Institut für Biochemie und Biologie
Nachwuchsgruppe Cell2Fab, AG Molekularbiologie

Development of a new DNA-assembly method and its application for the establishment of a red light-sensing regulation system

Kumulative Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Molekularbiologie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von Lena Hochrein

Potsdam, den 25.04.2017

TABLE OF CONTENTS

1 INDEX OF ABBREVIATIONS 1

2 SUMMARY 3

3 INTRODUCTION 6

3.1 From Cell to Fabrication 6

3.2 Cloning Strategies for Synthetic Biology – Stitching and Patching 8

3.3 Genetic Engineering – Paving the Way to Efficient Fabrication 12

3.4 Transcription Factors – Workhorses of Synthetic Biology 15

3.5 Light-Responsive Systems – Advanced Process Control 18

3.6 Goals of this Study 19

4 MANUSCRIPTS 21

4.1 AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies 21

4.2 PhiReX – a Programmable and Red Light-regulated Protein Expression Switch for Yeast 51

4.3 L-SCRaMbLE: a Tool for Light-controlled Cre-mediated Recombination in Yeast 90

5 DISCUSSION 120

6 REFERENCES 131

I. CURRICULUM VITAE 142

II. SCIENTIFIC ACHIEVEMENTS 143

PUBLICATIONS 143

CONFERENCE CONTRIBUTIONS 143

III. DECLARATIONS 144

DECLARATION OF INDEPENDENT WORK 144

DECLARATION OF OWN CONTRIBUTIONS TO SINGLE PUBLICATIONS 145

IV. ACKNOWLEDGEMENTS 146

1 INDEX OF ABBREVIATIONS

aa	Amino acids
AD	Activation domain
aTc	Anhydrotetracycline
bp	Base pairs
BR	Biliverdin reductase
Cas	CRISPR-associated
CCA	Circadian clock-associated protein
CDS	Coding DNA sequence
CPEC	Circular Polymerase Extension Cloning
CRISPR	Clustered regularly interspaced short palindromic repeats
DBD	DNA-binding domain
DBS	DNA-binding site
dCas9	Dead Cas9
DNA	Deoxyribonucleic acid
ds	Double-strand
dNTP	Deoxynucleotide triphosphate
E. coli	*Escherichia coli*
GOI	Gene of interest
HO	Heme oxygenase
IPTG	Isopropyl-β-D-thiogalactopyranoside
kbp	Kilo base pairs
kDa	Kilo Dalton
LCR	Ligase Cycling Reaction
LHY	Late elongated hypocotyl
LIC	Ligase Independent Cloning
Mb	Megabase pair
M. genitalium	*Mycoplasma genitalium*
M. mycoides	*Mycoplasma mycoides*
MODAL	Modular Overlap-Directed Assembly with Linkers
mRNA	Messenger RNA

NES	Nuclear export signal
NLS	Nuclear localization signal
nt	Nucleotide
OE-PCR	Overlap extension PCR
PAM	Protospacer adjacent motif
PCB	Phycocyanobilin
PCR	Polymerase chain reaction
Phy	Phytochrome
PIC	Preinitiation complex
PIF	Phytochrome interacting factor
PΦB	Phytochromobilin
RNA	Ribonucleic acid
RNAPII	RNA-polymerase II
RVD	Repeat variable diresidue
S. cerevisiae	*Saccharomyces cerevisiae*
SCRaMbLE	Synthetic Chromosome Rearrangement and Modification by *LoxPsym*-mediated Evolution
sgRNA	Single-guide RNA
SLIC	Sequence- and ligation-independent cloning
SLiCE	Seamless Ligation Cloning Extract
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
TAR	Transformation-associated recombination
tetO	Tetracycline operator
tetR	Tetracycline repressor
TF	Transcription factor
TU	Transcriptional unit
sp.	Species
yGG	Yeast Golden Gate
ZFN	Zinc finger nuclease
ZFP	Zinc finger protein

2 SUMMARY

With *Saccharomyces cerevisiae* being a commonly used host organism for synthetic biology and biotechnology approaches, the work presented here aims at the development of novel tools to improve and facilitate pathway engineering and heterologous protein production in yeast. Initially, the multi-part assembly strategy AssemblX was established, which allows the fast, user-friendly and highly efficient construction of up to 25 units, e.g. genes, into a single DNA construct. To speed up complex assembly projects, starting from sub-gene fragments and resulting in mini-chromosome sized constructs, AssemblX follows a level-based approach: Level 0 stands for the assembly of genes from multiple sub-gene fragments; Level 1 for the combination of up to five Level 0 units into one Level 1 module; Level 2 for linkages of up to five Level 1 modules into one Level 2 module. This way, all Level 0 and subsequently all Level 1 assemblies can be carried out simultaneously. Individually planned, overlap-based Level 0 assemblies enable scar-free and sequence-independent assemblies of transcriptional units, without limitations in fragment number, size or content. Level 1 and Level 2 assemblies, which are carried out via predefined, computationally optimized homology regions, follow a standardized, highly efficient and PCR-free scheme. AssemblX follows a virtually sequence-independent scheme with no need for time-consuming domestication of assembly parts. To minimize the risk of human error and to facilitate the planning of assembly projects, especially for individually designed Level 0 constructs, the whole AssemblX process is accompanied by a user-friendly webtool. This webtool provides the user with an easy-to-use operating surface and returns a bench-protocol including all cloning steps. The efficiency of the assembly process is further boosted through the implementation of different features, e.g. *ccdB* counter selection and marker switching/reconstitution. Due to the design of homology regions and vector backbones the user can flexibly choose between various overlap-based cloning methods, enabling cost-efficient assemblies which can be carried out either in *E. coli* or yeast. Protein production in yeast is additionally supported by a characterized library of 40 constitutive promoters, fully integrated into the AssemblX toolbox. This provides the user with a starting point for protein balancing and pathway

engineering. Furthermore, the final assembly cassette can be subcloned into any vector, giving the user the flexibility to transfer the individual construct into any host organism different from yeast.

As successful production of heterologous compounds generally requires a precise adjustment of protein levels or even manipulation of the host genome to e.g. inhibit unwanted feedback regulations, the optogenetic transcriptional regulation tool PhiReX was designed. In recent years, light induction was reported to enable easy, reversible, fast, non-toxic and nearly gratuitous regulation, thereby providing manifold advantages compared to conventional chemical inducers. The optogenetic interface established in this study is based on the photoreceptor PhyB and its interacting protein PIF3. Both proteins, derived from *Arabidopsis thaliana,* dimerize in a red/far-red light-responsive manner. This interaction depends on a chromophore, naturally not available in yeast. By fusing split proteins to both components of the optical dimerizer, active enzymes can be reconstituted in a light-dependent manner. For the construction of the red/far-red light sensing gene expression system PhiReX, a customizable synTALE-DNA binding domain was fused to PhyB, and a VP64 activation domain to PIF3. The synTALE-based transcription factor allows programmable targeting of any desired promoter region. The first, plasmid-based PhiReX version mediates chromophore- and light-dependent expression of the reporter gene, but required further optimization regarding its robustness, basal expression and maximum output. This was achieved by genome-integration of the optical regulator pair, by cloning the reporter cassette on a high-copy plasmid and by additional molecular modifications of the fusion proteins regarding their cellular localization. In combination, this results in a robust and efficient activation of cells over an incubation time of at least 48 h. Finally, to boost the potential of PhiReX for biotechnological applications, yeast was engineered to produce the chromophore. This overcomes the need to supply the expensive and photo-labile compound exogenously. The expression output mediated through PhiReX is comparable to the strong constitutive yeast *TDH3* promoter and - in the experiments described here - clearly exceeds the commonly used galactose inducible *GAL1* promoter.

The fast-developing field of synthetic biology enables the construction of complete synthetic genomes. The upcoming Synthetic Yeast Sc2.0 Project is currently underway

to redesign and synthesize the *S. cerevisiae* genome. As a prerequisite for the so-called "SCRaMbLE" system, all Sc2.0 chromosomes incorporate symmetrical target sites for Cre recombinase (*loxPsym* sites), enabling rearrangement of the yeast genome after induction of Cre with the toxic hormonal substance β-estradiol. To overcome the safety concern linked to the use of β-estradiol, a red light-inducible Cre recombinase, dubbed L-SCRaMbLE, was established in this study. L-SCRaMbLE was demonstrated to allow a time- and chromophore-dependent recombination with reliable off-states when applied to a plasmid containing four genes of the β-carotene pathway, each flanked with *loxPsym* sites. When directly compared to the original induction system, L-SCRaMbLE generates a larger variety of recombination events and lower basal activity. In conclusion, L-SCRaMbLE provides a promising and powerful tool for genome rearrangement.

The three tools developed in this study provide so far unmatched possibilities to tackle complex synthetic biology projects in yeast by addressing three different stages: fast and reliable biosynthetic pathway assembly; highly specific, orthogonal gene regulation; and tightly controlled synthetic evolution of *loxPsym*-containing DNA constructs.

3 INTRODUCTION

The American theoretical physicist Richard Phillips Feynman once said: "What I cannot create I do not understand." To be quite honest, synthetic biology made me realize, even understanding what you created can propose serious challenges.

In 2010, synthetic biology caused a stir around the world when researchers of the J. Craig Venter Institute synthesized the *Mycoplasma mycoides* genome and created a synthetic living cell by transplanting the 1 Mb synthetic chromosome into a bacterial host[1]. However, with the book entitled “La Biologie Synthétique”[2], Stephane Leduc first mentioned this discipline long before, in 1912. Since then, the field has been drawing increased attention with various research highlights like an expanded genetic code as a basis to produce new synthetic amino acids (aa) and proteins for industrial or medical applications[3-5]. By engineering photoautotrophic cyanobacteria to produce and secrete sugar, which was in turn used as sole carbon source for growth of *Escherichia coli,* the lab of Pamela Silver opened a new path towards dramatically decreased fermentation cost[6]. Two other groundbreaking studies were accomplished by the lab of Martin Fussenegger: the report of a blue light-inducible penile erection[7] and a mind-controlled gene network[8].

These examples make one realize that there is no ceiling to the imagination of scientists, understandably inflaming serious ethical discussions about weighing the benefits against the risks and harms of this new scientific discipline[9]. This debate recently reached a climax, with the upcoming idea to synthesize a human genome[10, 11].

3.1 From Cell to Fabrication

The aim of synthetic biology is to gain a deeper understanding of cellular processes by the reconstruction and redesign of natural biological systems. The acquired knowledge is then used to develop new artificial components, regulatory networks and biosynthetic pathways to improve their native counterparts. The establishment of manifold well characterized devices allows their fast and easy rewiring, to e.g. create large artificial circuits for integration in heterologous organisms and/or to control cellular processes by

manipulating the host genome. The Tet On/Tet Off system[12], which is based on a combination of hybrid transcription factors (TFs), is a good example for an artificial and highly regulated expression system, which can be part of larger regulatory networks. A second impressive example is the so-called "repressilator", constructed from artificial transcriptional repressor systems, to yield a time-dependent genetic oscillator as a synthetic biological clock[13]. Moreover, a genetic toggle switch was built based on two differentially induced repressor proteins to allow the easy swap between two stable states[14]. While the original system is based on the chemical inducers isopropyl-β-D-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc), a newly developed toggle switch is responsive to light pulses of two different wavelengths[15]. Finally, the construction of logic gates[16, 17] allows the integration of two signals into a distinct cellular response, e.g. by AND (two signals are required for a response), OR (one of two signals is required for a response) and NAND (two signals must be present to suppress a response) gates.

Practical applications of synthetic biology demonstrate the overlap with classical biotechnology and allow for the possibility to circumvent time-consuming or cost-intensive chemical synthesis methods. Examples are the heterologous production of the antimalarial drug precursor artemisinic acid[18], the taxol precursor taxadiene[19] or biofuels[20-22] in *Saccharomyces cerevisiae*.

The selection of the host organism is an important aspect for large-scale protein production, as this requires a highly efficient "factory" enabling cheap, fast and reliable "fabrication". The budding yeast *S. cerevisiae* is an often used host for synthetic biology purposes, as it is the best studied eukaryotic model organism, has a short generation time of 90 min in rich medium and can grow to very high cell densities[23, 24]. Moreover, its cultivation and manipulation is easy, cheap and well established. A large collection of characterized yeast promoters covers a wide range of expression levels, which is advantageous for the efficient engineering of biosynthetic pathways[25, 26]. The availability of strong constitutive and inducible promoters allows high-level production of heterologous genes. As an essential footing for genomic manipulation, the sequence of the whole yeast genome is available since 1996[27]. Beyond that, the international Synthetic Yeast Project (Sc2.0) is underway to redesign and fully synthesize all

16 chromosomes of the budding yeast *S. cerevisiae* in order to create the first artificial eukaryotic genome[28-34]. This provides the opportunity to answer profound questions about the structure and content of eukaryotic genomes and cellular processes.

3.2 Cloning Strategies for Synthetic Biology – Stitching and Patching

While the cloning of a single gene could make up large parts of a PhD thesis just a few decades ago, nowadays next generation DNA assembly strategies enable easy and fast cloning of complex biosynthetic pathways or even whole genomes. Although performance and cost efficiency of gene synthesis rapidly increased over the last years, the size of directly synthetized DNA sequences is still limited. This constitutes the requirement for reliable and highly efficient DNA assembly strategies to generate large and complex multi-gene constructs, even if single parts are synthesized. In the last decades, several techniques simplifying complex DNA assembly projects were reported. One of those, dubbed BioBrick[35], relies on conventional cloning methods based on endonucleases and ligases. BioBrick makes use of four 6-base cutters flanking the BioBrick parts in a given order to allow repeatable assembly rounds. In each step, two building blocks and a linearized target vector are ligated and, due to compatible overhangs, result in one plasmid containing two parts. The BioBrick procedure is not reversible and limited to a step-by-step cloning process, as the restriction sites downstream of the first part and upstream of the second part are deleted through ligation. However, the resulting, larger BioBrick fragment is again flanked by restriction sites allowing for repeatable assembly steps. Nevertheless, the method is not well suited for complex multi-gene assemblies. Another drawback of sequence dependent cloning methods like BioBrick is the requirement to domesticate assembly parts before usage. All restriction sites required for the respective cloning method need to be deleted or mutated in every assembly part prior to cloning. As a target sequence for a 6-base cutter statistically appears every 4 kbp, it is likely that a "forbidden" recognition site occurs in a given assembly part. Another downside is the occurrence of left-over scar sequences at the junctions between the assembled parts, which may affect the way parts interact. This aspect is particularly important for assembly parts acting on sub-gene level, e.g. promoters, CDS and terminators.

Methods like Golden Gate cloning[36] and MoClo[37] employ type IIS restriction enzymes. These endonucleases cut the DNA sequence several bases away from their recognition site. In this way, a single enzyme can generate different overhangs. By exploiting this special feature, the number of required enzymes and consequently the number of "forbidden" DNA-sequences can be minimized. Moreover, overhangs generated by type IIS restriction enzymes can be designed in a way to reduce interference with biological functions. Efficiencies of 95-100 % were reported for a one-pot assembly of ten pieces by Golden Gate cloning[36]. MoClo[37], an extension of Golden Gate, which is specialized on predefined gene assemblies, enabled an efficient assembly of eleven transcriptional units (TUs) consisting of 44 subunits. Nevertheless, the definition of the parts' position in the final assembly requires the time-consuming subcloning of each sub-gene fragment into special positioning vectors. Besides, the need for domestication of parts and the non-reversible assembly character remain, as restriction sites are lost through the assembly process. On the contrary, the loss of restriction sites enables a highly efficient and fast assembly strategy based on repetitive ligation-restriction cycles. The iBrick[38] assembly strategy reduces the likelihood of "forbidden" recognition sites by using homing endonucleases (also called meganucleases), which have large, asymmetric recognition sites (14 to 40 bp)[39]. Although these enzymes tolerate single base changes within their recognition site, the use of homing endonucleases virtually eliminates the risk of unintended cut sites within a sequence of interest. This practically excludes the need for parts domestications. However, scar sequences are still present in the assembly product and are much longer than the 4-bp scars produced by type IIS restriction enzymes.

Sequence independent cloning methods can overcome some drawbacks of restriction enzyme-based cloning methods. These methods allow the scar-free assembly of multiple DNA fragments in a predetermined order. Each assembly part is flanked by homology regions providing overlap to the neighboring DNA fragment or the vector backbone. During the assembly process, the overlaps are combined either by PCR or by recombination, complicating the usage of these methods for the assembly of repetitive sequences. One such method is Overlap Extension Polymerase Chain Reaction (OE-PCR)[40, 41] which allows PCR-based cloning of one fragment into a vector backbone. To

amplify inserts, primers giving homology to the backbone sequence at the 5' end and annealing to the insert sequence at the 3' end are used. Afterwards, the amplified fragment is used as a "mega-primer" for the amplification of the vector backbone. A variation of OE-PCR, called Circular Polymerase Extension Cloning (CPEC)[42], was employed for the successful assembly of up to four fragments. The polymerase independent Ligase Cycling Reaction (LCR)[43] eliminates the risk of PCR-born errors. LCR uses single-stranded bridging oligonucleotides with homology to the neighboring DNA fragments, which are then covalently linked by a thermostable ligase and assembled via multiple denaturation-annealing-ligation steps. LCR was announced to achieve a one-step assembly of 20 parts. An important group of recombination based, sequence independent cloning methods are chew back and anneal methods. These rely on 3'-5' DNA exonuclease activity to create single stranded overhangs, which are then suitable for annealing to neighboring assembly fragments sharing homology regions. Two methods that make use of 3'-5' exonuclease activity of T4 DNA polymerase are the early developed Ligase Independent Cloning (LIC)[44] and the later established Sequence- and Ligation-Independent Cloning method SLIC[45]. As a limiting step for exonuclease activity, LIC uses only one dNTP type in the reaction mix and incorporates homology regions lacking this specific dNTP. Thus, T4 DNA polymerase chews back double-stranded DNA until the first nucleotide (nt) matching the dNTP in the reaction mix is reached. This changes exonuclease activity to polymerase function and stalls the reaction at this position; neither further chew back nor further polymerase activity can occur. SLIC circumvents the design requirements of the LIC overlap regions by addition of one dNTP type after a reaction time sufficiently long for the generation of adequate single stranded overlaps. Gibson DNA assembly[46, 47] is another very powerful method, which uses an optimized enzyme mixture consisting of a heat labile T5 exonuclease to obviate the dNTP addition step, Phusion® DNA polymerase to fill in the gaps, and a thermostable ligase to link single-stranded overhangs. The SLiCE method (Seamless Ligation Cloning Extract)[48] circumvents the supply of expensive enzymes by using raw *E. coli* cell extract for *in vitro* assembly of up to six parts. Yet another cost-efficient method, Transformation-Associated Recombination (TAR) cloning[49], uses *in vivo* homologous recombination mechanisms of *S. cerevisiae* for the assembly of a large number of fragments. Daniel Gibson and his colleagues used this method for

synthesizing the *Mycoplasma genitalium* genome[50] and further demonstrated that yeast is able to uptake and recombine 38 overlapping single-stranded oligonucleotides in one step[51]. To decrease costs for oligonucleotide synthesis, required for the insertion of overlaps to assembly fragments, companies like Clontech and New England Biolabs developed highly efficient assembly methods tolerating shorter overlap regions. Clontech developed the In-Fusion® cloning and NEB the NEBuilder® HiFi DNA Assembly. Both methods require the same DNA starting materials like SLIC[52], SLiCE[48] or Gibson assembly[46] and yield in the same final product with the main difference that overlap sequences of only 15 bp are sufficient for efficient assemblies.

Most of the recombination-based methods described above, enable efficient one step assemblies of multiple parts and are indispensable for many synthetic biology projects. Nevertheless, the individually designed cloning process neglects two important principles of synthetic biology: standardization[53] and modularity[54]. The methods permit individual assemblies but also demand time consuming planning and individual designs of single assembly steps. Parts used in these assemblies are not reusable in different sequence contexts and the methods do not directly support iterative assemblies. For this reason, many higher-order assembly strategies have been developed, mostly based on the above presented cloning methods. An example for this is the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy[55], which uses two different adapter regions fused to the 3' and 5' ends of DNA fragments. These adapter sequences allow the standardized PCR-based addition of homology regions for directed and overlap-based assembly, thereby supporting easy shuffling and modular assembly of adapter equipped parts in different order and orientation. To enhance the cloning efficiency, homology regions used for MODAL are optimized by the R2oDNA designer software[56] with respect to orthogonality to the host organism and among each other, guanine-cytosine content and secondary structures. Despite the reported advantages of MODAL, the user must accept large scar-sequences, flanking each assembly part and the risk of unintended recombination events due to repetitive sequences. Yeast Golden Gate (yGG)[57] is an adaption of the Golden Gate cloning method[36], optimized for the assembly of TUs in *S. cerevisiae*. The VEGAS system[58] uses TUs pre-assembled by yGG[57] and adds so-called "VEGAS adapters" to direct the homology-based assembly into multi-

gene constructs. Flexibility with respect to the order of the parts is gained by simply adding new homology regions to yGG-assembled TUs via PCR amplification. This imposes the risk of PCR-born errors, especially for large DNA fragments. To avoid this risk and to circumvent the need for mutagenesis of internal restriction sites, PaperClip[59], which generates "Clips" in a PCR-free manner to link two assembly parts previously lacking any end homology, was established. To this end, two double-stranded oligonucleotides (half-Clips) are generated, each with homology to one of the assembly parts. To generate a full-Clip, giving homology to both assembly parts, two half-Clips are ligated via a predesigned single-stranded 3-nt overhang. The assembled Clip can then be used in any overlap-based cloning strategy to stitch two fragments together. If an assembly part is shared in different projects, the possibility to reuse half-Clips reduces primer costs.

The collection of many different assembly methods gives researchers the possibility to choose from a well-stocked toolbox and to apply and combine different strategies to tackle specific challenges. However, most of the described methods still require a well-established synthetic biology laboratory with expertise in multi-part assemblies.

3.3 Genetic Engineering – Paving the Way to Efficient Fabrication

The multi-gene assembly strategies introduced above can be used to construct complex biosynthetic pathways up to mini-chromosome size. The protein amounts, resulting from each gene of a pathway, and the ratios between different enzymes often have an immense influence on product yield. Hence, pathway optimization to balance expression output is a very important, but challenging task for sustainable heterologous production. Currently, most biosynthetic pathways for production of technical, pharmaceutical or medical products are not entirely clarified; single enzymes or the relevant ratios between proteins may be unknown. In such situations, combinatorial approaches open the possibility to rapidly test a multitude of different constructs to optimize expression output, e.g. by shuffling multiple promoters of different strengths or CDS sequences coding for different enzyme variants. A combinatorial approach to improve the output of the vitamin A precursor β-carotene was reported, based on the aforementioned VEGAS method[58]. Here, the combination of four different CDS with ten promoters and five terminators

each was used to determine the best producing combination. Another approach in yeast applies combinatorial cloning, supported by computational model-based predictions, to optimize the production of the pigment violacein[60]. While these examples aim at optimization of transcript levels, a different *E. coli* based approach targets translational efficiency. A set of ribosomal binding sites was combinatorically paired with genes for the production of the antioxidant astaxanthin, resulting in high levels of the desired product[61].

In many cases, it is not sufficient to solely optimize the production of heterologous genes to achieve efficient production. Even if the selected host offers an efficient expression system, the host metabolism does not necessarily support high product levels. It is often necessary to improve the metabolic flux through the engineered pathway by manipulating the host organism to reduce byproduct formation, enhance precursor availability, or to overcome unwanted feedback mechanisms. Such approaches were successfully applied to improve production of the natural fragrance β-ionone[62] or the antimalarial drug precursor artemisinic acid[18] in *S. cerevisiae*. However, this usually requires genome editing tools allowing the specific insertion, deletion, or replacement of DNA sequences in the host genome. In recent years, some very promising genome editing approaches arose along with the upcoming designer nucleases. These proteins contain programmable and sequence-specific DBDs and a nuclease domain, inducing double-strand (ds) DNA breaks at the desired genomic target site. Depending on the host organism and the experimental design, these breaks are subsequently repaired through homologous recombination or nonhomologous end-joining[63]. One example for such designer nucleases is the aforementioned group of meganucleases[35]. Their natural binding specificity towards long and rarely occurring recognition sites can be modified towards new target sites, but this requires labor-intensive methods like site-directed mutagenesis and elaborate screening procedures[64, 65]. Further approaches to engineer programmable DNA binding proteins aim for the modification of Cys_2-His_2 zinc finger nucleases (ZFNs) based on DNA-binding zinc finger motifs[66, 67]. Each motif, comprising approximately 30 aa, usually binds three DNA basepairs[68]. Within each zinc finger, only a few specified amino acids were found to determine the binding of a certain base triplet. Therefore, it is basically possible to develop prediction algorithms for DBS sequences.

However, a reliable prediction of an exact binding correlation is complicated by e.g. steric effects, which additionally influence DNA binding of zinc finger proteins[69]. Nevertheless, libraries of zinc finger modules have been developed along with protocols for screening procedures, enabling the selection and linkage of predefined modules to target customized DNA sequences with lengths of 9 to 18 bp[70-73].

A less complicated and more straightforward option for precise genomic manipulation became possible with the invention of Transcription Activator–Like Effector nucleases (TALENs)[74, 75]. Their natural counterpart, the TALE family from *Xanthomonas sp.*, features a common region in the middle of the protein, which consists of multiple, highly conserved repeat units[75-77]. The single repeat units, usually 34-aa long, differ only in the amino acids at positions 12 and 13, the so-called repeat variable diresidues (RVDs)[75, 77]. Repeat units with different RVDs show a simple and strong affinity towards specific nucleotides within the DBS. Hence, a simple code was generated, linking the different RVDs to the four different nucleotides[77, 78]. By using the RVD code, it is possible to modularly construct customized DBDs, targeting nearly every desired DBS[79]. The only limitation comes with the requirement of a thymine residue directly upstream of the target site[77]. The synthesis of the DBD, usually containing 18 to 24 repeat units[79], is based on strategies that tolerate repetitive sequences, including Golden Gate cloning[75], LIC[80] or high-throughput solid phase assembly[81]. Like ZFNs, TALENs are based on the fusion of a nuclease to the TALE-DBD. The significantly larger size of TALENs (~1000 aa) compared to ZFNs (~330 aa) hampers their usage for therapeutic applications[82, 83].

The latest type of programmable nucleases, the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system[84, 85], overcomes laborious cloning methods or screening procedures. The CRISPR/Cas method originates from a bacterial defense mechanism against invading foreign DNA that has been slightly modified for genome editing[86]. In this system, a synthetic single-guide RNA (sgRNA) directs the Cas9 nuclease to a desired 20-nt target site by complementary base pairing between the sgRNA and the DNA target[84, 87]. The Cas9 nuclease then generates a double strand break at the intended site. The only limitation, with respect to the chosen DBS, lies within the requirement of a directly contiguous protospacer

adjacent motif (PAM sequence, NGG in case of Cas9 from *Streptococcus pyogenes*)[84, 87]. When compared to TALENs, it becomes obvious that usage of the CRISPR/Cas9 system is more straightforward. In spite of the larger size of the Cas9 protein (~1400 aa), the simplicity of sgRNA cloning and the possibility for easy multiplexing[87] render this system the state-of-the-art technology. Nevertheless, ongoing research addresses the occurrence of undesired off-target cutting events. Although the CRISPR/Cas9 system has a low overall off-target rate[88], in direct comparison to TALENs, the specificity of CRISPR/Cas9 tends to be lower[89]. This may be problematic, especially for therapeutic applications[82].

The far end of host genome modification strategies and one of the major topics in synthetic biology is the generation of a minimal genome/minimal cell by deleting non-essential genes[90-92]. A reduced cell provides an excellent base for the integration of new genes, gene combinations and whole biosynthetic pathways. The synthetic background can be modified to yield a cell with specific properties for industrial, pharmaceutical or medical applications. In 2016, the first quasi minimal bacterial genome was reported. The reduced *M. mycoides* genome contains roughly half as many genes (473 genes) as its natural counterpart, reducing its size by 50 % to 531 kbp[93]. The "SCRaMbLE"-system[94] (Synthetic Chromosome Rearrangement and Modification by *LoxPsym*-mediated Evolution) is a current approach towards a eukaryotic minimal cell and part of the aforementioned Synthetic Yeast Sc2.0 Project. As a requirement for this system, each synthetic Sc2.0 yeast chromosome incorporates symmetrical *loxP* recombination sites (*loxPsym*) downstream of each non-essential gene and at specific landmarks[29]. This enables "inducible evolution" by reorganizing the whole genome, through duplications, inversions, translocations and deletions, upon expression of a Cre recombinase[29, 95, 96].

3.4 Transcription Factors – Workhorses of Synthetic Biology

As mentioned before, the heterologous biosynthetic production of relevant substances requires tight and specific expression control of each single gene. The prerequisite for designing artificial gene regulatory networks is the understanding of their natural counterparts. The highly-regulated expression of thousands of genes is essential for the

multitude of orchestrated processes taking place in each cell. The fact, that the various cell types in a multicellular organism contain the same genetic information, but differ gravely in their gene expression pattern, makes the complexity and importance of gene regulation obvious[97]. Differential gene expression is not limited to different cell types; even clonal cells can differ in their gene expression pattern, depending on environmental and internal signals. Gene expression is regulated on the transcriptional level by controlling the amount of produced mRNA, and post-transcriptionally by regulating the mRNA stability and its translation into proteins[98]. In eukaryotes, the regulation of mRNA production requires the combined action of many proteins, including RNA-polymerase II (RNAPII) and general, as well as specific transcription factors[99]. The expression of a given gene depends on the rate of transcription initiation, which is influenced by binding of activator or repressor proteins to specific or basal *cis*-elements in the promoter region[99, 100]. Different types of promoter elements exist, ranging from core elements like the TATA box[101], which is often described as the minimal promoter element, to more distal, general elements like the CAAT box[102], as well as cell specific and response elements. Each promoter contains a specific combination of *cis*-elements and drives gene expression, depending on the presence of transcriptional regulators. Prior to transcription initiation, the basal eukaryotic TFs need to bind the core promoter to recruit the components of the preinitiation complex (PIC)[103-105]. Initiation of transcription depends on the activation of RNAPII by phosphorylation of its C-terminal domain. The rate of transcription is controlled by transcriptional regulators, which bind to upstream *cis*-elements and interact with the mediator complex, which acts as signal transducer between regulators and PIC[106].

Certain genes are constitutively expressed, meaning these genes are perpetually transcribed, because their gene product is required in each cell type, at any time. Other genes are activated or repressed in response to certain spatial, temporal or environmental conditions. Several inducible promoters have been characterized in various species and adapted for biotechnology applications. For example, the chemically induced yeast *MET3*, *MET25* and *PHO5* promoters, which are negatively regulated by methionine or inorganic phosphate, the *CUP1* promoters, regulated by the

supply of copper to the media and the *GAL1* and *GAL10* promoters, that are activated by galactose and repressed by glucose[107-113].

Profound insight into the mechanism of transcriptional activation was gained by studies on the *S. cerevisiae* GAL4 protein, done by the Ptashne laboratory[114]. This research clarified that TFs consist of a DNA-binding domain (DBD) and an activation domain (AD), which stimulates the basal transcriptional machinery[114, 115]. Furthermore, the conserved character of activation mechanisms in eukaryotes was shown by demonstrating the functionality of the yeast-derived transcriptional activator protein GAL4 in mammalian cells[116]. Moreover, the functionality of hybrid proteins, consisting of a DBD and an AD derived from two different proteins or even two different species, was proven[117]. This principle was used to generate chimeric transcriptional regulators by combining diverse ADs and DBDs, thereby achieving a wide range of expression strengths for the fine-tuning of gene expression. A common example for artificial hybrid regulators is the dual Tet-Off/Tet-On system established in *S. cerevisiae*[12]. Here, the gene of interest (GOI) is placed under the control of an engineered promoter, containing the *Tn10* transposon-derived tetracycline-responsive operator (*tetO)* element. In the absence of tetracycline, the DBD of the tetracycline repressor (tetR), which is fused to the herpes simplex virus activation domain VP16, binds to *tetO* and activates gene expression. Supplied tetracycline forms a biologically active complex with a divalent magnesium cation[118]. This chelate mediates a conformational switch of tetR, releases it from the promoter region and thereby shuts off gene expression[118]. To reduce basal activity and consequently achieve tighter control of gene expression, this tetracycline-inactivable system was combined with a tetracycline-induced repressor in the same cell. The latter contains a mutated tetR-DBD (tetR'), which is fused to a repressor domain and binds *tetO* only in the presence of tetracycline. The addition of tetracycline to the medium results in an exchange of the *tetO*-bound activator for the tetR'-repressor.

In contrast to native DBDs, which bind to their specific DBS, synthetic programmable DBDs can be targeted to any plasmid- or genome-located DNA-sequence, as described above. DNA-binding proteins like zinc finger proteins or TALEs can be used as transcription factors[70, 119, 120], by fusing an activation domain or a repression domain to the DBD. A catalytically dead Cas9 (dCas9) protein, lacking nuclease activity, can also

be linked to an effector domain, to generate a sgRNA-directed artificial transcription factor[121, 122].

3.5 Light-Responsive Systems – Advanced Process Control

Analogous to the differential regulation of endogenous genes, it is often helpful to externally control regulators of gene expression or components needed for genome editing. Therefore, native or artificial induction systems were established, allowing the activation of a certain function at a defined point in time, like the aforementioned Tet-On/Tet-Off system. In most cases, these systems rely on chemical inducers that may cause toxic effects at elevated concentrations[12, 123]. Furthermore, a spatial, dose-dependent and reversible control of gene expression is not straightforward, as chemicals cannot be spatially targeted or traceless removed from the medium.

The establishment of diverse light-responsive regulator systems over the last years provides a promising alternative to conventional induction systems. Such regulators were used e.g. to control TFs[124, 125], nucleases[126], recombinases[127-129] or protein localization[130, 131]. Many of these systems are based on “photocaged” or “photolabile” chemicals, which are irreversibly released after illumination with certain wavelengths[132-134], or on biological photoreceptors from plants or microorganisms[135]. Plants, as sessile organisms, need to adapt to changing environmental conditions. Light is one of the most important parameters for sensing their environment. Therefore, plants evolved different photoreceptors to monitor the quality and quantity as well as the direction of light and the duration of the irradiation. The four physiologically most important classes of photoreceptors in the model plant *Arabidopsis thaliana* are the red/far-red light sensing phytochromes[136-138], the blue and UVA light sensitive cryptochromes[139, 140], the phototropins[141, 142], and the UVB sensing receptor UVR8[143, 144]. The phytochrome family, which plays a role in seed germination, plant growth and development, consists of five members, phyA to phyE[138, 145]. These dimeric proteins consist of two polypeptides, each folding into two main structural domains: a carboxy-terminal domain which mediates dimerization and an amino-terminal photosensory domain, which binds the chromophore phytochromobilin (PΦB)[146]. Red light illumination mediates a conformational switch of the covalently bound chromophore, which triggers photoconversion of the phytochrome

from the inactive Pr to the active Pfr form[138, 147]. The active phytochrome translocates to the nucleus and affects TFs, such as phytochrome interacting factors (PIFs). PIFs are basic helix-loop-helix transcription factors, which specifically dimerize with the active form of phytochromes[148]. For example, the Pfr form of PhyB, the phytochrome largely responsible for seed germination[137], was shown to inhibit the binding of negatively acting PIF3 to its target promoters by mediating phosphorylation, ubiquitination and thereby degradation of PIF3[149, 150]. PIF3 binds to G-box DNA sequence motifs present in various light-regulated gene promoters, like for example the CIRCADIAN CLOCK-ASSOCIATED protein 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY), both acting at the core of the circadian oscillator[151]. The two components of a natural optical dimerizer can be fused to split proteins, e.g. a split TF, consisting of a DBD and an AD. Independent of the light conditions, the DBD binds to its DBS in the promoter region. Light application induces dimerization of the photo-sensitive components and reconstitutes the split TF, thereby stimulating gene expression. Many expression systems for different host organisms have been reported in the last years, based on different light-sensing photoreceptors[152-155]. Even multichromatic systems were established for the specific and differential control of several genes with light of different wavelengths. Two examples for this invention are the combination of a red and a green light sensing system for control of gene expression in *E. coli*[156], and a UVB, red and blue light sensitive system reported for multiple gene control in mammalian cells[157, 158].

3.6 Goals of this Study

A major goal of the doctoral study presented here, was to establish a modular cloning system, specifically designed for multi-gene assemblies in *S. cerevisiae* (**Figure 1**). The system is supposed to feature easy and fast assembly of reusable parts, without the requirement of domestication of parts before usage. Furthermore, a scar free assembly of sub-gene fragments is intended, preferably with a minimum of PCR-amplification steps. Besides the focus on ease of use, the strategy should not rely on expensive or sophisticated cloning procedures, to ensure accessibility even to non-specialized laboratories.

The second major goal was to design light regulated tools for metabolic pathway engineering by employing the multi-gene assembly strategy established before (**Figure 1**). It was intended to design a novel gene-expression regulator, fulfilling the following criteria: programmability towards any desired target site, tunability of expression strength as well as cheap and easy handling. In the course of this study, the idea for an additional tool, enabling light-dependent genome rearrangement, was developed.

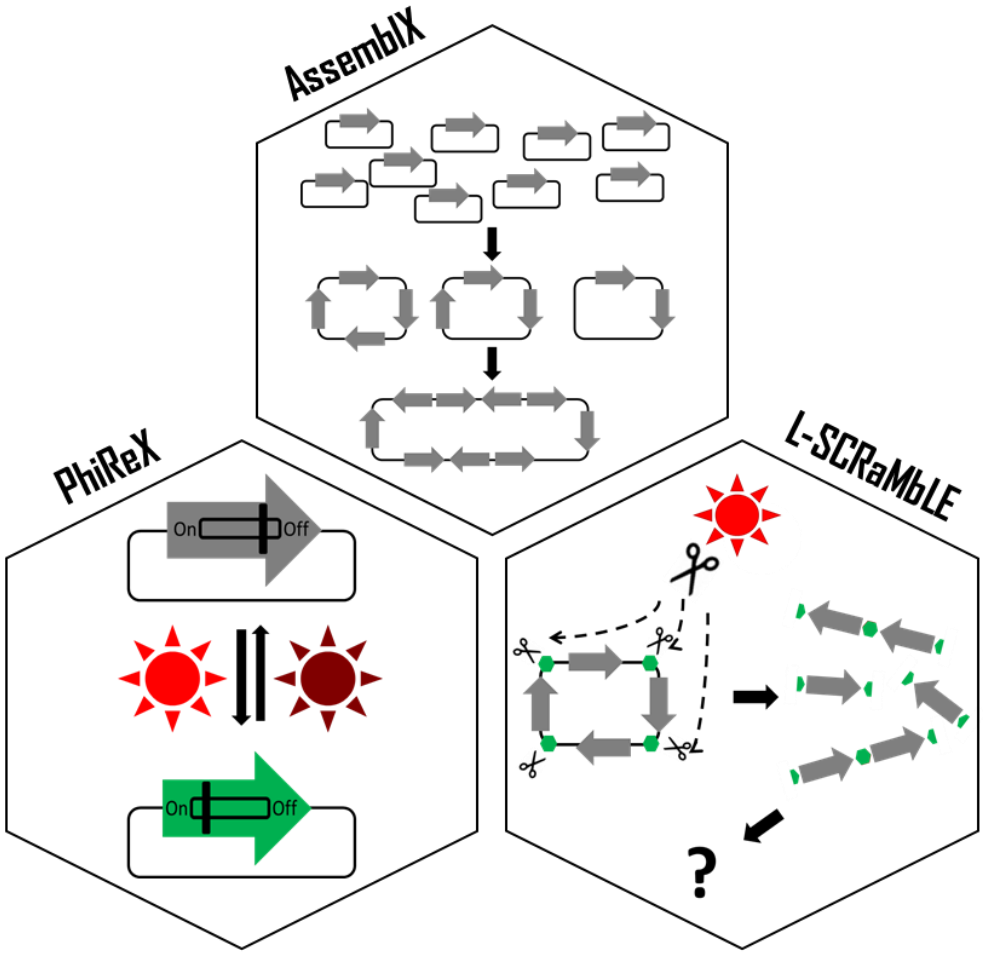


Figure 1: Overview of the synthetic biology tools developed in this work. Together, the multi-gene assembly toolkit AssemblX (Chapter 4.1), the light-responsive gene expression system PhiReX (Chapter 4.2) and L-SCRaMbLE for light-mediated recombination (Chapter 4.3), provide a versatile toolkit for synthetic biology.

4 MANUSCRIPTS

4.1 AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies

Lena Hochrein[+1], Fabian Machens[+1], Juergen Gremmels[3], Karina Schulz[1], Katrin Messerschmidt[#,*,1] and Bernd Mueller-Roeber[#,*,2,3]

[1]University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[2]University of Potsdam, Department of Molecular Biology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[3]Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany

[*] To whom correspondence should be addressed.

[+] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

[#] The authors wish to be known that, in their opinion, the last two authors should be regarded as joint last authors.

The manuscript has been published in Nucleic Acids Research:
Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Research*, gkx034 (2017).

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx034

AssemblX: a user-friendly toolkit for rapid and reliable multi-gene assemblies

Lena Hochrein[1,†], Fabian Machens[1,†], Juergen Gremmels[2], Karina Schulz[1], Katrin Messerschmidt[1,*] and Bernd Mueller-Roeber[2,3,*]

[1]University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Strasse 24–25, 14476 Potsdam, Germany, [2]Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany and [3]University of Potsdam, Department of Molecular Biology, Karl-Liebknecht-Strasse 24–25, 14476 Potsdam, Germany

Received August 11, 2016; Revised January 10, 2017; Editorial Decision January 11, 2017; Accepted January 13, 2017

ABSTRACT

The assembly of large DNA constructs coding for entire pathways poses a major challenge in the field of synthetic biology. Here, we present AssemblX, a novel, user-friendly and highly efficient multi-gene assembly strategy. The software-assisted AssemblX process allows even unexperienced users to rapidly design, build and test DNA constructs with currently up to 25 functional units, from 75 or more subunits. At the gene level, AssemblX uses scar-free, overlap-based and sequence-independent methods, allowing the unrestricted design of transcriptional units without laborious parts domestication. The assembly into multi-gene modules is enabled via a standardized, highly efficient, polymerase chain reaction-free and virtually sequence-independent scheme, which relies on rare cutting restriction enzymes and optimized adapter sequences. Selection and marker switching strategies render the whole process reliable, rapid and very effective. The assembly product can be easily transferred to any desired expression host, making AssemblX useful for researchers from various fields.

INTRODUCTION

Synthetic biology is a relatively young, but fast developing, scientific discipline. The field combines aspects of life sciences, chemistry and engineering to design, build and test new biological systems. Many projects in synthetic biology require sophisticated cloning methods, suitable for the modular assembly of complex DNA constructs, for example for the programmable production of compounds of medical or technical interest, construction of synthetic genomes or the design of regulatory circuits (1–4 and https://doi.org/10.1101/041871).

While cloning of single DNA parts into target vectors has become a routine technique in laboratories long ago, the assembly of multi-gene constructs or even genomes by *de novo* assembly of DNA fragments from different sources still remains a challenge (5). To enable and standardize such assembly projects, different methods have been reported. These can be roughly separated into two groups: namely sequence-dependent and sequence-independent methods. One of the first multi-part assembly strategies described is the sequence-dependent BioBrick (6). It is based on iterative restriction and ligation steps, using a combination of four restriction endonucleases (REs) with 6 bp-long recognition sites and compatible overhangs. BioBrick allows only step-by-step assemblies precluding the rapid assembly of large multi-gene constructs. Furthermore, recognition sites of six-base cutters statistically occur every 4 kilo base pairs (kb), making it necessary to 'domesticate' assembly parts by eliminating undesired cutting sites. This is a drawback many sequence-dependent methods, like Golden Gate cloning, MoClo and Golden Braid, have in common (7–9).

To overcome the drawbacks of RE-based cloning methods, sequence-independent cloning methods have been developed. Such cloning methods avoid REs and are often based on short overlapping sequences shared by the assembly parts, thereby allowing the directed assembly of multiple DNA fragments. These methods include the *in vitro* Gibson assembly (10), the bacterial cell extract-based SLiCE (11) and *in vivo* transformation-associated recombination (TAR) in *Saccharomyces cerevisiae* (12,13). The overlapping regions for each assembly reaction have to be designed individually, depending on the sequence of neighboring assembly parts. This results in a lack of modularity and an increased design and planning effort. Nevertheless, these methods are widely used to assemble scar-free constructs.

An example for a modular, but still sequence-independent assembly strategy is MODAL, which in particular focuses on high modularity for easy shuffling of single parts (14). Universal prefix and suffix adapters

*To whom correspondence should be addressed. Tel: +49 331 977 2810; Fax: +49 331 977 2512; Email: bmr@uni-potsdam.de
Correspondence may also be addressed to Katrin Messerschmidt. Tel: +49 331 977 6184; Fax: +49 331 977 6337; Email: messer@uni-potsdam.de
†These authors contributed equally to the paper as first authors.

for upstream and downstream primer binding are fused to individual assembly parts, which are then kept in storage plasmids. Specific linker sequences, computationally optimized for the overlap-guided assembly with neighboring fragments, are then fused to these adapters via a polymerase chain reaction (PCR) step. While MODAL allows the assembly of parts in different orders and orientations, the user has to accept large scar sequences, the risk of internal recombination (as every part contains the same adapter sequence), and, due to the necessary amplification steps, the risk of PCR-born errors with every new assembly reaction.

There are many more highly efficient assembly methods available, for example BASIC, yeast Golden Gate, VEGAS or MoClo (8,15–17). However, in most cases the user has to decide between scar-free, sequence-independent but non-modular methods or modular approaches, which generate scar sequences and require laborious parts domestication. We believe that this decision hinders many biology labs from outside the synthetic biology community to explore new grounds.

With these problems in mind, we have developed a new DNA assembly strategy, which makes use of several highly efficient cloning methods and combines the advantages in a streamlined process. The user-oriented AssemblX workflow allows scar-less, sequence-independent *de novo* assemblies from multiple parts at the single gene level and subsequent standardized, modular and PCR-free assemblies into multi-gene constructs. The whole process was designed for use with cost efficient cloning methods, enables multi parallel assemblies to speed up the process and allows the user to freely choose the final host organism for gene expression. The user is supported in all steps by the AssemblX web tool and gets a protocol including all primers, PCR conditions, digestions and assembly reactions. Taken together, AssemblX can be used in any lab and by any user to assemble complex genetic constructs for various purposes including pathway engineering.

MATERIALS AND METHODS

Strains, growth conditions, reagents and plasmids

Escherichia coli strains NEB5α, NEB10β (New England Biolabs, Frankfurt am Main, Germany) and DH5α were used for cloning purposes. All strains were grown in Luria-Bertani medium at 37°C, with kanamycin or ampicillin at 50 mg/ml as selection markers for transformed plasmids. For all yeast experiments, *S. cerevisiae* strains YPH500 (ATCC® 76626™) or BY4741 (ATCC® 201388™) were used and grown at 30°C in yeast extract peptone dextrose adenine (YPDA)-rich medium or in appropriate synthetic dextrose (SD) media lacking one or more amino acids to allow selection for transformed cells. Dominant selection markers were used in YPDA medium in the indicated final concentrations: G418 (200 μg/ml), hygromycin B (200 μg/ml), phleomycin (20 μg/ml) and nourseothricin (100 μg/ml). PCR purification and gel elution were done with NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Dueren, Germany) and plasmid purification was done with NucleoSpin Plasmid EasyPure kit (Macherey-Nagel). Gel elution of large fragments (>10 kb) was done with NucleoTrap kit (Macherey-Nagel). Plasmids pLOG1, pTPG1 and pUOGB are gifts from Tom Ellis, Imperial College London (18). Plasmids pCEV-G1-Ph and pCEV-G2-Km are gifts from Lars Nielsen and Claudia Vickers (Addgene plasmid # 46814 and 46815) (19). pAG36 is a gift from John McCusker (Addgene plasmid # 35126) (20). pHIS3p:mRuby2-Tub1+3′UTR::HPH is a gift from Wei-Lih Lee (Addgene plasmid # 50633). EBFP2-pBAD is a gift from Michael Davidson (Addgene plasmid # 54542). pJC104_violacein is a gift from Jeff Boeke, NYU School of Medicine (17). p426-SNR52p-gRNA.CAN1.Y-SUP4t is a gift from George Church (Addgene plasmid # 43803) (21). p2GWL7,0 and pKGWFS7 were provided by VIB-Gent University (https://gateway.psb.ugent.be) (22,23). pYC6Lys-TRP1URA3 is a gift from Francisco Malagon (Addgene plasmid # 11010). pdCas9-bacteria is a gift from Stanley Qi (Addgene plasmid # 44249) (24). Sequences of plasmids generated in this work are included in the Supplementary Data of this manuscript. The sequences were also deposited at NCBI GenBank (Accession numbers KY131987–KY132089) and are available via www.assemblx.org.

PCR methods

PCR amplifications were done with Phusion (Thermo Fisher Scientific, Waltham, MA, USA) or Q5 DNA Polymerases (New England Biolabs), following the manufacturer's recommendations. Sequences of all primers and oligonucleotides used for cloning are listed in Supplementary Tables S3 and 4. Amplified DNA parts were analyzed by agarose gel electrophoresis and column- or gel-purified prior to further use.

In vitro DNA assembly and cell transformation protocols

For overlap-based *in vitro* DNA assembly methods like SLiCE, Gibson and NEBuilder HiFi DNA Assembly, DNA fragments containing 25- to 36-bp-long overlap regions were generated by PCR amplification and were assembled with a linearized plasmid. SLiCE assembly was done as previously reported (11,25). Gibson assembly and NEBuilder HiFi DNA Assembly (both New England Biolabs) were done according to the manufacturer's recommendations. Commercially available chemically competent *E. coli* NEB5α or NEB10β cells (New England Biolabs) or homemade competent DH5α cells were transformed with 2 μl of the assembly reaction. Transformations were performed following the High Efficiency Transformation Protocol from New England Biolabs. Assemblies were verified by sequencing.

In vivo DNA assembly and cell transformation protocols

Saccharomyces cerevisiae transformations were done according to the LiAc/SS carrier DNA/PEG method by Gietz and Schiestl (26). For the assembly of multiple fragments, 100 ng of each DNA part were mixed with water to a final volume of 34 μl and transformed as described. Plasmid rescue from yeast colonies was done with the Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research Corporation, Freiburg, Germany) and the entire plasmid DNA

was subsequently transformed into chemically competent NEB10β cells. Level 2 constructs were transformed into electrocompetent NEB10β cells.

Assembly of Level 0 vector sets

To generate the Level 0 vector backbone, the pUC19 replication origin and the *nptII* kanamycin resistance gene from plasmid pHis2.1 (Takara Bio, Saint-Germain-en-Laye, France) were amplified by PCR using primer combinations P094/P095 and P096/P097, respectively. Primers contain overlaps for SLiCE cloning and introduce a SmaI site between the pUC19 origin and the *nptII* gene. The PCR products were combined in a SLiCE reaction to give plasmid pFM011. The vector pFM011 was linearized with SmaI and double-stranded oligonucleotides P251/P252 and P253/P254 were used in a SLiCE reaction to introduce two regions containing multiple 8-base cutter restriction sites, resulting in pFM012. Primers P288/P289 were used to amplify the 2-micron origin and the *URA*3 marker gene from pYES2/CT (Thermo Fisher Scientific). The resulting PCR product and AflIII-linearized pFM012 were used in a SLiCE reaction to give the final Level 0 backbone pFM032. To create individual Level 0 vectors (pL0 vectors) with different pairs of homology regions, pFM032 was cut by SmaI in between the 8-base cutter regions. Subsequently, two single-stranded oligonucleotides, each containing a different homology region and suitable overlaps for SLiCE cloning, were assembled into pFM032 together with a HindIII-flanked *ccdB* expression cassette, PCR-amplified from plasmid pDEST22 (Thermo Fisher Scientific) with primers P1165/P1166. The resulting Level 0 vectors contain both homology regions, separated by the HindIII-flanked *ccdB* cassette. For homology regions *AR*, *BR*, *CR*, *DR* and *ER*, PCR products containing the respective homology regions, coding for the first 17 codons of an auxotrophic marker gene, preceded by a promoter region for later control of the Level 1 marker gene, were generated (primer pairs P1173/P1174, P1207/P1208, P1240/P1241, P1274/P1275 or P1307/P1308). These PCR products were assembled together with the HindIII-flanked *ccdB* cassette and double-stranded oligonucleotides containing the other homology region in SmaI-linearized pFM032.

Assembly of Level 1 vectors

Two types of Level 1 vector backbones were constructed: a low-copy version with CEN/ARS region and a high-copy version with 2-micron origin. For the high-copy backbone, primers P104/P105 and P106/P107 were used to amplify the pBR322 bacterial origin of replication, the ampicillin resistance gene and the 2-micron origin from plasmid pYES2/CT (Thermo Fisher Scientific). Primers P108/P109 and P110/P111 were used to PCR-amplify stuffer fragments from pYES2/CT and to introduce PacI and NotI restriction sites. All four PCR fragments were combined in a SLiCE reaction to give plasmid pFM013. To create the low-copy plasmid pFM014, the PCR product obtained with primers P106/P107 was exchanged for a PCR product generated with primers P112/P113, containing the CEN/ARS origin from pDEST22 (Thermo Fisher Scientific). We further implemented a single-guide RNA (sgRNA) expression cassette for potential future CRISPR/Cas9 applications. Vectors pFM013 and pFM014 were linearized by NotI digestion. Primers P114/P115 and P199/P200 were used to PCR-amplify the crRNA expression cassette from plasmid p426-SNR52p-gRNA.CAN1.Y-SUP4t. The primer pairs were designed to amplify the *S. cerevisiae SNR52* promoter and the structural crRNA. Instead of the 20 bp sgRNA target region, a NotI site was introduced via the primer sequence, to allow later integration of any desired sgRNA target sequence. Linearized pFM013 or pFM014 were combined with the above PCR products in a SLiCE reaction to give plasmids pFM015 and pFM016, respectively. Vectors pFM015 and pFM016 were PCR-amplified with primer pair P402/P403, thereby adding I-SceI restriction sites on both termini. Level 1 vectors (pL1 vectors) with different combinations of homology regions and promoter-less marker genes were generated combining the PCR product P402/P403 with different PCR products (primers P1124–P1140, Supplementary Table S3) containing promoter-less auxotrophic marker genes equipped with 50-bp homology regions and a PacI site in a SLiCE reaction.

Assembly of Level 2 vectors

The basic Level 2 backbone is identical to the Level 1 backbone. To assemble different Level 2 vectors (pL2 vectors), either pFM015 or pFM016 were amplified with primer pair P402/P403 and subsequently used in a SLiCE reaction with two oligonucleotides, containing the desired homology regions separated by an EcoRI site and equipped with appropriate overlaps.

Promoter library

To generate a yeast promoter library, different native promoter and terminator sequences were PCR-amplified from *S. cerevisiae* BY4741. The primers used are detailed in Supplementary Table S4. Each corresponding promoter and terminator pair was individually cloned into the same Level 0 vector pL0A_0R (with A0-AR homology regions), separated by a NotI-flanked stuffer fragment. The stuffer contains the first and the last 36 bp of the yEGFP CDS, separated by an AsiSI restriction site. Resulting plasmids were linearized with AsiSI to allow insertion of PCR-amplified yEGFP. The plasmids were cut with PmeI or AscI to release the complete reporter gene with the regulatory sequences and flanking homology regions. The resulting fragments were cotransformed with the corresponding PacI-linearized Level 1 CEN/ARS vector pL1A-lc into *S. cerevisiae* YPH500 to give single-gene Level 1 vectors. Promoter activity for each construct was evaluated by flow cytometry-based yEGFP fluorescence measurements.

Flow cytometry

Cells were inoculated in 500 μl drop-out medium, containing either glucose or galactose as carbon source, in 48-well deep-well plates and incubated with shaking for 24 h (30°C, 240 rpm). A main culture in 500 μl fresh medium (supplemented with 20 mM IPTG or 1 μg/ml ATc, if applicable) was then inoculated to $OD_{600} \approx 0.1$ and incubated

for another 16 h, before fluorescence was determined via flow cytometry. Fluorescence output of single cells was measured with a BD FACSCalibur flow cytometer (BD Biosciences, Heidelberg, Germany). *S. cerevisiae* cells grown as described above were diluted in water and passed through the cytometer with less than 3000 counts per second. Yeast cells were identified and gated in a forward/sideward scatter dot plot. Within this gate, 20 000 cells were counted per measurement. Fluorescence was analyzed in a histogram and the geometrical mean fluorescence per cell was calculated for each measurement, using Flowing Software 2 (version 2.5.1, www.flowingsoftware.com). All results shown are mean values calculated from the indicated number of samples. The error bars indicate the standard deviation.

Twenty-five units demo assembly

All Level 0 assemblies were performed via SLiCE or HiFi DNA Assembly in a preliminary AssemblX version, which used non-optimized homology regions. All individual parts and their source plasmids or organisms are given in Supplementary Table S2. To make these constructs compatible with the final AssemblX platform, all 25 Level 0 units were PCR-amplified with primers that exchanged the non-optimized homology regions for the R2o-optimized versions thereof. Subsequently, from these 25 PCR products, five Level 1 modules were assembled using TAR with PacI-linearized Level 1 vector backbones (CEN/ARS versions). Positive candidates were identified via yeast colony PCR and plasmid DNA of selected clones was isolated with the Zymo Yeast Plasmid II kit (Zymo Research). Plasmid DNA was then transformed into *E. coli* NEB10β cells. For further verification, plasmids were isolated and analysed by restriction digestion and verified by sequencing of assembly junctions. Positive plasmids were processed following the AssemblX protocol and assembled in the EcoRI-digested CEN/ARS Level 2 vector pL2_AE_lc. After verification of successful assembly by colony PCR the expression of incorporated genes was analyzed via different reporter assays. The final construct was subsequently isolated from yeast following a protocol for isolation of circular YACs (27). For further analysis by sequencing and restriction digestion, the construct was transformed into *E. coli* NEB10β cells by electroporation and isolated using the QIAGEN Large-Construct Kit (Qiagen, Hilden, Germany). Sequencing was done by LGC Genomics (Berlin, Germany).

Pulsed-field gel electrophoresis (PFGE)

Gel electrophoretic analysis of the full-length 25-gene construct was done with a CHEF MAPPER XA System (Bio-Rad Laboratories, Muenchen, Germany) on a 1% pulsed-field gel electrophoresis (PFGE)-grade agarose gel in 0.5 × TBE at 14°C. Instrument settings were: 6 V/cm, 120° pulse angle, initial switch time 0.22 s, final switch time 7.67 s with linear ramp, run time 15:16 h.

β-Galactosidase and β-glucuronidase overlay assays

Plate overlay assays to detect β-galactosidase (*LacZ*) or β-glucuronidase (*GUS*) were done according to the Dual Bait Hybrid Hunter Yeast Two-Hybrid System manual (Version C, Invitrogen). Results shown are from a 45 min incubation at 37°C and subsequent incubation at 4°C overnight.

Dual luciferase assay

Dual luciferase assays were done as described using the dual luciferase reporter assay system (Promega GmbH, Mannheim, Germany) and a GlowMax 20/20 luminometer (Promega) (28). For easy comparison, FLuc values are normalized by corresponding RLuc values and *vice versa*. Results shown are mean values with standard deviation from three biologically independent samples, each measured in triplicate during a single experiment.

Online tool

The web tool was implemented with the web application framework Struts 2 (Apache Software Foundation, Los Angeles, CA, USA) using the integrated development environment MyEclipse (Genuitec, Flower Mound, TX, USA). As a servlet container we use Tomcat (Apache Software Foundation). The communication with the web service j5 works by using an XML-RPC interface (29). The AssemblX web tool will be made available for a minimum full 2 years following publication, via www.assemblx.org and via GitHub (https://github.com/AssemblX/AssemblXWeb) under a GNU General Public License as published by the Free Software Foundation.

RESULTS

The AssemblX strategy

The assembly strategy presented here allows the *de novo* assembly of up to 25 (transcriptional) units from 75 or even more subunits into a single construct in a predefined order. AssemblX relies on overlap-based cloning methods, like *in vitro* SLiCE and HiFi DNA assembly (NEB) or *in vivo* TAR (10,11,13). These methods utilize overlapping homology regions of various lengths flanking the DNA parts to allow the scar-less assembly of multiple parts into a single DNA construct. To facilitate a fast assembly process, AssemblX is designed in three successive cloning Levels, each Level allowing multiple parallel assemblies: entry Level 0 for assembly of units from individual parts like promoter, terminator and coding sequence (CDS), or any other type of DNA sequence; Level 1 for the assembly of up to five Level 0 units into Level 1 modules, and Level 2 for the final assembly of up to five Level 1 modules into a single construct (Figure 1). Furthermore, AssemblX provides a j5-based web tool (www.assemblX.org), which assists the user with all design and assembly steps and therefore greatly reduces the time required to complete complex assemblies (29).

Assembly Level 0

Level 0 assemblies allow the construction of assembly units from several subunits, e.g. promoter, CDS, terminator. For this purpose, we designed a vector backbone, which is similar for all Level 0 vectors and just differs in two variable homology regions that will later define the position of the assembly unit within the final construct (Figure 2). The Level

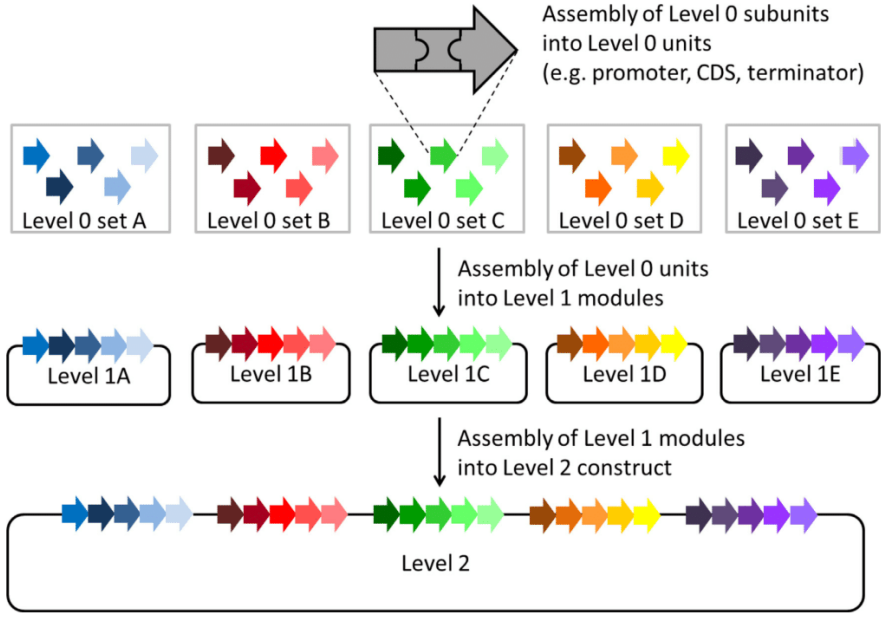


Figure 1. The AssemblX workflow for the assembly of 25 units. Level 0 subunits, here with three DNA parts each, are assembled into five different sets of Level 0 vectors resulting in 25 Level 0 units. In the next step, Level 0 units from one Level 0 set are assembled in one corresponding Level 1 module. The five Level 1 modules, each containing five Level 0 units, are then assembled into one single Level 2 vector.

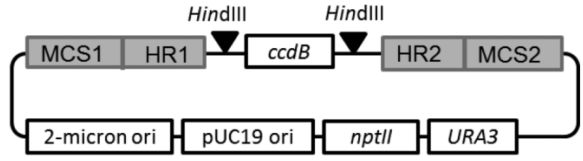


Figure 2. Design of Level 0 vectors. The backbone is a yeast shuttle vector with a pUC19 replication origin, the kanamycin resistance gene *nptII* for selection in *Escherichia coli*, a high-copy 2-micron origin for replication and the *URA3* marker gene for selection in *Saccharomyces cerevisiae*. Each Level 0 vector possesses a different combination of homology regions (HR). HR1 gives homology to the Level 1 vector backbone or to the HR2 of the adjacent Level 0 vector. HR2 gives homology to the next Level 0 vector or to the Level 1 vector backbone. These regions are separated by a HindIII-flanked *ccdB* expression cassette. The HindIII sites are used to release the *ccdB* cassette and to allow the subsequent insertion of the individual subunits in-between the homology regions by scar-free, overlap-based cloning. The homology regions are flanked by two multiple cloning sites (MCS), containing the five 8-base cutters AscI, SbfI, SwaI, FseI, PmeI to release Level 0 constructs for further cloning into Level 1.

0 backbone consists of the *E. coli* pUC19 replication origin, the kanamycin resistance gene *nptII*, the 2-micron origin for replication in yeast and the yeast *URA3* marker gene. Each Level 0 vector corresponds to a given position (position 1–25) in the final Level 2 assembly product and therefore acts as a positioning vector. The upstream (left) homology region of a vector defines the upstream neighbor, whereas the downstream (right) homology region defines the downstream neighbor. To this end, each downstream homology region is used as an upstream region in the neighboring downstream Level 0 vector. The homology regions are separated by a HindIII-flanked *ccdB* expression cassette. For the Level 0 assembly of a given assembly unit from multiple subunits, the *ccdB* cassette is released by HindIII digestion, to allow the overlap-based cloning in-between the homology regions, e.g. by well-established methods like SLiCE, TAR or HiFi DNA assembly. To this end, all subunits need to be equipped with customized overlaps to their neighboring subunits or to the vector sequence, respectively. This is usually achieved by introducing the overlaps with primers during PCR amplification (Figure 3). The primer design can be done manually or, for more convenience, with the help of the AssemblX web tool (see below). All PCR products necessary for the assembly are then combined with the linearized Level 0 backbone in an appropriate assembly reaction. The size, number of subunits and overall design of a Level 0 unit is completely user defined and only limited by the chosen assembly method. During the Level 0 cloning process, the *ccdB* gene reduces empty vector background and obviates purification of the linearized Level 0 backbone. A successfully assembled Level 0 unit can be released from the backbone for subsequent Level 1 assembly by using one of five rare 8-base cutter recognition sites (AscI, SbfI, SwaI, FseI, PmeI) flanking the homology regions. This makes it unnecessary to PCR-amplify the assembly units prior to Level 1 assemblies and eliminates the risk of PCR-born errors.

Level 0 vectors are grouped into five sets (set *A*, *B*, *C*, *D* and *E*), each containing five positioning vectors. All units, cloned in one set of Level 0 vectors, are assembled into one specific Level 1 vector, e.g. pL1A for Level 0 set *A*, or pL1B for Level 0 set *B*. The resulting Level 1 module holds up to five Level 0 units in the predefined order (Figure 4). The

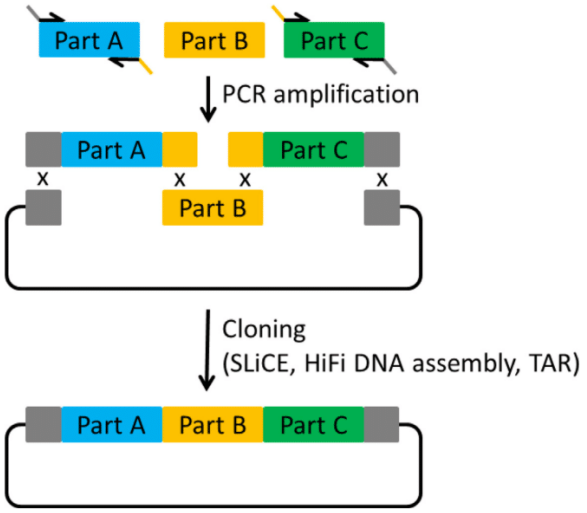


Figure 3. Level 0 assembly with overlap-based cloning methods. The overlaps between the individual subunits can be added via PCR, using primers overlapping with the neighboring fragment in their 5′ region. After amplification of Part A and Part C with appropriate primers the three parts can be assembled into a vector backbone.

first (or left) homology region of a Level 0 vector set is designated as type *X0 (X = A, B, C, D* or *E)*. These regions are 50 bp long and provide the left-arm homology to a specific Level 1 vector. The inlying homology regions of type *X1* to *X4* are 36 bp long each and will later provide the necessary overlaps for a directed assembly of Level 0 units into a single Level 1 module. The last homology region of a vector set, designated type *XR*, is 51 bp long and provides right arm homology to a Level 1 vector. The longer overlaps to the Level 1 vector support the TAR mediated assembly of large fragments in the later Level 2 assembly step. All homology regions of type *XR* were defined manually: each represents the first 17 codons from one out of five different auxotrophic yeast marker genes and is preceded by the corresponding promoter region and followed by a stop codon. Despite providing overlap to a Level 1 vector via *XR,* this region also drives the expression of an auxotrophic marker, in successfully assembled Level 1 constructs.

To design the sequences for homology regions of type *X0* and type *X1* to *X4* for each vector set, we employed the web tool R2o designer (30). The tool generates orthogonal sequences, intended to be biologically neutral and well suited for isothermal overlap-based assembly methods (40% GC content, Supplementary Table S1) (14).

As the *X0* and *XR* regions of a given Level 0 set provide overlap to the Level 1 vector, they are strictly required for the next assembly level. Therefore, the vector collection described so far would only allow the cloning of exactly five Level 0 units per Level 1 module. To allow the assembly of less than five units, we created additional Level 0 vectors that contain the *XR* homology region of a vector set, in combination with *X0*, *X1*, *X2* or *X3*. For example *A1-AR* vectors allow Level 1 assembly of two Level 0 units by using *A0-A1* and *A1-AR* vectors from vector set *A*. Table 1 shows a complete list of all Level 0 vectors.

Design of Level 1 vectors

Level 1 vectors allow the assembly of up to five Level 0 units, released from their Level 0 backbone by digestion, into one Level 1 module. All Level 0 fragments for one module are combined with the appropriate, linearized Level 1 vector in an *in vitro* or *in vivo* overlap-based assembly reaction. Each Level 1 vector backbone is available in two versions, with either a high-copy 2-micron origin or a low-copy CEN/ARS region for replication in yeast. This boosts the flexibility for the direct expression of small assembly projects consisting of up to five units in a single Level 1 module. To create Level 1 vectors compatible with the different Level 0 vector sets, we inserted homology regions *X0* and *XR*. The left homology region of type *X0* is followed by a PacI site for linearization, and a promoterless auxotrophic marker gene, which contains a *XR* type homology region, defined as the first 17 codons of the respective gene. The resulting vectors are listed in Table 1. The marker gene is preceded by a third homology region *Y0* (first homology region of the next Level 1 vector) downstream of the terminator, e.g. *B0* for the pL1A vector. Homology regions X0 and Y0 are flanked by 18 bp long recognition sites for the homing endonuclease I-SceI (31,32). Digestion with I-SceI allows the release of Level 1 modules from the backbone and their further cloning into Level 2 without PCR amplification. During this step the auxotrophic marker genes integrated in the Level 1 constructs are also transferred into the Level 2 construct, which allows stringent selection of successful assemblies. Level 1 assembly may be done by *in vitro* overlap-based methods or by *in vivo* assembly in yeast. In contrast to the Level 0 vectors, the Level 1 destination vectors confer ampicillin resistance, allowing selection against undigested Level 0 backbones in *E. coli.*

Design of Level 2 vectors

After I-SceI mediated release from the Level 1 backbone, up to five Level 1 modules can be combined into a single Level 2 destination vector, preferably by using TAR. Level 2 vectors can be used directly to maintain and express the assembled units in *S. cerevisiae* either in high or low copy number. The basic Level 2 vector backbones are identical to the Level 1 backbones and differ only in the homology regions inserted between the two I-SceI sites. The different Level 2 vectors are designed to hold different numbers of Level 1 modules. A Level 2 vector for only one Level 1 module is not necessary, as it would be identical to the corresponding Level 1 construct. All Level 2 vectors contain homology region *A0* in combination with a second region, providing overlap to the homology region *Y0* from the last Level 1 module that is part of the assembly (Figure 5 and Table 2). The successful assembly can be monitored by selection of transformed yeast on SD dropout medium selecting for the presence of all incorporated Level 1 modules.

If the final construct is intended for use in host organisms different from yeast, any desired (expression) vector, e.g. a T-DNA vector for plant transformation, can be converted into an AssemblX compatible vector (Supplementary Figure S1). To this end, the target vector is linearized either with a suitable restriction enzyme, or by PCR amplification.

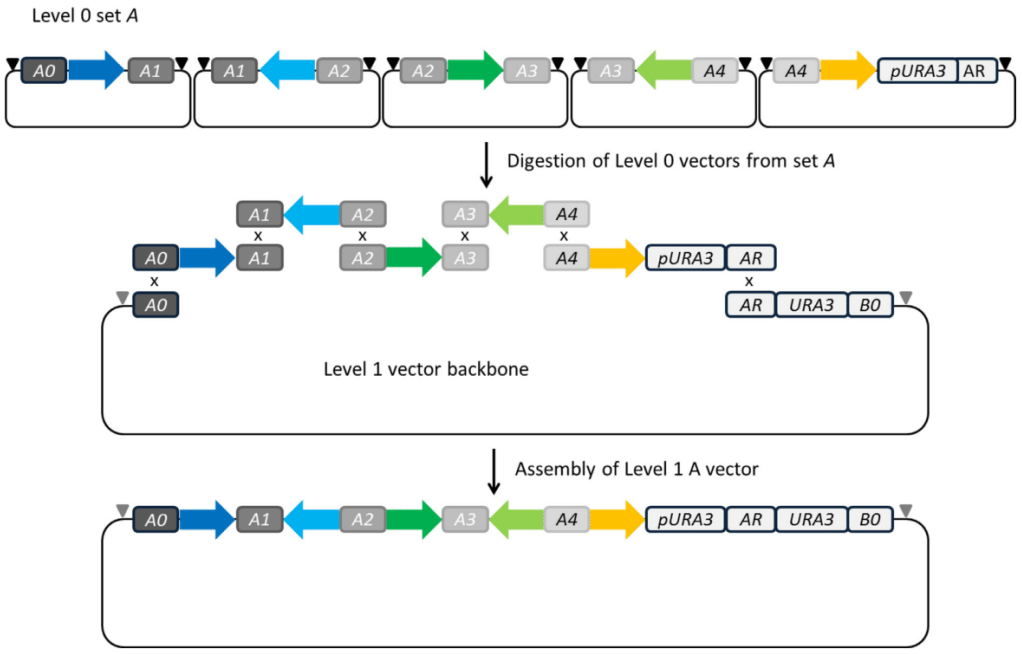


Figure 4. Level 1 assembly scheme. Five Level 0 units are released from Level 0 vectors by digestion with one of the designated 8-base cutters (black triangle). During *in vivo* or *in vitro* assembly, overlaps between the different Level 0 units and the Level 1 destination vector ensure a directed assembly into a single circular DNA construct. Located in front of the homology region *AR* is the *URA3* promoter. *AR* itself represents the first 50 nucleotides of the *URA3* CDS. The complete *URA3* CDS and the appropriate terminator are located on the Level 1 vector backbone. In this way a functional selection marker is created only by a successful Level 1 assembly. Complete Level 1 modules are flanked by I-SceI sites (gray triangle).

Thereafter, the user has the following three options: (i) insertion of a vector conversion cassette, amplified from a pL2 vector containing the required homology regions. This cassette contains the homology regions and a yeast replication origin. The newly created vector can be used as backbone in the Level 2 assembly reaction in yeast. The resulting construct is suitable for transfer to the desired host organism. (ii) Insertion of an appropriate homology region cassette into the target vector for overlap-based subcloning of I-SceI digested Level 1 or Level 2 constructs. This cassette can be amplified from a Level 2 AssemblX vector containing the required homology regions. (iii) The insertion of a single I-SceI site or an I-SceI *–ccdB-* I-SceI cassette into the target vector for the subcloning of an I-SceI digested Level 1 or Level 2 construct by ligation.

The AssemblX web tool

The assembly of up to 25 genes with AssemblX requires a thorough design of primer sequences for Level 0 assemblies. To facilitate planning and design we established a web tool, which supports the user in every step of the assembly. The web tool is accessible via www.assemblx.org. The help function guides the user through every step of the intuitive planning and design workflow: beginning with the number of units to be assembled, organization of the individual assembly units into distinct modules, planning the variable parts of the final assembly, detailed primer design, assembly suggestions for Level 0 and ending with individualized protocols for higher-level assemblies. After the user has defined a schematic overview of the final assembly product, the Level 0 assemblies are addressed: the user provides DNA sequences for all individual subunits, in either GenBank or plain text format. The sequence files are then used to design primers for Level 0 assemblies. To this end, the user's input is converted into input files for the online tool j5, which is an automated design software for DNA assemblies (29). The web tool hands over the data to the j5 server and integrates the j5 results into the further process. Basically, the j5 output contains suggestions for PCR primers for amplification and attachment of homology regions to Level 0 subunits. From these suggestions, individual protocols for the assembly of Level 0 units are compiled by the AssemblX web tool and combined into a written bench protocol for all steps of the AssemblX process leading the user to the final product.

Highly efficient assembly of 25 expression cassettes using AssemblX

For the successful construction and expression of multigene pathways it is necessary to control the expression level of many individual genes. To provide a first starting point for researchers working with *S. cerevisiae*, we cloned and characterized a panel of 40 constitutive yeast promoters, each in combination with the terminator region from the same gene (Supplementary Results 1 and Supplementary Figure S2). To make use of this promoter collection and to demonstrate the capability of AssemblX, we designed and successfully cloned a panel of 25 individually

Table 1. Complete list of available Level 0 and Level 1 vectors

Level 0 vector set	Vector name	HR1	HR2	Corresponding Level 1 vectors (HR1/HR2)
A	pL0A_0-1	A0	A1	pL1A-hc/pL1A-lc (A0/AR)
	pL0A_1-2	A1	A2	
	pL0A_2-3	A2	A3	
	pL0A_3-4	A3	A4	
	pL0A_0-R	A0	AR	
	pL0A_1-R	A1	AR	
	pL0A_2-R	A2	AR	
	pL0A_3-R	A3	AR	
	pL0A_4-R	A4	AR	
B	pL0B_0-1	B0	B1	pL1B-hc/pL1B-lc (B0/BR)
	pL0B_1-2	B1	B2	
	pL0B_2-3	B2	B3	
	pL0B_3-4	B3	B4	
	pL0B_0-R	B0	BR	
	pL0B_1-R	B1	BR	
	pL0B_2-R	B2	BR	
	pL0B_3-R	B3	BR	
	pL0B_4-R	B4	BR	
C	pL0C_0-1	C0	C1	pL1C-hc/pL1C-lc (C0/CR)
	pL0C_1-2	C1	C2	
	pL0C_2-3	C2	C3	
	pL0C_3-4	C3	C4	
	pL0C_0-R	C0	CR	
	pL0C_1-R	C1	CR	
	pL0C_2-R	C2	CR	
	pL0C_3-R	C3	CR	
	pL0C_4-R	C4	CR	
D	pL0D_0-1	D0	D1	pL1D-hc/pL1D-lc (D0/DR)
	pL0D_1-2	D1	D2	
	pL0D_2-3	D2	D3	
	pL0D_3-4	D3	D4	
	pL0D_0-R	D0	DR	
	pL0D_1-R	D1	DR	
	pL0D_2-R	D2	DR	
	pL0D_3-R	D3	DR	
	pL0D_4-R	D4	DR	
E	pL0E_0-1	E0	E1	pL1E-hc/pL1E-lc (E0/ER)
	pL0E_1-2	E1	E2	
	pL0E_2-3	E2	E3	
	pL0E_3-4	E3	E4	
	pL0E_0-R	E0	ER	
	pL0E_1-R	E1	ER	
	pL0E_2-R	E2	ER	
	pL0E_3-R	E3	ER	
	pL0E_4-R	E4	ER	

The Level 0 vectors are grouped into five sets. Each set is compatible with a distinct Level 1 vector, which is available in either high- or low-copy format (hc/lc).

Table 2. Available Level 2 vectors

Level 1 modules used for assembly	Corresponding Level 2 vectors (HR1/HR2)	Auxotrophic markers for selection
A, B	pL2_AB_hc/pl2_AB_lc (A0_C0)	*URA3, LEU2*
A, B, C	pL2_AC_hc/pL2_AC_lc (A0_D0)	*URA3, LEU2, HIS3*
A, B, C, D	pL2_AD_hc/pL2_AD_lc (A0_E0)	*URA3, LEU2, HIS3, LYS2*
A, B, C, D, E	pL2_AE_hc/pl2_AE_lc (A0_F0)	*URA3, LEU2, HIS3, LYS2, TRP1*

Depending on the number of Level 1 modules to assemble, different Level 2 vectors must be used. Each vector is available in high- and low-copy format (hc/lc). Depending on the assembled Level 1 modules, the resulting Level 2 construct can be selected with up to five auxotrophic markers.

regulated genes, consisting of 69 different DNA parts in total, into a single Level 2 construct, in about 4 weeks. Besides several fluorescent and enzymatic reporter genes (*yEGFP*, *AsRed2*, *EBFP2*, *lacZ*, *uidA*, *FLuc*, *RLuc*), a set of dominant yeast selection markers (G418, hygromycin, phleomycin and nourseothricin resistance genes) for easy and direct validation of successful assemblies in yeast has been incorporated. *LacI* and *TetR* genes and the corresponding modified *GAL1 UAS*-based promoters enable the IPTG- or ATc-controlled expression of yEGFP and AsRed2, respectively. Furthermore, the complete Level 2 assembly product contains the biosynthetic pathway for

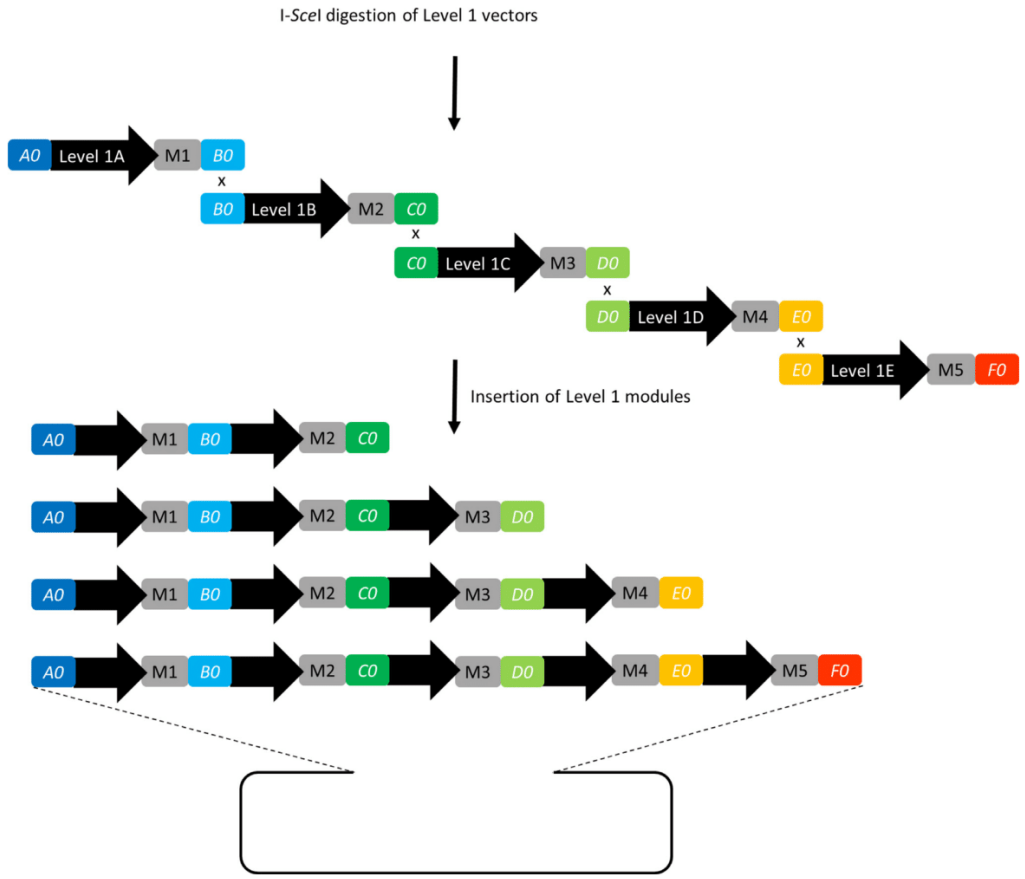


Figure 5. Level 2 assembly scheme. Level 1 modules are released from Level 1 vectors by digestion with I-SceI. During assembly, overlaps between the different modules and the Level 2 destination vector ensure a directed assembly into a single circular DNA construct. Depending on the number of Level 1 modules to be assembled, different Level 2 vectors for the assembly of two to five Level 1 modules are used. M1: *URA3*, M2: *LEU2*, M3: *HIS3*, M4: *LYS2*, M5: *TRP1*.

the purple pigment violacein consisting of five yeast codon optimized genes (*vioA, vioB, vioC, vioD* and *vioE*) from *Chromobacterium violaceum*, and five *rpo* genes for the production of *E. coli* RNA polymerase subunits (alpha, beta, beta', omega and sigma70 factor) in yeast. Finally, two *E. coli* resistance markers (spectinomycin and chloramphenicol resistance) for easy monitoring of plasmid integrity, after transfer to *E. coli,* are included. Supplementary Table S2 shows a complete list of all genes incorporated into the assembly, the subunits from which the Level 0 units have been assembled, and their position in the final construct. All Level 0 assemblies have been done with a preliminary version of AssemblX, using a set of Level 0 vectors with non-optimized homology regions and lacking the *ccdB* cassette. Overall, the cloning efficiency for Level 0 assemblies greatly varied with complexity and choice of the assembly method (efficiencies between 2 and 100%, data not shown). Therefore, a meaningful general efficiency cannot be calculated. To increase the cloning efficiency and reduce empty vector background, we redesigned the Level 0 vectors by optimizing the homology regions using R2o designer (30) and inserting the *ccdB* cassette, as described above. To assess the efficiency of these changes, the four Level 0 assemblies that performed worst with the previous non-optimized AssemblX version (cloning efficiencies below 5%) were reassembled in the optimized pL0 vectors. The cloning efficiencies determined by colony PCR and restriction digestion increased to 33% for one construct and 100% for the remaining three constructs. This demonstrates the positive effect of *ccdB* counterselection and optimized homology regions. To change the old homology regions to the optimized ones in the remaining constructs, the complete Level 0 units were amplified with primers containing the new R2o-optimized homology regions. Subsequently, the PCR products were directly used in Level 1 assemblies. All Level 1 assemblies were done with TAR cloning in YPH500 yeast cells. Yeast colonies were pre-screened by colony PCR and potential positive plasmids were transferred to *E. coli* for subsequent analysis by restriction digestion and sequencing of assembly junctions (Table 3). For Level 1 modules A, B, C and D, 100% of clones analyzed by colony PCR were identified as potential positives and all clones further investigated by restriction digestion showed the expected fragment pattern. Sequencing of selected clones further confirmed the

expected assembly junctions. Assembly of Level 1E was less efficient, as only 15/24 clones (62%) were identified as potential positives by colony PCR. Four of these clones were further analyzed by digestion and all were confirmed to be positive by sequencing. The R2o-optimized homology regions are probably not the reason for the lower assembly efficiency in Level 1E, as we experienced a much lower cloning efficiency for this module with non-optimized homology regions (4/47 clones, data not shown). Most likely, random internal homologies between the *vioE* CDS and the *ACT1* and *TDH1* promoters, which are parts of Level 1 module E, are the cause for the reduced cloning efficiency. Subsequent to the Level 1 assemblies, the individual Level 1 modules were assembled into the pL2AE-lc backbone using TAR. Analysis by multiplex colony PCR (24/24 clones positive), assays for the reporter genes *lacZ* and *uidA* (42/42 clones positive) and growth assays on plates containing G418, hygromycin, phleomycin and nourseothricin (42/42 clones positive) indicate a cloning efficiency of 100% and confirm the expression of six reporter genes (Table 3 and Supplementary Figure S3a–d). For further analysis, a single clone was randomly selected. The DNA was isolated and successfully transformed into *E. coli*. Selection of transformants was done with spectinomycin and chloramphenicol, indicating the presence of the respective resistance genes, which are part of the assembly. The DNA was isolated from *E. coli* and integrity was confirmed via gel electrophoresis and restriction mapping (Supplementary Figure S4). The size of the final assembly product was determined to be between 70 and 75 kb, which fits the expected size of 72 kb, and restriction with I-SceI releases a 68.5 kb fragment, which was confirmed with PFGE. Also, all expected restriction fragments from a SalI/PacI digestion were observed. Sequencing of the assembly junctions between all Level 1 modules and the Level 2 vector backbone gave further proof for a complete and error-free assembly. Functional analysis of the reporter genes *FLuc* and *RLuc* was done by dual luciferase measurements, confirming the expression of these genes in yeast (Supplementary Figure S5a). Similarly, flow cytometry based measurements for yEGFP or AsRed2 fluorescence with or without the inducers IPTG or ATc indicate the functionality of *LacI* and *TetR*, regulating the expression of yEGFP and AsRed2 in yeast, respectively (Supplementary Figure S5b). The above mentioned experiments confirm the function of 14 (reporter) genes. To investigate the successful expression of the five-gene violacein pathway, the construct was retransformed into BY4741 yeast cells. This was done, because the original assembly host YPH500 is auxotroph for tryptophane, which is converted to violacein by the *vioABCDE* pathway. Transformed BY4741 cells are purple, indicating the successful production of the purple pigment violacein by the five biosynthetic genes *vioA*, *vioB*, *vioC*, *vioD* and *vioE* (Supplementary Figure S3e and f). The five assembly units, containing *E. coli rpo* genes for expression in yeast, were verified by complete sequencing. Equally, the *EBFP2* gene was confirmed by sequencing, because, for unknown reasons, the fluorescence signal in microscopic analyses was very weak. Taken together, the results from the demo assembly, which was completed in about 4 weeks, confirm the effectiveness of the assembly strategy presented here. Several other, but not yet published assembly projects, have already been completed using AssemblX. The testimonial section available at www.assemblX.org provides an overview on the different topics addressed so far.

DISCUSSION

The AssemblX toolkit provides a user-friendly strategy for reliable *de novo* multi-part assemblies in *S. cerevisiae* and the subsequent expression in yeast or any other host organism. The toolkit does not require parts domestication, allows performing all necessary Level 0 assemblies in parallel to save time and avoids PCR amplification for higher assembly levels. Furthermore, AssemblX enables the functional analysis and reuse of individual Level 0 units or complete Level 1 modules, is highly efficient, uses inexpensive methods, is flexible with regard to the user's preference for different overlap based cloning methods, and is accompanied by an online tool. The latter provides easy access for users regardless of their scientific background and minimizes experimental errors. In this way, undergraduate students and experienced researchers alike will be able to create complex genetic constructs, benefiting from the easy-to-follow three-step workflow.

The novelty of AssemblX lies in an elaborate combination of different methods to facilitate and optimize cloning workflow and expression of multi-gene constructs, rather than an optimization or modification of already functional cloning methods. The overall AssemblX strategy incorporates many useful tweaks that include reduced background through *ccdB* counter selection, *E. coli* marker switching between Level 0 and Level 1, and selection of successful *in vivo* assemblies in yeast through marker gene reconstitution. The homology regions used in AssemblX are R2o-optimized, supporting highly efficient assemblies (up to 100% correct clones in Level 1 and 2 assemblies) and can be used with most overlap-based assembly methods. AssemblX combines the advantages of overlap-based, scar-free and sequence-independent assemblies in Level 0 vectors and subsequent standardized adapter-based assemblies in Levels 1 and 2. This flexibility allows an unrestricted architecture for every Level 0 unit and minimizes the risk of running into intractable conflicts during cloning.

The sequence independency of AssemblX relies on the release of Level 0 units from the positioning vectors by restriction digestion with a set of five rare-cutting 8-base cutters for further Level 1 assemblies. Laborious deletion of unintended cut sites, or PCR amplifications which bear the risk of introducing sequence errors, thereby become unnecessary. This is also true for I-SceI-based release of Level 1 modules from their backbones, as the I-SceI recognition site is 18 bp long, extremely rare and does for example not cut within the entire yeast genome (33). The avoidance of PCR amplification facilitates Level 1 and 2 construct verification, as diagnostic restriction digestion or sequencing of assembly junctions is usually sufficient.

The features, uniquely combined in AssemblX, allow completing a 25-parts assembly in about a month, starting from scratch, as demonstrated in the present study by generating a 25-gene assembly resulting in a 72 kb plasmid. An even faster multi-gene assembly is possible when pre-existing modules are employed. While several cloning

Table 3. Efficiency of Level 1 and 2 assemblies during the 25-gene demo assembly

Assembly step	Positive in colony PCR	Confirmed by restriction analysis	Confirmed by sequencing
1A	24/24	4/4	2/2
1B	24/24	4/4	2/2
1C	24/24	4/4	2/2
1D	24/24	4/4	2/2
1E	15/24	4/4	2/2
2	24/24	1/1	1/1

strategies enabling multi-gene cloning have been published, none of them aims at constructs of this complexity and simultaneously allows sequence-independent, scar-free gene-level assemblies in a user oriented workflow (8,14–17,34).

We chose *S. cerevisiae* as host for the *in vivo* cloning steps, as homologous recombination in yeast is an efficient and cost-effective alternative to other recombination-based cloning methods. Furthermore, *S. cerevisiae* is a well characterized organism with reported success in the production of important natural substances like taxol and artemisinic acid (35–37). The library of constitutive promoters that we established and characterized here will assist users in establishing expression constructs for complex biosynthetic pathways in yeast. Researchers working with a different host or even 'non-model' organisms can easily transfer every Level 1 or Level 2 AssemblX construct into a new host by either including the necessary markers and/or replication origins into the assembly itself or by I-SceI- or homology region-based subcloning into appropriate vectors. We believe that this will encourage researchers from various fields to make use of the AssemblX toolkit and to explore new grounds in synthetic biology.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

B.M.-R. thanks the Federal Ministry of Education and Research of Germany for funding. The authors greatly acknowledge Dr Pamela Holzlöhner (University of Potsdam, Germany) for introduction to the flow cytometer, Nathan Hillson (Joint BioEnergy Institute, Emeryville, CA, USA) for support with the implementation of j5 into the AssemblX web tool, and Jessica Behrend and Christoph Alt for testing the web tool.

Author contributions: L.H. and F.M. designed and developed the overall AssemblX strategy, conducted the experiments, analyzed the data and wrote the paper. J.G. programmed the online tool. K.S. constructed the different vector sets. B.M.-R. initiated the project and the overall research strategy, and edited the paper. K.M. and B.M.-R. jointly supervised the work. All authors take full responsibility for the content of the paper.

FUNDING

Federal Ministry of Education and Research of Germany (BMBF) [031A172 to B.M.-R.]. Funding for open access charge: BMBF [031A172].

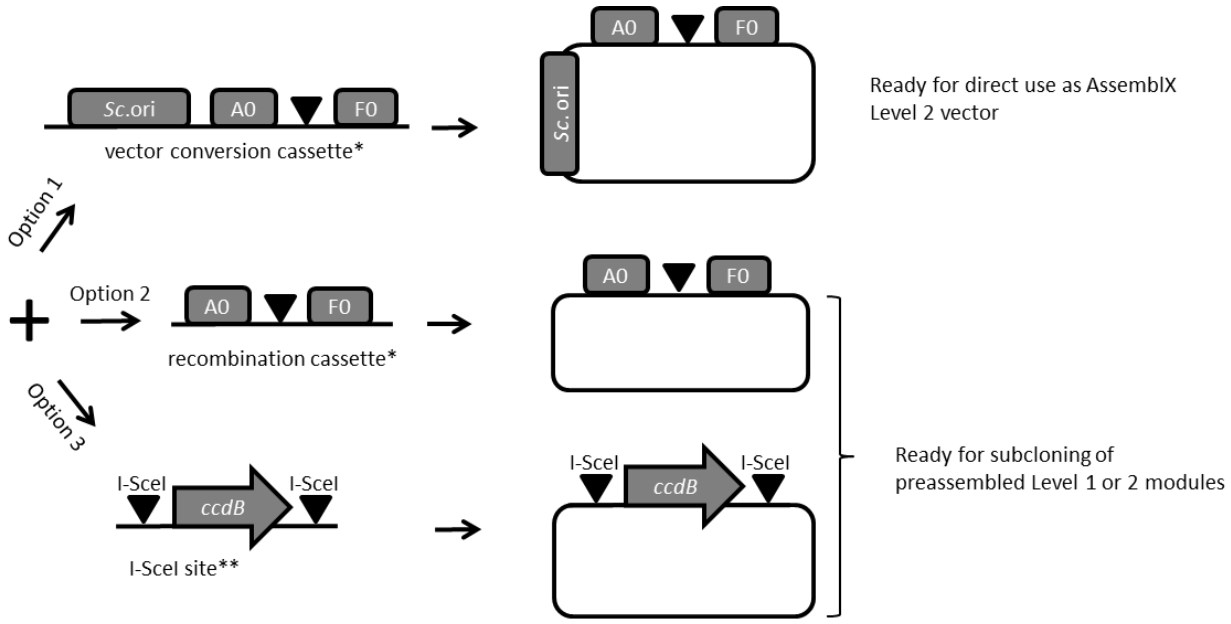
Conflict of interest statement. The authors declare competing financial interests, due to a patent application for the AssemblX cloning strategy, filed by the University of Potsdam with L.H., F.M., K.M. and B.M.-R. as inventors (Patent number EP16188155.2).

REFERENCES

1. Si,T., Luo,Y., Xiao,H. and Zhao,H. (2014) Utilizing an endogenous pathway for 1-butanol production in *Saccharomyces cerevisiae*. *Metab. Eng.*, **22**, 60–68.
2. Runguphan,W. and Keasling,J.D. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.*, **21**, 103–113.
3. Annaluru,N., Muller,H., Mitchell,L.A., Ramalingam,S., Stracquadanio,G., Richardson,S.M., Dymond,J.S., Kuang,Z., Scheifele,L.Z., Cooper,E.M. *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science*, **344**, 55–58.
4. Kiani,S., Beal,J., Ebrahimkhani,M.R., Huh,J., Hall,R.N., Xie,Z., Li,Y. and Weiss,R. (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods*, **11**, 723–726.
5. Cohen,S.N., Chang,A.C., Boyer,H.W. and Helling,R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 3240–3244.
6. Knight,T. (2003) *Idempotent vector design for standard assembly of biobricks*. DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group, Massachusetts Institute of Technology. Cambridge, MA.
7. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**, e3647.
8. Weber,E., Engler,C., Gruetzner,R., Werner,S. and Marillonnet,S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One*, **6**, e16765.
9. Sarrion-Perdigones,A., Falconi,E.E., Zandalinas,S.I., Juárez,P., Fernández-del-Carmen,A., Granell,A. and Orzaez,D. (2011) GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One*, **6**, e21622.
10. Gibson,D.G. (2011) Enzymatic assembly of overlapping DNA fragments. In: Voigt,C (ed). *Methods in Enzymology*. Academic Press, San Diego, CA, Vol. **498**, pp. 349–361.
11. Zhang,Y., Werling,U. and Edelmann,W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.*, **40**, e55.
12. Gibson,D.G. (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.*, **37**, 6984–6990.
13. Kouprina,N. and Larionov,V. (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.*, **3**, 371–377.
14. Casini,A., MacDonald,J.T., De Jonghe,J., Christodoulou,G., Freemont,P.S., Baldwin,G.S. and Ellis,T. (2014) One-pot DNA construction for synthetic biology: the modular overlap-directed assembly with linkers (MODAL) strategy. *Nucleic Acids Res.*, **42**, e7.
15. Storch,M., Casini,A., Mackrow,B., Fleming,T., Trewhitt,H., Ellis,T. and Baldwin,G.S. (2015) BASIC: A new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.*, **4**, 781–787.
16. Agmon,N., Mitchell,L.A., Cai,Y., Ikushima,S., Chuang,J., Zheng,A., Choi,W.-J., Martin,J.A., Caravelli,K., Stracquadanio,G. *et al.* (2015)

Yeast Golden Gate (yGG) for the efficient assembly of *S. cerevisiae* transcription units. *ACS Synth. Biol.*, **4**, 853–859.
17. Mitchell,L.A., Chuang,J., Agmon,N., Khunsriraksakul,C., Phillips,N.A., Cai,Y., Truong,D.M., Veerakumar,A., Wang,Y., Mayorga,M. *et al.* (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res.*, **43**, 6620–6630.
18. Ellis,T., Wang,X. and Collins,J.J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechol.*, **27**, 465–471.
19. Vickers,C.E., Bydder,S.F., Zhou,Y. and Nielsen,L.K. (2013) Dual gene expression cassette vectors with antibiotic selection markers for engineering in *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **12**, 96.
20. Goldstein,A.L. and McCusker,J.H. (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, **15**, 1541–1553.
21. DiCarlo,J.E., Norville,J.E., Mali,P., Rios,X., Aach,J. and Church,G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.*, **41**, 4336–4343.
22. Licausi,F., Weits,D. and Pant,B. (2011) Hypoxia responsive gene expression is mediated by various subsets of transcription factors and miRNAs that are determined by the actual oxygen availability. *New Phytol.*, **190**, 442–456.
23. Karimi,M., Inzé,D. and Depicker,A. (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.*, **7**, 193–195.
24. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
25. Messerschmidt,K., Hochrein,L., Dehm,D., Schulz,K. and Mueller-Roeber,B. (2016) Characterizing seamless ligation cloning extract for synthetic biological applications. *Anal. Biochem.*, **509**, 24–32.
26. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
27. Noskov,V.N., Chuang,R.Y., Gibson,D.G., Leem,S.H., Larionov,V. and Kouprina,N. (2011) Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat. Protoc.*, **6**, 89–96.
28. McNabb,D.S., Reed,R. and Marciniak,R.A. (2005) Dual luciferase assay system for rapid assessment of gene expression in *Saccharomyces cerevisiae*. *Eukaryot. Cell*, **4**, 1539–1549.
29. Hillson,N.J., Rosengarten,R.D. and Keasling,J.D. (2011) j5 DNA assembly design automation software. *ACS Synth. Biol.*, **1**, 14–21.
30. Casini,A., Christodoulou,G., Freemont,P.S., Baldwin,G.S., Ellis,T. and MacDonald,J.T. (2014) R2oDNA Designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.*, **3**, 525–528.
31. Colleaux,L., D'Auriol,L., Galibert,F. and Dujon,B. (1988) Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6022–6026.
32. Monteilhet,C., Perrin,A., Thierry,A., Colleaux,L. and Dujon,B. (1990) 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.*, **18**, 1407–1413.
33. Thierry,A., Perrin,A., Boyer,J., Fairhead,C., Dujon,B., Frey,B. and Schmitz,G. (1991) Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-SceI. *Nucleic Acids Res.*, **19**, 189–190.
34. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palaci,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.*, **162**, 1618–1631.
35. Engels,B., Dahm,P. and Jennewein,S. (2008) Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metab. Eng.*, **10**, 201–206.
36. DeJong,J.M., Yule,L., Bollon,A.P., Long,R.M., Jennewein,S., Williams,D. and Croteau,R.B. (2006) Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.*, **93**, 212–224.
37. Ro,D.-K., Paradise,E.M., Ouellet,M., Fisher,K.J., Newman,K.L., Ndungu,J.M., Ho,K.A., Eachus,R.A., Ham,T.S. and Kirby,J. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940–943.

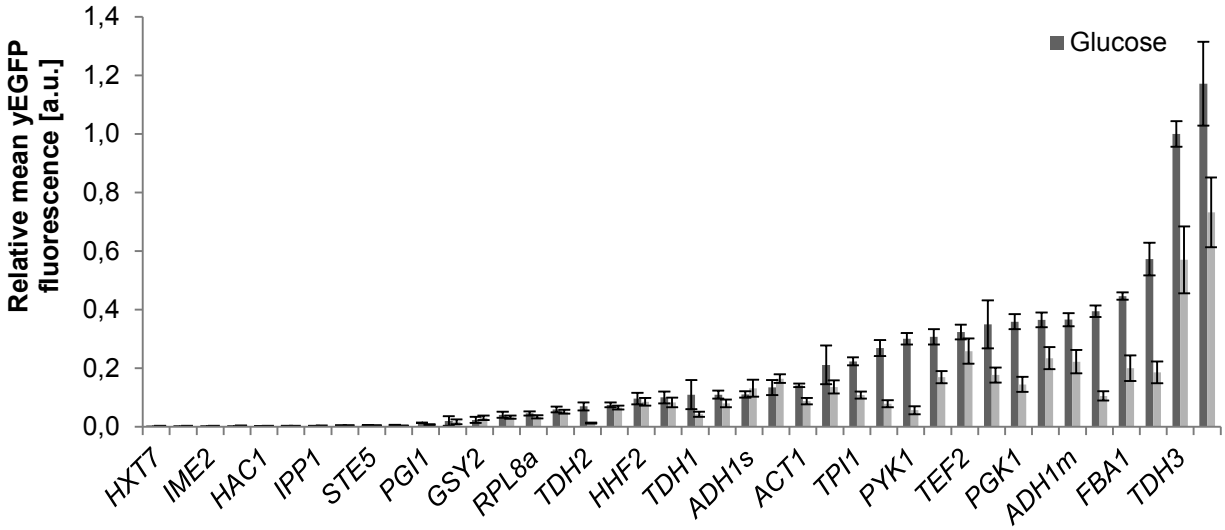
SUPPLEMENTARY DATA



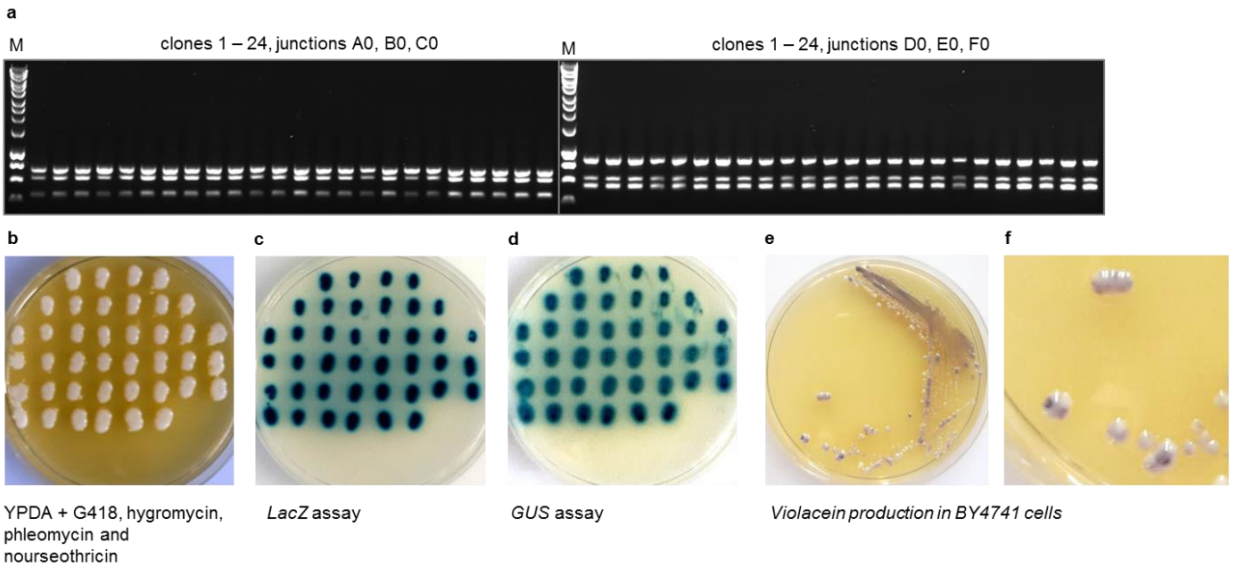
*amplified from appropriate pL2 vector

**amplified from pL1A_12

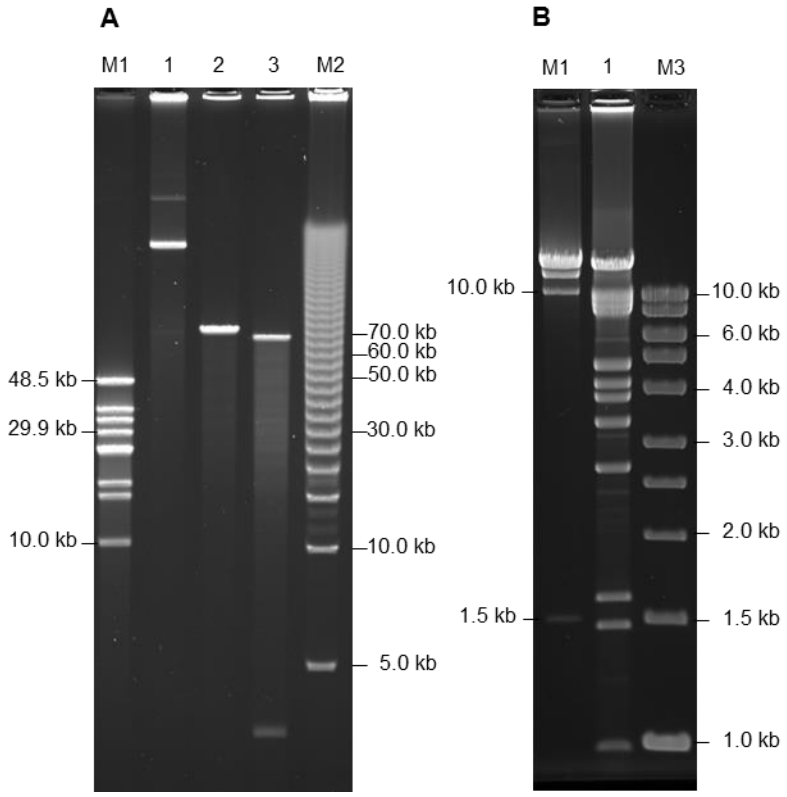
Supplementary Figure 1: Three different possibilities to convert any target vector into an AssemblX compatible vector. Option 1: Equip expression vector with a yeast replication origin and use the converted vector directly in a TAR mediated Level 2 assembly. The appropriate cassette can be amplified from the existing AssemblX Level 2 vectors. Option 2: Equip expression vector with a recombination cassette compatible to your Level 2 construct. The appropriate recombination cassette can be amplified from the existing AssemblX Level 2 vectors. Following the vector conversion, the modified vector can be used to subclone an *I-SceI* released Level 2 module via overlap based cloning. Option 3: Equip expression vector with an I-*Sce*I-*ccdB*-I-*Sce*I cassette amplified from vector pL1A_12. The modified vector can be digested with I-*Sce*I and ligated to any I-*Sce*I released fragment from any AssemblX Level 1 or 2 construct.



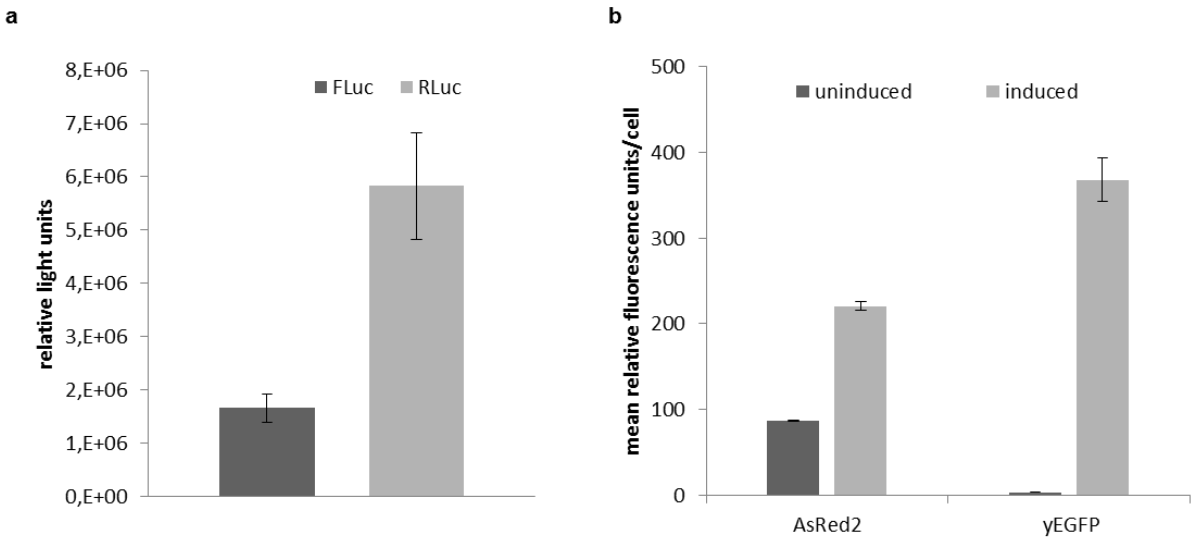
Supplementary Figure 2: Comparison of 40 different constitutive promoter-terminator pairs from *S. cerevisiae*. The geometrical mean of the yEGFP fluorescence per cell was determined using flow cytometry after 16 h of batch culture in either glucose- or galactose-containing SD medium. Fluorescence values are given relative to the *TDH3* promoter-terminator pair. Values are mean values from three independent experiments, each with two technical replicates. Error bars represent standard deviation.



Supplementary Figure 3: **Analysis of assembled Level 2 construct.** (a) Amplification of Level 2 assembly junctions A0 to F0 by yeast multiplex colony PCR. All 24 clones analyzed show the expected product sizes. (b - d) Plate assays with 42 clones to confirm expression off all four dominant yeast selection markers, LacZ and GUS activity. (e+f) BY4741 cells transformed with the 25-gene construct yield purple colonies due to violacein production.



Supplementary Figure 4: Gel electrophoretic analyses of the final 25-gene construct. M1: λ DNA-Mono Cut Mix (NEB), M2: 5 kb Ladder (BioRad), M3: HyperLadder 1 kb (Bioline). (a) PFGE analysis of undigested (lane 1), *Not*I-linearized (lane 2) and I-*Sce*I-digested (lane 3) construct. (b) Gel electrophoretic separation of *Sal*I/*Pac*I-digested construct (lane 1).



Supplementary Figure 5: Functional analysis of reporter genes incorporated into the final Level 2 construct. (a) Mean *Firefly* (FLuc) and *Renilla* (RLuc) luciferase activity determined for the final 25-gene construct. (b) Mean AsRed2 and yEGFP fluorescence per cell, determined by flow cytometry. Induction of AsRed2 or yEGFP was done by application of 1 µg/ml Atc or 20 mM IPTG, respectively, in galactose-containing SD medium. Values are mean values from three biological replicates, measured during one experiment. Error bars represent standard deviation.

Supplementary Table 1: Sequences of all AssemblX homology regions.

Name	*Sequence (5'-3')*
A0	*AAATAGGAAAGTCTGGTAACTCAGGACACTGTATCTGCTACGCTGTTTAT*
A1	*TAAAGTGAGATTCGCACAATAAACGCTGCTTCTGTT*
A2	*TAGATTCTGTGAAACTCTCAATAAAGGGACTTCGTC*
A3	*TCCTTCCCAGTGAATAGAAAAGTAGTTGCTGTTACA*
A4	*TAATAAGTTCAGTTTCGGATACCTGCCCAATAGAGT*
AR	*ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTT*
B0	*TAAGGTTTCAGTCACACTACGGATACTTTTACAACGGGAGCAGTTATTCA*
B1	*CAATAAGTGCTATCCTGAATCTCTGGTTGTGAAACA*
B2	*AGTTGAGAATCTATTTTACTACCCGAAGACTGCGTA*
B3	*ATTACGGTCTGTGGATTTCTCACTGAACGCAATAAA*
B4	*TTGGAAGTCTACTCCCTGAACTGTGAAGAATACTAT*
BR	*ATGTCTGCCCCTAAGAAGATCGTCGTTTTGCCAGGTGACCACGTTGGTCAA*
C0	*TATTGCGTGTGATACTTCAGGAGAACCGTTATTCCGTAAAACTACAGTCA*
C1	*GGTATCGTAATCGCAACTCACTGTAAGGACTATTTA*
C2	*TAGTAAGTTATTATCGCACTCACGAGAAGCGTATTC*
C3	*ATAAACTCTGAATACTCGGGATTTCCTGGCAATAGT*
C4	*TGGTGATTCAATAAAGCGTCGTTACACAGATACTTC*
CR	*ATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGAT*
D0	*CACGAGCGTTGAGGCGGATTTCAGATTATTTACAGACACTATTCTACAAT*
D1	*AAATAACTGCCACTACTGAAGATTCGGTGATTGTCT*
D2	*ACCACTGTAAAGGTATTATTCGGAAGTTCTCTACGA*
D3	*AACTGAGATTGGATAGACTACCTGTTCGTGTAATAC*
D4	*ATTTATCGGAGAAAACCTCGTCGTTGAACCAGATTT*
DR	*ATGACTAACGAAAAGGTCTGGATAGAGAAGTTGGATAATCCAACTCTTTCA*
E0	*TCTCTTACAATAAAATAGAATACTCTGCGAAGGGCGTGACACTGCTGTTT*
E1	*AGTTTATCAATCCGCTACAGGACGGTATTTCGTTAT*
E2	*TTGATACGAAGTGTGACAATCCTTTTACTGAACCGA*
E3	*TAAAAGTAAACTCTCACCCGTAGGATTGGTATCTGT*
E4	*TAGTGATAGCGTTCAATACTTCCTGGCACGATAATA*
ER	*ATGTCTGTTATTAATTTCACAGGTAGTTCTGGTCCATTGGTGAAAGTTTGC*
F0	*AAAGTATTGGTGTGAGTAATCGCTATCTCTTCTACGACTGGCTACAACAA*

Supplementary Table 2: Components of the 25-gene demo assembly. The subunits from which every gene has been assembled on Level 0, as well as the gene function in the final assembly are given.

Assembly position (HR1-HR2)	***Level 0 assembly components/source***			***Function***
A0-A1	Spectinomycin adenyltransferase expression cassette (*aadA*) / pCR8/GW			*E. coli* spectinomycin resistance
A1-A2	*PDC1* promoter / *S. cerevisiae* BY4741	lac repressor (*lacI)* CDS / pLOG1	*UBX6* terminator / *S. cerevisiae* BY4741	*lacI* regulation of *yEGFP* expression
A2-A3	*GPH1* promoter / *S. cerevisiae* BY4741	beta-galactosidase (*lacZ*) CDS / pYES3/CT/*lacZ*	*GPH1* terminator / *S. cerevisiae* BY4741	*lacZ* expression
A3-A4	*GAL1-lacO* promoter / pLVGI	yeast enhanced yellow fluorescent protein (*yEGFP*) CDS / pUOGB	*CYC1* terminator / *S. cerevisiae* BY4741	yEGFP expression
A4-AR	*ADH1*(m) promoter / *S. cerevisiae* BY4741	aminoglycoside-3'-phosphotransferase (*neo*)CDS / pCEV-G2-Km	*ADH1* terminator / *S. cerevisiae* BY4741	*S. cerevisiae* G418 resistance
B0-B1	*PGK1* promoter / *S. cerevisiae* BY4741	bleomycin resistance protein (*ble)* CDS / pCEV-G1-Ph	*PGK1* terminator / *S. cerevisiae* BY4741	*S. cerevisiae* phleomycin resistance
B1-B2	*TPI1* promoter / *S. cerevisiae* BY4741	tetracylcine repressor protein (*tetR*) CDS / pTPG1	*TPI1* terminator / *S. cerevisiae* BY4741	*tetR* regulation of RFP expression
B2-B3	*EFT2* promoter / *S. cerevisiae* BY4741	*E. coli* RNAP subunit beta' (*rpoC)* CDS / *E. coli* DH5alpha	*EFT2* terminator / *S. cerevisiae* BY4741	*rpoC* expression
B3-B4	*RPL3* promoter / *S. cerevisiae* BY4741	*Renilla* luciferase (*Rluc*) CDS / pRL-null	*RPL3* terminator / *S. cerevisiae* BY4741	*Rluc* expression

Assembly position (HR1-HR2)	***Level 0 assembly components/source***			***Function***
B4-BR	*GAL1-tetO* promoter / pTVGI	Red fluorescent protein (AsRed2) CDS / pAsRed2-N1	*UBC6* terminator / *S. cerevisiae* BY4741	ATc induced RFP expression
C0-C1	Hygromycin B resistance cassette / pHIS3p:mRuby2-Tub1+3'UTR::HPH			*S. cerevisiae* hygromycin B resistance
C1-C2	*HHF2* promoter / *S. cerevisiae* BY4741	*E. coli* RNAP subunit alpha (*rpoA*) CDS / *E. coli* DH5alpha	*HHF2* terminator / *S. cerevisiae* BY4741	*rpoA* expression
C2-C3	*HTA2* promoter / *S. cerevisiae* BY4741	*E. coli* RNAP subunit omega (*rpoZ)* CDS / *E. coli* DH5alpha	*HTA2* terminator / *S. cerevisiae* BY4741	*rpoZ* expression
C3-C4	*ENO2* promoter / *S. cerevisiae* BY4741	enhanced blue fluorescent protein (*EBFP2*) CDS / EBFP2-pBAD	*ENO2* terminator / *S. cerevisiae* BY4741	*EBFP2* expression
C4-CR	*RPL8a* promoter / *S. cerevisiae* BY4741	*E. coli* RNAP subunit beta (*rpoB*) CDS / *E. coli* DH5alpha	*RPL8a* terminator / *S. cerevisiae* BY4741	*rpoB* expression
D0-D1	*FBA1* promoter / *S. cerevisiae* BY4741	Nourseothricin acetyltransferase (*nat1*) CDS / pAG36	*FBA1* terminator / *S. cerevisiae* BY4741	*S. cerevisiae* nourseothricin resistance
D1-D2	*TEF1* promoter / *S. cerevisiae* BY4741	beta-glucuronidase (*uidA*) CDS / pKGWFS7	*TEF1* terminator / *S. cerevisiae* BY4741	*uidA* expression
D2-D3	*CCW12* promoter / *S. cerevisiae* BY4741	Firefly luciferase (*Fluc*) CDS / p2GWL7,0	*CCW12* promoter / *S. cerevisiae* BY4741	Firefly luciferase expression
D3-D4	Complete chloramphenicol acetyltransferase expression cassette (c*amR*)/ pDCas9			*E. coli* chlormaphenicol resistance
D4-DR	*GPM1* promoter / *S. cerevisiae* BY4741	*E. coli* sigma70 factor (r*poD*) CDS / *E. coli* DH5alpha	*GPM1* terminator / *S. cerevisiae* BY4741	*rpoD* expression

Assembly position (HR1-HR2)	***Level 0 assembly components/source***			***Function***
E0-E1	*TDH3* promoter / *S. cerevisiae* BY4741	Yeast codon optimized *vioA* CDS / pJC104_violacein	*TDH3* terminator / *S. cerevisiae* BY4741	Expression of violoacein biosynthesis pathway (*vioABCDE*)
E1-E2	*TEF2* promoter / *S. cerevisiae* BY4741	Yeast codon optimized *vioB* CDS / pJC104_violacein	*TEF2* terminator / *S. cerevisiae* BY4741	
E2-E3	*ACT1* promoter / *S. cerevisiae* BY4741	Yeast codon optimized *vioC* CDS / pJC104_violacein	*ACT1* terminator / *S. cerevisiae* BY4741	
E3-E4	*TDH1* promoter / *S. cerevisiae* BY4741	Yeast codon optimized *vioD* CDS / pJC104_violacein	*TDH1* terminator / *S. cerevisiae* BY4741	
E4-ER	*ZEO1* promoter / *S. cerevisiae* BY4741	Yeast codon optimized *vioE* CDS / pJC104_violacein	*ZEO1* terminator / *S. cerevisiae* BY4741	

Supplementary Table 3: List of primers used for cloning and assembly of Level 0, Level 1, and Level 2 vectors.

Name	*Sequence (5´-3´)*
P094	CACGCGCCTCGTCCCGGGCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
P095	CCACTGAGCGTCAGACCCTCTTCCGCTTCCTCGCTCACTGA
P096	GAGCGAGGAAGCGGAAGAGGGTCTGACGCTCAGTGGAACG
P097	CTTTGATCTTTTCTACGGCCCGGGACGAGGCGCGTGTAAGTTACAGGC
P104	ATCCACTAGTTTAATTAACGGCCCTGCATTAATGAATCGG
P105	AACTGTATTATAAGTAAATGCATGTATACTAAACTC
P106	GAGTTTAGTATACATGCATTTACTTATAATACAGTTTTATCGATGATAAGCTGTCAAAGATG AG
P107	AAAAATTTAACGCGAATTTTAACAAAATATTAACGTCATTGCGAATACCGCTTCCACA
P108	ACGTTAATATTTTGTTAAAATTCGCGTT
P109	TACGGCACCTGCGGCCGCGACCCCAAAAAACTTGATTAGGGTG
P110	ATCAAGTTTTTTGGGGTCGCGGCCGCAGGTGCCGTAAAGCAGTAAATCGGAAGG
P111	TTCATTAATGCAGGGCCGTTAATTAAACTAGTGGATCATCCCCACGCGC
P112	GAGTTTAGTATACATGCATTTACTTATAATACAGTTGACGGATCGCTTGCCTGT
P113	AAAAATTTAACGCGAATTTTAACAAAATATTAACGTGGTCCTTTTCATCACGTGCT
P114	ACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCTCTTTGAAAAGATAATGTATGATTA TGCTTTCA
P115	GCTCTAAAACGCGGCCGCGATCATTTATCTTTCACTGCGGAGA
P199	ACTTCTCCGCAGTGAAAGATAAATGATCGCGGCCGCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGC
P200	CCGTTTACCCTTCCGATTTACTGCTTTACGGCACCTAAAGCCTTCGAGCGTCCC
P251	GGATCGCTTGCCTGTAACTTACACGCGCCTCGTGGCGCGCCCCTGCAGGATTTAAATGG CCGGCCGTTTAAACCCCGGGATTTAAATGGCGCGCCGTTT
P252	TAAACGGCGCGCCATTTAAATCCCGGGGTTTAAACGGCCGGCCATTTAAATCCTGCAGG GGCGCGCCACGAGGCGCGTGTAAGTTACAGGCAAGCGATCC
P253	TAAATGGCCGGCCGTTTAAACCCCGGGATTTAAATGGCGCGCCGTTTAAACCCTGCAGG GGCCGGCCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
P254	GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGCCGGCCCCTGCAGGGTTTAAACGG CGCGCCATTTAAATCCCGGGGTTTAAACGGCCGGCCATTTA
P288	TTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCATGATATAATTAAATTGAAGCTCTAA TTTGTGAGT
P289	GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAATTACAATTTAAATATTTGCTTATAC AATCTTCCTG
P402	ATTACCCTGTTATCCCTACGGCCCTGCATTAATGAATCGGCC
P403	ATTACCCTGTTATCCCTAACTAGTGGATCATCCCCACGCGCC
P1124	AGTTAGGGATAACAGGGTAATAAATAGGAAAGTCTGGTAACTCAGGACACTGTATCTGCT ACGCTGTTTATTTAATTAAATGTCGAAAGCTACATATAAGGAACGTGC
P1125	TAATGCAGGGCCGTAGGGATAACAGGGTAATTGAATAACTGCTCCCGTTGTAAAAGTATC CGTAGTGTGACTGAAACCTTAACTCTTCCTTTTTCAATGGGTAATAACTG
P1128	GTTAGGGATAACAGGGTAATTAAGGTTTCAGTCACACTACGGATACTTTTACAACGGGAG CAGTTATTCATTAATTAAATGTCTGCCCCTAAGAAGATCGTCG
P1129	CAGGGCCGTAGGGATAACAGGGTAATTGACTGTAGTTTTACGGAATAACGGTTCTCCTGA AGTATCACACGCAATACTACCCTATGAACATATTCCATTTTGTAATTTCG
P1132	AGGGATAACAGGGTAATTATTGCGTGTGATACTTCAGGAGAACCGTTATTCCGTAAAACT ACAGTCATTAATTAAATGACAGAGCAGAAAGCCCTAGTAAAGCG
P1133	TAATGCAGGGCCGTAGGGATAACAGGGTAATATTGTAGAATAGTGTCTGTAAATAATCTG AAATCCGCCTCAACGCTCGTGGGTATTTTCTCCTTACGCATCTGTGCGGT

Name	*Sequence (5´-3´)*
P1136	GGGATAACAGGGTAATCACGAGCGTTGAGGCGGATTTCAGATTATTTACAGACACTATTC TACAATTTAATTAAATGACTAACGAAAAGGTCTGGATAG
P1137	CAGGGCCGTAGGGATAACAGGGTAATAAACAGCAGTGTCACGCCCTTCGCAGAGTATTC TATTTTATTGTAAGAGAATTAAATTGAAGCTCTAATTTGTGAGTTTAGTAG
P1139	AGGGATAACAGGGTAATTCTCTTACAATAAAATAGAATACTCTGCGAAGGGCGTGACACT GCTGTTTTTAATTAAATGTCTGTTATTAATTTCACAGGTAGTTCTGG
P1140	TAATGCAGGGCCGTAGGGATAACAGGGTAATTTGTTGTAGCCAGTCGTAGAAGAGATAG CGATTACTCACACCAATACTTTTGATGCGGTATTTTCTCCTTACGC
P1165	AAGCTTGGCTTACTAAAAGCCAGATAACAGTATGCG
P1166	AAGCTTACTGGCTGTGTATAAGGGAGCC
P1173	ACAGCCAGTAAGCTTTAGCTTTTCAATTCAATTCATCATTTTTTT
P1174	GGGTTTAAACGGCGCGCCATTTAAATTCAAACAGGACTAGGATGAGTAGCAGCACGT
P1207	AGCCAGTAAGCTTAACTGTGGGAATACTCAGGTATCGTAAGATGC
P1208	GGGTTTAAACGGCGCGCCATTTAAATTCATTGACCAACGTGGTCACCTGGCA
P1240	ACAGCCAGTAAGCTTAATTCCCGTTTTAAGAGCTTGGTG
P1241	GGGTTTAAACGGCGCGCCATTTAAATTCAATCTTGGTTTCATTTGTAATACGCTTTAC
P1274	AGCCAGTAAGCTTTAGAGGCATCGCACAGTTTTAGCGAGG
P1275	GGGTTTAAACGGCGCGCCATTTAAATTCATGAAAGAGTTGGATTATCCAACTTCTCTATCC AGAC
P1307	ACAGCCAGTAAGCTTAACGACATTACTATATATATAAT
P1308	GGGTTTAAACGGCGCGCCATTTAAATTCAGCAAACTTTCACCAATGGACCAG

Supplementary Table 4: Primers used for amplification and overlap-based cloning of native yeast promoter-terminator pairs.

Name	*Sequence (5´-3´)*	*Target*
P614	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCC	*ADH1* promoter
P615	TTCTTCACCTTTAGACATGCGGCCGCTGAGATAGTTGATTGTATGCTTGGTATAGC	
P618	GATGAATTGTACAAATAAGCGGCCGCGCGAATTTCTTATGATTTATG	*ADH1* terminator
P619	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGTGTGGTCAATAAGAGC	
P620	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGTAACAAAAATCACGATCTGGGTGGG	*PGI1* promoter
P621	TTCTTCACCTTTAGACATGCGGCCGCTTTTAGGCTGGTATCTTGATTCTAAATCG	
P624	GATGAATTGTACAAATAAGCGGCCGCACAAATCGCTCTTAAATATATACCTAAAGAAC	*PGI1* terminator
P625	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGTAGTTTAGTGTTTTTCTTCCAGTGCG	
P626	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTATTGGTTTTTCCAGTGAATGATTATTTGTCGT	*TDH2* promoter
P627	TTCTTCACCTTTAGACATGCGGCCGCTTTGTTTTGTTTGTTTGTGTGATGAATTTAATTTGA	
P630	GATGAATTGTACAAATAAGCGGCCGCATTTAACTCCTTAAGTTACTTTAATGATTTAG	*TDH2* terminator
P631	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGCGAAAAGCCAATTAGTGTG	
P632	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTACTTCTCGTAGGAACAATTTCGGG	*HXT7* promoter
P633	TTCTTCACCTTTAGACATGCGGCCGCCATTTTTTGATTAAAATTAAAAAAACTTTTTGTTT	
P636	GATGAATTGTACAAATAAGCGGCCGCTTTGCGAACACTTTTATTAATTCATGATCACGC	*HXT7* terminator
P637	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAATAACCTGACTCATTAGACACTTTTTGAAGCGGG	
P638	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTACGCACAGATATTATAACATCT	*PGK1* promoter
P639	TTCTTCACCTTTAGACATGCGGCCGCTGTTTTATATTTGTTGTAAAAAGTAGA	
P642	GATGAATTGTACAAATAAGCGGCCGCATTGAATTGAATTGAAATCGATAGATCA	*PGK1* terminator
P643	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAGGCATTAAAAGAGGAGCG	
P644	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGGGGCCGTATACTTACATATAGTAGATG	*TEF2* promoter
P645	TTCTTCACCTTTAGACATGCGGCCGCGTTTAGTTAATTATAGTTCGTTGACCGTATATTC	
P648	GATGAATTGTACAAATAAGCGGCCGCGAGTAATAATTATTGCTTCCATATAATAT	*TEF2* terminator
P649	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGGGTAGCGACGGATTAAT	
P650	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTAATGCTAGTATTTTGGAGATTAATCTCAGTACA	*PYK1* promoter
P651	TTCTTCACCTTTAGACATGCGGCCGCTGTGATGATGTTTTATTTGTTTTGATTGGTGTCT	

Name	*Sequence (5´-3´)*	*Target*
P654	GATGAATTGTACAAATAAGCGGCCGCAAAAAGAATCATGATTGAATGAAGATATT	*PYK1* terminator
P655	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGCATTTATGTACCCATGTATAACCTTCC	
P656	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGTGTCGACGCTGCGGGTATAG	*ENO2* promoter
P657	TTCTTCACCTTTAGACATGCGGCCGCTATTATTGTATGTTATAGTATTAGTTGCTTGGTG	
P660	GATGAATTGTACAAATAAGCGGCCGCAGTGCTTTTAACTAAGAATTATTAGTCTTTTCTGC	*ENO2* terminator
P661	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAATCTCCATCTCCCATATGCATATCACTGTGG	
P662	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCATGCGACTGGGTGAGCATATGTTCCG	*PDC1* promoter
P663	TTCTTCACCTTTAGACATGCGGCCGCTTTGATTGATTTGACTGTGTTATTTTGCGTGAGG	
P666	GATGAATTGTACAAATAAGCGGCCGCGCGATTTAATCTCTAATTATTAGTTAAAGT	*PDC1* terminator
P667	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGCAGTTTTGAATTGAGTAACCA	
P668	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTATAACAATACTGACAGTACTAAATAATTGCC	*FBA1* promoter
P669	TTCTTCACCTTTAGACATGCGGCCGCTTTGAATATGTATTACTTGGTTATGGTTATATATGA	
P672	GATGAATTGTACAAATAAGCGGCCGCGTTAATTCAAATTAATTG	*FBA1* terminator
P673	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGCTATCAAAAACGATAGATC	
P674	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTAGTTTATCATTATCAATACTGCCATTTCAAAG	*TDH3* promoter
P675	TTCTTCACCTTTAGACATGCGGCCGCTTTGTTTGTTTATGTGTGTTTATTCGAAAC	
P678	GATGAATTGTACAAATAAGCGGCCGCGTGAATTTACTTTAAATCTTGCA	*TDH3* terminator
P679	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGAATCTGTGTATATTACTGCATCT	
P680	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTAGTCGTGCAATGTATGACTTTAAGATTTG	*GPM1* promoter
P681	TTCTTCACCTTTAGACATGCGGCCGCTATTGTAATATGTGTGTTTGTTTGGATTATTAAG	
P684	GATGAATTGTACAAATAAGCGGCCGCGTCTGAAGAATGAATGATTTGATGATTTCTTTTTCC	*GPM1* terminator
P685	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATATTCGAACTGCCCATTCAGC	
P686	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTATATCTAGGAACCCATCAGG	*TPI1* promoter
P687	TTCTTCACCTTTAGACATGCGGCCGCTTTTAGTTTATGTATGTGTTTTTTGTAG	
P690	GATGAATTGTACAAATAAGCGGCCGCGATTAATATAATTATATAAAAATATTATCTTCTTT	*TPI1* terminator
P691	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATATATAACAGTTGAAATTTGGATAAGAACATC	

Name	*Sequence (5´-3´)*	*Target*
P692	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTATA	*TEF1* promoter
	GCTTCAAAATGTTTCTACTCCTTTTT	
P693	TTCTTCACCTTTAGACATGCGGCCGCTTTGTAATTAAAAC	
	TTAGATTAGATTGCTATGCT	
P696	GATGAATTGTACAAATAAGCGGCCGCGGAGATTGATAAG	*TEF1* terminator
	ACTTTTCTAGTTGC	
P697	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAGATA	
	GCGCCGATCAAAGTATTTG	
P698	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTACA	*ACT1* promoter
	AGCGCGCCTCTACCTTG	
P699	TTCTTCACCTTTAGACATGCGGCCGCTGTTAATTCAGTAA	
	ATTTTCGATCTTGGG	
P702	GATGAATTGTACAAATAAGCGGCCGCTCTCTGCTTTTGTG	*ACT1* terminator
	CGCGT	
P703	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATACAC	
	GGTCCAATGGATAAACATT	
P704	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTAGC	*CYC1* promoter
	GTTGGTTGGTGGATCAAGCC	
P705	TTCTTCACCTTTAGACATGCGGCCGCTATTAATTTAGTGT	
	GTGTATTTGTGTTTGTGTGTC	
P708	GATGAATTGTACAAATAAGCGGCCGCACAGGCCCCTTTT	*CYC1* terminator
	CCTTTGTCGA	
P709	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAGCTT	
	GCAAATTAAAGCCTTCGAGCGT	
P710	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGAT	*IME2* promoter
	ACCTATTCTCTCAGTAGCTGTTTCACC	
P711	TTCTTCACCTTTAGACATGCGGCCGCAAATGACCTATTAA	
	GTTAAGCTTAGTACTCTTC	
P714	GATGAATTGTACAAATAAGCGGCCGCTAATTTGTCATTTG	*IME2* terminator
	ACTGACGAATCGT	
P715	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATCGTA	
	TGCAGTAATAGGGTTTTTTGAGAATATACCG	
P716	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTTA	*TDH1* promoter
	GTAAAAATGTGCGCACCAC	
P717	TTCTTCACCTTTAGACATGCGGCCGCTTTGTTTTGTGTGT	
	AAATTTAGTGAAGTAC	
P719	GATGAATTGTACAAATAAGCGGCCGCATAAAGCAATCTTG	*TDH1* terminator
	ATGAGGATAATGATTT	
P720	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATATTG	
	ATCCGTACTAATATTGTTAGATGAGCTGC	
P721	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGAA	*ADH1* promoter (s)
	GTTTCACTACCCTTTTTCCATTTGCC	
P615	TTCTTCACCTTTAGACATGCGGCCGCTGAGATAGTTGATT	
	GTATGCTTGGTATAGC	
P722	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTACA	*ADH1* promoter (m)
	ATATGGAAGGGAACTTTACACT	
P615	TTCTTCACCTTTAGACATGCGGCCGCTGAGATAGTTGATT	
	GTATGCTTGGTATAGC	
P723	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTAG	*STE5* promoter
	AAGGCGTATTGCTCAATAGTAGAAAGCAGGC	
P724	TTCTTCACCTTTAGACATGCGGCCGCTTAAAAGTTGTTTC	
	CGCTGTATCCTGTATCTTCC	

Name	***Sequence (5´-3´)***	***Target***
P727	GATGAATTGTACAAATAAGCGGCCGCAGTATACACTAAATTTTATGCAATAATAAAAAGA	*STE5* terminator
P728	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAATAATATTGACCCACCTTCTGCTCCGG	
P729	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGTTGGAATGAAATTTTAATATCATCTATTTCGCAGC	*IPP1* promoter
P730	TTCTTCACCTTTAGACATGCGGCCGCTAGTAAATTGCGCGGCGGG	
P733	GATGAATTGTACAAATAAGCGGCCGCAATATTTTGAATTGAAAATGAGACACCTACCACG	*IPP1* terminator
P734	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATTTGTTGATTGTCATCAGTTTTAGTAAAAAA	
P735	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTACGCTCAACGGAAGAGGAAG	*GPH1* promoter
P736	TTCTTCACCTTTAGACATGCGGCCGCTGTTCAAAATTAAATTAAGTTGAAAGCTGCT	
P739	GATGAATTGTACAAATAAGCGGCCGCCTGGGAGTTTTTTTTATCTCTTTGGCACCC	*GPH1* terminator
P740	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGTTATACTATTCTCTCTTCCTTTCTGGCCTCG	
P741	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTAGAGTGACCGATGAGTTGCATG	*GSY2* promoter
P742	TTCTTCACCTTTAGACATGCGGCCGCTCAAAATTTTTCTCTGAGGTAGTCAAAAC	
P745	GATGAATTGTACAAATAAGCGGCCGCATCCTATGAGGATATAAACAGTATTAAAAAAATCT	*GSY2* terminator
P746	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTACGACCGCCTTAATTCAAAATTTATCTTTCGT	
P747	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTATACAGCTTGTGATTTCGCGCGCC	*SGA1* promoter
P748	TTCTTCACCTTTAGACATGCGGCCGCTTTGCTGCTGTTACTTATTTGAAATCTTGC	
P751	GATGAATTGTACAAATAAGCGGCCGCACAAAAAAAAAAATAAAAGGAAAGCGAGAAG	*SGA1* terminator
P752	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATATCTAATCAATTTATAATATCAGAGTTTTGT	
P753	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTGACCTACCTTAAAACACGTTTCGTTTAAG	*ALG9* promoter
P754	TTCTTCACCTTTAGACATGCGGCCGCCGTTCCCTTATAATTGTGCTTCAGGG	
P757	GATGAATTGTACAAATAAGCGGCCGCTATACAGGTTCTCTAAAACAAGCAGC	*ALG9* terminator
P758	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGGCGTTGGACATCCATGAC	
P759	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTATTCAACGAAACTTACGACAAGATAGTCTTTGG	*UBC6* promoter
P760	TTCTTCACCTTTAGACATGCGGCCGCTACTATTGTACGTACTTTGTTTGCAATTTGC	
P763	GATGAATTGTACAAATAAGCGGCCGCAGCCATTAAATGAACTAAACTTTAGATAAATTTTG	*UBC6* terminator
P764	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAATGGGAGTTTCTGTCTCCCTAGGT	

Name	*Sequence (5´-3´)*	*Target*
P765	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCCA	*KRE11* promoter
	CTTTTTGTCTTATGAGTGCTTTCTTGTATATGC	
P766	TTCTTCACCTTTAGACATGCGGCCGCGGTCAATGAGTTGT	
	GATTTATCAATTGCGC	
P769	GATGAATTGTACAAATAAGCGGCCGCGCGGAACAAGTAT	*KRE11* terminator
	ACATTTTATTAACATAGAC	
P770	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTACGCA	
	ACGCCTAAGTAAAGCTG	
P771	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTAAA	*TFC1* promoter
	TTATTTTTCTTCTTCCTTTTTCTCATACATATC	
P772	TTCTTCACCTTTAGACATGCGGCCGCTTTCACGATTTGGA	
	AGATATTTTCGC	
P775	GATGAATTGTACAAATAAGCGGCCGCTAAGTAGGTTTTCT	*TFC1* terminator
	TTTATTTCAATATCAAAA	
P776	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTACCGTT	
	CTTTCTGGAGGCATTCT	
P808	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCTA	*CWP2* promoter
	ATAGACAAGGTGCTATGAGTGAA	
P809	TTCTTCACCTTTAGACATGCGGCCGCTTTTTTTCTTGTTAG	
	TGTGTAGCGAA	
P810	GATGAATTGTACAAATAAGCGGCCGCGAAATCTCTGATTT	*CWP2* terminator
	TTTATAATATCTATATGGCT	
P811	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAAAAA	
	GATCCCACCGTAGCACACC	
P812	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCCT	*RPL3* promoter
	TTCCTGCGGGATAGCTTTTGC	
P813	TTCTTCACCTTTAGACATGCGGCCGCGATTGATTGTTGTA	
	GTAACTGTGTTGTTC	
P814	GATGAATTGTACAAATAAGCGGCCGCGAAGTTTTGTTAGA	*RPL3* terminator
	AAATAAATCA	
P815	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGCAAT	
	ATAATGGACGGGT	
P816	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCAA	*RPL8a* promoter
	CATAAATAATTTCTATTAACAATGTAATTTCCA	
P817	TTCTTCACCTTTAGACATGCGGCCGCTTCGAATTAGTTGT	
	TTTGATGTGATAATAGAGGG	
P818	GATGAATTGTACAAATAAGCGGCCGCATTGAAAATGAGAA	*RPL8a* terminator
	ATTTTGCATAAAAAA	
P819	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAAATGA	
	TCGTGGAGTTTCAAACATCG	
P820	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTAAC	*CCW12* promoter
	CAGGGCAAAGCAAAATAAAAG	
P821	TTCTTCACCTTTAGACATGCGGCCGCTATTGATATAGTGT	
	TTAAGCGAATGACAGAAG	
P822	GATGAATTGTACAAATAAGCGGCCGCACTTAGTTTATTAT	*CCW12* terminator
	TATTTATACATTCTAAATTT	
P823	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTAATGCT	
	ATGTAATAGACAATAAAACCATGT	
P824	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTGT	*HHF2* promoter
	GGAGTGTTTGCTTGGATCC	
P825	TTCTTCACCTTTAGACATGCGGCCGCTATTTTATTGTATTG	
	ATTGTTGTTTTTGCTACTC	

Name	***Sequence (5´-3´)***	***Target***
P826	GATGAATTGTACAAATAAGCGGCCGCACAATCGGTGGTTAAACAATCG	*HHF2* terminator
P827	AATAAAAAAAAAATGATGAATTGAATTGAAAAGCTATTTCTTTTATTGAGACTTATTTTTAATATATATAG	
P828	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTCTCAAATCCATTGGTGTTTCCATTAGGCC	*EFT2* promoter
P829	TTCTTCACCTTTAGACATGCGGCCGCCTTGCAATTATGTTATAGTTTGTGTTTGTGG	
P830	GATGAATTGTACAAATAAGCGGCCGCGAATGGTTAAACAATTTTTAATTATT	*EFT2* terminator
P831	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTAGCTTTCATGCTAGGAATGTG	
P832	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTAGTTGTAGAGTAAGTTGTTGTTGGTTTG	*HTA2* promoter
P833	TTCTTCACCTTTAGACATGCGGCCGCTATATATTAAATTTGCTCTTGTTCTGTACTTTCC	
P834	GATGAATTGTACAAATAAGCGGCCGCGAACTGAGTTGAAAAGAAACAAAGCA	*HAT2* terminator
P835	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTACTGTAATATATCTTTATAACATGTATACTAATGA	
P836	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTTTTAGAGTCTTCCTTGACCAATCATCTTATTCG	*UBI4* promoter
P837	TTCTTCACCTTTAGACATGCGGCCGCAATCTATTAGTTAAAGTAAAGTGGGAGGG	
P838	GATGAATTGTACAAATAAGCGGCCGCTCAGTCCTCGCAATATTTTCATTATGTC	*UBI4* terminator
P839	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTATTAACTGTTCTTAATTTTCCTTTTTTCTGCC	
P840	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTCTGTCTTCCTCTACTGGGCTTATC	*HAC1* promoter
P841	TTCTTCACCTTTAGACATGCGGCCGCAGTGGCGGTTGTTGTCGTAGG	
P842	GATGAATTGTACAAATAAGCGGCCGCAGACAATCGCAAGAGGGTATAATTTTTTTTTTCTCG	*HAC1* terminator
P843	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTACTCATTGGCCCGGAGTTAGGGGG	
P844	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTACTTAAGAGCTAAACGAAAAGATAAGTATCTCGC	*ARF1* promoter
P845	TTCTTCACCTTTAGACATGCGGCCGCTTTTAATTCTATTAAACCTGTTTGAGTTCTTTC	
P846	GATGAATTGTACAAATAAGCGGCCGCAAATTCTAGAATATGGATCAAATACGCTTG	*ARF1* terminator
P847	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTAATTAATACAGTCCCCAGATTGAGAAAG	
P849	GTCACCAGCCAACAATGTGCGGATGGCGTTACAACTTTAAACGTGTGGTTTATGGGTGCACCAGGG	*ZEO1* promoter
P850	TTCTTCACCTTTAGACATGCGGCCGCTATTAATTGATATAAACG	
P851	GATGAATTGTACAAATAAGCGGCCGCTCACACTTGAATGTGCTGC	*ZEO1* terminator
P852	AATAAAAAAAAAAATGATGAATTGAATTGAAAAGCTATATTTTTACTTTCTTATCAAACTTTTTTTCCTCCGG	

Supplementary Result 1: A library of constitutive promoters for gene expression in yeast

For the successful construction and expression of multi-gene pathways it is often necessary to adjust the expression level of individual genes. To provide a first starting point for researchers working with *S. cerevisiae*, we cloned a panel of 40 constitutive yeast promoters, each in combination with the terminator region from the same gene. To easily assess the promoter strengths, we cloned promoter-yEGFP-terminator cassettes of each native promoter-terminator pair into low-copy Level 1 vector pL1-lc and transformed the constructs into *S. cerevisiae*. By using flow cytometry, the relative yEGFP fluorescence for each promoter-terminator pair was determined after 16 h of batch culture. To address two widely used growth conditions, we used the carbon sources glucose or galactose, respectively. Our results show a wide range of yEGFP fluorescence intensities mediated by the 40 different promoter-terminator pairs (**Supplementary Figure 2**). While the strong *CCW12* and *ENO2* promoters shape the upper end of the expression scale with a fluorescence value of 117% and 57% relative to *TDH3*, the yEGFP fluorescence mediated by e.g. *HXT7*, *SGA1*, and *IME2* is less than 1%, relative to *TDH3*. In our system, these promoters confer fluorescence comparable to, or only slightly above background fluorescence from wild-type cells. However, most promoters driving the yEGFP expression result in fluorescence values between 1% and 50% relative to *TDH3* and should therefore be suitable for many applications. For a new assembly, the selected promoter and the corresponding terminator can be amplified by PCR from the provided plasmids with the necessary overlaps for Level 0 cloning incorporated in the primers. To reduce PCR steps and primer expenses, it is also possible to amplify and clone the complete promoter-terminator cassette into a given Level 0 vector. The resulting plasmid can afterwards be linearized using *Not*I and the desired CDS can be inserted by ligation- or recombination-based cloning.

Sequences of plasmids generated in this work

GenBank files for all Level 0, Level 1 and Level 2 plasmids and the promoter/terminator collection are available via the journals homepage and at www.assemblx.org.

4.2 PhiReX – a Programmable and Red Light-regulated Protein Expression Switch for Yeast

Lena Hochrein[1], Fabian Machens[1], Katrin Messerschmidt[1], Bernd Mueller-Roeber[2,3,*]

[1]University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[2]University of Potsdam, Department of Molecular Biology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[3] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany

* To whom correspondence should be addressed.

The manuscript has been submitted to Nucleic Acids Research as a standard manuscript. After the submission of this dissertation, it was published by Nucleic Acids Research in a revised version after reviewer and editor assessments:
Lena Hochrein, Fabian Machens, Katrin Messerschmidt, Bernd Mueller-Roeber; PhiReX: a programmable and red light-regulated protein expression switch for yeast. *Nucleic Acids Research*, Volume 45, Issue 15, Pages 9193–9205, https://doi.org/10.1093/nar/gkx610 (2017).

ABSTRACT

Highly regulated induction systems enabling dose-dependent and reversible fine-tuning of protein expression output are beneficial for engineering complex biosynthetic pathways. To address this, we developed PhiReX, a novel red/far-red light-regulated protein expression system for use in *Saccharomyces cerevisiae.* PhiReX is based on the combination of a customizable synTALE DNA-binding domain, the VP64 activation domain and the light-sensitive dimerization of the photoreceptor PhyB and its interacting partner PIF3 from *Arabidopsis thaliana.* Robust gene expression and high protein levels are achieved by combining genome integrated red light-sensing components with an episomal high-copy reporter construct. The gene of interest as well as the synTALE DNA-binding domain can be easily exchanged, allowing the flexible regulation of any desired gene by targeting endogenous or heterologous promoter regions. To allow low-cost induction of gene expression for industrial fermentation processes, we engineered

yeast to endogenously produce the chromophore required for the effective dimerization of PhyB and PIF3. Time course experiments demonstrate high-level induction over a period of at least 48 h.

INTRODUCTION

In recent years, *Saccharomyces cerevisiae* has proven to be a valuable host for diverse synthetic biology applications. Examples range from the production of medically relevant substances like taxadiene (1), artimisinic acid (2) or opioids (3) to compounds of technical importance, e.g. biofuels (4-6). Efficient protein production often requires robust molecular tools for the transcriptional regulation of both, heterologous and endogenous genes.

Chemically inducible promoters, available in a large number in yeast (7-12), do not satisfy the high demands of synthetic biology: reversible or dose-dependent induction is in most cases not easily accomplished and certain chemicals are expensive or show toxic effects in higher concentrations. To overcome these drawbacks, different light-responsive expression systems have been implemented in *S. cerevisiae,* mostly based on plant-derived optical dimerizers, combining a DNA-binding domain (DBD) and a transcriptional activation domain (AD) upon light application (13-16). However, none of the previously reported light-inducible expression systems for yeast allows programmable targeting of DNA-binding sites (DBS), which is a requirement for the orthogonal regulation and the specific control of endogenous genes. Chemically inducible synthetic transcription factors (synTFs) with customizable DBDs, based on zinc finger proteins (ZFPs), transcription activator like effectors (TALEs) or CRISPR/Cas9 have previously been reported in yeast (17-19).

To fill this gap, we developed PhiReX (**P**rogrammable, **h**ighly **i**nducible **Re**d light e**X**pression system). The system is based on the photoreceptor PhyB and its interacting partner, the helix-loop-helix protein PIF3, from *Arabidopsis thaliana* (13). The photoreceptor binds to a chromophore, which triggers a photon-dependent switch between two conformations: the red light (λ=660 nm) absorbing Pr form and the far-red

light (λ=730 nm) absorbing Pfr form. The active PhyB Pfr dimerizes specifically with PIF3 (20). This interaction can be used to design light-sensitive synTFs, by fusing the programmable DBD of a synthetic TALE (synTALE) to PhyB and an AD to PIF3. TALEs are natural type III effector proteins secreted by *Xanthomonas sp.* and contain a region consisting of highly conserved tandem repeat modules, each usually 34 amino acids (aa) long (21). Positions 12 and 13 of each repeat module, the so-called repeat variable di-residues (RVDs), are hypervariable. Recent studies showed a strong correlation between unique RVDs and the corresponding nucleotide in the targeted DBS. This results in a simple code that allows predicting the DNA-target site for any given combination of RVDs (22,23). Consequently, it is possible to design customized synTALE-DBDs, which specifically target any plasmid-born or genome located promoter region upstream of the gene of interest (GOI).

Phytochromobilin (PϕB), the chromophore of PhyB, is naturally not available in yeast. Therefore, phycocyanobilin (PCB), a PϕB analog that can serve as a cofactor of PhyB, needs to be supplied exogenously (24-26). To overcome this, we engineered yeast to produce both chromophores in sufficient amounts. Comparable approaches have already been reported in bacteria, mammalian cells and *Pichia pastoris*, but not in yeast (27-30). With this, we guarantee a nearly gratuitous gene induction system, perfectly suited for large-scale applications.

We established two versions of the PhiReX system enabling either the tight control of gene expression with low basal expression levels and moderate expression output or slightly elevated background expression with high expression output, comparable to the strong yeast *TDH3* promoter. As PhiReX is compatible with the previously reported multi-gene assembly strategy AssemblX (31) it represents a perfect tool for pathway engineering.

MATERIAL AND METHODS

Strains, growth conditions and reagents

Escherichia coli strains DH5α and NEB5α NEB10β (both New England Biolabs, Frankfurt am Main, Germany) were employed for DNA cloning. Cells were grown at 37°C and 230 rpm in Luria-Bertani (LB) medium, with kanamycin or ampicillin (50 mg/ml) as selection marker. For yeast experiments, *S. cerevisiae* strain YPH500 (ATCC 76626) was cultured at 30°C and 230 rpm in Yeast extract Peptone Dextrose Adenine (YPDA)-rich medium or in appropriate Synthetic Dextrose (SD) media for selection of transformed cells.

Constructs, DNA assembly and cell transformation protocols

Plasmid pLOGI, used as template for PCR amplification of the reporter gene yEGFP, was a gift from Tom Ellis, Imperial College London (32). Plasmids p414-TEF1p-Cas9-CYC1t and p426-SNR52p-gRNA.CAN1.Y-SUP4t were gifts from George Church (Addgene plasmids # 43802 and 43803). pMLM3705 was a gift from Keith Joung (Addgene plasmid # 47754) (33). All plasmids reported in this study were established using AssemblX (31) in combination with the overlap-based DNA assembly methods SLiCE (34), Gibson assembly (35), NEBuilder HiFi DNA Assembly (New England Biolabs, Frankfurt am Main, Germany) and TAR (36). *In vitro* overlap-based cloning methods were used as previously reported (34,35). To this end, DNA fragments harboring 25- to 36-bp long homology regions were produced by PCR and assembled with linearized and dephosphorylated plasmids. *E. coli* transformations were performed according to the High Efficiency Transformation Protocol from New England Biolabs. The correctness of DNA assemblies was verified by sequence analysis (LGC Genomics, Berlin, Germany). For *in vivo* DNA assembly, *S. cerevisiae* transformations were carried out following the LiAc/SS carrier DNA/PEG method by Gietz and Schiestl (37). Information about all yeast strains obtained in this study is provided in **Table 1**. The AssemblX toolkit enables level-based multi-gene assemblies (31). While Level 0 plasmids are generally employed to assemble transcriptional units (TUs), Level 1 vectors allow the combination of up to five Level 0 modules. AssemblX was used in a preliminary version with vector backbones and homology regions slightly different from

the final version (sequences available in **Supplementary Data File 1**). **Supplementary Table 1** lists all parts used in this study with appropriate sources and, if applicable, sequence modifications. Sequences of all constructs described here are given in **Supplementary Data File 1**. Sequences of oligonucleotides are available in **Supplementary Table 2**

Generation of Level 0 and Level 1 constructs containing the light-inducible split synTF

For the construction of each multi-gene containing Level 1 plasmid, three appropriate Level 0 vectors (A1, A2 and A3) were constructed. Each Level 0 vector obtains homology regions for recombination with the Level 1 vector backbone or the neighboring Level 0 fragment. Each Level 1 vector contains three cassettes: (i) the synTALE-DBD-PhyB-fusion, (ii) the PIF3-AD-fusion, (iii) the reporter cassette, containing the reporter gene *yEGFP*, located downstream of a synthetic promoter, which consists of the synTALE-DBS and a *Cyc1* minimal promoter. Multi-gene containing Level 1 constructs were generated by combining three Level 0 fragments, released from the preassembled AssemblX Level 0 constructs A1, A2 and A3 with the Level 1 vector backbone pL1A_Leu_lc, a CEN/ARS plasmid with selection marker *LEU2* and homology regions A0 and B0. The different Level 1 plasmids and the appropriate Level 0 fragments are given in **Supplementary Table 3.**

For Level 0 plasmid pL0_RL_A1, the CDS for *PIF3*, the NLS of the simian virus 40 (SV40-NLS), and a quadruple tandem repeat of the herpes simplex virus VP16-AD (VP64-AD) were cloned as a fusion transcript under control of the strong constitutive yeast *FBA1* promoter and terminator into the AssemblX Level 0 A1 vector backbone pL0_A0-A1. For construct pL0_RL_A2_CT the CDS for PhyBNT (aa 1-621 of PhyB), the yeast Cdc1-NLS and the empty synTALE-backbone without RVD region, and for construct pL0_RL_A2_NT the synTALE-backbone, Cdc1-NLS and PhyBNT were cloned as fusion transcripts in the indicated order under control of the *TDH3* promoter and terminator into the AssemblX Level 0 A2 plasmid pL0_A1-A2. Syn-TALE1-RVD region was assembled to target sequence *Jub1.1* (TCTATAAGATCTTGTGTGC, a region of the *JUB1* promoter from *Arabidopsis thaliana*; AGI code: At2g43000*)* following the protocol described in Morbitzer *et al.* (38), and subcloned as a *Bam*HI and *Age*I fragment into

equally digested plasmids pL0_RL_A2_CT and pL0_RL_A2_NT, resulting in constructs pL0_RL_A2_CT_TALE and pL0_RL_A2_NT_TALE, respectively. For vector pL0_RL_A3 the synTALE-DBS containing four repeats of *Jub1.1* were cloned upstream of the *Cyc1* minimal promoter. The CDS of *yEGFP* and the *Cyc1* terminator were cloned downstream of the resulting synthetic promoter into the AssemblX Level 0 A3 plasmid pL0_A2-B0. Plasmid pL0_RL_A3_w/o_DBS was cloned by inserting a stuffer sequence instead of the synTALE-DBS. Plasmid pL0_RL_A3_ADH1p was cloned by inserting the *ADH1* promoter instead of the synTALE-DBS and the *Cyc1* minimal promoter with overlap based cloning methods.

Generation of constructs required for the genome integration of the red light-inducible split synTF

For genome integration of the red light-inducible split synTF, Level 0 fragments were released from Level 0 vectors pL0_RL_A1 and pL0_RL_A2_CT_TALE and cloned together with the *LEU2* selection marker into plasmid pFM75, resulting in plasmid pRL_CT_GI. Plasmid pFM75 is designed for genome integration into the *ura3-52* locus of the yeast genome, after *Pme*I digestion. The reporter cassette remained on plasmid. For this, Level 0 fragments pL0_RL_A3, pL0_RL_A3_w/o_DBS and pL0_RL_A3_ADH1p were released from their backbones and cloned into pL0_A2-A3_Ura, a Level 0 vector with 2-micron origin and selection marker *URA3*, resulting in plasmids pRL_yEGFP_hc, pRL_yEGFP_ADH1p_hc and pRL_yEGFP_w/o_DBS_hc.

Generation of constructs for minimization of basal expression levels

To investigate reasons for the high basal expression levels of pRL_NT and pRL_CT, Level 1 constructs pRL_CT_w/o_TALE, pRL_CT_w/o_PhyBNT_w/o_PIF3-AD and pRL_CT_w/o_PIF3-AD were constructed. For this, Level 0 plasmids pL0_RL_A2_TALE_w/o_PhyB and pL0_RL_A2_CT_w/o_TALE were cloned. For construction of pL0_RL_A2_TALE_w/o_PhyB the yeast Cdc1-NLS and the synTALE-backbone were cloned as fusion transcript under control of the *TDH3* promoter and *TDH3* terminator into the AssemblX Level 0 A2 vector backbone pL0_A1-2. SynTALE1 and plasmid pL0_A1-2 were digested with restriction endonucleases *Bam*HI and *Age*I and the released synTALE-repeats were cloned into the linearized backbone. For

pL0_RL_A2_CT_w/o_TALE the Cdc1-NLS and the CDS of PhyBNT were cloned under control of the *TDH3* promoter and *TDH3* terminator into the AssemblX Level 0 A2 plasmid pL0_A1-2. Level 1 constructs were generated by combining Level 0 fragments, released from the different AssemblX Level 0 constructs A1, A2 and A3, and Level 1 vector backbone pL1A_Leu_lc, as described in **Supplementary Table 3**. To lower the basal expression levels of pRL_CT_GI, the Cdc1-NLS incorporated in the PhyB-synTALE fusion and also the NLS sequences present in the synTALE-backbone were deleted. For this, plasmids pRL_CT_GI_del1 and pRL_CT_GI_del2 were constructed (**Supplementary Table 4**). All primers used for mutagenesis are listed in **Supplementary Table 2**. For the construction of pRL_CT_GI_del1 a PCR product obtained with primers L772/L762 and template pL0_RL_A2_CT_TALE was inserted into *Swa*I/*Aat*II-linearized pRL_CT_GI. To obtain construct pRL_CT_GI_del2, a PCR product obtained with primers L662/L767 and template pRL_CT_GI was inserted into *Age*I-linearized Level 1 vector pRL_CT_GI. Furthermore, the nuclear export signal (NES) of the protein kinase inhibitor peptide (PKI) was added in between PhyBNT and the synTALE of plasmids pRL_CT_GI_del1 and del2, resulting in plasmids pRL_CT_GI_del1_1 and del2_1. For these constructs, PCR products obtained with primers L795/L796 and template pRL_CT_GI were inserted into *Bam*HI/*Sac*I-digested plasmids pRL_CT_GI_del1 and pRL_CT_GI_del2, respectively. For the construction of pRL_CT_GI_del2_2, which contains a PKI-NES at the C-terminal domain of the PhyBNT-synTALE fusion, a PCR product obtained with primers L767/L797 and template pRL_CT_GI was inserted into *Age*I-digested plasmid pRL_CT_GI_del2. Plasmids pRL_CT_GI_del1, pRL_CT_GI_del2, pRL_CT_GI_del1_1, del2_1 and pRL_CT_GI_del2_2 were genome integrated and cotransformed with the reporter construct pRL_yeGFP_hc to yield strains RL_CT_GI_del1/del2, RL_CT_GI_del1_1/del2_1 and strain RL_CT_GI_del2_2.

A Cas9-based genome editing approach was followed for the construction of additional variants of the strains, given in detail in **Supplementary Table 4 and Figure 4B**: (i) Addition of a PKI-NES to the C-terminus of the synTALE-PhyBNT-fusion of strain RL_CT_GI_del1 to yield RL_CT_GI_del1_2*. (ii) Addition of a Cdc1-NLS to the N-terminus of the PIF3-SV40-NLS-AD-fusion of strains RL_CT_GI_del1 and RL_CT_GI_del2 to obtain RL_CT_GI_del1_3* and RL_CT_GI_del2_3*. (iii) Deletion of

the SV40-NLS of the PIF3-AD-fusion of strains RL_CT_GI_del1 and RL_CT_GI_del2 to yield strains RL_CT_GI_del1_4* and RL_CT_GI_del2_4*. (iv) Addition of a Cdc1-NLS to the N-terminus of the PIF3-SV40-NLS-AD-fusion of strains RL_CT_GI_del1_2* and RL_CT_GI_del2_2* to obtain RL_CT_GI_del1_5* and RL_CT_GI_del2_5*. (v) Deletion of the SV40-NLS of strains RL_CT_GI_del1_2* and RL_CT_GI_del2_2* to yield RL_CT_GI_del1_6* and RL_CT_GI_del2_6*.

Shortly described, respective yeast cells, given in **Supplementary Table 4**, were transformed with the Cas9 expression plasmid p414-TEF1p-Cas9-CYC1t (39), the appropriate sgRNA expression plasmid and the donor DNA as a repair template. To generate sgRNA cloning and expression plasmid pFM021, we modified plasmid p426-SNR52p-gRNA.CAN1.Y-SUP4t (39) by exchanging the *URA3* marker for the *LEU2* marker gene and by introducing a *Bam*HI-*Eco*RI cloning site between the structural sgRNA backbone and the *SNR52* promoter. Individual sgRNA expression cassettes were generated via NEBuilder HiFi DNA Assembly (New England Biolabs) of double-stranded oligonucleotides into pFM021. As previously described, the donor DNA was generated from two overlapping oligonucleotides (40). Oligonucleotides used for sgRNA cloning and sgRNA sequences are given in **Supplementary Table 2 and 4**. Strains RL_CT_GI_del1_2*/3*/4*/5*/6* and RL_CT_GI_del2_3*/4*/5*/6* were cotransformed with the reporter construct pRL_yEGFP_hc to yield strains RL_CT_GI_del1_2/3/4/5/6 and RL_CT_GI_del2_3/4/5/6. Details about the generated strains are given in **Table 1**.

Generation of constructs for chromophore biosynthesis in yeast

Two enzymes are required for the production of the PhyB chromophores PCB and PΦB*:* a heme oxygenase (HO) and a bilin reductase (BR). Different combinations of native and modified sequences of HOs, i.e., *HY1* from *A. thaliana* and *Hmx1* from *S. cerevisiae*, and BRs, i.e., *HY2* from *A. thaliana* and *PcyA* from *Synechocystis sp*., were tested. Details are given in **Supplementary Table 1**. HOs were set under control of the yeast *PGK1* promoter and terminator and cloned into linearized Level 0 vector pL0_A3-A4 resulting in Level 0 plasmids pPCB1, pPCB2, pPCB3 and pPCB5. BRs were cloned under control of the yeast *ADH1* promoter and the *ADH1* terminator and cloned into linearized AssemblX Level 0 vector backbone pL0_A4-AR, resulting in Level 0 constructs pPCB6b, pPCB8b and pPCB5b. For the construction of plasmids

pRL_GI_PCB_27 to pRL_GI_PCB_34, different combinations of Level 0 vectors, given in **Supplementary Table 3**, were inserted into linearized vector backbone pRL_GI_PCB1. Plasmid pRL_GI_PCB1 allows genome integration of Level 1 cassettes into the *his1-Δ200* locus after release of the Level 1 cassette by restriction with *Pme*I or *Sbf*I, respectively.

Generation of plasmid pYES2CT_yEGFP and expression experiments

In order to compare PhiReX1.0+ with the well-established, galactose-inducible *GAL1* promoter, the commercially available expression vector pYES2/CT (Thermo Fisher Scientific, Waltham, MA, USA) was digested with *Hin*dIII and *Pme*I and the CDS of the reporter gene yEGFP was cloned in between the two recognition sites to yield plasmid pYES2/CT_yEGFP. Galactose induction was done according to the Invitrogen user manual for part no. 25-0304 (Rev. date: 30 November 2009). Precultures and uninduced samples were maintained in glucose containing medium.

PCR methods

PCR amplifications were done using either Q5 DNA polymerase (New England Biolabs) or Phusion polymerase (Life Technologies, Darmstadt, Germany), according to the manufacturer's instructions.

Light induction experiments

If not mentioned otherwise, cells were inoculated in 1 ml appropriate SD-medium, containing 2% glucose, and incubated for 24 h at 30°C and 230 rpm in 12-well cell culture plates (TPP Techno Plastic Products AG, Trasadingen, Switzerland; order number 92412). Cultures were immediately irradiated with a 30-sec far-red light pulse and grown for 6 h at 30°C and 230 rpm in the dark. Thereafter, induced cells were treated with a 30-sec red-light pulse, followed by 10-sec red light pulses every 30 min over a period of 16 h, while uninduced cells were maintained in darkness. Protein production was stopped by adding 50 µl 10 mg/ml cycloheximide. Fluorescence output of single cells was measured as described before (31). If not stated differently, the presented results are mean values calculated from three biologically independent

measurements, each with three technical replicates. Standard deviation is indicated by error bars.

Light sources

LED light sources for red light (λ = 640 – 680 nm, peak at 661 nm) and far-red light (λ = 720 – 760 nm, peak at 740 nm) were installed by COMPLED Solutions GmbH (Dresden, Germany) into an INFORS HT Multitron Pro (INFORS, Bottmingen, Switzerland). For each wavelength, 40 LEDs were placed at the inner ceiling of the incubator. Light-reflecting foil was installed on the walls to ensure even illumination at a light intensity of 2.8 mW/cm^2 for red light, and 6.9 mW/cm^2 for far-red light. All light-sensitive manipulations were done in a dark room under green safelight using a standard LED strip.

RESULTS

Establishment of a red light-induced split synTF

The PhiReX system described here is a red/far-red light-dependent expression system based on two fusion proteins: a synTALE-DBD linked to an N-terminal truncation of the light-sensitive photoreceptor PhyB (PhyBNT) and a fusion of the PhyB interacting partner PIF3 with the VP64-AD (**Figure 1A**). The synTALE-DBD was programmed to specifically target the synTALE-DBS, which is located upstream of a minimal promoter and the reporter gene *yEGFP*. The synTALE-PhyBNT fusion protein binds the synTALE-DBS, regardless of the light condition (**Figure 1B**). However, dimerization with the PIF3-AD fusion protein and subsequent activation of gene expression are red light-induced. We designed two multi-gene CEN/ARS plasmids, pRL_NT and pRL_CT, which differ with respect to the PhyBNT-synTALE-fusion: the synTALE-DBD was fused to the N-terminus (NT) or C-terminus (CT) of PhyBNT, respectively, whereby Cdc1-NLS links the two components (**Figure 1A**).

As the natural chromophore of the plant photoreceptor PhyB, PϕB, is not synthesized in yeast, the analogous compound PCB which can act as a chromophore of PhyB, was supplied to the medium (24,25). To show PCB-dependent and red light-specific

activation of protein expression, we measured yEGFP fluorescence of RL_NT and RL_CT cells grown in the dark or treated with red light pulses in drop-out medium containing different PCB concentrations (0–50 µM). For both cell types, we observed growing red light-mediated yEGFP output with increasing PCB concentrations, reaching a plateau at concentrations of 25 µM (**Figure 2**). All following experiments were therefore carried out with media containing 25 µM PCB. A very low yEGFP signal was observed for the negative controls lacking the synTALE-DBS. The absolute values of yEGFP fluorescence and the fold induction (~3.3 fold with 25 µM PCB) were similar in RL_NT and RL_CT cells.

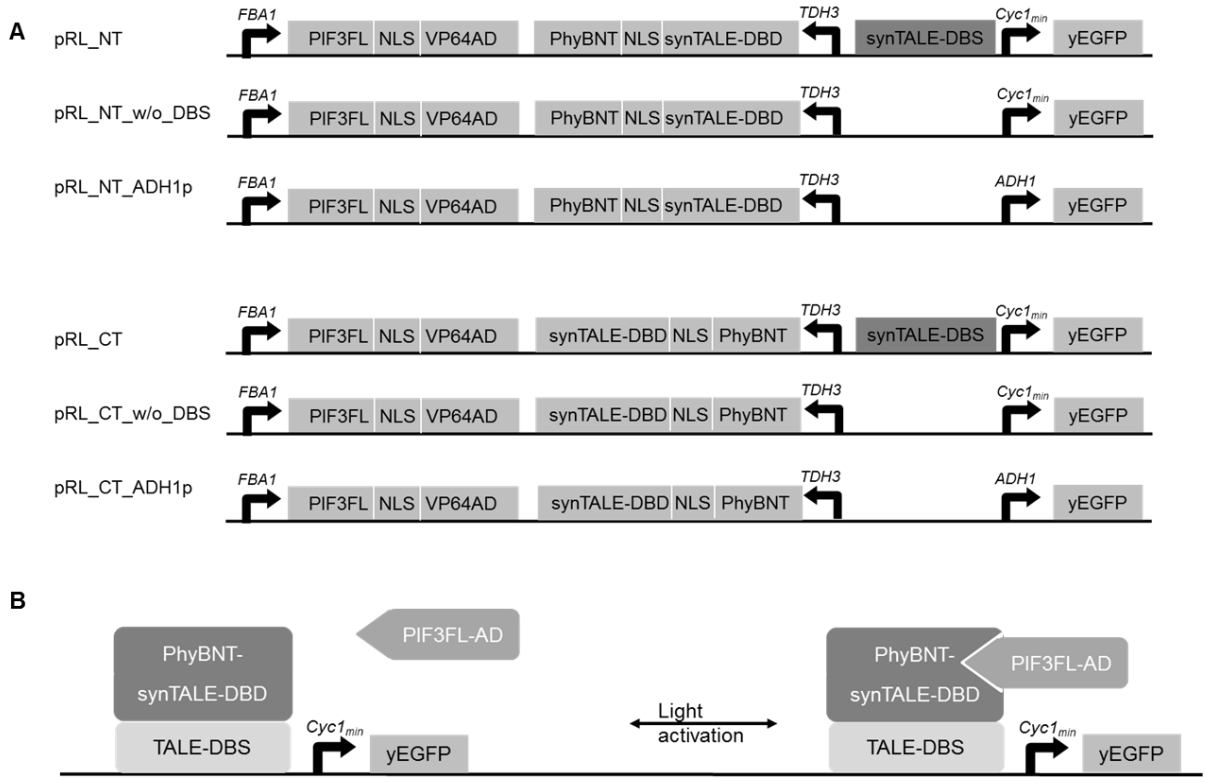


Figure 1: Schematic overview and design of the PhiReX-system. A Design of light-responsive constructs pRL_NT and pRL_CT, negative control plasmids lacking the synTALE-DBS (pRL_NT_w/o_DBS and pRL_CT_w/o_DBS), and plasmids with *yEGFP* under control of the constitutive yeast *ADH1* promoter as a technical control (pRL_CT_ADH1p and pRL_NT_ADH1p). **B** Mode of action of pRL_NT and pRL_CT. In the dark or in far-red light (λ=740 nm) the PhyB-synTALE-DBD-fusion binds to its DBS. Stimulation with red light (λ=660 nm) induces dimerization of the split transcription factor leading to activation of yEGFP expression.

However, RL_NT showed no light-induced yEGFP fluorescence in the absence of PCB, while RL_CT exhibited red light-dependent yEGFP signal (fold induction ~2.4) (**Figure 2**). This observation is consistent with a previously reported phytochrome A (PhyA) - based system, which also showed red light-triggered expression without exogenous supplied chromophore in *S. cerevisiae*. Considering these observations, it was proposed that an unidentified cellular metabolite functions as a chromophore for the photoreceptor (16). As chromophore-independent light induction was only observed for RL_CT, this system may be more sensitive than RL_NT and was therefore used for further experiments.

Combining genome integrated regulators and an episomal reporter cassette results in boosted fluorescence output upon red light induction

To achieve stable and robust protein expression, we integrated the C-terminal version of the PhyBNT-synTALE-fusion and the PIF3-AD-fusion into the *ura3-52* locus of the yeast genome. The expression cassette, containing the target promoter and the reporter, remained on a plasmid to enable an easy and flexible exchange of the GOI (**Figure 3A**). To design a system generating high expression output, a high-copy 2-micron origin of replication was selected for the reporter construct, as opposed to plasmid pRL_CT, which harbors the low-copy CEN/ARS replication origin. The optimized system RL_CT_GI yields a reliable induction in usually more than 80% of cells. In comparison, the percentage of activated cells in RL_CT varied gravely between less than 10% to much higher values reaching 80%. yEGFP fluorescence obtained for RL_CT_GI was 55-fold higher (fold induction ~6.3), compared to RL_CT (fold induction ~3.9) (**Figure 3B**). However, also basal expression levels increased gravely. As RL_CT_GI_ADH1p and RL_CT_GI_w/o_DBS (**Table 1**) show an obvious increase as well, the boost of reporter output is most likely due to the generally higher expression output from high-copy plasmids.

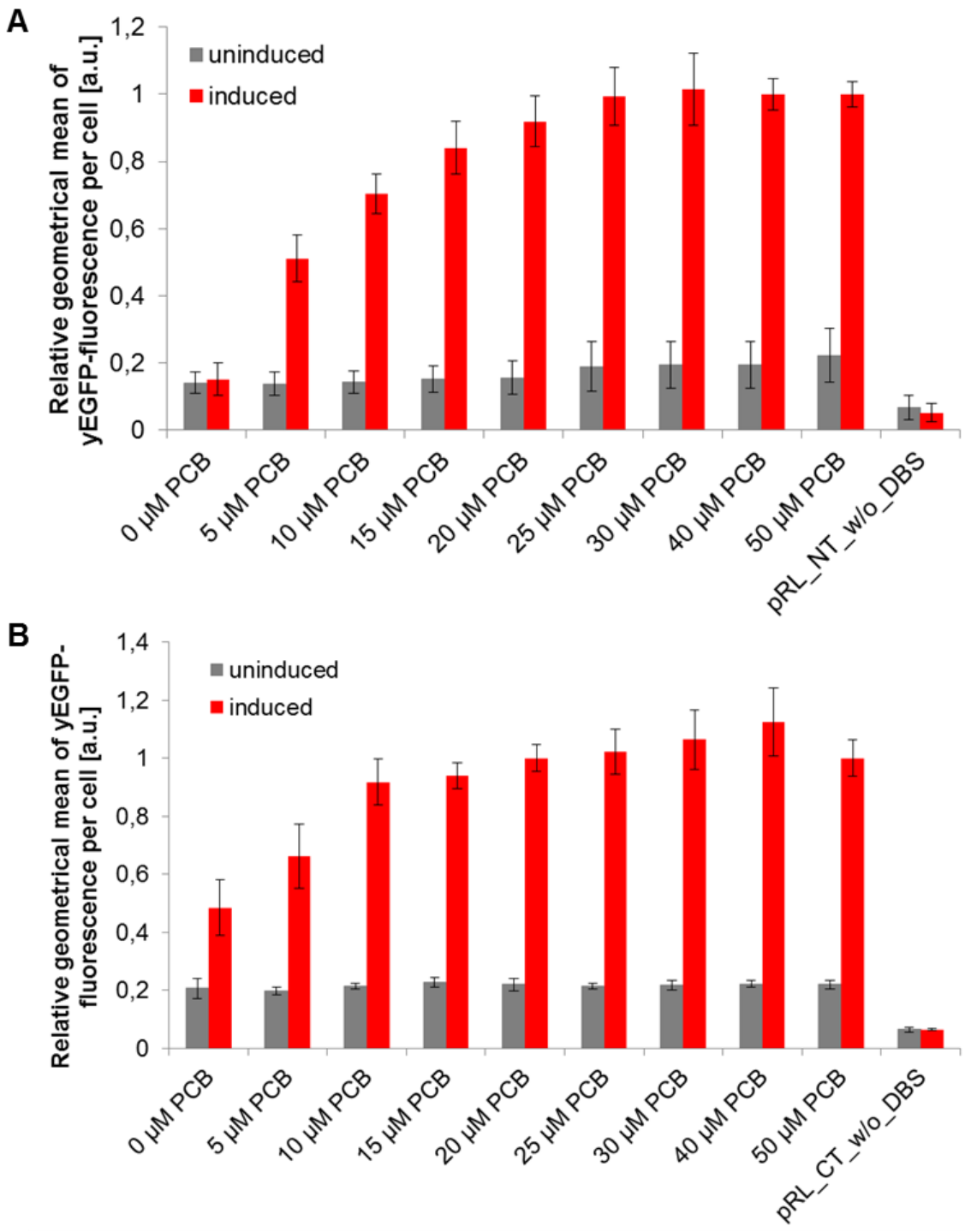


Figure 2: PCB-dependent yEGFP output of the red light sensing expression systems. A RL_NT and **B** RL_CT yeast cells were grown in SD-Leu medium with 0 - 50 µM PCB for 6 h in darkness. Induced cells were treated with 10-sec red light pulses every 30 min for 16 h. Uninduced cells remained in the dark. Yeast cells RL_NT_w/o_DBS and RL_CT_w/o_DBS were grown in SD-Leu medium with 25 µM PCB. Output of yEGFP was measured via flow cytometry. A.u.: Arbitrary units.

Improving fold induction by minimizing basal expression levels

The high basal expression levels observed for RL_CT_GI may be due to different reasons, including e.g. the following: yeast endogenous TFs binding to the synTALE-DBS, thereby activating gene expression; a cryptic DBD in the PIF3-AD-fusion, that binds to the synTALE-DBS; unspecific interaction between the PIF3-AD- and the PhyBNT-synTALE-fusion, activating transcription independent of light conditions; intrinsic ADs in the synTALE-PhyB-fusion that activate transcription in a light-independent way. To determine the actual cause of (high) basal expression levels, different plasmids were constructed and tested with regard to yEGFP fluorescence (**Supplementary Figure 1** and **Supplementary Table 3**). Our results show that the synTALE-DBD-fusion is largely responsible for the high yEGFP levels observed in the uninduced state. However, basal expression levels of RL_CT are higher than those containing solely the synTALE-DBD- or the PhyBNT-synTALE-fusion. This leads to the interpretation that unspecific and light-independent interaction of PhyBNT with PIF3, which cannot be reliably tested, causes an additional effect, as all other reasons can be excluded considering the results obtained in this experiment.

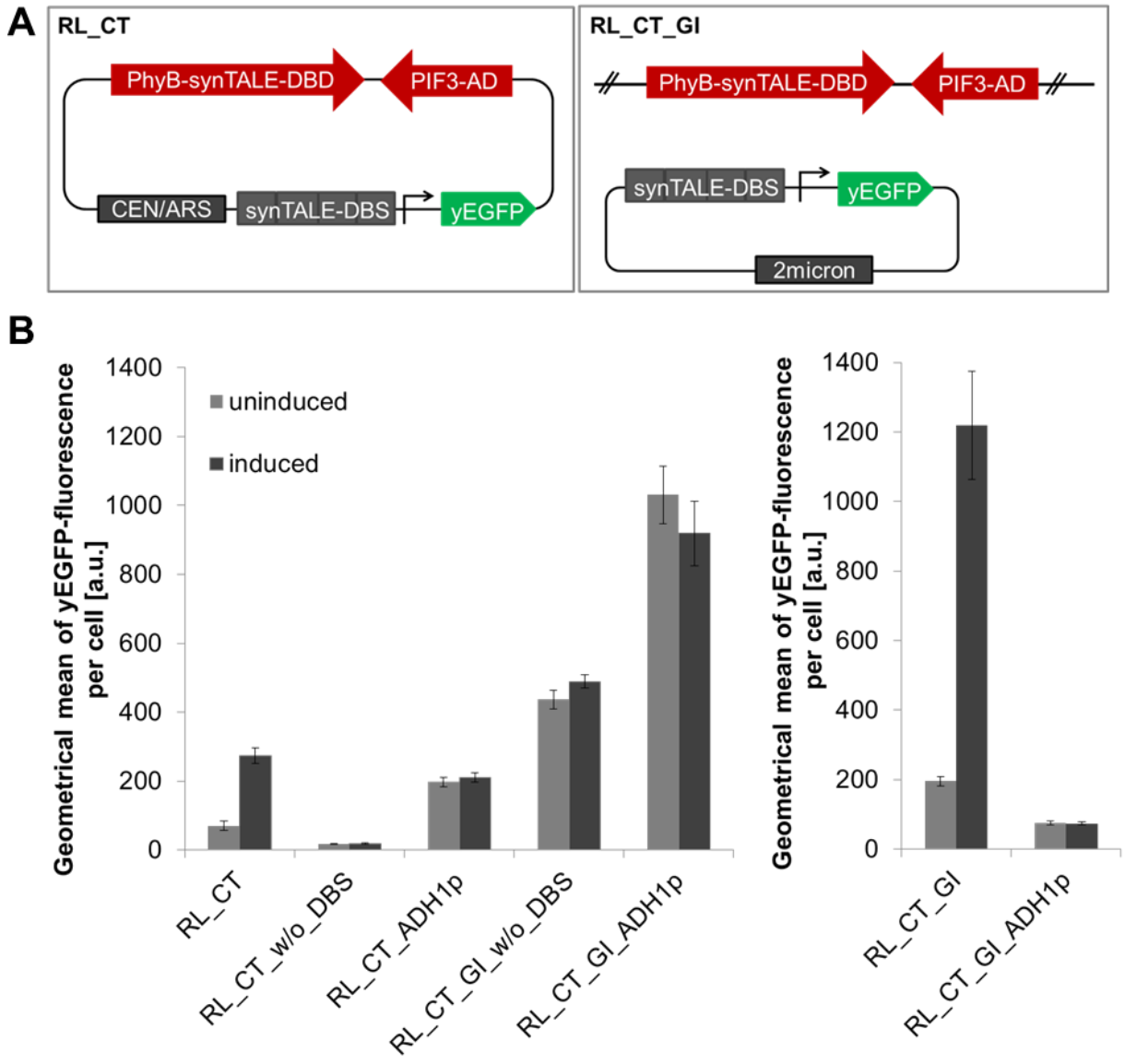


Figure 3: Comparison of red light-induced yEGFP fluorescence of yeast cells RL_CT and RL_CT_GI. A Schematic overview of RL_CT and RL_CT_GI showing the difference of the episomal (RL_CT) and genome integrated system (RL_CT_GI). In the episomal system, all three parts are located on one CEN/ARS plasmid. For the genome integrated system, only the reporter cassette is located on a 2-micron plasmid while the PhyB-synTALE-DBD- and the PIF3-AD-fusions were genome integrated. **B** Measurement of light-induced gene expression to compare RL_CT and RL_CT_GI. RL_CT_w/o_DBS and RL_CT_GI_w/o_DBS served as negative controls without synTALE-DBS. RL_CT_ADH1p and RL_CT_GI_ADH1p served as technical controls and enable comparison with the medium-strong yeast *ADH1* promoter. Yeast cells were grown in SD-Leu and SD-Leu-Ura selection media with 25 µM PCB for 6 h in darkness. Induced cells were treated with a 30-sec red light pulse and grown for another 16 h with 10-sec red light pulses every 30 min. Uninduced cells remained in darkness. The left graph represents measurements with high detector sensitivity, while the right graph represents measurements with lower detector sensitivity. Detector sensitivity was modified due to strong yEGFP fluorescence achieved by RL_CT_GI. A.u.: Arbitrary units.

Table 1: Complete list of strains generated. All yeast strains were generated from YPH500 cells.

Yeast strain	Genomic integration/ locus	Episomal plasmid	Selection
RL_CT	-	pRL_CT	Leu
RL_CT_w/o_DBS	-	pRL_CT_w/o_DBS	Leu
RL_CT_ADH1p	-	pRL_CT_ADH1p	Leu
RL_NT	-	pRL_NT	
RL_NT_w/o_DBS	-	pRL_NT_w/o_DBS	
RL_NT_ADH1p	-	pRL_NT_ADH1p	
RL_CT_GI	pRL_CT_GI/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_w/o_DBS	pRL_CT_GI/ *ura3-52*	pRL_yEGFP_w/o_DBS_hc	Leu/Ura
RL_CT_GI_ADH1p	pRL_CT_GI/ *ura3-52*	pRL_yEGFP_ADH1p_hc	Leu/Ura
RL_GI_PCB27	pRL_CT_GI/ *ura3-52* pRL_GI_PCB27/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB28	pRL_CT_GI/ *ura3-52* pRL_GI_PCB28/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB29	pRL_CT_GI/ *ura3-52* pRL_GI_PCB29/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB30	pRL_CT_GI/ *ura3-52* pRL_GI_PCB30/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB31	pRL_CT_GI/ *ura3-52* pRL_GI_PCB31/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB32	pRL_CT_GI/ *ura3-52* pRL_GI_PCB32/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB33	pRL_CT_GI/ *ura3-52* pRL_GI_PCB33/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_GI_PCB34	pRL_CT_GI/ *ura3-52* pRL_GI_PCB34/ *his1-Δ200*	pRL_yEGFP_hc	Leu/Ura/ Trp
RL_CT_w/o_PhyBNT_ w/o_PIF3-AD	-	pRL_CT_w/o_PhyBNT_ w/o_PIF3-AD	Leu
RL_CT_w/o_PIF3-AD	-	pRL_CT_w/o_PIF3-AD	Leu
RL_CT_w/o_TALE	-	pRL_CT_w/o_TALE	Leu
RL_CT_GI_del1 = PhiReX1.0	pRL_CT_GI_del1/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del2	pRL_CT_GI_del2/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del1_1	pRL_CT_GI_del1_1/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del2_1	pRL_CT_GI_del2_1/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del1_2 = PhiReX1.1	RL_CT_GI_del1_2*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del2_2	RL_CT_GI_del2_2*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del1_3	RL_CT_GI_del1_3*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del1_4	RL_CT_GI_del1_4*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del1_5	RL_CT_GI_del1_5*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del2_5	RL_CT_GI_del2_5*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
RL_CT_GI_del2_6	RL_CT_GI_del2_6*/ *ura3-52*	pRL_yEGFP_hc	Leu/Ura
PhiReX1.0+	RL_CT_GI_del1/ *ura3-52* pRL_GI_PCB28/ *his1-Δ200*	pRL_yeGFP_hc	Leu/Ura/ Trp

To lower the synTALE-mediated background activity, we modified the PhyBNT-synTALE-fusion to prevent translocation of the protein to the nucleus before red light irradiation. Upon light application, interaction of PhyB-synTALE with the NLS-containing PIF3-AD-fusion allows translocation of the protein complex to the nucleus and activation of gene expression. To this end, the following modifications were introduced to the optical dimerizer (**Figure 4, Supplementary Table 4**): (i) Deletion of the Cdc1-NLS incorporated in the PhyBNT-synTALE-fusion. (ii) Deletion of previously reported internal NLS sequences in the synTALE to prevent nuclear localization before induction (41) (**Figure 4A**). (iii) Addition of a nuclear export signal (NES) to the PhyBNT-synTALE-fusion, to localize the protein to the cytoplasm (**Figure 4B**). (iv) Insertion of an additional Cdc1-NLS to the PIF3-SV40-NLS-AD-fusion to support efficient nuclear import of the large complex consisting of the PhyB-synTALE- and the PIF3-AD-fusion. (v) Deletion of the SV40-NLS of the PIF3-AD-fusion to allow undisrupted native interaction between PhyB and PIF3 resulting in nuclear import of the dimerized fusion protein only after light application (42-44). The constructs described above were analyzed for red light-induced reporter gene expression (**Supplementary Figure 2**). Best results with regard to basal expression levels and fold induction were observed for RL_CT_GI_del1 lacking all potential NLS sequences in the PhyB-synTALE-fusion (PhiReX 1.0) and RL_CT_GI_del1_2 (PhiReX 1.1) containing an additional PKI-NES (**Figure 4B and C**). In medium containing 25 µM PCB, PhiReX 1.0 confers 11-fold induction and high expression output, similar to the strong yeast *TDH3* promoter. PhiReX 1.1 shows nearly no background activity and high fold induction of up to 41. The expression levels upon light induction are comparable to those of the medium strong yeast *ADH1* promoter.

Engineering yeast to produce the chromophores PCB and PΦB

Even if RL_CT_GI shows low red light-mediated reporter expression in the absence of PCB, we proposed that expression output can be increased by engineering a yeast strain that endogenously produces the chromophores PCB or PΦB from the precursor heme, which is naturally present in the host. To achieve this, two enzymes are required: an HO which catalyzes the conversion of heme to biliverdin, and a BR which converts biliverdin to PϕB or PCB, depending on its source organism (**Figure 5A**) (27).

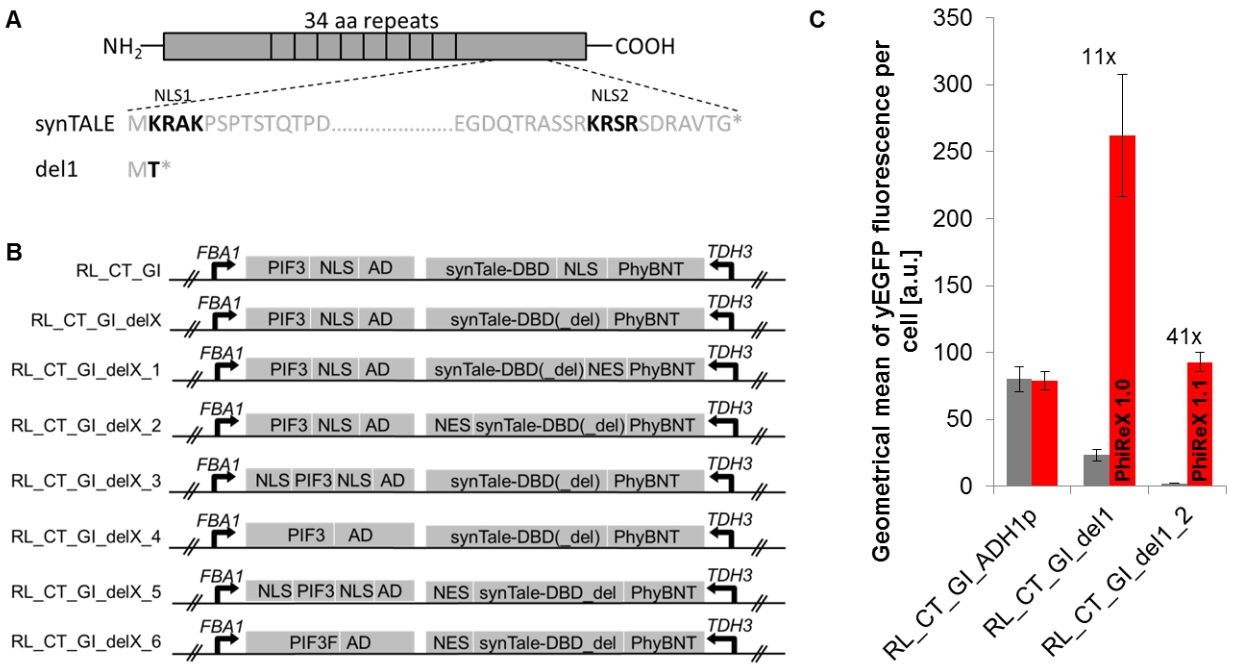


Figure 4: Strategy for minimizing basal expression levels. A Deletion of NLS sequences in the TALE backbone. **B** Schematic overview of different RL_CT_GI modifications. **C** Comparison of RL_CT_GI_del1 and RL_CT_GI_del1_2, dubbed PhiReX 1.0 and PhiReX1.1. Yeast cells were grown in SD-Ura-Leu medium with 25 µM PCB for 6 h in darkness. Induced cells were treated with a 30-sec red light pulse, followed by 10-sec red light pulses every 30 min for 16 h. Uninduced cells and RL-CT_GI_ADH1p cells remained in the dark. Output of yEGFP was measured via flow cytometry. Fold induction is indicated. A.u.: Arbitrary units.

We designed and tested different combinations of native and modified HOs and BRs derived from *A. thaliana* (*HY1*, *HY2*), *Synechocystis sp.* (*PcyA*) and *S. cerevisiae* (*Hmx1*) (**Figure 5B**). The different expression cassettes were genome integrated into the *his1-Δ200* locus of RL_CT_GI cells (**Figure 5C**). Previous studies in mammalian cells (28) and *Pichia pastoris* (29) revealed optimized chromophore biosynthesis when the two biosynthesis enzymes were targeted to mitochondria, the location of heme biosynthesis in eukaryotes (45). Accordingly, we deleted the proteins' native target sequence, if applicable, and replaced it by a yeast-derived mitochondrial targeting sequence (MTS) (**Supplementary Table 1**). As reporter output was shown to depend on cellular chromophore concentration (**Figure 2)**, the most efficient enzyme combination for holo-phytochrome production was determined by measuring yEGFP fluorescence of RL_GI_PCB27-34 in the absence of exogenous PCB**.** All combinations showed increased expression levels compared to RL_CT_GI, with the highest values observed for RL_GI_PCB28 (HY1, PcyA), RL_GI_PCB31 (MTS-mHY1, MTS-mHY2),

RL_GI_PCB32 (MTS-mHY1, MTS-PcyA), and RL_GI_PCB33 (MTS-mHmx1, MTS-mHY2), reaching approximately 3.5-fold induction in all combinations (**Figure 5D**). As the red light-mediated fold induction was distinctly higher in cells harboring genome integrated chromophore biosynthesis genes than in cells lacking HO and BR, we concluded that yeast was successfully engineered to produce PCB and PΦB regardless of the localization of the enzymes. The expression cassette of pRL_GI_PCB28 for endogenous chromophore production was genome integrated into PhiReX1.0, to yield strain PhiReX1.0+, which is independent of chromophore supply.

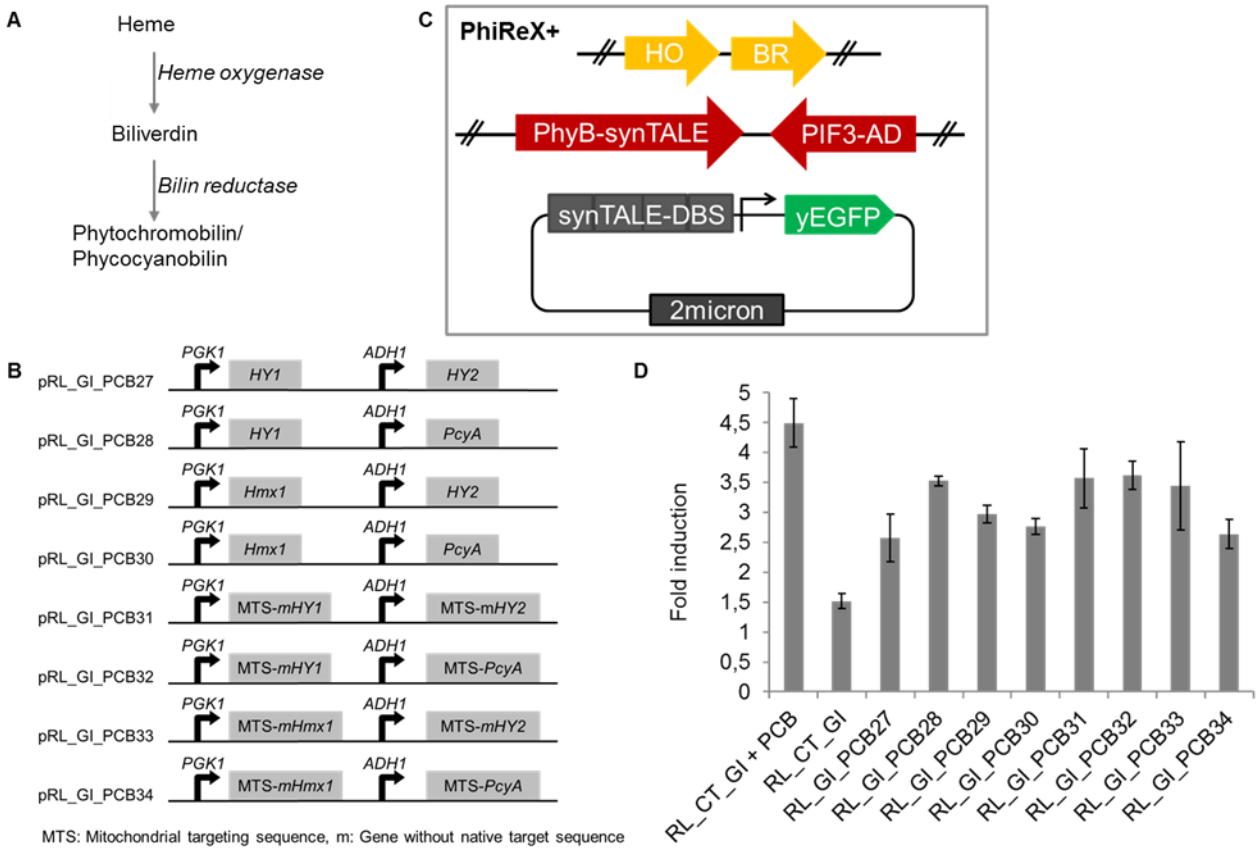


Figure 5: Chromophore biosynthesis in *S. cerevisiae*. A Overview of the conversion of heme to the chromophore via biliverdin. **B** Design of constructs pRL_GI_PCB27-34, created to achieve chromophore biosynthesis after genome integration into RL_CT_GI. **C** Scheme of the chromophore-independent PhiReX+ system. Red light-sensing synTF and chromophore biosynthesis components are genome integrated. The reporter construct is located on an episomal high-copy plasmid. **D** Fold induction of yEGFP output upon red light induction of RL_CT_GI with PCB concentrations of 25 µM (RL_CT_GI + PCB) and 0 µM (RL_CT_GI) and of RL_GI_PCB27-34 without PCB in the medium.

Characterization of PhiReX

In a time-course experiment, the PCB producing PhiReX 1.0+ was analyzed in comparison to the strong constitutive yeast *TDH3* promoter and the galactose-inducible *GAL1* promoter (**Figure 6A**). Our results demonstrate that PhiReX 1.0+ mediates a robust expression already 4 h after induction, with expression levels reaching those of the *TDH3* promoter after 8 h. The reporter output is stable for at least 48 h. The achieved fold induction is comparable to that of the *GAL1* promoter, while PhiReX 1.0+ mediates a much higher expression output. However, also basal expression levels of PhiReX 1.0+ exceed those of the *GAL1* promoter. Notably, the PhiReX 1.1 system described above confers very low background activity and therefore is a suitable alternative for applications requiring reduced basal expression. To provide a reliable basis for researchers intending to work with PhiReX, we further proved the PhiReX 1.0+ system to show light dose dependency and reversibility (**Figure 6B and C**): We applied single light pulses of various length and determined reporter output 4 h post-induction. Notably, a short light pulse of only five seconds is sufficient to induce reporter expression. A stepwise prolongation of the light pulse up to 120 sec results in increasing yEGFP fluorescence. Thereby, PhiReX 1.0+ provides an easy way to fine-tune gene expression output. To test the reversibility of induction by red/far-red light application, PhiReX 1.0+ was treated with light pulses of different wavelength. While illumination with red light of 660 nm induced gene expression, far-red light (730 nm) reliably switched off protein production. Re-induction was readily achieved by a second round of red light application.

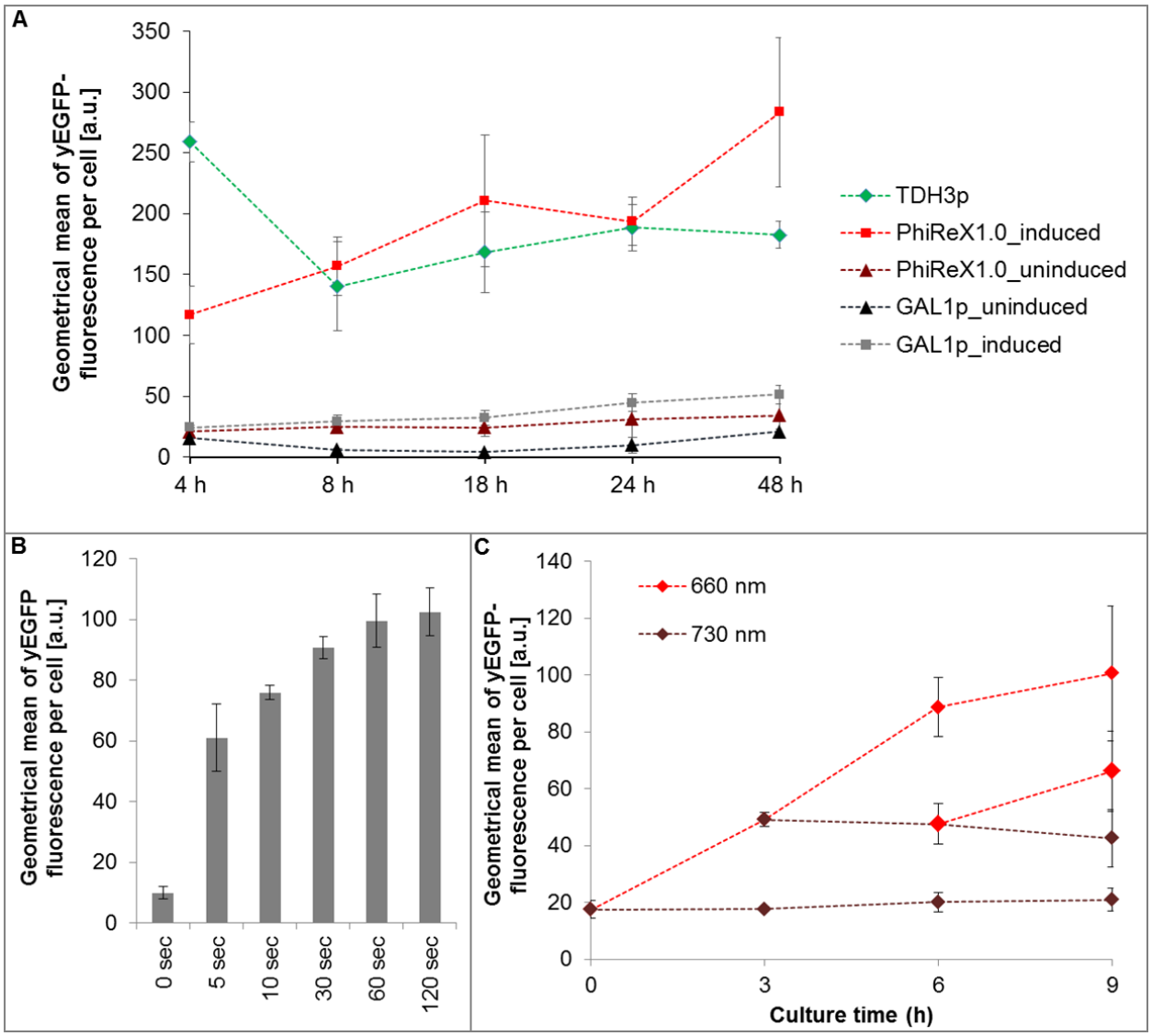


Figure 6: Characterization of PhiReX 1.0+. Yeast cells were grown in SD-Ura-Leu-Trp medium for 6 h in darkness prior to induction. **A** Time course of light-dependent reporter gene induction, compared to the constitutive *TDH3* promoter and the *GAL1* induction system (pYES2/CT). Cells were grown in a volume of 20 ml SD medium in 100-ml flasks. Output of yEGFP was measured via flow cytometry in a 48-h time course experiment. **B** Light dose dependency of PhiReX 1.0+ was measured by activating cells with a single light pulse of the indicated duration. Post-induction, cells were grown for 4 h in the dark and yEGFP was measured via flow cytometry for two biological replicates, each in duplicate. **C** Reversibility of protein production via PhiReX 1.0+. Cells were grown in a volume of 20 ml SD medium in 100-ml flasks. Uninduced cells were grown in the dark, while induced cells were treated with a single 30-sec red light pulse, followed by 5-sec pulses every 30 sec. Reversion of gene expression was mediated by a 5-min far-red light pulse. Cultures were divided into two sub-cultures, of which one was treated with red light as indicated, while the other was further maintained in unchanged condition. yEGFP output was measured via flow cytometry for two biological replicates, each with three technical replicates. Line colors indicate light conditions of the respective samples: red lines indicate growth under inducing conditions; dark-red lines indicate growth in darkness after far-red light application. A.u.: Arbitrary units.

DISCUSSION

The work presented here describes the development of a user-friendly, programmable, dose-dependent and robust red/far-red light-regulated gene expression switch, dubbed PhiReX, for application in the yeast *Saccharomyces cerevisiae*. Most important, PhiReX is independent of the supply of exogenous chromophore, making it perfectly suitable for large scale production. PhiReX employs the N-terminal part of the photoreceptor PhyB (PhyBNT), which dimerizes with its interacting partner PIF3 in a red/far-red light dependent manner. The implementation of a synTALE-DBD allows its flexible programming to specifically target any DBS, synthetically generated or available in the host genome. Blue light-dependent regulation of gene expression, based on synTALE-DBDs, was previously reported for mammalian cells (46). However, to our knowledge no customizable light induction system has been developed for yeast so far. The synTALE-mediated flexibility is a great advantage over DBDs with invariable binding specificity, limiting the regulation to promoters including their native DBS, like e.g. the GAL4-DBD, currently employed in several optogenetic expression systems (13-15). However, the synTALE-DBD (around 1,000 aa) is much larger than the GAL4-DBD (around 150 aa), which may hinder the conformational switch of the photoreceptor and the dimerization with PIF3. We therefore tested two different architectures of the PhyB-synTALE-fusion, where we linked the synTALE-DBD to either the N- or C-terminus of PhyBNT, resulting in plasmids pRL_NT and pRL_CT (**Figure 1A**). We found that both versions show comparable functionality regarding light- and PCB-dependent reporter gene activation (**Figure 2**), with the difference that solely RL_CT induced yEGFP expression in a light-dependent manner, also in the absence of PCB. Red light activation without chromophore supply has already been reported in *S. cerevisiae*, indicating the presence of an unidentified cellular compound that functions as a chromophore (16). At this stage of knowledge, it can just be assumed that steric effects are responsible for the selective activation of exclusively RL_CT.

Even if light- and PCB-dependency could be proofed by the initially developed light induction systems RL_CT and RL_NT, the constructs show great variability with respect to the number of activated cells and high basal activity, indicating that further optimization is needed.

To achieve stable and robust expression, we integrated the PhyB-synTALE- and the PIF3-AD-fusions into the yeast genome, as this has often been reported to yield a more uniform expression. The reporter expression cassette was placed on a 2-micron high-copy plasmid, to enable an easy exchange of the GOI and to guarantee high expression levels. The first step of optimization yielded a great boost in yEGFP fluorescence output and a reliable and robust regulation of gene expression **(Figure 3B)**. Furthermore, we demonstrated that the high basal expression levels result from the synTALE-DBD, which binds to the DBS independent of light conditions in the initial system (**Supplementary Figure 1**). The TALE *avrBS3* backbone, used in this study in a truncated version, incorporates a single acidic AD at its C-terminus, not included in the synTALE-DBD employed here (47). It may be assumed, that an additional non-characterized internal AD, which is at least functional in *S. cerevisiae,* causes basal expression. However, Konermann *et al.* (2013) also reported high background levels for a synTALE-based blue light expression system in mammalian cells and described different strategies to lower the undesired expression by preventing nuclear localization of the synTALE-containing domain before light activation (46). Based on their results we redesigned our system to prevent nuclear import of the synTALE-PhyB-fusion before dimerization with the PIF3-AD-fusion, containing the strong SV40-NLS (**Figure 4**). Therefore, the Cdc1NLS included in the synTALE-PhyB-fusion was deleted and potential NLS sequences in the synTALE backbone were mutated or deleted according to previous approaches (41). As PhyB was reported to contain NLS-like motifs only in the C-terminus, there was no need to mutate or delete sequences in the photoreceptor PhyBNT, used in this study (48). The resulting PhiReX 1.0 strain showed reduced basal expression levels with high expression output in induced state compared to the *TDH3* promoter and a fold induction of up to 11 (**Figure 4C**). The addition of the strong PKI-NES to the PhyBNT-synTALE-fusion of PhiReX 1.0, resulting in PhiReX 1.1, minimizes basal expression level and allows tight control of protein production with a fold induction of 41, while red light-controlled expression output is comparable to that of the medium strong *ADH1* promoter. Due to its low background, this system is ideally suited for the expression of toxic proteins. Additional manipulations of the PIF3-AD-fusion did not result in further improvements of the system (**Figure 4B, Supplementary Figure 2**).

The exogenous supply of the chromophore PCB devaluates the advantages of a light system as a cheap and easy-to-handle system. Even if RL_CT_GI cells cause light-induced yEGFP output also in the absence of PCB, the obtained fold induction (~1.5) was not satisfying and prompted us to engineer *S. cerevisiae* to produce the required chromophore by expressing the enzymes HO and a BR catalyzing the conversion of heme to biliverdin, and to PCB or PϕB, respectively (27). The successful production of holo-phytochrome was already shown in bacteria, mammalian cells and the yeast *Pichia pastoris* (27-29), but not yet in *S. cerevisiae.* To change this, we integrated different combinations and versions of HOs and BRs into RL_CT_GI cells to detect the best chromophore-producing combination by measuring the respective reporter output (**Figure 5B**). The highest fold induction (~3.5) was achieved with the combinations of HY1 and PcyA, MTS-mHY1/MTS-mHY2, MTS-mHY1/MTS-PcyA, and MTS-mHmx1/MTS-mHY2 (**Figure 5D**). In contrast to published data (28,29), targeting HY1 and PcyA to mitochondria did not gravely boost induction rates compared to cells expressing the two enzymes in the cytoplasm. Still, by combining native HY1 and PcyA we achieved robust expression and doubled the fold induction compared to RL_CT_GI without PCB supply. Furthermore, the difference to RL_CT_GI cells grown in medium supplemented with 25 µM PCB (fold induction of 4.5) was minor, proving that a system was achieved, that is independent of exogenous chromophore supply.

Chromophore biosynthesis genes *HY1* and *PcyA* were integrated into the genome of PhiReX 1.0 cells, resulting in the red light-dependent expression system PhiReX 1.0+, which is independent of exogenous chromophore supply. With respect to expression output and fold induction, PhiReX 1.0+ shows a similar behavior as PhiReX 1.0 with supplied PCB. As compared to the galactose-inducible production of yEGFP, PhiReX 1.0+ results in similar fold induction values, but by far exceeds the absolute fluorescence values (**Figure 6A**). Users wishing to tightly control GOI expression in uninduced conditions may refer to PhiReX 1.1, which offers greatly reduced basal activity. Our analyses further demonstrate that the applied red light dose is a user-friendly option to easily fine-tune expression output, which is an important advantage over e.g. the *GAL1* system (**Figure 6B**). The reversible control of gene expression by red/far-red light application has many potential implications for the construction of expression switches within synthetic regulatory networks (**Figure 6C**).

Taken together, PhiReX allows for a tightly regulated and red light-dependent control of gene expression with a strong and robust fold induction. With a single cloning step the user can exchange the plasmid-located reporter with the CDS of any GOI and thereby adopt the system easily to individual demands. Furthermore, the synTALE-based approach enables the user to flexibly program the PhiReX system towards any desired promoter region. As *S. cerevisiae* was engineered to produce sufficient amounts of the chromophore, the system is independent of the supply of toxic or expensive chemicals. PhiReX allows tunable and reversible induction via simple light exposure. We chose the well characterized budding yeast as host organism, which is eminently suitable for pathway engineering and high protein production. We envision that the fast, flexible, easy-to-use, nontoxic and low-cost nature of PhiReX makes it even suitable for industry-scale protein production.

FUNDING

This work was funded by the Federal Ministry of Education and Research of Germany (BMBF; grant number 031A172) and the Potsdam Graduate School.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This research was funded by the Federal Ministry of Education and Research of Germany (BMBF; grant number 031A172). L.H. thanks the Potsdam Graduate School for providing a scholarship. The authors greatly acknowledge Dr. Pamela Holzlöhner (University of Potsdam, Germany) for introduction to flow cytometer.

AUTHOR CONTRIBUTIONS

L.H. designed and developed the overall strategy with contributions from F.M. L.H. conducted the experiments, analyzed the data and wrote the manuscript. F.M. provided and characterized the synTALE-DBD/BS pair and assisted with Cas9 genome editing. B.M.-R. initiated the project and the overall research strategy. B.M.-R. and F.M. edited the paper. All work was carried out in the Cell2Fab lab, led by K.M. All authors take full responsibility for the content of the paper.

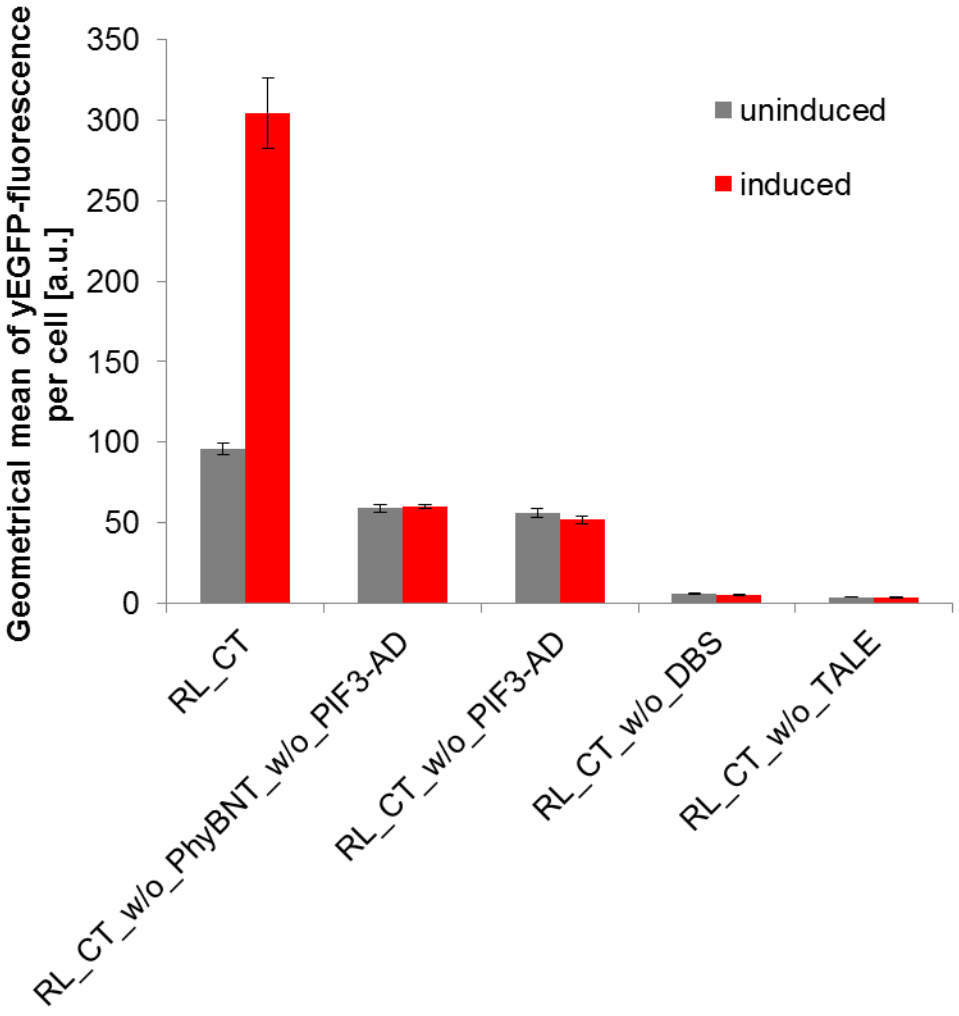
REFERENCES

1. Engels, B., Dahm, P. and Jennewein, S. (2008) Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic engineering*, **10**, 201-206.
2. Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S. and Kirby, J. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940-943.
3. Galanie, S., Thodey, K., Trenchard, I.J., Interrante, M.F. and Smolke, C.D. (2015) Complete biosynthesis of opioids in yeast. *Science*, **349**, 1095-1100.
4. Runguphan, W. and Keasling, J.D. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic engineering*, **21**, 103-113.
5. Si, T., Luo, Y., Xiao, H. and Zhao, H. (2014) Utilizing an endogenous pathway for 1-butanol production in *Saccharomyces cerevisiae*. *Metabolic engineering*, **22**, 60-68.
6. Brat, D., Weber, C., Lorenzen, W., Bode, H.B. and Boles, E. (2012) Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast *Saccharomyces cerevisiae*. *Biotechnology for biofuels*, **5**, 65.
7. Macreadie, I.G., Horaitis, O., Verkuylen, A.J. and Savin, K.W. (1991) Improved shuttle vectors for cloning and high-level Cu 2+-mediated expression of foreign genes in yeast. *Gene*, **104**, 107-111.
8. Kramer, R.A., DeChiara, T.M., Schaber, M.D. and Hilliker, S. (1984) Regulated expression of a human interferon gene in yeast: control by phosphate concentration or temperature. *Proceedings of the National Academy of Sciences*, **81**, 367-370.
9. Bellí, G., Garí, E., Piedrafita, L., Aldea, M. and Herrero, E. (1998) An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Research*, **26**, 942-947.
10. Quintero, M.J., Maya, D., Arévalo-Rodríguez, M., Cebolla, Á. and Chávez, S. (2007) An improved system for estradiol-dependent regulation of gene expression in yeast. *Microbial cell factories*, **6**, 1.
11. Butt, T.R., Sternberg, E.J., Gorman, J.A., Clark, P., Hamer, D., Rosenberg, M. and Crooke, S.T. (1984) Copper metallothionein of yeast, structure of the gene, and regulation of expression. *Proceedings of the National Academy of Sciences*, **81**, 3332-3336.
12. Cherest, H., Kerjan, P. and Surdin-Kerjan, Y. (1987) The *Saccharomyces cerevisiae MET3* gene: nucleotide sequence and relationship of the 5′ non-coding region to that of *MET25*. *Molecular and General Genetics MGG*, **210**, 307-313.
13. Shimizu-Sato, S., Huq, E., Tepperman, J.M. and Quail, P.H. (2002) A light-switchable gene promoter system. *Nature biotechnology*, **20**, 1041-1044.
14. Hughes, R.M., Bolger, S., Tapadia, H. and Tucker, C.L. (2012) Light-mediated control of DNA transcription in yeast. *Methods*, **58**, 385-391.
15. Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D. and Tucker, C.L. (2010) Rapid blue-light-mediated induction of protein interactions in living cells. *Nature methods*, **7**, 973-975.
16. Sorokina, O., Kapus, A., Terecskei, K., Dixon, L.E., Kozma-Bognar, L., Nagy, F. and Millar, A.J. (2009) A switchable light-input, light-output system modelled and constructed in yeast. *Journal of biological engineering*, **3**, 1.
17. Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K. and Collins, J.J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, **150**, 647-658.
18. Farzadfard, F., Perli, S.D. and Lu, T.K. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology*, **2**, 604-613.
19. Blount, B.A., Weenink, T., Vasylechko, S. and Ellis, T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PloS one*, **7**, e33279.
20. Ni, M., Tepperman, J.M. and Quail, P.H. (1999) Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature*, **400**, 781-784.
21. Boch, J. and Bonas, U. (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Phytopathology*, **48**, 419.

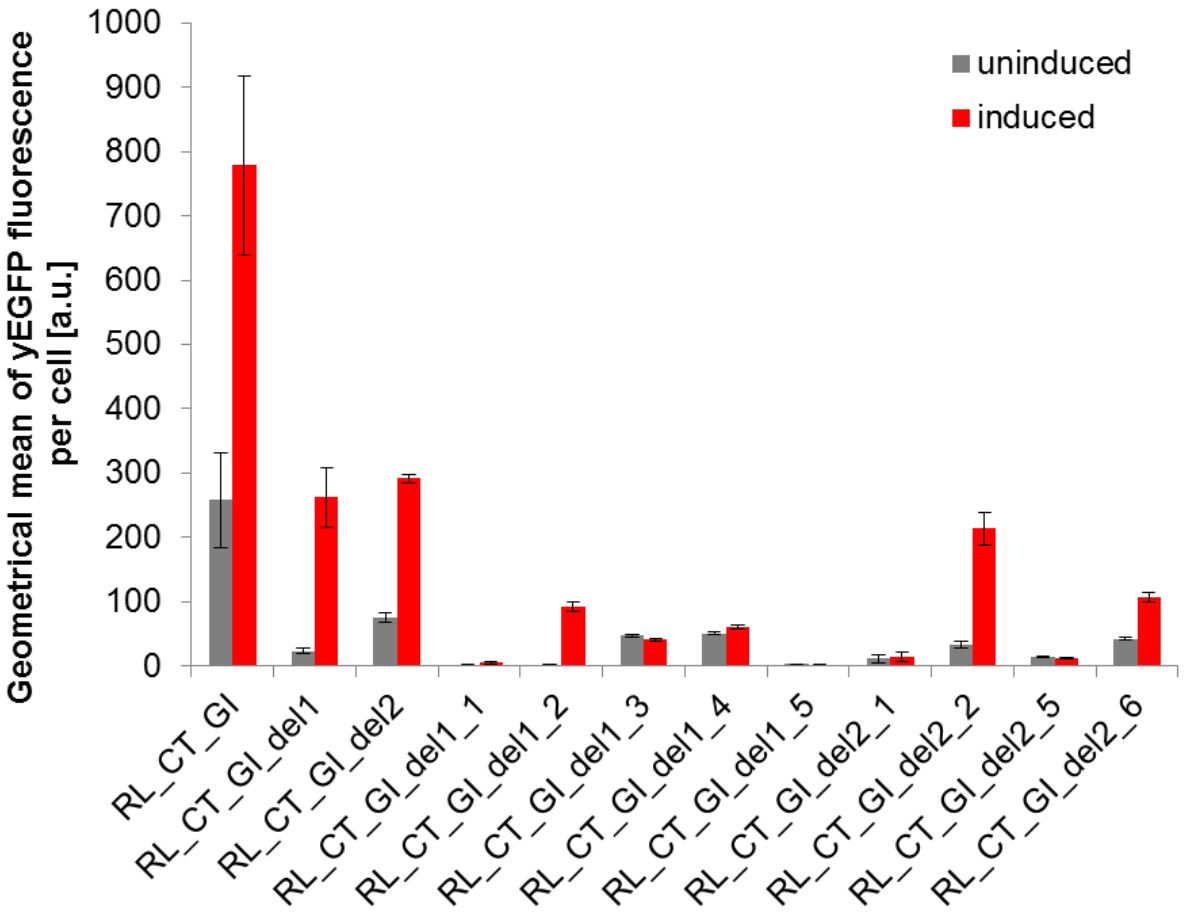
22. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.
23. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501-1501.
24. Kunkel, T., Neuhaus, G., Batschauer, A., Chua, N.H. and Schäfer, E. (1996) Functional analysis of yeast-derived phytochrome A and B phycocyanobilin adducts. *The Plant Journal*, **10**, 625-636.
25. Kunkel, T., Speth, V., Büche, C. and Schäfer, E. (1995) *In Vivo* Characterization of Phytochrome-Phycocyanobilin Adducts in Yeast. *Journal of Biological Chemistry*, **270**, 20193-20200.
26. Li, L. and Lagarias, J.C. (1994) Phytochrome assembly in living cells of the yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, **91**, 12535-12539.
27. Gambetta, G.A. and Lagarias, J.C. (2001) Genetic engineering of phytochrome biosynthesis in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 10566-10571.
28. Müller, K., Engesser, R., Timmer, J., Nagy, F., Zurbriggen, M.D. and Weber, W. (2013) Synthesis of phycocyanobilin in mammalian cells. *Chemical communications*, **49**, 8970-8972.
29. Shin, A.Y., Han, Y.J., Song, P.S. and Kim, J.I. (2014) Expression of recombinant full-length plant phytochromes assembled with phytochromobilin in *Pichia pastoris*. *FEBS letters*, **588**, 2964-2970.
30. Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M. *et al.* (2005) Synthetic biology: Engineering *Escherichia coli* to see light. *Nature*, **438**, 441-442.
31. Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. (2017) AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Research*, gkx034.
32. Ellis, T., Wang, X. and Collins, J.J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature biotechnology*, **27**, 465-471.
33. Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nature methods*, **10**, 977-979.
34. Zhang, Y., Werling, U. and Edelmann, W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research*, **40**, e55.
35. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, **6**, 343-345.
36. Kouprina, N. and Larionov, V. (2006) TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics*, **7**, 805-812.
37. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, **2**, 31-34.
38. Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Research*, **39**, 5790-5799.
39. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, gkt135.
40. Rodríguez-López, M., Cotobal, C., Fernández-Sánchez, O., Bravo, N.B., Oktriani, R., Abendroth, H., Uka, D., Hoti, M., Wang, J. and Zaratiegui, M. (2016) A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast. *Wellcome Open Research*, **1**.
41. Van den Ackerveken, G., Marois, E. and Bonas, U. (1996) Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell*, **87**, 1307-1316.
42. Pfeiffer, A., Nagel, M.-K., Popp, C., Wüst, F., Bindics, J., Viczián, A., Hiltbrunner, A., Nagy, F., Kunkel, T. and Schäfer, E. (2012) Interaction with plant transcription factors can mediate nuclear import of phytochrome B. *Proceedings of the National Academy of Sciences*, **109**, 5892-5897.
43. Beyer, H.M., Juillot, S., Herbst, K., Samodelov, S.L., Müller, K., Schamel, W.W., Römer, W., Schäfer, E., Nagy, F. and Strähle, U. (2015) Red light-regulated reversible nuclear localization of proteins in mammalian cells and zebrafish. *ACS synthetic biology*, **4**, 951-958.
44. Chen, M., Tao, Y., Lim, J., Shaw, A. and Chory, J. (2005) Regulation of phytochrome B nuclear localization through light-dependent unmasking of nuclear-localization signals. *Current Biology*, **15**, 637-642.

45. Ponka, P. (1999) Cell biology of heme. *The American journal of the medical sciences*, **318**, 241-256.
46. Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M. and Zhang, F. (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, **500**, 472-476.
47. Szurek, B., Marois, E., Bonas, U. and Van den Ackerveken, G. (2001) Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *The Plant Journal*, **26**, 523-534.
48. Matsushita, T., Mochizuki, N. and Nagatani, A. (2003) Dimers of the N-terminal domain of phytochrome B are functional in the nucleus. *Nature*, **424**, 571-574.

SUPPLEMENTARY DATA



Supplementary Figure 1: Investigation of basal expression levels. Yeast cells RL_CT, RL_CT_w/o_PhyBNT_w/o_PIF3-AD, RL_CT_w/o_PIF3-AD, RL_CT_w/o_DBS and RL_CT_w/o_TALE were grown in SD-Leu medium with 25 µM PCB for 6 h in darkness. Induced cells were treated with 30-sec red light pulse and grown for another 16 h with 10-sec red light pulses every 30 min. Uninduced cells remained in darkness. A.u.: Arbitrary units.



Supplementary Figure 2: Investigation of constructs with different modifications with respect to RL_CT_GI. Yeast cells were grown in SD-Ura-Leu-Trp medium with 25 µM PCB for 6 h in darkness. Induced cells were treated with 30-sec red light pulse and grown for another 16 h with 10-sec red light pulses every 30 min. Uninduced cells remained in darkness. A.u.: Arbitrary units.

Supplementary Table 1: List of parts with source, sequence modifications and function.

Assembly components	Source	Modifications	Function
PhyBNT	*A. thaliana*	AA 1-621 of PhyB/ Mutation nucleotide 576 C-->A	Photoreceptor
PIF3	*A. thaliana*		Interacting factor
PKI-NES	Oligo synthesis		Nuclear export sequence
SV40-NLS	Oligo synthesis		Nuclear localization sequence
Cdc1-NLS	Oligo synthesis		Nuclear localization sequence
FUM1-MTS	*S. cerevisiae*		Mitochondrial targeting sequence from *FUM1* gene
VP64-AD	pMLM3705		Activation domain
yEGFP	pLOGI		Reporter gene
LEU2	pGAD424		Selection marker
TRP1			Selection marker
synTALE-DBD			Transcription factor binding domain
synTALE-DBD_del1		Mutation AA 450 K-->T, Deletion of AA 451-506	Transcription factor binding domain
Jub1.1 DBS			Transcription factor binding sequence
HY1	*A. thaliana*		Heme oxygenase 1
HY2	*A. thaliana*		Long hypocotyl 2
mHY1	*A. thaliana*	HY1 lacking AA 1-54	Heme oxygenase 1 without transit peptide
mHY2	*A. thaliana*	HY2 lacking AA 1-44	Long hypocotyl 2 without transit peptide
Hmx1	*S. cerevisiae*		Heme oxygenase
mHmx1	*S. cerevisiae*	Hmx1 lacking AA 290-310	Heme oxygenase without membrane anchor
PcyA	Synthesized	Codon optimized for *S. cerevisiae*	Phycocyanobilin:ferredoxin oxidoreductase
FBA1 promoter	pPC_010		Constitutive promoter from *S. cerevisiae*
TDH3 promoter	pPC_012		
PDC1 promoter	pPC_009		
ENO2 promoter	pPC_008		
ADH1 promoter	pPC_020		
ACT1 promoter	pPC_015		
TDH1 promoter	pPC_018		
PGK1 promoter	pPC_005		
TPI1 promoter	pPC_013		
CYC1 minimal promoter	pHIS2.1[1]		Minimal promoter from *S. cerevisiae*

Assembly components	Source	Modifications	Function
FBA1 terminator	pPC_010		Terminator from *S. cerevisiae*
TDH3 terminator	pPC_012		
Cyc1 terminator	pHIS2.1[1]		
ENO2 terminator	pPC_008		
ADH1 terminator	pPC_020		
ACT1 terminator	pPC_015		
TDH1 terminator	pPC_018		
PGK1 terminator	pPC_005		
TPI1 terminator	pPC_013		
PDC1 terminator	pPC_009		

[1] Takara Bio, Saint-Germain-en-Laye, France

Supplementary Table 2: List of primers used for cloning.

Name	Sequence (5´-3´)
L662	CGTAAACGGTCCCGATCGGATCGTGCTGTCACCGGTTAGGTGAATTTACTTTAAATCTTGCATTTA
L762	GTCATAAAGCTATAAAAAGAAAATTTATTTAAATGCAAGATTTAAAGTAAATTCACCTATGTCATCCCTGATGCCTGGAGG
L767	TAGGGATAACAGGGTAATTTGTACACCTACCGGTCACC
L772	ACCACCTCGTCGCCTTGG
L795	GCTCGCGGGCAGGACTTGGTGTGCGCGAACGAATGGGATCCATAGTTTTATTTATATCTAACCCGGCTAATTTCAAGGCAAGTTCGTTCCCCCCGCCAGATTCTTTAAAAGAGTCTC
L796	ATGGGAACTGATGTGAGATC
L797	GTAAACGGTCCCGATCGGATCGTGCTGTCACCGGTGGCGGGGGGAACGAACTTGCCTTGAAATTAGCCGGGTTAGATATAAATAAAACTTAAGTGAATTTACTTTAAATCTTG
sgRNA_RL_1_F	GCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGATCGTCGCCCAAGAAGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
sgRNA_RL_1_R	AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTTCTTCTTGGGCGACGATCGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGC
sgRNA_RL_2_F	GCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCAGCCTGAAAAGCTCAAACAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
sgRNA_RL_2_R	AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTGTTTGAGCTTTTCAGGCTGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGC
sgRNA_RL_3_F	GCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCATGTCATCCCTGATGCCTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
sgRNA_RL_3_R	AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCAGGCATCAGGGATGACATGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGC
donor_RL_1_F	CAAACAACGACGAATAACAACAGTAACTTGAAACCAATAAAGAGAAAACAGGGGTCTTCTGATCAGTTTTGTGGATCGTCGGGATCGAGCGACTACAAAG
donor_RL_1_R	CCCGCTTCCACCGCCTCCTGCAGCCTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGCTCGATCCC
donor_RL_2_F	CGTTATTCTTCTGTTCTTCTTTTTCTTTTGTCATATATAACCATAACCAAGTAATACATATTCAAAATGAGGCCATTAAAGCGTAAGAAAGGATCGTCGC
donor_RL_2_R	GTGGAGAAGGGTTCCTGTCTTGAGCAGATTCAAGCTTAGCTTTGGTTAGCCTGAATAGCTCAAATAGAGGCGACGATCCTTTCTTACGCTTTAATGGCCT

Name	Sequence (5´-3´)
donor_RL_3_F	AAACTAAGTCATAAAGCTATAAAAAGAAAATTTATTTAAATGCAAGATTTAAAGT AAATTCACCTAAGTTTTATTTATATCTAACCCGGCTAATTTCAAGGCAAGTTCGT TCCCCCCGCC
donor_RL_3_R	TTTCGCAGAGTGGGCGTCACCGAACTCGAAGCCCGCAGTGGAACGCTCCCCC CAGCCTCGCAGCGTTGGGACCGTATTCTCCAGGCATCAGGGATGACAGGCGG GGGGAACGAACTTGCC

Supplementary Table 3: List of Level 1 constructs.

Level 1 assembly	Backbone/homology regions	Level 0 vector	Level 0 homology regions	Assembly components
Constructs containing the red light-inducible split TF				
pRL_NT	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_NT_TALE	A1/A2	synTALE-DBD, Cdc1-NLS, *PhyBNT*
		pL0_RL_A3	A2/AR	synTALE-DBS, *yEGFP*
pRL_CT	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_CT_TALE	A1/A2	*PhyBNT*, Cdc1-NLS, synTALE-DBD
		pL0_RL_A3	A2/AR	synTALE-DBS, *yEGFP*
pRL_NT_w/o_DBS	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_NT_TALE	A1/A2	synTALE-DBD, Cdc1-NLS, *PhyBNT*
		pL0_RL_A3_w/o_DBS	A2/AR	*yEGFP*
pRL_NT_ADH1p	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_NT_TALE	A1/A2	synTALE-DBD, Cdc1-NLS, *PhyBNT*
		pL0_RL_A3_ADH1p	A2/AR	*yEGFP*
pRL_CT_w/o_DBS	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_CT_TALE	A1/A2	*PhyBNT*, Cdc1-NLS, synTALE-DBD
		pL0_RL_A3_w/o_DBS	A2/AR	*yEGFP*
pRL_CT_ADH1p	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_CT_TALE	A1/A2	*PhyBNT*, Cdc1-NLS, synTALE-DBD
		pL0_RL_A3_ADH1p	A2/AR	*yEGFP*

Level 1 assembly	Backbone/homology regions	Level 0 vector	Level 0 homology regions	Assembly components
Constructs for genome integrated red light-inducible TF				
pRL_CT_GI	pFM75/ A0-C0	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_CT_TALE	A1/A2	*PhyBNT*, Cdc1-NLS, synTALE-DBD
		Leu marker	A2/C0	
pRL_yEGFP_hc	pL0_A2-A3_Ura/A2-A3	pL0_RL_A3	A2/A3	synTALE-DBS, *yEGFP*
pRL_yEGFP_w/o_DBS_hc	pL0_A2-A3_Ura/A2-A3	pL0_RL_A3_w/o_DBS	A2/A3	*yEGFP*
pRL_yEGFP_ADH1p_hc	pL0_A2-A3_Ura/A2-A3	pL0_RL_A3_ADH1p	A2/A3	*yEGFP*
Constructs for genome integrated chromophore biosynthesis				
pRL_GI_PCB27	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB1	A3/A4	*HY1*
		pPCB6b	A4/HR_ADH1t	*HY2*
pRL_GI_PCB28	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB1	A3/A4	*HY1*
		pPCB5b	A4/HR_ADH1t	*PcyA*
pRL_GI_PCB29	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB2	A3/A4	*Hmx1*
		pPCB6b	A4/HR_ADH1t	*HY2*
pRL_GI_PCB30	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB2	A3/A4	*Hmx1*
		pPCB5b	A4/HR_ADH1t	*PcyA*
pRL_GI_PCB31	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB3	A3/A4	MTS-*mHY1*
		pPCB8b	A4/HR_ADH1t	MTS-*mHY2*
pRL_GI_PCB32	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB3	A3/A4	MTS-*mHY1*
		pPCB7b	A4/HR_ADH1t	MTS-*PcyA*
pRL_GI_PCB33	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB4	A3/A4	MTS-*mHmx1*
		pPCB8b	A4/HR_ADH1t	MTS-*mHY2*
pRL_GI_PCB34	pRL_GI_PCB1/ A3-HR_ADH1t	pPCB4	A3/A4	MTS-*mHmx1*
		pPCB7b	A4/HR_ADH1t	MTS-*PcyA*

Level 1 assembly	Backbone/homology regions	Level 0 vector	Level 0 homology regions	Assembly components
Constructs for investigation of basal expression levels				
pRL_CT_w/o_TALE	pL1A_Leu_lc/ A0-AR	pL0_RL_A1	A0/A1	*PIF3*, SV40-NLS, VP64-AD
		pL0_RL_A2_CT_w/o_TALE	A1/A2	*PhyBNT*, Cdc1-NLS
		pL0_RL_A3	A2/AR	synTALE-DBS, *yEGFP*
pRL_CT_w/o_PhyBNT_w/o_ PIF3-AD	pL1A_Leu_lc/ A0-AR	A0-A1 stuffer	A0/A1	
		pL0_RL_A2_TALE_w/o_PhyB	A1/A2	Cdc1-NLS, synTALE-DBD
		pL0_RL_A3	A2/AR	synTALE-DBS, *yEGFP*
pRL_CT_w/o_PIF3-AD	pL1A_Leu_lc/ A0-AR	A0-A1 stuffer	A0/A1	
		pL0_RL_A2_CT	A1/A2	*PhyBNT*, Cdc1-NLS, synTALE-DBD
		pL0_RL_A3	A2/AR	synTALE-DBS, *yEGFP*

Supplementary Table 4: List of constructs and strains for minimization of basal expression levels.

Level 1 construct/ strain	Backbone/ template	sgRNA target sequence (cloning oligonucleotides)	oligonucleotides	Included components
pRL_CT_GI_del1 §	pRL_CT_GI	-	L772/L762	PhyBNT, synTALE-DBD_del1/ PIF3, SV40-NLS, VP64-AD
pRL_CT_GI_del2 §	pRL_CT_GI	-	L662/L767	PhyBNT, synTALE-DBD/ PIF3, SV40-NLS, VP64-AD
pRL_CT_GI_del1_1 §	pRL_CT_GI_del1	-	L795/L796	PhyBNT, PKI-NES, synTALE-DBD_del1/ PIF3, SV40-NLS, VP64-AD
pRL_CT_GI_del2_1 §	pRL_CT_GI_del2	-	L795/L796	PhyBNT, PKI-NES, synTALE-DBD/ PIF3, SV40-NLS, VP64-AD
pRL_CT_GI_del2_2 §	pRL_CT_GI_del2	-	L767/L797	PhyBNT, synTALE-DBD, PKI-NES/ PIF3, SV40-NLS, VP64-AD
RL_CT_GI_del1_2*	RL_CT_GI_del1	ATGTCATCCCTGATGCC TGG (sgRNA_RL_3_F/R)	donor_RL_3_F/R	PhyBNT, synTALE-DBD_del1, PKI-NES/ PIF3, SV40-NLS, VP64-AD
RL_CT_GI_del1_3*	RL_CT_GI_del1	AGCCTGAAAAGCTCAAA CAG (sgRNA_RL_2_F/R)	donor_RL_2_F/R	PhyBNT, synTALE-DBD_del1/ Cdc1-NLS, PIF3, SV40-NLS, VP64-AD
RL_CT_GI_del1_4*	RL_CT_GI_del1	GATCGTCGCCCAAGAAG AAG (sgRNA_RL_1_F/R)	donor_RL_1_F/R	PhyBNT, synTALE-DBD_del1/ PIF3, VP64-AD
RL_CT_GI_del1_5*	RL_CT_GI_del1_2	AGCCTGAAAAGCTCAAA CAG (sgRNA_RL_2F/R)	donor_RL_2_F/R	PhyBNT, synTALE-DBD_del1, PKI-NES/ Cdc1-NLS, PIF3, SV40-NLS, VP64-AD
RL_CT_GI_del2_5*	RL_CT_GI_del2_2	AGCCTGAAAAGCTCAAA CAG (sgRNA_RL_2F/R)	donor_RL_2_F/R	PhyBNT, synTALE-DBD, PKI-NES/ Cdc1-NLS, PIF3, SV40-NLS, VP64-AD
RL_CT_GI_del2_6*	RL_CT_GI_del2_2	GATCGTCGCCCAAGAAG AAG (sgRNA_RL_1_F/R)	donor_RL_1_F/R	PhyBNT, synTALE-DBD, PKI-NES/ PIF3, VP64-AD

§ Produced by mutagenesis of template DNA indicated in column 2 and PCR products from oligonucleotides given in column 4.

* Generated by CRISPR/Cas9-based genome editing.

4.3 L-SCRaMbLE: a Tool for Light-controlled Cre-mediated Recombination in Yeast

Lena Hochrein[1], Leslie A. Mitchell[2], Karina Schulz[1], Katrin Messerschmidt[1] and Bernd Mueller-Roeber[*,3,4]

[1]University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[2]Institute for Systems Genetics, New York University Langone School of Medicine, New York City, NY 10016, USA.
[3]University of Potsdam, Department of Molecular Biology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[4]Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany

* To whom correspondence should be addressed.

The manuscript has been initially submitted as a letter to Nature Biotechnology/ Nature Communications. After submission of this thesis, it was accepted for publication by Nature Communications in a revised version after reviewer and editor assessments.

Here, we present L-SCRaMbLE, a chromophore-dependent, red light-controlled, split Cre recombinase system acting on *loxPsym* sites[1-4] in a reversible, dose-dependent and robust induction manner with very low basal activity. To demonstrate the recombination capacity of L-SCRaMbLE we tested it in yeast using a low-copy plasmid harboring four genes of a β-carotene pathway[5], each flanked by *loxPsym* sites. Recombination events induced by red light exposure (λ=660 nm) span a variety of possible outcomes, ranging from inversions and deletions to more complex rearrangements. The extent of recombination depends on induction time and concentration of the chromophore phycocyanobilin (PCB), which can be easily adjusted. When SCRaMbLE-ing[6, 7] a β-carotene pathway, L-SCRaMbLE provides improved recombination control over the previously reported estradiol-dependent SCRaMbLE induction system and circumvents the use of toxic chemicals. Thereby, L-SCRaMbLE boosts the potential for further

customization and provides a facile application for use in the *Saccharomyces cerevisiae* genome re-engineering project Sc2.0[8, 9] or in other recombination-based systems.

L-SCRaMbLE combines an optical dimerizer from *Arabidopsis thaliana*[10] and a split Cre recombinase from bacteriophage P1[11, 12], to be functional in yeast. The light-sensitive system builds on an N-terminal version of the photoreceptor phytochrome B[13, 14] (PhyBNT), which binds the chromophore-like compound phycocyanobilin (PCB)[15, 16]. Upon red light illumination PhyBNT undergoes a conformational switch from the Pr to the Pfr form, which is reversible by illumination with far-red light[17]. The Pfr form of PhyBNT binds to phytochrome interacting factor 3 (PIF3)[18, 19]. The light-dependent switch of the PhyBNT-PIF3 interaction is employed here to allow light-controlled reconstitution of functional Cre recombinase from otherwise non-functional N- and C-terminal halves of the protein (split Cre recombinase). The N-terminal domain of the Cre protein (CreN) was fused to PhyBNT (creating PhyBNT-CreN protein), while its C-terminal domain (CreC) was linked to PIF3 (creating PIF3-CreC). The sequences of the split Cre recombinase and the linkers were adopted from a blue light-responsible split Cre system, which is based on the CRY2-CIB1 system from *A. thaliana* and was reported to be functional in mammalian cells[11, 20]. We tested two different light-dependent proteins in this study, encoded by plasmids pLH_Scr15 and pLH_Scr16 (**Fig. 1**). Both proteins harbor a nuclear localization signal of the simian virus 40 (SV40-NLS) within the PIF3-CreC-fusion, but only pLH_Scr16 incorporates an additional NLS (from yeast Cdc1) within the PhyBNT-CreN-fusion. In the pLH_Scr15 system, PIF3-CreC protein is expected to be nuclearly localized in the dark, while PhyBNT-CreN is predicted to be located in the cytoplasm. After red light illumination, PhyBNT-CreN binds PIF3-CreC and the protein complex formed moves to the nucleus (**Fig. 1A**). In the pLH_Scr16 system, both, PIF3-CreC and PhyBNT-CreN are present in the nucleus in the dark. Upon red light illumination, PhyBNT-CreN binds PIF3-CreC which leads to reconstitution of functional Cre recombinase from the split halves (**Fig. 1B**). In both cases, reconstituted Cre initiates recombination at *loxPsym* sites present in the target DNA.

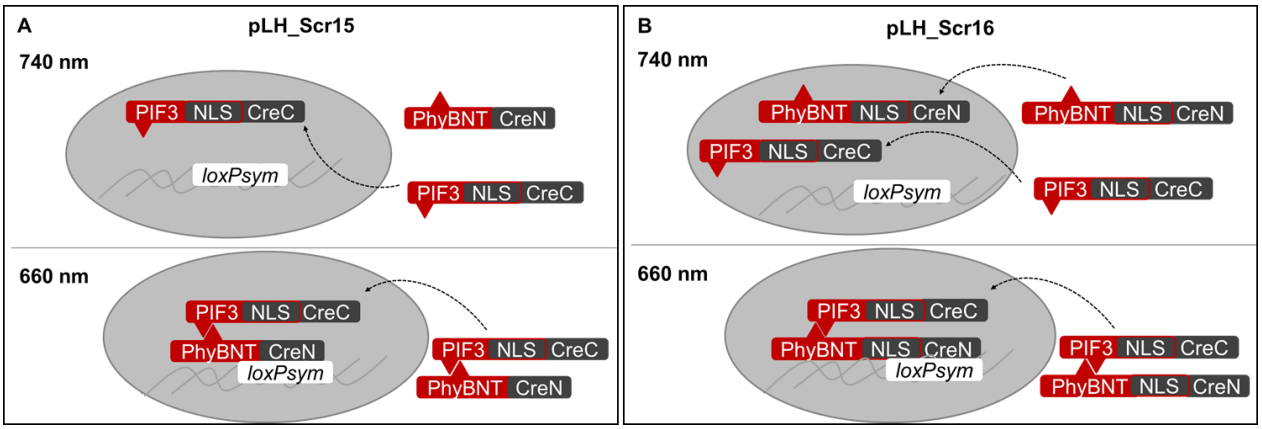


Figure 1: Schematic overview of L-SCRaMbLE. A Mode of action in cells harboring pLH_Scr15. In darkness or far-red light (λ = 740 nm), PIF3-CreC and PhyBNT-CreN are located in the nucleus and cytoplasm, respectively. Upon red light illumination (λ = 660 nm) PhyBNT-CreN binds PIF3-CreC, thereby reconstituting Cre recombinase from its split halves. Reconstituted Cre moves to the nucleus where it induces recombination between the *loxPsym* sites. **B** Mode of action in cells harboring pLH_Scr16. In darkness or far-red light (λ =740 nm) PIF3-CreC and PhyBNT-CreN are separately located in the nucleus. Upon red light illumination (λ =660 nm) PhyBNT-CreN binds PIF3-CreC, thereby reconstituting Cre recombinase from its split parts, allowing recombination at *loxPsym* sites.

To test L-SCRaMbLE, we used the low-copy plasmid pLM494 (CEN/ARS), which harbors four functional genes of the β-carotene biosynthesis pathway each flanked by *loxPsym* sites. Yeast cells carrying this pathway produce β-carotene and its precursors resulting in yellow colonies, as previously shown for the β-carotene-producing parental plasmid pJC178 lacking the *loxPsym* sites[5]. In the presence of active Cre, the symmetric *loxPsym* sites can lead to inversions, deletions, and/or duplications on pLM494[6, 7]. While Cre-driven inversions of the *loxPsym*-flanked genes should not affect β-carotene production, deletion of *crtI* and/or *crtYB* is expected to result in a colorless product and, consequently, a white yeast colony[21]. Deletion of only *tHMG1* would cause orange colonies and deletion of *crtE* and *tHMG1* is expected to result in faintly yellow colonies[5, 21]. Nevertheless, white colonies can be used to readout recombination events, and the ratio of white to yellow colonies to approximate recombination frequency. However, the fraction of white colonies observed post-SCRaMbLE induction is expected to be an underestimate for the following reasons: First, as CEN/ARS plasmids occur in 2-5 copies per cell[22] white colonies are only expected if all copies of pLM494 in a cell are recombined and second, the formation of dicentric plasmids due to recombination between two copies of pLM494 should cause cell death.

To test red light-inducible Cre efficiency, plasmid pLM494 was cotransformed into yeast with either plasmid pLH_Scr15 or pLH_Scr16, or - to compare the light-inducible systems with the already established β-estradiol-inducible Cre system[8] - with plasmid pLM006. The resulting strains were designated Light-Cre1, Light-Cre2 and EST-Cre, respectively. To evaluate L-SCRaMbLE, the light-controlled Cre recombinase systems were tested with regard to induction time- and PCB-dependency. Cells carrying Light-Cre1, Light-Cre2 and EST-Cre were pre-cultured for 6 h in the dark. Subsequently, Light-Cre1 and Light-Cre2 cultures were induced with a red light pulse (660 nm) of 5 min, and grown for another 4 h or 16 h with 10-sec red light pulses every 5 min. EST-Cre cultures were induced by adding β-estradiol and subsequently grown for another 4 h or 16 h. Non-induced controls were kept in the dark (Light-Cre1/2) or in the absence of β-estradiol (EST-Cre). Cells were then plated on appropriate selective media without PCB or β-estradiol and incubated for 2 days at 30°C in the dark. For cultures plated after four hours of induction, 14% and 27% of colonies from Light-Cre1 and Light-Cre2, respectively, were white, while 94% of EST-Cre colonies were white. In the non-induced samples, 2% and 10% of Light-Cre1 and Light-Cre2 colonies were white, compared to 48% of EST-Cre colonies (**Fig. 2A; Supplementary Fig. 1A**). While these results demonstrate highly efficient recombination mediated by the β-estradiol-induced Cre recombinase, it is also clear that the system exhibits high basal activity in the absence of inducer in this experiment. Notably, such high basal activity has not been observed for the EST-Cre system in the context of a circular Sc2.0 chromosome arm, *synIXR*[23]. In contrast, L-SCRaMbLE exhibits only low background activity and yields colonies with recombination events after 4 h of red light induction. Thus, L-SCRaMbLE is nearly inactive in its off-state, while it has a moderate activity in the on-state. Of note, Light-Cre1 exhibits lower basal activity than Light-Cre2, consistent with the fact that in the Light-Cre1 system PIF3-CreC resides in the nucleus, while PhyBNT-CreN is located in the cytosol before induction, thereby avoiding spurious reconstitution of the split Cre recombinase. For samples plated after 16 h of induction, 100% of the EST-Cre colonies were white (45% in non-induced samples) (**Fig. 2B, Supplementary Fig. 1B**). In the case of Light-Cre1 and Light-Cre2, 1% and 7% of the colonies, respectively, were white in non-induced cultures, similar to the 4 h experiment. However, the percentage of white

colonies significantly increased in light-induced cultures (32% in Light-Cre1, 47% in Light-Cre2), consistent with ongoing light-dependent activation of L-SCRaMbLE-mediated recombination.

To raise the percentage of white colonies produced by L-SCRaMbLE even further, we extended the induction time to 24 h (**Fig. 2C, Supplementary Fig. 1C**). While the basal activity remained low in both systems (5% for Light-Cre1, 8% for Light-Cre2), the number of white colonies increased to 52% after red light induction in Light-Cre1 but remained largely unchanged (compared to the 16 h experiment) for Light-Cre2 (53% white).

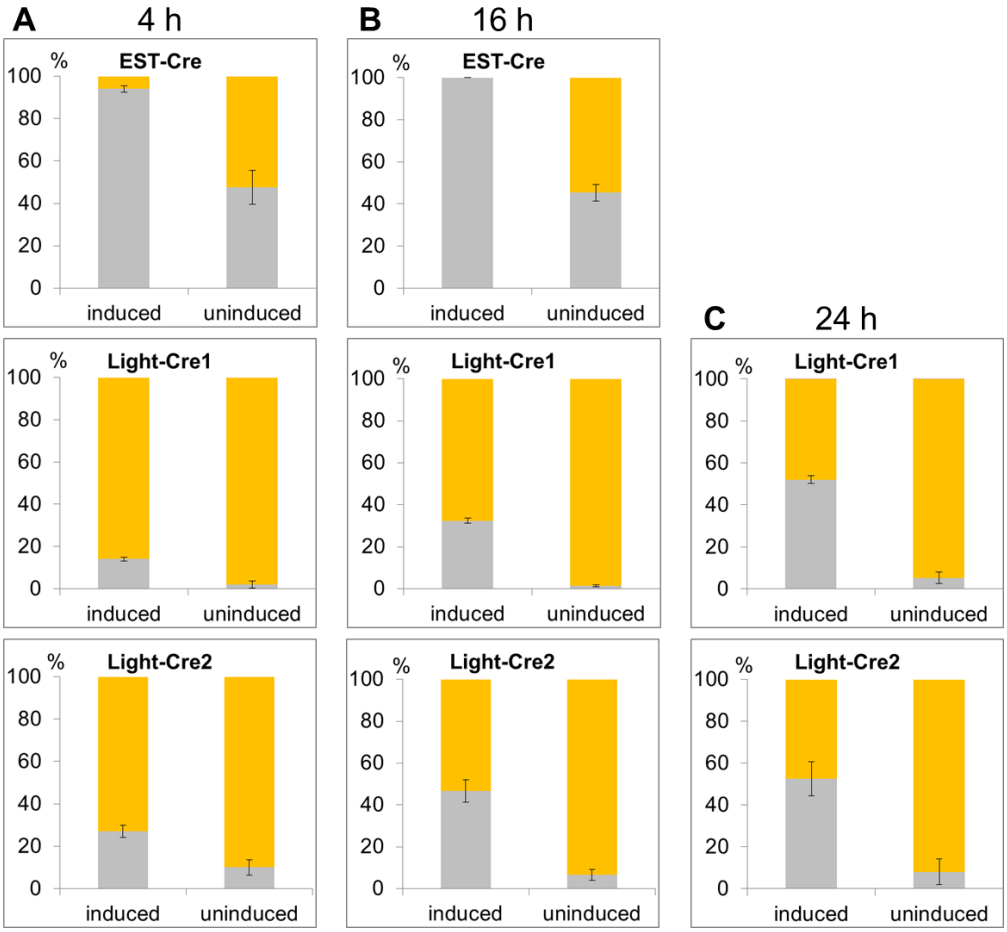


Figure 2: Recombination experiment with incubation times of 4 h, 16 h or 24 h. Light-Cre1, Light-Cre2 and EST-Cre cells were grown for 6 h in darkness. After induction with a 5-min red light pulse or by addition of β-estradiol (2 µM final concentration), samples were grown for further **A** 4 h, **B** 16 h and **C** 24 h at 30°C and 230 rpm. One ml of each culture was pelleted, diluted and plated on appropriate SD media. Each experiment was performed in three independent replicates. Bar blots show the mean values of the percentages of white and yellow colonies in induced and uninduced samples for strains Light-Cre1, Light-Cre2 and EST-Cre. Grey, white colonies; yellow, yellow colonies. Error bars indicate SD.

Collectively, our data reveal L-SCRaMbLE as a reliable tool with low basal activity in the absence of light, and high recombination capacity after red light induction. Importantly, recombination frequency can be controlled in a time- and light dose-dependent fashion. As light-controlled activation was more pronounced using Light-Cre1 compared to Light-Cre2, we used the Light-Cre1 system for further experiments.

The Sc2.0 project incorporates SCRaMbLE as an inducible system to investigate genome evolution-related aspects and to generate reduced or even minimal genomes. Beneficial in such studies are numerous small-scale modifications of the native genome, allowing the investigation of many different genomic states, while large deletions encompassing essential genes are expected to cause cell death. To this aim we had a deeper look at the rearrangements to the β-carotene pathway caused by Light-Cre1 and EST-Cre. We isolated plasmids from white and yellow colonies from induced and non-induced plates of the 16 h experiment. Plasmids were retransformed into *Escherichia coli*, isolated after culture, and digested with restriction enzymes *Pst*I and *Sac*I. The resulting fragments were analyzed to identify individual recombination events. The restriction patterns were matched to those of pLM494 and different hypothetical plasmids resulting from possible recombination events. One representative plasmid of each pattern was sequenced. For Light-Cre1, we isolated plasmids from two white and two yellow colonies from a non-induced plate. As expected, restriction patterns matched those of unmodified pLM494 for the yellow colonies, and different gene deletions for the white colonies. This result is consistent with the assumption that yellow colonies carry genetically intact β-carotene pathways, while white colonies contain recombined plasmids with gene deletions dependent on basal Cre activity since all rearrangements occurred specifically at *loxPsym* sites (**Fig. 3**). The patterns of 12 plasmids from light induced colonies (2 yellow, 10 white) all differ from that of pLM494 and mostly also differ from each other. This demonstrates that L-SCRaMbLE yields a wide spectrum of unique deletion and inversion events after light induction. Moreover, that most plasmids yielded different digestion patterns is consistent with the recovery of plasmids from unrelated yeast cells, indicating that the 16 h induction culture is not dominated by siblings carrying recombined pathways from a small number of mother cells. We did not observe any restriction patterns consistent with duplication events in this analysis, possibly

because we focused largely on studying events in white colonies post-induction. Plasmids recovered from two yellow colonies post-light induction unexpectedly had restriction patterns consistent with deletion of β-carotene pathway genes (**Fig. 3**, plasmids 5 and 6). Specifically, digestion of plasmid 6 suggests deletion of both *tHMG1* and *crtYB* genes and plasmid 5 is consistent with deletion of all four pathway genes; both plasmids should yield white colonies. We hypothesized that the original yeast colonies encoded multiple uniquely recombined pathways or that we had recovered the plasmid from a mixed colony. To investigate further, four additional independent *E. coli* colonies for both plasmids 5 and 6 were analyzed by DNA restriction. The restriction patterns confirmed that both parental colonies contain a mixture of different plasmids (**Supplementary Fig. 2A**). While plasmids originating from colony 6 collectively show to contain all four genes required for β-carotene biosynthesis, the plasmids isolated from colony 5 are expected to yield a white colony. To distinguish whether the parental yeast colony contained a mixture of different cells or whether cells of the parental colony are identical but contain multiple variants of the β-carotene plasmids, the cells were streaked out to obtain single colonies (**Supplementary Fig. 2B**). For colony 5 we observed a mixed population of white and yellow colonies. Plasmids of a white and yellow yeast colony were isolated and retransformed into *E. coli*, and four independent *E. coli* colonies were tested by restriction analysis, confirming uniform plasmid population for the white colony. For the yellow colony, a mixture of three plasmids with a single *crtI* deletion and one plasmid with a *crtE* inversion was obtained. Both results fit the observed colony color (**Supplementary Fig. 2C**). For colony 6, almost no white colonies were detected upon re-streaking, suggesting cells of colony 6 contained different β-carotene plasmids. For EST-Cre, we investigated plasmids isolated from seven non-induced colonies (2 yellow, 5 white) and found that none of them showed restriction patterns matching pLM494. Sequencing confirmed that plasmid 17 contained inverted sequences of two genes, *crtI* and *tHMG1*, while plasmid 18 had an inversion of *crtI* and a deletion of *tHMG1*. Each of the seven EST-Cre colonies of the uninduced experiment contained recombined plasmids, indicating that the basal activity is even higher than expected from counting white colonies. Twelve out of 12 plasmids isolated from white

colonies of the EST-Cre induced experiment (16 h) showed restriction patterns of recombined pLM494 with deletions of always all four β-carotene biosynthesis genes.

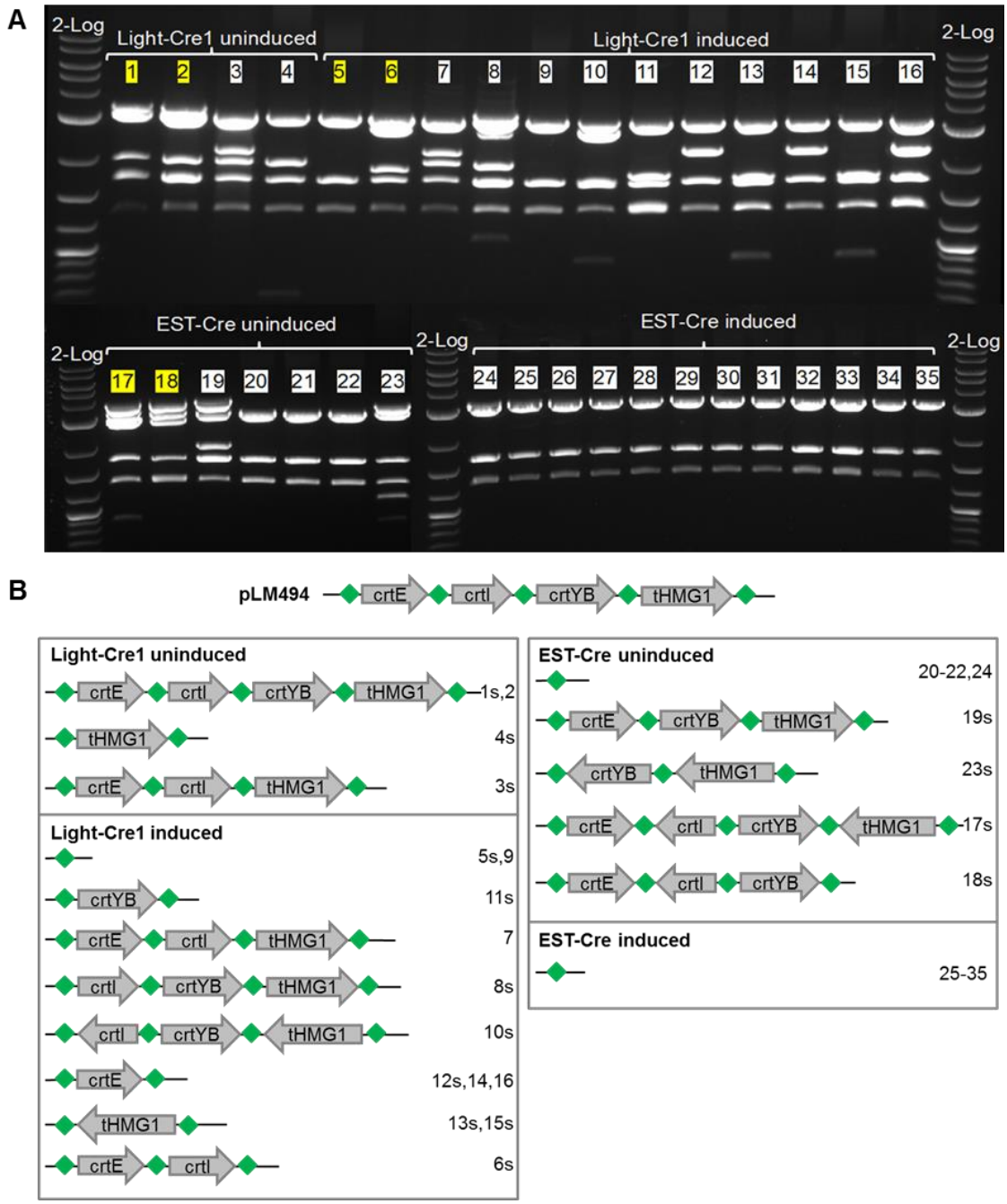


Figure 3: Restriction patterns of pLM494 after Cre-mediated recombination. Plasmids were isolated from different yeast colonies of induced and uninduced Light-Cre1 and EST-Cre cultures (16 h) and after passage through *E. coli* digested with *Pst*I and *Sac*I. **A** Gel electrophoresis of digested plasmids. Numbers shaded in yellow show patterns of plasmids isolated from yellow yeast colonies, while numbers shaded in white show patterns of plasmids isolated from white colonies. 2-Log, DNA Ladder (NEB). **B** Overview of the recombination events. Gene names are given. Numbers on the right refer to the corresponding

plasmid restrictions shown in (A). "s" indicates sequence-verified plasmids. Green diamonds indicate *loxPsym* sites.

The activity of the EST-Cre system in the absence of estradiol appears to be significantly boosted in the context of pLM494 (**Supplementary Fig. 3**) compared to SCRaMbLE of synthetic yeast chromosome *synIXR*, a 90-kb circular yeast chromosome arm with 43 *loxPsym* sites[23]. It is not currently clear what underlies this result.

As L-SCRaMbLE allows light-triggered recombination to be controlled in a time-dependent manner, we next investigated whether PCB-dependent regulation is also feasible. PCB, the chromophore of the PhyB photoreceptor, enables the light-responsive conformational switch of the holoenzyme from the Pr to the Pfr form. PCB dose-dependency was previously reported for a different gene expression control system in *S. cerevisiae*[24]. The concentration of PCB supplied to SD-Ura-Leu media was varied from 0 – 25 µM. Light-Cre1 cells were pre-cultured for 6 h in the dark, subsequently induced with a red light pulse of 5 min, and grown for another 16 h with 10-sec red light pulses every 5 min before plating. Non-induced samples were grown in the dark throughout. Irrespective of the PCB concentration used, cultures grown in the dark always yielded fewer than 1% white colonies (**Supplementary Fig. 4 and 5**); the same result was observed for illuminated cells grown in 0 or 5 µM PCB. In contrast, light-induced cultures grown in 15 and 25 µM PCB yielded plates with 10% and 47% white colonies, respectively. For recombination experiments, termination of Cre activity is an important aspect to ensure stability of the newly generated genetic construct. The results shown here demonstrate that virtually no recombination events occur in the absence of the chromophore PCB, regardless of the light condition used. This suggests that Cre can be inactivated by transferring cells to medium lacking PCB, in addition to inactivation by far-red light application. To confirm this, 10 yellow and 10 white colonies resulting from samples induced in medium with 25 µM PCB (**Supplementary Fig. 5D**) were re-streaked on PCB-free medium (**Supplementary Fig. 6A, B**). After re-streaking, all 10 white colonies produced white sibling colonies, as expected. Of the 10 yellow colonies, eight produced sibling colonies of identical color as the parental ones. Only two colonies (numbers 4 and 8; **Supplementary Fig. 6A**) resulted in very few white colonies, while

the vast majority of sibling colonies retained the same color as the parental ones. The plate assay thus indicated low Cre recombination activity after removal of PCB from the medium. To verify this, plasmids were isolated from four single colonies derived from two yellow (A1 and A5, **Supplementary Fig. 6A**) and two white parental colonies (B2 and B5, **Supplementary Fig. 6B**) and retransformed into *E. coli*. Plasmids extracted from four independent *E. coli* colonies were analyzed by enzyme restriction. All 16 plasmids isolated from four sibling colonies of parental colony B5 showed deletions of all four pathway genes. Consequently, these plasmids cannot be used to monitor undesired recombination events in the off-state. However, also all 16 plasmids derived from four sibling colonies of parental colony B2 harbored the same plasmid (tHMG1 only) without further recombination events (**Supplementary Fig. 6C**). For each of the yellow parental colonies A1 and A5, 15 out of 16 plasmids passaged through *E. coli* showed restriction patterns matching those of plasmid pLM494. Only one plasmid from A1 (colony 3_2) and A5 (colony 4_4) differ from the parental colony resulting from gene deletions.

Taken together, L-SCRaMbLE allows for tight control of recombination pre-induction and high Cre-mediated recombination after light induction requiring only brief light pulses. The fact that L-SCRaMbLE can additionally be controlled by PCB concentration makes it an excellent tool for fine-tuning recombination frequency in yeast, according to experimental needs. Furthermore, withdrawal of PCB from the medium allows a complete shut-down of L-SCRaMbLE-mediated recombination. When applied to Sc2.0 synthetic strains[25-30], which will encode 1000's of *loxPsym* sequences across synthetic chromosomes and therefore demand exquisitely fine control of Cre activity, L-SCRaMbLE will be an excellent tool.

ACKNOWLEDGMENTS

This research was funded by the Federal Ministry of Education and Research of Germany (BMBF; FKZ 031A172). L.H. thanks the Potsdam Graduate School for providing a scholarship and Dr. Fabian Machens for helpful discussions.

AUTHOR CONTRIBUTIONS

L.H. designed and developed the overall L-SCRaMbLE strategy with contributions from L.A.M. L.H. conducted the experiments and analyzed the data. L.H. wrote the manuscript, with contributions from B.M.-R. and L.A.M. L.A.M. designed and constructed plasmids pLM494 and pLM006. K.S. constructed plasmids pLH_SCR12 to pLH_SCR16. K.M. and B.M.-R. jointly supervised the Cell2Fab group. K.M. edited the paper. All authors take full responsibility for the content of the paper.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interest.

MATERIALS & CORRESPONDENCE

All correspondence and material requests should be sent to B.M.-R.

REFERENCES

1. Hoess, R.H., Ziese, M. & Sternberg, N. P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proceedings of the National Academy of Sciences* **79**, 3398-3402 (1982).
2. Hoess, R.H. & Abremski, K. Mechanism of strand cleavage and exchange in the Cre-*lox* site-specific recombination system. *Journal of molecular biology* **181**, 351-362 (1985).
3. Sternberg, N. & Hamilton, D. Bacteriophage P1 site-specific recombination: I. Recombination between *loxP* sites. *Journal of molecular biology* **150**, 467-486 (1981).
4. Hoess, R.H., Wierzbicki, A. & Abremski, K. The role of the *loxP* spacer region in PI site-specific recombination. *Nucleic Acids Research* **14**, 2287-2300 (1986).
5. Mitchell, L.A. et al. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Research* **43**, 6620-6630 (2015).
6. Dymond, J. & Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered* **3**, 170-173 (2012).
7. Jovicevic, D., Blount, B.A. & Ellis, T. Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. *BioEssays* **36**, 855-860 (2014).
8. Dymond, J.S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
9. Richardson, S.M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).

10. Shimizu-Sato, S., Huq, E., Tepperman, J.M. & Quail, P.H. A light-switchable gene promoter system. *Nature biotechnology* **20**, 1041-1044 (2002).
11. Kennedy, M.J. et al. Rapid blue-light-mediated induction of protein interactions in living cells. *Nature methods* **7**, 973-975 (2010).
12. Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *genesis* **26**, 99-109 (2000).
13. Quail, P.H. An emerging molecular map of the phytochromes. *Plant, Cell & Environment* **20**, 657-665 (1997).
14. Smith, H. Phytochromes and light signal perception by plants-an emerging synthesis. *Nature* **407**, 585-591 (2000).
15. Kunkel, T., Neuhaus, G., Batschauer, A., Chua, N.H. & Schäfer, E. Functional analysis of yeast-derived phytochrome A and B phycocyanobilin adducts. *The Plant Journal* **10**, 625-636 (1996).
16. Kunkel, T., Speth, V., Büche, C. & Schäfer, E. *In Vivo* Characterization of Phytochrome-Phycocyanobilin Adducts in Yeast. *Journal of Biological Chemistry* **270**, 20193-20200 (1995).
17. Rockwell, N.C., Su, Y.-S. & Lagarias, J.C. Phytochrome structure and signaling mechanisms. *Annual Review of Plant Biology* **57**, 837-858 (2006).
18. Ni, M., Tepperman, J.M. & Quail, P.H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* **95**, 657-667 (1998).
19. Ni, M., Tepperman, J.M. & Quail, P.H. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature* **400**, 781-784 (1999).
20. Taslimi, A. et al. Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nature chemical biology* **12**, 425-430 (2016).
21. Verwaal, R. et al. High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Applied and Environmental Microbiology* **73**, 4342-4350 (2007).
22. Karim, A.S., Curran, K.A. & Alper, H.S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS yeast research* **13**, 107-116 (2013).
23. Shen, Y. et al. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome research* **26**, 36-49 (2016).
24. Hughes, R.M., Bolger, S., Tapadia, H. & Tucker, C.L. Light-mediated control of DNA transcription in yeast. *Methods* **58**, 385-391 (2012).
25. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).
26. Mitchell, L.A. et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355** (2017).
27. Shen, Y. et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355** (2017).
28. Wu, Y. et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355** (2017).
29. Xie, Z.-X. et al. “Perfect” designer chromosome V and behavior of a ring derivative. *Science* **355** (2017).
30. Zhang, W. et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355** (2017).

METHODS

Strains, Growth Conditions and Reagents

Escherichia coli strain DH5α was used for cloning purposes. For experiments with *Saccharomyces cerevisiae,* strain BY4742 was cultured at 30°C in Yeast extract Peptone Dextrose Adenine (YPDA)-rich medium or in appropriate Synthetic Dextrose (SD) media lacking one or more amino acids to select for the maintenance of transformed plasmids. NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel, Düren, Germany) were used for PCR purification and gel elution, and plasmid purification was done with NucleoSpin Plasmid EasyPure kit (Macherey Nagel).

Cloning and DNA Manipulation Techniques

For overlap-based *in vitro* DNA assembly, SLiCE cloning was done as previously reported[1, 2]. DNA fragments containing 36 bp-long homology regions were amplified by PCR. Correct assemblies were verified by DNA sequencing. For dephosphorylation of digested vector backbones Antarctic Phosphatase (New England Biolabs, Frankfurt am Main, Germany) was used. Phusion polymerase (Life Technologies GmbH, Darmstadt, Germany) was used for PCR amplifications. Restriction enzymes were purchased from New England Biolabs.

In vivo DNA Assembly and Cell Transformation Protocols

Saccharomyces cerevisiae sequential transformations were carried out using the LiAc/SS carrier DNA/PEG method by Gietz and Schiestl[3]. For multiple-fragment assemblies, 100 ng of each part were used. Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research) was used for plasmid preparation from yeast colonies.

Constructs

Plasmid pLM006 was made by Gibson assembly, amplifying the pSCW11-Cre-EBD transcription unit from pDL12[4] and inserting it into the *Sma*I site of pRS413[5]. Plasmid pLM494 was built using a previously assembled construct encoding the β-carotene pathway, pJC178, as starting material[6]. Briefly, PCR amplicons spanning each of the

four β-carotene transcription units (TU) were generated using overhang primers to add a *loxPsym* sequence to one end of each TU and *Bsa*I sites to both ends. The *Bsa*I sites were oriented to cut themselves out, leaving non-palindromic, unique 4-bp overhangs for directional assembly with a yeast Golden Gate acceptor vector[7]. Golden gate reactions were carried out as previously described[7].

Plasmids pLH_Scr15 and pLH_Scr16 were constructed with the overlap-based DNA assembly methods SLiCE or TAR cloning and by using the AssemblX toolkit[1, 8, 9]. To this end, pLH_Scr12, pLH_Scr13 and pLH_Scr14 were constructed as precursor (Level 0) plasmids according to the AssemblX cloning procedure[8]. All parts used in this study to clone pLH_Scr12 to pLH_Scr16, the appropriate sources and sequence modifications, if applicable, are listed in **Supplementary Table 1**, while sequences of PCR primers are provided in **Supplementary Table 2**. Sequences and plasmid maps of all constructs described here are given in **Supplementary Data File 1** and **Supplementary Figures 7 and 8**.

To construct Level 0 plasmids pLH_Scr12 and pLH_Scr13, plasmid pRot4 containing a fusion of an N-terminal version of the PhyB (PhyBNT) and the nuclear localization signal Cdc1-NLS under control of the *TDH3* promoter and *TDH3* terminator was linearized by *Bam*HI/*Age*I restriction. Subsequently, the DNA fragments encoding CreN, resulting from PCR on pLM006 with primers L714/L777 and L777/L778, respectively, were inserted via SLiCE cloning. We mutated a few nucleotides in the CreC encoding sequence to match the CreC amino acid sequence encoded by plasmid pmCherry-CIBN-CreC[10] (see **Supplementary Table 1**). A linker from plasmid pmCherry-CRY2-CreN was used to fuse CreN to PhyBNT, and CreC to PIF3, respectively[10]. To construct pLH_Scr14, plasmid pRot1, containing the yeast *FBA1* promoter and terminator, was linearized with *Not*I. Thereafter, a PCR fragment encoding the PIF3-SV40-NLS-fusion was generated with primers L280/L773 and template pRot2. A DNA fragment encoding CreC was obtained by PCR on template pLM006, with primers L712/L774 and L775/L776. All three PCR products were inserted via SLiCE cloning into linearized pRot1 backbone. To construct Level 1 plasmid pLH_Scr15, Level 0 plasmids pLH_Scr12 and pLH_Scr14 were digested with *Pme*I and the resulting fragments containing homology regions A0/A1 and A1/A2, respectively, were combined together with annealed oligonucleotides L706/L707 into *Pac*I-linearized pL1A0*B0*_Leu via TAR

cloning[8]. To construct Level 1 plasmid pLH_Scr16, Level 0 plasmids pLH_Scr13 and pLH_Scr14 were digested with *Pme*I and the resulting large fragments containing homology regions A0/A1 and A1/A2, respectively, were combined together with annealed oligonucleotides L706/L707 into *Pac*I-linearized pL1A0*B0*_Leu via TAR cloning.[8]

Recombination Experiments

For recombination experiments, Light-Cre1, Light-Cre2, and EST-Cre cells were inoculated in 3 ml appropriate drop-out medium and cultured overnight at 30°C on a rotary shaker (230 rpm). Hundred µl of overnight grown Light-Cre1 and Light-Cre2 cultures were inoculated in 10 ml SD-Ura-Leu medium containing 25 µM PCB (Livchem, Frankfurt am Main, Germany) in 100-ml baffled Erlenmeyer flasks. Similarly, 100 µl of EST-Cre overnight culture were inoculated in 10 ml SD-Ura-His medium in 100-ml baffled flasks. For inactivation, Light-Cre1 and Light-Cre2 cultures were irradiated with a 1-min far-red light pulse and afterwards grown for 6 h in darkness at 30°C and 230 rpm. Thereafter, cultures were irradiated with a 5-min red-light pulse and then grown for 4, 16 or 24 h with 10-sec red-light pulses applied every 5 min. Non-induced samples were maintained in darkness throughout. EST-Cre cells were grown in darkness for 6 h at 30°C and 230 rpm, then induced with 2 µM β-estradiol (Sigma-Aldrich, Munich, Germany) and grown for 4, 16 or 24 h. Non-induced samples were maintained in medium without β-estradiol. For all samples 1 ml of cells was pelleted, redissolved in 1 ml ddH_2O, diluted 1:100 and 1:1,000 in ddH_2O, and plated on SD-Ura-Leu (Light-Cre1/2) or SD-Ura-His (EST-Cre).

To test PCB-dependent recombination, overnight cultures of strain Light-Cre1 were inoculated and treated as described above, except that PCB concentration was varied (0, 5, 15 and 25 µM) in the 10-ml SD-Ura-Leu media.

Light Sources

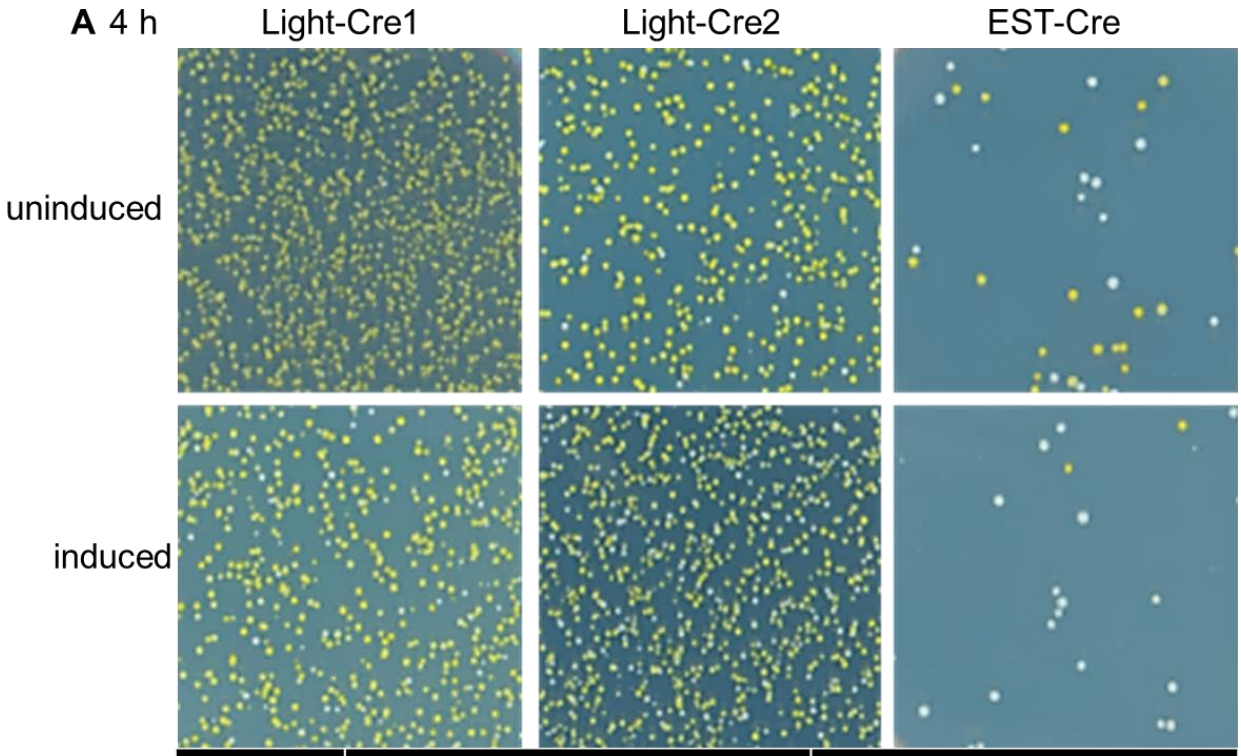
COMPLED Solutions GmbH (Dresden, Germany) installed LED light sources for red light (λ = 640 - 680 nm, peak at 661 nm) and far-red light (λ = 720 – 760 nm, peak at 740 nm) into an INFORS HT Multitron Pro incubator (INFORS AG, Bottmingen, Switzerland). 40 LEDs per wavelength were installed on the ceiling of the incubator; light

intensity was 2.8 mW/cm^2 for red light and 6.9 mW/cm^2 for far-red light. For uniform illumination, the incubator walls were covered with light-reflecting foil. Light-sensitive experiments in PCB-containing medium were done under green safelight using a standard LED strip.

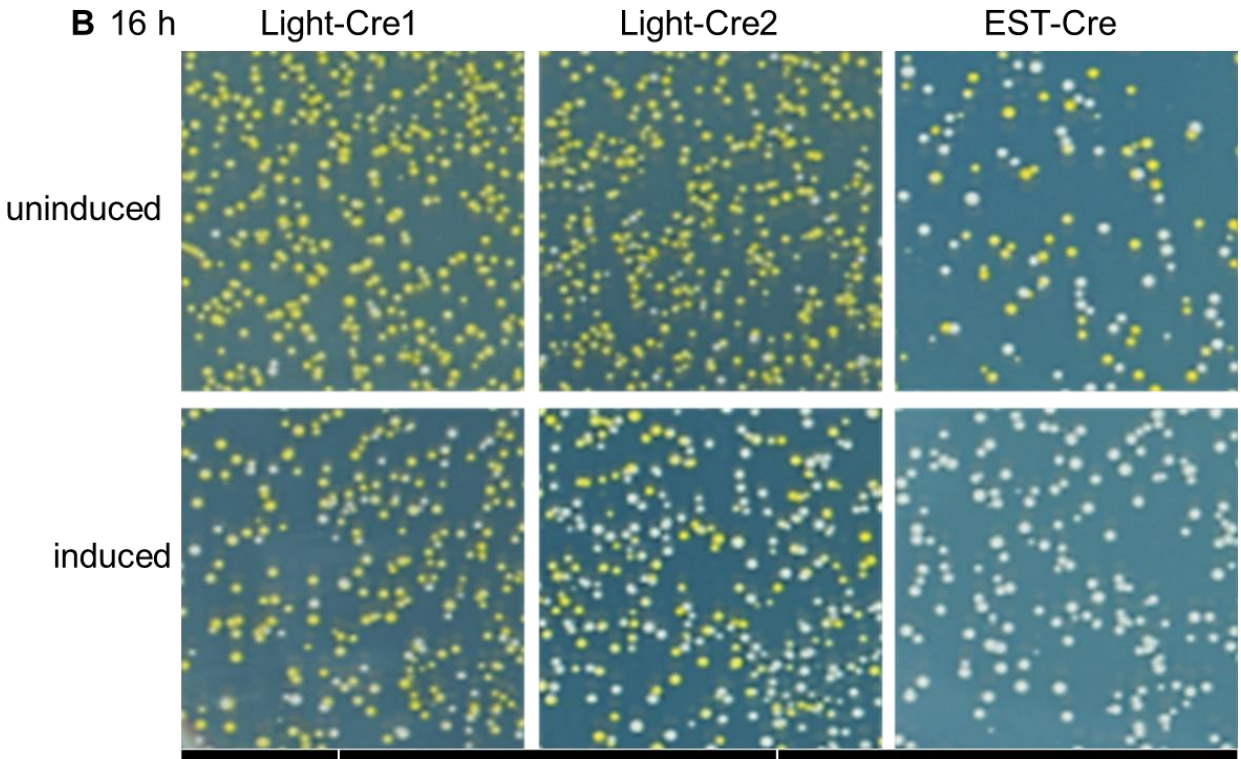
METHODS-ONLY REFERENCES

1. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research* **40**, e55 (2012).
2. Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K. & Mueller-Roeber, B. Characterizing seamless ligation cloning extract for synthetic biological applications. *Analytical Biochemistry* **509**, 24-32 (2016).
3. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols* **2**, 31-34 (2007).
4. Lindstrom, D.L. & Gottschling, D.E. The mother enrichment program: a genetic system for facile replicative life span analysis in *Saccharomyces cerevisiae*. *Genetics* **183**, 413-422 (2009).
5. Sikorski, R.S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19-27 (1989).
6. Mitchell, L.A. et al. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Research* **43**, 6620-6630 (2015).
7. Agmon, N. et al. Yeast Golden Gate (yGG) for the Efficient Assembly of *S. cerevisiae* Transcription Units. *ACS Synthetic Biology* **4**, 853-859 (2015).
8. Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Research*, gkx034 (2017).
9. Kouprina, N. & Larionov, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics* **7**, 805-812 (2006).
10. Kennedy, M.J. et al. Rapid blue-light-mediated induction of protein interactions in living cells. *Nature methods* **7**, 973-975 (2010).

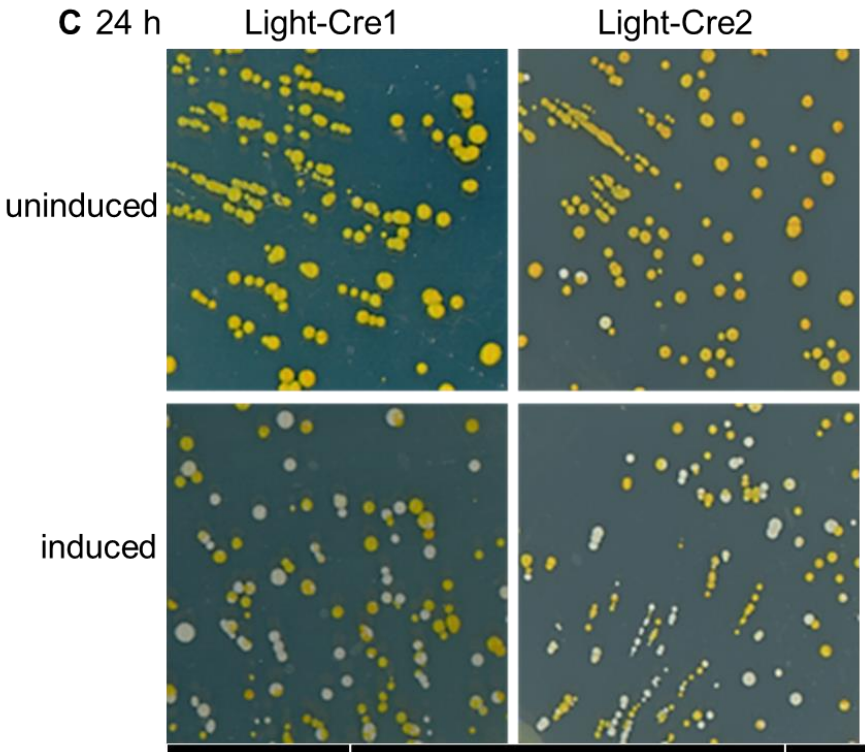
Supplementary Data



	uninduced		induced	
	white	yellow	white	yellow
Light-Cre1	0/ 5/ 2 0%/ 4%/ 2%	216/ 138/ 134	40/ 37/ 25 13%/ 15%/ 14%	266/ 212/ 159
Light-Cre2	10/ 16/ 7 6%/ 15%/ 9%	152/89/67	49/ 41/ 36 23%/ 30%/ 28%	165/95/91
EST-Cre	36/ 64/ 31 43%/ 59%/ 41%	47/ 45/ 45	123/ 47/ 44 96%/ 92%/ 94%	5/ 4/ 3

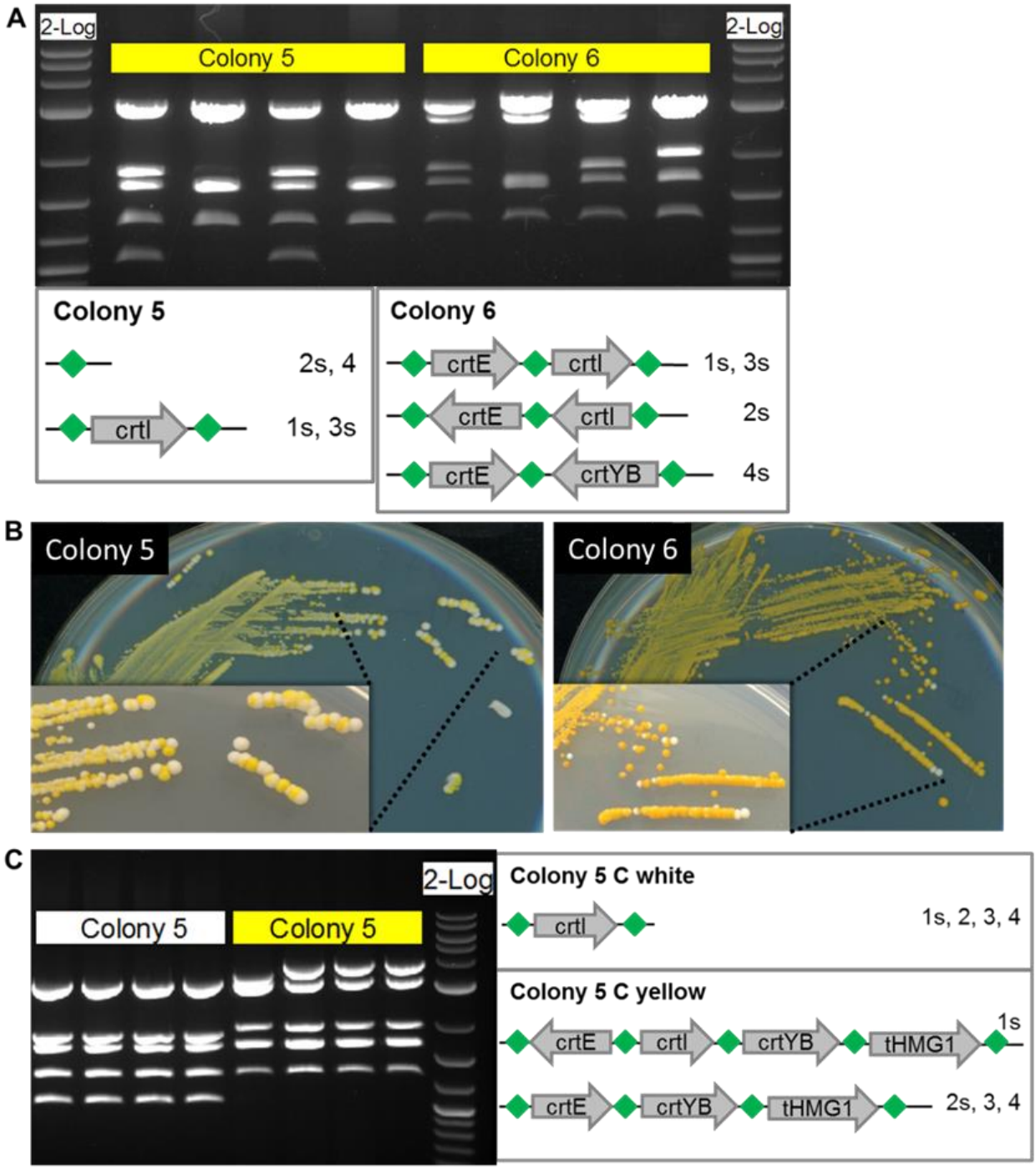


	uninduced		induced	
	white	yellow	white	yellow
Light-Cre1	12/ 20/ 2 1%/ 2%/ 1%	1584/ 848/ 200	492/ 312/ 44 34%/ 31%/ 32%	952/ 696/ 94
Light-Cre2	84/ 32/ 18 8%/ 3%/ 9%	976/ 1224/ 182	528/ 496/ 70 42%/ 44%/ 54%	728/ 640/ 60
EST-Cre	360/ 512/ 376 38%/ 46%/ 52%	584/ 596/ 344	352/ 500/ 484 100%/ 100%/ 100%	0/ 0/ 0

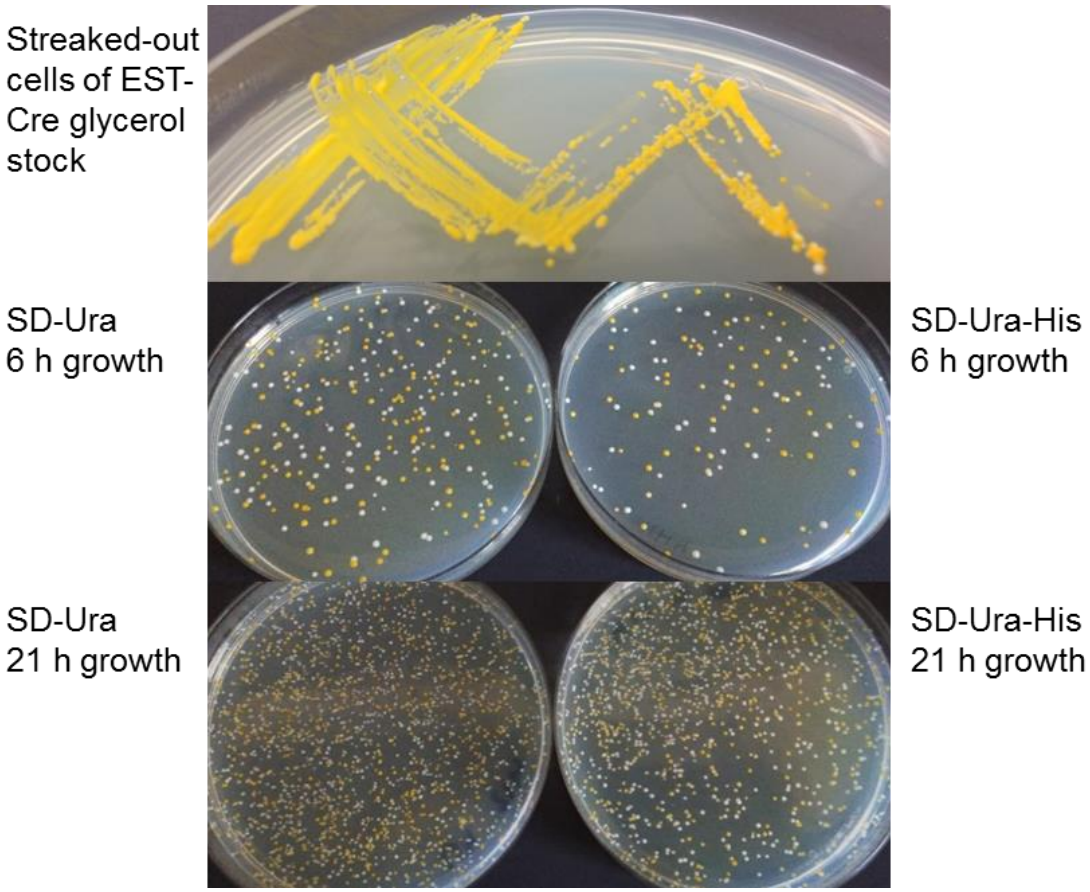


	uninduced		induced	
	white	yellow	white	yellow
Light-Cre1	8/ 5/ 3 9%/ 4%/ 3%	83/ 115/ 112	52/ 81/ 68 50%/ 52%/ 54%	52/ 76/ 57
Light-Cre2	4/ 21/ 5 5%/ 17%/ 3%	79/ 104/ 180	92/ 34/ 64 62%/ 43%/ 53%	56/ 46/ 57

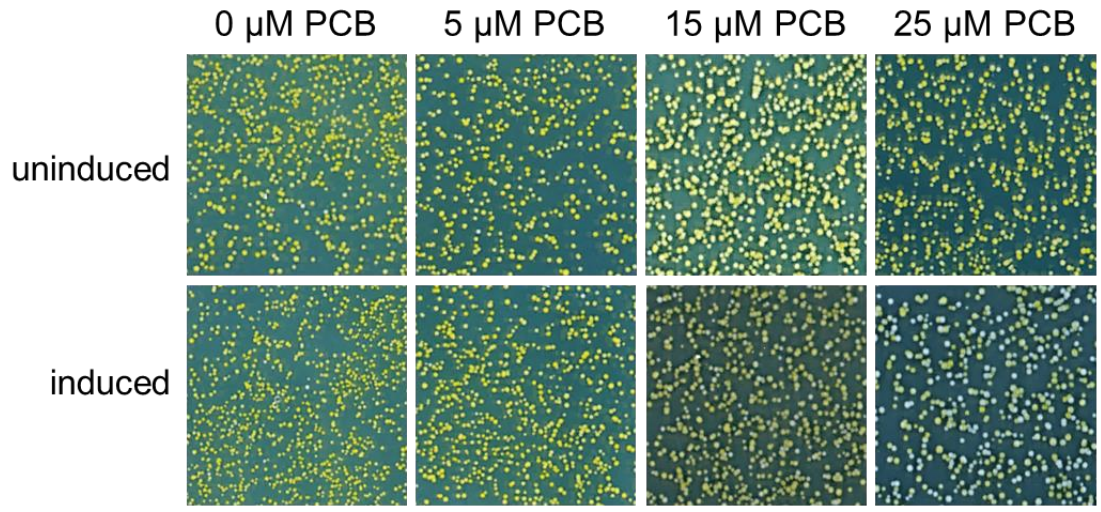
Supplementary Figure 1: Recombination experiment with incubation times of 4 h, 16 h and 24 h. Light-Cre1, Light-Cre2 and EST-Cre cells were grown for 6 h in darkness. After induction with a 5-min red light pulse (Light-Cre1/2) or with 2 µM β-estradiol (EST-Cre), samples were grown for further **A** 4 h, **B** 16 h, and **C** 24 h at 30°C and 230 rpm with 10-sec red light pulses every 5 min (Light-Cre1/2) or in the dark (EST-Cre). One ml of each culture was pelleted, diluted, plated on appropriate SD medium, and grown for another 2 d at 30°C in the dark. Each experiment was performed in three independent replicates. Plates show the result of one representative replicate of each strain. The tables below the figures show values for the counted colonies (first row) and the percentage of white colonies (second row) for all three replicates.



Supplementary Figure 2: Reanalysis of colonies 5 and 6. A Parental colonies 5 and 6 were analyzed by isolating plasmids and subsequent transformation into *E. coli.* From each transformation, four individual colonies were chosen for plasmid isolation and restriction analysis with *Pst*I and *Sac*I. Below the gel picture, a schematic overview of the recombination events based on restriction pattern and sequencing results (indicated by 's') is shown. Numbers refer to the lanes on the gel picture. Green diamonds indicate *loxPsym* sites. **B** Parental colonies 5 and 6 were streaked out to obtain single colonies. **C** White and yellow sibling colonies from parental colony 5 (B, left panel) were analyzed by plasmid preparation and restriction digestion with *Pst*I and *Sac*I. Overview of recombination events is given as in **A**.

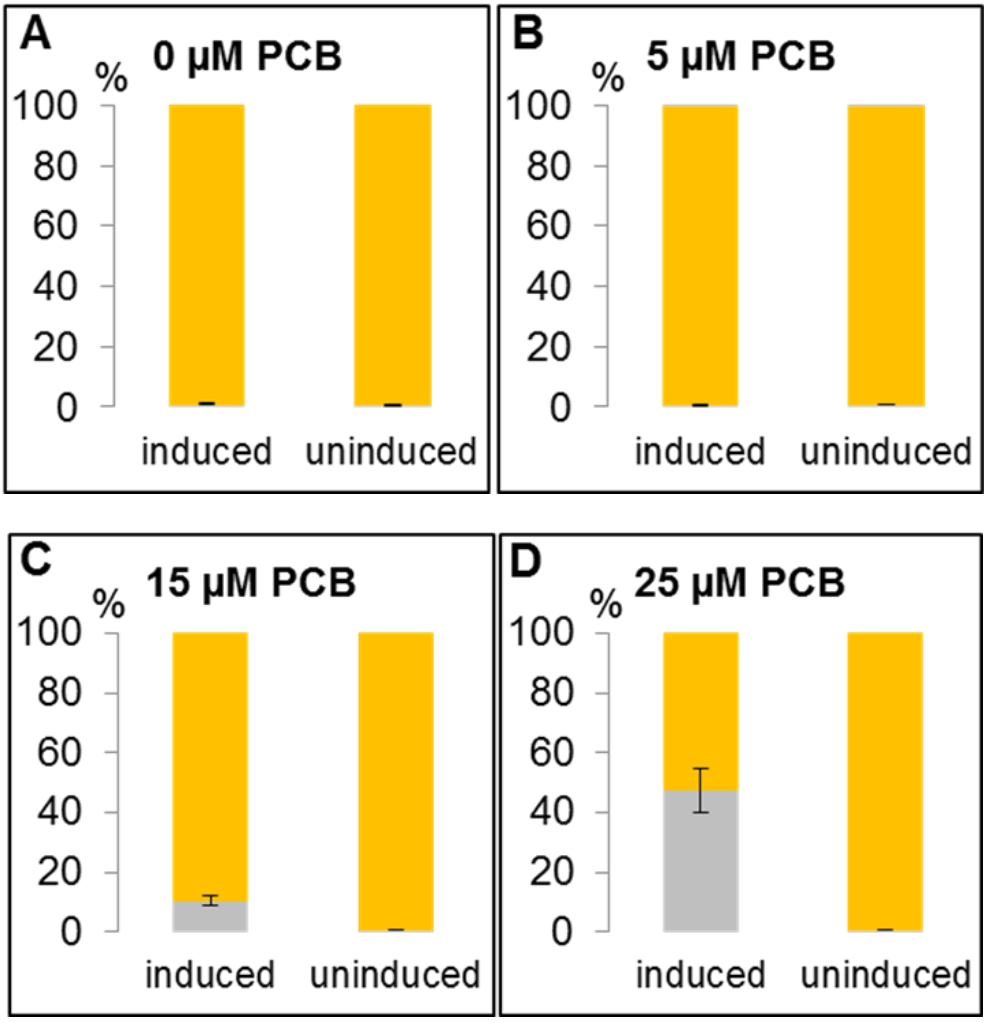


Supplementary Figure 3: Investigations of basal activity of EST-Cre. A glycerol stock of yeast strain EST-Cre was streaked out on SD-Ura-His media. Overnight cultures were inoculated from four individual yellow colonies and grown for 6 h and 21 h, respectively. Cells (1 ml) were pelleted, resuspended in 1 ml ddH_2O, diluted 1:10 in ddH_2O, and 100 µl were plated on SD-Ura and SD-Ura-His plates.



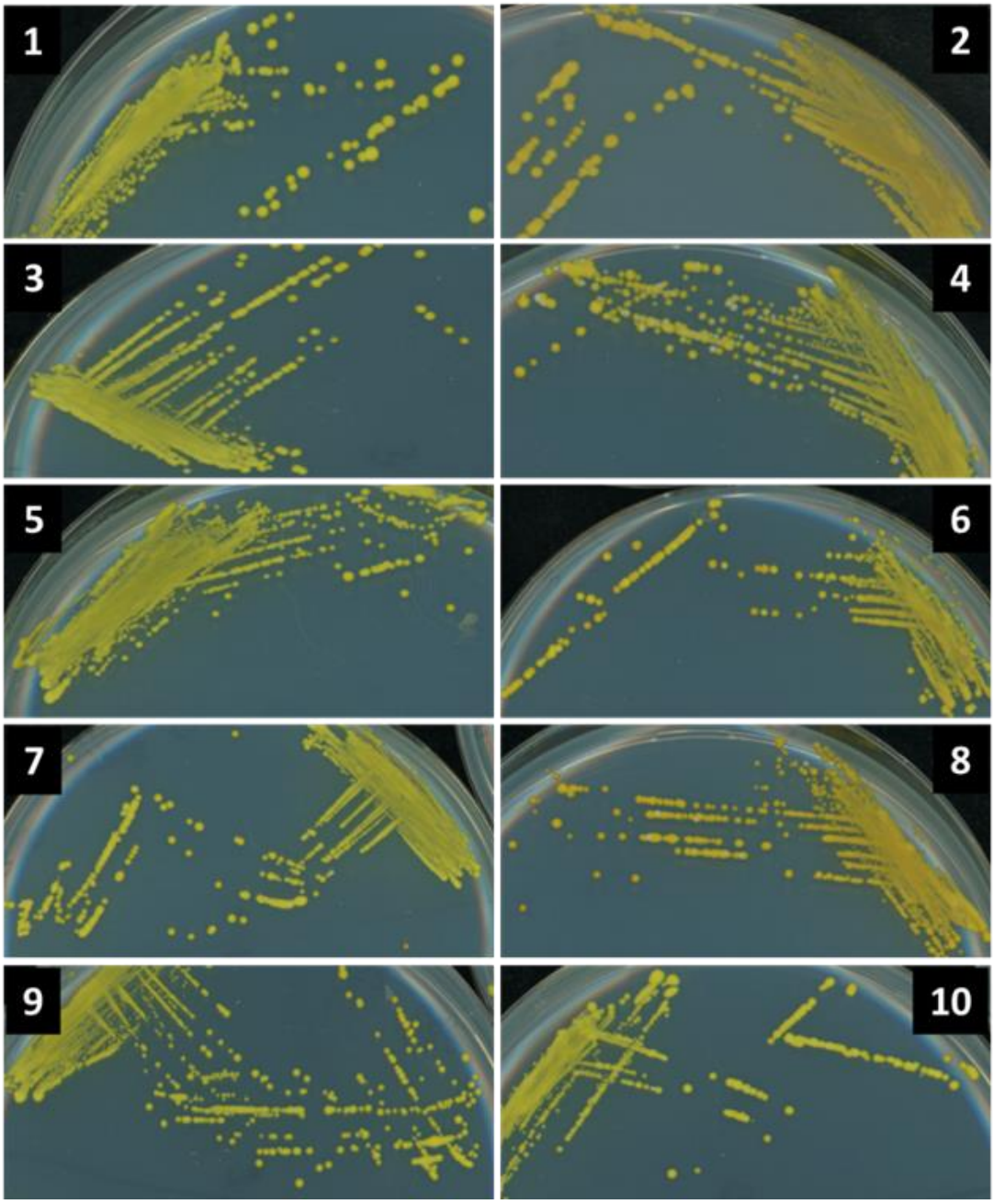
	uninduced		induced	
	white	yellow	white	yellow
0 µM PCB	8/11/7 0,4%/0,7%/0,5%	2232/1660/1456	7/20/12 0,4%/1%/1%	1796/1408/1136
5 µM PCB	14/13/13 0,7%/0,7%/0,7%	1996/1900/1656	9/6/13 0,6%/0,4%/1%	1620/1444/1288
15 µM PCB	11/15/6 0,8%/0,8%/0,3%	1412/1932/1440	244/184/120 12%/11%/8%	1776/1508/1326
25 µM PCB	8/18/7 0,6%/1%/0,5%	1276/1788/1496	696/584/1072 45%/40%/57%	844/896/808

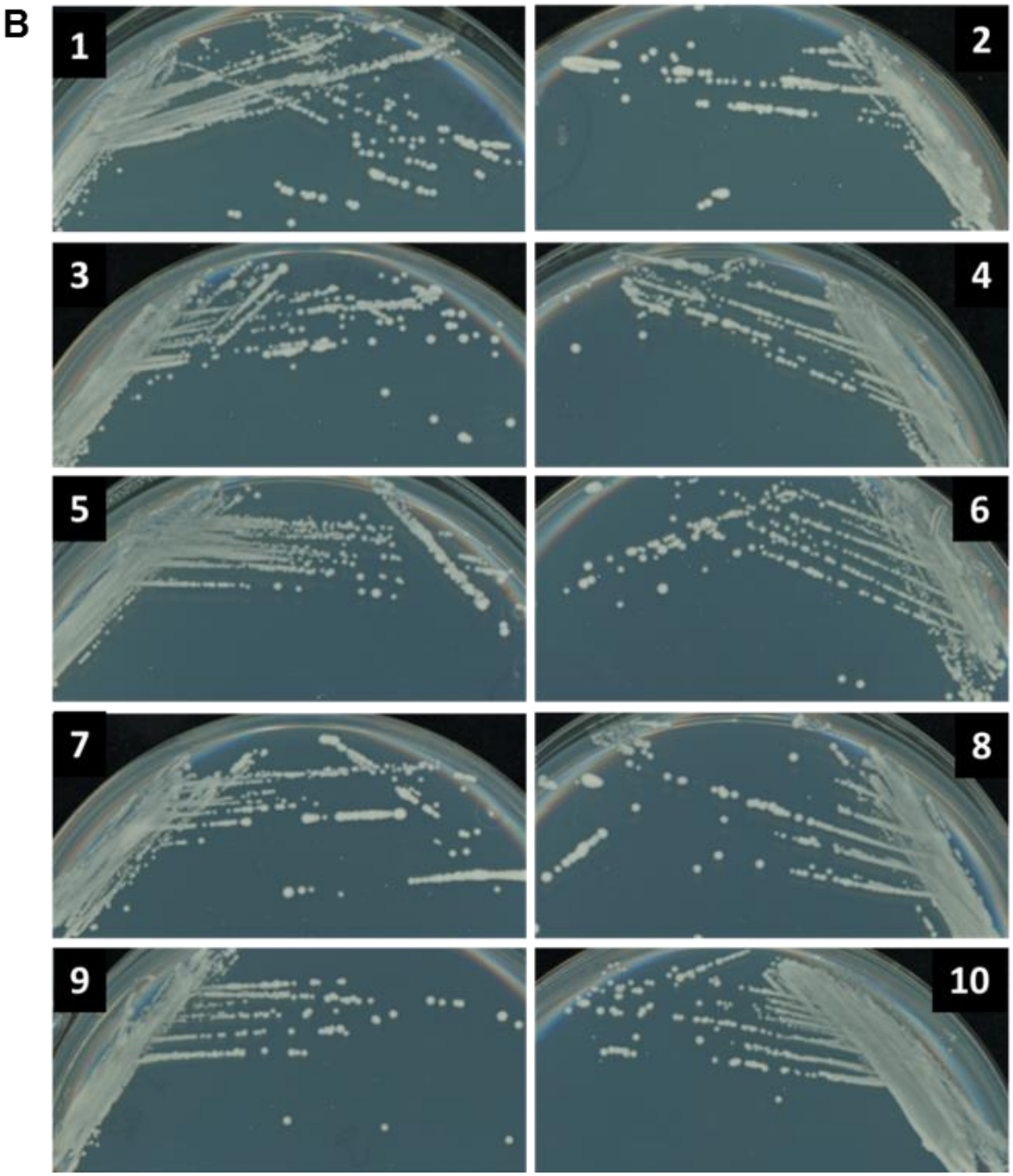
Supplementary Figure 4: Recombination experiment with varying PCB concentrations. Light-Cre1 was grown in SD-Ura-Leu medium with 0, 5, 15 or 25 µM PCB for 6 h in darkness after a far-red light pulse of 1 min. Thereafter, cells were induced with a 5-min red light pulse and grown for further 16 h with red light pulses of 10 sec, applied every 5 min at 30°C and 230 rpm. Uninduced cells were grown in the dark for 16 h. Hundred µl of each culture was pelleted, diluted, plated on appropriate SD medium, and grown for another 2 d. Each experiment was performed in three independent replicates. Plates show the results of one representative replicate with each PCB concentration. The table below shows values for the counted colonies (first row) and the percentage of white colonies (second row) for all three replicates.

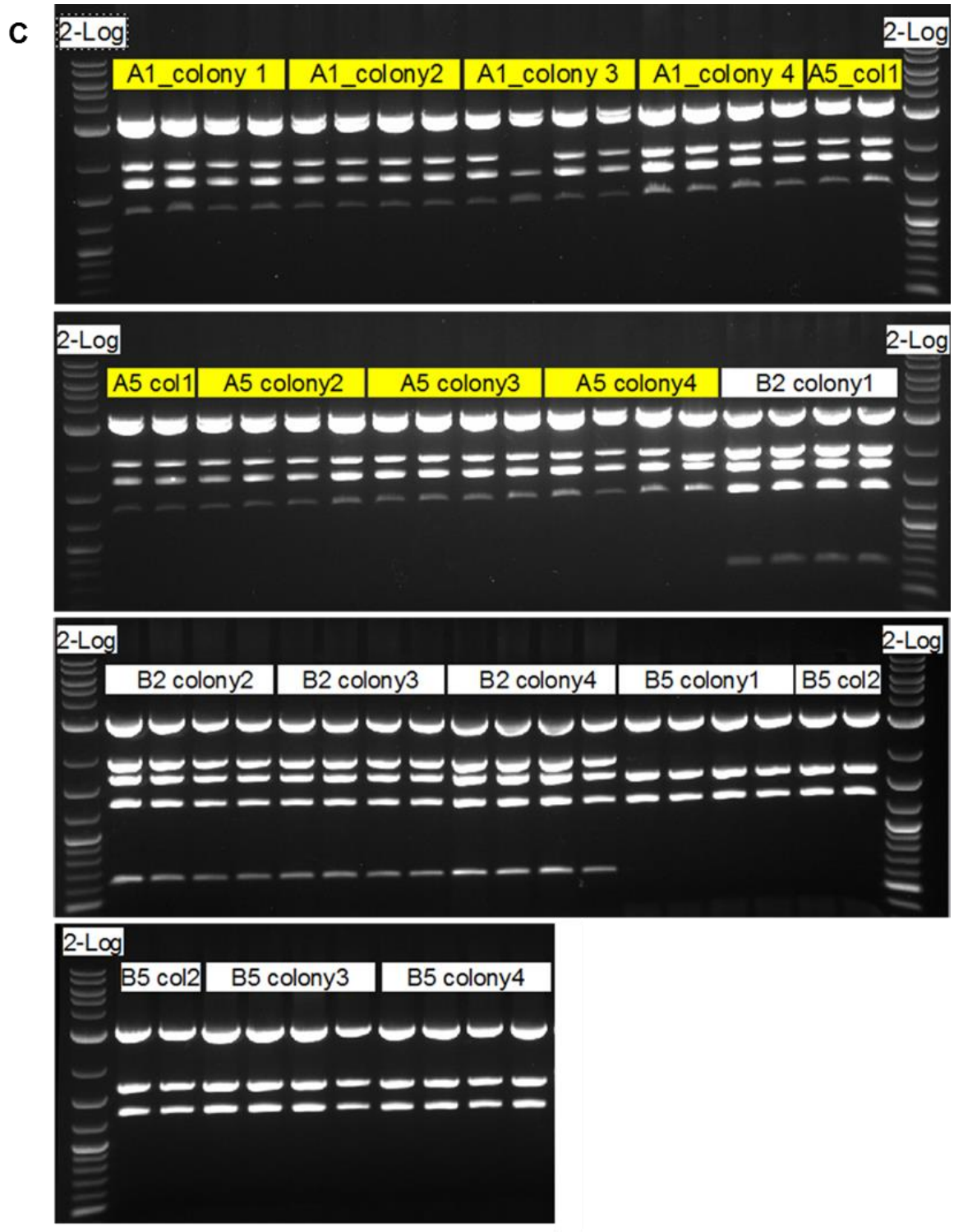


Supplementary Figure 5: L-SCRaMbLE-mediated recombination at varying PCB concentrations. Light-Cre1 cells were grown in SD-Ura-Leu media with **A** 0, **B** 5, **C** 15 and **D** 25 µM PCB for 6 h in darkness. Induced cells were treated with a 5-min red light pulse, followed by 16 h growth with 10-sec red light pulses every 5 min. Uninduced samples were grown for 16 h in darkness. Hundred µl of each culture were pelleted, diluted and plated on SD-Ura-Leu. Bar blots show the means of the percentages of white and yellow colonies in induced and uninduced samples. Grey, white colonies; yellow, yellow colonies. Data are the means of three independent replications ± SD.

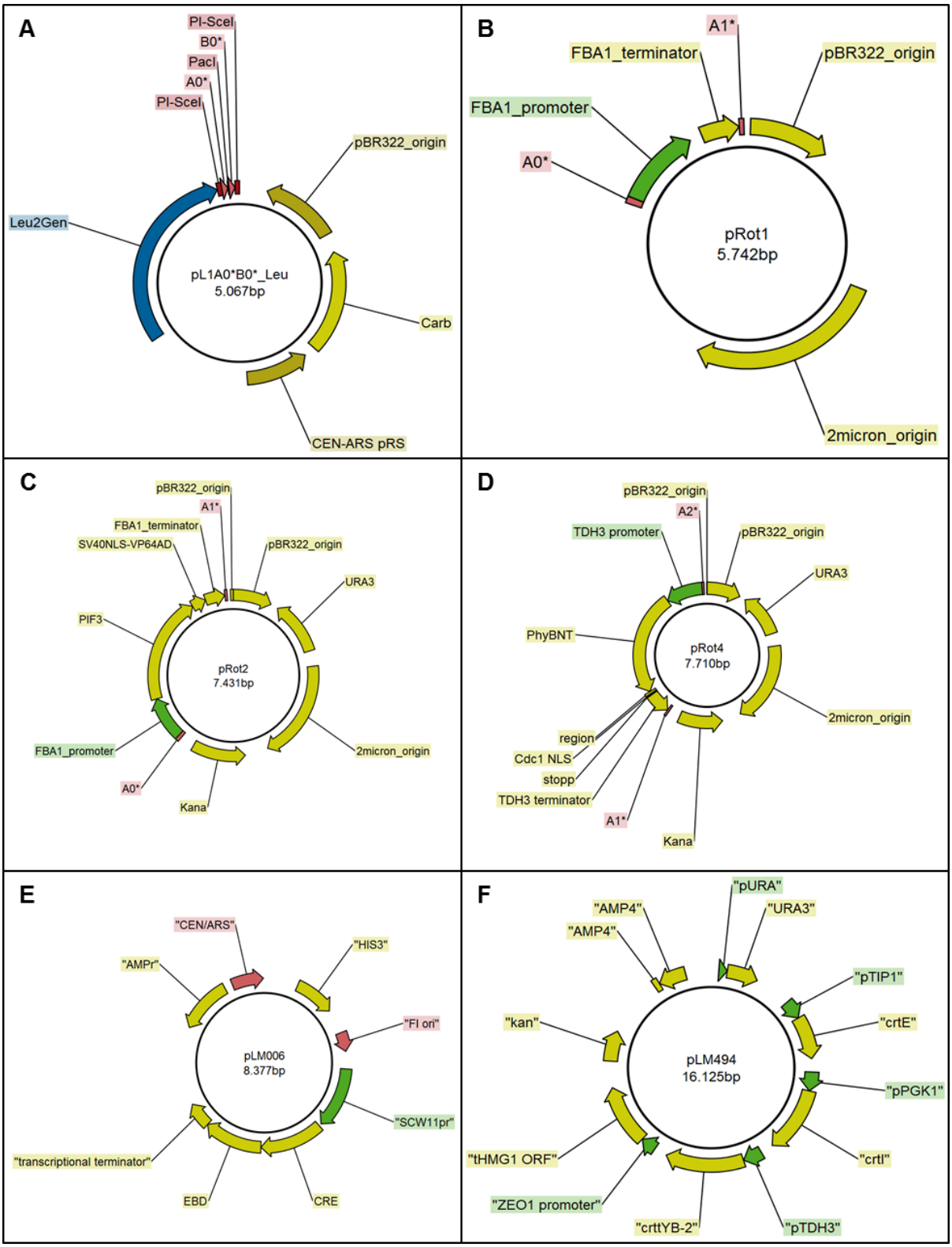
A



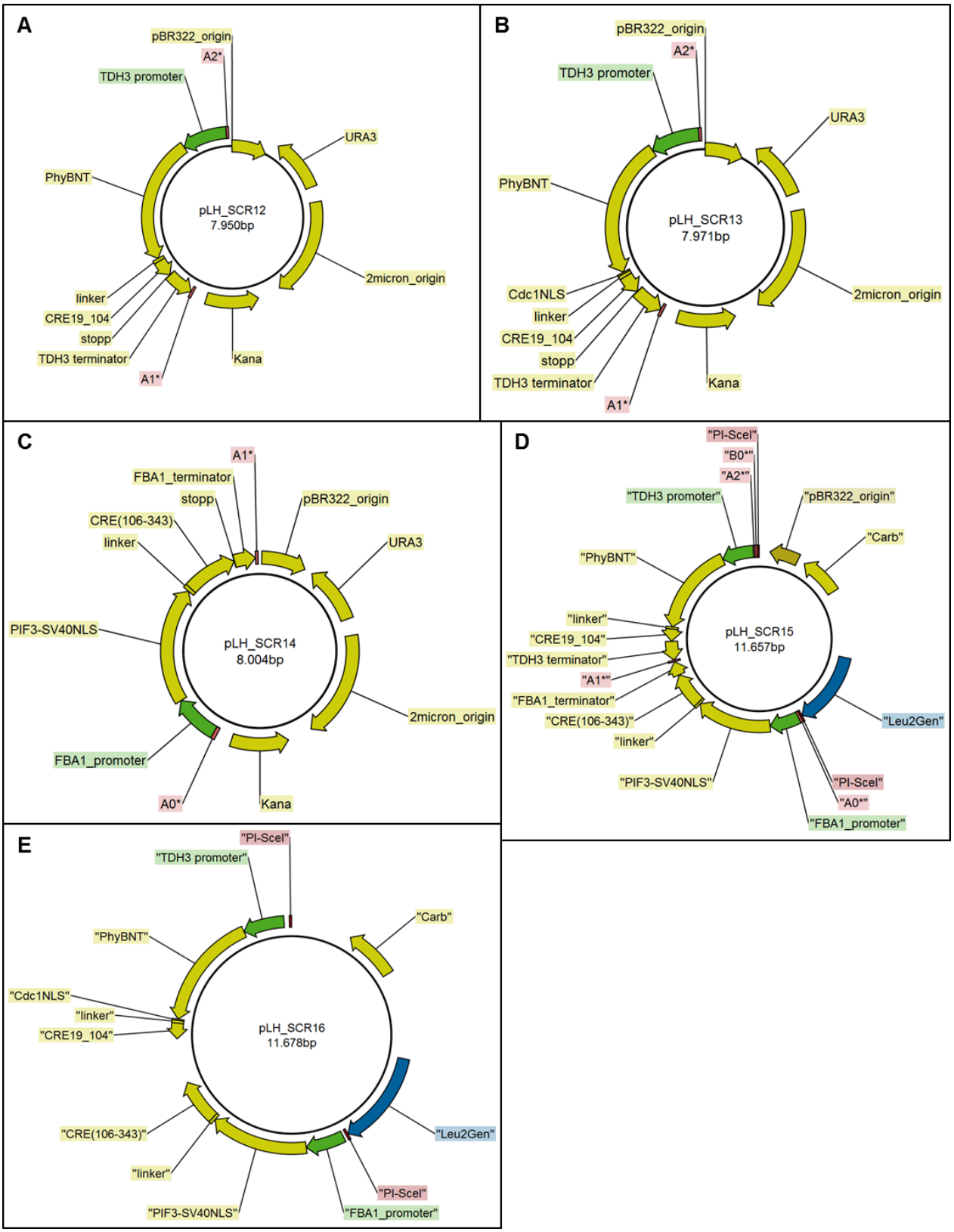

B
1
2
3
4
5
6
7
8
9
10



Supplementary Figure 6: L-SCRaMbLE can be inactivated by withdrawal of PCB. Sibling colonies from yellow (**A**) and white (**B**) colonies on PCB-free medium develop the same color as parental colonies. Parental colonies resulted from a red light-induced recombination experiment (25 µM PCB, 16 h growth under inducing conditions). **C** For A1, A5, B2 and B5, plasmids from four colonies each were isolated and transformed into *E. coli*. From each transformation, four independent *E. coli* colonies were analyzed by plasmid isolation and subsequent digestion with *Pst*I and *Sac*I.



Supplementary Figure 7: Maps of plasmids pL1A0*B0*_Leu, pRot1, pRot2, pRot4, pLM006, and pLM494. Maps are shown for **A** Level 1 backbone pL1A0*B0*_Leu, **B** pRot1, **C** pRot2, **D** pRot4, **E** plasmid pLM006 for β-estradiol inducible Cre and **F** plasmid pLM494 harboring β-carotene biosynthesis genes, flanked with *loxPsym* sites.



Supplementary Figure 8: Maps of plasmids pLH_Scr12 to pLH_Scr16. Maps are shown for Level 0 constructs **A** pLH_Scr12, **B** pLH_Scr13, **C** pLH_Scr14, and Level 1 constructs **D** pLH_Scr15 and **E** pLH_Scr16.

Supplementary Table 1: List of parts with source and sequence modifications.

Assembly components	***Source***	***Modifications***	***Function***
PhyBNT	*A. thaliana*	AA 1-621 of PhyB/ nucleotide 576 C-->A	Photoreceptor
PIF3	*A. thaliana*		Interacting factor
CreC	pLM006	AA 106-343 of Cre / AA 162 V-->R / AA 170E-->K / AA 231 V-->R	C-terminal part of Cre recombinase
CreN	pLM006	AA 19-104 of Cre	N-terminal part of Cre recombinase
Linker	pmCherry-CRY2-CreN		Linker
SV40-NLS	Oligo synthesis		Nuclear localization signal
Cdc1-NLS	Oligo synthesis		Nuclear localization signal
FBA1 promoter	*S. cerevisiae*		Constitutive promoter
TDH3 promoter	*S. cerevisiae*		Constitutive promoter
FBA1 terminator	*S. cerevisiae*		Terminator
TDH3 terminator	*S. cerevisiae*		Terminator

Supplementary Table 2: List of primers used for cloning.

Name	*Sequence (5´-3´)*
L280	TCATATATAACCATAACCAAGTAATACATATTCAAAATGCCTCTGTTTGAGCTTTTC
L706	TAGATTCTGTGAAACTCTCAATAAAGGGACTTCGTCTAAGGTTTCAGTCACACTACGGATACT TTTACAACGGGAGCAGTTATTCA
L707	TGAATAACTGCTCCCGTTGTAAAAGTATCCGTAGTGTGACTGAAACCTTAGACGAAGTCCCTT TATTGAGAGTTTCACAGAATCTA
L712	CCAAGAAGAAGAGAAAGGTGGGTGGCGGTGGCTCTGGAGGTGGTGGGTCCGGAGGAGGCG GCCGCCGACCAAGTGACAGCAATGCTGTTTCACTGG
L714	TCCAGCTTATTCTGAGAGACTCTTTTAAAGAATCTGGTGGCGGTGGCTCTGGAGGTGGTGGG TCCGGAGGAGGCGGCCGCACGAGTGATGAGGTTCGCAAG
L773	GAGCCACCGCCACcCACCTTTCTCTTCTTCTTGGGGCTCG
L774	CTTAGCGCCGTAAATCAGTCGATGAGTTGCTTCAAAAATCCCTTCC
L775	ACTCATCGACTGATTTACGGCGCTAAGGATGACTCTGGTCAGAGG
L776	AAAAAACTATATCAATTAATTTGAATTAACCTAATCGCCATCTTCCAGCAGGCGCACCATTGCC CCTGTTTC
L777	TTATTTAAATGCAAGATTTAAAGTAAATTCACCTACAGCCCGGACCGACGATG
L778	CAGCTTATTCTGAGAGACTCTTTTAAAGAATCTAGGCCATTAAAGCGTAAGAAAGGTGGCGGT GGCTCTGGAGGTGGTGGGTCCGGAGGAGGCGGCCGCACGAGTGATGAGGTTCGCAAG

5 DISCUSSION

With the rapidly increasing interest in synthetic biology, the demand for assembly strategies, supporting the construction of complex biosynthetic pathways, increased gravely. To meet the growing requirements, several highly efficient multi-part assembly strategies have been reported in recent years[159-161]. However, most of these methods are based on restriction enzymes, necessitating laborious parts domestications to delete or mutate spurious cleavage sites before usage. Furthermore, the established strategies are often specialized on the assembly of either single TUs or multiple genes, prompting researchers to combine different strategies.

The previously reported one-step generation of the whole *M. genitalium* genome from 25 fragments via TAR cloning[49, 162] may suggest that the development of multi-part assembly strategies is superfluous, at least for yeast. However, complex one-step assemblies demand detailed and time-consuming planning of individual assembly junctions. A very prominent recent example for a complicated assembly of 23 genes, the opioid biosynthesis in yeast, was carried out through a combination of different strategies, like TOPO cloning, Gateway cloning, Gibson assembly and TAR[163]. It can just be assumed that the decision for this elaborate design results from the lack of an appropriate, universal assembly strategy.

To fill this gap, we designed a flexible and robust cloning strategy, dubbed AssemblX[164], which allows the assembly of currently up to 25 parts, e.g. TUs, each consisting of several sub-units. The maximum number of assembly parts can be easily extended and is only limited by the capacity of the final host to uptake large constructs and to some extent by the availability of yeast selection markers. To ensure a fast and straightforward cloning procedure, AssemblX is based on three successive cloning levels, which enable parallel Level 0 and Level 1 assemblies. Furthermore, AssemblX is not restricted to a certain host as the whole final assembly cassette can easily be subcloned into any expression vector.

The assembly of functional units, usually TUs, poses the most challenging step in multi-gene assembly strategies because it requires a high degree of flexibility. On this account, different design principles were defined for AssemblX Level 0 assemblies.

First, the number and lengths of sub-units should not be limited. Many TU assembly strategies follow strict assembly syntax and are therefore less flexible regarding design of expression cassettes[57, 165-167]. A further precondition for AssemblX was the scar-free assembly of sub-gene fragments, because left-over DNA sequences at the junctions of assembly parts change the genetic context. For example, scar sequences between a promoter and a CDS or at the junction of a fusion protein may affect translational initiation and concomitant gene expression in unpredictable ways[168]. To reduce this risk, some strategies succeed in minimizing left-over scars between assembly junctions by using Type IIS endonucleases and carefully designed cutting sites[57]. A third criterion, which was addressed with AssemblX, was to allow the user to start from scratch without previous domestication or subcloning of parts. Beside the design principles defined in this work, a cloning process for use in synthetic biology should ideally follow a predefined and standardized workflow, with a minimum effort on planning or designing and the option to reuse and share individual assembly parts. However, standardization is not easily compatible with the AssemblX design principles, described here.

As mentioned above, most standardized approaches for establishing predefined TU assemblies were realized through cloning via restriction enzymes and ligases[35, 37, 165, 166]. This supports the establishment of libraries of e.g. promoter, gene and terminator sequences and in this way enables reusing or sharing of single parts and simplifies combinatorial assemblies[37, 165]. Consequently, such strategies are well suited for pathway engineering. However, the disadvantages of endonuclease-based cloning methods, for example the limitation in part number, the need for parts domestications and the presence of scar-sequences at assembly junctions, remain. Pairwise selection assembly (PSA)[169] and MASTER[170] use methylation-sensitive enzymes to circumvent the need of parts domestications, at least for PCR products. Other methods virtually eliminate the need to mutate recognition sites in assembly parts by incorporating rare-cutting homing endonucleases[38, 171], but trade this advantage for long scar sequences. Additionally, the limited number of these meganucleases restricts the complexity of multi-part assemblies. These aspects conflict with the requirements for AssemblX Level 0 assemblies. Therefore, endonuclease-based cloning methods were not included in the AssemblX strategy. Overlap-based assembly strategies like MODAL and BASIC allow

standardization by fusing prefix and suffix linkers to each assembly part which are further used to add overlap regions, determining the assembly order[55, 172]. This results in large scar sequences and again a limitation in fragment number, restricting their applicability for TU assemblies. Even though some methods involve computationally designed linker regions to minimize effects on gene expression, no universal prediction for inter-part compatibility can be made[55, 172, 173]. Moreover, linkers partly possessing the same sequence imply the risk for unwanted recombination events, especially in yeast.

Balancing the advantages and disadvantages of standardized and non-standardized overlap-based assembly strategies for use in AssemblX resulted in the combination of individually designed sub-gene Level 0 and standardized, linker-based Level1/Level2 assemblies. To simplify the planning procedure of multi-part constructs, AssemblX is accompanied by a user-friendly webtool (http://assemblx.org). The webtool minimizes the time required for experimental planning and reduces human error, especially for individual and scar-free Level 0 assemblies. The usage of overlap-based cloning methods in combination with rare cutting (homing) endonucleases ensures a domestication-free assembly process. Furthermore, it eliminates the requirement for further PCR amplification steps after Level 0 assemblies, as each fragment can be released from the vector backbone via the appropriate digestion. The elimination of PCR amplification steps for Level 1 and Level 2 assemblies is an important feature of AssemblX. Despite the risk of PCR-born errors, PCR-amplification of large Level 1 modules covering up to 5 genes is challenging. Moreover, sequencing of assembly junctions is sufficient and reduces expenses of AssemblX projects.

AssemblX builds on the fact that most overlap-based cloning methods, like e.g. SLiCE, Gibson assembly and TAR[46, 48, 49], share the same requirements for parts preparation and differ only slightly in the required overlap lengths. The lengths of standardized AssemblX homology regions directing Level 1 and 2 assemblies were defined to be compatible with a wide range of overlap-based cloning methods. This gives the user the freedom of choice according to personal preferences or experimental constraints. For example, a lab without yeast experience can carry out the whole assembly process in *E. coli* by using SLiCE[48], Gibson assembly[46] or any other suitable overlap-based cloning method. Two advantages of SLiCE and TAR cloning are the nearly gratuitous set up and

the tolerance for non-homologous sequences present at the ends of assembly pieces (e.g. overhangs from restriction digestion). To further support and simplify AssemblX cloning steps, the protocol for SLiCE cloning was improved in our laboratory. The cloning efficiencies were carefully tested and characterized with regard to lengths and content of homology regions, influence of parts preparations and tolerance for heterologous vector sequences[174].

AssemblX focuses on a user-friendly handling and high reliability, even for complex assemblies. This is achieved by a range of useful tweaks and helpful support features like *ccdB* counter selection[175], optimized overlap regions[55, 56], software assisted planning and a collection of ready-to-use promoter-terminator combinations for gene expression in yeast. AssemblX realizes a unique degree of flexibility through the unrestricted design of functional units, the possibility of choosing the preferred assembly method and assembly host, as well as the free choice of the final target organism. The level-based strategy supports fast parallel assemblies. PCR-free and standardized Level 1 and Level 2 assemblies allow reliable and efficient multi-gene assemblies. The novelty of AssemblX is not an optimization or modification of already functional basic cloning methods, but the elaborate combination of these different methods to facilitate and optimize cloning workflow and further expression of multi-gene constructs. It can be assumed that using AssemblX will help to reduce the amount of time and money required to realize the heterologous expression of very complex pathways, like the one for opioid biosynthesis in yeast, mentioned above[163]. However, there is no doubt that AssemblX can be further developed and optimized regarding standardization, reusability of parts and the complexity of combinatorial Level 0 assemblies. A cloning method supporting these aspects, along with a domestication- and scar-free assembly approach, is currently not available but would be highly valuable. While the possibility for combinatorial cloning of AssemblX Level 0 units and the reusability of sub-gene fragments is limited, Level 0 and Level 1 modules can be flexibly combined and reused in different projects. Two additional and easier to implement optimization possibilities are the removal of the large yeast selection markers in between the Level 1 modules to support approaches in host organisms different from *S. cerevisiae* and an integration of multi-gene assembly cassettes into the yeast genome to allow robust gene expression.

The elimination of yeast selection markers would also support the extension of the maximal assembly part number of AssemblX, currently limited by the availability of auxotrophic markers in yeast. The insertion of an additional homology region (B0-F0), followed by a second type of homing endonuclease restriction site upstream of the marker cassette, would enable the user to remove yeast markers after positive selection of successful Level 1 assemblies. This would require a simple modification of Level 0 backbones carrying homology regions AR-ER. Genome integration could be easily implemented by using a CRISPR-Cas9-based approach, allowing the simultaneous release of the assembly module and integration into a specified genomic locus.

The constitution of complex DNA constructs covers only the first step towards successful heterologous production of relevant substances. A further necessary approach is the regulation and balancing of gene expression levels in a biosynthetic pathway, to optimize product output. Optogenetic interfaces were reported to achieve robust and tight regulation of many cellular processes like gene expression[124, 125, 176] or recombinase activity[127, 128, 153, 177]. Many of these systems rely on plant-derived optical dimerizers, which have been shown to enable highly efficient, fast and reversible regulation in diverse host organisms[130, 152, 178-181]. Compared to chemical inducers, light induction is beneficial regarding handling, toxicity, cost effort, dose-dependency and reversibility. Chemical inducers cannot be easily and fully removed from growth media and trace amounts, which may cause toxic or unpredictable effects to the host organism, remain. Even though blue and green light were reported to have negative effects on yeast metabolism at elevated concentrations, cell growth quickly recovers post-illumination[182]. Red light does not severely affect yeast metabolism[182]. A further advantage of the red/far-red light sensing phytochromes entail the fast dimerization upon a single red light pulse, which persists over several hours in darkness. Importantly, this interaction can be rapidly reversed via a short far-red light pulse[148]. In comparison, the blue light-sensing cryptochrome CRY2 from *A. thaliana* reverses within minutes in the dark[153]. This is a disadvantage for technical applications, as activation of gene expression would require high light intensities and nearly continuous illumination. Of note, the aforementioned high levels of constantly applied blue light were reported to cause cellular stress responses[183]. However, in contrast to the blue light sensitive

cryptochromes, phytochromes require the supply of the expensive and highly photo-labile chromophore PCB, limiting their use for biotechnology applications.

In this work, two light-regulated induction systems were established in yeast with the help of AssemblX: PhiReX, a gene expression system for pathway engineering based on a split synTF and L-SCRaMbLE, a recombination system based on a split Cre recombinase for application in the SCRaMbLE system[94]. The optogenetic interface, which is part of both tools, is based on the N-terminal domain of PhyB (PhyBNT) and its interacting protein PIF3[152, 178].

The light-regulated gene expression system was intended to allow targeting of heterologous, as well as endogenous promoters. Therefore, a programmable synTALE-based DBD was incorporated in the PhiReX system. Highly efficient gene-expression systems have been reported in yeast before, however they are based on invariant DBDs, like the GAL4-DBD and do not support programmable targeting of specific DNA sequences[152, 153, 178, 179]. For mammalian cells, several blue light-regulated gene expression systems, based on customizable DBDs like zinc finger proteins (ZFPs), synTALEs or CRISPR/Cas9 have already been presented[126, 154, 155, 184]. Especially the latter two generated very promising results[126, 155]. Even if the CRISPR/Cas9 strategy facilitates the targeting of various DNA sequences, as the DNA specificity of the dCas9 protein is only determined by a simple-to-clone sgRNA, synTALEs have been reported to allow more efficient gene regulation than dCas9[185]. In the study presented here, a plasmid located combination of the PhyB-synTALE-DBD-fusion, the PIF3-AD-fusion and the reporter cassette was demonstrated to allow PCB-dependent activation of gene expression. Light-mediated gene activation was independent of whether the synTALE-DBD was fused to the N- (pRL_NT) or C-terminal domain of the photoreceptor (pRL_CT). Interestingly, only pRL_CT shows red light-dependent activation of the reporter gene even in the absence of the chromophore PCB. Red light-mediated gene expression, independent of PCB supply, was previously reported in *S. cerevisiae* for the plant-derived photoreceptor PhyA and its interacting partner FHY1[179]. As a reason for this, Sorokina *et al.*[179] supposed a functional chromophore-like substance, present in yeast. This leads to the assumption that pRL_CT is more sensitive than pRL_NT, most likely due to the different architectures of the fusion proteins.

Even though the plasmid-born system shows light- and PCB-dependency, it was optimized in several points to achieve tighter regulation, lower basal activity, and high and robust expression levels. Genome integration of expression cassettes is a common approach to yield more robust and efficient expression[186, 187], while the overall expression level is usually higher on episomal plasmids[165]. Accordingly, a combination of a genome-located optical dimerizer and a reporter expression cassette, located on a high-copy plasmid, boosted expression levels and yielded robust light-dependent yEGFP induction in usually more than 80% of cells. Another possibility to raise the expression output, not executed in this study, would be to test for a stronger AD or an additional PhyBNT region fused to the synTALE-DBD in order to localize another AD near the promoter, comparable to previously reported approaches[155]. However, despite enhanced expression levels in induced state, the genome-integrated system also exhibits very high basal expression levels, mainly caused by the synTALE-DBD. Similar observations were reported before for the blue light-sensing, synTALE-based LITE expression system[154]. Konermann *et al*.[154] solved the problem by deleting inherent NLS regions in the CIB1 protein, fused to the synTALE-DBD and adding a second NLS at the CRY2-AD-fusion, in order to prevent nuclear import before light-mediated dimerization. In the study presented here, the synTALE-DBD is fused to an N-terminal version of PhyB, which was reported to be free of internal NLS sequences[188]. Pfeiffer *et al.* (2012) showed that neither the full-length version nor a C-terminal or an N-terminal version of PhyB is targeted to the nucleus in a cell-free system[189]. The protein was translocated to the nucleus, only after dimerization of PhyB with PIF3. Beyer *et al*. (2015) confirmed these results by demonstrating that PhyBNT (aa 1-908) is located in the cytoplasm in mammalian cells[130]. However, NLS sequences inherent to the synTALE-backbone have been characterized previously and different combinations of NLS deletions were reported to be exclusively located in the cytoplasm[190]. Here, in this thesis, two promising systems were established, namely PhiReX 1.0 and PhiReX 1.1, by deleting all potential NLS sequences and the insertion of a nuclear export signal (NES) in the PhyB-synTALE-DBD-fusion. The two systems mainly differ in two points: while PhiReX 1.0 mediates high expression levels in induced state (11-fold induction) and retains some background activity in uninducing conditions, PhiReX 1.1 shows nearly no basal activity and expression levels comparable to the medium strong yeast *ADH1* promoter (41-fold

induction). The performance of PhiReX 1.1 with respect to fold induction is superior compared to the blue light-sensing synTALE-based LITE 2.0 system in mammalian cells[154].

To boost the potential of PhiReX for biotechnological applications, yeast was for the first time successfully engineered to produce the required chromophore. Different combinations of native and modified heme oxygenases (HOs) and bilin reductases (BRs) tested with PhiReX can support red light-induced gene expression in the absence of exogenous PCB to various extends, thereby obviating the need to apply chromophore to the culture medium.

Summing up, it was demonstrated that PhiReX allows dose-dependent and reversible activation of gene expression without any chromophore supply. It achieves robust expression for at least 48 h and expression levels comparable to that of the strong constitutive yeast *TDH3* promoter. PhiReX users only need to handle a relatively small plasmid containing the target promoter upstream of the GOI (~ 5 kb without reporter). With the possibility to flexibly exchange the CDS by a single cloning step the system can be easily adapted to the customer's demands. Cloning of a different TALE-DBD would even allow the light-dependent control of endogenous promoters. Replacing the AD with a nuclease is also straightforward using the AssemblX method[164] and would result in a light-regulated tool, e.g. for genome editing.

Another possibility for genome editing is the "SCRaMbLE" system, which is designed to induce randomized recombination events within synthetic chromosomes generated in the Synthetic Yeast Sc2.0 Project[28, 31-34]. Each synthetic Sc2.0 chromosome incorporates *loxPsym* sites downstream of every non-essential gene and at specific landmarks. The functionality of SCRaMbLE was confirmed by the generation of diverse genome rearrangements in yeast strains carrying one or two synthetic chromosomes[29, 95, 96]. The light-regulated L-SCRaMbLE system presented here provides an alternative to the recombinase induction system used in the original SCRaMbLE, which relies on the toxic hormone β-estradiol[29, 95]. The red light-inducible Cre recombinase was realized by fusing a previously described split Cre to the C-termini of PhyBNT or PIF3, respectively[127, 153]. Two different configurations of the fusion proteins were tested and

compared to β-estradiol induced Cre recombinase with respect to recombination rates. A plasmid borne β-carotene pathway with four *loxPsym* tagged genes, which results in yellow colonies, was employed as test system[58, 186]. Gene deletions, mediated by the recombinase, can be readily observed based on the colony color. As a measure for the recombination rate, the ratio of white to yellow colonies was determined. It should be noted that recombination mediated inversions, translocations, duplications, as well as the deletion of *tHMG1* cannot be detected with this method[58, 186]. Furthermore, the real recombination rate is most likely higher than the observed rate, because the pathway encoding plasmid occurs in 2-5 copies per cell[191]. Thus, colonies appear yellow, if one copy of intact plasmid, or a combination of plasmids expressing all four biosynthesis genes, is available in the cell. However, the ratio of yellow to white cells can be used as a first measure for recombination frequency. As expected, the recombination efficiency of L-SCRaMbLE was shown to increase with prolonged incubation time after light induction with a maximum of 52% white colonies for Light-Cre1 after 24 h. Remarkably, background levels in uninduced samples are very low, especially compared to β-estradiol induced Cre recombinase, which exhibits roughly 22-fold higher recombination in uninduced state. This is rather surprising, because leaky expression has so far not been reported to be an issue of the original SCRaMbLE system[95]. A possible explanation for this may be higher sensitivity of the plasmid-based recombination monitoring used in this study, in contrast to *loxPsym* site scattered throughout a complete chromosome in an actual SCRaMbLE experiment. Light-Cre2 consistently shows slightly higher recombination frequencies than Light-Cre1. This is most likely due to the NLS, incorporated in the PhyBNT-CreN-fusion of Light-Cre2, but not of Light-Cre1. Neither PhyBNT nor CreN contain NLS sequences and nuclear diffusion is unlikely for an 80-kDa protein[189, 192]. Thus, prior to induction the two components of the split Cre in Light-Cre1 are located in the cytosol and the nucleus, respectively. This is most likely the reason for the lower basal recombination rate.

Analysis of recombination events at a molecular level confirmed the notion that Light-Cre1 confers a substantial lower background than EST-Cre. The sequence analyses revealed that the background of EST-Cre is even higher than estimated on the base of colony color. Remarkably, Light-Cre1 produces a more diverse set of recombination

events than the EST-Cre system, which results only in a single type of recombined plasmid (deletion of all four CDS). This is an important advantage of L-SCRaMbLE, as too many simultaneous recombination events within the Sc2.0 yeast genome would cause harsh deletions or rearrangements of the genome, most likely resulting in cell death. Furthermore, a variety of different recombination events is advantageous for synthetic evolution studies.

Another important requirement for the synthetic yeast project is the isogenic character of colonies resulting from a defined recombination experiment, allowing investigations of stable yeast strains. To this end, recombination activity needs to be switched off in a reliable manner, which prevents secondary recombination events. In this study, it was demonstrated that L-SCRaMbLE mediated recombination can be controlled on an additional level besides far-red light irradiation, i.e., by restricting PCB availability. Without sufficient PCB supply the recombination frequencies are below 1%, even under red light conditions, and plasmid integrity is maintained after prolonged sub-culturing. Thus, PCB-free medium is a simple and safe way to stably maintain recombined strains without having to avoid light exposure. In this context, it becomes obvious that an endogenous expression of the PhyB chromophore, comparable to the expression system PhiReX, is not beneficial, as this would exclude this additional fail-safe. Furthermore, while PhiReX is designed to allow expression even at a technical scale, SCRaMbLE is carried out in small volumes and in this way, costs for chromophore supply are negligible.

Taken together, L-SCRaMbLE offers a robust system for tight control of Cre recombinase. It mediates diverse recombination events, very low basal activity and can be easily switched off by plating yeast cells on PCB-free medium irrespective of the light condition. Application of the light-responsive Cre recombinase presented here is not limited to SCRaMbLE-ing yeast chromosomes. It can also be used e.g. to regulate Cre-mediated gene expression[128, 193], to create knockout mutants, or to add, translocate or exchange any *loxP*/*loxPsym* site flanked DNA sequence. Until now, only blue and UV light responsible Cre recombination systems have been reported[128, 177, 194, 195]. In yeast, no light-inducible recombinase has been presented at all.

The tools generated in this study facilitate pathway engineering and genome editing on different levels. First, as a basis of heterologous protein production the multi-gene assembly strategy AssemblX was established, which allows the easy, fast, reliable and (cost-) efficient generation of large DNA constructs, not limited to biosynthetic pathways. By using this strategy, an optogenetic interface was designed, based on a red/far-red light sensing optical dimerizer to support specific, time-dependent and tunable control of split regulation tools. This system was combined with a programmable split transcription factor and applied for the regulation of gene expression. The optimization and further development of this system yielded the PhiReX system. The fusion of the optogenetic interface to a split Cre recombinase resulted in L-SCRaMbLE, promoting host genome reengineering to generate improved strains for heterologous protein production. The generation of this complete toolbox enables even non-specialized labs to assemble large and complex constructs and to easily and precisely control gene expression by an artificial and programmable light-induced transcription factor. The L-SCRaMbLE system adds a huge range of possible applications to this toolbox.

6 REFERENCES

1. Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).
2. Leduc, S., La biologie synthétique. Paris: Poinat, 1912.
3. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500 (2001).
4. Wang, L., Xie, J., Schultz, P.G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**, 225-249 (2006).
5. Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., Schultz, P.G. An expanded eukaryotic genetic code. *Science* **301**, 964-967 (2003).
6. Niederholtmeyer, H., Wolfstädter, B.T., Savage, D.F., Silver, P.A., Way, J.C. Engineering cyanobacteria to synthesize and export hydrophilic products. *Applied and Environmental microbiology* **76**, 3462-3466 (2010).
7. Kim, T., Folcher, M., Baba, M.D.E., Fussenegger, M. A Synthetic Erectile Optogenetic Stimulator Enabling Blue-Light-Inducible Penile Erection. *Angewandte Chemie International Edition* **54**, 5933-5938 (2015).
8. Folcher, M., Oesterle, S., Zwicky, K., Thekkottil, T., Heymoz, J., Hohmann, M., Christen, M., El-Baba, M.D., Buchmann, P., Fussenegger, M. Mind-controlled transgene expression by a wireless-powered optogenetic designer cell implant. *Nature communications* **5** (2014).
9. Kaebnick, G.E., Murray, T.H., Synthetic biology and morality: Artificial life and the bounds of nature, MIT Press, 2013.
10. Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L. The genome project-write. *Science* **353**, 126-127 (2016).
11. Callaway, E. Plan to synthesize human genome triggers mixed response. *Nature* **534**, 163 (2016).
12. Bellí, G., Garí, E., Piedrafita, L., Aldea, M., Herrero, E. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Research* **26**, 942-947 (1998).
13. Elowitz, M.B., Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).
14. Gardner, T.S., Cantor, C.R., Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339-342 (2000).
15. Müller, K., Engesser, R., Metzger, S., Schulz, S., Kämpf, M.M., Busacker, M., Steinberg, T., Tomakidi, P., Ehrbar, M., Nagy, F., Timmer, J., Zubriggen, M.D., Weber, W. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Research* **41**, e77 (2013).
16. Buchler, N.E., Gerland, U., Hwa, T. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences* **100**, 5136-5141 (2003).
17. Silva-Rocha, R., de Lorenzo, V. Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS letters* **582**, 1237-1244 (2008).
18. Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).
19. Engels, B., Dahm, P., Jennewein, S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic engineering* **10**, 201-206 (2008).

20. Runguphan, W., Keasling, J.D. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic engineering* **21**, 103-113 (2014).
21. Brat, D., Weber, C., Lorenzen, W., Bode, H.B., Boles, E. Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast *Saccharomyces cerevisiae*. *Biotechnology for biofuels* **5**, 65 (2012).
22. Buijs, N.A., Siewers, V., Nielsen, J. Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Current opinion in chemical biology* **17**, 480-488 (2013).
23. Bergman, L.W. Growth and maintenance of yeast. *Two-Hybrid Systems: Methods and Protocols*, 9-14 (2001).
24. Alberghina, L., Porro, D., Martegani, E., Ranzi, B.M. Efficient production of recombinant DNA proteins in *Saccharomyces cerevisiae* by controlled high-cell-density fermentation. *Biotechnology and Applied Biochemistry* **14**, 82-92 (1991).
25. Sun, J., Shao, Z., Zhao, H., Nair, N., Wen, F., Xu, J.H., Zhao, H. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering* **109**, 2082-2092 (2012).
26. Struhl, K. Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. *Molecular and Cellular Biology* **6**, 3847-3853 (1986).
27. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G. Life with 6000 genes. *Science* **274**, 546, 563-547 (1996).
28. Mitchell, L.A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J.A., Zhao, Y., Walker, R., Luo, Y., Dai, H., Dong, K., Tang, Z., Yang, Y., Cai, Y., Heguy, A., Ueberheide, B., Fenyö, D., Dai, J., Bader, J.S., Boeke, J.D. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355** (2017).
29. Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J., Lindstrom, D.L., Boeke, A.C., Gottschling, D.E., Chandrasegaran, S., Bader, J.S., Boeke, J.D. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
30. Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).
31. Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., Luo, Y., Müller, C.A., Paul-Dubois-Taine, A., Alver, B., Stracquadanio, G., Mitchell, L.A., Luo, Z., Fan, Y., Zhou, B., Wen, B., Tan, F., Wang, Y., Zi, J., Xie, Z., Li, B., Yang, K., Richardson, S.M., Jiang, H., French, C.E., Nieduszynski, C.A., Koszul, R., Marston, A.L., Yuan, Y., Wang, J., Bader, J.S., Dai, J., Boeke, J.D., Xu, X., Cai, Y., Yang, H. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355** (2017).
32. Wu, Y., Li, B.-Z., Zhao, M., Mitchell, L.A., Xie, Z.-X., Lin, Q.-H., Wang, X., Xiao, W.-H., Wang, Y., Zhou, X., Liu, H., Li, X., Ding, M.-Z., Liu, D., Zhang, L., Liu, B.-L., Wu, X.-L., Li, F.-F., Dong, X.-T., Jia, B., Zhang, W.-Z., Jiang, G.-Z., Liu, Y., Bai, X., Song, T.-Q., Chen, Y., Zhou, S.-J., Zhu, R.-Y., Gao, F., Kuang, Z., Wang, X., Shen, M., Yang, K., Stracquadanio, G., Richardson, S.M., Lin, Y., Wang, L., Walker, R., Luo, Y., Ma, P.-S., Yang, H., Cai, Y., Dai, J., Bader, J.S., Boeke, J.D., Yuan, Y.-J. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355** (2017).
33. Xie, Z.-X., Li, B.-Z., Mitchell, L.A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B.-X., Liu, H.-M., Wu, X.-L., Feng, Q., Zhang, W.-Z., Liu, W., Ding, M.-Z., Li, X., Zhao, G.-R., Qiao, J.-J., Cheng, J.-S., Zhao, M., Kuang, Z., Wang, X., Martin, J.A., Stracquadanio, G., Yang, K., Bai, X., Zhao, J., Hu, M.-L., Lin, Q.-H., Zhang, W.-Q., Shen, M.-H., Chen, S., Su, W., Wang, E.-X., Guo, R., Zhai, F., Guo, X.-J., Du, H.-X., Zhu, J.-Q., Song, T.-Q., Dai, J.-J., Li, F.-F., Jiang, G.-Z., Han, S.-L., Liu, S.-Y., Yu, Z.-C., Yang, X.-N., Chen, K., Hu, C., Li, D.-S., Jia, N., Liu, Y., Wang, L.-T., Wang, S., Wei, X.-T., Fu, M.-Q., Qu, L.-M., Xin,

S.-Y., Liu, T., Tian, K.-R., Li, X.-N., Zhang, J.-H., Song, L.-X., Liu, J.-G., Lv, J.-F., Xu, H., Tao, R., Wang, Y., Zhang, T.-T., Deng, Y.-X., Wang, Y.-R., Li, T., Ye, G.-X., Xu, X.-R., Xia, Z.-B., Zhang, W., Yang, S.-L., Liu, Y.-L., Ding, W.-Q., Liu, Z.-N., Zhu, J.-Q., Liu, N.-Z., Walker, R., Luo, Y., Wang, Y., Shen, Y., Yang, H., Cai, Y., Ma, P.-S., Zhang, C.-T., Bader, J.S., Boeke, J.D., Yuan, Y.-J. "Perfect" designer chromosome V and behavior of a ring derivative. *Science* **355** (2017).

34. Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., Wang, Y., Hou, S., Huang, J., Li, T., Qin, Y., Dong, J., Qin, Q., Zhang, J., Zou, X., He, X., Zhao, L., Xiao, Y., Xu, M., Cheng, E., Huang, N., Zhou, T., Shen, Y., Walker, R., Luo, Y., Kuang, Z., Mitchell, L.A., Yang, K., Richardson, S.M., Wu, Y., Li, B.-Z., Yuan, Y.-J., Yang, H., Lin, J., Chen, G.-Q., Wu, Q., Bader, J.S., Cai, Y., Boeke, J.D., Dai, J. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355** (2017).
35. Knight, T., Idempotent vector design for standard assembly of biobricks, DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group, Massachusetts Institute of Technology. Cambridge, MA., 2003.
36. Engler, C., Kandzia, R., Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**, e3647 (2008).
37. Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PloS one* **6**, e16765 (2011).
38. Liu, J.K., Chen, W.H., Ren, S.X., Zhao, G.P., Wang, J. iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. *PloS one* **9**, e110852 (2014).
39. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., Pâques, F. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Current gene therapy* **11**, 11-27 (2011).
40. Urban, A., Neukirchen, S., Jaeger, K.-E. A rapid and efficient method for site-directed mutagenesis using one-step overlap extension PCR. *Nucleic Acids Research* **25**, 2227-2228 (1997).
41. Bryksin, A.V., Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463 (2010).
42. Quan, J., Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PloS one* **4**, e6441 (2009).
43. Kok, S.d., Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., Platt, D., Shapland, E.B., Serber, Z., Dean, J. Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synthetic Biology* **3**, 97-106 (2014).
44. Aslanidis, C., de Jong, P.J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research* **18**, 6069-6074 (1990).
45. Li, M.Z., Elledge, S.J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature methods* **4**, 251-256 (2007).
46. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).
47. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498**, 349-361 (2011).
48. Zhang, Y., Werling, U., Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research* **40**, e55 (2012).
49. Kouprina, N., Larionov, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics* **7**, 805-812 (2006).
50. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215-1220 (2008).

51. Gibson, D.G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research* (2009).
52. Li, M., Elledge, S. in Gene Synthesis: Methods and Protocols (ed. J. Peccoud) 51-59 (2012).
53. Müller, K.M., Arndt, K.M. Standardization in synthetic biology. *Synthetic Gene Networks: Methods and Protocols*, 23-43 (2012).
54. Porcar, M., Latorre, A., Moya, A. What symbionts teach us about modularity. *Frontiers in bioengineering and biotechnology* **1** (2013).
55. Casini, A., MacDonald, J.T., Jonghe, J.D., Christodoulou, G., Freemont, P.S., Baldwin, G.S., Ellis, T. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Research* **42**, e7-e7 (2014).
56. Casini, A., Christodoulou, G., Freemont, P.S., Baldwin, G.S., Ellis, T., MacDonald, J.T. R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synthetic Biology* **3**, 525-528 (2014).
57. Agmon, N., Mitchell, L.A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., Choi, W.J., Martin, J.A., Caravelli, K., Stracquadanio, G., Boeke, J.D. Yeast Golden Gate (yGG) for the Efficient Assembly of *S. cerevisiae* Transcription Units. *ACS Synthetic Biology* **4**, 853-859 (2015).
58. Mitchell, L.A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N.A., Cai, Y., Truong, D.M., Veerakumar, A., Wang, Y., Mayorga, M., Blomquist, P., Sadda, P., Trueheart, J., Boeke, J.D. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Research* **43**, 6620-6630 (2015).
59. Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., French, C.E. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Research* **42**, e154 (2014).
60. Lee, M.E., Aswani, A., Han, A.S., Tomlin, C.J., Dueber, J.E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research*, gkt809 (2013).
61. Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Liebermeister, W., Flamholz, A., Noor, E., Amram, S., Brandis, A., Bareia, T., Yofe, I., Jubran, H., Milo, R. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Research* **41**, e98-e98 (2013).
62. López, J., Essus, K., Kim, I.-k., Pereira, R., Herzog, J., Siewers, V., Nielsen, J., Agosin, E. Production of β-ionone by combined expression of carotenogenic and plant *CCD1* genes in *Saccharomyces cerevisiae*. *Microbial cell factories* **14**, 84 (2015).
63. Symington, L.S., Gautier, J. Double-strand break end resection and repair pathway choice. *Annual review of genetics* **45**, 247-271 (2011).
64. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.-C., Stricher, F., Petit, A.-S., Patin, A., Guillier, S. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *Journal of molecular biology* **355**, 443-458 (2006).
65. Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R., Stoddard, B.L., Seligman, L.M. Isolation and characterization of new homing endonuclease specificities at individual target site positions. *Journal of molecular biology* **342**, 31-41 (2004).
66. Carroll, D., Morton, J.J., Beumer, K.J., Segal, D.J. Design, construction and in vitro testing of zinc finger nucleases. *Nature protocols* **1**, 1329-1341 (2006).
67. Cathomen, T., Joung, J.K. Zinc-finger nucleases: the next generation emerges. *Molecular Therapy* **16**, 1200-1207 (2008).
68. Nagaoka, M., Sugiura, Y. Artificial zinc finger peptides: creation, DNA recognition, and gene regulation. *Journal of inorganic biochemistry* **82**, 57-63 (2000).
69. Grover, A., Pande, A., Choudhary, K., Gupta, K., Sundar, D. Re-programming DNA-binding specificity in zinc finger proteins for targeting unique address in a genome. *Systems and synthetic biology* **4**, 323-329 (2010).

70. Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., Collins, J.J. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).
71. Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* **31**, 294-301 (2008).
72. Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F., Joung, J.K. Oligomerized pool engineering (OPEN): an'open-source'protocol for making customized zinc-finger arrays. *Nature protocols* **4**, 1471-1501 (2009).
73. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature methods* **8**, 67-69 (2011).
74. Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., Yang, B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Research*, gkr188 (2011).
75. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., Voytas, D.F. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, gkr218 (2011).
76. Boch, J., Bonas, U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Phytopathology* **48**, 419 (2010).
77. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512 (2009).
78. Moscou, M.J., Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501-1501 (2009).
79. Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., Zhang, F. A transcription activator-like effector toolbox for genome engineering. *Nature protocols* **7**, 171-192 (2012).
80. Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Höning, K., Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nature biotechnology* **31**, 76-81 (2013).
81. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., Joung, J.K. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology* **30**, 460-465 (2012).
82. Gaj, T., Gersbach, C.A., Barbas, C.F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**, 397-405 (2013).
83. Gupta, R.M., Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**, 4154-4161 (2014).
84. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
85. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308 (2013).
86. Horvath, P., Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170 (2010).
87. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
88. Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., Kim, J.-S. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods* **12**, 237-243 (2015).
89. Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J., Yee, J.-K. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature biotechnology* **33**, 175-178 (2015).

90. Shuler, M.L., Foley, P., Atlas, J. Modeling a minimal cell. *Microbial Systems Biology: Methods and Protocols*, 573-610 (2012).
91. Forster, A.C., Church, G.M. Towards synthesis of a minimal cell. *Molecular systems biology* **2**, 45 (2006).
92. Gil, R., Silva, F.J., Peretó, J., Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews* **68**, 518-537 (2004).
93. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
94. Dymond, J., Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered* **3**, 170-173 (2012).
95. Shen, Y., Stracquadanio, G., Wang, Y., Yang, K., Mitchell, L.A., Xue, Y., Cai, Y., Chen, T., Dymond, J.S., Kang, K. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome research* **26**, 36-49 (2016).
96. Mercy, G., Mozziconacci, J., Scolari, V.F., Yang, K., Zhao, G., Thierry, A., Luo, Y., Mitchell, L.A., Shen, M., Shen, Y., Walker, R., Zhang, W., Wu, Y., Xie, Z.-x., Luo, Z., Cai, Y., Dai, J., Yang, H., Yuan, Y.-J., Boeke, J.D., Bader, J.S., Muller, H., Koszul, R. 3D organization of synthetic and scrambled chromosomes. *Science* **355** (2017).
97. Phillips, T. Regulation of transcription and gene expression in eukaryotes. *Nature Education* **1**, 199 (2008).
98. Day, D., Tuite, M.F. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *Journal of Endocrinology* **157**, 361-371 (1998).
99. Roeder, R.G. The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences* **21**, 327-335 (1996).
100. Roeder, R.G. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends in biochemical sciences* **16**, 402-408 (1991).
101. Smale, S.T., Kadonaga, J.T. The RNA polymerase II core promoter. *Annual review of biochemistry* **72**, 449-479 (2003).
102. Bucher, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of molecular biology* **212**, 563-578 (1990).
103. Carey, M. The enhanceosome and transcriptional synergy. *Cell* **92**, 5-8 (1998).
104. Conaway, R.C., Conaway, J.W. General initiation factors for RNA polymerase II. *Annual Review of Biochemistry* **62**, 161-190 (1993).
105. Kornberg, R.D. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences* **104**, 12955-12961 (2007).
106. Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., Conaway, J.W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends in Biochemical Sciences* **30**, 250-255 (2005).
107. Macreadie, I.G., Horaitis, O., Verkuylen, A.J., Savin, K.W. Improved shuttle vectors for cloning and high-level Cu 2+-mediated expression of foreign genes in yeast. *Gene* **104**, 107-111 (1991).
108. Mountain, H.A., Byström, A.S., Larsen, J.T., Korch, C. Four major transcriptional responses in the methionine/threonine biosynthetic pathway of *Saccharomyces cerevisiae*. *Yeast* **7**, 781-803 (1991).
109. Cherest, H., Kerjan, P., Surdin-Kerjan, Y. The *Saccharomyces cerevisiae MET3* gene: nucleotide sequence and relationship of the 5' non-coding region to that of *MET25*. *Molecular and General Genetics MGG* **210**, 307-313 (1987).

110. Kramer, R.A., DeChiara, T.M., Schaber, M.D., Hilliker, S. Regulated expression of a human interferon gene in yeast: control by phosphate concentration or temperature. *Proceedings of the National Academy of Sciences* **81**, 367-370 (1984).
111. West, R., Yocum, R.R., Ptashne, M. *Saccharomyces cerevisiae GAL1-GAL10* divergent promoter region: location and function of the upstream activating sequence UASG. *Molecular and Cellular Biology* **4**, 2467-2478 (1984).
112. Rudolph, H., Hinnen, A. The yeast PHO5 promoter: phosphate-control elements and sequences mediating mRNA start-site selection. *Proceedings of the National Academy of Sciences* **84**, 1340-1344 (1987).
113. Koller, A., Valesco, J., Subramani, S. The *CUP1* promoter of *Saccharomyces cerevisiae* is inducible by copper in *Pichia pastoris*. *Yeast* **16**, 651-656 (2000).
114. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683-689 (1988).
115. Ma, J., Ptashne, M. Deletion analysis of *GAL4* defines two transcriptional activating segments. *Cell* **48**, 847-853 (1987).
116. Kakidani, H., Ptashne, M. *GAL4* activates gene expression in mammalian cells. *Cell* **52**, 161-167 (1988).
117. Sadowski, I., Ma, J., Triezenberg, S., Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563-564 (1988).
118. Chopra, I., Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and molecular biology reviews* **65**, 232-260 (2001).
119. Klug, A. Zinc finger peptides for the regulation of gene expression. *Journal of molecular biology* **293**, 215-218 (1999).
120. Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., Arlotta, P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149-153 (2011).
121. Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K. CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977-979 (2013).
122. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).
123. Quintero, M.J., Maya, D., Arévalo-Rodríguez, M., Cebolla, Á., Chávez, S. An improved system for estradiol-dependent regulation of gene expression in yeast. *Microbial cell factories* **6**, 1 (2007).
124. Salinas, F., Rojas, V., Delgado, V., Agosin, E., Larrondo, L.F. Optogenetic switches for light-controlled gene expression in yeast. *Applied Microbiology and Biotechnology*, 1-12 (2017).
125. Drepper, T., Krauss, U., Zu Berstenhorst, S.M., Pietruszka, J., Jaeger, K.-E. Lights on and action! Controlling microbial gene expression by light. *Applied microbiology and biotechnology* **90**, 23-40 (2011).
126. Nihongaki, Y., Kawano, F., Nakajima, T., Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature biotechnology* **33**, 755-760 (2015).
127. Taslimi, A., Zoltowski, B., Miranda, J.G., Pathak, G.P., Hughes, R.M., Tucker, C.L. Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nature chemical biology* **12**, 425-430 (2016).
128. Schindler, S.E., McCall, J.G., Yan, P., Hyrc, K.L., Li, M., Tucker, C.L., Lee, J.-M., Bruchas, M.R., Diamond, M.I. Photo-activatable Cre recombinase regulates gene expression *in vivo*. *Scientific reports* **5**, 13627-13627 (2014).
129. Edwards, W.F., Young, D.D., Deiters, A. Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS chemical biology* **4**, 441-445 (2009).

130. Beyer, H.M., Juillot, S., Herbst, K., Samodelov, S.L., Müller, K., Schamel, W.W., Römer, W., Schäfer, E., Nagy, F., Strähle, U. Red light-regulated reversible nuclear localization of proteins in mammalian cells and zebrafish. *ACS Synthetic Biology* **4**, 951-958 (2015).
131. Levskaya, A., Weiner, O.D., Lim, W.A., Voigt, C.A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997-1001 (2009).
132. Binder, D., Bier, C., Grünberger, A., Drobietz, D., Hage-Hülsmann, J., Wandrey, G., Büchs, J., Kohlheyer, D., Loeschcke, A., Wiechert, W. Photocaged Arabinose: A Novel Optogenetic Switch for Rapid and Gradual Control of Microbial Gene Expression. *ChemBioChem* **17**, 296-299 (2016).
133. Kusen, P.M., Wandrey, G., Probst, C., Grünberger, A., Holz, M., Meyer zu Berstenhorst, S., Kohlheyer, D., Büchs, J., Pietruszka, J.r. Optogenetic Regulation of Tunable Gene Expression in Yeast Using Photo-Labile Caged Methionine. *ACS Chemical Biology* **11**, 2915-2922 (2016).
134. Wandrey, G., Bier, C., Binder, D., Hoffmann, K., Jaeger, K.-E., Pietruszka, J., Drepper, T., Büchs, J. Light-induced gene expression with photocaged IPTG for induction profiling in a high-throughput screening system. *Microbial cell factories* **15**, 63 (2016).
135. Pathak, G.P., Vrana, J.D., Tucker, C.L. Optogenetic control of cell function using engineered photoreceptors. *Biology of the Cell* **105**, 59-72 (2013).
136. Quail, P., Boylan, M., Parks, B., Short, T., Xu, Y., Wagner, D. Phytochromes: photosensory perception and signal transduction. *Science* **268**, 675-680 (1995).
137. Quail, P.H. Phytochrome photosensory signalling networks. *Nature Reviews Molecular Cell Biology* **3**, 85-93 (2002).
138. Smith, H. Phytochromes and light signal perception by plants - an emerging synthesis. *Nature* **407**, 585-591 (2000).
139. Ahmad, M., Jarillo, J.A., Cashmore, A.R. Chimeric proteins between cry1 and cry2 Arabidopsis blue light photoreceptors indicate overlapping functions and varying protein stability. *The Plant Cell* **10**, 197-207 (1998).
140. Cashmore, A.R., Jarillo, J.A., Wu, Y.-J., Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284**, 760-765 (1999).
141. Briggs, W., Beck, C., Cashmore, A., Christie, J., Hughes, J., Jarillo, J., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A. The phototropin family of photoreceptors. *The Plant Cell* **13**, 993-997 (2001).
142. Briggs, W.R., Christie, J.M., Salomon, M. Phototropins: a new family of flavin-binding blue light receptors in plants. *Antioxidants and Redox Signaling* **3**, 775-788 (2001).
143. Kliebenstein, D.J., Lim, J.E., Landry, L.G., Last, R.L. Arabidopsis UVR8 regulates ultraviolet-B signal transduction and tolerance and contains sequence similarity to human regulator of chromatin condensation 1. *Plant Physiology* **130**, 234-243 (2002).
144. Rizzini, L., Favory, J.-J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., Baumeister, R., Schäfer, E., Nagy, F., Jenkins, G.I. Perception of UV-B by the Arabidopsis UVR8 protein. *Science* **332**, 103-106 (2011).
145. Mathews, S., Sharrock, R. Phytochrome gene diversity. *Plant, Cell & Environment* **20**, 666-671 (1997).
146. Quail, P. An emerging molecular map of the phytochromes. *Plant, Cell & Environment* **20**, 657-665 (1997).
147. Hendricks, S.B., Borthwick, H.A. The function of phytochrome in regulation of plant growth. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 2125-2130 (1967).
148. Ni, M., Tepperman, J.M., Quail, P.H. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature* **400**, 781-784 (1999).
149. Park, E., Park, J., Kim, J., Nagatani, A., Lagarias, J.C., Choi, G. Phytochrome B inhibits binding of phytochrome-interacting factors to their target promoters. *The Plant Journal* **72**, 537-546 (2012).

150. Franklin, K.A., Quail, P.H. Phytochrome functions in Arabidopsis development. *Journal of experimental botany* **61**, 11-24 (2010).
151. Martínez-García, J.F., Huq, E., Quail, P.H. Direct targeting of light signals to a promoter element-bound transcription factor. *Science* **288**, 859-863 (2000).
152. Shimizu-Sato, S., Huq, E., Tepperman, J.M., Quail, P.H. A light-switchable gene promoter system. *Nature biotechnology* **20**, 1041-1044 (2002).
153. Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D., Tucker, C.L. Rapid blue-light-mediated induction of protein interactions in living cells. *Nature methods* **7**, 973-975 (2010).
154. Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., Zhang, F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476 (2013).
155. Polstein, L.R., Gersbach, C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* **11**, 198-200 (2015).
156. Tabor, J.J., Levskaya, A., Voigt, C.A. Multichromatic Control of Gene Expression in *Escherichia coli*. *Journal of molecular biology* **405**, 315-324 (2011).
157. Müller, K., Engesser, R., Schulz, S., Steinberg, T., Tomakidi, P., Weber, C.C., Ulm, R., Timmer, J., Zurbriggen, M.D., Weber, W. Multi-chromatic control of mammalian gene expression and signaling. *Nucleic Acids Research* **41**, e124 (2013).
158. Müller, K., Engesser, R., Timmer, J., Zurbriggen, M.D., Weber, W. Orthogonal Optogenetic Triple-Gene Control in Mammalian Cells. *ACS Synthetic Biology* **3**, 796-801 (2014).
159. Casini, A., Storch, M., Baldwin, G.S., Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology* (2015).
160. Ellis, T., Adie, T., Baldwin, G.S. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology* **3**, 109-118 (2011).
161. Chao, R., Yuan, Y., Zhao, H. Recent advances in DNA assembly technologies. *FEMS yeast research* (2014).
162. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., Hutchison, C.A. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20404-20409 (2008).
163. Galanie, S., Thodey, K., Trenchard, I.J., Interrante, M.F., Smolke, C.D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095-1100 (2015).
164. Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Research*, gkx034 (2017).
165. Lee, M.E., DeLoache, W.C., Cervantes, B., Dueber, J.E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology* (2015).
166. Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., Orzaez, D. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PloS one* **6**, e21622 (2011).
167. Pereira, F., de Azevedo, F., Parachin, N.S., Hahn-Hägerdal, B., Gorwa-Grauslund, M.F., Johansson, B. The Yeast Pathway Kit: a method for metabolic pathway assembly with automatically simulated executable documentation. *ACS Synthetic Biology* (2016).
168. Crook, N.C., Freeman, E.S., Alper, H.S. Re-engineering multicloning sites for function and convenience. *Nucleic Acids Research* **39**, e92-e92 (2011).
169. Blake, W.J., Chapman, B.A., Zindal, A., Lee, M.E., Lippow, S.M., Baynes, B.M. Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Research*, gkq123 (2010).

170. Chen, W.-H., Qin, Z.-J., Wang, J., Zhao, G.-P. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Research* **41**, e93-e93 (2013).
171. Li, M.V., Shukla, D., Rhodes, B.H., Lall, A., Shu, J., Moriarity, B.S., Largaespada, D.A. HomeRun vector assembly system: a flexible and standardized cloning system for assembly of multi-modular DNA constructs. *PloS one* **9**, e100948 (2014).
172. Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T., Baldwin, G.S. BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synthetic Biology* (2015).
173. Torella, J.P., Boehm, C.R., Lienert, F., Chen, J.-H., Way, J.C., Silver, P.A. Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Research* **42**, 681-689 (2014).
174. Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K., Mueller-Roeber, B. Characterizing seamless ligation cloning extract for synthetic biological applications. *Analytical Biochemistry* **509**, 24-32 (2016).
175. Hartley, J.L., Temple, G.F., Brasch, M.A. DNA cloning using in vitro site-specific recombination. *Genome research* **10**, 1788-1795 (2000).
176. Deiters, A. Light activation as a method of regulating and studying gene expression. *Current opinion in chemical biology* **13**, 678-686 (2009).
177. Kawano, F., Okazaki, R., Yazawa, M., Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nature Chemical Biology* (2016).
178. Hughes, R.M., Bolger, S., Tapadia, H., Tucker, C.L. Light-mediated control of DNA transcription in yeast. *Methods* **58**, 385-391 (2012).
179. Sorokina, O., Kapus, A., Terecskei, K., Dixon, L.E., Kozma-Bognar, L., Nagy, F., Millar, A.J. A switchable light-input, light-output system modelled and constructed in yeast. *Journal of biological engineering* **3**, 1 (2009).
180. Müller, K., Siegel, D., Rodriguez Jahnke, F., Gerrer, K., Wend, S., Decker, E.L., Reski, R., Weber, W., Zurbriggen, M.D. A red light-controlled synthetic gene expression switch for plant systems. *Molecular BioSystems* **10**, 1679-1688 (2014).
181. Müller, K., Zurbriggen, M.D., Weber, W. Control of gene expression using a red- and far-red light–responsive bi-stable toggle switch. *Nature Protocols* **9**, 622-632 (2014).
182. Robertson, J.B., Davis, C.R., Johnson, C.H. Visible light alters yeast metabolic rhythms by inhibiting respiration. *Proceedings of the National Academy of Sciences* **110**, 21130-21135 (2013).
183. Bodvard, K., Wrangborg, D., Tapani, S., Logg, K., Sliwa, P., Blomberg, A., Kvarnström, M., Käll, M. Continuous light exposure causes cumulative stress that affects the localization oscillation dynamics of the transcription factor Msn2p. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1813**, 358-366 (2011).
184. Polstein, L.R., Gersbach, C.A. Light-Inducible Spatiotemporal Control of Gene Activation by Customizable Zinc Finger Transcription Factors. *Journal of the American Chemical Society* **134**, 16480-16483 (2012).
185. Lebar, T., Jerala, R. Benchmarking of TALE-and CRISPR/dCas9-based transcriptional regulators in mammalian cells for the construction of synthetic genetic circuits. *ACS Synthetic Biology* **5**, 1050-1058 (2016).
186. Verwaal, R., Wang, J., Meijnen, J.-P., Visser, H., Sandmann, G., van den Berg, J.A., van Ooyen, A.J. High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Applied and Environmental Microbiology* **73**, 4342-4350 (2007).
187. Flagfeldt, D.B., Siewers, V., Huang, L., Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545-551 (2009).

188. Matsushita, T., Mochizuki, N., Nagatani, A. Dimers of the N-terminal domain of phytochrome B are functional in the nucleus. *Nature* **424**, 571-574 (2003).
189. Pfeiffer, A., Nagel, M.-K., Popp, C., Wüst, F., Bindics, J., Viczián, A., Hiltbrunner, A., Nagy, F., Kunkel, T., Schäfer, E. Interaction with plant transcription factors can mediate nuclear import of phytochrome B. *Proceedings of the National Academy of Sciences* **109**, 5892-5897 (2012).
190. Van den Ackerveken, G., Marois, E., Bonas, U. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* **87**, 1307-1316 (1996).
191. Karim, A.S., Curran, K.A., Alper, H.S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS yeast research* **13**, 107-116 (2013).
192. Le, Y., Gagneten, S., Tombaccini, D., Bethke, B., Sauer, B. Nuclear targeting determinants of the phage P1 cre DNA recombinase. *Nucleic Acids Research* **27**, 4703-4709 (1999).
193. Kanegae, Y., Lee, G., Sato, Y., Tanaka, M., Nakal, M., Sakaki, T., Sugano, S., Saito, I. Efficient gene activation in mammalian cells by using recombinant adenovirus expressing site-specific Cre recombinase. *Nucleic Acids Research* **23**, 3816-3821 (1995).
194. Faal, T., Wong, P.T., Tang, S., Coulter, A., Chen, Y., Tu, C.H., Baker, J.R., Choi, S.K., Inlay, M.A. 4-Hydroxytamoxifen probes for light-dependent spatiotemporal control of Cre-ER mediated reporter gene expression. *Molecular BioSystems* **11**, 783-790 (2015).
195. Inlay, M.A., Choe, V., Bharathi, S., Fernhoff, N.B., Baker, J.R., Weissman, I.L., Choi, S.K. Synthesis of a photocaged tamoxifen for light-dependent activation of Cre-ER recombinase-driven gene modification. *Chemical communications* **49**, 4971-4973 (2013).

I. CURRICULUM VITAE

PERSONAL INFORMATION

Name	Lena Hochrein
Date of Birth	January 29th 1988
Place of Birth	Bad Kissingen

EDUCATION AND RESEARCH EXPERIENCE

06/2007	Jack-Steinberger-Gymnasium, Bad Kissingen University entrance qualification
10/2007 – 02/2012	Technical University of Berlin Study of Chemistry Degree earned: Diploma in Chemistry
05/2012 – 12/2012	Technical University of Berlin Diploma thesis Title: Characterization of kynureninase genes in the actinomycin biosynthetic gene cluster of *Streptomyces chrysomallus.*
05/2013 – 10/2016	University of Potsdam, Junior Research Group Cell2Fab PhD student
11/2016 - 04/2017	Scholarship holder of the Potsdam Graduate School (Completion grant) PhD student

II. SCIENTIFIC ACHIEVEMENTS

PUBLICATIONS

Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K., Mueller-Roeber, B. (2016). Characterizing Seamless Ligation Cloning Extract for Synthetic Biology Applications. *Analytical Biochemistry* 509: 24-32.

Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. (2017). AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res.,* gkx034.

Hochrein, L., Machens, F., Messerschmidt, K., Mueller-Roeber, B. PhiReX – a Programmable and Red Light-regulated Protein Expression Switch for Yeast. *Submitted to Nucleic Acids Research.* (Status: In peer-review)

Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K., Mueller-Roeber, B. L-SCRaMbLE: a Tool for Light-controlled Cre-mediated Recombination in Yeast. *Submitted to Nature Biotechnology.* (Status: Under consideration)

CONFERENCE CONTRIBUTIONS

Hochrein, L., Machens, F., Mueller-Roeber, B., Messerschmidt, K. (2015). Controlling multiple gene expression in *Saccharomyces cerevisiae* by optogenetic switches. SEED 2015, Jun 10-13, Boston, USA.

Hochrein, L., Machens, F., Mueller-Roeber, B., Messerschmidt, K. (2014). Establishing a red light-inducible protein expression system with artificial transcription factors. DELATEC, Jun 1-4, Potsdam, Germany.

Hochrein, L., Machens, F., Mueller-Roeber, B., Messerschmidt, K. (2013). Comparison of alternative cloning methods. Molecular Interactions, Aug 14-16, Berlin, Germany.

Hochrein, L., Crnovcic, I., Keller, U. (2012) Characterization of kynureninase genes in the actinomycin biosynthetic gene cluster of *Streptomyces chrysomallus*. Symposium: Actinobacteria within soils: Capacities for mutualism, symbiosis and pathogenesis, Oct 25-28, Münster, Germany.

III. DECLARATIONS

DECLARATION OF INDEPENDENT WORK

I hereby confirm that I have written the attached thesis on my own and that I did not use any resources than those specified above. This work has not been previously submitted, either in the same or in a similar form to any other examination committee and has not yet been published.

Potsdam, April 25th 2017

Lena Hochrein

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter Verwendung keiner anderen als den von mir angegebenen Quellen und Hilfsmitteln verfasst habe.

Ferner erkläre ich, dass ich bisher weder an der Universität Potsdam noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Potsdam, den 25.04.2017

Lena Hochrein

DECLARATION OF OWN CONTRIBUTIONS TO SINGLE PUBLICATIONS

AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies

L.H. and F.M. jointly designed and developed the overall strategy, conducted the experiments and analyzed the data with equal contribution. L.H. and F.M. jointly planned diverse optimization steps and designed the final AssemblX-system. K.S. constructed the different vector sets under supervision of L.H. and F.M. The online tool was programmed by J.G. according to specifications defined by L.H. and F.M. L.H. and F.M. wrote the manuscript with equal contribution. B.M.-R. initiated the project and the overall research strategy and edited the manuscript. B.M.-R and K.M. jointly supervised the work.

PhiReX – a Programmable and Red Light-regulated Protein Expression Switch for Yeast

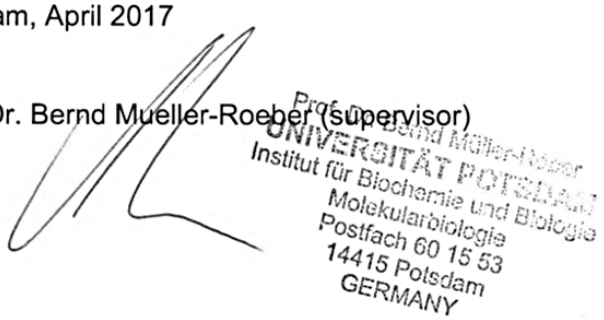
L.H. designed and developed the overall strategy with contributions from F.M. L.H. planned and conducted the experiments, analyzed the data and wrote the manuscript. The synTALE-DBD/BS pair was provided and characterized by F.M., who also assisted with Cas9-based genome editing. B.M.-R. initiated the project and the overall research strategy. B.M.-R. supervised the work. F.M. and B.M.-R. edited the manuscript. All work was carried out in the Cell2Fab lab, led by K.M.

L-SCRaMbLE: a Tool for Light-controlled Cre-mediated Recombination in yeast

L.H. initiated the project of a light-inducible 'SCRaMbLE'-system, developed the overall strategy and designed the appropriate constructs and experiments with contributions from L.A.M. L.H. conducted the experiments for the characterization of the light-responsible Cre recombinase and analyzed the data. L.A.M. designed and constructed plasmids pLM494 and pLM006. Cloning of constructs pLH_SCR12-pLH_SCR16 and yeast transformations were done by K.S. under supervision of L.H. L.H. wrote the paper with contributions from B.M.-R. and L.A.M.

Potsdam, April 2017

Prof. Dr. Bernd Mueller-Roeber (supervisor)



IV. ACKNOWLEDGEMENTS

First and foremost, I want to thank Prof. Dr. Bernd Müller-Röber for giving me the opportunity to carry out my PhD under his supervision; not just for many interesting discussions and a great cooperation, but also for his infectious enthusiasm for science. Thank you for all the time you invested in helping me.

Another thank you goes to Prof. Dr. Katja Arndt and Dr. Katrin Messerschmidt for accompanying the work as second supervisor and mentor.

I also want to express my gratitude to Dr. Arren Bar-Even and Prof. Dr. Helge B. Bode for accepting to be referees of this work.

Furthermore, I want to thank all members of the Cell2Fab group and the whole Müller-Röber group. I would like to express my special thanks to Dr. Fabian Machens, who supported me and all of this work. Thanks for all the ideas we developed, the great discussions we had and everything I could learn from you. Thank you for lifting me up every time I fell and for your confidence that in the end all will be well. I also want to especially thank Karina Schulz, who made every day's work fun (at least after her first coffee). Thank you for cloning assistance and many encouraging words. I also want to express my gratitude to Ines Nowak for her kind support with any administrative affairs.

My personal gratitude goes to my parents, who supported and believed in me my whole life. Thank you for always accepting my decisions and for reminding me of the truly important things in life. Another special thank you goes to my brother Martin, not only for editing of this work and the design of the Cell2Fab-, AssemblX- and PhiReX-logos, but also for just being the refreshing and inspiring person you are. A huge thank you goes to my aunty Doris for always being there for me.

And finally, thanks to my friends, for your understanding and encouragements.