2
3 **Screening of accurate clones for gene synthesis in yeast**

4
5
6 Running title: Screening of accurate clones for gene synthesis

7
8 Tohru Yarimizu,^{1,§} Mikiko Nakamura,² Hisashi Hoshida,^{1,*} and Rinji Akada¹

9
10 *Department of Applied Molecular Bioscience, Graduate School of Medicine,¹ Innovation*
11 *Center,² Yamaguchi University, Tokiwadai, Ube 755-8611, Japan.*

12
13 *Corresponding author at: Tokiwadai 2-16-1, Ube 755-8611, Japan. Tel.: +81 836 85 9293,
14 Fax: +81 836 85 9201. *E-mail address:* hoshida@yamaguchi-u.ac.jp (H. Hoshida)

15
16 §Present address: National Institute for Agro-Environmental Sciences, Environmental
17 Biofunction Division, 3-1-3 Kan-nondai, Tsukuba, Ibaraki 305-8604, Japan

18
19 **[Key words:** gene synthesis; *Kluyveromyces marxianus*; NHEJ-mediated DNA cloning;
20 luciferase; glucoamylase; codon optimization]

1 Methods for error-less gene synthesis are desired because synthesized genes often contain
2 mutations. By cloning PCR-assembled oligonucleotide fragments fused to a selection marker
3 in yeast, we developed a novel method to screen accurate clones in gene synthesis. As a
4 model case, the 555-bp luciferase gene from *Gaussia princeps* (GLuc) was synthesized to
5 contain yeast-optimized codons (called yGLuc hereafter). After standard PCR-mediated
6 oligonucleotide assembly, many clones showed no luciferase activity. Of these clones, most
7 contained randomly located nucleotide deletions that produced frameshifts and resulted in
8 premature termination. To exclude clones with premature termination, the synthesized yGLuc
9 gene was cloned in-frame to fuse with the *URA3* coding sequence, which served as a
10 selection marker in the yeast *Kluyveromyces marxianus*. Ura⁺ transformation selection was
11 expected to eliminate clones with frameshift errors. The results showed that in-frame marker
12 selection increased the frequency of active yGLuc gene to 79%. We used this strategy to
13 synthesize the 1812-bp gene from *Rhizopus oryzae* that encodes glucoamylase. Five out of
14 seven Ura⁺ clones exhibited amylase activity. Of the functional clones, one contained the
15 correct sequence, and four contained sequences with nucleotide changes, suggesting that
16 in-frame selection frequently produced functional mutants. The *K. marxianus*
17 non-homologous end joining mediated cloning method for gene synthesis will be useful for
18 synthetic biological studies.

Gene synthesis through the assembly of synthetic oligonucleotides is used in genetic engineering. For cloning genes or cDNA, gene synthesis is a more attractive strategy than PCR amplification from template DNA because organisms are not required for source material. In higher eukaryotes, cDNA can be difficult to prepare if expression is low or restricted to certain developmental stages (1, 2). In addition, alternative splicing produces mRNA variants, some of which are not easily obtained when mRNA is prepared from tissue samples or from cDNA libraries (3-5). A similar situation is encountered in microbiology. Metagenome analyses using next-generation sequencers are generating a vast amount of sequence data without cultured microorganisms (6-8). Genes identified from this data cannot be easily obtained by PCR amplification from template DNA samples if the template microorganisms are rare or unculturable. In addition, genes from highly hazardous organisms and viruses, such as Ebola, Lassa, and Variola viruses, cannot be easily prepared because the pathogens must be handled in a high-level biosafety facility. Therefore, gene synthesis is a useful strategy for obtaining DNA of interest without living organisms.

Another useful feature of gene synthesis is codon optimization for gene expression, which enables higher expression of recombinant genes in different host organisms. For example, higher expression of codon-optimized green fluorescent protein (yEGFP) (9), red fluorescent protein (yEmRFP) (10), human growth hormone (11), lipase from *Rhizopus oryzae* (12), glucoamylase from *Aspergillus awamori* (13), and viral proteins (14, 15) has been demonstrated in yeast hosts. The DNA sequences of the codon-optimized genes differ from the original sequences and require gene synthesis for preparation.

The whole genome of *Mycoplasma genitalium* has been synthesized (16-19). Since this milestone event, the pace of synthetic biological studies that design and construct biological parts, devices, and systems has accelerated. The goal of the Synthetic Yeast 2.0 project is the chemical synthesis of the whole genome of the yeast *Saccharomyces cerevisiae* (20, 21).

1 Therefore, gene synthesis has great potential in life sciences and biotechnological
2 applications.

3 The basic technological concept of gene synthesis is the assembly of synthetic
4 oligonucleotides. The polymerase chain assembly method is a primary strategy (22-24).
5 Ligase is also used for assembly in a process called ligase chain reaction (LCR) (25).
6 However, these strategies are not reliable because they often generate error fragments that
7 contain incorrect sequences. The dominant source of error is incorrect oligomers (26).
8 Therefore, highly purified oligomers obtained from high-performance liquid chromatography
9 or polyacrylamide gel electrophoresis are often required to decrease the error frequency. As
10 an alternative way to reduce errors, an error correction strategy has been developed, which
11 utilizes mismatch repair mechanisms. Bacterial mismatch-cleaving enzymes have been used
12 to remove errors in chloramphenicol-acetyltransferase gene synthesis (27). In addition, the
13 mismatch detection and binding protein MutS from *Thermus aquaticus* has been used for error
14 correction during green fluorescent protein synthesis (28). However, the use of mismatch
15 repair proteins for error correction has been limited, possibly because enzyme conditions
16 suitable for repair are difficult to achieve (26). To overcome these problems, another
17 approach has been developed that detects clones with correctly synthesized fragments. In this
18 approach, green fluorescence protein (GFP) is fused in frame to the synthesized gene (29).
19 Kim *et al.* (29) successfully used this post-synthesis error detection approach to synthesize
20 six genes in *Escherichia coli*.

21 In this study, we used a simple yeast cloning system (30) to improve the post-synthesis
22 error detection approach. We cloned PCR-assembled fragments into the yeast *Kluyveromyces*
23 *marxianus* by fusing them in-frame with the *URA3* coding sequence, placed downstream as a
24 selection marker. Premature termination due to sequence errors resulted in the loss of *URA3*
25 function, which inhibited colony growth on selection plates. Thus, selected transformants

were expected to express fragments without premature terminations. We applied this strategy to secretory luciferase and glucoamylase genes and obtained correct clones and active mutant clones at a high frequency. The direct functional selection of synthetic genes in yeast will be useful for genetic engineering and synthetic biological studies.

MATERIALS AND METHODS

Yeast strains, media, and primers The *Saccharomyces cerevisiae* and *Kluyveromyces marxianus* yeast strains used in this study are listed in Table 1. Cells were grown in YPD medium (1% yeast extract, 2% polypeptone, and 2% glucose) and synthetic drop-out media (0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 2% glucose, and required nutrients) (30). If necessary, 2% agar was added. Starch plates contained 1% soluble starch in YPD. To check for plasmid loss, 5-fluoroorotic acid (FOA) plates were prepared as described previously (31).

Gene synthesis and cloning in *S. cerevisiae* The luciferase gene from *Gaussia princeps* (GLuc: 558 bp with stop codon) (11) and a glucoamylase gene from *Rhizopus oryzae* (RoGLUI: 1815 bp with stop codon) (32) were selected as synthesis targets. The DNA sequences of yeast codon-optimized GLuc (yGLuc) and RoGLUI (yRoGLUI) were determined from the amino acid sequences (Supplementary Fig. S1). The yeast codons used for sequence design are shown in Supplementary Table S1. The oligonucleotides used are listed in Supplementary Table S2. The oligonucleotides were dissolved in sterilized water to a concentration of 10 μ M. For yGLuc synthesis, 80-mer or 40-mer oligonucleotides from both strands were designed and purchased from Fasmac Co., Ltd. (Atsugi, Japan). The same 40-mer oligonucleotides were also purchased from Hokkaido System Science Co., Ltd. (Sapporo, Japan). For assembly, three different oligonucleotide mixtures, named F80/C40, F80/C80, and F40/C40, were prepared (Fig. 1A). F80/C40 was a mixture of 80-mer forward

oligonucleotides and 40-mer complementary oligonucleotides containing 20 base overlaps in the forward strand. F80/C80 was a mixture of 80-mer forward and 80-mer complementary oligonucleotides. F40/C40 was a mixture of 40-mer forward and 40-mer complementary oligonucleotides. The concentration of each oligonucleotide was adjusted to 0.5 μ M in the mixtures. For yRoGLUI synthesis, 50-mer oligonucleotides of both strands without gaps were designed and purchased from Hokkaido System Science. The concentration of each oligonucleotide in the mixture was 0.27 μ M.

For oligonucleotide assembly, a PCR mixture was prepared by mixing 1 μ l of 10 $\times$ KOD -Plus- buffer, 1 μ l of 2 mM dNTPs, 0.4 μ l of 25 mM of MgSO₄, ~~6.4 μ l~~ 5.4 μ l of water, and 0.2 μ l of KOD -Plus- polymerase (Toyobo, Osaka, Japan). Subsequently, 2 μ l of F80/C40, F80/C80, or F40/C40 mixture was added. The standard assembly reaction was initiated at 94°C for 1 min, followed by 20 cycles of 94°C for 20 s, 55°C for 30 s, and 68°C for 1 min. For synthesis experiments with the F40/C40 mixture, the annealing and extension temperatures were changed, as indicated in the Results section. Phusion High-Fidelity DNA polymerase (New England BioLabs Inc., Ipswich, MA, USA) was also used for PCR assembly for the F40/C40 mixture and yRoGLUI synthesis. The Phusion polymerase reaction mixture consisted of 2.0 μ l of 5 $\times$ Phusion HF buffer, 1.0 μ l of 2 mM dNTPs, 0.2 μ l of MgCl₂, 0.3 μ l of dimethyl sulfoxide, 5.4 μ l of water, 0.1 μ l of DNA polymerase, and 1 μ l of an oligonucleotide mixture. The reaction was initiated at 98°C for 30 s, followed by 30 cycles of 98°C for 7 s, 55°C for 20 s, and 72°C for 1 min. These reactions were performed in an iCycler thermal cycler (Bio-Rad, Hercules, CA, USA).

The PCR-assembled oligonucleotides were used as a template, and the DNA fragment to be cloned was amplified by PCR using KOD -Plus- or KOD FX Neo DNA polymerase (Toyobo). The reaction mixture for KOD -Plus- was prepared by mixing 1 μ l of 10 $\times$ KOD -Plus- buffer, 1 μ l of 2 mM dNTPs, 0.4 μ l of 25 mM of MgSO₄, 5.8 μ l of water, and 0.2 μ l of

KOD -Plus- DNA polymerase. Thereafter, 1 μ l of the assembled oligonucleotide and 0.3 μ l each of 10 μ M forward and reverse primers were added. The reaction was initiated at 94°C for 1 min, followed by 30 cycles of 94°C for 20 s, 60°C for 30 s, and 68°C for 1–4 min. The KOD FX Neo reaction mixture was prepared by mixing of 5 μ l of 2 $\times$ KOD FX Neo buffer, 2 μ l of 2 mM dNTPs, 0.2 μ l of DNA polymerase, 0.3 μ l each of 10 μ M forward and reverse primers, and 1 μ l of the assembled oligonucleotides in a total volume of 10 μ l. The PCR reaction was initiated at 94°C for 2 min, followed by 30 cycles of 98°C for 10 s, 65.0°C for 30 s, and 68°C for 1–3 min. For *yGLuc* and *yRoGLUI* amplification, TDH3p40-*yGLuc*/15G-*yGLuc* and TDH3p40-*glucoR*/15G-*glucoR* were used as the primer pairs, respectively. For cloning in *S. cerevisiae*, the amplified DNA fragments were fused with the transformation marker *ScLEU2* using fusion PCR (33, 34). *ScLEU2* was amplified from the chromosomal DNA of RAK3614 (35) using the 15C-LEU2-1 and URA3-300c primers. The synthesized gene and the *ScLEU2* fragments were mixed and amplified using the TDH3p40 and URA3-200c primers. In this case, *ScLEU2* and the synthesized fragment were not fused in-frame. The fragment was used to transform the *S. cerevisiae* RAK4314 strain for integration into the *ura3 Δ 0::ScTDH3p_{15C}ScURA3* locus (35). *Leu*⁺ and *Ura*⁻ colonies were picked as transformants with the correct sequence (35).

Transformation of *K. marxianus* for cloning The *K. marxianus* RAK3908 strain (36) was cultured in 30 ml of YPD medium in a 250-ml flask with shaking at 150 rpm for 24 h at 30°C. The cells were collected by centrifugation at 3000 rpm for 3 min and suspended in 900 μ l of transformation buffer [TFB: 40% polyethylene glycol 3350 (Sigma-Aldrich, Tokyo, Japan), 0.1 M dithiothreitol (Wako, Osaka, Japan), 0.2 M lithium acetate (Kishida Chemical, Osaka, Japan)] (37). The cells were centrifuged again and suspended in 600 μ l of fresh TFB. DNA fragments were mixed with 50 μ l of the cell suspension and incubated at 42°C for 30 min. The mixture was spread onto a synthetic dropout plate and incubated at 28°C for 2–3 d.

***K. marxianus* plasmid construction**

For in-frame marker selection, the pKM149 plasmid was prepared from pKM019. The plasmid was constructed in *K. marxianus*, as described previously (30). To construct the pKM019 plasmid, the Sc*URA3* gene was amplified from pRS316KmARS7 (30) by PCR using the pRS316URA3-226 and URA3+880c primers. This fragment was used as a template for the second PCR with the KmARS7(201-260)316URA3-226 and 15C-URA3+880c primers. The Sc*ADE2* gene was amplified from the chromosomal DNA of the *S. cerevisiae* BY4743 strain using the ADE2-797 and 15G-ADE2+1916c primers. The two DNA fragments were fused by fusion PCR through the 15C/15G sequence using the KmARS7(201-260)316URA3-226 and ADE2-797 primers (33). The fused DNA fragment was used to transform the *K. marxianus* RAK3908 strain, generating the RAK5986 strain expressing pKM019 (Supplementary Fig. S2).

From pKM019, the pKM030 plasmid was constructed. A centromere sequence, KmCenD (30, 38), was amplified from the chromosomal DNA of *K. marxianus* DMKU3-1042 (39) using the 9C-CenDCDEIII-211 and CenDCDEIII+36c primers. To the end of this fragment, a short C-terminal sequence of Sc*URA3* was attached through a second PCR with the URA3+772-9c(118) and CenDCDEIII+36c primers. Vector DNA was amplified from the total DNA of the RAK5986 strain containing pKM019 using the ADE2+1716c and URA3+771c primers. This amplified DNA fragment contained truncated, non-functional Sc*URA3*. The two DNA fragments were mixed and used to transform the *K. marxianus* RAK3908 strain (36). Non-homologous end joining (NHEJ) between the truncated Sc*URA3* and the Sc*URA3* C-terminus produced functional *URA3* and generated a recombinant plasmid in *K. marxianus* (30), resulting in construction of pKM030 in the RAK6140 strain (Supplementary Fig. S2).

Finally, the pKM149 plasmid was constructed from pKM030. First, the yCLuc gene

(Atto, Tokyo, Japan) was amplified from the chromosomal DNA of *S. cerevisiae* strain RAK4296 (35) using the TDH3p40yCLuc+1 and 15G-yCLuc+1662c primers. The ScLEU2 gene was amplified from the chromosomal DNA of *S. cerevisiae* RAK3625 (35) using the 15C-LEU2-1 and URA3-300c primers. The two fragments were fused by PCR with TDH3p40yCLuc+1 and URA3-300c, and the fused fragment was used to transform *S. cerevisiae* strain RAK4314 (35) to construct *ura3Δ0::ScTDH3p-yCLuc-ScLEU2* (*S. cerevisiae* RAK4920 strain). The ScLEU2 marker of RAK4920 was replaced with ScURA3, generating the RAK4960 strain. From the chromosomal DNA of RAK4960, the ScTDH3p-yCLuc-ScURA3ΔC fragment containing a C-terminal deletion of ScURA3 was amplified using the TDH3-698 and URA3+771c primers. The linearized PCR fragment of pKM030 was prepared from the total DNA of RAK6140 using the URA3+771c and URA3+772 primers. The second PCR was performed using the URA3+772 and KmARS7(201-260)-ADE2-797 primers and the PCR product as template. The two DNA fragments were mixed and used to transform *K. marxianus* strain RAK3908 (36). The Ade⁺ Ura⁺ CLuc⁺ transformants were examined for growth on FOA plates (31) to verify plasmid replication. Thus, the pKM149 plasmid constructed in the *K. marxianus* RAK6202 strain contained ScADE2 and ScURA3 as *K. marxianus* selection markers, ScTDH3 promoter-driven yCLuc, an autonomously replicating sequence from *K. marxianus* (KmARS7), and a centromere sequence from *K. marxianus* (KmCenD) (Supplementary Fig. S2).

In-frame marker selection in *K. marxianus* To prepare vector DNA, total DNA from RAK6202 containing pKM149 was used as a template, and PCR was performed using URA3+21 and yGLuc+3-TDH3-1c primers. For the N-terminal deletion of yGLuc=Ura3 (“=” is used to indicate in-frame fusion in this paper), the chromosomal DNA of RAK5125 containing the wild-type yGLuc was used as a template, and PCR was performed using the

yGLuc+4, yGLuc+28, yGLuc+52, or yGLuc-4 (from +79) primer and the URA3(+4-20)c-yGLuc+555c primer. The vector fragment and the N-terminus deleted yGLuc fragment containing the short *URA3* N-terminus sequence were mixed and used to transform the RAK3908 strain (36). Then, Ura⁺ transformation efficiency was examined. For in-frame marker selection, the same PCR-assembled yGLuc fragments prepared for *S. cerevisiae* cloning were used as templates, and DNA fragments were amplified using the URA3(+4-20)c-yGLuc+555c and yGLuc+4 primers. The vector and the yGLuc fragment were mixed and used to transform the RAK3908 strain (36). Similarly, the assembled yRoGLU1 fragments were used as a template and amplified using the URA3+30c-yRoGLU1(1793-1812)c primer and the yRoGLU1-4 or yRoGLU1+43 primer. The vector and the yRoGLU1 fragments were mixed and used to transform the RAK3908 strain. The N-terminal-deleted yRoGLU1 was sequenced.

For the addition of the N-terminus of yRoGLU1 and the removal of in-frame *URA3*, the clones were amplified by PCR using the yRoGLU1+1(75) and 3CG9-yRoGLU1+1815c primers. The second PCR was performed using the yRoGLU1+1(75) and URA3+772term3CG9 primers through a 3CG9 minimum annealing sequence (40). The vector DNA, pKM288, contained the *ScTDH3* promoter, *ScURA3*, *ScADE2*, and KmCEND/KmARS7 (RAK9817 strain; details of the construction process will be published elsewhere). This vector was used as template and amplified using TDH3-1c and URA3+771c. The two fragments were mixed and used to transform RAK3908. The resulting transformants were suspended and spotted on starch plates. After incubation for 2 d, the plates were exposed to iodine vapor, as described previously (41).

Luciferase assay Yeast cells were incubated in 1 ml of YPD with shaking at 28°C overnight. To detect luciferase activity, a BioLux *Gaussia* Luciferase Assay Kit (New England BioLabs Inc.) was used. To prepare the substrate solution, 10 µl of 100× BioLux

1 GLuc substrate and 1000 μ l of BioLux GLuc Assay Buffer were mixed. Subsequently, 20 μ l
2 of the substrate solution was mixed with 10 μ l of the yeast culture. The relative luminescence
3 unit (RLU) for 1 sec was obtained using a GloMax 20/20 Luminometer (Promega, Tokyo,
4 Japan) or a Centro LB 960-YC Microplate Luminometer (Berthold Technologies, Wildbad,
5 Germany).

6 The sequences of the *yGLuc* and *yRoGLUI* have been submitted to the
7 DDBJ/EMBL/GenBank databases under accession numbers AB971538 and AB971539,
8 respectively.

9 RESULTS

10 **Assembly PCR for gene synthesis and cloning in *S. cerevisiae*** For the analysis of
11 gene synthesis accuracy, three types of oligonucleotides were designed and purchased from
12 different suppliers. Forward and complementary oligonucleotides were synthesized and
13 mixed as shown in Fig. 1. The F80/C40 mixture contained forward 80-mer oligonucleotides
14 and complementary 40-mer oligonucleotides that assembled with gaps in the complementary
15 strand. The F80/C80 and F40/C40 consisted of 80-mer and 40-mer oligomers, respectively,
16 that assembled without gaps. Using the F80/C40, F80/C80, and F40/C40 mixtures, PCR
17 assembly was performed with annealing at 55°C and extension at 68°C. After the assembly,
18 fusion PCR was performed to attach the *ScLEU2* marker gene to the fragment, and the
19 construct was inserted into the *Scura3 Δ 0::ScTDH3p-ScURA3* locus of the *S. cerevisiae*
20 RAK4314 strain (35) (Fig. 1B). The colonies that acquired *Leu*⁺ and *Ura*⁻ phenotypes by gene
21 replacement were picked, and their luciferase activity was measured. However, few GLuc
22 positive (GLuc⁺) colonies were obtained (Table 2, experiments 1–3, *S. cerevisiae* fusion
23 PCR).

24 To increase the accuracy of synthesis, the annealing/extension temperatures in the
25 assembly PCR were increased with the F40/C40 mixture (Nos. No. 4–7 in Table 2). A slight

1 increase in the frequency of positive clones was observed, but the frequency did not exceed 4
2 in 20 colonies. Oligonucleotides purchased from another supplier were also used in the
3 F40/C40 mixture, but the error frequency was similar (Nos. 8–9 in Table 2). We then replaced
4 KOD -Plus- DNA polymerase with Phusion High Fidelity DNA polymerase. The Phusion
5 polymerase resulted in greater accuracy (Table 2, No. 10: 11/20 at 55.0°C and No. 11: 8/20 at
6 72.0°C). These results indicated that the DNA polymerase used for the assembly step affected
7 the frequency of correct gene synthesis. However, approximately half of the clones contained
8 incorrect sequences.

9 To determine what kinds of errors occurred in the inactive clones, we sequenced 15
10 GLuc⁻ clones (Table 3). All sequence errors were randomly located, indicating that there was
11 no preference in error positions (Supplementary Fig. S3). Most of the errors were deletions,
12 and a few substitutions and insertions were found (Table 3). Overall, 13 of 15 GLuc⁻ clones
13 contained coding sequences with frameshifts that caused premature termination
14 (Supplementary Fig. S4). Kim *et al.* reported that deletion and insertion are major error in
15 gene synthesis and in-frame GFP fusion was used for post-synthesis error detection (29).
16 High frequency of frameshifts in GLuc⁻ clones in this study strongly suggested that in-frame
17 fusion of synthesized yGLuc with URA3 coding sequence allowed screening of accurate
18 clones without frameshift. A positive clone from experiment 6 (see Table 2 for details) with
19 no errors, the RAK5125 strain, was used as a source of the wild-type yGLuc gene in
20 subsequent experiments.

21 In-frame marker selection for accurate-clone screening in *K. marxianus*

22 Previously, we described a simple DNA cloning strategy that employs NHEJ in the yeast *K.*
23 *marxianus* (30). The method uses a truncated, non-functional marker gene and a DNA
24 fragment containing a primer-introduced truncated marker region. The highly efficient NHEJ
25 of *K. marxianus* enabled end joining between the truncated marker and the truncated region,

1 which generated a functional marker during transformation. We applied this cloning system
2 to gene synthesis (Fig. 1C).

3 For this experiment, we were concerned that fusion between Ura3 and secretory GLuc
4 would be deleterious to the activities of both proteins. To examine this possibility, wild-type
5 and signal sequence-deleted yGLuc genes were used for in-frame marker selection. The
6 wild-type and N-terminal-deleted yGLuc fragments were amplified from the chromosomal
7 DNA of the RAK5125 strain and used for in-frame marker selection in *K. marxianus*. When
8 wild-type yGLuc was fused to URA3, the number of Ura⁺ transformants was similar to the
9 number obtained when signal sequence-deleted yGLuc was fused to URA3 (Supplementary
10 Table S3). The activity of wild-type yGLuc=Ura3 in the supernatant was weakened (0.3% of
11 the activity of wild-type yGLuc), but still measurable. Therefore, we used the in-frame
12 marker selection method for yGLuc synthesis without signal sequence deletion.

13 The vector pKM149 is a CEN/ARS plasmid in *K. marxianus*, in which ScURA3 is
14 downstream of the ScTDH3 promoter (ScTDH3p) in the same direction. PCR amplification
15 using pKM149 with the ScURA3+21 and yGLuc+3-TDH3-1c primers generated a vector
16 fragment containing an N-terminal-truncated ScURA3 maker and a ScTDH3 promoter with a
17 start codon. The N-terminal 20-bp deletion in the ScURA3 coding sequence completely
18 disrupted Ura3 function (data not shown). As an insert, the assembled yGLuc fragment was
19 amplified using the yGLuc+4 and URA3(+4-20)c-yGLuc+555c primers, which generated a
20 yGLuc fragment without a start codon and with the N-terminus of ScURA3 at the C-terminus
21 (Fig. 1C). The two DNA fragments were then mixed and used to transform the *K. marxianus*
22 *ura3 ade2* mutant strain RAK3908. The Ura⁺ colonies were picked, and their luciferase
23 activities were measured.

24 The assembled oligonucleotides used for cloning in *S. cerevisiae* were used for
25 in-frame marker selection in *K. marxianus* (Table 2, *K. marxianus* in-frame). The frequency

of GLuc⁺ clones from the *S. cerevisiae* fusion PCR strategy was very low in experiments 1–3. However, with these DNA templates, in-frame marker selection produced GLuc⁺ clones at a frequency of 25%–45%. All other in-frame selection experiments produced a higher frequency of GLuc⁺ clones, up to 79% with the template in experiment 9. The results indicated that cloning with in-frame marker selection in *K. marxianus* was effective for screening of accurate genes in gene synthesis.

yRoGLU1 gene synthesis by in-frame marker selection To examine the effectiveness of in-frame marker selection, we synthesized a yeast codon-optimized glucoamylase coding sequence (yRoGLU1) of 1812 bp. The sequence was derived from *Rhizopus oryzae* (42) and cloned in *S. cerevisiae* and *K. marxianus*. PCR assembly was performed with Phusion DNA polymerase at an annealing temperature of 55°C. The assembled fragment was first fused with the ScLEU2 selection marker by fusion PCR and then introduced into the *S. cerevisiae* RAK4314 strain. In 25 transformants picked, no colonies exhibited amylase activity. Four of the inactive transformants were sequenced (Table 4); all contained sequences with premature stop codons, as expected.

The same assembled fragment was used as a template in PCR with the URA3+20c-yRoGLU1(1793-1812)c and TDH3p40-glucoR primers for in-frame marker selection. The fragment was mixed with a pKM149 vector fragment, amplified using the URA3+21 and TDH3p-1c primers, and used to transform the *K. marxianus* RAK3908 strain. Ten Ura⁺ colonies were obtained, but none exhibited amylase activity on starch plates. Analysis of the DNA sequences of two clones (RAK7889 and RAK7890) revealed that both contained N-terminal deletions; the first ATG was the codon for Met of 185th amino acid in yRoGLU1 (Table 4). In the Met185 frame, no premature stop codons were present. On the basis of this result, we assumed that the secretory activity of yRoGLU1 was higher than that of yGLuc and that the former reduced the localization of fused Ura3 to the cytoplasm.

1 In light of these results, the N-terminus-deleted yRoGLUI fragment was amplified
2 from the assembled template using the URA3+30c-yRoGLU1(1793-1812)c and
3 yRoGLU1+42 primers. This deleted the N-terminal +4 to +42 region, coding for 13 amino
4 acids, in yRoGLUI. The fragment was mixed with a vector fragment and used to transform *K.*
5 *marxianus* strain RAK3908. Seven Ura⁺ colonies were obtained and sequenced
6 (Supplementary Fig. S5). All of them showed in-frame fusion of yRoGLUI with URA3
7 (Table 4, Insert: Δ42), indicating that in-frame marker selection worked faultlessly. However,
8 six contained several types of errors (Table 4, Supplementary Fig. S6). The RAK10281 clone
9 contained a correct yRoGLUI sequence without the N-terminus.

10 Because the seven clones contained yRoGLUI with an N-terminal deletion fused to
11 URA3, we attached the wild-type N-terminal sequence and removed the fused URA3
12 sequence (Fig. 2). N-terminus- and stop codon-containing primers were used to amplify
13 yRoGLUI from the seven clones. The vector and these fragments were mixed and used to
14 transform *K. marxianus*. Ura⁺ transformants were picked, and their amylase activities were
15 examined. Five transformants derived from the seven clones, including one with the correct
16 sequence (from RAK10281), exhibited amylase activity on starch plates (Fig. 2B), indicating
17 that the four clones with errors were functional mutants. The RAK10280 active clone had a
18 long amino acid sequence change from the 105th to 152nd codons as a result of internal
19 frameshifts. On the other hand, the RAK10282 inactive clone contained a single amino acid
20 deletion of Trp338, indicating that Trp338 residue is important for amylase activity.

21 22 DISCUSSION

23 Methods for accurate gene synthesis are needed because sequence errors frequently
24 occur during synthesis. The main reason for errors may be the presence of mis-synthesized
25 oligonucleotides (26, 43). In this study, even when assembly conditions such as

annealing/extension temperature, oligonucleotide length, oligonucleotide supplier, and DNA polymerase were varied, the frequency of correct gene synthesis was low (Table 2, 3). In this study and in others (26), the nucleotide errors were randomly located, and deletions were commonly observed. Thus, oligonucleotide quality is important for accurate gene synthesis. However, high quality oligonucleotide synthesis and purification are difficult and expensive at present. Consequently, error correction and detection strategies are required. When the *S. cerevisiae* clones produced in this study were sequenced, 13 of the 15 inactive clones contained frameshifts that resulted in premature termination (Table 3). If these frameshift errors could be eliminated, the majority of the selected clones should contain correct sequences. Therefore, we developed a cloning method that eliminated frameshifts at the C-terminus through in-frame selection of a transformation marker gene (Fig. 1C). The in-frame marker selection method resulted in a much higher frequency of active clones, even when the same assembled DNA sources were used as templates (Table 2).

Error correction methods for synthesized genes have been developed that use DNA repair enzymes (43). The enzymatic error correction methods are effective in generating correct clones. However, in addition to cloning, preparation and adjustment of the enzyme treatment conditions are required. Our in-frame marker selection method is simpler (Fig. 1). However, a drawback of the in-frame method is the need to later eliminate the fused marker gene. Nevertheless, NHEJ cloning in *K. marxianus* offers a simple approach for subsequent modifications, such as deletion of a fused marker or addition of sequences (30).

Selection markers in yeast are usually auxotrophic or drug-resistance proteins that are functional when localized in the cytoplasm. If a synthesized gene product is targeted to another location, such as the extracellular space or nucleus, in-frame marker selection does not appear to work. However, in our case, in-frame fusion of the *URA3* marker to secretory *yGLuc* was successful. Interestingly, the *yGLuc=URA3* fusion gene can be used for both

1 *URA3*⁺ selection (Supplementary Table S3) and the secreted luciferase assay. This suggests
2 that some yGLuc=Ura3 localized to the cytoplasm and some was secreted extracellularly,
3 possibly because the activity of the yGLuc signal sequence was weak. In contrast, a similar
4 construct, yRoGLUI=*URA3*, could not be used for gene cloning without deletion of its signal
5 sequence (Table 4). The yRoGLUI signal sequence might be highly active, thus preventing
6 any localization to the cytoplasm. These results suggest that even if the product of a
7 synthesized gene does not localize to the cytoplasm, in-frame marker selection might be used
8 if the localization signal is weak.

9 When synthesizing yRoGLUI, all seven clones obtained from *URA3*⁺ selection
10 contained correct in-frame constructs, indicating that in-frame marker selection is efficient
11 when synthesizing larger genes. However, only one of the clones (RAK10281) contained the
12 correct sequence. Therefore, the frequency of correct clones was low when a gene of 1815 bp
13 was synthesized. Nevertheless, when the clones with nucleotide changes were used to
14 construct full-length yRoGLUI in *K. marxianus*, four of the clones generated active mutants
15 (Fig. 2). Two mutants contained sequences with multiple long amino acid substitutions
16 caused by two successive frameshifts (Supplementary Fig. S5 and S6). This suggests that
17 active mutants could be obtained frequently if premature termination caused by mutagenesis
18 was eliminated. Usually, random mutagenesis is achieved through error-prone PCR or
19 chemical treatment of DNA, which also produces frameshifts. If in-frame marker selection is
20 used for random mutagenesis, it may be possible to obtain functional mutants more
21 frequently.

22
23 Supplementary data related to this article can be found at <http://>
24
25

ACKNOWLEDGMENTS

We would like to thank Yukie Misumi for technical assistance. This study was supported in part by JSPS KAKENHI (Grant No. 24658096), the Adaptable and Seamless Technology Transfer Program through Target-driven R&D, and the Advanced Low Carbon Technology R&D Program (JST, Japan).

References

1. **Khan, M. A., Kar, M., Mittal, S., Kumar, S., Bharagava, V. L., Sengupta, J. and Ghoshi, D.:** Small scale transcript expression profile of human first trimester placental villi analyzed by a custom-tailored cDNA array, *Indian J. Physiol. Pharmacol.*, **54**, 235-254 (2010).
2. **Ko, M. S., Kitchen, J. R., Wang, X., Threat, T. A., Wang, X., Hasegawa, A., Sun, T., Grahovac, M. J., Kargul, G. J., Lim, M. K. and other 13 authors:** Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, *Development*, **127**, 1737-1749 (2000).
3. **Lareau, L. F., Green, R. E., Bhatnagar, R. S. and Brenner, S. E.:** The evolving roles of alternative splicing, *Curr. Opin. Struct. Biol.*, **14**, 273-282 (2004).
4. **He, C., Zuo, Z., Chen, H., Zhang, L., Zhou, F., Cheng, H. and Zhou, R.:** Genome-wide detection of testis- and testicular cancer-specific alternative splicing, *Carcinogenesis*, **28**, 2484-2490 (2007).
5. **Dhir, A. and Buratti, E.:** Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies, *FEBS J.*, **277**, 841-855 (2010).
6. **Kotik, M.:** Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications, *J. Biotechnol.*, **144**, 75-82 (2009).
7. **Lorenz, P., Liebeton, K., Niehaus, F. and Eck, J.:** Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space, *Curr. Opin. Biotechnol.*, **13**, 572-527 (2002).
8. **Rajendhran, J. and Gunasekaran, P.:** Strategies for accessing soil metagenome for desired applications, *Biotechnol. Adv.*, **26**, 576-590 (2008).
9. **Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A., Falkow, S. and Brown, A. J.:** Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*, *Microbiology*, **143 (Pt 2)**, 303-311 (1997).
10. **Keppler-Ross, S., Noffz, C. and Dean, N.:** A new purple fluorescent color marker for genetic studies in *Saccharomyces cerevisiae* and *Candida albicans*, *Genetics*, **179**, 705-710 (2008).
11. **Apte-Deshpande, A., Rewanwar, S., Kotwal, P., Raiker, V. A. and Padmanabhan, S.:** Efficient expression and secretion of recombinant human growth hormone in the methylotrophic yeast *Pichia pastoris*: potential applications for other proteins,

- Biotechnol. Appl. Biochem., **54**, 197-205 (2009).
12. **Chen, M. L., Guo, Q., Wang, R. Z., Xu, J., Zhou, C. W., Ruan, H. and He, G. Q.:** Construction of the yeast whole-cell *Rhizopus oryzae* lipase biocatalyst with high activity, J. Zhejiang Univ. Sci. B, **12**, 545-551 (2011).
 13. **Favaro, L., Jooste, T., Basaglia, M., Rose, S. H., Saayman, M., Gorgens, J. F., Casella, S. and van Zyl, W. H.:** Codon-optimized glucoamylase sGAI of *Aspergillus awamori* improves starch utilization in an industrial yeast, Appl. Microbiol. Biotechnol., **95**, 957-968 (2012).
 14. **Cheng, C. Y., Wu, C. W., Lin, G. J., Lee, W. C., Chien, M. S. and Huang, C.:** Enhancing expression of the classical swine fever virus glycoprotein E2 in yeast and its application to a blocking ELISA, J. Biotechnol., **174**, 1-6 (2014).
 15. **Seo, J. Y., Chung, H. J. and Kim, T. J.:** Codon-optimized expression of fish iridovirus capsid protein in yeast and its application as an oral vaccine candidate, J. Fish Dis., **36**, 763-768 (2013).
 16. **Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A. and other 7 authors:** Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome, Science, **319**, 1215-1220 (2008).
 17. **Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M. and other 14 authors:** Creation of a bacterial cell controlled by a chemically synthesized genome, Science, **329**, 52-56 (2010).
 18. **Lartigue, C., Vashee, S., Algire, M. A., Chuang, R. Y., Benders, G. A., Ma, L., Noskov, V. N., Denisova, E. A., Gibson, D. G., Assad-Garcia, N. and other 7 authors:** Creating bacterial strains from genomes that have been cloned and engineered in yeast, Science, **325**, 1693-1696 (2009).
 19. **Benders, G. A., Noskov, V. N., Denisova, E. A., Lartigue, C., Gibson, D. G., Assad-Garcia, N., Chuang, R. Y., Carrera, W., Moodie, M., Algire, M. A. and other 8 authors:** Cloning whole bacterial genomes in yeast, Nucleic Acids Res., **38**, 2558-2569 (2010).
 20. **Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L. and other 5 authors:** Synthetic chromosome arms function in yeast and generate phenotypic diversity by

- design, *Nature*, **477**, 471-476 (2011).
21. **Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M. and other 70 authors:** Total synthesis of a functional designer eukaryotic chromosome, *Science*, **344**, 55-58 (2014).
 22. **Jayaraman, K., Fingar, S. A., Shah, J. and Fyles, J.:** Polymerase chain reaction-mediated gene synthesis: synthesis of a gene coding for isozyme c of horseradish peroxidase, *Proc. Natl. Acad. Sci. USA*, **88**, 4084-4088 (1991).
 23. **Jayaraman, K. and Puccini, C. J.:** A PCR-mediated gene synthesis strategy involving the assembly of oligonucleotides representing only one of the strands, *Biotechniques*, **12**, 392-398 (1992).
 24. **Cherry, J., Nieuwenhuijsen, B. W., Kaftan, E. J., Kennedy, J. D. and Chanda, P. K.:** A modified method for PCR-directed gene synthesis from large number of overlapping oligodeoxyribonucleotides, *J. Biochem. Biophys. Methods*, **70**, 820-822 (2008).
 25. **Au, L. C., Yang, F. Y., Yang, W. J., Lo, S. H. and Kao, C. F.:** Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli*, *Biochem. Biophys. Res. Commun.*, **248**, 200-203 (1998).
 26. **Ma, S., Saaem, I. and Tian, J.:** Error correction in gene synthesis technology, *Trends Biotechnol.*, **30**, 147-154 (2012).
 27. **Duarte, V., Gasparutto, D., Jaquinod, M. and Cadet, J.:** *In vitro* DNA synthesis opposite oxazolone and repair of this DNA damage using modified oligonucleotides, *Nucleic Acids Res.*, **28**, 1555-1563 (2000).
 28. **Carr, P. A., Park, J. S., Lee, Y. J., Yu, T., Zhang, S. and Jacobson, J. M.:** Protein-mediated error correction for *de novo* DNA synthesis, *Nucleic Acids Res.*, **32**, e162 (2004).
 29. **Kim, H., Han, H., Shin, D. and Bang, D.:** A fluorescence selection method for accurate large-gene synthesis, *Chembiochem.*, **11**, 2448-2452 (2010).
 30. **Hoshida, H., Murakami, N., Suzuki, A., Tamura, R., Asakawa, J., Abdel-Banat, B. M., Nonklang, S., Nakamura, M. and Akada, R.:** Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*, *Yeast*, **31**, 29-46 (2014).
 31. **Akada, R., Kitagawa, T., Kaneko, S., Toyonaga, D., Ito, S., Kakiyama, Y., Hoshida, H., Morimura, S., Kondo, A. and Kida, K.:** PCR-mediated seamless gene deletion and

- marker recycling in *Saccharomyces cerevisiae*, *Yeast*, **23**, 399-405 (2006).
32. **Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., Shibano, Y., Tanaka, T., Amachi, T. and Yoshizumi, H.:** *Rhizopus* raw-starch-degrading glucoamylase: Its cloning and expression in yeast, *Agric. Biol. Chem.*, **50**, 957-964 (1986).
33. **Cha-aim, K., Fukunaga, T., Hoshida, H. and Akada, R.:** Reliable fusion PCR mediated by GC-rich overlap sequences, *Gene*, **434**, 43-49 (2009).
34. **Cha-Aim, K., Hoshida, H., Fukunaga, T. and Akada, R.:** Fusion PCR via novel overlap sequences, *Methods Mol. Biol.*, **852**, 97-110 (2012).
35. **Fukunaga, T., Cha-Aim, K., Hirakawa, Y., Sakai, R., Kitagawa, T., Nakamura, M., Nonklang, S., Hoshida, H. and Akada, R.:** Designed construction of recombinant DNA at the *ura3Δ0* locus in the yeast *Saccharomyces cerevisiae*, *Yeast*, **30**, 243-253 (2013).
36. **Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., Sutthikhumphu, S., Pukahuta, C., Kitagawa, T., Nakamura, M. and other 4 authors:** Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*, *Yeast*, **30**, 485-500 (2013).
37. **Abdel-Banat, B. M., Nonklang, S., Hoshida, H. and Akada, R.:** Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*, *Yeast*, **27**, 29-39 (2010).
38. **Iborra, F. and Ball, M. M.:** *Kluyveromyces marxianus* small DNA fragments contain both autonomous replicative and centromeric elements that also function in *Kluyveromyces lactis*, *Yeast*, **10**, 1621-1629 (1994).
39. **Nonklang, S., Abdel-Banat, B. M., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M. and Akada, R.:** High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042, *Appl. Environ. Microbiol.*, **74**, 7514-7521 (2008).
40. **Nakamura, M., Suzuki, A., Hoshida, H. and Akada, R.:** Minimum GC-rich sequences for overlap extension PCR and primer annealing, *Methods Mol. Biol.*, **1116**, 165-181 (2014).
41. **Hoshida, H., Fujita, T., Cha-aim, K. and Akada, R.:** N-Glycosylation deficiency enhanced heterologous production of a *Bacillus licheniformis* thermostable alpha-amylase in *Saccharomyces cerevisiae*, *Appl. Microbiol. Biotechnol.*, **97**, 5473-5482 (2013).

- 1 42. **Hata, Y., Kitamoto, K., Gomi, K., Kumagai, C., Tamura, G. and Hara, S.:** The
2 glucoamylase cDNA from *Aspergillus oryzae*: its cloning, nucleotide sequence, and
3 expression in *Saccharomyces cerevisiae*, *Agric. Biol. Chem.*, **55**, 941-949 (1991).
- 4 43. **Fuhrmann, M., Oertel, W., Berthold, P. and Hegemann, P.:** Removal of mismatched
5 bases from synthetic genes by enzymatic mismatch cleavage, *Nucleic Acids Res.*, **33**, e58
6 (2005).
- 7 44. **Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P. and Boeke, J.**
8 **D.:** Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set
9 of strains and plasmids for PCR-mediated gene disruption and other applications, *Yeast*,
10 **14**, 115-132 (1998).

Figure Legends

Fig. 1. *yGLuc* gene synthesis through oligonucleotide assembly with fusion PCR or in-frame marker selection. (A) Oligonucleotide mixtures for assembly. The F80/C40 mixture contained forward 80-mer and complementary 40-mer oligonucleotides with gaps. The F80/C80 and F40/C40 mixtures contained forward and complementary 80-mer and 40-mer oligonucleotides, respectively, without gaps. (B) Fusion PCR and cloning at the *ura3* locus in *S. cerevisiae*. The assembled DNA was amplified and connected to the *ScLEU2* gene using fusion PCR. The fragment was used for the transformation of RAK4314 to replace *ScURA3* with the *yGLuc-ScLEU2* fragment, which was not an in-frame selection. Leu^+ and Ura^- colonies were picked and used for the GLuc activity assay. (C) In-frame marker selection in *K. marxianus*. The assembled DNA was amplified using the *yGLuc*+4 forward primer lacking the start codon and the reverse primer containing the *ScURA3*-N-terminus (+4 to +20) and lacking the stop codon of *yGLuc*. The vector fragment was prepared by PCR using the *yGLuc*+3-TDH3-1c and *ScURA3*+21 primers and the pKM149 plasmid (RAK6202) as template, which contained a start codon downstream of the *ScTDH3* promoter and non-functional *ScURA3* truncated at the N-terminus. The two fragments were mixed and used to transform *K. marxianus* RAK3908. Correct NHEJ-mediated joining at the *URA3* N-terminus produced functional *URA3*, resulting in colony formation on uracil dropout plates.

Fig. 2. Addition of a secretory signal sequence and stop codon to synthesized *yRoGLU1* and detection of amylase activity. (A) Signal sequence-deleted and *URA3*-fused *yRoGLU1* was amplified using primers that contained a signal sequence and stop codon. The *yRoGLU1* and vector fragments were mixed and used to transform RAK3908 to produce the complete *yRoGLU1* sequence. (B) The transformants were spotted on starch plates. Iodine vapor was

1 used to detect the starch degradation halo. Predicted mutation positions are indicated, with
2 the initiation methionine counted as 1. Regional mutations caused by frameshifts are
3 indicated by parentheses. W338Δ indicates deletion of W at amino acid 338. The amino acid
4 sequences are shown in Supplementary Fig. S6.

Fig. 1. Yarimizu *et al*

Fig 2. Yarimizu *et al*

TABLE 1. Yeast strains used in this study

Strain name	Species ^a	Genotype	References
BY4743	Sc	<i>MATa/a his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0</i>	(44)
RAK3614	Sc	<i>MATa ade2Δ0::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0::LEU2</i>	(35)
RAK3625	Sc	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::ScTDH3p_{AscI}AoTAA_{NotI}PGK1ter-LEU2</i>	(35)
RAK3908	Km	<i>ura3-1 ade2-1</i>	(36)
RAK4296	Sc	<i>MATa his3Δ200 leu2Δ0 met15Δ0 trpΔ63 ura3Δ0::ScGAL10p-yCLuc_{15C}URA3</i>	(35)
RAK4314	Sc	<i>MATa ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::ScTDH3p_{15C}URA3</i>	(35)
RAK4920	Sc	<i>MATa ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::ScTDH3p-yCLuc_{15C}LEU2</i>	This study
RAK4960	Sc	<i>MATa ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::ScTDH3p-yCLuc_{15C}URA3</i>	This study
RAK5125	Sc	<i>MATa ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::ScTDH3p-yGLuc_{15C}LEU2</i>	This study
RAK5986	Km	<i>ura3-1 ade2-1 pKM019 [ScADE2-KmARS7-ScURA3]</i>	This study
RAK6140	Km	<i>ura3-1 ade2-1 pKM030 [KmCenD-ScADE2-KmARS7-ScURA3]</i>	This study
RAK6202	Km	<i>ura3-1 ade2-1 pKM149 [KmCenD-ScADE2-KmARS7-ScTDH3p-yCLuc_{15C}ScURA3]</i>	This study
RAK9817	Km	<i>ura3-1 ade2-1 pKM288 [ScADE2-ScTDH3p-PhEG-ScURA3-KmCenD-KmARS7]</i>	This study
RAK7889	Km	<i>ura3-1 ade2-1 [KmCenD-ScADE2-KmARS7-ScTDH3p-yRoGLU1^{ΔN-1}=ScURA3]</i>	This study
RAK7890	Km	<i>ura3-1 ade2-1 [KmCenD-ScADE2-KmARS7-ScTDH3p-yRoGLU1^{ΔN-2}=ScURA3]</i>	This study
RAK10276	Km	<i>ura3-1 ade2-1 pKM407 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1^{T138P A139L T140P}=ScURA3]^d</i>	This study
RAK10277	Km	<i>ura3-1 ade2-1 pKM408 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1^{[S84L-D120T] L418F}=ScURA3]^{b,d}</i>	This study

(Continues)

TABLE 1. (Continued)

Strain name	Species ^a	Genotype	References
RAK10278	Km	<i>ura3-1 ade2-1</i> pKM409 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1 ^{T136S} =ScURA3] ^d	This study
RAK10279	Km	<i>ura3-1 ade2-1</i> pKM410 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1 ^{L422W I504V N506T} =ScURA3] ^d	This study
RAK10280	Km	<i>ura3-1 ade2-1</i> pKM411 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1 ^{I101N [K105R-T152P]} =ScURA3] ^{c,d}	This study
RAK10281	Km	<i>ura3-1 ade2-1</i> pKM412 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1 wild type=ScURA3] ^d	This study
RAK10282	Km	<i>ura3-1 ade2-1</i> pKM413 [KmCenD-ScADE2-KmARS7-ScTDH3p-Δ42yRoGLU1 ^{W338A} =ScURA3] ^d	This study
RAK10470	Km	<i>ura3-1 ade2-1</i> pKM429 [ScADE2-ScTDH3p-yRoGLU1 wild type-ScURA3-KmCenD-KmARS7]	This study
RAK13613	Km	<i>ura3-1 ade2-1</i> pKM935 [ScADE2-ScTDH3p- yRoGLU1 ^{T138P A139L T140P} -ScURA3-KmCenD-KmARS7]	This study
RAK13615	Km	<i>ura3-1 ade2-1</i> pKM936 [ScADE2-ScTDH3p-yRoGLU1 ^{[S84L-D120T] L418F} -ScURA3-KmCenD-KmARS7] ^b	This study
RAK13617	Km	<i>ura3-1 ade2-1</i> pKM937 [ScADE2-ScTDH3p-yRoGLU1 ^{T136S} -ScURA3-KmCenD-KmARS7]	This study
RAK13619	Km	<i>ura3-1 ade2-1</i> pKM938 [ScADE2-ScTDH3p-yRoGLU1 ^{L422W I504V N506T} -ScURA3-KmCenD-KmARS7]	This study
RAK13621	Km	<i>ura3-1 ade2-1</i> pKM939 [ScADE2-ScTDH3p- yRoGLU1 ^{I101N [K105R-T152P]} -ScURA3-KmCenD-KmARS7] ^c	This study
RAK13623	Km	<i>ura3-1 ade2-1</i> pKM940 [ScADE2-ScTDH3p-yRoGLU1 ^{W338A} -ScURA3-KmCenD-KmARS7]	This study

^aSc: *Saccharomyces cerevisiae*, Km: *Kluyveromyces marxianus*.

^bAmino acid sequence from 84 to 120: LLQSLVLTNTGPSLLLSTVSRNSTSSTKSLVRPTTT.

^cAmino acid sequence from 105 to 152:

RNSTSSTKSLVRPTTTTTLTTLTKSLPLSQPPPPLPLPPPPLHLPLPP.

^d= indicates in-frame fusion of coding sequences.

TABLE 2. Frequency of active yGLuc gene synthesis under various conditions

Oligonucleotide preparation			PCR cycle temperature (°C)		Number of GLuc positive/examined colonies		
Exp. no.	Mixture	Sup-plier ^a	DNA polymerase used	Anneal-ing (30 s)	Exten-sion (1 min)	<i>S. cerevisiae</i> fusion PCR	<i>K. marxianus</i> in-frame ^b
1	F80/C40	F	KOD -Plus-	55.0	68.0	1/36	8/24
2	F80/C80	F	KOD -Plus-	55.0	68.0	0/31	6/24
3	F40/C40	F	KOD -Plus-	55.0	68.0	0/32	9/20
4	F40/C40	F	KOD -Plus-	64.5	68.0	1/13	ND
5	F40/C40	F	KOD -Plus-	68.0	68.0	1/20	ND
6	F40/C40	F	KOD -Plus-	75.1	75.1	4/20	13/24
7	F40/C40	F	KOD -Plus-	79.0	79.0	0/20	ND
8	F40/C40	H	KOD -Plus-	55.0	68.0	4/20	14/24
9	F40/C40	H	KOD -Plus-	75.1	75.1	4/20	19/24
10	F40/C40	H	Phusion	55.0	72.0	11/20	15/24
11	F40/C40	H	Phusion	72.0	72.0	8/20	ND

^aF and H indicate Fasmac and Hokkaido System Science, respectively.

^bND: not done.

TABLE 3. DNA sequence analysis of yGLuc inactive clones

Sequenced clone	Exp. no. ^a	Mutation (A of the start codon was counted as +1) ^b			Premature termination by frameshift
		Deletion	Substitution	Insertion	
No. 1	3	A68Δ, T103Δ, C240Δ, T280Δ			+
No. 2	3	A68Δ, G88Δ, G150-A151Δ, T160-T161Δ, A223Δ, C285Δ, T396Δ			+
No. 3	6	G427Δ, T434Δ	G548A		+
No. 4	6		C308T		-
No. 5	6	C402Δ			+
No. 6	8	G222Δ			+
No. 7	8		G410T		-
No. 8	9			G373GG	+
No. 9	9	A299Δ		G529GGG	+
No. 10	9	A11Δ		A533AA, G534GG	+
No. 11	9	A368Δ			+
No. 12	9	A490Δ, G542Δ			+
No. 13	9	T170Δ			+
No. 14	9			G134GG	+
No. 15	9	C15Δ, T26Δ, G296Δ			+

^aSee details for each experiment in Table 2.

^bΔ indicates deletion of the nucleotide.

TABLE 4. Sequence analysis of the synthesized yRo*GLU1* gene

Method	Insert ^a	Strain	Mutation ^b			Premature termination	Amylase activity ^c
			Deletion	Substitution	Insertion		
Fusion	WT	-	A912Δ	T264A	A971AC	+	Inactive
			T1660–G1602Δ	C269T	C1602CC		
			T1761Δ	G811C	T1813TT		
				A824C			
				T864G			
				G871C			
				A911C			
Fusion	WT	-	G1018–C1023Δ	A209C	G385GG	+	Inactive
			G1087Δ	C210A	C714CC		
				A1132C	C809CC		
Fusion	WT	-	A133Δ	T752C		+	Inactive
			T322Δ				
Fusion	WT	-	C1211–A1226Δ	G1298C	G385GG	+	Inactive
					G550GG		
					A962AA		
					A1057AA		
In-frame	WT	RAK7889	A1–G109Δ (loss of start codon)	C452A	G1510GG	-	ND ^d
In-frame	WT	RAK7890	G3Δ (loss of start codon)	T1299G T1300G		1st ATG: Met185	ND ^d
						1st ATG: Met185	
In-frame	Δ42	RAK10276	C411Δ		C420CC	-	Active
In-frame	Δ42	RAK10277	T250Δ	G1257C	C360CC	-	Inactive
In-frame	Δ42	RAK10278		A406T		-	Active
				T1770C			
In-frame	Δ42	RAK10279		T1265G		-	Active
				A1510G			
				A1517C			

(Continues)

TABLE 4. (Continued)

Method	Insert ^a	Strain name	Mutation ^b			Premature termination	Amylase activity ^c
			Deletion	Substitution	Insertion		
In-frame	Δ42	RAK10280	A314Δ	T302A C1023A	C456CC	-	Active
In-frame	Δ42	RAK10281				-	Active
In-frame	Δ42	RAK10282	T1012–G1014Δ			-	Inactive

^aWT and Δ42 indicate the wild-type and signal sequence-deleted (+4 to +42) constructs, respectively.

^bThe nucleotide numbers indicated are based on the wild-type yRoGLUI sequence, where A of the start codon was counted as +1.

^cThe activity of the signal sequence-deleted constructs (Δ42) was examined after the addition of the N-terminal sequence.

^dND: Not done