaffii* studies have only focused on *ADE1* and *ADE2*[ 27,28]. *K. phaffii*
122 LA2, a strain mutant for *ADE2* [29], was used as a starting point for the construction of
123 a strain that would allow plasmid stability verification. Deletion of *ADE2* results in cells
124 auxotrophic for adenine which accumulate a red pigment [26] while deletion of genes
125 located upstream, such as *ADE1* or *ADE3*, should prevent the formation of such
126 pigment [27]. As expected, the deletion of *ADE3* in LA3 strain results in white colonies
127 (Fig 1). Deletion of *ADE3* in *S. cerevisiae* has regulatory effects in the histidine
128 synthesis pathway [30]. Consequently, *ade2 ade3* strains are not only auxotrophic for
129 adenine, but also for histidine. In order to verify if this phenotype is applicable to *K.*
130 *phaffii*, we plated strains X-33, LA2 and LA3 on MD medium without supplementation,
131 comparing growth and colony color to cells plated on MD medium with adenine and
132 histidine (Fig 1). LA3 strain displayed the expected histidine auxotrophy phenotype,

133 showing that the adenine-histidine pathways in *K. phaffii* and *S. cerevisiae* have
134 common characteristics.

135 In order to assess plasmid stability, we first constructed plasmid pPICH-ADE3
136 bearing the *ADE3* gene (Fig 2). When transformed with pPICH-ADE3, LA3 cells
137 should return to being red and any changes on colony color would allow a simple
138 screening of plasmid loss [26]. Although adenine auxotrophy has been explored for
139 other purposes in *K. phaffii* [28], this particular color-based system has not yet been
140 used for measuring plasmid stability in this yeast.

141 pPICH-ADE3 was used for cloning of all four *K. phaffii* centromeres. Since it
142 revealed extremely difficult to amplify entire centromeric regions we designed a
143 strategy to amplify centromeres in halves in order to reduce fragment size and to avoid
144 primer annealing inside the inverted repeats (Fig 3). Amplified fragments exhibited in
145 their ends overlapping regions that would allow recombination between each other and
146 with vector pPICH-ADE3.

147 LA3 strain was individually transformed with pPICH-ADE3 and all four
148 constructed centromeric plasmids (pPICH-CEN1-4). Plasmids were verified for
149 autonomous replication by PCR analysis which confirmed their circular conformation
150 (data not shown) and by plasmid rescue in *E. coli*. In this case, only pPICH-CEN3 could
151 not be recovered, probably because CEN3 was the only centromere that could not be
152 cloned directly in *E. coli*.

153 Plasmid stability was firstly verified through colony color in non-selective
154 medium (Fig 4). When plated on YPD non-selective medium, colonies transformed with
155 pPICH-ADE3 lost their color rapidly and presented a red center with large white edges,

156 a result consistent with plasmid instability. In contrast, strains transformed with all
157 centromeric plasmids presented a uniform red coloration throughout the colony.
158 Further stability examination of the centromeric plasmids was performed by
159 growing cells in liquid YPD medium for 144 hours. After diluting and plating cultures
160 on non-selective medium, red and white colonies were counted and compared between
161 each construction (Fig 5). LA3 strain transformed with pPICH-ADE3 did not yield red
162 colonies in any growth period, indicating that the plasmid was mitotically unstable.
163 Conversely, centromeric plasmids presented a higher mitotic stability than pPICH-
164 ADE3. After 96 hours of growth, cells with pPICH-CEN1 started to present white
165 colonies, while all other centromeric plasmids remained stable. After 144 hours,
166 pPICH-CEN1 was lost in most colonies while the other centromeric plasmids were lost
167 in <10% cells. The reason for the instability of pPICH-CEN1 is still unclear but it could
168 be related to the absence of important regulatory regions.
169 Yeast centromeric plasmids knowingly have a higher mitotic stability under non-
170 selective conditions than common replicative vectors since they are equally segregated
171 between daughter cells and therefore provide a uniform culture of cells containing the
172 plasmid [26]. A centromeric vector containing *K. phaffii* CEN2 has been constructed
173 and it presented an enhanced stability when compared to a replicative plasmid [25]. In
174 addition to *K. phaffii* and *S. cerevisiae*, centromeric plasmids have been developed for
175 other yeasts such as *S. pombe, C. glabrata* and *Scheffersomyces stipitis* and in all cases
176 enhanced plasmid stability under non-selective conditions was verified [12,17].
177 Yeast replicative plasmids are normally replicated but are unevenly distributed
178 between daughter cells, which creates both multi-copy and plasmidless cells [26]. Under
179 selective conditions, cells lacking the plasmid are unable to survive and the result is a

180 population of multi-copy plasmid-containing cells. The construction of centromeric
181 plasmids should provide better plasmid segregation and stability and cells should
182 maintain a low and stable plasmid copy number during yeast growth [31]. Plasmid copy
183 number was assessed by qPCR after strains were grown in YPD medium containing
184 zeocin in order to ensure that all cells assayed were harboring the centromeric plasmids.
185 Results from qPCR (Fig 6) indicate that the strain transformed with centromeric
186 plasmids carried 1-2 copies per cell while the replicative plasmid was present at
187 approximately 25 copies per cell. The difference between plasmid copy number for the
188 replicative vector and all centromeric vectors was significant according to a t-test
189 ($p<0.05$). This result illustrates the expected segregation pattern described above for
190 growth in selective conditions and, together with the mitotic stability analysis, provides
191 a clear picture of *K. phaffii* genetic manipulation using centromeric plasmids.

192 In a previous study, *K. phaffii* transformed with a centromeric plasmid was
193 analyzed regarding plasmid copy number and compared to both a replicative plasmid
194 and an integrative strategy [25]. Results exhibited an equal low number of plasmids per
195 cell in all strategies, which does not correspond to our observation. This can be
196 explained by the fact that their analysis was performed in non-selective medium, a
197 condition in which replicative plasmids are readily lost as we have shown in Fig 5A
198 (pPICH-ADE3).

199 Overall, our results indicate that centromeric plasmids could be employed as a
200 new tool for genetic manipulation of *K. phaffii*. Plasmids were maintained for long
201 periods in non-selective medium, indicating that growth can be performed without the
202 addition of antibiotics or any form of selective pressure. The centromeric plasmids' low
203 copy numbers per cell characterize a stable and homogeneous culture that can provide

204 reliable expression results. Finally, their structure as a circular molecule allows *in vivo*
205 plasmid assembly with relatively short homologous sequences when compared to
206 genomic integration techniques where sequences have to be much longer for directed
207 homologous recombination. Simpler assembly may also facilitate the construction of
208 larger and more sophisticated vectors such as yeast artificial chromosomes (YAC)
209 whose stability features may be also analysed by the color-based assay described in this
210 work.
211

212 # Methods

213

214 **Strains and Media**

215 DNA cloning was performed using chemically competent *Escherichia coli* XL-
216 10 Gold (Agilent Technologies) grown in LB medium (5 g $L^{-1}$ yeast extract, 10 g $L^{-1}$
217 peptone and 10 g $L^{-1}$ NaCl, pH 7,2). When needed, agar was added to a final
218 concentration of 1,5%. When zeocin (25 µg $mL^{-1}$) was used for bacterial antibiotic
219 selection, NaCl concentration was reduced to 5 g $L^{-1}$.

220 *K. phaffii* strains were derived from X-33 (Invitrogen). LA2 strain (*amd2 ade2*)
221 was described in a previous work [29]. Yeast was routinely grown in YPD medium (10 g
222 $L^{-1}$ yeast extract, 20 g $L^{-1}$ peptone and 20 g $L^{-1}$ glucose). Solid medium used 2% agar.
223 Zeocin and geneticin, when used, were added at 100 µg $mL^{-1}$ and 500 µg $mL^{-1}$,
224 respectively. Hygromycin B was used to a final concentration of 50 µg $mL^{-1}$. Minimal
225 medium (MD) used 0,34% Yeast Nitrogen Base, 1% $(NH_4)_2SO_4$, 2% glucose,
226 0,00004% biotin and 0,0002% adenine or 0,004% histidine, when needed.
227

228 **PCR**

229 DNA was amplified using Invitrogen Platinum Taq DNA Polymerase (High
230 Fidelity), Promega GoTaq Colorless Master Mix or Sigma-Aldrich Accutaq LA DNA
231 Polymerase. All primers used in this work are shown in Table 1.

232

233 **Table 1. Primers used in this work.**

| Primer | Sequence | Enzyme |
|---|---|---|
| ADE3up-F | GATAAGCTTGATATCGAATTCCTGCAGCCCCCCGGGACGTAATGGAATAACTGCTGAC | *Sma*I |
| ADE3up-R | GAAGTTATGGATCCTACGAGGTAATTGAAGGCTCAC | |
| ADE3lox-F | CTTCAATTACCTCGTAGGATCCATAACTTCGTATAATG | |
| ADE3lox-R | CAATCTCTCCCTTGTCATCGGATCCATAACTTCGTATAG | |
| ADE3dw-F | GAAGTTATGGATCCGATGACAAGGGAGAGATTGAAG | |
| ADE3dw-R | GGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCCCCGGGTGCAATGTACTGTTGAGTAGG | *Sma*I |
| ADE3conF | GGGGACCGGAGGTAAAAGAC | |
| ADE3conR | GTTGGAATAATTGCATGGTCTG | |
| MUT1(Hpa) | CCATTGACATGGTAAACAGTTGGA | |
| MUT2(Bam) | GAACACTGGGATCTTGGTTGAGG | |
| Cen1-F | GAGTTTAAACGGATCCACGAAGCAATGGATAGGCACT | *Bam*HI |
| Cen1-R | CCGCGAATTCGGATCCTGAAGTCTTTCAGAGAGGAGCA | *Bam*HI |
| Cen1c-F | CAAGTATGCGTGATCCCAGGT | |
| Cen1c-R | TACGAATTGTGGGGCTCTGT | |
| Cen2-F | GAGTTTAAACGGATCCATCTCCGTTGATACTCCCAAC | *Bam*HI |
| Cen2-R | CCGCGAATTCGGATCCATCGACAAGCAGAACACTAAG | *Bam*HI |

| | | |
|---|---|---|
| Cen2c-F | GAATGGAGGTGCTGGTGGTTA | |
| Cen2c-R | TGTAATGCTCGCTGGTGAGT | |
| Cen3-F | GAGTTTAAACGGATCCAAGTGGTACACCAGTCAGCG | *Bam*HI |
| Cen3-R | CCGCGAATTCGGATCCTCAGTATTCAACTGCAACTGC | *Bam*HI |
| Cen3c-F | TCAGCCGAATACCCACACTT | |
| Cen3c-R | TCAGCCGTCAGCGAAATGAT | |
| Cen4-F | GAGTTTAAACGGATCCCAAACGCACCGTCTTGTTCA | *Bam*HI |
| Cen4-R | CCGCGAATTCGGATCCAATTGATGTAGACGAGCAGC | *Bam*HI |
| Cen4c-F | TCAAGAATCGTACTGGCACCT | |
| Cen4c-R | CAAGCTCGTGAGATGGGATGT | |
| Cen370-F | CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTGGTCATGAGATCAGATCTAACATCCAAAGACGAAAGGTTGAATGAGTTTAAACGGATCCAAGTGG | |
| Cen370-R | GCTGGCCCTCTCTTCCCAGCTCACGAATCAGATCCTAAGTCCTACTCAACAGTACATTGCAGCGGCCGCGTTTAAACGAATTCGGATCCTCAGTATTCAA | |
| qZEO-F | CGACGTGACCCTGTTCATCA | |
| qZEO-R | TGGACACGACCTCCGACCA | |
| qHIS-F | GTGTATCCTGGCTTGGCATCT | |
| qHIS-R | GCCAAGTACGGTGTGACGTT | |

Restriction sites are underlined.

**DNA manipulation**

All basic DNA manipulation and analysis were performed as previously described [32]. Restriction digestion was performed in accordance to the manufacturer instructions (New England Biolabs), as well as vector dephosphorylation with Shrimp

240 Alkaline Phosphatase (Promega) and ligation with T4 DNA ligase (USB). In-Fusion
241 Cloning Kit (Clontech) was used for in-vitro assembly of plasmids. Site-directed
242 mutagenesis was performed using the Transformer Site-Directed Mutagenesis kit
243 (Clontech). PCR and gel purification used Promega Wizard SV Gel and PCR Clean-Up
244 System.
245
246 **Quantitative PCR (qPCR)**
247 Cells were grown to an O.D. of 1 in 10 mL YPD containing zeocin and then
248 collected by centrifugation at 2000 x g for 5 minutes. Cell pellet was resuspended with
249 1 mL 0,25% SDS and incubated at 98ºC for 8 minutes according to Nakamura (2018).
250 Finally, cell debris was removed by centrifugation and DNA was diluted 10-fold in
251 water before qPCR reactions.
252 Quantitative PCR reactions used primers qZEO-F and qZEO-R for plasmid
253 quantification and qHIS-F and qHIS-R as an internal single-copy control. Assays were
254 carried out with iTaq Universal SYBR Green Supermix (Bio-Rad) in a Rotor-Gene Q
255 (Qiagen) thermal cycler. Analysis used the absolute quantification method and standard
256 curves that ranged from $1x10^4$ to $1x10^8$ copies of the gene of interest. pPIC9
257 (Invitrogen) and pPICH linearized plasmids were used for construction of the standard
258 curves.
259
260 **Yeast transformation**
261 *K. phaffii* was electroporated following two different protocols. For integrative
262 cassettes, we followed the *Pichia* Expression Kit protocol (Invitrogen) and when using
263 replicative plasmids, we proceeded as described previously [33].

264

265 **Construction of an *ade2 ade3* strain for color-based stability assays**

266 Strain LA2 [29] was transformed with an *ADE3* deletion cassette and had the
267 marker recycled before moving on with the centromeric plasmid transformations.
268 Construction of the deletion cassette used PCR reactions assembled by an "In-Fusion"
269 cloning reaction. Briefly, primers ADE3up-F and R; ADE3dw-F and R were used for
270 PCR amplification of 491 bp and 582 bp, respectively, from *K. phaffii* genome. These
271 reactions amplified sequences used for directing homologous recombination and
272 substitution of the complete *ADE3* coding sequence. Meanwhile, primers ADE3lox-F
273 and ADE3lox-R amplified the *kanR* geneticin resistance cassette from plasmid pGKL
274 [34]. PCR fragments were assembled and cloned into pBluescript II $SK^+$ linearized with
275 *Sma*I. A final PCR reaction using primers ADE3up-F and ADE3dw-R amplified the
276 whole deletion cassette which was used for transformation of *K. phaffii* LA2. Cells were
277 selected in YPD containing geneticin.

278 The resulting *ade2 ade3* strain was later transformed with pYRCre2 [35] and
279 selected in YPD supplied with hygromycin B. This step promoted a Cre-mediated
280 excision of the *kanR* cassette eliminating geneticin resistance. After PCR confirmation
281 of marker recycling using primers ADE3conF and ADE3conR, the resulting strain was
282 plated in non-selective YPD medium, causing loss of the pYRCre2 plasmid. The
283 resulting strain was named LA3.

284

285 **Construction of centromeric plasmids containing *ADE3***

286 Plasmid pPICH [29], which is derived from pPICHOLI (MoBiTec), was digested
287 with *Not*I for cloning of the *K. phaffii* native *ADE3*. The complete gene was amplified

288 from X-33 DNA using primers ADE3up-F and ADE3dw-R following digestion with
289 *Not*I. After vector dephosphorylation, fragments were ligated and transformed into *E.*
290 *coli* XL-10 Gold. One positive clone was then submitted to site-directed mutagenesis
291 using primers Mut1(Hpa) and Mut2(Bam) for removal of the *Bam*HI restriction site
292 present within the *ADE3* coding sequence. The final plasmid containing ARS1, the *Sh*
293 *ble* resistance marker and *ADE3* was named pPICH-ADE3.

294 pPICH-ADE3 was digested with *Bam*HI for cloning of all four *K. phaffii*
295 centromeres. These were amplified from *K. phaffii* X-33 genomic DNA using two PCR
296 reactions for each centromeric sequence. Primers Cen1/2/3/4-F and Cen1/2/3/4c-R
297 amplified the first inverted repeat of each centromere while primers Cen1/2/3/4c-F and
298 Cen1/2/3/4-R amplified the other half of the sequences. In order to promote *in vitro*/*in*
299 *vivo* assembly amplicons had approximately 80 bp homology between each other and
300 15 bp with pPICH-ADE3. Firstly, we attempted an "In-Fusion" cloning reaction for
301 each of the four centromeres using linearized pPICH-ADE3 and the two PCR
302 fragments. Plasmids were extracted and analyzed by restriction digestion. Centromeres
303 1, 2 and 4 were successfully assembled and cloned into the plasmid through this
304 strategy.

305 Centromere 3 did not yield any *E. coli* clones following the "In-Fusion"
306 reaction; therefore, we proceeded to an *in vivo* assembly strategy. Primers Cen370-F
307 and Cen3c-R; Cen3c-F and Cen370-R amplified both inverted repeats adding 70 bp of
308 homologous sequences between the fragments and pPICH-ADE3. Finally, we
309 transformed *K. phaffii* LA3 using the linearized vector and both centromeric fragments,
310 using 85 bp of homology for directing recombination. Clones were selected in YPD
311 supplied with zeocin.

312 The resulting plasmids were named pPICH-CEN1, pPICH-CEN2, pPICH-CEN3
313 and pPICH-CEN4. All plasmids were transformed into *K. phaffii* LA3 for subsequent
314 stability and quantification assays.
315
316 **Stability analysis**
317 LA3 strain transformed with each of the four centromeric plasmids was grown in
318 20 mL YPD for 16 hours at 28ºC and 200 rpm. This culture was inoculated to 20 mL
319 YPD to an initial optical density of 0,1. After 24 h of growth under the same conditions
320 the culture was used as inoculum for another flask containing 20 mL YPD to an $OD_{600}$
321 of 0,1. This procedure was repeated every 24 h until a total of 144 hours. At 96 and at
322 144 hours of growth, a small amount of cells was diluted $10^6$-fold and 100 µL of this
323 dilution were plated on YPD. Plates were incubated at 30 ºC for 72 h.
324
325 **References**
326
327 1. Cregg, J. M., Barringer, K. J., Hessler, a Y., Madden, K. R. *Pichia pastoris* as a host
328 system for transformations. Mol. Cell. Biol. 1985; 5(12): 3376–85.
329 2. De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J.,
330 et al. Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat.
331 Biotechnol. 2009;27(6): 561–566.
332 3. Sturmberger, L., Chappell, T., Geier, M., Krainer, F., Day, K. J., Vide, U., et al.
333 Refined *Pichia pastoris* reference genome sequence. J. Biotechnol. 2016; 235: 121–131.
334 4. Schwarzhans, J. P., Luttermann, T., Geier, M., Kalinowski, J., Friehs, K. Towards
335 systems metabolic engineering in *Pichia pastoris*. Biotechnol. Adv. 2017;35(6): 681–

336 710.

337 5. Cereghino, J. L., Cregg, J. M. Heterologous protein expression in the methylotrophic

338 yeast *Pichia pastoris*. FEMS Microbiol. Rev. 2000; 24(1): 45–66.

339 6. Obst, U., Lu, T. K., Sieber, V. A modular toolkit for generating *Pichia pastoris*

340 secretion libraries. *ACS Synth. Biol.* 2017; 6 (6): 1016-1025.

341 7. Camattari, A., Goh, A., Yip, L. Y., Tan, A. H. M., Ng, S. W., Tran, A., et al.

342 Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia*

343 *pastoris* for recombinant protein production and synthetic biology applications. Microb.

344 Cell Fact. 2016;15(139). doi: 10.1186/s12934-016-0540-5

345 8. Schwarzhans, J.-P., Wibberg, D., Winkler, A., Luttermann, T., Kalinowski, J., Friehs,

346 K. Integration event induced changes in recombinant protein productivity in *Pichia*

347 *pastoris* discovered by whole genome sequencing and derived vector optimization.

348 Microb. Cell Fact. 2016;15( 84). doi: 10.1186/s12934-016-0486-7

349 9. Yu, X., Wang, R., Zhang, M., Xu, Y., Xiao, R. Enhanced thermostability of a

350 Rhizopus chinensis lipase by in vivo recombination in *Pichia pastoris*. Microb. Cell

351 Fact. 2012;11(102). doi : 10.1186/1475-2859-11-102

352 10. Lee, C. C., Williams, T. G., Wong, D. W. S., Robertson, G. H. An episomal

353 expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid

354 2005;54: 80–85.

355 11. Liachko, I., Youngblood, R. A., Tsui, K., Bubb, K. L., Queitsch, C., Raghuraman,

356 M. K., et al. GC-Rich DNA elements enable replication origin activity in the

357 methylotrophic yeast *Pichia pastoris*. PLOS Genet. 2014;10:3. doi:

358 10.1371/journal.pgen.1004169

359 12. Cao, M., Gao, M., Lopez-garcia, C. L., Wu, Y., Seetharam, A. S., Severin, A. J., et

360 al. Centromeric DNA facilitates nonconventional yeast genetic engineering. ACS Synth.
361 Biol. 2017; 6(8): 1545-1553.
362 13. Kouprina, N., Larionov, V. Transformation-associated recombination (TAR)
363 cloning for genomics studies and synthetic biology. Chromosoma 2016; 125(4): 621-
364 632.
365 14. Lyn Chan, F., Wong, L. H. Transcription in the maintenance of centromere
366 chromatin identity. Nucleic Acids Res. 2012;40(22): 11178–11188.
367 15. Malik, H. S., Henikoff, S. Major evolutionary transitions in centromere complexity.
368 Cell 2009;138(6): 1067–1082.
369 16. Morales, L., Noel, B., Porcel, B., Marcet-Houben, M., Hullo, M. F., Sacerdot, C., et
370 al. Complete DNA sequence of *Kuraishia capsulata* illustrates novel genomic features
371 among budding yeasts (Saccharomycotina). Genome Biol. Evol. 2013;5(12): 2524–
372 2539.
373 17. Kitada, K., Yamaguchi, E., Arisawa, M. Isolation of a *Candida glabrata* centromere
374 and its use in construction of plasmid vectors. Gene 1996;175(1-2): 105–108.
375 18. Ketel, C., Wang, H. S. W., McClellan, M., Bouchonville, K., Selmecki, A., Lahav,
376 T., et al. Neocentromeres form efficiently at multiple possible loci in *Candida albicans*.
377 PLOS Genet. 2009;5(3). doi: 10.1371/journal.pgen.1000400
378 19. Chatterjee, G., Sankaranarayanan, S. R., Guin, K., Thattikota, Y., Padmanabhan, S.,
379 Siddharthan, R., et al. Repeat-associated fission yeast-like regional centromeres in the
380 ascomycetous budding yeast *Candida tropicalis*. PLOS Genet. 2016;12(2). doi:
381 10.1371/journal.pgen.1005839
382 20. Smirnova, J. B., McFarlane, R. J. The unique centromeric chromatin structure of
383 *Schizosaccharomyces pombe* is maintained during meiosis. J. Biol. Chem. 2002;

384 277(22); 19817–19822.

385 21. Varoquaux, N., Liachko, I., Ay, F., Burton, J. N., Shendure, J., Dunham, M. J., et al.
386 Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic
387 Acids Res. 2015;43(11): 5331–5339.

388 22. Coughlan, A. Y., Hanson, S. J., Byrne, K. P., Wolfe, K. H. Centromeres of the yeast
389 *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. Genome
390 Biol. Evol. 2016;8(8): 2482–2492.

391 23. McCarroll, R. M., Fangman, W. L. Time of replication of yeast centromeres and
392 telomeres. Cell 1988;54(4): 505–513.

393 24. Koren, A., Tsai, H., Tirosh, I., Burrack, L. S., Barkai, N., Berman, J. Epigenetically-
394 inherited centromere and neocentromere DNA replicates earliest in s-phase. PLOS
395 Genet. 2010;6(8). doi: 10.1371/journal.pgen.1001068

396 25. Nakamura, Y., Nishi, T., Noguchi, R., Ito, Y., Watanabe, T., Nishiyama, T., et al.
397 Construction of a new stable, autonomously replicating plasmid vector containing
398 *Pichia pastoris* centromeric DNA. Appl. Environ. Microbiol. 2018;84(15). doi:
399 10.1128/AEM.02882-17

400 26. Hieter, P., Mann, C., Snyder, M., Davis, R. W. Mitotic stability of yeast
401 chromosomes : a colony color assay that measures nondisjunction and chromosome
402 loss. Cell 1985;40(2): 381–392.

403 27. Ugolini, S., Bruschi, C. V. The red/white colony color assay in the yeast
404 *Saccharomyces cerevisiae*: epistatic growth advantage of white ade8-18, ade2 cells over
405 red ade2 cells. Curr. Genet. 1996;30(6): 485–492.

406 28. Du, M., Battles, M. B., Nett, J. H. A color-based stable multi-copy integrant
407 selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as

408 auxotrophic markers. Bioeng. Bugs 2012;3(1): 32–37.

409 29. Piva, L. C., De Marco, J. L., de Moraes, L. M. P., Reis, V. C. B., Torres, F. A. G.
410 Acetamidase as a dominant recyclable marker for *Komagataella phaffii* strain
411 engineering. Appl. Microbiol. Biotechnol. 2018;102(6): 2753–2761.

412 30. Rébora, K., Laloo, B., Daignan-Fornier, B. Revisiting purine-histidine cross-
413 pathway regulation in *Saccharomyces cerevisiae*: a central role for a small molecule.
414 Genetics 2005;170(1): 61–70.

415 31. Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K.,
416 Verstrepen, K. J. Improving industrial yeast strains: exploiting natural and artificial
417 diversity. FEMS Microbiol. Rev. 2014;38(5): 947–995.

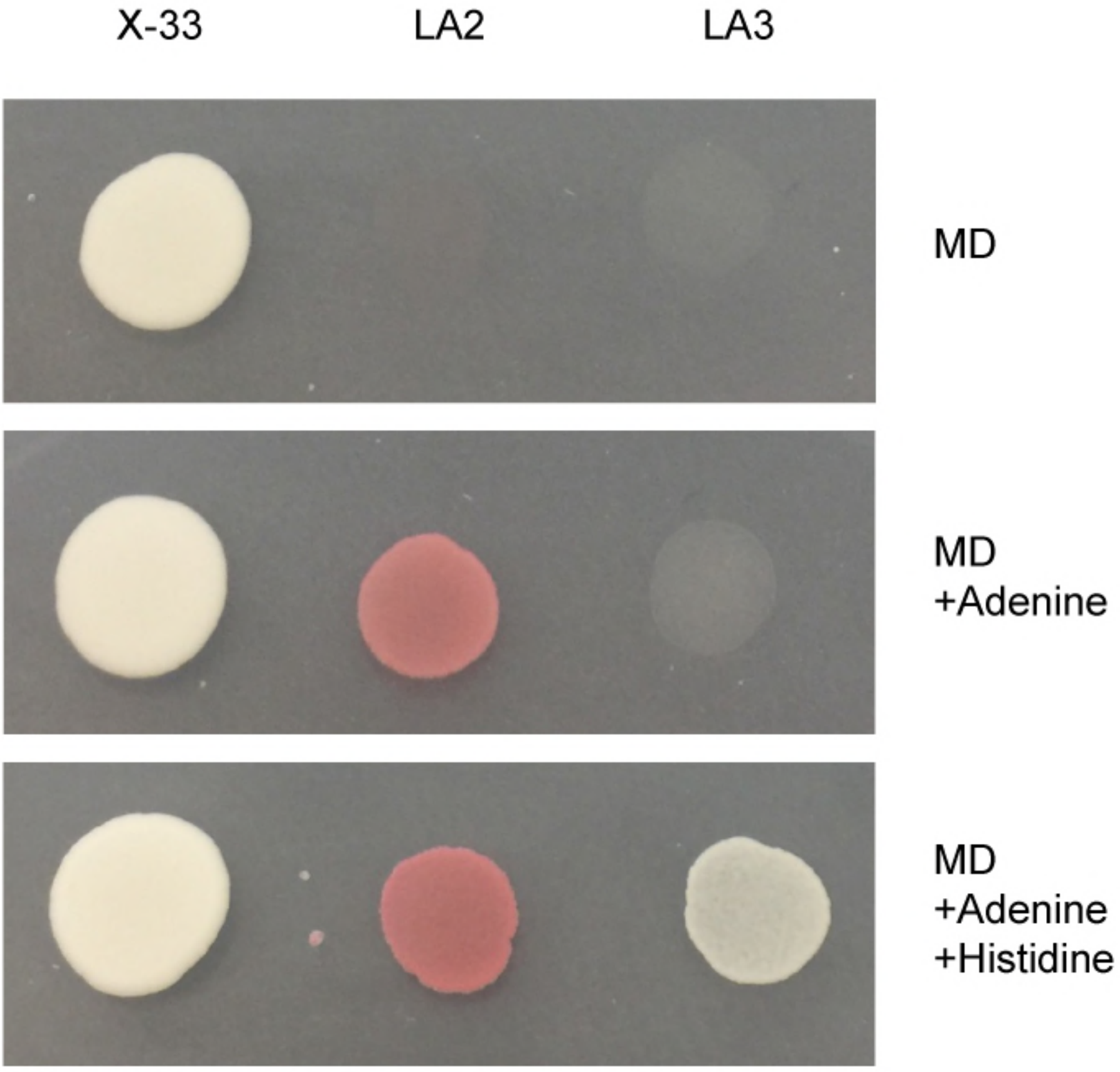
418 32. Sambrook, J., Fritsch, E., Maniatis, T. Molecular cloning: a laboratory manual. 2nd
419 ed. New York: Cold Spring Harbor Laboratory Press; 1989.

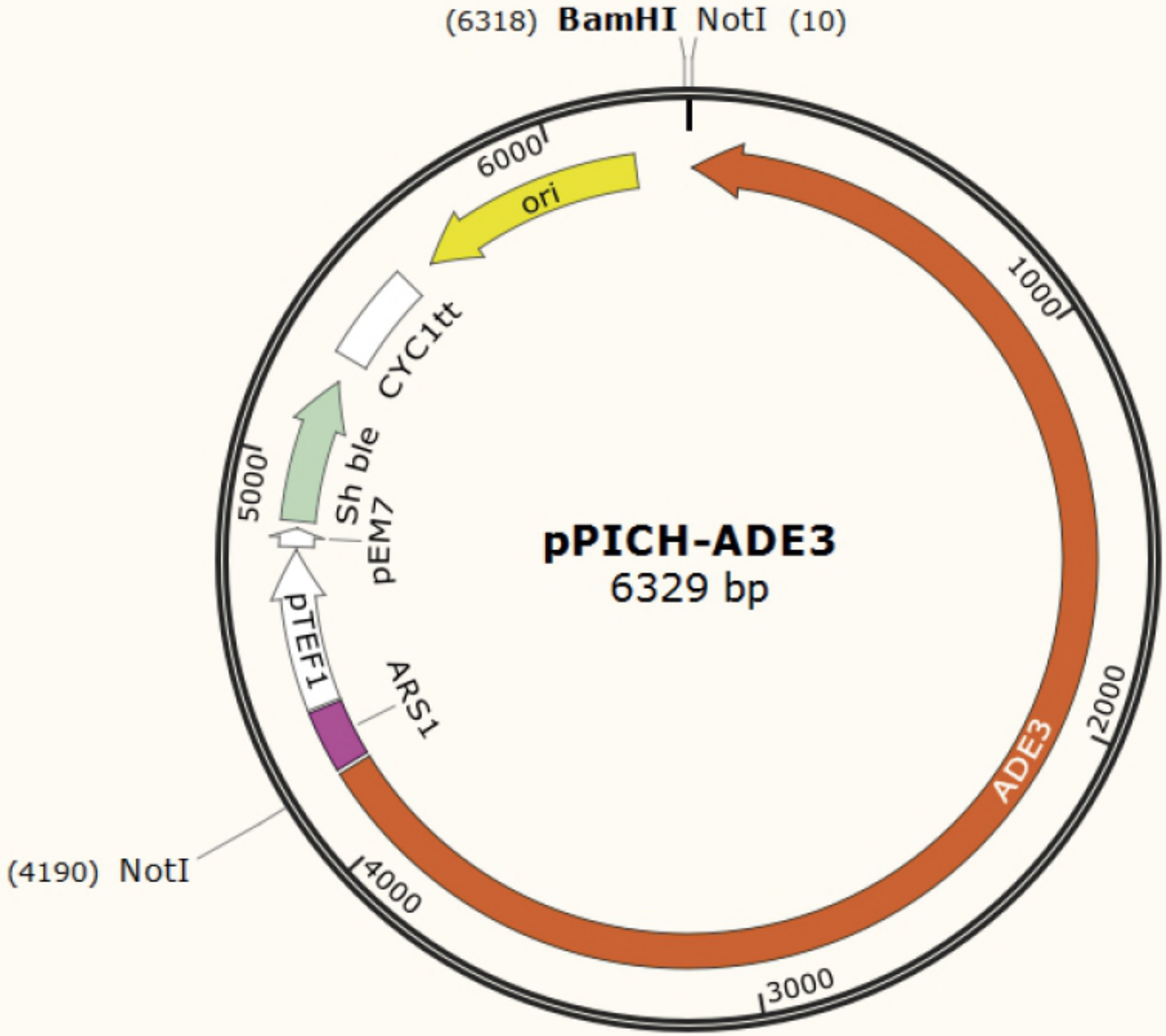
420 33. Wu, S., Letchworth, G. J. High efficiency transformation by electroporation of
421 *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. Biotechniques
422 2004;36(1): 152–154.

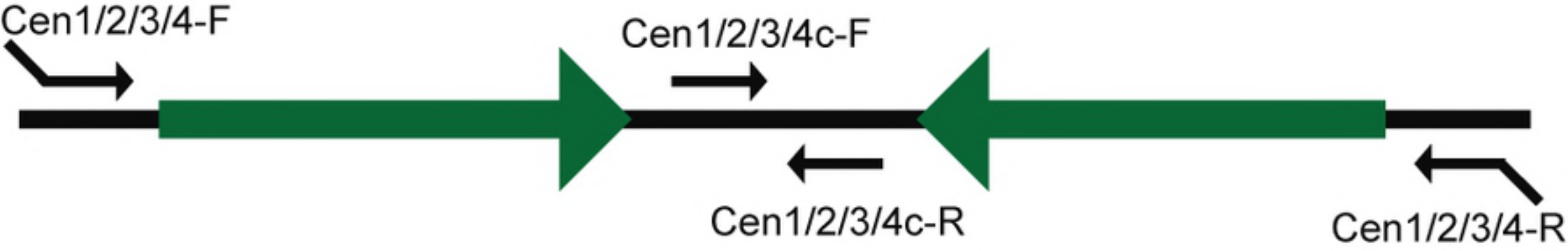
423 34. Reis, V. C. B., Nicola, A. M., de Souza Oliveira Neto, O., Batista, V. D. F., de
424 Moraes, L. M. P., Torres, F. A. G. Genetic characterization and construction of an
425 auxotrophic strain of *Saccharomyces cerevisiae* JP1, a Brazilian industrial yeast strain
426 for bioethanol production. J. Ind. Microbiol. Biotechnol. 2012;39(11): 1673–1683.

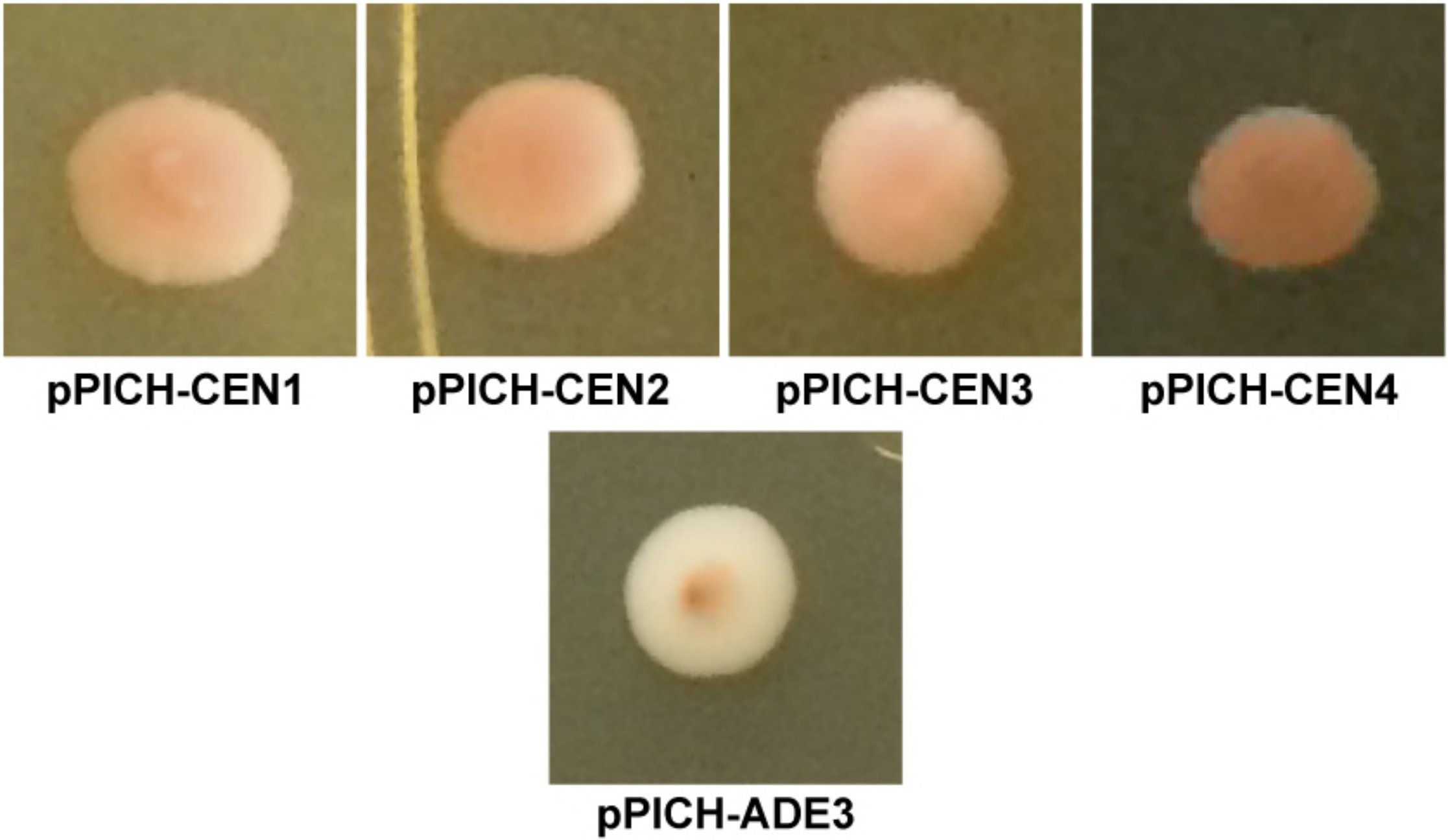
427 35. Betancur, M. O., Reis, V. C. B., Nicola, A. M., De Marco, J. L., de Moraes, L. M.
428 P., Torres, F. A. G. Multicopy plasmid integration in *Komagataella phaffii* mediated by
429 a defective auxotrophic marker. Microb. Cell Fact. 2017;16(99). doi: 10.1186/s12934-
430 017-0715-8

432 **Fig 1. Strain phenotypic analysis on defined media.** *K. phaffii* X-33 (wild-type), LA2
433 and LA3 cultures were spotted on MD medium with or without supplementation.
434 **Fig 2. Map of vector pPICH-ADE3.** The *Not*I site was used for cloning *K. phaffii*
435 *ADE3* gene.
436 **Fig 3. Strategy for amplification of *K. phaffii* centromeres.** Schematic representation
437 of a typical *K. phaffii* centromere. Inverted repeats are represented by green arrows.
438 Primer annealing regions are shown by small arrows.
439 **Fig 4. Plasmid stability analysis.** LA3 strain transformed with pPICH-ADE3 and the
440 four centromeric plasmids was grown on YPD medium for 3 days until color
441 development.
442 **Fig 5. Plasmid stability test.** (A) Aliquots of the liquid cultures were collected at after
443 96 and 144 hours of growth and plated on YPD medium. Red colonies represent cells
444 that maintained the *ADE3*-containing plasmid, while white colonies have lost it. Red
445 portions of the bars represent red colonies; light pink bars represent white colonies. (B)
446 A plate representing a typical result after 144 h growth.
447 **Fig 6. Plasmid copy number determination.** The number of plasmids in each cell was
448 estimated by qPCR. LA3 was used as negative control. Statistical analysis comparing
449 each of the centromeric plasmids with the replicative vector was performed through a t-
450 test using GraphPad Prism 5. ($p<0.05$). Error bars depict the standard deviation of the
451 mean ($n = 3$).
452

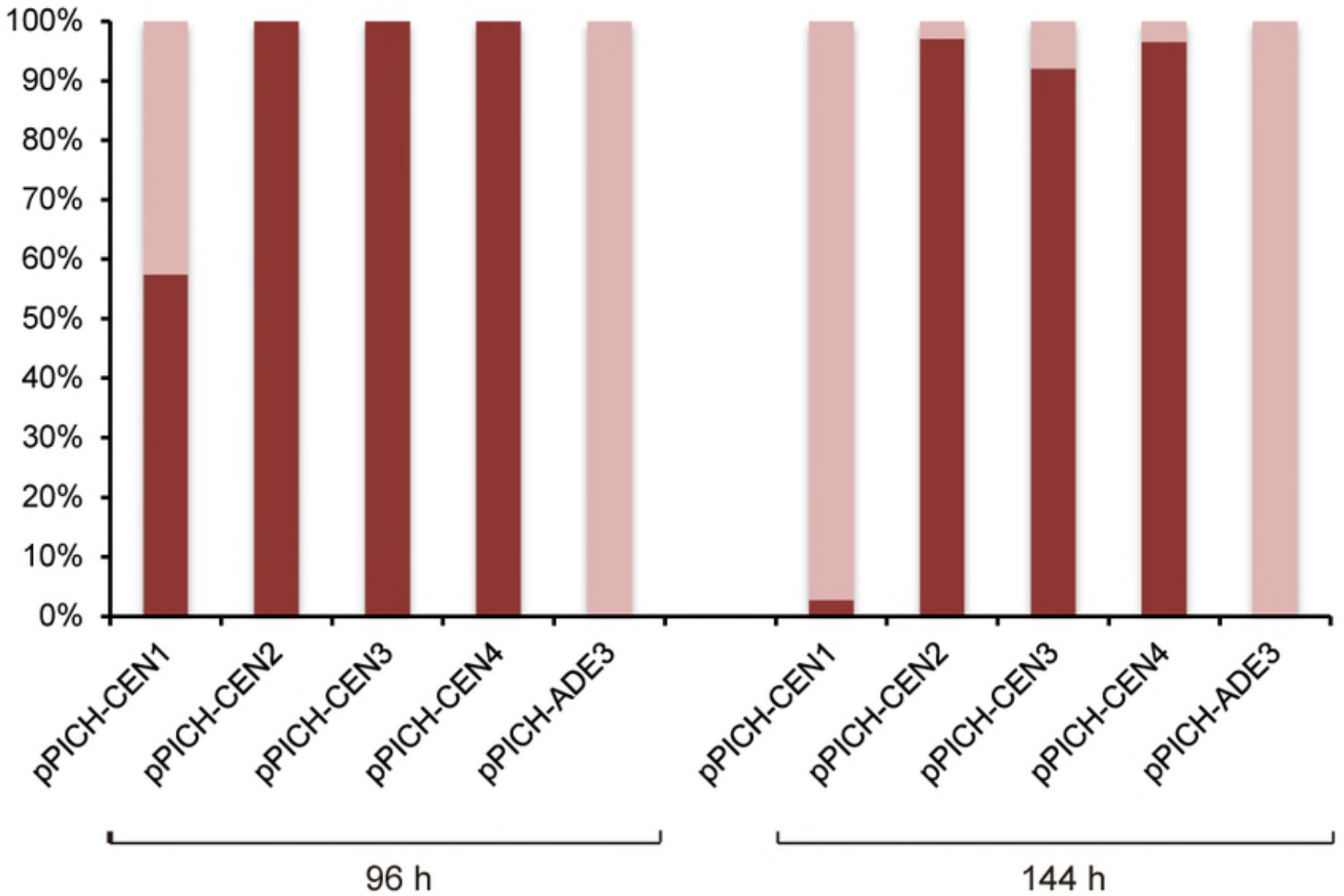

X-33
LA2
LA3
MD
MD
+Adenine
MD
+Adenine
+Histidine


(6318) BamHI NotI (10)
6000
ori
1000
CYC1tt
Sh ble
5000
pEM7
pPICH-ADE3
6329 bp
pTEF1
ARS1
2000
ADE3
(4190) NotI
4000
3000

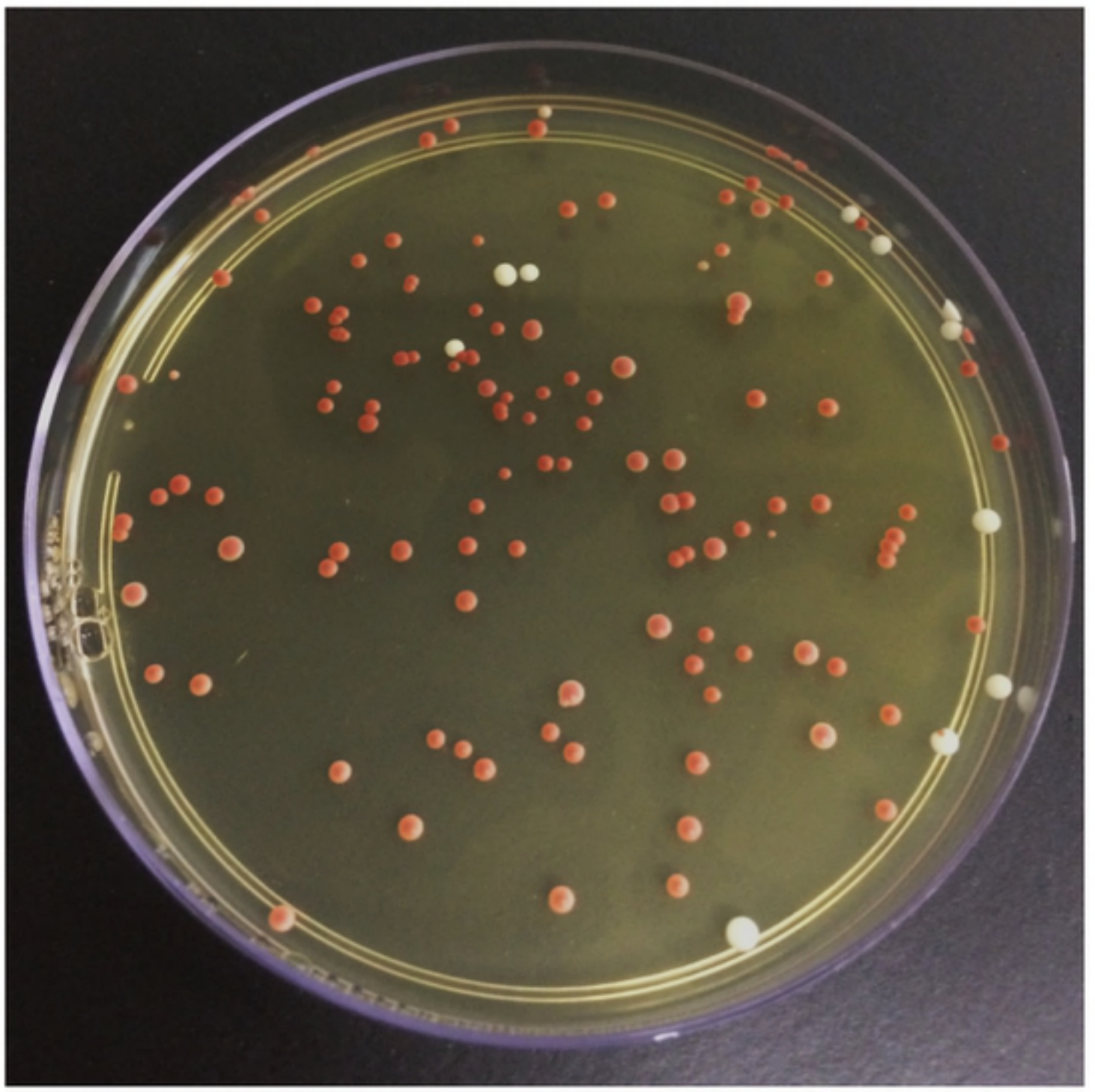

Cen1/2/3/4-F
Cen1/2/3/4c-F
Cen1/2/3/4c-R
Cen1/2/3/4-R

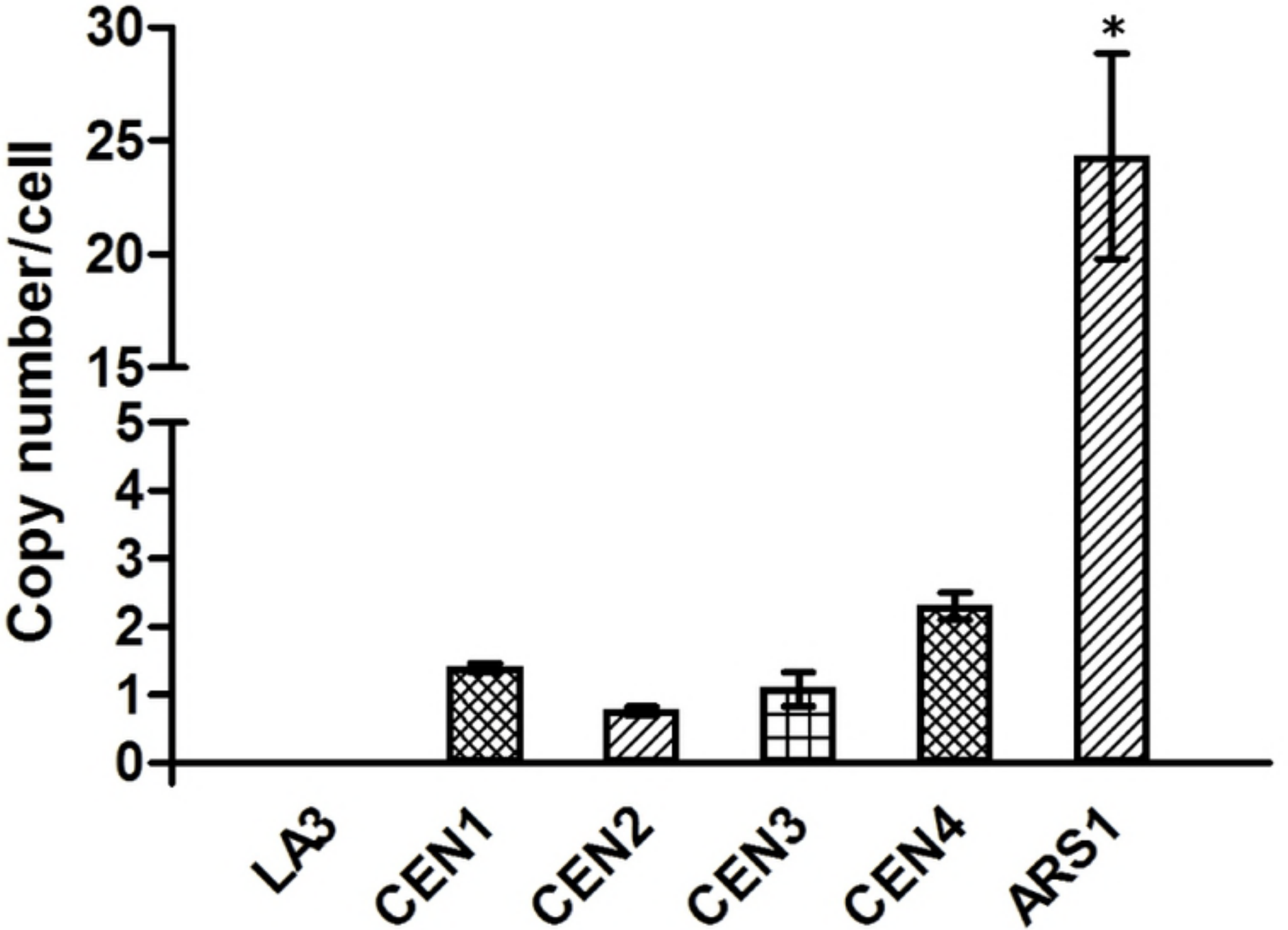

pPICH-CEN1
pPICH-CEN2
pPICH-CEN3
pPICH-CEN4
pPICH-ADE3

A


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
pPICH-CEN1
pPICH-CEN2
pPICH-CEN3
pPICH-CEN4
pPICH-ADE3
pPICH-CEN1
pPICH-CEN2
pPICH-CEN3
pPICH-CEN4
pPICH-ADE3
96 h
144 h

B




Copy number/cell
30
25
20
15
5
4
3
2
1
0
*
LA3
CEN1
CEN2
CEN3
CEN4
ARS1