



EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*

Holkenbrink, Carina; Dam, Marie I.; Kildegaard, Kanchana R.; Beder, Johannes; Dahlin, Jonathan; Belda, David Doménech; Borodina, Irina

Published in:
Biotechnology Journal

Link to article, DOI:
[10.1002/biot.201700543](https://doi.org/10.1002/biot.201700543)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Holkenbrink, C., Dam, M. I., Kildegaard, K. R., Beder, J., Dahlin, J., Belda, D. D., & Borodina, I. (2018). EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. *Biotechnology Journal*, 13(9), [1700543]. <https://doi.org/10.1002/biot.201700543>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Biotech Method

**EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast
Yarrowia lipolytica[†]**

Carina Holkenbrink¹

Marie Inger Dam¹

Kanchana Rueksomtawin Kildegaard¹

Johannes Beder^{1, a}

Jonathan Dahlin¹

David Doménech Belda ^{1, b}

Irina Borodina¹

¹The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

Correspondence: Dr. Irina Borodina, The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, DK-2800 Kongens Lyngby, Denmark.

E-mail: irbo@biosustain.dtu.dk

^a Present address: ALK-Abelló Nordic A/S, Bøge Allé 6-8, DK-2970 Hørsholm, Denmark

^b Present address: Inbiose NV, Incubatie-en Innovatiecentrum UGent, Technologiepark 3, bus 41, B-9052, Zwijnaarde, Belgium

[†]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: [10.1002/biot.201700543].

This article is protected by copyright. All rights reserved
Received: September 5, 2017 / Revised: January 11, 2018 / Accepted: January 26, 2018

Abstract

The oleaginous yeast *Yarrowia lipolytica* is an emerging host for production of fatty acid-derived chemicals. To enable rapid iterative metabolic engineering of this yeast, there is a need for well-characterized genetic parts and convenient and reliable methods for their incorporation into yeast. Here, we present the EasyCloneYALI genetic toolbox, which allows streamlined strain construction with high genome editing efficiencies in *Y. lipolytica* via the CRISPR/Cas9 technology. The toolbox allows marker-free integration of gene expression vectors into characterized genome sites as well as marker-free deletion of genes with the help of CRISPR/Cas9. Genome editing efficiencies above 80% were achieved with transformation protocols using non-replicating DNA repair fragments (such as DNA oligos). Furthermore, the toolbox includes a set of integrative gene expression vectors with prototrophic markers conferring resistance to hygromycin and nourseothricin.

Keywords: CRISPR/Cas9, integrative vectors, metabolic engineering, *Yarrowia lipolytica*, yeast

Abbreviations: **CRISPR**, Clustered Regularly Interspaced Short Palindromic Repeats; **Cas9**, CRISPR-associated nuclease; **NHEJ**, non-homologous end joining; **HR**, homologous recombination.

1 Introduction

The oleaginous yeast *Yarrowia lipolytica* is an attractive host for industrial production of fatty-acid derived chemicals, organic acids, and enzymes [1]-[3]. *Y. lipolytica* is commercially used for production of omega-3 eicosapentaenoic acid [4], and a number of other processes are emerging, such as production of lipids from unsterile feedstocks [5] and bioremediation of oil-contaminated environments [6].

Development of a strain into an efficient industrial cell factory requires multiple rounds of metabolic engineering. The engineering efforts so far have been hampered by limited genome targeting efficiencies in *Y. lipolytica* and by the requirement for selection markers, which need to be recycled. The low genome targeting efficiencies are due to a high rate of non-homologous end joining (NHEJ) for repair of DNA double-strand (ds) breaks in contrast to *Saccharomyces cerevisiae*, where the homologous recombination (HR) mechanism is the dominating repair pathway [7][8].

To overcome the above-mentioned limitations of marker-based genome editing, the CRISPR/Cas9 system has been successfully implemented in several yeast species [9]. The most widely applied CRISPR/Cas9 system for genome editing combines three parts: (1) a sgRNA [10], composed of a site-specific crRNA fused to a tracrRNA, which binds Cas9 endonuclease, (2) the Cas9 endonuclease, capable of creating dsDNA breaks, and (3) a dsDNA repair template, which is used by the HR pathway to repair the dsDNA break.

Three previous studies have successfully applied the CRISPR/Cas9 technology in *Y. lipolytica* for single knockouts of genes and insertion of expression constructs, reaching 90-100% correct genome edit efficiencies after 2-4 days recovery phase after transformation [11]-[13]. Due to the long recovery, during which the cells divide through multiple generations, it is imperative to use replicating repair templates, e.g., with the repair elements cloned into episomal vectors. We have developed a CRISPR/Cas9 method,

which achieves high genome editing efficiency using non-replicating DNA repair templates. For gene deletions/mutations, we use synthetic double-stranded oligos as repair templates, while for gene insertions, we use linearized non-replicating vectors.

The genome engineering toolbox EasyCloneYALI, presented in this study, comprises a set of integrative expression vectors which allow expression of one or two genes per vector and integration into highly expressed intergenic genome sites. Five vectors can be integrated with the help of the CRISPR/Cas9 technology, reaching efficiencies above 80%. 11 vectors can be integrated with the help of dominant selection markers, with 30-100% efficiency. Further, we provide vectors and protocols for knockout of up to two genes simultaneously using the CRISPR/Cas9 system. Using 90bp double-stranded oligos as DNA repair templates, we obtained efficiencies of 90% for individual knockouts and 6-66 % for double knockouts. To ease and accelerate plasmid construction, all plasmids in this toolbox are standardized and allow recycling of biobricks. The vectors are assembled by USER® cloning. The vectors can be obtained via AddGene. We provide detailed protocols in the supplementary material.

2 Materials and methods

A detailed step-by-step protocol for the EasyCloneYALI toolbox is provided as supplementary material. All the vectors can be obtained individually or as a set from AddGene.

2.1 Strains and media

Yarrowia lipolytica GB20 ^[14] was a kind gift of Volker Zickermann. The genotype of the GB20 is *mus51Δ (=ku70)*, *nugm-Htg2*, *ndh2i*, *lys11⁻*, *leu2⁻*, *ura3⁻*, *MatB*. All the strains in this study were derived from *Y. lipolytica* GB20, they are listed in Table S1. *Y. lipolytica*

strains were grown at 30°C and 250 rpm in standard yeast peptone dextrose (YPD, Sigma-Aldrich), synthetic complete (SC, Sigma-Aldrich) or mineral medium [15]. For growth on solid media, YPD and SC medium were supplemented with 20 g/L agar. When necessary for selection, the YPD medium was supplemented with 250 mg/L nourseothricin or 50-100 mg/L hygromycin B. To screen for successful *thi6* knockout strains, the cells were spotted on mineral medium agar without addition of thiamine. *Escherichia coli* strain DH5 α was used for cloning and plasmid propagation. *E.coli* was grown in Lysogeny Broth (LB) supplemented with 100 mg/L ampicillin and when needed with 15 g/L agar.

2.2 Design of crRNAs and assembly of gRNA vectors

Optimal target-specific crRNA sequences (20 bp) were identified with the help of the online tool “CHOPCHOP” (<http://chopchop.cbu.uib.no/>) [16]. To predict suitable crRNAs, the genomic target region/gene was entered, and the *Y. lipolytica* CLIB122 genome sequence and standard settings were chosen. One high-ranking crRNA sequence was chosen for plasmid construction. The crRNA-encoding DNA was ordered as two complementary primers encoding overhangs compatible to the promoter and terminator biobrick. The sense primer comprised the 20 bp crRNA sequence (as predicted above) and was extended at its 3'end with linker L1 (5'-gttttagagct-3'), the antisense primer comprised the 20 bp reverse complement crRNA sequence extended with linker L2 (5'-taaccaacct-3') at its 3'end (Figure S2). gRNA vectors harboring a single gRNA expression cassette were assembled by USER® cloning (NEB) and directly transformed into *E.coli*. To do so the following DNA fragments were included into a USER® reaction: The sense and antisense primers (as described above), the tRNA (YALI0A04565r) promoter biobrick (BB1635), the crRNA-terminator biobrick (BB1636), and the vector backbone pCfB3405. The USER® reaction was performed as described in “2.3 Plasmid construction”. For assembly of vectors composed of multiple gRNA expression cassettes, the gRNA cassettes

were pre-assembled by PCR as follows: 1 μ L of BB1635, 1 μ L of BB1636 and 1 μ L 10 μ mol of each the sense and antisense primer were mixed, treated with USER[®] enzyme and ligated with T4 ligase (Thermo Fisher Scientific) according to manufacture's instructions. 2 μ L of the ligation reaction was then used as a template for PCR using a primer pair as described in the detailed manual. The PCR products were gel purified. Multiple cassettes were assembled by USER reaction into plasmid pCfB3405.

2.3 Plasmid construction

All biobricks used in this study are listed in Table S3. Biobricks were amplified by PCR with Phusion U (Thermo Fisher Scientific) under the following conditions: 98°C for 5 min, 30 cycles of 98°C for 10 s, 52°C for 10 s, 72°C for 30 s/1 kb, 72°C for 10 min., and gel purified. The templates and primers used for PCR amplification of each biobrick are listed in Table S3.

All vectors were assembled by USER[®] cloning according to Table S2. A typical USER[®] reaction was composed of 0.5 μ L CutSmart buffer (NEB), 0.5 μ L USER[®] enzyme and 0.8 to 1 μ L of each biobrick/linearized parent vector/oligonucleotide (10 μ mol). If necessary, water was added to adjust the volume to a total of 5 μ L. The reaction was incubated at 37°C for 25 min and 25°C for 10 min followed by transformation into *E.coli*. Correct assembly was verified by DNA sequencing. The parent vectors for USER[®] reaction were prepared by digestion with restriction endonuclease FastDigest AsiSI (Life Technologies) and sticky ends obtained by nicking with Nb.BsmI (New England Biolabs). The vectors were then gel-purified.

To construct plasmids pCfB4132 and pCfB4589, biobricks BB1243 and BB1388, respectively, were circularized by self-ligation with T4 ligase.

Biobrick BB1135 was treated with the nicking enzyme Nb.BsmI after PCR amplification and PCR purified prior to USER[®] cloning.

2.4 Strain construction

DNA was transformed into the parent strain using a lithium-acetate-based protocol [17]. In detail, the parent strain was incubated for 24 hours at 30°C on solid medium prior to transformation. Integration plasmids were linearized with the endonuclease NotI (Thermo Fisher Scientific) and gel purified. 5×10^7 cells of the parent strain were resuspended in 1 mL sterile water in a 1.5-mL Eppendorf tube. The cells were collected by centrifugation at $3000 \times g$ for 5 min and the water decanted. This washing step was repeated once more. DNA was added to the cell pellet, and mixed by carefully flicking the tube. Then 100 μ L transformation mix was added to the tube and mixed by flicking the tube. For integrating vectors with selection marker cassettes, 500 ng DNA was used per transformation. For CRISPR/Cas9 mediated transformations, 500-1000 ng of gRNA plasmid and either 2 nmole of 90 bp-double stranded oligonucleotide (if multiple 2 nmole of each) or 500 ng of linearized marker-free integration construct was added. The transformation mix was composed of 87.5 μ L sterile PEG (50% v/v), 5 μ L lithium acetate (2 M), 2.5 μ L boiled single stranded DNA from salmon testes (10 g/L; Sigma-Aldrich) and 5 μ L of sterile dithiothreitol (2 M). The cell suspension was incubated for 60 min at 39°C and afterwards pelleted by centrifugation for 5 min at $3000 \times g$ at room temperature. The supernatant was discarded and the cells were carefully resuspended in 500 μ L YPD medium and incubated for two hours at 30°C with 250 rpm shaking for recovery. The cells were pelleted by centrifugation for 5 min at $3000 \times g$ at room temperature, resuspended in 100 μ L sterile water and plated on selective plates. The plates were incubated at 30°C until colonies appeared, usually two days after transformation. The correct integration of DNA constructs into the *Y. lipolytica* genome was confirmed by colony PCR.

2.5 Design of double-stranded 90mer repair template for gene knockouts via CRISPR/Cas9

To obtain gene knockouts via the CRISPR/Cas9 technology, cells were transformed with a 90 bp double-stranded DNA repair template. The repair template encoded 45 bp up- and 45 bp downstream of the Cas9 cleavage site or of the ORF, and was ordered from Integrated DNA Technologies. The 90mers were designed either to introduce a stop codon and a frame-shift mutation (replacing the PAM motif (NGG)) or for a complete removal of the ORF.

2.6 Phenotyping of Δgfp , Δhph and $\Delta thi6$

Strains were tested for the Δgfp phenotype by resuspending cells in Milli-Q water and measuring biomass and fluorescence in a microplate reader (Biotek). Biomass was measured as absorbance at 600 nm, and fluorescence with excitation at 485 nm and emission at 528 nm. Strains were tested for the Δhph and $\Delta thi6$ phenotypes by resuspending cells in Milli-Q water and plating on solid YPD medium with hygromycin and on mineral medium without thiamine, respectively (Figure S4).

2.7 Fluorescence measurements with FACS

Y. lipolytica strains expressing humanized recombinant green fluorescent protein (*hrgfp*) under the control of different promoters were grown overnight in 5 mL SC medium in 13-ml tubes at 30°C and 250 rpm agitation. An adequate volume was then used to inoculate 500 μ L of mineral medium in a 96 deep-well microtiter plate with air-penetrable lid (EnzyScreen) to an initial OD₆₀₀ of 0.1. Cells were grown at 30°C and 300 rpm at 5 cm orbit cast for 24-72 hours.

For FACS analysis, 30 μ L culture was added to 150 μ L phosphate buffered saline (PBS) buffer. Cells were then analyzed by flow cytometry using a BD Biosciences Fortessa flow

867 cytometer (Becton Dickinson) with a blue laser (488 nm). For each strain, 10,000 single-cell events were recorded. Flow cytometry data sets were analyzed and interpreted by FlowJo software (Tree Star Inc.). Outliers were removed from the forward-scattered light (FSC) and side-scattered light (SSC) data sets plot, with the rule for outliers set at 90% quantile region. Cells were analyzed for their mean values. A two-tailed *T*-tests (*p*-value <0.05) prove significant difference in fluorescence between Int_B and the remaining 11 integration sites.

2.8 Determination of growth rates

A pre-culture of each *Y. lipolytica* strain was grown in 5 ml SC medium in 13-ml tubes overnight at 30°C and 250 rpm. The growth rates were determined in 96-well microtiter well plates (Greiner Bio-One). Each well was filled with 150 µl mineral medium and inoculated with 2.5 µL pre-culture. The plates were sealed with Breathe-Easy sealing membrane (Z380059; Sigma-Aldrich) to allow online optical density measurements. The plate was cultivated in a microtiter plate reader (ELx808, BioTek) at 30°C for 42 hours with agitation. The optical density was measured at 600 nm every 30 min.

3 Results and Discussion

3.1 Construction of the vectors for the EasyCloneYALI toolbox

The EasyCloneYALI toolbox enables three key genome-editing operations (Figure 1). A) The integration of one or two gene expression cassettes into a defined genomic locus using auxotrophic/resistance markers for selection (Figure 1A). B) The marker-free integration of one or two gene expression cassettes using the CRISPR/Cas9 technology (Figure 1B). C) The marker-free knockout/mutation of genes using the CRISPR/Cas9 technology (Figure 1C). To facilitate the integration of gene expression cassettes, we chose to construct integrative expression vectors that target specific intergenic sites. Initially, we choose 11

target regions according to the following criteria. The regions of ca. 5,000 bp should be unique, not containing ORFs or other non-coding RNA elements, and the adjacent five ORFs must have high expression levels in both exponential growth phase and under nitrogen limitation [18]. The exact genome location of the integration sites are listed in Table S5. The integration sites were named after their chromosome location, e.g. IntA_1. We constructed 11 integration vectors containing an EasyClone cloning site for standardized cloning [15], flanked by the *pex20* and *lip2* terminator sequences (Figure 2A). The two terminator sequences are in turn each flanked by a 500 bp region homologous to the selected integration site, named up- and downstream region. The set of vectors described in Figure 1A additionally contain a selection marker cassette between the *lip2* terminator and the downstream homologous region. The marker cassettes are flanked by loxP sites, allowing marker removal by Cre recombinase [19]. Integration vectors were constructed with auxotrophic *URA3* marker, as well as *nat* and *hph* markers that confer resistance to nourseothricin and hygromycin, respectively. Vector maps for all the plasmids are available on AddGene. The whole integration cassette is flanked by NotI recognition sites allowing excision for yeast transformation. The genome editing via the CRISPR/Cas9 technology required DNA repair templates, either in form of marker-free integration vectors or double-stranded oligonucleotides, respectively (Figure 1B and C). The design of the marker-free integration vectors is similar to the integration vectors with markers but devoid of selection markers (Figure 2A). Further, both applications require the expression of a gRNA from an episomal vector (Figure S1). The episomal vector pCfB3405 was inspired by the P-POT1 plasmid [20]. The vector is composed of the Pot1 promoter fused to the CEN1 sequence for replication in *Y. lipolytica* [21], adjacent to the EasyClone cloning site [15]. Further, the plasmid contains a *nat* resistance marker for selection on the antibiotic nourseothricin, as well as an ampicillin resistance cassette and puC ori for propagation in *E.coli*. Next, the gRNA expression cassette was cloned into the

linearized vector pCfB3405. The gRNA expression cassette is composed of four elements: the tRNA-Gly (YALI0A04565r), functioning as RNA III polymerase promoter (encoded by BB1635), the crRNA sequence (target specific 20 bp), followed by the tracrRNA sequence fused to the RPR1 terminator sequence (encoded by BB1636) (Figure S1A) [22]. As the crRNA had to be designed specifically for each knockout target, the crRNA was introduced as two complementary oligonucleotides compatible with BB1635 and BB1636 (Figure S2). The gRNA vector can be designed to target one or more genomic sites. Another toolbox component, essential for CRISPR/Cas9-mediated genome editing, is the integrative Cas9 expression vector (Figure 1B and C). In pCfB4906, a *Y. lipolytica* codon-optimized Cas9 gene from *Streptococcus pyogenes* is under control of the Tef promoter and terminator and has been integrated into the IntB integration site [23] with the help of a hygromycin resistance marker. The toolbox also contains a Cas9 integration vector, pCfB6364, which integrates into the *ku70* locus using a *dsdA* marker cassette, allowing growth on D-serine [24]. This plasmid has been used for genome engineering of *Y. lipolytica* strains resistant to hygromycin (unpublished data). We integrated the Cas9 gene as we were unable to maintain to episomal plasmids in one cell. A detailed description of the molecular cloning workflow can be found in the supplementary handbook.

3.2 Evaluation of genome integration sites

To validate how the integration of a gene expression cassette into the genome integration sites affects gene expression levels and cell growth, *hrgfp* was cloned into each of the 11 integration vectors under the control of the GPD promoter and transformed into *Y. lipolytica* ST3683. GFP fluorescence was used to evaluate *hrgfp* expression levels (Figure 2B). All 11 strains (ST5143-ST5149, ST5235-5238) displayed higher fluorescence levels, with 44%-139% increase, compared to the control strain (ST4296) which express *hrgfp* from a previously described integration site IntB [23]. The majority replicate clones differed

not more than 11 % in emitted fluorescence from each other, which indicates that no multiple integrations have occurred. Only two out of 28 clones showed a 30 and 50% increase in fluorescence compared to their respective replicate clones. The growth rate of the strains was not affected by the integration of the gene expression cassette (Table S6).

3.3 Compatibility of *Y. lipolytica* promoters with the standardized EasyCloneYALI system

To ease the usage of the EasyCloneYALI toolbox and to decrease the lab workload, promoter and gene biobricks contain standardized overhangs. 12 frequently used *Y. lipolytica* promoters were adapted to the EasyClone system and cloned to control *hrgfp* expression. The vectors were integrated into the *Y. lipolytica* genome and GFP fluorescence was used to evaluate *hrgfp* expression levels (Figure 2C). We detected promoter activities for the *fba1*, *icl1*, *yat1*, *ilv5*, *tef1*, *exp*, *gpd*, and *tef1in* promoters. The DNA sequences of the promoters can be found in supplementary data. No fluorescence could be detected for the strains expressing *hrgfp* from the *gpat*, *dga1*, *gpm1* or *fba1in* promoters. The *tef1in* promoter (*tef* promoter combined with its intron) has previously shown to provide a 17-fold increase in expression compared to the *tef* promoter alone [25]. However, when we introduced the EasyClone linker between the *tef* promoter's intron and *hrgfp*, fluorescence level was similar to the TEF promoter without intron. To test whether the EasyClone linker influenced the promoter activity, we fused *hrgfp* directly to the *tef1in* promoter, omitting the start codon of *hrgfp* as described previously [25]. This strain showed fluorescence levels 7.4-fold higher than the strain expressing *hrgfp* from the EasyClone-adapted *tef1in* promoter. To test whether the linker also influenced promoter strength of a non-intronic promoter, we constructed a strain expressing *hrgfp* from the *gpd* promoter without EasyClone linker. In this case the EasyClone-adapted *gpd* promoter led to 24% higher fluorescence levels than the GPD promoter without the EasyClone linker. As

standardization of biobricks facilitates the workflows, we decided to proceed with the positively evaluated EasyClone-standardized promoters.

3.4 Selection marker-mediated integration of gene expression cassettes

The *Yarrowia lipolytica* strains used in this study lack Ku70p, which is involved in DNA double-strand repair by non-homologous end-joining (NHEJ) [7][8]. Deletion of *ku70* gene in *Y. lipolytica* has been shown to increase the rate of DNA double-strand breaks repaired by homologous recombination[7][8]. The strain is slightly higher sensitive to UV light⁷. We constructed a set of integration vectors harboring auxotrophic or antibiotic selection markers (Fig. 1A). Then we determined the integration efficiencies of the vectors containing a nourseothricin selection marker targeting different integration sites (Fig. 3A). The vectors integrated with more than 50% efficiency, with exception of vector targeting site IntE1, for which the efficiency was ca. 30%.

3.5 Validation of genome editing via CRISPR/Cas9 technology

Vectors constructed for marker-free integration into the genome were evaluated by transforming the linearized integration vectors together with their corresponding gRNA vector into strain ST5010 ($\Delta ku70$ *hrgfp hph* Cas9) (Figure 1B and Table S7). Strain ST5010 grows at a maximum specific growth rate of 0.21 ± 0.04 h⁻¹ similar to its parent strain (0.19 ± 0.06 h⁻¹) and did not seem to be affected by constitutive Cas9 expression. Also, the Cas9 expression vector pCfB4906 integrated with an efficiency of 67% similar to the empty IntB integration vector. For marker free integration, we tested two gRNA vectors for each integration site. Three days post transformation, colonies were screened for correct integration of the EasyCloneYALI vectors by colony PCR. The integration efficiencies ranged from 0 to 95% (Table S7). For five out of the 11 sites, IntC_2, IntC_3, IntD_1, IntE_1 and IntE_3, integration efficiencies of above 80% could be obtained, and

30% for IntE_4 (Figure 3B). For the remaining sites, only 0-12% of the colonies had the vector correctly integrated. Interestingly, the two gRNA sequences tested for integration site IntE_3 lead to 95% and 12% efficiency, respectively, although they target the same genomic region overlapping in 15 out of 20 nucleotides (Figure S3). We are currently investigating this difference in more detail. The simultaneous integration of two integration vectors was not successful.

We evaluated the EasyCloneYALI toolbox for knockout of genes, by targeting *hrgfp*, *hph* and *thi6* allowing an easy phenotypic screening. Knockout of *hrgfp* leads to loss of fluorescence, knockout of *hph* loss of resistance to hygromycin, and without *thi6* the cells are unable to grow without thiamine supplementation in the medium. The repair templates were designed to introduce a premature stop-codon and a frame-shift into *hrgfp* and *hph*, and to delete the whole ORF of *thi6* (Figure 1C). We validated the toolbox for single, double and triple knockout of genes (Figure 3C). For all three single genes knockouts, efficiencies of around 90% could be achieved, while the knockout-efficiencies for two genes simultaneously varied depending on the gene combination – between 6% and 66% (Figure 3C). The knockout of three genes simultaneously was not successful. To confirm that the dsDNA break generated by CRISPR/Cas9 was repaired via the HR pathway, colony PCR was performed on all confirmed Δ *thi6* colonies of a transformation targeting *gfp* and *thi6* simultaneously. 100% of the 14 Δ *thi6* colonies had repaired the dsDNA break by using the supplied double-stranded oligonucleotide.

Following the EasyCloneYALI toolbox manual, a new strain can be obtained within 6 days, comparable to the protocol described by [13]. The transformation is performed on day one (four hours required), followed by 2 to 3 days incubation on selective plates. On day three the colonies are genotyped by colony PCR and correct clones are inoculated in non-selective media for gRNA vector removal overnight. A diluted suspension of the overnight culture is spread on a non-selective plate overnight and the strain is ready for a new round

of transformation on day 6. Optionally the strain can be tested for sensitivity towards nourseothricin to confirm gRNA vector loss.

4 Discussion

We demonstrated that the EasyCloneYALI toolbox can be used for reliable and fast strain engineering of *Y. lipolytica*. We identified 11 intergenic sites with high gene expression levels and validated that integration of gene expression cassettes into these sites did not affect the growth. For marker-mediated integration of the expression vectors, we determined integration efficiencies of 33 to 100 percent. Similar efficiencies, 56 and 85% have been previously reported for $\Delta ku70$ strains [7][8] or 73% for *ku70*-repressed strains [26]. The CRISPR/Cas9-mediated integration of gene expression constructs into sites IntC_2, IntC_3, IntD_1, IntE_1, and IntE_3 occurred with efficiencies above 80%. A previous study reported efficiencies of 47 to 69% [13]. For single genome edits, efficiencies above 80% could be achieved, compared to 58% and 100% knockout efficiency reported for a $\Delta ku70$ strain by Gao et., 2016 and Schwartz et al, 2016 (Figure 3) [11][12]. When deleting different combinations two genes variable efficiencies of 5.9 to 66% were obtained while a triple knockout was not successful (Fig. 3C). Gao et al. achieved efficiencies of 36 % an 19% for a double and triple gene knockout, respectively, using a plasmid-based CRISPR/Cas9 system in a *ku70+* strain and four days of recovery time [11]. Comparisons of these knockout efficiencies should be taken with caution as the studies target different genes and experimental protocols. In the EasyCloneYALI protocols, linear DNA fragments, as commercial double-stranded oligonucleotides or PCR products, can be used as repair templates instead of episomal vectors, and time-consuming cloning of plasmids can be avoided. Additionally, the toolbox allows the knockout of two genes or integration of two gene expression cassettes simultaneously (Figure 2 and 3). The EasyCloneYALI integration vectors have been used for multiple rounds of genome engineering and no loss of already

integrated cassettes could be detected^[27] By standardizing molecular cloning, promoter and gene biobricks as well as linearized parent plasmids can be reused and easily exchanged between lab members.

Acknowledgement

The work was funded by the Novo Nordisk Foundation (Grant agreement NNF15OC0016592). We would like to thank Indrė Jaglinskaitė and Trine Ann-Mirll Pedersen for excellent technical assistance and Vratislav Stovicek for discussions on the manuscript.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

5 References

- [1] J.L. Adrio, *Biotechnol. Bioeng.* **2017**, *114*, 1915.
- [2] R. Ledesma-Amaro, J.-M. Nicaud, *Prog. Lipid Res.* **2016**, *61*, 40.
- [3] H.-H. Liu, X.-J. Ji, H. Huang, *Biotechnol. Adv.* **2015**, *33*, 1522.
- [4] Z. Xue, P.L. Sharpe, S.-P. Hong, N.S. Yadav, D. Xie, D.R. Short, H.G. Damude, R.A. Rupert, J.E. Seip, J. Wang, D.W. Pollak, M.W. Bostick, M.D. Bosak, D.J. Macool, D.H. Hollerbach, H. Zhang, D.M. Arcilla, S.A. Bledsoe, K. Croker, E.F. McCord, B.D. Tyreus, E.N. Jackson, Q. Zhu, *Nat. Biotechnol.* **2013**, *31*, 734.
- [5] A.J. Shaw, F.H. Lam, M. Hamilton, A. Consiglio, K. MacEwen, E.E. Brevnova, E. Greenhagen, W.G. LaTouf, C.R. South, H. van Dijken, G. Stephanopoulos, *Science* **2016**, *353*, 583.
- [6] S. Zinjarde, M. Apte, P. Mohite, A.R. Kumar, *Biotechnol. Adv.* **2014**, *32*, 920.
- [7] A. Kretzschmar, C. Otto, M. Holz, S. Werner, L. Hübner, G. Barth, *Curr. Genet.* **2013**, *59*, 63.
- [8] J. Verbeke, A. Beopoulos, J.-M. Nicaud, *Biotechnol. Lett.* **2013**, *35*, 571.
- [9] V. Stovicek, C. Holkenbrink, I. Borodina, *FEMS Yeast Res.* **2017**, *17*.
- [10] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816.
- [11] S. Gao, Y. Tong, Z. Wen, L. Zhu, M. Ge, D. Chen, Y. Jiang, S. Yang, *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 1085.
- [12] C.M. Schwartz, M.S. Hussain, M. Blenner, I. Wheeldon, *ACS Synth. Biol.* **2016**, *5*, 356.
- [13] C. Schwartz, M. Shabbir-Hussain, K. Frogue, M. Blenner, I. Wheeldon, *ACS Synth. Biol.* **2016**.
- [14] H. Angerer, M. Radermacher, M. Mańkowska, M. Steger, K. Zwicker, H. Heide, I. Wittig, U. Brandt, V. Zickermann, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5207.
- [15] N.B. Jensen, T. Strucko, K.R. Kildegaard, F. David, J. Maury, U.H. Mortensen, J. Forster, J. Nielsen, I. Borodina, *FEMS Yeast Res.* **2014**, *14*, 238.
- [16] K. Labun, T.G. Montague, J.A. Gagnon, S.B. Thyme, E. Valen, *Nucleic Acids Res.* **2016**, *44*, W272.
- [17] D.C. Chen, *Appl. Microbiol. Biotechnol.* **1997**, *48*, 232.
- [18] N. Morin, J. Cescut, A. Beopoulos, G. Lelandais, V. Le Berre, J.-L. Uribelarrea, C. Molina-Jouve, J.-M. Nicaud, *PloS One* **2011**, *6*, e27966.
- [19] P. Fickers, M. Le Dall, C. Gaillardin, P. Thonart, J. Nicaud, *J. Microbiol. Methods* **2003**, *55*, 727.
- [20] L. Liu, P. Otoupal, A. Pan, H.S. Alper, *FEMS Yeast Res.* **2014**, *14*, 1124.
- [21] T. Yamane, H. Sakai, K. Nagahama, T. Ogawa, M. Matsuoka, *J. Biosci. Bioeng.* **2008**, *105*, 571.
- [22] C. Neuvéglise, C. Gaillardin, C. Marck, in *Yarrowia Lipolytica*, Springer, Berlin, Heidelberg **2013**, 79.
- [23] F. Matthäus, M. Ketelhot, M. Gatter, G. Barth, *Appl. Environ. Microbiol.* **2014**, *80*, 1660.
- [24] M.K. Vorachek-Warren, J.H. McCusker, *Yeast* **2004**, *21*, 163.
- [25] M. Tai, G. Stephanopoulos, *Metab. Eng.* **2013**, *15*, 1.
- [26] C. Schwartz, K. Frogue, A. Ramesh, J. Misa, I. Wheeldon, *Biotechnol. Bioeng.* **2017**, *114*, 2896.

- [27] K.R. Kildegaard, B. Adiego-Pérez, D. Doménech Belda, J.K. Khangura, C. Holkenbrink, I. Borodina, *Synth. Syst. Biotechnol.* **2017**.

Figure 1. EasyCloneYALI toolbox applications

A. The EasyCloneYALI toolbox contains a set of 27 standardized vectors for integration of gene expression cassettes into defined genome loci (named Int(chromosome)_#). **B.** A marker-free integration of gene expression cassettes into defined genomic loci is mediated by CRISPR/Cas9. **C.** Marker-free gene deletions with CRISPR/Cas9. A Cas9-expressing strain is transformed with a target-specific gRNA vector and a DNA repair template. The DNA repair template can be designed to introduce point mutations (frame-shift, premature stop-codon, amino acid changes) or to delete the entire ORF.

Figure 2. Comparison of integration sites and promoters

A. Vector map of the EasyCloneYALI integration vectors for integration of gene expression cassettes into defined genome loci. G, gene; Pr, promoter; T, terminator; up/down, regions homologous to *Y. lipolytica* genome **B.** Mean fluorescence of strains expressing *hrgfp* from EasyCloneYALI genome integration sites and the IntB integration site. Strain ST3683, not expressing *hrgfp*, served as negative control. **C.** Mean fluorescence of strains expressing *hrgfp* from the IntB sites controlled by a set of *Yarrowia lipolytica* promoters. Strain ST3683, not expressing *hrgfp*, served as negative control. in, promoter combined with native intron.

Figure 3. Genome editing efficiencies for integration of EasyCloneYALI marker-free vectors and for gene knockouts mediated by CRISPR/Cas9

A. Integration efficiencies of linearized EasyCloneYALI marker-free integration vectors when transformed into strain ST5010 together with their corresponding gRNA vectors (see also Figure 1B). Each bar represents an independent experiment. For each experiment 20 colonies were tested. Vectors with efficiencies above 30% are displayed. **B.** Knockout efficiencies of the target genes *hrgfp*, *hph*, and *thi6* using the CRISPR/Cas9 system described in Figure 1C. The genes were knocked out in strain ST5010 by transforming the indicated target-specific gRNA plasmid and a dsoligo(s) as repair template. 43 colonies were tested for each transformation. The bars represent the mean of three biological replicates. The error bars represent the standard deviation.

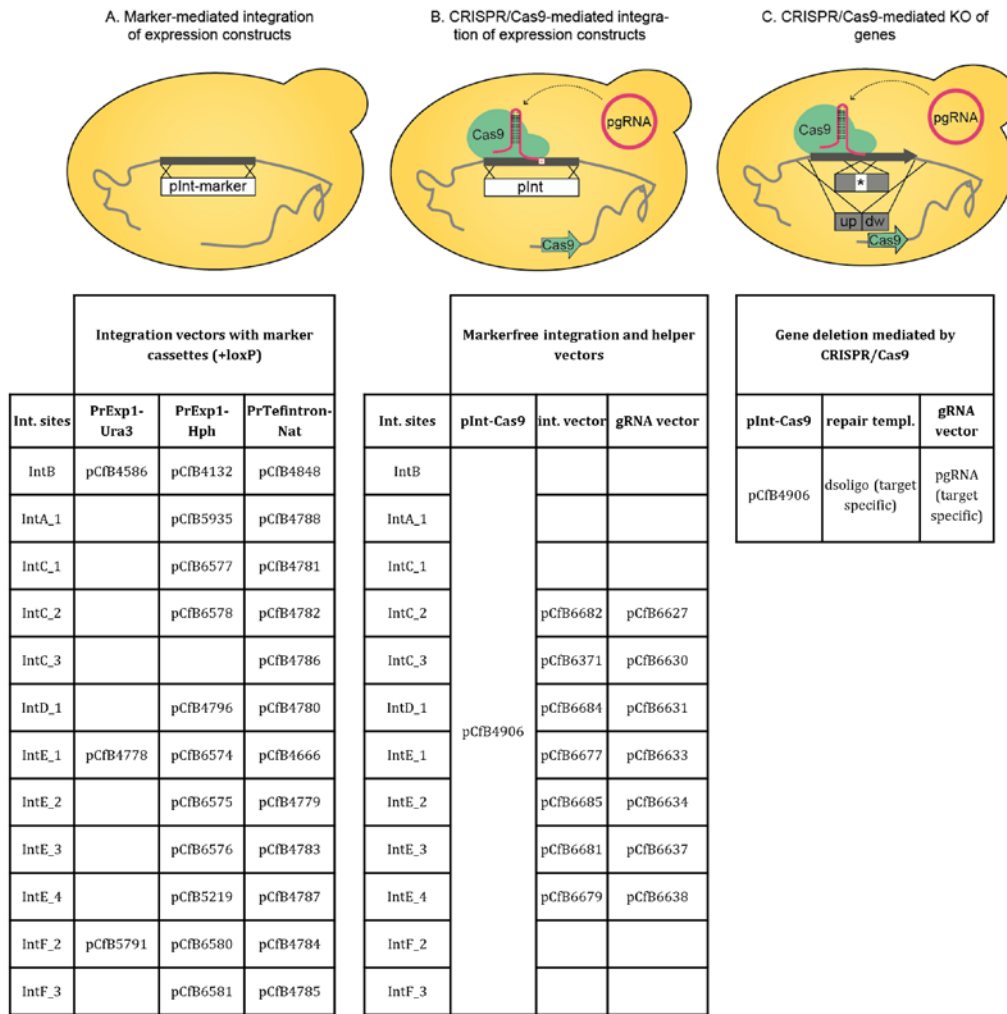


Figure 1

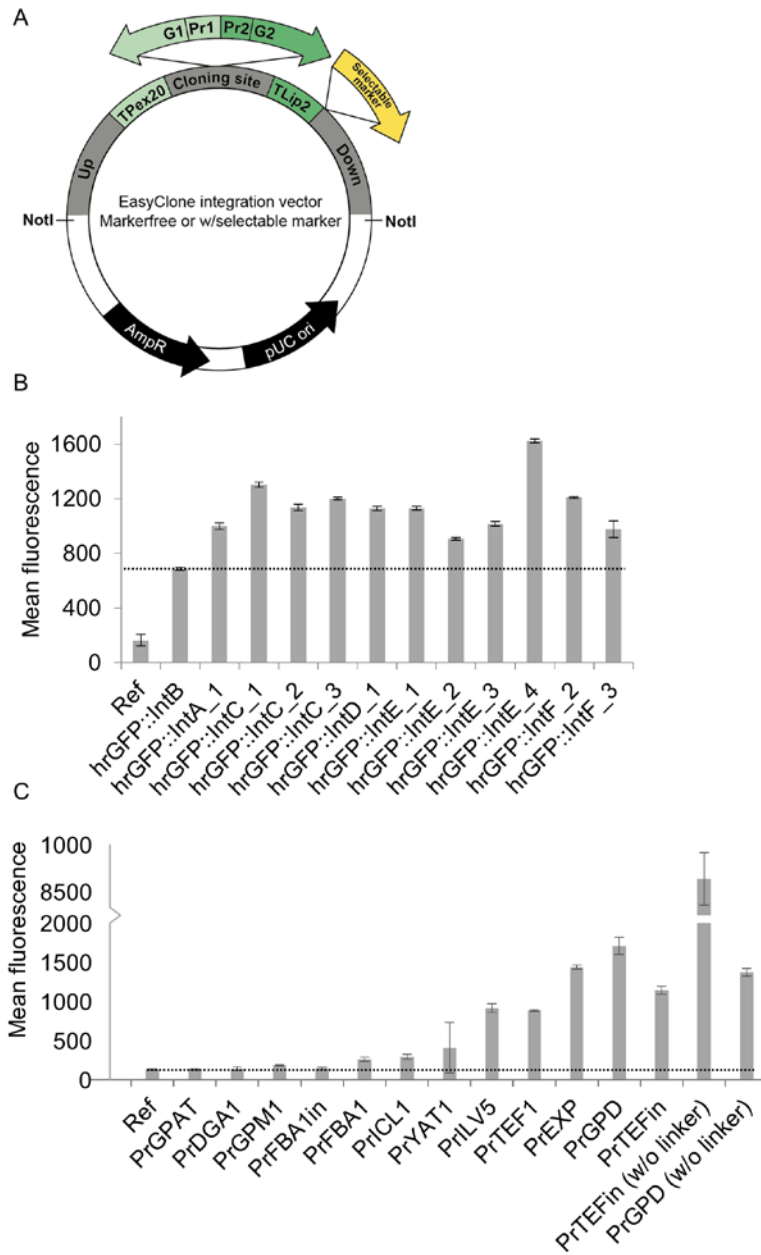


Figure 2

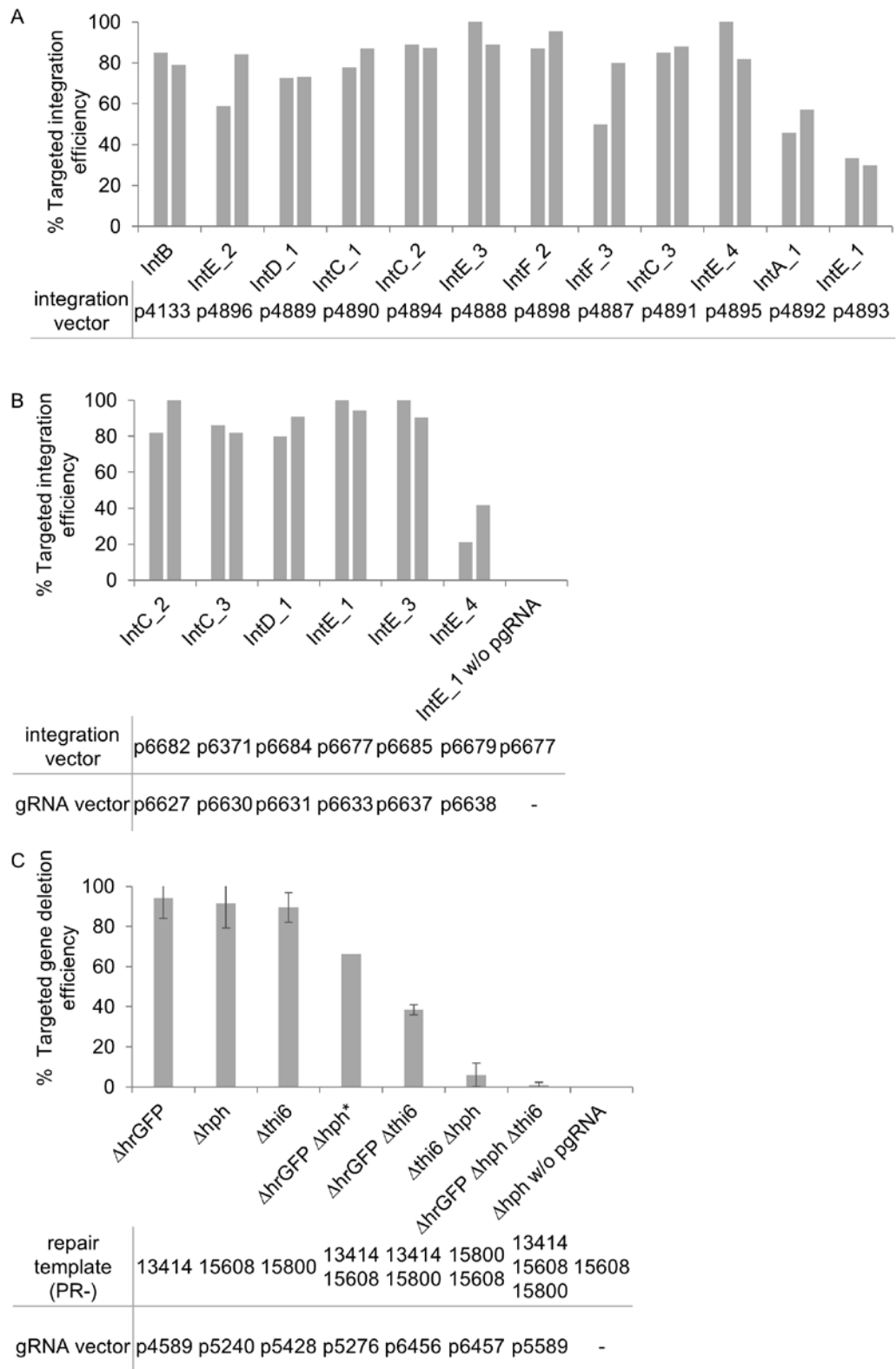


Figure 3