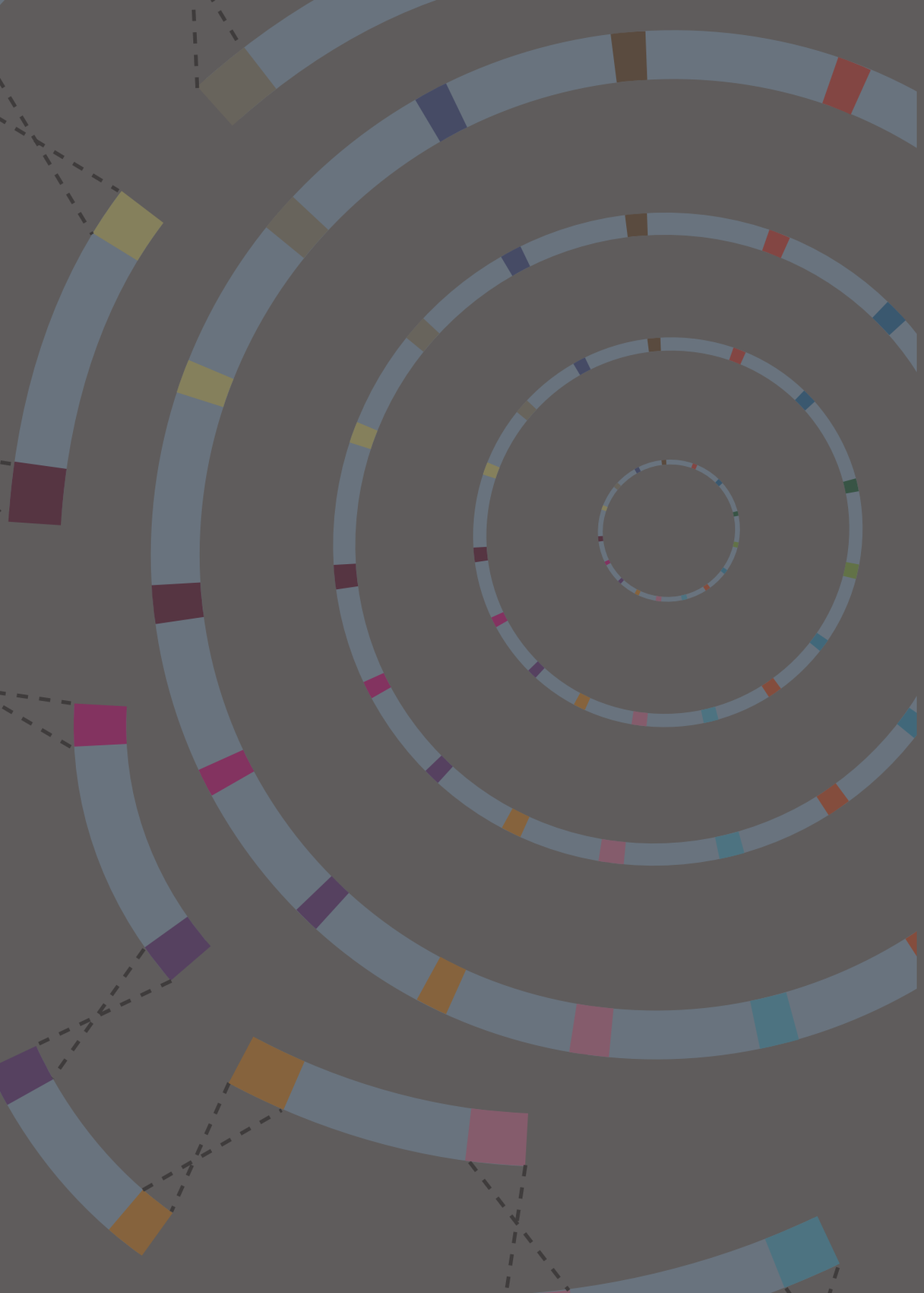
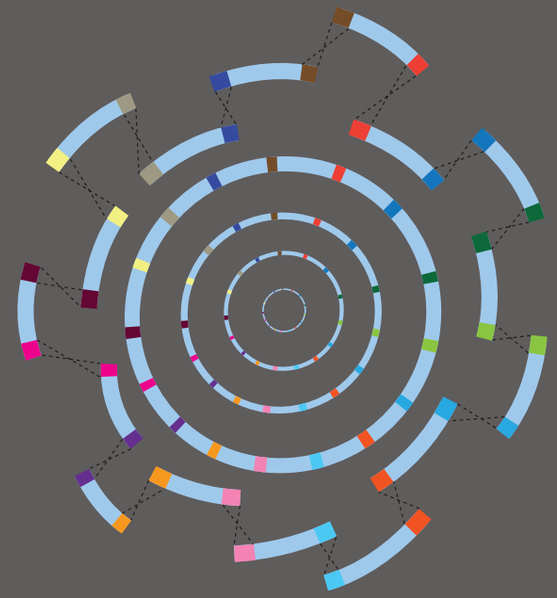


ENGINEERING SYNTHETIC GLYCOLYTIC PATHWAYS IN *Saccharomyces cerevisiae* | **Niels Kuijpers**



ENGINEERING SYNTHETIC GLYCOLYTIC PATHWAYS IN *Saccharomyces cerevisiae*

Niels Kuijpers

Engineering synthetic glycolytic pathways in *Saccharomyces cerevisiae*

Proefschrift

ter verkrijging van de graad van doctor aan de Technische Universiteit Delft, op gezag van de Rector Magnificus prof. ir. K.Ch.A.M Luyben, voorzitter van het College voor Promoties, in het openbaar te verdedigen op woensdag 20 april 2016 om 12:30 uur

door

Niels Gerard Adriaan Kuijpers

Ingenieur in Life Science & Technology
geboren te Apeldoorn

Dit proefschrift is goedgekeurd door de promotor: Prof. dr. J.T. Pronk

Copromotor: Dr. ir. P.A.S Daran-Lapujade

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. J.T. Pronk	Technische Universiteit Delft
Dr. ir. P.A.S. Daran-Lapujade	Technische Universiteit Delft
Dr. ir. J.-M. Daran	Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. W.R. Hagen	Technische Universiteit Delft
Prof. dr. K.-D. Entian	Goethe-Universität
Prof. dr. J. van der Oost	Wageningen UR
Dr. I. Borodina	Technical University of Denmark
Prof. dr. U. Hanefeld	Technische Universiteit Delft, reservelid

The studies presented in this thesis were performed at the Industrial Microbiology section, Department of Biotechnology, Delft University of Technology, the Netherlands and financed by the Ministry of Education, Culture and Science via the Netherlands Organisation for Scientific Research (NOW) through the Technology foundation S T W (Vidi Grant 10776).

Layout by Mirjam Leppers, persoonlijkproefschrift.nl
Printed by Ipskamp Drukkers, Enschede

TU Delft

NWO
Netherlands Organisation for Scientific Research

“beatus homo qui
invenit sapientiam”

Waar het allemaal begon...

Content

CHAPTER 1 Introduction **1**

CHAPTER 2 **35**
A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences

CHAPTER 3 **61**
One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*

CHAPTER 4 **87**
Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in *Saccharomyces cerevisiae*

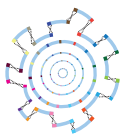
CHAPTER 5 **119**
Pathway swapping: a new approach to radically remodel essential, multigene cellular processes

SAMENVATTING/SUMMARY **185**

CURRICULUM VITAE **197**

LIST OF PUBLICATIONS **199**

DANKWOORD/ ACKNOWLEDGEMENTS **201**



Chapter 1

Introduction

1.1 INDUSTRIAL BIOTECHNOLOGY

An increasing demand for large-scale genetic engineering | Fermentation has been used for thousands of years to conserve and improve the sensory qualities of a large variety of foods and beverages. For example, the ability of yeast cells to ferment sugars into alcohol and carbon dioxide was already exploited in ancient times to produce alcoholic drinks as wine and beer [1]. Similarly, fermentation of milk by lactic acid bacteria has, over the centuries, led to an enormous variety of fermented dairy products [2]. In the last decades, the interest in fermentation products has broadened and intensified. This increased interest originates from the fact that fermentation products, derived from renewable carbohydrate feedstocks, can provide sustainable functional replacements for many oil-derived compounds. Some fermentation products, such as 'bio'-ethanol produced by yeast, can be directly used as substitutes for petrochemistry-derived transport fuels [3]. Others are excellent, renewable precursors for the chemical industry. For example, lactic acid is the building block for synthesis of the biodegradable plastic poly-lactic acid (PLA)[4].

To become economically competitive with petrochemistry, microbial production of chemicals requires product yields on the feedstock that are close to the theoretical limits specified by conservation laws and thermodynamics. Moreover, conversion rates need to be maximized to shorten production times and the robustness of industrial microorganisms needs to be compatible with the harsh conditions in large-scale industrial processes and with the use of low grade feedstocks, which often contain inhibitors of microbial performance. Reaching these goals requires a relentless drive to develop microbes with improved performance.

Classical, non-targeted approaches for microbial strain improvement, such as natural selection and random mutagenesis, remain powerful and, over the past half century, have substantially improved the performance of key industrial 'cell factories'. An iconic example is the over 1000-fold increase in biomass-specific productivity of penicillin by the fungus *Penicillium chrysogenum* [5]. An important drawback of these classical approaches is, however, that improvements are often small and incremental, making strain improvement a slow process. Moreover, classical strain improvement is limited by the natural complement of genetic information present in an industrial microorganism.

Rapid developments in molecular tools for genetic engineering have, over the past few decades, opened up new ways of strain improvement, which are based on rational design and targeted genetic intervention. The resulting, rapidly expanding field of metabolic engineering has been defined as "the improvement of cellular activities by manipulation of enzymatic, transport and regulatory functions of the cell, with the use of recombinant DNA technology" [6]. Metabolic engineering, often in combination with classical strain improvement, is now intensively used to improve current industrial production strains and even to develop completely new processes that can use novel feedstocks and/or enable the production of novel compounds [7, 8]. In the case of bio-ethanol production, metabolic engineering has yielded yeast strains with the

ability to ferment an increased range of sugars [9], with reduced by-product formation [10], increased ethanol yields on substrate [11] and improved resistance to toxic levels of compounds present in the industrial feedstocks [12].

Recombinant-DNA techniques not only allow for improved production of naturally occurring products of microbial metabolism, but also for the production of compounds that are totally new to the production host. Examples include the production, by metabolically engineered yeast cells, of the plant-derived antimalarial compound artimisinic acid [13] and the plant flavonoid naringenin [14]. Implementation of such newly designed and complex metabolic pathways in an industrial microorganism requires extensive engineering of the host genome. For instance, naringenin production by yeast required functional expression of eight heterologous genes, while four native yeast genes had to be inactivated [14]. Such extensive modifications of a specific organism by rational design and engineering principles is increasingly labeled as synthetic biology. Synthetic biology, a fashionable term used to describe a wide range of scientific and engineering activities, has been defined as "the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature" [15]. While this definition clearly overlaps with that of metabolic engineering, it is clear that the past five years have seen a dramatic acceleration of metabolic engineering research that is to a large extent due to the development of highly efficient, cost-effective techniques for the *de novo* synthesis and modification of synthetic DNA molecules. The power of the current synthetic biology toolbox is exemplified by the milestone construction and functional expression of a completely synthetic *Mycoplasma* genome, which required the assembly of no fewer than 1078 DNA fragments [16].

The challenges of engineering living systems are tremendous compared to most engineering strategies outside biology. Whereas the core machinery of a car can easily be replaced while the car is out of order, engineering of living cells can be compared to modifying a car engine while the car is driving. Cells require a constant supply of energy and cellular building blocks to stay alive and a disruption in the supply of these requirements will be fatal. Consequently, essential cellular processes have to remain in operation while the modifications are engineered. This requirement makes it extremely challenging to drastically modify the core machinery of cells and necessitates the development of precise and creative engineering strategies. In this thesis, such a new strategy is explored in order to engineer a key part of the core machinery of carbon metabolism in the yeast *Saccharomyces cerevisiae*: the glycolytic pathway. In this first chapter, an overview of the current toolbox for genetic engineering of *S. cerevisiae* is presented, together with a short description of the glycolytic pathway and an introduction to the approach undertaken in this thesis to engineer it to an extent that has not been achieved before.

1.2 THE GENETIC TOOLBOX FOR GENETIC ENGINEERING OF *S. CEREVISIAE*

The yeast *Saccharomyces cerevisiae* is commonly known as baker's yeast due to its use in the production of leavened bread and is also extensively used in beer and wine fermentation. Moreover, *S. cerevisiae* has been developed into a widely used metabolic engineering platform, due to its excellent accessibility to genetic modification, the availability of robust processes for large-scale industrial cultivation and the extensive knowledge of its physiology and genomics [17, 18]. The excellent amenability of *S. cerevisiae* to molecular genetic modification is related to its extremely efficient homologous recombination machinery, in which homologous sequences of 40 bp are already sufficient to promote high-frequency homologous recombination. This convenient capability allows for accurate targeting and integration of linear DNA fragments into the genome, enabling gene deletions and addition of novel genetic elements [19-22]. The availability of high-quality, annotated sequences of different *S. cerevisiae* strains [17] has further contributed to the immense popularity of *S. cerevisiae* as a versatile, industrially relevant platform for metabolic engineering [3, 13, 14]. In order to engineer synthetic pathways into *S. cerevisiae*, a wide diversity of tools is available. This section describes the current toolbox for genetic engineering of *S. cerevisiae*, which will be used to its full extent in the following chapters.

General principles of genetic modification in *S. cerevisiae* | Expression of a gene in the cell starts with the assembly or chemical synthesis of an expression cassette, which minimally consists of three genetic elements: a promoter, the coding region (on average 1346 bp in eukaryotes [23]) and a terminator. The promoter contains all sequences required to enable the regulated initiation of transcription of a gene into the corresponding mRNA, the terminator contains sequences that signal the end of the transcript to the RNA polymerase, while the coding region represents the DNA sequence that is eventually translated into a protein. Promoters and terminators are organism specific, but may still be recognized by the transcription machinery in a heterologous host when this is genetically closely related to the donor organism [24]. Promoter sequences play an important role in determining the expression level of a gene. In *S. cerevisiae*, several strong, constitutive promoters such as *GPDp, ADH1p, TEF1p, PGK1p* and *PYK1p* are routinely used in expression cassettes to achieve high transcript levels [25]. Alternatively, inducible promoters can be used, whose activity requires activation by a certain compound. A frequently used inducible promoter in *S. cerevisiae* is *GAL1p*, which is activated by galactose [26]. In industrial contexts, inducible (or repressible) promoters provide the option to express genes in a controlled, time dependent manner, for example to prevent product toxicity during an initial biomass propagation phase. The coding sequence itself can also influence the expression level of an expression cassette due to species-specific codon preferences, which are reflected in the codon adaptation index (CAI) [27]. Several algorithms are available to compute the optimal coding sequence for optimal expression of a gene in *S. cerevisiae* yeast [28].

To achieve stable replication of an expression cassette in the yeast cell during cell division, it has to be incorporated in a self-replicating plasmid or integrated in the

genome. Plasmids can be efficiently introduced into yeast cells by transformation and their sizes can range from 300 bp to 2.4 Mbp [29]. *S. cerevisiae* plasmids can be grouped into episomal plasmids (YEp) and centromeric plasmids (YCp). Episomal plasmids rely on sequences derived from the native yeast *2μ* plasmid for their replication and are present in multiple copies per cell [30], while centromeric plasmids contain a centromeric sequence (CEN) and an autonomous replication sequence (ARS), resulting in a single copy per cell [29]. For plasmid-based expression, the expression cassette is first incorporated into the plasmid *in vitro*, after which the plasmid is transformed to *S. cerevisiae*. For chromosomal expression, homologous flanks are added to the expression cassette. Subsequently, the cassette is transformed into *S. cerevisiae* for targeted integration into the genome via homologous recombination. Since the process of transformation has a rather low efficiency, co-transformation of a marker gene is essential to enable selection of positive transformants [31]. Marker genes are therefore indispensable for yeast genetic engineering and used in each cycle of genetic modification. Examples of selection marker genes for *S. cærevisiae* are dominant marker genes such as *KanMX* [32] and *hph* [33], which provide resistance to specific antibiotics, and 'auxotrophic marker genes' such as *LEU2* and *URA3* [34], which complement a specific auxotrophy that is present in the host strain.

Engineering entire pathways into *S. cerevisiae* requires the introduction and functional expression of multiple expression cassettes. The challenge is therefore not only to efficiently transform *S. cerevisiae*, but also to efficiently assemble the genes expressing the whole pathway on a plasmid or to efficiently introduce them into the yeast genome in a targeted manner.

Classical genetic modification: restriction and ligation based assembly of plasmids | In classical genetic modification strategies, DNA fragments are assembled by restriction and ligation of plasmids, after which the ligated product is transferred into *E. coli* for fast and efficient replication [35]. Commonly used plasmids for yeast genetic modification therefore also contain sequences for selection and propagation in *E. coli* [36]. Acting as 'scissors and glue' for DNA, restriction enzymes and ligases have, for almost four decades, been indispensable for the assembly of DNA fragments. The first restriction enzymes were identified in the early 1960s in *Haemophilus influenzae* and naturally act as a defense mechanism against foreign DNA [37]. Heterologous (e.g. phage) DNA sequences that enter bacterial cells are cut at specific palindromic recognition sequences, while the same recognition sequences in the bacterial genome are protected from restriction by methylation. This discovery and the discovery of T4 DNA ligase in 1967 [38, 39], ultimately resulted in the use of this powerful enzyme combination for genetic engineering [40]. Nowadays, around 4000 different restriction enzymes have been identified and deposited in the restriction enzyme database REBASE [41]. Most restriction enzymes used for cloning generate short single-stranded overhangs, often referred to as cohesive ends. Those ends can be ligated to compatible ends of a different DNA fragment by a ligase. In 1987, the polymerase chain reaction (PCR) was added to the genetic toolbox [42], which made it possible to add restriction sites to oligonucleotides, resulting in PCR-products containing the desired restriction sites at both ends [43] (Fig. 1.1).

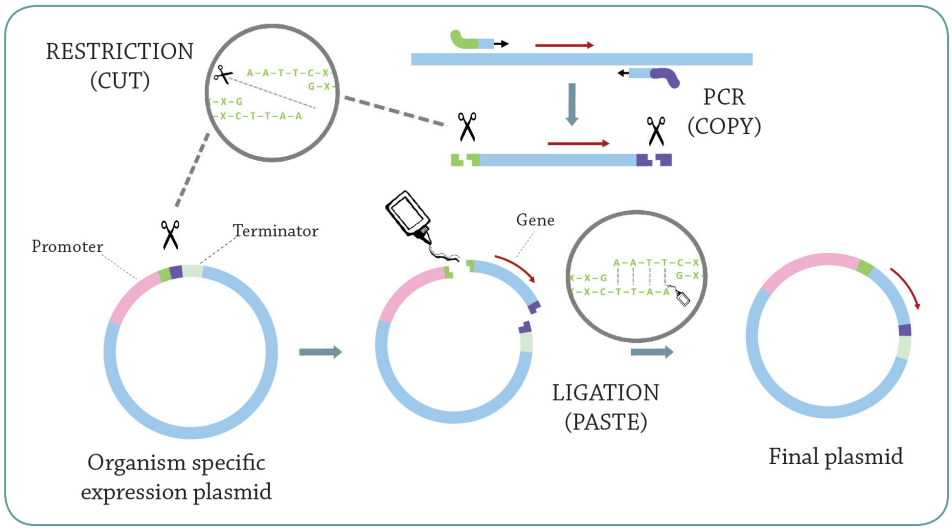


FIGURE 1.1 | 'Classical' strategy for genetic modification based on restriction and ligation. The DNA fragment of interest is obtained by PCR, thereby adding the desired restriction sites to the product. Digestion of the PCR product and digestion of a suitable expression plasmid with the appropriate restriction enzymes results in two DNA fragments with compatible cohesive ends. Those two fragments are assembled into the final plasmid by a ligase.

Although restriction and ligation have dominated genetic modification in the past 30 years, this system has clear limitations. The major shortcoming is the low throughput of DNA parts per assembly. As it is generally difficult and inefficient to simultaneously assembly multiple fragments into a plasmid in a single step, the assembly of multi-gene constructs by restriction and ligation is a laborious and time-consuming effort. Furthermore, the number of unique restriction sites available to clone a fragment into a plasmid decreases as the constructs become larger. Engineering and successful implementation of new and even more complex concepts in biological systems therefore requires restriction-independent genetic modification that allows for assembly of multiple DNA fragments [44].

Methods for simultaneous assembly of multiple DNA fragments | Most currently available restriction-independent DNA assembly techniques for genetic modification of *S. cerevisiae* rely on homology of the terminal sequences of different fragments ('overlapping sequences'). These overlapping sequences can be easily added by extension-PCR and, depending on the method, their length can range from 15 to 80 bp [44]. In general, the basic concept that underlies different recombination-based cloning methods is highly similar: multiple overlapping fragments, of which one fragment contains all sequences required for amplification in *E. coli,* are *in vitro* assembled into a plasmid by using specific recombination enzymes such as exonucleases, polymerases and ligases. The resulting assembly mix is then transformed to *E. coli* for amplification of circular products, followed by analysis of plasmid DNA from

different transformants for correct assembly. One of the first methods applying such a strategy was called SLIC (Sequence and Ligation Independent Cloning) [45]. In this method, DNA fragments with 25bp overlapping sequences are treated with T4 DNA polymerase, an enzyme which harbors 3'exonuclease activity in the absence of dNTP's. The resulting single-strand overhangs facilitate *in vitro* annealing of the fragments in the subsequent annealing step, after which the mixture is transformed into *E. coli* for amplification and *in vivo* ligation of the nicks (Fig. 1.2, first panel). SLIC efficiently and with high fidelity assembles constructs of up to five DNA fragments, but the efficiency of correct assembly strongly decreases as the number of DNA fragments increases. With 10 fragments, as few as 17% of the tested transformants were found to harbor a correctly assembled construct [45]. A similar method for multiple fragment assembly is In-Fusion (Clontech) [46], of which the exact mechanism is kept confidential. This method requires overlapping sequences of 15bp and incorporation of one fragment into a linearized backbone has an efficiency of correct assembly that ranges from 62% to 85%. While this method could be used for simultaneous assembly of multiple fragments, this low efficiency for assembly of a single fragment in a plasmid backbone suggests that In-Fusion is not particularly suited for assembly of large constructs from multiple fragments. A very convenient *in vitro* recombination-based assembly method is Gibson Assembly (Fig. 1.2, second panel). This method is highly analogous to SLIC and uses T5 exonuclease to create the 3' single stranded overhangs [47]. However, Gibson assembly uses *in vitro* nick sealing by Phusion polymerase and Taq ligase, which probably contributes to its high assembly efficiency. The advantage of this method is that it is based on a 'one-pot' reaction at 50°C and that the efficiency of correct assembly is relatively high compared to that of previous methods, since three fragments could be correctly assembled into the expected plasmid with 90% efficiency [47].

Another approach to assemble multiple DNA fragments in a single reaction mixture is the PCR-based Circular Polymerase Extension Cloning (CPEC) method, in which overlapping fragments are assembled by overlap-extension PCR [48] (Fig. 1.2, third panel). In CPEC, the overlapping inserts are mixed with the linearized plasmid backbone and the DNA is melted to single strands, which anneal on the overlapping sequences. In this way, each fragment serves as an oligonucleotide for amplification of the desired plasmid in the subsequent extension thermo-cycle with Phusion polymerase. When a single fragment is inserted in a plasmid backbone, a few thermo-cycles are sufficient to obtain the final plasmid. Transformation to *E. coli* is required to seal the nicks by *in vivo* ligase activity. This method works also with high efficiency for multiple inserts (95% with 4 fragments), but the additional thermo-cycles required for multi-fragment assemblies increase the risk of mutations due to amplification errors introduced by the polymerase.

The recently optimized Ligase Cycling Reaction (LCR) method for assembly of multiple DNA fragments is another very promising strategy for assembly of multiple fragments (Fig. 1.2, fourth panel). In contrast to recombination-based methods, this method does not require overlapping fragments, but uses bridging oligonucleotides to assemble all fragments into a plasmid [49]. Assembly fragments are mixed with the bridging

oligonucleotides and subjected to thermo-cycles, in which the DNA is first melted to enable subsequent annealing of the fragments to the bridging oligonucleotides, after which a thermo-stable ligase seals the nicks. A disadvantage of LCR is that it requires 5'-phosphorylated fragments and that it introduces an additional step into the procedure. The key advantage of this method is that it does not require overlapping sequences and that a polymerase step, which can be prone to mutations, is not involved. LCR can assemble DNA constructs from 12 different DNA fragments with sizes up to 20kb, with efficiencies above 75% of correct assembly [49].

All DNA assembly methods discussed above are essentially *in vitro* methods, which use well defined enzyme kits to obtain the desired product. Nature supplies a wide variety of DNA assembly machineries, since DNA repair mechanisms are essential for survival of cells due to the lethality of double-strand DNA breaks. Use of complete cellular DNA repair mechanisms for assembly of DNA fragments has been explored as alternative for defined enzyme mixtures. An interesting approach is the SLiCE method, which uses cell extracts of laboratory strains of *E. coli* to recombine *in vitro* overlapping DNA fragments [50]. The use of crude cell extracts implies that the exact mechanism of repair is unknown, but SLiCE yields high efficiencies for assembly of a single fragment with a plasmid backbone. However, efficiencies of correct assembly of multiple fragments have not been reported.

Another, intensively explored assembly method that makes use of a full cellular DNA repair machinery is *in vivo* assembly via homologous recombination in *S. cerevisiae* (Fig. 1.2, fifth panel). *S. cerevisiae* and its native homologous recombination system have proved to be an exceptionally efficient DNA assembly factory. As stated before, *S. cerevisiae* has an extremely efficient homologous recombination machinery, whose natural function is to repair double-strand DNA breaks [20, 51]. After early pioneering studies in the 1980's, in which the use of homologous recombination for plasmid assembly in *S. cerevisiae* was explored [21], a 1996 study labeled the technique as 'transformation associated recombination (TAR) cloning' [52]. The technique was developed to isolate human DNA fragments by simultaneously transforming a linearized plasmid and human chromosomal DNA to yeast spheroplasts. The terminal sequences of the linearized plasmid were designed to correspond with sequences homologous to the target region present on the co-transformed human DNA. Homologous recombination of this target region with the linearized plasmid subsequently resulted in a circular plasmid in which the targeted region was incorporated. The full potential of yeast homologous recombination for the purpose of DNA assembly came to light when Gibson *et al.* made intensive use of TAR cloning to assemble a fully synthetic *Mycoplasma genitalium* genome [53]. In this milestone study, it was demonstrated that *S. cerevisiae* was able to assemble the 592 kb *M. genitalium* genome from 25 DNA-fragments with overlapping ends [54]. In another study, the efficiency of this method was further investigated and it was demonstrated that efficiencies of 70% of correct assembly could be achieved for the *in vivo* assembly of 20 kb constructs from 9 overlapping fragments [55]. In **Chapter 2** of this thesis, the potential of *in vivo* DNA assembly in *S. cerevisiae*, as a tool for standardized assembly of large DNA fragments, is further investigated.

Selecting the optimal assembly method for assembly of large synthetic constructs, sometimes harboring the genes for entire metabolic pathways, is challenging. Few quantitative data has been reported on efficiencies of correct assembly and on possible boundaries and limitations of the methods. A study by De Kok *et al.* quantitatively compared LCR, homologous recombination in yeast, CPEC and Gibson assembly [49]. This study clearly demonstrated that Gibson assembly and CPEC are not suited for assembly of more than five fragments, since efficiencies of correct assembly were below 25% for assemblies based on a larger number of starting fragments. Furthermore, these authors concluded that the fidelity of LCR and homologous recombination in yeast were similar for assemblies up to 20 kb in size. This size, however, represents the upper limit for LCR, since fragments over 2 kb in length are poorly assembled when the number of fragments exceeds 10. For one-step assembly of fragments over 20 kb in length, only *in vivo* assembly by homologous recombination in *S. cerevisiae* has been reported to deliver correct assemblies with sufficiently high efficiencies [54].

Editing the genome by engineering double-strand DNA breaks | The DNA assembly methods described in the previous paragraph generate plasmids that contain the construct of interest. When appropriate replication sequences and a marker gene are present, these plasmids can be directly transformed into the host of interest for functional expression. However, plasmid-borne gene expression is generally not favored for industrial-scale production, as plasmids are notoriously unstable, and maintaining the selection pressure necessary to prevent plasmid loss is not always compatible with economic and technological constraints in large-scale processes [56]. Therefore, stable integration of expression cassettes into chromosomal DNA is generally the preferred strategy to ensure faithful replication of the introduced expression cassettes. In this section, the genetic tools for genome editing of *S. cerevisiae* are discussed.

Despite the extensive molecular genetic toolbox for *Saccharomyces cerevisiae*, sequential modifications of the genome have long remained challenging due to the limited number of available marker genes [57] and the relatively low efficiency of integration of transformed, linear DNA fragments. Repeated rounds of transformation, alternated with their recycling of marker genes, is therefore required for the introduction of large numbers of expression cassettes. Marker-gene recycling is also desirable from a physiological perspective, since expression of (multiple) marker genes may have deleterious effects on the performance of yeast cells in industrial processes [34, 58]. Several approaches have been developed to recycle marker genes from the yeast genome. A widely used strategy involves the use of bacteriophage-P1 Cre recombinase, which catalyses recombination at specific recognition sites. When these sites are placed on each side of a marker gene, Cre-mediated recombination excises the marker from the genome [59]. However, this system leaves a 'scar' in the form of a single remaining target site, which is still susceptible to Cre recombinase activity when, during subsequent rounds of genetic modification and marker recycling, additional target sites are introduced into the genome. This situation can lead to unintended deletions and inversions and, when recognition sites are present on different chromosomes, the activity of Cre recombinase has even been shown to lead to chromosomal translocations [60, 61]. Another strategy is excision of the marker

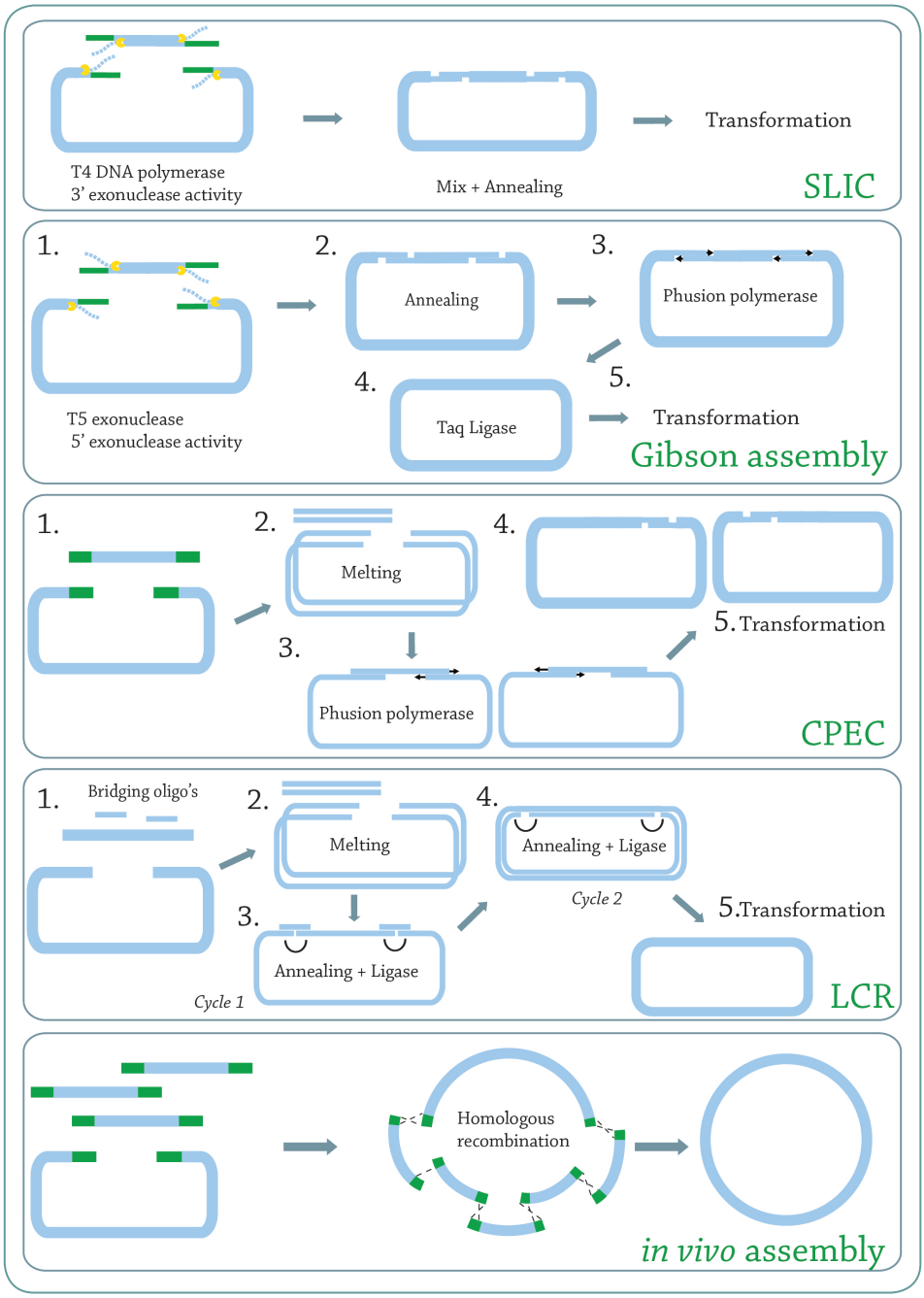


FIGURE 1.2 | Schematic overview of different DNA assembly methods as discussed in this Chapter: SLIC (Sequence and Ligation Independent Cloning)[45], Gibson Assembly [47], CPEC (Circular Polymerase Extension Cloning)[48], LCR (Ligase Cycling Reaction) [49] and *in vivo* assembly by homologous recombination in yeast [54]. For description of methods see text.

by homologous recombination, which is facilitated by placing direct repeats on both sides of the marker gene. Although this system is able to remove the marker gene, the frequency of marker gene removal is lower than in the case of Cre recombinase and requires a method to (counter)select for absence of the marker gene [62].

It has been clearly established that double-strand breaks (DSB) in chromosomal DNA promote homologous recombination [63]. This knowledge has led to the realization that induction of artificial DNA breaks may favor marker gene removal by recombination and, at the same time, could contribute to increased integration efficiency of large DNA fragments into the genome. In view of the lethality of DSBs, the challenge here is to precisely and exclusively introduce the DSB at a very specific sequence. Several methods are available to introduce a specific DSB in the yeast genome. One method exploits the ability of highly specific endonucleases with a long and extremely rare, recognition sequence. A well-studied example is the mitochondrial homing endonuclease I-SceI from *S. cerevisiae* itself [64, 65]. When expressed in the nucleus, I-SceI is able to cut at its 18 bp recognition sequence when this is artificially introduced somewhere in the nuclear genome. The use of I-SceI is further investigated in **Chapter 3** for integration of large DNA constructs and in **Chapter 4** for the simultaneous removal of multiple marker genes. Other popular systems to induce specific DSBs in the genome are zinc-finger nucleases (ZFNs)[66] and Transcription Activator-Like Effector Nucleases (TALENs) [67]. These systems have been especially used in organisms which, in contrast to *S. cerevisiae*, have poor homologous recombination efficiencies as they allow for the deletion of genes by simply cutting them out of the genome. The resulting DSB are then repaired by non-homologous end joining (NHEJ), which is the predominant DSB repair mechanism in many organisms, including mammals [68-70].

Recently, a new system has been developed that may outcompete all previously discussed methods on introducing specific DSBs and which has already unleashed a small revolution in genome editing [71]. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) systems in Bacteria and Archaea use RNA-guided nuclease activity to provide adaptive immunity against foreign DNA fragments [72, 73]. The CRISPR-Cas system from *Streptococcus pyogenes* has been adapted for site specific introduction of DSBs in other organisms [74, 75]. This system consists of the endonuclease Cas9, which is guided to a specific target sequence by a separate small guide-RNA (gRNA) molecule. Functionality of this system, which requires the simultaneous expression of Cas9 and presence of a small guide-RNA (gRNA), has been demonstrated in *S. cerevisiae* [76]. The gRNA, which confers sequence specificity, is less than 100 bp in size and contains a 20 bp recognition sequence. In *S. cerevisiae*, the gRNA can be constitutively expressed from the polymerase-III promoter snr52. The recognition sequence can be any sequence downstream of a genomically encoded NGG triplet. This triplet is called the PAM (protospacer adjacent motif) sequence. Expression of Cas9 and a gRNA targeting the *CAN1* gene in *S. cerevisiae* was demonstrated to be lethal, with only 0.01% - 0.07% of the cells surviving due to mutation of the targeted sequence. When a linear 'repair' DNA fragment, which bridges the targeting site within the *CAN1* locus and contains a mutated PAM sequence, was co-transformed with the gRNA to a Cas9 expressing

strain, homologous recombination occurred in nearly 100% of the transformed cells, leading to exchange of the PAM sequence for the mutated version. This result implied that it would be possible to make gene knockouts in a *CAS9* expressing strain by simply co-transforming a plasmid expressing the appropriate gRNA and a marker-free deletion cassette of 120 bp, which facilitates repair of the introduced DSB by homologous recombination, thereby deleting the gene from the genome. This strategy has recently been proven to be extremely efficient in *S. cerevisiae* and enables the simultaneous and precise editing of multiple genomic loci in a single transformation [77, 78]. Since the system is so efficient, a marker gene is only required for selection of the plasmid from which the gRNA is expressed. Use of this technique in *S. cerevisiae* is further explored in **Chapter 5** for gene deletions and for one-step integration and removal of multi-gene constructs.

Genetic modification 2.0 | The developments in synthetic biology tools discussed in this chapter open up a whole new world of possibilities for engineering genomes to an extent that was unimaginable a decade ago. Boosted by the strongly reduced cost of DNA sequencing and chemical DNA synthesis, the past five years have seen the emergence of several large and ambitious DNA engineering programs. The synthesis and successful implementation of the complete *Mycoplasma genitalium* genome in 2010 and the complete synthesis and assembly of chromosome III of *S. cerevisiae* in 2014 are important technological milestones in synthetic biology [16, 79]. However, while these examples present unprecedented technical achievements, they do not in themselves bring new insight into metabolic pathways and their regulation. In contrast, the combination of synthetic biology, metabolic engineering and systems biology does have the potential to generate such new insights in complex biological systems, for example via the elucidation of regulatory mechanisms and cellular processes that are up to now not fully understood. One of these complex systems, which moreover has a high economical relevance, is a key part of the core machinery of carbon metabolism in many living cells: glycolysis.

1.3 THE GLYCOLYTIC PATHWAY

Because of its central role in the majority of organisms and its economic relevance, the glycolytic pathway is arguably the most intensively studied metabolic pathway. The name 'glycolysis' originates from the Greek words *γλυκύς* (glukus = sugar) and *λύσις* (lysis) and literally means 'unbinding of sugar'. In the glycolytic pathway, glucose is oxidized to two molecules of pyruvate, a process that generates two molecules each of ATP and NADH from ADP/phosphate and NAD^+, respectively. Although glycolysis is highly conserved among organisms, multiple variants exists that contain small modifications or bypasses. The main variant of glycolysis is the Embden-Meyerhof-Parnas pathway (usually referred to as 'glycolysis'), which is the predominant variant in nature [80, 81]. Another well-known variant is the Entner-Doudoroff pathway, which is present in certain prokaryotes and is very rare in eukaryotic cells. Although, just like the Embden-Meyerhof-Parnas pathway, the Entner-Doudoroff pathway encompasses ten reactions, some of the initial reactions are different. The Entner-

Doudoroff pathway yields 1 ATP and 2 NAD(P)H per glucose molecule, implying that the Embden-Meyerhof-Parnas pathway is superior in terms of ATP yield. Some heterofermentative bacteria exploit the phosphoketolase pathway, which yields 1 molecule of ethanol and 1 molecule of lactic acid per glucose, thereby producing net 1 ATP [82]. Another variant in bacteria is the bypass of hexokinase in the PEP-phosphotransferase system (PTS). In this system, phospho-*enol*-pyruvate is dephosphorylated and the phosphate is transferred by the PTS-complex to glucose during its transport across the cytoplasmic membrane [83, 84]. Other less frequently encountered variants of glycolysis are the semi-phosphorylative [85] and phosphorylative [86] variants of the Entner-Doudoroff pathway. Recently, it has been demonstrated that the bacterium *Clostridium thermocellum* possesses an atypical glycolytic pathway in which pyruvate kinase is bypassed [87]. Also many Archaea exploit slightly different glycolytic pathways [88].

During respiratory growth, the pyruvate produced by glycolysis can be further oxidized to carbon dioxide and water. This catabolic process is initiated by oxidative decarboxylation of pyruvate to acetyl-CoA, which can subsequently be completely oxidized in the Krebs cycle. The resulting redox equivalents (NADH and FADH) are then reoxidized by a respiratory chain, thereby generation additional ATP via oxidative phosphorylation. Under anaerobic conditions and in the absence of a suitable external electron acceptor for reoxidizing the NADH generating in glycolysis, pyruvate is reduced to a fermentation product. Organisms capable of sugar fermentation have adapted to different environments, resulting in different fermentation products [89]. For example, where yeast produces ethanol and carbon dioxide from pyruvate (ethanol fermentation), lactic acid bacteria produce lactate (homolactic acid fermentation) or equimolar mixtures of lactate and ethanol (via the phosphoketolase pathway). Other examples of the many fermentation products of bacteria include butyric acid, propionic acid, acetone, butanol, and acetic acid [89].

The glycolytic and fermentative pathways in *S. cerevisiae* | The glycolytic pathway of *S. cerevisiae* has been extensively studied. Wild-type *S. cerevisiae* strains rapidly ferment hexose sugars to ethanol and carbon dioxide [93]. Although *S. cerevisiae* has a fully functional system for respiratory sugar dissimilation, the mode of sugar dissimilation in *S. cerevisiae* in sugar-grown cultures is predominantly fermentative. As a consequence, a low biomass yield is invariably observed in aerobic, glucose-grown batch cultures and a completely respiratory mode of sugar dissimilation is only observed in relatively slow growing, sugar-limited cultures. This phenomenon of aerobic fermentation is referred to as the Crabtree effect [94]. Some authors have proposed that the Crabtree effect in *S. cerevisiae* evolved as a competitive mechanism, in which the antiseptic nature of ethanol inhibits growth of competitors [95].

The first step in the metabolism of hexose sugars is transport from the extracellular environment into the cell. *S. cerevisiae* harbors a wide variety of plasma-membrane hexose transporters, with different characteristics with respect to substrate specificity and affinity for glucose [93]. The predominant transporters are Hxt1 to Hxt7, but it has been demonstrated that at least 21 transporters in *S. cerevisiae* are able

Elucidating the glycolytic pathway | The intermediates and reaction sequence of the Embden-Meyerhof Parnass pathway were only fully elucidated in the 20th century [90]. However, research into this pathway had already been initiated by Louis Pasteur who, in 1860, discovered that living cells are responsible for alcoholic fermentation by unequivocally demonstrating that fermentation does not occur under sterile conditions. In 1897, evidence was first presented that cell free extracts of yeast could also ferment glucose [90]. This result marked the start of decades of intensive research to elucidate the biochemical processes involved in glycolysis. In this quest, scientists focused on yeast cell extracts and on muscle tissue as the two major experimental systems. Much of the work at the start of the 20th century involved the development of techniques for obtaining cell-free extracts and of improved procedures for measuring the rate of carbon dioxide production [90]. Furthermore, the role of inorganic phosphate which, already in 1901, was shown to accelerate fermentation, was intensively investigated. Soon it was discovered that fructose-1,6-biphosphate was present in glucose-converting cell extracts, leading to an hypothetical reaction scheme (i) in which the inorganic phosphate was continuously added to (i) and removed from (ii) 6-carbon sugars (hexoses).

(i) $2\ C_6H_{12}O_6 + 2P_i = 2CO_2 + 2C_2H_5OH + 2H_2O + C_6H_{12}O_6(P_i)_2$

(ii) $C_6H_{12}O_6(P_i)_2 + 2H_2O = C_6H_{12}O_6 + 2P_i$

Early models postulated that the phosphorylated hexose somehow catalysed the degradation of a non-phosphorylated hexose into ethanol and carbon dioxide.

In 1911, pyruvate was first proposed as a likely intermediate in alcoholic fermentation and cell free extracts were shown to be able to convert pyruvate to ethanol and carbon dioxide. Additionally, small amounts of glycerol were observed in sugar-fermenting cell extracts of yeast. These observations led to different hypotheses on the reactions and intermediates involved in the fermentation process. In 1913, Neuberg proposed that methylglyoxal, a highly reactive C3-compound, is a key intermediate in glycolysis, which is subsequently converted into glycerol and pyruvate. Oppenheimer suggested that glycerol originated from dihydroxyacetone, but also supported the idea of the methylglyoxal intermediate. It was Embden who abandoned these earlier models and, in 1933, proposed a brilliant scheme that involved multiple enzymatic reactions. In this scheme, fructose-1,6-biphosphate was split into two triose phosphates. Building on this first basic scheme of the glycolytic pathway, a 1934 study by Parnas demonstrated that the phosphoryl groups were transferred from ATP to hexose, thereby firmly establishing the link between glycolysis and ATP.

In 1951, Entner and Doudoroff described a novel pathway of glucose metabolism in the bacterium *Pseudomonas saccharophila* [91, 92]. The use of radioactively labelled glucose indicated that different reactions contributed to pyruvate production. The key intermediate KDPG (2-keto-3-deoxy-phosphogluconate) was crystallized shortly after and the enzymatic steps of the Entner-Doudoroff pathway of glycolysis were rapidly identified.

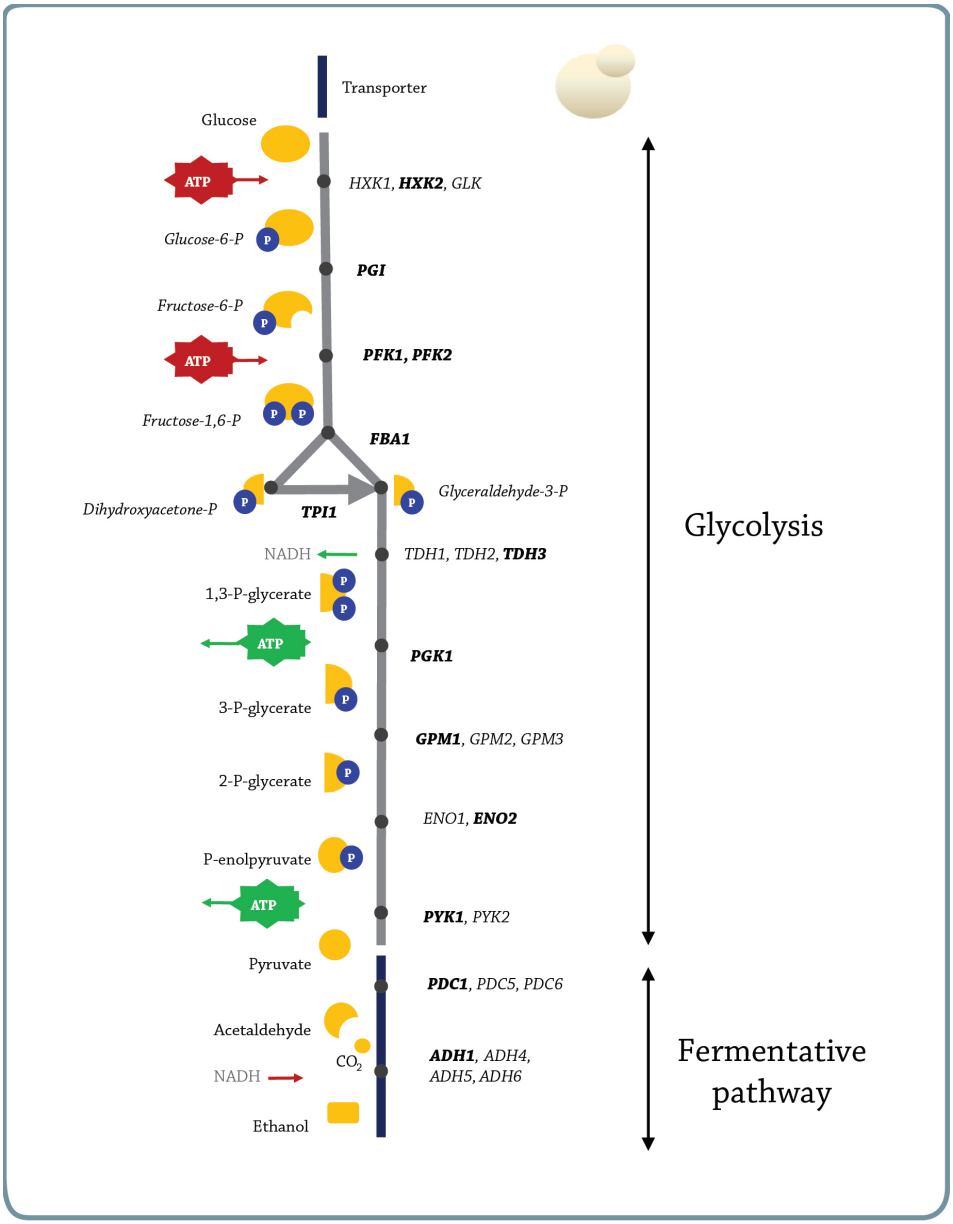


FIGURE 1.3 | Schematic representation of the glycolytic and fermentative pathway in *S. cerevisiae*. For all enzyme-catalysed reactions, the genes encoding the corresponding (iso-)enzyme(s) are indicated. The genes chosen for the 'minimal glycolysis' are indicated in bold (see chapter 1.4 and the PhD thesis of Solis-Escalante [131]).

to transport glucose inside the cell [96]. These differences in uptake characteristics enable *S. cerevisiae* to deal with a large range of extracellular glucose concentrations. Once transported into the cells, intracellular glucose enters the glycolytic pathway and is oxidized to pyruvate, thereby producing 2 ATP molecules per glucose. During fermentation, pyruvate is converted to ethanol and carbon dioxide to regenerate NAD^+. Since glycolytic intermediates are used for biomass synthesis and some carbon dioxide is generated in biosynthetic reactions, glycerol is formed as an additional product for maintaining redox balance during fermentative, anaerobic growth [97].

The 12 reactions of the fermentative pathway (from here on named 'glycolysis') are carried out by 12 different enzymes. For some of these enzymes multiple variants exist, referred to as iso-enzymes. In total, the glycolytic pathway of *S. cerevisiae* involves 26 enzymes, which are encoded by 26 structural genes [98] (Fig. 1.3). Phosphofructokinase represents a special case, as it consists of heterooctamers that contain Pfk1 as well as Pfk2 proteins and deletion of either *PFK1* or *PFK2* strongly reduces growth on glucose [99]. Not all genes encoding iso-enzymes are constitutively expressed. For example, *PDC1,* which encodes pyruvate decarboxylase, is under most cultivation conditions the main enzyme catalyzing the decarboxylation of pyruvate to acetaldehyde. However, under sulfur-limited conditions, yeast cells switch to Pdc6, an iso-enzyme which has a much lower content of sulfur-containing amino acids [100]. For most other iso-enzymes, such distinctive regulation patterns or roles have not been found and their biological relevance in *S. cerevisiae* remains unknown. However, some of the glycolytic enzymes have been reported to fulfil a secondary, non-glycolytic function in the cell, referred to as a 'moonlighting function'. Three cases of moonlighting have been reported for glycolytic enzymes in *S. cerevisiae*: Hxk2 [101] is involved in glucose repression, Fba1 has an essential role in the activity and assembly of vacuolar ATPase [102], and Eno2 is involved in vacuolar fusion [103].

Despite all research efforts to understand the glycolytic pathway in *S. cerevisiae*, the precise regulatory mechanisms controlling the flux through this key pathway remain to be elucidated. Most glycolytic enzymes are regulated at the transcriptional level and, to a significant extent, are constitutively expressed [98]. Moreover, no clear correlation exists between transcript levels and the glycolytic flux [104, 105]. This implies that post-transcriptional regulation mechanisms are dominant in the regulation of the activity of the glycolytic genes [106]. For example, recent studies revealed that about two-thirds of the *S. cerevisiae* metabolic proteins are subject to the phosphatase signaling network [107-109] and *in vivo* activities of glycolytic enzymes are strongly influenced by intracellular concentrations of low-molecular-weight substrates, products and effectors. This is exemplified by the complex allosteric regulation of Pfk1,2 and Pyk1 [110, 111]. Recently, it has also been demonstrated that acetylation of lysine residues in metabolic enzymes of human liver tissue plays a major role in metabolic regulation [112]. Although this phenomenon has not yet been thoroughly investigated in *S. cerevisiae*, there are some strong indications that acetylation is also a regulatory mechanism active in yeast [113-115].

Engineering glycolysis in *Saccharomyces cerevisiae:* economical relevance | Many industrially relevant chemicals produced by *S. cerevisiae* are directly or indirectly derived from glycolysis. The ability to control the glycolytic flux, and thereby to increase productivity, therefore has a great economical relevance. Despite decades of research, the mechanisms regulating the glycolytic flux are still not fully unraveled and therefore flux prediction based on models is still very hard to achieve, especially under dynamic conditions [116]. Attempts to regulate the glycolytic flux by genetic engineering have, up to now, met with very limited success. Several attempts to increase the glycolytic flux were based on individual or combined overexpression of glycolytic genes. In one now classical study, eight glycolytic genes (*HXK2*, *PGI1*, *PFK1*/*PFK2*, *PGK1*, *PGM*, *PYK1*, *PDC1* and *ADH1*) were individually overexpressed on multi-copy plasmids. While this yielded increased enzyme activities of up to 14 times relative to the reference strain, no significant effects on the ethanol production rate or specific growth rate were observed [117]. In another study, seven genes encoding enzymes of the lower part of glycolysis (*TPI1*, *PGK1*, *ENO1*, *GPM1*, *PYK1*, *PDC1* and *ADH1*) were simultaneously overexpressed from the yeast genome, resulting in an at least 2-fold overexpression [118]. Also in this case, extensive metabolic engineering of glycolysis did not have a clear impact on ethanol production or growth rates.

In several studies, it has been investigated whether glycolytic genes from other organisms can complement the corresponding mutations in *S. cerevisiae*. Many heterologous glycolytic enzymes have been shown to complement their *S. cerevisiae* ortholog. For example, the phosphofructokinase mutants can be complemented by expression of phosphofructokinase genes from *E. coli* [119], *Homo sapiens* [119], *Kluyveromyces lactis* [120] and *Dictyostelium discoideum* [121]. Simultaneous replacement of multiple enzymes of yeast glycolysis, in order to more drastically modify the pathway, has not yet been attempted.

Another important economic parameter is the product yield on substrate, especially for commodity chemicals, in which the price of substrate accounts for a large part for the overall process costs [122]. It would therefore also be interesting to replace the native yeast glycolytic pathway for variants that have a different stoichiometry, in particular with respect to ATP yield. Ethanol production with *S. cerevisiae* has a net ATP yield of 2 ATP per glucose, which is invested in maintenance and biomass formation. During industrial production of ethanol, excess microbial biomass is an undesired byproduct, whose formation goes at the expense of ethanol yield on substrate. In this case, a reduced ATP yield would be beneficial for industrial production strains and is an interesting target for metabolic engineering. This might, for example, be achieved by replacing the Embden Meyerhof-Parnas glycolysis in *S. cerevisiae* by a bacterial Entner-Doudoroff pathway, which has a lower ATP on substrate yield. Serious efforts to achieve this goal have, however, so far not been successful [123].

Since, unlike many other industrial microbes, *S. cerevisiae* is tolerant to low pH, it is an extremely attractive host for the production of weak organic acids such as lactic acid, succinic acid and malic acid. Production of these compounds at low pH reduces the need for base titration during fermentation and enables the microbial production of the acid in its undissociated form ($pH<pK_a$). This is very beneficial for the recovery of the acid, since less acidifying agents are required in downstream processing and the production of large quantities of salts as by-products can be avoided [124]. The main hurdle for the production of weak organic acids with genetically modified *S. cerevisiae* is that export of these fermentation products at low pH requires ATP [125]. In the case of engineered homolactic strains of *S. cerevisiae*, this ATP-requirement resulted in a net ATP yield of the fermentative pathway of zero. As a consequence, lactic acid production could only be achieved when a sufficient amount of ATP for maintenance was formed through respiration, which requires oxygen supply [126, 127]. Aeration of industrial-scale fermentations is expensive and leads to loss of substrate as CO_2, thereby increasing process costs and lowering the product yield on substrate. This particular case illustrates the need to increase the net ATP yield on substrate for organic acids and other chemicals that lead to a 'zero ATP' scenario, to enable product formation under anaerobic conditions [122].

Synthetic glycolytic pathways | As discussed above, the regulation of glycolytic flux in *S. cerevisiae* depends on multiple levels of cellular regulation. The complex interaction of the kinetic properties of the individual enzymes and their allosteric and post-translation regulation make it very hard to predict and tune glycolytic flux. The option to replace the complete system by an array of synthetic variants, which allows the expression of any desired combination of mutations and variants, would be of tremendous value for unravelling regulatory mechanisms and assess their quantitative contribution to regulation of the glycolytic flux. While synthetic biology tools have been developed to engineer complete synthetic pathways into yeast, these have predominantly been used for the iterative addition of new functions to the cell and not for simultaneous engineering of all steps in a central metabolic pathway [55]. The latter is also far more challenging, especially in the case of glycolysis, which is an essential pathway for survival of the cell. The lack of viability of many glycolytic null mutants implies that simple screening platforms for individual glycolytic genes are not available. Moreover, the presence of iso-enzymes complicates the formulation of unambiguous mathematical models of the pathway [128]. Therefore, modelling generally rests heavily on *in vitro* data on individual iso-enzymes and does not take into account the interplay of iso-enzymes with different regulatory and kinetic properties [129]. An *in vivo* screening platform for *in silico* designed glycolytic pathways could therefore be a valuable tool to validate mathematical models of the kinetics and regulation of yeast glycolysis. Enabling the replacement of the complete glycolytic pathway, instead of relying on iterative, step-by-step modifications, requires that the glycolytic genes, which are scattered over the genome, have to be localized to a single chromosomal locus from which they can be removed in a single step. As will be outlined below, the engineering of such a valuable experimental platform for research on the glycolytic pathway in yeast is the overarching goal of the research described in this thesis.

1.4 SCOPE OF THIS THESIS

This PhD thesis is part of a larger research project, funded from a VIDI-grant to Dr. Pascale Daran-Lapujade. The central goal of this project is to construct a yeast platform in which the glycolytic pathway can be easily and rapidly exchanged for synthetic variants. Such a platform, in which a sheer endless variety of configurations of glycolysis can be tested, would offer unique novel possibilities to study the fundamental design of this essential and near-ubiquitous pathway. In order to achieve this goal, a research plan with several key milestones was written (see Fig. 1.4 for an outline). The first challenging goal was to reduce the complexity of the glycolytic system in yeast by minimizing the number of genes required for a functional pathway. For each step, one major enzyme was selected and the structural genes encoding the other iso-enzymes were deleted from the genome [130]. In the next phase, which will be described in this thesis, the remaining glycolytic genes were to be relocalized to a single chromosomal locus, resulting in a **Sin**gle **Lo**cus **G**lycolysis (SinLoG) yeast. This target represented a key stepping stone towards the replacement of the complete glycolytic pathway. After construction of this platform, the next target was therefore to demonstrate the introduction of synthetic variants of the glycolytic pathway, followed by removal of the native SinLoG cassette (Fig. 1.4). In order to achieve its goals, the project was divided in two research lines:

(i) Simplification of the native yeast glycolytic pathway of *S. cerevisiae* by removing all glycolytic iso-enzymes and elucidate their role in glycolysis. This work is described in the PhD thesis of Daniel Solis-Escalante: Reducing the generic complexity of glycolysis in *Saccharomyces cerevisiae* [131].

(ii) Construct and validate a 'plug-and-play' system in which redesigned versions of the glycolytic pathway can be easily implemented in yeast. This goal represents the major objective of the research presented in this thesis.

Daniel Solis-Escalante [131] demonstrated that reduction of the genetic complexity of the glycolytic pathway in *S. cerevisiae* from 26 to 13 paralogous genes did not, under a wide variety of laboratory cultivation conditions, affect yeast physiology and glycolytic flux. To construct the envisaged 'plug and play' platform, this minimal set of glycolytic genes had to be relocated to a single chromosomal locus. This required engineering of a synthetic construct of approximately 35 kb, which could be assembled and specifically targeted to a selected chromosomal locus of interest. **Chapter 2** explores how *in vivo* assembly by homologous recombination can be standardized and used at a scale that is compatible with the requirements for assembly of the number of fragments that is required for a 'plug and play' glycolysis. **Chapter 3** investigates whether this assembly system, which enabled highly efficient assembly of multigene constructs, can be combined with similarly efficient, targeted chromosomal integration. Specifically, this Chapter explores the targeted introduction of artificial DSB at a selected locus in the chromosome with the meganuclease I-SceI as a means to facilitate simultaneous assembly and integration of multi-gene constructs.

Once the glycolytic pathway was introduced in to a *S. cerevisiae* strain containing the minimalized glycolytic pathway, all remaining native glycolytic genes had to be removed. The sequential removal of 13 genes first appeared to be cumbersome, time-consuming process hindered by the limited availability of selection marker genes and the inevitable necessity to recover selection markers after deletion. To address this recurrent problem, **Chapter 4** describes a new method for simultaneous recycling of multiple (dominant) marker genes by I-SceI-facilitated recombination. This technique, which greatly shortened the time-scale of sequential deletion programs, was subsequently used to construct the 'plug-and-play' synthetic glycolysis platform.

The goal of the research described in **Chapter 5** is into integrate the different strands of research in the project by providing a proof of principle of the 'plug and play' system. It describes the engineering of a platform strain (GlycoSwitch) that contains a minimal, but fully functional, cluster of glycolytic genes placed at a single chromosomal locus. Furthermore, this Chapter explores the complete replacement of the native *S. cerevisiae* glycolytic pathway with a minimalized synthetic variant composed of genes originating from *S. cerevisisae*, *S. kudriavzevii* and *Homo sapiens*.

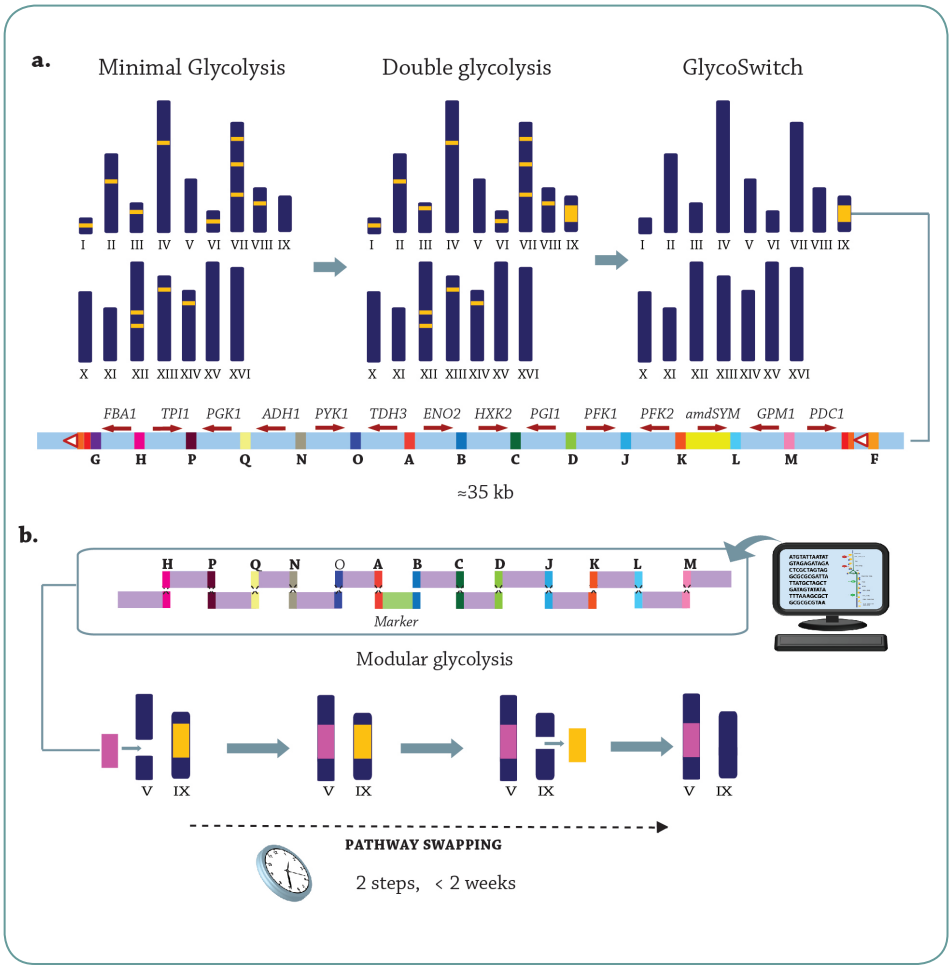


FIGURE 1.4 | Schematic overview of the approach to develop a platform for engineering synthetic glycolytic pathways in *S. cerevisiae*. The recently engineered 'minimal glycolysis' strain [131] will be used as a starting strain to introduce a synthetic glycolytic pathway module on chromosome IX, resulting in a single chromosomal locus that harbors all genes required for a functional glycolytic pathway. In the next phase, all native chromosomal counterparts of the glycolytic genes will be deleted, resulting in the desired platform strain (GlycoSwitch) (A). In this platform strain, a synthetic modular glycolysis can be engineered into a target site on chromosome V, after which the original clustered glycolytic genes on chromosome IX can be removed in a single step (B).

REFERENCE LIST

1. Sewell SL: **The spatial diffusion of beer from its Sumerian origins to today.** In *The Geography of Beer.* Edited by Patterson M, Hoalst-Pullen N: Springer Netherlands; 2014: 23-29.

2. Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, Evershed RP: **Earliest evidence for cheese making in the sixth millennium bc in northern Europe.** *Nature* 2013, **493:**522-525.

3. Nielsen J, Larsson C, van Maris A, Pronk J: **Metabolic engineering of yeast for production of fuels and chemicals.** *Curr Opin Biotechnol* 2013, **24:**398-404.

4. Garlotta D: **A Literature Review of Poly(Lactic Acid).** *Journal of Polymers and the Environment* 2001, **9:**63-84.

5. Thykaer J, Nielsen J: **Metabolic engineering of beta-lactam production.** *Metab Eng* 2003, **5:**56-69.

6. Bailey JE: **Toward a science of metabolic engineering.** *Science* 1991, **252:**1668-1675.

7. Hong KK, Nielsen J: **Metabolic engineering of *Saccharomyces cerevisiae* : a key cell factory platform for future biorefineries.** *Cell Mol Life Sci* 2012, **69:**2671-2690.

8. Chen X, Zhou L, Tian K, Kumar A, Singh S, Prior BA, Wang Z: **Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production.** *Biotechnol Adv* 2013, **31:**1200-1223.

9. Kuyper M, Hartog MM, Toirkens MJ, Almering MJ, Winkler AA, van Dijken JP, Pronk JT: **Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation.** *FEMS Yeast Res* 2005, **5:**399-409.

10. Guadalupe Medina V, Almering MJ, van Maris AJ, Pronk JT: **Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor.** *Appl Environ Microbiol* 2010, **76:**190-195.

11. Basso TO, de Kok S, Dario M, do Espirito-Santo JC, Muller G, Schlolg PS, Silva CP, Tonso A, Daran JM, Gombert AK, et al: **Engineering topology and kinetics of sucrose metabolism in *Saccharomyces cerevisiae* for improved ethanol yield.** *Metab Eng* 2011, **13:**694-703.

12. Swinnen S, Fernandez-Nino M, Gonzalez-Ramos D, van Maris AJ, Nevoigt E: **The fraction of cells that resume growth after acetic acid addition is a strain-dependent parameter of acetic acid tolerance in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2014, **14:**642-653.

13. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al: **High-level semi-synthetic production of the potent antimalarial artemisinin.** *Nature* 2013, **496:**528-532.

14. Koopman F, Beekwilder J, Crimi B, van Houwelingen A, Hall RD, Bosch D, van Maris AJ, Pronk JT, Daran JM: ***De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*.** *Microb Cell Fact* 2012, **11:**155.

15. Serrano L: **Synthetic biology: promises and challenges.** *Mol Syst Biol* 2007, **3:**158.

16. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, et al: **Creation of a bacterial cell controlled by a chemically synthesized genome.** *Science* 2010, **329:**52-56.

17. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al: **Life with 6000 Genes.** *Science* 1996, **274:**546-567.

18. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis.** *Proc Natl Acad Sci U S A* 1997, **94:**13057-13062.

19. Resnick M, Martin P: **The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control.** *Mol Gen Genet* 1976, **143:**119-129.

20. Orr-Weaver TL, Szostak JW, Rothstein RJ: **Yeast transformation: a model system for the study of recombination.** *Proc Natl Acad Sci USA* 1981, **78:**6354-6358.

21. Ma H, Kunes S, Schatz PJ, Botstein D: **Plasmid construction by homologous recombination in yeast.** *Gene* 1987, **58:**201-216.

22. Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, Stubbs L, Kouprina N, Larionov V: **Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning.** *Nucleic Acids Res* 2001, **29:**E32-E32.

23. Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW: **Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms.** *Mol Biol Evol* 2006, **23:**1107-1108.

24. Schirmaier F, Philippsen P: **Identification of two genes coding for the translation elongation factor EF-1 alpha of *Saccharomyces cerevisiae*.** *Embo j* 1984, **3:**3311-3315.

25. Partow S, Siewers V, Bjørn S, Nielsen J, Maury J: **Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*.** *Yeast* 2010, **27:**955-964.

26. Johnston M, Flick JS, Pexton T: **Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1994, **14:**3834-3841.

27. Sharp PM, Li WH: **The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.** *Nucleic Acids Res* 1987, **15:**1281-1295.

28. Gritsenko A, Reinders MT, de Ridder D: **Using Predictive Models to Engineer Biology: A Case Study in Codon Optimization.** In *Pattern Recognition in Bioinformatics. Volume* 7986. Edited by Ngom A, Formenti E, Hao J-K, Zhao X-M, van Laarhoven T: Springer Berlin Heidelberg; 2013. [*Lecture Notes in Computer Science*].

29. Kado CI: **Origin and evolution of plasmids.** *Antonie Van Leeuwenhoek* 1998, **73:**117-126.

30. Broach JR, Hicks JB: **Replication and recombination functions associated with the yeast plasmid, 2 mu circle.** *Cell* 1980, **21:**501-508.

31. Gietz RD, Woods R: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Methods Enzymol* 2002, **350:**87-96.

32. Wach A, Brachat A, Pohlmann R, Philippsen P: **New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*.** *Yeast* 1994, **10:**1793-1808.

33. Gritz L, Davies J: **Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia* coli and *Saccharomyces cerevisiae*.** *Gene* 1983, **25:**179-188.

34. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD: **Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.** *Yeast* 1998, **14:**115-132.

35. Cohen SN, Chang AC, Boyer HW, Helling RB: **Construction of biologically functional bacterial plasmids *in vitro*.** *Proc Natl Acad Sci U S A* 1973, **70:**3240-3244.

36. Chee MK, Haase SB: **New and redesigned pRS plasmid shuttle vectors for genetic manipulation of *Saccharomyces cerevisiae*.** *G3 (Bethesda)* 2012, **2:**515-526.

37. Dussoix D, Arber W: **Host specificity of DNA produced by *Escherichia coli:* II. Control over acceptance of DNA from infecting phage lambda.** *J Mol Biol* 1962, **5:**37-49.

38. Zimmerman SB, Little JW, Oshinsky CK, Gellert M: **Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide.** *Proc Natl Acad Sci U S A* 1967, **57:**1841-1848.

39. Weiss B, Richardson CC: **Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage.** *Proc Natl Acad Sci U S A* 1967, **57:**1021-1028.

40. Danna K, Nathans D: **Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*.** *Proc Natl Acad Sci U S A* 1971, **68:**2913-2917.

41. Roberts RJ, Vincze T, Posfai J, Macelis D: **REBASE--a database for DNA restriction and modification: enzymes, genes and genomes.** *Nucleic Acids Res* 2010, **38:**D234-236.

42. Mullis KB, Faloona FA: **Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction.** *Methods Enzymol* 1987, **155:**335-350.

43. Kaufman DL, Evans GA: **Restriction endonuclease cleavage at the termini of PCR products.** *Biotechniques* 1990, **9:**304-306.

44. Ellis T, Adie T, Baldwin GS: **DNA assembly for synthetic biology: from parts to pathways and beyond.** *Integr Biol (Camb)* 2011, **3:**109-118.

45. Li MZ, Elledge SJ: **Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC.** *Nat Methods* 2007, **4:**251-256.

46. Benoit RM, Wilhelm RN, Scherer-Becker D, Ostermeier C: **An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids.** *Protein Expr Purif* 2006, **45:**66-71.

47. Gibson DG, Young L, Chuang RY, Venter JC, Iii CAH, Smith HO, America N: **Enzymatic assembly of DNA molecules up to several hundred kilobases.** *Nat Methods* 2009, **6:**12-16.

48. Quan J, Tian J: **Circular polymerase extension cloning of complex gene libraries and pathways.** *PLoS One* 2009, **4:**e6441.

49. de Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, Platt D, Shapland EB, Serber Z, Dean J, et al: **Rapid and reliable DNA assembly via ligase cycling reaction.** *Nucleic Acids Res* 2014, **3:**97-106.

50. Zhang Y, Werling U, Edelmann W: **Seamless Ligation Cloning Extract (SLiCE) cloning method.** *Methods Mol Biol* 2014, **1116:**235-244.

51. Jasin M, Rothstein R: **Repair of strand breaks by homologous recombination.** *Cold Spring Harb Perspect Biol* 2013, **5:**a012740.

52. Larionov V, Kouprina N, Graves J, Resnick MA: **Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning.** *Proc Natl Acad Sci U S A* 1996, **93:**13925-13930.

53. Gibson DG, Benders G, Andrews-Pfannkoch C, Denisova E, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire M, et al: **Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome.** *Science* 2008, **319:**1215-1220.

54. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA, 3rd: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome.** *Proc Natl Acad Sci U S A* 2008, **105:**20404-20409.

55. Shao Z, Zhao H, Zhao H: **DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways.** *Nucleic Acids Res* 2009, **37:**e16.

56. Zhang Z, Moo-Young M, Chisti Y: **Plasmid stability in recombinant *Saccharomyces cerevisiae*** *Biotechnol Adv* 1996, **14:**401-435.

57. Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P: **amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:**126-139.

58. Baganz F, Hayes A, Farquhar R, Butler PR, Gardner DC, Oliver SG: **Quantitative analysis of yeast gene function using competition experiments in continuous culture.** *Yeast* 1998, **14:**1417-1427.

59. Sauer B: **Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1987, **7:**2087-2096.

60. Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ, Oliver SG: **Exploring redundancy in the yeast genome: an improved strategy for use of the cre-loxP system.** *Gene* 2000, **252:**127-135.

61. Solis-Escalante D, van den Broek M, Kuijpers NGA, Pronk JT, Boles E, Daran JM, Daran-Lapujade P: **The genome sequence of the popular hexose-transport-deficient *Saccharomyces cerevisiae strain* EBY. VW4000 reveals LoxP/Cre-induced translocations and gene loss.** *FEMS Yeast Res* 2015, **15**.

62. Alani E, Cao L, Kleckner N: **A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains.** *Genetics* 1987, **116:**541-545.

63. Storici F, Durham CL, Gordenin DA, Resnick MA: **Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast.** *Proc Natl Acad Sci U S A* 2003, **100:**14994-14999.

64. Plessis A, Perrin A, Haber JE, Dujon B: **Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus.** *Genetics* 1992, **130:**451-460.

65. Choulika A, Perrin A, Dujon B, Nicolas JF: **Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*** *Mol Cell Biol* 1995, **15:**1968-1973.

66. Kim YG, Cha J, Chandrasegaran S: **Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.** *Proc Natl Acad Sci U S A* 1996, **93:**1156-1160.

67. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF: **Targeting DNA double-strand breaks with TAL effector nucleases.** *Genetics* 2010, **186:**757-761.

68. Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X: **Targeted genome modification in mice using zinc-finger nucleases.** *Genetics* 2010, **186:**451-459.

69. Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T: **Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases.** *PLoS One* 2010, **5:**e8870.

70. Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, et al: **Knockout rats generated by embryo microinjection of TALENs.** *Nat Biotechnol* 2011, **29:**695-696.

71. Hsu PD, Lander ES, Zhang F: **Development and applications of CRISPR-Cas9 for genome engineering.** *Cell* 2014, **157:**1262-1278.

72. Carroll D: **A CRISPR approach to gene targeting.** *Mol Ther* 2012, **20:**1658-1660.

73. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.** *Science* 2012, **337:**816-821.

74. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339:**823-826.

75. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339:**819-823.

76. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM: **Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.** *Nucleic Acids Res* 2013, **41:**4336-4343.

77. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJ, Daran JM: **CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2015, **15**.

78. Jakociunas T, Bonde I, Herrgard M, Harrison SJ, Kristensen M, Pedersen LE, Jensen MK, Keasling JD: **Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*.** *Metab Eng* 2015, **28:**213-222.

79. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, et al: **Total synthesis of a functional designer eukaryotic chromosome.** *Science* 2014, **344:**55-58.

80. Romano AH, Conway T: **Evolution of carbohydrate metabolic pathways.** *Res Microbiol* 1996, **147:**448-455.

81. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R: **Glycolytic strategy as a tradeoff between energy yield and protein cost.** *Proc Natl Acad Sci U S A* 2013, **110:**10039-10044.

82. Heath EC, Hurwitz J, Horecker BL, Ginsburg A: **Pentose fermentation by *Lactobacillus plantarum*. I. The cleavage of xylulose 5-phosphate by phosphoketolase.** *J Biol Chem* 1958, **231:**1009-1029.

83. Deutscher J, Francke C, Postma PW: **How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria.** *Microbiol Mol Biol Rev* 2006, **70:**939-1031.

84. Kundig W, Ghosh S, Roseman S: **Phosphate Bound to histidine in a protein as an intermediate in a novel phospho-transferase system** *Proc Natl Acad Sci U S A* 1964, **52:**1067-1074.

85. Elzainy TA, Hassan MM, Allam AM: **New pathway for nonphosphorylated degradation of gluconate by *Aspergillus niger*.** *J Bacteriol* 1973, **114:**457-459.

86. Szymona M, Doudoroff M: **Carbohydrate metabolism in *Rhodopseudomonas sphreoides*.** *J Gen Microbiol* 1960, **22:**167-183.

87. Zhou J, Olson DG, Argyros DA, Deng Y, van Gulik WM, van Dijken JP, Lynd LR: **Atypical glycolysis in *Clostridium thermocellum*.** *Appl Environ Microbiol* 2013, **79:**3000-3008.

88. Verhees CH, Kengen SW, Tuininga JE, Schut GJ, Adams MW, De Vos WM, Van Der Oost J: **The unique features of glycolytic pathways in Archaea.** *Biochem J* 2003, **375:**231-246.

89. Müller V: **Bacterial Fermentation.** In *eLS* Edited by Finazzi-Agró A. John Wiley & Sons, Ltd; 2001.

90. Barnett JA: **A history of research on yeasts 5: the fermentation pathway.** *Yeast* 2003, **20:**509-543.

91. Wiame JM, Doudoroff M: **Oxidative assimilation by *Pseudomonas saccharophilia* with C14-labeled substrates.** *J Bacteriol* 1951, **62:**187-193.

92. Entner N, Doudoroff M: **Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*.** *J Biol Chem* 1952, **196:**853-862.

93. Otterstedt K, Larsson C, Bill RM, Stahlberg A, Boles E, Hohmann S, Gustafsson L: **Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*.** *EMBO Rep* 2004, **5:**532-537.

94. De Deken RH: **The Crabtree effect: a regulatory system in yeast.** *J Gen Microbiol* 1966, **44:**149-156.

95. Hagman A, Sall T, Compagno C, Piskur J: **Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication.** *PLoS One* 2013, **8:**e68734.

96. Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP, Boles E: **Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*.** *FEBS Lett* 1999, **464:**123-128.

97. Bakker BM, Overkamp KM, van Maris AJ, Kotter P, Luttik MA, van Dijken JP, Pronk JT: **Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*.** *FEMS Microbiol Rev* 2001, **25:**15-37.

98. Fraenkel DG: **The top genes: on the distance from transcript to function in yeast glycolysis.** *Curr Opin Microbiol* 2003, **6:**198-201.

99. Breitenbach-Schmitt I, Heinisch J, Schmitt HD, Zimmermann FK: **Yeast mutants without phosphofructokinase activity can still perform glycolysis and alcoholic fermentation.** *Mol Gen Genet* 1984, **195:**530-535.

100. Boer VM, de Winde JH, Pronk JT, Piper MD: **The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur.** *J Biol Chem* 2003, **278:**3265-3274.

101. Ahuatzi D, Herrero P, de la Cera T, Moreno F: **The glucose-regulated nuclear localization of hexokinase 2 in *Saccharomyces cerevisiae* is Mig1-dependent.** *J Biol Chem* 2004, **279:**14440-14446.

102. Lu M, Ammar D, Ives H, Albrecht F, Gluck SL: **Physical interaction between aldolase and vacuolar H+-ATPase is essential for the assembly and activity of the proton pump.** *J Biol Chem* 2007, **282:**24495-24503.

103. Decker BL, Wickner WT: **Enolase activates homotypic vacuole fusion and protein transport to the vacuole in yeast.** *J Biol Chem* 2006, **281:**14523-14528.

104. Jansen ML, Diderich JA, Mashego M, Hassane A, de Winde JH, Daran-Lapujade P, Pronk JT: **Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity.** *Microbiology* 2005, **151:**1657-1669.

105. Mashego MR, Jansen ML, Vinke JL, van Gulik WM, Heijnen JJ: **Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats.** *FEMS Yeast Res* 2005, **5:**419-430.

106. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, Heck AJ, Daran JM, de Winde JH, Westerhoff HV, et al: **The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels.** *Proc Natl Acad Sci USA* 2007, **104:**15753-15758.

107. Oliveira AP, Ludwig C, Picotti P, Kogadeeva M, Aebersold R, Sauer U: **Regulation of yeast central metabolism by enzyme phosphorylation.** *Mol Syst Biol* 2012, **8:**623.

108. Portela P, Moreno S, Rossi S: **Characterization of yeast pyruvate kinase 1 as a protein kinase A substrate, and specificity of the phosphorylation site sequence in the whole protein.** *Biochem J* 2006, **396:**117-126.

109. Portela P, Howell S, Moreno S, Rossi S: ***In vivo* and *in vitro* phosphorylation of two isoforms of yeast pyruvate kinase by protein kinase A.** *J Biol Chem* 2002, **277:**30477-30487.

110. Blazquez MA, Lagunas R, Gancedo C, Gancedo JM: **Trehalose-6-phosphate, a new regulator of yeast glycolysis that inhibits hexokinases.** *FEBS Lett* 1993, **329:**51-54.

111. Rider MH, Bertrand L, Vertommen D, Michels PA, Rousseau GG, Hue L: **6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: head-to-head with a bifunctional enzyme that controls glycolysis.** *Biochem J* 2004, **381:**561-579.

112. Zhao S, Xu W, Jiang W, Yu W, Lin Y, Zhang T, Yao J, Zhou L, Zeng Y, Li H, et al: **Regulation of cellular metabolism by protein lysine acetylation.** *Science* 2010, **327:**1000-1004.

113. Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, Tao SC, Qian J, Zhao Y, Boeke JD, et al: **Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis.** *Cell* 2009, **136:**1073-1084.

114. Oliveira AP, Sauer U: **The importance of post-translational modifications in regulating *Saccharomyces cerevisiae* metabolism.** *FEMS Yeast Res* 2012, **12:**104-117.

115. Tripodi F, Nicastro R, Reghellin V, Coccetti P: **Post-translational modifications on yeast carbon metabolism: Regulatory mechanisms beyond transcriptional control.** *Biochim Biophys Acta* 2015, **1850:**620-627.

116. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL: **Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry.** *Eur J Biochem* 2000, **267:**5313-5329.

117. Schaaff I, Heinisch J, Zimmermann FK: **Overproduction of glycolytic enzymes in yeast.** *Yeast* 1989, **5:**285-290.

118. Hauf J, Zimmermann FK, Muller S: **Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*.** *Enzyme Microb Technol* 2000, **26:**688-698.

119. Heinisch JJ: **Expression of heterologous phosphofructokinase genes in yeast.** *FEBS Lett* 1993, **328:**35-40.

120. Heinisch J, Kirchrath L, Liesen T, Vogelsang K, Hollenberg CP: **Molecular genetics of phosphofructokinase in the yeast *Kluyveromyces lactis*.** *Mol Microbiol* 1993, **8:**559-570.

121. Estevez AM, Heinisch JJ, Aragon JJ: **Functional complementation of yeast phosphofructokinase mutants by the non-allosteric enzyme from *Dictyostelium discoideum*.** *FEBS Lett* 1995, **374:**100-104.

122. de Kok S, Kozak BU, Pronk JT, van Maris AJ: **Energy coupling in *Saccharomyces cerevisiae:* selected opportunities for metabolic engineering.** *FEMS Yeast Res* 2012, **12:**387-397.

123. Benisch F, Boles E: **The bacterial Entner-Doudoroff pathway does not replace glycolysis in *Saccharomyces cerevisiae* due to the lack of activity of iron-sulfur cluster enzyme 6-phosphogluconate dehydratase.** *J Biotechnol* 2014, **171:**45-55.

124. Abbott DA, Zelle RM, Pronk JT, van Maris AJ: **Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges.** *FEMS Yeast Res* 2009, **9:**1123-1136.

125. van Maris AJ, Konings WN, van Dijken JP, Pronk JT: **Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes.** *Metab Eng* 2004, **6:**245-255.

126. van Maris AJ, Winkler AA, Porro D, van Dijken JP, Pronk JT: **Homofermentative lactate production cannot sustain anaerobic growth of engineered *Saccharomyces cerevisiae*: possible consequence of energy-dependent lactate export.** *Appl Environ Microbiol* 2004, **70:**2898-2905.

127. Abbott DA, van den Brink J, Minneboo IM, Pronk JT, van Maris AJ: **Anaerobic homolactate fermentation with *Saccharomyces cerevisiae* results in depletion of ATP and impaired metabolic activity.** *FEMS Yeast Res* 2009, **9:**349-357.

128. Smallbone K, Messiha HL, Carroll KM, Winder CL, Malys N, Dunn WB, Murabito E, Swainston N, Dada JO, Khan F, et al: **A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes.** *FEBS Lett* 2013, **587:**2832-2841.

129. Bogorad IW, Lin T-S, Liao JC: **Synthetic non-oxidative glycolysis enables complete carbon conservation.** *Nature* 2013, **502:**693-697.

130. Solis-Escalante D, Kuijpers NGA, Barrajon-Simancas N, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P: **A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism.** *Eukaryotic Cell* 2015, **14:**804-816.

131. Solis-Escalante D: **Reducing the generic complexity of glycolysis in *Saccharomyces cerevisiae*.** *PhD Thesis.* Delft University of Technology, Biotechnology; 2015.

Chapter 2

A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences

Kuijpers NGA, Solis-Escalante D,
Bosman L, van den Broek M, Pronk JT,
Daran JM, Daran-Lapujade P

Published in Microbial Cell Factories (2013) 12:47

ABSTRACT

Background | *In vivo* recombination of overlapping DNA fragments for assembly of large DNA constructs in the yeast *Saccharomyces cerevisiae* holds great potential for pathway engineering on a small laboratory scale as well as for automated high-throughput strain construction. However, the current *in vivo* assembly methods are not consistent with respect to yields of correctly assembled constructs and standardization of parts required for routine laboratory implementation has not been explored. Here, we present and evaluate an optimized and robust method for *in vivo* assembly of plasmids from overlapping DNA fragments in *S. cerevisiae*.

Results | To minimize occurrence of misassembled plasmids and increase the versatility of the assembly platform, two main improvements were introduced; i) the essential elements of the vector backbone (yeast episome and selection marker) were disconnected and ii) standardized 60 bp synthetic recombination sequences non-homologous with the yeast genome were introduced at each flank of the assembly fragments. These modifications led to a 100 fold decrease in false positive transformants originating from the backbone as compared to previous methods. Implementation of the 60 bp synthetic recombination sequences enabled high flexibility in the design of complex expression constructs and allowed for fast and easy construction of all assembly fragments by PCR. The functionality of the method was demonstrated by the assembly of a 21 kb plasmid out of nine overlapping fragments carrying six glycolytic genes with a correct assembly yield of 95%. The assembled plasmid was shown to be a high fidelity replica of the *in silico* design and all glycolytic genes carried by the plasmid were proven to be functional.

Conclusion | The presented method delivers a substantial improvement for assembly of multi-fragment expression vectors in *S. cerevisiae*. Not only does it improve the efficiency of *in vivo* assembly, but it also offers a versatile platform for easy and rapid design and assembly of synthetic constructs. The presented method is therefore ideally suited for the construction of complex pathways and for high throughput strain construction programs for metabolic engineering purposes. In addition its robustness and ease of use facilitates the construction of any plasmid carrying two genes or more.

BACKGROUND

Restriction and ligation, complemented with the creative application of PCR, has long been the universal method for gene cloning in fundamental research and metabolic engineering. However, the increasing size and complexity of today's constructs in metabolic engineering has made design and construction of plasmids by these classical techniques increasingly complicated and time consuming. Several *in vitro* techniques have been developed to deal with these issues. Methods such as SLIC [3], InFusion™ [4], and Gibson's isothermal assembly [5], enable efficient assembly of up to six overlapping DNA fragments into a plasmid. However, the efficiencies of these systems decrease at higher numbers of fragments and commercial kits are required to obtain the necessary recombinases. In contrast, *in vivo* assembly of multiple overlapping DNA fragments by homologous recombination in *Saccharomyces cerevisiae* does not exhibit these limitations [6-9]. In this method, yeast is transformed with a mixture of multiple linear DNA fragments, which assemble through homologous recombination of overlapping terminal sequences [10]. Although pioneering work in the 1980s already made use of this method to assemble circular plasmids [11], its application remained limited, probably due to the difficulties in the generation of the terminal homologous sequences required for recombination of the linear fragments. Later, *in vivo* assembly (also known as transformation associated recombination (TAR)) was used for the cloning of large DNA fragments that were resisting traditional methods based on restriction-ligation [12].

With the development of fast and cost effective chemical DNA synthesis, the method readily took off. It was shown that even single-strand 80 bp 'stitching' oligonucleotides overlapping the ends of adjacent fragments could be used to join DNA sequences by *in vivo* assembly in *S. cerevisiae* [13]. The full implication of these developments for TAR cloning was realized when Gibson *et al*. turned to *S. cerevisiae* to assemble four quarter genomes into a fully synthetic 583 kb *Mycoplasma genitalium* genome to overcome the size limitations of *in vitro* assembly resulting from the requirement of *E. coli* transformation [14]. This successful demonstration led to further research on *S. cerevisiae* as DNA assembly platform. It was subsequently demonstrated that a whole *M. genitalium* genome could be successfully assembled out of 25 overlapping DNA fragments in a single step [7]. In follow-up studies it was shown that as many as 38 single-stranded 200 bp oligonucleotides with 20 bp sequence overlaps could be incorporated into a linearized plasmid, thereby creating a whole new platform for gene synthesis [15]. This unparalleled efficiency of homologous recombination in *S. cerevisiae*, harnessed for high-efficiency *in vivo* assembly of linear DNA fragments, soon caught the interest of metabolic engineers [8,9]. Although TAR cloning showed many advantages, published versions of the method still yield false positive transformants at frequencies ranging from 10% to 80%, an aspect that has hitherto received comparatively little attention [16-18]. One of the main sources of incorrect assembly resides in a high incidence of transformants that contain re-circularized plasmid backbones, which contain all genetic elements required for selection and propagation [19,20]. To prevent backbone self-closure, selection procedures based

on dual markers and counter-selection have been proposed [20]. Recent published protocols do not adequately deal with this incorrect assembly problem and still rely on single linearized plasmid backbones that are co-transformed to *S. cerevisiae* with a number of overlapping DNA fragments. Efficiencies measured as the percentage of clones containing the desired plasmid range from 20% to 90% and are thought to depend on the length of the homologous regions and the number of fragments to be assembled [8,9,18].

The existing methods show the potential to use *in-vivo* assembly as a standard tool for assembly of large and complex DNA constructs, but two main points should be addressed to make the system more robust and suited for large scale metabolic engineering; (i) the presence of undesired subassemblies due to regeneration of the plasmid should be reduced; and (ii) *in vivo* assembly systems should be designed in such way that replacing or swapping fragments should be feasible without extensive DNA modifications.

To meet the above requirements, the aim of the present study was to reduce the incidence of incorrect plasmid assembly and to make a robust, versatile *in vivo* assembly strategy for multi-component plasmids. To this end, the concept of a single linearized vector backbone was abandoned and replaced by separated key genetic elements involved in plasmid selection and propagation. Furthermore, specially designed 60 bp synthetic homologous recombination sequences (SHR-sequences) were implemented to enhance the versatility of the method. As a proof of principle, the method was used to assemble a 21 kb plasmid from 9 overlapping fragments, using only PCR and yeast transformation. Key factors for successful and highly efficient assembly of DNA by homologous recombination in *S. cerevisiae* are discussed.

RESULTS

The use of single-fragment plasmid backbones results in frequent incorrect assembly | Current *in vivo* plasmid assembly methods in yeast use a linearized vector containing two elements essential for survival; i) an episome (centromere plus autonomously replicating sequence (CEN/ARS) or 2-micron origin) and ii) one or more selection marker genes. Presence of these 'survival elements', essential for replication and selection of the plasmid in yeast, on a single fragment, implies that re-circularization of this fragment will always generate a plasmid conferring viability to transformants in selective medium, and therefore in false positives. Circularization of such plasmid backbones can occur via two mechanisms: homologous recombination and non-homologous end joining (NHEJ) [21,22].

In the case of homologous recombination, the open end of the backbone recombines with an homologous region present in the backbone itself, which can be as short as 15 bp [23]. To estimate the frequency of false-positive transformants resulting from utilization of linearized plasmid backbones, we took plasmid backbones used in two recently reported *S. cerevisiae* based *in vivo* assembly methods [17,18]. Both methods

made use of backbones derived from the plasmid pRS416, although in one BamHI was used to linearize pRS416, leading to a 4898 bp backbone and the other used a 2064 bp backbone resulting from the digestion of pRS416 with SspI. Yeast transformations, using the same amount of linearized plasmid DNA as described in the previous studies (100 fmol), led to over 1000 and 245 ± 14 transformants with backbone DNA restricted by BamHI and the SspI respectively (Fig. 2.1). The higher number of clones obtained with BamHI could be explained by circularization of the backbone by cohesive-end ligation of the BamHI-cut DNA, while SspI leaves blunt ends [24]. Furthermore the BamHI fragment was 2.5 fold larger than the SspI fragment, leaving much more chance for circularization by internal recombination on short homologous sequences. These results showed that use of backbones obtained by restriction from standard yeast vectors are a serious factor in determining the fidelity of the system.

Separation of the survival elements is important to reduce plasmid self-closure | To enhance the versatility of the *in vivo* assembly platform, we designed specific overlapping sequences (Table 2.1). These unique synthetic 60 bp homologous recombination sequences (SHR-sequences) were obtained by randomly combining bar-code sequences used in the *Saccharomyces* Genome Deletion Project (Table 2.1) [25]. The resulting 60 bp SHR-sequences can be easily attached to any DNA fragment of interest by PCR. The SHR-sequences add versatility to the system, thereby creating a platform in which DNA cassettes can be easily interchanged and different combinations of genes can be effortlessly assembled. In the present study, the classical plasmid backbone was replaced by two separate cassettes flanked by SHR-sequences: one fragment containing the episome and one carrying the selection marker of the plasmid (Fig. 2.2). Since both elements are required for a viable clone and lack any homology to each other, two independent NHEJ events would require assembling a viable plasmid out of these fragments when co-transformed to *S. cerevisiae*. Moreover, since these fragments are flanked by SHR-sequences, interference and recombination with genomic DNA or internal regions of other assembly cassettes was expected to be less likely.

To quantify the impact of separated survival elements on the occurrence of undesired recombination events, three different cassettes were generated by PCR: (i) A *Kluyveromyces lactis URA3* cassette flanked by SHR-sequences *A* and *B* (*K.l.URA3*$_{AB}$), (ii) a *CEN6/ARS4* cassette flanked by SHR-sequences *B* and *C* (*CEN6/ARS4*$_{BC}$), and (iii) a *CEN6/ARS4* cassette flanked by SHR-sequences *F* and *G* (*CEN/ARS*$_{FG}$). A fourth cassette was obtained by linking *K.l.URA3*$_{AB}$ and *CEN6/ARS4*$_{BC}$ by fusion PCR, leading to cassette *K.l.URA3/ CEN6/ARS4*$_{AC}$. Different fragment combinations containing 100 fmol of each cassette were transformed to the *S. cerevisiae* strain CEN.PK 113-5D, which is auxotrophic for uracil. The fragment *K.l.URA3/ CEN6/ ARS4*$_{AC,}$ used to mimic a linearized plasmid backbone, was also transformed alone in yeast. As expected, this fragment, which can be circularized in one single NHEJ event, resulted in a substantial number of clones (Fig. 2.1). Still, this short fragment displayed a more than five-fold reduction in clone formation as compared to a linearized backbone (Fig. 2.1). When the overlapping survival elements *K.l.URA3*$_{AB}$

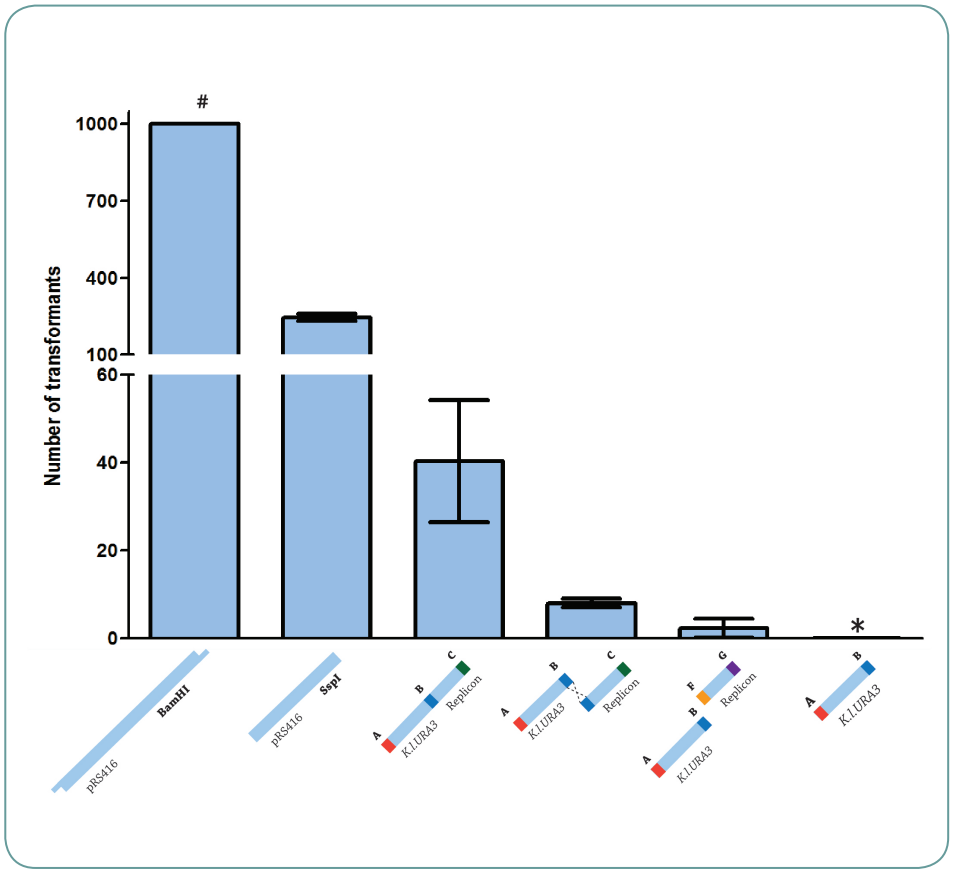


FIGURE 2.1 | Influence of the plasmid backbone structure on the in vivo assembly efficiency. The quantification of the number of transformants obtained after transformation of 100 fmol of each of the corresponding fragments sets was based on triplicate experiments and the data presented are average ± standard deviation. (#) Transformation of the pRS416 backbone linearized by BamHI (1st bar) gave so many transformants that the exact number of transformants could not be determined due to the colony density on the plates, but it exceeded 1000 transformants in all three transformations. (*) Transformation of the *K. lactis URA3* fragment only (rightmost bar) did not yield any transformants.

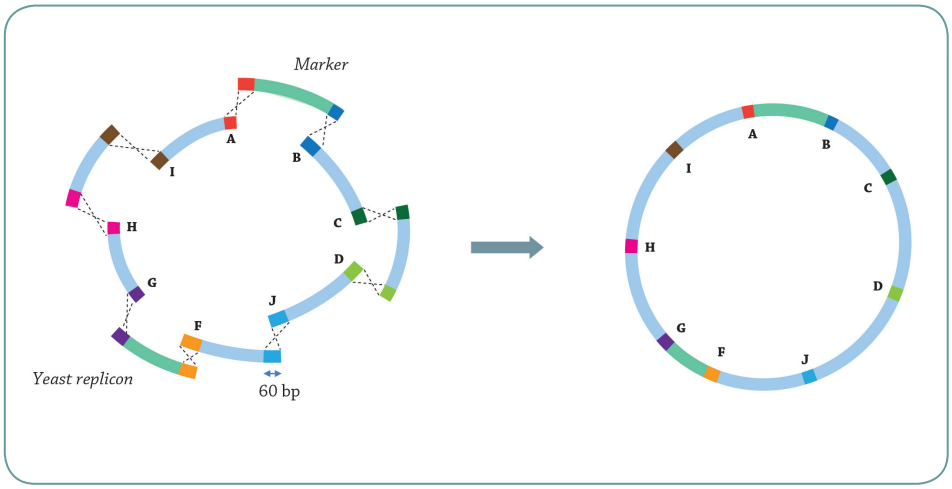


FIGURE 2.2 | Schematic representation of in vivo assembly of plasmids using 60 bp synthetic homologous recombination sequences. The green survival fragments are essential for replication and selection.

and *CEN6/ARS4*$_{BC}$ were co-transformed, a 30-fold reduction was observed in the number of clones as compared to transformation with linearized backbones (Fig. 2.1). This strong decrease can be explained by the requirement of two recombination events (if integration of the marker in chromosomal DNA is not taken into account) to generate a viable plasmid: the recombination of the separated fragments by homologous recombination and the circularization by NHEJ. When combining the non-overlapping *K.l.URA3*$_{AB}$ and *CEN6/ARS4*$_{FG}$ fragments hardly any clones (2±2) were obtained, which is consistent with the need for two recombination events via NHEJ to obtain a viable plasmid. Finally, transformation of only the *K.l.URA3*$_{AB}$ cassette did not yield any clones, showing that integration of this cassette in the genome is extremely rare (Fig. 2.1). These results support the hypothesis that separation of the survival elements on non-overlapping fragments reduces plasmid regeneration by at least 100-fold as compared to a linearized plasmid situation (Fig. 2.1). Those results enabled the design of a simple, efficient *in vivo* assembly platform (Fig. 2.2). All fragments were flanked with 60 bp SHR sequences and two survival elements, both essential for replication and selection of the plasmid, were placed physically opposite each other in the design.

High efficiency and fidelity of *in vivo* assembly of a 21 kb plasmid from nine overlapping fragments | To test the proposed system, assembly of a 21 kb plasmid from nine DNA fragments was attempted. The fragments were amplified by PCR to add the desired SHR-sequences designed for recombination of the overlapping fragments. The nine fragments consisted of two *S. cerevisiae* survival elements, an *E. coli* amplification fragment and six expression cassettes, each containing a yeast glycolytic gene fully homologous to its genomic counterpart. The yeast survival elements *K.l.URA3*$_{AB}$ and *CEN6/ARS4*$_{FG}$ were constructed as described above and the *E.coli* amplification cassette *E.coli*$_{IA}$ was obtained from pRS416 in the same way. The six glycolytic expression cassettes were amplified by PCR from genomic DNA of

TABLE 2.1 | Overlapping sequences for homologous recombination. Sequences were compared to the S. cerevisiae CEN.PK 113-7D strain [43] whole genome shotgun contigs (Accession number: PRJNA52955) by BLASTN analysis [44] using BLOSUM 62 substitution matrix. The statistical significance of matches found (Blast score and E-value) are reported.

		Comparison to *S.cerevisiae* 113-7D genome		
SHR-sequence	**Sequence 5' à 3'**	**Blast score**	**E-value**	**GC content**
Optimized SHR-sequences				
A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCCAGATCATCAATAGGCAC	28.3	1.3	50.0
B	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACCAGCCTAAGAATGTTCAAC	30.1	0.37	48.3
C	ACGTCTCACGGATCGTATATGCCGTAGCGACAATCTAAGAACTATGCGAGGACACGCTAG	26.5	4.5	50.0
D	ACGCATCTACGACTGTGGGTCCCGTGGAGAAATGTATGAAACCCTGTATGGAGAGTGATT	28.3	1.3	48.3
F	CATACGTTGAAACTACGGCAAAGGATTGGTCAGATCGCTTCATACAGGGAAAGTTCGGCA	28.3	1.3	46.7
G	GCCAGAGGTATAGACATAGCCAGACCTACCTAATTGGTGCATCAGGTGGTCATGGCCCTT	28.3	1.3	51.7
H	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGAATTGCTGAGAACCCGTGAC	30.1	0.37	48.3
I	TATTCACGTAGACGGATAGGTATAGCCAGACATCAGCAGCATACTTCGGGAACCGTAGGC	28.3	1.3	50.0
J	GGCCGTCATATACGCGAAGATGTCCAAGCAGGTAGAACACATAGTCTGAGCATCTCGTCG	26.5	4.5	51.7
Endogenous sequences				
A#	GTCGACAACCCTTAATATAACTTCGTATAATGTATGCTATACGAAGTTATTAGGTCTAGA	28.3	1.3	33.3
B#	GAGTGTTTAGAACATAATCAGTTTATCCATGGTCTATCTCTTCTTGTCGCTTTTTCTCCT	28.3	1.3	36.7
C#	TTAATTTTAAATTTTTTTGGTAGTAAAAGATGCTTATATAAGGATTTCGTATTTATTGTT	109	5e-25	18.3
D#	TAATATTTTTTCTTTTGAAAGTACTACCCACATCCGAACATTGCCACTTACATAGCGATG	109	5e-25	35.0
J#	GAACAAAGTATTTAACGCACATGTATAAATATTGTATTAAAAGGGTACCTTTATAAATAT	109	5e-25	23.3
F#	GAATAGTCTTTACACCCACAGTTTTTCGTGTGGCAGTTACTATATATTAGTAGGATATTC	109	5e-25	35.0
G#	CTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTT	30.1	0.37	35.0
H#	ACGTAGGATTATTATAACTCAAAAAAATGGCATTATTCTAAGTAAGTTAAATATCCGTAA	109	5e-25	25.0
I#	TTGGCAATTTTTTGCTCTTCTATATAACAGTTGAAATTTGAATAAGAACATCTTCTCAAA	109	5e-25	26.7

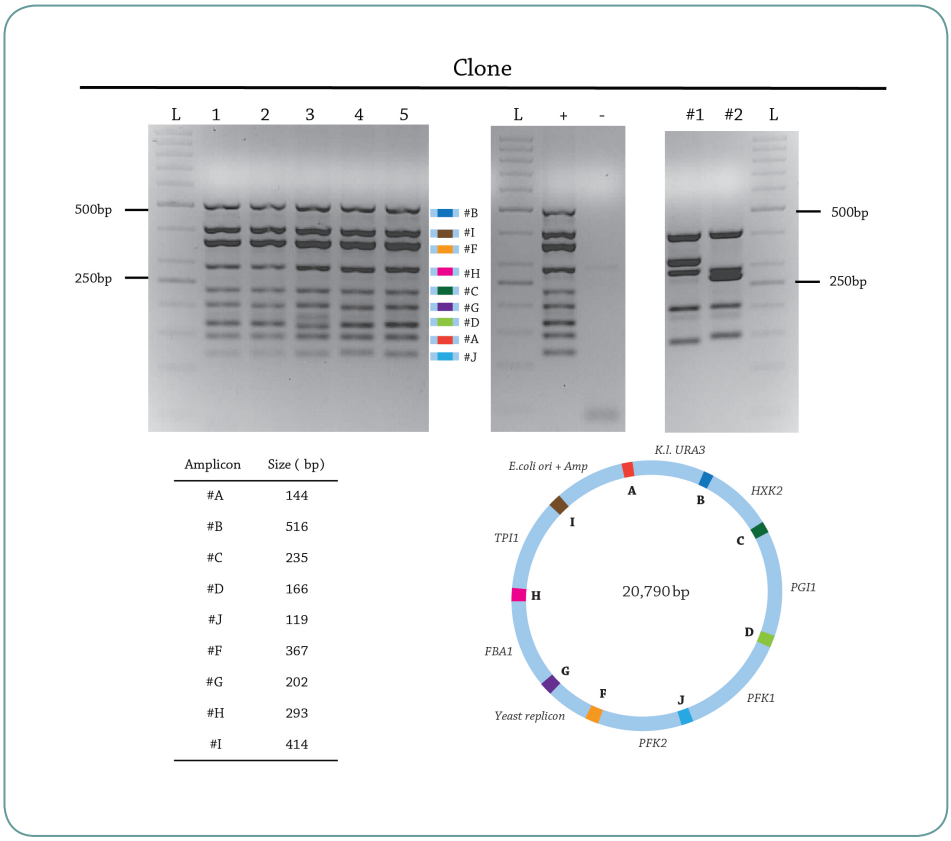


FIGURE 2.3 | Multiplex PCR analysis of clones obtained after co-transformation of nine overlapping fragments in *S. cerevisiae* and of clones obtained from control experiments. The multiplex primer mix was designed to produce nine amplicons, ranging in size from 119-516 bp. Each amplicon covered a specific SHR. Amplicons were separated on a 2% agarose gel by electrophoresis. Lanes 1-5 represent clones obtained after transformation of a full set of fragments. As a negative control genomic DNA of CEN.PK 113-5D was used (-); The later fully analyzed plasmid pUDC074 is added as a positive control (+). All nine bands were obtained in clones 1-5. The clones obtained from transformation of an incomplete mix show a completely different multiplex pattern (#1 and #2). In the lanes labeled 'L' a 50 bp Gene Ruler ladder was loaded; sizes are indicated. In total 40 clones were analyzed and 38 multiplex patterns matched the positive control.

S. cerevisiae strain CEN.PK113-7D. Based on concentration and size measurements, ca. 100 fmol of the survival elements and 200 fmol of each of the other fragments were pooled and transformed to ca. 10^8 yeast cells. Several controls were performed to estimate the efficiency and reliability of the technique, including: (i) transformation of the marker fragment alone, to estimate the frequency of integration of the fragment into the yeast genome, and (ii) transformation with a control mix in which a single fragment was omitted to estimate the amount of miss-assemblies. After incubation for three days on selective medium, over 1000 clones were obtained for the cells transformed with the complete set of fragments. Conversely, no clones were obtained

for the cells transformed with the marker fragment only. The cells transformed with the incomplete control mix yielded six clones. High-fidelity assembly was confirmed by multiplex PCR analysis of 40 clones, randomly picked after two independent transformations. PCR with primers specifically designed to cover the SHR-sequences, produced the expected nine amplicons for 38 out of the 40 clones (Fig. 2.3). Clones obtained by transformation with the incomplete control pool displayed aberrant multiplex PCR profiles. These results provided a strong indication for the presence of the correct assembly in 38 of the 40 tested clones, which corresponded to an extremely high efficiency of *in vivo* assembly of 95% correct assemblies.

To ensure that correct multiplex profiles indeed reflected the desired assembly, a single clone was randomly selected and its plasmid was isolated and named pUDC074. After amplification in *E. coli*, the sequence of pUDC074 was determined by Illumina next generation sequencing and *de novo* sequence assembly. Sequence analysis confirmed the correct assembly of the plasmid, thereby supporting the practicality of the multiplex PCR approach to screen for correct assemblies. Among the 21000 bases of pUDC074, as few as four nucleotides were different from the original design. Of these four mutations, three single base-pair deletions were localized in the SHR-sequences. These mutations could result from erroneous homologous recombination in yeast itself, but more likely these errors were introduced prior to assembly during primer synthesis. Although HPLC purified primers were used in this study to attach the SHR-sequences, the use of PAGE purified oligonucleotides could reduce mutations resulting from synthesis. However, since the SHR-sequences are not coding, potential mutations in these overlaps are of minor concern for functionality of the assembly. The single mutation found outside the SHR-sequences was in fragment *PFK2* and caused an amino acid substitution in Pfk2 (N822K). Finally, to determine whether the presence of SHR-sequences could have an effect on the biological functionality of the proteins encoded by the six plasmid-borne glycolytic genes, a complementation study was performed (Fig. 2.4a and b). In *S. cerevisiae*, deletion of *PGI1* [26], *TPI1* [27] and *FBA1* [28] leads to lethality and deletion of *PFK1* or *PFK2* results in severe growth impairment [29] when cultivated on glucose. The heterozygous diploids of *PGI1*, *TPI1*, *FBA1*, *PFK1* and *PFK2* [30] (Table 2.2) were therefore transformed with the plasmid pUDC074 and subsequently incubated in sporulation medium. After tetrad dissection, the spores containing the deletion could be selected for by the G418 resistance marker (*AgTEF2*$_p$*-nptII-AgTEF2*$_t$) while the presence of the plasmid was ensured by selecting for the *K.l.URA3* marker. The growth of spores in the absence of uracil and in the presence of G418 demonstrated the ability of the plasmid-borne *PGI1*, *TPI1*, *FBA1*, *PFK1* and *PFK2* genes to complement the deletion of the corresponding chromosomal gene (Fig. 2.4a). Functionality of the assembled *HXK2* was demonstrated by restoration of growth on glucose of a glucose phosphorylation-deficient strain (IMX188, *hxk1hxk2glk1*; Table 2.2) upon transformation with pUDC074. (Fig. 2.4b). These results demonstrated highly efficient assembly of a 21 kb plasmid out of nine fragments and that the presence of SHR-sequences have no detectable impact on the functionality of the assembled plasmid.

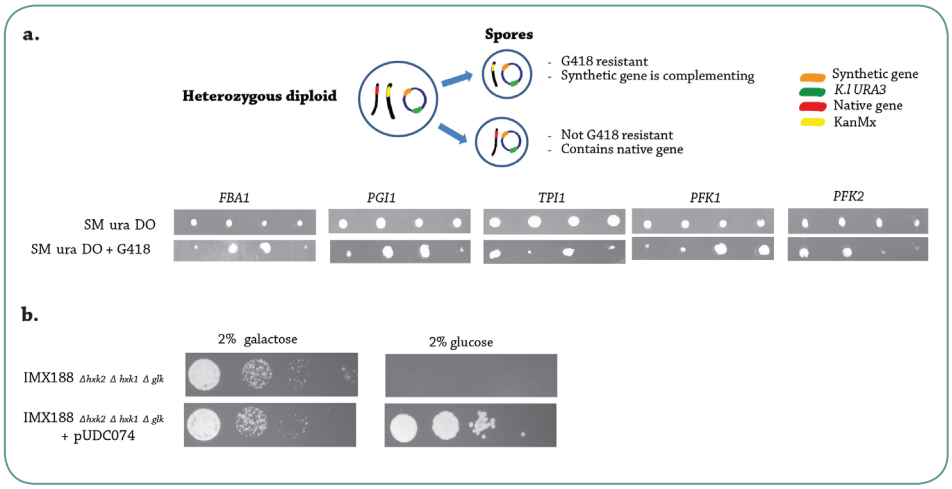


FIGURE 2.4 | Complementation studies of six glycolytic gene deletions with pUDC074. a) On synthetic medium complemented with all amino acids except uracil (SM ura DO) all isolated spores from a single tetrad of a heterozygous diploid of the indicated gene can grow, proving the presence of the plasmid in all spores. Those spores were replicated to SM ura DO medium complemented with G418. Only spores containing the chromosomal deletion of the represented gene can grow due to the selection for the KanMX marker. Spores growing on both media confirmed the presence of a functional copy of the gene on pUDC074. b) Complementation study of HXK2 with pUDC074 in a strain incapable of phosphorylating glucose. Spot plates are shown (10,000, 1000,100,10 cells/µl). Introduction of the plasmid restored the ability to grow on glucose as the sole carbon source.

Optimization of SHR-sequences is not critical for high efficiency of *in vivo* assembly | In published *in vivo* assembly studies, the design of the overlapping sequences across which recombination occurs does not receive much attention and often just depends on the ends of the assembled fragments. Therefore, overlaps are likely to differ in GC content and binding capacities could vary strongly between constructs. Moreover, when fragments are specifically designed for metabolic engineering in yeast, overlapping ends often share sequence identity with chromosomal sequences, since expression cassettes typically end with yeast promoter and terminator sequences. The previous experiments demonstrated that combining the separation of essential elements and the use of SHR-sequences resulted in high efficiency of *in vivo* assembly. To more precisely evaluate the contribution of the customized SHR-sequences on this high efficiency, *in vivo* assembly was also performed with fragments harboring non-optimized overlapping sequences homologous to *S. cerevisiae* genomic DNA with relatively low G/C content (18.3% to 36.7%, Table 2.1). Co-transformation of cassettes carrying the same glycolytic genes as in the previous experiment, but with endogenous overlapping sequences, resulted in similar colony numbers as when SHR-sequences were used. In addition, analysis of 10 randomly picked transformants by multiplex PCR did not reveal significant differences in the fidelity of *in vivo* assembly. These results demonstrated that optimization of the 60 bp overlapping sequences is not required to obtain high *in vivo* assembly efficiencies with the platform described in this study.

DISCUSSION

Although uncovered nearly three decades ago [11], the high efficiency of *S. cerevisiae* homologous recombination is only beginning to reveal its full potential for the assembly of large DNA constructs. *In vivo* assembly in yeast is predicted to have a large impact on laboratory practice, ranging from simple plasmid construction to engineering of complex pathways via automated high-throughput strain construction [7,8]. Despite those promising prospects, *in vivo* assembly has not yet become a standard technique in most academic laboratories. This offers unique possibilities for standardization and, simultaneously, for further optimization. While reported efficiencies of correct assembly of larger (over 15 kb) DNA constructs do not exceed 70%, efficiencies of 95% were reached in the present work for the assembly of a 21 kb construct [8,9]. Physical separation of essential elements of the plasmid backbone contributed to a strong reduction of the frequency of plasmid mis-assembly.

The high efficiency obtained with relatively short 60 bp overlaps demonstrates that, in contrast to practices and claims from recent reports [18,31], longer overlaps are not essential for efficient *in vivo* assembly. This conclusion is supported by earlier studies in which 30 to 60 bp homologous sequences were shown to lead to high recombination efficiencies in *S. cerevisiae* [30,32]. Although we anticipated that the high GC content, and therefore optimal binding properties, of the optimized SHR-sequences contributed to the high assembly efficiency obtained with the present platform, our results clearly indicate that the nature of these SHR-sequences is not an essential factor for achieving efficiencies above 90%. Therefore the increase in efficiency compared to previous studies essentially originates from the implementation of a backbone-free approach, and more specifically in the physical separation of the genetic elements on a plasmid assembly that are essential for its propagation and selection in the recipient yeast cell. Earlier studies have shown that reassembly of the backbone could make up to 95% of the obtained clones [20]. Placing the essential yeast elements on different fragments flanked by SHR-sequences and co-transforming them to *S. cerevisiae* reduced erroneous assemblies by plasmid regeneration by a factor of at least 100 (Fig. 2.1), thereby substantially increasing the fidelity of *in vivo* assembly. Other yet unidentified factors, such as yeast strain-dependent efficiency of homologous recombination, could also have contributed to the high efficiency of assembly in the present platform and should be considered for further development of the method.

A potential downside of the proposed system was the inherent increase in the number of fragments to be transformed. However, in a standardized transformation protocol, assembly of a plasmid from 16 fragments, still generated hundreds of clones (data not shown), which is more than sufficient for metabolic engineering purposes. This result indicates that increasing the number of fragments is not a serious limitation and that use of two survival elements instead of one single backbone does not have a high impact on the overall transformation efficiency.
The considerable potential of *S. cerevisiae* for pathway assembly has been recently proposed [8,9] for short pathways up to eight genes. To obtain highly productive and balanced synthetic pathways it is evident that finding the optimal combination

of (heterologous) genes and expressing them at the right levels is essential [33]. Combinatorial approaches are therefore necessary and hundreds to thousands of constructs carrying different alleles expressed behind various promoters will have to be constructed in high throughput platforms. The presented approach will facilitate these strain construction programs, since assembly efficiency and robustness are bound to be a key variable for high throughput strain construction, as it determines how many clones will be correctly assembled. The SHR-sequences can be designed and tested for these requirements and thus contribute to the development of these systems. Moreover, the use of SHR-sequences offers an unprecedented versatility. It is a goal of synthetic biology to create versatile platforms with libraries of interchangeable parts and pieces, as exemplified by the Biobricks concept [34]. Using the SHR-sequences libraries of standardized parts, ranging from individual gene expression cassettes to fragments that carry entire pathways, can be generated and used for combinatorial assembly and subsequent screening for high-performing strains. In contrast to existing *in vivo* assembly approaches, no extensive re-designing has to be performed. Libraries of survival elements, genes and SHR-sequences will enable easy *in silico* design, straightforward *in vitro* synthesis of the fragments by PCR and efficient *in vivo* assembly.

While implementing *in vivo* assembly in our research, we have been surprised by its simplicity, ease of implementation and high efficiency. Within two years, *in vivo* assembly has almost completely replaced standard restriction/ligation protocols for construction of plasmids carrying two or more genes in our laboratory, thereby greatly accelerating strain construction and opening possibilities for strain modification that previously would have been deemed too complicated. Looking into the future, *S. cerevisiae* has the potential to be developed into an even more powerful platform. Similarly to popular *E. coli* strains, which have been extensively optimized to become extremely efficient hosts for plasmid transformation and replication, specific modifications of the yeast chassis, such as removal of the NHEJ machinery or enhancement of DNA uptake, could further extend the efficiency and fidelity of the *in vivo* assembly method.

CONCLUSIONS

The presented method for *in-vivo* assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* delivers a substantial improvement in terms of fidelity and flexibility as compared to existing methods. This improvement, achieved by replacing the plasmid backbone by standardized survival elements and by implementing the use of standardized 60 bp synthetic recombination sequences, was demonstrated by the correct assembly of a 21 kb plasmid from nine fragments with an efficiency of 95%. Ideal for the assembly of large constructs, the presented approach delivers a straightforward method for the assembly of any DNA construct carrying two or more genes and can be implemented in any molecular biology laboratory. It is our hope that the present work will contribute to standardization of *in vivo* assembly of plasmids, artificial chromosomes and synthetic genomes in *S. cerevisiae*.

METHODS

Strains, media and DNA templates | The *S. cerevisiae* strains used in this study are listed in Table 2.2. Cultures for transformation were cultivated in complex media containing 10 $g \cdot l^{-1}$ Bacto yeast extract, 20 $g \cdot l^{-1}$ Bacto peptone and 20 $g \cdot l^{-1}$ glucose as carbon source. Synthetic medium (SM) contained per liter of demineralized water 2% agar (w/v) 5g $(NH_4)_2SO_4$, 3g KH_2PO_4, 0.5 g $MgSO_4.7 \cdot H_2O$, and trace elements according to [35]. Vitamins were added after heat sterilization of the medium at 120

TABLE 2.2 | Strains used in this study

Strain	Relevant Genotype	Source
CEN.PK113-7D	*MATa MAL2-8c SUC2*	[43,45]
CEN.PK113-5D	*MATa ura3-52 MAL2-8c SUC2*	[45]
IMX188	*MATa ura3-52 his3-Δ1 leu2-3,112 MAL2-8c SUC2 glk1Δ::LoxPhxk1Δ::LoxP hxk2Δ::LoxP-KanMX-LoxP*	This study
IMX214	*MATa ura3-52 his3-Δ1 leu2-3,112 MAL2-8c SUC2 glk1Δ::LoxP hxk1Δ::LoxP hxk2Δ::LoxP-KanMX-LoxP + PUDC074*	This study
BY4743-Y23336	*MAT a/a; ura3 Δ 0/ura3 Δ 0; pgi1Δ::KanMX/PGI1*	Euroscarf[a]
BY4743-Y25893	*MAT a/a; ura3 Δ 0/ura3 Δ 0; pfk1Δ::KanMX/PFK1*	Euroscarf[a]
BY4743-Y20791	*MAT a/a; ura3 Δ 0/ura3 Δ 0; pfk2 Δ::KanMX/PFK2*	Euroscarf[a]
BY4743-Y23986	*MAT a/a; ura3 Δ 0/ura3 Δ 0; tpi1 Δ::KanMX/TPI1*	Euroscarf[a]
BY4743-Y24909	*MAT a/a; ura3 Δ 0/ura3 Δ 0; fba1Δ::KanMX/FBA1*	Euroscarf[a]
Spore from Y23336 + pUDC074	*ura3 Δ 0; pgi1::kanMX pUDC074 (K.L.URA3(pUG72)TPI HXK2 FBA1 PGI PFK1 PFK2)*	This study
Spore from Y25893 + pUDC074	*ura3 Δ 0; pfk1::kanMX pUDC074 (K.L.URA3(pUG72)TPI HXK2 FBA1 PGI PFK1 PFK2)*	This study
Spore from Y20791 + pUDC074	*ura3 Δ 0; pfk2::kanMX pUDC074 (K.L.URA3(pUG72)TPI HXK2 FBA1 PGI PFK1 PFK2)*	This study
Spore from Y24909 + pUDC074	*ura3 Δ 0; tpi1::kanMX pUDC074 (K.L.URA3(pUG72)TPI HXK2 FBA1 PGI PFK1 PFK2)*	This study
Spore from Y23986 + pUDC074	*ura3 Δ 0; fba1::kanMX pUDC074 (K.L.URA3(pUG72)TPI HXK2 FBA1 PGI PFK1 PFK2)*	This study

°C for 20 min. Glucose or galactose were separately sterilized at 110 °C and added to a final concentration of 20 $g \cdot l^{-1}$. When required, the medium was supplemented with appropriate amounts of auxotrophic requirements [36]. Sporulation medium contained, per liter of demineralized water, 10 g potassium acetate, 1 g Bacto yeast extract and 0.5 g glucose. To rescue strain auxotrophy 10 mg leucine, 5 mg histidine, 5 mg lysine and 5 mg methionine were added prior to heat sterilization. Solid medium was prepared by adding 2% (w/v) agar to the media prior to heat sterilization. Spores were grown on solid SM supplemented with Yeast Synthetic Drop-out Medium supplements without uracil (Sigma, St Louis, MO) and replica plated on the same medium supplemented with G418 (200 $mg \cdot l^{-1}$).

Plasmids pUG72 [37] and pRS416 [38] were maintained in *E.coli* DH5α and isolated with the GenElute™ Plasmid Miniprep Kit (Sigma). Genomic DNA was isolated from *S. cerevisiae* 113-7D using the Qiagen 100/G kit (Qiagen, Hilden, Germany).

Production of DNA fragments and transformation | Fragments for *in vivo* assembly were obtained from either genomic or plasmid template DNA by extension PCR using Phusion® Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA). Primers were HPLC purified (Sigma, St Louis, MO) and are given in table 2.3. To improve the PCR efficiency, we modified the conditions recommended by the supplier, by decreasing the primer concentration from 500 nM to 200 nM and increasing the Phusion™ Hot Start High Fidelity polymerase concentration from 0.02 $U.\mu l^{-1}$ to 0.03 $U.\mu l^{-1}$. All other conditions were chosen according to standard manufacturer instructions. The primers were designed in such a way that the annealing temperature was >65 °C to minimize non-specific product formation caused by false priming. Subsequently, the amplified fragments were concentrated by chromatography using Vivacon® 500 spin columns (Sartorius Stedim, Aubagne, France). Conversely, fragments amplified from plasmid templates were submitted to an extra purification step by gel extraction to avoid contamination of the fragments by the linearized template plasmid and the ensuing formation of false positive clones. The fragments $CEN6/ARS4_{FG}$ and $E.coli_{IA}$ were amplified from pRS416. Fragment $K.l.URA3_{AB}$ was amplified from pUG72. Fragment $K.l.URA3/\ CEN6/ARS4_{AC}$ was obtained by fusion PCR from fragments $K.l.URA3_{AB}$ and $CEN6/ARS4_{BC}$ using primers Fus1 and Fus2. The fragments containing the glycolytic genes were all amplified from CEN.PK113-7D genomic DNA. DNA concentrations were measured by the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and 200 fmol of each glycolytic gene cassette and the $E.coli_{IA}$ fragment were pooled with 100 fmol of the $K.l.URA3_{AB}$ and $CEN6/ARS4_{FG}$ fragments in a final volume of 50 µl. A control pool lacking the *PFK2* cassette was created in the same way. Both pools were transformed to *S. cerevisiae* strain CEN.PK 113-5D using the lithium acetate protocol [39]. After transformation cells were selected on synthetic medium for 3-4 days at 30˚C.

Analysis by multiplex PCR | Colonies were randomly picked and incubated overnight in appropiate medium maintaining selection pressure for the plasmid. The assemblies were isolated from 1 ml of exponentially growing culture using the GenElute™ Plasmid Miniprep Kit by adding an extra step to the supplied protocol; after harvesting the cells by centrifugation, the pellet was resuspended in 200 µl resuspension buffer supplemented with 3 µl 1000 $U \cdot ml^{-1}$ zymolyase (Amsbio, Abingdon, United Kingdom) and incubated for 30 min at 37 ˚C to digest the cell walls of the yeast cells. Further steps were performed as described by the manufacturer's recommendations. Multiplex PCR was performed with DreamTaq PCR Master Mix (2x) (Thermo Fisher Scientific). Primers were used at a concentration of 150 nM and given in Table 3. Cycling parameters were 94˚C for 3 min, then 35 cycles of 94 ˚C for 30 s, 55˚C for 90 s, and 72 ˚C for 60 s, followed by a 10 min incubation at 72 ˚C. Of each reaction 10 µl was loaded on a 2% agarose gel and gel electrophoresis was performed in 0.5x TBE buffer at 120 V for 40 min.

DNA Isolation and sequencing | One positive assembly, as determined by multiplex PCR, was transformed to Electro Ten-Blue Electro-Competent cells (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. From a resulting clone, plasmid DNA was isolated and analyzed by multiplex PCR. This isolated plasmid was named pUDC074. To isolate enough plasmid DNA for sequencing the plasmid was amplified in a 100 ml *E.coli* cell culture and extracted using the method of Birnboim et al. [40] The resulting plasmid was further purified using the Zyppy Miniprep kit (Zymoresearch, Irvine, CA).

For the sequencing of pUDC074, a library of 250-bp insert was constructed and paired end sequenced (100 base pair reads) using an Illumina HISeq 2000 sequencer (Illumina, Eindhoven, The Netherlands) were provided by Baseclear BV (Leiden, The Netherlands), generating 2 million reads. A subset of 4000 randomly picked reads, which represents a 20-fold coverage of the plasmid sequence was de novo assembled using IDBA [41]. IDBA was used with the following parameters: i) the paired end information was not used to scaffold the contig and ii) iterations were performed with k-mers ranged from kmin=21, kmax=99. One contig with a size of 20.8 kbase pair was assembled in which the ends were duplicated and could be merged into a circular sequence. Finally, the assembled contig was aligned to the in silico designed pUDC074 plasmid sequence using Clustal X in Clone Manager 9 (Sci-Ed Software, Cary, NC).

TABLE 2.3 | Primers used in this study

Primers	Purification	Sequence 5' --> 3'
		To add SHR-sequences
E.coli Rv +A	HPLC	GTGCCTATTGATGATCTGGCGGAATGTCTGCCGTGCCATAGC CATGCCTTCACATATAGTTGCGCGGAACCCCTATTTG
E.coli Fw +I	HPLC	TATTCACGTAGACGGATAGGTATAGCCAGACATCAGCAGCAT ACTTCGGGAACCGTAGGCGAGAGGCGGTTTGCGTATTGG
TPI1 Rv +H	HPLC	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGA ATTGCTGAGAACCCGTGACTAGTGTGAGCGGGATTTAAACTGTG
TPI1 Fw +I	HPLC	GCCTACGGTTCCCGAAGTATGCTGCTGATGTCTGGCTAT ACCTATCCGTCTACGTGAATAGCGAAAATGACGCTTGCAGTG
FBA1 Rv +H	HPLC	GTCACGGGTTCTCAGCAATTCGAGCTATTACCGATGATGGCTGA GGCGTTAGAGTAATCTAAAATCTCAAAAATGTGTGGGTCATTACG
FBA1 Fw +G	HPLC	GCCAGAGGTATAGACATAGCCAGACCTACCTAATTGGTGCAT CAGGTGGTCATGGCCCTTAGTGCATGACAAAAGATGAGCTAGG
Cen6 Ars4 Rv +G	HPLC	AAGGGCCATGACCACCTGATGCACCAATTAGGTAGGTCT GGCTATGTCTATACCTCTGGCGACGGATCGCTTGCCTGTAAC
Cen6 Ars4 Fw +F	HPLC	CATACGTTGAAACTACGGCAAAGGATTGGTCAGATCGCTTCAT ACAGGGAAAGTTCGGCAGTGCCACCTGGGTCCTTTTC
Cen6 Ars4 Rv +B	HPLC	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGAC CAGCCTAAGAATGTTCAACGTGCCACCTGGGTCCTTTTC
Cen6 Ars4 Fw +C	HPLC	CTAGCGTGTCCTCGCATAGTTCTTAGATTGTCGCTACGGC ATATACGATCCGTGAGACGTGACGGATCGCTTGCCTGTAAC
PFK2 Rv +F	HPLC	TGCCGAACTTTCCCTGTATGAAGCGATCTGACCAATCCTTTG CCGTAGTTTCAACGTATGATAGCCATTCTCTGCTGCTTTGTTG
PFK2 Fw +J	HPLC	GGCCGTCATATACGCGAAGATGTCCAAGCAGGTAGAACACATAG TCTGAGCATCTCGTCGGAGATCCGAGGGACGTTTATTGG
PFK1 Rv +D	HPLC	ACGCATCTACGACTGTGGGTCCCGTGGAGAAATGTATGAAACCC TGTATGGAGAGTGATTTCGAGATTCCTCAATCCATACACCATTATAG
PFK1 Fw +J	HPLC	CGACGAGATGCTCAGACTATGTGTTCTACCTGCTTGGACAT CTTCGCGTATATGACGGCCTGTCGTCTTCGTGAACCATTGTC
PGI1 Rv +D	HPLC	AATCACTCTCCATACAGGGTTTCATACATTTCTCCACGGGAC CCACAGTCGTAGATGCGTCTGAAGAAGGCATACTACGCCAAG
PGI1 Fw +C	HPLC	ACGTCTCACGGATCGTATATGCCGTAGCGACAATCTAAGAAC TATGCGAGGACACGCTAGTTCGCGACACAATAAAGTCTTCACG
HXK2 Rv +C	HPLC	CTAGCGTGTCCTCGCATAGTTCTTAGATTGTCGCTACGGCATATACGA TCCGTGAGACGTGCAAGAGAAAAAAACGAGCAATTGTTAAAAG
HXK2 Fw +B	HPLC	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACCAG CCTAAGAATGTTCAACGACGGCACCGGGAAATAAACC
URA3K.l. Rv +B	HPLC	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGCAT CGTCCTCTCGAAAGGTGCTCAGAAGCTCATCGAACTGTCATC
URA3K.l. Fw + A	HPLC	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCC AGATCATCAATAGGCACGATCCCAATACAACAGATCACGTGATC

Primers	Purification	Sequence 5' --> 3'
FUS1	HPLC	ACTATATGTGAAGGCATGGCTATGG
FUS2	HPLC	CTAGCGTGTCCTCGCATAGTTC
Amp-rv + A-ctrl	HPLC	TCTAGACCTAATAACTTCGTATAGCATACATTATACGAAGTTAT ATTAAGGGTTGTCGACTGCGCGGAACCCCTATTTG
Amp-fw + I-ctrl	HPLC	TTGGCAATTTTTTGCTCTTCTATATAACAGTTGAAATTTGAAT AAGAACATCTTCTCAAAGAGAGGCGGTTTGCGTATTGG
CEN/6ARS4-fw + F-ctrl	HPLC	GAATAGTCTTTACACCCACAGTTTTTCGTGTGGCAGTTACTAT ATATTAGTAGGATATTCGTGCCACCTGGGTCCTTTTC
CEN6ARS4-rv + G-ctrl	HPLC	AAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGA TAATAATGGTTTCTTAGGACGGATCGCTTGCCTGTAAC
FBA1-fw + G-ctrl	HPLC	CTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG CGTATCACGAGGCCCTTAGTGCATGACAAAAGATGAGCTAGG
FBA1-rv + H-ctrl	HPLC	TTACGGATATTTAACTTACTTAGAATAATGCCATTTTTTTGAGT TATAATAATCCTACGTAAAATCTCAAAAATGTGTGGGTCATTACG
HXK2-fw + B-ctrl	HPLC	GAGTGTTTAGAACATAATCAGTTTATCCATGGTCTATCTCTTCT TGTCGCTTTTTCTCCTGACGGCACCGGGAAATAAACC
HXK2-rv + C-ctrl	HPLC	AACAATAAATACGAAATCCTTATATAAGCATCTTTTACTACCAAAAA AATTTAAAATTAAGCAAGAGAAAAAAACGAGCAATTGTTAAAAG
K.l.URA3 -fw + A-ctrl	HPLC	GTCGACAACCCTTAATATAACTTCGTATAATGTATGCTATACGAAG TTATTAGGTCTAGAGATCCCAATACAACAGATCACGTGATC
K.l.URA3-rv + B-ctrl	HPLC	AGGAGAAAAAGCGACAAGAAGAGATAGACCATGGATAAACTGAT TATGTTCTAAACACTCCTCAGAAGCTCATCGAACTGTCATC
PFK1-fw + J-ctrl	HPLC	ATATTTATAAAGGTACCCTTTTAATACAATATTTATACATGTGCGT TAAATACTTTGTTCTGTCGTCTTCGTGAACCATTGTC
PFK1-rv + D-ctrl	HPLC	TAATATTTTTTCTTTTGAAAGTACTACCCACATCCGAACATTGC CACTTACATAGCGATGTCGAGATTCCTCAATCCATACACCATTATAG
PFK2-fw + J-ctrl	HPLC	GAACAAAGTATTTAACGCACATGTATAAATATTGTATTAAAAGGG TACCTTTATAAATATGAGATCCGAGGGACGTTTATTGG
PFK2-rv + F-ctrl	HPLC	GAATATCCTACTAATATATAGTAACTGCCACACGAAAAACTGT GGGTGTAAAGACTATTCATAGCCATTCTCTGCTGCTTTGTTG
PGI-fw + C-ctrl	HPLC	TTAATTTTAAATTTTTTTGGTAGTAAAAGATGCTTATATAAG GATTTCGTATTTATTGTTTTCGCGACACAATAAAGTCTTCACG
PGI-rv + D-ctrl	HPLC	CATCGCTATGTAAGTGGCAATGTTCGGATGTGGGTAGTACTTTC AAAAGAAAAAATATTACTGAAGAAGGCATACTACGCCAAG
TPI-fw + I-ctrl	HPLC	TTTGAGAAGATGTTCTTATTCAAATTTCAACTGTTATATAGAAG AGCAAAAAATTGCCAAGCGAAAATGACGCTTGCAGTG
TPI-rv + H-ctrl	HPLC	ACGTAGGATTATTATAACTCAAAAAAATGGCATTATTCTAAG TAAGTTAAATATCCGTAATAGTGTGAGCGGGATTTAAACTGTG

Primers	Purification	Sequence 5' --> 3'
		For multipex PCR
A Ctrl Fw	Desalted	AAATAAACAAATAGGGGTTCCGC
A Ctrl Rv	Desalted	GCAACACTCACTTCAACTTCATC
B Ctrl Fw	Desalted	TTACCACCATCCAATGCAGAC
B Ctrl Rv	Desalted	ACGGAATAGAACACGATATTTGC
C Ctrl Fw	Desalted	TCACGGGATTTATTCGTGACG
C Ctrl Rv	Desalted	GCGTCCAAGTAACTACATTATGTG
D Ctrl Fw	Desalted	ACTCGCCTCTAACCCCACG
D Ctrl Rv	Desalted	ACGGACTATAATGGTGTATGGATTG
J Ctrl Fw	Desalted	GCTTAATCTGCGTTGACAATGG
J Ctrl Rv	Desalted	CAATAAACGTCCCTCGGATCTC
F Ctrl Fw	Desalted	GACGCCATTTGGAACGAAAAAAAG
F Ctrl Rv	Desalted	ATAGCACGTGATGAAAAGGAC
G Ctrl Fw	Desalted	GCGTGTAAGTTACAGGCAAGC
G Ctrl Rv	Desalted	GCTCTTTTCTTCTGAAGGTCAATG
H Ctrl Fw	Desalted	GTTACGTGCTCAGTTGTTAGATATG
H Ctrl Rv	Desalted	GCAGAAGTGTCTGAATGTATTAAGG
I Ctrl Fw	Desalted	TGAGCCACTTAAATTTCGTGAATG
I Ctrl Rv	Desalted	GCCTTTGAGTGAGCTGATACC

Complementation studies | The required heterozygous deletion mutants (EUROSCARF, Frankfurt, Germany) were transformed with pUDC074 and clones were screened by multiplex PCR for presence of the plasmid. For each heterozygous deletion mutant a clone containing the plasmid was transferred to solid sporulation medium and incubated for 5 days. Tetrads dissection was performed as described previously [42] using a MSM 400 micromanipulator (Singer instrument, Watchet, United Kingdom). Digested asci were plated on URA drop-out medium to rescue auxotrophy and to select for spores containing the plasmid. After incubation of 3 days, colonies were replica plated on synthetic URA drop-out medium + G418 to select for presence of the KanMX marker. Plates were incubated for 2 days and checked for growth. For complementation of HXK2, S. cerevisiae strain IMX188 was transformed with pUDC074. IMX188 is deficient in glucose phoshorylation. Transformants were plated on SM with glucose as the sole carbon source and incubated for 3 days. A colony was picked and checked for presence of the plasmid by multiplex PCR. The resulting strain was named IMX214. IMX188 and IMX214 were grown overnight in SM 2% galactose. Dilutions were made for a spot plate experiment (1000, 100, 10, 1 cells / 10µl) and 10 µl of each dilution were spotted on a SM with either 2% glucose (w/v) or 2% (w/v) galactose as carbon source. The plates were incubated at 30 °C for 2 days and checked for growth.

ACKNOWLEDGEMENTS

We thank Pilar de la Torre for her technical assistance. We thank Dr. Dan Gibson (JCVI) for his kind advice during the initial phase of implementing *in vivo* assembly in our lab and Dr. Daniel González Ramos for critically reading this manuscript.

REFERENCE LIST

1. Cohen SN, Chang AC, Boyer HW, Helling RB: **Construction of biologically functional bacterial plasmids *in vitro*.** *Proc Natl Acad Sci U S A* 1973, **70:** 3240-3244.

2. Mullis KB, Faloona FA: **Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction.** *Method Enzymol* 1987, **155:** 335-350.

3. Li MZ, Elledge SJ: **Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC.** *Nat Methods* 2007, **4:** 251-256.

4. Benoit RM, Wilhelm RN, Scherer-Becker D, Ostermeier C: **An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids.** *Protein Expres Purif* 2006, **45:** 66-71.

5. Gibson DG, Young L, Chuang RY, Venter JC, Iii CAH, Smith HO *et al.*: **Enzymatic assembly of DNA molecules up to several hundred kilobases.** *Nat Methods* 2009, **6:** 12-16.

6. Raymond CK, Pownder TA, Sexson SL: **General method for plasmid construction using homologous recombination.** *BioTechniques* 1999, **26:** 134-138.

7. Gibson DG, Benders G, Axelrod KC, Zaveri J, Algire M, Moodie M *et al.*: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome.** *Proc Natl Acad Sci U S A* 2008, **105:** 20404-20409.

8. Shao Z, Zhao H, Zhao H: **DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways.** *Nucleic Acids Res* 2009, **37:** e16.

9. Shao Z, Luo Y, Zhao H: **Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler.** *Mol BioSystems* 2011, **7:** 1056-1059.

10. Resnick M, Martin P: **The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control.** *Mol Gen Genet* 1976, **143:** 119-129.

11. Ma H, Kunes S, Schatz PJ, Botstein D: **Plasmid construction by homologous recombination in yeast.** *Gene* 1987, **58:** 201-216.

12. Larionov V, Kouprina N, Graves J, Resnick MA: **Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning.** *Proc Natl Acad Sci U S A* 1996, **93:** 13925-13930.

13. Raymond CK, Pownder TA, Sexson SL: **General method for plasmid construction using homologous recombination.** *BioTechniques* 1999, **26:** 134-+.

14. Gibson DG, Benders G, Andrews-Pfannkoch C, Denisova E, Baden-Tillson H, Zaveri J *et al.*: **Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome.** *Science* 2008, **319:** 1215-1220.

15. Gibson DG: **Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides.** *Nucleic Acids Res* 2009, **37:** 6984-6990.

16. Gibson DG: **Gene and genome construction in yeast.** *Curr Protoc Mol Biol* 2011, **94:** 3.22.1-3.22.17.

17. Shao Z, Luo Y, Zhao H: **DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*.** *Methods Mol Biol* 2012, **898:** 251-262.

18. Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, Cooper EM *et al.*: **Assembling large DNA segments in yeast.** *Methods Mol Biol* 2012, **852:** 133-150.

19. Raymond CK, Sims EH, Olson MV: **Linker-mediated recombinational subcloning of large DNA fragments using yeast.** *Genome Res* 2002, **12:** 190-197.

20. Anderson PR, Haj-Ahmad Y: **Counter-selection facilitated plasmid construction by homologous recombination in *Saccharomyces cerevisiae*.** *BioTechniques* 2003, **35:** 692-4, 696, 698.

21. Fishman-Lobell J, Rudin N, Haber JE: **Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated.** *Mol Cell Biol* 1992, **12:** 1292-1303.

22. Boulton S: **Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance.** *Nucleic Acids Res* 1996, **24:** 4639-4648.

23. Manivasakam P, Weber SC, McElver J, Schiestl RH: **Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 1995, **23:** 2799-2800.

24. Suzuki K, Imai Y, Yamashita I, Fukui S: ***In vivo* ligation of linear DNA molecules to circular forms in the yeast *Saccharomyces cerevisiae*.** *J Bacteriol* 1983, **155:** 747-754.

25. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW: **Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.** *Nat Genet* 1996, **14:** 450-456.

26. Aguilera A: **Deletion of the phosphoglucose isomerase structural gene makes growth and sporulation glucose dependent in Saccharomyces cerevisiae.** *Mol Gen Genet* 1986, **204:** 310-316.

27. Compagno C, Brambilla L, Capitanio D, Boschi F, Ranzi BM, Porro D: **Alterations of the glucose metabolism in a triose phosphate isomerase-negative *Saccharomyces cerevisiae* mutant.** *Yeast* 2001, **18:** 663-670.

28. Schwelberger HG, Kohlwein SD, Paltauf F: **Molecular cloning, primary structure and disruption of the structural gene of aldolase from *Saccharomyces cerevisiae*.** *Eur J Biochem* 1989, **180:** 301-308.

29. Breitenbach-Schmitt I, Heinisch J, Schmitt HD, Zimmermann FK: **Yeast mutants without phosphofructokinase activity can still perform glycolysis and alcoholic fermentation.** *Mol Gen Genet* 1984, **195:** 530-535.

30. Oldenburg KR, Vo KT, Michaelis S, Paddon C: **Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast.** *Nucleic Acids Res* 1997, **25:** 451-452.

31. Shao Z, Luo Y, Zhao H: **DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*.** *Method Mol Biol* 2012, **898:** 251-262.

32. Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH *et al.*: **Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning.** *Nucleic Acids Res* 2001, **29:** E32.

33. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM *et al.*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440:** 940-943.

34. Shetty RP, Endy D, Knight TF, Jr.: **Engineering BioBrick vectors from BioBrick parts.** *J Biol Eng* 2008, **2:** 5.

35. Verduyn C, Postma E, Scheffers W, van Dijken JP: **Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.** *Yeast* 1992, **8:** 501-517.

36. Pronk JT: **Auxotrophic yeast strains in fundamental and applied research.** *Appl Environ Microbiol* 2002, **68:** 2095-2100.

37. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.** *Nucleic Acids Res* 2002, **30:** e23.

38. Sikorski RS, Hieter P: **A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*.** *Genetics* 1989, **122:** 19-27.

39. Gietz RD, Woods R: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Method Enzymol* 2002, **350:** 87-96.

40. Birnboim HC, Doly J: **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** *Nucleic Acids Res* 1979, **7:** 1513-1523.

41. Peng Y, Leung H, Yiu SM, Chin F: **IDBA - practical iterative de Bruijn graph *de novo* assembler.** In *Research in Computational Molecular Biology*. 6044 edition. Edited by Berger B. Berlin: Springer; 2010:426-440.

42. Burke D, Dawson D, Stearns T: **Methods in Yeast Genetics: a Cold Spring Harbor Laboratory course manual**, NY: Cold Spring Harbor Laboratory Press; 2000.

43. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik M *et al.*: ***De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology.** *Microb Cell Fact* 2012, **11:** 36.

44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:** 403-410.

45. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C *et al.*: **An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains.** *Enzyme Microb Technol* 2000, **26:** 706-714.

Chapter 3

One step assembly and targeted integration of multi-gene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*

Kuijpers NGA, Chroumpi S, Vos T,
Solis-Escalante D, Bosman L, Pronk JT,
Daran JM, Daran-Lapujade P

Published in FEMS Yeast Research (2013) 13:8 126-139

ABSTRACT

In vivo assembly of overlapping fragments by homologous recombination in *Saccharomyces cerevisiae* is a powerful method to engineer large DNA constructs. Whereas most *in vivo* assembly methods reported to date result in circular vectors, stable integrated constructs are often preferred for metabolic engineering as they are required for large-scale industrial application. The present study explores the potential of combining *in vivo* assembly of large, multi-gene expression constructs with their targeted chromosomal integration in *S. cerevisiae*. Combined assembly and targeted integration of a ten-fragment 22-kb construct to a single chromosomal locus was successfully achieved in a single transformation process, but with low efficiency (5% of the analyzed transformants contained the correctly assembled construct). The meganuclease I-SceI was therefore used to introduce a double-strand break at the targeted chromosomal locus, thus to facilitate integration of the assembled construct. I-SceI-assisted integration dramatically increased the efficiency of assembly and integration of the same construct to 95%. This study paves the way for the fast, efficient, and stable integration of large DNA constructs in *S. cerevisiae* chromosomes.

INTRODUCTION

The yeast *Saccharomyces cerevisiae* is intensively explored and applied as a platform for the industrial production of a wide range of endogenous and heterologous compounds. Its success in this role can be readily explained from its robustness, simple nutritional requirements, and easy genetic accessibility. This last feature has recently propelled the popularity of *S. cerevisiae* as a preferred platform in synthetic biology, especially for the assembly of large DNA constructs [1]. Improvement of performance of industrial organisms (higher productivity and yield, increased robustness, expression of complex heterologous pathways, etc.) by metabolic engineering requires simultaneous expression of dozens of genes. Handling such numbers of genes by classical cloning methods is extremely time-consuming. Over the last decade, several methods have been developed for fast and efficient assembly of large DNA constructs. The most promising of these are recombination-based methods in which multiple linear DNA fragments with overlapping terminal sequences are recombined into a single vector [2]. While *in vitro* recombination-based methods such as SLIC [3], InFusion [4] and Gibson's isothermal assembly [5] are undeniably valuable, *in vivo* recombination using *S. cerevisiae* is proving to be the method of choice for large constructs assembled from many fragments [1,6]. This method can be used for efficient and accurate plasmid assembly [7]. However, plasmid-borne gene expression is not favored for industrial-scale production as plasmids are notoriously unstable, and maintaining the selection pressure necessary for the cells to retain the plasmid is typically difficult to achieve in industrial settings [8]. Conversely, chromosomal integration results in stable expression of genes, but methods capable of rapid and accurate assembly and integration of large constructs are currently not available. A recent pioneering study has demonstrated that *S. cerevisiae* can, in a single step, assemble multiple fragments in a 23.7-Kb construct and integrate this construct in the yeast chromosome [6]. To increase the probability of integration events, the abundant δ-sites were chosen as integration loci. Although reasonable efficiencies were obtained (up to 70% correct transformants), targeting to δ-sites randomizes the number and location of integration sites and therefore results in unpredictable copy numbers and integration loci. For strain engineering programs, where high integration efficiencies are required, control of the locus and the number of integration events is of paramount importance. Considering the relatively low chromosomal integration efficiencies of linear DNA fragments in *S. cerevisiae* (*c.* 10^{-6} transformants per viable cell using 50-bp flanks homologous to the integration site [9]), it is not surprising that targeting a single specific chromosomal site for combined assembly and integration of a multigene fragment results in low efficiencies. The aim of the present study was to evaluate the potential of Combined *in vivo* Assembly and Targeted chromosomal Integration (from now on referred to as CATI) of large DNA constructs in *S. cerevisiae*. Preliminary results identified integration as the bottleneck for the CATI approach. Use of the meganuclease I-SceI was therefore explored to create double-strand DNA breaks and thereby enhance the integration efficiency. I-SceI and its homolog I-SceII are native to *S. cerevisiae*, in which they are encoded by mitochondrial introns [10-12]. These meganucleases, also named homing endonucleases, are responsible for intron mobility in the mitochondria of yeast, in which they initiate a site-specific gene conversion [13]. Much like the well-studied HO

meganuclease, I-SceI initiates a double-strand break at a specific recognition site. The recognition site of I-SceI extends over a 18-bp nonsymmetrical sequence and generates a cut with a 4-bp overhang within its recognition site [12]. In the early 90s, it was demonstrated that I-SceI, when expressed in the nucleus, was active on nuclear targets and, as predicted from the absence of I-SceI cutting sites in the genomic DNA, was not toxic upon expression in wild-type *S. cerevisiae* [13]. In the presented work, I-SceI was implemented and investigated to develop a robust system for combined assembly and targeted chromosomal integration of multigene constructs in *S. cerevisiae*.

MATERIALS AND METHODS

Strains and media | The *S. cerevisiae* strains used in this study are derived from the CEN.PK family (Table 3.1.) [14,15]. Cultures for transformation were grown in complex medium containing 10 g l^{-1} Bacto yeast extract, 20 g l^{-1} Bacto peptone, and 20 g l^{-1} glucose as carbon source. When galactose induction of *SCEI* was required, cultures were transferred to synthetic medium (SM) containing galactose as the sole carbon source and grown for 4 h on that medium prior to transformation. SM contained, per liter of demineralized water, 5 g $(NH_4)_2SO_4$, 3 g KH_2PO_4, 0.5 g $MgSO_4.7{\cdot}H_2O$, and trace elements [16]. Vitamins [16] were added after heat sterilization of the medium at 120 °C for 20 min. Glucose or galactose were separately sterilized at 110 °C and added to a final concentration of 20 g l^{-1}. Where required, the medium was supplemented with appropriate amounts of auxotrophic requirements [17]. Solid medium was prepared by adding 2% (w/v) agar to the media prior to heat sterilization. Selective medium for the *amdS* marker was prepared as previously described [21].

Molecular biology techniques | PCR amplification was performed using Phusion® Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA). To improve PCR efficiency, the conditions in the PCR as recommended by the supplier were modified by decreasing the primer concentration from 500 to 200 nM and increasing the Phusion™ Hot Start High-Fidelity polymerase concentration from 0.02 to 0.03 U μl^{-1}. All other conditions followed the manufacturer's instructions. Genomic template DNA was isolated from *S. cerevisiae* CEN.PK113-7D using the Qiagen 100/G kit (Qiagen, Hilden, Germany). Plasmids maintained in *E. coli* DH5α were isolated with the GenElute™ Plasmid Miniprep Kit (Sigma, St. Louis, MI). DNA fragments were separated on 1% (w/v) agarose (Sigma) gels in 1x TAE (40 mM Tris-acetate pH 8.0 and 1 mM EDTA) and in 2% (w/v) agarose in 0.5x TBE (45 mM Tris-borate pH 8.0 1 mM EDTA) when fragments were smaller than 500-bp. Fragments were isolated from gel using the Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA). The glycolytic gene fragments for assembly were not gel-purified, but concentrated directly after PCR amplification by Vivacon® 500 spin columns (Sartorius Stedim, Aubagne, France). DNA concentrations were measured in a NanoDrop 2000 spectrophotometer (wavelenght 260 nm; Thermo Fisher Scientific). Genomic DNA of transformants was isolated using the YeaStar™ Genomic DNA kit (Zymo Research). Multiplex PCR was performed with primers (Table 3.2.) at a concentration of 150 nM. Cycling parameters were 94°C for 3 min, then 35 cycles of 94 °C for 30 s, 60°C for 90 s, and 72 °C for 60 s, followed by a 10-

min incubation at 72 ˚C. Prior to transformation, fragments were pooled, maintaining equimolar concentrations with the marker fragment. Transformation to yeast was performed with the LiAc/ssDNA method [18].

TABLE 3.1 | Strains used in this study

Strain	Relevant Genotype	Source
CEN.PK113-7D	*MATa MAL2-8c SUC2*	[14,39]
CEN.PK113-5D	*MATa ura3-52 MAL2-8c SUC2*	[39]
CEN.PK102-3A	*MATa ura3-52 leu2-3 MAL2-8c SUC2*	[39]
IMX212	*MATa ura3-52 leu2-3 MAL2-8c SUC2 spr3::(P_{GAL1}-SCEI-T_{cyc1}; KlURA3)*	This study
IMX221	*MATa ura3-52 MAL2-8c SUC2 spr3::(TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*	This study
IMX222	*MATa ura3-52 leu2-3 MAL2-8c SUC2 spr3::(TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*	This study
IMX224	*MATa ura3-52 MAL2-8c SUC2 spr3::(TagG-amdSYM-TagF)*	This study

TABLE 3.2 | Oligonucleotide primers used in this study

Primers	Sequence 5' --> 3'
	To add SHR-sequences
TPI1 Rv +H	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGAAT TGCTGAGAACCCGTGACTAGTGTGAGCGGGATTTAAACTGTG
TPI1 Fw +I	GCCTACGGTTCCCGAAGTATGCTGCTGATGTCTGGCTATACCT ATCCGTCTACGTGAATAGCGAAAATGACGCTTGCAGTG
FBA1 Rv +H	GTCACGGGTTCTCAGCAATTCGAGCTATTACCGATGATGGCT GAGGCGTTAGAGTAATCTAAAATCTCAAAAATGTGTGGGTCATTACG
FBA1 Fw +G	GCCAGAGGTATAGACATAGCCAGACCTACCTAATTGGTGCAT CAGGTGGTCATGGCCCTTAGTGCATGACAAAAGATGAGCTAGG
Amds GPD Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCG CCAGATCATCAATAGGCACGCTGGAGCTCTTCGA

Primers	**Sequence 5' --> 3'**
	To add SHR-sequences
Amds GPD Rv+ B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACG GGCATCGTCCTCTCGAAAGGTGGGCCGCAAATTAAAGCCTTCGAG
Amds TEF Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATT CCGCCAGATCATCAATAGGCACGCGACATGGAGGCCCAGAATACC
Amds TEF RV+ B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGC ATCGTCCTCTCGAAAGGTGAGTATAGCGACCAGCATTCACATACG
KlLEU2-Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGC CAGATCATCAATAGGCACAGAGATCCGCAGGCTAACCG
KlLEU2-Rv + B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGG CATCGTCCTCTCGAAAGGTGGCTGTGAAGATCCCAGCAAAGG
PFK2 Rv +F	TGCCGAACTTTCCCTGTATGAAGCGATCTGACCAATCCT TTGCCGTAGTTTCAACGTATGATAGCCATTCTCTGCTGCTTTGTTG
PFK2 Fw +J	GGCCGTCATATACGCGAAGATGTCCAAGCAGGTAGAA CACATAGTCTGAGCATCTCGTCGGAGATCCGAGGGACGTTTATTGG
PFK1 Rv +D	ACGCATCTACGACTGTGGGTCCCGTGGAGAAATGTATG AAACCCTGTATGGAGAGTGATTTCGAGATTCCTCAATCCATACACCATTATAG
PFK1 Fw +J	CGACGAGATGCTCAGACTATGTGTTCTACCTGCTTGGAC ATCTTCGCGTATATGACGGCCTGTCGTCTTCGTGAACCATTGTC
PGI1 Rv +D	AATCACTCTCCATACAGGGTTTCATACATTTCTCCACGGG ACCCACAGTCGTAGATGCGTCTGAAGAAGGCATACTACGCCAAG
PGI1 Fw +C	ACGTCTCACGGATCGTATATGCCGTAGCGACAATCTAAGAAC TATGCGAGGACACGCTAGTTCGCGACACAATAAAGTCTTCACG
HXK2 Rv +C	CTAGCGTGTCCTCGCATAGTTCTTAGATTGTCGCTACGGC ATATACGATCCGTGAGACGTGCAAGAGAAAAAAACGAGCAATTGTTAAAAG
HXK2 Fw +B	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCG ACCAGCCTAAGAATGTTCAACGACGGCACCGGGAAATAAACC
PGK1 Fw + I	GTGCCTATTGATGATCTGGCGGAATGTCTGCCGTGCC ATAGCCATGCCTTCACATATAGTCCTGCATTTAAAGATGCCGATTTGG
PGK1 Rv + A	GTAGACGGATAGGTATAGCCAGACATCAGCAGCATACT TCGGGAACCGTAGGCATTTTAGCGTAAAGGATGGGGAAAGAG
	For the construction of pUDC073
SCEI-Fw	GCTGCCACTAGTATAATGCATCAAAAAAACCAGGTAATG
SCEI –Rv	TTATCACTCGAGTTATTACTTAAGGAAAGTTTCGGAGGAGATAG

Primers	**Sequence 5' --> 3'**
	For fusion-PCR of the ISCEI-URA3-cassette
URA3-Fw	GAGCCATCCATTCGTAATTCACTACTGCCTGAGGGTTGTTCTC AGAAGCTCATCGAACTGTCATC
URA3-Rv	CCATTCTGTAGCCACCTTATCCATGACCGTTTTATTAATTATTTCATAGCA CTTGTAATTATATTACCCTGTTATCCCTAGCGAAGTGAGTGTTGCACCGTGCCAATG
SCEI -Fw(2)	GCTGCATCCTTCCCATGCAAAGTGTCTTCGTATTTAGTGAT GTTTTGTTAGCGACACAAAGCTAGGGATAACAGGGTAATAT GCAGTGAGCGCAACGCAATTAATG
SCEI-Rv(2)	AACAACCCTCAGGCAGTAGTGAATTACGAATGGAT GGCTCCGACTCACTATAGGGCGAATTGG
FUS1	GCTGCATCCTTCCCATGCAAAGTG
FUS2	CCATTCTGTAGCCACCTTATCC
SCEI+URA-Fw	GCAGTGAGCGCAACGCAATTAATG
SCEI+URA-Rv	GAAGTGAGTGTTGCACCGTGCCAATG
Tag F-REC-fw	CCATTCTGTAGCCACCTTATCCATGACCGTTTTATTAATTATTTC ATAGCACTTGTAATTTGCCGAACTTTCCCTGTATGAAGC GATCTGACCAATCCTTTGCCG
Tag F-REC-rv	CCTGCATTGGCACGGTGCAACACTCACTTCGCTAGGGATAACAGG GTAATATCATACGTTGAAACTACGGCAAAGGATTGGTC AGATCGCTTCATACAGGG
Tag G-REC-fw	GCTGCATCCTTCCCATGCAAAGTGTCTTCGTATTTAG TGATGTTTTGTTAGCGACACAAAGCCAGAGGTATAG ACATAGCCAGACCTACCTAATTGGTGCATC
Tag G-REC-rv	GTAACTCACATTAATTGCGTTGCGCTCACTGCATATTACCCT GTTATCCCTAGCAAGGGCCATGACCACCTGATGCAC CAATTAGGTAGGTCTGGCTATGTCTATACC
	For construction of the fragments targeting the CAN1 locus
H1-Fw	TTCTAGGTTCGGGTGACGTGAAG
H1-Rv	AAGGGCCATGACCACCTGATGCACCAATTAGGTAGGTCTGGCTATGT CTATACCTCTGGCATTACCCTGTTATCCCTATTAATCACATTCCCACGCCATTTCG
Y1	CATACGTTGAAACTACGGCAAAGGATTGGTCAGATCGCTTCATACAGGGAAAGTTCGGC ATAGGGATAACAGGGTAATGCTCATTGATCCCTTAAACTTTCTTTTCGGTGTATGAC
Y2	CCAGTTTTCAATCTGTCGTCAATCGAAAGTTTATTTTAATCACATTCCCACGCCAT TTCGCATTCTCACCCTCATAAGTCATACACCGAAAAGAAAGTTTAAGGGATCAATGAGC

Primers	Sequence 5' --> 3'
	For construction of the fragments targeting the CAN1 locus
H2-Fw	AATAAACTTTCGATTGACGACAGATTG
H2-Rv	GTTTCCGGGTGAGTCATACG
FUS3	CATACGTTGAAACTACGGCAAAGG
	For analytical PCR: glycolytic genes integrated in the CAN1 Locus
G- fw	CCGTCATCGGAGTCGTTATCAG
G- rv	GCTCTTTTCTTCTGAAGGTCAATG
F-fw	GACGCCATTTGGAACGAAAAAAAG
F-rv	TAACGGCAAACAGCAAAGGC
H-fw	GTTACGTGCTCAGTTGTTAGATATG
H-rv	GCAGAAGTGTCTGAATGTATTAAGG
I-fw	TGAGCCACTTAAATTTCGTGAATG
I-rv	TTTCTCTTTCCCCATCCTTTACG
A-fw	AAGGATTCGCGCCCAAATCG
A-rv	CTTCCCAAGATTGTGGCATGTC
B-fw	TGGCTATCGCTGAAGAAGTTGG
B-rv	ACGGAATAGAACACGATATTTGC
C-fw	TCACGGGATTTATTCGTGACG
C-rv	CCCACGATGCTTCTACCAAC
D-fw	ACTCGCCTCTAACCCCACG
D-rv	AATCATGTTGATGACGACAATGG

Primers	**Sequence 5' --> 3'**
	For analytical PCR: glycolytic genes integrated in the CAN1 Locus
J-fw	GCTTAATCTGCGTTGACAATGG
J-rv	CAATAAACGTCCCTCGGATCTC
	For multiplex PCR: glycolytic genes integrated in the SPR3 locus
G- fw	CTTGGCTCTGGATCCGTTATCTG
G- rv	GCTCTTTTCTTCTGAAGGTCAATG
F-fw	GACGCCATTTGGAACGAAAAAAAG
F-rv	TTGGGCTGGACGTTCCGACATAG
H-fw	GTTACGTGCTCAGTTGTTAGATATG
H-rv	GCAGAAGTGTCTGAATGTATTAAGG
I-fw	TGAGCCACTTAAATTTCGTGAATG
I-rv	TTTCTCTTTCCCCATCCTTTACG
A-fw	AAGGATTCGCGCCCAAATCG
A-rv	CTTCCCAAGATTGTGGCATGTC
B-fw	TGGCTATCGCTGAAGAAGTTGG
B-rv	ACGGAATAGAACACGATATTTGC
C-fw	TCACGGGATTTATTCGTGACG
C-rv	CCCACGATGCTTCTACCAAC
D-fw	ACTCGCCTCTAACCCCACG
D-rv	AATCATGTTGATGACGACAATGG
J-fw	GCTTAATCTGCGTTGACAATGG
J-rv	CAATAAACGTCCCTCGGATCTC

TABLE 3.3 | Plasmids used in this study

Plasmid	Characteristic	Source
pUG72	PCR template for *Kluyveromyces lactis URA3 (KlURA3)*	[40]
pUG73	PCR template for *Kluyvermyces lactis LEU2 (KlLEU2)*	[40]
pUG*amdSYM*	PCR template for *amdSYM* under the control of the *AgTEF2* promoter	[21]
pUDE158	PCR template for *amdSYM* under the control of the *TDH3* promoter	[21]
pAG416GAL-ccdB	*CEN6/ARS4* ori, *URA3*, P_{GAL1}-*ccdB*-T_{CYC1}	[41]
pUDC073	*CEN6/ARS4* ori, *URA3*, P_{GAL1}-*SCEI*-T_{CYC1}	This study

Construction of a platform strain for I-SceI assisted integration | The plasmids used in this study are listed in Table 3.3. Plasmid pUDC073 was obtained by cloning the *SCEI* ORF into pAG416GAL-ccdB. The *SCEI* ORF was amplified from Biobrick BBa_K175041 (http://parts.igem.org/Part:BBa_K175041) with primers *SCEI*-Fw and *SCEI*-Rv. The resulting fragment was restricted by SpeI and XhoI and ligated into pAG416GAL-ccdB, yielding pUDC073.

The integration site was constructed by integration of the TagG-*SCEI-URA3*-TagF cassette into the yeast genome. Construction of the TagG-*SCEI-URA3*-TagF cassette was performed in multiple steps. First, *SCEI* was amplified from pUDC073 with primers *SCEI*-Fw(2) and *SCEI*-Rv(2), and *KlURA3* was amplified from pUG72 with primers *URA*-Fw and *URA*-Rv. The resulting cassettes were gel-purified, and 100 ng of each cassette was used for fusion-PCR [19] using primers FUS1 and FUS2. Cycling parameters were 98˚C for 1 min, then 8 cycles of 98 ˚C for 30 s, 58˚C for 30 s, and 72 ˚C for 120 s, followed by 27 cycles of 98 ˚C for 30 s, 70˚C for 30 s, and 72 ˚C for 120 s, followed by a 10-min incubation at 72 ˚C. The intermediate strain IMX212 was constructed by integration of the resulting product at the *SPR3* locus of CEN.PK102-3A, yielding strain IMX212 (Fig. 3.1.a1). Secondly, genomic DNA of IMX212 served as a template for amplification of the *SCEI/URA3* cassette with primers *SCEI+URA*-Fw and *SCEI+URA*-Rv, resulting in fragment X1 (Fig. 3.1.a2). Thirdly, the flanking fragments X2 and X3 carrying 3 regions: (1) the regions homologous to the *SPR3* locus necessary for integration of the final cassette, (2) the I-SceI recognition site and (3) the F and G synthetic homologous recombination sequences (SHR-sequences) required for integration of cassettes during I-SceI-assisted integration were obtained by PCR. Fragment X2 (Fig. 3.1.a2) was obtained by annealing oligonucleotides TagG-REC-Fw and TagG-REC-Rv in a 50 µl PCR mix at a concentration of 1 µM. This mix was subjected to 10 cycles in a thermocycler using the following conditions: 98 ˚C for 30 s, 65˚C for 30 s, and 72 ˚C for 15 s, followed by a 10-min incubation at 72 ˚C. Fragment X3 was obtained by the same procedure using oligonucleotides TagF-REC-Fw and TagF-

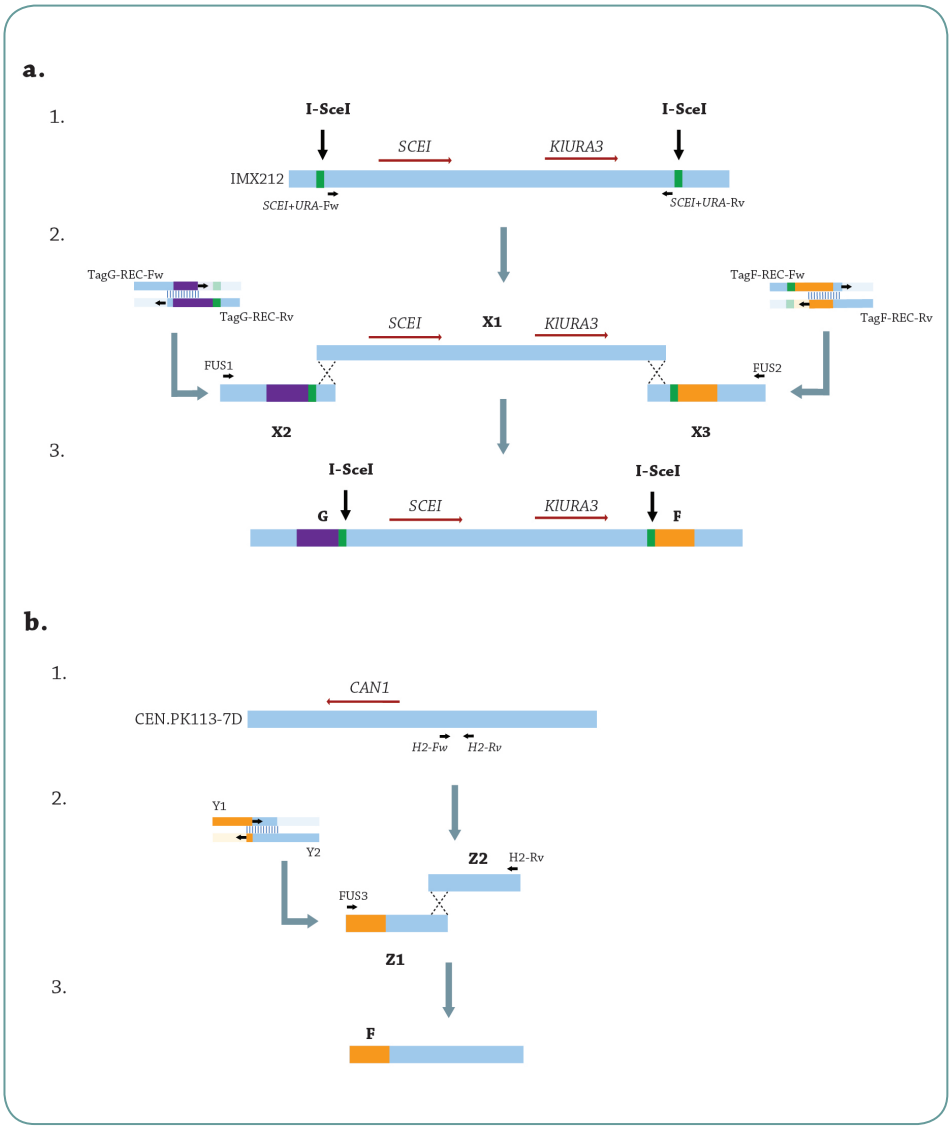


FIGURE 3.1 | Construction of the TagG-*SCEI-KlURA3*-TagF and the H2 cassette. (a) First the *SCEI/URA3* cassette was obtained by PCR on genomic DNA of IMX212 with primers *SCEI+URA*-Fw and *SCEI+URA*-Rv, resulting in fragment X1 (a1). Fragment X2 was obtained by fusing oligos TagG-REC-Fw and TagG-REC-Rv in an independent PCR. Fragment X3 was obtained in the same way, using oligos TagF-REC-fw and TagF-REC-Rv (a2). Fragments X1, X2 and X3 were fused in a fusion PCR with primers FUS1 and FUS2, resulting in the TagG-*SCEI-KlURA3*-TagF cassette (a3). (b) Fragment Z2 was obtained by PCR on genomic DNA of CEN.PK113-7D with primers H2-fw and H2-rv (Bb). Fragment Z1 was obtained by fusing oligos Y1 and Y2 in a PCR (b2). Fragments Z1 and Z2 were fused in a fusion-PCR with oligos FUS3 and H2-rv, resulting in fragment H2, which contains SHR-sequence F and 300-bp homology to the *CAN1* locus (b3).

REC-Rv (Fig. 3.1.a2). Fourthly, the DNA fragments X1, X2 and X3 were gel purified and fused by fusion-PCR using primers FUS1 and FUS2, using the same cycling parameters as for the pevious fusion-PCR (Fig. 1a2). The resulting product was the TagG-*SCEI*-*URA3*-TagF cassette (Fig. 3.1.a3), which was gel-purified and transformed using the LiAc/ssDNA method [18] to CEN.PK113-5D, leading to IMX221, and to CEN.PK102-3A, leading to IMX222. In both strains, the regions containing the I-SceI recognition sites and the SHR-sequences F and G were PCR-amplified and sequenced using Sanger sequencing (BaseClear, Leiden, the Netherlands).

Preparation of fragments for *in vivo* assembly | Fragments for *in vivo* assembly were obtained by PCR from either genomic or plasmid template DNA. The amplified fragments were stocked in TE buffer (10 mM Tris, pH8, 1 mM EDTA). Fragments amplified from plasmid templates were subjected to gel extraction to prevent false-positive transformants that might arise from contamination with linearized template plasmid. Fragment $amdSYM_{AB}$ carrying the counter selectable *amdS* marker behind the *TDH3* promoter was amplified from pUDE158 (Table 3.3.) in the experiment targeting the *CAN1* locus using primers *Amds*-GPD-Fw+A and *Amds*-GPD-Rv+B. In the other experiments, an $amdSYM_{AB}$ cassette carrying the *amdS* marker behind the Ag*TEF2* promoter was used, to eliminate sequence homology between the marker cassette and the yeast genome. This cassette was amplified from pUG*amdSYM* with primers *Amds*-TEF-Fw+A and *Amds*-TEF-Rv+B. The marker fragment $KlLEU2_{AB}$ was obtained from pUG73 with primers *KlLEU2*-Fw+A and *KlLEU2*-Rv+B. The fragments containing the glycolytic genes were all amplified from CEN.PK113-7D genomic DNA [14] using the primers with the corresponding glycolytic gene names listed in Table 3.2. Fragments H1 and H2 homologous to the *CAN1* locus and used for targeted integration of *in vivo*-assembled constructs at that locus were obtained by PCR amplification from CEN.PK113-7D genomic DNA. H1 was amplified with primers H1-Fw and H1-Rv. Fragment H2 was obtained in two steps (Fig. 1b). First fragment Z1 was obtained by fusion of oligos Y1 and Y2 in the same way as described for fragment X2 (Fig. 3.1.b2). Fragment Z2 was obtained by PCR on genomic DNA using primers H2-Fw and H2-Rv (Fig. 3.1.b1). Fragment Z1 and Z2 were gel-purified and fused by fusion-PCR using the same method as described before using primers H2-rv and FUS3, resulting in fragment H2 (Fig. 3.1.b3). Fragments H1 and H2 were gel-purified before addition to the transformation mix.

RESULTS

Poor efficiency of simultaneous assembly and targeted integration of seven glycolytic genes into the *CAN1* locus | We demonstrated previously that *S. cerevisiae* can assemble nine fragments very efficiently and with high fidelity into a 21-kb plasmid carrying six glycolytic genes [7]. To test whether combined assembly and targeted integration at a specific locus of a multiple gene construct could be achieved, the six previously designed glycolytic gene fragments [7] and one additional glycolytic gene fragment were used to assemble and integrate a total of seven glycolytic genes in a single step (Fig 3.2). A set of ten fragments was obtained by adding two flanking fragments designed to target the *CAN1* locus [20] and one marker fragment carrying the *amdS* dominant marker (Fig. 3.2., [21]) to the seven glycolytic gene cassettes. All fragments were designed to overlap by 60-bp synthetic homologous recombination sequences (from now on referred to as SHR-sequences) that do not share homology with the yeast genome [7]. After transformation with these ten fragments, yeast cells were grown on glucose synthetic medium. To identify transformants in which genomic integration of *amdS* had occurred, acetamide was used as the sole nitrogen source. Thirty-five transformants were obtained, of which 20 were picked and plated on medium containing L-canavanine, to select for integration of the assembled construct in the *CAN1* locus. Only two out of the 20 transformants were able to grow

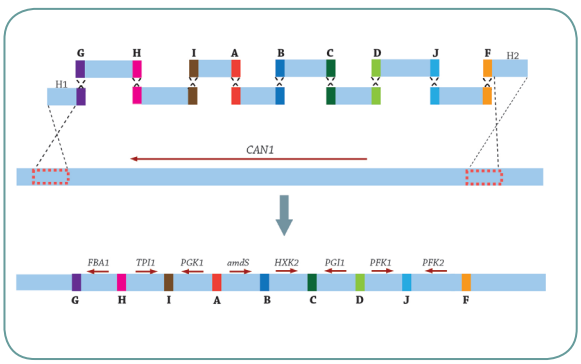
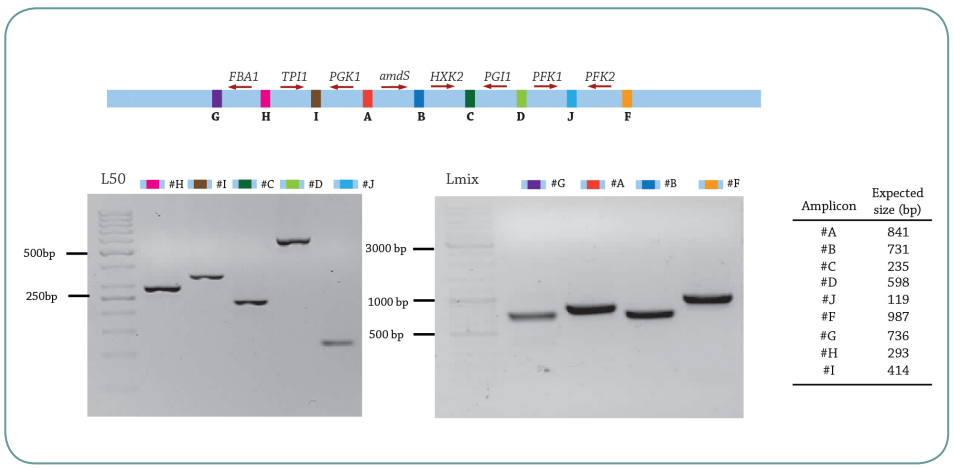


FIGURE 3.2 | Combined assembly and integration of seven glycolytic genes in the *CAN1* locus of *Saccharomyces cerevisiae*. Ten overlapping DNA fragments, containing seven glycolytic genes, the *amdS* selection marker, and the two flanking fragments H1 and H2, carrying 300-bp sequences homologous to the *CAN1* integration locus, were co-transformed to *S. cerevisiae* and assembled in yeast via homologous recombination into a single large integration cassette. 60-bp SHR-sequences were used to promote *in vivo* assembly of the fragments.

in presence of L-canavanine, indicating that *CAN1* was disrupted in only 10% of the tested transformants. Of these two L-canavanine-resistant transformants, only one was correctly assembled and carried all transformed fragments (Fig. 3.3). Although simultaneous assembly and targeted integration of a multigene construct in a single chromosomal locus was achieved, it was extremely inefficient (one out of 20 of the tested transformants). As we previously established efficiencies of plasmid assembly of 95% with the same overlapping sequences [7], these results pointed at the integration step as the main bottleneck for the CATI approach.

Substantial improvement of integration efficiency of an *amdS* marker cassette into the *SPR3* locus using I-SceI-induced double-strand DNA breaks | The cellular function of homologous recombination is to repair double-strand DNA breaks (DSBs). While *in vivo* assembly supplies DNA fragments with open ends, readily accessible for the homologous recombination machinery, chromosomal integration of these fragments requires recombination of DNA fragments with intact genomic DNA and is therefore far less likely to occur [22,23]. A way to enhance the efficiency of integration would therefore be to artificially introduce a DSB at the integration site. Rare-cutting endonucleases can be used to introduce DSBs, thereby



Amplicon	Expected size (bp)
#A	841
#B	731
#C	235
#D	598
#J	119
#F	987
#G	736
#H	293
#I	414

FIGURE 3.3 | PCR analysis of a positive transformant after cotransformation of ten overlapping fragments to *Saccharomyces cerevisiae*. The PCRs were designed to produce amplicons covering the indicated junctions. PCR products covering junctions H, C, D, J and I were separated on a 2% agarose gel, and PCR products covering junctions A,F,G and B were separated on a 1% agarose gel by electrophoresis. In the lane labeled 'L50' a 50-bp Gene Ruler ladder was loaded; in the lane labeled 'Lmix' a Gene Ruler Mix ladder was loaded; sizes are indicated. All amplicons matched the expected size, thereby indicating correct assembly and integration of seven glycolytic genes in the *CAN1* locus.

drastically increasing the efficiency of integration by homologous repair [9,24]. The well-studied I-SceI meganuclease, originally encoded by the *S. cerevisiae* mitochondrial gene *SCEI*, has a 18-bp unique recognition sequence and has been previously functionally expressed in the nucleus of *S. cerevisiae* [13]. To investigate whether introduction of DSBs might eliminate or alleviate the bottleneck in chromosomal integration, a platform was engineered to combine *in vivo* assembly with I-SceI facilitated integration. IMX221 was constructed by integration of a cassette containing *SCEI* under the control of the inducible *GAL1* promoter and a uracil marker in the *SPR3* locus of the uracil auxotroph *S. cerevisiae* CEN.PK113-5D (Fig 3.4a). The resulting locus carried two 22-bp I-SceI recognition sequences flanked by 60-bp SHR-sequences G and F (Fig. 3.4a). A single cassette, carrying the *amdS* marker and SHR-sequences G and F at its 5' and 3' ends, respectively, was constructed to integrate at this synthetic locus (Fig 3.4b). This cassette was transformed to IMX221 pre-incubated in galactose medium to induce expression of *SCEI*. To quantify the effect of the I-SceI-induced DSB, IMX221 cells not induced on galactose were transformed with the same *amdSYM* cassette and used as a negative control. Previous reports indicated that incubation of strains containing *SCEI* under control of the *GAL1* promoter in the presence of galactose resulted in induction of DSBs within several hours [9]. In the present study *SCEI* was induced by growing the yeast cells for four hours in galactose medium prior to transformation. While transformation of I-SceI-expressing cells resulted in *c*. 10^4 transformants, the negative control yielded only 15 transformants. PCR analysis of three colonies of each of the I-SceI-expressing and nonexpressing transformants revealed that they all contained the *amdSYM* cassette, correctly integrated at the *SPR3* locus. One correct clone resulting from the transformation of induced cells was named IMX224. Sequencing of the *SPR3* locus of IMX224 showed that the region between the SHR-sequences G and F was successfully replaced by the *amdSYM* cassette without leaving any scar of the I-SceI recognition sequences. These results demonstrate that induction of a DSB is a critical step for integration of DNA fragments in yeast chromosomes and suggested that I-SceI-assisted integration should improve the efficiency of CATI.

I-SceI-assisted integration of seven glycolytic genes at a synthetic chromosomal locus | To test whether I-SceI-assisted integration could be combined with *in vivo* assembly of multiple genes, the same seven overlapping glycolytic gene cassettes used in the first experiment and the *amdSYM* marker cassette were cotransformed to I-SceI expressing cells of the platform strain IMX221 (Fig. 3.4c). Two control sets of fragments were also tested. An incomplete set of cassettes (lacking $HXK2_{BC}$) was used to estimate the occurrence of nonhomologous recombination events within the construct. Secondly, a single cassette carrying the selection marker but without homology to the integration site was used to evaluate the possible integration of nonhomologous fragments. Transformation of I-SceI-expressing cells with these two control sets of fragments did not yield transformants. Conversely, transformation of I-SceI-expressing cells with the complete set of fragments resulted in 336 transformants capable of using acetamide as sole nitrogen source. Analysis by multiplex PCR of ten randomly picked clones demonstrated the integration of a full set

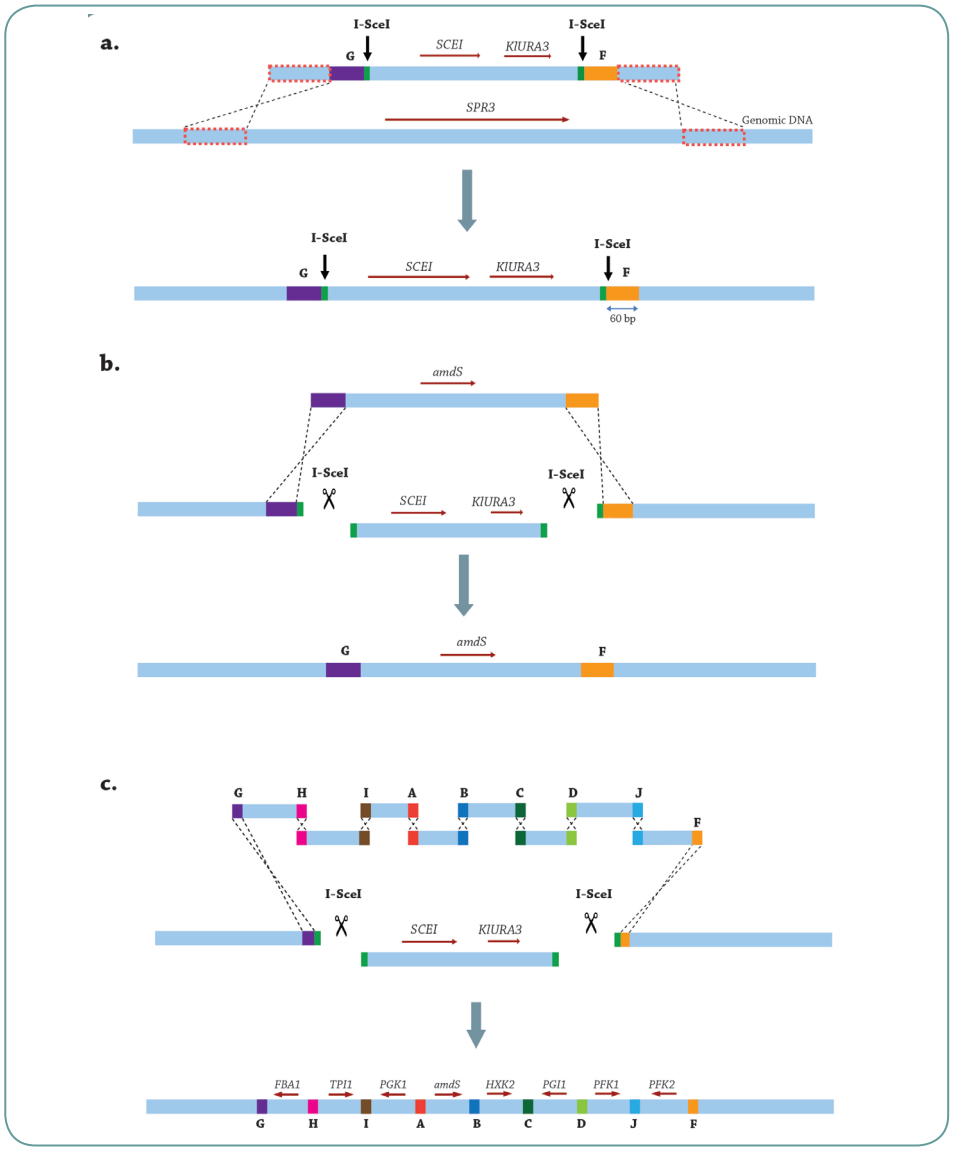


FIGURE 3.4 | Design of the I-*Sce*I-facilitated CATI method. (a) First the platform strain was obtained by introducing a cassette containing *SCEI* and *KlURA3* flanked by three regions, the I-SceI recognition site, synthetic recombination sequences G and F, and flanking regions homologous to the targeted locus *SPR3*. (b) Induction of plasmid-borne *SCEI* in the platform strain prior to transformation, causing excision of the *SCEI*/*URA3* fragment and leaving the 60-bp SHR-sequences F and G exposed for recombination. Transformation of the induced yeast cells with the *amdSYM* cassette flanked by SHR-sequences G and F led to integration of the cassette at the I-SceI-restricted locus. (c) integration of multiple overlapping fragments, using the same integration approach described in (b), leading to I-SceI-assisted integration of seven glycolytic genes and a *KlLEU2* marker cassette into the *SPR3* locus.

of fragments in the *SPR3* locus of *S. cerevisiae* for nine clones (Fig. 3.5a). To evaluate the robustness of the CATI approach, the same experiment was repeated by replacing the dominant *amds* marker by the widely used auxotrophic selection marker *LEU2*. A new platform strain was constructed by introducing the same synthetic locus used in IMX221 in the leucine auxotrophic strain *S. cerevisiae* CEN.PK102-3A, resulting in strain IMX222. Subsequently, the above-described cassettes carrying the seven glycolytic genes were cotransformed to IMX222 together with the *Kluyveromyces lactis LEU2* orthologous marker cassette. The *KlLEU2*-based CATI resulted in 470 clones, of which ten clones were analyzed by multiplex PCR. These ten clones harbored all expected amplicons, indicating the correct integration of all eight fragments at the targeted locus (Fig. 3.5b). These results demonstrate the high efficiency (*c.* 95%) and robustness of I-SceI-assisted simultaneous assembly and chromosomal integration of eight overlapping DNA cassettes, comprising a 22-kb construct, in *S. cerevisiae*.

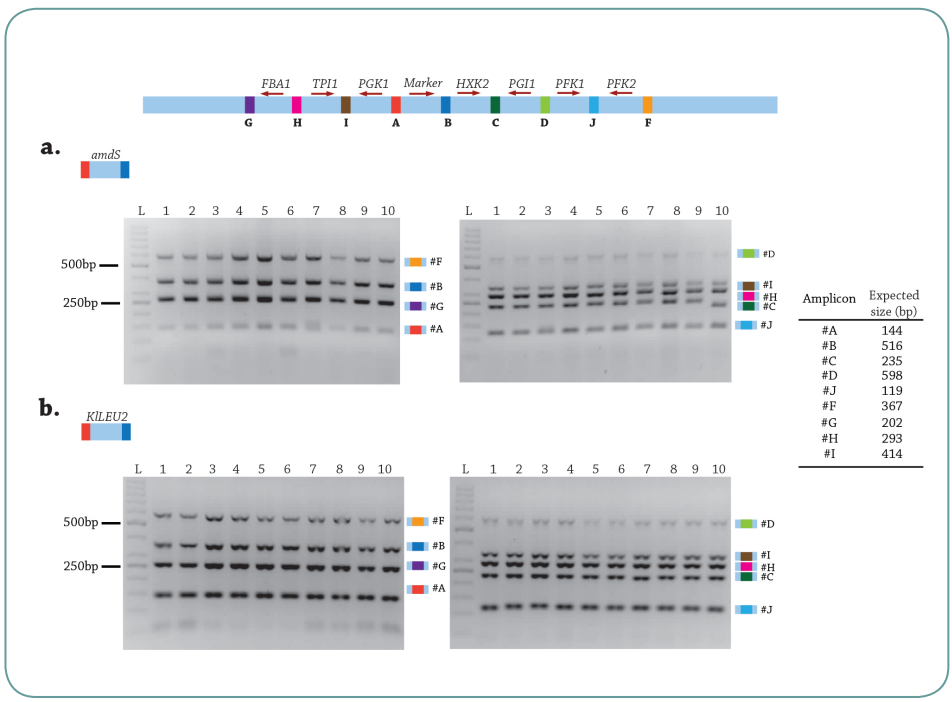


Figure 3.5 | Characterization of positive clones isolated after I-SceI-assisted CATI of ten fragments by multiplex PCR. Panel (a) PCR patterns of ten clones resulting from cotransformation of the glycolytic genes with the *amdS* selection marker, panel (b) PCR patterns of ten clones obtained by replacing *amdS* by the *KlLEU2* selection marker in the cotransformation with the glycolytic genes. Transformants were randomly picked and analyzed by multiplex PCR producing amplicons covering the indicated junctions. PCR products were separated on a 2% agarose gel by electrophoresis. In lanes labeled 'L' a 50-bp Gene Ruler ladder was loaded; sizes are indicated. From these 20 tested clones, a single one [(a), transformant number 10, amplicon C] did not display the expected pattern.

DISCUSSION

The current demand from the biotechnology industry for strains able to produce complex synthetic pathways with ever increasing productivity and robustness requires the construction of strains simultaneously expressing dozens of homologous and heterologous genes [25,26]. Genetic tools enabling the fast and efficient construction of chromosome-borne large synthetic constructs are therefore urgently needed. Comparatively little attention has been given to the development of such methods to date. One study explored the possibilities of combining recombination-based cloning with chromosomal integration in *S. cerevisiae* [6] and provided a clear proof of principle. Eight genes were assembled *in vivo* into a 23.7-Kb construct and successfully integrated into a δ-site. Using this approach, variable number of clones (30-50 clones) and efficiencies (10% to 70%) were obtained depending on the length of the overlapping sequences used, the lowest efficiency being obtained using short overlaps of 50-bp. In the present study, implementation of I-SceI-assisted combined *in vivo* assembly and targeted chromosomal integration in *S. cerevisiae* led to consistently high efficiencies of *c.* 95% for a construct of a similar size using 60-bp overlapping sequences and a similar number of fragments. Furthermore, while transformation to δ-sites inherently randomizes the location and number of integration sites, the present work is, to the best of our knowledge, the first published report of targeted integration of an *in vivo*-assembled multigene construct in *S. cerevisiae*. The SHR-sequences used in the presented platform for the assembly of the fragments add versatility to the system, which allows for parallel construction of replicative and integrative constructs. These achievements present a new step towards reliable and robust high-throughput strain construction. We are currently using the I-SceI-assisted CATI approach to assemble and integrate complete pathways up to 35-kb from 15 fragments, and the efficiency of correct assembly is similar to the efficiencies in this study. While the number of transformants seems to decrease with the number of assembled fragments (*c.* 400 clones with nine fragments and *c.* 200 clones with 15 fragments), the number of transformants obtained is sufficiently high to be compatible with high-throughput strain construction programs.

The present study indicates that the critical step in chromosomal integration and the key to the high efficiencies obtained for combined assembly and integration of a multifragment pathway is the introduction of a double-strand DNA break at the integration locus targeted. Because of its high specificity and resulting lack of toxicity in most tested organisms, I-SceI has been employed in many systems to induce site-specific double-strand breaks, such as induction of homologous recombination in higher eukaryotes [27,28], seamless gene modifications in yeast [29,30], and sequential pathway engineering in yeast [24]. In recent years, several approaches have been explored to engineer synthetic endonucleases for any recognition sequence of interest: the zinc-finger nucleases (ZFNs) [31], the TAL effector nucleases (TALENs) [32] and the recent RNA-guided CRISPR/Cas nucleases [33]. Those synthetic 'genomic scissors' could greatly contribute to further development of the CATI method for chromosomal modifications. Screening of those synthetic endonucleases for site-specific nuclease

activity in *S. cerevisiae* might reveal even more efficient DSB inducers, thereby further improving the targeting efficiency for integration. Furthermore, design of synthetic meganucleases for specific recognition sequences already present in the yeast genome could abolish the need for a defined synthetic locus to target for integration. While in the present approach the meganuclease was expressed by the host organism, it has been previously shown that endonucleases can enter yeast cells and reach the nucleus during transformation [34]. Cotransformation of cells with the desired fragments and custom-made meganucleases would make the presented method applicable to any strain without any prior modification of the genome. We therefore anticipate that the coming years will see further increases in the flexibility and ease of endonuclease-assisted CATI for strain engineering.

While *S. cerevisiae* is known for its high efficiency of homologous recombination [1,34], the present work demonstrates that integration of DNA fragments in its genome can be substantially increased by introduction of a DSB. Microbes in which it is notoriously difficult to achieve highly efficient targeted integration of DNA fragments, such as the yeast *Kluyveromyces lactis* and filamentous fungi [35-38], may similarly benefit from endonuclease-assisted integration. Although the presented method has been engineered for *S. cerevisiae*, the principle could be applied to any organism with an efficient homologous repair mechanism. Alternatively *S. cerevisiae* and its outstanding recombination efficiency could be exploited to assemble *in vivo* the DNA constructs prior to transformation to the final host.

The presented I-SceI-assisted CATI approach has drastically changed the strain engineering procedures in our laboratory and opened new possibilities for large scale metabolic engineering of *S. cerevisiae*. It is our hope that this work will further contribute to the development of *S. cerevisiae* as a valuable platform for the production of many industrially relevant compounds.

ACKNOWLEDGMENTS

We thank the TU Delft iGEM team 2009 for supplying the I-SceI BioBrick. We thank Dr. Kirsten Benjamin (Amyris, Emeryville CA) for helpful discussions. We thank the Netherlands Organisation for Scientific Research (NWO) for financial support (VIDI grant).

REFERENCES

1. Gibson DG, Benders G, Axelrod KC, Zaveri J, Algire M, Moodie M *et al.*: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome.** *Proc Natl Acad Sci USA* 2008, **105:** 20404-20409.

2. Ellis T, Adie T, Baldwin GS: **DNA assembly for synthetic biology: from parts to pathways and beyond.** *Integr Biol (Camb)* 2011, **3:** 109-118.

3. Li MZ, Elledge SJ: **Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC.** *Nat Methods* 2007, **4:** 251-256.

4. Benoit RM, Wilhelm RN, Scherer-Becker D, Ostermeier C: **An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids.** *Protein Expres Purif* 2006, **45:** 66-71.

5. Gibson DG, Young L, Chuang RY, Venter JC, Iii CAH, Smith HO *et al.*: **Enzymatic assembly of DNA molecules up to several hundred kilobases.** *Nat Methods* 2009, **6:** 12-16.

6. Shao Z, Zhao H, Zhao H: **DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways.** *Nucleic Acids Res* 2009, **37:** e16.

7. Kuijpers NGA, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran JM *et al.*: **A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences.** *Microb Cell Fact* 2013, **12:**47.

8. Zhang Z, Moo-Young M, Chisti Y: **Plasmid stability in recombinant *Saccharomyces cerevisiae*.** *Biotechnol Adv* 1996, **14:** 401-435.

9. Storici F, Durham CL, Gordenin DA, Resnick MA: **Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast.** *Proc Natl Acad Sci USA* 2003, **100:** 14994-14999.

10. Watabe H, Iino T, Kaneko T, Shibata T, Ando T: **A new class of site-specific endodeoxyribonucleases. Endo.Sce I isolated from a eukaryote, *Saccharomyces cerevisiae*.** *J Biol Chem* 1983, **258:** 4663-4665.

11. Shibata T, Watabe H, Kaneko T, Iino T, Ando T: **On the nucleotide sequence recognized by a eukaryotic site-specific endonuclease, Endo. SceI from yeast.** *J Biol Chem* 1984, **259:** 10499-10506.

12. Monteilhet C, Perrin A, Thierry A, Colleaux L, Dujon B: **Purification and characterization of the *in vitro* activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron.** *Nucleic Acids Res* 1990, **18:** 1407-1413.

13. Plessis A, Perrin A, Haber JE, Dujon B: **Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus.** *Genetics* 1992, **130:** 451-460.

14. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik M *et al.*: ***De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology.** *Microb Cell Fact* 2012, **11:**36.

15. Entian KD, Kotter P: **Yeast genetic strain and plasmid collections.** *In* Stansfield I, Stark MJR (ed) *Methods in microbiology*. Amsterdam: Academic Press; 2007:629-666.

16. Verduyn C, Postma E, Scheffers W, van Dijken JP: **Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.** *Yeast* 1992, **8:** 501-517.

17. Pronk JT: **Auxotrophic yeast strains in fundamental and applied research.** *Appl Environ Microbiol* 2002, **68:** 2095-2100.

18. Gietz RD, Woods R: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Method Enzymol* 2002, **350:** 87-96.

19. Shevchuk NA, Bryksin AV, Nusinovich YA, Cabello FC, Sutherland M, Ladisch S: C**onstruction of long DNA molecules using long PCR-based fusion of several fragments simultaneously.** *Nucleic Acids Res* 2004, **32(2)**:e19. 2004.

20. Whelan WL, Gocke E, Manney TR: **The *CAN1* Locus of *Saccharomyces cerevisiae*: Fine-structure analysis and forward mutation rates.** *Genetics* 1979, **91:** 35-51.

21. Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT *et al.*: ***amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:** 126-139.

22. Orr-Weaver TL, Szostak JW, Rothstein RJ: **Yeast transformation: a model system for the study of recombination.** *Proc Natl Acad Sci USA* 1981, **78:** 6354-6358.

23. Leem SH, Noskov VN, Park JE, Kim SI, Larionov V, Kouprina N: **Optimum conditions for selective isolation of genes from complex genomes by transformation-associated recombination cloning.** *Nucleic Acids Res* 2003, **31(6)**:e29.

24. Wingler LM, Cornish VW: **Reiterative Recombination for the *in vivo* assembly of libraries of multigene pathways.** *Proc Natl Acad Sci USA* 2011, **108:** 15135-15140.

25. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM *et al.*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440:** 940-943.

26. Koopman F, Beekwilder J, Crimi B, van HA, Hall RD, Bosch D *et al.*: ***De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*.** *Microb Cell Fact* 2012, **11:**155.

27. Choulika A, Perrin A, Dujon B, Nicolas JF: **Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1995, **15:** 1968-1973.

28. Rouet P, Smih F, Jasin M: **Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.** *Mol Cell Biol* 1994, **14:** 8096-8106.

29. Khmelinskii A, Meurer M, Duishoev N, Delhomme N, Knop M: **Seamless gene tagging by endonuclease-driven homologous recombination.** *PLoS One* 2011, **6(8):**e23794.

30. Noskov VN, Segall-Shapiro TH, Chuang RY: **Tandem repeat coupled with endonuclease cleavage (TREC): a seamless modification tool for genome engineering in yeast.** *Nucleic Acids Res* 2010, **38:** 2570-2576.

31. Kandavelou K, Mani M, Durai S, Chandrasegaran S: **"Magic" scissors for genome surgery.** *Nat Biotechnol* 2005, **23:** 686-687.

32. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A *et al.*: **Targeting DNA double-strand breaks with TAL effector nucleases.** *Genetics* 2010, **186:** 757-761.

33. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N *et al.*: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339:** 819-823.

34. Schiestl RH, Petes TD: **Integration of DNA fragments by illegitimate recombination in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 1991, **88:** 7585-7589.

35. Kooistra R, Hooykaas PJ, Steensma HY: **Efficient gene targeting in *Kluyveromyces lactis*.** *Yeast* 2004, **21:** 781-792.

36. Ninomiya Y, Suzuki K, Ishii C, Inoue H: **Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining.** *Proc Natl Acad Sci USA* 2004, **101:** 12248-12253.

37. Krappmann S, Sasse C, Braus GH: **Gene targeting in *Aspergillus fumigatus* by homologous recombination is facilitated in a nonhomologous end-joining-deficient genetic background.** *Eukaryot Cell* 2006, **5:** 212-215.

38. Snoek IS, van der Krogt ZA, Touw H, Kerkman R, Pronk JT, Bovenberg RA *et al.*: **Construction of an hdfA *Penicillium chrysogenum* strain impaired in non-homologous end-joining and analysis of its potential for functional analysis studies.** *Fungal Genet Biol* 2009, **46:** 418-426.

39. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C *et al.*: **An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains.** *Enzyme Microb Technol* 2000, **26:** 706-714.

40. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.** *Nucleic Acids Res* 2002, **30(6):**e23. 2002.

41. Alberti S, Gitler AD, Lindquist S: **A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*.** *Yeast* 2007, **24:** 913-919.

Chapter 4

Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand breaks in *Saccharomyces cerevisiae*.

Solis-Escalante D, Kuijpers NGA, van der Linden FH, Pronk JT, Daran JM, Daran-Lapujade P

Published in FEMS Yeast Research (2014) 14:5 741–754

ABSTRACT

Large strain construction programs and functional analysis studies are becoming commonplace in *Saccharomyces cerevisiae* and involve construction of strains that carry multiple selectable marker genes. Extensive strain engineering is, however, severely hampered by the limited number of recyclable marker genes and by the reduced genome stability that occurs upon repeated use of heterologous recombinase-based marker removal methods. The present study proposes an efficient method to recycle multiple markers in *S. cerevisiae* simultaneously, thereby circumventing shortcomings of existing techniques and substantially accelerating the process of selection-excision. This method relies on artificial generation of double strand breaks around the selection marker cassette by the meganuclease I-SceI and the subsequent repair of these breaks by the yeast homologous recombination machinery, guided by direct repeats. Simultaneous removal of up to three marker cassettes was achieved with high efficiencies (up to 56%), suggesting that I-SceI-based marker removal has the potential to co-excise an even larger number of markers. This locus- and marker-independent method can be used for both dominant and auxotrophy-complementing marker genes. Seven pDS plasmids carrying various selectable markers, which can be used for PCR-based generation of deletion cassettes suited for I-SceI marker recycling, are described and made available to the scientific community.

INTRODUCTION

Over the past decades, fundamental and application-oriented research on the budding yeast *Saccharomyces cerevisiae* has seen a tremendous intensification. Popularity of this yeast is readily explained from its robustness, ease of cultivation and remarkable genetic accessibility. The rapidly expanding molecular genetic toolbox for *S. cerevisiae* has been successfully applied for the production of a large variety of endogenous and heterologous products of interest [1-6].

Strain construction requires selection marker genes to screen for successful introduction of mutations after each genetic manipulation. Large strain construction programs and functional analysis studies, which are becoming common practice in *S. cerevisiae*, often involve strains that carry multiple selection marker genes [7]. In such cases, progress is hampered by the relatively small number of selection markers available [8]. Furthermore, while it is of paramount importance that selection markers do not interfere with yeast physiology, expression of selectable markers can have deleterious effects on the host [9-11]. To avoid interference by selection markers, marker-free strains are preferred in both academia and industry.

To answer these needs, clever systems for marker removal have been designed in the past decades and fall in two categories. One approach to marker recycling relies on the expression of heterologous recombinase systems [12,13], the most intensively used being the Cre/*Lox*P system [12,14,15]. Specific recognition sites surrounding the selection marker are recombined by P1 bacteriophage Cre recombinase, thereby leading to marker excision. While external recombinase-based systems are highly efficient and have considerably enhanced *S. cerevisiae*'s molecular toolbox, they have an important limitation. Each recombination catalysed by Cre leaves a scar composed of the recombinase recognition site. When used repeatedly, for instance in serial gene deletion experiments, scars spread over the chromosomes promote recombination upon Cre induction, resulting in chromosomal translocations [16]. While a few mutated recognition sequences have been engineered to prevent the occurrence of unwanted genomic rearrangements [17], this instability nevertheless limits the potential of external recombinase-based systems for extensive strain construction programs. A second approach relies on the homologous recombination (HR) machinery of the host. HR and nonhomologous end joining (NHEJ) are the two processes for maintenance of genome integrity after DNA damage, such as double strand breaks (DSB), in most eukaryotic cells, HR being the preferred repair mechanism in *S. cerevisiae* [18,19]. Starting decades ago, the power of yeast HR has been harnessed by the scientific community for targeted genome modifications [20,21], and was used in the first method for marker recycling in *S. cerevisiae* [22]. By flanking a marker gene with repeated sequences and cultivating mutants in nonselective media, it was observed that mitotic recombination could remove the marker, albeit at a low frequency (10^{-4} to 10^{-3}). Cells that underwent this process can be easily screened using negative selection or counterselection [23]. This approach requires the availability of a growth condition under which the presence of the selection marker is lethal and the presence of direct repeats flanking the marker to enable mitotic recombination and thereby marker

excision. For instance, incubation of mutants carrying the most popular recyclable marker *URA3*, encoding orotidine-5′-phosphate decarboxylase involved in pyrimidine biosynthesis, in medium containing 5-fluoroorotic acid (5-FOA) is lethal and suitable for direct selection of marker excision. Although over 20 selectable markers have been developed for budding yeast, only a handful of them can thus be counterselected: *URA3* [23], *LYS2* [24], *MET15* [25], *TRP1* [26] and *amdSYM* [8]. Methods enabling the efficient and scarless removal of any marker are therefore highly relevant for advanced genetic engineering of *S. cerevisiae*.

Based on the knowledge that DSB promote recombination, artificially created DNA breaks were used to remove the auxotrophic marker *URA3* [27]. Artificial DSB can be created *in vivo* by expression of highly specific endonucleases and upon integration of their recognition sequence in the host genome. HO and I-SceI are the most widely used [28], however, as HO can induce the mating type switch [29], I-SceI is the preferred nuclease. I-SceI is a mitochondrial homing endonuclease encoded by the *S. cerevisiae* mitochondrial genome and recognizes an 18 bp recognition sequence with high affinity [30,31]. Fairhead and co-workers flanked the *URA3* gene with I-SceI recognition sequences, thereby creating DSBs around the marker upon *in vivo* expression of *SCEI* gene [27]. As DSB is one of the most lethal DNA damages, only cells that can repair the DSBs via the NHEJ pathway will survive in this approach. Indeed, while few clones survived the DSBs (5% of the clones), the vast majority of these survivors had lost *URA3*. HR being the prevalent mechanism for DSB repair in *S. cerevisiae*, the presence of repeats nearby the created DSB should promote homologous repair and increase the cells survival rate. Accordingly, it has been shown that flanking the marker *URA3* by direct repeats and inducing the formation of an artificial double strand break close to a single repeat increased the efficiency of *URA3* counter-selection [32-34]. While hitherto never reported, this efficient combination of DSB and direct repeats-aided HR opened the way to marker excision in the absence of negative selection and to the potentiality of excising multiple markers in a single step.

This present study describes the construction of a new set of standardized plasmids compatible with the widely used pUG series [14,35], which exploits the potential of I-SceI induced artificial DSB combined with HR using direct repeats to promote selection marker excision without the need of counter-selection and minimizing the risk of chromosomal translocation. Furthermore, while current HR-based systems are restricted to the excision of a single marker, methodologies enabling the excision of multiple markers in one step would tremendously accelerate strain construction programs. Multiple marker excision has until now only being achieved using Cre/LoxP system, yielding low efficiencies, eight out of 96 clones [36], and promoting chromosome translocations [16]. Thus, the ability of I-SceI-induced DSB to recycle multiple markers simultaneously was investigated.

MATERIALS AND METHODS

Strains and media | Plasmid propagation was performed with chemically competent *Escherichia coli* DH5α (Z-competent™ transformation kit, Zymo research, Orange, CA) cultivated in lysogeny broth (LB) media [37,38], supplemented with 100 mg L^{-1} ampicillin (LBAmp) when required. *Escherichia coli* strains were kept in 30% (v/v) glycerol vials at -80°C. All *S. cerevisiae* strains used in this study were derived from the CEN.PK strain family [39-41] and are listed in Table 4.1. and Table S1. All yeast strains were grown in nonselective media for *c.* 16h and then stored in 30% (v/v) glycerol vials and kept at -80°C. Under non-selective conditions, complex media containing 10 g L^{-1} yeast extract, 20 g L^{-1} peptone and 20 g L^{-1} glucose (YPD) were used. For the selection of mutants and replica plating, YPD was supplemented with one of the following selectable agents, 200 mg L^{-1} G418 (YPD+G418), 100 mg L^{-1} nourseothricin (YPD+Clonat) or 200 mg L^{-1} hygromycin (YPD+Hyg). Selection of strains harboring the plasmid pUDC073 (*$GAL1_{pr}$-SCEI, URA3*) [42] were obtained by growing the transformation mix on synthetic medium (SM) containing 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4.7H_2O$, 5 g L^{-1} $(NH_4)_2SO_4$, 1 mL L^{-1} of a trace element solution as previously described [43], 1 mL L^{-1} of a vitamin solution [43] and supplemented with 20 g L^{-1} glucose (SMGlu). SMGlu was supplemented with 20 mg L^{-1} adenine when required. All yeast transformations were performed with the lithium acetate protocol [44]. When solid medium was required, 20 g L^{-1} of agar was added to the different media. Incubation of liquid-medium cultures was performed in orbital shakers at 30°C and at 250 rpm.

Construction of the pDS plasmid series | As deletion cassettes need to harbor several features specific to I-SceI-aided marker removal, a series of plasmids named pDS, carrying various tailor-made cassettes with several auxotrophic and dominant selection markers, was constructed. In addition to a selection marker, the deletion cassettes in these plasmids carried the I-SceI recognition site and repeats to facilitate HR after marker excision.

To construct these cassettes, two sets of PCRs were performed on the popular pUG plasmid series. A first set of PCRs, performed with the primer pair pDS-FW/pDS-RV, was used to amplify the marker genes from pUG6 [14], pUG-amdSYM [8], pUG73 [35], pUG66 [35], pUG-natNT1 [45] and pUG-hphNT1 [46] respectively (Fig. 4.1.). From 5' to 3', pDS-FW contained four regions, (i) the BglII recognition site, (ii) a 40 bp sequence obtained by combining the tags 5'-ATGACAAGAGGGTCGAACTC-3' and 5'-GCCTAAGTCGTAATTGAGTC-3' from the bar-coded sequences in the yeast deletion project [47] to produce the section B of the synthetic repeats, (iii) a 30 bp sequence containing the 18 bp I-SceI recognition site 5'-AGTTACGCTAGGGATAACAGGGTAATATAG-3' and (iv) a sequence for binding on a pUG plasmid upstream of the marker's promoter. pDS-RV also contained four regions, (i) the XbaI recognition site, (ii) a 40 bp sequence obtained by combining the tags 5'-CCGCCAAGCGAATTGAAGGA-3' and 5'-CCGTGCGTAGAATGAAGAAC-3' from the bar-coded sequences in the yeast deletion project [47] to produce the section A of the synthetic repeats, (iii) a 30 bp sequence containing the 18 bp I-SceI recognition site 5'-AGTTACGCTAGGGATAACAGGGTAATATAG-3' and (iv) a sequence for binding in a

TABLE 4.1 | Strains used in the study.

Strain Name	Features	Genotype	Reference
CEN.PK113-7D	Control strain	*MAT*a *MAL2-8c SUC2 URA3 ADE2 CAN1*	[39-41]
IMY088	One marker, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
IMY089	One marker, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
IMY099	Two markers, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB ade2::AB-ISceI$_{rs}$-natMX-ISceI$_{rs}$-AB* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
IMY100	Two markers, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$ ade2::ISceI$_{rs}$-natMX-ISceI$_{rs}$* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
IMY090-0	Three markers, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB ade2::AB-ISceI$_{rs}$-natMX-ISceI$_{rs}$-AB can1::AB-ISceI$_{rs}$-nph-ISceI$_{rs}$-AB* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
IMY100-1	Three markers, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$ ade2::ISceI$_{rs}$-natMX-ISceI$_{rs}$ can1::ISceI$_{rs}$-nph-ISceI$_{rs}$* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study
CEN.PK113-5D	Control strain	*MAT*a *MAL2-8c SUC2 ura3-52ADE2 CAN1*	[39,41]
IMC078	Control strain	*MAT*a *MAL2-8c SUC2 ura3-52* pUDC073*(CEN6/ARS4 ori URA3 GAL1$_{pr}$-SCEI-CYC1$_{ter}$)*	This study

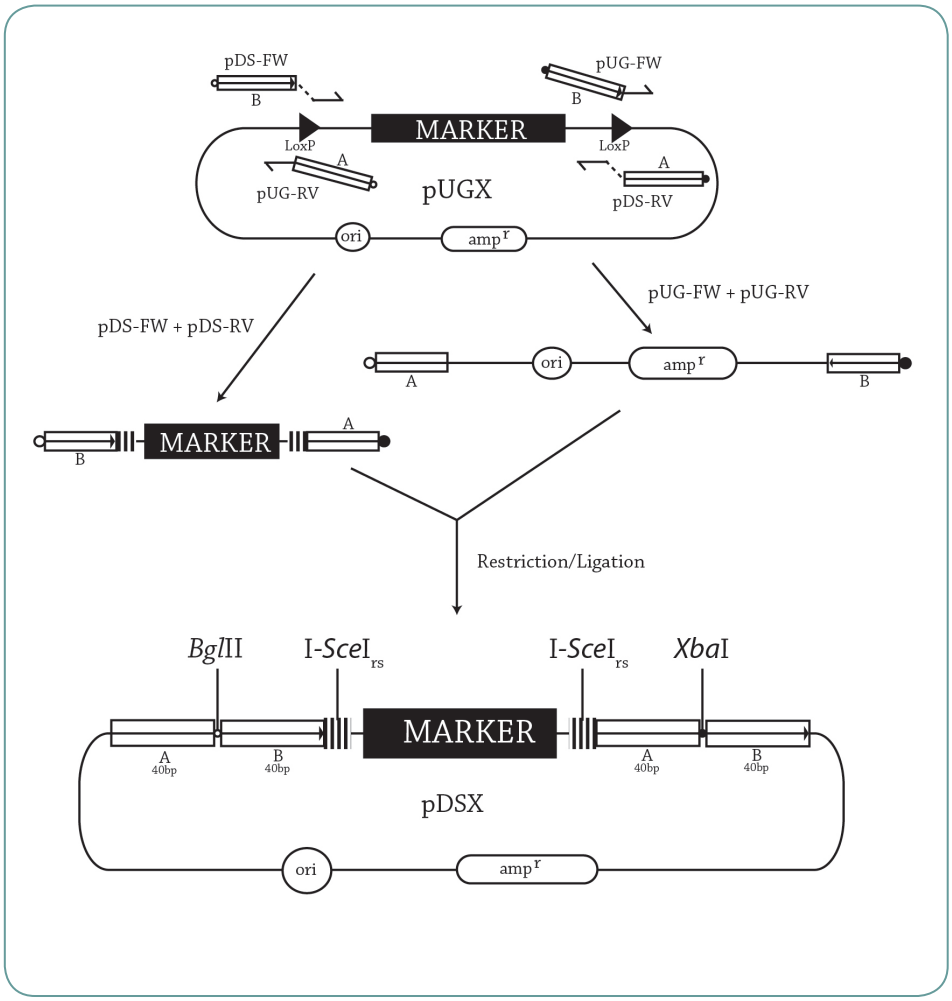


FIGURE 4.1 | Schematic representation of the construction of the pDS plasmid series. The primer pair pDS-FW/pDS-RV (or pDS4-FW/pDS4-RV in the case of pDS4) was used to amplify the selectable marker and add the I-SceI recognition sequence together with the 40 bp A and B sequences and to incorporate the restriction sites for BglII and XbaI (BamHI and EcoRI in the case of pDS4). The primer pair pUG-FW/pUG-RV (pUG4-FW/pUG-RV for pDS4) was used to amplify the plasmid backbone, add BglII and XbaI (BamHI and EcoRI in the case of pDS4) restriction sites and incorporate the 40 bp A and B sequences. The templates for PCR were the corresponding pUG plasmids [8,35,45,46]. After restriction and ligation seven different pDS plasmids were generated. All plasmids contain a selectable marker flanked by I-SceI restriction sites and the synthetic repeats AB for marker excision. The main function of the pDS plasmids is to serve as PCR template for the construction of deletion cassette for I-SceI-aided marker excision. Depending of the primer design, the deletion cassettes can contain synthetic AB repeats or seamless (Sm) repeats (see materials and methods section).

pUG plasmid downstream the selection marker.
A second set of PCRs, performed with the primer set pUG-FW/pUG-RV on the same plasmid series, aimed at amplifying the plasmid backbone while incorporating the necessary additional sequences. pUG-FW contained the XbaI recognition site, the section A of the synthetic repeats and a sequence for binding to a pUG plasmid downstream the *Lox*P site, while pUG-RV contained the BglII recognition site, respectively, the section B of the synthetic repeats and a sequence for binding to a pUG plasmid upstream a *Lox*P site.

All pDS plasmids were constructed using the mentioned primer sets with the exception of pDS4. For pDS4, primers sets pDS4-FW/pDS4-RV and pUG4-FW/pUG4-RV were used and resulted in the integration of the BamHI and EcoRI restriction sites instead of the BglII and XbaI sites, respectively. In all other aspects the pDS4 plasmid was identical to the other plasmids from the pDS series. All the above-mentioned PCRs were performed with Phusion Hot Start Polymerase (Finnzymes, Espoo, Finland).
Gel-purified PCR fragments (Zymoclean™ Gel DNA Recovery Kit, Zymo research) were restricted (FastDigest, Thermo Scientific, Waltham, MA) and ligated (T4 DNA polymerase, Thermo Scientific). The ligation mix was used to transform *E. coli* DH5α chemically competent cells. Transformed cells were plated on LBAmp for selection. Single colonies were used for plasmid isolation (GenElute™ Plasmid Miniprep Kit, Sigma-Aldrich, St. Louis, MO). Plasmids were verified by restriction analysis and Sanger 24H full sequencing (Baseclear, Leiden, The Netherlands). All primers for plasmid construction and sequencing are listed in Table S2. The plasmid series pDS is composed of pDS1 (*KanMX*), pDS2 (*amdSYM*), pDS3 (*KlURA3*), pDS4 (*KlLEU2*), pDS5 ($AgTEF2_{pr}$-*ble*-$AgTEF2_{ter}$), pDS6 ($AgTEF2_{pr}$-*natNT1*-$AgTEF2_{ter}$) and pDS7 ($AgTEF2_{pr}$-*hphNT1*-$AgTEF2_{ter}$) (Fig. 4.2.). A list of all plasmids used in this study is in Table S3. The pDS plasmid series is available to the research community at Euroscarf (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/) with accession numbers P30769 (pDS1), P30770 (pDS2), P30771 (pDS3), P30772 (pDS4), P30773 (pDS5), P30774 (pDS6) and P30775 (pDS7).

Deletion cassettes construction | All cassettes for gene deletion were constructed by PCR amplification using a plasmid from the pDS series as template (Phusion Hot Start Polymerase, Finnzymes). Using different primer sets, two types of deletion cassettes were constructed, termed AB and seamless (Sm) cassettes (Fig. S1). For marker excision using synthetic AB repeats, PCRs were performed with primer sets binding outside the AB repeats. The resulting PCR-amplified cassette therefore contained seven elements: 1) A selection marker, 2) two I-SceI recognition sequences ($ISceI_{rs}$) flanking the selection marker, 3) two AB repeats flanking the marker and recognition sites 4) 98 or 100 bp flanking sequences homologous to the targeted deletion site. To perform seamless marker removal, deletion cassettes were amplified with different primers sets binding just outside the I-SceI recognition site and were therefore devoid of AB repeats (Fig. S1). The Sm cassettes contained six elements: 1) A selection marker, 2) two $I\text{-}SceI_{rs}$ flanking the selection marker, 3) a 40 bp sequence homologous to the upstream region of the targeted locus and 4) 60 or 100 bp sequences homologous to the targeted genomic locus. All primers used for deletion cassette construction are listed in Table S2.

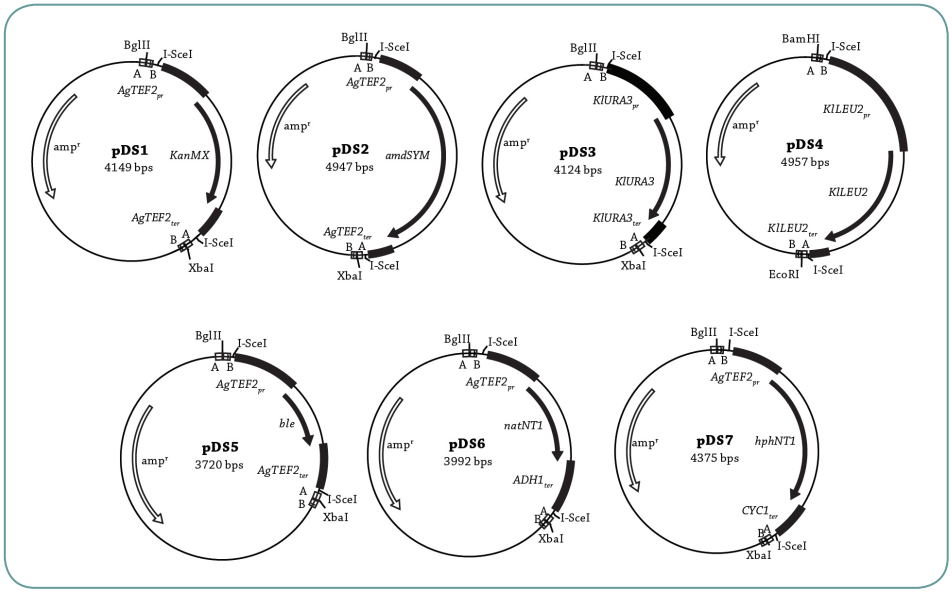


FIGURE 4.2 | Maps of the pDS plasmid series. All plasmids contain a marker gene flanked by I-SceI recognition sites and the synthetic AB repeats. Since all plasmids backbones are the same, the primers designed for a pDS plasmid can be used in the other pDS plasmids. The markers available are *KanMX* (pDS1), *amdSYM* (pDS2), *KlURA3* (pDS3), *KlLEU2* (pDS4), *ble* (pDS5), *natNT1* (pDS6) and *hphNT1* (pDS7).

Marker excision induced by artificial DSB | In strains containing one, two or three gene deletions and harboring the plasmid pUDC073, expression of *SCEI* was induced by incubating mutant yeast cells overnight in 100 mL of liquid SM medium supplemented with 20 g l^{-1} galactose (SMGal), and incubated at 30°C and 250 rpm. *C.* 200 cells from that culture were inoculated on SMGal agar plates and incubated for two days at 30°C. For seamless marker excision, prolonged incubation on galactose was performed by an additional transfer to new SMGal agar plates followed by incubation for another 2 days.

To repress *SCEI* expression, 100 colonies from the agar plates were pooled and suspended in 1 mL sterile demineralized water. An aliquot of this suspension with *c.* 100 cells was spread on SMGlu agar plates and incubated for 2 days at 30°C. Subsequently 50 single colonies were incubated in liquid complex medium (YPD) and restreaked on SMGlu agar plates containing 1 g L^{-1} 5-fluoroorotic acid (5-FOA) and 150 mg L^{-1} uracil (SM-5FOA) to select for colonies lacking plasmid pUDC073 (*GAL1*$_{pr}$*-SCEI, URA3*). Efficiency of marker excision was calculated by replica plating 50 single colonies from the SM-5FOA agar plate onto agar plates containing YPD and one or several of the following media: YPD+G418, YPD+Clonat and YPD+Hyg. Each individual colony was suspended in 100 µL sterile demineralized water and replica plated in the appropriate medium. All plates were incubated at 30°C for two days. Ten colonies for which the

plates suggested full marker removal were analyzed by PCR and a single colony was sequenced (Baseclear, Leiden, The Netherlands). As a negative control, the strain CEN. PK113-5D was transformed with the plasmid pUDC073 harboring *SCEI*, generating strain IMC078. This strain was grown as described for the mutant strains. SMGal, SM-5FOA and SMGlu were supplemented with 20 mg L^{-1} adenine when required.

Chromosome separation using Contour Clamped Homogeneous Electric Field (CHEF) electrophoresis | Agarose plugs for strains E3, E2, G3 and IMY090-0 were prepared (CHEF yeast genomic DNA plugs Kit, Bio-Rad, Richmond, CA) following the manufacturer's recommendations and used for CHEF electrophoresis. The plugs were placed in a 1% megabase agarose in TBE buffer (5.4 g trizma base, 2.75 g boric acid, 2 mL of 5 M EDTA pH 8.0 and 1 L demineralized water) gel. For chromosome separation, the CHEF-DRIII Pulsed Field Electrophoresis System (Bio-Rad) was used following the manufacturer's recommendations.

RESULTS

Design of the I-SceI-based marker excision method and construction of a pDS plasmid series | The proposed methodology for marker excision relies on the artificial and targeted creation of DSB by I-SceI in the genomic DNA surrounding a selection marker and the subsequent repair of this DSB by HR of dedicated repeats built into the deletion cassette. To achieve this, the selection marker needs to be flanked on each side by the I-SceI recognition site and by DNA repeats. To reduce the complexity of deletion cassette construction with different marker genes and to standardize the process, a series of seven plasmids termed pDS was designed and constructed. The dominant markers *KanMX* [22], *amdSYM* [8], *ble* [48], *natNT1*[49] and *hphNT1* [50] and the auxotrophy-complementing markers *KlURA3* [23,51,52] and *KlLEU2* [53,54], were selected for the pDS series (Fig. 4. 2.). Eighteen base pair I-SceI recognition sites ($ISceI_{rs}$) were incorporated to create DSBs upon induction of the homing endonuclease. In order to facilitate DSB repair by HR, 86 bp direct repeats termed AB repeats were designed and incorporated into the pDS series. AB repeats are located at both flanks of the marker module and are composed of two sections, A and B, separated by BglII or XbaI restriction sites. Up and downstream AB repeats present a high level of identity (93%) that will allow repair by HR because complete identity is not necessary for this process [55]. Seamless repair can easily be achieved by leaving out the AB repeats from the deletion cassette and adding a 40 bp sequence homologous to the upstream section of the targeted locus. This approach will generate direct repeats upon integration (Sm repeats, Fig. S1) as previously described [8,32,56]. As the pDS plasmids are based on the widely used pUG series, primers designed for pUG plasmids [14,35] can also be used for the pDS plasmid series. Expression of *SCEI*, carried by a centromeric vector, was driven by the galactose inducible *GAL1* promoter (pUDC073, [42]).

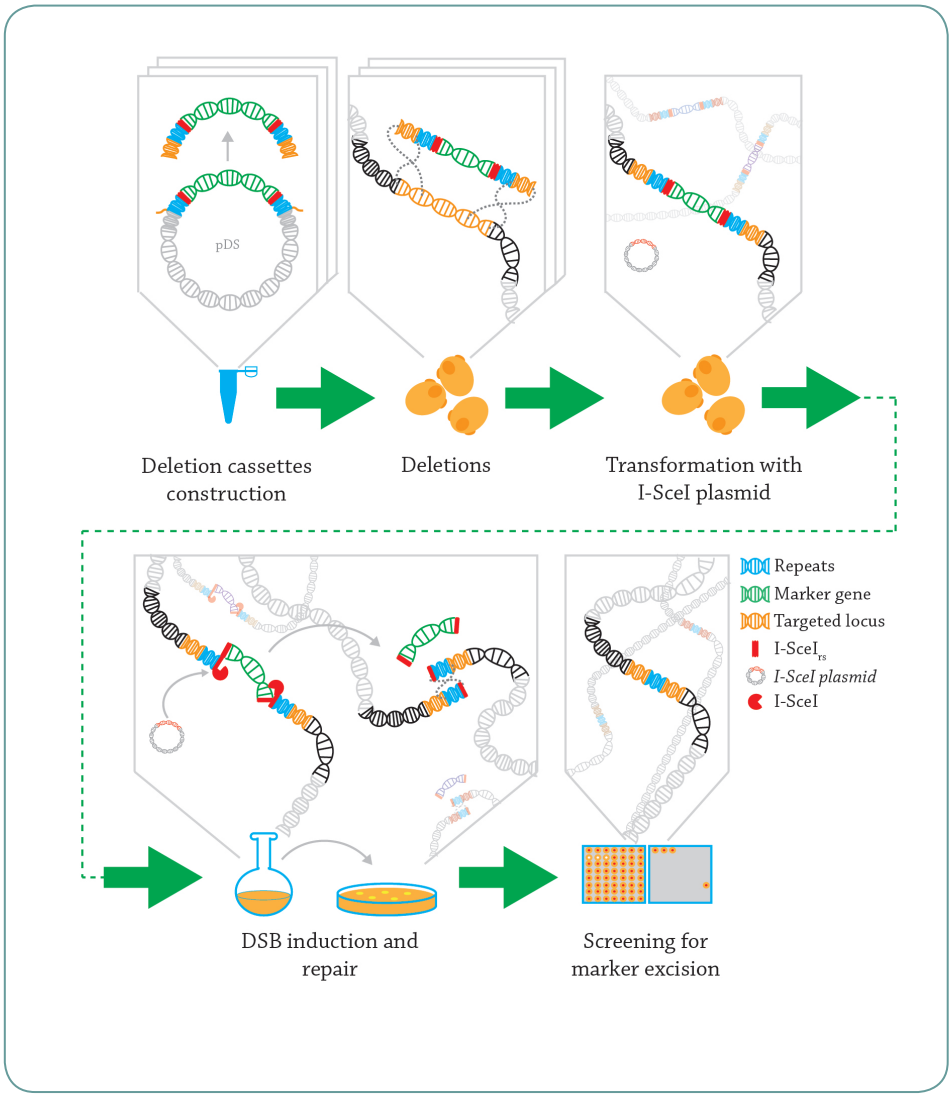


FIGURE 4.3 | Schematic overview of the I-SceI-aided marker removal. The five stages of the I-SceI-aided efficient marker excision: 1) Construction, by PCR, of deletion cassettes containing three key elements: a marker gene, I-SceI recognition sites (I-SceI$_{rs}$) and repeats by, using the pDS plasmids as templates, 2) yeast transformation to generate deletion mutants, 3) yeast transformation with the *SCEI*-harbouring plasmid, 4) induction of *SCEI* by growth on galactose, thereby inducing double strand breaks flanking the marker gene. The yeast DSB repair machinery will recognize the repeats and repair the resulting gap by homologous recombination between the provided repeats. Thus, the marker cassette will be replaced by a synthetic repeat or be seamlessly removed from the genome. 5) screening for marker-free strains by phenotypic analysis and PCR.

Using the above-described plasmids, the I-SceI-dependent marker recovery approach consisted of five steps (Fig. 4.3.). First, the deletion cassettes (containing AB or Sm repeats) are constructed by PCR amplification using a plasmid of the pDS series as template. Secondly, *S. cerevisiae* is transformed with a deletion cassette, leading to the deletion of the targeted locus via HR. These steps can be repeated using different selection markers in the different pDS plasmids, leading to the deletion of multiple loci and the integration in the genome of multiple markers. Thirdly, the deletion mutants are transformed with pUDC073 carrying *SCEI*. Fourthly, these strains are incubated in media containing galactose as sole carbon source so as to induce *SCEI* expression and thereby trigger DSB formation at the I-SceI restriction sites (I-SceI$_{rs}$). During this step, I-SceI generates DSB on both sides of the selectable marker gene, between the marker and each direct repeat, leading to marker excision. The endogenous yeast HR machinery is recruited at the DSB site and repairs the DSB by recombining AB or Sm direct repeats. Finally, mutants devoid of markers are screened on selective media.

Efficient marker excision induced by artificial DSB | To evaluate the efficiency of marker excision induced by artificial DSB in combination with direct repeats, two cassettes for deletion *URA3* in *S. cerevisiae* strain CEN.PK113-7D were constructed using pDS1 (*KanMX*) as template. One deletion cassette termed AB contained the synthetic AB repeats for DSB repair, while a second cassette, Sm, was constructed to enable seamless repair. Yeast cells were transformed with one of these cassettes and grown on selective agar plates. After selection and verification by PCR of gene deletion, the obtained strains were transformed with the plasmid pUDC073 [42] containing the *SCEI* gene. This generated the mutants IMY089 (*ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB* pUDC073(*GAL1$_{pr}$-I-SCEI, URA3*)) and IMY088 (*ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$* pUDC073(*GAL1$_{pr}$-I-SCEI, URA3*)). To generate DSBs *in vivo*, expression of *SCEI* was induced by growing IMY089 and IMY088 directly from glycerol stocks in media containing galactose as sole carbon source. Induction of *SCEI* in IMY089 and IMY088 should lead to the excision of the *KanMX* marker and therefore sensitivity to G418. After growth on galactose, 50 colonies were selected for each of the two transformations to check for marker excision and to evaluate the frequency of this event. Following the removal of the pUDC073 plasmid (*GAL1$_{pr}$-SCEI, URA3*), galactose-induced colonies of IMY089 and IMY088 were analysed by replica plating on selective and nonselective media (Fig. 4.4.). While noninduced controls fully retained G418 tolerance (Fig. 4.4., green box), 98% (49 of 50) and 54% (27 of 50) of the colonies tested with AB and Sm repeats respectively showed a loss of G418 resistance, suggesting the removal of the marker module *KanMX*. Marker excision was confirmed in ten randomly picked colonies (Fig. 4.4., circled colonies) by PCR. By sequencing the *URA3* locus of two colonies, one for AB and one for Sm repeats, the replacement of the marker by the corresponding repeat was confirmed. These results demonstrated that a dominant selectable marker gene can be efficiently excised without the need for counter-selection.

The length of exposure to galactose medium influenced the efficiency of marker removal. As few as five of 25 clones showed the phenotype characteristic of marker excision after incubation in liquid galactose medium for *c.* 16 h. However, prolonged incubation with galactose by inoculation on solid media resulted in substantially higher efficiencies, with 24 of 25 clones having lost their G418 resistance. In addition,

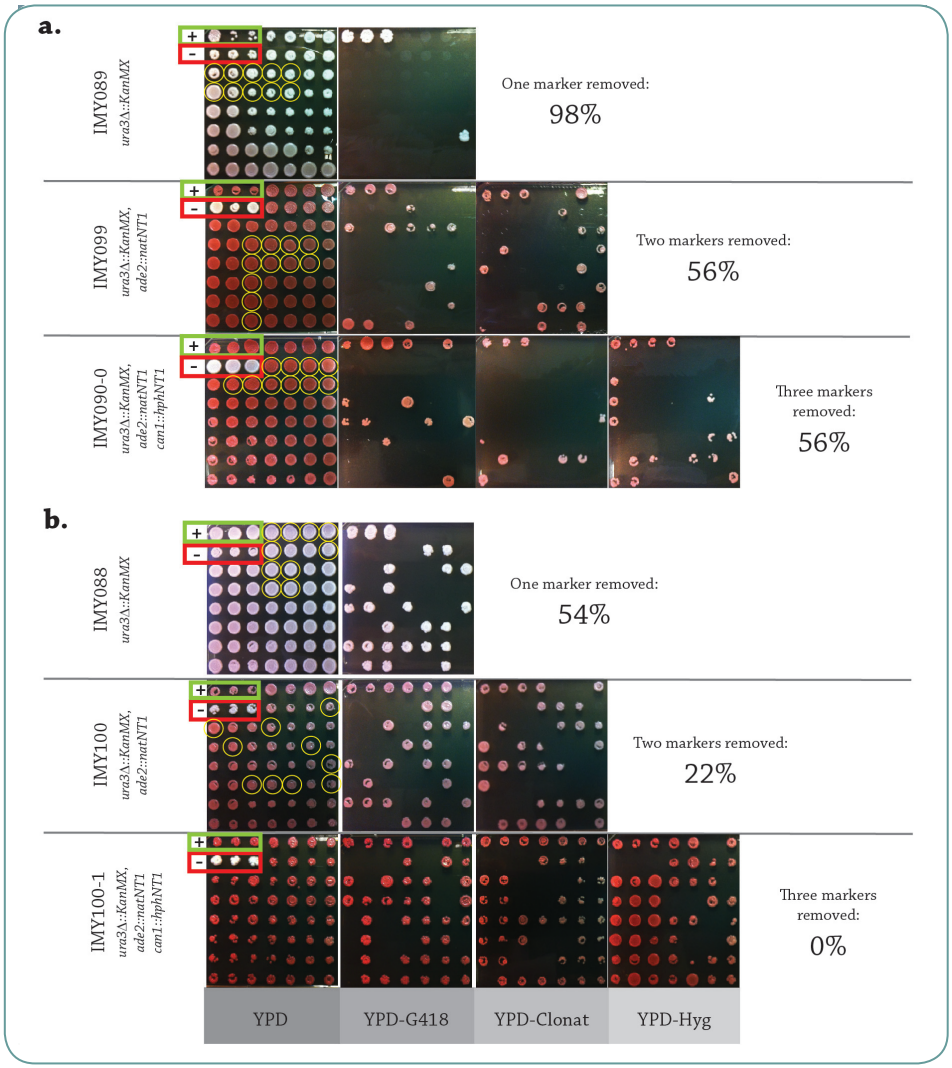


FIGURE 4.4 | Phenotypic evaluation of the efficiency of single, double and triple marker excision. Panel a, strains carrying AB repeats. Panel b, strains with seamless repeats. This analysis estimated the number of colonies in which marker excision occurred after induction of I-SceI. 50 colonies of strains with one, two or three deletions incubated in the presence of galactose were replica plated on complex non-selective media (YPD, 1st column) and selective media: YPD-G418 containing G418 (screening medium for *KanMX*, 2d column), YPD-Clonat containing nourseothricin (Screening medium for *natNT1*, 3rd column) and YPD-Hyg containing hygromycin (Screening medium for *hphNT1*, 4th column). Inability to grow on selective media indicates the absence of the corresponding selectable marker. For each strain, three colonies of non-induced cultures were used as positive controls (+, green). Ten colonies of each strain that showed a phenotype linked to the removal of all markers after induction were analysed by PCR to confirm the marker excision (circled colonies). Three colonies from the strain IMC078 containing the plasmid pUDC073 carrying I-SceI (*GAL1*$_{pr}$*-SCEI URA3*) but lacking I-SceI recognition sites were used as negative control (-, red).

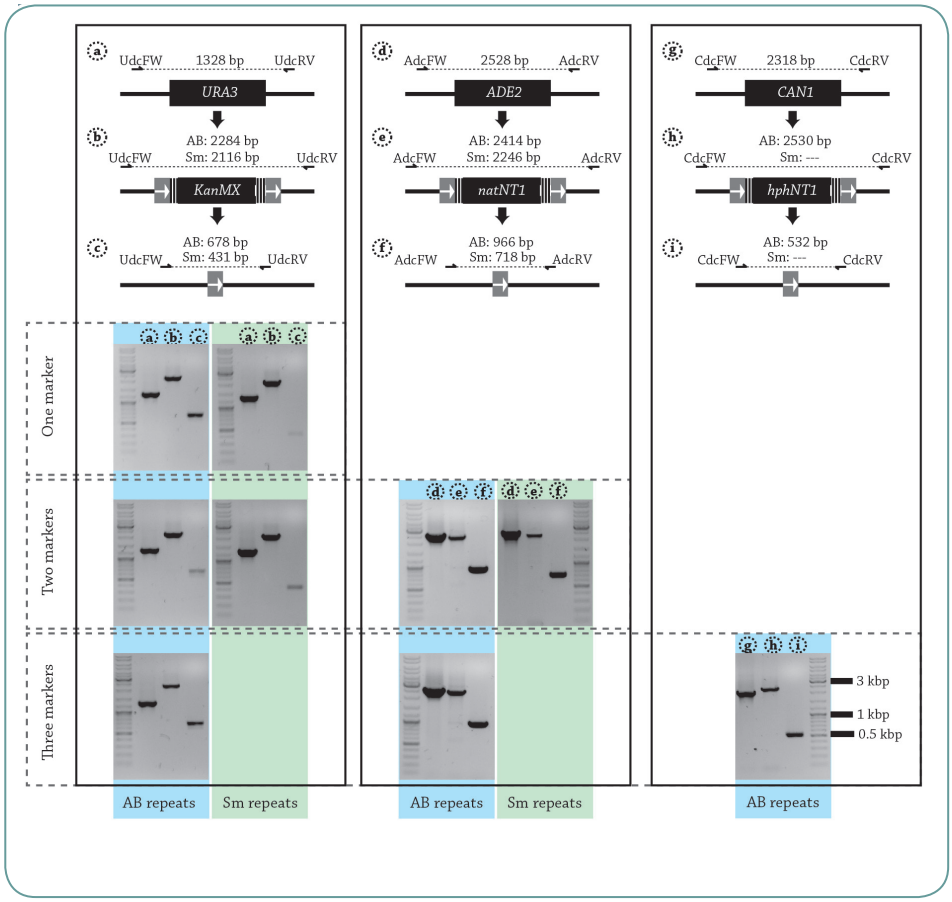


FIGURE 4.5 | PCR analysis on clones with one, two or three removed marker genes. Primers binding up- and down-stream the targeted loci were used to test gene deletion and marker excision via AB or seamless (Sm) repeats. Expected PCR products for the *URA3*, *ADE2* and *CAN1* loci in CEN. PK113-7D were obtained for the native loci (a, d and g). Integration of the constructed deletion cassettes in the targeted loci was also confirmed by presence of the expected PCR products (b, e and h). Lanes c, f and i display correct PCR patterns for the removal of one, two and three markers using AB repeats and of one and two markers simultaneously with Sm repeats.

when seamless repeats were used, a similar efficiency was only obtained after a second transfer on solid media. In the present work, galactose induction was therefore performed on a combination of liquid and solid media, experiments for seamless repair requiring two sequential transfers to solid media.

Efficient excision of multiple markers in a single step induced by artificial DSB | To explore the potential of I-SceI-based marker removal for excising multiple markers in a single step, mutants carrying multiple deletions were constructed (Table 4.1.). IMY099 and IMY100, carrying two deletions, were constructed from IMY089 and IMY088 respectively by disrupting *ADE2* using *natNT1* as selection marker (dominant marker conferring resistance to nourseothricin [49]). Subsequently, strains IMY090-0 and IMY100-1 were constructed from IMY099 and IMY100, respectively, by disrupting *CAN1* using *hphNT1* as selection marker (dominant marker conferring resistance to hygromycin [50]). All deletions in strains IMY089, IMY099 and IMY090-0 were performed using AB cassettes, while strains IMY088, IMY100 and IMY100-1 were constructed with seamless deletion cassettes.

Marker removal in the strains harboring two and three deletions was triggered by growing the mutants on liquid and solid media containing galactose as sole carbon source as described above. Plating on selective media of strains IMY099 (two markers, AB repeats) after induction of *SCEI* revealed that the excision of two markers occurred in 56% of the colonies (28 of 50, Fig. 4.4.a). Marker excision was less efficient in seamless IMY100 (two markers, Sm repeats) as only 22% of the colonies (11 of 50, Fig. 4.4.b) lost their resistance to both G418 and nourseothricin. While efficiency of marker removal was substantially lower in IMY100 as compared to IMY099, the frequency of *KanMX* and *natNT1* excision was similar within each strain (*KanMX* 76% and *natNT1* 72% in IMY099, and *KanMX* 50% and *natNT1* 44% in IMY100), indicating the absence of marker- and locus-specificity of I-SceI-based marker removal. Remarkably, excision of three markers in a single step using AB repeats (strain IMY090-0) remained high, as 56% of the colonies (28 out of 50, Fig. 4.4.a) lost their ability to grow on media containing all three antibiotics. Conversely, none of the colonies from the induced IMY100-1 (three markers, Sm repeats) showed a phenotype characteristic of a marker-free strain (Fig. 4.4.b). To confirm marker removal, ten colonies of the mutants suspected to have lost two or three markers (Fig. 4.4., circled colonies) based on phenotypic analysis were checked by PCR for the presence of marker (Fig. 4.5.). For all 20 tested mutants carrying two deletions, simultaneous excision of both markers was confirmed, irrespective of the sequence used for homologous recombination repair (i.e. AB or Sm repeats). Simultaneous excision of three markers also proved to be very efficient, as only one colony out of the ten analyzed after induction of I-SceI in IMY090-0 (three markers, AB repeats) did not show the correct PCR band pattern. This colony, which phenotypically indicated the absence of three markers, but could not be confirmed by PCR was termed E3 and subjected to further investigation.

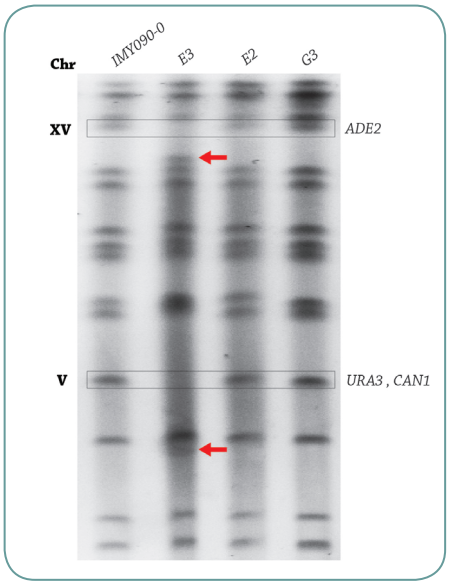


FIGURE 4.6 | Karyotyping of IMY090-0, carrying three deletions and AB repeats, and three clones after induction of *SCEI*. Deletions in IMY090-0 were targeted to *ADE2*, located on chromosome XV, and *URA3* and *CAN1* both located on chromosome V (framed areas). Clones E2 and G3 both displayed the same karyotype as their ancestor IMY090-0. Clone E3 had undergone chromosomal translocations as indicated by the absence of the expected bands corresponding to chromosomes V and XV (framed areas) and the appearance of neochromosomes (red arrows).

Occurrence of I-SceI induced chromosomal rearrangements during marker excision | While colony E3 lost its triple antibiotic resistance, PCR failed to amplify the targeted loci and thereby confirm the removal of all three selection markers. Before galactose induction, the parental strain of E3 (IMY090-0) contained six AB repeats distributed by groups of two over the genome, all adjacent to I-SceI recognition sites. Upon galactose cultivation, DSBs generated next to these AB repeats will free them for recognition and recombination by the DSB repair machinery. It is well documented that HR preferentially occurs between the closest repeats [33], and it was therefore anticipated that the AB repeats from the same deletion loci would be used for DSB repair. However, it cannot be excluded that recombination occurs between AB repeats from different deletion loci. Although presumably rare [33,57], this event may occur and trigger chromosomal translocations. In the parental strain CEN.PK113-7D the three genes targeted for deletion were *ADE2* located on chromosome XV, and *URA3* and *CAN1* both located on the left arm of chromosome V. The region of *ca.* 83 Kb located between *URA3* and *CAN1* contains several essential genes (*e.g. POL5* and *MCM3*), making recombination between these two loci lethal and therefore unlikely. Conversely recombination between chromosome XV and V could result in viable chromosomal rearrangements. To determine whether *SCEI* expression leads to chromosomal rearrangements in strain E3, the karyotyping of E3, its parental strain IMY090-0 and two other randomly picked transformants with confirmed triple marker excision, termed E2 and G3, was carried out (Fig. 4.6.). While all other strains displayed the expected bands corresponding to chromosomes V and XV, these bands were absent from E3, thereby confirming rearrangement of both chromosomes V and XV. While this analysis is insufficient to characterize the precise nature of the chromosomal rearrangement, two new bands appeared, indicating the presence of neochromosomes of *c.* 1000 and 400 Kbp (Fig. 4.6.). Conversely, there were no indications of chromosomal rearrangements in strains E2 and G3.

DISCUSSION

The present work demonstrates that simultaneous recovery of up to three markers can be easily and efficiently achieved without heterologous recombinase, by simply relying on the coordinated action of *S. cerevisiae* I-SceI meganuclease and HR machinery guided by direct repeats [27,58]. Upon excision by I-SceI, the recovery of three markers was achieved with *c.* 50% efficiency. In addition to this technical achievement, a series of plasmids harbouring seven dominant and auxotrophic selectable markers flanked with I-SceI sites was constructed and made available to the scientific community (Euroscarf http://web.uni-frankfurt.de/fb15/mikro/euroscarf/).

Multiple marker excision has been previously achieved in *S. cerevisiae* using the Cre/*Lox*P system [16]. In this earlier study, four markers were recycled in one step with efficiencies similar to the ones measured here. However the described approach presents a number of advantageous features compared to the *Lox*P-based approach. While *Lox*P-aided marker removal revealed disparities in recycling efficiency between markers, the I-SceI-based approach is locus- and marker-independent. Much like AB-repeats based excision, the *Lox*P-based approach offers multiple identical repeats (*Lox*P sites) scattered over the yeast genome that can be targeted by the Cre recombinase and thereby result in chromosomal rearrangements. The *Lox*P system triggered the frequent occurrence of chromosomal rearrangements as 50% of the marker-free transformants displayed altered PCR patterns and karyotypes [16]. In the present study, only one out of ten strains subjected to the simultaneous removal of three marker cassettes displayed a PCR pattern reflecting chromosomal rearrangements. Furthermore, use of Cre/*Lox*P system leaves cryptic *Lox*P sites that can be recognized and recombined by the Cre recombinase during additional deletion rounds. This precludes utilization of the Cre/*Lox*P system for studies in which sequential deletion rounds are mandatory. Conversely, the I-SceI-based approach generates stable strains as the AB scars left are inert and the probability that they recombine is as low as for natural repeats in the genome (10^{-7} to 10^{-8}) [59,60]. The I-SceI-based approach is therefore perfectly suited for strain engineering studies that require multiple gene deletions.

To minimize the occurrence of chromosomal rearrangements during both marker recovery and subsequent deletion rounds a seamless approach is preferable. With seamless marker recovery the repeats are different for each marker and will therefore not lead to interlocus recombination by HR. In addition, this approach is totally scar-free upon marker removal and will therefore not promote recombination during successive deletion rounds. Deletion cassettes for seamless marker removal can be easily PCR-amplified from the supplied plasmids. Using a seamless approach, two markers could be co-excised in *S. cerevisiae*. As observed for AB repeats, marker removal was marker- and locus-independent. However, the efficiency of marker recovery was lower and simultaneous excision of three markers was not achieved. The most probable explanation for this low efficiency is the repeats size. While AB repeats were 86 bp long, Sm repeats were only 40 bp and it is well documented that HR efficiency depends on the length of homologous sequences [34]. Despite the lower efficiency of seamless repair, the results showed that this method is capable of generating markerfree strains.

Furthermore, PCR analysis of the marker-free transformants gave no indication of the occurrence of chromosomal rearrangement with the seamless approach. Seamless marker removal therefore offers a great potential for large scale deletion programs.
While the presented approach already offers a superior alternative to current methods, several features can be further improved to make marker recovery faster and even more efficient. As suggested above, the first step for improvement is the elongation of the Sm repeats to enable seamless marker excision. Deletion cassettes with longer Sm repeats can be easily PCR-amplified from the supplied plasmids. In the present study *SCEI* was expressed from a centromeric vector behind a galactose inducible promoter. Expression of *SCEI* was found to substantially affect the efficiency of marker removal. While this aspect was not studied in depth in the present study, the design of the *SCEI* expression cassette and of the conditions for expression can be further fine tuned. Finally, to reduce the number of steps and time required for marker recovery, an elegant approach would be to perform the last deletion from the series with a cassette carrying *SCEI* in addition to a selectable marker.

Other methods for advanced genome editing based on the artificial creation of DSB are currently being developed, such as TALEN's [61] and the CRISPR/Cas system [62,63]. Recently, multiple targeted integrations were achieved [36], the combination of multiple deletions/integrations with the power of DSB-guided marker excision will enormously accelerate the process of strain engineering. Harnessing the amazing efficiency of yeast HR, combined with creative systems for controlled induction of DSB will undoubtedly contribute to a rapid further expansion of the *S. cerevisiae* synthetic biology toolbox.

ACKNOWLEDGEMENTS

This work was supported by the Technology Foundation STW (Vidi Grant 10776). J.T.P. and J.-M.D. were also supported by the 'Platform Green Synthetic Biology' program (http://www.pgsb.nl/). We are grateful to Lizanne Bosman for her experimental contribution to the development of the system.

REFERENCES

1. Hong K, Nielsen J: **Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries.** *Cell Mol Life Sci* 2012, **69:** 2671-2690.

2. Ignea C, Cvetkovic I, Loupassaki S, Kefalas P, Johnson CB, Kampranis SC *et al.*: **Improving yeast strains using recyclable integration cassettes, for the production of plant terpenoids.** *Microb Cell Fact* 2011, **10**:4.

3. Koopman F, Beekwilder J, Crimi B, van Houwelingen A, Hall RD, Bosch D *et al.*: ***De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*.** *Microb Cell Fact* 2012, **11**:155.

4. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM *et al.*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440:** 940-943.

5. Szczebara FM, Chandelier C, Villeret C, Masurel A, Bourot S, Duport C *et al.*: **Total biosynthesis of hydrocortisone from a simple carbon source in yeast.** *Nat Biotechnol* 2003, **21:** 143-149.

6. Kozak BU, van Rossum HM, Benjamin KR, Wu L, Daran JMG, Pronk JT *et al.*: **Replacement of the *Saccharomyces cerevisiae* acetyl-CoA synthetases by alternative pathways for cytosolic acetyl-CoA synthesis.** *Metab Eng* 2014, **21:** 46-59.

7. Suzuki Y, St Onge RP, Mani R, King OD, Heilbut A, Labunskyy VM *et al.*: **Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection.** *Nat Methods* 2011, **8:** 159-164.

8. Solis-Escalante D, Kuijpers NG, Bongaerts N, Bolat I, Bosman L, Pronk JT *et al.*: ***amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:** 126-139.

9. Baganz F, Hayes A, Farquhar R, Butler PR, Gardner DCJ, Oliver SG: **Quantitative analysis of yeast gene function using competition experiments in continuous culture.** *Yeast* 1998, **14:** 1417-1427.

10. Gopal CV, Broad D, Lloyd D: **Bioenergetic consequences of protein overexpression in *Saccharomyces cerevisiae*.** *Appl Environ Microb* 1989, **30:** 160-165.

11. Nacken V, Achstetter T, Degryse E: **Probing the limits of expression levels by varying promoter strength and plasmid copy number in *Saccharomyces cerevisiae*.** *Gene* 1996, **175:** 253-260.

12. Sauer B: **Functional expression of the *Cre-Lox*P site-specific recombination system in the yeast *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1987, **7:** 2087-2096.

13. Tohe A: **Construction of a marker gene cassette which is repeatedly usable for gene disruption in yeast.** *Curr Genet* 1995, **27:** 293-297.

14. Guldener U, Heck S, Fiedler T, Beinhauer J, Hegemann JH: **A new efficient gene disruption cassette for repeated use in budding yeast.** *Nucleic Acids Res* 1996, **24:** 2519-2524.

15. Hoess RH, Abremski K: **Mechanism of strand cleavage and exchange in the *Cre*-Lox site-specific recombination system.** *J Mol Biol* 1985, **181:** 351-362.

16. Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ *et al.*: **Exploring redundancy in the yeast genome: an improved strategy for use of the *Cre-Lox*P system.** *Gene* 2000, **252:** 127-135.

17. Carter Z, Delneri D: **New generation of loxP-mutated deletion cassettes for the genetic manipulation of yeast natural isolates.** *Yeast* 2010, **27:** 765-775.

18. Jasin M, Rothstein R: **Repair of strand breaks by homologous recombination.** Cold Spring Harb Perspect Biol 2013, **5**:11.

19. Aylon Y, Kupiec M: **DSB repair: the yeast paradigm.** *Dna Repair* 2004, **3:** 797-815.

20. Orrweaver TL, Szostak JW, Rothstein RJ: **Yeast transformation: A model system for the study of recombination.** *Proc Natl Acad Sci USA* 1981, **78:** 6354-6358.

21. Scherer S, Davis RW: **Replacement of chromosome segments with altered DNA sequences sonstructed *in vitro*.** *Proc Natl Acad Sci USA* 1979, **76:** 4951-4955.

22. Wach A, Brachat A, Pohlmann R, Philippsen P: **New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*.** *Yeast* 1994, **10:** 1793-1808.

23. Alani E, Cao L, Kleckner N: **A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains.** *Genetics* 1987, **116:** 541-545.

24. Chattoo BB, Sherman F, Azubalis DA, Fjellstedt TA, Mehnert D, Ogur M: **Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilization of alpha-aminoadipate.** *Genetics* 1979, **93:** 51-65.

25. Singh A, Sherman F: **Association of methionine requirement with methyl mercury resistant mutants of yeast.** *Nature* 1974, **247:** 227-229.

26. Toyn JH, Gunyuzlu PL, White WH, Thompson LA, Hollis GF: **A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance.** *Yeast* 2000, **16:** 553-560.

27. Fairhead C, Llorente B, Denis F, Soler M, Dujon B: **New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination.** *Yeast* 1996, **12:** 1439-1458.

28. Paques F, Haber JE: **Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*.** *Microbiol Mol Biol R* 1999, **63:** 349-404.

29. Nickoloff JA, Singer JD, Heffron F: ***In vivo* analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis.** *Mol Cell Biol* 1990, **10:** 1174-1179.

30. Plessis A, Perrin A, Haber JE, Dujon B: **Site-specific recombination determined by I-SceI, a mitochondrial group-I intron-encoded endonuclease expressed in the yeast nucleus.** *Genetics* 1992, **130:** 451-460.

31. Monteilhet C, Perrin A, Thierry A, Colleaux L, Dujon B: **Purification and characterization of the *in vitro* activity of I-SceI, a novel and highly specific endonuclease encoded by a group-I intron.** *Nucleic Acids Res* 1990, **18:** 1407-1413.

32. Noskov VN, Segall-Shapiro TH, Chuang RY: **Tandem repeat coupled with endonuclease cleavage (TREC): a seamless modification tool for genome engineering in yeast.** *Nucleic Acids Res* 2010, **38:** 2570-2576.

33. Rudin N, Haber JE: **Efficient repair of HO-induced chromosomal breaks in *Saccharomyces cerevisiae* by recombination between flanking homologous sequences.** *Mol Cell Biol* 1988, **8:** 3918-3928.

34. Sugawara N, Haber JE: **Characterization of double strand break induced recombination - homology requirements and single stranded DNA Formation.** *Mol Cell Biol* 1992, **12:** 563-575.

35. Guldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of *Lox*P marker cassettes for *Cre*-mediated multiple gene knockouts in budding yeast.** *Nucleic Acids Res* 2002, **30:** e23.

36. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH *et al.*: **EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2014, **14:** 238-248.

37. Bertani G: **Lysogeny at mid-twentieth century: P1, P2, and other experimental, systems.** *J Bacteriol* 2004, **186:** 595-600.

38. Bertani G: **Studies on Lysogenesis. 1. The mode of phage liberation by lysogenic *Escherichia coli*.** *J Bacteriol* 1951, **62:** 293-300.

39. Entian KD, Kotter P: **Yeast genetic strain and plasmid collections.** *In* Stansfield I, Stark MJR (ed) *Methods in microbiology*. Amsterdam: Academic Press; 2007:629-666.

40. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik MA *et al.*: ***De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology.** *Microb Cell Fact* 2012, **11:**36.

41. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C *et al.*: **An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains.** *Enzyme Microb Tech* 2000, **26:** 706-714.

42. Kuijpers NGA, Chroumpi S, Vos T, Solis-Escalante D, Bosman L, Pronk JT *et al.*: **One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:** 769-781.

43. Verduyn C, Postma E, Scheffers WA, Vandijken JP: **Effect of benzoic acid on metabolic fluxes in yeasts - A continuous culture study on the regulation of respiration and alcoholic fermentation.** *Yeast* 1992, **8:** 501-517.

44. Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Method Enzymol* 2002, **350:** 87-96.

45. de Kok S, Nijkamp JF, Oud B, Roque FC, de Ridder D, Daran JM *et al.*: **Laboratory evolution of new lactate transporter genes in a *jen1D* mutant of *Saccharomyces cerevisiae* and their identification as *ADY2* alleles by whole-genome resequencing and transcriptome analysis.** *FEMS Yeast Res* 2012, **12:** 359-374.

46. de Kok S, Yilmaz D, Suir E, Pronk JT, Daran JM, van Maris AJA: **Increasing free-energy (ATP) conservation in maltose-grown *Saccharomyces cerevisiae* by expression of a heterologous maltose phosphorylase.** *Metab Eng* 2011, **13:** 518-526.

47. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW: **Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.** *Nat Genet* 1996, **14:** 450-456.

48. Gatignol A, Baron M, Tiraby G: **Phleomycin resistance encoded by the *ble* gene from transposon Tn5 as a dominant selectable marker in *Saccharomyces cerevisiae*.** *Mol Gen Genet* 1987, **207:** 342-348.

49. Goldstein AL, McCusker JH: **Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*.** *Yeast* 1999, **15:** 1541-1553.

50. Gritz L, Davies J: **Plasmid encoded Hygromycin-B resistance the sequence of Hygromycin-B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*.** *Gene* 1983, **25:** 179-188.

51. Langlerouault F, Jacobs E: **A method for performing precise alterations in the yeast genome using a recyclable selectable marker.** *Nucleic Acids Res* 1995, **23:** 3079-3081.

52. Shuster JR, Moyer D, Irvine B: **Sequence of the *Kluyveromyces lactis URA3* gene.** *Nucleic Acids Res* 1987, **15:** 8573.

53. Brachmann CB, Davies A, Cost GJ, Caputo E, Li JC, Hieter P *et al.*: **Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.** *Yeast* 1998, **14:** 115-132.

54. Zhang YP, Chen XJ, Li YY, Fukuhara H: ***LEU2* gene homolog in *Kluyveromyces lactis*.** *Yeast* 1992, **8:** 801-804.

55. Koren P, Svetec IK, Mitrikeski PT, Zgaga Z: **Influence of homology size and polymorphism on plasmid integration in the yeast *CYC1* DNA region.** *Curr Genet* 2000, **37:** 292-297.

56. Akada R, Kitagawa T, Kaneko S, Toyonaga D, Ito S, Kakihara Y *et al.*: **PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*.** *Yeast* 2006, **23:** 399-405.

57. Agmon N, Liefshitz B, Zimmer C, Fabre E, Kupiec M: **Effect of nuclear architecture on the efficiency of double-strand break repair.** *Nat Cell Biol* 2013, **15:** 694-699.

58. Benjamin KR: **Nucleic acids, compositions and methods for the excision of target nucleic acids.** *Pat. no. US US7919605 B1,* 2010.

59. Lichten M, Haber JE: **Position effects in ectopic and allelic mitotic recombination in *Saccharomyces cerevisiae*.** *Genetics* 1989, **123:** 261-268.

60. Jinks-Robertson S, Michelitch M, Ramcharan S: **Substrate length requirements for efficient mitotic recombination in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1993, **13:** 3937-3950.

61. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A *et al.*: **Targeting DNA double-strand breaks with TAL effector nucleases.** *Genetics* 2010, **186:** 757-761.

62. Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N *et al.*: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339:** 819-823.

63. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM: **Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.** *Nucleic Acids Res* 2013, **41:** 4336-4343.

SUPPLEMENTARY DATA

TABLE S1 | Intermediate strains constructed in the study.

Strain Name	Features	Relevant genotype	Reference
IMX487	One marker, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$*	This study
IMX506	One marker, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB*	This study
IMX488	Two markers, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB ade2::AB-ISceI$_{rs}$-natMX-ISceI$_{rs}$-AB*	This study
IMX489	Two markers, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$ ade2::ISceI$_{rs}$-natMX-ISceI$_{rs}$*	This study
IMX504	Three markers, AB repeats	*MAT*a *MAL2-8c SUC2 ura3::AB-ISceI$_{rs}$-KanMX-ISceI$_{rs}$-AB ade2::AB-ISceI$_{rs}$-natMX-ISceI$_{rs}$-AB can1::AB-ISceI$_{rs}$-nph-ISceI$_{rs}$-AB*	This study
IMX505	Three markers, Sm repeats	*MAT*a *MAL2-8c SUC2 ura3::ISceI$_{rs}$-KanMX-ISceI$_{rs}$ ade2::ISceI$_{rs}$-natMX-ISceI$_{rs}$ can1::ISceI$_{rs}$-nph-ISceI$_{rs}$*	This study

TABLE S2 | Primers used in this study.

Primer name	Primer sequence 5' to 3'	Application
pDS-FW	GGG*AGATCT***ATGACAAGAGGGTCGAACTCGC CTAAGTCGTAATTGAGTC**AGTTACGCTAGGGAT AACAGGGTAATATAGCTGTTTAGCTTGCCTCGTCC	pDS plasmid construction
pDS4-FW	GGG*GGATCC***ATGACAAGAGGGTCGAACTCGCCTAAGT CGTAATTGAGTC**AGTTACGCTAGGGATAACAGGGTAAT ATAGAGAGCTCGCTGTGAAGATCC	pDS plasmid construction
pDS-RV	GGG*TCTAGA***CCGCCAAGCGAATTGAAGGACCGTGCGT AGAATGAAGAAC**CTATATTACCCTGTTATCCCTAGCGTAA CTTTAAGGGTTCTCGAGAGCTC	pDS plasmid construction
pDS4-RV	GGG*GAATTC***CCGCCAAGCGAATTGAAGGACCGTGCGTA GAATGAAGAAC**CTATATTACCCTGTTATCCCTAGCGTAAC TAGAGATCCGCAGGCTAACCG	pDS plasmid construction
pUG-FW	GGG*TCTAGA***ATGACAAGAGGGTCGAACTCGCCTAAGTCG TAATTGAGTC**CAGATCCACTAGTGGCCTATG	pDS plasmid construction
pUG4-FW	GGG*GAATTC***ATGACAAGAGGGTCGAACTCGCCTAA GTCGTAATTGAGTC**CAGATCCACTAGTGGCCTATG	pDS plasmid construction
pUG-RV	GGG*AGATCT***CCGCCAAGCGAATTGAAGGACCGTG CGTAGAATGAAGAAC**ATTAAGGGTTGTCGACCTGC	pDS plasmid construction
pUG4-RV	GGG*GGATCC***CCGCCAAGCGAATTGAAGGACCGTGCG TAGAATGAAGAAC**ATTAAGGGTTGTCGACCTGC	pDS plasmid construction
*ura3*_Sm_fw	**ATATATACGCATATGTAGTGTTGAAGAAACATGAAATTG CCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGC AGGAAACGAAGATAAATCATG**TTGAGTCAGTTACGCTAGGG	*URA3* Sm deletion cassette
*ura3*_Sm_rv	**AATTGAAGCTCTAATTTGTGAGTTTAGTATACATG CATTTACTTATAATACAGTTTTTTA***ACCAATCTAAGTCT GTGCTCCTTCCTTCGTTCTTCCTTCT*CGTGCGTAG AATGAAGAACC	*URA3* Sm deletion cassette
*ade2*_Sm_fw	**TGTATAAATTGGTGCGTAAAATCGTTGGATCTCTCTT CTAAGTACATCCTACTATAACAATCAAGAAAAACAAGA AAATCGGACAAAACAATCAAGTATG**TTGAGTCAG TTACGCTAGGG	*ADE2* Sm deletion cassette
*ade2*_Sm_rv	**AATAGGTATATCATTTTATAATTATTTGCTGTACAAGT ATATCAATAAACTTATATATTA***TGAAATGCTCCATAATATT GTCCATTTAGTTCTTAATAAA*CGTGCGTAGAATGAAGAACC	*ADE2* Sm deletion cassette
*can1*_Sm_fw	**GGATCCAGTTTTCAATCTGTCGTCAATCGAAAGTTTA TTTCAGAGTTCTTCAGACTTCTTAACTCCTGTAAAAA CAAAAAAAAAAAAAGGCATAGCAATG**TTGAGTCAGTTACGCTAGGG	*CAN1* Sm deletion cassette

Primer name	Primer sequence 5' to 3'	Application
*can1*_Sm_rv	**GAGGGTGAGAATGCGAAATGGCGTGGGAATGTGATTAA AGGTAATAAAACGTCATATCTA***AACTATAAGTATAATAGTAA CTTATATATTTCTGTTCCAG*CGTGCGTAGAATGAAGAACC	*CAN1* Sm deletion cassette
*ura3*_AB_fw	**ATATATACGCATATGTAGTGTTGAAGAAACATGAAATTG CCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTG CAGGAAACGAAGATAAATCATG**CAGCTGAAGCTT CGTACGC	*URA3* AB deletion cassette
*ura3*_AB_rv	**TTACGACCGAGATTCCCGGGTAATAACTGATATAATTA AATTGAAGCTCTAATTTGTGAGTTTAGTATACATGCATTT ACTTATAATACAGTTTTTTA**GCATAGGCCACTAG TGGATCTG	*URA3* AB deletion cassette
*ade2*_AB_fw	**TGTATAAATTGGTGCGTAAAATCGTTGGATCTCTCTT CTAAGTACATCCTACTATAACAATCAAGAAAAACAAGAAAA TCGGACAAAACAATCAAGTATG**CAGCTGAAGCTTCGTACGC	*ADE2* AB deletion cassette
*ade2*_AB_rv	**GAAGTCCACATTTGATGTAATCATAACAAAGCCTAAAAAAT AGGTATATCATTTTATAATTATTTGCTGTACAAGTATATCA ATAAACTTATATATTA**GCATAGGCCACTAGTGGATCTG	*ADE2* AB deletion cassette
*can1*_AB_fw	**TCCAATAGGTGGTTAGCAATCGTCTTACTTTCTAACTTTT CTTACCTTTTACATTTCAGCAATATATATATATATATTTCA AGGATATACCATTCTAATG**CAGCTGAAGCTTCGTACGC	*CAN1* AB deletion cassette
*can1*_AB_rv	**CTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAG TTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAA CCTTGTTCAGGTCTAACACTA**GCATAGGCCACTAGT GGATCTG	*CAN1* AB deletion cassette
UdcFW	GCTACTGCGCCAATTGATGAC	Confirmation of deletion and marker excision in *URA3* locus
KanA	CGCACGTCAAGACTGTCAAG	Confirmation of *URA3*, *ADE2* and *CAN1* deletions
KanB	TCGTATGTGAATGCTGGTCG	Confirmation of *URA3* deletion
UdcRV	CGAGATTCCCGGGTAATAACTG	Confirmation of deletion and marker excision in *URA3* locus
AdcFW	AAAGGACACCTGTAAGCGTTG	Confirmation of deletion and marker excision in *ADE2* locus
NatRV	CTGTAGGTCAGGTTGCTTTC	Confirmation of *ADE2* deletion

Primer name	Primer sequence 5' to 3'	Application
AdcRV	AACGCCGTATCGTGATTAAC	Confirmation of deletion and marker excision in *ADE2* locus
CdcFW	TTCTAGGTTCGGGTGACGTGAAG	Confirmation of deletion and marker excision in *CAN1* locus
hygroFW	GGACGCTCGAAGGCTTTAAC	Confirmation of *CAN1* deletion
CdcRV	GGTTGCGAACAGAGTAAACC	Confirmation of deletion and marker excision in *CAN1* locus
FK157-Fb1	GCGGATAAAGTTGCAGGAC	Confirmation of yeast transformation with pUDC073
I-SceI inside rv	GAACCAGTATGCCAGAGACATC	Confirmation of yeast transformation with pUDC073
UdcFW	GCTACTGCGCCAATTGATGAC	Sequencing of *URA3* locus
UdcRV	CGAGATTCCCGGGTAATAACTG	Sequencing of *URA3* locus
AdcRV	AACGCCGTATCGTGATTAAC	Sequencing of *ADE2* locus
AdcFW	AAAGGACACCTGTAAGCGTTG	Sequencing of *ADE2* locus
CdcRV	GGTTGCGAACAGAGTAAACC	Sequencing of *CAN1* locus
CdcFW	TTCTAGGTTCGGGTGACGTGAAG	Sequencing of *CAN1* locus
3988	CAATTCAACGCGTCTGTGAG	Sequencing of pDS1 and pDS6
3890	GTTCTTCATTCTACGCACGGTCC	Sequencing of pDS2

Primer name	Primer sequence 5' to 3'	Application
3811	CTCGGTGAGTTTTCTCCTTCAT	Sequencing of pDS1
3751	GGTCAGCAGTACAGAACCGTCG	Sequencing of pDS3
3725	CCTCAGTGGCAAATCCTAAC	Sequencing of pDS2 and pDS7
3219	GTATCACGAGGCCCTTTC	Sequencing of pDS1, 3, 4, 5, 6 and 7
2528	TCTTTCCTGCGTTATCCC	Sequencing of pDS1-7
2498	ATACGACCCAGAAGCTTACC	Sequencing of pDS2
2363	TTACCACCATCCAATGCAGAC	Sequencing of pDS3
1937	CGAATTGCTTGCAGGCATCTC	Sequencing of pDS5
1936	AATCTCGTGATGGCAGGTTGG	Sequencing of pDS5
1166	GCTGGAGGTCACCAACGTCAAC	Sequencing of pDS6
1077	AGCTTCCCTACCTGACACTAAC	Sequencing of pDS4
1076	CACGTGACTGCGCTGAATTG	Sequencing of pDS4
526	CTCGCCGATAGTGGAAACCG	Sequencing of pDS7

*Italic: XbaI or BglII or BamHI or EcoRI recognition sites. Bold: A or B fragments for AB synthetic repeats. Underlined: I-SceI recognition sequence. Bold and underlined: sequence for targeted homologous recombination. Italic and underlined: seamless repeats for marker excision.

TABLE S3 | Plasmids used in this study.

Plasmid name	Characteristics	Reference
pUG6	*KanMX*	[35]
pUGamdSYM	*amdSYM*	[8]
pUG72	*KlURA3*	[35]
pUG73	*KlLEU2*	[35]
pUG66	*$AgTEF2_{pr}$-ble-$AgTEF2_{ter}$*	[35]
pUG-natNT1	*$AgTEF2_{pr}$-natNT1-$AgTEF2_{ter}$*	[45]
pUG-hphNT1	*$AgTEF2_{pr}$-hphNT1-$AgTEF2_{ter}$*	[46]
pDS1	*AB-$ISceI_{rs}$-KanMX-$ISceI_{rs}$-AB*	This study
pDS2	*AB-$ISceI_{rs}$-amdSYM-$ISceI_{rs}$-AB*	This study
pDS3	*AB-$ISceI_{rs}$-KlURA-$ISceI_{rs}$-AB*	This study
pDS4	*AB-$ISceI_{rs}$-KlLEU2-$ISceI_{rs}$-AB*	This study
pDS5	*AB-$ISceI_{rs}$-$AgTEF2_{pr}$-ble-$AgTEF2_{ter}$-$ISceI_{rs}$-AB*	This study
pDS6	*AB-$ISceI_{rs}$-$AgTEF2_{pr}$-natNT1-$AgTEF2_{ter}$-$ISceI_{rs}$-AB*	This study
pDS7	*AB-$ISceI_{rs}$-$AgTEF2_{pr}$-hphNT1-$AgTEF2_{ter}$-$ISceI_{rs}$-AB*	This study
pUDC073	*GAL1pr-SCEI, URA3*	[42]

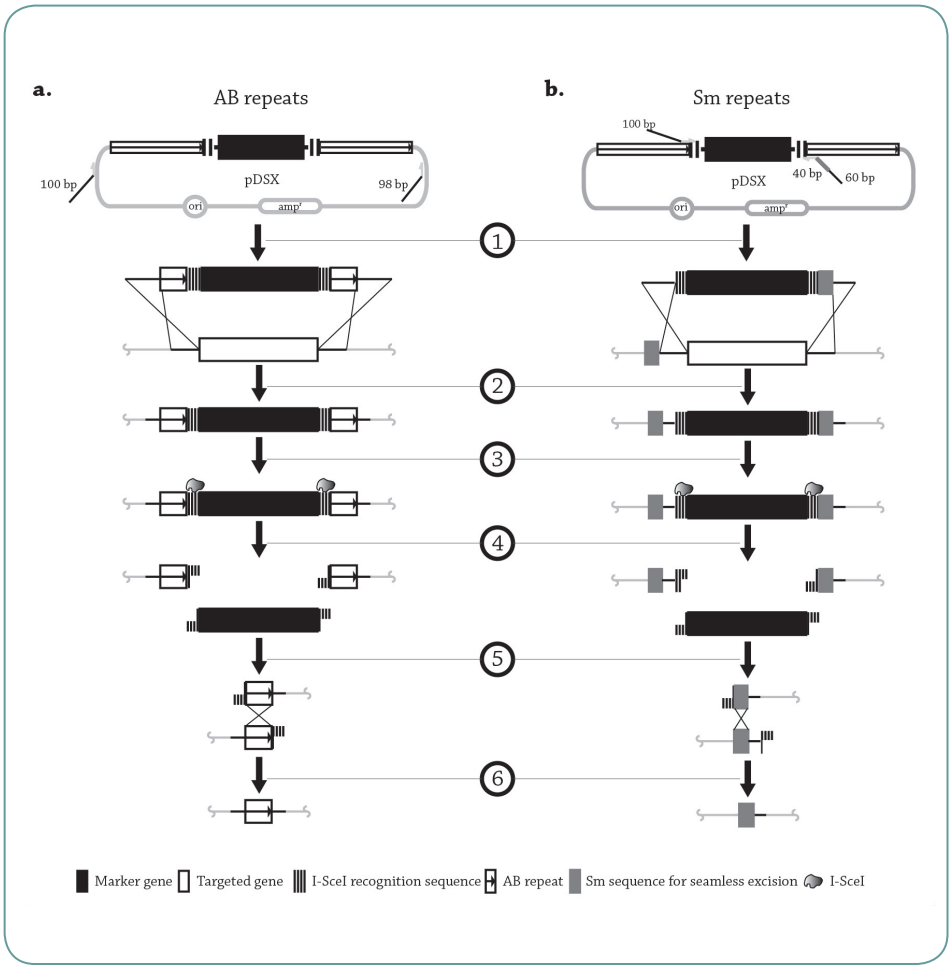


FIGURE S1 | Schematic representation of the I-SceI-facilitated marker excision with AB and Sm homologous recombination repeats. (a) use of synthetic AB repeats for homologous recombination repair and (b), use of seamless repair. 1) Using a pDS plasmid as template for PCR, a deletion cassette containing the key elements for marker excision by DSB repair can be easily obtained. 2) Targeted integration of deletion cassette into yeast genome via homologous recombination allows the selection of mutants. 3) Expression of I-SceI generate DSB flanking the selectable marker. 5) The DSB repair machinery localize the homologous adjacent regions and repairs the DSB and the marker is replaced by a synthetic repeat or seamlessly removed from the genome.

Chapter 5

Pathway swapping: towards modular engineering of essential cellular processes

Kuijpers NGA, Solis-Escalante D, Luttik MAH, Pronk JT, Daran JM, Daran-Lapujade P

ABSTRACT:

Recent developments in synthetic biology enable one-step implementation of entire metabolic pathways in industrial microorganisms. A similarly radical remodeling of central metabolism, required for fast optimization of such 'cell factories', is impeded by the mosaic organization of microbial genomes. To eliminate this limitation, we explore the concept of 'pathway swapping', with glycolysis, a near-ubiquitous pathway for sugar metabolism, as experimental model. A 'single-locus glycolysis' *Saccharomyces cerevisiae* platform was constructed to enable quick and easy replacement of this yeast's entire 26-isoenzyme complement of glycolytic enzymes by any alternative, functional glycolytic pathway configuration. The viability of this approach was demonstrated by characterization of *S. cerevisiae* strains whose growth depended on two non-native glycolytic pathways: a complete glycolysis from the related yeast *S. kudriavzevii* and a mosaic glycolysis consisting of yeast and human enzymes. This work paves the way for a new, modular approach to engineering and analysis of core cellular processes.

INTRODUCTION

Replacement of petrochemistry by bio-based processes, a key element in scenarios for sustainable development, requires microbes equipped with novel-to-nature capabilities. Recent developments in synthetic biology enable introduction of entire metabolic pathways and, thereby, new functionalities for product formation and substrate consumption, into microbial cells [1]. However, industrial viability of the resulting strains critically depends on optimal interaction of the newly introduced pathways with the microbial host's central metabolism. Central metabolic pathways such as glycolysis, TCA-cycle and pentose-phosphate pathway are essential for synthesis of precursors, for providing free energy (ATP), and for redox cofactor balancing. Optimization of productivity, product yield and robustness will therefore also require modifications in the configuration and/or regulation of core metabolic functions.

Engineering of central metabolism is, in many cases, more challenging than the mere introduction of heterologous pathways. Millions of years of evolution of microbial genomes and metabolic networks have resulted in a level of complexity that cannot be efficiently reengineered by iterative, single-gene modifications. Simple inactivation and subsequent replacement of structural genes is complicated by the essential role of many of the enzymes involved in central metabolism. Moreover, the enzymes of central metabolism are encoded by hundreds of genes that are scattered across microbial genomes. Microbial platforms in which entire pathways in central metabolism can be remodeled at will and in a combinatorial manner would provide an invaluable asset to fundamental research and engineering of central metabolism.

While rapid, cost-effective assembly of entire synthetic genomes is becoming a realistic perspective for small bacterial genomes [2,3], routine synthesis and expression of entire eukaryotic genomes is unlikely to be implemented in the next few years. Here, we propose a new, modular approach to the engineering of central metabolism that involves versatile, synthetic microbial strain platforms in which entire metabolic pathways can, in a few simple steps, be swapped for any functional, newly designed configuration. As proof of principle, we set out to construct a platform enabling swapping of the entire Embden-Meyerhoff-Parnas pathway of glycolysis, a largely conserved metabolic highway for sugar utilization, in the model eukaryote and industrial yeast *Saccharomyces cerevisiae*. Including the reactions leading to the formation of ethanol, the main fermentation product of *S. cerevisiae*, yeast glycolysis encompasses a set of 12 reactions, catalyzed by no fewer than 26 cytosolic isoenzymes [4]. Several of these (e.g. Tpi1, Tdh3, Adh1) are among the most abundant proteins in yeast cells. The genes encoding these isoenzymes are scattered all over the 16 yeast chromosomes. Construction of a platform for glycolysis swapping involved a two-step approach (Fig. 5.1.). First, as described in a recent study by our group, the genetic complexity of the glycolytic pathway was reduced by deleting the structural genes for 13 of the 26 glycolytic genes. Remarkably, a detailed systems analysis revealed that, under laboratory conditions, the phenotype of this 'minimal glycolysis' (MG) strain was virtually identical to that of the parental strain carrying the full set of

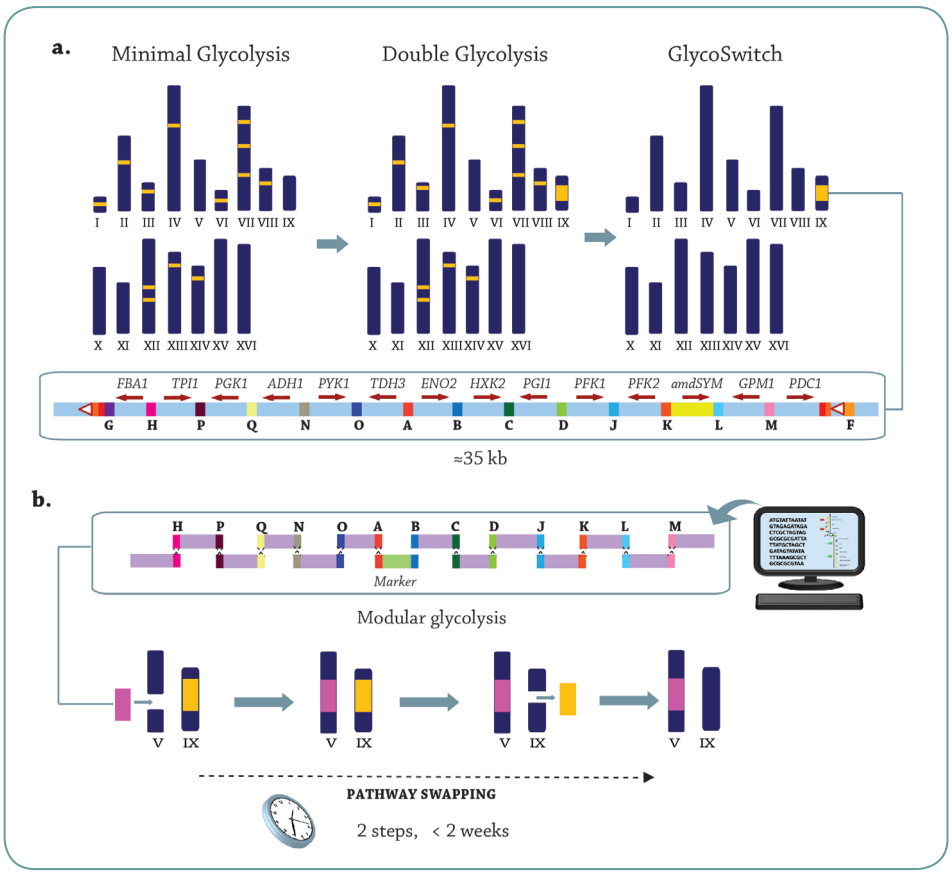


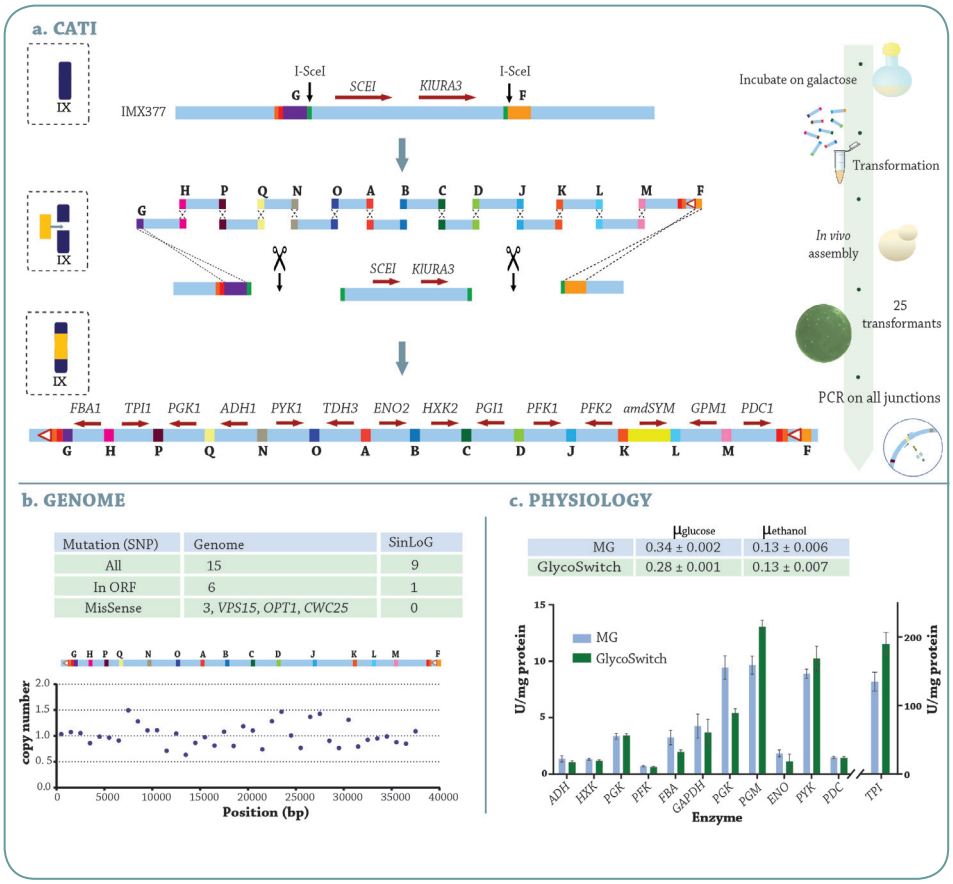
FIGURE 5.1 | Schematic overview of the glycolysis swapping approach. First a strain containing a double set of glycolytic genes was constructed from the minimal glycolysis strain by combined integration and *in vivo* assembly of the minimal set of endogenous glycolytic genes at a single locus on chromosome IX. Then all redundant endogenous glycolytic genes scattered over the genome were removed from their original locus (a). The resulting GlycoSwitch contains a single locus endogenous glycolysis and can be used for swapping the glycolytic pathway. An *in silico* designed glycolysis is assembled and integrated *in vivo* on chromosome V. After validation of correct assembly and integration, the endogenous glycolysis on chromosome IX can be removed in a single transformation step, leading to a strain with a redesigned glycolysis (b). The modular design enables a shear-endless number of glycolytic configurations.

glycolytic genes [5]. In a next step, the remaining 13 glycolytic genes in the MG strain were relocalized to a single chromosomal locus. Finally, the remaining scattered native genes were removed from their original locus, leading to 'GlycoSwitch', a strain carrying a native Single Locus Glycolysis (SinLoG). In the GlycoSwitch strain, glycolysis can be swapped in two steps by integration of a new, synthetic SinLoG, followed by removal of the SinLoG that was initially present (Fig. 5.1.).

RESULTS

Engineering of GlycoSwitch, the synthetic yeast platform for glycolysis swapping | The single-locus native glycolysis was assembled from 'glycoblocks' (Fig. 5.2.). These consisted of 13 DNA cassettes that each consisted of a *S. cerevisiae* glycolytic gene including its native promoter and terminator and flanked by 60-bp Synthetic Homologous Recombination (SHR) sequences [6]. SHR-sequences, which share no homology with the *S. cerevisiae* genome, were used for efficient *in vivo* assembly and integration of the glycoblocks and to enable flexible design and combinatorial assembly of synthetic glycolytic pathway variants. The single-locus native glycolysis was composed of 13 glycoblocks (corresponding to the 13 genes remaining in the MG strain) and of an ancillary block (A-block) containing a dominant counterselectable marker gene (amdSYM, [7]) flanked by SHR-sequences. The single-locus native glycolysis was integrated in the yeast genome rather than expressed from a plasmid to promote stable, single copy expression. The single-locus native glycolysis was introduced by Combined Assembly and Targeted Integration (CATI) [8], Fig. 5.2.). To this end, the MG strain was engineered to introduce recognition sequences for the I-SceI homing endonuclease at the *SGA1* locus on chromosome IX (Fig. S1). The 13 glycoblocks and the amdSYM cassette were pooled and co-transformed to the modified MG strain, in which *SCEI* was induced by growth on galactose to introduce a double-strand DNA break at the *SGA1* locus, thereby promoting integration of the glycoblocks (Fig. 5.2a, Fig. S1). Out of five tested colonies, four harbored the complete 35 kb single-locus native glycolysis integrated at the *SGA1* locus, demonstrating the efficiency of the CATI approach for chromosomal multi-gene DNA constructs. In a selected transformant (IMX382) the synthetic glycolytic genes were analyzed for single nucleotide polymorphisms (SNPs) as compared to the *in silico* design. The *in vivo*-assembled synthetic glycolysis was virtually identical to its *in silico* blueprint. As previously observed by Annaluru and co-workers during the assembly of SynIII, a synthetic version of chromosome III, in yeast [9], the majority of the nine deviations in nucleotide sequence were found at the HR loci linking the glycoblocks, which may either reveal recombinase-based errors or simply errors in the primers used to construct the HR sequences. Of the glycolytic genes only *ADH1* was found to contain a sense mutation (A180A).

S. cerevisiae IMX382 was further engineered by deleting the 13 remaining native glycolytic genes from their native loci, which were scattered over 10 yeast chromosomes (Fig. 5.1., for overview see Fig. S2). The first five deletions, targeting *PYK1*, *PGI1*, *TPI1*, *TDH3* and *PGK1*, were performed with standard deletion cassettes, using I-SceI-based marker removal to recycle multiple selection markers simultaneously without leaving



Mutation (SNP)	Genome	SinLoG
All	15	9
In ORF	6	1
MisSense	3, *VPS15*, *OPT1*, *CWC25*	0

	$\mu_{glucose}$	$\mu_{ethanol}$
MG	0.34 ± 0.002	0.13 ± 0.006
GlycoSwitch	0.28 ± 0.001	0.13 ± 0.007

FIGURE 5.2 | Construction and validation of GlycoSwitch. A minimal set of glycolytic genes resulting in a functional glycolytic pathway was integrated on chromosome IX using the Combined Assembly and Targeted Integration (CATI) approach [8]. First a *SCEI/KlURA3* cassette was introduced in the targeted locus on chromosome IX. In a subsequent experiment, *SCEI* was induced and the overlapping glycolytic expression cassettes were co-transformed for assembly and I-SceI facilitated integration of the endogenous glycolysis cassette (a). The remaining native glycolytic genes were removed from their native loci in a series of sequential deletions. The resulting auxotrophic GlycoSwitch strain (IMX589) was analysed by next-gen sequencing. In total 15 single nucleotide variations were detected compared to the Minimal Glycolysis strain, of which three resulted in a MisSense mutation. The synthetic construct did not contain a MisSense mutation. Gene copy number analysis (Magnolya algorithm, [31]) demonstrated that each glycolyic gene was present in a single copy and was therefore not duplicated in GlycoSwitch (b). To test the effect of clustering of the glycolytic genes on a single locus, the maximum specific growth rate in chemically defined medium with glucose as carbon source was determined. Furthermore, the *in vitro* enzymatic activity of the glycolytic enzymes from IMX606 (prototrophic GlycoSwitch) were measured and compared to the Minimal Glycolysis strain (MG). Growth rate and enzyme activities determinations result from at least two independent culture replicates (c).

scars in the genome [10] (Fig. S3). For subsequent engineering, an expression cassette encoding CRISPR endonuclease Cas9 was integrated at the *PFK2* locus, thereby deleting *PFK2* (Fig. S4). The remaining glycolytic genes, *PFK1*, *GPM1*, *HXK2*, *FBA1*, *ADH1* and *PDC1*, were deleted from their native loci with the CRISPR/Cas9 system [11-13]. Except for *ENO2*, all genes located on the single-locus native glycolysis were able to complement a null mutation in the corresponding native gene and therefore to carry the glycolytic flux. The persistent failure of attempts to delete *ENO2* from its native locus suggested that the synthetic *ENO2* variant was non-functional. Indeed, replacement of this *ENO2* allele with another glycoblock with a longer promoter region (1012 bp instead of 411 bp) enabled the deletion of the native gene from its original locus. This incident underlines the limited knowledge on promotor structure and function in yeast, even for glycolytic genes, and highlights the need for systematic design of synthetic promoters. This last genetic manipulation yielded the GlycoSwitch strain (IMX589). Whole genome sequencing of this strain confirmed: i) the correct sequence of the single-locus native glycolysis and its integration at the *SGA1* locus, ii) proper deletion of the native glycolytic genes from their original loci (Fig. S5) and iii) the absence of duplicated glycolytic genes in the single-locus native glycolysis and in the genome (Fig. S6). Relative to the ancestor 'minimal glycolysis' strain just six open reading frames (ORF's) in the complete genome were identified to contain a nucleotide difference. Three of those resulted in an amino acid substitution in the translated protein (Fig. 5.2b., Table S1). None of these mutations affected glycolytic genes or genes that could be otherwise associated with glycolysis.

Glycolytic genes are controlled by some of the strongest promoters in yeast and, consistent with the essentiality of the encoded proteins, are constitutively transcribed. Co-localization of a set of genes that are heavily loaded with RNA polymerase II might affect the conformation of DNA and thereby locally affect transcription. Moreover, Pol II disruption of chromatin has been proposed to be more sensitive to a variety of stresses [14], but it also affects the binding of a number of proteins such as cohesin that play an important role in genetic stability [15]. Nevertheless, the activity of glycolytic enzymes in cell extracts of strains expressing scattered (MG) and co-localized (GlycoSwitch) glycolytic genes were similar (Fig. 5.2c.). Furthermore, the specific growth rate of the prototrophic GlycoSwitch strain (IMX606) on chemically defined medium, was only slightly lower than that of the parental MG strain (17% lower, p-value=0.0001, Fig. 5.2c.).

Proof of principle: chromosome hopping of native yeast glycolysis | To test the feasibility of glycolysis swapping, we first attempted to exchange the single-locus native glycolysis integrated on chromosome IX by a nearly identical copy integrated on another locus on chromosome V. As, during this project, it became evident that Cas9 is superior to I-SceI in terms of efficiency and versatility of DNA editing, the CRISPR-Cas system was used instead of I-SceI to promote insertion of the glycoblocks of the new single-locus native glycolysis. The outer glycoblocks were modified to integrate at the *CAN1* locus on chromosome V, and the A-block carried KanMX as dominant selection marker instead of amdSYM. The GlycoSwitch strain was transformed with a complete set of glyco- and A-blocks, and with a CRISPR plasmid carrying the guide

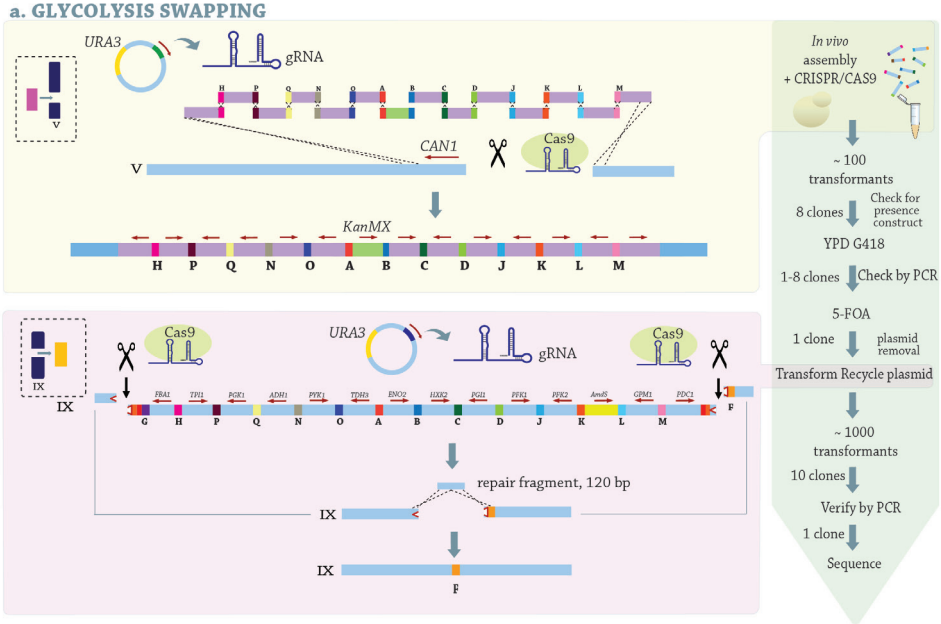
RNA required for targeting Cas9 to the *CAN1* locus, leading to the *in vivo* assembly and integration of a second glycolytic cassette (Fig. 5.3a.). Out of 12 G418-resistant transformants, colony-PCR showed that at least two carried the complete single-locus native glycolysis inserted in the *CAN1* locus. A clone was selected, cured of the *URA3*-carrying CRISPR plasmid, and transformed with a repair fragment and a new CRISPR plasmid in order to remove the remaining single-locus native glycolysis cassette located on chromosome IX (Fig. 5.3a.). This was achieved by targeting Cas9 specifically to sequences positioned at each end of the SinLoG cassette leading to excision of the cassette from the genome after which the dsDNA break could be repaired by homologous recombination with the 120 bp repair DNA fragment. All of the three selected transformants were shown to lack the single-locus native glycolysis on chromosome IX and to have retained the newly inserted single-locus native glycolysis on chromosome V. One clone ($SinLoG_V$ IMX605) was fully sequenced which further confirmed the successful relocalization of the entire glycolytic pathway. Analysis of copy number variation showed that no recombination occurred between the glycoblocks or the excised single-locus native glycolysis insert and the genome during glycolysis swapping (Fig. S6). While it has been demonstrated that the chromosomal localization can have a strong impact on gene expression, IMX605 grew as fast as the original GlycoSwitch strain in chemically defined medium (Fig. 5.3c.) and displayed the same activity of the glycolytic enzymes in cell extracts. In IMX605, the position of the A-block and the *ENO2* glycoblock were reversed as compared to the single-locus native glycolysis present in the GlycoSwitch strain. However, this different organization did not affect *ENO2* expression, as demonstrated by the enolase activity in cell extracts (Fig. 5.3c). This lack of locus-specific effect demonstrated that the *CAN1* locus on chromosome V was suitable for further testing the pathway swapping concept. This work presents the first demonstration that a ca. 35 kb construct can be integrated followed by the CRISPR-Cas9-mediated removal of an equal-sized large fragment that shows substantial sequence homology with the incoming fragment. Sequencing results confirmed that no unintended recombination events occurred, neither between the engineered genetic elements themselves nor between the engineered genetic elements and the native yeast genome. This technical milestone therefore demonstrates the applicability, ease of use and efficiency of pathway swapping.

***Saccharomyces cerevisiae* expressing foreign and mosaic glycolyses |** Demonstration of the technical feasibility of pathway swapping opened up the way to test whether it is possible to integrally replace yeast glycolysis, an essential, tightly controlled metabolic pathway, by heterologous or synthetic variants. For this purpose, we first selected a donor of glycolytic genes from within the *Saccharomyces* genus. *S. kudriavzevii* is a cold-tolerant close relative of *S. cerevisiae*, recently identified as an important contributor to wine making in cool climates, mostly in the form of *S. cerevisiae* x *S. kudriavzevii* interspecific hybrids [16, 17]. While the glycolytic genes and enzymes of *S. kudriavzevii* have not been characterized, its genome sequence is available [18]. In *S. kudriavzevii* the complement of putative glycolytic genes and their sequences are different from the set of established glycolytic genes in *S. cerevisiae*, however putative *S. kudriavzevii* glycolytic genes with substantial homology (above

89% at the protein level) with their *S. cerevisiae* orthologs were found (Table S2). Major glycolytic isoenzymes in *S. kudriavzevii* were selected based on their high expression level of their structural genes during wine fermentation (Eladio Barrio, personal communication). *S. kudriavzevii* does not contain a *ScTDH3* homolog, while this is the most highly expressed glycolytic gene in *S. cerevisiae* [19]. However, *S. kudriavzevii* does harbor two putative *TDH* genes, *SkTDH1*, which closely resembles *ScTDH1,* and *SkTDH2*, which is more similar to *ScTDH2* and *ScTDH3*. Based on its high expression level during wine fermentation the gene homologous to *ScTDH1* was assumed to be the predominant homolog in this yeast and was selected to be part of the synthetic *S. kudriavzevii* glycolysis. Since many promoters within the formerly *sensu stricto Saccharomyces* family have been demonstrated to cross function within this family, [20] *S. kudriavzevii* glycoblocks were constructed with their native promoters and terminators. Following the previous methodology, the single-locus native glycolysis in GlycoSwitch was replaced by the *S. kudriavzevii* glycolysis. Transformants that only retained the *S. kudriavzevii* glycolysis displayed poor growth on chemically defined medium. A selected clone, *Sk*SinLoG IMX637, showed a strongly reduced growth rate on glucose, suggesting that the glycolytic function of the integrated set of *S. kudriavzevii* genes was suboptimal (Fig. 5.3b). Since *skTDH1* was not an orthologue of *ScTDH3*, we hypothesized that insufficient glyceraldehyde dehydrogenase activity was the cause for this decrease in maximum specific growth rate. Indeed growth was almost completely restored in comparison to SinLoG$_V$ IMX605 by transforming *Sk*SinLoG IMX637 with a 2μ plasmid containing the *SkTDH1* glycoblock resulting in an overexpression of *SkTDH1* (strain IMX652, Fig. 5.3b). The enzyme activities (Fig. 5.3c) confirmed that glyceraldehyde dehydrogenase activity was lower in the *S.k.*SinLoG strain in comparison to SinLoG$_V$ IMX605 and that overexpression of *skTDH1* did indeed boost glyceraldehyde dehydrogenase activity. Interestingly, the enzyme activities also revealed that the *in vitro* phosphofructokinase activity in the *S.k.* SinLoG strain was far too low to support a growth rate comparable to SinLoG$_V$ IMX605. Estimating the *in vivo* flux channeled by phosphofructokinase based on its *in vitro* activity (0.2 mmol.g dry biomass^{-1}.h^{-1} assuming a cellular soluble protein content of 33% of the dry weight), revealed that this *in vitro* enzyme activity was ca. 60 fold too low to support even the glycolytic flux necessary for the growth of *S.k.* SinLoG IMX637 at 0.15 h^{-1} (ca. 6 mmol.g$_{\text{dry biomass}}^{-1}$.h^{-1}). Remarkably, overexpression of *skTDH1*

FIGURE 5.3 | Construction and characterization of yeast strains with a remodeled glycolysis. The synthetic glycolytic pathways were introduced to chromosome V in a single step via *in vivo* assembly and targeted integration with the use of CRISPR/Cas9. The endogenous glycolytic cassette was subsequently removed using the CRISPR/Cas9 system in combination with a repair fragment of 120 bp (a). The constructed strains were analysed by Next-Gen Sequencing and analysed for mutations as compared to GlycoSwitch. Furthermore, the maximum specific growth rate on chemically defined medium with glucose as carbon source was measured. (b). The *in vitro* enzymatic activity of the glycolytic enzymes was measured in cell extracts (c). IMX652 was constructed by overexpression of the *SkTDH1* gene in the *Sk*SinLoG strain IMX637. Growth rate and enzyme activities determinations results from at least two independent culture replicates.

a. GLYCOLYSIS SWAPPING



b. SYNTHETIC GLYCOLYTIC VARIANTS

Native SinLog$_V$ (IMX605)

- *ADH1, HXK2, PGI1, PFK1, PFK2, FBA1, TDH3, PGK1, GPM1, ENO2, PYK1, PDC1, TPI1*

Mutation (SNP)	Genome	SinLoG
All	8	7
In ORF	1	2
MisSense	1, *IES2*	1, *ADH1*

	$\mu_{glucose}$
IMX605	0.28 ± 0.000

S. k. SinLoG (IMX637 & IMX652)

- *ADH1, HXK2, PGI1, PFK1, PFK2, FBA1, TDH1, PGK1, GPM1, ENO2, PYK1, PDC1, TPI1*
- IMX652: + *overexpression TDH1*

Mutation (SNP)	Genome	SinLoG
All	3	16
In ORF	2	1
MisSense	1, *GCN1*	0

	$\mu_{glucose}$
IMX637	0.13± 0.000
IMX652	0.25± 0.000

Mosaic SinLoG (IMX645)

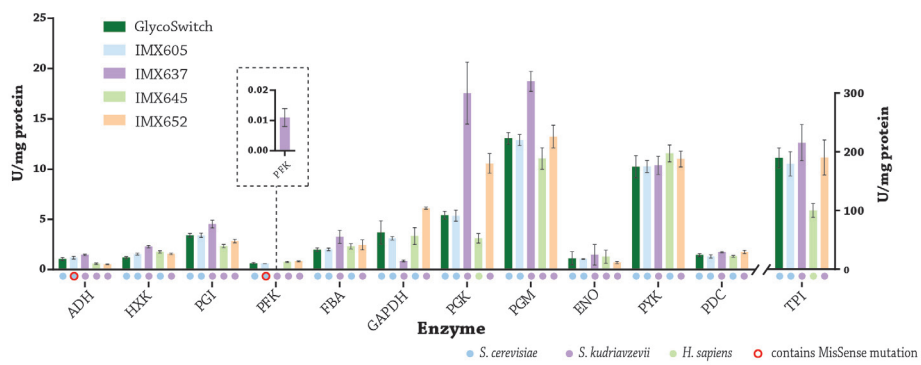
- *HXK2, FBA1, TDH3, PYK1, PDC1*
- *ADH1, PGI1, PFK1, PFK2, GPM1, ENO2*
- *TPI, PGK*

Mutation (SNP)	Number	SinLoG
All	1	4
In ORF	1	0
MisSense	0	0

	$\mu_{glucose}$
IMX645	0.28 ± 0.002

● *S. cerevisiae* ● *S. kudriavzevii* ● *H. sapiens*

c. ENZYMATIC ACTIVITIES



did fully restore the phosphofructokinase activity. This not previously observed response of phosphofructokinase to a glycolytic imbalance might point to an unknown regulatory mechanism. Although *SkPFK1* and *SkPFK2* are highly similar to their *S. cerevisiae* counterpart (ca. 98% at protein level, Fig. S7), they display several amino acid variations, and particularly one in an important nucleotide binding site [21]. We can speculate that the suboptimal SkTDH activity results in *S. cerevisiae* in a glycolytic imbalance and in modifications of intracellular glycolytic intermediates and effectors that, in turn, specifically affect SkPFK, but not ScPFK, activity. Alternatively, a similar response may exist in *S. cerevisiae*, but has not yet been uncovered. In either case, this work uncovered a compelling new regulatory mechanism. The replacement of the complete *S. cerevisiae* glycolysis composed of 26 genes of which a significant number is essential for survival of the cell by a synthetic heterologous variant is a milestone for research on essential metabolic pathways and paves the way for not just modifications to the core processes in the cell, but for complete redesign of those systems.

To illustrate this further, we constructed a mosaic SinLoG composed of a combination of five *S. cerevisiae*, five *S. kudriavzevii* and two *Homo sapiens* genes. As previously shown, the human *HsTPI1* and *HsPGK1* are able to fully complement their *S. cerevisiae* orthologue [22-24] and were therefore chosen for the mosaic SinLoG. The most abundant splicing variant of *TPI1* [25] and the single splicing variant of *PGK1* from human muscular tissue were codon-optimized (supplementary data 1) and each stitched to the yeast promoter and terminator of their respective orthologue. The resulting human glycoblocks were pooled with *HXK2*, *TDH3*, *PYK1*, *FBA1* and *PDC1* glycoblocks from *S. cerevisiae* and *PGI1*, *PFK1*, *PFK2*, *ENO2*, *GPM1* and *ADH1* from *S. kudriavzevii* and this mix was used to transform GlycoSwitch using the previous methodology and after removal of the native SinLoG this resulted in a strain in which solely the mosaic SinLoG was in charge of the glycolytic flux. A single colony was isolated (strain mosaic SinLoG IMX645) and sequenced, revealing the absence of the single-locus native glycolysis insert and the presence of the complete mosaic glycolysis. Just a single, non-sense nucleotide variation within ORFs of the mosaic SinLoG IMX645 strain was detected, and the mosaic SinLoG cassette was faithful to the *in silico* design (Fig. 5.3b, Fig. S6). Although *HsTPI1* and *HsPGK1* expression was driven by their orthologous *ScTPI1* and *ScPGK1* promoters, the *in vitro* enzyme activity of *Hs*TPI and *Hs*PGK was ca. 50% lower in the mosaic glycolysis strain as compared to the native $SinLoG_v$ IMX605 strain (Fig. 5.3c, t-test p-value < 0.01). Also the activity of SkADH was ca. 50% lower in IMX645 as compared to IMX605 (Fig. 3C, t-test p-value below 0.01). Remarkably, the strain carrying the mosaic SinLoG grew just as well as $SinLoG_v$ IMX605. While these observations are in line with the notion that the natural abundance of most glycolytic enzymes enable a glycolytic capacity far above the flux channeled *in vivo* [19], it still reflects the amazing robustness of yeast to major perturbations of an essential catabolic route.

DISCUSSION

Towards modular genome engineering | The high numbers of correct transformants obtained in the pathway swapping experiments demonstrate the efficiency and versatility of *in vivo* assembly and CRISPR-Cas9-facilitated genome editing in *S. cerevisiae* [6, 13]. Pathway swapping involved the transient, simultaneous presence in the yeast nucleus of a synthetic SinLoG on chromosome V and 35 kb of native SinLoG fragment(s) excised from chromosome IX, which share up to 100% identity. Nevertheless, unintended genome rearrangements caused by homologous recombination between these sequences were not observed in the selected strains.

Glycolytic genes are among the highest expressed genes in glucose-grown yeast cultures [26]. Construction of the GlycoSwitch platform strain therefore generated, on a single chromosomal locus, a 35 kb transcriptional hotspot. This co-localization of glycolytic genes had remarkably little impact on the expression of the glycolytic enzymes. In higher eukaryotic systems, it is well established that the expression of individual genes can be greatly affected by chromosomal position [27]. While the impact of genomic context on gene expression has not been established in *S. cerevisiae*, the location of a complete set of glycolytic genes on either of two different chromosomes did not strongly affect expression levels. The GlycoSwitch platform offers an excellent tool to systematically explore the impact of genomic context and spacing of genes on gene expression and, thereby, to guide the design of synthetic yeast chromosomes.

Functional replacement of the entire *S. cerevisiae* glycolysis by that of its close relative, the cold-tolerant yeast *S. kudriavzevii*, provided a proof of principle that pathway swapping can be used to rapidly express and study metabolic pathways in a heterologous context. *S. kudriavzevii* and *S. cerevisiae* are sympatric and both show fast, fermentative sugar dissimilation in glucose-rich media [28]. Pathway swapping demonstrated that a set of *S. kudriavzevii* glycolytic enzymes can support glycolysis and fast growth of *S. cerevisiae*. Co-evolution in the same ecological niches may have led to similar optima in term of expression level and transcriptional regulation and explain the highly similar activities observed for most glycolytic enzymes upon replacement of all *S. cerevisiae* glycolytic genes by their *S. kudriavzevii* counterparts, controlled by their native promoters. However, prior to this study, little was known on the kinetics and regulation of glycolytic enzymes in this cold-tolerant yeast. The pathway swapping experiments identified interesting leads for follow-up studies on *S. kudriavzevii* glycolytic enzymes, including a strong sensitivity of *Sk*PFK to the expression level of glyceraldehyde dehydrogenase.

Humanized yeast strains can be used as powerful models to explore effects of therapeutics, gene dosage and of wild-type or disease-causing variants of human genes on protein function [29]. Hitherto, replacement of entire pathways remains a technically challenging and time consuming activity, as exemplified by studies on humanization of protein glycosylation in the yeast *Pichia pastoris* [30]. Recent large-scale studies on the ability of human genes to complement native genes in *S. cerevisiae* demonstrated that complementation of haploid yeast gene knockouts is a reliable approach for the functional characterization of human gene variants [23]. However elegant, such studies are limited to single-gene complementation and require the generation of multiple yeast strains that each contain only a single orthologue of a studied heterologous gene. Pathway swapping enables the systematic analysis of heterologous complementation of entire pathways and should enable humanization of the complete glycolytic pathway. Availability of strains containing a fully or partially humanized glycolytic pathway will enable to test the impact of mutations or drugs on human proteins in their natural glycolytic context, and thereby to identify potential synergetic effects between native human proteins.

The modular pathway swapping approach opens up unprecedented possibilities. A myriad of applications ranging from functional analysis of heterologous proteins, testing of kinetic modeling (now hindered by the multiplicity of paralogs) or screening drugs, to more technical aspects such as exploring the effect of genomic location of highly expressed native pathways, are now within reach. Furthermore, future improvements in CRISPR-removal of scattered genes [13] should enable the functional clustering and fast, modular swapping of key pathways/processes. A large worldwide effort has led to the first synthetic yeast chromosome and is progressing towards the synthesis of the entire yeast genome [9]. The present study indicates that modular design of complete synthetic yeast genomes offers unprecedented possibilities for fast, combinatorial analysis of pathway configurations and development of new experimental platforms for fundamental and applied research.

ACKNOWLEDGMENTS

We thank M. Niemeijer, R. Brinkman and M. Zaidi for their valuable contribution to the synthesis and confirmation of the synthetic *S. kudriavzevii* and *H. sapiens* glycoblocks. We thank Eladio Barrio Esparducer for kindly providing *S. kudriavzevii* CR85 and for advice on selecting the *S. kudriavzevii* genes for the *Sk*SinLoG strain.

This work was supported by the Technology Foundation STW (Vidi grant 10776).

REFERENCES:

1. Paddon CJ, Keasling JD: **Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development.** *Nature Rev Microbiol* 2014, **12:**355-367.

2. Serrano L: **Synthetic biology: promises and challenges.** *Mol Syst Biol* 2007, **3:**158.

3. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA, 3rd: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome.** *Proc Natl Acad Sci U S A* 2008, **105:**20404-20409.

4. Barnett JA: **A history of research on yeasts 5: the fermentation pathway.** *Yeast* 2003, **20:**509-543.

5. Solis-Escalante D, Kuijpers NGA, Barrajon-Simancas N, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P: **A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism.** *Eukaryotic Cell* 2015, **14:**804-816.

6. Kuijpers NGA, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P: **A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences.** *Microb Cell Fact* 2013, **12:**47.

7. Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P: **amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:**126-139.

8. Kuijpers NGA, Chroumpi S, Vos T, Solis-Escalante D, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P: **One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2013, **13:**769-781.

9. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, et al: **Total synthesis of a functional designer eukaryotic chromosome.** *Science* 2014, **344:**55-58.

10. Solis-Escalante D, Kuijpers NGA, van der Linden FH, Pronk JT, Daran JM, Daran-Lapujade P: **Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2014, **14:**741-754.

11. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339:**819-823.

12. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM: **Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems.** *Nucleic Acids Res* 2013, **41:**4336-4343.

13. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJ, Daran JM: **CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2015, **15**.

14. Ide S, Miyazaki T, Maki H, Kobayashi T: **Abundance of ribosomal RNA gene copies maintains genome integrity.** *Science* 2010, **327:**693-696.

15. Bausch C, Noone S, Henry JM, Gaudenz K, Sanderson B, Seidel C, Gerton JL: **Transcription alters chromosomal locations of cohesin in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 2007, **27:**8522-8532.

16. Lopes CA, Barrio E, Querol A: **Natural hybrids of *S. cerevisiae* x *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*.** *FEMS Yeast Res* 2010, **10:**412-421.

17. Arroyo-Lopez FN, Orlic S, Querol A, Barrio E: **Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae*, *S. kudriavzevii* and their interspecific hybrid.** *Int J Food Microbiol* 2009, **131:**120-127.

18. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT: **The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus.** *G3 (Bethesda)* 2011, **1:**11-25.

19. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, Heck AJ, Daran JM, de Winde JH, Westerhoff HV, et al: **The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels.** *Proc Natl Acad Sci USA* 2007, **104:**15753-15758.

20. Oliveira BM, Barrio E, Querol A, Perez-Torrado R: **Enhanced enzymatic activity of glycerol-3-phosphate dehydrogenase from the cryophilic *Saccharomyces kudriavzevii*.** *PLoS ONE* 2014, **9:**e87290.

21. Banaszak K, Mechin I, Obmolova G, Oldham M, Chang SH, Ruiz T, Radermacher M, Kopperschläger G, Rypniewski W: **The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle.** *J Mol Biol* 2011, **407:**284-297.

22. Gruning NM, Du D, Keller MA, Luisi BF, Ralser M: **Inhibition of triosephosphate isomerase by phosphoenolpyruvate in the feedback-regulation of glycolysis.** *Open Biol* 2014, **4:**130232.

23. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM: **Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity.** *Science* 2015, **348:**921-925.

24. Hamza A, Tammpere E, Kofoed M, Keong C, Chiang J, Giaever G, Nislow C, Hieter P: **Complementation of yeast genes with human genes as an experimental platform for functional testing of human genetic variants.** *Genetics* 2015, **201:**1263-1274.

25. Maquat LE, Chilcote R, Ryan PM: **Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man.** *J Biol Chem* 1985, **260:**3748-3753.

26. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278:**680-686.

27. Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF, van Lohuizen M, van Steensel B: **Chromatin position effects assayed by thousands of reporters integrated in parallel.** *Cell* 2013, **154:**914-927.

28. Goncalves P, Valerio E, Correia C, de Almeida JM, Sampaio JP: **Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species.** *PLoS One* 2011, **6:**e20739.

29. Dunham MJ, Fowler DM: **Contemporary, yeast-based approaches to understanding human genetic variation.** *Curr Opin Genet Dev* 2013, **23:**658-664.

30. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, et al: **Humanization of yeast to produce complex terminally sialylated glycoproteins.** *Science* 2006, **313:**1441-1443.

31. Nijkamp JF, van den Broek MA, Geertman JM, Reinders MJ, Daran JM, de Ridder D: ***De novo* detection of copy number variation by co-assembly.** *Bioinformatics* 2012, **28:**3195-3202.

32. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik M, Daran-Lapujade P, Vongsangnak W, Nielsen J, Heijne WH, et al: ***De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology.** *Microb Cell Fact* 2012, **11:**36.

33. Entian KD, Kotter P: **Yeast genetic strain and plasmid collections.** *In* Stansfield I, Stark MJR (ed) *Methods in microbiology*. Amsterdam: Academic Press; 2007:629-666.

34. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, Giuseppin ML, Heijnen JJ, Hoare M, Lange HC, et al: **An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains.** *Enzyme MicrobTechnol* 2000, **26:**706-714.

35. Verduyn C, Postma E, Scheffers W, van Dijken JP: **Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.** *Yeast* 1992, **8:**501-517.

36. Pronk JT: **Auxotrophic yeast strains in fundamental and applied research.** *Appl Environ Microbiol* 2002, **68:**2095-2100.

37. Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH: **A new efficient gene disruption cassette for repeated use in budding yeast.** *Nucleic Acids Res* 1996, **24:**2519-2524.

38. Goldstein AL, McCusker JH: **Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*.** *Yeast* 1999, **15:**1541-1553.

39. Gritz L, Davies J: **Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia* coli and *Saccharomyces cerevisiae*.** *Gene* 1983, **25:**179-188.

40. Gietz RD, Woods R: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Methods Enzymol* 2002, **350:**87-96.

41. Gritsenko A, Reinders MT, de Ridder D: **Using Predictive Models to Engineer Biology: A Case Study in Codon Optimization.** *In* Ngom A, Formenti E, Hao J-K, Zhao X-M, van Laarhoven T (ed) *Pattern Recognition in Bioinformatics*. Springer Berlin Heidelberg; 2013.

42. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.** *Nucleic Acids Res* 2002, **30:**e23.

43. Mumberg D, Muller R, Funk M: **Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.** *Gene* 1995, **156:**119-122.

44. Shevchuk NA, Bryksin AV, Nusinovich YA, Cabello FC, Sutherland M, Ladisch S: **Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously.** *Nucleic Acids Res* 2004, **32:**e19.

45. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM: **Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.** *PLoS One* 2014, **9:**e112963.

46. Postma E, Verduyn C, Scheffers WA, Van Dijken JP: **Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae.** *Appl Environ Microbiol* 1989, **55:**468-477.

47. Jansen ML, Diderich JA, Mashego M, Hassane A, de Winde JH, Daran-Lapujade P, Pronk JT: **Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity.** *Microbiology* 2005, **151:**1657-1669.

48. Cruz LA, Hebly M, Duong GH, Wahl SA, Pronk JT, Heijnen JJ, Daran-Lapujade P, van Gulik WM: **Similar temperature dependencies of glycolytic enzymes: an evolutionary adaptation to temperature dynamics?** *BMC Syst Biol* 2012, **6:**151.

49. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ: **Protein measurement with the Folin phenol reagent.** *J Biol Chem* 1951, **193:**265-275.

50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

51. Solis-Escalante D: **Reducing the generic complexity of glycolysis in *Saccharomyces cerevisiae*.** *PhD Thesis*. Delft University of Technology, Biotechnology; 2015.

SUPPORTING INFORMATION

SECTION 1: MATERIAL & METHODS

Strains & Media | The *Saccharomyces kudriavzevii* strain used in this study was a wild isolate obtained from oak bark (Ciudad Real, Spain) [16] (Table S3) and was grown on complex (YPD) medium containing 10 $g \cdot l^{-1}$ Bacto Yeast extract, 20 $g \cdot l^{-1}$ Bacto Peptone and 20 $g \cdot l^{-1}$ glucose as carbon source. Due to the lower optimal growth temperature of *S. kudriavzevii* as compared to *Saccharomyces cerevisiae*, this strain was cultivated at 16˚C. The *S. cerevisiae* strains used in this study are all derived from the CEN.PK family (Table S3) [32-34]. Cultures for transformation were grown in YPD medium. When galactose induction of *SCEI* was required, cultures were transferred to YPGal medium containing 20 $g \cdot l^{-1}$ galactose and grown for 4 hours on that medium prior to transformation [8]. Synthetic media (SM) contained, per liter of demineralized water, 5 g $(NH_4)_2SO_4$, 3 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7 \cdot H_2O$, and trace elements [35]. The pH was set at 6.0 by 1M KOH. Vitamins as previously described [35] were added after heat sterilization of the medium at 120 °C for 20 min. Glucose was separately sterilized at 110 °C and added to a final concentration of 20 $g \cdot l^{-1}$. When required, the medium was supplemented with 150 $mg \cdot L^{-1}$ uracil to rescue the auxotrophy [36]. SM without nitrogen source (SMwn) was prepared by replacing $(NH_4)_2SO_4$ by 6.6 $g \cdot l^{-1}$ K_2SO_4. In media for determination of growth rates, SMU was prepared by supplementing SMwn with 2.3 $g \cdot l^{-1}$ urea (filter sterilized) to reduce acidification of the medium during growth as compared to medium with ammonium as nitrogen source. The pH was set to 6. For strain selection using amdSYM [7], 1.8 $g \cdot l^{-1}$ acetamide was added to SMwn. For counter-selection of the *URA3* and *KlURA3* marker gene, SMwn was supplemented with 3.53 $g \cdot l^{-1}$ proline, 0.010 $g \cdot l^{-1}$ uracil and 0.20 $g \cdot l^{-1}$ 5-fluoroorotic acid (5-FOA) (Sigma Aldrich, St. Louis, MO). For the selection of mutants carrying the marker genes kanMX [37], natNT1 [38] or hphNT1 [39], 200 $mg \cdot l^{-1}$ G418, 100 $mg \cdot l^{-1}$ nourseothricin or 200 $mg \cdot l^{-1}$ hygromycin, respectively, were added to complex medium. Solid versions of the media described above were prepared by adding 2% (w/v) agar prior to heat sterilization.

Molecular Biology techniques | PCR amplification for cloning purposes was performed using Phusion® Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA). To improve PCR efficiency, the conditions in the PCR reaction as recommended by the supplier were modified by decreasing the primer concentration from 500 nM to 200 nM and increasing the Phusion™ Hot Start High Fidelity polymerase concentration from 0.02 U μl^{-1} to 0.03 U μl^{-1}. All other conditions were in agreement with the manufacturer's instructions. Analytical PCR's were performed using the DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to manufacturer's recommendations. Genomic DNA as template for the glycoblocks was isolated from *S. cerevisiae* CEN.PK113-7D and *S. kudriavzevii* CR85 using the Qiagen 100/G kit (Qiagen, Hilden, Germany). Genomic DNA isolation for analytical purposes was performed with the YeaStar kit (Zymo Research, Irvine, CA). Plasmids maintained in *E. coli* DH5α were isolated with the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich). PCR products were separated in 1% (w/v) agarose (Sigma) gels in 1x TAE (40 mM Tris-acetate pH 8.0 and 1 mM EDTA), 2% (w/v) agarose in 0.5x TBE (45

mM Tris-borate pH 8.0 1 mM EDTA) was used when fragments were smaller than 500 bp. Glycoblocks were isolated from gel using the Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA). Prior to transformation, fragments were pooled, maintaining equimolar concentrations (150 fmol/fragment). with the DNA fragments containing the marker [8]. Yeast transformation was performed with the LiAc/ssDNA method [40].

Plasmids used in this study are given in (Table S4). Primers for the construction of plasmids are given in (Table S5).

Construction of glycoblocks and marker cassettes | The glycolytic gene cassettes containing the SHR-sequences (glycoblocks) were obtained by extension PCR. Genomic DNA of *S. cerevisiae* CEN.PK113-7D and *S. kudriavzevii* CR85 was used as PCR template for glycolytic genes. Each glycolytic gene was cloned with its own promoter and terminator. Promoter sizes for the *S. cerevisiae* genes were chosen using the following criteria: 1) approximately 800 bp of upstream sequence was considered as promoter sequence and, 2) If this sequence would overlap with an upstream located gene, the promoter size was limited to not contain any coding sequence of the neighboring gene. For the *S. kudriavzevii* genes approximately 800 bp upstream of the gene was picked as promoter sequence. Terminator sequences were in all cases approximately 200 bp of the downstream sequence of the genes. All primers are given in (Table S6), whereby the names indicate the gene and SHR-sequences of the resulting glycoblocks.

To add extra restriction sites for HO and I-CreI endonucleases, enabling later excision of the single locus synthetic glycolysis, the *PDC1* glycoblock was prepared differently. *PDC1* was obtained by PCR amplification from CEN.PK113-7D genomic DNA using primers PDC1 Fw+RES and PDC1 Rv+M (Table S6). The fragment SYN2 was obtained by fusion PCR of the oligo's Syn2 Fw and Syn2 Rv using primers FUS2 Fw and FUS2 Rv (Fig. S7). The resulting product was cloned in a pCR™4Blunt-TOPO® vector and verified by restriction/digestion, resulting in pUD336. The glycoblock PDC1-$SYN_{FM,}$ was obtained from pUD336, using primers FUS2 Fw and FUS2 Rv.

The *Homo sapiens* genes *TPI1* (muscle, splicing variant 1) and *PGK1* (muscle, splicing variant 1) (Table S2), were codon optimized [41] and chemically synthesized (GeneArt, Life Technologies). The resulting plasmids pSYN-TPI1 and pSYN-PGK1 were used as a template to PCR amplify the codon optimized ORFs with primers given in (Table S5). For each gene the promoter and terminator of the *S. cerevisiae* orthologous gene was amplified from CEN.PK113-7D genomic DNA thereby adding overlapping sequences to the accompanying synthesized ORF, using primers given in (Table S5). The promoter, ORF and terminator were mixed in equimolar amounts, normalized to 100 ng of the ORF, and stitched by fusion-PCR. The resulting products were cloned in pCR™4Blunt-TOPO®vectors and verified by restriction/digestion, yielding pUD329 (p*TPI1-HsTPI1-tTPI1*) and pUD331 (p*PGK1-HsPGK1-tPGK1*). Plasmids pUD329 and pUD331 were used as a template for the human glycoblocks *TPI1* and *PGK1* respectively.

The amdSYM and kanMX marker cassettes were obtained by PCR with pUGamdSYM [7] and pUG6 [42] as templates respectively. Primers are given in (Table S6).

All cassettes were gel-purified prior to transformation and the concentrations were measured in a NanoDrop 2000 spectrophotometer (wavelenght 260 nm) (Thermo Fisher Scientific).

Construction of deletion cassettes and CRISPR-Cas plasmids | The endogenous *PYK1, PGI1, TPI1, TDH3* and *PGK1* were deleted using standard deletion techniques. Cassettes for the deletion of these genes were obtained as previously described [10] using the pDS-plasmid series (Table S4). Primers to obtain the deletion cassettes are given in (Table S6). Depending on the desired marker, the appropriate plasmid template was used. Cassettes were gel-purified and 500 ng of each cassette was used for transformation.

To enable CRISPR-Cas mediated genome editing, the gene encoding Cas9, driven by the constitutive *TEF1* promoter, was integrated in the genome of strain IMX511. Two fragments were constructed to replace the native locus of the deleted *pfk2* gene with *cas9* (Fig. S4). A cassette containing *cas9* was obtained by PCR with p414-TEF1p-cas9-CYC1t [12] as template and primers CAS9 Fw+pfk2 and CAS9 Rv+link (Table S7). A second cassette containing the natNT1 marker gene was obtained by PCR on plasmid pUGnatNT1 with primers nat Fw+link and nat Rv+Rpt+pfk2 (Table S7). Both cassettes were gel purified and pooled in equimolar amounts. Of this mixture 500 ng was used to transform IMX511, leading to IMX535.

Consecutively *HXK2*, *FBA1*, *ENO2*, *GPM1*, *PFK1*, *PDC1* and *ADH1* were deleted using CRISPR-Cas9 editing [13]. To rescue the double strand DNA break (DSB) introduced by Cas9, 120 bp marker-free deletion cassettes, called repair fragments, were sufficient to generate deletions. dsDNA repair fragments were constructed by annealing complementing oligo's listed in Table S7 [13].

The expression cassettes for the guide RNA (gRNA) to introduce Cas9-mediated DSB in *HXK2*, *FBA1*, *ENO2*, *GPM1* and *PFK1*, flanked by SHR-sequences [6] were chemically synthesized (GeneArt). The plasmids containing the synthesized DNA as supplied by the manufacturer were used as template to obtain the gRNA expression cassettes including the SHR-sequences by PCR. Primers are given in (Table S5). To incorporate the gRNA cassettes in a yeast expression vector, p426-GPD [43] was linearized by PCR with primers adding SHR-sequences corresponding to the SHR-sequences of the gRNA cassettes (Table S5). The gRNA cassettes were assembled into the p426 backbone by Gibson assembly (New England Biolabs, Ipswich, MA) following the manufacturer's recommendations (Fig. S9A). Each plasmid contained a single gRNA. For each deletion 100 ng of the appropriate CRISPR-plasmid was co-transformed with 1.5 μg of the corresponding repair fragment. The *ENO2* gene in its endogenous locus is closely flanked by other genes. Hence, using deletion sites outside the *ENO2* gene may lead to the interruption of these adjacent genes and may affect the strain phenotype, deletion sites therefore had to be designed on a sequence also present in the *ENO2* glycoblock.

To prevent deletion of the *ENO2* gene on the single locus synthetic glycolysis two different repair fragments were used (Fig. S10).

PDC1 and *ADH1* were deleted simultaneously using Cas9. The two plasmids carrying the gRNA targeting *PDC1* and *ADH1* were constructed using *in vivo* assembly. Plasmid p426-SNR52p-gRNA.CAN1.Y-SUP4t [12] was linearized with primers p426-crispr Fw and p426-crispr Rv. The 120 bp targeting fragments (*crPDC1* and *crADH1*) were obtained by annealing complementing oligo's as previously described [13] (Table S5). A mix consisting of 100 ng of the linearized CRISPR-backbone, 300 ng of *crPDC1*, 300 ng of *crADH1* and 1.8 µg of each of the appropriate repair fragments (obtained as described above, Table S7) was used for transformation.

Two additional CRISPR-plasmids, one targeting the deletion cassette amdSYM and the other targeting the flanking regions of the synthetic glycolysis construct, were constructed. Plasmids were designed as previously described [13] (Fig. S9B). As described above, the linearized plasmid backbone was obtained with primers p426-crispr Fw and p426-crispr Rv from p426-SNR52p-gRNA.CAN1.Y-SUP4t [12] and the 120 bp targeting fragments (cramdSYM and crRECYCLE) were obtained by annealing complementing 120 bp oligo's (Table S5). The backbone and the desired targeting fragments were assembled into the CRISPR plasmids by Gibson assembly resulting in pUDE337 carrying cramdSYM and pUDE342 carrying crRECYCLE (Fig. S9B).

Construction of the GlycoSwitch strain | An overview of the construction of the GlycoSwitch strain is given in Fig. S2. To prepare the locus for chromosomal integration of the synthetic glycolysis, an I-SceI restriction site was introduced on chromosome IX, at the *SGA1* locus. The cassette carrying the I-SceI recognition site targeted to *SGA1* also carried the *SCEI* gene which encodes an intron-encoded homing endonuclease, under the control of the galactose inducible promoter *GAL1*, and the selection marker *KlURA3*. First the *SCEI/KlURA3* cassette was obtained by PCR using IMX221 genomic DNA as template [8] and the primers Tag G Fw and SGA1 Rv (Fig. S11A). Fragment SYN1 (Fig. S1B), was obtained by mixing the oligonucleotides Syn1 Fw and Syn1 Rv. The resulting fragment SYN1 and the *SCEI/KlURA3* cassette were gel-purified and fused by fusion-PCR [44] using primers FUS1 Fw and FUS1 Rv (Fig. S11B and S11C). The resulting product was cloned in a pCR™4Blunt- TOPO®vector (Invitrogen, Life Technologies), resulting in pUD335 and verified by restriction/digestion. The *KlURA3-SCEI* cassette was obtained by PCR from pUD335 using primers FUS1 Fw and FUS1 Rv (Table S6) (Fig. S1A).

The *S. cerevisiae* strain IMX370 (Fig. S2)[5], carrying a minimal set of glycolytic genes, was transformed with 100 ng of the *KlURA3-SCEI* cassette (Fig. S1A), resulting in IMX377. IMX377 also harbors in the integrated *KlURA3-SCEI* cassette additional restriction sites recognized by the HO and I-CreI endonucleases, and homologous flanking regions to promote recombination upon excision of the endogenous Single Locus Glycolysis (SinLoG) cassette (Fig. S1A).

The endogenous SinLoG cassette was assembled and integrated in IMX377 using the

Combined *in vivo* Assembly and Targeted chromosomal *I*ntegration (CATI) approach [8]. IMX377 was transformed with a mix consisting of the *S. cerevisiae* glycoblocks and the *amdS* marker cassette (*FBA1*$_{GH}$, *TPI1*$_{HP}$, *PGK1*$_{PQ}$, *ADH1*$_{QN}$, *PYK1*$_{NO}$, *TDH3*$_{OA}$, amdSYM$_{AB}$, *HXK2*$_{BC}$, *PGI*$_{CD}$, *PFK1*$_{DJ}$, *PFK2*$_{JK}$, *ENO2*$_{KL}$, *GPM1*$_{LM,}$ *PDC1-SYN*$_{MF}$) (the subscript letters indicate the SHR-sequences, Fig. S1B). The molar ratio of transformed fragments was 1:1 normalized to 150 ng of the amdSYM$_{AB}$ cassette. Transformants were selected on medium containing acetamide as sole nitrogen source. Clones were analyzed for presence of all junctions between glycoblocks and selection markers with primers given in (Table S8). One colony that showed correct PCR patterns was selected and named IMX382. This strain was further analyzed by sequencing a set of 14 PCR products obtained with primer pairs 1 to 14 (Table S8) (Fig. S4D). All PCR products were pooled in a molar ratio of 1:1. From this set of 14 products a library of 300 bp insert was constructed and paired end sequenced (100 bp paired end reads) using an Illumina HiSeq 2500 sequencer (BaseClear, Leiden, The Netherlands). The sequence reads were mapped onto the synthetic construct using Burrows-Wheeler Alignment tool (using "BWA mem" command; version 0.7.10-r789) and the resulting Alignment file (BAM file) was further processed by Pilon (version 1.10; using "--vcf --fix all,breaks" parameter[45]) for variant detection which were stored in VCF (Variant Call Format) file.

To construct the GlycoSwitch strain, the 13 genes made redundant by the newly added synthetic locus were removed from IMX382 in the following order: *PYK1, PGI1, TPI1, TDH3, PFK2, PGK1, GPM1, FBA1, HXK2, PFK1, ADH1, PDC1, ENO2* (Fig. S2). The natNT1, kanMX and hphNT1 marker were used for the deletion of *PYK1, PGI1* and *TPI1* respectively. Those markers were excised using I-SceI as previously described [10] by transforming strain IMX493 with plasmid pUDC073 carrying *SCEI* (Fig. S3B). *PFK2* was deleted by a cassette containing *cas9* and the natNT1 marker cassette. The *KlURA3* and kanMX markers used for the subsequent deletion of *TDH3* and *PGK1* were recycled by the same I-SceI facilitated marker removal, by transforming the *SCEI* expressing plasmid pUDE206 to IMX557. The deletion of *GPM1, FBA1, HXK2* and *PFK1* was performed by the CRISPR/Cas9 system by transforming the appropriate CRISPR-plasmid and accompanying repair fragment. *ADH1* and *PDC1* were simultaneously deleted using the CRISPR/Cas9 cloning free deletion method as previously described [13]. Transformants were selected on SM and the CRISPR-plasmids were recycled by growing the strain overnight on YPD medium followed by plating on SM medium with 5-FOA. In order to restore a functional *ENO2* glycoblock to the single locus glycolysis, a glycoblock containing *ENO2* with a longer promoter sequence (*ENO2-LONG*$_{AB}$) was introduced to the synthetic construct by replacing the *amdSYM*$_{AB}$ marker cassette in IMX583 resulting in strain IMX586 (Fig. S2). This was achieved by transforming CRISPR-plasmid pUDE337 together with the *ENO2-LONG*$_{AB}$ glycoblock. Transformants were selected on SM. Subsequently, the endogenous *ENO2* gene could be deleted by co-transforming the CRISPR-plasmid pUDE326 and the corresponding repair fragments in IMX586 resulting in IMX587. Transformants were selected on SM. Finally the dysfunctional glycoblock ENO2$_{KL}$ was replaced by transforming 500 ng of marker cassette amdSYM$_{KL}$. Transformants were selected on SMwn with acetamide and one clone displaying the correct PCR profile was plated on medium with 5-FOA to recycle

the CRISPR-plasmid and stocked as IMX589 (auxotrophinc GlycoSwitch strain). To be able to perform growth experiments on SM, the uracil auxotrophy was repaired by transforming IMX589 with pUDE325, resulting in the prototrophic GlycoSwitch strain IMX606.

Construction of synthetic glycolytic pathways in the *CAN1* locus | The synthetic glycolytic pathways introduced in the *CAN1* locus were obtained by transforming IMX589 with a mix of glycoblocks for assembly and targeted integration of the desired synthetic glycolytic pathway. To facilitate the targeted integration into the genome, a similar approach to the CATI approach was chosen, but the CRISPR/Cas9 system was used instead of I-SceI to promote the formation of a double strand break and therefore integration of synthetic construct at the targeted locus. Therefore 300 ng of p426-SNR52p-gRNA.CAN1.Y-SUP4t plasmid coding for the gRNA targeting the *CAN1* locus [12] was co-transformed with the glycoblocks, as well as a cassette carrying the selection marker kanMX. The glycoblocks for the native $SinLoG_V$ were: $FBA1_{can1H}$, $TPI1_{HP}$, $PGK1_{PQ}$, $ADH1_{QN}$, $PYK1_{NO}$, $TDH3_{OA}$, $ENO2_{AB}$, $HXK2_{BC}$, PGI_{CD}, $PFK1_{DJ}$, $PFK2_{JK}$, $GPM1_{LM,}$ $PDC1_{Mcan1}$. For the *S.k.* SinLoG were used: $skFBA1_{can1H}$, $skTPI1_{HP}$, $skPGK1_{PQ}$, $skADH1_{QN}$, $skPYK1_{NO}$, $skTDH1_{OA}$, $skHXK2_{BC}$, $skPGI_{CD}$, $skPFK1_{DJ}$, $skPFK2_{JK}$, $skENO2_{KL}$, $skGPM1_{LM,}$ $skPDC1_{Mcan1.}$ For the mosaic SinLoG the following mixture was transformed: $FBA1_{can1H}$, $hsTPI1_{HP}$, $hsPGK1_{PQ}$, $skADH1_{QN}$, $PYK1_{NO}$, $TDH3_{OA}$, $skHXK2_{BC}$, $skPGI_{CD}$, $skPFK1_{DJ}$, $skPFK2_{JK}$, $skENO2_{KL}$, $skGPM1_{LM,}$ $PDC1_{Mcan1}$. Cassettes were mixed in a 1:1 molar ratio normalized to 140 ng of the kanMX cassette. Selection was on SM for presence of the CRISPR-plasmid, which contained the *URA3* marker. For each transformation eight clones were plated to medium selective for kanMX. Resistant clones were analyzed by PCR for presence of the full synthetic glycolysis cassette with primers given in Table S7. For each synthetic glycolysis variant, a correctly assembled strain was grown on complex medium and plated on SM proline with 5-FOA and uracil to recycle the CRISPR-plasmid. The resulting strains were stocked on SM acetamide supplemented with uracil (IMX591, IMX607, IMX633).

Excision of the native SinLoG cassette from chromosome IX | Strains IMX591, IMX607 and IMX633 contained the native SinLoG at the *SGA1* locus. This cassette was removed using CRISPR-Cas9. To remove the native SinLoG 100 ng of the CRISPR-plasmid pUDE342 was transformed to IMX591, IMX607 and IMX633 together with 1.5 µg of the recycle repair fragment (Table S7)(Fig. S12). Transformants were plated on SM glucose and were analyzed for removal of the endogenous SinLoG cassette by PCR with primers SGA1 Fw and SGA1 Rv (Table S9).

***skTDH1* overexpression in IMX637 |** A plasmid backbone obtained by PCR using the plasmid p426-GPD as template with the primers p426-rv+O and p426-fw+A (Table S5) and the *skTDH1* glycoblock were assembled *in vitro* using Gibson assembly, resulting in the plasmid pUDESkTDH1 (Table S4). IMX637 was plated on complex medium with 5-FOA to counterselect the pUDE342 plasmid. A selected colony was then transformed with 100 ng of the pUDEskTDH1 plasmid and transformants were selected on SM. One transformant was stocked as IMX652.

Sequencing | Strains IMX589 (auxotrophic GlycoSwitch), IMX605 (endogenous SinLoG on chromosome V), IMX637 (*Sk*SinLoG) and IMX645 (mosaic SinLoG) were sequenced. Ilumina Nextera libraries were constructed (300 bp inserts size) and paired end sequenced (100 bp reads) using an Illumina HISeq 2500 sequencer at Baseclear BV (Leiden, The Netherlands). A minimum quantity of 750 Mb was generated, representing a minimum 60-fold coverage. The genomes were *de novo* assembled using the gsAssembler (version 2.6) software package, also known as the Newbler software package (454 Life Sciences, Branford, CT). To verify the deletions in IMX589, all contigs were mapped to the *in silico* design after gene removal using Clustal X in Clone Manager 9 (Sci-Ed Software, Cary, NC). To verify the integration of the different synthetic glycolytic pathways, contigs were mapped to the *in silico* design. To exclude possible duplications of glycolytic genes, a copy number variation analysis was performed using the Magnolya algorithm [31] (Fig. S6).

In order to compare the strains on nucleotide level, all sequence libraries of samples IMX372 [5], IMX589, IMX605, IMX637 and IMX645 were processed by an in-house pipeline hosted in Galaxy (https://galaxyproject.org/). The samples were mapped to CEN.PK113-7D [32] for whole genome comparison and to the *in silico* design of the SinLoG present in the specific strain. The Burrows-Wheeler Alignment tool (BWA, version 0.7.10-r789) was used and the resulting binary alignment file (BAM file) was further processed using SAMtoolsmpileup (version 0.1.18) and bcftools (from the SAMtools package) to compute the genotype likelihood and stores these likelihoods in Binary variant call format (BCF). The script vcfutils.pl was used, with parameter varFilter and maximum read depth 400, to filter and convert to variant call format (VCF). The resulting VCF files were annotated and effects of variants on genes were predicted by the snpEff package (version 3.4). To compare IMX589 to IMX372, The called and annotated variants in both IMX589 and IMX372 samples were subtracted from sample IMX589 with the "subtract whole dataset from another dataset" tool in Galaxy. The same procedure was followed to compare IMX605, IMX637 and IMX645 to IMX589.

Determination of growth rates | Glycerol stocks from strains IMX372 (MG), IMX606 (prototrophic GlycoSwitch), IMX605 (endogenous $SinLoG_V$), IMX637 (*Sk*SinLoG), IMX652 (*Sk*SinLoG with *SkTDH1* overexpression), IMX645 (mosaic SinLoG) were inoculated in 100 ml SM urea + 2% glucose (w/v) in 500 ml shake flasks and grown to late exponential phase. Cells were harvested and immediately transferred to pre-warmed 500 ml flasks containing the same medium at an OD660 of 0.2. Biomass formation was followed by measuring the OD660. Concentration of extracellular metabolites in culture supernatants was measured by HPLC using a Aminex HPX-87H ion exchange column operated at 60 °C with 5 mM H_2SO_4 as mobile phase at an isocratic flow rate of 0.6 ml· min^{-1}. The data reported in the results section are calculated based on at least two independent culture replicates.

Determination of *in vitro* enzyme activities | Samples equivalent to 62.5 mg of dry weight biomass taken at the mid-exponential phase of shake flask grown cultures were used to obtain cell extracts as previously described [46]. Measurement of the activity of the glycolytic enzymes was carried out as previously described [47], except for phosphofructokinase, of which the activity was determined as described in [48]. Enzyme activities are expressed as units/mg of protein. Total protein concentrations in the cell extracts were determined as described in [49] with bovine serum albumin as a standard. The data reported in the results section are calculated based on at least two analytical replicates for each enzyme assayed and at least two independent culture replicates.

SECTION 2: TABLES & FIGURES

TABLE S1 | Single nucleotide variations identified in the constructed strains.

Single nucleotide variations (SNV's)			
Systematic name	Name	Type	Amino acid change
GlycoSwitch (IMX589) vs. Minimal Glycolysis (IMX372)			
Genome			
YBR079W	*VPS15*	MisSense	E474K
YJL212C	*OPT1*	MisSense	I463T
YNL245C	*CWC25*	MisSense	P62L
YDL079C	*MRK1*	Sense	I190I
YLR180W	*SAM1*	Sense	V217V
YNL262W	*POL2*	Sense	F1536F
Construct			
YOL086C	*ADH1*	Sense	A180A
Native SinLoG on chromosome V (IMX605) vs GlycoSwitch (IMX589)			
Genome			
YNL215W	*IES2*	MisSense	E160G
Construct			
YOL086C	*ADH1*	MisSense	R212G
YGR240C	*PFK1*	MisSense	T118A
S.k. SinLoG (IMX637) vs GlycoSwitch (IMX589)			
Genome			
YGL195W	*GCN1*	MisSense	G427C
Construct			
YGR240C like	*SkPYK1*	Sense	A167G
Mosaic SinLoG (IMX645) vs GlycoSwitch (IMX589)			
Genome			
YDR539W	*FDC1*	MisSense	P117S

Table S2 | Identities between the glycolytic genes of *S. cerevisiae* and *S. kudriavzevii/H. sapiens* with respect to the nucleotide sequence (nt) and the coded protein sequences (aa).

Gene		**Nucleotide query**	**Amino acid query**
S. cerevisisae	*S. kudriavzevii*	% identity	% identity
HXK2	*HXK2*	90	96
PGI1	*PGI1*	91	98
PFK1	*PFK1*	89	98
PFK2	*PFK2*	90	98
FBA1	*FBA1*	95	95
TPI1	*TPI1*	95	97
TDH3	*TDH1*	88	89
PGK1	*PGK1*	97	99
GPM1	*GPM1*	96	97
ENO2	*ENO2*	97	98
PYK1	*PYK1*	95	97
PDC1	*PDC1*	95	98
ADH1	*ADH1*	95	96

S. cerevisisae	*H. sapiens*		
TPI1	TP1 (NP_000356.1*)	Codon optimized	53
PGK1	PGK1 (NP_000282.1*)	Codon optimized	66

*accession numbers NCBI

The *S. cerevisiae* CEN.PK 113-7D nucleotide sequences were compared to the *S. kudriavzevii* IFO1802 nucleotide sequences by BLASTN and BLASTX analysis [50].

Table S3 | Strains used in this study

Strain	Relevant genotype
IMX372 (Minimal Glycolysis, MG)	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis, hxk1::KlLEU2 tdh1::KlURA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4*
IMX221	*MATa ura3-52 MAL2-8c SUC2 spr3::(TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*
S. kudriavzevii CR85	*Wild isolate*
IMX370	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4*
IMX377	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1:: (TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*
IMX382	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$)*
IMX457	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$), pyk1::*natNT1
IMX492	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc, pdc, adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1::*NatNT1 *pgi1::*kanMX
IMX493	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$), pyk1::*NatNT1 *pgi1::*kanMX *tpi1::*hphNT1
IMX509	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1::*NatNT1, *pgi1::*kanMX, *tpi1::*hphNT1 pUDC073 *(CEN6/ARS4 ori URA3 GAL1pr-SCEI-CYC1ter)*
IMX510	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1*

Strain	Relevant genotype
IMX511	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2, tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1$_{MP}$) pyk1 pgi1 tpi1 tdh3*::kanMX
IMX535	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3*::kanMX *pfk2::(pTEF-cas9-tCYC1* natNT1*)*
IMX557	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3*::kanMX *pfk2::(pTEF-cas9-tCYC1*natNT1*) pgk1::KlURA3*
IMX561	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1*
IMX566	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1*
IMX568	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFcas9-tCYC1* natNT1*) pgk1 gpm1 fba1*
IMX570	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2*
IMX571	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1*
IMX583	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1*
IMX586	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1*

Strain	Relevant genotype
IMX587	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6, adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*
IMX589 (GlycoSwitch)	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*
IMX606	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2* pUDE325
IMX591	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_{can1H}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ KanMX$_{KL}$ GPM1$_{LM}$ PDC1$_{Mcan1}$)*
IMX607	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(SkFBA1$_{can1H}$ SkTPI1$_{HP}$ SkPGK1$_{PQ}$ SkADH1$_{QN}$ SkPYK1$_{NO}$ SkTDH1$_{OA}$ KanMX$_{AB}$ SkHXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ SkPDC1$_{Mcan1}$)*
IMX633	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MP}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_{can1H}$ pTPI1-HsTPI1-tTPI1$_{HP}$ pPGK1-HsPGK1-tPGK1$_{PQ}$, SkADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ KanMX$_{AB}$ HXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ PDC1$_{Mcan1}$)*
IMX605	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2:: (pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1:: (FBA1$_{can1H}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ KanMX$_{KL}$ GPM1$_{LM}$ PDC1$_{Mcan1}$)* pUDE342
IMX637	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2:: (pTEF-cas9-tCYC1* natNT1*), pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1:: (SkFBA1$_{can1H}$ SkTPI1$_{HP}$ SkPGK1$_{PQ}$ SkADH1$_{QN}$ SkPYK1$_{NO}$ SkTDH1$_{OA}$ KanMX$_{AB}$ SkHXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ SkPDC1$_{Mcan1}$) pUDE342*

Strain	Relevant genotype
IMX645	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2, can1::(FBA1*$_{can1H}$ *pTPI1-HsTPI1-tTPI1*$_{HP}$ *pPGK1-HsPGK1-tPGK1*$_{PQ}$*, SkADH1*$_{QN}$ *PYK1*$_{NO}$ *TDH3*$_{OA}$ *KanMX*$_{AB}$ *HXK2*$_{BC}$ *SkPGI1*$_{CD}$ *SkPFK1*$_{DJ}$ *SkPFK2*$_{JK}$ *SkENO2*$_{KL}$ *SkGPM1*$_{LM}$ *PDC1*$_{Mcan1}$*)* pUDE342
IMX652	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(SkFBA1*$_{can1H}$ *SkTPI1*$_{HP}$ *SkPGK1*$_{PQ}$ *SkADH1*$_{QN}$ *SkPYK1*$_{NO}$ *SkTDH1*$_{OA}$ *KanMX*$_{AB}$ *SkHXK2*$_{BC}$ *SkPGI1*$_{CD}$ *SkPFK1*$_{DJ}$ *SkPFK2*$_{JK}$ *SkENO2*$_{KL}$ *SkGPM1*$_{LM}$ *SkPDC1*$_{Mcan1}$*)* pUDESkTDH1

TABLE S4 | Plasmids used in this study

Plasmid	Characteristic	Source
For construction of KO cassettes:		
pDS1	I-SceI$_{rec}$-*Ag*$_{p}$*TEF2*-kanMX-Ag$_{t}$*TEF2*-I-SceI$_{rec}$	[10]
pDS3	I-SceI$_{rec}$-*KlURA3*-I-SceI$_{rec}$	[10]
pDS6	I-SceI$_{rec}$-*natNT1*-SceI$_{rec}$	[10]
pDS7	I-SceI$_{rec}$-*hphNT1*-I-SceI$_{rec}$	[10]
pUGAmdSYM	*amdSYM*	[7]
pUG6	*kanMX*	[42]
pUGnatNT1	*natNT1*	[42]

Plasmid	Characteristic	Source
For construction of assembly cassettes:		
pSYN-TPI1	Human codon optimized *TPI1* in pMA-T vector	this study
pSYN-PGK1	Human codon optimized *PGK1* in pMK-RQ vector	this study
pUD335	pCR™4BluntTOPO + TagG-*SCEI/KLURA3*-TagF	this study
pUD336	pCR™4BluntTOPO + *PDC1*-SYN_{MF}	this study
pUD331	pCR™4BluntTOPO + *HsPGK1*	this study
pUD329	pCR™4BluntTOPO + *HsTPI1*	this study
For CRISPR/Cas9:		
p414-TEF1p-cas9-CYC1t	*TEF1*p-*cas9-CYC1t*	[12]
p426-SNR52p-gRNA.CAN1.Y-SUP4t	*SNR52*p-gRNA.CAN1.Y-*SUP4*	[12]
P426-GPD	Episomal plasmid	[43]
pUDE324	*SNR52*p-gRNA.*GPM1-SUP4*	this study
pUDE325	*SNR52*p-gRNA.*FBA1-SUP4*t	this study
pUDE326	*SNR52*p-gRNA.*ENO2-SUP4*t	this study
pUDE327	*SNR52*p-gRNA.*HXK2-SUP4*t	this study

Plasmid	Characteristic	Source
pUDE342	*SNR52*p-gRNA.*SGA1-SUP4*t RECYCLE SinLoG	this study
pUDE337	*SNR52p*-gRNA.*AmdSYM-SUP4t*	this study
pUDE329	*SNR52p*-gRNA.*PFK1-SUP4t*	this study
For Marker recycling		
pUDC073	*pGAL1-SCEI-tCYC1 CEN6/ARS4 URA3*	[8]
PUDE206	*TPI1*p-*SCEI-TEF1*, episomal, *AgTEF2-hphNT1-CYC1*t	this study
For overexpression of *SkTDH1*		
pUDESkTDH1	*SkTDH1*, episomal, *URA3*	this study

TABLE S5 | Primers used to construct plasmids.

Plasmid	Name	Sequence 5'--> 3'
pUDE327	p426 Fw + V	GCATCTCTATAACTGGTGTCGCTGAACTACCATGTACTGCCCATGCGGCAAAT GAATCCAGCATCTGTGCGGTATTTCACACC
	p426 Rv + U	GGAATCTGTGTAGTATGCCTTCATTTGAGTTCTGGAGGACCACACACT GCGGGCTAATGACTCAAAGGCGGTAATACGGTTATCC
	U Fw	TCATTAGCCCGCAGTGTGTGGTCC
	V Rv	TGGATTCATTTGCCGCATGGGC
pUDE325	p426 Fw + U	TCATTAGCCCGCAGTGTGTGGTCCTCCAGAACTCAAATGAAGGCAT ACTACACAGATTCCGCATCTGTGCGGTATTTCACACC
	p426 Rv + S	CACAAGCTGAGGCTGCCATAGATTCTCCGAACTTAGTCTCATCGAGGGCTCT GACGCAATCTCAAAGGCGGTAATACGGTTATCC
	S Fw	ATTGCGTCAGAGCCCTCGATGAGAC
	U Rv	GGAATCTGTGTAGTATGCCTTCATTTG
pUDE326	p426 Fw + X	TCTGGCAGTCCATTGGCATGCCAGCCCTGCGATTATTTGTTCATACCGGCC AGTAGGATGGCATCTGTGCGGTATTTCACACC
	p426 Rv + W	CTAGTTAGGGAAGTGTCCTTTCCATGTGTTCTGTCGGGCACGGAATTAAC ACTGCTTCGACTCAAAGGCGGTAATACGGTTATCC
	W Fw	TCGAAGCAGTGTTAATTCCGTGC
	X Rv	CATCCTACTGGCCGGTATGAAC
pUDE324	p426 Fw + W	TCGAAGCAGTGTTAATTCCGTGCCCGACAGAACACATGGAAAGGACACTT CCCTAACTAGGCATCTGTGCGGTATTTCACACC
	p426 Rv + V	TGGATTCATTTGCCGCATGGGCAGTACATGGTAGTTCAGCGACACCAGTTA TAGAGATGCCTCAAAGGCGGTAATACGGTTATCC
	V Fw	GCATCTCTATAACTGGTGTCGCTGAAC
	W Rv	CTAGTTAGGGAAGTGTCCTTTCCATG

Plasmid	Name	Sequence 5'--> 3'
pUDE329	p426 Fw + S	ATTGCGTCAGAGCCCTCGATGAGACTAAGTTCGGAGAATCTATGGCAGC CTCAGCTTGTGGCATCTGTGCGGTATTTCACACC
	p426 Rv + R	CGTTATATGAATGGCTGGCCACTCAGCGTCCGTAGAATCGCAACACGATG CAGGGTCGATCTCAAAGGCGGTAATACGGTTATCC
	R Fw	ATCGACCCTGCATCGTGTTG
	S rv	CACAAGCTGAGGCTGCCATAG
p426 crispr-backbone	p426-crispr Fw	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
	p426-crispr Rv	GATCATTTATCTTTCACTGCGGAGAAG
crADH1	crADH1 Fw	GAGTTAGCATATCTACAATTGGGTGAAATGGGGAGCGATTTGCAGGC ATTTGCTCGGCATGCTCTATTGTTCGCACCACCGGCAAACTCGCGTCT CGCAAGTCTTGGCTCATTCTTCTAG
	crADH1 Rv	CTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGT GCGAACAATAGAGCATGCCGAGCAAATGCCTGCAAATCGCTCCCCAT TTCACCCAATTGTAGATATGCTAACTC
crPDC1	crPDC1 Fw	ATCGAGGTGTCTAGTCTTCTATTACGCTAATGCAGTTTCAGGGTTTT GGAAACCACACTGTCAAGTTGAAGACTATATATTTTATTGAGTTTAT GTTATGGGGAGGCTACCCTTTACGTC
	crPDC1 Rv	GACGTAAAGGGTAGCCTCCCCATAACATAAACTCAATAAAATATATAG TCTTCAACTTGACAGTGTGGTTTCCAAAACCCTGAAACTGCATTAGC GTAATAGAAGACTAGACACCTCGAT
crAmdSYM	crAmdSYM Fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAAT GATCTGGTTGAACAAGTACGACGAGTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCCGTTATCAAC
	crAmdSYM Rv	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAA ACTCGTCGTACTTGTTCAACCAGATCATTTATCTTTCACTGCGGAGAA GTTTCGAACGCCGAAACATGCGCA
crRECYCLE	crRECYCLE Fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAAT GATCTTACAATATAGTGATAATCGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAAC
	crRECYCLE Rv	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAA ACCGATTATCACTATATTGTAAGATCATTTATCTTTCACTGCGGAGAA GTTTCGAACGCCGAAACATGCGCA

Plasmid	Name	Sequence 5'--> 3'
pUD329	pTPI1 Fw	TAGTGTGAGCGGGATTTAAACTGTG
	pTPI1 Rv + link	TCCAGTTACCACCGACGAAGAACTTTCTAGATGGAGCCATTTTTAGT TTATGTATGTGTTTTTTGTAGTTATAGATTTAAGC
	tTPI1 Fw + link	GAAGCCAGAATTCGTCGACATCATTAACGCTAAGCAATAAGATTAA TATAATTATATAAAAATATTATCTTCTTTTCTTTATATCTAGTG
	tTPI1 Rv	GCGAAAATGACGCTTGCAGTG
	TPI1 Fw	ATGGCTCCATCTAGAAAGTTCTTCG
	TPI1 Rv	TTATTGCTTAGCGTTAATGATGTCG
pUD331	pPGK1 Fw	CCTGCATTTAAAGATGCCGATTTGG
	pPGK1 Rv + link	CGTCCAACTTGTCCAAAGTCAACTTGTTAGACAAAGACATTGTTT TATATTTGTTGTAAAAAGTAGATAATTACTTC
	tPGK1 Fw + link	TAAGGTCTTGCCAGGTGTCGACGCTTTGTCTAACATTTAAATTGAA TTGAATTGAAATCGATAGATCAATTTTTTTC
	tPGK1 Rv	ATTTTAGCGTAAAGGATGGG
	PGK1 Fw	TAAATGTTAGACAAAGCGTCGACACC
	PGK1 Rv	ATGTCTTTGTCTAACAAGTTGACTTTGG
p426SkTDH1 backbone	p426 rv + O	ATACTCCCTGCACAGATGAGTCAAGCTATTGAACACCGAGAACGCG CTGAACGATCATTCCTCAAAGGCGGTAATACGGTTATCC
	p426 fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCCAG ATCATCAATAGGCACGCATCTGTGCGGTATTTCACACC

TABLE S6 | Primers used to construct the glycolytic gene cassettes and the *SCEI* cassette.

Product	Name	Sequence 5'--> 3'
SCEI cassette	Syn1 Fw	ACATTTACAATATAGTGATAATCGTGGACTAGAGCAAGATTTCAAATAAGTAA CAGCAGCAGTTTCAGCTTTCCGCAACAGTATAATTTCAAAACGTCGTAC
	Syn1 Rv	AAGGGCCATGACCACCTGATGCACCAATTAGGTAGGTCTGGCTATGTCT ATACCTCTGGCTCAAAACGTCGTACGACGTTTTGAAATTATACTGTTGC GGAAAGCTGAAACTGCTGC
	Tag G Fw	GCCAGAGGTATAGACATAGCCAGAC
	SGA1 Rv	TCTACAAACTCTGTAAAACTTCTTGTCTTATTTGATAGGCATCCCA GAATGAAGTATAGGGCCGAACTTTCCCTGTATGAAGC
	FUS1 Fw	TTACAATATAGTGATAATCGTGGACTAGAG
	FUS1 Rv	CAAACTCTGTAAAACTTCTTGTCTTATTTG
Assembly cassettes		
Marker cassettes		
$AmdSYM_{AB}$	AmdSYM Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCCAGATCAT CAATAGGCACGCGACATGGAGGCCCAGAATACC
	AmdSYM Rv + B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGCATCGTCC TCTCGAAAGGTGAGTATAGCGACCAGCATTCACATACG
$AmdSYM_{KL}$	AmdSYM Fw + K	AAGATAGTCGCCGAACTCGCAAGAGTCATTAACACCTCGCAATTGAT GGGAAGTCCTCGCGCGACATGGAGGCCCAGAATACC
	AmdSYM Rv + L	GCCGTAGCTTCCGCAAGTATGCCGTAGTTGAAGAGCATTTGCCGT CGGTTCAGGTCATATAGTATAGCGACCAGCATTCACATACG
$kanMX_{AB}$	kanMX Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCCAGATCAT CAATAGGCACGCGACATGGAGGCCCAGAATACC
	kanMX Rv + B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGCATCGTC CTCTCGAAAGGTGAGTATAGCGACCAGCATTCACATACG
$kanMX_{KL}$	kanMX Fw + K	AAGATAGTCGCCGAACTCGCAAGAGTCATTAACACCTCGCAATTGATG GGAAGTCCTCGCGCGACATGGAGGCCCAGAATACC
	kanMX Rv + L	GCCGTAGCTTCCGCAAGTATGCCGTAGTTGAAGAGCATTTGCCGTCGGTT CAGGTCATATAGTATAGCGACCAGCATTCACATACG

Product	Name	Sequence 5'--> 3'
Saccharomyces cerevisiae cassettes:		
$TPI1_{HP}$	TPI Fw + P	CTGATAGTGCTGTAAGTCGCCTCCATCTTAGCAGAGCTGTCCCTGAATGCG TACTCGTGAGCGAAAATGACGCTTGCAGTG
	TPI Rv + H	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGAATTGCTGAG AACCCGTGACTAGTGTGAGCGGGATTTAAACTGTG
$PGK1_{QP}$	PGK1 Fw + Q	GAGCTGAATGTATATGCTGCGGGATCATTGCACAGCTCTGAGAGCCCT GCAACGCGATATCCTGCATTTAAAGATGCCGATTTGG
	PGK1 Rv + P	TCACGAGTACGCATTCAGGGACAGCTCTGCTAAGATGGAGGCGACTTA CAGCACTATCAGATTTTAGCGTAAAGGATGGGGAAAGAG
$PYK1_{NO}$	PYK1 Fw + N	GATCAGCAGCCACGATTGAGTCCTAACGAAGATATGTGGACCTTGCAT CAAAGCCTAGAAAAATAGCCGCCATGACCTCG
	PYK1 Rv + O	ATACTCCCTGCACAGATGAGTCAAGCTATTGAACACCGAGAACGCGCT GAACGATCATTCTGCATTTATGTACCCATGTATAACCTTCC
$ADH1_{QN}$	ADH1 Fw + Q	ATATCGCGTTGCAGGGCTCTCAGAGCTGTGCAATGATCCCGCAGCATATA CATTCAGCTCGCCGGTAGAGGTGTGGTCAATAAG
	ADH1 Rv + N	TTCTAGGCTTTGATGCAAGGTCCACATATCTTCGTTAGGACTCAATCGT GGCTGCTGATCGAGGAAACAGCAATAGGGTTGCTAC
$TDH3_{OA}$	TDH3 Fw + O	GAATGATCGTTCAGCGCGTTCTCGGTGTTCAATAGCTTGACTCATCTGTGC AGGGAGTATAAATTTCACTCAGCATCCACAATGTATCAG
	TDH3 Rv + A	GTGCCTATTGATGATCTGGCGGAATGTCTGCCGTGCCATAGCCATGCCTTC ACATATAGTGAATACGTAAATAATTAATAGTAGTGATTTTCCTAAC
$HXK2_{BC}$	HXK2 Fw + B	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACCAGCCTAAG AATGTTCAACGACGGCACCGGGAAATAAACC
	HXK2 Rv + C	CTAGCGTGTCCTCGCATAGTTCTTAGATTGTCGCTACGGCATATACGAT CCGTGAGACGTGCAAGAGAAAAAAACGAGCAATTGTTAAAAG
$PGI1_{CD}$	PGI1 Fw + C	ACGTCTCACGGATCGTATATGCCGTAGCGACAATCTAAGAACTATGCG AGGACACGCTAGTTCGCGACACAATAAAGTCTTCACG
	PGI1 Rv + D	AATCACTCTCCATACAGGGTTTCATACATTTCTCCACGGGACCCACA GTCGTAGATGCGTCTGAAGAAGGCATACTACGCCAAG
$FBA1_{GH}$	FBA1 Fw + G	GCCAGAGGTATAGACATAGCCAGACCTACCTAATTGGTGCATCAGGTG GTCATGGCCCTTAGTGCATGACAAAAGATGAGCTAGG
	FBA1 Rv + H	GTCACGGGTTCTCAGCAATTCGAGCTATTACCGATGATGGCTGAGGCGT TAGAGTAATCTAAAATCTCAAAAATGTGTGGGTCATTACG

Product	Name	Sequence 5'--> 3'
$FBA1_{can1H}$	FBA1 Fw + can1	GGTGTATGACTTATGAGGGTGAGAATGCGAAATGGCGTGGGAATGTGAT TAAAGGTAATAAGTGCATGACAAAAGATGAGCTAGG
	FBA1 Rv + H	GTCACGGGTTCTCAGCAATTCGAGCTATTACCGATGATGGCTGAGG CGTTAGAGTAATCTAAAATCTCAAAAATGTGTGGGTCATTACG
$PFK1_{DJ}$	PFK1 Fw + J	CGACGAGATGCTCAGACTATGTGTTCTACCTGCTTGGACATCTTCG CGTATATGACGGCCTGTCGTCTTCGTGAACCATTGTC
	PFK1 Rv + D	ACGCATCTACGACTGTGGGTCCCGTGGAGAAATGTATGAAACCCT GTATGGAGAGTGATTTCGAGATTCCTCAATCCATACACCATTATAG
$PFK2_{JK}$	PFK2 Fw + J	GGCCGTCATATACGCGAAGATGTCCAAGCAGGTAGAACACATAGTCTGAGC ATCTCGTCGGAGATCCGAGGGACGTTTATTGG
	PFK2 Rv + K	GCGAGGACTTCCCATCAATTGCGAGGTGTTAATGACTCTTGCGAGTTC GGCGACTATCTTATAGCCATTCTCTGCTGCTTTGTTG
$ENO2_{KL}$	ENO2 Fw + K	AAGATAGTCGCCGAACTCGCAAGAGTCATTAACACCTCGCAATT GATGGGAAGTCCTCGCGAAGCCCACTTTCGTGGACTTTG
	ENO2 Rv + L	GCCGTAGCTTCCGCAAGTATGCCGTAGTTGAAGAGCATTTGCC GTCGGTTCAGGTCATATCCTTCCAGTGCATTATGCAATAGACAG
$ENO2\text{-}LONG_{AB}$	ENO2 Fw + A	ACTATATGTGAAGGCATGGCTATGGCACGGCAGACATTCCGCCAGAT CATCAATAGGCACAAGTGCTACAGAAATCCTACTCTTGCC
	ENO2 Rv + B	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGCATCGTC CTCTCGAAAGGTGCCAGCTGATTGAAGGTTCTCAAAGTGAC
$GPM1_{LM}$	GPM1 Fw + L	ATATGACCTGAACCGACGGCAAATGCTCTTCAACTACGGCATACTTGCG GAAGCTACGGCCATGTCATGTCACCATTAATTACCACTC
	GPM1 Rv + M	ACGAGAGATGAAGGCTCACCGATGGACTTAGTATGATGCCATGCTGG AAGCTCCGGTCATGGTATATTTCTTAATGTGGAAAGATACTAGCG
$PDC1_{Mcan1}$	PDC1 Fw + can1	GATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCGACG AGAGTAAATGGCGAGTTTAAACAGTGTTCCTTAATCAAGGATAC
	PDC1 Rv + M	ATGACCGGAGCTTCCAGCATGGCATCATACTAAGTCCATCGGTGAGC CTTCATCTCTCGTGCCGAAATGCATGCAAGTAACC

Product	Name	Sequence 5'--> 3'
PDC1-SYN$_{MF}$	PDC1 Fw + RES	CGATTTCTTGTGTAACAGAAGTTTCAGCTTTCCGCAACAGTATAATTTCAAA ACGTCGTACGACGTTTTGATTTAAACAGTGTTCCTTAATCAAGGATAC
	PDC1 Rv + M	ATGACCGGAGCTTCCAGCATGGCATCATACTAAGTCCATCGGTGAGCCTT CATCTCTCGTGCCGAAATGCATGCAAGTAACC
	Syn2 Fw	TGCCGAACTTTCCCTGTATGAAGCGATCTGACCAATCCTTTGCCGTAGTT TCAACGTATGGCTGCTGTTACTTATTTGAAATCTTGCTCTAGTCCACGAT TATCACTATATTGTAAATG
	Syn2 Rv	TCAAAACGTCGTACGACGTTTTGAAATTATACTGTTGCGGAAAGCTG AAACTTCTGTTACACAAGAAATCGTACATTTACAATATAGTGATAATCGT GGACTAGAGCAAGATTTC
	FUS2 Fw	ATGACCGGAGCTTCCAGCATG
	FUS2 Rv	TGCCGAACTTTCCCTGTATGAAGC
Saccharomyces Kudriavzevii cassettes:		
skTPI1$_{HP}$	skTPI Fw + P	CTGATAGTGCTGTAAGTCGCCTCCATCTTAGCAGAGCTGTCCCTGAATGCG TACTCGTGACAAATCCCAATTTTTCACGGACGGTAATC
	skTPI Rv + H	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGAATTGCTGA GAACCCGTGACGACAAGAGAGAAGACCCAGGGATG
skPGK1$_{QP}$	skPGK1 Fw + Q	GAGCTGAATGTATATGCTGCGGGATCATTGCACAGCTCTGAGAGCCCTG CAACGCGATATGGATCTTAGCTTCAACTCAAGATGTACAG
	skPGK1 Rv + P	TCACGAGTACGCATTCAGGGACAGCTCTGCTAAGATGGAGGCGACTT ACAGCACTATCAGGCCTAAATAAATGAAGTAAATGCGAGGTAAGC
skPYK1$_{NO}$	skPYK1 Fw + O	ATACTCCCTGCACAGATGAGTCAAGCTATTGAACACCGAGAACGCGCTG AACGATCATTCTAGCATAAGATGCTACATCTTAGGATTCTG
	skPYK1 Rv + N	GATCAGCAGCCACGATTGAGTCCTAACGAAGATATGTGGACCTTGCAT CAAAGCCTAGAACCTTATGTTATGGATATTCTTCTTCTTGCG
skADH1$_{QN}$	skADH1 Fw + Q	ATATCGCGTTGCAGGGCTCTCAGAGCTGTGCAATGATCCCGCAGCA TATACATTCAGCTCAACTCGTTGCTGGAGCTAGCATAC
	skADH1 Rv + N	TTCTAGGCTTTGATGCAAGGTCCACATATCTTCGTTAGGACTCAAT CGTGGCTGCTGATCGCATAACCGGTAGAGTACTTTGGAGTC
skTDH3$_{OA}$	skTDH1 Fw + O	GAATGATCGTTCAGCGCGTTCTCGGTGTTCAATAGCTTGACT CATCTGTGCAGGGAGTATTGGAAAAGAGGATAGGAAGGAGGAGAAG
	skTDH1 Rv + A	GTGCCTATTGATGATCTGGCGGAATGTCTGCCGTGCCATAGCCATGC CTTCACATATAGTTAAAGCACATTTAACCTTTCTCGCTACC

Product	Name	Sequence 5'--> 3'
$skHXK2_{BC}$	skHXK2 Fw + B	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACCAGCC TAAGAATGTTCAACGACCTTAGTCCATTGACGTCTGTATTTG
	skHXK2 Rv + C	CTAGCGTGTCCTCGCATAGTTCTTAGATTGTCGCTACGGCATATACG ATCCGTGAGACGTGAGGTCAATCATACACCGGAAGAAAG
$skPGI1_{CD}$	skPGI1 Fw + C	ACGTCTCACGGATCGTATATGCCGTAGCGACAATCTAAGAACTATG CGAGGACACGCTAGTGTTCCAAGACACCAAGAATGTCATAC
	skPGI1 Rv + D	AATCACTCTCCATACAGGGTTTCATACATTTCTCCACGGGACCCAC AGTCGTAGATGCGTGCTTGATAATCAAAGCAGCGCACAG
$skFBA1_{can1H}$	skFBA1 Fw + can1	GGTGTATGACTTATGAGGGTGAGAATGCGAAATGGCGTGGGAAT GTGATTAAAGGTAATAATGCCGACACGCGTTATGCAAAG
	skFBA1 Rv + H	GTCACGGGTTCTCAGCAATTCGAGCTATTACCGATGATGGCT GAGGCGTTAGAGTAATCTACGGCTTGAACAACAATGCCAACC
$skPFK1_{DJ}$	skPFK1 Fw + J	CGACGAGATGCTCAGACTATGTGTTCTACCTGCTTGGACATCTTCG CGTATATGACGGCCTTCATTGCTCATTGTTATGTGTATCATATCG
	skPFK1 Rv + D	ACGCATCTACGACTGTGGGTCCCGTGGAGAAATGTATGAAACCCTG TATGGAGAGTGATTCCTTTATATTTTATGACACCATCTTCCGTACAC
$skPFK2_{JK}$	skPFK2 Fw + J	GGCCGTCATATACGCGAAGATGTCCAAGCAGGTAGAACACATAGTCT GAGCATCTCGTCGGATTCGAAGGACGTTTATTGGGAATATC
	skPFK2 Rv + K	GCGAGGACTTCCCATCAATTGCGAGGTGTTAATGACTCTTGCGAGTT CGGCGACTATCTTCTTCGAATGCACGGCAATAATGATACG
$skENO2_{KL}$	skENO2 Fw + K	AAGATAGTCGCCGAACTCGCAAGAGTCATTAACACCTCGCAATTGAT GGGAAGTCCTCGCGTCATCTGGATCCCATACTTTACGAGAAAC
	skENO2 Rv + L	GCCGTAGCTTCCGCAAGTATGCCGTAGTTGAAGAGCATTTGCCGT CGGTTCAGGTCATATACTTCAGGAAGTCCCGCCGTGTG
$skGPM1_{LM}$	skGPM1 Fw + L	ATATGACCTGAACCGACGGCAAATGCTCTTCAACTACGGCATACTT GCGGAAGCTACGGCGCAAGGAGTCCCAGGCCTTAATTTTC
	skGPM1 Rv + M	ACGAGAGATGAAGGCTCACCGATGGACTTAGTATGATGCCATGCT GGAAGCTCCGGTCATGGACACTTTAACTGGGGCCATATC
$skPDC1_{Mcan1}$	skPDC1 Fw + can1	GATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCG ACGAGAGTAAATGGCGAGTGCAGCTATCAGGTTTTGCTTTACAATTG
	skPDC1 Rv + M	ATGACCGGAGCTTCCAGCATGGCATCATACTAAGTCCATCGGTG AGCCTTCATCTCTCGTTCCAGACGGAAAAACCGCACGAG

Product	Name	Sequence 5'--> 3'
Homo sapiens cassettes:		
$hsTPI1_{HP}$	hsTPI Fw + H	AGATTACTCTAACGCCTCAGCCATCATCGGTAATAGCTCGAATTGCTG AGAACCCGTGACTAGTGTGAGCGGGATTTAAACTGTG
	hsTPI Rv + P	CTGATAGTGCTGTAAGTCGCCTCCATCTTAGCAGAGCTGTCCCTGAAT GCGTACTCGTGAGCGAAAATGACGCTTGCAGTG
$hsPGK1_{QP}$	hsPGK1 Fw + P	TCACGAGTACGCATTCAGGGACAGCTCTGCTAAGATGGAGGCG ACTTACAGCACTATCAGATTTTAGCGTAAAGGATGGGG
	hsPGK1 Rv + Q	GAGCTGAATGTATATGCTGCGGGATCATTGCACAGCTCTGAG AGCCCTGCAACGCGATATCCTGCATTTAAAGATGCCGATTTGG

TABLE S7 | Primers used to construct the deletion cassettes.

Product	Name	Sequence 5'--> 3'
Deletion cassettes		
PYK1 KO	PYK1 KO Fw	TTTCCCCCCTTATTTTTTTTTGTTAGAATTGATCCAAATGTAAAT AAACAATCACAACGGTCGTAATTGAGTCAGTTACGCTAGG
	PYK1 KO Rv	GATTAAACCACCAAACGAAGGCCAGAAGCTGAACATAGTTCACTG GCATCCGTTGTGATTGTTTATTTACATTTGGATCAATTCTAACAAC GTGCGTAGAATGAAGAACCTATATTACC
PGI1 KO	PGI KO Fw	ACAGTTGATGAGAACCTTTTTCGCAAGTTCAAGGTGCTCTAATTTTT AAAATTTTTACTTGTCGTAATTGAGTCAGTTACGCTAGG
	PGI KO Rv	TTAACTTACTTAGAATAATGCCATTTTTTTGAGTTATAATAATCCTACGTTTT GAGAAGATGTTCTTATTCAAATTTCAACTGTTATATACGTGCGTAGAATGAA GAACCTATATTACC
TPI1 KO	TPI KO Fw	TTGGCAATTTTTTGCTCTTCTATATAACAGTTGAAATTTGAATAAGAACAT CTTCTCAAAGTCGTAATTGAGTCAGTTACGCTAGG
	TPI KO Rv	TTAACTTACTTAGAATAATGCCATTTTTTTGAGTTATAATAATCCTACGTTTT GAGAAGATGTTCTTATTCAAATTTCAACTGTTATATACGTGCGTAGAATGA AGAACCTATATTACC

Product	Name	Sequence 5'--> 3'
TDH3 KO	TDH3 KO Fw	AACTTTAAAAAAAAAAGCCAATATCCCCAAAATTATTAAGAGCGCCTCC ATTATTAACTAGTCGTAATTGAGTCAGTTACGCTAGG
	TDH3 KO Rv	AACACGCTTTTTCAGTTCGAGTTTATCATTATCAATACTGCCATTTCAA ATAGTTAATAATGGAGGCGCTCTTAATAATTTTGGGGATATCGTGCGTAGAAT GAAGAACCTATATTACC
PGK1 KO	PGK1 KO Fw	CATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGGAAAGAGTGAGG AACTATCGCATAGTCGTAATTGAGTCAGTTACGCTAGG
	PGK1 KO Rv	TATATATACGTATATAAATAAAAAATATTCAAAAAATAAAATAAACTATT TATGCGATAGTTCCTCACTCTTTCCTTACTCACGAGTAATCGTGCGTA GAATGAAGAACCTATATTACC
pfk2::CAS9/natNT1		
CAS9/nat	CAS9 Fw + pfk2	TGTATAAATATTGTATTAAAAGGGTACCTTTATAAATATGAGATCCGAGG CATAGCTTCAAAATGTTTCTACTCCTTTTTTACTC
	CAS9 Rv + link	CGAAGTTATATTAAGGGTTGTCGACCTGCAGCGTACGAAGCTTCAGC TGGGTACCGGCCGCAAATTAAAGCCTTCG
	nat Fw + link	TGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCC GGTACCCAGCTGAAGCTTCGTACGCTGCAG
	nat Rv + Rpt +pfk2	CAGCATTGTAAGTACTACCATCTATATAGATCGAATATCCTACTAATAT ATGTGCGTTAAATACTTTGTTCTCTTATTATCTATGAACTTGCATAGGC CACTAGTGGATCTGATATCAC
Repair fragments		
GPM1	GPM1 repair Fw	CACATGCAGTGATGCACGCGCGATGGTGCTAAGTTACATATATATAT ATATATATATATATTATAGCAATAACCTTGATGTTTACATCGTAGTTT AATGTACACCCCGCGAATTCGTTCA
	GPM1 repair Rv	TGAACGAATTCGCGGGGTGTACATTAAACTACGATGTAAACATCAAGG TTATTGCTATAATATATATATATATATATATATATGTAACTTAGCACCAT CGCGCGTGCATCACTGCATGTG
HXK2	HXK2 repair Fw	TTTCTAATGCCTTTTCCATCATGTTACTACGAGTTTTCTGAACCTCCT CGCACATTGGTAGCTTAATTTTAAATTTTTTTGGTAGTAAAAGATGC TTATATAAGGATTTCGTATTTATTG
	HXK2 repair Rv	CAATAAATACGAAATCCTTATATAAGCATCTTTTACTACCAAAAAAATTT AAAATTAAGCTACCAATGTGCGAGGAGGTTCAGAAAACTCGTAGTAA CATGATGGAAAAGGCATTAGAAA

Product	Name	Sequence 5'--> 3'
PFK1	PFK1 repair Fw	AGGCCGACAAATAAACCAAACGGTATTCGTAGACCGATGACAATACGA CTACAATTAAGGCATGTTTTTCCATCGTTTTCAACGATGACTGTAA CCCGTAGATTGAACCAGGCATGCCAA
	PFK1 repair Rv	TTGGCATGCCTGGTTCAATCTACGGGTTACAGTCATCGTTGAAAA CGATGGAAAAACATGCCTTAATTGTAGTCGTATTGTCATCGGTCTA CGAATACCGTTTGGTTTATTTGTCGGCCT
FBA1	FBA1 repair Fw	ACTCCAAAATGAGCTATCAAAAACGATAGATCGATTAGGATGACTTT GAAATGACTCCGCAACTATTACGTATTACGATAATCCTGCTGTCATTA TCATTATTATCTATATCGACGTAT
	FBA1 repair Rv	ATACGTCGATATAGATAATAATGATAATGACAGCAGGATTATCGTAA TACGTAATAGTTGCGGAGTCATTTCAAAGTCATCCTAATCGATCTAT CGTTTTTGATAGCTCATTTTGGAGT
ADH1	ADH1 repair Fw	GAGTTAGCATATCTACAATTGGGTGAAATGGGGAGCGATTTGCAGGC ATTTGCTCGGCATGCTCTATTGTTCGCACCACCGGCAAACTCGCGTCT CGCAAGTCTTGGCTCATTCTTCTAG
	ADH1 repair Rv	CTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGTG CGAACAATAGAGCATGCCGAGCAAATGCCTGCAAATCGCTCCCCATTT CACCCAATTGTAGATATGCTAACTC
PDC1	PDC1 repair Fw	ATCGAGGTGTCTAGTCTTCTATTACGCTAATGCAGTTTCAGGGTTTTGG AAACCACACTGTCAAGTTGAAGACTATATATTTTATTGAGTTTATGTTAT GGGGAGGCTACCCTTTACGTC
	PDC1 repair Rv	GACGTAAAGGGTAGCCTCCCCATAACATAAACTCAATAAAATATATAGT CTTCAACTTGACAGTGTGGTTTCCAAAACCCTGAAACTGCATTAGCGT AATAGAAGACTAGACACCTCGAT
ENO2-A	ENO2-A repair Fw	CCAAAACTGGCATCCACTAATTGATACATCTACACACCGCACGCCTTTTTT CTGAAGCCCGGAAAAAAAAAGGTGCACACGCGTGGCTTTTTCTTGAATTT GCAGTTTGAAAAATAACTAC
	ENO2-A repair Rv	GTAGTTATTTTTCAAACTGCAAATTCAAGAAAAAGCCACGCGTGTGCACCT TTTTTTTCCGGGCTTCAGAAAAAAGGCGTGCGGTGTGTAGATGTATCAAT TAGTGGATGCCAGTTTTGG
ENO2-B	ENO2-B repair Fw	ATTTAGGTTTAAAAATTGATACAGTTTTATAAGTTACTTTTTCAAAGACTC GTGCTGTCTCACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACC AGCCTAAGAATGTTCAAC
	ENO2-B repair Rv	GTTGAACATTCTTAGGCTGGTCGAATCATTTAGACACGGGCATCGTCCTCT CGAAAGGTGAGACAGCACGAGTCTTTGAAAAAGTAACTTATAAAACTGTAT CAATTTTTAAACCTAAAT
RECYCLE	RECYCLE Fw	TTTTTCTCATCTCTTGGCTCTGGATCCGTTATCTGTTCTGTTACACAAGAAAT CGTACATACTAGAGCAAGATTTCAAATAAGTAACAGCAGCCATACGTTGAAA CTACGGCAAAGGATT
	RECYCLE Rv	AATCCTTTGCCGTAGTTTCAACGTATGGCTGCTGTTACTTATTTGAAATCTT GCTCTAGTATGTACGATTTCTTGTGTAACAGAACAGATAACGGATCCAGAG CCAAGAGATGAGAAAAA

TABLE S8 | Primers used to check the integration of the endogenous SinLoG in chromosome IX.

Amplicon	Primers	Sequence 5'--> 3'
G	G Fw	CTTGGCTCTGGATCCGTTATCTG
	G Rv	GCTCTTTTCTTCTGAAGGTCAATG
H	H Fw	GTTACGTGCTCAGTTGTTAGATATG
	H Rv	GCAGAAGTGTCTGAATGTATTAAGG
P	P Fw	TGAGCCACTTAAATTTCGTGAATG
	P Rv	TTTCTCTTTCCCCATCCTTTACG
Q	Q Fw	GCCCAAATCGGCATCTTTAAATG
	Q Rv	GTCAGGTTGCTTTCTCAGGTATAG
N	N Fw	AGTGTTGTATGTACCTGTCTATTTATACTG
	N Rv	GTCATGGCGGCTATTTTTCTAGG
O	O Fw	TTCCCAAGAACTAACTTGGAAGG
	O Rv	CTCTACCAGAGTTGTCGACTTG
A	A Fw	CCAGGCAGGTTGCATCACTC
	A Rv	CGCACGTCAAGACTGTCAAG
B	B Fw	TCGTATGTGAATGCTGGTCG
	B Rv	ACGGAATAGAACACGATATTTGC
C	C Fw	TCACGGGATTTATTCGTGACG
	C Rv	GCGTCCAAGTAACTACATTATGTG
D	D Fw	ACTCGCCTCTAACCCCACG
	D Rv	ACGGACTATAATGGTGTATGGATTG
J	J Fw	GCTTAATCTGCGTTGACAATGG
	J Rv	CAATAAACGTCCCTCGGATCTC
K	K Fw	GACGCCATTTGGAACGAAAAAAAG
	K Rv	TATGCTGACTTGGTATCACACTTC
L	L Fw	CAAAGACTCGTGCTGTCTATTGC
	L Rv	AATGAGTGGTAATTAATGGTGACATGAC
M	M Fw	ACGGAAAGTGGAATCCCATTTAG
	M Rv	ACCCTCATGAAACATGTATGAGATATTAC
F	F Fw	AAGCTAAGTTGACTGCTGCTACC
	F Rv	TTGGGCTGGACGTTCCGACATAG

Amplicon	Primers	Sequence 5'--> 3'
Sequencing fragment:		Primers used to obtain fragment:
1	A Fw	CCAGGCAGGTTGCATCACTC
	C Rv	GCGTCCAAGTAACTACATTATGTG
2	N Fw	AGTGTTGTATGTACCTGTCTATTTATACTG
	A Rv	CGCACGTCAAGACTGTCAAG
3	L Fw	CAAAGACTCGTGCTGTCTATTGC
	SGA1 2 Rv	TGGTCGACAGATACAATCCTGG
4	D 2 Fw	ACGCTGGCACAACATAGTTC
	J Rv	CAATAAACGTCCCTCGGATCTC
5	P Fw	TGAGCCACTTAAATTTCGTGAATG
	N Rv	GTCATGGCGGCTATTTTTCTAGG
6	O Fw	TTCCCAAGAACTAACTTGGAAGG
	B Rv	ACGGAATAGAACACGATATTTGC
7	K Fw	TGTCTTACCCTGGACGGTATC
	M Rv	ACCCTCATGAAACATGTATGAGATATTAC
8	H Fw	GTTACGTGCTCAGTTGTTAGATATG
	Q Rv	GTCAGGTTGCTTTCTCAGGTATAG
9	M Fw	ACGGAAAGTGGAATCCCATTTAG
	SGA1 2 Rv	TGGTCGACAGATACAATCCTGG
10	C 2 Fw	TCACGGGATTTATTCGTGACG
	J Rv	CAATAAACGTCCCTCGGATCTC
11	Q Fw	GCCCAAATCGGCATCTTTAAATG
	O Rv	CTCTACCAGAGTTGTCGACTTG
12	J Fw	GCTTAATCTGCGTTGACAATGG
	L Rv	AATGAGTGGTAATTAATGGTGACATGAC
13	SGA1 2 Fw	ACTCGTACAAGGTGCTTTTAACTTG
	P Rv	TTTCTCTTTCCCCATCCTTTACG
14	B Fw	TCGTATGTGAATGCTGGTCG
	D 2 Rv	AATCATGTTGATGACGACAATGG

Table S9 | Primers used to check the integration of the SinLoGs in can1.

S. cerevisiae SinLoG in *CAN1*:

Amplicon	primers	Sequence 5'--> 3'
1	CAN1 Fw	TCGGGAGCAAGATTGTTGTG
	SC1 Rv	TTTCTCTTTCCCCATCCTTTACG
2	SC2 Fw	GCCCAAATCGGCATCTTTAAATG
	SC2 Rv	GTAGTTATTTTTCAAACTGCAAATTCAAG
3	SC3 Fw	GGTGCACACGCGTGGCTTTTTCTTGAATTTGC
	SC3 Rv	AATCATGTTGATGACGACAATGG
4	SC4 Fw	GCTTAATCTGCGTTGACAATGG
	SC4 Rv	AAACTCACCGAGGCAGTTCCATAGG
5	SC5 Fw	TCGTATGTGAATGCTGGTCG
	CAN1 Rv	AGAAGAGTGGTTGCGAACAGAG

S. kudriavzevii & Mosaic SinLoGs in *CAN1:*

Amplicon	primers	Sequence 5'--> 3'
1	CAN1 Fw	TCGGGAGCAAGATTGTTGTG
	SK1 Rv	GTTCGGCAAATGCCTGCAAATC
2	SK2 Fw	CGTTTACCATGGCCTATGTAGC
	SK2 Rv	CGCACGTCAAGACTGTCAAG
3	SK3 Fw	ATGGGAAGCCCGATGCGCCAGAG
	SK3 Rv	TGACAATATGCGCCTTGGCGATTTC
4	SK4 Fw	AGCTGAAGTGGCCGCTTCAACCACC
	SK4 Rv	CATCGTTGCTTGCAGGATGTTC
5	SK5 Fw	TTAATCGATGACAGCGTAGGG
	CAN1 Rv	AGAAGAGTGGTTGCGAACAGAG

Removaling of the endogenous SinLoG from the *SGA1* locus on chromosome IX

F	SGA1 Fw	ACTCGTACAAGGTGCTTTTAACTTG
	SGA1 Rv	TTGGGCTGGACGTTCCGACATAG

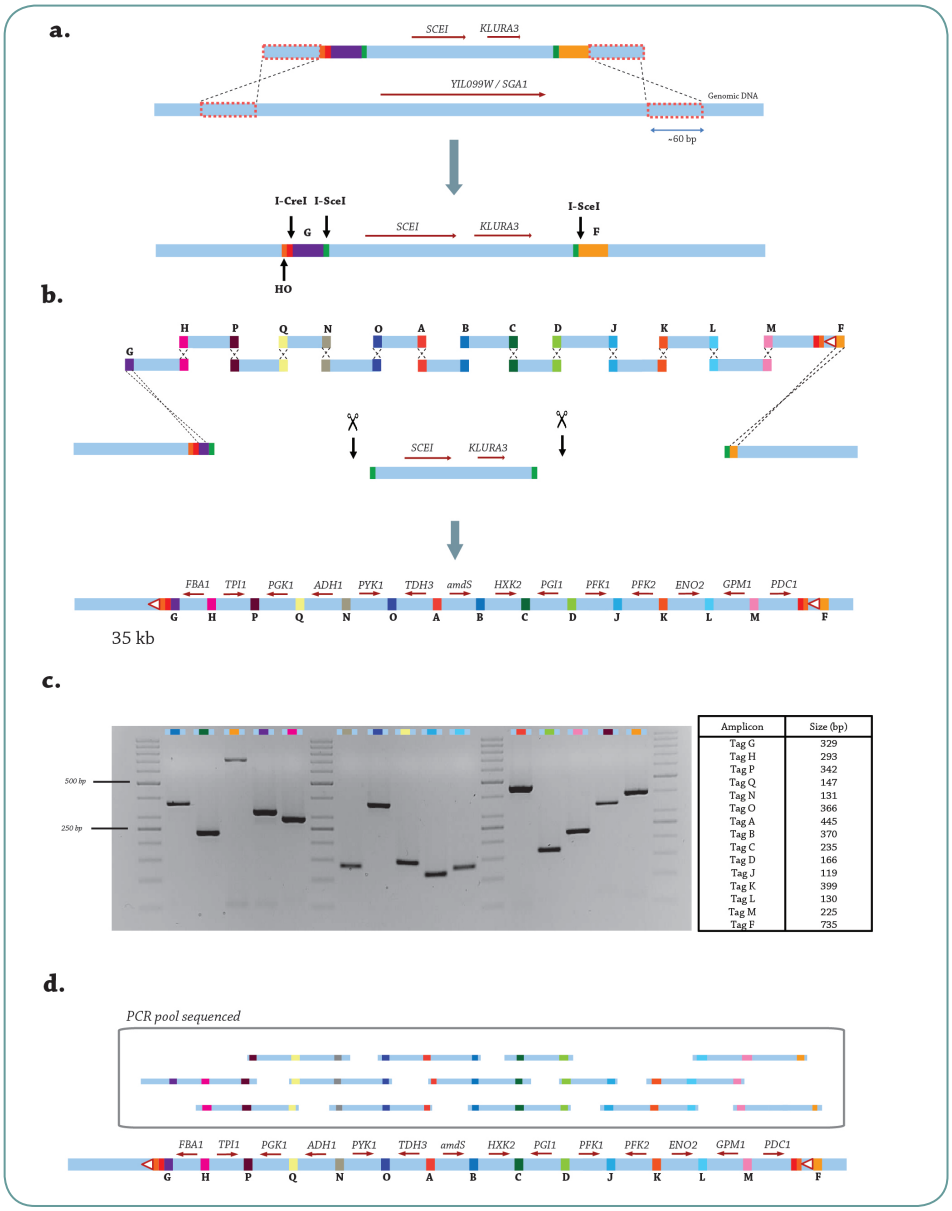
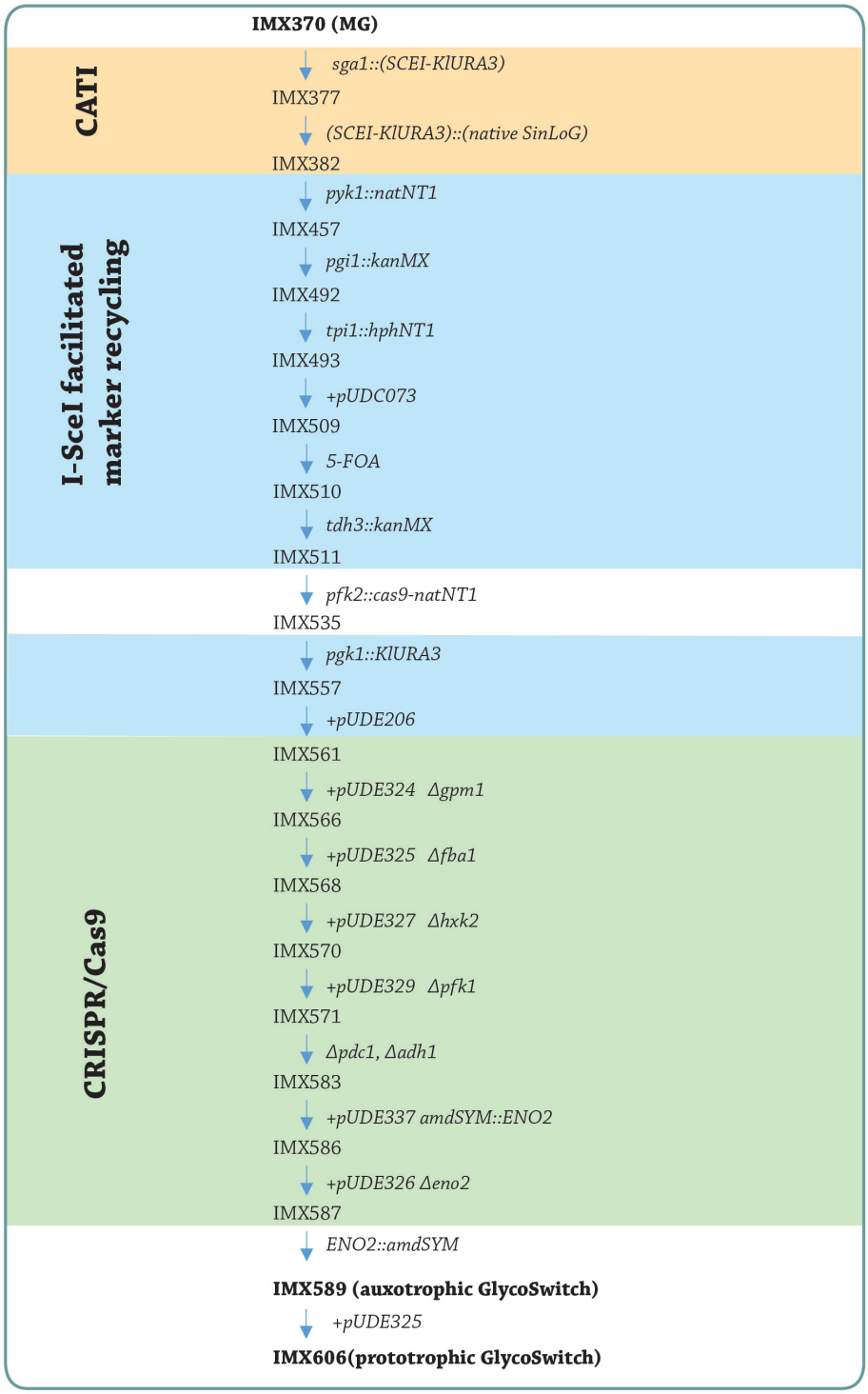


FIGURE S1 | Two-step construction and verification of IMX382 carrying the endogenous SinLoG. Introduction of the *SCEI/KlURA3* cassette in the *SGA1* locus of IMX370 to produce IMX377 (b). One-step *in vivo* assembly and integration of the endogenous SinLoG at the *SGA1* locus on chromosome IX (b). PCR confirmation of the correct assembly and integration of the complete endogenous SinLoG construct at the *SGA1* locus. PCRs were designed to produce amplicons covering the junctions of the assembly cassettes (c). To identify potential mutations within the native SinLoGconstruct, 14 overlapping fragments were amplified by PCR, pooled in a molar ratio of 1:1 and sequenced by next-generation sequencing (d).


IMX370 (MG)
CATI
sga1::(SCEI-KlURA3)
IMX377
(SCEI-KlURA3)::(native SinLoG)
IMX382
I-SceI facilitated marker recycling
pyk1::natNT1
IMX457
pgi1::kanMX
IMX492
tpi1::hphNT1
IMX493
+pUDC073
IMX509
5-FOA
IMX510
tdh3::kanMX
IMX511
pfk2::cas9-natNT1
IMX535
pgk1::KlURA3
IMX557
+pUDE206
IMX561
CRISPR/Cas9
+pUDE324 Δgpm1
IMX566
+pUDE325 Δfba1
IMX568
+pUDE327 Δhxk2
IMX570
+pUDE329 Δpfk1
IMX571
Δpdc1, Δadh1
IMX583
+pUDE337 amdSYM::ENO2
IMX586
+pUDE326 Δeno2
IMX587
ENO2::amdSYM
IMX589 (auxotrophic GlycoSwitch)
+pUDE325
IMX606(prototrophic GlycoSwitch)

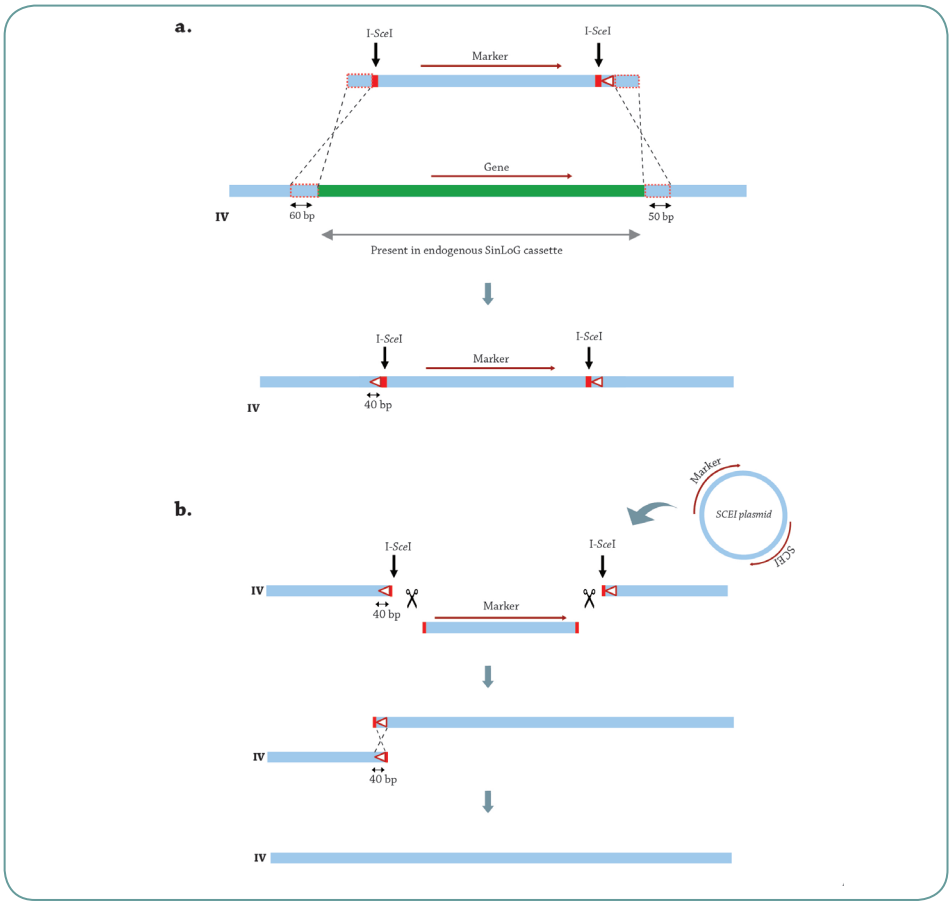
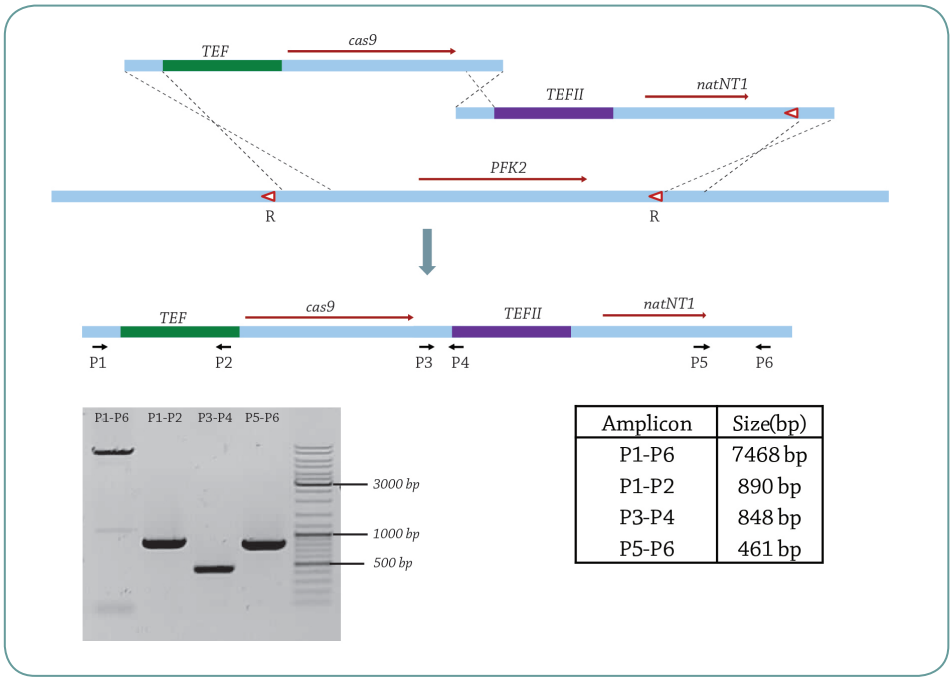


FIGURE S3 | Schematic overview of deletions with recyclable marker cassettes. Each deletion cassette was targeted to sequences not present on the native SinLoG cassette and contained two recognition sequences for I-SceI surrounding the marker and a 40-nucleotides sequence homologous to the DNA surrounding the deletion locus (repeat) (a). After three successive deletions, selection markers were removed by expression of I-SceI from a plasmid. The gap resulting from I-SceI restriction was repaired by homologous recombination of the repeats (b).

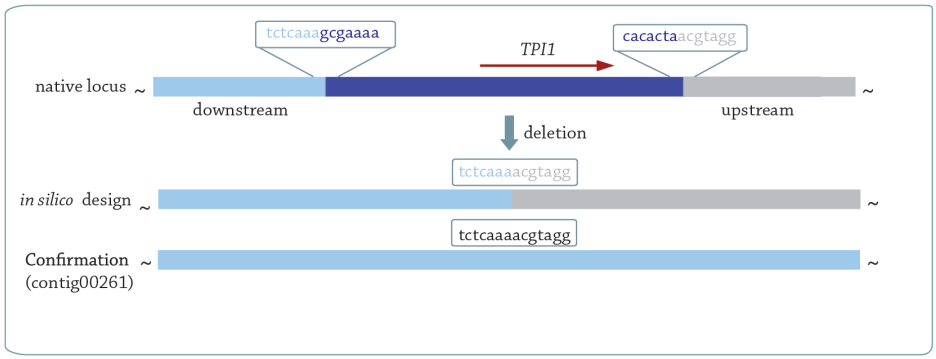
FIGURE S2 | Process flow of the construction of GlycoSwitch. The GlycoSwitch strain was constructed from the Minimal Glycolysis strain IMX372 in 19 rounds of modification [51]. First the native SinLoG cassette, containing the synthetic glycolytic pathway, was introduced using the CATI approach at the *SGA1* locus on chromosome IX [8]. The endogenous glycolytic genes *PYK1, PGI1, TPI1, TDH3* and *PGK1* were deleted using specific marker cassettes, which were recycled using the I-SceI facilitated marker recycling method [10]. *PFK2* was deleted by a 'classical' knockout cassette containing *cas9* and the marker gene natNT1. The genes *GPM1, FBA1, HXK2, PFK1, PDC1* and *ADH1* were deleted using CRISPR/Cas9 [13]. The synthetic *ENO2* gene appeared to be dysfunctional, so an additional synthetic *ENO2* variant was introduced into the native SinLoG cassette, after which the endogenous *ENO2* was removed from its original locus using the CRISPR/Cas9 system. Finally, the dysfunctional copy of *ENO2* was replaced by amdSYM.



Amplicon	Size(bp)
P1-P6	7468 bp
P1-P2	890 bp
P3-P4	848 bp
P5-P6	461 bp

FIGURE S4 | Integration of *cas9* in the *PFK2* locus. A cassette containing *cas9,* obtained by PCR from p414-TEF1p-cas9-CYC1t, and a second cassette containing the natNT1 marker gene, obtained by PCR from plasmid pUGnatNT1, were co-transformed to IMX511 for assembly and integration in the *PFK2* locus. Analysis with different primer sets, resulting in amplicons covering the complete construct, demonstrated the successful replacement of *pfk2* by *cas9*/natNT1.

FIGURE S5 | Confirmation of deletion of glycolytic genes from their native loci in the auxotrophic GlycoSwitch strain IMX589 based on whole genome sequencing. The *in silico* designs of the seamless deletions were blasted against the assembled contigs of the GlycoSwitch strain. For *tpi, fba1, gpm1, hxk2, pfk1, pdc1, pyk1, pgk1, pgi1* and *tdh3* the predicted *in silico* sequences were present on single contigs, thereby confirming their removal from the native loci. The loci of *eno2* and *adh1* were not present on a single contig due to the presence of the *adh1* terminator in the *cas9* cassette and the double presence of part of the *ENO2* promoter in the genome as visualized in Fig. S6. The deletion of *pfk2* was not included in this analysis, since this gene was removed by replacing it with the *cas9* cassette as represented in Fig. S5.



tpi

WT: aataagaacatcttctaaa.....acgtaggattattataactc
in silico design: aataagaacatcttctaaacgtaggattattataactc
Contig00261: aataagaacatcttctaaacgtaggattattataactc

fba1

WT: tgactttgaaatgactccgc.....aactattacgtattacgata
in silico design: tgactttgaaatgactccgcaactattacgtattacgata
Contig00222: tgactttgaaatgactccgcaactattacgtattacgata

gpm1

WT: atcaaggttattgctataa.....tatatatatatatatatata
in silico design: atcaaggttattgctataatatatatatatatatata
Contig00085: atcaaggttattgctataatatatatatatatatata

hxk2

WT: aacctcctcgcacattggta.....gcttaattttaaatttttt
in silico design: aacctcctcgcacattggtagcttaattttaaatttttt
Contig00361: aacctcctcgcacattggtagcttaattttaaatttttttt *insertion*

pfk1

WT: caatacgactacaattaagg.....catgtttttccatcgttttc
in silico design: caatacgactacaattaaggcatgtttttccatcgttttc
Contig00001: caatacgactacaattaaggcatgtttttccatcgttttc

pdc1

WT: gggttttggaaaccacactg.....tcaagttgaagactatatat
in silico design: gggttttggaaaccacactgtcaagttgaagactatatat
Contig00046: gggttttggaaaccacactgtcaagttgaagactatatat

pyk1

WT: gtaaataaacaatcacaacg.....gatgccagtgaactatgttc
in silico design: gtaaataaacaatcacaacggatgccagtgaactatgttc
Contig00017: gtaaataaacaatcacaacggatgccagtgaactatgttc

pgk1

WT: gagtgaggaactatcgcata.....aatagtttatttttatttttt
in silico design: gagtgaggaactatcgcataaatagtttatttttatttttt
Contig00248: gagtgaggaactatcgcataaatagtttatttttatttttt

pgi1

WT: aatttttaaaatttttacttt.....taatatttttttcttttgaaa
in silico design: aatttttaaaatttttacttaatatttttttcttttgaaa
Contig00061: aatttttaaaatttttacttaata-tt----ttttgaaa *deletion*

tdh3

WT: agcgcctccattattaacta.....tttgaaatggcagtattgat
in silico design: agcgcctccattattaactatttgaaatggcagtattgat
Contig00136: agcgcctccattattaactagtcgtaattgagtcagttacgctagggataacagggtaatatagttcttc
attctacgcacgatatccccaaaattattaagagcgcctccttattaactatttgaaatggcagtattgat
insertion

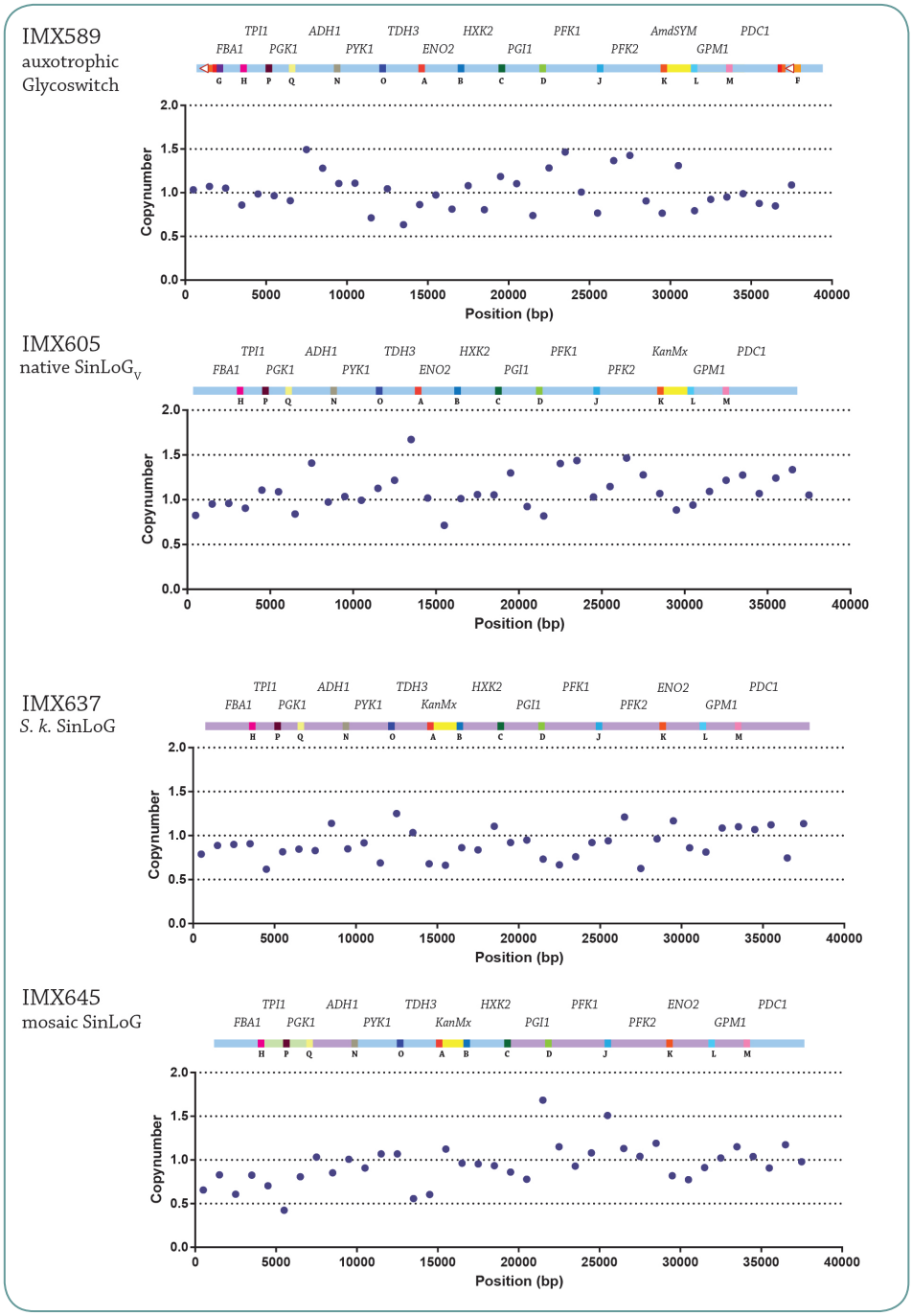


FIGURE S6 | Copy number variation within the synthetic glycolysis constructs of the engineered strains. The graphs represent the copy number variation as generated by the Magnolya algorithm [31] of contigs that were *de novo* assembled by Newbler (www.454.com) and aligned to the *in silico* designs of the synthetic glycolysis constructs.

Alignment of PFK subunitα from *Saccharomyces cerevisiae* (top line) and*S. kudriavzevii* (bottom line)

```
ScPFK1aa              3 sqdscygvafrsiitndealfkktihfyhtlgfatvkdfnkfkhgensllssgtsqdslrevwlesfklsevdasgfrip
PFK1aa                3 ....................................................................................

ScPFK1aa             83 qqeatnkaqsqgallkirlvmsapidetfdtnetatityfstdlnkivekfpkqaeklsdtlvflkdpmgnnitfsglan
PFK1aa               83 ...............................................q....................................

ScPFK1aa            163 atdsaptskdafleatsedeiisrassdasdllrqtlgssqkkkkiavmtsggdspgmnaavravvrtgihfgcdvfavy
PFK1aa              163 ..........t.....................................................................

ScPFK1aa            243 egyeqllrggkylkkmawedvrgwlseggtligtarsmeirkregrrqaagnlisqgidalyvcgqdgsltgadlfrhew
PFK1aa              243 ..........................................k.....................i.................

ScPFK1aa            323 pslvdelvaegrftkeevapyknlsivglvgsidndmsgtdstigayslericemvdyidataksshsrafvvevmgrhc
PFK1aa              323 ....e.............f.....................................................f.....

ScPFK1aa            403 qwlalmagiatqadyifiperavphqkwqdelkevcqrhrskgrrnntiivaeqalddqlnpvtandvkdalielgldtk
PFK1aa              403 ...................................................l..............vd.....

ScPFK1aa            483 vtilghvqrggtavahdrwlatlqgvdavkavleftpetpspligilenkiirmplvesvkltksvataienkdfdkais
PFK1aa              483 ....................................................................................

ScPFK1aa            563 lrdtefielyenflsttvkddgselpvsdrlnigivhvgapsaalnaatraatlyclshghkpyaimngfsgliqtgev
PFK1aa              563 ........................f...........................................................

ScPFK1aa            643 kelswidvenwhnlggseigtnrsvasedlgtiayyfqknkldgliilggfegfrslkqlrdgrtqhpifnipmclipat
PFK1aa              643 ...........................................................k........................

ScPFK1aa            723 vsnnvpgteyslgvdtclnalvnytddikqsasatrrrvfvcevqgghsgyiasftglitgavsvytpekkidlasired
PFK1aa              723 ....................................................................................

ScPFK1aa            803 itllkenfrhdkgenrngkllvrneqassvystqlladiiseaskgkfgvrtaipghvqqggvpsskdrvtasrfavkci
PFK1aa              803 ....q...............................................................................

ScPFK1aa            883 kfieqwnkkneaspntdakvlrfkfdthgekvptveheddsaavicvngshvsfkpianlwenetnvelrkgfevhwaey
PFK1aa              883 ....k...........................................................h.......

ScPFK1aa            963 nkigdilsgrlklraevaalaaenk
PFK1aa              963 ............m.......stt...
```

Alignment of PFK subunitβ from *Saccharomyces cerevisiae* (top line) and *S. kudriavzevii* (bottom line)

```
ScPFK2aa              1 mtvttpfvngtsyctvtaysvqsykaaidfykflslenrsspdenstllsndsislkillrpdekinknveahlkelnsitktqdwrsh
PFK2aa                1 ......................p.........e.....................e...........g........v.....

ScPFK2aa             91 atqslvfntsdilavkdtlnamnaplqgyptelfpmqlytldplgnvvgvtstknavstkptpppapeasaesglsskvhsytdlayrmk
PFK2aa               91 .s.....................................................................t..r.....

ScPFK2aa            181 ttdtypslpkpsnrpqkaiavmtsggdapgmnsnvraivrsaifkgcrafvvmeqyeglvrggpeyikefhwedvrgwsaeggtnigtar
PFK2aa              181 ............kk......................................................................

ScPFK2aa            271 cmefkkregrllgaqhlieagvdalivcggdgsltgadlfrsewpslieellktnrisneqyermkhlnicgtvgsidndmsttdatiga
PFK2aa              271 ...................................................d...q..............................

ScPFK2aa            361 ysaldrickaidyveatanshsrafvvevmgrncgwlallagiatsadyifipekpatssewqdqmcdivskhrsrgkrttivvvaegai
PFK2aa              361 ....................................................................................

ScPFK2aa            451 sadltpispsdvhkvlvdrlgldtrittlghvqrggtavaydrilatlqgleavnavlestpdtpspliavnenkivrkplmesvkltka
PFK2aa              451 g.....................................................s..............................

ScPFK2aa            541 vaeaiqakdfkramslrdtefiehlnnfmainsadhnepklpkdkrlkiaivnvgapagginsavysmatycmsqghrpyaiyngwsgla
PFK2aa              541 ....................................................................................

ScPFK2aa            631 rhesvrslnwkdmlgwqsrggseigtnrvtpeeadlgmiayyfqkyefdgliivggfeafeslhqleraresypafripmvlipatlsnn
PFK2aa              631 ....................................................................................

ScPFK2aa            721 vpgteyslgsdtalnalmeycdvvkqsasstrgrafvvdcqggnsgylatyaslavgaqvsyvpeegisleqlsedieylaqsfekaegr
PFK2aa              721 ..........................................s.........................................

ScPFK2aa            811 grfgklilkstnaskalsatklaevitaeadgrfdakpaypghvqqqglpspidrtratrmaikavgfiednqaaiaearaaeenfnadd
PFK2aa              811 ........................d...........................................................d.....

ScPFK2aa            901 ktisdtaavvgvkgshvvynsirqlydyetevsmrmpkvihwqatrliadhlvgrkrvd
PFK2aa              901 ....................................................
```

FIGURE S7 | Protein alignment of the α and β subsunits of phosphofructokinase of *S. cerevisiae* and *S. kudriavzevii* using Clustal X in Clonemanager 9. Arg283 of subunit α (shown in the red box) is important for enzymatic activity [21] and is not present in the α subunit of phosphofructokinase in *S. kudriavzevii*.

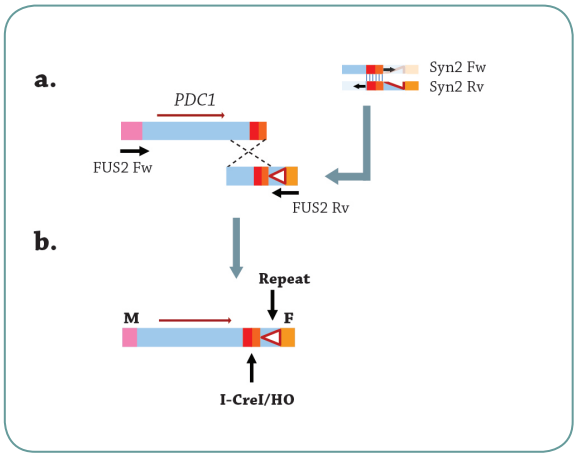
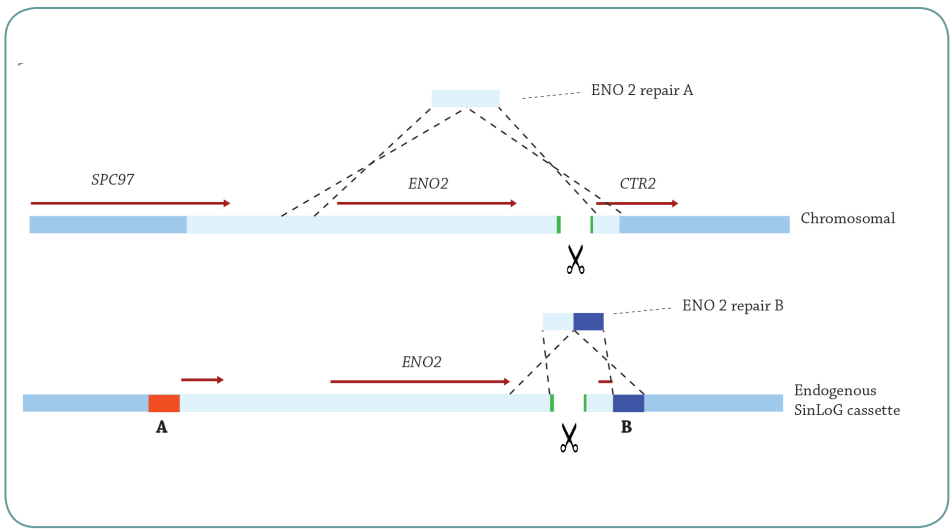


FIGURE S8 | Construction scheme of the *SYN-PDC1*$_{MF}$ glycoblock. The SYN2 fragment was obtained by fusing oligo's Syn2-Fw and Syn2-Rv in an independent PCR (A). *PDC1* from *S. cerevisiae*, including its own promoter and terminator, and SYN2 were fused in a fusion-PCR using primers FUS2-Fw and FUS2-Rv, resulting in the *SYN-PDC1*$_{MF}$ cassette (B).

FIGURE S10 | Overview of the *ENO2* deletion scheme from its endogenous locus. The target sequence of the *ENO2* gRNA was present in the native SinLoG cassette and on the endogenous chromosomal locus of *ENO2*. Co-transformation of two different repair fragments targeting the two loci enabled to specifically delete *ENO2* from its endogenous locus, but not from the endogenous SinLoG construct.



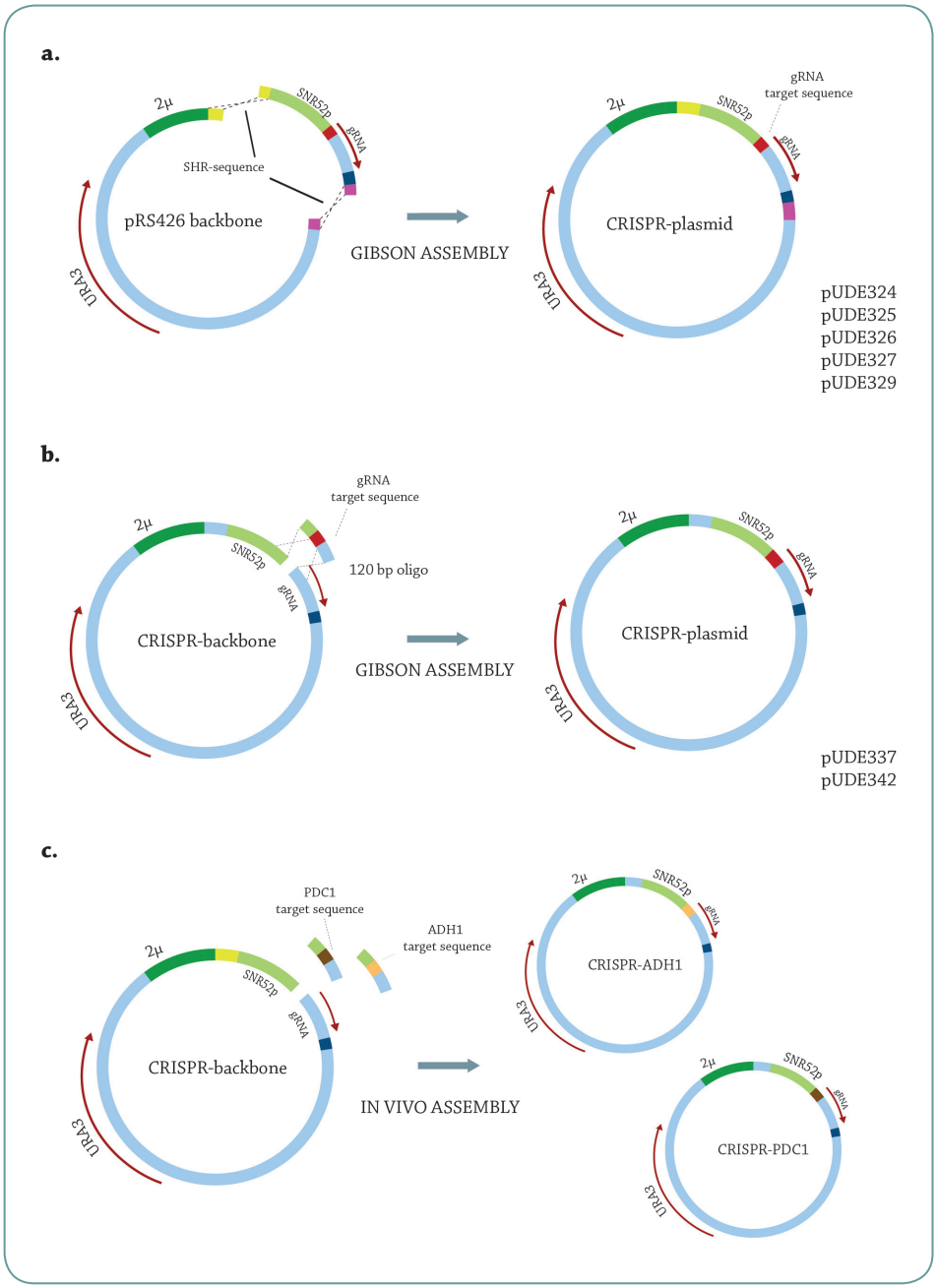


FIGURE S9 | Overview of the construction scheme of the CRISPR-plasmids. Gibson assembly-based construction of the CRISPR-plasmids carrying the gRNA targeting *GPM1* (pUDE324), *FBA1* (pUDE325), *ENO2* (pUDE326), *HXK2* (pUDE327) and *PFK1* (pUDE329) (A). Gibson assembly of the CRISPR-plasmids carrying the amdsYM cassette (pUDE337) and the native SinLoG flanking regions (pUDE342) (B). *In vivo* assembly of CRISPR-plasmids for the simultaneous double deletion of *ADH1* and *PDC1* (C).

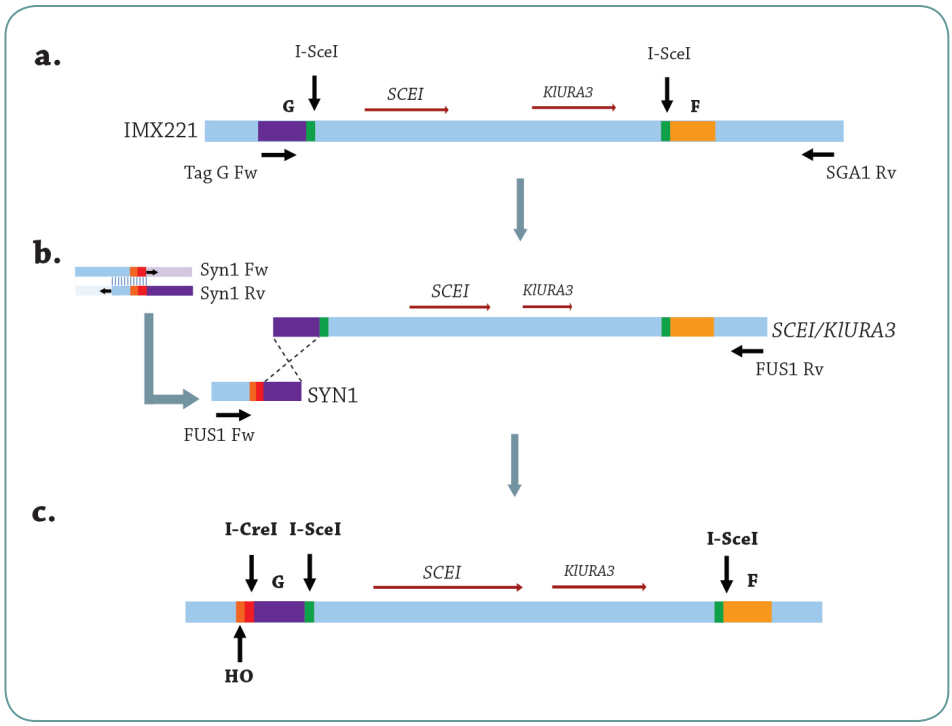
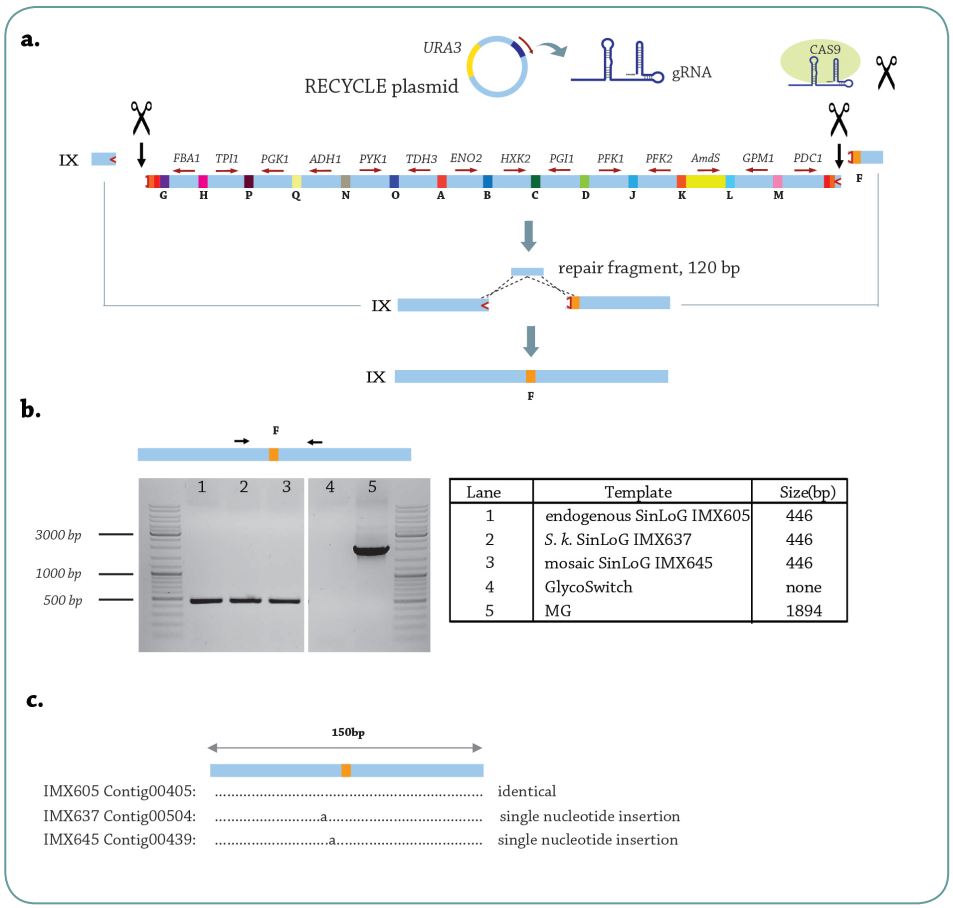


FIGURE S11 | Construction scheme of the *SCEI/KlURA3* cassette for combined *in vivo* assembly and integration (CATI). PCR amplification of the *SCEI/KlURA3* cassette from IMX221 (A). Addition of extra restriction sites and of the flanking regions targeting the *SGA1* locus (B). Final *SCEI/KlURA3* cassette used to promote the integration of the endogenous SinLoG cassette in IMX370, resulting in IMX377 (C).



Lane	Template	Size(bp)
1	endogenous SinLoG IMX605	446
2	*S. k.* SinLoG IMX637	446
3	mosaic SinLoG IMX645	446
4	GlycoSwitch	none
5	MG	1894

FIGURE S12 | Schematic overview of the deletion of the native SinLoG cassette from the *SGA1* locus. Co-transformation of strains IMX591, IMX607 and IMX633 with the CRISPR plasmid pUDE342 targeting the native SinLoGinsert with the corresponding repair fragment resulted in excision of the native SinLoGcassette from their genome. PCR analysis and sequencing of the resulting strains (IMX605, IMX637 and IMX645 respectively) demonstrated the absence of native SinLoG cassette at the SGA1 locus.

SECTION 3: SYNTHESIZED DNA CONSTRUCTS

Sequence codon optimized *Homo sapiens TPI1*, muscle, splicing variant 1.

1ATGGCTCCATCTAGAAAGTTCTTCGTCGGTGGTAACTGGAAGATGAACGGTAGAA
AGCAATCTTTAGGTGAATTGATCGGTACCTTGAACGCTGCTAAGGTCCCAGCTGAC
ACCGAAGTTGTCTGTGCTCCACCAACCGCTTACATCGACTTCGCTAGACAAAAGTT
GGACCCAAAGATCGCTGTCGCTGCTCAAAACTGCTACAAGGTCACCAACGGTGCT
TTCACCGGTGAAATCTCTCCAGGTATGATTAAGGACTGCGGTGCTACCTGGGTCG
TTTTGGGTCACTCTGAAAGAAGACACGTCTTCGGTGAATCTGACGAATTGATCG
GTCAAAAGGTCGCTCACGCTTTGGCTGAAGGTTTGGGTGTCATTGCTTGTATCG
GTGAAAAGTTGGACGAAAGAGAAGCTGGTATCACCGAAAAGGTCGTCTTCGAA
CAAACCAAGGTCATTGCTGACAACGTCAAGGACTGGTCTAAGGTCGTCTTGGCT
TACGAACCAGTCTGGGCTATTGGTACCGGTAAGACCGCTACCCCACAACAAGCTC
AAGAAGTCCACGAAAAGTTGAGAGGTTGGTTGAAGTCCAACGTCTCTGACGCTG
TCGCTCAATCCACCAGAATCATTTACGGTGGTTCTGTCACCGGTGCTACCTGCAA
GGAATTGGCTTCTCAACCAGACGTCGACGGTTTCTTGGTCGGTGGTGCTTCTTT
GAAGCCAGAATTCGTCGACATCATTAACGCTAAGCAATAA

Sequence codon optimized *Homo sapiens PGK1*, muscle, splicing variant 1.

ATGTCTTTGTCTAACAAGTTGACTTTGGACAAGTTGGACGTTAAGGGTAAGAGA
GTCGTCATGAGAGTTGACTTCAACGTTCCAATGAAGAACAACCAAATTACCAAC
AACCAAAGAATTAAGGCTGCTGTCCCATCTATTAAGTTCTGTTTGGACAACGGT
GCTAAGTCTGTCGTCTTGATGTCTCACTTGGGTAGACCAGACGGTGTTCCAATG
CCAGACAAGTACTCCTTGGAACCAGTTGCTGTTGAATTGAAGTCTTTGTTGGGTA
AGGACGTCTTGTTCTTGAAGGACTGCGTCGGTCCAGAAGTCGAAAAGGCTTGTG
CTAACCCAGCTGCTGGTTCCGTCATTTTGTTGGAAAACTTGAGATTCCACGTCGA
AGAAGAAGGTAAGGGTAAGGACGCTTCTGGTAACAAGGTTAAGGCTGAACCAG
CTAAGATTGAAGCTTTCAGAGCTTCTTTGTCTAAGTTGGGTGACGTTTACGTCA
ACGACGCTTTCGGTACCGCTCACAGAGCTCACTCTTCTATGGTTGGTGTTAACT
TGCCACAAAAGGCTGGTGGTTTCTTGATGAAGAAGGAATTGAACTACTTCGCT
AAGGCTTTGGAATCTCCAGAAAGACCATTCTTGGCTATCTTGGGCGGTGCTAA
GGTCGCTGACAAGATTCAATTGATCAACAACATGTTGGACAAGGTCAACGAAA
TGATTATTGGTGGTGGTATGGCTTTCACCTTCTTGAAGGTCTTGAACAACATGG
AAATTGGTACCTCCTTGTTCGACGAAGAAGGTGCTAAGATCGTCAAGGACTTGA
TGTCTAAGGCTGAAAAGAACGGTGTTAAGATCACCTTGCCAGTTGACTTCGTCA
CCGCTGACAAGTTCGACGAAAACGCTAAGACCGGTCAAGCTACCGTCGCTTCCG
GTATTCCAGCTGGTTGGATGGGTTTGGACTGTGGTCCAGAATCTTCTAAGAAGT
ACGCTGAAGCTGTCACCAGAGCTAAGCAAATTGTCTGGAACGGTCCAGTCGGTG
TCTTCGAATGGGAAGCTTTCGCTAGAGGTACCAAGGCTTTGATGGACGAAGTCGT
CAAGGCTACCTCTAGAGGTTGTATTACCATTATTGGTGGTGGTGACACCGCTACCT
GTTGTGCTAAGTGGAACACCGAAGACAAGGTTTCCCACGTCTCTACCGGTGGTGGT
GCTTCTTTGGAATTGTTAGAAGGTAAGGTCTTGCCAGGTGTCGACGCTTTGTCTAA
CATTTAA

FBA1 gRNA + SHR

ATTGCGTCAGAGCCCTCGATGAGACTAAGTTCGGAGAATCTATGGCAGCCTCAGCT
TGTGTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAACTT
GATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTA
GATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTAC
AATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGTGCGCA
TGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC**AATTAACGGCCA**
GTCCACTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCATT
AGCCCGCAGTGTGTGGTCCTCCAGAACTCAAATGAAGGCATACTACACAGATTCC

ENO2 gRNA + SHR

TCGAAGCAGTGTTAATTCCGTGCCCGACAGAACACATGGAAAGGACACTTCCCTAA
CTAGAGACATAAAAAACAAAAAAAGCACCACCGACTCGGTGCCACTTTTTCAAG
TTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC**CTTCCAG**
TGCATTATGCAATGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGCCGAAA
CATGCGCACCAACTTTCACTTCTACAGCGTTTGACCAAAATCTTTTGAACAGAACA
TTGTAGGGTGTGAAAAAATGCGCACCTTTACCGCTAGCCCAAGAGGGCACTACAA
AATCTAGAGTTGTACTTCAAACGTACATGTAATCACCTTGTATATACTCGAAAGA
AAACATCAAGTTTCTGTATAAATATGAGTGAAAGCATAATCATACATTATCTTTT
CAAAGATCTGGCAGTCCATTGGCATGCCAGCCCTGCGATTATTTGTTCATACCG
GCCAGTAGGATG

HXK2 gRNA + SHR

TCATTAGCCCGCAGTGTGTGGTCCTCCAGAACTCAAATGAAGGCATACTACACAGA
TTCCAGACATAAAAAACAAAAAAAGCACCACCGACTCGGTGCCACTTTTTCAAGTT
GATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC**GCATGATAGC**
CATTTCTAGGGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGCCGAAACATG
CGCACCAACTTTCACTTCTACAGCGTTTGACCAAAATCTTTTGAACAGAACATTGTA
GGGTGTGAAAAAATGCGCACCTTTACCGCTAGCCCAAGAGGGCACTACAAAATCTA
GAGTTGTACTTCAAACGTACATGTAATCACCTTGTATATACTCGAAAGAAAACATC
AAGTTTCTGTATAAATATGAGTGAAAGCATAATCATACATTATCTTTTCAAAGAGCA
TCTCTATAACTGGTGTCGCTGAACTACCATGTACTGCCCATGCGGCAAATGAATCCA

GPM1 gRNA + SHR

GCATCTCTATAACTGGTGTCGCTGAACTACCATGTACTGCCCATGCGGCAAATGAA
TCCATCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAAC
TTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAA
CTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCAC
ACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTT
GGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC**TGAT**
GTCTAAGTAACCTTTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT
CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTA
TGTCTTCGAAGCAGTGTTAATTCCGTGCCCGACAGACACATGGAAAGGACACTTC
CCTAACTAG

PFK1 gRNA + SHR

ATCGACCCTGCATCGTGTTGCGATTCTACGGACGCTGAGTGGCCAGCCATTCATAT
AACGAGACATAAAAAACAAAAAAAGCACCACCGACTCGGTGCCACTTTTTCAAGT
TGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC**CATCATAG**
TGAGGCGCGCTAGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGCCGAAAC
ATGCGCACCAACTTTCACTTCTACAGCGTTTGACCAAAATCTTTTGAACAGAACA
TTGTAGGGTGTGAAAAAATGCGCACCTTTACCGCTAGCCCAAGAGGGCACTACAA
AATCTAGAGTTGTACTTCAAACGTACATGTAATCACCTTGTATATACTCGAAAGAAA
ACATCAAGTTTCTGTATAAATATGAGTGAAAGCATAATCATACATTATCTTTTCAA
AGAATTGCGTCAGAGCCCTCGATGAGACTAAGTTCGGAGAATCTATGGCAGCCTC
AGCTTGTG

Samenvatting

Al millennia lang maakt de mensheid gebruik van fermentatieprocessen voor de productie van onder andere alcoholische dranken en zuivelproducten als yoghurt en kaas. In de afgelopen decennia is er, onder andere vanwege zorgen omtrent klimaatverandering, gewerkt aan biotechnologische alternatieven voor petrochemische processen. Op deze manier proberen wetenschappers een alternatief te bieden voor het gebruik van fossiele brandstoffen. Voorbeelden van dit soort biotechnologische processen zijn de productie van bio-alcohol, een biobrandstof, met gist en de productie van melkzuur, een bouwsteen voor de biologisch afbreekbare bioplastic polymelkzuur, met behulp van melkzuurbacteriën. Om deze processen economisch competitief te maken is er een intensieve zoektocht gaande naar verbeterde, efficiëntere productie-organismen. Naast het klassiek veredelen van productiestammen worden ook doelgerichte, rationele mutaties in het DNA van micro-organismen aangebracht om gewenste eigenschappen te verkrijgen. De laatstgenoemde "metabolic engineering"-benadering voor het genetische modificeren van micro-organismen kan resulteren in hogere productopbrengsten en robuustere productiestammen. In het afgelopen decennium heeft het genetisch modificeren van micro-organismen een enorme vlucht genomen, waarbij soms complete stofwisselingsroutes zijn toegevoegd aan een organisme. Een goed voorbeeld hiervan is de productie van het malariamedicijn artimisinine, dat oorspronkelijk uit planten werd gewonnen, maar nu ook en veel goedkoper met behulp van genetisch gemodificeerde gist kan worden gemaakt. Het op zo'n drastische schaal wijzigen van de natuurlijke eigenschappen van levende cellen wordt ook wel aangeduid als synthetische biologie.

Het op grote schaal modificeren van het genoom van levende cellen vormt een grote technische uitdaging, omdat de essentiële processen voor overleving van de cel niet onderbroken kunnen worden. Dit heeft tot gevolg dat modificatie veelal beperkt blijft tot het verwijderen van genen die niet essentieel zijn voor het voortbestaan van de cel en tot het introduceren van additionele stofwisselingsroutes die leiden tot de vorming van nieuwe producten zoals het eerder genoemde artimisinine. Om echter productopbrengsten en productiesnelheden (fluxen) verder te optimaliseren zal er ook ingegrepen moeten worden in de centrale stofwisseling en in centrale regelprocessen van de cel. Het radicaal aanpassen of vervangen van een sleutelroute in de stofwisseling vereist een strategie waarbij de essentiële cellulaire functie van deze route niet wordt onderbroken terwijl deze wordt vervangen door een alternatieve route. In dit proefschrift wordt een strategie voor dit type van radicale herprogrammering van de stofwisseling ontwikkeld en toegepast op de suikerstofwisseling van bakkersgist (*Saccharomyces cerevisiae*).

De keuze voor *S. cerevisiae* als organisme voor dit promotieproject is een vanzelfsprekende. *S. cerevisiae* is een belangrijk organisme in het fundamentele onderzoek en is bovendien een populair industrieel platform voor de biotechnologische productie van tal van suiker-afgeleide chemicaliën. Deze wijdverbreide populariteit van *S. cerevisiae* komt mede door de beschikbaarheid van een goed geannoteerde sequentie van het genoom en een excellente toegankelijkheid voor genetische modificatie. Deze laatste eigenschap komt mede doordat deze gist extreem efficiënt is in het "lijmen" van DNA door middel van homologe recombinatie, wat tal van mogelijkheden biedt voor

het maken van genetische aanpassingen en dus voor het veranderen van cellulaire processen.

Zoals in bijna alle cellen, vindt het eerste deel van de suikerafbraak in *S. cerevisiae* plaats via een serie van aaneengeschakelde enzymatische reacties die samen de glycolyse worden genoemd. De glycolyse is aanwezig in bijna iedere levende cel en is daarmee een van de meest geconserveerde stofwisselingsroutes in het leven op aarde. In de glycolyse wordt glucose in 10 enzymatische stappen geoxideerd tot twee moleculen pyrodruivenzuur, waarbij energie wordt vastgelegd in de vorm van ATP. Naast de rol van de glycolyse bij het beschikbaar maken van energie, fungeren tussenproducten van deze stofwisselingsroute ook als essentiële bouwstenen voor het maken van celbestanddelen. Het in de glycolyse gevormde pyrodruivenzuur kan in aanwezigheid van zuurstof verder worden geoxideerd via de citroenzuurcyclus. In afwezigheid van zuurstof wordt pyruvaat door *S. cerevisiae* gedecarboxyleerd, waarna het gevormde aceetaldehyde wordt gereduceerd tot ethanol. In dit proefschrift is er voor gekozen om ook de twee enzymatische omzettingen van pyrodruivenzuur in ethanol tot de glycolyse te rekenen.

Ondanks de uitstekende toegankelijkheid van *S. cerevisiae* voor genetische modificatie vormt het vervangen van de gehele glycolyse een enorme uitdaging. De in totaal 12 cytosolische glycolytische reacties worden in *S. cerevisiae* gekatalyseerd door 26 enzymen, die worden gecodeerd door 26 corresponderende genen. Dit betekent dat sommige omzettingen door meerdere enzymen kunnen worden uitgevoerd. De eerste stap in de richting van een "synthetische glycolyse" is daarom het verwijderen van overbodige iso-enzymen om zo, idealiter, de complexiteit van de glycolyse terug te brengen tot een enkel enzym (en dus een enkel gen) per reactie. Een geslaagde poging om dit ambitieuze doel te realiseren is uitvoerig beschreven in het proefschrift van Daniel Solis-Escalante. De door hem geconstrueerde "minimale glycolyse" staat aan de basis van het onderzoek dat wordt beschreven in dit proefschrift.

Het doel van het huidige proefschrift is het construeren van een *S. cerevisiae* stam waarin de complete glycolyse in enkele eenvoudige stappen kan worden vervangen door een synthetische variant. Een belangrijke tussenstap bij het bereiken van dit doel was het clusteren van de glycolytische genen in de "minimale glycolyse"-stam op een enkele plek in het gistgenoom. Het clusteren van de gehele glycolyse heeft tot gevolg dat de gehele glycolyse eenvoudig in een enkele transformatie verwijderd kan worden doordat alle genen fysiek naast elkaar liggen op het genoom. Om tot een geclusterde glycolyse te komen vanaf de "minimale glycolyse"-stam moesten dus 13 genen worden geïntroduceerd in het genoom en moesten tevens de 13 originele genen worden verwijderd. Het binnen een kort tijdsbestek verwijderen van deze 13 genen vormde een serieuze uitdaging. Met het resulterende platform kan vervolgens aan "glycolyse wisselen" worden gedaan, waarbij een synthetische glycolyse wordt geïntroduceerd in de cel, waarna de "originele" geclusterde glycolyse in een enkele stap verwijderd kan worden.

Het aanbrengen van een dusdanig groot aantal veranderingen in het gistgenoom

betekende een nieuwe schaalgrootte van genetische modificatie. Aan het begin van dit project waren nog niet alle technieken voorhanden om deze ambitieuze doelen binnen een vierjarig project te realiseren. Voor het in een enkele transformatie introduceren van 13 genen waren klassieke technieken voor genetische modificatie niet toereikend. Het assembleren van 13 genen in een plasmide, met behulp van de klassieke restrictie- en ligatie-technieken is een tijdrovend repetitief proces dat bovendien wordt gelimiteerd door beschikbaarheid van unieke restrictie-sequenties in grote DNA-constructen.

In het afgelopen decennium zijn alternatieve methoden ontwikkeld voor het simultaan assembleren van meerder DNA-fragmenten. Een zeer veelbelovende techniek is het *in vivo* assembleren van DNA-fragmenten door middel van homologe recombinatie in *S. cerevisiae*. Deze methode, die bekend staat als 'TAR-cloning', is door Dan Gibson en zijn team gebruikt voor het assembleren van het eerste complete synthetische genoom in 2010. In deze methode worden overlappende DNA-fragmenten getransformeerd naar *S. cerevisisae*. In de gistcel worden de fragmenten vervolgens, gestuurd door homologe recombinatie van overlappende uiteinden, geassembleerd tot een circulair plasmide. Hoewel deze techniek snel werd opgepikt door andere laboratoria, was de gerapporteerde efficiëntie van correcte assemblage niet consistent en was standaardisering van deze methode niet grondig onderzocht. In **hoofdstuk 2** is daarom onderzocht hoe het *in vivo* assembleren van plasmiden uit overlappende DNA-fragmenten kan worden geoptimaliseerd. Op basis van dit onderzoek zijn twee belangrijke verbeteringen aangebracht in de methode: i) het markergen en de sequenties voor replicatie van het plasmide zijn op verschillende fragmenten geplaatst en ii) er zijn speciaal ontworpen synthetische DNA sequenties van 60 basenparen gebruikt voor homologe recombinatie (aangeduid als SHR-sequenties). Deze modificaties resulteerden in een honderdvoudige verlaging van het aantal fout geassembleerde plasmiden, doordat zelfsluiting van het ontvangende basisplasmide werd bemoeilijkt. Tevens voegde de standaardisatie van de SHR-sequenties flexibiliteit toe aan het systeem, door fragmenten uitwisselbaar te maken. Om deze verbeteringen experimenteel te testen, werd de geoptimaliseerde methode gebruikt om zes glycolytische genen in een enkele transformatie te assembleren tot een plasmide van 21 kb. Gemeten naar het aantal correct geassembleerde plasmiden, leverde dit een efficiëntie van 95%. Door bepaling van de DNA-volgorde van een geassembleerd plasmide kon bovendien worden vastgesteld dat deze overeenkwam met het *in-silico* ontwerp. Daarnaast bleek uit complementatiestudies dat alle zes de glycolytische genen op het plasmide functioneel tot expressie kwamen in *S. cerevisiae*. Deze methode bood, in de context van dit promotieonderzoek, een oplossing voor het efficiënt assembleren van de 13 DNA-fragmenten die nodig zijn voor een geclusterde glycolyse. In meer algemene zin laat het gepresenteerde werk zien dat *S. cerevisiae* gebruikt kan worden als universeel platform voor snelle en accurate vervaardiging van grote plasmiden.

Hoewel het *in vivo* assembleren van DNA-fragmenten in *S. cerevisiae* een oplossing bood voor het introduceren van een complete stofwisselingsroute in een enkele transformatie, is het verkregen product altijd een plasmide. Plasmiden zijn echter niet stabiel tijdens replicatie en zonder selectieve druk verdwijnen plasmiden

geleidelijk uit een populatie. Daarom heeft chromosomale integratie gewoonlijk de voorkeur boven expressie van een plasmide. In **hoofdstuk 3** is onderzocht of het *in vivo* assembleren van meerdere DNA-fragmenten gecombineerd kan worden met integratie op een specifieke plek in het genoom. Hierbij worden de overlappende DNA-fragmenten niet geassembleerd tot een circulair plasmide, maar tot een lineaire DNA-cassette, waarvan de uiteinden homoloog zijn met de DNA-volgorde van de locatie op het genoom waar de cassette uiteindelijk terecht dient te komen (zoals dit ook bij een klassieke integratiecassette het geval is). Deze strategie is toegepast voor de gelijktijdige *in vivo* assemblage en integratie van zeven glycolytische genen op een specifieke chromosomale locatie. Hoewel succesvol, was de efficiëntie van deze strategie, gemeten naar het aantal correcte transformanten, slechts 5 %. Daarom is onderzocht of het aanbrengen van een gerichte dubbelstrengs DNA-breuk in het genoom, precies op de beoogde integratieplek en met behulp van het meganuclease I-SceI, de effectiviteit en accuratesse van deze strategie kon verhogen. De door I-SceI gefaciliteerde integratie van dezelfde set van zeven glycolytische genen resulteerde in een drastische toename van het aantal verkregen transformanten. Van de geteste transformanten bevatte zelfs 95 % de beoogde configuratie van de zeven genen, op de beoogde plek in het gistgenoom. Dit resultaat toonde duidelijk aan dat het *in vivo* assembleren en tegelijkertijd integreren van de genen voor een complete stofwisselingsroute, de zogenaamde CATI benadering, een uitstekende methode is voor ingrijpende "metabolic engineering"-studies in gist.

Naast het snel en accuraat introduceren van een complete set van glycolytische genen, vereiste het beoogde platform ook een aanpak voor het verwijderen van 13 genen uit het genoom. Een belangrijke hindernis bij het verwijderen van een zo groot aantal genen is de noodzaak om gebruik te maken van selecteerbare genetische markers. Hoewel er een behoorlijk aantal markers voor *S. cerevisiae* beschikbaar is, is het niet wenselijk om voor iedere gendeletie een nieuwe marker te gebruiken. De aanwezigheid van markers in het genoom kan namelijk de fysiologie van de giststam beïnvloeden. Er was daarom een methode vereist om markers snel en efficiënt weer uit het genoom te verwijderen, zodat ze opnieuw konden worden gebruikt. Er waren maar enkele van de hiervoor benodigde contra-selecteerbare markers beschikbaar en bovendien was het proces van markerverwijdering uitermate tijdrovend. Als alternatief kan het zogenaamde LoxP /Cre-systeem worden gebruikt, maar dit systeem kan leiden tot chromosomale recombinatie en is daarom niet optimaal geschikt voor herhaald gebruik. In **hoofdstuk 4** is een methode onderzocht voor het tegelijkertijd verwijderen van meerdere markers met behulp van het meganuclease I-SceI. Door het markergen op de deletiecassettes te flankeren met I-SceI herkeningssequenties, kan de marker uit het genoom worden "geknipt". In hoofdstuk 4 worden nieuwe deletiecassettes beschreven, die zo zijn ontworpen dat na de "knip" de resulterend chromosomale uiteinden homoloog zijn aan elkaar. Hierdoor kan de ontstane breuk in het DNA eenvoudig worden gerepareerd door middel van homologe recombinatie. Deze strategie resulteerde in het tegelijkertijd verwijderen van drie markers met een efficiëntie van tot wel 56%. Doordat niet iedere deletieronde meer hoefde te worden gevolgd door een markerverwijderingsstap, versnelde deze nieuwe techniek het uitvoeren van achtereenvolgende deleties aanzienlijk. Daarnaast verruimt deze methode de

keuzemogelijkheid van markers, doordat de keuze van de marker niet langer beperkt wordt door de mogelijkheid om deze te contra-selecteren.

Hoofdstuk 5 beschrijft vervolgens het gebruik van de in de eerdere hoofdstukken beschreven methoden voor het construeren van een platform voor het wisselen van de gehele glycolyse in *S. cerevisiae*. Hiervoor werd eerst de I-SceI methode uit hoofdstuk 2 gebruikt om de geclusterde glycolyse te introduceren in de "minimale glycolyse" stam. Daarna werden, onder andere met behulp van de methode voor markerverwijdering die is beschreven in hoofdstuk 4, de 13 originele glycolytische genen verwijderd uit het genoom van de resulterende stam. Dit leverde een platformstam waarin alle glycolytische genen tot expressie komen vanaf een enkel chromosomaal locus. Deze stam werd vervolgens grondig onderzocht. De maximale specifieke groeisnelheid van de platformstam, waarin de genen voor de gehele glycolyse geclusterd waren in een enkel locus, was slechts een fractie lager dan die van de "minimale glycolyse" stam waarmee dit project werd gestart. Om te testen of dit platform inderdaad gebruikt kan worden voor het snel wisselen van de complete glycolyse, werd eerst een replica van de geclusterde *S. cerevisiae* glycolyse elders geïntroduceerd in het genoom, waarna de "originele geclusterde glycolyse" werd verwijderd met behulp van het CRISPR/CAS9 systeem. Dezelfde procedure werd gevolgd voor het introduceren van een 'minimale glycolyse' van de verwante gist *Saccharomyces kudriavzevii* en een "mozaïek-glycolyse" bestaande uit enzymen van *S. cerevisiae*, *S. kudriavzevii* en *Homo sapiens*. De verkregen stammen bleken levensvatbaar en zijn grondig geanalyseerd met behulp van 'next-gen sequencing'. Waar het 'verhuizen' van de geclusterde *S. cerevisiae*-glycolyse geen effect had op de maximale specifieke groeisnelheid, lag de maximale specifieke groeisnelheid van de *S. cerevisiae*-stam met de *S. kudriavzevii*-glycolyse beduidend lager. Overexpressie van *S. kudriavzevii TDH1* in deze stam leidde tot een toename van de maximale groeisnelheid tot een waarde die correspondeerde met die van de *S. cerevisiae*-variant. De maximale specifieke groeisnelheid van de stam met de mozaïek-glycolyse was, verrassenderwijs, gelijkwaardig aan die van de *S. cerevisiae*-variant. Deze resultaten tonen aan dat het mogelijk is om een essentiële stofwisselingsroute, die onontbeerlijk is voor de levensvatbaarheid van de cel, integraal te vervangen. Het in dit proefschrift beschreven onderzoek levert daarmee een basis voor een nieuwe, modulaire benadering van metabolic engineering, waarbij gehele routes snel kunnen worden vervangen en geoptimaliseerd. De mogelijkheid om de gistglycolyse snel te vervangen door alternatieve routes vormt bovendien een prachtig gereedschap voor fundamenteel onderzoek naar de regulatie van de glycolytische "flux". Dit kan op termijn bijdragen aan de verdere ontwikkeling van *S. cerevisiae* tot een nog efficiënter industrieel platform voor de productie van chemicaliën uit suikers. Het hier beschreven onderzoek levert daarnaast een illustratie van de razendsnelle ontwikkelingen en nieuwe mogelijkheden op het gebied van genetische modificatie.

Summary

For millennia mankind has used fermentation processes for the production of alcoholic beverages and dairy products like yogurt and cheese. In the past few decades, concerns on climate change have contributed to the development of biotechnological alternatives for petrochemical processes. In this way, scientists attempt to provide an alternative to the use of fossil fuels. Examples of this type of biotechnological processes are the production of bioethanol with yeast and the production of lactic acid with lactic acid bacteria, which is used as a starting compound for the production of the biodegradable plastic polylactic acid (PLA). In order to make these processes economically competitive, there has been an intensive search for improved, more efficient production organisms. In addition to classical breeding methods and random mutagenesis strategies, a more rational approach to obtain desired strain properties is by engineering well defined mutations in the DNA of micro-organisms. This "metabolic engineering" approach for genetic modification of microorganisms can result in higher product yields and more robust production strains. In the past decade, genetic engineering of microorganisms has undergone a tremendous development resulting in sometimes complete novel metabolic routes being engineered into an organism. A good example is the production of the malaria drug artimisinine, which is originally won from plants, but is now also produced by genetically modified yeast, enabling a cheaper process and a more consistent supply. Engineering such a drastic change in the natural properties of living cells is also referred to as "synthetic biology".

The large-scale modification of the genome of living cells is a major technical challenge, because interruption of the essential processes for survival of the cell would result in cell dead. As a consequence, modification of microorganisms has been limited to the removal of genes not essential for survival and to the introduction of additional metabolic pathways as for the previously mentioned artimisinine production. However, in order to further optimize product yields and production rates (fluxes), intervention in the central metabolism and central control processes of the cell is a prerequisite. The radical modification or replacement of a key route in metabolism requires a strategy in which the essential cellular function of this route is not interrupted while it is being replaced. In this thesis, such a strategy for radical reprogramming of metabolism is developed and applied to the sugar metabolism of baker's yeast (*Saccharomyces cerevisiae*).

The choice of *S. cerevisiae* as organism for this PhD project is obvious. *S. cerevisiae* is an important organism in fundamental research and is also a popular industrial platform for the biotechnological production of many sugar-derived chemicals. The widespread popularity of *S. cerevisiae* is partly due to the availability of a well-annotated genome sequence and to its excellent accessibility to genetic modification. This latter property is partly because this yeast is extremely efficient in "stitching" of DNA fragments by means of homologous recombination, which offers numerous possibilities for making genetic modifications and thus for the change of cellular processes.

As in almost all cells, the first part of sugar degradation in *S. cerevisiae* takes place through a series of concatenated enzymatic reactions that together are referred to as glycolysis. Glycolysis is present in almost every living cell and is one of the most

conserved metabolic pathways in life on earth. In glycolysis, glucose is oxidized in 10 enzymatic steps into two molecules of pyruvic acid, whereby energy is captured in the form of ATP. In addition, glycolysis also supplies many essential building blocks for making cell constituents via the intermediate metabolites in the pathway. The produced pyruvic acid can be further oxidized in the presence of oxygen via the citric acid cycle. In the absence of oxygen, pyruvate is decarboxylated by *S. cerevisiae,* and the resulting acetaldehyde is subsequently reduced to ethanol. In this thesis, it was chosen to also count the two enzymatic conversions of pyruvic acid in ethanol as glycolysis.

Despite the excellent accessibility of *S. cerevisiae* for genetic modification, the replacement of the whole glycolysis constitutes a huge challenge. In *S. cerevisiae,* the total of 12 cytosolic glycolytic reactions are catalyzed by 26 enzymes, which are encoded by 26 corresponding genes. This means that some conversions can be carried out by multiple enzymes. The first step towards a "synthetic glycolysis" is therefore ideally the removal of those excess iso-enzymes to reduce the complexity of the glycolysis to a single enzyme (and thus a single gene) per reaction. A successful effort to achieve this ambitious objective is described in the thesis of Daniel Solis Escalante. The "minimal glycolysis" engineered by him forms the basis of the research described in this thesis.

The aim of the present thesis is the construction of a *S. cerevisiae* strain in which the complete glycolysis can be replaced by a synthetic variant in a few simple steps. An important intermediate step towards this goal has been the clustering of the glycolytic genes in the "minimum glycolysis" strain at a single location in the yeast genome. The clustering of the whole glycolysis facilitates easy removal of the pathway in a later stage by a single transformation, because all genes are physically next to each other on the genome. In order to obtain a clustered glycolysis from the "minimum glycolysis" strain 13 genes had to be introduced into the genome and also the 13 original genes had to be deleted. Removal of these 13 genes within a short period of time represented a serious challenge. With the resulting platform the "glycolysis swapping" can be done, whereby a synthetic glycolysis is introduced into the cell, after which the "original" clustered glycolysis is removed in a single step.

The application of such a large number of changes in the yeast genome implied the exploration of a new scale of genetic modification. Evaluation of the available techniques at the start of the project led to the conclusion that those techniques were not sufficient to achieve these ambitious goals within a time-frame of four years. Classical cloning techniques were not suited for the introduction of 13 genes in a single transformation. The assembly of 13 genes in a plasmid by restriction and ligation is after all a repetitive time-consuming process which is, moreover, limited by the availability of unique restriction sequences in the inevitable large DNA constructs.

In the past decade, alternative methods have been developed for the simultaneous assembly of multiple DNA fragments. A very promising technique is *in vivo* assembly of DNA fragments by homologous recombination in *S. cerevisiae*. This method, also known as "TAR cloning," was used by Dan Gibson and his team for the assembly of the first complete synthetic genome in 2010. In this method, overlapping DNA fragments

are transformed to *S. cerevisisae*, in which the fragments are then assembled into a circular plasmid by homologous recombination of the overlapping ends. Although this technique was quickly picked up by other laboratories, the reported efficiencies of correct assembly were not consistent and standardization of this method has not been thoroughly investigated. In **Chapter 2** it is therefore investigated how the *in vivo* assembly of plasmids from overlapping DNA fragments can be optimized. Based on this study two significant improvements have been implemented to the method: i) the marker gene and sequences for replication of the plasmid were placed on different fragments and ii) specially designed synthetic DNA sequences of 60 base pairs (referred to as SHR-sequences) were used for homologous recombination. These modifications resulted in a one hundred-fold decrease in the number of incorrect assembled plasmids, because self-closure of the receiving plasmid-backbone was complicated. Also, implementation of the standardized SHR-sequences added flexibility to the system by making fragments interchangeable. To test these improvements experimentally, the optimized method was used to assemble six glycolytic genes into a plasmid of 21 kb in a single transformation. Based on the number of correctly assembled plasmids, this resulted in an efficiency of 95%. Determination of the DNA sequence of one of the assembled plasmids proved that the obtained plasmid corresponded to the *in-silico* design. Complementation studies demonstrated that all of the six glycolytic genes on the plasmid were functionally expressed in *S. cerevisiae*. This method offered, in the context of this thesis, a solution for the efficient assembly of the 13 DNA fragments, which are required to engineer a clustered glycolysis. In a more general sense, the presented work shows that *S. cerevisiae* can be used as an universal platform for rapid and accurate assembly of large plasmids.

Although the *in vivo* assembly of DNA fragments in *S. cerevisiae* offered a solution for the introduction of a complete metabolic route in a single transformation, the product obtained is always a plasmid. Plasmids, however, are not stable during replication and without selective pressure plasmids disappear gradually from a population. Therefore, expression of genes from the chromosome is usually preferred over plasmid-based gene expression. In **Chapter 3** it is investigated whether *in vivo* assembly of multiple DNA fragments could be combined with integration of the assembled construct at a specific location in the genome. In this case, the overlapping DNA fragments are not assembled into a circular plasmid, but into a linear DNA cassette, of which the ends are homologous to the DNA sequence of the location on the genome where the cassette should eventually integrate (such as is the case for a conventional integration cassette). This strategy has been applied for the simultaneous *in vivo* assembly and integration of seven glycolytic genes at a specific chromosomal locus. While successful, the efficiency of this strategy, measured by the number of correct transformants, was only 5%. Therefore it was investigated whether the addition of a targeted double-stranded DNA break at the intended place of integration in the genome with the aid of the meganuclease I-SceI was able to increase the effectiveness and accuracy of this strategy. The I-SceI facilitated integration of the same set of seven glycolytic genes resulted in a dramatic increase in the number of transformants obtained. Furthermore, 95% of the analysed transformants contained the intended configuration of the seven genes at the targeted site in the yeast genome. This result clearly demonstrated

that combined *in vivo* assembly and targeted integration of the genes of a complete metabolic pathway, the so-called CATI approach, is an excellent method for large and complex metabolic engineering projects in yeast.

In addition to the rapid and accurate introduction of a complete set of glycolytic genes, the intended platform required also an approach for the removal of the 13 native glycolytic genes from the genome. A major obstacle in removing a large number of genes is the need for a variety of selectable genetic markers. Although there are quite a number of markers for *S. cerevisiae* available, it is not desirable to use a new marker for each gene deletion, because the presence of markers in the genome can affect the physiology of the yeast strain. Therefore a method was required for quick and efficient removal of markers from the genome. The required counter-selectable markers were limited in number and moreover the process of marker removal was extremely time consuming. Alternatively, the so-called loxP / Cre-system could be used, but this system might result in chromosomal recombinations and is therefore not optimal for repeated use. In **Chapter 4**, a method has been studied for simultaneous removal of multiple markers, using the meganuclease I-SceI. By flanking the marker gene on the deletion cassettes with I-SceI recognition sequences, the marker can be "cut" from the genome. Chapter 4 describes those new deletion cassettes, which are designed in such way that after the "cut" the resulting chromosomal ends are homologous to each other. As a result, the break in the DNA can be easily repaired by homologous recombination. This strategy resulted in the simultaneous removal of three markers with an efficiency of up to 56%. Since no longer every deletion round had to be followed by a marker removal step, this new technique accelerated the process of successive deletions considerably. In addition, this method expands the choice of markers, since the choice of marker is no longer limited by the ability to counter-select.

Finally in **Chapter 5** it is described how the methods discussed in the previous chapters are used for the construction of a platform for the exchange of the complete glycolysis in *S. cerevisiae*. In order to achieve this, first the I-SceI method of Chapter 2 was used to introduce the clustered glycolysis in the "minimum glycolysis" strain. Subsequently, with the use of the method for marker removal as described in Chapter 4, the 13 original glycolytic genes were deleted from the genome of the resulting strain. This led to the platform strain in which all the glycolytic genes were expressed from a single chromosomal locus. This strain was then thoroughly investigated. The maximum specific growth rate of the platform strain, in which the genes for the whole glycolysis were clustered in a single locus, was only slightly lower than that of the "minimum glycolysis" strain which was the starting strain of this project. To test whether this platform could indeed be used for a quick exchange of the entire glycolysis, a replica of the clustered *S. cerevisiae* glycolysis was introduced elsewhere into the genome, after which the 'original clustered glycolysis' was removed with the aid of the CRISPR / CAS9 system. The same procedure was followed for the introduction of a "minimal glycolysis' of the cognate yeast *Saccharomyces kudriavzevii* and a "mosaic-glycolysis" consisting of enzymes of *S. cerevisiae*, S. *kudriavzevii* and *Homo sapiens*. The resulting strains proved to be viable and were thoroughly analyzed using next-gen sequencing. The 'relocation' of the clustered *S. cerevisiae* glycolysis had no effect on

the maximum specific growth rate, while the maximum specific growth rate of the *S. cerevisiae* strain with the *S. kudriavzevii*-glycolysis turned out to be significantly lower. Overexpression of *S. kudriavzevii TDH1* in this strain resulted in an increase in the maximum specific growth rate to a value that corresponded to that of the *S. cerevisiae*-variant. The maximum specific growth rate of the strain with the mosaic-glycolysis was, surprisingly, equivalent to that of the *S. cerevisiae*-variant. These results show that it is possible to integrally replace an essential metabolic pathway, which is essential for the viability of the cell. The research described in this thesis therefore provides a basis for a new, modular approach for metabolic engineering, in which entire routes can be quickly replaced and optimized. The ability to quickly replace the yeast glycolysis by alternative routes supplies also a wonderful tool for fundamental research into the regulation of the glycolytic "flux". This may eventually contribute to the further development of *S. cerevisiae* to a more efficient industrial platform for the production of chemicals from sugars. In addition, the present study provides an illustration of the rapid developments and new possibilities in the field of genetic modification.

Curriculum vitae

NIELS GERARD ADRIAAN KUIJPERS

Niels Gerard Adriaan Kuijpers was born on January 25th , 1984 in Apeldoorn (The Netherlands). There he attended pre-University education at the Stedelijk Gymnasium Apeldoorn, where he graduated in 2002. Afterwards Niels enrolled in the joined Life Science & Technology program of Delft University of Technology and Leiden University. Beside the Life Science & Technology study program he was active within several student committees and organizations. So did he participate in the organizing committee of a study tour to Denmark and Sweden and was for a year board member of internship agency Integrand Delft. After getting his Bachelor's degree, Niels continued at Delft University by taking on the Life Science & Technology Master's track "Cells as Factory" in 2007. During the same year he went with three fellow-students for 3 months to Quelimane, (Mozambique) for an internship investigating the production of bio-diesel with algae. Niels finished his Master's in 2010 with a research project performed in the Industrial Microbiology Section of Delft University of Technology of Professor Jack Pronk, focusing on lactic acid production with genetically engineered yeast.
Intrigued by genomics and industrial fermentation, he proceeded in the Industrial Microbiology Section as a PhD student within the VIDI-project of Pascale Daran-Lapujade, which had as aim to engineer a synthetic glycolysis in yeast. As from September 2014, Niels is working as Scientist Fermentation & Yeast at HEINEKEN.

LIST OF PUBLICATIONS

1. Solis-Escalante D, **Kuijpers NGA**, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P: amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS yeast research* 2013, 13(1):126-139.

2. **Kuijpers NGA**, Chroumpi S, Vos T, Solis-Escalante D, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P: One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS yeast research* 2013, 13(8):769-781.

3. **Kuijpers NGA**, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P: A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microbial Cell Factories* 2013, 12:47.

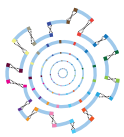
4. Solis-Escalante D, **Kuijpers NGA**, van der Linden FH, Pronk JT, Daran JM, Daran-Lapujade P: Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in *Saccharomyces cerevisiae*. *FEMS yeast research* 2014, 14(5):741-754.

5. Solis-Escalante D, van den Broek M, **Kuijpers NGA**, Pronk JT, Boles E, Daran JM, Daran-Lapujade P: The genome sequence of the popular hexose-transport-deficient *Saccharomyces cerevisiae strain* EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss. *FEMS yeast research* 2015, 15(2).

6. Mans R, van Rossum HM, Wijsman M, Backx A, **Kuijpers NGA**, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJ, Daran JM: CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS yeast research* 2015, 15(2).

7. Solis-Escalante D, **Kuijpers NGA**, Barrajon-Simancas N, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P: A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryotic cell* 2015, 14(8):804-816.

8. **Kuijpers NGA**, Solis-Escalante D, Luttik MAH, Pronk JT, Daran JM, Daran-Lapujade P: Pathway swapping: a new approach to radically remodel essential, multigene cellular processes. *Submitted*



Dankwoord / Acknowledgements

This thesis is the final product of a great time and great adventure for over a decade in Delft. My gratitude goes to a large number of people of which I will certainly forget some to mention in the next paragraphs. My apologies to all of those.

Jack, thank you for letting me part of the Industrial Microbiology group and for your always positive guidance and great help in my personal development. Pascale, I am so happy that you trusted me to take on this crazy project with you. Thank you for your open attitude, patience and guidance, it was a fantastic journey. You were a fantastic supervisor, merci beaucoup! Lots of gratitude go to the other members of the VIDI team: Jean-Marc, thank you for introducing me in the world of genetics, it was great! Prof. Carlos and Prof. Eckhard, it was a honour to have you on the team as advisors, the yearly meetings were definitely highlights during my PhD project. No synthetic glycolytic pathways without a minimal glycolysis. Daniel, you were the best colleague I could have wished for, thank you for all the moments of fun and the great cooperation. Lizanne & Nurnia, your contributions were essential, thanks for all your input and great team spirit.

Erik, Marijke, and Marinka, I can't imagine how the lab and my lab skills would have been without you. Marijke, I really hope you will never have to go through so many enzyme assays again, you were fantastic and should be in the Guinness Book of records for this.

There is nothing so inspiring as teaching and I want to thank Ton for involving me in the practical fermentation course and for linking me to the most fantastic students who took part in this project: Tim, Matthijs, Tania, Mehek and Rudy, you were the best! Thanks for keeping me sharp and detecting when coffee was essential. In that respect, I want to thank Stefan for the most perfect coffee protocol in history and for the many coffee breaks we had.

Tim and Matthijs, thanks for being my paranimfs. I have been lucky that you both continued in IMB after your large contribution to this thesis. Tim, you were always there to discuss the broader aspects of science and you were always enthusiastic and full of confidence, thanks! Matthijs, you were an example for me on organization and planning and still always in for a coffee, some nice Ajax talks and legendary lab cleanings with Robert, thanks for everything!

No DNA assembly without reading it. Marcel, Jurgen, Dick, and Alexy, without you I would still be lost in all those datasets. Thanks for giving me the source code of yeast!

Keeping the speed in such a large project is only possible with such dedicated people taking care of all the media, chemicals, autoclaves and many more. Apilena, Astrid and Jannie thanks for your devotion, help on last minute requests and for being great persons to work with!

The secret ingredient of a PhD thesis is the environment you have to work in. And that environment was great due to you all:

Roomies Barbara , Irina, Mark, Pilar and Robert, you made our aquarium the best office of the lab! Thanks to you I know where to shop, how to conquer the world, how to handle cats, how to speak Spanish and how to never trust your computer mouse…

Frank, Susan and Jules, exploring the molecular techniques with you was inspiring and fun! Harmen, Barlimon's best was incredible, and so was our time together in the lab. Many thanks for the greatest assay ever 'on the history of plasmids', I am sure you will recognize your hand in the introduction of this thesis. Dani, thanks for all the laughs and for teaching me how to express frustrations in the Spanish language. Bart, AJAX! GabrieLe, (your name spelled correct at last ;-) grazi!), Victor, Chi Chi Chi, Le Le Le, Viva Chile! Gracias! Nick, great scientists have a great moustache, Thx! Lots of thanks to all the senior PhD's for guiding me in the first steps of this big project: Rintze, Eline, Leonie, Stefan and Tania, you were great examples. I also would like to thank all PostDocs: Angel, Benjamin, Filipa, Ishwar, Thiago, Wouter and Yulia, thx for all the discussions and fun times.

Alle leden van de feestcommissie, dank voor het mij laten rondrennen in een roze konijnenpak door het Kluyver en voor vele andere essentiële bezigheden! Jullie enthousiasme heeft zeker bijgedragen aan de geweldige sfeer in het gebouw, dank!

Finishing your thesis while being in your new job can be challenging. Thanks Jan-Maarten, Tom and all other colleagues at HEINEKEN for your patience and support.

Zonder vrienden kom je nergens, Mozamboys, dank voor jullie zetje in deze richting en voor een geweldige studietijd.

Dr. Prutser, 13 jaar hebben we in hetzelfde schuitje gezeten, zonder jou had ik nooit: mijn practica zo snel gehaald, de versheid van een ei kunnen meten met een zaklamp, fysische transportverschijnselen kunnen maken, hoofdpijn gehad bij het tentamen statistische thermodynamica, zoveel koffie gedronken, posters kunnen maken, succesvol microarrays kunnen doen, zoveel naar Den Haag gefietst, de meeuwen makrelen loop gelopen, zoveel lol gehad in het maken van deze thesis. Buddies for Life indeed ;-) , heel veel dank!

Twaalf reacties, 999 mogelijkheden, boys van 9B, dank voor jullie vriendschap en steun.

Pa en ma, zonder jullie was ik hier geeneens aan begonnen, dank voor jullie steun en toewijding de afgelopen 32 jaar! Nerd, ook jij bedankt voor jouw support in het zetten van deze stap.

Fransje, het mooiste resultaat behaalde ik aan de start, bedankt voor dat smsje op mijn eerste dag als promovendus. En uiteraard voor alle steun, toewijding en geduld van de afgelopen jaren…