Improvement of Synthetic Biology Tools for DNA Editing

Cavaleiro, Mafalda; Nørholm, Morten

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Improvement of Synthetic Biology Tools for DNA Editing

Ph.D. thesis

Ana Mafalda Cavaleiro

Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

Supervisor: Morten Nørholm

Improvement of Synthetic Biology Tools for DNA Editing

Novo Nordisk Center For Biosustainability
Department of Systems Biology
Technical University of Denmark
Cover art: Illustration by Se Hyeuk Kim, Ph.D.

Todo o gesto é um acto revolucionário.
Every gesture is a revolutionary act.

Fernando Pessoa

Preface

This thesis is written as a partial fulfilment of the requirements to obtain a Ph.D. degree at the Technical University of Denmark. This thesis was carried out at the Novo Nordisk Foundation Center For Biosustainability, Technical University of Denmark, from March 2013 to February 2016 under the supervision of Morten Nørholm. This Ph.D. project was funded by The Novo Nordisk Foundation and a Ph.D. grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme [FP7-People-2012-ITN], under grant agreement no. 317058, "BACTORY".

Ana Mafalda Cavaleiro

Hørsholm, February 2016

Abstract

The unpredictability and complexity of biological systems limit the development of economically efficient bio-based production processes that rely on renewable carbon sources and are essential for biosustainability and environmental protection. Synthetic biology (synbio) aims at making biology easier to engineer and addresses these challenges.

The ability to systematically construct, modify and tune biological systems from fully characterized biological components, or parts, is crucial to the success of synbio projects. This thesis aims at contributing to standardization and part sharing with the development and improvement of DNA editing strategies, compatible with other DNA assembly methodologies, genome engineering and, eventually, automation processes.

Expanding and optimizing the synbio toolkit has important applications in pathway optimization for metabolic engineering, design and characterization of gene circuits, synthesis of whole genomes and natural product discovery. In line with this, it is also described in this thesis how discovery of new cytochromes P450 (CYPs) from marine bacteria could benefit industrial processes.

Dansk resumé

Uforudsigeligheden og kompleksiteten af biologiske systemer begrænser udviklingen af økonomisk effektive biobaserede produktionsprocesser, og disse er afhængige af vedvarende kulstofkilder og er afgørende for biobæredygtighed og miljøbeskyttelse. Syntetisk biologi (synbio) har til formål at muliggøre og standardisere udnyttelse og design af nye biologiske systemer.

Evnen til systematisk at konstruere, modificere og optimere biologiske systemer fra velkarakteriserede biologiske komponenter er afgørende for succes i synbio projekter. Denne afhandling bidrager til at standardisere og forbedre DNA redigeringsteknologier til modelbakterien *Escherichia coli* - og i sidste ende muliggøre automatisering af sådanne processer.

Udvidelse og optimering af en synbio værktøjskasse er desuden vigtig for biosynteseoptimering, design af cellefabrikker og opdagelsen af nye naturstoffer. I tråd med dette undersøges marine bakteriers potentiale for opdagelsen af nye enzymer til anvendelse i industrielle processer.

Acknowledgements

Being the first person to start in the Bactory training program was challenging but I met a lot of super talented and amazing people along the way that made it an easier task.

I would like to start by thanking my mentor Dr. Morten Nørholm for giving me the opportunity to speak my mind freely, for putting up with my brutal honesty, for never saying NO to my ideas, and for being an extremely enthusiastic and fun supervisor to work with (not for). You were the best "Boss" I have ever had. ☺

Thanks to all the current and former members of the Membrane Synthetic Biology group: Susanna Seppälä, Maja Rennig, Roberto Ferro, Darío Vázquez Albacete, Ida Lauritsen, Ulla Christensen, Emil Christian Fisher, Tonja Wolff, Karina Marie Søgaard and Victor Forman, for the amazing group environment and brainstorming, Sofie Wendel for the inspiring and super positive attitude, Morten Thrane Nielsen for guidance and for helping me focus, Virginia Martinez and specially Se Hyeuk Kim for being the most generous, fun and helpful post-docs a Ph.D. student can wish for. You rock! ☺

I would also like to express my gratitude to Professor Søren Molin for the opportunity to be part of the Bactory Training program and Dr. Ana Rute Neves, screening department manager at Christian Hansen, for believing and encouraging me to apply to the program in the first place.

I think the Bactory was a success in a lot of aspects and I'm pretty sure most of it was because of us, the fourteen chosen ones ☺ I would like to thank each one of you for all the fun moments, brainstorming, advice and even fights that made us grow as individuals but also as a group. I owe a huge and special thanks to my great dear friend and collaborator in several projects Henrique Machado for being an amazing person to work with, for all the very long discussions and brainstorming, for all the laughs, for putting up with my complaints, for pushing me forward and for being always there when I needed the most. Obrigada do fundo do coração! ☺ A very special thanks to Isotta D'Arrigo, my friend and travel companion, for all the support and "scientific trust", for all the discussions, for making me laugh 24/7 and for listening (but not talking that much). ☺ Remember Stockholm, Tel-Aviv, Seoul and China in…? Another huge thanks to my friend and party partner, Klara Bojanovič for all the help and brainstorming, for being contagiously happy and responsible for an amazing environment in and out of the office, and for always having a nice word to say in the worst situation.

During my three-year period at CFB, I met a lot of amazing people and was fortunate enough to make a lot of friends. I cannot list you all but thank you CFB people! I owe a really huge thanks to my colleagues and friends Mikkel Lindgaard, João Cardoso, Rosa Aragão, Gheorghe Borja, Nabin Aryal, Christian Lieven and specially Carlotta Ronda for all the advice, help and for making it so easy for me to go to CFB everyday. You guys are amazing!

My Ph.D. also took me to Evolva for two months, where I was lucky to meet amazing people. I would like to thank particularly Carlos Casado and Ângela Carvalho for fruitful discussions and for making my stay at Evolva so inspiring.

As challenging as a Ph.D. might be, I think the most challenging task belonged to my boyfriend Gustavo Seleiro. Thanks for all the help, encouragement and positive attitude, for putting up with me in general and for being present through my darkest times. ☺

I would like to thank my mother for being my role model and for supporting me blindly no matter what, and my twin sister for listening to my complaints and believing in me. I could not have done it without you!

To my father for caring and actually trying to understand what I was working on, to my big brother and the rest of my family for encouragement, and my grandparents for being always with me.

A very heart-felted thanks to my godmother, Professor Luísa Morgado, for wise advice, encouragement and support throughout the whole Ph.D. process.

Finally, I would like to thank my friends Maria Joao Teixeira, Paulo Fernandes, Vanessa Nunes, Margarida Jerónimo Barbosa, Carolina Albuquerque, Sofia Fernandes, Ana Conduto, Francisca Monteiro, Catarina Santos, André Grilo, Cristiana Madaleno, Filipa Corrêa D'Almeida, Inês Luís, Cristiana Faria, Ana Maria Vaz, Sara Jayne and all the others I cannot mention for caring and being present despite the physical distance.

Contents

List of publications

Paper I

Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T., Nørholm, M. H. H. (2015) **Accurate DNA assembly and genome engineering with optimized uracil excision cloning**. *ACS Synth Biol* **4**, 1042-1046.

Paper II

Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppälä, S., Nørholm, M. H. H. (2015) **Uracil excision for assembly of complex pathways**. In *Springer Protocols Handbooks*, p1-11, Humana Press, Springer-Verlag Berlin Heidelberg.

Paper III

Kim, S. H., Cavaleiro, A. M., Rennig, M., Nørholm, M. H. H. (2016) **SEVA linkers: a versatile and automatable DNA backbone exchange standard for synthetic biology**. *ACS Synth Biol* (*In press*).

Paper IV

Cavaleiro, A. M.[*], Machado, H.[*], Nørholm, M. H. H.[§], Gram, L. (2016) **Marine Bacterial Cytochromes P450 and their potential in biotechnology**.

Manuscript submitted to *Plos One*

*(These authors contributed equally)

Not included in this thesis

Vazquez, D. A., Cavaleiro, A. M., Christensen, U., Seppälä, S., Møller, B. L., Nørholm, M. H. H. (2016) **Exploiting the N-terminal tag toolbox for functional expression of medicinal cytochromes P450**.
Manuscript in preparation and planned for submission to *Microbial Cell Factories*

Machado, H.*, Cavaleiro, A. M.*, D'Arrigo, I., Bojanovič, K., Gram, L., Nørholm, M. H. H. (2016) **Exploring marine environments to unravel tolerance mechanisms to relevant compounds**.
Manuscript in preparation.
*(These authors contributed equally)

Abbreviations

A – Deoxyadenosine

AMUSER – Automated DNA Modifications with USER Cloning

att – Attachment Sequence Motifs

BASIC – Biopart Assembly Standard for Idempotent Cloning

BIOFAB – International Open Facility Advancing Biotechnology

bp – Base Pair

C – Deoxycytidine

CAD – Computer-Aided Design

Cas9 – CRISPR associated gene

CFU – Colony Forming Units

CPEC – Circular Polymerase Extension Cloning

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

CYPs – Cytochromes P450

DIYBIO – Do-it-yourself Biology

DNA – Deoxyribonucleic Acid

dNTPs – Deoxyribonucleotides triphosphate

dsDNA – Double Stranded DNA

G – Deoxyguanosine

GFP – Green Fluorescent Protein

GRAS – Generally Recognized As Safe

gRNA – Guide RNA (synthetic)

HVAS – Homerun Vector Assembly System

IDT – Integrated DNA Technologies

iGEM – International Genetically Engineered Machine student competition

IP – Intellectual Property

IHF – Integration Host Factor

JBEI-ICEs – Joint BioEnergy Institute Inventory of Composable Elements

kb – Kilobase

LIC – Ligation-Independent Cloning

LCR – Ligase Cycling Reaction

M – Molar

MASTER – Methylation-Assisted Tailorable Ends Rational

Mg^{+} – Magnesium ion

mMoClo – Mammalian Modular Cloning System

MoClo – Modular Cloning System

MODAL – Modular Overlap-Directed Assembly with Linkers

N – Random Deoxyribonucleotide

NGS – Next-Generation Sequencing

NOMAD – Nucleic acid Ordered Assembly with Directionality

OE-PCR – Overlap Extension Polymerase Chain Reaction

PCA – Polymerase Chain Assembly

PCR – Polymerase Chain Reaction

Pfu – *Pyrococcus furiosus*

PHUSER – Primer Help for USER

PLICing – Phosphorothioate-based Ligase-Independent gene Cloning

PSA – Pairwise Selection Assembly

RE – Restriction Enzymes

RNA – Ribonucleic Acid

SEVA – Standard European Vector Architecture

SLIC – Sequence and Ligation-Independent Cloning

SLiCE – Seamless Ligation Cloning Extract

SIRA – Serine Integrase Recombinational Assembly

SRAS – Single-selective-marker Recombination Assembly System

ssDNA – Single Stranded DNA

SSRTA – Site-Specific Recombination-based Tandem Assembly

Synbio – Synthetic biology

T – Deoxythymidine

Taq – *Thermus aquaticus*

TALEs – Transcription Activator-like Effectors

T_m – Melting Temperature

U – Deoxyuridine

µL – Microliter

UDG – Uracil DNA Glycosidase

USER – Uracil Specific Reagent

VEGAS – Versatile Genetic Assembly System

List of figures

Introduction and thesis overview

To reduce the world's dependency on oil, our society has to change from a fossil-based industry to more biosustainable production strategies. Environmental protection and biosustainability are currently heavily dependent on economically efficient bio-based production processes from renewable carbon sources[1]. The problem with this approach is that biological systems are complex and unpredictable. Synthetic biology (synbio) addresses these challenges.

One definition of synbio is the attempt to make biology easier to engineer by understanding and making sense of composition and rational logic of biological systems[2–4]. Its potential was recognized in 2012 when synbio was included in the World Economic Forum's Global Agenda Council on Emerging Technologies as one of the top 10 emerging technologies (http://www.chemistryviews.org/details/news/1478647/The_2012_Top_10_Emerging_Technologies.html). Important accomplishments and the available synbio toolkit have been reviewed[5–7].

The vision of synbio relies on key concepts such as multidisciplinarity and standardization. The ultimate goal of the field is the targeted design and rational construction of complex biological systems for industrial purposes.

In 1988, Richard Feynman said: "What I cannot create, I do not understand". This statement has become a foundational concept of synbio. The success of future synbio and biotechnology will depend on the ability to systematically construct,

modify and tune biological systems from fully characterized biological components, or parts[8]. Parts are defined as DNA fragments that perform a specific function in a genetic circuit and an extended catalogue of well-studied biological parts is currently lacking. Bringing people together in a highly dynamic synbio community for the common goal of gathering knowledge in a systematic way is one of the adopted strategies to overcome this limitation. Several groups are currently involved in shaping the field such as funding agencies, research centres, companies or amateur "biohackers" like the International Genetically Engineered Machine (iGEM) student competition or the do-it-yourself biology (DIYBIO) movement[9,10].

When starting a new synbio project, there are typically at least four parameters to consider when choosing the molecular tools and the cellular container that accommodates and executes the necessary cellular functions (the chassis), for production of specific proteins, chemicals and genetic circuits[11]:

a) Chassis stability and growth conditions;

b) Available DNA editing technologies;

c) Ability to manipulate homologous and heterologous biosynthetic pathways (e.g. promoters and vectors);

d) Software tools for assisted planning and debugging.

Thesis outline

This thesis focuses on several important tools used in synbio. For simplification purposes, a glossary with the synbio concepts used throughout the thesis is provided (on page 4). The thesis is focused mainly on three of the above-mentioned parameters: chassis selection, available DNA editing strategies and software tools. Additionally, the most recent advances in DNA synthesis and sequencing, as well as examples of some of the most successful applications using synbio tools are briefly discussed. The need for standardization and more efficient DNA assembly technologies fuelled the work described in three publications: Paper I - *Accurate DNA assembly and genome engineering with optimized uracil excision cloning*; Paper II – *Uracil excision for assembly of complex pathways*; Paper III – *SEVA linkers: a versatile and automatable DNA backbone exchange standard for synthetic biology*. Paper IV - *Marine Bacterial Cytochromes P450 and their potential in biotech* is a good example on how powerful *in silico* tools can be for new enzyme discovery with industrial applications. Finally, concluding remarks and future perspectives are described, summarizing the contribution of this project to the synbio field.

1. Glossary – Important synbio concepts

Abstraction - the idea that complex biological functions can be organized across levels and conceptually separated from the original sequence context;

BioBricks - standardized biological components;

Chassis - cellular container that accommodates and executes the required cellular functions;

Combinatorial - assembly with any number of parts at any number of defined positions;

Destination vector - the vector into which the DNA from entry vector will be sub-cloned;

Entry vector – the vector that carries the DNA segment to be transferred to destination vector;

Forbidden sites – restriction sites recognized by restriction enzymes that cannot be present in parts to be assembled;

Genetic circuit - functional clusters of genes that impact each other's expression

Modularity – the ability of a component or system to function in a context-independent way;

Operon - group of genes under the control of a single promoter or multiple promoters;

Orthogonality – independent behaviour of biological parts;

Standardization – process of developing or implementing technical standards (fully characterized parts);

Part - DNA sequence defined by a specific function it performs in a genetic circuit;

Pathway – group of genes or operons, which may perform related functions;

Transcriptional units - DNA sequences that encode a single RNA molecule together with sequences required for its transcription - usually promoter, open reading frame and terminator.

2. The pre-synbio Era: History of recombinant DNA engineering

The elucidation of the molecular structure of DNA in 1953 by James Watson, Francis Crick and Rosalin Franklin was an important first step in the history of recombinant DNA engineering. Since then, several achievements have joined the hall of fame in DNA engineering - from oligonucleotide and gene synthesis to the discovery of restriction enzymes (Figure 1). These tools created the basis for the first successful recombinant DNA engineering[12–15] and for the first heterologously expressed gene for production of a recombinant pharmaceutical protein (human insulin) in the early 80's[16].

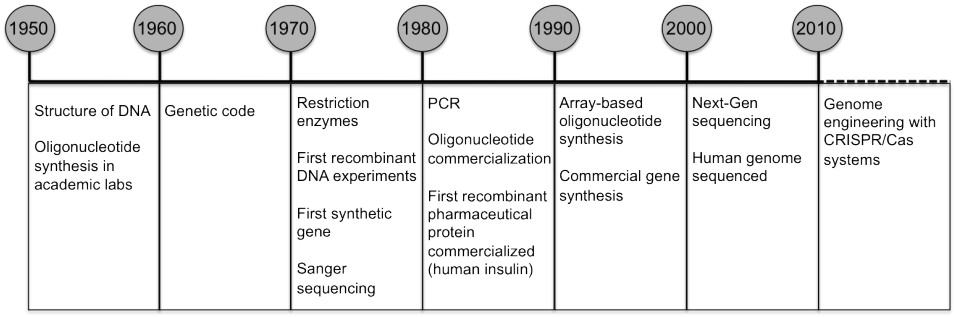


Figure 1 – Timeline on history of recombinant DNA engineering.

Less than a decade later, the development of the Polymerase Chain Reaction (PCR) was a major breakthrough. The simple and exponential nature of PCR, which consists of the combined action of two synthetic oligonucleotides with a DNA template strand and a thermophilic DNA polymerase, is a prime example of

a human-made innovation that revolutionized molecular biology. PCR simply made it easy to work with DNA. Shortly after the advent of PCR, a plethora of novel methods for DNA editing were developed. DNA reading and writing technologies followed and around the new millennium, commercial gene synthesis by Blue Heron Biotechnologies, Inc and Next Generation sequencing hit the market.

Recent major achievements that will transform synbio are related to the development of CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats) engineering systems. Just like restriction enzymes, CRISPR is a bacterial immune mechanism used by many prokaryotes for protection against invading nucleotide sequences[17–20]. Transformation from a natural immune system to a genome engineering tool happened when two independent reports demonstrated for the first time that type II CRISPR/Cas9 nucleases (from *Streptococcus pyogenes*) could be programmed using short synthetic guide RNAs (gRNAs) for sequence-specific targeting and cleavage in human and mouse cells[21,22]. In the same year (2013), this was shown for bacteria using Cas9 from *S. pneumoniae*[23]. The ability to reprogram Cas9 to target any sequence, as well as the accuracy, efficiency, simplicity and cost of CRISPR/Cas systems turned CRISPR/Cas-based genome engineering into the most popular technology currently available. In essence, CRISPR is doing to genome engineering, what restriction enzymes and PCR did to DNA engineering.

Overall, the constant innovation and improvement of most of these technologies help build the foundations of synbio.

3. The value of standardization and part sharing

Biological systems are generally complex and unpredictable, which makes them difficult to engineer. Furthermore, combinatorial studies can easily become too laborious or even unfeasible to perform when assembly of hundreds of parts is required. Finally, increased complexity frequently comes with a price tag. These fundamental challenges were a call to action for development of more efficient characterization strategies and open sharing of information and parts, which would be beneficial for the whole scientific community[24,25]. There is general agreement that the success of building processes depends on more efficient and design-driven use of predictable standards.

The engineering of biology needs standardization since it enables selection and recycling of required DNA elements from a vast list of previously characterized parts, as well as part exchange between different users. With this perception came the attempt to establish large registries of parts[26].

The NOMAD (Nucleic acid Ordered Assembly with Directionality) was probably the first attempt to standardize DNA construction and happened 20 years ago[27]. More than a decade later, the Registry of Standard Biological Parts appeared as a joint effort involving the International Genetically Engineered Machine (iGEM)

student competition (http://www.igem.org) and the BioBricks Foundation[28]. It is still widely used by the scientific community, despite several discrepancies being reported in quality and published sequences. The Standard European Vector Architecture (SEVA) platform[29–31] is yet another large repository of parts that includes broad host range origins of replication. Several other repositories were then reported like the JBEI-ICEs registry[32] or the BIOFAB collection[33,34], now available in the Addgene repository[35] that contains more than 31,000 unique plasmids available for the scientific community. More recently, the GenoLIB database[36] aimed at getting rid of problems regarding ambiguous plasmid annotations.

The open source nature of these repositories is both advantageous and limiting. The public accessibility and free sharing of information and parts enables faster, cheaper and less laborious experimental design and testing. Standardization goes together with simplicity and should be accompanied by as few rules as possible. Despite the facilitated sharing process, there are still intellectual property (IP) issues and costs for shipping and general operating maintenance. Part sharing can have additional drawbacks and users should keep in mind that there might be some lack of consistency in parts description and that unpredictable interactions between different elements may occur[25]. In other words, universality and orthogonality should not be taken for granted[37].

Predictability of parts is hence a major challenge for synbio projects. The idea that complex biological functions can be organized across different levels and conceptually separated from the original sequence context defines the use of abstraction in synbio[24]. Together with powerful software tools and programming to achieve automation, abstraction is crucial for the attempt to predict and optimize reliable part behaviour[10]. Additionally, advances in metrology, the science of measurement, will increase confidence in part characterization-associated measurements, contribute to reproducibility between different laboratories and, thus, reliable characterization of parts or standardization[38].

Overall, the green, clean and cheap vision of synbio will depend heavily on standardization and predictability of standards, which will only be accomplished with the development of DNA editing technologies and software tools, as well as full characterization and optimization of the host chosen for expression of synthetic circuits.

4. Model cell factories – chassis selection

In synbio, the chassis is the cellular container (or host) that provides structures to accommodate and execute the necessary cellular functions. The choice of chassis is thus a crucial step in bio-production of any chemical or protein, since it will influence the function and behaviour of genetic elements[39,40]. The production host, or cell factory, should be non-pathogenic, genetically stable and well-

characterized, be able to survive under the desired process conditions (such as specific pH and temperature or low nutrient requirements) and ideally possess a catalogue of available engineering tools[11]. Additionally, one should also consider whether the desired metabolic pathway exits or can be reconstructed in that particular host[41]. Several options are currently available for chassis selection[10]. However, two microbial cell factories, *Escherichia coli* and *Saccharomyces cerevisiae,* have dominated as workhorses in synbio tool development as well as bio-production of chemicals or proteins with applications in pharma, food or biotech industry.

4.1. *Escherichia coli*

The Gram negative, facultatively anaerobic, rod-shaped bacterium *E. coli* is by far the most widely studied prokaryotic model organism – and in fact probably the best studied living organism. While some strains might be pathogenic, most *E. coli* strains are part of the natural flora of the gut of warm-blooded organisms. This robust bacterium possesses a number of properties that make it an ideal candidate for metabolic engineering and synbio such as fast doubling time, metabolic versatility, well-known metabolism and genome, and available genetic tools for strain manipulation and engineering[10,41–43]. Despite being the host of choice for development of DNA engineering technologies, non-pathogenic *E. coli*'s fame is due to the considerable amount of generally recognized as safe (GRAS) recombinant pharmaceutical proteins that are being produced in this

model organism as well as bio-fuels, amino acids, sugar alcohols, diols and polymers[42].

Major shortcomings of using *E. coli* as a chassis include the inability to perform post-translational modifications such as glycosylation, common in eukaryotic proteins, the susceptibility to phage attacks and production of endotoxins[41,44–46].

4.2. *Saccharomyces cerevisiae*

The baker's yeast is the most intensively studied unicellular eukaryote and the first genetically modified organism to be approved for use in food production. As bacteria, *S. cerevisiae* is a fast growing, well-characterized organism with a vast array of engineering tools that justify its frequent use[47]. Besides being tolerant to low pH and high sugar and ethanol concentrations, *S. cerevisiae* shares more similarity with higher organisms, in terms of complexity of cellular structure, and is able to perform post-translational modifications that are required for production of most eukaryotic proteins. Furthermore, the GRAS status of both *S. cerevisiae* and yeast-derived products helped build its reputation as a host of choice[48]. Apart from centuries of use in baking, brewing, wine making and bio-ethanol production, this microorganism has proven its value as a cell factory for production of amorphadiene, vanillin, polyketides, isoprenoids, steviol components and opiates[47–49].

Synbio tools in yeast are lagging behind the *E. coli* ones and one major drawback when choosing the first chassis is the small number of inducible promoters available[47,50,51].

4.3. Alternative chassis

There is an obvious trade-off between chassis selection and successful expression of synthetic systems. When special traits are required for production or specific technology development, alternative hosts with attractive peculiarities might be advantageous over the established model cell factories. Currently, there is a list of organisms that goes from bacteria and yeast to insect and mammalian cells, as well as cell-free systems that in specific situations offer advantages to the previously described model cell factories[52,53]. Two of them deserve special attention, *Bacillus subtilis* and *Pseudomonas putida,* mainly due to their recent contributions to the synbio field[54].

The model Gram-positive *B. subtilis* is also an extensively studied and used GRAS organism with a diverse molecular biology toolkit available. Apart from rapid and inexpensive growth, the most appealing features of this host are the heat-resistance spore formation and secretion of proteins, explaining why it is frequently chosen for recombinant protein production[55–57]. This organism has also shown its potential when handling large DNA construct assembly[58]. Limitations on its use are related to plasmid stability and very active proteolytic degradation systems[10].

One example of non-established but emerging chassis is *P. putida.* It emerged some years ago as an alternative to the previously mentioned chassis. This soil and plant-root associated, ubiquitous and saprophytic Gram-negative bacteria has gained some attention due to its metabolic versatility and tolerance to many xenobiotic compounds as well as pH and temperature variations[59,60]. Its potential for applications like industrial biocatalysis, bioplastic production or *in situ* bioremediation resulted in rapid increase in reports with the aim of expanding the tools available for *P. putida* engineering[29–31,61]. As for all emerging or potential chassis, the major shortcomings of choosing *P. putida* are related to the still limited synbio toolkit.

Despite the potential of alternative (established or emerging) chassis, *E. coli* and *S. cerevisiae* are still the preferred choices. Both systems have advantages and limitations in their use and its common to combine them even in simple projects. *E. coli* serves as cloning host, while *S. cerevisiae* is responsible for production. A recent trend has taken advantage of yeast's homologous recombination and ability to handle large DNA fragments, where big-sized constructs are assembled in yeast and transferred to *E. coli* for expression and production[62].

5. DNA assembly

Assembly of DNA is at the core of synbio and biotechnology[11,63,64]. Consequently, advances in DNA editing technologies have been game changing for most biological research projects[63]. Tailoring of parts into genes, genes into pathways and ultimately whole genomes would not be possible without the tools of molecular biology developed over the past half century such as restriction enzymes, homologous recombination, PCR and more recently CRISPR/Cas.

Currently, the steadily decreasing price of synthetic DNA is increasingly allowing for outsourcing DNA editing[65]. However DNA synthesis does not prevent the need for DNA assembly (even for gene synthesis companies) and synbio still relies on effective methods to facilitate exchange of individual parts and minimize *de novo* design.

The increased ability to synthetize/PCR amplify different parts stretched our boundaries and is allowing for challenging DNA assemblies[58,62,66,67]. Hence, considerable efforts have been devoted to the development of new DNA assembly methods or improvement of the existing toolbox to achieve modularity, recycling of pre-existing parts and simplicity at a cheaper price and in a less laborious fashion[63,68–70]. Consequently, the growing throughput of assembly projects does not call exclusively for new improved DNA assembly technologies but also for new software packages that allow researchers to cope with increasing complexity and scale[71]. *In vitro* (Figure 2A) and *in vivo* (Figure 2B) DNA

assembly technologies are herein reviewed (summarized in Table 1) and grouped according to the strategy employed. The software tools that support these DNA assembly approaches will be further described and can also be found in Table 1. Genome engineering strategies are based on editing of existing sequences instead of combining parts together like DNA assembly technologies[68]. Thus, genome-based approaches will not be reviewed in this section, except when in combination with DNA assembly strategies.

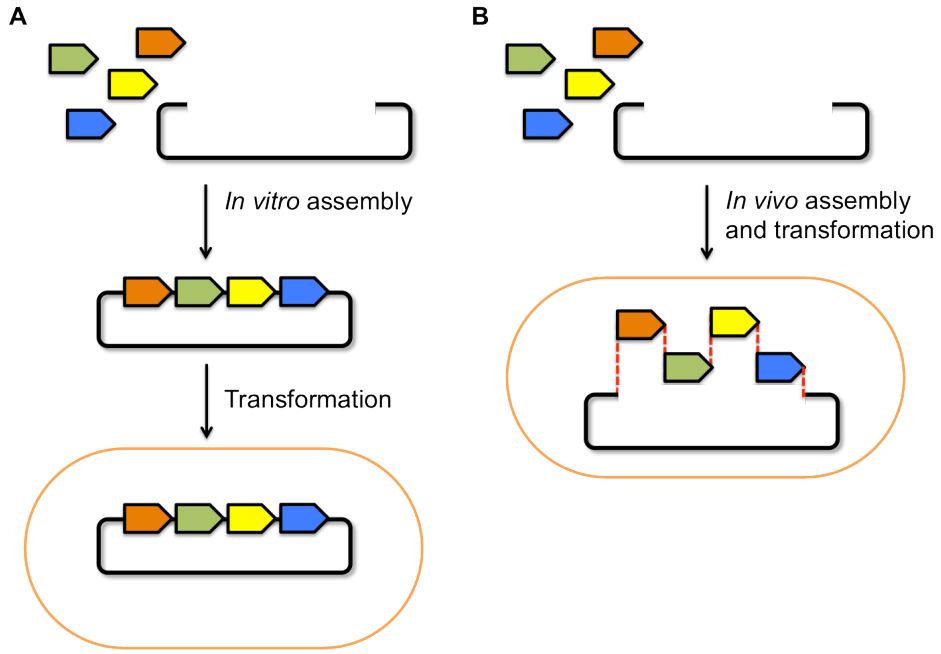


Figure 2 – Schematic representation of (A) *in vitro* and (B) *in vivo* five-fragment DNA assembly of a four-gene pathway (colourful shapes) and backbone (black line), and transformation into the host cell (orange shape).

Table 1 - DNA assembly tools and supporting software.

Strategy	Mechanism	Sequence independency	Specific requirements/ comments	Supporting Software
BioBricks[72,73]	Type IIP RE	No	Restriction site at the ends; forbidden sites within parts	Registry of standard biological parts[28]; Raven[71]
BglBricks[74]	Type IIP RE	No	Restriction site at the ends; forbidden sites within parts	Under development
BioScaffold[75]	Type IIB RE	No	Restriction site at the ends; forbidden sites within parts	-
Golden Gate[76,77]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts	J5[78,79]
PSA[80]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts; attachment tags	-
GoldenBraid[8,81,82]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts	-
MoClo[83]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts	Raven[71]
mMoClo[84]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts	-
GreenGate[85]	Type IIS RE	No	Restriction site at the ends; forbidden sites within parts; methylated sites to prevent restriction	-

MASTER[86]	Type IIM RE	No	Methylated restriction site at the ends; forbidden sites within parts	-
SEVA[29–31]	RE	No	Restriction site at the ends; forbidden sites within parts	SEVA-DB platform[31]
iBrick[87]	RE	No	Homing endonucleases restriction site at the ends	-
Gateway®[88,89]	SSR	No	λ recombination *att* sites	-
SIRA[90]	SSR	No	φC31 recombination *att* sites	Under development[90]
SSRTA[91]	SSR		ϕBT1 recombination *att* sites	-
OE-PCR[92,93]	Overlap	Yes	Inefficient for long constructs	-
CPEC[94,95]	Overlap	Yes	Inefficient for short sequences	J5[78,79]
LIC-PCR[96]	Overlap	No	12 bp overhang with no predetermined dNTP	-
SLIC[97]	Overlap	Yes	Inefficient for short sequences	J5[78,79]
SLiCE[98]	Overlap	Yes	Inefficient for short sequences	J5[78,79]
Uracil excision[99–103]	Overlap	Yes	AN_xT sequence, incorporated uracil	PHUSER[104,105] AMUSER[106]
PLICing[107]	Overlap	Yes	Phosphotioate bonds, iodine solution	-
In Fusion®[108–110]	Overlap	Yes	Inefficient for short sequences	Convert PCR Primers Into In-Fusion ® Primers website*

Gibson Assembly[111–113]	Overlap	Yes	Inefficient for short sequences	J5[78,79]; Raven[71]
LCR[114]	Bridging oligos	Yes	5' phosphorylation; bridging oligonucleotides	Gene2Oligo[115]
PaperClip[116]	Bridging oligos	No	Half clips with GCC/GGC at 5'end; 5' phosphorylation; bridging oligonucleotides	-
MODAL[117]	Mixed	No	45 bp linker regions plus 15 bp adapter sequences	R2oDNA Designer[118]
BASIC[119]	Mixed	No	Restriction site at the ends; forbidden sites within parts	R2oDNA Designer[118]
HVAS[120]	Mixed	No	λ recombination *att* sites; homing endonucleases restriction site	-
Guye *et al*[121]	Mixed	No	λ recombination *att* sites; homing endonucleases restriction site	-
Torella *et al*[122]	Mixed	No	Restriction site at the ends; forbidden sites within parts	-

*http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Cloning_Kits/xxclt_onlineToolsLoad.jsp?citemId=http://bioinfo.clontech.com/infusion/convertPcrPrimersInit.do&xxheight=750

5.1. Restriction enzyme-based approaches

Restriction enzymes (RE) cut DNA at (or near) a specific recognition site and, like CRISPR, originate from bacterial defence systems against invading

nucleotide sequences. More than 40 years ago, the discovery of restriction digestion and ligation revolutionized the field of molecular biology[12]; so much that, in 1978, a Nobel Prize was jointly awarded to Werner Arber, Dan Nathans and Hamilton Smith for the discovery of "restriction enzymes and their application to problems in molecular genetics". Since then, several methods have been developed using RE.

There are four types (type I, II, III and IV) of RE classified based on cleavage position, sequence specificity, co-factor requirements and subunit composition. Type II RE are a more heterogeneous group that, in general, cut DNA within or close to restriction sites, require Mg^{+} and have only endonuclease activity[123]. They are also the only group used in laboratorial routine DNA cloning and analysis.

The main requirement relies on restriction sites exclusively flanking the parts to be joined that are recognized by specific RE. Restriction sites are thus "forbidden sites" within the sequences to be assembled. RE-based methods are still popular nowadays due to robustness and long use, in spite of the inherent constraints by the sequence context ("forbidden sites") and the ligation step that is generally renowned to be inefficient[68].

By today's criteria, modularity, recycling of parts and standardization become a crucial part in complex engineering projects. The first step towards modularity was accomplished with the BioBrick standard[72,73], where BioBricks refer to

standardized biological components. This lego-like concept requires standardized restriction sites that compose the so-called prefix (*Eco*RI *Xba*I) and suffix (*Spe*I *Pst*I) that flank each BioBrick. Digestion and ligation allows for assembly of two parts per reaction. Continuous addition of BioBricks is possible, and several enzymes might be used to remove sites, produce compatible ends or adapt to different cloning strategies. However, the biggest shortcomings of this approach are the presence of scars that might affect gene expression and its limitation in terms of multipart assembly[72]. Improvements to the method towards a more flexible setup include the so-called BglBricks[74] that introduced a six bp scar (using *Bgl*II and *Bam*HI) encoding a glycine-serine innocuous peptide linker for most protein fusions. Shortly afterwards the BioScaffold[75] overcame the issue of scars or insertion of additional inserts previously incompatible with the BioBrick standard assembly resorting to *Prs*I, a type IIB RE that cleaves both sides of target DNA sequence in both strands.

Overall, the previously mentioned strategies still require the removal of excessive restriction sites from the parts to be joined prior to assembly. Homing endonucleases offer an alternative solution to this problem since these RE recognize long sequences (12 - 40 bp) with extremely rare restriction sites[124,125]. These sequences are also non-palindromic allowing for directional cloning[125]. Recently, the iBrick standard was developed using two homing endonucleases that recognize >18 bp DNA sequences[87]. Notwithstanding the advantage of

enabling assembly of large sized parts, a 21 bp scar is created between each part, which might affect gene expression.

Nowadays, modularity and simplicity are key elements to success in complex assembly projects. The Standard European Vector Architecture (SEVA) platform[29,31] offers a new level of parts standardization with regards to expansion of the available toolbox from enteric bacteria to other hosts like *P. putida*[29] or Mycobacteria[126]. It consists of three interchangeable modules: (1) antibiotic resistance markers, (2) origins of replication and (3) the cargo, which generally contains the genetic elements necessary for the end-application of the cell factory (very often a promoter driving expression of a gene), flanked by the rare RE sites *Pac*I, *Spe*I, *Swa*I, *Psh*AI, *Fse*I, *Asc*I (rare because they recognize a sequence not frequently found in natural DNA sequences). This system currently allows for 54 combinations of nine origins of replications with six antibiotic resistance markers[31]. Recently, the gadget, a fourth module encoding for the *hok/sok* system that ensures killing of cells that lose the plasmid, was included, with the additional introduction of a *San*DI rare site[30]. Apart from supporting modularity and promoting standardization, the SEVA platform has some limitations related to the constraints of restriction enzyme-based assembly. To overcome this limitation, the SEVA linker backbone exchange standard (Paper III in this thesis), was developed to be compatible with rare cutters but also nicking enzymes.

5.1.1. Nicking endonucleases

Acting usually as homodimers, restriction enzymes perform site-specific double-stranded breaks via hydrolysis of both phosphodiester bonds of double-stranded DNA (dsDNA)[127]. So-called nicking enzymes (NE), like other RE, recognize short specific DNA sequences and cleave DNA at a defined position; however, unlike normal REs, NEs cleave only one predetermined DNA strand in the dsDNA[128]. Suggestions that NE are naturally mutated restriction enzymes that do not possess dimerization ability might explain their origin[129,130].

The discovery of NE happened almost a decade ago[128,131]. Nevertheless, the limiting number of available NE was an impelling cause that led to protein engineering efforts to create new NE[130,132]. Since then, this group of enzymes has been extensively used either in ligation-independent cloning methods or for the generation of compatible overhangs that can be coupled with another DNA assembly technology[133–136]. The biggest advantages of these NE-based methods when compared to normal RE-based approaches lie in their compatibility with other cloning strategies to create compatible overhangs and the fact that nicked DNA can endure after ligation or transformation.

5.1.2. Type IIS restriction enzymes

The use of RE experienced a renaissance with the type IIS-based methodologies. Type IIS RE are a class of endonucleases that cleaves outside of its recognition site generating variable overhang sequences[76]. The biggest advantage of using

type IIS RE comes from their ability to assemble multiple DNA fragments in a defined linear order, since the same enzyme can generate different overhangs. Moreover, the freedom to choose the overhang sequence offers the possibility for scarless DNA assembly. Seamless assembly is accomplished by flanking the DNA parts with recognition sites in inverse orientation, which are then removed during the cleavage process.

Previously, although not used routinely in cloning, type IIS RE have been used for assembling a 32 kb cluster[137] or cloning repetitive sequences[138,139]. Golden Gate cloning[76,77] has gained popularity because it exploits the advantages of type IIS RE-based DNA assembly, particularly with *Bsa*I (Figure 3). Golden Gate parts, generated by either DNA synthesis or PCR, are sub-cloned into entry vectors, which carry the DNA segment to be transferred, before digested and ligated into destination vectors, into which the DNA will be sub-cloned[76,77]. What makes this strategy so simple is that several plasmids can be directly mixed into one single tube, together with one single RE that will create different overhangs and a DNA ligase that will join all parts. This mixture can be directly transformed into a cloning host without additional steps and screening of desired clones is accomplished by using the antibiotic to which correct transformants will be resistant to.

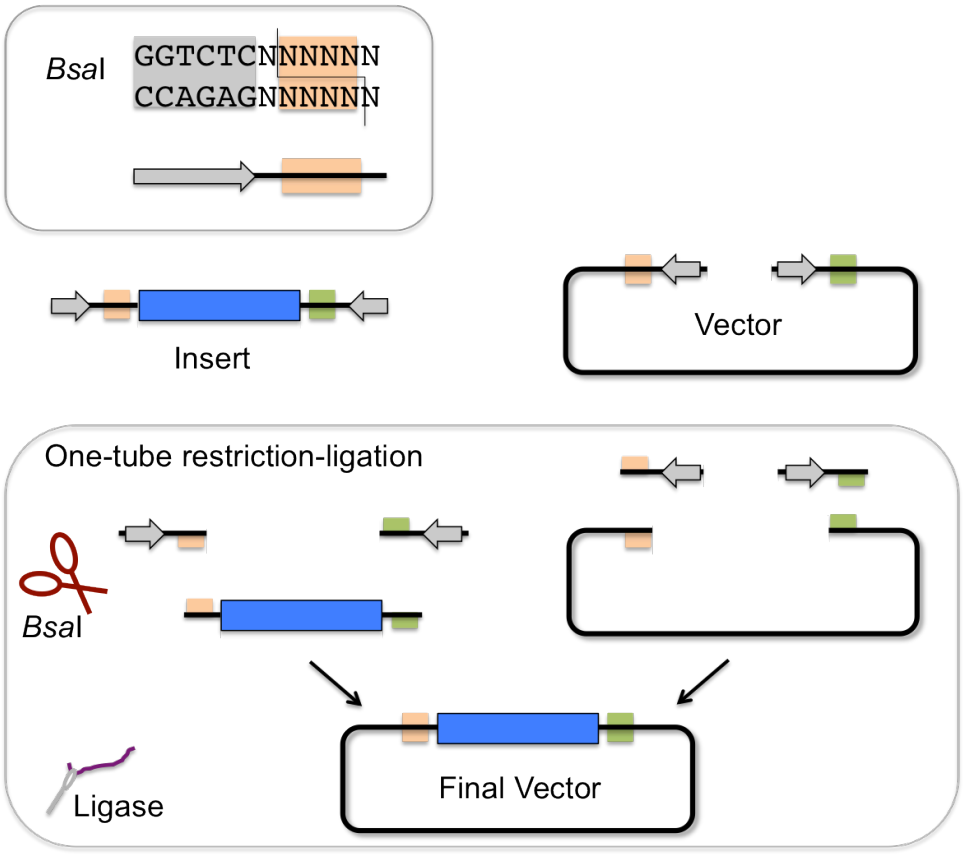


Figure 3 – Overview of Golden Gate cloning. PCR-amplified parts (insert and vector) with *Bsa*I-recognition sites (grey arrows) flanking each part in reverse orientation are mixed together with *Bsa*I and DNA ligase. *Bsa*I cleavage allows the formation of different overhangs (orange and green) that are four bp in length and complementary with the corresponding end in the other part (in this case, complementary overhangs are represented in the same colour). After restriction digestion, the DNA ligase, present in the mixture, ensures complementary overhangs are stitch together. This strategy is repeated for several cycles of restriction-ligation in a one-tube reaction. Final vectors are not subjected to further restriction digestion since they do not possess a *Bsa*I-restriction site. For

re-use of parts, inserts are usually cloned into an entry vector first and then mixed with a destination vector that possesses a different antibiotic resistance marker.

Golden Gate cloning was described eight years ago and, since then, several improvements have been made to the technology. The first appeared with the Modular Cloning System (MoClo)[83] where a 33 kb construct containing 11 eukaryotic transcriptional units (DNA sequences that encode a single RNA molecule together with sequences required for its transcription - usually promoter, open reading frame and terminator) was assembled from 44 modules in only three successive cloning steps. It was further extended to a 50 kb construct corresponding to 68 DNA fragments and 17 eukaryotic transcriptional units[140]. Mammalian Modular Cloning (mMoClo)[84] has also been described, where six transcriptional units corresponding to 27 kb and 42 parts were chromosomally integrated. The shortcomings of these optimizations are the large number of entry and destination vectors. The GoldenBraid standard[81] and its optimized version GoldenBraid 2.0[82], other improvements to the Golden Gate cloning, overcome this issue with the reduction of the number of required vectors; however, additional rounds of assembly are required for larger constructs. Other reports on Golden Gate cloning developments have been reported both for yeast and plants[141–144].

The Pairwise Selection Assembly (PSA)[80] is a related strategy using type IIS RE that was published shortly after Golden Gate cloning. The premises are similar to

Golden Gate cloning; however, PSA requires the use of two divergently oriented non-functional antibiotic resistance markers and the attachment of corresponding recyclable activation tags flanking the fragments[80]. Activation flags contain type IIS recognition sequences that allow for cloning in the next round of DNA assembly, besides allowing for screening of positive transformants due to the encoded antibiotic resistance markers. The use of sequence-specific blocking oligonucleotides that prevent DNA methylation at desired restriction sites has solved the "forbidden site" issue. A 91 kb completely synthetic right arm of *S. cerevisiae* chromosome IX was constructed using this method[80]. Programmed DNA methylation for preventing digestion by certain RE has been described also in the Methylation-assisted Tailorable Ends Rational (MASTER)[86] ligation method and in the Green Gate system[85]. The latter still uses *Bsa*I and methylated oligonucleotides while MASTER uses *Msp*JI that shares properties from type IIM and type IIS RE, recognizing only methylation specific sites and cutting outside of its asymmetric recognition sequence, respectively.

Besides allowing for combinatorial studies, creation of libraries and assembly of large constructs, the most important benefit of choosing Golden Gate cloning and successors regards assembly of repetitive sequences as exemplified by cloning of Transcription Activator Like Effectors (TALEs)[145–148]. A major problem that has not been solved is the fact that type IIS restriction sites occur frequently in

mammalian promoters and genes[121]. Furthermore, the ligase dependency is also a limitation.

5.2. Site-specific recombination-based approaches

Various nucleotide recombination mechanisms are nature's way to embrace diversity and evolve. Enzymes involved in recombination have also been exploited for decades to artificially and rationally recombine DNA molecules, both *in vivo* and *in vitro*. The typical process of recombination involves one or more proteins that recognize and catalyse the shuffling of homologous sequences[149,150]. The homologous sequence stretches are typically in the range of 30-100 nucleotides and this can represent a methodological limitation (e.g. long oligonucleotides are more expensive, might favour secondary structures and create difficulties during PCR)[151]. On the other hand, recombination-based methods have high fidelity and efficiency and have proven superior for formation of very large constructs such as for whole genome construction[58,111].

Site-specific recombination mediated by phage integrases is conservative and highly specific since they recognize versions of attachment (*att*) sequence motifs, *attP* (P originally for phage – donor DNA) and *attB* (B for bacteria – receptor) DNA sequences found in the phage and host genomes, respectively[152]. The phage-coded integrase and the bacterial integration host factor (IHF) bind tightly to the *attP* site. The complex then couples with *attB* in the host genome. The strand exchange happens after DNA nicks are produced at the ends of the core

sequence of both *attP* and *attB* sites forming the two-hybrid *att* sites, *attL* (left) and *attR* (right). Gateway® cloning[88,89] takes advantage of a λ integrase to mediate site-specific recombination (Figure 4).

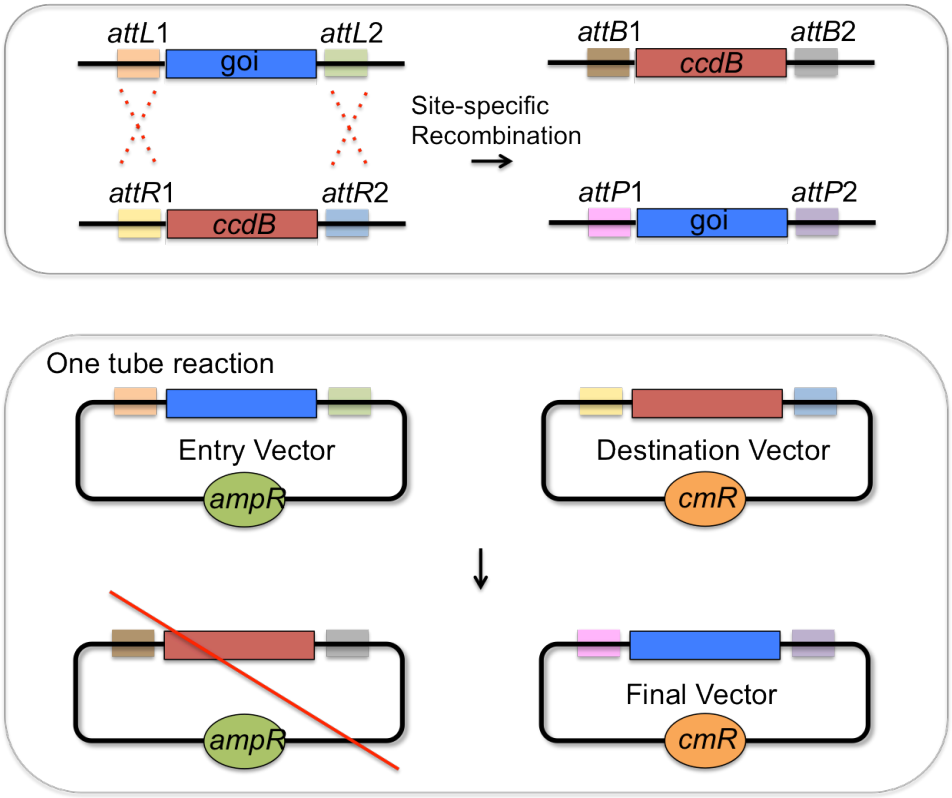


Figure 4 – Overview of the Gateway® cloning system. The gene of interest (goi - blue) is cloned into the entry vector flanked by *attL* sequences. The goi can then be transferred to the destination vector, the application-specific vector into which the DNA will be sub-cloned. This destination vector carries a *ccdB* gene (red) flanked by *attR* sequences. Both vectors are mixed, *in vitro*, with recombination proteins. Recombination of *attL*1 (light orange) with *attR*1 (yellow) and *attL*2 (light green) with *attR*2 (light blue) creates *attB*1 (brown) with *attP*1 (pink) and *attB*2 (grey) with *attP*2 (purple) sites, respectively. The product of this reaction is

the final vector (or expression clone), which carries the DNA of interest flanked by *attP* sites, and the donor vector (crossed with a red line), which contains the toxic *ccdB* gene flanked by *attB* sites. Selection of final vector after transformation is done by using the desired antibiotic (in this example, chloramphenicol – cmR in orange) and guaranteeing that cells containing the *ccdB*-encoding plasmid do not survive.

The method requires an entry clone (*attL1*-goi-*attL2*), that contains the gene of interest intended for cloning, and a destination vector (*attR1-ccdB-attR2*), in which the DNA will be sub-cloned[88]. The *ccdB* gene, encoded in the destination vector, inhibits *E. coli* growth[153,154], and together with different antibiotic resistance markers, allows for a more efficient screening[88]. This assembly strategy relies on several mutated *att* sites that exhibit high specificity and virtually no cross talk. This means *attL1* recombines only with *attR1* and not *attR2* and so on. The creation of multiple variants of the *attL*/*attR* pair has allowed for multisite gateway cloning to stich up to four DNA fragments in a single reaction[155,156].

Gateway® cloning is reliable, efficient and widely used in both prokaryotes and eukaryotes[69,157,158] and a plethora of destination vectors have been constructed by the Gateway® research community[159,160]. Being a commercial available tool means that there is a working–for–profit centre responsible for maintenance, benchmarking of protocols and dissemination of vectors, which greatly enables standardization.

The recently published Serine Integrase Recombinational Assembly (SIRA)[90] relies on ϕC31 integrase to assemble pre-designed pathways from DNA parts while Site-Specific Recombination-based Tandem Assembly (SSRTA)[91] uses an alternative integrase, ϕBT1. As Gateway® cloning, these methods depend on recombination at *att* sites that can be reversible by addition of an excisionase (a phage protein that allows for excision of DNA sequences) or by a combination of integrase and recombination directionality factor (accessory protein that reverses the reaction together with the integrase)[161]. The latter was shown to be extremely useful for re-cycling of parts in SIRA[90].

Both Gateway® cloning and SSRTA have a major disadvantage when compared to SIRA or other non site-specific recombination methods, which is the need for an entry clone that adds an extra cloning step and, hence, more complexity. Moreover, recombinase-based cloning leaves long palindromic scar sequences that might interfere with DNA integrity or gene expression. However, the high efficiency and accuracy of site-specific recombination-based methods make them a popular choice for both plasmid construction and genome engineering. A good example of site-specific recombination-based genome integration is the recently described clonetegration[162]. Clonetegration vectors were constructed with integrase-encoding gene and *attP* sequence in the same plasmid, which simplified the whole process[162]. The belief in the full potential of combining this technology with a simple DNA assembly strategy was the reasoning behind Paper I, in this

thesis. Direct genome integration of a six-fragment whole biosynthetic pathway in one-tube uracil excision reaction was described by combining uracil excision (see next section) with clonetegration. This powerful combination made the process more efficient and simple since it prevents the use of restriction-ligation or additional less efficient and straightforward methods.

5.3. Long overlap-based assembly

Some of the most popular alternatives to the hitherto described methodologies fall into the category of long overlap-based assembly. This group includes all the *in vivo* or *in vitro* DNA assembly strategies that enable the stitching of DNA parts with homologous sequences at their ends, allowing for formation of longer compatible single-stranded overhangs when compared to the short overhangs generated by restriction methods[63,68]. Homologous recombination has important advantages over site-specific recombination since no specific sequences are required. There is also the possibility of specific-order assembly in a one-pot reaction. These arguments are also the biggest disadvantages of these methods since they prevent combinatorial studies, modularity and re-usage of parts for an alternative assembly strategy, unless the overlap sequence is general[117].

The Overlap Extension Polymerase Chain Reaction (OE-PCR)[92,93] was described more than 25 years ago and is one of the first examples of overlap-depend cloning. PCR-generated fragments flanked by homologous sequences are mixed together in a second round of PCR, where they anneal and enable extension by

DNA polymerase[92,93,163]. The amplicon can then be inserted into a plasmid using another cloning strategy[92,93,163]. The Circular Polymerase Extension Cloning (CPEC)[94,95] is an improvement to OE-PCR. The homology-sharing parts anneal and circularize generating nicked sequences that are then fixed in *E. coli* after transformation. The simplicity of the methods limits their potential for more complex DNA assembly projects, mainly due to all the shortcomings that come with PCR-based methods such as being error prone (especially when exponential amplification of misannealed sequences occurs) difficulty to amplify large DNA parts and need for case-by-case optimization.

Several other methods rely on PCR for part amplification but with the additional use of alternative enzymes to create overlapping single-stranded overhangs that catalyse efficient and accurate assembly. One example is the so-called uracil excision cloning (or uracil-specific excision reagent - USER - cloning)[99–103,134,164] (Figure 5). It is based on the use of uracil-containing oligonucleotides where the uracil (U) replaces selected thymines (T) in a non-mutagenic and PCR-tolerated alternative base pairing with an adenine nucleotide on the complementary strand. This PCR-tolerance is accomplished with non proof-reading enzymes like Taq DNA polymerase or special, engineered proof-reading DNA polymerases that read through uracils such as PfuX7[165] or Phusion-U DNA polymerase (Thermo Scientific). Following PCR, uracils can be selectively removed by treatment with a uracil DNA glycosidase (UDG) that removes the uracil base and leaves a

chemically unstable phosphoribose backbone (abasic site). The lyase activity of Endonuclease VIII breaks the phosphodiester backbone at the 3′ and 5′ sides of the abasic site guaranteeing that the base-free deoxyribose is released, and creating a nick in each of the homology regions. Dissociation of the upstream sequence allows the formation of a cohesive DNA end, the U-overhang, and the consequent annealing of complementary sequences[103].

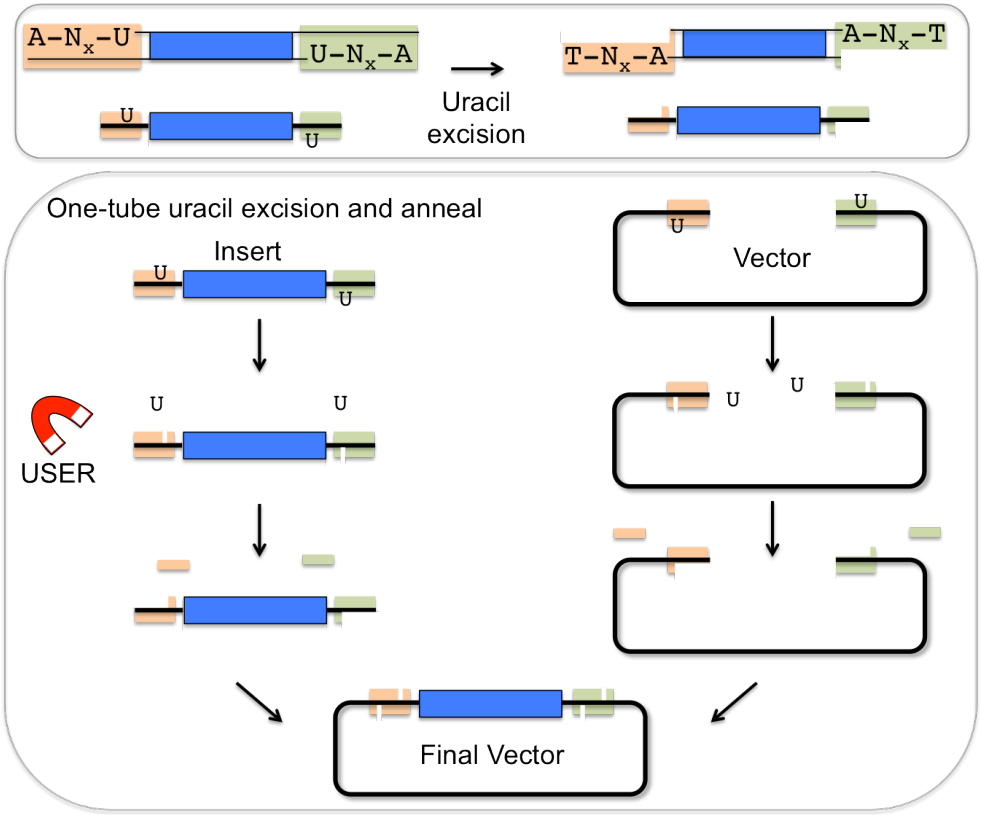


Figure 5 – Overview of uracil excision cloning. The assembly of PCR fragments (insert and vector) is mediated by 3′-TN$_x$A-overhangs (orange and green), where X typically denotes 7-12 nucleotides. The uracil in the 5′-AN$_x$U

sequence is incorporated in the PCR oligonucleotide used to amplify the parts. Uracil is selectively removed by the combined action of uracil DNA glycosylase (UDG) and the lyase activity of Endonuclease VIII (USER – represented by a red magnet), creating a nick. A 3´-TN_xA-overhang is generated when the complementary nicked sequence is dissociated, which allows for annealing of parts at proper temperature.

Papers I and II, included in this thesis, describe the optimization of this technology with increased efficiency and accuracy for both plasmid and genome-based assemblies in *E. coli* in a one-pot uracil excision reaction. Apart from *E. coli*, this technology has been also developed for Gram positives like *Lactococcus lactis*[166], yeast[167,168] and mammalian systems[169]. Despite its PCR-dependency, the most frequently pointed shortcoming of this strategy is the high cost of uracil-containing oligonucleotides, an argument that is gradually loosing its strength as the cost of these oligonucleotides is currently decreasing. Additionally, it has recently become possible to order uracil-containing synthetic genes that allow for PCR-independent assembly and overcome the PCR disadvantages (Anja Martinez, Thermo Fisher Scientific, personal communication). The drawback of this method is the requirement of at least one thymine (T) in the assembly junctions to be replaced by a uracil (U). This might be an issue specially for seamless cloning in high GC content-parts but degeneracy of the genetic code is usually a simple solution.

Alternatively, a whole different branch of overlap-based assembly strategies aims at establishing enzyme-free DNA assembly procedures. One of the most recent examples is the phosphorothioate-based ligase-independent gene cloning (PLICing)[107] that relies on the chemical cleavage reaction of phosphorothioate bonds in an iodine/ethanol alkaline solution. Homology-sharing parts are PCR amplified with oligonucleotides containing phosphorothioated nucleotides, in which phosphodiester bonds are replaced for phosphothioester bonds at the 5'-end, and then cleaved in an iodine/ethanol alkaline solution to produce the cohesive ends[107,170]. Preventing the need for enzymes might be an advantage but the cost of phosphorothiated oligonucleotides is still high and from our experience appears to be highly error prone.

"Chew-back and anneal" methods that rely on repair enzymes with exonuclease activity, such as T4 DNA polymerase, T5 exonuclease, exonuclease III or lambda exonuclease, offer another alternative to the previously described technologies. These approaches are based in single-stranded degradation of an exposed end of a part creating an overhang that anneals with a complementary sequence from another DNA part.

One of the first reports of a DNA assembly strategy exploiting the exonuclease activity was the ligation-independent cloning of PCR products (LIC-PCR)[96]. The procedure relies on the 3'- 5' exonuclease activity of T4 DNA polymerase in the

presence of a predetermined dNTP (e.g. dCTP for vector and dGTP for insert) to generate single-stranded overhangs, complementary between the parts to be joined. Different 12-nucleotide sequences without guanines (G) and with a G in the 13th position are incorporated in the oligonucleotides used to PCR-amplify and linearize the vector while complementary sequences with C in the 13th position are used for PCR-amplification of the insert. Defined sequence length is accomplished by the presence of the specific dNTP that controls exonuclease activity of T4 DNA polymerase. Moreover, the 12 nucleotide-overlapping sequence shared between the parts eliminates the need for a ligation step[96,171]. Much like uracil excision cloning, this method relies on one nucleotide, which can be viewed as a drawback. The sequence and ligation-independent cloning (SLIC)[97] is, to some extent, a variation of LIC which removes sequence constraints by generating longer single-strand overhangs (>20 bp). Overhang annealing might be accomplished with or without RecA, relying on the bacterial recombination/repair systems to complete the assembly after transformation.

The most popular methods belonging to this group of DNA assembly strategies are Gibson isothermal assembly[111–113] and the commercial In-Fusion® cloning[108–110] system from Clontech®. The first relies on the joint activity of three enzymes: T5 exonuclease to create a >20 bp overhang that allows for annealing of complementary sequences, Phusion DNA polymerase for gap filling, and Taq DNA ligase to seal the nicks. The second, despite being proprietary, probably

relies on 3'-5' exonuclease activity of poxvirus DNA polymerase and requires a 15 bp homology with a slight increase in efficiency with longer sequences[108].

Some of the most impressive DNA assembly constructs were accomplished using Gibson Isothermal Assembly[111,172] showing that "chew-back and anneal" strategies are suited for assembly of large constructs. Roughly, the overhang length can be estimated based on the duration and temperature control of the DNA treatment with DNA exonucleases[96]. However, the uncontrolled chew-back of DNA exonucleases leads to the formation of ssDNA overlaps with variable length, which is probably one of the biggest shortcoming of exonuclease-dependent strategies[97,173].

In vivo recombination-based methods have also been described and have a cost-benefit of preventing the need for commercial enzymes, similarly to enzyme-free cloning approaches like the already described PLICing. The Seamless Ligation Cloning Extract (SLiCE)[98] is one example that replaces the repair enzyme cocktail for the cellular DNA repair systems by utilizing bacterial cell extracts to assemble multiple fragments. The Single-selective-marker Recombination Assembly System (SRAS) is another example that explores the endogenous homologous recombination systems of *E. coli* to assemble parts *in vivo*[174]. However, the most successful examples take advantage of *in vivo* and native homologous recombination of *S. cerevisiae* and *B. subtilis* that are able to take up

linear DNA fragments with shared end-overlap and assemble them relying exclusively on cellular repair systems[58,111,175,176].

From a general perspective, this heterogeneous group of DNA editing tools might have some limitations inherent to the manner ssDNA overhangs are formed. Even so, the simplicity and freedom of overhang-design as well as their potential to be coupled with other strategies, make them extremely attractive and, hence, a popular choice amongst researchers.

5.4. Bridging oligonucleotide-based methods

In the early 90s, a new amplification method termed Ligase Chain Reaction was first reported[177]. The method relies on the annealing of four oligonucleotides complementary to a specific template and their ligation through the action of a thermostable ligase. Recently, a new concept arose based on the premise of the previous method in which DNA assembly and synthesis of short sequences can be accomplished simultaneously[178]. Shortly after, a very promising approach termed Ligase Cycling Reaction (LCR)[114] was described for DNA assemblies of up to 20 parts (Figure 6). It consists of a series of denaturation-annealing-ligation cycles that allow for DNA assembly via bridging oligonucleotides complementary to both ends of DNA parts that are joined by a thermostable DNA ligase.

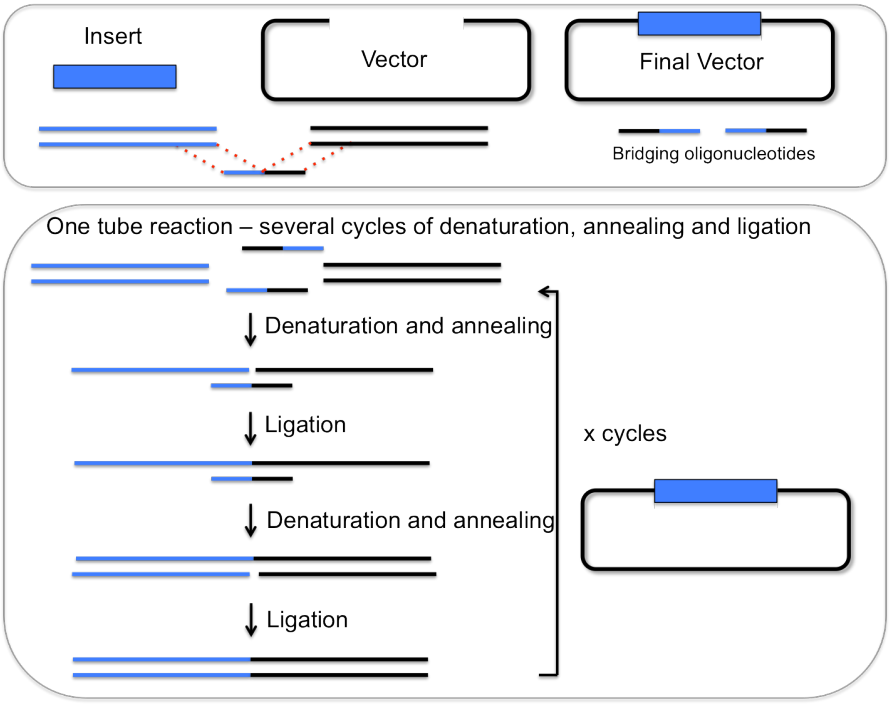


Figure 6 – Overview of Ligase Cycling Reaction (LCR). Single-stranded bridging oligonucleotides, complementary to the ends of the parts to be joined, serve as a template to bring the upper strands of denatured 5'-phosphorylated-parts together. Then, a thermostable ligase stitches the parts that serve as template for ligation of the lower strand, in the second and following cycles. Denaturation–annealing–ligation cycles allow for assembly of the two parts into the final vector. With this protocol, it is possible to assemble many parts into complex DNA constructs.

A consistent drawback of these methods lies on the fact that *de novo* design of bridging oligonucleotides might be required when changing parts. The PaperClip strategy[116] overcame this issue since it relies on double-stranded bridging

oligonucleotides, clips, that allow for multipart assembly. Each clip is designed by joining two half clips that are nothing but two annealed oligonucleotides with a GCC overhang that allows them to be ligated. Modularity is accomplished based on the order in which the clips were created and a three bp scar is created. Additionally, half clips are re-usable as long as the corresponding part is needed. Despite their potential, these methods require the 5'-phosphorylation of parts. Either purchasing phosphorylated oligonucleotides or phosphorylating in-house oligonucleotides or parts makes the process more expensive and adds some complexity to the whole assembly project.

5.5. Mixed approaches

Struggling to find the best-suited assembly strategy is common among researchers since all of the previously described methods have their limitations. The increasing complexity in assembly DNA projects gives rise to DNA assemblies that are more laborious and error prone. Hence, a recent trend is to combine two or more cloning approaches in tandem to overcome these limitations.

Several studies are resorting to linker regions that remain immutable with different assembly projects. Recently, Torella *et al.*[122] described a linker-based strategy that combines a first cloning round with restriction endonucleases and a second round with Gibson isothermal assembly. Alternatively, the Modular

Overlap-Directed Assembly with Linkers (MODAL)[117] was developed with standardized linker sequences to enable modular construction of different plasmids in a one-pot reaction that can call upon different long overlap-directed DNA assembly strategies like Gibson isothermal assembly, CPEC or *S. cerevisiae in vivo* DNA assembly. The Biopart Assembly Standard for Idempotent Cloning (BASIC)[119] is an upgrade of the MODAL strategy associated with type IIS restriction enzyme-based mechanisms that allows the orthogonal linkers to be added to DNA parts without resorting to PCR. This improvement also enables the user to choose between multipart assembly and hierarchical assembly via methylation of the linkers.

Taking advantage of linker-based assemblies, an extension to yeast Golden Gate[142] was made with the Versatile Genetic Assembly System (VEGAS)[179] where linker regions are added by yeast Golden Gate to allow for homologous recombination of several modules in yeast.

Another example of mixed approaches is the Homerun Vector Assembly System (HVAS) that combines the multisite Gateway® cloning system with homing nucleases[120]. Despite overcoming the issue of "forbidden sites" imposed by the use of most restriction enzymes, this method creates long scars that may pose a problem in gene expression. Additionally, the number of commercially available homing endonucleases is currently limiting the methodology to four modules.

Finally, HVAS still possesses the limitations of Gateway® cloning with regards to the added complexity by the need for a sub-cloning step to an entry vector.

A similar approach was developed by Guye *et al*[121] where the combination of the multisite Gateway® system and homing endonucleases with Gibson isothermal assembly allowed for modular construction of up to 11 transcriptional units in mammalian cells. The biggest advantage when comparing to HVAS is the fact that it enables further rounds of assembly.

Following the linker-based assembly trend, Paper III included in this thesis describes the SEVA linker backbone exchange standard, a SEVA-compatible strategy that enables backbone swapping with 20 combinations of classical enzymatic restriction/ligation, Gibson isothermal assembly, uracil excision cloning and nicking enzyme-based methodology. This simplistic one-tube protocol for backbone swapping directly from plasmid stock solutions (so-called SEVA cloning) is also freely available to the synbio community, promoting standardization.

Despite possible expression effects the addition of linker sequences might have in synthetic circuits, these modular strategies will most likely contribute to a more consistent part sharing and increasing complexity of combinatorial studies. The freedom to choose a favourite assembly technology while still having "back-up" technologies turns them into attractive options. Also, what these mixed approaches have in common is translated into a crescent scientific awareness that

different strategies work best at different stages of complex DNA assembly projects.

6. Software tools for DNA assembly

The constant appearance of faster and more efficient DNA assembly strategies is being mirrored by the rapid development of new software packages (or assembly algorithms) that assist with the whole planning process. Nevertheless, our lack of knowledge on important biological mechanisms and parameters is the major challenge to achieve complete automated and *in silico*-based approaches for DNA assembly experimental design and optimization[180].

Software tools available for synbio design have been reviewed elsewhere[26,181]. The synbio toolkit employs software resources for genome engineering, part visualization and circuit design, among others, but in this thesis the focus will be exclusively on the recent developments in DNA assembly planning and optimization as well as oligonucleotide design. Table 1 lists several DNA assembly strategies and the software tools available for each of them.

Several computational tools are currently available for oligonucleotide design. One example is the PHUSER (Primer Help for USER)[104,105] and its improved version AMUSER (Automated DNA modifications with USER cloning)[106] platforms that become very useful when designing oligonucleotides specifically for uracil excision cloning[99–103,134,164]. Another example is the GeneDesign 3.0[182]

that enables oligonucleotide design and sequence manipulation and is also compatible with uracil excision cloning.

Oligonucleotide design can be time-consuming, laborious and error-prone, especially when dealing with large combinatorial projects. Software-assisted experimental planning to minimize cloning rounds and human errors, recycle parts and reduce costs enables optimization of the whole process. DNA assembly strategies that employ fixed sequences on assembly regions facilitate the automated protocol design since there is no requirement for junction design or optimization. The process becomes more complex when assisting in sequence-independent DNA assembly procedures[79].

Algorithms were previously reported for automated DNA assembly with BioBrick and BglBrick technologies[180] to minimize assembly steps and improve the recycling of parts. However, the web-based computer-aided design (CAD) software J5[78,79] for DNA assembly design automation was a pioneer in assessing the cost-efficiency of DNA assembly by having DNA synthesis and in-house available parts as parameters to consider. Additionally, this tool is compatible with scarless multipart DNA assembly strategies such as the previously described Gibson isothermal assembly[111,112], CPEC[94,95], SLIC[97] and Golden Gate cloning[76,77].

Since linker-based DNA assembly technologies are in fashion, more recently, R2o DNA designer[117,118] was developed to help with the design of linker

sequences compatible with MODAL[117]. A major breakthrough in assembly algorithms emerged with Raven (http://www.ravencad.org), the software tool that allows for optimized multipart DNA assembly planning and troubleshooting for six cloning strategies[71]. It is the first interactive DNA assembly tool that recalculates alternative assembly plans if the first option fails[71]. Finally, the Clotho v2.0[183] is a database for managing all the information related to parts.

Besides the obvious advantage of helping to eliminate human error, particularly in high-throughput projects, computational methods can better deal with the inherent stochasticity and complexity of biological systems. Some of these computational methods have already been used to control robotic liquid handlers which have the capacity to carry out relatively accurate pipetting and transportation of labware as well as automate multiple complex modular assemblies[184–186]. Software tool development will enable combinatorial DNA assembly studies with unprecedented scale. Additionally, the growing synbio software toolkit also extends to computer-assisted DNA synthesis and sequencing.

7. DNA synthesis and sequencing

DNA sequencing and synthesis are out of the scope of this thesis and have been extensively reviewed elsewhere[65,187–189]. However, since DNA writing, editing and reading are inherently complementary technologies and, hence, directly

intertwined, this section will briefly sum up the current state of play on DNA synthesis and sequencing to understand how developments in both fields can affect DNA assembly and synbio.

A plethora of applications like whole-genome sequencing, metagenomics and transcriptomics, is now available thanks to Next-Generation Sequencing (NGS) that has revolutionized our ability to read DNA at exponentially decreasing prices[189]. Currently, Pacific Biosciences instruments is producing the longest reads, up to 14 kb, though with high error rates[190–192]. Highest accuracy is accomplished with the short reads produced by the Illumina/Solexa technology[193]. In spite of new sequencing technologies being in rapid development[189], our ability to write DNA is still on a less mature level[65]. Several companies are synthetizing DNA at reasonable prices and most of them are using the Polymerase Chain Assembly (PCA) strategy to do so[65,194,195]. PCA employs a thermostable DNA polymerase to assemble many oligonucleotides (with overlapping regions of 15-25 nucleotides) into longer dsDNA fragments[195]. Larger size synthetic DNA fragments however, are stitched from these smaller parts using one of the previously described DNA assembly strategies. Nowadays, DNA pieces from 200 bp to 3,000 bp are being synthetized from five to 50 oligonucleotides[65]. Until now, synthesis of *Mycoplasma genitalium* genome (582,970 bp) was the largest accomplished DNA synthesis[111].

GeneArt®, Genscript Biotech Corporation, DNA 2.0 and Blue Heron Biotechnologies Inc. dominate the gene synthesis market right now, with oligonucleotide synthesis technologies like Integrated DNA technologies (IDT) also providing the service[188]. An important upgrade was accomplished by GeneArt®, which, besides the Gibson isothermal assembly-compatible synthetic genes, is also synthetizing modified parts compatible with uracil excision cloning by incorporation of uracils (Anja Martinez, Thermo Fisher Scientific, personal communication).

The major bottleneck in the gene synthesis process is the synthesis of oligonucleotides, since it is an error-prone process[188]. The recently published Sniper cloning[196] overcomes this issue by coupling synthesis and high throughput sequencing, which enables screening of the correct DNA sequences.

Synbio projects have benefited greatly from the rapid decline of DNA synthesis cost[197]. Particularly for RE-dependent strategies, DNA synthesis offers the advantage of removing possible forbidden sites, overcoming the biggest shortcoming of these methodologies. As a whole, developments on DNA synthesis and sequencing present the opportunity to manipulate and analyse a larger set of parts, which in turn reflects on the throughput level of DNA assembly projects.

8. Applications

Applications that come with the fast development of all synbio technologies are important for pathway optimization for metabolic engineering, design and characterization of gene circuits, synthesis of whole genomes and natural product discovery. In line with this idea, Paper IV, included in this thesis, shows how software approaches can be powerful tools for natural product and enzyme discovery, particularly for a group of enzymes with biotechnological interest, the cytochromes P450 (CYPs). CYPs are a superfamily of enzymes found in many different taxonomic groups. Natural product synthesis in microorganisms, plants and fungi, includes several unique chemical reactions, some of which are catalyzed by these CYPs[198]. Despite their ubiquity and diversity in plants, the study and engineering of eukaryotic CYPs is facing several challenges. In higher organisms, CYPs are membrane associated, representing a major obstacle in expressing and purifying this class of enzymes[199]. Bacterial CYPs, although present in lower number as compared to numbers in eukaryotes, are soluble, more stable and exhibit higher activity[200,201]. Thus, they are easier to engineer, overexpress, purify and crystallize[199]. Paper IV, in this thesis, unravels putative CYP-encoding genes from the unexplored marine bacterial genomes and demonstrates that all the tested CYPs are successfully expressed in *E. coli* and *S. cerevisiae*.

The expanding synbio toolkit has allowed for certain applications such as microbial production of biofuels, biomaterials and pharmaceuticals, among others[202]. However, for specific applications like plant engineering, this toolkit is still limited[70]. A specific example that supports the relevance of CYPs as well as the usefulness of the synbio toolkit is the successful synthesis and commercial production of semi-synthetic artemisinin[203,204]. While the whole biosynthetic pathway remains unknown, the production of the intermediate artemisinic acid followed by chemical conversion to artemisinin enabled its commercialization[203,205,206].

Re-designing of biological systems is an important goal of synbio[3–5]. The synthetic yeast genome project and, specifically, the total synthesis of a functional designer chromosome (272,871 bp) of *S. cerevisiae* is probably one of the biggest accomplishments of this field[67,207]. The fact that uracil excision cloning was used to assemble the starting smaller DNA fragments demonstrates how this technology and its improvements (described in Paper I and II in this thesis) can be helpful to the synbio community[66,208].

These examples prove the value of synbio tools, in this case for production of pharmaceutical agents and re-designing new biological systems, although several others exist.

Concluding remarks and future perspectives

Synbio is a direct bridge between biology and engineering and its success relies on its multidisciplinary nature.

The unpredictability and complexity of biological systems is undoubtedly the biggest challenge that synbio has been struggling with. Engineering of even simple biological systems is still a frequently laborious process with several trial-and-error steps. Despite the efforts of the whole synbio community, standardization is also a major challenge that remains to be solved. This thesis focused on the development and improvement of strategies that could potentially contribute to standardization. Paper I describes our efforts in optimizing the uracil excision cloning approach, showing that increased efficiency (20-100%) and accuracy especially for longer U-overhangs can be obtained. The versatility of this method allows for combination with genome engineering strategies like clonetegration. With the right set of plasmids, direct genome integration of a whole biosynthetic pathway was accomplished at high accuracy in a one-pot uracil excision cloning reaction. Paper II adds to this story a detailed set of tips on how to plan and troubleshoot uracil excision-based projects. It also provides protocols for site-directed mutagenesis, multigene assembly, one-step cloning and genome integration with uracil excision, and for standardized linker-dependent uracil excision-based DNA editing pipelines. Future work on uracil excision cloning would benefit from compatible software tools like J5, to aid with the

planning process, troubleshooting and recalculating the assembly strategy in case of failure.

We took the trendy linker strategy to a new level with SEVA linkers for SEVA cloning described in Paper III. This strategy allows for direct plasmid backbone exchange and is compatible with several popular choices like uracil excision cloning, Gibson isothermal assembly, restriction-ligation and nicking enzyme-based methodology. The flexibility of the strategy and compatibility with the popular SEVA resource makes SEVA cloning an excellent choice for automated multiple complex modular assemblies. Besides associating it with robotics, future work should focus on expanding the entry and donor vector libraries, and adapt the strategy for one-pot multigene assembly.

Finally, Paper IV shows how software tools can make a difference in scientific discovery projects. In this manuscript we exploit the potential of marine bacteria for discovery of new cytochromes P450 (CYPs). This group of enzymes is responsible for challenging chemistry, making it of great biotechnological interest. Genome mining of 19 marine bacteria revealed 26 putative CYPs. In this paper, it was demonstrated that five of these enzymes can be produced in the two favourite model cell factories, *E. coli* and *S. cerevisiae*. Identifying specific functions to confirm the *in silico* predictions as well as optimizing CYP expression in yeast should be considered as future work. Additionally, engineering these enzymes to use a broader range of substrates could facilitate

production of relevant compounds.

Overall, the work described in this thesis reflects the steps taken towards the goal of making DNA assembly less laborious, more efficient and compatible with other synbio tools. Although there are still many challenges to overcome, I believe that my contribution to the synbio toolkit will benefit several research projects.

Bibliography

1. Sauer, M. & Mattanovich, D. Construction of microbial cell factories for industrial bioprocesses. *J. Chem. Technol. Biotechnol.* **87,** 445–450 (2012).

2. Anderson, J. *et al.* Engineering and ethical perspectives. *EMBO Rep.* **13,** 584–590 (2012).

3. Jefferson, C., Lentzos, F. & Marris, C. Synthetic biology and biosecurity: challenging the 'myths'. *Front. public Heal.* **2,** 1–15 (2014).

4. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10,** 410–422 (2009).

5. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go. *Cell* **157,** 151–161 (2014).

6. Arpino, J. A. J. *et al.* Tuning the dials of Synthetic Biology. *Microbiology* **159,** 1236–1253 (2013).

7. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat. Publ. Gr.* **15,** 95–107 (2014).

8. Yeh, B. J. & Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **3,** 521–525 (2007).

9. Landrain, T., Meyer, M., Perez, A. M. & Sussan, R. Do-it-yourself biology: challenges and promises for an open science and technology movement. *Syst. Synth. Biol.* **7,** 115–126 (2013).

10. Kelwick, R., Macdonald, J. T., Webb, A. J. & Freemont, P. Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* **2,** 1–23 (2014).

11. Keasling, J. Synthetic Biology for Synthetic Chemistry. *ACS Chem. Biol.* **3,** 64–76 (2008).

12. Cohen, S. N., Chang, a C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70,** 3240–3244 (1973).

13. Jackson, D. A., Symonst, R. H. & Berg, P. Biochemical Method for Inserting New Genetic Information into DNA of. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2904–2909 (1972).

14. Mertz, J. E. & Davis, R. W. Cleavage of DNA by RI Restriction Endonuclease Generates Cohesive Ends. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 3370–3374 (1972).

15. Lobban, P. E. & Kaiser, A. D. Enzymatic End-to-end Joining of DNA Molecules. *J. Mol. Biol.* **78,** 453–471 (1973).

16. Johnson, I. S. Human Insulin from Recombinant DNA Technology. *Science (80-.).* **219,** 632–637 (1983).

17. Jansen, R., Embden, J. D. A. Van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

18. Mojica, F. J. M., Díez-Villasnor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **60,** 174–182 (2005).

19. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9,** 467–477 (2011).

20. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

21. Mali, P. *et al.* RNA-guided Human Genome Engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

22. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339,** 819–823 (2013).

23. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

24. Endy, D. Foundations for engineering biology. *Nat. Rev.* **438,** 449–453 (2005).

25. Arkin, A. Setting the standard in synthetic biology. *Nat. Biotechnol.* **26,** 771–774 (2008).

26. McArthur Iv, G. H., Nanjannavar, P. P., Miller, E. H. & Fong, S. S. Integrative metabolic engineering. *bioengineering* **2,** 93–103 (2015).

27. Rebatchouk, D., Daraselia, N. & Naritat, J. NOMAD: A versatile strategy for in vitro DNA manipulation applied. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10891–10896 (1996).

28. Smolke, C. D. Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.* **27,** 1099–1102 (2009).

29. Durante-Rodríguez, G., de Lorenzo, V. & Martínez-Garcia, E. The Standard European Vector Architecture (SEVA) Plasmid Toolkit. *Methods Mol. Biol.* **1149,** 469–478 (2014).

30. Martinez-Garcia, E., Aparicio, T., Goni-Moreno, a., Fraile, S. & de Lorenzo, V.

SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43,** D1183–D1189 (2014).

31. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41,** 666–675 (2013).

32. Ham, T. S. *et al.* Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* **40,** 1–8 (2012).

33. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).

34. Mutalik, V. K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **10,** 347–353 (2013).

35. Kamens, J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.* **43,** 1–6 (2014).

36. Adames, N. R. *et al.* GenoLIB: a database of biological parts derived from a library of common plasmid features. *Nucleic Acids Res.* **43,** 4823–4832 (2015).

37. Vilanova, C. *et al.* Standards not that standard. *J. Biol. Eng.* **9,** 1–4 (2015).

38. Plant, A. L., Locascio, L. E., May, W. E. & Gallagher, P. D. Improved reproducibility by assuring confidence in measurements in biomedical research. *Nat. Publ. Gr.* **11,** 895–898 (2014).

39. Acevedo-rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W. & Danchin, A. From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet.* **29,** 273–279 (2013).

40. Danchin, A. & Sekowska, A. The logic of metabolism and its fuzzy consequences. *Environ. Microbiol.* **16,** 19–28 (2014).

41. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science (80-.).* **330,** 1355–1358 (2010).

42. Chen, X. *et al.* Metabolic engineering of Escherichia coli: A sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* **31,** 1200–1223 (2013).

43. Wright, O., Delmans, M., Stan, G. & Ellis, T. GeneGuard: A Modular Plasmid System Designed for Biosafety. *ACS Synth. Biol.* **4,** 307–316 (2015).

44. Jenkins, N. Modifications of therapeutic proteins: challenges and prospects. *Cytotechnology* **53,** 121–125 (2007).

45. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and

higher organisms. *Biotechnol. Adv.* **27,** 297–306 (2009).

46. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71,** 635–700 (2002).

47. Kavšček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. Yeast as a cell factory: current state and perspectives. *Microb. Cell Fact.* **14,** 94 (2015).

48. Nevoigt, E. Progress in Metabolic Engineering of Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **72,** 379–412 (2008).

49. Li, M. & Borodina, I. Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae. *FEMS Yeast Res.* **15,** 1–12 (2015).

50. Blount, B. A., Weenink, T. & Ellis, T. Construction of synthetic regulatory networks in yeast. *FEBS Lett.* **586,** 2112–2121 (2012).

51. Weinhandl, K., Winkler, M., Glieder, A. & Camattari, A. Carbon source dependent promoters in yeasts. *Microb. Cell Fact.* **13,** 1–17 (2014).

52. Hodgman, C. E. & Jewett, M. C. Cell-free synthetic biology: Thinking outside the cell. *Metab. Eng.* **14,** 261–269 (2012).

53. Kelwick, R., Macdonald, J. T., Webb, A. J. & Freemont, P. Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* **2,** 1–23 (2014).

54. Troeschel, S. C. *et al.* Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. *J. Biotechnol.* **161,** 71–79 (2012).

55. Schallmey, M., Singh, A. & Ward, O. P. Developments in the use of Bacillus species for industrial production. *Can. J. Microbiol.* **17,** 1–17 (2004).

56. Westers, L., Westers, H. & Quax, W. J. Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochem. Biophys. Acta* **1694,** 299–310 (2004).

57. Li, W., Zhou, X. & Lu, P. Bottlenecks in the expression and secretion of heterologous proteins in Bacillus subtilis. *Res. Microbiol.* **155,** 605–610 (2004).

58. Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat. Methods* **5,** 41–43 (2008).

59. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ. Microbiol.* **4,** 799–808 (2002).

60. Jiménez, J. I., Miñambres, B., Luis, J. & Díaz, E. Genomic analysis of the aromatic

catabolic pathways from Pseudomonas putida KT2440. *Environ. Microbiol.* **4,** 824–841 (2002).

61. Martínez-garcía, E. & Lorenzo, V. De. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440. *Environ. Microbiol.* **13,** 2702–2716 (2011).

62. Merryman, C. & Gibson, D. G. Methods and applications for assembling large DNA constructs. *Metab. Eng.* **14,** 196–204 (2012).

63. Chao, R., Yuan, Y. & Zhao, H. Recent advances in DNA assembly technologies. *FEMS Yeast Res.* 1–9 (2014). doi:10.1111/1567-1364.12171

64. Cobb, R. E., Ning, J. C. & Zhao, H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **41,** 469–477 (2014).

65. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11,** 499–507 (2014).

66. Muller, H. *et al.* Assembling Large DNA Segments in Yeast. *Methods Mol. Biol.* **852,** 133–150 (2012).

67. Annaluru, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science (80-.).* **344,** 55–58 (2014).

68. Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* **9,** 1–9 (2015).

69. Marsischky, G. & LaBaer, J. Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res.* **14,** 2020–8 (2004).

70. Liu, W., Yuan, J. S. & Stewart, C. N. Advanced genetic tools for plant biotechnology. *Nat. Rev. Genet.* **14,** 781–93 (2013).

71. Appleton, E., Tao, J., Haddock, T. & Densmore, D. Interactive assembly algorithms for molecular cloning. *Nat. Methods* (2014). doi:10.1038/nmeth.2939

72. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2,** 5 (2008).

73. Phillips, I. & Silver, P. A new biobrick assembly strategy designed for facile protein engineering. *DSpace@MIT* 1–6 (2006). doi:http://hdl.handle.net/1721.1/32535

74. Anderson, J. C. *et al.* BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng.* **4,** 1 (2010).

75. Norville, J. E. *et al.* Introduction of customized inserts for streamlined assembly and

optimization of BioBrick synthetic genetic circuits. *J. Biol. Eng.* **4,** 17 (2010).

76. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3,** e3647 (2008).

77. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4,** e5553 (2009).

78. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA Assembly Design Automation Software. *ACS Synth. Biol.* **1,** 14–21 (2012).

79. Hillson, N. J. j5 DNA Assembly Design Automation. *Methods Mol. Biol.* **1116,** 245–269 (2014).

80. Blake, W. J. *et al.* Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Res.* **38,** 2594–2602 (2010).

81. Sarrion-Perdigones, A. *et al.* GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* **6,** e21622 (2011).

82. Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162,** 1618–31 (2013).

83. 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).

84. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42,** 13440–13451 (2014).

85. Lampropoulos, A. *et al.* GreenGate - A novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One* **8,** (2013).

86. 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 Res.* **41,** (2013).

87. 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).

88. Hartley, J. L. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.* **10,** 1788–1795 (2000).

89. Walhout, A. J. M. *et al.* Gateway Recombinational Cloning: Application to the Cloning of Large Numbers of Open Reading Frames or ORFeomes. *Methods Enzymol.* **328,** 575–592 (2000).

90. Colloms, S. D. *et al.* Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* 1–10 (2013). doi:10.1093/nar/gkt1101

91. Zhang, L., Zhao, G. & Ding, X. Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci. Rep.* **1,** 1–5 (2011).

92. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77,** 61–68 (1989).

93. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59 (1989).

94. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* **4,** e6441 (2009).

95. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6,** 242–51 (2011).

96. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18,** 6069–74 (1990).

97. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4,** 251–256 (2007).

98. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40,** e55 (2012).

99. Nour-eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34,** (2006).

100. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T. & Halkier, B. a. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* **35,** e55 (2007).

101. Nour-eldin, H. H., Geu-flores, F. & Halkier, B. A. USER Cloning and USER Fusion: The Ideal Cloning Techniques for Small and Big Laboratories. *Methods Mol. Biol.* **643,** 185–200 (2010).

102. Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppälä, S. & Nørholm, M. H. H. Uracil excision for assembly of complex pathways. *Springer Protoc. Handbooks* (2015). doi:10.1007/8623

103. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H.

Accurate DNA assembly and genome engineering with optimized uracil excision cloning. *ACS Synth. Biol.* **4,** 1042–1046 (2015).

104. Olsen, L. R. *et al.* PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. *Nucleic Acids Res.* **39,** W61–7 (2011).

105. Salomonsen, B., Mortensen, U. H. & Halkier, B. A. USER-Derived Cloning Methods and Their Primer Design. *Methods Mol. Biol.* **1116,** 59–72 (2014).

106. Genee, H. J. *et al.* Software-Supported USER Cloning Strategies for Site-Directed Mutagenesis and DNA Assembly. (2014).

107. Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J. & Schwaneberg, U. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal. Biochem.* **406,** 141–6 (2010).

108. Zhu, B., Cai, G., Hall, E. O. & Freeman, G. J. In FusionTM assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* **43,** 354–359 (2007).

109. Sleight, S. C. & Sauro, H. M. BioBrick Assembly Using the In-Fusion PCR cloning Kit. *Methods Mol. Biol.* **1073,** 19–30 (2013).

110. Sleight, S. C., Bartley, B. a, Lieviant, J. a & Sauro, H. M. In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Res.* **38,** 2624–36 (2010).

111. Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20404–9 (2008).

112. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 12–16 (2009).

113. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* **37,** 6984–90 (2009).

114. Kok, S. De *et al.* Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* **3,** 97–106 (2014).

115. Rouillard, J., Lee, W., Truan, G., Gao, X. & Zhou, X. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res.* **32,** 176–180 (2004).

116. Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, a. & French, C. E. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res.* **42,** e154–e154 (2014).

117. Casini, A. *et al.* One-pot DNA construction for synthetic biology: The Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res.* **42,** 1–

13 (2014).

118. Casini, A. *et al.* R2oDNA Designer: Computational Design of Biologically Neutral Synthetic DNA Sequences. *ACS Synth. Biol.* **3,** 525–8 (2014).

119. Storch, M. *et al.* BASIC: a new Biopart Assembly Standard provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.* **4,** 781–787 (2014).

120. Li, M. V. *et al.* HomeRun vector assembly system: A flexible and standardized cloning system for assembly of multi-modular DNA constructs. *PLoS One* **9,** (2014).

121. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41,** 3–8 (2013).

122. Torella, J. P. *et al.* Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res.* **42,** 681–689 (2014).

123. Smith, H. O. & Nathans, D. A Suggested Nomenclature for Bacterial Host Modikation and Restriction Systems and their Enzymes. *J. Mol. Biol.* **81,** 419–423 (1973).

124. Marcaida, M. J., Muñoz, I. G., Blanco, F. J., Prieto, J. & Montoya, G. Homing endonucleases: From basics to therapeutic applications. *Cell. Mol. Life Sci.* **67,** 727–748 (2010).

125. Belfort, M. & Roberts, R. J. Homing endonucleases: Keeping the house in order. *Nucleic Acids Res.* **25,** 3379–3388 (1997).

126. García-Fernández, E., Medrano, F. J., Galán, B. & García, J. L. Deciphering the transcriptional regulation of cholesterol catabolic pathway in mycobacteria: Identification of the inducer of KstR repressor. *J. Biol. Chem.* **289,** 17576–17588 (2014).

127. Heitman, J. On the origin, structure and functions of restriction-modification enzymes. *Genet. Eng. Princ. Methods* **15,** 57–108 (1993).

128. Zheleznaya, L. A. *et al.* Nicking Endonucleases. *Biochem.* **74,** 1457–1466 (2009).

129. Higgins, L. S., Besnier, C. & Kong, H. The nicking endonuclease N.Bst NBI is closely related to Type IIs restriction endonucleases MlyI and PleI. *Nucleic Acids Res.* **29,** 2492–2501 (2001).

130. Zhu, Z., Samuelson, J. C., Zhou, J., Dore, A. & Xu, S. Engineering Strand-specific DNA Nicking Enzymes from the Type IIS Restriction Endonucleases BsaI, BsmBI and BsmAI. *J. Mol. Biol.* **337,** 573–583 (2004).

131. Xu, S. *et al.* Discovery of natural nicking endonucleases Nb.BsrDI and Nb.BtsI and engineering of top-strand nicking variants from BsrDI and BtsI. *Nucleic Acids Res.* **35,**

4608–4618 (2007).

132. Too, P. H., Zhu, Z., Chan, S. & Xu, S. Engineering Nt.BtsCI and Nb.BtsCI nicking enzymes and applications in generating long overhangs. *Nucleic Acids Res.* **38,** 1294–1303 (2010).

133. Wang, R., Shi, Z., Guo, Y., Chen, J. & Chen, G. DNA Fragments Assembly Based on Nicking Enzyme System. *PLoS One* **8,** 1–12 (2013).

134. Bitinaite, J. *et al.* USER TM friendly DNA engineering and cloning method by uracil excision. **35,** 1992–2002 (2007).

135. Yang, J., Zhang, Z., Zhang, X. a & Luo, Q. A ligation-independent cloning method using nicking DNA endonuclease. *Biotechniques* **49,** 817–21 (2010).

136. Luzzietti, N. *et al.* Efficient preparation of internally modified single-molecule constructs using nicking enzymes. *Nucleic Acids Res.* **39,** (2011).

137. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15573–15578 (2004).

138. Berger, S. L., Manrow, R. E. & Lee, H. Y. Phoenix Mutagenesis: One-step reassembly of multiply cleaved plasmids with mixtures of mutant and wild-type fragments. *Anal. Biochem.* **214,** 571–579 (1993).

139. Padgett, K. A. & Sorge, J. A. Creating seamless junctions independent of restriction sites in PCR cloning. *Gene* **168,** 31–35 (1996).

140. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden gate cloning and the MoClo system. *Bioeng. Bugs* **3,** 38–43 (2011).

141. Binder, A. *et al.* A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. *PLoS One* **9,** (2014).

142. Agmon, N. *et al.* Yeast Golden Gate (yGG) for the Efficient Assembly of S. cerevisiae Transcription Units. *ACS Synth. Biol.* **4,** 853–859 (2015).

143. Kakui, Y. *et al.* Module-based construction of plasmids for chromosomal integration of the fission yeast Schizosaccharomyces pombe. *Open Biol.* **5,** 150054 (2015).

144. Lee, M. E., Deloache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Chem. Biol.* **4,** 975–986 (2015).

145. Sakuma, T. *et al.* Efficient TALEN construction and evaluation methods for human cell and animal applications. *Genes to Cells* **18,** 315–326 (2013).

146. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39,** 1–11 (2011).

147. Morbitzer, R., Elsaesser, J., Hausner, J. & Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* **39,** 5790–5799 (2011).

148. Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLoS One* **6,** e19722 (2011).

149. West, S. C. Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.* **4,** 1–11 (2003).

150. Cox, M. M. Recombinational DNA repair of damaged replication forks in *Escherichia coli*: questions. *Annu. Rev. Genet.* **35,** 53–82 (2001).

151. Liang, X. *et al.* Recombination-Based DNA Assembly and Mutagenesis Methods for Metabolic Engineering. *Methods Mol. Biol.* **834,** 93–109 (2012).

152. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58,** 913–49 (1989).

153. Bernard, P. & Couturier, M. Cell Killing by the F Plasmid CcdB Protein Involves Poisoning of DNA-Topoisomerase II Complexes. *J. Mol. Biol.* **226,** 735–745 (1992).

154. Miki, T., Park, J. A., Nagao, K., Murayama, N. & Horiuchi, T. Control of Segregation of Chromosomal DNA by Sex Factor F in Escherichia coli Mutants of DNA Subunit A Suppress letD (ccdB) Product Growth Inhibition. *J. Mol. Biol.* **225,** 39–52 (1992).

155. Azaro, M. A. & Landy, A. The isomeric preference of Holliday junctions influences resolution bias by λ integrase. *EMBO J.* **16,** 3744–3755 (1997).

156. Sasaki, Y., Sone, T., Yoshida, S., Yahata, K. & Hotta, J. Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. *J. Biotechnol.* **107,** 233–243 (2004).

157. Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. *Yeast* **24,** 913–919 (2007).

158. Dalal, J. *et al.* A novel gateway-compatible binary vector series (PC-GW) for flexible cloning of multiple genes for genetic transformation of plants. *Plasmid* **81,** 55–62 (2015).

159. Katzen, F. Gateway® recombinational cloning: a biological operating system. *Expert Opin. Drug Discov.* **2,** 571–589 (2007).

160. Hilson, P. Cloned sequence repertoires for small- and large-scale biology. *Trends*

Plant Sci. **11,** 133–141 (2006).

161. Khaleel, T., Younger, E., Mcewan, A. R., Varghese, A. S. & Smith, M. C. M. A phage protein that binds f C31 integrase to switch its directionality. *Mol. Microbiol.* **80,** 1450–1463 (2011).

162. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. (2013).

163. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–932 (2007).

164. Bitinaite, J. & Nichols, N. M. DNA Cloning and Engineering by Uracil Excision. 1–16 (2009). doi:10.1002/0471142727.mb0321s86

165. Nørholm, M. H. H. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* **10,** 1–7 (2010).

166. Puri, P. *et al.* Lactococcus lactis YfiA is necessary and sufficient for ribosome dimerization. *Mol. Microbiol.* **91,** 394–407 (2014).

167. Jensen, N. B. *et al.* EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae. *FEMS Yeast Res.* **14,** 238–248 (2014).

168. Stovicek, V., Borja, G. M., Forster, J. & Borodina, I. EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial Saccharomyces cerevisiae strains. *J. Ind. Microbiol. Biotechnol.* **42,** 1519–1531 (2015).

169. Lund, A. M. *et al.* A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. *PLoS One* **9,** e96693 (2014).

170. Nakamaye, K. L., Gish, G., Eckstein, F. & Vosberg, H. Direct sequencing of polymerase chain reaction amplified DNA fragments through the incorporation of deoxynucleoside alpha-thiotriphosphates. *Nucleic Acids Res.* **16,** 9947–9959 (1988).

171. Aslanidis, C., de Jong, P. J. & Schmitz, G. Minimal Length Requirement of the Single-stranded Tails forLigation-independent Cloning (LIC) of PCR Products. *Genome Res.* **4,** 172–177 (1994).

172. Noskov, V. N. *et al.* Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth. Biol.* **1,** 267–73 (2012).

173. Tachibana, A. *et al.* Preparation of long sticky ends for universal ligation-independent cloning: Sequential T4 DNA polymerase treatments. *J. Biosci. Bioeng.* **107,** 668–669 (2009).

174. Shi, Z., Wedd, A. G. & Gras, S. L. Parallel In Vivo DNA Assembly by Recombination: Experimental Demonstration and Theoretical Approaches. *PLoS One* **8,** e56854 (2013).

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

176. Raymond, C. K., Pownder, T. A. & Sexson, S. L. General Method for Plasmid Construction Using Homologous Recombination. *Biotechniques* **26,** 134–141 (1999).

177. Barany, F. The ligase chain reaction in a PCR world. *PCR Methods Appl.* **1,** 5–16 (1991).

178. Paetzold, B., Carolis, C., Ferrar, T., Serrano, L. & Lluch-Senar, M. In situ overlap and sequence synthesis during DNA assembly. *ACS Synth. Biol.* **2,** 750–755 (2013).

179. Mitchell, L. A. *et al.* Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. *Nucleic Acids Res.* **43,** 6620–6630 (2015).

180. Densmore, D. *et al.* Algorithms for automated DNA assembly. *Nucleic Acids Res.* **38,** 2607–2616 (2010).

181. Macdonald, J. T. *et al.* Computational design approaches and tools for synthetic biology. *Integr. Biol.* **3,** 97–108 (2011).

182. Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D. & Bader, J. S. GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic Acids Res.* **38,** 2603–2606 (2010).

183. Xia, B. *et al. Developer's and User's Guide to Clotho v2.0: A Software Platform for the Creation of Synthetic Biological Systems*. *Methods Enzymol.* **498,** (Elsevier Inc., 2011).

184. Kong, F., Zheng, Y. F. & Chen, W. Automatic Liquid Handling for Life Sciences – A Critical Review of the Current State-of-the-Art. *Proc. 2008 IEEE* 480–486 (2009).

185. Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32,** 1241–1249 (2014).

186. Linshiz, G. *et al.* PR-PR: Cross-Platform Laboratory Automation System. *ACS Synth. Biol.* **3,** 515–524 (2014).

187. Shendure, J. & Aiden, E. L. The expanding scope of DNA sequencing. *Nat. Biotechnol.* **30,** 1084–1094 (2012).

188. Czar, M. J., Anderson, J. C., Bader, J. S. & Peccoud, J. Gene synthesis demystified. *Trends Biotechnol.* **27,** 63–72 (2008).

189. Nagarajan, N. & Pop, M. Sequence assembly demystified. *Nat. Rev. Genet.* **14,** 157–67 (2013).

190. Koren, S. *et al.* Hybrid error correction and de novo assembly of single-molecule

sequencing reads. *Nat. Biotechnol.* **30,** 693–700 (2012).

191. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science (80-.).* **323,** 133–138 (2009).

192. Ribeiro, F. J. *et al.* Finished bacterial genomes from shotgun sequence data. *Genome Res.* **22,** 2270–2277 (2012).

193. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30,** (2012).

194. Marchand, J. & Peccoud, J. Building Block Synthesis Using Polymerase Chain Assembly Method. *Methods Mol. Biol.* **852,** 3–10 (2012).

195. Mamedov, T. G., Padhye, N. V, Viljoen, H. & Subramanian, A. Rational de novo Gene Synthesis by Rapid Polymerase Chain Assembly (PCA) and Expression of Endothelial Protein-C and Thrombin receptor Genes. *J. Biotechnol.* **131,** 379–387 (2007).

196. Lee, H. *et al.* A high-throughput optomechanical retrieval method for sequence-verified clonal DNA from the NGS platform. *Nat. Commun.* **2,** 1–7 (2015).

197. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27,** 1091–1094 (2009).

198. Podust, L. M. & Sherman, D. H. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep.* **29,** 1251–66 (2012).

199. Jung, S. T., Lauchli, R. & Arnold, F. H. Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* **22,** 809–817 (2011).

200. Kelly, S. L. & Kelly, D. E. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120476 (2013).

201. Bernhardt, R. & Urlacher, V. B. Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl. Microbiol. Biotechnol.* **98,** 6185–6203 (2014).

202. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Publ. Gr.* **11,** 367–379 (2010).

203. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).

204. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Focus Synth. Biol.* **12,** 355–367 (2014).

205. Jensen, M. K. & Keasling, J. D. Recent applications of synthetic biology tools for

yeast metabolic engineering. *FEMS Yeast Res.* **15,** 1–10 (2015).

206. Kitaoka, N., Lu, X., Yang, B. & Peters, R. J. The Application of Synthetic Biology to Elucidation of Plant Mono-, Sesqui-, and Diterpenoid Metabolism. *Mol. Plant* **8,** 6–16 (2015).

207. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477,** 471–477 (2011).

208. Annaluru, N. *et al.* Assembling DNA fragments by USER Fusion. *Methods Mol. Biol.* **852,** 77–95 (2012).

Publications

Paper I

Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning

ACS SyntheticBiology

Technical Note

pubs.acs.org/synthbio

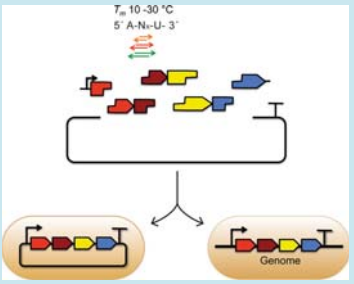
Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning

Ana Mafalda Cavaleiro, Se Hyeuk Kim, Susanna Seppälä, Morten T. Nielsen, and Morten H. H. Nørholm*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

Supporting Information

ABSTRACT: Simple and reliable DNA editing by uracil excision (a.k.a. USER cloning) has been described by several research groups, but the optimal design of cohesive DNA ends for multigene assembly remains elusive. Here, we use two model constructs based on expression of *gfp* and a four-gene pathway that produces β-carotene to optimize assembly junctions and the uracil excision protocol. By combining uracil excision cloning with a genomic integration technology, we demonstrate that up to six DNA fragments can be assembled in a one-tube reaction for direct genome integration with high accuracy, greatly facilitating the advanced engineering of robust cell factories.



KEYWORDS: *molecular cloning, DNA assembly, uracil excision cloning, genome engineering*

Synthetic biology encompasses the combining of genes from multiple sources into optimized or novel biosynthetic pathways. In order to do so, the synthetic biology community needs to have easy access to genes, genomic parts, and tools for their assembly.[1,2] Owing to the steadily decreasing price of synthetic DNA, outsourcing DNA editing has become a popular option.[3] However, novel assembly methods or improved cloning techniques are still frequently reported.[2,4−8] This is likely because large genes, whole pathways, and combinatorial libraries are still relatively expensive to synthesize and because repetitive or complex sequences are difficult to produce with standard operating procedures. Thus, synthetic biology continues to rely on effective methods to assemble DNA to minimize *de novo* design and to facilitate exchange of individual parts at reasonable cost. These methods must be compatible with a high-throughput format and ideally must be simple, reliable, flexible, seamless, and cheap.[2] The uracil excision cloning technology (a.k.a. USER cloning) was invented more than 20 years ago[9,10] and increased in popularity with the development of compatible high-fidelity DNA polymerases, like PfuTurbo CX Hotstart and PfuX7.[11−14] Uracil excision cloning is a highly versatile, sequence-independent DNA assembly technology that facilitates simple manipulations such as site-directed mutagenesis, complex multigene assemblies, and standardized biobrick assembly pipelines as well as any combination thereof with simple one-tube protocols.[15,16] The recently developed PHUSER and AMUSER software[17,18] further facilitate the *in silico* design processes.[19] The present study aims to improve the established uracil excision cloning methodology and to show its potential in combination with genome engineering.

RESULTS

Effect of Using Assembly Junctions with Different Melting Temperatures. In uracil excision cloning, the assembly of PCR fragments is mediated by 3′-TN$_x$A overhangs, where *x* typically denotes 7−12 nucleotides.[14,20] A 3′-TN$_x$A overhang is generated when the complementary 5′-AN$_x$U sequence that is incorporated in the PCR oligonucleotide is selectively removed by uracil DNA glycosylase (Figure 1a). To explore the optimal design parameters of TN$_x$A overhangs, we amplified a plasmid encoding constitutively expressed *gfp* with oligonucleotides containing AN$_x$U sequences with predicted T_m's of *circa* 10, 20, 30, 50, and 60 °C (Figure 1b,c, using oligonucleotide nos. 2−11 for creating one PCR fragment and nos. 2−41 for four fragments; see Supporting Information Table S1). The template (pET_Duet_GFP_stop, Supporting Information Table S2) contains an internal stop codon in the *gfp* ORF that is removed upon reassembly (Figure 1d). Counting the number of colonies after transformation allowed the efficiency of assembling one or four fragments to be assessed, and the accuracy was judged by the green-to-white colony ratio (defined as the percentage of green colonies). For both the one- and four-fragment assemblies, junctions with a T_m of 10−30 °C were assembled with 85−96% accuracy (Figure 1e,f, white diamonds). With these junctions, the efficiency was 42 000 to 65 000 colonies per microgram of DNA for one-fragment and 4400 to 19 300 colonies per microgram of DNA for four fragments (Figure 1e,f, white bars).

Received: April 7, 2015
Published: August 11, 2015

DOI: 10.1021/acssynbio.5b00113
ACS Synth. Biol. 2015, 4, 1042−1046

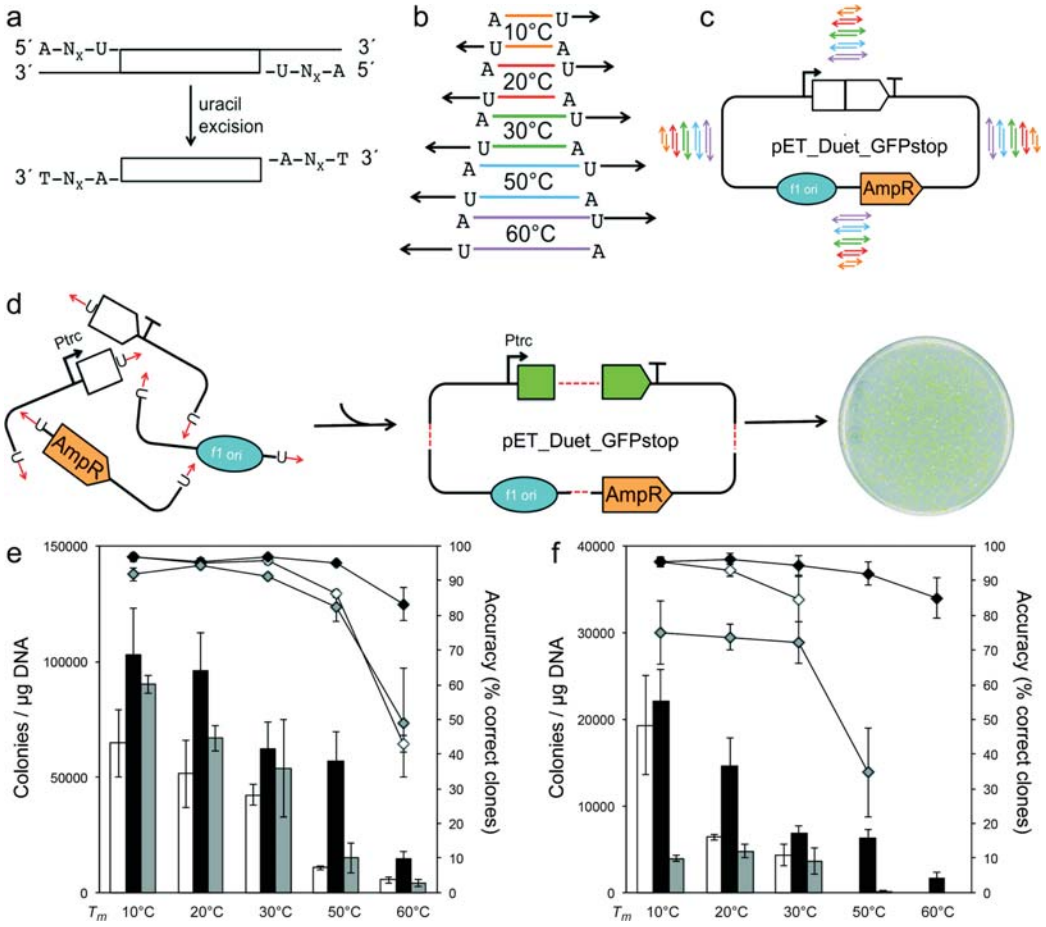


Figure 1. Illustration of uracil excision cloning, model assembly assay, and optimization. (a) TN_xA overhangs are generated by uracil excision when AN_xU sequences are incorporated in the oligonucleotides used in the PCR. (b) Oligonucleotides were designed to vary in the length of AN_xU overhangs corresponding to melting temperatures (T_m) from 10 to 60 °C. (c) The plasmid pET Duet_ GFPstop (*gfp* ORF with an internal stop codon is represented in white) was used as a template for PCR amplification of one or four DNA fragments with uracil-containing oligonucleotides (colored arrows correspond to the oligonucleotides illustrated in (b)); (d) uracil-containing fragments were assembled in the uracil excision reaction, resulting in an intact *gfp* expressible from the leaky P_{trc} promoter; efficiency of (e) one- and (f) four-fragment DNA assemblies via uracil excision cloning in colonies/μg of DNA for chemical transformation of 2.5 μL of the uracil excision mixture into *Escherichia coli*. Results represent mean values of at least three independent experiments with standard error. The original protocol is shown in white bars, and the optimized protocol with purified and unpurified DNA parts, in black and gray bars, respectively. Accuracy as a percentage of correct clones is represented as white diamonds for the original protocol and black or gray diamonds for the optimized protocol with purified and unpurified fragments, respectively.

Increasing the T_m of the AN_xU sequences to 50 and 60 °C resulted in fewer correct recombinants for the one-fragment assembly, and it completely impaired the four-fragment assembly.

Optimizing the Annealing Step in Uracil Excision Cloning. Previously described uracil excision DNA assembly protocols typically include an initial incubation at 37 °C for enzymatic removal of the uracil base and optimal cleavage of the phospho-ribose backbone and a second incubation at 25 °C or room temperature (RT) for assembly of the cohesive ends.[14,15,18,21] However, an assembly junction with a T_m below RT may not assemble efficiently, and in a junction with a T_m above 37 °C, the nucleotides upstream from the excised uracil may not be released efficiently prior to fragment assembly. Thus, we hypothesized that the efficiency of uracil excision cloning could be improved by including an incubation step around the T_m of the most stable assembly junction, followed by a 10 °C incubation step to ensure efficient assembly prior to transformation. This change in the protocol significantly increased both the cloning efficiency and accuracy for all AN_xU sequences tested, independent of the number of fragments assembled (Figure 1e,f, black bars). Strikingly,

DOI: 10.1021/acssynbio.5b00113
ACS Synth. Biol. 2015, 4, 1042−1046

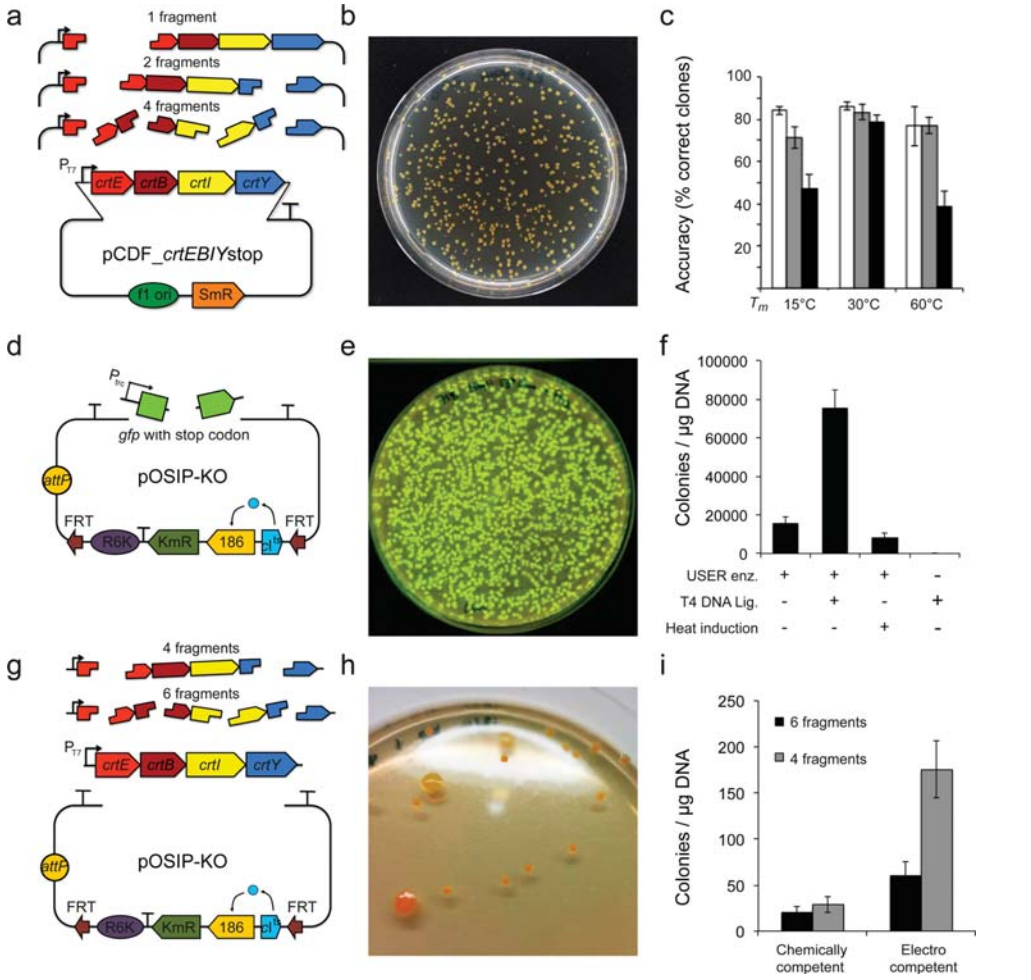


Figure 2. One-step uracil excision cloning for pathway assembly and genomic integration (u-clonetegration). (a) Schematic representation of the pCDF_Duet_crtEBIYstop plasmid that contains the β-carotene biosynthetic genes from *Pantoea ananatis* and illustration of the reassembly with one, two, or four fragments. (b) Representative plate with colonies containing the assembled carotenoid biosynthetic pathway. (c) Accuracy as a percentage of correct clones for one- (white bars), two- (gray bars), and four-fragment (black bars) assemblies with 15, 30, and 60 °C U-overhangs. (d) Schematic representation of the assembly of the pOSIP-KO plasmid and fragments of GFP used to assay u-clonetegration. (e) Representative plate illustrating the efficiency of the direct assembly and genomic integration of the construct leading to *gfp* expression. (f) Optimization by addition of ligase or heat induction for u-clonetegration. (g) Schematic representation of the pOSIP-KO plasmid and four/six fragments of the genes of the β-carotene pathway to assay multigene u-clonetegration. (h) Representative plate with colonies containing the assembled β-carotene pathway in the genome. (i) Comparison of different transformation protocols for u-clonetegration of four fragments (gray bars) and six fragments (black bars). Results are represented as mean values of at least three independent experiments with standard error.

using this protocol, both one and four fragments can be assembled with almost 100% accuracy using AN_xU sequences with a T_m < 30 °C. With the 50 and 60 °C overhangs, the four-fragment assembly was successful only with the newly optimized protocol at an accuracy of *circa* 90%.

Assembling DNA Fragments Directly from the PCR. Uracil excision cloning enables assembly of unpurified PCR products because the generated 3′ ends are not filled in by excessive DNA polymerase activity (similar to the principle of Gibson assembly, where a 5′−3′ exonuclease is preferred over a 3′−5′ exonuclease to avoid competition from the activity of DNA polymerase[7]) and because the USER enzyme is active in most PCR buffers.[11] This feature is attractive, as it is time-saving and enables automation. As shown in Figure 1f (gray bars), four fragments can be efficiently assembled with approximately 80% accuracy directly from unpurified PCR products; for one fragment, the efficiency and accuracy are even

DOI: 10.1021/acssynbio.5b00113
ACS Synth. Biol. 2015, 4, 1042−1046

higher (Figure 1e). Larger overhangs perform poorly, suggesting that fragment purification is a better choice for these.

Accuracy of Uracil Excision Cloning for Assembly of Complex Metabolic Pathways. Our uracil excision protocol enables highly efficient assembly of four DNA parts of variable size (from 540 to 2700 bp) (Figure 1). To test whether the optimized protocol enables accurate assembly of several intact genes in a biosynthetic pathway, we designed a strategy based on *de novo* parts' assembly (sizes varying from 1 to 4.6 kb, using oligonucleotide nos. 42−66 in Supporting Information Table S1) inside the four genes, *crtE*, *crtB*, *crtI*, and *crtY*, of the carotenoid biosynthetic pathway from *Pantoea ananatis*.[22] The four genes were expressed in an operon from the T7 promoter (P_{T7}, Figure 2a). Similar to the *gfp*-based system described above, we introduced a stop codon on the template plasmid within the ORF of *crtY* (pCDF_Duet_crtEBIYstop, Supporting Information Table S2). CrtY converts lycopene (red color) into *β*-carotene (orange color), and this setup therefore facilitates a simple colorimetric screen that distinguishes among correct *β*-carotene-producing clones (orange), misassembled clones (white), and lycopene-producing clones derived from the template plasmid (or constructs misassembled in *crtY*, both red). To generate a compatible, highly competent cloning strain that is capable of expressing *crtEBIY* from P_{T7}, we developed uracil excision combined with clonetegration[23] (see below) to integrate genes encoding four variants of T7 RNA polymerase in the genome of the NEB5*α* strain. The different T7 RNA polymerases were integrated in the same locus (Supporting Information Figure S1); each variant of T7 polymerase (T7*, T7*(T3), T7*(N4), and T7*(K1F)) is under the control of a lac promoter derivative, P_{tac}, and leads to different expression levels of genes (high to low, respectively) controlled by P_{T7}.[24] The four strains, NEB5*α*T7*, NEB5*α*T3, NEB5*α*N4, and NEB5*α*K1F, were transformed with the carotenoid pathway-encoding plasmid, and phenotype robustness was evaluated by colony homogeneity (data not shown). The NEB5*α*K1F strain produced the most homogeneous colonies and was selected for further studies. Assembly accuracy was then determined by counting the percentage of orange colonies (Figure 2b). One-, two-, and four-fragment assemblies were performed with three different AN_xT overhang melting temperatures of *circa* 15, 30, and 60 °C (Figure 2c). For both the one- and two-fragment assemblies, accuracy was more than 70% for all assembly junctions, but only the 30 °C AN_xT overhang resulted in highly accurate assembly of four fragments (≈80% orange colonies).

Uracil Excision Cloning for Genome Engineering. Genomic integration of heterologous genes and pathways is attractive, particularly for the generation of stable production strains for industrial scale-up. However, until recently, integration of large complex pathways in the genome of *Escherichia coli* and other bacteria was a laborious task.[23] The recently described clonetegration technology elegantly simplifies one-step DNA cloning and direct genomic integration into several phage integration sites available in common *E. coli* strains.[23] Motivated by the efficiency of the optimized uracil excision protocol, we wanted to combine the simplistic features of the two technologies. By amplifying the P_{trc} promoter and *gfp* in two fragments and combining them with the pOSIP-KO plasmid as a third PCR fragment (Figure 2d, using oligonucleotide nos. 67, 68, and 71−73 in Supporting Information Table S1 and the template plasmid pET_Duet_GFPstop in Supporting Information Table S2), the sequence was successfully assembled with one-pot uracil excision and subsequently integrated in the genome with the aid of the integrase encoded on the pOSIP vector with high efficiency and accuracy, as judged from the number of fluorescent colonies (Figure 2e). Correct genomic integration in the same genomic location was confirmed by PCR in 12 out of 12 tested colonies (data not shown). Adding T4 DNA ligase increased the efficiency by nearly 4-fold and resulted in >60 000 colonies per microgram of DNA (Figure 2f). We also tested the effect of heat induction for enhancing expression of the integrase by incubating the cells at 42 °C for 15 min after transformation, but this decreased the efficiency by up to 30%. T4 ligase did not catalyze assembly in the absence of uracil excision (Figure 2f). Next, we attempted to assess the accuracy of uracil excision clonetegration (u-clonetegration) by integrating multiple genes in the *E. coli* genome in a one-tube reaction using the *crtEBIY* model pathway (Figure 2g). The four-gene pathway was successfully integrated using the optimized uracil excision protocol with T4 ligase from four and six assembled fragments, albeit with very low efficiency (Figure 2h,i), suggesting that four or more fragments are more efficiently transferred to the genome by rounds of subassemblies by PCR- or plasmid-based amplifications prior to genome integration. In contrast to previous reports,[14] we were able to combine uracil excision cloning with electroporation, and this increased the efficiency of u-clonetegration (Figure 2i, gray bars).

Uracil excision cloning is one of the most versatile DNA assembly technologies available and can be used to perform scarless assemblies, deletions, insertions of up to 100 bp, and multiple simultaneous site-directed mutageneses.[14,18,21] In our experience, the short end-homology requirement in the uracil excision technology is an advantage when assembly junctions are in sequences with high secondary structure propensity, such as when two genes are assembled with a terminator in between. Here, we studied the design parameters for uracil excision DNA assembly and demonstrated that melting temperatures of the AN_xU sequences between 10 and 30 °C enable highly efficient and accurate assembly of up to four unpurified PCR fragments of different sizes. When approaching 10 °C, assembly appears to be more efficient, but less accurate, and thus the chosen design is a compromise between these two parameters. Furthermore, we describe a simple one-tube protocol for assembling up to six DNA fragments for direct genome integration in *E. coli*, greatly facilitating the complex engineering of multiple genes on the genome.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00113.

Complete methods, validating integration of T7* polymerase variants in the genome of *E. coli* (Figure S1), strains and plasmids used in this study (Table S1), and oligonucleotides used in this study (Table S2) (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +45 217-99184. Fax: +45-353-33300. E-mail: morno@biosustain.dtu.dk.

Author Contributions

A.M.C. and S.H.K. performed the experiments. A.M.C., S.H.K., S.S, M.T.N., and M.H.H.N. designed the experiments. The

DOI: 10.1021/acssynbio.5b00113
ACS Synth. Biol. 2015, 4, 1042−1046

manuscript was prepared by A.M.C, S.H.K., and M.H.H.N. with contributions from all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Christopher A. Voigt and his group for providing plasmids encoding the orthogonal T7 RNA polymerases, Keith A. Shearwin for making the pOSIP plasmids available through Addgene, and Sheila Ingemann Jensen for the plasmid pSIJ31B. Additionally, we thank the members of Membrane Synthetic Biology group at The Novo Nordisk Foundation Center for Biosustainability at Technical University of Denmark. This work was supported by The Novo Nordisk Foundation and a Ph.D. grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme [FP7-People-2012-ITN], under grant agreement no. 317058, "BACTORY". S.S. is the recipient of the VILLUM Foundation's Young Investigator Programme grant [VKR023128].

■ ABBREVIATIONS

DNA, deoxyribonucleic acid; bp, base pair; U, deoxyuridine; A, deoxyadenine; T, deoxythymidine; ORF, open reading frame; μL, microliter; M, molar; ng, nanogram; μg, microgram; PCR, polymerase chain reaction; USER, uracil specific reagent; T_m, melting temperature; CFU, colony forming units; IPTG, isopropyl β-D-1-thiogalactopyranoside; *E. coli*, *Escherichia coli*; GFP, green fluorescence protein; *crt*, carotenoid biosynthetic gene; NEB, New England Biolabs

■ REFERENCES

(1) Ellis, T., Adie, T., and Baldwin, G. S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. (Camb)* 3, 109−18.

(2) Marsischky, G., and LaBaer, J. (2004) Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res. 14*, 2020−8.

(3) Kosuri, S., and Church, G. M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods 11*, 499−507.

(4) Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., and Schwaneberg, U. (2010) Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal. Biochem. 406*, 141−6.

(5) Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C. M., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res. 42*, e23.

(6) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One 3*, e3647.

(7) 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., and Hutchison, C. A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A. 105*, 20404−9.

(8) Zhu, B., Cai, G., Hall, E. O., and Freeman, G. J. (2007) In FusionTM assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *BioTechniques 43*, 354−359.

(9) Nisson, P. E., Rashtchian, A., and Watkins, P. C. (1991) Rapid and efficient cloning of Alu-PCR products using uracil DNA glycosylase. *Genome Res. 1*, 120−123.

(10) Smith, C., Day, P. J., and Walker, M. R. (1993) Generation of cohesive ends on PCR products by UDG-mediated excision of dU, and application for cloning into restriction digest-linearized vectors. *Genome Res. 2*, 328−332.

(11) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol. Biol.*, 3.21.

(12) Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res. 35*, e55.

(13) Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol. 10*, 21.

(14) Nour-eldin, H. H., Geu-flores, F., and Halkier, B. A. (2010) USER cloning and USEr fusion: the ideal cloning techniques for small and big laboratories. *Methods Mol. Biol. 643*, 185−200.

(15) Nielsen, M. T., Ranberg, J. A., Christensen, U., Christensen, H. B., Harrison, S. J., Olsen, C. E., Hamberger, B., Møller, B. L., and Nørholm, M. H. H. (2014) Microbial synthesis of the forskolin precursor manoyl oxide in enantiomerically pure form. *Appl. Environ. Microbiol. 80*, 7258−65.

(16) Nielsen, M. T., Madsen, K. M., Seppälä, S., Christensen, U., Riisberg, L., Harrison, S. J., Møller, B. L., and Nørholm, M. H. H. (2015) Assembly of highly standardized gene fragments for high-level production of porphyrins in *E. coli*. *ACS Synth. Biol. 4*, 274−282.

(17) Olsen, L. R., Hansen, N. B., Bonde, M. T., Genee, H. J., Holm, D. K., Carlsen, S., Hansen, B. G., Patil, K. R., Mortensen, U. H., and Wernersson, R. (2011) PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. *Nucleic Acids Res. 39*, 61−7.

(18) Genee, H. J., Bonde, M. T., Bagger, F. O., Jespersen, J. B., Sommer, M. O. A., Wernersson, R., and Olsen, L. R. (2015) Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synth. Biol. 4*, 342−349.

(19) Salomonsen, B., Mortensen, U. H., and Halkier, B. A. (2014) Chapter 5 USER-derived cloning methods and their primer design. *Methods Mol. Biol. 1116*, 59−72.

(20) Bitinaite, J., Rubino, M., Varma, K. H., Schildkraut, I., Vaisvila, R., and Vaiskunaite, R. (2007) USER TM friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res. 35*, 1992−2002.

(21) Nour-eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res. 34*, e122.

(22) Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., and Harashima, K. (1990) Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J. Bacteriol. 172*, 6704−6712.

(23) St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol. 2*, 537−541.

(24) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res. 40*, 8773−81.

DOI: 10.1021/acssynbio.5b00113
ACS Synth. Biol. 2015, 4, 1042−1046

SUPPLEMENTARY MATERIALS

METHODS

Strains, media and plasmids. *Escherichia coli* strain NEB5α and its derivatives (see results section) were used for cloning and propagation of plasmids (Supplementary Table 1). Chemically competent cells of NEB5α and derivatives were obtained as described elsewhere.[1] Competency of the cells was 2.6-3.3 × 10^7 CFU/μg DNA. NEB5αK1F cells were made electro competent using manufacturers protocol (New England Biolabs, Ipswich, USA). Bacteria were propagated on Luria-Bertani (LB) agar plates or liquid 2xYT media supplemented with spectinomycin (50 μg mL^{-1}) or ampicillin (100 μg mL^{-1}) when required. Plasmids were isolated using the NucleoSpin® Plasmid QuickPure Kit (Macherey-Nagel, Durën, Germany). A plasmid encoding the carotenoid biosynthetic pathway (Genbank accession number D90087) from *Pantoea ananatis*[2] was generously provided by Sheila Ingemann Jensen. Constitutively expressed *gfp* was cloned into the pET-Duet-1 plasmid[3] (Novagen, Darmstadt, Germany), in which the T7 promoter (P_{T7}) was replaced with a leaky P_{trc} promoter. A stop codon was introduced in the *gfp* sequence by site-directed mutagenesis.

PCR conditions and DNA quantification. DNA parts were amplified using the proof-reading PfuX7 DNA polymerase[4] (PfuX7 clone available upon request) in Cloned Pfu DNA polymerase Buffer (buffer available from Agilent technologies, Santa Clara, USA). PCR products were obtained with 20 cycles in 50 μL reaction mixtures using a C1000 Touch™ Thermal Cycler (BioRad, Hercules, USA). PCR products were purified using a PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Foster City, USA). Buffers for PCR and cloning reactions were purchased from Agilent technologies (Santa Clara, USA) and Invitrogen (Paisley, UK). PCR products were quantified using a NanoDrop 8000 (Thermo Scientific, Wilmington, USA).

Oligonucleotide design. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA) (Supplementary Table 1) and designed with melting temperatures (T_m) of *circa* 60°C (http://www.thermoscientificbio.com/webtools/tmc/). Uracil excision compatible PCR products were obtained using oligonucleotides with overhangs designed with T_m varying from 10°C to 60°C.

Assembly with uracil excision cloning. Uracil excision cloning was adapted from the protocol described by Geu Flores *et al* [5]. The uracil excision reaction contained 100 ng of each purified PCR product, 5x Phusion HF buffer (Invitrogen, Paisley, UK) or T4 DNA ligase buffer (New England Biolabs, Ipswich, USA), 1 U of USER™ enzyme mix (New England Biolabs, Ipswich, USA) in a final

volume of 10 µL. The mixture was incubated for 15 min at 37°C, followed by 15 min at 10°C, 15°C, 20°C, 30°C, 50°C or 60°C and finally 10 min at 10°C, in a C1000 Touch™ Thermal Cycler (BioRad, Hercules, USA). Optionally, 2.5 U of T4 DNA ligase was added after the steps described above and incubated at room temperature for 15 min. Reaction mixtures were kept on ice prior to transformation into chemically competent *E. coli* NEB5α (for GFP assemblies) and chemical or electro competent *E. coli* NEB5αK1F cells (for carotenoid pathway assembly). For electroporation, 2 µL of USER mixture was added to 50 µL electro competent cells and transferred to ice-cold 2 mm cuvettes. After pulsing by 2.5 kV in an electroporator, 950 µL of SOC was added and cells recovered in 37°C for 1 hour. After transformation, plates were incubated at 37°C for 16 hours and when relevant for several days at room temperature for color development. All oligonucleotides used in this study are summarized in Supplementary Table 1.

Uracil excision cloning for genome engineering with clonetegration. Clonetegration was performed as previously described.[6] In short, the pOSIP backbones and genes to be integrated on the genome were amplified with the oligonucleotides listed in Supplementary Table 2 and purified DNA parts were assembled with uracil excision cloning in T4 ligase buffer as described above. For construction of NEB5α strains containing orthogonal T7 polymerase variants, DNA parts were obtained by amplification from pOSIP-KL backbone and each plasmid harboring the T7 RNA polymerase variants (Supplementary Table 2). Subsequently, the assembled parts were transformed into NEB5α and the integrase modules flipped out as described previously.[6] For the GFP assembly test, the pET_Duet_GFPstop plasmid was used as template for PCR amplification with 20°C AN_xU-overhang-containing oligonucleotides. For assembly and integration of the carotenoid pathway, DNA fragments were amplified with oligonucleotides containing 30°C AN_xU-overhangs from pCDF_Duet_crtEBIYstop. Reactions were performed in T4 DNA ligase buffer (New England Biolabs, Ipswich, USA) (as described above for assembly with uracil excision cloning) and 2 µL were transformed into *E. coli* NEB5αK1F chemically or electro competent cells. After transformation, plates were incubated at 30°C for 20 hours and subsequently at room temperature for color development.

Supplementary Figures

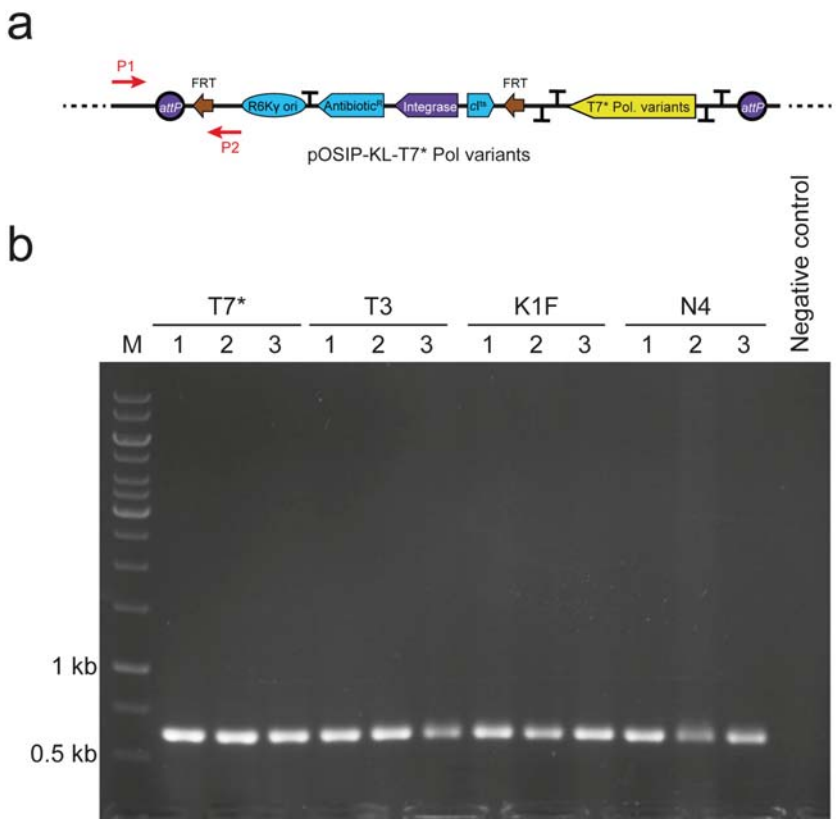


Figure 1S. Validating integration of T7* polymerase variants on the genome of *E. coli* NEB5a. (a) Integration of variants of the T7 RNA polymerase, using uracil excision in combination with clonetegration [1], was validated with PCR using a P1 oligonucleotide that anneals on the genome and a P2 oligonucleotide that anneals in the integrated DNA. (b) PCR was performed on three colonies from each T7 variant constructed and analyzed by agarose gel electrophoresis. The expected PCR product size (631 bp) was observed in all the tested colonies.

Supplementary Tables

	Property	Reference
Strains		
MG1655		
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB
NEBRP01	NEB5α with a T7* RNA polymerase integrated, Kan^R	This study
NEBRP02	NEB5α with a T7*(T3) RNA polymerase integrated, Kan^R	This study
NEBRP03	NEB5α with a T7*(K1F) RNA polymerase integrated, Kan^R	This study
NEBRP04	NEB5α with a T7*(N4) RNA polymerase integrated, Kan^R	This study
NEBRP05	NEBRP01 without integrase module	This study
Plasmids		
pUC19	Cloning vector, Amp^R	NEB
pET_Ptrc-GFP	GFPopt expressed from Ptrc promoter in pET-Duet-1	This study
pET_Duet_GFP stop	pET_Ptrc-GFP with stop codon in *gfp* ORF	This study
N249	T7* RNA polymerase, Spc^R	[7]
N377:115	T7*(T3) RNA polymerase, Spc^R	[7]
N421:115	T7*(K1F) RNA polymerase, Spc^R	[7]
W74	T7*(N4) RNA polymerase, Spc^R	[7]
pOSIP-KL	Clonetegration plasmid with kanamycin resistance and lambda integrase, Kan^R	[6]
pOSIP-KO	Clonetegration plasmid with kanamycin resistance and 186 integrase, Kan^R	[6]
pE-FLP	FLP recombinase-expressing plasmid, Amp^R	[6]
pSIJ31B	*P. ananatis crtEBIY* in pCDF_Duet, Spc^R	unpublished
pCDF_Duet_crtEBIYstop	pSIJ31B with stop codon in *crtY*	This study

Supplementary Table 1 Strains and plasmids used in this study

Number	Sequence	Strategy
1	AACAAGGGUCAATCACCTTCAAACTTGACTACAGC	
2	ACCCTTGUTAATCGTATCGAGTTAAAGGGTAC	
3	ACAAGGGUATCACCTTCAAACTTGACTACAGC	
4	ACCCTTGTUAATCGTATCGAGTTAAAGGG	
5	AACAAGGGUATCACCTTCAAACTTGACTACAGC	
6	ACCCTTGTTAAUCGTATCGAGTTAAAGGGTAC	
7	ATTAACAAGGGUATCACCTTCAAACTTGACTACAGC	
8	ACCCTTGTTAATCGTAUCGAGTTAAAGGGTACTG	
9	ATACGATTAACAAGGGUATCACCTTCAAACTTGACT ACAGC	
10	ACCCTTGTTAATCGTATCGAGUTAAAGGGTACTGAT TTTAAAG	
11	ACTCGATACGATTAACAAGGGUATCACCTTCAAACT TGACTACAGC	
12	ACGGTATCUCAAAAAGGATCTTCACCTAGATC	
13	AGATACCGUATAATCTCATGCCATCGTGC	
14	ACGGTATCTUCAAAAAGGATCTTCACCTAGATC	
15	AAGATACCGUATAATCTCATGCCATCGTGC	
16	ACGGTATCTTAGUCAAAAAGGATCTTCACCTAGATC	
17	ACTAAGATACCGUATAATCTCATGCCATCGTGC	
18	ACGGTATCTTAGTCCAGUCAAAAAGGATCTTCACCT AGATC	
19	ACTGGACTAAGATACCGUATAATCTCATGCCATCGT GC	Uracil excision cloning
20	ACGGTATCTTAGTCCAGTGACTGUCAAAAAGGATCT TCACCTAGATC	
21	ACAGTCACTGGACTAAGATACCGUATAATCTCATGC CATCGTGC	
22	AGGTACTAAUCAGCAGCAGTCGCTTCAC	
23	ATTAGTACCUCAAAACGTCTGCGACCTG	
24	AGGTACTAATGUCAGCAGCAGTCGCTTCAC	
25	ACATTAGTACCUCAAAACGTCTGCGACCTG	
26	AGGTACTAATGTCTUCAGCAGCAGTCGCTTCAC	
27	AAGACATTAGTACCUCAAAACGTCTGCGACCTG	
28	AGGTACTAATGTCTTACGGUCAGCAGCAGTCGCTT CAC	
29	ACCGTAAGACATTAGTACCUCAAAACGTCTGCGACC TG	
30	AGGTACTAATGTCTTACGGTCGATUCAGCAGCAGTC GCTTCAC	
31	AATCGACCGTAAGACATTAGTACCUCAAAACGTCTG CGACCTG	
32	AGCATCAGUGATGTCGGCGATATAGGCG	
33	ACTGATGCUACCGATGGGGAAGATCG	
34	AGCATCAGTAUGATGTCGGCGATATAGGCG	
35	ATACTGATGCUACCGATGGGGAAGATCG	

36	AGCATCAGTATCUGATGTCGGCGATATAGGCG	
37	AGATACTGATGCUACCGATGGGGAAGATCG	
38	AGCATCAGTATCTGCAAUGATGTCGGCGATATAGGCG	
39	ATTGCAGATACTGATGCUACCGATGGGGAAGATCG	
40	AGCATCAGTATCTGCAATCCAUGATGTCGGCGATATAGGCG	
41	ATGGATTGCAGATACTGATGCUACCGATGGGGAAGATCG	
42	ACGCACUATATTGATAATGCGTGATTAGATC	
43	ATGAATGGUAGGGCGTCCGC	
44	ACCATTCAUTCTCATTACGGAGAG	
45	AACGCGUAAGCCGGGGCGA	
46	ACGCGTUTGATCATCTGGAAGGCTTC	
47	ATAAATGGAUGAGGTGGCGAAGG	
48	ATCCATTTAUACGTTGATACACGCGCTG	
49	AGTGCGUGTCTTCAATTAACAATCTGG	
50	ACGCACUATATTGATAATGCGACATTAGATC	
51	AGAATGAATGGUAGGGCGTC	
52	ACCATTCATTCUCATTACGGAGAG	
53	ATCAAACGCGUAAGCCGGGGCGATATC	
54	ACGCGTTTGAUCATCTGGAAGGCTTC	
55	ACGTATAAATGGAUGAGGTGGCGAAG	
56	ATCCATTTATACGUTGATACACGCGCTG	
57	ATATAGTGCGUGTCTTCAATTAACAATCTG	
58	ACGCACTATAUTGATAATGCGACATTAGATC	
59	ATGCTCTCCGTAATGAGAATGAATGGU	
60	ACCATTCATTCTCATTACGGAGAGCAU	
61	AGCCTTCCAGATGATCAAACGCGU	
62	ACGCGTTTGATCATCTGGAAGGCU	
63	AGCGCGTGTATCAACGTATAAATGGAU	
64	ATCCATTTATACGTTGATACACGCGCU	
65	ATCTAATGTCGCATTATCAATATAGTGCGU	
66	ACGCACTATATTGATAATGCGACATTAGAU	
67	AGATGCAUGGCGCCTAACC	
68	AGCCCTCUAGAGGATCCCCGGGTAC	
69	AGAGGGCUGTTCTGGCAAATATTCTGAAATGAGCTG	
70	ATGCATCUCTAACTAACTAACCCTTAGTGACTCCTG	U-clonetegration
71	AGAGGGCUATGCGTCCGGCGTAGAGG	
72	ATGCATCUGATTATGCGGCCGTGTACAA	
73	AGAGGGCUGCGACTCCTGCATTAGGAAAT	

Supplementary Table 2. Oligonucleotides used in this study

Supporting references

(1) Inoue, H., Nojima, H., and Okayama, H. (1990) Optimization of transformation buffer. *Gene*, 96, 23–28.

(2) Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., and Harashima, K. (1990) Elucidation of the *Erwinia uredovora* Carotenoid Biosynthetic Pathway by Functional Analysis of Gene Products Expressed in *Escherichia coli*. *J. Bacteriol.*, 172, 6704–6712.

(3) Toddo, S., Söderström, B., Palombo, I., von Heijne, G., Nørholm, M. H. H., and Daley, D. O. (2012) Application of split-green fluorescent protein for topology mapping membrane proteins in *Escherichia coli*. *Protein Sci.* 21, 1571–6.

(4) Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* 10, 21.

(5) Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* 35, e55.

(6) St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-step cloning and chromosomal integration of DNA. 2, 537-541.

(7) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* 40, 8773–81.

Paper II

Uracil Excision for Assembly of Complex Pathways

Uracil Excision for Assembly of Complex Pathways

Ana Mafalda Cavaleiro, Morten T. Nielsen, Se Hyeuk Kim, Susanna Seppälä, and Morten H.H. Nørholm

Abstract

Despite decreasing prices on synthetic DNA constructs, higher-order assembly of PCR-generated DNA continues to be an important exercise in molecular and synthetic biology. Simplicity and robustness are attractive features met by the uracil excision DNA assembly method, which is one of the most inexpensive technologies available. Here, we describe four different protocols for uracil excision-based DNA editing: one for simple manipulations such as site-directed mutagenesis, one for plasmid-based multigene assembly in *Escherichia coli*, one for one-step assembly and integration of single or multiple genes into the genome, and a standardized assembly pipeline using benchmarked oligonucleotides for pathway assembly and multigene expression optimization.

Keywords: BioBricks, DNA editing, Metabolic engineering, Molecular cloning, Synthetic biology, Uracil excision cloning

1 Introduction

The polymerase chain reaction (PCR) [1] is a simple yet incredibly powerful technology that revolutionized molecular biology. Shortly after the advent of PCR, a handful of methods for assembly of PCR-amplified DNA into larger constructs was developed. PCR generates double-stranded DNA flanked by sequences that are defined by the two PCR primers, and several methods exist that facilitate the formation of cohesive ends for specific higher-order assemblies (Fig. 1). Simple features can be added when the oligonucleotides are chemically synthesized. As an example, uracil excision DNA assembly makes use of oligonucleotides where selected thymines are replaced by uracils. This is a non-mutagenic and PCR-tolerated replacement, as the uracil is able to form base pairs with adenine nucleotides on the complementary strand [2–4]. Following PCR, the uracils are selectively removed by treatment with uracil DNA glycosidase, leaving a chemically unstable phosphoribose backbone. At elevated temperatures, the upstream sequence

T.J. McGenity et al. (eds.), *Hydrocarbon and Lipid Microbiology Protocols*, Springer Protocols Handbooks,
DOI 10.1007/8623_2015_133, © Springer-Verlag Berlin Heidelberg 2015

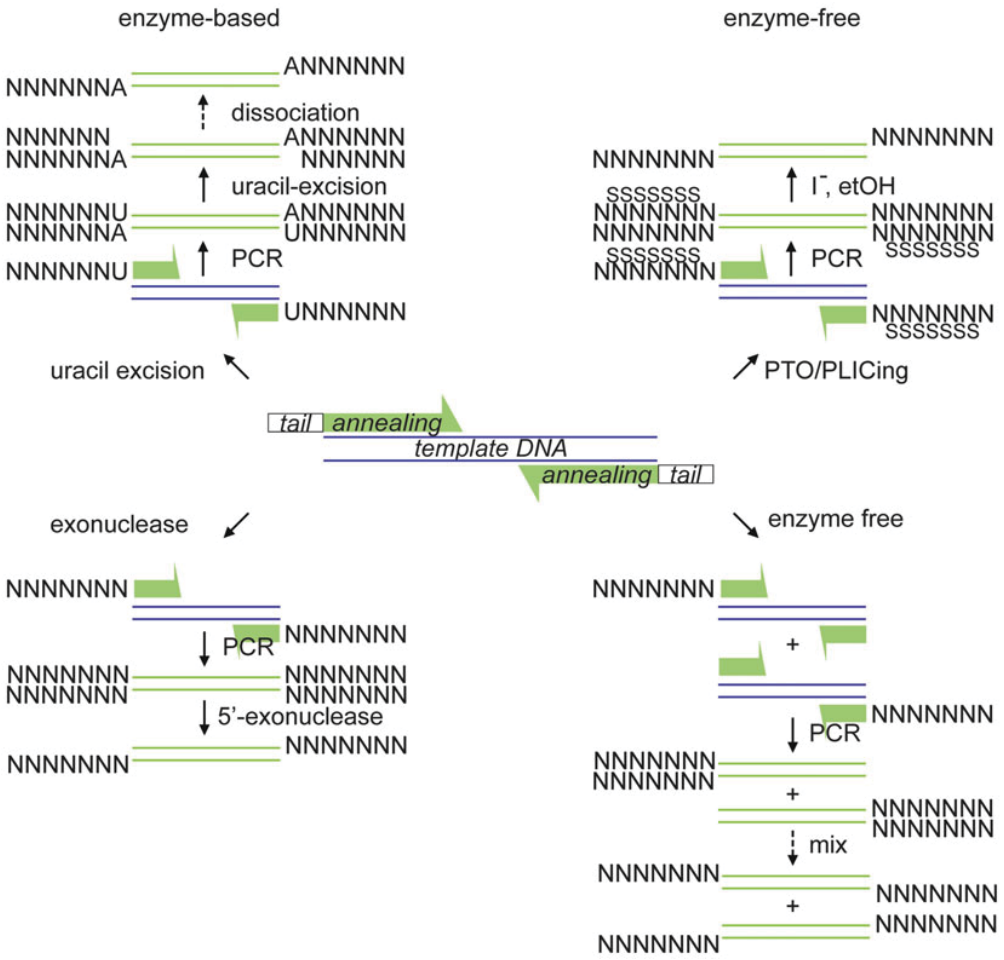


Fig. 1 Illustration of different methods to create cohesive ends on PCR fragments for specific higher-order assemblies. In the schematic examples, all four methods can generate the same 3′ cohesive ends that are not filled in by excess DNA polymerase activity from the PCR. Thus, all methods can in principle be employed directly after PCR with no prior purification. *S* denotes the phosphorothioate modification employed in the PTO/PLICing cloning technology. I^- denotes iodine and etOH denotes ethanol. For more information and references, see main text

dissociates, generating a single-stranded DNA overhang. A recently developed similar approach uses phosphorothioate (PTO)-modified synthetic oligonucleotides [5]. PTO-modified DNA is converted to single-stranded DNA by treatment with a solution of iodide and ethanol. Thus, in the case of PTO, the formation of cohesive ends is enzyme-free. Another enzyme-free route to cohesive ends on PCR products involves the use of two pairs of highly similar oligonucleotides, but of slightly different length, for amplification of the same DNA template [6]. When the resulting two

PCR products are mixed, denatured, and reannealed, the two products recombine and form single-stranded ends defined by the length difference of the two oligonucleotide pairs. However, this approach complicates the PCR setup (and doubles the price tag) and does not seem to be extensively used. Finally, exonuclease-catalyzed recessions of the ends of the DNA are heavily used alternatives, e.g., in the form of ligase-independent cloning (LIC) [7] or the commercially available cloning kit Gibson Assembly [8].

In our experience, uracil excision excels in robustness, simplicity, and price tag. This may be explained by the relatively short overlap sequence that uracil excision requires (typically 7–12 nucleotides [9], compared to, e.g., 12 nucleotides for PTO-based cloning [5] and 40 nucleotides for Gibson Assembly [10]). Theoretically, DNA fragments with cohesive ends should recombine with the same efficiency independently of how the single-stranded ends were generated. However, the protocol, purity, and quality of the DNA overhangs make all the difference. The quality and yield of synthetic oligonucleotides is typically low when approaching a size of 100 nucleobases [11]. Therefore, PCR-based assembly technologies that use short oligonucleotides are probably less error prone and more efficient. Moreover, short functional elements, such as promoters or ribosome binding sites, can easily be correctly incorporated directly in oligonucleotides that are assembled using short overlap sequences, because the total length of the oligonucleotide is kept relatively short. In our experience, 5′ "tails" (sequence added at the 5′ end of the oligonucleotides that do not anneal to the template DNA in the first PCR cycles) up to more than 100 nucleotides are possible, but often negatively affect the PCR yield.

Another way to ensure oligonucleotide quality is to build a molecular cloning pipeline that reuses benchmarked oligonucleotides. This was recently demonstrated for the uracil excision assembly and engineering of a six-gene biosynthetic pathway for porphyrin production [12] and a seven-gene heterologous pathway for production of a diterpene in *Escherichia coli* [13]. This type of standardization perfectly fits large collaborative efforts, much like BioBricks in the global iGEM project [14], and reuse of parts also enables better comparison of data.

Protocols for simple and seamless assembly of PCR products (also known as USER fusion), and the corresponding primer design, have been described and reviewed previously [15, 16]. Here, we provide protocols for simple manipulations and more complex assembly pipelines, including site-directed mutagenesis, multigene assembly, one-step cloning, and genome integration with uracil excision, and for a standardized, BioBrick uracil excision-based DNA editing pipeline.

2 Materials

2.1 Strains, Media, and Antibiotic Selection

1. Bacterial strains: *E. coli* strain NEB5α (New England Biolabs, Ipswich, USA) is used as a cloning host. *E. coli* BL21, K12 MG1655, and KRX (Promega, Madison, USA) are used for uracil excision combined with genomic integration (see below).
2. Growth media: SOC (20 g Bacto-Tryptone, 5 g yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM $MgSO_4$, 20 mM glucose, water up to 1 L), 2×YT (16 g Bacto-Tryptone, 10 g yeast extract, 5 g NaCl, water up to 1 L), and LB (10 g Bacto-Tryptone, 5 g yeast extract, 10 g NaCl, water up to 1 L) (all reagents can be purchased from Sigma-Aldrich, St. Louis, USA).
3. Antibiotics: chloramphenicol (25 μg/mL), kanamycin (50 μg/mL), and tetracycline (50 μg/mL) (Sigma-Aldrich, St. Louis, USA). For clonetegration, half concentration is used with all antibiotics.

2.2 PCR Components

1. DNA polymerase: uracil excision-compatible PCR products are amplified using the proofreading PfuX7 DNA polymerase [17] (*see* **Note 1**). Cloned Pfu DNA Polymerase Buffer (Agilent Technologies, Santa Clara, USA) is used to buffer the reaction mixture.
2. Oligonucleotides (Integrated DNA Technologies, Inc., Coralville, USA) are designed with melting temperatures (T_m) of ca. 60°C. Additionally, all oligonucleotides contain one uracil, typically placed 7–12 nucleotides from the 5′ end (*see* **Note 2**). Upon uracil excision, the generated single-stranded ends should have melting temperatures between 10 and 30°C [18] (*see* **Note 3**).
3. Template DNA: plasmid DNA is isolated using the NucleoSpin® Plasmid QuickPure Kit (Macherey-Nagel, Bethlehem, USA). Plasmid aliquots are kept at −20°C (*see* **Note 4**).
4. PCR purification: PCR products are purified using a PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Thermo Fisher Scientific Inc., Waltham, USA).
5. Template DNA removal: *DpnI* (20,000 U/mL) (New England Biolabs, Ipswich, USA) is used to degrade methylated template DNA after the PCR.

2.3 USER Cloning

1. USER™ enzyme mix (New England Biolabs, Ipswich, USA).
2. USER reaction is performed in 5× Phusion HF Buffer (Life Technologies, Grand Island, USA) or Cloned Pfu DNA Polymerase Buffer (Agilent Technologies, Santa Clara, USA).

2.4 Plasmid DNA

1. Vectors: a series of pOSIP vectors is described in St-Pierre et al. [19] and can be obtained from Addgene (Addgene, Cambridge, USA). Duet vectors are available from Merck Millipore (EMD Millipore, Billerica, USA) or Addgene.

3 Methods

The protocols described here showcase the versatility of the uracil excision methodology and include protocols for (1) simple introductions of mutations, deletions, and insertions in DNA, (2) multigene assembly, (3) direct assembly and genome integration, and (4) using standardized BioBricks for assembly of pathways. The first uracil excision protocol describes the introduction of mutations, insertions, or deletions by one-fragment whole-plasmid synthesis and is largely based on the overall principles described by Nørholm [17]. Multigene assembly is performed as described previously [20] with some modifications. The third uracil excision protocol adds direct genome integration (clonetegration [19]) to the uracil excision portfolio. The optimal design parameters for multigene assembly and uracil excision combined with clonetegration have recently been explored [18]. Detailed information on clonetegration including vectors and an oligonucleotide list for colony PCR is described in St-Pierre et al. [19]. The fourth uracil excision protocol describes two operations of a fully standardized assembly procedure. The first standardized operation encompasses cloning of genes of interest into an entry vector using gene-specific oligonucleotides with fixed extensions mediating cloning. This vector contains all elements required for protein production in *E. coli* and can therefore be used straightaway for monitoring proper transcription and translation. The second standardized operation is assembly of entry fragments into multigene constructs using pairs of oligonucleotides with generic annealing parts, but distinct cloning mediating extensions. These oligonucleotides facilitate directional and specific assembly of any number of fragments. For detailed description of the options and limitations of such a standardized design, please refer to Nielsen et al. [12].

3.1 PCRs

The PCRs are performed using 1 μL PfuX7 DNA polymerase (the optimal concentration is typically batch dependent and should be empirically determined when purifying the polymerase – after desalting of his-tagged-purified PfuX7 [17], we typically determine the optimal concentration by titrating the amount of PfuX7 in a standard PCR reaction), 5 μL 10× Cloned Pfu Polymerase Buffer, 5 μL dNTP mix (25 mM each of dATP, dTTP, dGTP, dCTP), 2 μL DNA template (150 ng μL), 5 μL forward primer (5 μM), 5 μL reverse primer (5 μM), 1.2 μL $MgCl_2$ (50 mM) (it may be advantageous to optimize the $MgCl_2$ concentration from batch to batch PfuX7 by titrating the final concentration from 1 to 5 mM), and

29.8 μL nuclease-free water. The PCR involves an initial denaturation step at 98°C for 2 min, then 20 cycles of 98°C for 20 s, 58°C for 20 s, and 72°C for 45 s/kbp. Finally, the thermocycler is programmed for 72°C for 8 min and stored at 12°C.

3.2 Analysis and Purification of PCR Results

PCR products are analyzed by standard agarose gel electrophoresis. The resulting PCR products may be purified using any PCR cleanup kit.

3.3 Simple Protocol for Site-Directed Mutagenesis, Insertions, or Deletions

Mutations, deletions, or insertions in plasmid constructs are made by amplifying the whole plasmid with uracil-containing oligonucleotides that incorporate these new features. The extraordinarily simple protocol involves adding USER™ enzyme mix and *DpnI* directly to the PCR reaction mix described above; incubate for 1 h at 37°C and 20 min at 16°C in a thermocycler followed by direct transformation of 3 μL of the reaction mixture into 17 μL chemically competent cells (*see* Sect. 3.5). Oligonucleotide design is very flexible, but general guidelines can be found in Sect. 2.2, and it is recommended to try software-assisted design tools such as AMUSER [21].

3.4 Uracil Excision-Assisted Multigene Assembly

3.4.1 Simple Multigene Assembly with Non-purified Fragments

For assembly of two or more fragments, equal volumes of each PCR reaction are mixed in a total volume of 10 μL and buffered using the 5× Phusion HF Buffer (*see* **Note 5**). For template removal, *DpnI* is added prior to USER™ enzyme mix and incubated for 1 h at 37°C. The *DpnI* enzyme is deactivated by incubation at 65°C for 10 min. After 5 min on ice, 1 μL of USER™ enzyme mix is added to the reaction tubes, and uracil excision is accomplished by incubating the sample at 37°C for 15 min. Subsequently, DNA assembly is executed by cooling down the reaction to below the melting temperature of the cohesive ends for at least 15 min.

3.4.2 Multigene Assembly with Purified Fragments

Purified DNA fragments (*see* **Note 6**) are assembled as described for the non-purified DNA fragments except that 100 ng of each fragment is used and the *DpnI*-assisted template elimination step can be omitted.

3.5 Chemical Transformation of *E. coli* NEB5α Cells

17 μL of chemically competent *E. coli* NEB5α cells are mixed with 3 μL of the assembly mix described above and incubated for 15 min on ice followed by a heat shock at 42°C for 1 min (*see* **Note 7**). Following the heat shock, 1 mL of LB medium is added, and the cells are incubated for 1 h at 37°C, followed by plating on solid LB medium with the appropriate antibiotic selection for 16 h at 37°C. For selection with antibiotics like ampicillin or carbenicillin, the cells can be spread without a 1 h recovery step.

3.6 One-Step Uracil Excision Assembly and Genome Integration

Amplify one of the pOSIP backbones (*see* **Note 8** and [19]) with the oligonucleotides 5′-AGATGCAUGGCGCCTAACC-3′ and 5′-AGCCCTCUAGAGGATCCCCGGGTAC-3′ and the DNA to be integrated on the genome with 5′-AGAGGGCU-3′ followed by a gene-specific forward annealing sequence and 5′-ATGCATCU-3′ followed by a gene-specific reverse annealing sequence using the PCR conditions described above. Gel purify the amplified DNA, and make an assembly mix as described above except for using a molar ratio of 3:1 between insert and vector. Transform *E. coli* cells as described above. Recover the cells in SOC medium at 37°C for 1 h, spread the cells on LB agar plate containing the appropriate antibiotic, and incubate the plate at 30°C for 20 h. Perform a standard colony PCR to confirm the clones are integrated as described in St-Pierre et al. [19].

3.7 Standardized BioBrick Bioengineering Pipeline with Uracil Excision

Make initial entry clones by PCR amplifying the pET-Duet-1 vector using the oligonucleotides 5′-AGCACTGGUCATTGCTA ATGCTTAAGTCGAACAG-3′ and 5′-ACCACTGGUCATTGC TTATCTCCTTCTTAAAGT-3′ (*see* **Note 9**). PCR amplify gene-coding sequences with 5′-ACCAGTGGU-3′ followed by a gene-specific forward annealing sequence and 5′-ACCAGTGCU-3′ followed by a gene-specific reverse annealing sequence. In the standardized entry clones, 5′-ATGACCAGTGGT-3′ that translates into MTSG is added to the 5′ end, and 5′-AGCACTGGTCA TTGC-3′ that translates into TSGHC is added to the open reading frame. Make sure that the oligonucleotides anneal in frame with the coding sequence. At this stage, genes of interest can be tested for proper transcription and translation using selective ^{35}S-methionine labeling of gene products in the presence of rifampicin (*see* **Note 10** and [22]). The standardized 5′ end may facilitate a more predictable translational initiation rate, as previously described for similar translational fusions [23, 24], and the standardized 5′ and 3′ sequences serve as anneal sites for collections of standardized oligonucleotides for higher-order assemblies, independent of the specific genes inserted in the entry vectors. Higher-order assemblies are generated with oligonucleotides with the same overall design: linker + control element + annealing sequence. When generating the pET-Duet-1-based entry vector as described above, the forward annealing sequence for downstream multigene assembly is 5′-ATAAGCAAT GACCAGTGGT-3′, and the reverse annealing sequence is 5′-TAATGTAAGTTAGCTCACTCATTAG-3′. The principle is schematically illustrated in Fig. 2. The setup will allow the buildup of a library of benchmarked oligonucleotides where differently designed linkers have been validated for correct assembly. Examples of validated linkers are 5′-ACACCGACU-3′/5′-AGTCGGTGU-3′, 5′-ACGCTGCTU-3′/5′-AAGCAGCGU-3′, 5′-AGACGTCAU-3′/5′-ATGACGTCU-3′, 5′-AGGTCTGAGU-3′/5′-ACTCAGACCU-3′, 5′-ATAGGCTTU-3′/5′-AAAGCCTAU-3′, and 5′-AACGTGGAU-3′/5′-ATCCACGTU-3′ [12, 13]. Examples of control elements are

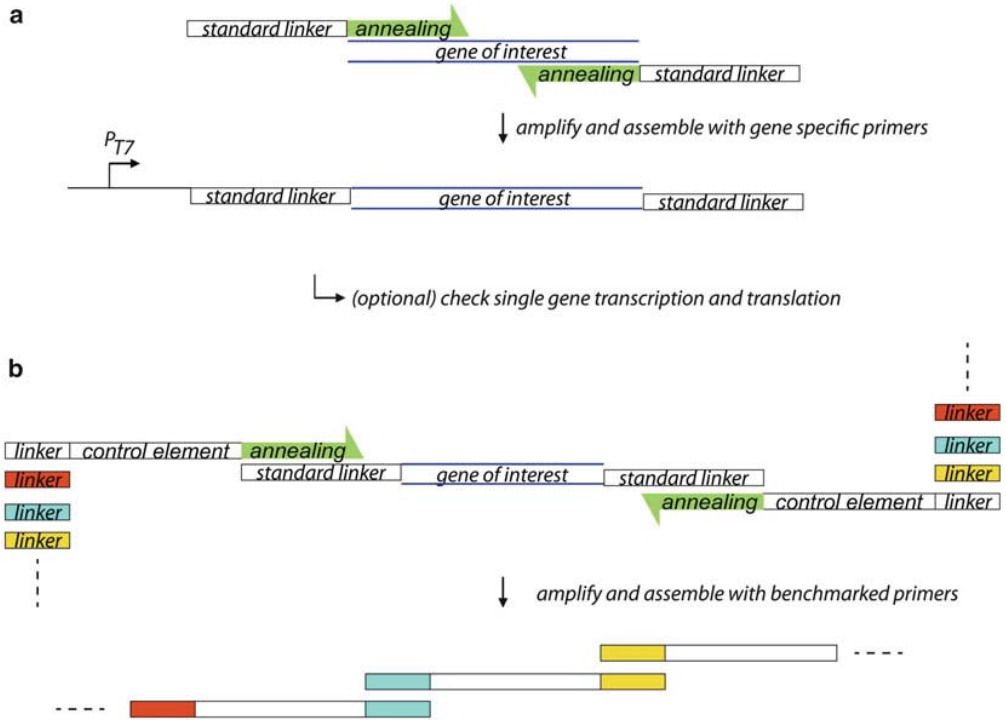


Fig. 2 Illustration of the two-step, uracil excision-based, standardized pipeline for multigene engineering. (**a**) In the first step, genes of interest are cloned with standardized linkers into an entry vector. In the entry vector, an orthogonal T7 phage promoter allows for assessment of proper transcription and translation by ^{35}S-methionine labeling in the presence of rifampicin (rifampicin blocks transcription of endogenous genes by inhibiting the endogenous *E. coli* RNA polymerase). (**b**) The standardized linkers allow the use of standardized oligonucleotides for re-amplification and construction of multigene constructs with benchmarked linkers and functional elements such as promoters and ribosome binding sites. Linkers for uracil excision are relatively short, thus allowing for larger control elements to be incorporated in standard oligonucleotides

constitutive promoters such as P_{trc} followed by randomized Shine-Dalgarno sequences (for details, *see* 12] and the phage promoter P_{T7} followed by the lac operator and a consensus Shine-Dalgarno sequence (for details, *see* 13]. The protocols for assembly are as described above (*see* **Note 11**).

4 Notes

1. Commercially available proofreading DNA polymerases with similar characteristics are available as Phusion U Hot Start DNA Polymerase (Thermo Fischer Scientific, Pittsburgh, USA) and KAPA HiFi Uracil+ (Kapa Biosystems, Inc., Wilmington, USA).

2. Oligonucleotides can be designed using the PHUSER or AMUSER software [21, 25].
3. The melting temperature of the overhangs can be calculated by online software tools such as the Tm calculator from Thermo Fischer Scientific.
4. Plasmid aliquots should contain a small volume (max. 50 μL) to avoid repeated cycles of freeze thawing.
5. According to the supplier (New England Biolabs, Ipswich, USA), the USER™ enzyme is active in all standard reaction buffers. We routinely use buffers such as Phusion HF, NEB4, cloned Pfu buffer, and T4 ligase buffer.
6. In our experience, purification in some cases enhances the efficiency and fidelity of the assembly reaction, possibly due the removal of interfering oligonucleotides [18], but it also complicates the protocol.
7. We routinely use between 30 s and 2 min for heat shock – the optimal incubation time depends on the plasticware and the heat block and can be optimized empirically.
8. Clonetegration is highly dependent on the kind of integrase in the pOSIP vector and the efficiency of the competent cells. Before you select the strain and vector for integration, check if the strain contains the *attB* site in the genome corresponding to the integrase and *attP* site in the vector. For example, in the case of pOSIP-KO (containing phage 186 integrase), MG1655 contains two corresponding *attB* sites, whereas BL21 (DE3) contains only one.
9. The protocol is described for uracil excision cloning, since this is the technique most often applied in our lab. The concept and principles of standardized assembly, however, are by no means limited to this cloning technique. On the contrary, the principles can be implemented with any PCR-based cloning technique as well as several restriction enzyme-based techniques as described in Nielsen et al. [12].
10. While his technique should be applicable to all *E. coli* strains containing T7-RNA polymerase, it is our experience that BL-21 (DE3) is superior regarding the 35-S labeling of proteins. We cannot say whether this is attributed to increased uptake and incorporation of labeled methionine, efficiency of cell lysis, or another parameter, but in side-by-side comparisons, BL-21 (DE3) consistently gives us the strongest labeling signals. Any defined media can be used, but we have found that the PASM-51 media developed by Studier (2005) yields robust expression of many different protein types in various *E. coli* expression

strains. By depleting the media of methionine, more efficient labeling is achieved.

11. The described oligonucleotides facilitate directional and specific assembly of any number of fragments, although efficiency decreases as the number of fragments increases. In our lab, 3–5 fragments (including the vector backbone) are routinely assembled using this protocol.

References

1. Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335–350
2. Nisson PE, Rashtchian A, Watkins PC (1991) Rapid and efficient cloning of Alu-PCR products using uracil DNA glycosylase. PCR Methods Appl 1:120–123
3. Smith C, Day PJ, Walker MR (1993) Generation of cohesive ends on PCR products by UDG-mediated excision of dU, and application for cloning into restriction digest-linearized vectors. PCR Methods Appl 2:328–332
4. Nour-Eldin HH, Hansen BG, Nørholm MHH et al (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res 34:e122
5. Blanusa M, Schenk A, Sadeghi H et al (2010) Phosphorothioate-based ligase-independent gene cloning (PLICing): an enzyme-free and sequence-independent cloning method. Anal Biochem 406:141–146
6. Tillett D (1999) Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. Nucleic Acids Res 27:26e
7. Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074
8. Gibson DG, Young L, Chuang R-Y et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Publ Group 6:343–345
9. Bitinaite J, Nichols NM (2001) DNA cloning and engineering by uracil excision. Wiley, Hoboken
10. Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349–361
11. LeProust EM, Peck BJ, Spirin K et al (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res 38:2522–2540
12. Nielsen MT, Madsen KM, Seppälä S et al (2014) Assembly of highly standardized gene fragments for high-level production of porphyrins in *E. coli*. ACS Synt Biol 4 (3):274–282
13. Nielsen MT, Ranberg JA, Christensen U et al (2014) Microbial synthesis of the forskolin precursor manoyl oxide in enantiomerically pure form. Appl Environ Microbiol 80 (23):7258–7265
14. Shetty RP, Endy D, Knight TF Jr (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2(5)
15. Salomonsen B, Mortensen UH, Halkier BA (2014) USER-derived cloning methods and their primer design. Methods Mol Biol (Clifton, NJ) 1116:59–72
16. Nour-Eldin HH, Geu-Flores F, Halkier BA (2010) USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. Methods Mol Biol (Clifton, NJ) 643:185–200
17. Nørholm MHH (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol 10:21
18. Cavaleiro AM, Kim SH, Seppälä S et al (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. ACS Synth Biol. doi:10.1021/acssynbio.5b00113
19. St-Pierre F, Cui L, Priest DG et al (2013) One-step cloning and chromosomal integration of DNA. ACS Synt Biol 2(9):537–541
20. Geu-Flores F, Nour-Eldin HH, Nielsen MT et al (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucl Acids Res 35, e55
21. Genee HJ, Bonde MT, Bagger FO et al (2014) Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. ACS Synt Biol 4(3):342–349

22. Nevin DE, Pratt JM (1990) A coupled in vitro transcription-translation system for the exclusive synthesis of polypeptides from the T7-promoter. FEBS Lett 291:259–263
23. Kudla G, Murray AW, Tollervey D et al (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. Science (New York, NY). 324:255–258
24. Goodman DB, Church GM, Kosuri S (2013) Causes and effects of N-terminal codon bias in bacterial genes. Science (New York, N.Y.). 342, 475–479
25. Olsen LR, Hansen NB, Bonde MT et al (2011) PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. Nucleic Acids Res 39:W61–W67

Paper III

SEVA linkers: a Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology

ACS SyntheticBiology

Technical Note

pubs.acs.org/synthbio

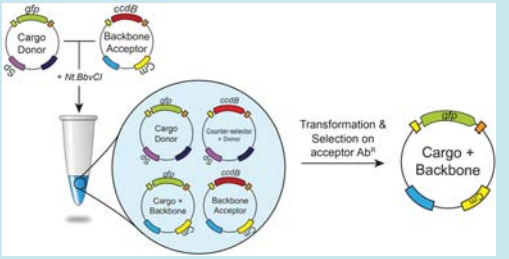
SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology

Se Hyeuk Kim, Ana Mafalda Cavaleiro, Maja Rennig, and Morten H. H. Nørholm*

Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm DK-2970, Denmark

Supporting Information

ABSTRACT: DNA vectors serve to maintain and select recombinant DNA in cell factories, and as design complexity increases, there is a greater need for well-characterized parts and methods for their assembly. Standards in synthetic biology are top priority, but standardizing molecular cloning contrasts flexibility, and different researchers prefer and master different molecular technologies. Here, we describe a new, highly versatile and automatable standard "SEVA linkers" for vector exchange. SEVA linkers enable backbone swapping with 20 combinations of classical enzymatic restriction/ligation, Gibson isothermal assembly, uracil excision cloning, and a nicking enzyme-based methodology we term SEVA cloning. SEVA cloning is a simplistic one-tube protocol for backbone swapping directly from plasmid stock solutions. We demonstrate the different performance of 30 plasmid backbones for small molecule and protein production and obtain more than 10-fold improvement from a four-gene biosynthetic pathway and 430-fold improvement with a difficult-to-express membrane protein. The standardized linkers and protocols add to the Standard European Vectors Architecture (SEVA) resource and are freely available to the synthetic biology community.



KEYWORDS: *synthetic biology standards, plasmid backbone exchange, standard parts characterization, cell factory design*

The early steps in engineering of microbial cell factories typically involve a choice of vector for gene expression. This vector enables maintenance (replication) of the genetic elements of interest, e.g., by containing an origin of replication or elements ensuring transfer to the microbial genome, and often includes a selectable trait, in bacteria typically in the form of antibiotic resistance. These initial choices can have a major impact on the performance of the cell factory, and balancing these factors is imperative to optimize production.[1,2]

Plasmids are extrachromosomal DNA elements that are nonessential, can replicate autonomously and are easily modified *in vitro* and thus represent an extremely powerful toolbox for molecular biology. The number of plasmid molecules in a single cell specifies the amount of gene copies available for expression, and this copy number is determined by different genetic elements at the origin of replication.[3] Toxicity of plasmid-encoded proteins is usually the highest metabolic burden for a production host,[4] and gene overexpression is a stress for the organism that needs to cope with this metabolic overload and prevent the system's breakdown.[2] Moreover, extra DNA elements and gene expression will likely compete with the native DNA for essential resources, e.g., the native RNA polymerase.[5,6]

In a typical laboratory setup, horizontal gene transfer events are selected for with the aid of antibiotic resistance genes. Antibiotics either inhibit bacterial cell growth (bacteriostatics) or cause bacterial cell death (bactericidals).[7] Thus, antibiotic selection is inherently linked to metabolic burden, fitness costs and physiological changes.[8−11]

Most synthetic biology and metabolic engineering projects require a first step of DNA assembly,[12−14] and with increasingly advanced design requirements, simple methods for genetic elements exchange are highly attractive. Struggling to find the best-suited assembly and exchange strategies is common among researchers since all of the described DNA cloning methods possess different limitations. Moreover, a consensus methodology is hard to agree on: best exemplified with the paradoxical high number of assembly strategies available at the Registry of Standard Biological Parts.[15] Nevertheless, standardized genetic parts and methods for their assembly are very important for continuous progress in the synthetic biology field.[16,17] For example, standards surely will enable more systematic and reliable approaches to assay the performance and robustness of genetic elements.

Here we attempt to address the standardization paradox by designing small multifunctional DNA fragments designed to link together genetic elements often used in bacterial cell factories. The resulting linkers contain elements enabling continuous reassembly with several of the most common

Special Issue: Synthetic Biology in Europe

Received: November 30, 2015

DOI: 10.1021/acssynbio.5b00257
ACS Synth. Biol. XXXX, XXX, XXX−XXX

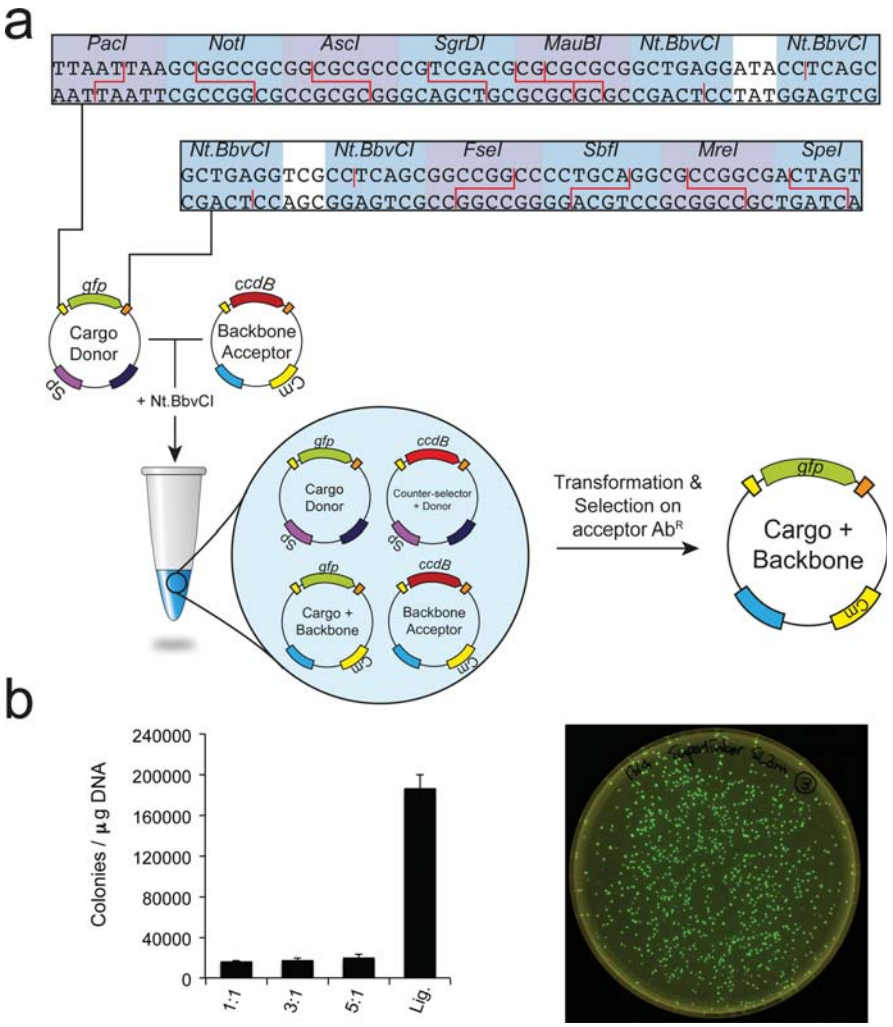


Figure 1. Illustration and optimization of SEVA-linker-based backbone exchange. (a) The cargo, here illustrated with a plasmid-encoded *gfp*, is flanked by two multifunctional linker sequences altogether hosting a total of nine rare restriction sites and four *Nt.BbvcI* nicking enzyme recognition sites (see expanded view). When mixed with the *Nt.BbvcI* enzyme and a backbone acceptor plasmid, hosting the toxic *ccdB* gene flanked with the same linkers, the recombined cargo can be selected on the antibiotic defined by the acceptor backbone. The expanded view of the reaction shows all assembly possibilities. (b) Left panel: efficiency of recombination with different ratios of backbone acceptor and cargo donor plasmids, and with the addition of T4 DNA ligase to the mixture. Right panel: representative picture showing the efficient recombination of a *gfp*-expressing cargo into a new backbone acceptor plasmid.

DNA assembly methodologies (e.g., restriction enzymes, Gibson assembly and uracil excision cloning) and with a new extraordinary simple protocol for plasmid backbone exchange. Finally, we demonstrate the usefulness of this resource by systematically comparing bacterial production of the membrane protein NarK and the food coloring pigment β-carotene produced from a four-gene biosynthetic pathway, each with a total of 30 combinations of origins of replication and antibiotic resistance markers.

In the Standard European Vector Architecture (SEVA),[18,19] different rare restriction sites flank three basic genetic elements: (1) antibiotic selection markers, (2) origins of replication and (3) the so-called cargo that contains the genetic elements necessary for the end-application of the cell factory (very often a promoter driving expression of a gene, Figure S1). Importantly, these genetic elements have well-defined borders that are useful for parts exchange and a catalogue of all 54 combinations of nine origins of replication with six antibiotic resistance markers is available.

DOI: 10.1021/acssynbio.5b00257
ACS Synth. Biol. XXXX, XXX, XXX−XXX

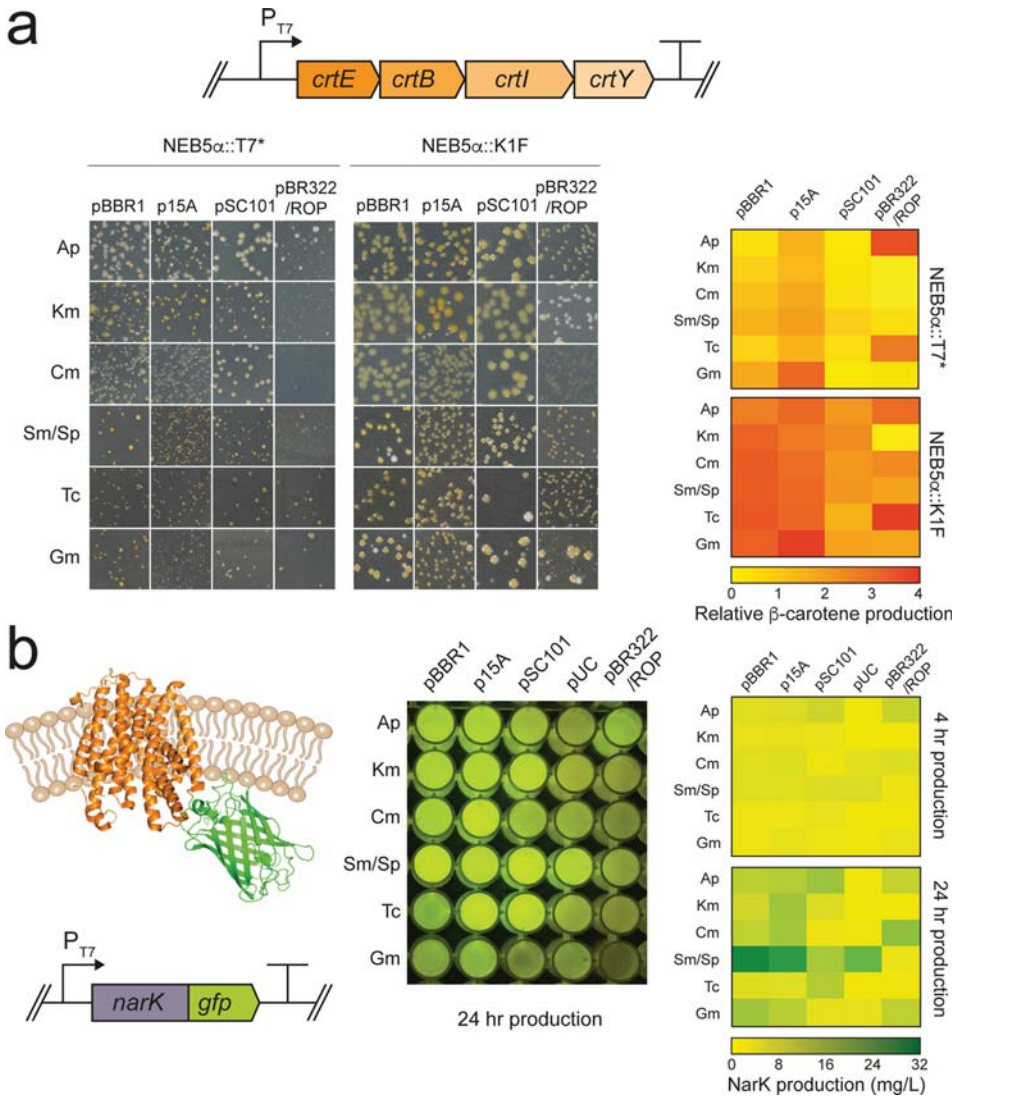


Figure 2. Production of β-carotene and the membrane protein NarK with 30 different combinations of origins of replication and antibiotic resistance markers. (a) Upper panel: illustration of the cargo with T7 promoter driven expression of the four-gene crtEBIY biosynthetic pathway. Lower left panel: colony phenotypes of the crtEBIY cargo combined with ampicillin (Ap), kanamycin (Km), chloramphenicol (Cm), spectinomycin/streptomycin (Sm/Sp), tetracycline (Tc) or gentamycin selection; and the pBBR1, p15A, pSC101 or pBR322/ROP origins of replication. All 30 combinations were transformed into both *E. coli* NEB5α::T7* (left side) and NEB5α::K1F (right side) hosting two different variants of the T7 RNA polymerase. No combinations with the high copy pUC origin of replication yielded surviving colonies. Lower right panel: heat map representation of carotenoid levels measured by acetone extraction and absorbance at 453 nm on the 24 viable backbones variants in NEB5α::T7* and NEB5α::K1F. (b) Left panel: illustration of the NarK-GFP protein (based on the pdb files: 1EMA and 4U4V) and cargo constructs. Middle panel: Fluorescence from *E. coli* BL21 (DE3) transformed with 30 different backbones in combination with the T7-narK-gfp cargo. Right panel: heat map representation of fluorescence levels quantified in a microplate reader after four and 24 h expression.

To make this great resource compatible with a range of state-of-the-art DNA assembly methods, we designed two multifunctional SEVA linker sequences that hosted five and four additional rare restriction sites, respectively (*Pac*I, *Not*I, *Asc*I, *Sgr*DI, *Mau*BI on one side and *Fse*I, *Sbf*II, *Mre*I, *Spe*I on the other side, Figure 1a). This enables backbone swapping with a total of 20 combinations of these enzymes. Additionally, all cargo elements, flanked with these linkers, are easily inserted in

DOI: 10.1021/acssynbio.5b00257
ACS Synth. Biol. XXXX, XXX, XXX−XXX

the existing SEVA system using the outermost *Pac*I and *Spe*I sites (but violates the SEVA design rules by reusing the *Fse*I and *Asc*I sites, Figure S1). The saturation with rare restriction sites means that it should always be possible to find a suitable pair of restriction enzymes for backbone swapping. In other words this minimizes the consequence of "forbidden sites" in the cargo. Moreover, these unique linkers sequences make it possible to design highly specific oligonucleotides (for examples see Supporting Information Table S1) that anneal in these regions, while hosting features compatible with state-of-the-art DNA assembly methods such as Gibson isothermal assembly[20] and uracil excision[21] (Supporting Information Table S1, Figure S2). Such oligonucleotides can be performance benchmarked and kept in the freezer for continuous reuse when assembling new parts with these technologies.

Another recently popularized assembly method is Golden Gate cloning.[22] One of the biggest advantages of Golden Gate cloning is that DNA can be exchanged directly from (compatible) plasmid stock solutions with a simple protocol, whereas a drawback is the frequent occurrence of the type IIS restriction sites typically used: "forbidden sites". Inspired by some of the features in Golden Gate cloning, we designed an extension to the SEVA linkers enabling backbone exchange with a very simple protocol directly from plasmid stocks. Instead of type IIS restriction sites, our design uses two pairs of nicking restriction sites (Figure 1a) that together form two different 7 bp cohesive ends (Figure 1a and Supporting Information Figure S3). We initially used the *Nb.Bts*I nicking enzyme, but changed to *Nt.Bbv*CI because the recognition site is 7 bp and only occurs rarely in standard sized DNA constructs. Furthermore, because the mutant enzyme only cuts one strand, occurrences outside the SEVA linker will probably not affect cloning efficiency significantly. This solves the forbidden site issue.

We designed two SEVA-linker-flanked cargo elements, one expressing *lacZ* and one expressing *gfp*, with different antibiotic selection markers, and checked their ability to recombine by mixing the plasmids together with the nicking enzyme and plating on media with the different antibiotics (Supporting Information Figure S4). Ultimately, the idea is that a cargo "donor" plasmid should be transferred to a backbone "acceptor" plasmid at high efficiency and specificity. To this end, we incorporated the toxic *ccdB* gene[23] in the backbone acceptor and selected recombinants by transforming constructs into standard (i.e., *ccdB*-incompatible) cloning strains with the backbone-defined antibiotic selection (Figure 1a). As an initial proof of concept, we examined *gfp* flanked by the two SEVA linkers in the commercially available pCDF backbone (CloDF13 origin and spectinomycin resistance) transferred to an acceptor plasmid containing *ccdB* in the pACYC backbone (p15A origin and chloramphenicol resistance). We then observed fluorescent colonies forming on chloramphenicol-containing plates with high efficiency (Figure 1b). Optimization of the ratio of donor and acceptor plasmids in the reaction increased efficiency approximately 2-fold, whereas adding T4 DNA ligase to the mixture enhanced efficiency more than 10-fold (Figure 1b). The optimized and simple protocol is described in detail in the Supporting Information. Notably, rather than getting it right the first time, several similar nicking enzyme-based designs were tested for efficiency and specificity before we settled on the sequence presented here (see Supporting Information Figure S5).

We based our further work on the comprehensive pSEVA collection as a backbone acceptor series. By combining five origins of replication (pBBR1: #3, p15A: #6, pSC101: #7, pUC: #8, pBR322/ROP: #9) and six antibiotic resistance markers (ampicillin: #1, kanamycin: #2, chloramphenicol: #3, spectinomycin: #4, tetracycline: #5, gentamycin: #6), we created 30 different backbone acceptors with the counter-selection marker *ccdB* flanked by SEVA linkers as the initial cargo.

To assay the performance of our 30 standardized plasmid backbones in two typical, cell factory-type experimental settings, we swapped-in the four-gene *crtEBIY* biosynthetic pathway for *β*-carotene from *Pantoea ananatis* and the membrane protein-encoding *narK-gfp* in all 30 constructs. *crtEBIY* is industrially relevant and a convenient model pathway mainly due to the simple product output (orange color), but also because the robustness of the phenotype seems sensitive to the cell factory design parameters.[24] Indeed, we were unable to obtain surviving colonies by swapping the pathway into vectors with the high-copy pUC origin of replication and we obtained highly variable phenotypes when T7 polymerase was used to drive expression from the construct in comparison with the weaker K1F variant (Figure 2a).[24−26] Overall, we observed a 10-fold difference in the *β*-carotene product titers, from the lowest to the highest performing cell factories (Supporting Information Table S3) and several of the combinations showed clear toxicity and population bias effects (e.g., pBR322/ROP in combination with chloramphenicol or pSC101 in combination with tetracycline, Figure 2a). In contrast, the pBR322/ROP origin in combination with both ampicillin and tetracycline was highly performing in both expression strains. With the membrane protein NarK,[27] the variation in expression yield was even more prominent (Figure 2b); the difference between the highest and the lowest performing combination of parts was in this case an impressive 430-fold (Supporting Information Table S3). Generally, using tetracycline selection, gentamycin selection or the high copy pUC origin had a negative impact on *narK* expression, whereas the p15A origin (low copy number) and spectinomycin selection seemed to positively impact the expression level.

Comparing the two different test cases, small molecule and protein production, showcases the value of a synthetic biology approach (i.e., systematic studies with standardized parts) and provides future design guidelines and a toolbox for similar experiments. For example, the high copy pUC origin of replication is likely a poor choice for anything but DNA production. The negative impact of the tetracycline selection may in contrast only reflect the fact that the resistance gene encodes a membrane protein that could compete for factors involved in membrane translocation important for production of NarK, and thus may be a particularly poor choice for membrane protein production. In many cases we observed clear population bias effects by simple visual inspection on agar plates, and the different robustness of the T7- and the K1F-based bacterial hosts highlights the value of adding promoter tuning as an extra dimension in the cell factory performance screen. These observations could be supported by an array of omics studies leading to a highly informed theoretical framework for rational cell factory design.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00257.

DOI: 10.1021/acssynbio.5b00257
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Supporting tables, figures, and methods. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: morno@biosustain.dtu.dk. Phone: +45 217-99184. Fax: +45-353-33300.

Author Contributions

S.H.K. and M.H.H.N. designed the experiments. S.H.K., A.M.C. and M.R. performed the experiments. S.H.K., A.M.C., M.R. and M.H.H.N. wrote the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Victor de Lorenzo and his research group for providing the SEVA collection, for fruitful discussions and for comments on the manuscript. Lone Riisberg and Pernille Smith are thanked for technical assistance. This work was supported by The Novo Nordisk Foundation and the Ph.D. grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme [FP7-People-2012-ITN], under grant agreement No. 317058, "BACTORY".

REFERENCES

(1) Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990) Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng. 35*, 668−681.

(2) Grabherr, R., Nilsson, E., Striedner, G., and Bayer, K. (2002) Stabilizing plasmid copy number to improve recombinant protein production. *Biotechnol. Bioeng. 77*, 142−147.

(3) Friehs, K. (2004) Plasmid copy number and plasmid stability. In *New Trends and Developments in Biochemical Engineering*, pp 47−82, Springer, Berlin.

(4) Corchero, J. L., and Villaverde, A. (1998) Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng. 58*, 625−632.

(5) Ceroni, F., Algar, R., Stan, G. B., and Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods 12*, 415−418.

(6) Weisse, A. Y., Oyarzun, D. A., Danos, V., and Swain, P. S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U. S. A. 112*, E1038−1047.

(7) Kohanski, M. a, Dwyer, D. J., and Collins, J. J. (2010) How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol. 8*, 423−435.

(8) Andersson, D. I., and Levin, B. R. (1999) The biological cost of antimicrobial resistance. *Curr. Opin. Microbiol. 2*, 489−493.

(9) Martínez, J. L., and Rojo, F. (2011) Metabolic regulation of antibiotic resistance. *FEMS Microbiol. Rev. 35*, 768−789.

(10) Schulz zur Wiesch, P., Engelstadter, J., and Bonhoeffer, S. (2010) Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrob. Agents Chemother. 54*, 2085−2095.

(11) Levin, B. R., and Rozen, D. E. (2006) Non-inherited antibiotic resistance. *Nat. Rev. Microbiol. 4*, 556−562.

(12) Chao, R., Yuan, Y., and Zhao, H. (2014) Recent advances in DNA assembly technologies. *FEMS Yeast Res. 15*, 1−9.

(13) Cobb, R. E., Ning, J. C., and Zhao, H. (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol. 41*, 469−477.

(14) Keasling, J. (2008) Synthetic biology for synthetic chemistry. *ACS Chem. Biol. 3*, 64−76.

(15) Vilanova, C., and Porcar, M. (2014) iGEM 2.0−refoundations for engineering biology. *Nat. Biotechnol. 32*, 420−424.

(16) Casini, A., Storch, M., Baldwin, G. S., and Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol. 16*, 568−576.

(17) Vilanova, C., Tanner, K., Dorado-Morales, P., Villaescusa, P., Chugani, D., Frías, A., Segredo, E., Molero, X., Fritschi, M., Morales, L., Ramón, D., Peña, C., Peretó, J., and Porcar, M. (2015) Standards not that standard. *J. Biol. Eng. 9*, 17.

(18) Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., Heras, A. D. L., Páez-Espino, A. D., Durante-Rodríguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res. 41*, 666−675.

(19) Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., and de Lorenzo, V. (2015) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res. 43*, D1183−D1189.

(20) 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., and Hutchison, C. A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U. S. A. 105*, 20404−20409.

(21) Nour-eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res. 34*, e122.

(22) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One 3*, e3647.

(23) Shetty, R. P., Endy, D., and Knight, T. F. J. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng. 2*, 5.

(24) Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T., and Nørholm, M. H. H. (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. *ACS Synth. Biol. 4*, 1042−1046.

(25) Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppälä, S., and Nørholm, M. H. H. (2015) Uracil excision for assembly of complex pathways. In *Springer Protocol Handbooks*, pp 1−11, Humana Press, Springer-Verlag, Berlin.

(26) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res. 40*, 8773−8781.

(27) Mirzadeh, K., Martínez, V., Toddo, S., Guntur, S., Herrgård, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector−coding sequence junction. *ACS Synth. Biol. 4*, 959−965.

NOTE ADDED AFTER ASAP PUBLICATION

In Figure 1 the red lines indicating how the restriction enzymes *Fse*I and *Sbf*I cut in the DNA were changed on March 9, 2016.

DOI: 10.1021/acssynbio.5b00257
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Supporting Information for "SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology"

Materials and Methods

Strains, media and plasmids

Escherichia coli NEB5α (New England Biolabs, Ipswich, MA, USA) was used for propagation of plasmids, backbone swapping optimization and as a general cloning host except when *E. coli* DB3.1 (Thermo Fisher Scientific, Invitrogen, Waltham, MA, USA) was used for handling of *ccdB*-containing plasmids. NEB5α::T7* and NEB5α::K1F[1] were used for β-carotene production. SOC media was used as a recovery media after transformation. Bacteria were propagated in Luria-Bertani (LB) liquid media or agar plates supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml), chloramphenicol (50 μg/ml), spectinomycin (50 μg/ml), tetracycline (10 μg/ml), or gentamycin (10 μg/ml) when required. In most cases LB media was used for liquid cultures except for carotenoid production assessment that was accomplished in 2×YT media supplemented with 0.5% glycerol.

Molecular biology reagents

T4 DNA ligase and restriction enzymes were purchased from Thermo Fischer Scientific (Waltham, MA, USA). Nicking enzymes were purchased from New England Biolabs (Ipswich, MA, USA). Plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany). PCR products were purified using a PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Foster City, USA). Buffers for PCR and cloning reactions were

purchased from Agilent Technologies (Santa Clara, CA, USA) and Thermo Fischer Scientific (Wilmington, USA). PCR was performed with the proofreading PfuX7 polymerase as previously described.[3]

Plasmid constructions

SEVA linker sequences (version #1, see Supporting Information Figure S3) were introduced into pACYCDuet-1 and pCDFDuet-1 (for references to plasmids see Supporting Information Table S2) by amplifying the backbones with oligonucleotides #1 and #2 (for numbering and sequences of oligonucleotides see Supporting Information Table S1), *gfp* from pETDuet-1-*gfp* with the oligonucleotides #3 and #4 and *lacZα* from pBluescript II KS (+) with oligonucleotides #5 and #6, followed by assembly by uracil excision as described previously.[1, 2] This created pACYC-sl1- *gfp* and pCDF-sl1-*gfp*.

The *ccdB* gene was amplified from pOSIP-KT using oligonucleotides #7 and #8 and mixed with a pACYC-sl1 fragment obtained from *Nb.BtsI*-digested pACYC-sl1-*gfp* creating pACYC-sl1-*ccdB* upon transformation into *E. coli*.

pCDF-sl2-*gfp* was constructed by amplifying the vector backbone and *gfp* insert from pCDF-sl1-*gfp* with oligonucleotides #9 and #10, and #11 and #12, respectively, treating the PCR products with *Nb.BtsI* and transforming the fragments into *E. coli*. pACYC-sl2-*ccdB* was similarly obtained by combining *Nb.BtsI*-digested pACYC-sl1-*ccdB* with a backbone fragment amplified from pACYC-sl1-*ccdB* with oligonucleotides #9 and #10.

pCDF-sl3-*gfp* and pACYC-sl3-*ccdB* were created by amplifying the corresponding sl2 versions with oligonucleotides #13 and #14, *gfp* with #15 and #16, and *ccdB* with #17 and #18, followed by *Nt.BbvCI*-treatment and transformation.

pCDF-sl4-*gfp* and pACYC-sl4-*ccdB* were created by amplifying the corresponding sl2 versions with the oligonucleotides #19 and #20, *gfp* with #21 and #22, and *ccdB* with #23 and #24 followed by uracil excision cloning.

pSEVAXX-sl3-*ccdB* series was constructed by combining five origins of replication (pBBR1; #3, p15A; #6, pSC101; #7, pUC; #8, pBR322/ROP; #9), prepared by digestion with *PacI* and *PshAI* from the corresponding parts in the pSEVA collection, and six antibiotic resistance markers (ampicillin; #1, kanamycin; #2, chloramphenicol; #3, spectinomycin; #4, tetracycline; #5, gentamycin; #6), prepared by digestion with *PshAI* and *SpeI* from the corresponding parts in the pSEVA collection, with the sl3-*ccdB* cargo isolated after digestion by *PacI* and *SpeI* from pACYC-sl3-*ccdB*.

The β-carotene biosynthetic pathway was introduced into pACYC-sl3 by SEVA cloning (see below) after amplifying the *crtEBIY* operon from pSIJ31B with the oligonucleotides #25 and #26, creating pACYC-sl3-T7-*crtEBIY*. Combining pACYC-sl3-T7-*crtEBIY* with the pSEVAXX-sl3-*ccdB* series as described below created the pSEVAXX-sl3-T7-*crtEBIY* series.

The T7-*narK-gfp* cargo was amplified from pET28a-*narK*^{WT}P7-6 using oligonucleotides #29 and #30, and mixed with a pSEVAXX-sl3 backbone amplified with oligonucleotides #27 and #28 followed by uracil excision cloning. Combining the initial pSEVAXX-sl3-T7-*narK-gfp* clone with the pSEVAXX-sl3-*ccdB* series created the complete pSEVAXX-sl3- T7-*narK-gfp* series.

Nicking enzyme mediated one-tube backbone exchange (SEVA cloning)

0.06 pmol of each plasmid were added to a 10 or 20 µl total reaction volume containing CutSmart® buffer (New England Biolabs, Ipswich, MA, USA) and five

units of *Nt.BbvCI*. Mixtures were kept at 37°C for 1 h, followed by 25°C for 15 min, 10°C for 10 min, then incubation at 0°C using a C1000 Touch™ Thermal Cycler (Bio-Rad, Hercules, CA, USA). 2.5 U of T4 DNA ligase and buffer were added, followed by incubation at room temperature for 15 min and storage on ice prior to transformation.

Gibson assembly- and uracil excision-based backbone exchange

The GFP cargo was PCR amplified from pCDF-sl3-*gfp* using the oligonucleotides #31 and #32 for Gibson assembly and #35 and #36 for uracil excision; and the backbone was amplified from pACYC-sl3-*ccdB* using the oligonucleotides #33 and #34 (Gibson) or #37 and #38 (uracil excision). After PCR, *Dpn*I treatment was done at 37°C for 40 min followed by gel purification. Insert and vector ratio of 2:1 was applied for both methods. Gibson assembly was performed with 2X Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA, USA) following the manufacturer´s instructions. Assembly by uracil excision was as described previously.[1, 2]

Production and relative quantification of β-carotene

The NEB5αT7* and NEB5αK1F strains[1] were used to compare β-carotene productivity in the pSEVA-*crtEBIY* series. Corresponding strains with the pathway integrated in one copy on the genome (NEB5αT7*::EBIY and NEB5αK1F::EBIY) were used as reference strains.[1] Cells were grown in 2×YT media supplemented with 0.5% glycerol at 30°C for 72 hours with 300 rpm. 1 ml of each culture was harvested by centrifugation at 13000 rpm for 5 min. After discarding the supernatant, cells were washed once with 1 ml of water. 1 ml of acetone was added and the pellets re-suspended vigorously by vortexing, followed by incubation at 55°C for 20 min at

1000 rpm using a tabletop shaker. The remaining cell debris were removed by centrifugation (13000 rpm, 5 min) and absorbance at 453 nm was measured in a UV-1600PC spectrophotometer (VWR International, Radnor, PA, USA) on 500 µl extract using a quartz cuvette. The equation for calculating β-carotene production titers is $\frac{A \times Vol(mL) \times 10^4}{E_{1cm}^{1\%}}$ where A is the UV/Vis absorbance at 453 nm and $E_{1cm}^{1\%}$ is the specific absorption coefficient ($E_{1cm}^{1\%}$, β-carotene=2503).[4]

Production and quantification of NarK-GFP

The pSEVAXX-sl3- T7-*narK-gfp* series was transformed into BL21 (DE3) (Novagen, Madison, WI, USA) using a standard protocol. Overnight cultures were prepared by inoculating a single colony in 800 µL of LB liquid media containing the different pSEVA-defined antibiotics (100 µg/ml ampicillin; 50 µg/ml kanamycin; 34 µg/ml chloramphenicol; 50 µg/ml spectinomycin; 10 µg/ml tetracycline; 10 µg/ml gentamycin) in 96-deep well plates at 37 °C and 300 rpm in an Innova 44 incubator (Thermo Scientific, Waltham, MA, USA). For expression measurements, overnight cultures were back-diluted 1:50 in 3 ml LB media containing the different antibiotics in 24-well plates at 37 °C and 300 rpm. When exponential phase (an OD_{600} of approximately 0.5) was reached, expression was induced by addition of 1 mM IPTG and incubated at 25 °C and 300 rpm for 2 hours. Subsequently, 1 ml of culture was harvested at 2270 x g for 20 min, resuspended in a buffer (50 mM Tris·HCl (pH 8.0), 200 mM NaCl, and 15 mM EDTA) and incubated for 2 hours at room temperature. Fluorescence was measured in a 96-well plate in a Synergy™ Mx plate reader (BioTek) (excitation wavelength 485 nm; emission wavelength 512 nm). The amount of protein produced was estimated from a GFP standard curve. The standard curve was obtained from purified GFP mixed with BL21 (DE3) cells to account for

quenching effects. 22 h after induction the remaining culture was harvested, resuspended in buffer, incubated for 2 hours and fluorescence measured as described above.

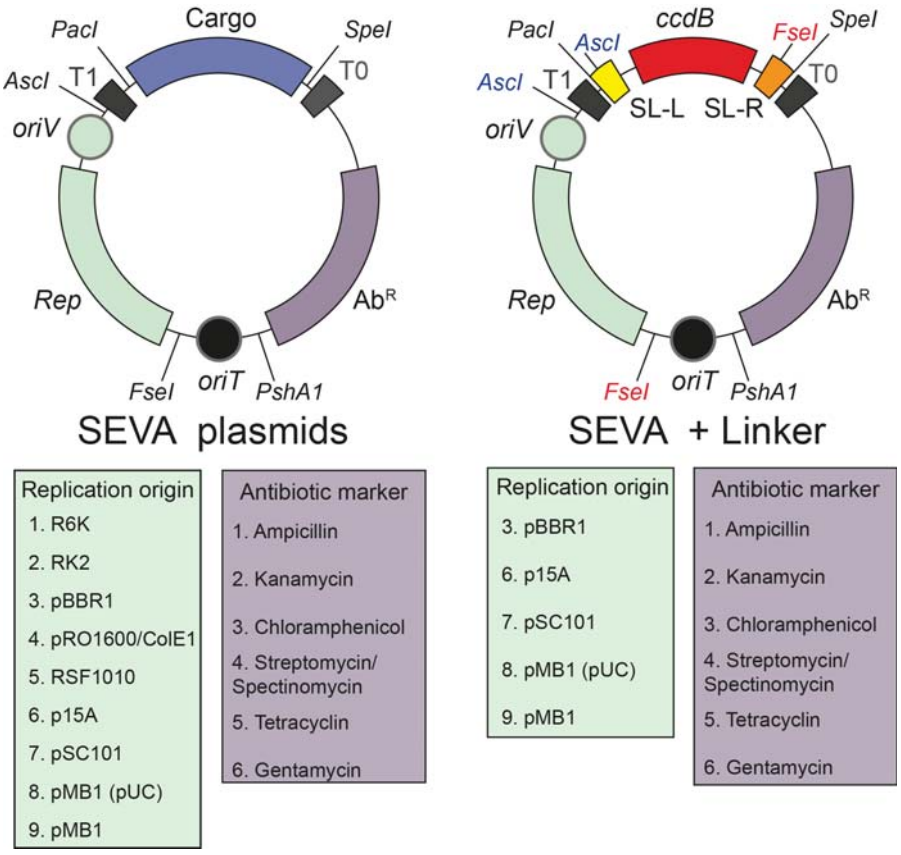


Figure S1. Illustration of the Standard European Vector Architecture (SEVA) system and the relationship to the SEVA linkers and plasmids described here. Left panel: In the original SEVA system, rare restriction sites flank the three basic components: origins of replication, antibiotic selection markers and the cargo. This basic design enables exchange of the basic components using classical restriction enzyme molecular cloning. Right panel: The new SEVA linkers enable simple, one-pot backbone shuffling by introducing two multifunctional linker sequences (highlighted in yellow and orange color) flanking the cargo in the SEVA system. In this example, the *ccdB* counterselection marker is the cargo. Any cargo that is flanked by SEVA linkers can be converted to the SEVA system by utilizing the *PacI* and *SpeI* restriction sites, but also by a range of other molecular cloning technologies (see main text). The SEVA linkers violate the basic SEVA design rules by reusing the *AscI* and *FseI* sites (highlighted in blue and red font).

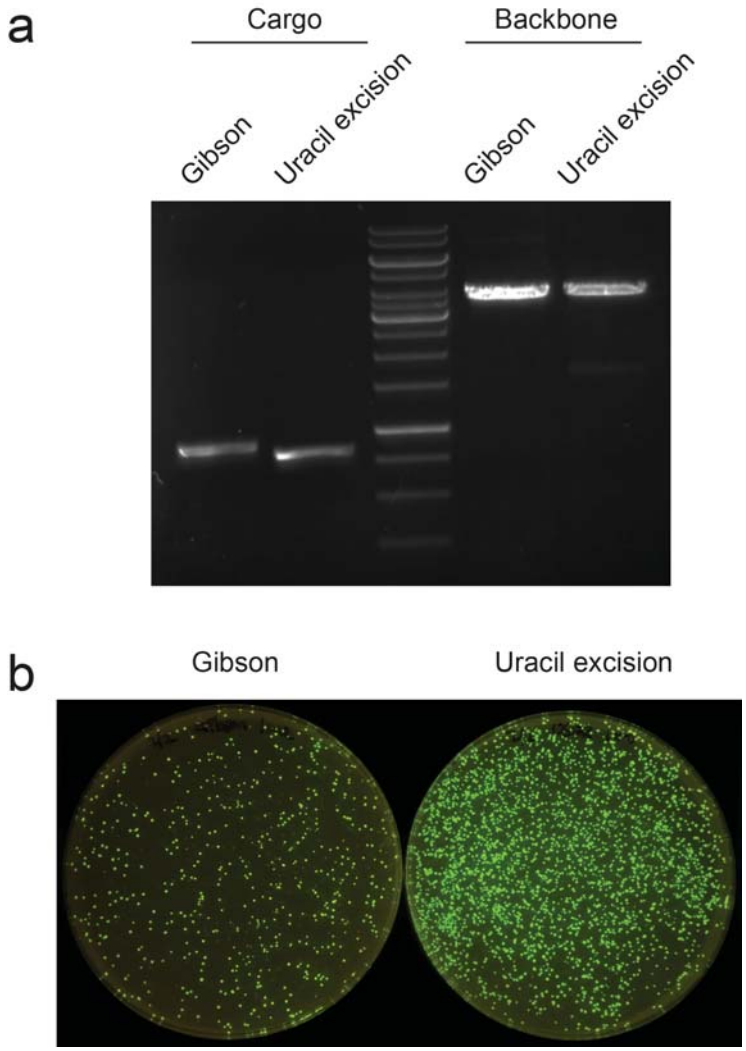


Figure S2. Demonstration of amplification and reassembly of SEVA linker flanked cargo and backbone with Gibson assembly and uracil excision cloning. (a) Agarose gel showing PCR amplified cargoes and backbones compatible with Gibson assembly or uracil excision cloning. In the cargo, SEVA linkers flanked a transcriptional unit (a leaky Ptrc promoter driving expression of *gfp*). For details on the protocol see Materials and Methods. (b) PCR-amplified cargo and backbone was reassembled with Gibson and uracil excision cloning and plated on agar plates supplemented with the antibiotic corresponding to the backbone selection marker. Green fluorescent colonies demonstrate the presence of the cargo.

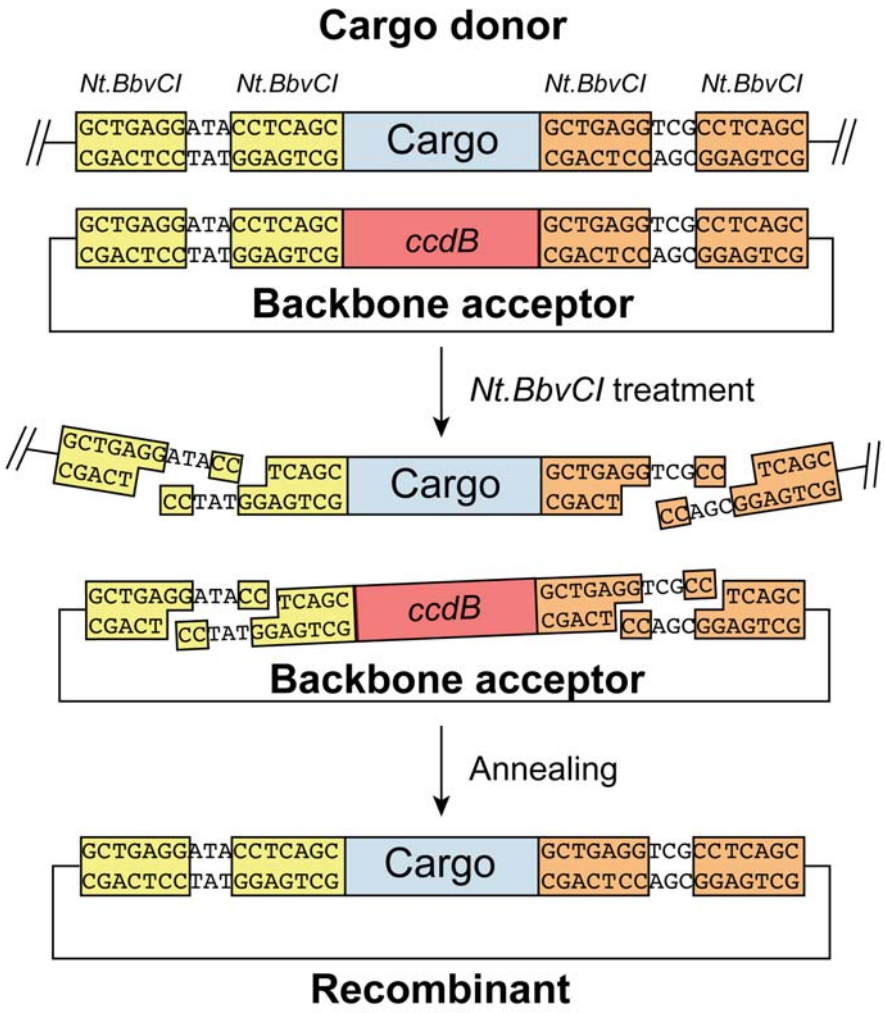


Figure S3. Exchange of DNA fragments using four *Nt.BbvCI*- nicking enzyme- sites in the SEVA linkers. Yellow and orange boxes mark the four *Nt.BbvCI* recognition sites. The two three-nucleotide spacers between the two double nicking sites are different from each other, thereby ensuring specificity and directionality in the parts exchange. Any cargo (light blue box) of choice can replace the *ccdB* (red box) counterselection marker.

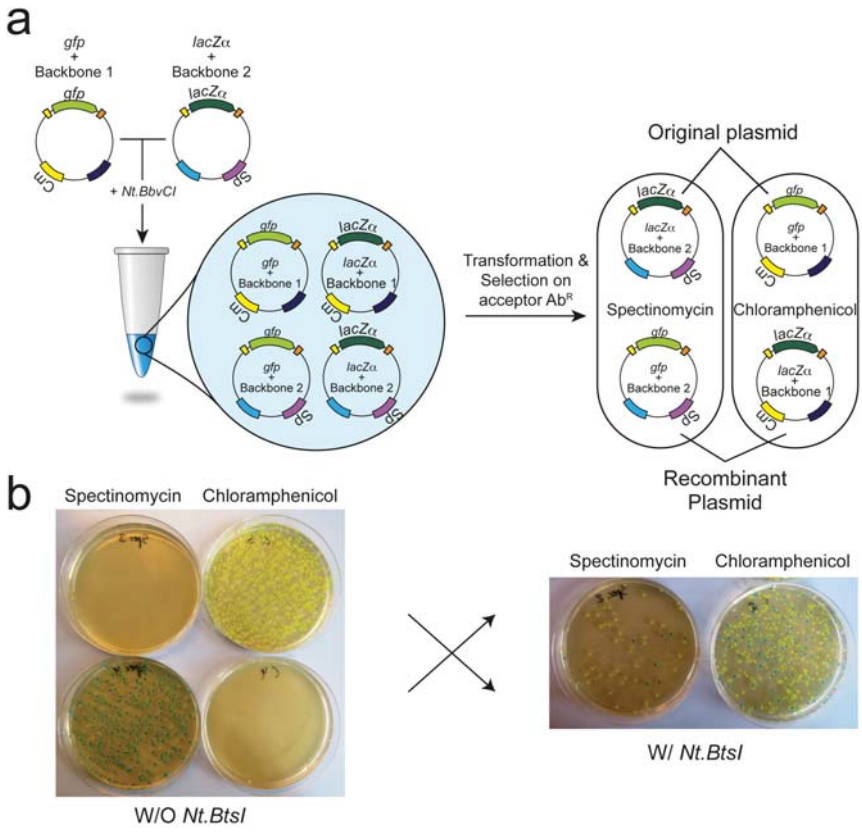


Figure S4. Demonstration of parts exchange between two plasmids that contain SEVA linkers. (a) Illustration of a *gfp* expressing SEVA linker cargo in a plasmid that confers chloramphenicol resistance, mixed with a *lacZα* expressing SEVA linker cargo in a plasmid conferring spectinomycin resistance, leading to four different recombinant plasmids. (b) Left panel: The two different plasmids were transformed into NEB5α and plated on LB agar supplemented with X-gal combined with spectinomycin or chloramphenicol. Right panel: The two plasmids were mixed with the nicking enzyme *Nb.BtsI*, transformed into NEB5α and plated on LB X-gal agar.

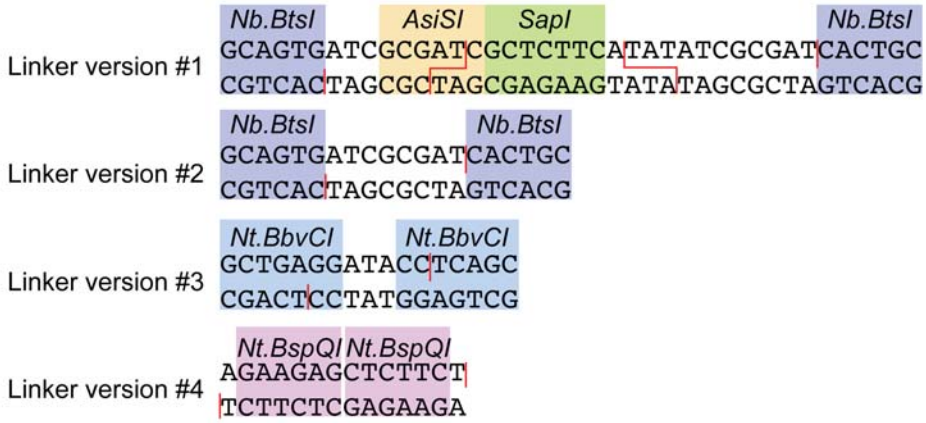


Figure S5. Different types of nicking enzyme-based linkers that were tested for cloning efficiency and accuracy. Version #2 outperforms version #1 in simplicity, but the *Nb.BtsI*-based designs generally performed poorly, probably due to the frequent occurrence of the six-nucleotide recognition sites outside the SEVA linkers. Version #4 likewise performed poorly, probably because the single-strand overhangs formed by the *Nt.BspQI* sites were too stable. Version #3 clearly outperformed the other designs in terms of cloning efficiency and accuracy and was the preferred choice for the SEVA linkers.

Table S1. Oligonucleotides used in this study

No.	NAME	SEQUENCE
1	Duet-sl1-rev	ATCGCGAUCACTGCCGCGCGCGTCGACGGGCGCGCCGCGGCCGCTTAATTA ACAAAATTATTTCTACAGGGGAATTGTTATCCGCTC
2	Duet-sl1-fwd	AGAGCGAUCGCACTCACTGCGGCCGGCCCCTGCAGGCGCCGGCGACTAGTCCT AGGCTGCTGCCACCGCTG
3	Sl1-PtrcGFP_fwd	ATCGCGAUCGCTCTTCATATATCGCGATCACTGCTTGACAATTAATCATCCGGC TCGTATAATG
4	Sl1-PtrcGFP_rev	ATCGCTCUTCATATATCGCGATCACTGCTTATTTGTAGAGCTCATCCATGCCAT GTG
5	Sl1-lacZa_fwd	ATCGCGAUCGCTCTTCATATATCGCGATCACTGCACCAGTGGNTCATCTCCAAG CAGTGGTTCGCGCAACGCAATTAATGTGAG
6	Sl1-lacZa_rev	AGTGCGAUCGCTCTTCATATAGTGCGATCACTGCACCAGTGCTACCTCCTGAAC CAC
7	Sl1-ccdB_fwd	ATC GCG AUC GCT CTT CAT ATA TCG CGA TCA CTG CTA CTA AAA GCC AGA TAA CAG TAT GCG TAT
8	Sl1-ccdB_rev	AGT GCG AUC GCT CTT CAT ATA GTG CGA TCA CTG CCG GGT TAT TAT ATT CCC CAG AAC ATC AG
9	Nb.BtsI_SL_F (vector)	ATCGCactCACTGCGGCCGGCCCCTG
10	Nb.BtsI_SL_R (vector)	ATCGCgatCACTGCCGCGCGCGCGTC
11	Nb.BtsI_PtrcGFP_F	ATCGCGATCACTGCTTGACAATTAATCATCCGGCTCGTATAATG
12	Nb.BtsI_PtrcGFP_R	AGTGCGATCACTGCTTATTTGTAGAGCTCATCCATGCCATGTG
13	Nt.BbvCI_V_F	GGTCGCCTCAGCGGCCGGCCCCTGCAGGCG
14	Nt.BbvCI_V_R	GGTATCCTCAGCCGCGCGCGCGTCGACGGG
15	Nt.BbvCI_GFP_F	GGATACCTCAGCTTGACAATTAATCATCCGGC
16	Nt.BbvCI_GFP_R	GGCGACCTCAGCTTATTTGTAGAGCTCATCCATGC
17	Nt.BbvCI_CCDB_F	GGATACCTCAGCTACTAAAAGCCAGATAACAGTATGC
18	Nt.BbvCI_CCDB_R	GGCGACCTCAGCCGGGTTATTATATTCCCCAG
19	Nt.BspQ1_V_R_U	AGA AGA GCT CTT CUC GCG CGC GCG TCG ACG GG
20	Nt.BspQ1_V_F_U	AGAAGAGCGCTCTTCUGGCCGGCCCCTGCAGGCG
21	Nt.BspQ1_GFP_F_U	AGAAGAGCTCTTCUTTGACAATTAATCATCCGGC
22	Nt.BspQ1_GFP_R_U	AGA AGA GCG CTC TTC UTT ATT TGT AGA GCT CAT CCA TGC
23	Nt.BspQ1_ccdB_F_U	AGAAGAGCTCTTCUtactaaaagccagataacagtatgc
24	Nt.BspQ1_ccdB_R_U	AGA AGA GCG CTC TTC UCG GGT TAT TAT ATT CCC CAG AAC ATC AG
25	Nt.BbvCI_EBIY_F	GGATACCTCAGCGGATCTCGACGCTCTCCC
26	Nt.BbvCI_EBIY_R	GGCGACCTCAGCGATTATGCGG
27	SL3-backbone_rev	AGGTATCCUCAGCCGCGCG
28	SL3-backbone_fwd	AGCTGAGGUCGCCTCAGC
29	NarK-sl3_fwd	AGGATACCUCAGCTAATACGACTCACTATAGGG
30	NarK-sl3_rev	ACCTCAGCUCAGTGGTGGTGG
31	SL-Gibson_ifwd	TTAATTAAGCGGCCGCGGCGCGCCCGTCGA
32	SL-Gibson_irev	ACTAGTCGCCGGCGCCTGCAGGGGCCGGCC
33	SL-Gibson_bbfwd	GGCCGGCCCCTGCAGGCGCCGGCGACTAGT
34	SL-Gibson_bbrev	TCGACGGGCGCGCCGCGGCCGCTTAATTAA
35	SL3m-Ptrc_U_F	AGGATACCUCAGCTTGACAATTAATC
36	SL3m-GFP_U_R	ACCTCAGCUTATTTGTAGAGCTC
37	pCDF_BB_U_F	AGCTGAGGUCGCCTCAGC
38	pCDF_BB_U_R	AGGTATCCUCAGCCGCGCG

Table S2. Strains and plasmids used in this study

Strain/plasmid	Property	Source/Reference
Strains		
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB
E. coli DB3.1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm^r) xyl5 Δleu mtl1	Thermo Fisher Scientific
NEB5α::T7*	NEB5α with a T7* RNA polymerase integrated	(1)
NEB5α::K1F	NEB5α with a T7*(K1F) RNA polymerase integrated	(1)
BL21 (DE3)	F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Novagen
Plasmids		
pCDFDuet-1	Cloning and expression vector, Sp^R	Novagen
pACYCDuet-1	Cloning and expression vector, Cm^R	Novagen
pETDuet-1-*gfp*	Constitutively expressed GFP, Amp^R	(1)
pBluescript II KS(+)	*lacZa* expressed from lac promoter	Agilent Technology
pCDF-sl1- *lacZa*	*lacZα* fragment flanked by SEVA linker ver1, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl1-*gfp*	*gfp* flanked by SEVA linker ver1, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl2-*gfp*	*gfp* flanked by SEVA linker ver2, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl2-*ccdB*	*ccdB* flanked by SEVA linker ver2, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl3- *gfp*	*gfp* flanked by SEVA linker ver3, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl3-*ccdB*	*ccdB* flanked by SEVA linker ver3, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl4- *gfp*	*gfp* flanked by SEVA linker ver4, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl4-*ccdB*	*ccdB* flanked by SEVA linker ver4, Cm^R, p15A origin from pACYCDuet-1	This study
pSIJ31B	*P. ananatis crtEBIY* operon, Sp^R, CloDF13 origin	Unpublished
pOSIP-KT	P21 Integration module, *ccdB*, Km^R, pUC origin	(5)
pET28a-*narK*^{WT}P7-6 P7-6	*narK*-gfp with T7 promoter, Km^R, pBR322/ROP origin	(6)
pSEVA13-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA18-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pUC origin	This study
pSEVA19-sl3-*ccdB*	*ccdB* flanked by SEVAlinkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA28-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pUC origin	This study
pSEVA29-sl3-	*ccdB* flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study

ccdB		
pSEVA33-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, p15A origin	This study
pSEVA37-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA38-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pUC origin	This study
pSEVA39-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pBBR1 origin	This study
pSEVA46-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA48-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pUC origin	This study
pSEVA49-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pBBR1 origin	This study
pSEVA56-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA58-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pUC origin	This study
pSEVA59-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pBBR1 origin	This study
pSEVA66-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA68-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pUC origin	This study
pSEVA69-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study
pSEVA13-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA19-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA29-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study
pSEVA33-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3-	*crtEBIY* operon flanked by SEVA linkers, Cm^R, p15A	This study

T7-*crtEBIY*	origin	
pSEVA37-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA39-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pBBR1 origin	This study
pSEVA46-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA49-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pBBR1 origin	This study
pSEVA56-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA59-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pBBR1 origin	This study
pSEVA66-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA69-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study
pSEVA13NarK/GFP	*narK-gfp* flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA18-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pUC origin	This study
pSEVA19-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA28-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pUC origin	This study
pSEVA29-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study
pSEVA33-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, p15A origin	This study
pSEVA37-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA38-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pUC origin	This study
pSEVA39-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3-	*narK-gfp* flanked by SEVA linkers, Sp^R, pBBR1 origin	This study

T7-*narK-gfp*		
pSEVA46-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA48-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pUC origin	This study
pSEVA49-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53 -sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pBBR1 origin	This study
pSEVA56 -sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA58-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pUC origin	This study
pSEVA59-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pBBR1 origin	This study
pSEVA66-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA68-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pUC origin	This study
pSEVA69-sl3-T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study

Table S3. β-carotene and NarK production overview

Strain/pSEVA construct	Resistance/origin	Production (mg/L)
NEB5α::T7*	/genome	1.29
NEB5α::T7*::Km^R	Km/genome	1.30
13	Amp/pBBR1	0.64
16	Amp/p15A	1.56
17	Amp/pSC101	0.54
19	Amp/pBR322/ROP	3.69
23	Km/pBBR1	0.97
26	Km/p15A	1.37
27	Km/pSC101	0.61
29	Km/pBR322/ROP	0.38
33	Cm/pBBR1	1.24
36	Cm/p15A	1.70
37	Cm/pSC101	0.63
39	Cm/pBR322/ROP	0.31
43	Sp/pBBR1	1.54
46	Sp/p15A	1.91
47	Sp/pSC101	1.03
49	Sp/pBR322/ROP	0.67
53	Tet/pBBR1	0.95
56	Tet/p15A	1.51
57	Tet/pSC101	0.66
59	Tet/pBR322/ROP	2.69
63	Gm/pBBR1	1.72
66	Gm/p15A	3.16
67	Gm/pSC101	0.51
69	Gm/pBR322/ROP	0.59
NEB5α::K1F		
13	Amp/pBBR1	2.58
16	Amp/p15A	3.15
17	Amp/pSC101	2.08
19	Amp/pBR322/ROP	3.03
23	Km/pBBR1	3.28
26	Km/p15A	2.72
27	Km/pSC101	2.33
29	Km/pBR322/ROP	0.43
33	Cm/pBBR1	3.46
36	Cm/p15A	2.97
37	Cm/pSC101	2.07
39	Cm/pBR322/ROP	2.36
43	Sp/pBBR1	3.41
46	Sp/p15A	3.13
47	Sp/pSC101	2.11
49	Sp/pBR322/ROP	1.83
53	Tet/pBBR1	3.51
56	Tet/p15A	3.17
57	Tet/pSC101	1.53
59	Tet/pBR322/ROP	4.10
63	Gm/pBBR1	3.31
66	Gm/p15A	4.15
67	Gm/pSC101	1.85
69	Gm/pBR322/ROP	1.73
BL21 (DE3), 4hr		
13	Amp/pBBR1	9.32
16	Amp/p15A	11.84
17	Amp/pSC101	17.21
18	Amp/pUC	0.64
19	Amp/pBR322/ROP	19.43

23	Km/pBBR1	5.50
26	Km/p15A	6.26
27	Km/pSC101	3.49
28	Km/pUC	0.21
29	Km/pBR322/ROP	0.54
33	Cm/pBBR1	8.76
36	Cm/p15A	10.46
37	Cm/pSC101	1.70
38	Cm/pUC	7.13
39	Cm/pBR322/ROP	10.91
43	Sp/pBBR1	9.25
46	Sp/p15A	9.08
47	Sp/pSC101	11.34
48	Sp/pUC	11.79
49	Sp/pBR322/ROP	1.92
53	Tet/pBBR1	2.48
56	Tet/p15A	2.28
57	Tet/pSC101	4.72
58	Tet/pUC	0.55
59	Tet/pBR322/ROP	0.70
63	Gm/pBBR1	2.52
66	Gm/p15A	5.61
67	Gm/pSC101	2.37
68	Gm/pUC	1.39
69	Gm/pBR322/ROP	3.77
BL21 (DE3), 24hr		
13	Amp/pBBR1	24.29
16	Amp/p15A	26.68
17	Amp/pSC101	36.69
18	Amp/pUC	0.95
19	Amp/pBR322/ROP	21.08
23	Km/pBBR1	15.94
26	Km/p15A	35.15
27	Km/pSC101	11.36
28	Km/pUC	0.47
29	Km/pBR322/ROP	0.72
33	Cm/pBBR1	22.26
36	Cm/p15A	34.44
37	Cm/pSC101	2.71
38	Cm/pUC	0.88
39	Cm/pBR322/ROP	42.22
43	Sp/pBBR1	91.58
46	Sp/p15A	78.48
47	Sp/pSC101	32.66
48	Sp/pUC	57.18
49	Sp/pBR322/ROP	3.16
53	Tet/pBBR1	9.63
56	Tet/p15A	9.29
57	Tet/pSC101	27.03
58	Tet/pUC	0.75
59	Tet/pBR322/ROP	0.96
63	Gm/pBBR1	35.26
66	Gm/p15A	25.26
67	Gm/pSC101	5.80
68	Gm/pUC	5.96
69	Gm/pBR322/ROP	23.14

References

(1) Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T., and Nørholm, M. H. H. (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. *ACS Synt. Biol. 4*, 1042-1046.

(2) Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppälä, S., and Nørholm, M. H. H. (2015) Uracil excision for assembly of complex pathways. In *Springer Protocol Handbooks*, pp1−11, Humana Press, Springer−Verlag Berlin Heidelberg.

(3) Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol. 10*, 21.

(4) Hiyama, T., Nishimura, M. and Chance, B. (1969) Determination of carotenes by thin-layer chromatography. *Anal. Biochem. 29*, 339–42.

(5) St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* 2, 537–541.

(6) Mirzadeh, K., Martínez, V., Toddo, S., Guntur S., Herrgård, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector–coding sequence junction. *ACS Synt. Biol. 4*, 959–965.

Paper IV

Marine Bacterial Cytochromes P450 and their Potential in Biotechnology

Marine bacterial cytochromes P450 and their potential in biotechnology

Ana Mafalda Cavaleiro[¶,1], Henrique Machado[¶,1,2], Carlos Casado[3], Stefan Kol[1], Morten H. H. Nørholm[1,*], Lone Gram[2]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allè 6, DK-2970 Hørsholm, Denmark

[2]Department of Systems Biology, Technical University of Denmark, Matematiktorvet bldg 301, DK-2800 Kgs Lyngby, Denmark

[3] Evolva Biotech A/S, DK-2200 Copenhagen, Denmark

[¶] **These authors contributed equally to this work**

[*] **Corresponding author**

E-mail: morno@biosustain.dtu.dk.

ABSTRACT

Cytochromes P450 (CYPs) are enzymes that modify several types of compounds, usually by substrate mono-oxygenation, and are of biotechnological interest due to their ability to perform challenging chemistry. In comparison with the membrane bound eukaryotic CYPs, the soluble bacterial CYPs are more stable and easier to express and purify. However, their use as biocatalysts in the biotechnological industry has not been fully explored. Marine bacteria are emerging as a yet unexplored source of natural products, many of which are dependent on CYP-modifications. We mined 19 genomes of bioactive marine bacteria and identified 26 distinct CYP open-reading frames (ORFs). These ORFs were compared to previously studied bacterial CYPs to identify their phylogenetic relationships and putative functions. Five CYPs were analyzed in more detail and expressed in the model cell factories *Escherichia coli* and *Saccharomyces cerevisiae*. Expression in yeast confirmed cytoplasmic localization of the bacterial CYPs. Furthermore, we showed that codon optimized bacterial CYPs were properly folded in yeast. To our knowledge, the present study represents the first successful mining for CYPs from marine bacterial genomes in a merger of white and blue biotechnology.

Keywords Cytochrome P450 (CYP), Marine bacterial CYPs, Biotechnological potential, Yeast expression

INTRODUCTION

Cytochromes P450 (CYPs) are a superfamily of enzymes found in many different taxonomic groups. These heme-thiolate proteins bind to carbon monoxide exhibiting an absorption spectrum at 450 nm, which is the phenomenon responsible for their designation. They use electrons from NAD(P)H to catalyze scission of molecular oxygen and an associated protein, the redox partner, to transfer the reducing equivalents to the heme prosthetic group. CYPs modify several substrates by mono-oxygenation and this is the most common function of these enzymes, although other functions have been described [1,2,3].

Several natural compounds are synthesized in microorganisms, plants and fungi through a plethora of chemical modifications catalyzed by CYPs [3]. Natural product synthesis includes several unique chemical reactions, some of which are performed by the associated CYPs. Therefore, understanding the mechanisms of action of CYP enzymes, associated with natural product synthetic clusters, will allow for catalytic innovation and can be of great use in biotechnological processes [3].

In Nature, eukaryotic CYPs are often part of membrane-associated multi protein complexes and their use as synthetic catalysts at an industrial scale is hampered by this complexity as well as their limited stability and activity when isolated [4,5]. Additionally, the requirements for a constant supply of NAD(P)H and a redox partner are challenges in establishing CYP-based biocatalysis. Despite these obstacles, CYPs have been successfully used in industrial settings. For example, in yeast, the co-expression of a CYP, its redox partner and cytochrome b5 from plants led to high yield production of the antimalarial drug artemisinin [6,7].

During the last decades, CYPs have received increased attention. Of the 905 total identified CYPs in 2009, 30% were of plant origin, but these only represented 11% of total family diversity [8]. Remarkably, bacterial CYPs accounted for only 6% of the total number of

CYPs, but 18% of family diversity [8]. By August 2013, approximately 20,000 CYP enzymes had been identified and catalogued (http://drnelson.uthsc.edu/CytochromeP450.html).

In contrast to eukaryotic CYPs, the bacterial CYPs offer several advantages. So far, all identified bacterial CYPs are soluble (in contrast to membrane-bound) enzymes that are faster and more stable catalysts than their eukaryotic counterparts [4]. The bacterial CYPs are usually associated with ferredoxins and ferredoxin reductases, which are used as redox partners [9]. One of the most extensively studied bacterial CYPs is CYP102A1 from *Bacillus megaterium*, commonly known as BM3 [10]. BM3 is a self-sufficient enzyme, as it is fused with a FAD/FMN reductase (differs from ferredoxin reductases since it has a flavin moiety instead of iron-sulfur core), and it uses C12-C20 saturated or unsaturated fatty acids as natural substrates [11]. For several years, efforts to achieve immobilization, cofactor regeneration and development of assays for CYPs have been focused on BM3 [12]. However, the major achievements with BM3 are mostly related to protein engineering towards, in particular, hydroxylation of BM3-unnatural substrates [5,13,14,15,16,17,18]. Additional efforts have been dedicated to engineering other prokaryotic CYPs into enzymes with the capability to modify a broad range of substrates aiming to exploit their unique ability to perform challenging chemistry in biotechnological processes [19].

Soil microorganisms have been successfully explored as sources of bioactive compounds used by the pharmaceutical and biotech industries [20], but the use of marine bacteria for similar purposes is on the rise [21,22,23]. Marine bacteria are believed to harbor a multitude of novel bioactive compounds due to the unique environmental conditions they have adapted to, such as high pressure, high salinity, high or low temperature or oligotrophic conditions [24,25]. Some of these bioactive compounds are produced by biosynthetic reactions catalyzed by CYPs, which are typically encoded in biosynthetic gene clusters [26]. However, there are very few studies on marine CYPs and bioprospecting marine bacteria would likely provide

novel CYPs. In the present study we have mined the genome sequences of several bioactive marine bacteria for CYP-encoding genes. Selected candidates were cloned and expressed in both *E. coli* and *S. cerevisiae* to explore their cellular localization and biotechnological potential in these model microbial cell factories. This study also explores the heterogeneity of CYP enzymes in marine bacteria and the process of discovering new CYP enzymatic functions.

MATERIALS AND METHODS

Strains and media

All strains used for cloning and expression are listed in Table 1. *Saccharomyces cerevisiae* strain PAP1503 was obtained from Per Amstrup Pedersen (Copenhagen University, Denmark). Yeast strains were grown either in synthetic complete (SC) drop-out media (0.67% Yeast Nitrogen Base with required amino acids from Sigma-Aldrich, St. Louis, MO, USA)) or standard yeast peptone dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% dextrose from Sigma-Aldrich, St. Louis, MO, USA) for both liquid cultures and agar plates. Yeast transformants were selected on SC media without the appropriate selection requirements corresponding to the plasmid-encoded auxotrophic markers. *Escherichia coli* strain NEB5α (New England Biolabs, Ipswich, MA, USA) was used for cloning and propagation of plasmids. *E. coli* strain BL21 (DE3) (Novagen, Madison, WI, USA) was used for radioactive labeling studies. Chemically competent cells of NEB5α and BL21 (DE3) were prepared as described elsewhere [27]. Competency of the cells was 2.6-3.3 × 10^7 CFU/μg DNA. Bacteria were propagated on Luria-Bertani (LB) agar plates (Oxoid, Altrincham, UK) or liquid 2xYT media (1.6% tryptone, 1% yeast extract from Sigma-Aldrich, St. Louis, MO, USA) supplemented with spectinomycin (50 μg/mL) or ampicillin (100 μg/mL) when required and Phosphate, Ammonium and Selenomethionine (PASM) media for ^{35}S-methionine labeling of proteins [28]. ^{35}S-methionine was purchased from Perkin Elmer (Waltham, Massachusetts, USA). Plasmids were isolated using the NucleoSpin® Plasmid QuickPure Kit (Macherey-Nagel, Durën, Germany).

Table 1 – Strains and plasmids used in this study

Strains/Plasmids	Property	Source
Saccharomyces cerevisiae		
PAP1503	*MATα ura3-52 trp1::GAL10-GAL4 lys2-801 leu2Δ1 his3Δ200 pep4::HIS3 prb1Δ1.6R can1 GAL*	[29,30]
PAP1503_P4	PAP1503 transformed with plasmid pEMBLyex4_P4	This study
PAP1503_P5	PAP1503 transformed with plasmid pEMBLyex4_P5	This study
PAP1503_P7	PAP1503 transformed with plasmid pEMBLyex4_P7	This study
PAP1503_P9	PAP1503 transformed with plasmid pEMBLyex4_P9	This study
PAP1503_P29	PAP1503 transformed with plasmid pEMBLyex4_P29	This study
PAP1503_BM3	PAP1503 transformed with plasmid pEMBLyex4_BM3	This study
PAP1503_CYP79A1	PAP1503 transformed with plasmid pEMBLyex4_CYP79A1	This study
Escherichia coli		
NEB5α	*fhuA2* Δ*(argF-lacZ)*U169 *phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
NEB5α_P5	NEB5α transformed with plasmid pCDF_P5	This study
NEB5α_P7	NEB5α transformed with plasmid pCDF_P7	This study
NEB5α_P9	NEB5α transformed with plasmid pCDF_P9	This study
NEB5α_P29	NEB5α transformed with plasmid pCDF_P29	This study
BL21 (DE3)	F^- *ompT gal dcm lon hsdS*$_B$(r_B^- m_B^-) λ(DE3 *[lacI lac*UV5-T7 gene *1 ind1 sam7 nin5])*	Novagen
BL21 (DE3)_P5	BL21 (DE3) transformed with plasmid pCDF_P5	This study
BL21 (DE3)_P7	BL21 (DE3) transformed with plasmid pCDF_P7	This study
BL21 (DE3)_P9	BL21 (DE3) transformed with plasmid pCDF_P9	This study
BL21 (DE3)_P29	BL21 (DE3) transformed with plasmid pCDF_P29	This study
Plasmids		
pEMBLyex4	High copy-number yeast expression vector controlled by the galactose-inducible *GAL1-CYC1* promoter, 2μ origin of replication, *URA3*, *leu2-d*	[31]
pEMBLyex4_P4	*P4* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_P5	*P5* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_P7	*P7* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_P9	*P9* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_P429	*P29* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_BM3	*BM3* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pEMBLyex4_CYP79A1	*CYP79A1* ORF expressed from $P_{GAL1-CYC1}$ promoter in pEMBLyex4	This study
pCDFDuet-1	Cloning and expression vector, Sp^R	Novagen
pCDF_P5	*P5* ORF expressed from P_{T7} promoter in pCDFDuet-1	This study
pCDF_P7	*P7* ORF expressed from P_{T7} promoter in pCDFDuet-1	This study
pCDF_P9	*P9* ORF expressed from P_{T7} promoter in pCDFDuet-1	This study
pCDF_P29	*P29* ORF expressed from P_{T7} promoter in pCDFDuet-1	This study

Genomic DNA isolation and sequencing

High purity genomic DNA was extracted by successive phenol:chloroform:isoamyl-alcohol purification steps followed by precipitation with isopropanol, treatment with RNase, and a

final purification and precipitation step [32]. Quality assessment and quantification of genomic DNA was done in 1% agarose gel electrophoresis, NanoDrop Spectrometer (Saveen Werner, Sweden), and Qubit 2.0 Analyzer (Invitrogen, United Kingdom). The genomic DNA was used as a template for PCR amplification of the genes of interest.

Bioinformatic analysis

The strains used in the genome mining (Fig. 1) were isolated during the Danish Galathea 3 global research expedition (http://www.galathea3.dk/uk) [33] and whole genome sequenced [23]. Prediction of cytochrome P450 genes was performed using an annotation and homology-based search after RAST genome annotation [34]. Identification of specific biosynthetic gene clusters was performed using antiSMASH 2.0 [35] and annotation-based analyses [34]. Possible identity of all predicted cytochromes P450 to known proteins was done by Basic Local Alignment Search (BLAST) analysis. Membrane protein topology predictions were made using the online tools Topcons http://topcons.cbr.su.se/ [36] and ΔG predictor http://dgpred.cbr.su.se/ [37]. An attempt to identify the families of the new CYPs was performed using the information in the "cytochrome P450 homepage" [8].

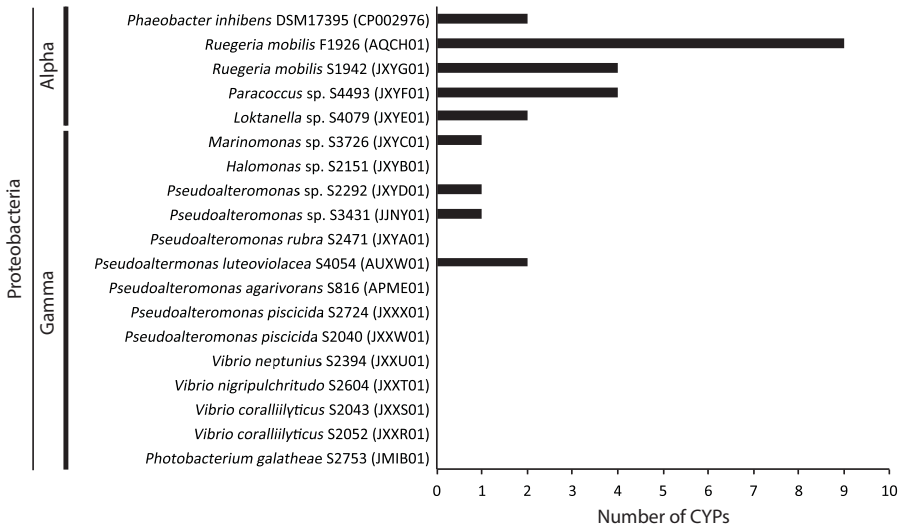


Figure 1. **Overview of the number of identified cytochromes P450 (CYP) in 19 marine bacterial genomes.**

Both members of alpha and gamma-proteobacteria were mined for putative CYPs. Strain names and numbers are referenced in the vertical axis with the NCBI whole genome sequence accession number in brackets. Black bars represent the number of CYPs identified per genome.

PCR, oligonucleotide design and uracil excision cloning

All oligonucleotides used for PCR amplification were purchased from Integrated DNA Technologies (Coralville, IA, USA) and are listed in Table S1. Oligonucleotides were designed with melting temperatures (T_m) of approximately 60°C. The uracil excision-specific oligonucleotides were designed with U-overhangs of T_m varying from 20 to 30°C. Oligonucleotides used for amplification of parts for homologous recombination-based yeast DNA assembly were designed with homology arms of >27 bp. DNA amplification and assembly in *E. coli* with uracil excision cloning were accomplished as described previously [38,39]. PCR products were purified using a PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Foster City, USA). PCR products were quantified using a NanoDrop 8000 (Thermo Scientific, Wilmington, USA).

Plasmids and strains construction

All plasmids used for CYP-ORF expression are listed in Table 1. For testing CYP expression in *E. coli,* PCR products containing CYP-predicted ORF and U-overhangs complementary to the entry vector were cloned in a pCDFDuet-1 plasmid via two-fragment uracil excision cloning. BL21 (DE3) was transformed and assayed as described below (see section ^{35}S-methionine labeling of proteins). In order to avoid negative effects of rare codons in *S. cerevisiae*, the genes were codon optimized using the codon optimization tool from Integrated DNA Technologies and then synthetized as gBlocks™ gene fragments (Integrated DNA Technologies, Coralville, IA, USA). The codon optimized sequences of P4, P5, P7, P9 and

P29 have been deposited on GenBank with accession numbers KU710259, KU710260, KU710261, KU710262, and KU710263, respectively. These codon optimized gene fragments were used as PCR templates for assembly in yeast. Three PCR products with long homology arms encoding for protein of interest (CYP), GFP (C-terminally fused to a His8 tag) and pEMBLyex4 plasmid previously digested with *Bam*HI and *Hin*dIII were transformed directly into PAP1503 and assembled in yeast by homologous recombination [40]. The resulting construct encodes for the protein of interest (CYP) C-terminally fused to GFP with a TEV protease cleavage site in between. The newly assembled plasmids were extracted from yeast and transformed into NEB5α for plasmid amplification. All constructs were confirmed by colony PCR and sequencing.

^{35}S-methionine labeling of proteins

Transcription and translation of cytochrome P450 enzymes in *E. coli* was confirmed using the rifampicin blocking technique and ^{35}S- methionine labeling [41] as described previously [42].

Assay for overexpression and production of GFP-fusions

Overexpression of CYP-GFP fusions in *S. cerevisiae* was performed in strain PAP1503 as described elsewhere [40] with minor modifications: synthetic complete (SC) medium was used instead of synthetic minimal (SD) medium. PAP1503 transformants were grown overnight (30°C, 200 rpm) in SC supplemented with glucose (20 g/L) without uracil. Overnight cultures were used to inoculate SC without uracil and leucine (SC-ura-leu) in a 1:50 ratio. Cultures were grown at 30°C for 48 hours before they were used to inoculate 250 mL of SC-ura-leu medium supplemented with glucose (5 g/L) with an initial optical density (OD) of 0.05. This was done in triplicate. Cultures were grown at 30°C, 200 rpm until OD 1.0 before GFP-fusion production was induced by addition of galactose at a final concentration of 2% (m/v). Cultures were then grown for 24 hours at 20°C with shaking at 200 rpm.

Microscopy

208 *S. cerevisiae* cells were grown as described above (see previous section). Fluorescence of
209 CYP-GFP fusions was visualized at 400 x magnification with a DM4000B fluorescence
210 microscope from Leica (Wetziar, Germany). CYP79A1 from *Sorghum bicolor* was used as a
211 membrane protein control [43].
212
213 **Protein purification and quantification**
214 Cell pellets from 0.75 L of CYP-GFP producing cultures were resuspended in 30 mL IMAC
215 buffer A (50 mM TRIS-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 1
216 tablet/50 mL complete, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, St. Louis,
217 MO, USA) and cells were lysed by three passes at 20.000 psi through an Emulsiflex-C5
218 (Avestin, Ottawa, Canada). After centrifugation for 15 min at 15.000*g*, the supernatant was
219 loaded onto a 5 mL HisTrap FF crude column (GE Healthcare, Buckinghamshire, UK)
220 equilibrated with IMAC buffer A. The columns were washed with 10 column volumes of
221 IMAC buffer A and the proteins of interest eluted with a 10 column volume gradient of 0-
222 100% IMAC buffer B (50 mM TRIS-HCl pH 7.5, 500 mM NaCl, 500 mM Imidazole).
223 Fractions containing the proteins of interest were identified by SDS-PAGE, pooled and
224 concentrated using Amicon Ultra-15 Centrifugal Filters (Millipore, Billerica, Massachusetts,
225 USA). The CYP proteins of interest (P4, P5, P7 and P29) were concentrated using a 30 kDa
226 cut-off filter, while P9 and BM3 were concentrated using a 10 kDa and 50 kDa cut-off filter,
227 respectively. Protein concentrations were measured with a NanoDrop ND-1000 (NanoDrop
228 Technologies, Inc., Wilmington, DE, USA). Samples were kept at 4°C until permanent
229 storage at -80°C.
230
231 **In-gel fluorescence**
232 In-gel fluorescence of purified GFP-fusions was carried out in a G:Box chemi XT4 (Syngene,
233 Cambridge, UK) using Mini-PROTEAN® TGX™ precast gels from Biorad (Hercules, USA)
234 and with 465 nm excitation and 535 nm emission.
235

Absorption Spectroscopy

Absorption spectra of CYP-GFP fusions at 450 nm were recorded with a SLM Aminco DW-2000 TM spectrophotometer (Spectronic Instruments, Rochester, NY, USA) as described elsewhere [44]. Purified protein-GFP fusions were used in concentrations between 0.07-1.2 mg/mL in Tris-HCl buffer pH 8.0.

RESULTS

Mining the genomes of marine bacteria for cytochromes P450 and redox partners

The genomes of 19 bioactive marine strains, previously identified as secondary metabolite producers [23,45,46], were mined for cytochrome P450 (CYP) encoding genes and a total of 26 putative CYP genes were identified (Table S2). The number of CYPs ranged from zero to nine per bacterial genome (Fig. 1). The identified CYP domain-containing open reading frames (ORFs) had GC contents between 39 and 73% and gene sizes from 300 to 1,400 bp (Table S2). Members of the alpha-proteobacteria had up to nine CYP-encoding genes, while a maximum of two was detected in a single gamma-proteobacterial strain (Fig. 1). CYP genes found in the gamma-proteobacteria were associated with putative specific pathways or biosynthetic clusters, such as a predicted gamma-butyrolactone biosynthetic cluster, a cluster involved in agar degradation or a non-ribosomal peptide/ polyketide synthase (Table 2).

Table 2 - Short-list of putative CYPs to be further studied and their redox partners

P450	Strain	GenBank Accession	Size (bp)	GC (%)	TM domains	Genomic context	CYP family	Ferredoxins	Ferredoxin Reductases
P4	S3431	KDC54395	1173	41	None	Agar degradation	CYP236A1	4	2
P5	S3726	KJZ11492	1209	39	None	Butyrolactone cluster	CYP107E1	3	1
P7	S4079	KJZ21287	1176	54	None	Exopolymer metabolism	CYP104A2	4	3
P9	S4493	KJZ30079	312	56	None	Cobalt-zinc-cadmium resistance	CYP152B1	2	2
P29	S4054	KKE82924	1341	39	None	NRPS-PKS	CYP197A1	3	3

The CYP redox-partners in bacteria are usually ferredoxins or flavoprotein reductases [4] and in line with this putative ferredoxins and ferredoxin reductases were identified in all the nineteen genomes. For CYPs P4 and P7, genes potentially encoding for redox partners were identified immediately downstream of the predicted CYP (Table 2).

Features of CYPs

We used UniProt (Universal Protein Resource), BLAST (Basic Local Alignment Search Tool) and the Cytochrome P450 homepage [8] to compare all predicted CYP sequences with previously studied CYPs. All the 26 identified CYPs contained cytochrome P450 domains, that were also identified by Pfam analyses [47], regardless of their gene size (data not shown). Although sequence conservation is low within this family of enzymes, their general topography and structural fold are highly conserved, allowing CYP domain identification [1]. Next, we compared the putative CYP protein sequences to known bacterial CYPs using the BLAST server on the Cytochrome P450 homepage (http://blast.uthsc.edu/) [8]. Several of the identified CYPs were conserved within the same species or genus. However, some diversity (with respect to CYP family) was found in CYP sequences across the different genera of studied marine bacteria (Fig. 2). The most similar bacterial CYPs, a recently reported marine CYP (AHA34040) and previously studied CYPs (AAC68886 and ADC79647) are also shown (Fig. 2).

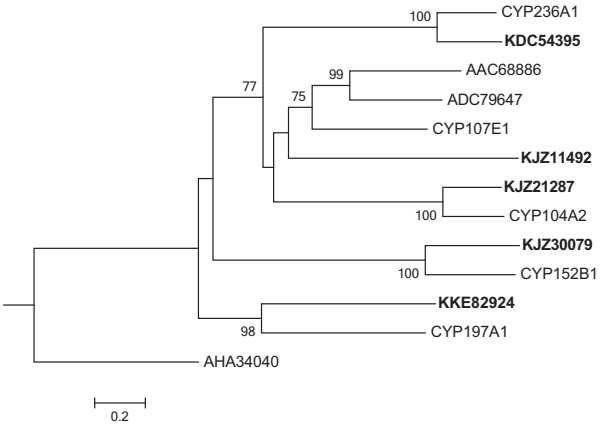


Figure 2. **Phylogenetic tree of the further studied CYPs.**

The five CYPs selected for further studies (highlighted in bold) shown together with the most similar bacterial CYPs, previously studied CYPs (AAC68886 and ADC79647) and a recently studied marine CYP (AHA34040). The nodes with bootstrap support of 70 or more are indicated (1000 replications). GenBank accession numbers are indicated for CYPs, except for the family representatives, where the CYP identification has been used.

Five CYPs were selected for further analyses based on high sequence homology to CYPs belonging to different families with previously identified distinct functions, genome clustering, pathway prediction, and microorganism diversity: one putative CYP from *Pseudoalteromonas* sp.*,* one from *Marinomonas sp*., one from *Loktanella sp*., one CYP from *Paracoccus sp*. and finally, one from *Pseudoalteromonas luteoviolacea* (Table 2). These CYPs not only differed in their primary sequences but were also located in distinct genetic environments, suggesting their participation in different metabolic reactions. The metabolic pathways and specific reactions, including substrate type were predicted using antiSMASH, RAST and BLAST analyses. CYPs P5 and P29 are part of two secondary metabolism clusters, predicted to be involved in gamma-butyrolactone and non-ribosomal peptide - polyketide synthesis, respectively (Table 2). CYPs P4, P7 and P9 are present in genomic

297 regions related to agar degradation, exopolymer metabolism and cobalt-zinc-cadmium
298 resistance, respectively (Table 2).
299
300 **Expression of putative CYPs in model cell factories**
301 To further study the five putative CYPs, all genes were PCR amplified from genomic DNA
302 using the oligonucleotides listed in Table S1 and cloned downstream from a phage T7
303 promoter (P_{T7}) in the pCDFDuet-1 plasmid. Despite several attempts, we were not able to
304 amplify the P4-encoding ORF from the genome of *Pseudoalteromonas* sp. (S3431). Proper
305 transcription and translation in the model bacterium *E. coli* BL21 (DE3) was monitored using
306 ^{35}S-methionine labeling of proteins followed by gel electrophoresis (see materials and
307 methods). BL21 (DE3) with an empty plasmid was used as negative control. P9 was only
308 produced in very low amount, but full-length proteins were observed for all tested CYPs (Fig.
309 S1).
310
311 To elucidate whether these putative CYPs would express in another popular model cell
312 factory, *Saccharomyces cerevisisae*, genes codon optimized for this eukaryotic host were
313 ordered form a commercial source (see materials and methods). Additionally, we included the
314 well studied bacterial CYP, BM3, as a control [10,11]. All five putative marine CYP-
315 encoding codon optimized DNA fragments and the control were fused with DNA encoding
316 GFP and a His8-tag in a pEMBLyex4 plasmid in the *S. cerevisiae* strain PAP1503 by yeast
317 homologous recombination (Fig. 3A). Subsequently the strains were starved for leucine to
318 increase plasmid copy number as described previously [40]. Starved cultures were used as
319 inoculum for higher volume cultures and gene expression was then induced with galactose
320 since CYP-GFP fusion expression is driven by the *CYC1-GAL1* promoter [40]. Expression
321 was confirmed by whole cell fluorescence. All CYP-GFP expressing cells appear green under
322 the microscope suggesting that all fusions are being produced in yeast, despite the different
323 fluorescent intensities observed (Fig. 3B).
324

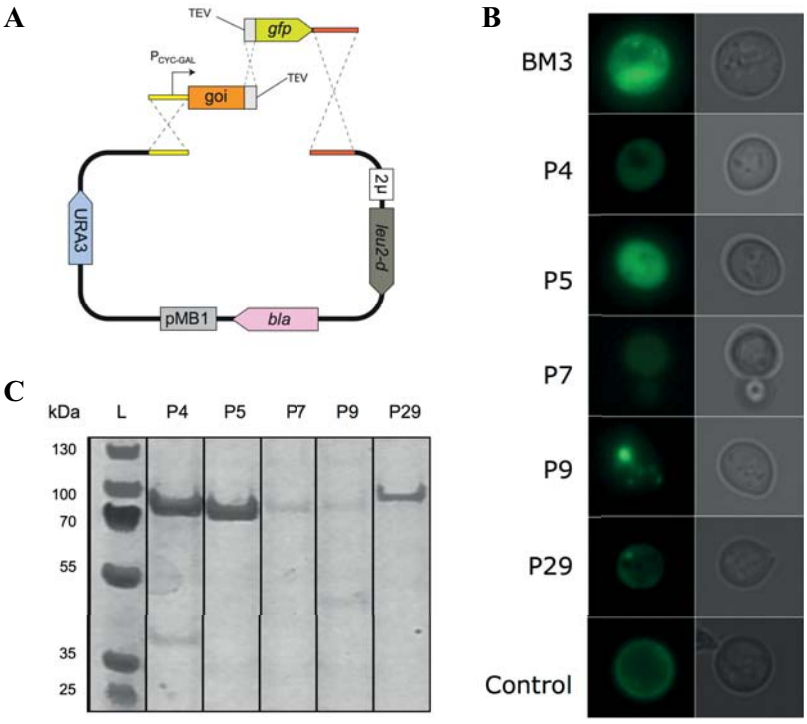


Figure 3. **CYPs production in yeast.**

(A) Schematic representation of the homologous recombination process that allows the pEMBLyex4 backbone plasmid and the DNA parts encoding for CYPs and GFP to be assembled in yeast. (B) Imaging of live yeast cells, induced with galactose for 24 hours at 20°C, producing the different analyzed CYP-GFP fusions P4, P5, P7, P9 and P29, the BM3 soluble control and the CYP79A1 membrane protein control. (C) SDS-PAGE analysis of purified CYP-GFP fusions P4, P5, P7, P9 and P29. The calculated molecular weights of CYP-GFP fusions are: P4, 71 kDa; P5, 72 kDa; P7, 71 kDa; P9, 38 kDa; P29, 78 kDa.

Cellular localization of bacterial CYPs expressed in *S. cerevisiae*

In contrast to bacterial CYPs, the hydrophobic nature of plant CYPs limits their potential in biotechnological applications. To explore the cellular localization of the marine bacterial CYPs expressed in yeast we monitored their appearance with a fluorescence microscope (Fig. 3B). Topology predictions using the online software Topcons (Topcons

http://topcons.cbr.su.se/ [36] and ΔG predictor http://dgpred.cbr.su.se/ [37] identified no N-terminal transmembrane segments in the CYPs. In line with this, the GFP signal derived from the heterologously expressed bacterial CYPs were homogeneously dispersed in the yeast cytoplasm, as compared to a plant membrane-bound CYP that localized to the periphery of the cells (Fig. 3B). Microscopy also showed that P9 localized in intensely fluorescent spots possibly representing misfolded aggregates (Fig. 3B).

Protein integrity and CO spectra

To evaluate the stability and integrity of the expressed CYP-GFP fusions, all five putative CYPs were purified by affinity using HisTrap columns, and assessed by SDS-PAGE (Fig. 3C) and in-gel fluorescence (Fig. S2). CYP-GFP fusions appeared to be produced as full-length fusions (Fig. 3C) with very little free GFP being detected (Fig. S2). In line with the microscopy data (Fig. 3B), P4 and P5 fusions were produced in the highest yields – the final concentrations being 0.184 and 0.175 mg/mL, respectively. The production levels of P7-GFP and P29-GFP were 0.065 and 0.066 mg/mL, respectively. Surprisingly, the concentration of P9-GFP was high at 0.151 mg/mL but only very little protein was detected when analyzed by SDS-PAGE and in-gel fluorescence (Fig. 3C and Fig. S2).

Proteins belonging to the P450 family exhibit light absorption maxima at 450 nm when correctly incorporated heme is bound to carbon monoxide. Hence, the most widely used method for assaying properly folded and active CYPs involves obtaining CO-binding difference spectra [48]. With this in mind, we used the purified CYP-GFP fusions to perform CO-binding difference spectra using bacterial CYP BM3 as a positive control (Fig. 4A). Again, P4 and P5 were well-behaved, exhibiting the characteristic peak at 450 nm confirming the CYP family predictions and suggesting that the enzymes are expressed in a correctly folded and active form. P7 and P29 were tested, but the CO-binding spectra were inconclusive probably due to low protein concentration (Fig. 4D and 4F). The P9 absorption

spectra (Fig. 4E) showed no peak at all, as expected by the absence of full-length protein and fluorescence (see previous result sections).

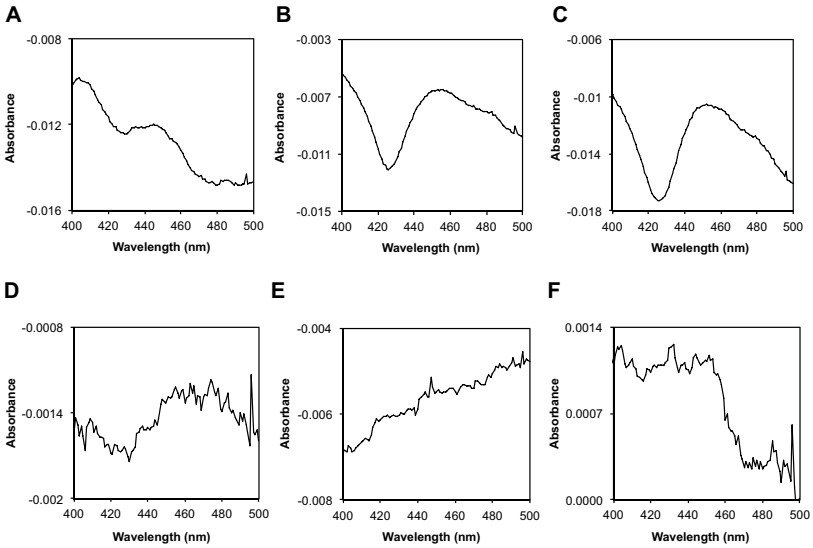


Figure 4. CO spectra different spectra of reduced CYPs.

The analyzed CYP-GFP fusions were kept in 10 mM Tris-HCl buffer pH 7.0 and used in protein concentrations of 0.065-0.184 mg/mL. Absorbance was measured ranging from 400 to 500 nm wavelength and appearance of the 450 nm characteristic peak (or the 420 nm peak for misfolded CYPs) was assessed for (A) BM3; (B) P4; (C) P5; (D) P7; (E) P9; (F) P29.

DISCUSSION

CYPs have received attention due to their association with specific and relevant metabolic pathways and their ability to perform challenging chemistry. Many studies have focused on plants that have a high amount of CYPs per genome [3,49,50,51]. Despite their ubiquity and diversity, the study and engineering of plant CYPs is facing several challenges. In higher organisms, CYPs are membrane associated, representing a major obstacle in expressing and purifying this class of enzymes [52]. Consequently, only two crystal structures of plant-derived CYPs have been published [53,54]. In addition, plant CYPs are less efficient than

385 their bacterial counterparts [55], meaning that high levels of expression would probably be
386 required for effective and fast conversion of substrates.
387
388 Bacterial CYPs, although present in lower number as compared to numbers in eukaryotes, are
389 soluble, more stable and exhibit higher activity [9,55]. Thus, they are easier to engineer,
390 overexpress, purify and crystallize [52]; and known structures of bacterial CYPs have been
391 used to assemble mammalian CYP structures based on homology modeling [56,57,58]. The
392 hydroxylation of camphor by CYP101 (P450cam) in *Pseudomonas putida* [59] was a major
393 breakthrough in the cytochrome P450 field and a few examples of studies on bacterial CYPs
394 have been unraveling important functions in the modification of natural compounds, such as
395 the antibiotic albaflavenone in *S. coelicolor* A3(2) [26] or polybrominated aromatic
396 compounds in *P. luteoviolacea* 2ta16 [60]. However, reports of new bacterial CYPs with new
397 functions/substrates are still scarce.
398
399 Marine bacteria are ideal sources of novel chemistry since they are understudied and capable
400 of performing a wide range of biological processes under environmentally challenging
401 conditions [61,62]. Here we present a study on genome mining and expression of bacterial
402 CYPs. Although the number of putative CYPs identified in the 19 analyzed genomes was
403 only 26, the diversity was considerable, with 19 different CYPs with protein identity lower
404 than 40% (Fig. S3). This diversity was also mirrored in their location in distinct genomic
405 environments and similarity to different CYP families, suggesting their participation in
406 different metabolic reactions.
407
408 The number of ORFs with CYP domains identified in the studied marine bacteria varied from
409 zero to nine per genome (Fig. 1). The high number identified for *Ruegueria mobilis* F1926
410 could be due to the poor quality of the genome sequence of this strain, which has more than
411 one thousand contigs. Indeed, the four CYPs found in the other *R. mobilis* strain (S1942)
412 suggest the average number of CYPs probably varies between zero and four in these marine

bacteria. The possible wrong annotation of ORFs with CYP domains can explain the poor results obtained for the 300 bp-CYP P9 (Fig. 3B and Fig. S1). P9 showed high similarity to CYP152B1, which is involved in fatty acid metabolism [63,64]. P9 function is likely different from the other selected CYPs. The P4 and P6 CYPs were present in the same genomic context, sharing 99,74% of amino acid identity (Fig. S3). These CYPs were found within a predicted agar degradation cluster where several agarases are present, pointing to a function in sugar modification. The closest related CYP family identified for these CYPs is CYP236, represented by CYP236A1. The latter was shown to have five agarases placed closely to it in a 101 kb plasmid (pSD15) from the agar degrading *Microscilla* sp. PRE1 [65].

AntiSMASH analysis has shown that P5 CYP is part of a predicted gamma-butyrolactone biosynthesis cluster in *Marinomonas* sp. S3726 [23,35]. Gamma-butyrolactone is a building block compound used in chemical industry and of great interest in bio-based production [66,67,68]. CYP family prediction places P5 in the CYP170 family. CYP170A1 is the family representative and it is responsible for the oxidation of the terpenoid epi-isozizaene to albaflavenone in *Streptomyces coelicolor* A3(2) [61]. CYP170A1 catalyzes a two-step full oxidation of epi-isozizaene to albaflavenone, similar to what could be expected for conversion of tetrahydrofuran to gamma-butyrolactone by oxidation of an aromatic ring (Fig. S4).

The predicted CYP P7 had high similarity to CYP104A2, an *Agrobacterium tumefaciens* CYP responsible for O-demethylation of plant phenolic compounds produced when infected with the plant-pathogen [69,70]. Other CYPs from the same CYP104 family are responsible for the O-demethylation of guaiacol during the degradation of the lignin polysaccharide [71]. The P7-coding gene is located next to predicted fructose transport and capsule polysaccharide export genes, pointing to a possible involvement in O-demethylation of sugars or polysaccharides. Sugar modifications might affect cell fitness and create a bias towards low producers, which could explain the low concentration of P7 in yeast (Fig. 3B and C). Despite

the inconclusive P7 CO-spectrum, the fact that this ORF clusters in the genome with ferredoxins/ferredoxin reductases supports its CYP prediction, as the confirmed P4 CYP (Table 2).

According to antiSMASH predictions, P29 is located right next to a NRPS/PKS gene cluster, suggesting its role as a tailoring enzyme involved in compound modification(s). Homology searches placed it in the CYP197 family, previously identified in *Streptomyces* and *Bacillus* species [72,73]. *Streptomyces* are known for being prominent producers of natural products and contain the highest number of structurally characterized CYPs [3].

All CYPs were successfully expressed in *E. coli* (Fig. S1) as commonly observed for other bacterial CYPs [12]. A growing field aims at introducing chemical modifications, not possible with standard organic chemistry [55]. One such example in *E. coli* is the synthesis of 1α,25-dihydroxyvitamin D3 from vitamin D3, where a bacterial CYP (CYP105A1) from *Streptomyces griseolus* was used [74]. Protein engineering efforts to modify P450cam [75] and BM3 [14] for *in vivo* oxidation of unnatural substrates demonstrate the importance of bacterial CYPs, and support the idea that new bacterial CYPs can be engineered to achieve broader specificity for biotechnological applications. Hence, the marine bacterial CYPs presented here appear as promising candidates for protein engineering.

Most of the successful commercial and industrially relevant CYP-dependent processes have been performed in *S. cerevisiae.* Production of a myriad of different compounds such as hydrocortisone, pregnenolone, amorphadiene (precursor of artemisinic acid), polyketides, isoprenoids, steviol components and opiates has been done in yeast, proving its value as a cell factory for these types of molecules [6,76,77,78]. Two major advantages of using yeast are a well-developed genetic engineering toolbox and a long history of use in the food industry.

Conclusions

To our knowledge, our work is the first demonstration of marine bacterial CYPs being expressed in yeast. This model cell factory offered a convenient assay allowing for cellular localization studies with GFP-fusions that are easily observed under the microscope. We also showed that it is possible, without further optimization besides codon optimization, to successfully express and produce bacterial CYPs in yeast. The stability and integrity of CYP-GFP fusions purified from yeast was evaluated by in-gel fluorescence and CO absorption spectra confirming that P4 and P5 are cytochromes P450 - and that they are properly folded in the yeast system (Fig. 4). Future work should focus on the further optimization of P7 and P29 expression in yeast.

Besides the obvious advantages of unraveling new functions that can be useful in biotechnology and replacing eukaryotic CYPs, this study is a first step towards exploiting the full potential of the bacterial CYPs present in marine environments. Future studies will aim at identifying the specific function of these enzymes.

Acknowledgements

The authors would like to acknowledge Darío Vázquez Albacete for fruitful discussions regarding CYPs, Helle Munck Petersen for technical assistance during protein purification and Per Amstrup Pedersen (Copenhagen University, Denmark) for kindly providing the PAP1503 strain and pEMBLyex4 plasmid. We would also like to thank Swee Hallwyl and Kasia Krzystanek (Evolva Biotech A/S).

Author contributions

Conceived and designed the experiments: AMC, HM, CC. Performed the experiments: AMC, HM, SK. Wrote the paper: AMC and HM with contributions from all the authors.

References

1. Werck-reichhart D, Feyereisen R. (2000) Cytochromes P450: a success story. Genome Biol 1(6): 3003.1–9.

2. Lamb DC, Waterman MR, Kelly SL, Guengerich FP. (2007) Cytochromes P450 and drug discovery. Curr Opin Biotechnol 18(6): 504–512.

3. Podust LM, Sherman DH. (2012) Diversity of P450 enzymes in the biosynthesis of natural products. Nat Prod Rep 29(10): 1251–66.

4. Urlacher V, Schmid RD. (2002) Biotransformations using prokaryotic P450 monooxygenases. Curr Opin Biotechnol 13(6): 557–564.

5. Peters MW, Meinhold P, Glieder A, Arnold FH. (2003) Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. J Am Chem Soc 125(44): 13442–13450.

6. Ro D, Paradise EM, Ouellet M, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 3–6.

7. Paddon CJ, Westfall PJ, Pitera DJ, et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496: 528–532.

8. Nelson DR. (2009) The cytochrome p450 homepage. Hum Genomics 4: 59–65.

9. Kelly SL, Kelly DE. (2013) Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? Philos Trans R Soc Lond B Biol Sci. 368(1612): 20120476.

10. Narhi LO, Fulco J. (1986) Characterization of a catalytically self-sufficient 119,000-Dalton Cytochrome P450 Monooxygenase Induced by Barbiturates *Bacillus megaterium*. J Biol Chem 261(16): 7160–7169.

11. Wen L, Fulco AJ. (1987) Cloning of the Gene Encoding a Catalytically Self-sufficient Fatty Acid Monooxygenase Induced by *Bacillus megaterium* and Its Functional

Expression Barbiturates in and Regulation in Heterologous (*Escherichia coli*) and Homologous (*Bacillus megaterium*) Hosts. J Biol Chem 262(14): 6676–6682.

12. Urlacher VB, Lutz-Wahl S, Schmid RD. (2004) Microbial P450 enzymes in biotechnology. Appl Microbiol Biotechnol. 64(3): 317–325.

13. Appel D, Lutz-wahl S, Fischer P, Schwaneberg U, Schmid RD. (2001) A P450 BM-3 mutant hydroxylates alkanes , cycloalkanes , arenes and heteroarenes. J Biotechnol 88:167–171.

14. Carmichael AB, Wong L-L. (2001) Protein engineering of *Bacillus megaterium* CYP102. Eur J Biochem 268(10): 3117–3125.

15. Farinas ET, Schwaneberg U, Glieder A, Arnold FH. (2001) Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. Adv Synth Catal 343: 601–606.

16. Li Q, Ogawa J, Schmid RD, Shimizu S. (2001) Engineering Cytochrome P450 BM-3 for Oxidation of Polycyclic Aromatic Hydrocarbons. Appl Environ Microbiol 67(12): 5735–5739.

17. Glieder A, Farinas ET, Arnold FH. (2002) Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat Biotechnol 20: 1135–1139.

18. Farwell CC, Zhang RK, Mcintosh JA, Hyster TK, Arnold FH. (2015) Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. ACS Cent Sci 1: 89–93.

19. Cirino PC, Arnold FH. (2002) Protein engineering of oxygenases for biocatalysis. Curr Opin Chem Biol 6(2): 130–135.

20. Bérdy J. (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. J Antibiot (Tokyo) 65: 441–441.

21. Xiong Z-Q, Wang J-F, Hao Y-Y, Wang Y. (2013) Recent advances in the discovery and development of marine microbial natural products. Mar Drugs 11(3): 700–17.

22. Wilson MC, Mori T, Rückert C, et al. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506: 58–62.

23. Machado H, Sonnenschein EC, Melchiorsen J, Gram L. (2015) Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. BMC Genomics 16(1): 1–12.

24. Fenical W, Jensen PR. (2006) Developing a new resource for drug discovery: marine actinomycete bacteria. Nat Chem Biol 2: 666–673.

25. Lozupone CA, Knight R. (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci U S A 104: 11436–11440.

26. Agarwal V, El Gamal AA, Yamanaka K, et al. (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. Nat Chem Biol 10(8): 640–647.

27. Inoue H, Nojima H, Okayama H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23–28.

28. Studier FW. (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41: 207–234.

29. Pedersen PA, Rasmussen JH, Jørgensen PL. (1996) Expression in High Yield of Pig a1b1 Na, K-ATPase and Inactive Mutants D369N and D807N in *Saccharomyces cerevisiae*. J Biol Chem 271(5): 2514–2522.

30. Steffensen L, Pedersen PA. (2006) Heterologous Expression of Membrane and Soluble Proteins Derepresses GCN4 mRNA Translation in the Yeast *Saccharomyces cerevisiae*. Eukaryot Cell 5(2): 248–261.

31. Cesareni G, Murray JAH. (1987) Plasmid vectors carrying the replication origin of filamentous single-stranded phages. Setlow JK, ed. Genet Eng Princ Methods 9: 135–154.

32. Sambrook J, Russel DW. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

574 33. Gram L, Melchiorsen J, Bruhn JB. (2010) Antibacterial activity of marine culturable
575 bacteria collected from a global sampling of ocean surface waters and surface swabs
576 of marine organisms. Mar Biotechnol (NY) 12(4): 439–51.

577 34. Aziz RK, Bartels D, Best AA, et al. (2008) The RAST Server: rapid annotations using
578 subsystems technology. BMC Genomics 9: 75.

579 35. Blin K, Medema MH, Kazempour D, et al. (2013) antiSMASH 2.0 - a versatile
580 platform for genome mining of secondary metabolite producers. Nucleic Acids Res
581 41: W204–12.

582 36. Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A. (2015) The TOPCONS web server
583 for consensus prediction of membrane protein topology and signal peptides. Nucleic
584 Acids Res 43: 401–407.

585 37. Hessa T, Meindl-Beinker NM, Bernsel A, *et al*. (2007) Molecular code for
586 transmembrane-helix recognition by the Sec61 translocon. Nature 450(7172): 1026–
587 1030.

588 38. Cavaleiro AM, Kim SH, Seppa S, Nielsen MT, Nørholm MHH. (2015) Accurate DNA
589 Assembly and Genome Engineering with Optimized Uracil Excision Cloning. ACS
590 Synth Biol 4: 1042–1046.

591 39. Cavaleiro AM, Nielsen MT, Kim SH, Seppälä S, Nørholm MHH. (2015) Uracil
592 excision for assembly of complex pathways. Springer Protoc 1–10.

593 40. Scharff-Poulsen P, Pedersen PA. (2013) *Saccharomyces cerevisiae*-Based Platform for
594 Rapid Production and Evaluation of Eukaryotic Nutrient Transporters and
595 Transceptors for Biochemical Studies and Crystallography. PLoS One 8(10): 1–12.

596 41. Studier FW, Rosenberg AH, Dunn JJ, Debendorff JW. (1990) Use of T7 RNA
597 Polymerase to Direct Expression of Cloned Genes. Methods Enzymol 185(1986): 60–
598 89.

599 42. Nielsen MT, Madsen KM, Seppa S, *et al*. (2014) Assembly of Highly Standardized

Gene Fragments for High-Level Production of Porphyrins in *E. coli*. ACS Synth Biol 4: 274–282.

43. Halkier BA, Sibbesen O, Koch B, Meiler BL. (1995) Characterization of cytochrome P450TYR, a multifunctional haem-thiolate7v-hydroxylase involved in the biosynthesis of the cyanogenic glucoside Dhurrin. Drug Metabol Drug Interact 12(3-4): 285–297.

44. Omura T, Sato R. (1964) The Carbon Monoxide-binding pigment of Liver Microsomes. J Biol Chem 239(7): 2370–2378.

45. Vynne NG. (2011) Bioactivity and phylogeny of the marine bacterial genus *Pseudoalteromonas*.

46. Mansson M, Gram L, Larsen TO. (2011) Production of bioactive secondary metabolites by marine *Vibrionaceae*. Mar Drugs 9(9): 1440–68.

47. Finn RD, Bateman A, Clements J, *et al*. (2014) Pfam: the protein families database. Nucleic Acids Res 42: D222–D230.

48. Guengerich FP, Martin MV, Sohl CD, Cheng Q. (2009) Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. Nat Protoc 4(9): 1245–1251.

49. Mizutani M, Sato F. (2011) Unusual P450 reactions in plant secondary metabolism. Arch Biochem Biophys 507(1): 194–203.

50. Hamberger B, Bak S. (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc Lond B Biol Sci 368(1612): 20120426.

51. Nelson D, Werck-Reichhart D. (2011) A P450-centric view of plant evolution. Plant J 66: 194–211.

52. Jung ST, Lauchli R, Arnold FH. (2011) Cytochrome P450: taming a wild type enzyme. Curr Opin Biotechnol 22(6): 809–817.

625 53. Lee D-S, Nioche P, Hamberg M, Raman CS. (2008) Structural insights into the
626 evolutionary paths of oxylipin biosynthetic enzymes. Nature 455(7211): 363–368.

627 54. Li L, Chang Z, Pan Z, Fu Z-Q, Wang X. (2008) Modes of heme binding and substrate
628 access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide
629 synthase. Proc Natl Acad Sci U S A 105(37): 13883–13888.

630 55. Bernhardt R, Urlacher VB. (2014) Cytochromes P450 as promising catalysts for
631 biotechnological application: chances and limitations. Appl Microbiol Biotechnol 98:
632 6185–6203.

633 56. Dai R, Pincus MR, Friedman FK. (1998) Molecular Modeling of Cytochrome P450
634 2B1: Mode of Membrane Insertion and Substrate Specificity. J Protein Chem 17(2):
635 121–129.

636 57. Belkina N V, Lisurek M, Ivanov AS, Bernhardt R. (2001) Modelling of three-
637 dimensional structures of cytochromes P450 11B1 and. J Inorg Biochem 87: 197–207.

638 58. Bathelt C, Schmid RD, Pleiss J. (2002) Regioselectivity of CYP2B6: homology
639 modeling , molecular dynamics simulation , docking. J Mol Model 8: 327–335.

640 59. Nelson DR, Kamataki T, Waxman DJ, *et al*. (1993) The P450 Superfamily: Update on
641 New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of
642 Enzymes, and Nomenclature. DNA Cell Biol 12(1): 1–51.

643 60. Moody SC, Zhao B, Lei L, *et al*. (2012) Investigating conservation of the
644 albaflavenone biosynthetic pathway and CYP170 bifunctionality in streptomycetes.
645 FEBS J 279(9): 1640–9.

646 61. Fenical W. (1993) Chemical studies of marine bacteria: developing a new resource.
647 Chem Rev 93: 1673–1683.

648 62. de Carvalho CCCR, Fernandes P. (2010) Production of metabolites as bacterial
649 responses to the marine environment. Mar Drugs 8(3): 705–27.

650 63. Matsunaga I, Yokotani N, Gotoh O, Kusunose E, Yamada M, Ichihara K. (1997)

Molecular cloning and expression of fatty acid alpha-hydroxylase from *Sphingomonas paucimobilis*. J Biol Chem 272(38): 23592–23596.

64. Fujishiro T, Shoji O, Nagano S, Sugimoto H, Shiro Y, Watanabe Y. (2011) Crystal Structure of H2O2-dependent Cytochrome P450SP{alpha} with Its Bound Fatty Acid Substrate: Insight into the regioselective hydroxylation of fatty acids at the alpha position. J Biol Chem 286(34): 29941–29950.

65. Zhong Z, Toukdarian A, Helinski D, *et al*. (2001) Sequence analysis of a 101-kilobase plasmid required for agar degradation by a *Microscilla isolate*. Appl Environ Microbiol 67(12): 5771–5779.

66. Hwang DW, Kashinathan P, Lee JM, *et al*. (2011) Production of γ-butyrolactone from biomass-derived 1,4-butanediol over novel copper-silica nanocomposite. Green Chem 13(7): 1672.

67. Martin CH, Dhamankar H, Tseng H-C, Sheppard MJ, Reisch CR, Prather KLJ. (2013) A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy-γ-butyrolactone. Nat Commun 4: 1414.

68. Dhamankar H, Tarasova Y, Martin CH, Prather KLJ. (2014) Engineering *E. coli* for the biosynthesis of 3-hydroxy-γ-butyrolactone (3HBL) and 3,4-dihydroxybutyric acid (3,4-DHBA) as value-added chemicals from glucose as a sole carbon source. Metab Eng 25: 72–81.

69. Kanemoto RH, Powell AT, Akiyoshi DE, *et al*. (1989) Nucleotide sequence and analysis of the plant-inducible locus pinF from *Agrobacterium tumefaciens*. J Bacteriol 171(5): 2506–2512.

70. Brencic A, Eberhard A, Winans SC. (2004) Signal quenching, detoxification and mineralization of vir gene-inducing phenolics by the VirH2 protein of *Agrobacterium tumefaciens*. Mol Microbiol 51: 1103–1115.

71. Filley TR, Cody GD, Goodell B, Jellison J, Noser C, Ostrofsky A. (2002) Lignin

demethylation and polysaccharide decomposition in spruce sapwood degraded by brown rot fungi. Org Geochem 33: 111–124.

72. Furuya T, Shibata D, Kino K. (2009) Phylogenetic analysis of Bacillus P450 monooxygenases and evaluation of their activity towards steroids. Steroids 74(12): 906–912.

73. Li Z-Z, Li X-F, Yang W, *et al*. (2013) Identification and functional analysis of cytochrome P450 complement in *Streptomyces virginiae* IBL14. BMC Genomics 14(1): 130.

74. Sawada N, Sakaki T, Yoneda S, *et al*. (2004) Conversion of vitamin D3 to 1alpha,25-dihydroxyvitamin D3 by *Streptomyces griseolus* cytochrome P450SU-1. Biochem Biophys Res Commun 320(1): 156–164.

75. Bell SG, Harford-cross CF, Wong L. (2001) Engineering the CYP101 system for *in vivo* oxidation of unnatural substrates. Protein Eng 14(10): 797–802.

76. Kavšček M, Stražar M, Curk T, Natter K, Petrovič U. (2015) Yeast as a cell factory: current state and perspectives. Microb Cell Fact 14(1): 94.

77. Duport C, Spagnoli R, Degryse E, Pompon D. (1998) Self-suffient biosynthesis of pregnenolone and progesterone in engineered yeast. Nat Biotechnol 16: 186–189.

78. Szczebara FM, Chandelier C, Villeret C, *et al*. (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. Nat Biotechnol 21: 143–149.

699 **Supporting information**

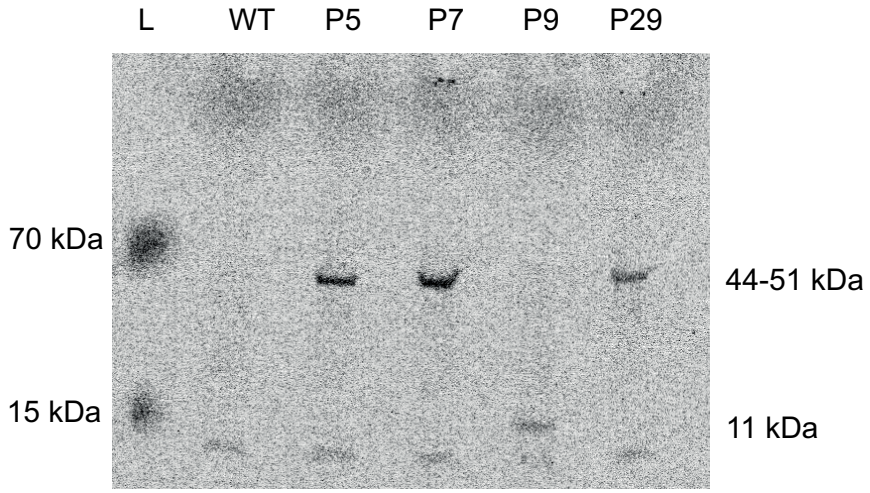
700 **Table S1.** List of oligonucleotides and plasmids.

Number	Name	Sequence
1	P4_*E.coli*_fw	AGATATACCUATGGCGAAAGTAATATTTATAACCAAAACAAACGA
2	P4_*E.coli*_rv	ATTATGCGUTTATTTTGATTTAAAGCTAACGTGCAGACTGT
3	P5_*E.coli*_fw	AGATATACCUATGAGTATCGAGTTGCATAGTAAGTTAAATAGT
4	P5_*E.coli*_rv	ATTATGCGUTTAGTTAATAACGATAGGTAGGCTATCTGGGC
5	P7_*E.coli*_fw	AGATATACCUATGCTGGATTTGCCAATCAATGAGAC
6	P7_*E.coli*_rv	ATTATGCGUTTAGTCGAGATAGACCGGCAGAT
7	P9_*E.coli*_fw	AGATATACCUATGCTCGACCTCTATGGTACGAAC
8	P9_*E.coli*_rv	ATTATGCGUATGCTCGACCTCTATGGTACGAAC
9	P29_*E.coli*_fw	AGATATACCUATGAAGCAAATACCAAAAGTAACCAACAAAG
10	P29_*E.coli*_rv	ATTATGCGUTTATTGCTCAATTAACTTATTCACTTTCATGACCAAC
11	pCDF_P450_fw	ACGCATAAUGCTTAAGCYGAACAGAAAGTAATCGTATTG
12	pCDF_P450_rv	AGGTATATCUCCTTATTAAAGTTAAACAAAATTATTTCTACAGGGGA ATTG
13	Seq_pCDF_fw	AGGTTTTGCGCCATTCGATGG
14	Seq_pCDF_rv	CGATTATGCGGCCGTGTACAA
15	P4_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATT ATGACTTCAAAGATTATAGGCGTTAGTGATTTTCC
16	P4_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCCTTAGACTTAAAAGAGACGT GCAAGGAA
17	P5_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATT ATGAGTATCGAGTTACATTCAAAGTTG
18	P5_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCGTTAATAACAATTGGCAAAG AATCAGGAC 3’
19	P7_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATT ATGCTAGACCTACCAATTAACGAAAC
20	P7_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCGTCTAAGTATACTGGTAGAT TTAACGG
21	P9_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATT ATGTTAGACTTATATGGCACGAACAC
22	P9_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCATCAAGACAACGTACATTTC TCATGATAAATC
23	P29_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATTA TGAAGCAGATCCCGAAG
24	P29_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCCTGCTCGATCAACTTGTTAA CTTTCATAA
25	BM3_yGFP_fw	ACACAAATACACACACTAAATTACCGGATCAATTCTAAGATAATTA TGACAATTAAAGAAATGCCTCAGCCAAAAACGT
26	BM3_yGFP_rv	AAATTGACTTTGAAAATACAAATTTTCCCCAGCCCACACGTCTTTTG CGT
27	CYC-GAL_long_seq_fw	TTACTATACTTCTATAGACACGCAAACAC
28	GFP_seq_rv	GTAGCATCACCTTCACCTTC
29	GFPup_fw	GAAAATTTGTATTTTCAAAGTCAATTTTCTAAAGGTGAAGAATTAT
30	GFPHISdo_rv	CTTCAATGCTATCATTTCCTTTGATATTGGATCATCTAATGGTGATGG TGATGGTGATGGTGTTTGTACAATTCATCCATACCAT

701

702

703

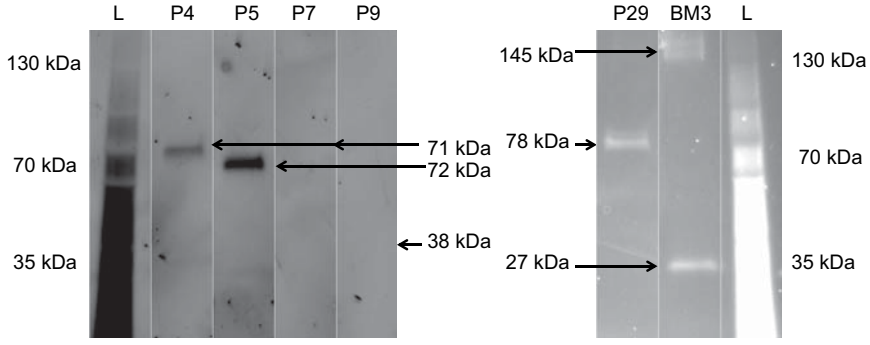


704

705 **Figure S1**. Bacterial CYPs expression in *E. coli* BL21 (DE3) (^{35}S-methionine labeling of

706 proteins).

707



708

709 **Figure S2.** Bacterial CYP-GFP fusions production in yeast *S. cerevisiae* (In-gel

710 fluorescence).

711

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ruegeria mobilis S1942	1	100,00	71,49	67,40	58,85	35,47	19,02	16,77	12,52	9,68	3,72	12,52	3,62	5,53	3,48	11,92	18,67	8,83	8,83	8,04	10,51	9,96	8,30	8,30	3,84	6,64
Ruegeria mobilis F1926	0	1	71,49	67,40	58,85	35,47	19,02	16,77	12,52	9,68	3,72	12,52	3,62	5,53	3,48	11,92	18,67	8,83	8,83	8,04	10,51	9,96	8,30	8,30	3,84	6,64
Phaeodactor inhibens DSM17395	134	134		64,63	56,38	34,01	19,43	16,67	12,55	9,79	3,37	12,18	2,73	5,33	3,77	12,32	17,76	8,79	8,79	8,80	11,80	9,69	8,09	8,09	3,70	5,35
Loktanella sp. S4079	149	149	168		56,61	32,48	17,96	17,04	11,75	8,45	4,05	13,46	3,06	5,73	4,47	13,27	20,00	10,15	10,15	8,24	10,57	8,33	7,41	7,41	4,19	6,15
Paracoccus sp. S4493	186	186	205	197		34,62	19,71	15,91	12,57	9,00	4,70	11,83	2,65	5,86	3,54	11,82	17,73	7,41	7,41	9,25	10,75	9,68	8,74	8,74	4,13	5,01
Paracoccus sp. S4493	302	302	315	318	306		17,85	15,98	11,71	8,14	4,61	13,84	3,21	6,17	5,15	11,79	18,09	8,21	8,21	9,33	11,83	9,04	8,67	8,67	4,25	5,52
Paracoccus sp. S4493	396	396	394	402	387	401		26,77	15,91	9,19	8,27	8,49	1,49	3,92	3,86	8,85	10,06	12,63	12,63	12,78	13,99	11,48	8,26	8,26	3,81	4,55
Ruegeria mobilis F1926	407	407	410	409	403	397	331		14,31	10,17	5,72	8,17	2,34	3,59	3,12	9,42	10,79	15,16	15,16	11,11	11,96	9,70	9,11	9,11	4,53	3,33
Ruegeria mobilis S1942	447	447	460	443	441	460	407	419		78,26	22,91	10,87	4,88	4,05	2,54	8,30	9,79	14,86	14,86	12,73	12,98	13,75	8,66	8,66	4,04	4,45
Ruegeria mobilis F1926	457	457	470	455	455	474	435	433	91		11,19	7,54	6,19	0,46	3,22	6,40	7,03	9,69	9,69	8,20	8,67	9,46	5,75	5,75	0,23	1,39
Ruegeria mobilis F1926	440	440	459	426	426	435	377	379	323	414		5,53	0,00	12,40	0,00	2,21	3,77	6,70	6,70	5,87	5,51	6,29	4,19	4,19	13,71	12,79
Ruegeria mobilis S1942	468	468	483	456	470	487	465	483	418	429	437		37,50	27,69	25,89	47,80	10,41	7,51	7,51	9,65	9,51	8,13	8,73	8,73	3,74	4,17
Ruegeria mobilis F1926	479	479	498	470	478	483	462	459	409	318	264	280		0,00	0,35	17,44	3,79	3,30	3,30	3,41	3,47	2,62	2,44	2,44	0,37	0,35
Ruegeria mobilis F1926	444	444	462	428	434	441	417	403	428	437	113	324	292		0,00	13,94	3,89	3,92	3,92	3,48	3,80	4,25	5,12	5,12	12,12	13,53
Ruegeria mobilis F1926	444	444	460	427	436	439	410	403	422	331	212	332	282	240		11,04	3,64	1,70	1,70	3,98	3,83	2,52	2,86	2,86	0,46	0,43
Phaeodactor inhibens DSM17395	473	473	484	451	470	479	482	481	431	439	443	237	374	389	403		9,68	6,87	6,87	8,51	9,53	7,95	7,44	7,44	3,05	4,57
Pseudoalteromonas luteoviolacea S4054	427	427	440	408	427	421	474	463	479	489	434	482	457	445	450	485		11,38	11,38	9,84	10,00	12,78	7,29	7,29	2,59	4,70
Pseudoalteromonas sp. S3431	485	485	488	478	487	481	429	386	424	438	376	480	439	392	404	488	448		99,74	10,21	14,59	13,71	9,09	9,09	4,23	5,20
Pseudoalteromonas sp. S2292	485	485	488	478	487	481	429	386	424	438	376	480	439	392	404	488	448	444		10,21	14,59	13,71	9,09	9,09	4,23	5,20
Marinomonas sp. S3726	492	492	489	490	481	476	416	432	432	448	385	468	425	416	410	473	462	422	422		23,08	11,83	6,25	6,25	3,71	3,68
Loktanella sp. S4079	477	477	471	474	473	462	412	427	429	442	377	466	417	405	402	465	459	396	396	320		11,74	6,99	6,99	3,57	3,04
Ruegeria mobilis F1926	479	479	483	473	476	473	432	428	414	421	358	463	409	383	387	463	423	384	384	395	391		6,15	6,15	3,00	4,67
Ruegeria mobilis S1942	530	530	534	537	522	527	489	469	506	509	412	523	480	408	441	535	534	450	450	495	492	488		100,00	21,69	5,91
Ruegeria mobilis F1926	530	530	534	537	522	527	489	469	506	509	412	523	480	408	441	535	534	450	450	495	492	488	0		21,69	5,91
Paracoccus sp. S4493	451	451	468	435	441	451	418	400	428	429	187	438	269	116	217	445	452	385	385	415	405	388	325	325		17,07
Pseudoalteromonas luteoviolacea S4054	436	436	460	427	438	445	420	407	429	426	116	437	282	115	230	439	446	383	383	422	414	388	398	398	102	

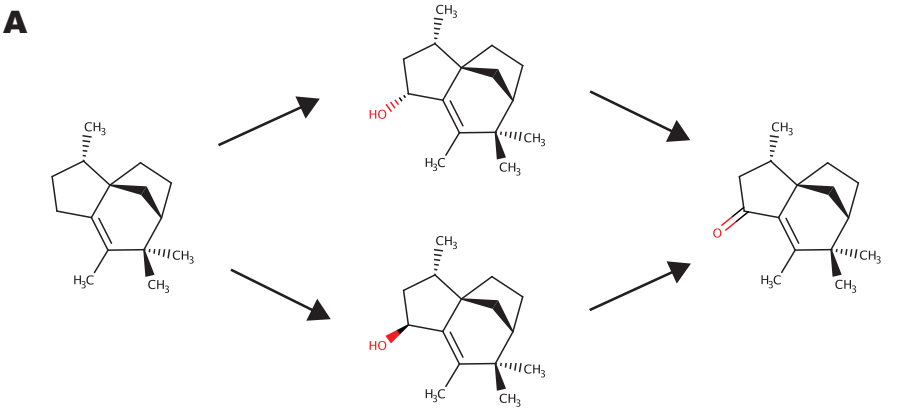
712

713 **Figure S3.** Amino-acid comparison of the 26 marine bacterial CYPs identified in this study.

714 The top quadrant presents the percentage of identity among the CYPs and the lower quadrant

715 the number of differences.

716



717

718 **Figure S4.** The chemical modifications performed by CYPs. (A) epi-isozizaene to

719 albaflavenone and (B) tetrahydrofuran to gamma-butyrolactone.

720