Philipps Universität Marburg

Konstruktion synthetischer sekundärer Chromosomen zur Charakterisierung von DNA-Reparatur und Segregation in *Escherichia coli*

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich der Biologie

der Philipps-Universität Marburg

vorgelegt

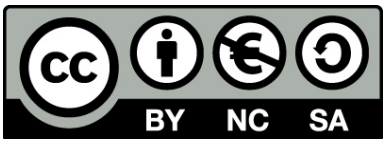
von

Daniel Schindler

Master of Science

aus Berlin-Spandau

Marburg November, 2016



Die Untersuchungen zur vorliegenden Arbeit wurden von Mai 2012 bis November 2016 am LOEWE Zentrum für Synthetische Mikrobiologie (SYNMIKRO) der Phillips-Universität Marburg unter der Leitung von Prof. Dr. Torsten Waldminghaus durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg
als Dissertation angenommen am:

Erstgutachter: Prof. Dr. Torsten Waldminghaus
Zweitgutachter: Prof. Dr. Martin Thanbichler

Tag der mündlichen Prüfung: 20.12.2016

Die in dieser Dissertation vorgestellten Ergebnisse sind in folgenden Originalpublikationen veröffentlicht, zur Veröffentlichung eingereicht oder zur Veröffentlichung vorbereitet:

Messerschmidt SJ, Kemter FS, **Schindler D**, Waldminghaus T (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. Biotechnol J 10(2):302-14 doi: 10.1002/biot.201400031

Messerschmidt SJ, **Schindler D**, Zumkeller CM, Kemter FS, Schallopp N, Waldminghaus T (under revision) Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*. Front Bioeng Biotechnol

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and assembly of DNA sequence libraries for chromosomal insertion in bacteria based on a set of modified MoClo vectors. ACS Synth Biol [Epub ahead of print]

Schindler D, Waldminghaus T (in preparation) Using experimentally constructed, rational synthetic chromosome designs to study functional interactions between segregation and DNA mismatch repair in *Escherichia coli*.

Schindler D, Lundius EG, Bruhn M, Murray S, Elf J, Waldminghaus T (in preparation) SeqA complexes in *Escherichia coli* exchange proteins rapidly and vary depending on replication patterns.

Weitere während der Promotion verfasste Veröffentlichungen, die nicht Gegenstand dieser Dissertation sind, sind nachfolgend aufgelistet:

Schindler D, Waldminghaus T (2015) Synthetic chromosomes. FEMS Microbiol Rev 39(6):871-91 doi: 10.1093/femsre/fuv030

Schindler D, Waldminghaus T (2013) "Non-canonical protein-DNA interactions identified by ChIP are not artifacts": response. BMC Genomics 14:638 doi: 10.1186/1471-2164-14-638

Inhaltsverzeichnis

1 Einleitung

Als Gregor Mendel 1856 seine Forschungen zur Vererbungslehre begann, war ihm noch nicht bewusst, dass die betrachteten Merkmale durch DNA vererbt werden. DNA konnte erstmals von Friedrich Miescher im Jahre 1869 isoliert werden. Es dauerte jedoch weitere 75 Jahre bis Oswaldo Avery experimentell zeigen konnte, dass DNA das Erbgut ist und nicht, wie zu diesem Zeitpunkt angenommen, Proteine (Avery *et al.* 1944). Diese Theorie konnte 1952 durch Alfred Hershey und Martha Chase bestätigt werden (Hershey and Chase 1952). Ein Jahr später konnten James Watson und Francis Crick die Struktur der DNA auflösen (Watson and Crick 1953). Seitdem wurden viele Entdeckungen auf der Ebene der DNA gemacht. Zuerst wurden nur kleinere Änderungen an der DNA durchgeführt, diese wurden aber im Laufe der Zeit immer komplexer. Heute ist es durch die Methoden der Synthetischen Biologie technisch möglich ganze Chromosomen nach natürlichem Vorbild zu synthetisieren und zu transplantieren (Hutchison *et al.* 2016; Gibson *et al.* 2010). In Zukunft könnte es möglich sein ganze Chromosomen *de novo* und damit „Designer-Organismen“ zu planen und zu synthetisieren, um diese abschließend zu erschaffen (Schindler and Waldminghaus 2015). 2016 hat sich ein Konsortium aus Wissenschaft und Industrie getroffen, um die Idee eines synthetischen, humanen Genoms als neue Herausforderung nach der Sequenzierung des humanen Genoms in Angriff zu nehmen. Dieses Projekt soll neben der Generierung stabiler Zelllinien als treibende Kraft genutzt werden, um neue Techniken in der Synthetischen Biologie zu entwickeln und DNA-Synthesekosten in Zukunft weiter zu senken (Boeke *et al.* 2016).

1.1 Was macht ein Chromosom zu einem Chromosom?

Das Erbgut einer jeden Zelle ist durch DNA kodiert und die Gesamtheit der DNA wird als Genom bezeichnet. Das Genom kann auf mehrere Moleküle, die sogenannten Chromosomen aufgeteilt sein. Bakterielle Genome bestehen in der Regel aus einem zirkulären Chromosom, welches durch extrachromosomale, autonom replizierende, zirkuläre DNA in Form von Plasmiden ergänzt sein kann. Chromosomen grenzen sich definitionsgemäß von Plasmiden durch ihre Größe und Kopienzahl, dem kodieren essentieller Gene sowie einer zellzyklusabhängigen DNA-Replikation ab (Krawiec and Riley 1990; Ochman 2002; Okada *et al.* 2005). Sekundäre Chromosomen, oder sogenannte Megaplasmide, sind ein Hybrid zwischen Chromosomen und Plasmiden. Sie besitzen plasmidähnliche Replikationsursprünge und Segregationssysteme, haben ansonsten jedoch die Eigenschaften eines Chromosoms. Aus diesem Grund wurde für diese Replikons versucht der Begriff des Chromids zu etablieren (Harrison *et al.* 2010). Allerdings wird die Bezeichnung Chromid wenig verwendet.

Chromosomen sind keine eigenständigen Systeme, denn sie benötigen die auf Ihnen kodierten Genprodukte, um sich selbst zu erhalten. Es werden unter anderem Mechanismen zur Vervielfältigung, Reparatur sowie für eine strukturierte und kompakte Organisation im dreidimensionalen Raum benötigt (Messerschmidt and Waldminghaus 2014; Touzain *et al.* 2011). Im Nachfolgenden wird die DNA-Replikation von bakteriellen Chromosomen anhand des Modellorganismus *Escherichia coli* dargelegt.

1.1.1 DNA-Replikation in Bakterien

Der Zellzyklus von *Escherichia coli* wird in drei Phasen gegliedert: Die B-Periode, in der die kritische Masse für die Initiation der DNA-Replikation akkumuliert wird, die C-Periode, die mit der Initiation der DNA-Replikation beginnt und mit dem Abschluss der DNA-Replikation in die D-Periode übergeht, in der aus der Ursprungszelle zwei Tochterzellen entstehen (Abb. 1a) (Stokke *et al.* 2012; Wallden *et al.* 2016). Im Näheren soll nun die C-Periode betrachtet werden, in der das Chromosom von *E. coli* mit einer Größe von 4,63 Megabasen (mb) in Form eines zirkulären Chromosoms verdoppelt wird (Blattner *et al.* 1997). Die Initiation der DNA-Replikation findet in Bakterien in einer spezifischen Region, dem Replikationsursprung (*oriC*) statt. Vom *oriC* ausgehend wird das zirkuläre Chromosom bidirektional repliziert, die Replikationsgabeln treffen in der Terminusregion aufeinander und beenden die DNA-Replikation mit der Duplikation des Chromosoms. Doch welche Eigenschaften machen den *oriC* zum Replikationsursprung?

oriC ist in der intergenischen Region zwischen *gidA* und *mioC* lokalisiert (Abb. 1b). Der minimale *oriC*, der in der Lage ist die DNA-Replikation zu initiieren, hat eine Länge von 245 Basenpaaren (bp) (Bates *et al.* 1995). Die Initiation der DNA-Replikation wird durch das Initiatorprotein DnaA bewerkstelligt. Dieses bindet die Konsensussequenz 5´-TTATNCACA-3´, welche als DnaA-Box bezeichnet wird (Messer 2002). Der Replikationsursprung weist eine hohe Dichte von DnaA-Boxen auf, wovon DnaA fünf mit hoher Affinität bindet (Abb. 1b) (Skarstad and Katayama 2013). Ein weiteres wichtiges Element des *oriC* ist eine AT-reiche Region, welche drei charakteristische 13mer DNA-Sequenzen aufweist und als *DNA unwinding element* (DUE) bezeichnet wird. Innerhalb des *oriC* gibt es weitere negativ oder positiv modulierende Bindestellen, wobei hier die GATC-Sequenz eine besondere Rolle spielt. Auf diese wird später im Zusammenhang mit der DNA-Adenin-Methyltransferase (Dam) und dem SeqA-Protein detailliert eingegangen.

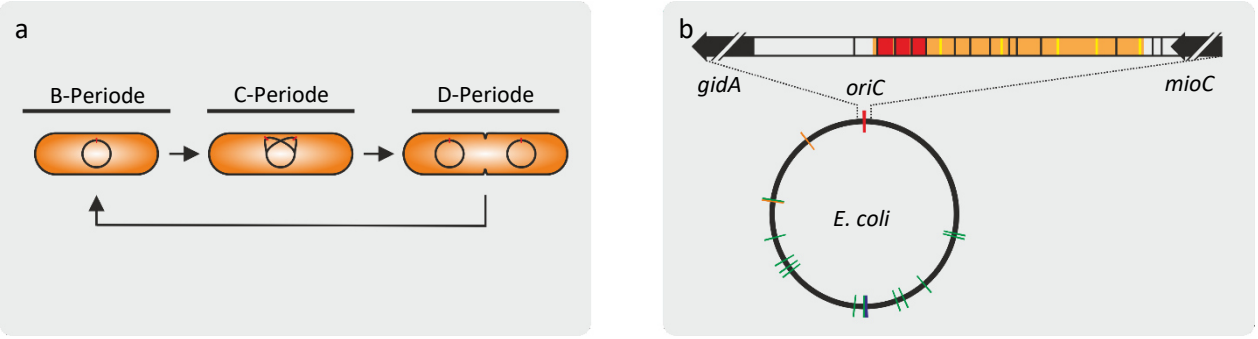


Abbildung 1: Zellzyklus und Aufbau des Chromosoms mit detaillierter Darstellung der Kernelemente des *E. coli oriC*. (a) Der Zellzyklus von *E. coli* ist in die B-, C- und D-Periode gegliedert, für Details siehe Text. (b) Darstellung des zirkulären Chromosoms von *E. coli* mit *oriC* (rot), *dif*-Sequenz (blau) und *ter* Sequenzen (grün), zwei *ter*-Sequenzen mit niedriger Bindeaffinität, die die falsche Orientierung aufweisen sind orange dargestellt. Zusätzlich ist der *oriC* mit den Kernelementen vergrößert dargestellt, wobei der minimal *oriC* beginnend bei dem äußersten der drei 13mere (rot) des DUE, orange hinterlegt ist. Die Anordnung der fünf elementaren DnaA-Boxen (gelb) und die Verteilung von GATCs (schwarz) ist ebenfalls indiziert.

Hat eine wachsende Zelle die kritische Masse für die Initiation der DNA-Replikation erreicht, ist der initiierende Schritt das Binden von DnaA an die DnaA-Boxen des *oriC* (Ozaki and Katayama 2009; Wolanski *et al.* 2014). DnaA multimerisiert anschließend auf der DNA, was zur Folge hat, dass die AT-reiche Region des DUE im *oriC* entwunden wird. Der entstehende offene Komplex dient zur Assemblierung der DNA-Helikase (DnaB) an der einzelsträngigen DNA durch eine Interaktion zwischen DnaA und DnaB. Dieser primäre Komplex bewirkt die Assemblierung der Replikationsmaschinerie mit der Primase DnaG und dem DNA-Polymerase III Holoenzym (Johnson and O'Donnell 2005). Aufgrund der bidirektionalen Replikation liegen zwei Replisomen in der Zelle vor, die in der Zellmitte lokalisiert sind (Wang *et al.* 2006). Die Replikation der DNA läuft mit einer Geschwindigkeit von etwa 750 bp/s ab, was zur Folge hat, dass die C-Periode eine Dauer von etwa 50 Minuten aufweist (McCarthy *et al.* 1976). Die Replikationsgabeln treffen am Ende der C-Periode in der Terminusregion, gegenüber von *oriC*, aufeinander und schließen damit die Replikation ab. Der eigentliche Prozess der Termination ist bisher unbekannt. Ein beschriebenes System, das einen Einfluss auf die Termination hat, ist das Tus Protein, das an gerichtete *ter*-Sequenzen bindet und für Replisomen nur in Richtung Terminusregion passierbar ist (Abb. 1b) (Hill *et al.* 1987; Kamada *et al.* 1996). Die *ter*-Sequenzen sind in der Terminusregion so arrangiert, dass die Replisomen die Terminusregion nicht verlassen können und eine Überreplikation unterbunden wird (Abb. 1b) (Duggin and Bell 2009; Duggin *et al.* 2008). Interessanterweise konnte für eine Deletion von *tus* bisher kein Phänotyp beschrieben werden, was auf ein redundantes System

hinweisen könnte. Andere Arbeiten deuten darauf hin, dass die Termination an der *dif*-Sequenz stattfindet (Hendrickson and Lawrence 2007). An dieser ausschließlich in der Terminusregion von Bakterien vorkommenden, hoch konservierten Sequenz können die sequenzspezifischen Rekombinasen XerC und XerD Chromosomen-Catenane auflösen (Kuempel *et al.* 1991; Blakely *et al.* 1993). Die DNA-Replikation ist abhängig von Protein-DNA-Interaktionen für die Regulation und Organisation. Diese Interaktionen werden unter dem Begriff des *Chromosome Maintenance* zusammengefasst; Im Nachfolgenden werden *Chromosome Maintenance* Systeme definiert und Beispiele vorgestellt.

1.1.2 *Chromosome Maintenance* Systeme in Bakterien

Das Chromosom muss nicht nur repliziert, sondern auch segregiert und strukturell organisiert werden, damit am Ende des Zellzyklus zwei Tochterzellen mit je einem Chromosom entstehen. Dies wird durch sogenannte *Chromosome Maintenance* Systeme bewerkstelligt, die aus Wechselwirkungen einzelner Proteine oder Proteinkomplexe mit einer mehr oder minder spezifischen DNA-Sequenz bestehen, wie das bereits erwähnte Beispiel der Tus *ter*-Sequenz Interaktion (Abb. 1b) (Touzain *et al.* 2011; Messerschmidt and Waldminghaus 2014). Bindemotive für *Chromosome Maintenance* Systeme weisen verschiedene Verteilungen über das Chromosom auf, zudem können die Sequenzen orientiert vorliegen (Abb. 2) (Schindler and Waldminghaus 2015; Touzain *et al.* 2011). Nachfolgend werden drei in *E. coli* wichtige *Chromosome Maintenance* Systeme exemplarisch vorgestellt.

Ein *Chromosome Maintenance* System zur Reparatur von DNA-Doppelstrangbrüchen basiert auf der Interaktion von RecBCD mit der *Chi* (*crossover hotspot instigator*) Sequenz 5´-GCTGGTGG-3´ (Smith *et al.* 1981; Taylor *et al.* 1985). Die *Chi*-Sequenzen sind in *E. coli* über das gesamte Chromosom verteilt und dort überrepräsentiert, da das Sequenzmotiv ca. alle 4500 bp (1008 Sequenzmotive) vorkommt, wobei bei einer Zufallsverteilung statistisch nur alle 65536 bp (70 Sequenzmotive) eine *Chi*-Sequenz erwartet werden würde (Abb. 2a) (El Karoui *et al.* 1999). Zudem zeigen die nachfolgenden, der *Chi*-Sequenz sehr ähnlichen Sequenzen, ebenfalls *Chi*-Aktivität: 5´-GCTAGTGG-3´ (38 % Aktivität), 5´-ACTGGTGG-3´ (11 % Aktivität), 5´-GTTGGTGG-3´ (6 % Aktivität) (El Karoui *et al.* 1999; Cheng and Smith 1984, 1987).

Ein weiteres *Chromosome Maintenance* System, das für die Segregation der DNA in *E. coli* verantwortlich ist, basiert auf der Interaktion der an das Divisom gebundenen Translokase FtsK, die die gerichtete *FtsK* orienting *polar sequences* (*KOPS*) 5´-GGGNAGGG-3´ erkennt und die DNA aktiv auf die beiden Tochterzellen verteilt (Bigot *et al.* 2005; Aussel *et al.* 2002; Errington *et al.* 2001; Levy *et al.* 2005). Interessanterweise ist das Motiv von *oriC* zur *dif*-Sequenz auf beiden Chromosomenhälften so angeordnet, dass FtsK die DNA in die korrekte Richtung transloziert (Abb. 2b). An der *dif*-Sequenz

interagiert FtsK mit der Topoisomerase TopoIV, um XerCD zu stimulieren und somit die Auflösung von Chromosomen-Catenanen zu bewirken (Ip *et al.* 2003; Hendrickson and Lawrence 2007; Grainge *et al.* 2007; Zechiedrich *et al.* 1997).

Als drittes Beispiel ist das GATC-Sequenzmotiv zu nennen, eine Sequenz deren Verteilung auf dem Chromosom einen Gradienten vom *oriC* zur *dif*-Sequenz aufweist (Abb. 2c). Die Sequenz ist im Chromosom von *E. coli* leicht überrepräsentiert und kommt alle 242 bp anstatt der statistisch erwarteten 256 bp vor. GATC-Sequenzen sind im *oriC*, in Promotoren und kodierenden Sequenzen einiger Gene, sowie in zwei symmetrisch angeordneten Regionen rechts und links des Replikationsursprungs deutlich überrepräsentiert (Strzelczyk *et al.* 2003; Riva *et al.* 2004; Sobetzko *et al.* 2016; Waldminghaus and Skarstad 2009; Barras and Marinus 1988). Da dieses *Chromosome Maintenance* System für die vorliegende Arbeit von großer Wichtigkeit ist, wird das GATC-Sequenzmotiv, die regulatorischen Funktionen und die interagierenden Proteine nachfolgend im Detail vorgestellt.

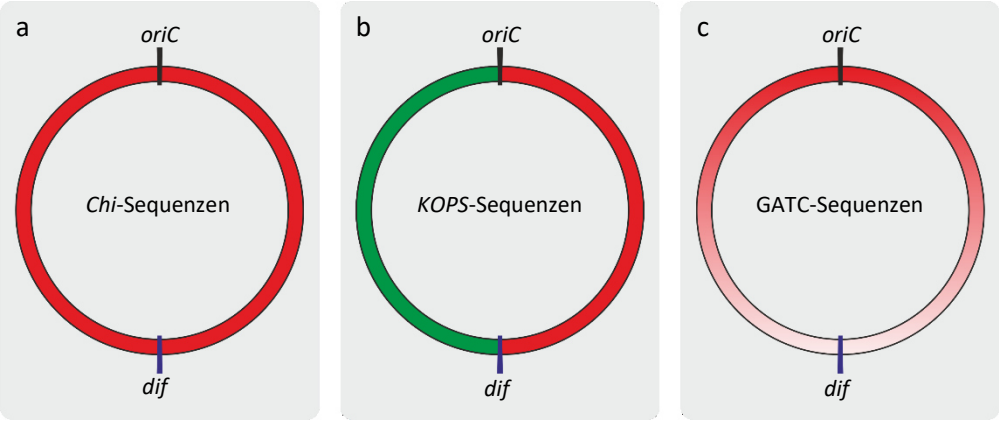


Abbildung 2: Darstellung des zirkulären Chromosoms von *E. coli* mit der Sequenzverteilung von drei verschiedenen *Chromosome Maintenance* Systemen. *oriC* (schwarz) liegt der *dif*-Sequenz (blau) im Chromosom gegenüber. (a) *Chi*-Sequenzen sind über das ganze Chromosom verteilt (rot). (b) *KOPS* zeigen eine gerichtete Verteilung vom Replikationsursprung zur Terminusregion die unterschiedlichen Orientierungen sind rot und grün indiziert. (c) Das GATC-Sequenzmotiv ist über das gesamte Chromosom von *E. coli* verteilt und zeigt einen von *oriC* zu *dif*-Sequenz verlaufenden Gradienten (rot = viel, weiß = wenig).

1.1.3 Das GATC-Sequenzmotiv und seine Bedeutung für *E. coli*

Das GATC-Sequenzmotiv ist ein Palindrom und wird auf beiden Seiten des DNA-Strangs an der Position N^6 des Adenins durch Dam methyliert (Abb. 3) (Geier and Modrich 1979; Marinus and Morris 1973; Marinus and Lobner-Olesen 2014). In *E. coli* sind nahezu alle der 19120 GATC-Sequenzen auf beiden DNA-Strängen methyliert (Marinus and Lobner-Olesen 2014). Während der DNA-Replikation werden

jedoch nur nichtmodifizierte Nukleotide in die DNA eingebaut und es entsteht ein Bereich hinter den Replikationsgabeln, in dem nur der parentale und nicht der neusynthetisierte DNA-Strang methyliert ist (Marinus 1987; Waldminghaus *et al.* 2012). Dieser Zustand wird als hemi-methyliert bezeichnet und bleibt so lange aufrecht bis Dam das Adenin wieder re-methyliert (Marinus and Lobner-Olesen 2014). Der hemi-methylierte Zustand dauert etwa ein bis zwei Minuten, was einem hemi-methylierten DNA-Abschnitt von 30 bis 120 Kilobasenpaare (kb) Länge hinter der Replikationsgabel entspricht (Campbell and Kleckner 1990; Ogden *et al.* 1988). Die hemi-methylierten GATCs dienen bei der Detektion einer Basenfehlpaarung dem MutSLH *DNA mismatch* Reparatursystem als Erkennungssequenz, an der alter und neuer DNA-Strang unterschieden werden können(Li 2008; Marinus and Lobner-Olesen 2014; Lenhart *et al.* 2016). Das hemi-methylierte GATC-Sequenzmotiv ist außerdem die Zielsequenz für SeqA, dass durch das Binden eine Re-methylierung durch Dam blockiert. In den nachfolgenden Kapiteln werden die einzelnen an dem beschriebenen Prozess beteiligten Proteine und ihre Funktionen im Detail dargelegt.

1.1.3.a GATC-Methylierung durch die DNA-Adenin Methyltransferase (Dam)

Die Dam-Methyltransferase katalysiert, als monomeres Protein, die Methylierung der GATC-Sequenz unter Verwendung von S-Adenosylmethionin (SAM) als Substrat, wobei als Produkt N^6-Methyladenin und S-Adenosylhomocystein (SAH) entsteht (Abb. 3) (Hattman *et al.* 1978; Chiang *et al.* 1996; Urig *et al.* 2002). In *E. coli* Zellen liegt nur eine geringe Anzahl an Dam Molekülen vor (Boye *et al.* 1992; Li *et al.* 2014; Szyf *et al.* 1984). Aufgrund der geringen Anzahl an Molekülen pro Zelle, aber der hohen Anzahl an GATC-Sequenzen, muss es sich bei Dam um ein effizientes Enzym handeln. Die Prozessivität konnte *in vitro* belegt werden und ist darauf zurückzuführen, dass Dam auf der DNA entlang gleitet, etwa 3000 GATC-Sequenzen zufällig scannt und während eines Bindevorgangs etwa 55 GATC-Sequenzen re-methyliert (Horton *et al.* 2005; Urig *et al.* 2002). Wenn *dam* in *E. coli* deletiert wird, liegt keines der GATCs methyliert vor. Da eine Detektion von Basenfehlpaarungen dadurch nicht mehr möglich ist, weisen die Zellen einen hypermutablen Phänotyp sowie eine Störung der Initiation der DNA-Replikation auf (Marinus *et al.* 1984; Boye *et al.* 1988; Marinus 2010). Interessanterweise führt auch eine Überexpression von *dam* und damit eine Verringerung der hemi-methylierten Phase zu einem hypermutablen Phänotyp (Herman and Modrich 1981; Yang *et al.* 2004).

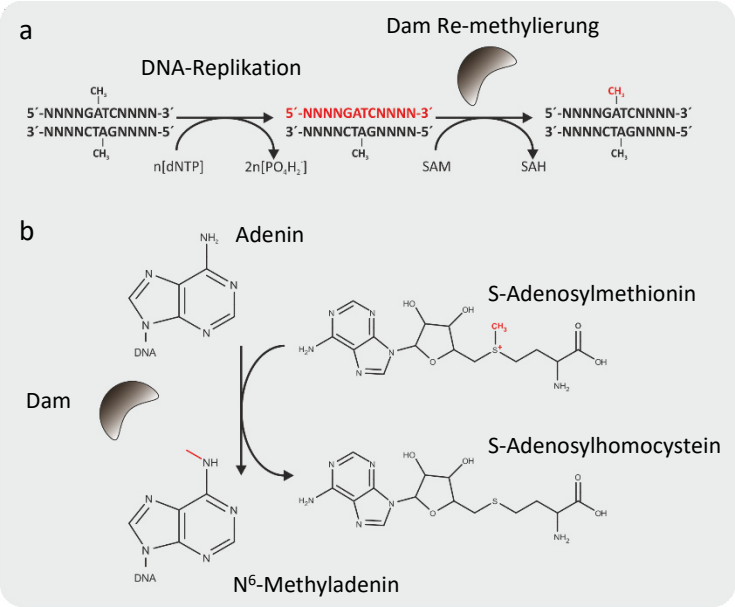


Abbildung 3: Entstehung von hemi-methylierten GATCs und der Prozess der Re-methylierung. (a) Während der DNA-Replikation werden nur nicht modifizierte Nukleotide eingebaut, die Folge sind hemi-methylierte GATC-Sequenzmotive, diese werden durch Dam re-methyliert. (b) Dam katalysiert die Methylierung an der Position N^6 des Adenins unter Verwendung des Substrats SAM. Als Produkte entstehen N^6-Methyladenin und der Methyldonor SAM wird zu SAH umgesetzt. Die übertragene Methylgruppe ist rot dargestellt.

1.1.3.b Sequestering Protein A (SeqA) und seine biologische Funktion

Es konnte in Experimenten gezeigt werden, dass das Einbringen von methylierten oder hemi-methylierten *oriC*-Replikons in *E. coli* Δ*dam* Stämme nur bedingt möglich ist und die Replikation dieser DNA unterbunden wird, un-methylierte *oriC*-Replikons können jedoch problemlos repliziert werden (Russell and Zinder 1987). Daraus wurde geschlussfolgert, dass es einen Sequestrierungsfaktor geben muss, der spezifisch mit den hemi-methylierten GATCs des Replikationsursprungs interagiert. Dieser Faktor konnte identifiziert werden und wurde SeqA genannt (Lu *et al.* 1994; Waldminghaus and Skarstad 2009). SeqA ist in zwei funktionelle Domänen gegliedert, eine C-terminale DNA-Bindedomäne und eine N-terminale Dimerisierungsdomäne, die über einen flexiblen Linker miteinander verbunden sind (Daghfous *et al.* 2009; Chung *et al.* 2009; Guarne *et al.* 2005; Guarne *et al.* 2002). SeqA kommt in großer Anzahl in *E. coli* Zellen vor und bindet als Dimer zwei benachbarte hemi-methylierte GATCs (Abb. 4a) (Slater *et al.* 1995). An die DNA gebundene SeqA Dimere können multimerisieren und bilden dadurch eine filamentartige Struktur die für die Organisation der neusynthetisierten DNA verantwortlich ist (Abb 4b) (Joshi *et al.* 2013; Chung *et al.* 2009; Odsbu *et al.* 2005). Zum einen bindet SeqA die hemi-methylierten GATCs, die hinter der Replikationsgabel entstehen und hält somit den

hemi-methylierten Zustand für ein bis zwei Minuten aufrecht (Sanchez-Romero *et al.* 2010; Waldminghaus *et al.* 2012; Campbell and Kleckner 1990; Ogden *et al.* 1988). Dadurch lokalisiert SeqA während der DNA-Replikation nahe den DNA-Replikationsgabeln und kann als Marker für neu synthetisierte DNA verwendet werden (Waldminghaus *et al.* 2012; Helgesen *et al.* 2015). Zum anderen sequestriert SeqA den *oriC* nach der DNA-Replikation und unterbindet somit eine erneute Initiation der DNA-Replikation durch DnaA, wodurch die Initiation der DNA-Replikation auf ein Mal pro Zellzyklus beschränkt wird (Taghbalout *et al.* 2000; Campbell and Kleckner 1990; Slater *et al.* 1995; Waldminghaus and Skarstad 2009; Boye *et al.* 2000). Eine Deletion von *seqA* führt zu einer erhöhten und asynchronen Initiation der DNA-Replikation (Boye *et al.* 1996; Lu *et al.* 1994). Eine *seqA* Überexpression hingegen verlängert die Dauer der hemi-methylierten Phase für GATCs und führt zu einer geringeren Initiationsrate der DNA-Replikation (Lu *et al.* 1994; Bach *et al.* 2003; Saint-Dic *et al.* 2008; Boye *et al.* 1996). Da SeqA hemi-methylierte GATCs bindet und sequestriert, kann Dam diese nicht binden und re-methylieren. Re-methyliert Dam die GATCs kann SeqA diese hingegen nicht mehr binden. Der genaue Zusammenhang zwischen diesen beiden Proteinen sowie die molekularen Prozesse dieses Wechselspiels sind bisher jedoch nicht ausreichend bekannt.

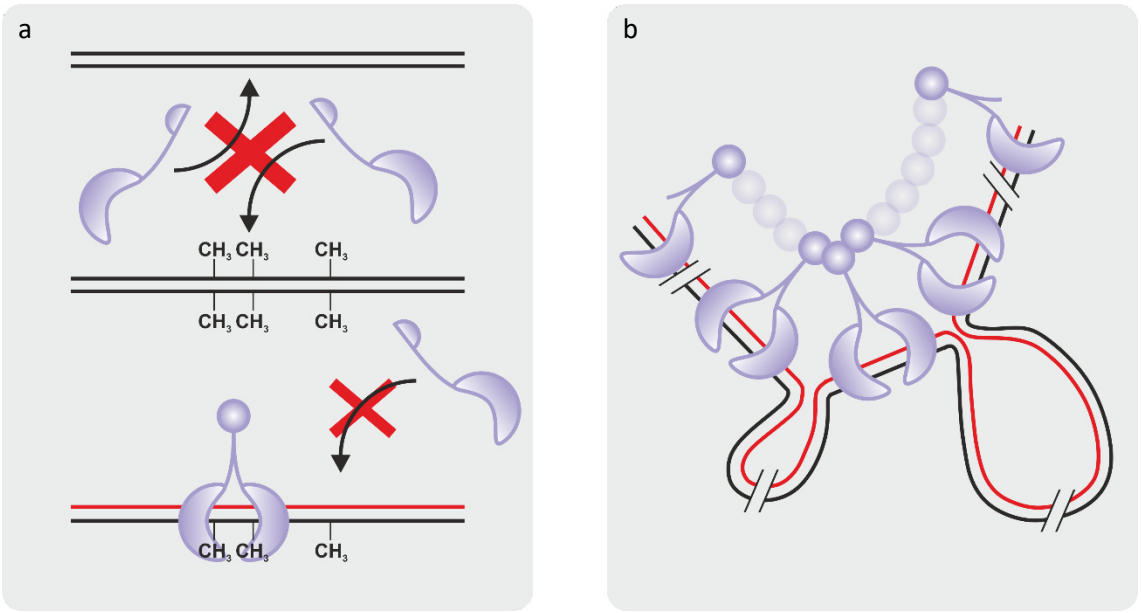


Abbildung 4: DNA-Bindeverhalten und übergeordnete Struktur von SeqA. (a) SeqA (lila) bindet als Dimer zwei benachbarte hemi-methylierte GATCs. SeqA bindet weder voll-methylierte oder un-methylierte GATC-Paare noch einzelne hemi-methylierte GATC-Sequenzen mit einer vergleichbaren Affinität (indiziert durch durchgestrichene Pfeile). (b) SeqA ist in der Lage zu multimerisieren und dadurch die DNA zu strukturieren. Multimerisiertes SeqA bildet filamentartige Strukturen, die durch transparente N-Termini des SeqA Proteins verdeutlicht werden. Neu synthetisierte DNA ist rot und parentale DNA ist schwarz dargestellt.

1.1.3.c Die Bedeutung der Hemi-methylierung für die *DNA mismatch* Reparatur durch MutSLH

Im Jahr 2015 hat Paul Modrich, ein Vorreiter in der Erforschung der *DNA mismatch* Reparatur, mit zwei Kollegen den Nobelpreis in Chemie *„for mechanistic studies of DNA repair"* erhalten (Cressey 2015; Radman 2016). Die DNA-Replikation verläuft nicht perfekt, im Durschnitt werden alle 2,6 x 10^6 Nukleotide Fehlpaarungen durch das Replisom generiert (Schaaper 1993). Durch die Korrekturfunktion der DNA-Polymerasen wird die Fehlerrate auf 4,5 x 10^{-8} reduziert, was in *E. coli* einer Rate von zehn Mutation pro Chromosom pro 1000 Generationen entspricht. Im Laufe der Evolution sind zusätzliche DNA-Reparatursystem entstanden, wodurch die Mutationsrate in *E. coli* bei 0,2 Mutationen pro Chromosom pro 1000 Generation liegt (Schaaper and Dunn 1991). Durch modernere Methoden des *Next-Generation Sequencing* und neutrale Mutations-Akkumulations-Experimente konnte gezeigt werden, dass die Mutationsrate in *E. coli* um den Faktor 5 bis 10 höher ist, als zuvor publiziert wurde und bei ein bis zwei Mutationen pro Chromosom pro 1000 Generationen liegt (Foster *et al.* 2015). Ein *DNA mismatch* Reparatur defizienter *E. coli* Stamm hat eine um den Faktor 100 bis 200-fach erhöhte Mutationsrate und somit eine Mutationsrate von 100 bis 200 Mutationen pro Chromosom pro 1000 Generationen, was die Bedeutung der *DNA mismatch* Reparatur für die Integrität des Chromosom zeigt (Foster *et al.* 2015; Marinus 2010; Lee *et al.* 2012).

Entstehen Fehlpaarungen in der DNA, die nicht durch das Replisom korrigiert werden, wird die *DNA mismatch* Reparatur durch einen heterotetrameren MutSL-Komplex ausgeführt. MutL und MutS sind hoch konserviert und in allen Domänen des Lebens zu finden, was ein Indiz für die Bedeutung und den frühen evolutiven Ursprung dieses Systems ist (Kolodner 1996; Kunkel and Erie 2005). In einer Reihe von γ-Proteobakterien kann zusätzlich noch MutH im Genom kodiert sein. MutH bindet hemi-methylierte GATC-Sequenzen und schneidet nach Aktivierung spezifisch den un-methylierten, neu synthetisierten Strang, um eine Neusynthese des fehlerhaften Bereichs zu ermöglichen (Lahue *et al.* 1987). Interessanterweise wird MutH nur in Organismen gefunden, die zusätzlich Dam und SeqA besitzen (Brezellec *et al.* 2006). In Organismen, die keine durch Dam methylierten GATCs besitzen, weist meist MutL eine Domäne mit einer Endonukleaseaktivität auf (Erdeniz *et al.* 2007; Kadyrov *et al.* 2006; Kadyrov *et al.* 2007; Kosinski *et al.* 2008). Im Nachfolgenden soll die methylorientierte *DNA mismatch* Reparatur in *E. coli* detailliert betrachtet werden (Abb. 5).

MutS erkennt im ADP gebundenen Zustand DNA-Basenfehlpaarungen und bindet diese als Dimer (Su and Modrich 1986; Acharya *et al.* 2003). Wird eine Basenfehlpaarung detektiert wird ADP zu ATP ausgetauscht und zwei MutL rekrutiert, die mit dem MutS Dimer nur im ATP-gebundenem Zustand interagieren können. Eine Interaktion dieses heterotetrameren Komplexes mit MutH bewirkt, dass MutL

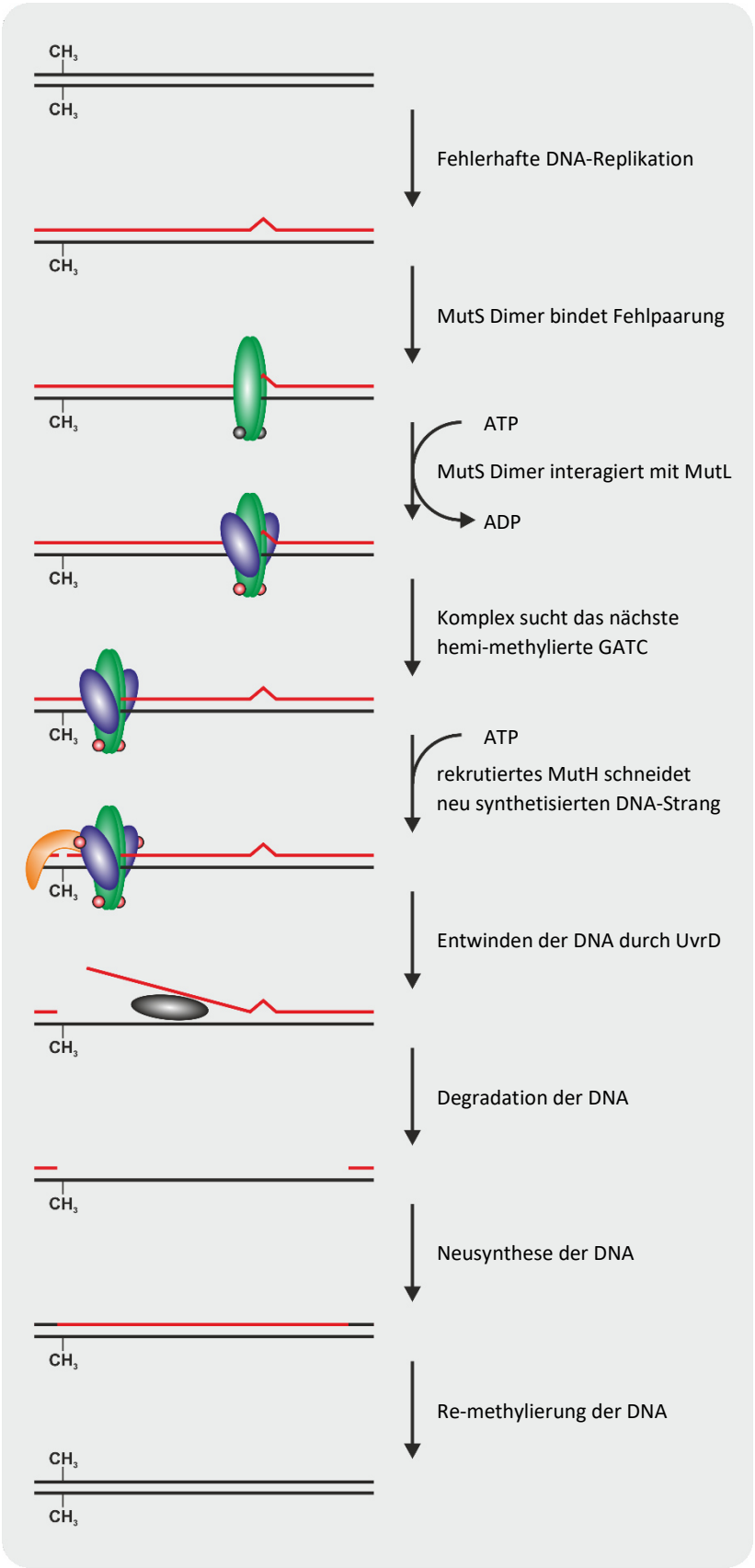


Abbildung 5: Darstellung des Ablaufes der *DNA mismatch* Reparatur in *E. coli*. Kommt es während der DNA-Synthese zu einer Basenfehlpaarung wird diese durch das MutS Dimer detektiert (grün) und es entsteht durch die Interaktion von MutS und MutL (blau) ein heterotetramerer Komplex. Der Komplex aktiviert MutH (orange), das spezifisch an hemi-methylierter DNA den neu synthetisierten Strang schneidet. Dieser wird degradiert und anschließend neu synthetisiert. Abschließend wird das GATC durch Dam re-methyliert. ATP ist durch rote Kreise und ADP durch schwarze Kreise dargestellt.

ATP bindet und dadurch die Endonukleaseaktivität von MutH verstärkt. MutH schneidet spezifisch den neu synthetisierten DNA-Strang am nächsten in 5´- oder 3´-Richtung liegenden hemi-methylierten GATC (Welsh *et al.* 1987; Cooper *et al.* 1993; Grilley *et al.* 1993). Wie das nächste GATC erreicht wird ist fraglich und es gibt verschiedene Modelle die auf Gleiten von MutSL auf der DNA oder der Bildung von DNA-Schlaufen basieren (Li 2008). Bei einer Schlaufenbildung ist es nicht sicher, ob der MutSL-Komplex an der Fehlpaarung fixiert ist oder diese Fehlpaarung in die DNA-Schlaufe transferiert wird. Im Anschluss wird die DNA durch die Helikase UvrD entwunden und eine der vier Exonukleasen (RecJ, ExoI, ExoVII und ExoX) degradiert den neusynthetisierten Strang ausgehend von der Schnittstelle (Matson 1986; Viswanathan and Lovett 1998; Yamaguchi *et al.* 1998). Die Läsion wird durch die DNA-Polymerase III aufgefüllt und die Lücke abschließend durch die DNA-Ligase LigA verbunden (Lahue *et al.* 1989; Lehman 1974; Nandakumar *et al.* 2007). Abschließend wird das GATC-Sequenzmotiv durch die Dam-Methyltransferase re-methyliert und neuer und alter Strang können nicht mehr voneinander unterschieden werden (Marinus and Lobner-Olesen 2014). Die Abhängigkeit der *DNA mismatch* Reparatur von der Existenz hemi-methylierter GATC-Sequenzen in *E. coli* lässt die Frage offen, warum Dam, SeqA und MutH evolutiv konserviert koexistieren. Der Zusammenhang zwischen SeqA und Dam sowie Dam und MutH ist durch die Konkurrenz um die Bindestellen offensichtlich, denn durch die Re-methylierung können SeqA und MutH nicht mehr binden. Es ist allerdings fraglich, ob SeqA funktionell mit der *DNA mismatch* Reparatur interagiert, oder ob die Proteine ebenfalls um die hemi-methylierten GATCs konkurrieren.

1.1.3.d Hypothesen zur Interaktion von Dam, SeqA und der *DNA mismatch* Reparatur

Die Abläufe hinter der Replikationsgabel sind sehr komplex und nicht vollständig verstanden. SeqA und Dam konkurrieren um die hemi-methylierten GATC-Bindestellen. Eine Veränderung der Molekülzahl sowohl für Dam als auch für SeqA hat eine phänotypische Ausprägung, was auf ein Equilibrium zwischen Dam und SeqA hinweisen könnte (Saint-Dic *et al.* 2008; Bach *et al.* 2003; Yang *et al.* 2004; Herman and Modrich 1981). Für die Re-methylierung durch Dam sind zwei verschiedene Prozesse denkbar: Ein kompetitiver Prozess in dem SeqA und Dam um jede Bindestelle konkurrieren oder ein geordneter Prozess indem erst SeqA bindet, nach einer gewissen Verweildauer wieder dissoziiert und die Bindestelle für Dam zur Re-methylierung freigibt (Abb. 6a). Wird die *DNA mismatch* Reparatur mit in den Prozess einbezogen, wäre es denkbar, dass die Funktion von SeqA darin besteht, das Zeitfenster der Hemi-methylierung zu verlängern, um eine effiziente *DNA mismatch* Reparatur durchführen zu können. Es ist fraglich ob zwischen SeqA und der *DNA mismatch* Reparatur eine funktionelle Interaktion vorliegt, oder ob es distinkte DNA-Regionen hinter der Replikationsgabel gibt, in denen die Mechanismen

separiert, sequenziell ablaufen. Hier ist zu bedenken, dass zwischen den DNA-Replikationsgabeln und der SeqA Struktur eine Distanz von 200 bis 300 nm vorliegt, wodurch eine strukturelle Gliederung der DNA hinter der Replikationsgabel in Reparatur – Sequestrierung – Re-methylierung denkbar wäre (Helgesen *et al.* 2015; Hasan and Leach 2015). Die existierenden Daten können keine der Hypothesen belegen und *in vitro* Experimente gestalten sich aufgrund der multimerisierenden Eigenschaften von SeqA sehr schwierig, das aggregiert Protein *in vitro*. Phänotypische Analysen von Mutanten sind nicht möglich, da die Deletion von *seqA, dam* und der einzelnen *mutSLH* Gene starke phänotypische Ausprägungen besitzen und somit die Analyse von Doppelmutanten nicht aussagekräftig ist. Für ein besseres Verständnis dieser molekularen Prozesse, bedarf es neuer Methoden und Konzepte, um dieses Wechselspiel zu analysieren.

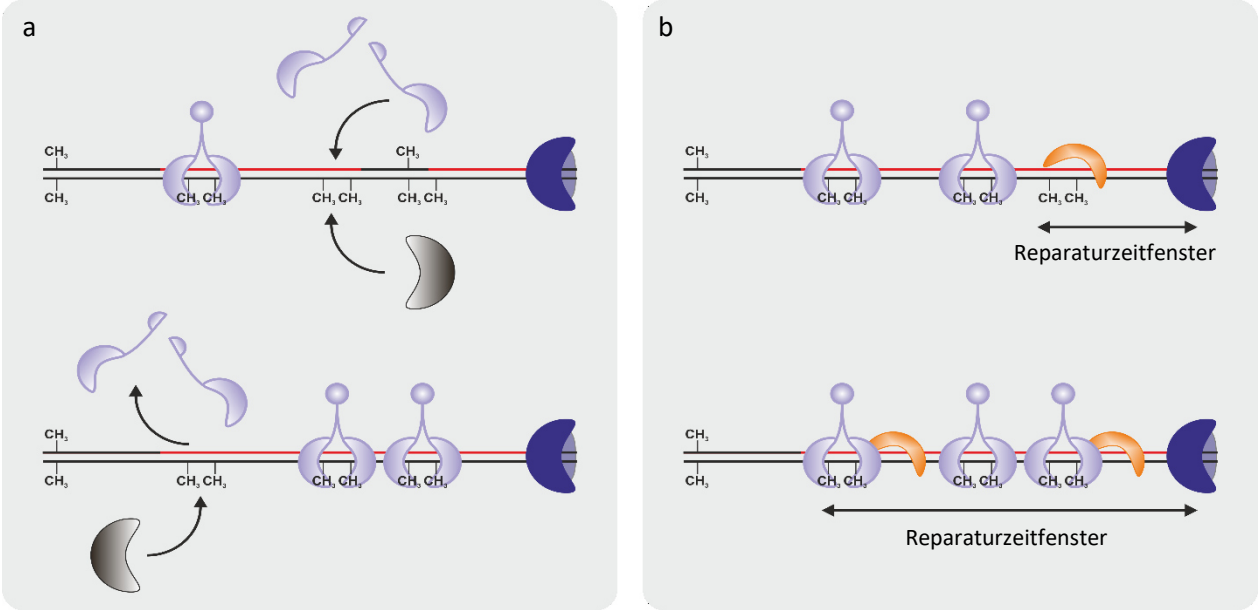


Abbildung 6: Darstellung zweier Re-methylierungs-Hypothesen sowie Modelle für eine mögliche Organisation von SeqA und MutH hinter dem Replisom. (a) Für die Re-methylierung der GATC-Sequenzen sind zwei Modelle denkbar: SeqA (lila) kann mit Dam (schwarz) um jede einzelne Bindestelle konkurrieren, wodurch die hemi-methylierte DNA nicht durchgängig ist (oben). Alternativ wäre ein geordneter Mechanismus denkbar, der erst SeqA binden lässt, gefolgt von der Dissoziation und anschließender Re-methylierung durch Dam (unten). Dadurch würde eine strikte Abfolge von hemi-methylierter und methylierter DNA hinter dem Replisom (blau) vorliegen. (b) Es sind zwei Hypothesen für die Organisation von SeqA und MutH denkbar: SeqA und MutH (orange) sind unabhängig voneinander und haben distinkte Regionen hinter den Replikationsgabeln (oben) oder interagieren funktionell an der hemi-methylierten DNA (unten). Assoziation und Dissoziation sind durch Pfeile dargestellt, Zeitspannen durch Doppelpfeile. Der Einfachheit halber ist nur der synthetisierte Leitstrang dargestellt, parentale DNA (schwarz) und neusynthetisierte DNA (rot) sind farblich unterschieden.

1.2 Multiple Chromosomensysteme am Beispiel von *Vibrio cholerae*

Rein nach dem deutschen Sprichwort „Keine Regel ohne Ausnahme“ gibt es in der Natur Bakterien, die mehr als ein Chromosom besitzen. Dies gilt beispielsweise für die gesamte Familie der *Vibrionaceae*, in der alle Arten zwei Chromosomen besitzen (Okada *et al.* 2005; Egan *et al.* 2005). Der Erreger der Cholera, *Vibrio cholerae*, hat sich als Modellsystem für die Forschung an Bakterien mit multiplen Chromosomen etabliert (Heidelberg *et al.* 2000; Val *et al.* 2014b; Schoolnik and Yildiz 2000; Jha *et al.* 2012). Die Funktionsweise und Struktur des Replikationsursprungs (*oriI*) des primären, 2,96 mb großen Chromosoms (*chrI*) ähnelt der des *oriC* bei *E. coli* (Heidelberg *et al.* 2000). Das zweite Chromosom (*chrII*) hat eine Größe von 1,07 mb und der Replikationsursprung (*oriII*) ähnelt dem Replikationsursprung von Plasmiden (Heidelberg *et al.* 2000; Gerding *et al.* 2015). Die Initiation von *oriII* ist durch ein eigenes Initiatorprotein geregelt: RctB (Egan and Waldor 2003; Heidelberg *et al.* 2000; Duigou *et al.* 2006). Es stellt sich die Frage, ob beide Chromosomen zeitgleich (Initiationssynchronie) oder versetzt initiieren, dafür aber möglicherweise zeitgleich die DNA-Replikation terminieren (Terminationssynchronie). Zu der Initiation des zweiten Chromosoms wurden kontroverse Daten publiziert (Rasmussen *et al.* 2007; Egan *et al.* 2004). In einer aktuellen Publikation konnte belegt werden, dass die Replikation einer regulativen Sequenz auf dem primären Chromosom, *crtS* (*chrII replication triggering site*) genannt, die Initiation der DNA-Replikation des zweiten Chromosoms bestimmt (Val *et al.* 2016). Erst wenn die *crtS* repliziert wurde, findet die Initiation der DNA-Replikation an *oriII* statt. Bedingt durch die räumliche Lage der *crtS*, wird die DNA-Replikation beider Chromosomen zeitgleich terminiert.

Aufgrund der Unabhängigkeit des zweiten Chromosoms vom Initiationskomplex des primären Chromosoms würde sich das zweite Chromosom von *V. cholerae* als ideales System anbieten, um in *E. coli* ein synthetisches sekundäres Chromosom als neue Plattform für die Synthetische Biologie zu etablieren. Die Funktionalität von Replikons basierend auf dem zweiten Chromosom von *Vibrio cholerae* in *E. coli* konnte zudem bereits belegt werden (Egan and Waldor 2003).

1.3 Synthetische Biologie - Gegenwart und Zukunft

Die Synthetische Biologie ist ein neues Feld der Naturwissenschaften, das versucht biologische Prozesse aus der Sichtweise eines Ingenieurs zu betrachten, Prozesse und Teile zu modularisieren und neu zusammenzusetzen (Serrano 2007). Die Europäische Union hat durch eine *NEST High-Level Expert Group* im Jahr 2005 folgende Definition formuliert:

> *"Synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems which display functions that do not exist in nature. This engineering perspective may be applied at all levels of the hierarchy of biological structures – from individual molecules to whole cells, tissues and organisms. In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way."* (European Commission 2005)

Ein wichtiger Aspekt der Synthetischen Biologie ist es neue Techniken zu etablieren und anzuwenden, um beispielsweise immer komplexere Fragestellungen in den Lebenswissenschaften zu beantworten bzw. komplexe Synthesewege von Feinchemikalien und pharmazeutisch relevanten Produkten zu ermöglichen.

1.3.1 Methoden der Synthetischen Biologie: DNA-Assemblierung und DNA-Sequenzanalyse

Eine grundlegende Technik in molekularbiologischen Laboren ist das Herstellen rekombinanter DNA. Diese Methode wurde in den 70ern erstmals durch das klassische Klonieren mittels enzymatischer Restriktion und Ligation von DNA-Fragmenten eingeführt (Cohen *et al.* 1972; Cohen *et al.* 1973; Bolivar *et al.* 1977). Das klassische Klonieren wird noch heute in molekularbiologischen Laboren durchgeführt, jedoch ist es aufgrund einer Limitierung an geeigneten Enzymen auf simple Klonierungen mit nur wenigen DNA-Fragmenten beschränkt. Diese Methode ist nicht sehr flexibel, häufig werden DNA-Klonierungen für ein Projekt geplant und können in neuen Projekten nicht neu kombiniert werden. Durch die modulare Denkweise in der Synthetischen Biologie kam es zu der Entwicklung neuer DNA-Assemblierungsmethoden für die Herstellung rekombinanter DNA-Fragmente, um diese möglichst effizient und wiederverwendbar assemblieren zu können (Ellis *et al.* 2011; Chao *et al.* 2014; Cobb *et al.* 2014). Im Nachfolgenden werden drei etablierte Methoden näher beschrieben: *Gibson Assembly*, homologe Rekombination in Hefe und das *Golden Gate Cloning* mit dem daraus resultierenden *Modular Cloning* (MoClo) System (Abb. 7).

Die *Gibson Assembly* ist eine Methode, in der DNA-Fragmente aufgrund von homologen DNA-Sequenzen an ihren Enden *in vitro* assembliert werden (Gibson *et al.* 2009). Es handelt sich um eine Reaktion in der drei Enzyme das gewünschte DNA-Produkt assemblieren, wofür eine DNA-Polymerase, eine hitzestabile DNA-Ligase und eine 5´-Exonuklease benötigt werden. Die Methode basiert darauf, dass die

5´-Exonuklease die Enden der DNA von 5´- in 3´-Richtung abbaut und sich dadurch die homologen DNA-Sequenzen aneinander anlagern können. Geschieht dies, kann die DNA-Polymerase vom 3´-Ende ausgehend die entsprechenden Lücken auffüllen, die abschließend von der DNA-Ligase verbunden werden. Die *in vitro* assemblierte DNA wird abschließend in einen Rezipienten eingebracht. Es gibt eine Abwandlung dieser Methode, in der auf die Ligase verzichtet wird und die endogene DNA-Reparatur des DNA-Rezipienten, meist *E. coli*, verwendet wird (Fu *et al.* 2014). Die Größe von Replikons, die in *E. coli* eingebracht und über viele Generationen stabil bleibt, ist limitiert und die Effizienz der DNA-Transformation nimmt deutlich mit zunehmender Größe ab (Gibson *et al.* 2008a; Sheng *et al.* 1995).

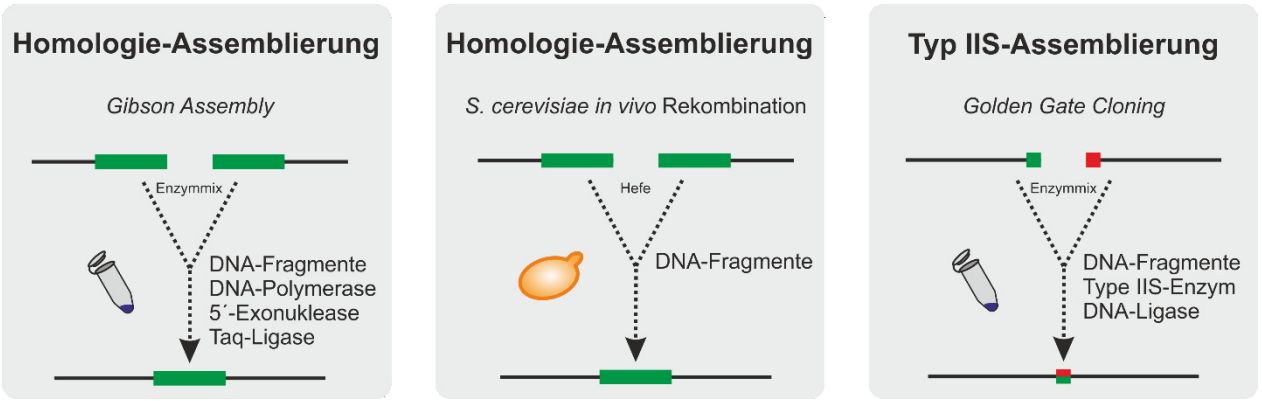


Abbildung 7: Darstellung dreier verschiedener DNA-Assemblierungsmethoden. Edukte, Produkte und die benötigten Elemente sind für die drei im Text erläuterten Methoden dargestellt. Ein Reaktionsgefäß indiziert eine *in vitro* Reaktion wohingegen die Hefezelle eine *in vivo* Reaktion indiziert. Für Details siehe die Beschreibung der jeweiligen Methode im Text. Grafik verändert nach Schindler und Waldminghaus 2016.

Die Hefe *Saccharomyces cerevisiae* kann deutlich größere Replikons über viele Generationen stabil aufrechterhalten. Ein weiterer Vorteil der Hefe ist, dass diese ein äußerst effizientes Reparatursystem für Doppelstrangbrüche besitzt: Das homologe Rekombinationssystem (Renkawitz *et al.* 2014). Die homologe Rekombination der Hefe ist hocheffizient und wurde als effizientes *in vivo* DNA-Assemblierungssystem etabliert (Oldenburg *et al.* 1997; Raymond *et al.* 1999; van Leeuwen *et al.* 2015a, b). Zur *in vivo* DNA-Assemblierung benötigen die zu assemblierenden DNA-Fragmente mindestens 20 bp homologe DNA-Sequenz und das finale Konstrukt muss einen Replikationsursprung sowie einen Selektionsmarker für die Hefe besitzen (Gibson 2009; Joska *et al.* 2014). Die Effizienz der DNA-Assemblierung in der Hefe ist größer als bei der *Gibson Assembly*, zudem können viel größere Produkte erstellt werden und deutlich mehr Fragmente in einem Schritt assembliert werden (de Kok *et al.* 2014; Gibson *et al.* 2008b). Anhand der homologen Rekombination in Hefe konnten bereits ganze bakterielle Chromosomen in Hefe assembliert oder in Hefe transferiert und anschließend manipuliert

werden (Karas *et al.* 2013a; Karas *et al.* 2014; Karas *et al.* 2015; Noskov *et al.* 2012; Tagwerker *et al.* 2012; Karas *et al.* 2012; Hutchison *et al.* 2016; Gibson *et al.* 2008a; Gibson *et al.* 2010). Ein Nachteil, den sowohl die *Gibson Assembly* als auch die homologe Rekombination der Hefe aufweisen, ist, dass sie auf die homologen Bereiche von mindestens 20 bp angewiesen sind.

Eine weitere Methode für die DNA-Assemblierung von großen und komplexen DNA-Fragmenten ist das *Golden Gate Cloning*. Diese Methode basiert nicht auf homologen DNA-Sequenzen, sondern auf einer Restriktions-Ligations-Reaktion. Jedoch werden hier anstelle von Typ II Endonukleasen, die innerhalb einer spezifischen Erkennungssequenz schneiden, Typ IIS-Endonukleasen verwenden (Engler *et al.* 2008; Weber *et al.* 2011). Typ IIS-Endonukleasen besitzen eine spezifische Erkennungssequenz, schneiden jedoch gerichtet eine unspezifische Sequenz in einem definierten Abstand zur Erkennungssequenz (Szybalski *et al.* 1991). Durch geschicktes planen und das Verwenden von Typ IIS-Enzymen in Klonierungsexperimenten können Restriktion und Ligation in einer Reaktion durchgeführt werden, da das Produkt keine Erkennungssequenzen mehr aufweist und somit nicht geschnitten werden kann (Engler *et al.* 2008). Zudem ist es durch entsprechendes Design der entstehenden Überhänge möglich Assemblierung von mehreren DNA-Fragmenten in einer Reaktion durchzuführen (Engler *et al.* 2008; Weber *et al.* 2011). Werden des Weiteren entsprechende hierarchische Vektorsets verwendet, können sequenzielle DNA-Assemblierungen durchgeführt werden (Weber *et al.* 2011; Werner *et al.* 2012). Erstellte DNA-Elemente können durch ein Alternieren von Typ IIS-Enzymen und Resistenzgenen zwischen den aufeinanderfolgenden Assemblierungsschritten zu immer komplexeren DNA-Assemblierungen kombiniert werden. Diese sequenzielle Strategie wird als *Modular Cloning System* (MoClo) bezeichnet (Weber *et al.* 2011; Werner *et al.* 2012). Ein Vorteil dieser Methodik ist, dass hergestellte DNA-Fragmente wiederverwendet, beziehungsweise die zur Herstellung dieses Fragments verwendeten DNA-Sequenzen neu kombiniert werden können. Der Nachteil dieser Methode ist, dass die zu assemblierenden DNA-Fragmente keine der verwendeten Typ IIS-Erkennungssequenzen aufweisen dürfen, allerdings lassen sich vorhandene Erkennungssequenzen in den DNA-Fragmenten durch stille Punktmutationen eliminieren. Zudem entstehen aufgrund der Restriktion und Ligation zwischen den verbundenen DNA-Fragmenten drei bis vier Basenpaare lange Narbensequenzen, die auf den benötigten DNA-Überhang für das Verknüpfen von DNA-Fragmenten zurückzuführen sind.

Rekombinant hergestellte DNA-Fragmente müssen verifiziert werden. Dies geschieht durch Sequenzierung der DNA. Die dazu verwendete Standardmethode ist die Sanger-Sequenzierung (Stranneheim and Lundeberg 2012; Sanger *et al.* 1977). Diese Methode erlaubt es mit jeder Reaktion

etwa 1000 bp einer DNA-Sequenz zu analysieren, was ausreichend für einen Großteil der im Labor hergestellten DNA-Fragmente ist. Werden jedoch viele Veränderungen an Chromosomen durchgeführt oder ist es Ziel evolutive Prozesse zu analysieren, bieten sich die Methoden des *Next-Generation Sequencing* an. Die Durchführung des Projekts zur Sequenzierung des humanen Genoms hat zu der Entwicklung vieler neuer DNA-Sequenzierungsmethoden geführt und dadurch konnten die Kosten für Genom Sequenzierungen massiv gesenkt werden (Bennett *et al.* 2005; Sboner *et al.* 2011; Venter *et al.* 2001; Lander *et al.* 2001). Bei den Methoden des *Next-Generation Sequencing* wird nicht, wie bei der Sanger-Sequenzierung, ein einzelnes DNA-Molekül sequenziert, sondern eine Vielzahl von unterschiedlichen DNA-Fragmenten, die auf eine ursprüngliche DNA-Probe zurückzuführen sind (DNA-Bibliothek) (Goodwin *et al.* 2016; Bankier 2001; Buermans and den Dunnen 2014). DNA-Bibliotheken werden meist durch zufälliges fragmentieren der zu sequenzierenden DNA-Probe hergestellt und werden je nach verwendeter Sequenzierungstechnik entsprechend den Vorgaben für die Sequenzierung vorbereitet. Die anschließende Sequenzierung liefert als Ergebnis eine Vielzahl von in der Regel 30 bis 600 bp langen DNA-Sequenzen, welche anschließend computergestützt zu einer Genomsequenz assembliert oder anhand eines vorhandenen Referenzgenoms abgeglichen werden (Reinert *et al.* 2015). Die Fehlerrate während einer Sequenzierung wird meist durch eine hohe Abdeckung der gesamten Genomsequenz mit mehreren hundert Sequenzen pro Base auf ein Minimum reduziert (Paszkiewicz and Studholme 2010; Reuter *et al.* 2015). Die benötigte Abdeckung pro Base ist jedoch aus Kostengründen je nach Experiment zu bedenken, denn für die Sequenzierung eines unbekannten Genoms ist eine deutlich größere Datenmenge notwendig im Vergleich zur Analyse von Mutationen in einem Stamm zu dem bereits ein Referenzgenom vorliegt (Sims *et al.* 2014). Die Anwendung von *Next-Generation Sequencing* Methoden ist jedoch nicht auf die Sequenzierung von Genomen beschränkt, es ist möglich jegliche Art von DNA-Bibliotheken zu sequenzieren (Buermans and den Dunnen 2014). Beispielsweise kann das Binden eines Proteins auf genomischer Ebene durch das Verwenden entsprechender Protokolle bestimmt und analysiert werden, was als *ChIP-Seq* (*Chromatin Immunoprecipitation DNA-Sequencing*) bezeichnet wird (Furey 2012; Barski and Zhao 2009; Park 2009). Es ist außerdem möglich, durch das Verwenden entsprechender Sequenzierungsmethoden, eine Methylierung einzelner Basen der DNA festzustellen (Korlach and Turner 2012; Krueger *et al.* 2012; Flusberg *et al.* 2010). In Zukunft werden sich DNA-Sequenzierungstechniken noch weiterentwickeln, insbesondere dahingehend, dass Sequenzierungen noch günstiger werden und dass die Länge der einzelnen sequenzierten DNA-Fragmente größer werden wird.

1.3.2 *Genome Engineering* und synthetische Chromosomen

Genome Engineering ist eine Herangehensweise, um eine größere Anzahl an Veränderungen an einem Genom durchzuführen und steht der kompletten Neusynthese und Transplantation eines Chromosoms methodisch gegenüber. Bahnbrechende Arbeiten im Gebiet des *Genome Engineering* kommen aus der Arbeitsgruppe von Georg Church, in der es gelungen ist in *E. coli* alle 314 TAG Stopp-Codons zu verändern sowie anschließend *prfA* zu deletieren, dessen Genprodukt für die Termination der Proteinbiosynthese an UAG Stopp-Codons verantwortlich ist (Isaacs *et al.* 2011; Lajoie *et al.* 2013). Anschließend konnte dieser Stamm durch das Einbringen einer tRNA für das UAG-Codon und der korrespondierenden Aminoacyl-tRNA Synthetase so verändert werden, dass er abhängig von einer nicht proteinogenen Aminosäure ist (Lajoie *et al.* 2013). Dies ist äußerst interessant, da dieser Organismus nur durch nicht natürlich vorkommende Zusätze in Kultur überlebensfähig ist. In einer aktuellen Arbeit zeigt die gleiche Arbeitsgruppe, dass es sehr wahrscheinlich möglich sein wird, die verwendete Menge von Codons von 64 auf 57 zu reduzieren (Ostrov *et al.* 2016). Dieser Organismus hätte in Zukunft ein weites Anwendungsspektrum, beispielsweise in der biotechnologischen Produktion von Feinchemikalien, da Kulturen beispielsweise nicht mehr durch Phagen kontaminiert werden könnten. Bei dieser Arbeit ist jedoch der Sprung vom *Genome Engineering* zur Herstellung eines synthetischen Chromosoms notwendig, da solch weitreichende Veränderungen technisch nicht mit dem Methodenspektrum des *Genome Engineering* möglich sind (Ostrov *et al.* 2016).

Die Synthese ganzer Chromosomen ist möglich und wurde durch Arbeiten am Craig Venter Institute an Organismen der Gattung *Mycoplasma* stark beeinflusst. 2008 konnte erstmals ein 582 kb Chromosom von *Mycoplasma mycoides de novo,* basierend auf Oligonukleotiden synthetisiert werden (Gibson *et al.* 2008a). Allerdings dauerte es zwei weitere Jahre, um den ersten synthetischen Organismus JCV-syn1.0 zu generieren, dessen Chromosom auf *Mycoplasma mycoides* basiert und in *Mycoplasma capricolum* transplantiert wurde (Gibson *et al.* 2010). In einem darauffolgenden Projekt wurde von zwei Teams am Craig Venter Institute unabhängig voneinander versucht ein minimales bakterielles Chromosom zu designen und zu transplantieren (Hutchison *et al.* 2016). Interessanterweise scheiterten beide Projekte und ein minimales Chromosom (JCVI-syn3.0) konnte nur durch sequenzielle Reduktion erstellt werden. JCVI-syn3.0 basiert auf JCVI-syn1.0 und weist eine Genomreduktion von 1079 kb auf 531 kb auf mit lediglich 473 Genen, was nah an dem postulierten Set des minimalen Genoms von *Mycoplasma genitalium* mit ca. 265 bis 350 Genen liegt (Hutchison *et al.* 1999; Juhas *et al.* 2014; Koonin 2000). Erstaunlicherweise ist von 149 in dem Chromosom von JCVI-syn3.0 kodierten Genen die

biologische Funktion bis heute unbekannt, was Raum für Spekulationen über bisher unentdeckte biologische Funktionen offenlässt.

Mit dem *S. cerevisiae Synthetic Genome 2.0 Project* wurde begonnen eukaryotische Chromosomen komplett zu synthetisieren und sukzessive gegen das synthetisch hergestellte und re-designte Genom zu ersetzen (Sliva *et al.* 2015; Cooper *et al.* 2012). Im Rahmen dieses Projektes werden mehrere Designregeln angewendet, es werden repetitive Sequenzen und mobile Elemente (Transposons) entfernt und alle nicht-essentiellen Gene werden von Rekombinase-Erkennungssequenzen flankiert. Das Flankieren mit diesen Erkennungssequenzen soll es ermöglichen Rekombinationsvorgänge zu induzieren, die zu Deletionen, Insertionen und Neuanordnungen der nicht essentiellen Gene führen, wodurch genetische Diversität hervorgerufen werden soll (Dymond and Boeke 2012). Das erste synthetische Hefe Chromosom konnte 2014 fertiggestellt und publiziert werden (Annaluru *et al.* 2014). Dabei handelt es sich um das dritte Chromosom der Hefe, welches neben zahlreichen Veränderungen eine Reduktion von 316 kb auf 273 kb (13,5 %) aufweist (Annaluru *et al.* 2014). Für die Zukunft ist ein neues richtungsweisendes Projekt in der Entstehung: Die Herstellung eines synthetischen humanen Genoms (Boeke *et al.* 2016).

Die bisherigen funktionellen synthetischen Chromosomen sind auf dem Prinzip des Nachbauens oder einer Reduktion eines vorhandenen, natürlichen Designs basiert und nicht auf einem rationalen Design an einem Reißbrett zurückzuführen. Dies trifft auch auf das Design des *Synthetic Yeast 2.0 Project* zu, auch wenn hier die DNA nach und nach ersetzt wird und es sich nicht um eine sukzessive Reduktion handelt. Das Wissen über die Funktionsweise einer minimalen Zelle reicht heute noch nicht aus, um einen Organismus *de novo* zu generieren und so greift an dieser Stelle das Zitat von Richard Feyman:

"What I cannot create, I do not understand." Richard Phillips Feyman (1918-1988)

Dies deutet darauf hin, dass neue Vorgehensweisen für das Design von synthetischen Chromosomen benötigt werden, da nicht nur Genprodukte, sondern auch Sequenzmotive essentielle Funktionen besitzen. Könnten anhand von sekundären, nicht essentiellen synthetischen Chromosomen generelle Regeln und Designs für das *de novo* Design von Chromosomen generiert werden?

2 Zielsetzung dieser Arbeit

Das Planen und Herstellen von synthetischen Chromosomen ist aufgrund der technischen Möglichkeiten, die heute bestehen nicht utopisch. Das *de novo* Design funktioneller, synthetischer Chromosomen, ohne ein natürliches Vorbild ist jedoch (noch) nicht möglich. Die meisten Studien konzentrieren sich auf die essentiellen Gene, aber Chromosomen sind mehr als das. In der vorliegenden Arbeit soll ein *in vivo* System zur Analyse von *Chromosome Maintenance* Systemen in *E. coli* etablieren werden, um in Zukunft allgemeingültige Regeln für die Konstruktion synthetischer Chromosomen generieren zu können.

Plasmide verhalten sich nicht wie Chromosomen, deshalb muss zuerst ein Replikon hergestellt und charakterisiert werden, das sich in *E. coli* wie ein sekundäres Chromosom verhält. Dazu soll das sekundäre Chromosom von *Vibrio cholerae* als Vorbild genutzt werden. Doch wie kann ein solches Replikon in *E. coli* angewendet werden, um *Chromosome Maintenance* Systeme zu analysieren?

Es muss die Möglichkeit bestehen Sets von Chromosomen herzustellen, die sich in der Verteilung ausgewählter DNA-Sequenzmotive unterscheiden, um vergleichende Analysen zu ermöglichen. Dazu müssen Methoden entwickelt werden, um möglichst einfach, große DNA-Fragmente herzustellen, so dass diese in das sekundäre Chromosom integriert werden können. Es werden zudem Methoden für Design, Herstellung und Assemblierung der entsprechenden DNA-Fragmente benötigt.

Als *Proof of Principle* soll ein Set von drei synthetischen sekundären Chromosomen hergestellt werden, die sich in der Verteilung des GATC-Sequenzmotivs unterscheiden. Das Set sollte es ermöglichen eine Interaktion zwischen *DNA mismatch* Reparatur und Segregation zu analysieren. Die gezielt angeordnete GATC-Verteilung ermöglicht in dem jeweiligen synthetischen Chromosom ein Binden von SeqA und MutH, nur von MutH oder keines der Proteine. Das Herstellen und die vergleichende Analyse der Mutationsraten der verschiedenen synthetischen Chromosomen sollte eine Aussage ermöglichen, ob eine funktionelle Interaktion zwischen SeqA und der *DNA mismatch* Reparatur vorliegt oder nicht.

Es ist viel zu der Lokalisation von SeqA bekannt, viele weitere Fragen sind jedoch noch nicht hinreichend beantwortet. Es wäre ideal durch quantitative Methoden und darauf basierenden Modellierungen ein besseres Verständnis von SeqA zu erlangen. Dazu müssen verschiedene Parameter wie beispielsweise die Molekülzahl und die Fraktion an gebundenem und ungebundenem SeqA während des Zellzyklus bestimmt werden. Eine solche Modellierung wäre eine gute Grundlage, um in Zukunft Fragen bezüglich des Prozesses der Re-methylierung durch Dam beantworten zu können. Ein erstes Puzzleteil könnte zudem die Beantwortung der Frage sein, ob SeqA und Dam ein bestimmtes Mengenverhältnis zueinander aufweisen.

3 Ergebnisse

Die Ergebnisse der vorliegenden Dissertation sind in fünf Kapitel gegliedert. Jedes Kapitel stellt eine individuelle wissenschaftliche Studie dar, die Teilaspekte dieser Dissertation behandelt. Die Manuskripte sind entweder publiziert (Kapitel 3.1 und 3.3), befinden sich in der Revision (Kapitel 3.2) oder für die Einreichung bei einem Fachjournal vorbereitet (Kapitel 3.4 und Kapitel 3.5).

Jedes Kapitel wird durch eine kurze Zusammenfassung eingeleitet, die die elementaren Ergebnisse der Studie hervorhebt und die zudem eine kurze Beschreibung der Beiträge der jeweiligen, einzelnen Autoren beinhaltet.

3.1 Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*.

Heutzutage ist die Synthese ganzer Chromosomen möglich, allerdings konnte bisher noch kein *de novo* designtes und synthetisiertes Chromosom erfolgreich transplantiert werden. Die Herstellung primärer Chromosomen unterliegt großen Risiken, da bereits kleine Fehler zu einem nichtfunktionellen Chromosom führen, zudem ist der wirtschaftliche Aufwand sehr hoch. Aus diesem Grund ist es wichtig grundlegende Regeln für das Design primärer Chromosomen durch Experimente an sekundären Chromosomen zu erforschen. Des Weiteren bieten synthetische sekundäre Chromosomen eine alternative Plattform für biotechnologische Anwendungen für die Generierung von Produktionsstämmen im Vergleich zu technisch aufwendigen chromosomalen Integrationen oder Plasmiden.

Die meisten Bakterienarten besitzen ein Chromosom, *Vibrio cholerae* sowie alle anderen Vertreter der *Vibrionaceae* besitzen zwei Chromosomen. In diesem Kapitel wird die Konstruktion und Charakterisierung eines synthetischen sekundären Chromosoms in *E. coli*, nachfolgend als synVicII bezeichnet, basierend auf dem Vorbild des zweiten Chromosoms von *V. cholerae* beschrieben. synVicII besteht aus dem Replikationsursprung (*oriII*), dem Gen für das Initiatorprotein RctB sowie den Genen für ParA und ParB, die für die Segregation benötigt werden. Für weitere biotechnologische Anwendungen wurde synVicII zusätzlich mit Elementen für *in vitro* und *in vivo* Klonierungen ausgestattet. Mittels Zellzyklusanalysen und Bestimmung der Kopienzahl für synVicII wurde belegt, dass synVicII in *E. coli* wie das sekundäre Chromosom in *V. cholerae* in einfacher Kopie vorliegt. Die Verteilung von synVicII auf die Tochterzellen, im Nachfolgenden als Replikonstabilität bezeichnet, konnte in *E. coli* durch eine in dieser Studie neu entwickelte, auf Durchflusszytometrie basierende Methode bestimmt werden. Als ein wichtiger Aspekt konnte gezeigt werden, dass synVicII eine deutlich höhere Replikonstabilität als ein vergleichbares, *oriC*-basiertes Minichromosom besitzt. synVicII konnte als sekundäres Chromosom in *E. coli* etabliert werden und kann dadurch in Zukunft zur Beantwortung von Fragestellungen der DNA-Replikation in *E. coli* oder beispielsweise als Expressionsplattform in der Biotechnologie angewendet werden.

Sonja Messerschmidt hat in Rücksprache mit Torsten Waldminghaus die Studie konzipiert und durchgeführt. Franziska Kemter hat die Kopienzahl des Minichromosoms mit qPCR und Microarray mit Torsten Waldminghaus und Daniel Schindler bestimmt. Daniel Schindler hat das Microarray Experiment in Rücksprache mit Torsten Waldminghaus konzipiert und etabliert. Alle Autoren haben zum Design des Minichromosoms beigetragen. Das Manuskript wurde von Sonja Messerschmidt und Torsten Waldminghaus verfasst unter Beteiligung von Franziska Kemter und Daniel Schindler.

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2015, 10, 302–314

DOI 10.1002/biot.201400031

Research Article

Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*

Sonja J. Messerschmidt, Franziska S. Kemter, Daniel Schindler and Torsten Waldminghaus

LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-Universität Marburg, Germany

Recent developments in DNA-assembly methods make the synthesis of synthetic chromosomes a reachable goal. However, the redesign of primary chromosomes bears high risks and still requires enormous resources. An alternative approach is the addition of synthetic chromosomes to the cell. The natural secondary chromosome of *Vibrio cholerae* could potentially serve as template for a synthetic secondary chromosome in *Escherichia coli*. To test this assumption we constructed a replicon named synVicII based on the replication module of *V. cholerae* chromosome II (*oriII*). A new assay for the assessment of replicon stability was developed based on flow-cytometric analysis of unstable GFP variants. Application of this assay to cells carrying synVicII revealed an improved stability compared to a secondary replicon based on *E. coli oriC*. Cell cycle analysis and determination of cellular copy numbers of synVicII indicate that replication timing of the synthetic replicon in *E. coli* is comparable to the natural chromosome II (ChrII) in *V. cholerae*. The presented synthetic biology work provides the basis to use secondary chromosomes in *E. coli* to answer basic research questions as well as for several biotechnological applications.

Received	06 MAY 2014
Revised	02 OCT 2014
Accepted	30 OCT 2014
Accepted article online	31 OCT 2014

Supporting information available online

Keywords: DNA replication · Replicon copy number · Replicon stability · Synthetic biology · Synthetic genomics

1 Introduction

Escherichia coli is a well-studied model organism and frequently used as production strain in biotechnology. In recent years the organism served, e.g. for the production of biofuels, protein therapeutics, and organic acids [1–3]. To obtain an efficient production strain, the bacterium is often genetically manipulated either with the help of plasmids or insertions into the primary chromosome. New synthetic biology methodologies allow a new quality in genetic engineering with the construction of entire synthetic replicons becoming a feasible goal.

Correspondence: Prof. Torsten Waldminghaus, LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-Universität Marburg, Hans-Meerwein-Str. 6, Mehrzweckgebäude, D-35043 Marburg, Germany
E-mail: torsten.waldminghaus@synmikro.uni-marburg.de

Abbreviations: CAA, casamino acids; **ChrI,** primary chromosome of *Vibrio cholerae*; **ChrII,** secondary chromosome of *Vibrio cholerae*; ***oriI*,** origin of replication of *Vibrio cholerae* chromosome I; ***oriII*,** origin of replication of *Vibrio cholerae* chromosome II

In the past years the successful assembly of synthetic chromosomes has been demonstrated. Venter and coworkers synthesized the complete 583 kb *Mycoplasma genitalium* chromosome [4]. Two years later a transplantation of a synthetic 1.08 Mbp *Mycoplasma mycoides* chromosome into a *M. capricolum* cell was accomplished [5]. Recently, Annaluru et al. set a milestone with the synthesis of the first eukaryotic chromosome of *Saccharomyces cerevisiae* (synIII) [6]. The native chromosome III, one of the smallest yeast chromosomes (273 kb), was replaced by the synthetic replicon in a stepwise fashion. An alternative approach to these re-constructions of primary chromosomes would be the addition of a synthetic secondary chromosome. Such a replicon could be gradually build up from a suitable backbone. Meeting the synthetic biology philosophy such a backbone should allow simple and standardized manipulation. Naturally occurring secondary chromosomes could serve as the respective template.

Properties of bacterial secondary chromosomes are that they represent an extra replicon within the cell that contains essential genes needed for growth and that they are smaller in comparison to primary chromosomes. Like primary chromosomes their replication starts at a specif-

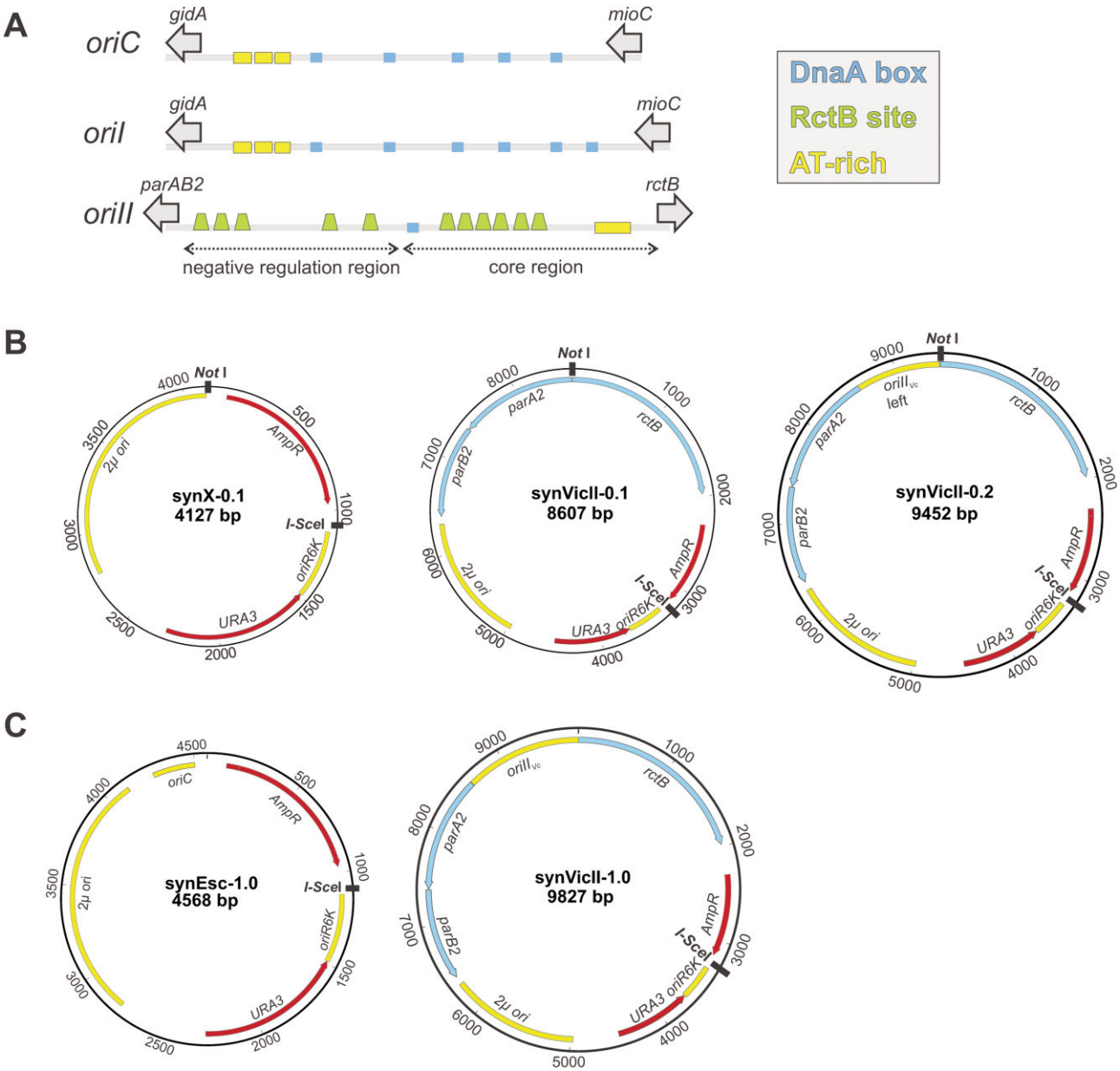


Figure 1. (**A**) Scheme of replication origins from *E. coli* and *V. cholerae*. Colored symbols present relevant sequence motifs as indicated. (**B**, **C**) Design of synthetic secondary chromosomes constructed in this study. Genes are indicated by colored arrows originating from *V. cholerae* (blue) or selection markers (red). Replication origins are shown in yellow. (**A**) Backbones of synthetic chromosomes contain a *Not*I site to insert chromosomal replication origins and an *I-Sce*I site for integration of larger DNA fragments. (**B**) Secondary synthetic chromosomes based on *oriC* (synEsc-1.0) and *oriII* (synVicII-1.0). Replicon maps are not drawn to scale.

ic point during the cell cycle. The genome of most bacteria is encoded on a single (primary) chromosome. However, secondary chromosomes were found in representatives of proteobacteria such as *Rhodobacter sphaeroides*, *Agrobacterium tumefaciens*, *Ralstonia eutropha*, and all members of the *Vibrionaceae* [7–10]. Moreover, secondary chromosomes were also discovered in representatives of other bacterial phylla, e.g., *Leptospira interrogans* and *Deinococcus radiodurans* [11, 12].

One of the best studied secondary chromosomes is found in *Vibrio cholerae*. This human pathogen is a model organism among the rare multi-chromosome bacteria. *E. coli* and *V. cholerae* are closely related members of the γ-Proteobacteria. The genome of *V. cholerae* is divided into two different sized chromosomes (primary chromosome of *Vibrio cholerae* (ChrI) 2.96 Mbp and secondary chromosome of *Vibrio cholerae* (ChrII) 1.07 Mbp) [13]. ChrII is regarded a secondary chromosome since it

encodes essential genes and replication takes place at a specific point during the cell cycle [13–15]. The origin of replication of *Vibrio cholerae* chromosome I (*oriI*) shows similarities to *oriC* in *E. coli* and both origins are initiated by DnaA (Fig. 1A) [16, 17]. In contrast, the complex origin of replication of *Vibrio cholerae* chromosome II (*oriII*) differs from *oriI* and *oriC* [17]. The protein RctB binds to an array of so-called iterons in the core origin region to initiate replication (Fig. 1A) [16, 18]. Additional RctB binding sites in the left part of the *oriII* region are involved in negative regulation of initiation [19–21]. ChrII possesses its own segregation system with the two proteins ParA2 and ParB2, which are homologous to plasmid partitioning proteins [22]. Beside its role in segregation, ParB2 has been shown to contribute to the regulation of replication initiation [23]. Interestingly, the genes *rctB*, *parA2*, and *parB2* are located to the left and right of *oriII* forming a compact replication-segregation module. Fragments of different length including the minimal *oriII* region have been shown to promote replication of small replicons in the heterologous host *E. coli* [17, 18, 24–26].

Our current study describes the design and construction of a synthetic, secondary chromosome based on *oriII* of *V. cholerae* and an initial characterization of this replicon in *E. coli*. The new secondary chromosome, termed synVicII, is designed to serve as scaffold for the assembly of larger replicons. We believe that *V. cholerae oriII* has great potential for this purpose because nature has selected this replication origin for stable maintenance of a mega-base sized replicon. In biotechnology, the use of gene synthesis has increased enormously over the last years. The constant decrease of costs per synthesized bp suggests that ordering DNA fragments encoding whole pathways or even entire replicons will be common in the future. The work presented here is considered a first step to explore the potential of *V. cholerae oriII* based replicons as backbone to integrate new genetic features to the cell and their carefully monitored maintenance within the host.

2 Material and methods

2.1 Bacterial strains, plasmids, oligonucleotides, and culture conditions

All used strains and plasmids are listed in Supporting information, Table S1, Oligonucleotides are listed in Supporting information, Table S2. Pre-cultures of *E. coli* and *V. cholerae* were grown in 3 mL LB medium overnight at 37°C. For the determination of replicon copy numbers the reference strains were grown in AB medium [27] supplemented with 10 μg/mL thiamine, 25 μg/mL uridine, and either 0.2% glucose alone or a combination of 0.2% glucose and 0.5% casamino acids (CAA). Antibiotic selection was used at the following concentrations: ampicillin 100 μg/mL, chloramphenicol 30 μg/mL, and kanamycin 35 μg/mL. For *gfp* induction, 0.2 mM IPTG was used. Pre-cultures of *S. cerevisiae* were grown in 2× YPAD at 30°C. Frozen competent cells of *S. cerevisiae* were obtained and transformed as described [28]. The *URA-3* gene that complements mutations found in *S. cerevisiae* PJ69-4a was used as selection marker by plating on SD–URA.

2.2 Construction of secondary replicons and strains

To construct the replicon synVicII-0.1, *rctB* and *parAB2* were PCR amplified from genomic DNA of *V. cholerae* N16961 with the primers 1/2 and 9/10, respectively (Supporting information, Table S2). *bla* (AmpR) was amplified from pNEB193 with primers 3/4, *ori*R6K from pPS11 with primers 5/6 and 2 μ *ori*/*URA-3* from pRSII426 with the primers 7/8. All primers used for assembly in yeast contain overlapping sequences (at least 26 bp) to the neighboring fragment. They also add a half 3′ *Sma*I site to facilitate release after sub-cloning. The amplified fragments were sub-cloned into *Sma*I-cut pUC57kan and verified by sequencing. The fragments were assembled by homologous recombination in *S. cerevisiae* pJ69-4a according to a method described [29], followed by transformation as described [28]. The resulting synVicII-0.1 was isolated and transformed into *E. coli* DH5α λ *pir*. Obtained clones indicated replication of synVicII-0.1 based on *ori*R6K. The sequence was verified by sequencing. synX-0.1 was constructed by amplifying synVicII-0.1 with the primers 117/118. *Not*I-digestion and ligation resulted in circular synX-0.1, which was transformed into *E. coli* DH5α λ *pir*. For the construction of synVicII-0.2, the negative regulation region of *oriII* (Fig. 1A) was amplified with the primers 14/16 from genomic DNA of *V. cholerae* N16961. This DNA served as a template for a second PCR with the primers 11/14. The amplified *oriII* was sub-cloned into pUC57kan and verified by sequencing. *Sma*I-released *oriII* and *Not*I-digested synVicII-0.1 were assembled to synVicII-0.2 by homologous recombination in yeast. To construct synEsc-1.0, *oriC* (which we define as the intergenic region between *gidA* incl. 41 bp and *mioC* incl. 30 bp), was PCR amplified from genomic DNA of *E. coli* MG1655 with oligonucleotides 124/125 and inserted into the *Not*I-digested synX-0.1 by homologous recombination in yeast. For construction of synVicII-1.0 *oriII* was PCR amplified from genomic DNA of *V. cholerae* N16961 with oligonucleotides 14/15 and cloned into the *Not*I-digested synVicII-0.1 by homologous recombination in yeast. *gfp*-LAA, -LVA, -AAV, -ASV, and *gfp*mut3* with the synthetic ptac promoter named A1/04/03 promoter [30] were PCR amplified from pJBA110, 111, 112, 113, and 27, respectively, with primers 26/27. The *gfp* variants were inserted into *I-Sce*I digested synVicII-1.0 by Gibson Assembly [31] resulting in the series synVicII-1.1- synVicII-1.5. synEsc-1.3 was constructed accordingly with insertion of *gfp*mut3* into the *I-Sce*I site of synEsc-1.0. To

construct synVicII-1.6, *thyA* was PCR amplified with the primers 126/127 from pAMD135. Then, *thyA* was inserted into the *I-Sce*I site of synVicII-1.3 by homologous recombination in *S. cerevisiae*. To construct synX-0.3, *gfp*-AAV was inserted into synX-0.1 as described above for synVicII-1.1-1.5 by homologous recombination in yeast. The synF-plasmid replicon was constructed by amplification of the F replication *origin* and *sopABC* by PCR in two fragments from pBeloBAC11 (NEB) with the primers 608–611. The two fragments were integrated into *Not*I digested synX-0.3 by homologous recombination in yeast and confirmed by sequencing. To construct strain FSK18 we first cloned sequences corresponding to the three microarray probes for synVicII into pBR-FRT-kan-FRT, giving pMA411. The microarray cassette including the KanR marker was PCR amplified with primers 354/355 containing chromosomal homology ends. λ red recombination was used for chromosomal integration into *tnaA* in *E. coli* TB13 [32]. The successful integration was confirmed by sequencing. P1 transduction was used to transfer the microarray cassette into *E. coli* MG1655 resulting in strain FSK18.

2.3 Flow cytometry measurements

The different experiments were carried out as described in the figure legends. Unless described otherwise, after harvesting, the cells were washed two times in TBS (0.1 M Tris-HCl pH 7.5, 0.75 M NaCl). For the quantification of $GFP^{+/-}$ cells, the cells were fixed for 10 min in 1% paraformaldehyde-PBS pH 7.2 and collected by centrifugation for 4 min at 15 000 × *g*. Following one wash step with TBS, samples were diluted in 500 μL of TBS and stained with 500 μL of Hoechst 33258 (9 μg/mL TBS) 30 min before flow cytometry analysis in a BD LSRII Flow Cytometer. The GFP fluorescence was measured through a 530/30 nm bandpass filter. The software FlowJo (Treestar, Ashland, USA) was used to analyze the GFP fluorescence. Hoechst positive cells were gated and the respective population divided into $GFP^{+/-}$ cells by gating. Respective gates were kept unchanged for each experimental series.

For the cell cycle analysis, 1 mL of exponential LB grown cells (OD_{600} = 0.15) was treated with rifampicin and cephalexin as described [33]. The samples were centrifuged at 15 000 × *g* 4°C for 4 min, washed 2× with 1 mL TBS, resuspended in 100 μL TBS and mixed with 1 mL 77% ethanol. They were stored for at least 20 min at 4°C. Samples were washed again with 1 mL TBS, resuspended in 500 μL TBS, and stained with 500 μL Hoechst-solution (9 μg/mL in TBS).

E. coli MG1655 cells grown in AB Glucose were fixed in ethanol, washed in 0.1 M phosphate buffer (PB) pH 9 and stained with FITC solution (3 μg/mL in PB) overnight at 4°C. Afterwards the cells were washed with TBS and stained with Hoechst as described above. These cells served as an internal standard for calibration of the flow cytometer measurement (see Supporting information, Fig. S1). Flow cytometry measurements were carried out as described [33].

2.4 Plate count assays

For monitoring plasmid loss, strains SMS18 and SMS62 were cultivated in LB with ampicillin at 37°C to an OD_{600} of around 0.15. After taken 100 μL sample (T_0), the cultures were transferred to pre-warmed LB without antibiotic and cultivated for 6 h at 37°C. To keep the cultures in log growth they were diluted 1:10 into fresh pre-warmed non-selective media at least three times. At 3 and 6 h after transfer, 100 μL samples were taken and three suitable dilutions were plated onto non-selective plates. After over night incubation, obtained colonies were re-streaked parallel onto ampicillin and non-selective LB agar plates. The fraction of plasmid-containing cells in each population was determined by re-streaking a total of at least 250 colonies from three biological replicates onto selective and non-selective LB agar plates.

2.5 Quantitative PCR (qPCR) and determination of *oriC*/*ter* and synVicII/*oriC* ratio

In exponential growth phase, 1 mL culture of the *E. coli* strain SMS28 grown in LB was harvested by centrifugation at 15 000 × *g* for 4 min. The cells were stored at −20°C. After thawing on ice, the sample was re-suspended in 1 mL water and incubated at 95°C for 10 min. According to the formula $OD_{600}\,1 = 8 \times 10^5$ cells/μL, the samples were diluted to 1.25×10^4 cells/μL.

The *E. coli* strains SMS28 and FSK9 were used as reference cells. SMS28 was cultivated in AB Glucose CAA until stationary phase and was then diluted to 6.25×10^5 cells/μL. FSK9 was grown in AB Glucose CAA to an OD_{450} of around 0.15 and then diluted to 1.25×10^3 cells/μL.

Each reaction was carried out in triplicates of 10 μL with average SDs of 22%. Primer sets for *oriC* (3921366fw, 3921366rv, 3921366pr), *ter* (ter-fw, ter-rv, ter-p) and synVicII (AmpR fv, AmpR rv, AmpR Sonde or ori2fw, ori2rv, ori2probe) were used in separate reactions. Three biological replicates were analyzed three times each. *OriC*/*ter* ratios were calculated relative to stationary phase cells of SMS28 (see above), which should have an *oriC*/*ter* ratio of one. Ratios of synVicII to *oriC* were calculated relative to plasmid pMA407 containing the genomic regions that are template for the *oriC* and synVicII primer sets (see above). To measure the reference in whole cell extracts we used strain FSK9, which lacks the genomic region for the *oriC* primer binding (in the *gidB* gene) but carries plasmid pMA407. qPCR reactions are composed of 5 μL KAPA Probe Fast QPCR mastermix Universal 2× (peqlab, Erlangen, Germany), 1 μL primer mix and 4 μL cell suspension. The 250 μL primer mix was prepared for each set of primer

and contains 22.5 μL primer fw (100 pmol/μL), 22.5 μL primer rv (100 pmol/μL), 6.25 μL probe (5′-Fam/3′-Tamra, 100 pmol/μL), 50 μL Rox Reference Dye Low 50× (peqlab) and 148.75 μL water. qPCR reactions were performed in the real-time cycler qTower (Analytik Jena AG, Jena, Germany) with the following program: 1, 95°C for 3 min; 2, 95°C for 3 s; 3, 55°C for 20 s; and 4, fluorescence read. Steps 2–4 were repeated 45 times. The determination of the CT-values was carried out with the software qPCRsoft (Analytik Jena AG, Jena, Germany) without using the rox reference.

2.6 Cell cycle analysis

The generation time τ was determined by plotting the time versus log values of OD_{600} (considering only OD_{600} values between 0.04 and 2.0). The slope of the respective regression line represents the generation time of the culture. The *oriC*/*ter* ratio was determined via qPCR as described above. The C-period, the time of one complete replication round of the chromosome, was calculated by the formula $C = \log_2\left(\frac{oriC}{ter}\right) \times \tau$. The initiation age a_i, the time between cell division and initiation of replication, was determined by the formula $a_i = \tau - \log(2 - F) \times \tau$. F is the percentage of cells, which have not initiated yet as obtained by flow cytometry analysis described above (in our case: all cells with eight chromosomes). The D-period is given by the formula $D = 4 \times \tau - C - a_i$. To determine the time point of synVicII replication the logarithmic interpolation of the copy number distribution of primary chromosome loci was calculated for *ter* = 1 and *oriC* = 3.2 (measured by qPCR) in a distance of 2 319 838 bp from *ter*. The respective formula was used to calculate ratios of locus copy numbers relative to *ter*. The synVicII/*oriC* ratio of 0.54 corresponds to a synVicII/*ter* ratio of about 1.7 (0.54 × 3.2). According the log interpolation this corresponds to a genomic position 1 240 000 bp distant from *oriC* indicating that synVicII replication happens after 53% of the primary chromosome is replicated.

2.7 Microarray construction and comparative genomic hybridization

The custom microarray probes (50 bp) were designed to cover the whole *E. coli* chromosome with a spacing of around 45 kB. In addition, three probes match regions on the synthetic secondary chromosomes synVicII. Probe sequences are provided in Supporting information, Table S2. The DNA probes contain a 5′ C6-amino linker and were printed using a BioRobotics MicroGrid microarray printer as described [34]. Each probe is printed eight times onto the array. To isolate the DNA, exponential grown cells from SMS18 were harvested at an OD_{600} of 0.15 and FSK18 in stationary phase (reference strain). Cells were incubated in 300 μg/mL lysozyme, 50 mM Tris-HCl, and 50 mM EDTA pH 8 for 15 min on ice followed by incubation at RT in 1% SDS. DNA was isolated with phenol/chloroform and an ethanol precipitation. It was treated with 2 μL RNaseA (10 mg/mL) for 2 h at 37°C and purified again with phenol/chloroform and ethanol precipitation. DNA was sonicated in 35 μL H_2O for five cycles of 30 s with 30 s breaks at 4°C with a Bioruptor (Diagenode, Denville, NJ, USA) to obtain fragments of about 300–1000 bp. Processing of microarrays prior to hybridization was carried out as described [34]. 150 or 300 ng DNA were labeled as described [35]. 50 μL sample and reference were mixed and denatured for 3 min at 94°C. DNA was mixed with hybridization solution (0.96× MES, 0.96 M NaCl, 0.192% formamide, 20 mM EDTA, 0.96% Triton X-100) and applied to the microarray using Agilent hybridization chambers. Hybridization took place overnight at 55°C and 5 rpm. After hybridization arrays were washed for 5 min each with 6× SSPE and 0.005% *N*-Lauroylsarcosin and 0.06× SSPE and 0.0018% PEG 200, respectively. Slides were dried and scanned with a Genepix 4000B and the GenePix Pro 6.0 software (Molecular Devises, Silicon Valley, USA). Spot detection, image segmentation and signal quantification was carried out with ImaGene 8.0 software (Biodiscovery Inc., Los Angeles, USA). For data processing the average of the Cy3- and Cy5-intensities for each probe were calculated and the ratios of sample (Cy3) to control (Cy5) determined after background (spotting buffer) subtraction. Values were printed relative to their genomic position. Exponential curves were fitted to the left and right chromosomal arm separately. The respective formula was used to normalize values to a value of 1 for the terminus of the left chromosome arm and to calculate *oriC/ter* ratios. For synVicII the average for all three probes was used to calculate its copy number relative to the primary chromosome. The *oriII*/*oriC* ratio was calculated for each arm of the genome plot. *OriC/ter* ratios and the synVicII/*oriC* ratios were used for calculation of the time point at which synVicII replication takes place in relation to replication of the primary chromosome as described in the previous section.

3 Results

3.1 Design and construction of secondary chromosomes

The first step toward synthetic secondary chromosomes was the construction of three backbones (Fig. 1B). For the design, specific modules were selected. The first one comprises the yeast origin 2 μ *ori* and yeast marker *URA-3* gene. They are needed for the efficient assembly of synthetic chromosomes via homologous recombination in *S. cerevisiae*. The method provides advantages over the conventional cloning in *E. coli*. First, it allows

scar-free cloning, omitting extra sites for restriction enzymes in the final construct. Second, yeast homologous recombination allows the accurate parallel assembly of several large fragments, which is important for future size extensions of synthetic chromosomes [5]. The second element is the R6K origin that allows replication in *E. coli* strains, which harbor the *pir* gene encoding the respective replication protein π. This conditional replication is important to be able to clone replicons lacking other functional *E. coli* replication origins. However, in a *pir* minus background the respective replicon will only be established if a functional origin beside *ori*R6K is included, allowing the testing of sequences for their ability to function as autonomously replicating sequence. The *bla* gene (AmpR) conferring ampicillin resistance served as selective marker in *E. coli*. The backbones are designed to be extended in two respects. First, *Not*I digestion of the backbones enables the integration of different origins of choice (Fig. 1A). The respective 8 nt recognition site should occur about every 65 kbp only, and serve by decreasing the probability of the inserted origin containing a respective site. The second integration site is *I-Sce*I, which can be used to integrate large fragments. *I-Sce*I possesses a recognition site of 18 nt and should thus occur only about once in 70 Gbp by chance, thereby reducing the chances of a cut in the fragments to be inserted. An important concept of synthetic biology is the refactoring of biological modules, allowing the specific redesigning of relevant parts without disturbing other parts. To facilitate refactoring of *V. cholerae oriII* we constructed two versions of the synthetic chromosome backbones. One lacks the native sequence of the whole *oriII* from *parA2* to *rctB* (synVicII-0.1, Fig. 1B) and the other one lacks only the core region of *oriII* (synVicII-0.2, Fig. 1A and B). For the addition of an extra chromosome to the native chromosome in *E. coli*, one solution could be the use of the *E. coli oriC* as additional copy. Another possibility would be to use the *oriII* of *V. cholerae*. Thus, we integrated *oriC* which we defined as the intergenic region between the *gidA* and *mioC*, into the *Not*I-digested backbone synX-0.1 resulting in synEsc-1.0. Similarly, the construction of synVicII-1.0 was achieved by integrating *oriII*, the intergenic region between *parA2* and *rctB*, into *Not*I-digested synVicII-0.1. Transformation of these synthetic chromosomes into *E. coli* MG1655 successfully produced clones, which demonstrated their ability to replicate, consistent with previous studies [17, 36]. However, the sole property of replication in *E. coli* is not sufficient for the use of such synthetic chromosomes in biotechnology. Further characterization of these replicons is obligatory and we therefore compare the respective performance of *E. coli oriC* versus *V. cholerae oriII* in the following.

3.2 A flow-cytometry-based assay to assess secondary chromosome stability

One important property of chromosomes is that they are stably maintained in the cell population. To characterize the synthetic chromosomes regarding stability we developed a new flow-cytometry assay based on the measurement of a fluorescence reporter. For this purpose *gfp* was integrated into the *I-Sce*I site of synVicII-1.0 (Fig. 2A). The assay is based on the following principle: If the replicon is stable in the cell, the cell shows GFP fluorescence. If the replicon is not stable in the cell, it will be lost during cell division and the cells will show no fluorescence. The respective strain is cultivated without selection and the relative amount of GFP positive cells over time is a measure of stability. Before applying the assay, we tested if we are able to quantify GFP fluorescence in fixed *E. coli* cells. Strains NZ23 (GFP^-) and SMS20 (GFP^+) were grown in LB medium supplemented with IPTG to induce *gfp* expression. During exponential growth phase, both cultures were harvested and fixed using 1% paraformaldehyde. Cultures were then mixed in different ratios of $GFP^{+/-}$ cells. Staining the cells with Hoechst allowed the discrimination between cells and debris in the following quantification of $GFP^{+/-}$ cells in the flow cytometer. For three independent cultures the measured ratios of $GFP^{+/-}$ cells showed excellent correlations to expected ratios (Fig. 2B). Flow-cytometry histograms demonstrate the changing ratios of $GFP^{+/-}$ cells (Fig. 2C). The results show that our setup allows precise quantification of $GFP^{+/-}$ cells by flow cytometry.

Another requirement for the flow-cytometry-based assay is that the GFP fluorescence should be a sensitive indicator for the replicon in the cell. However, GFP is known to be a stable protein [37]. The cells could thus still be fluorescent even after loss of *gfp* encoded on the secondary chromosome. To avoid this, we integrated genes encoding unstable GFP variants as developed by Andersen and colleagues [30] into synVicII-1.0. The GFP variants have different C-tail *ssrA*-degradation tags to vary degradation by specific proteases. Degradation kinetics of the GFP variants have only been measured based on hole populations and not at single-cell level as needed for our assay. We therefore treated exponentially growing cells with chloramphenicol to stop protein synthesis. Samples were harvested at various time points and ratios of $GFP^{+/-}$ cells determined by flow cytometry. As expected, a high percentage of cells harboring the original GFP (synVicII-1.5) showed fluorescence, which remained constant for 30 min while the negative control (synVicII-1.0) showed no fluorescence (Supporting information, Fig. S2). Variants GFP-LAA and GFP-ASV are slowly degraded (Supporting information, Fig. S2). In contrast, fluorescence measurements of GFP-LVA and GFP-AAV indicate fast degradation. In general the data agree with stabilities determined on the population level [30]. For the replicon

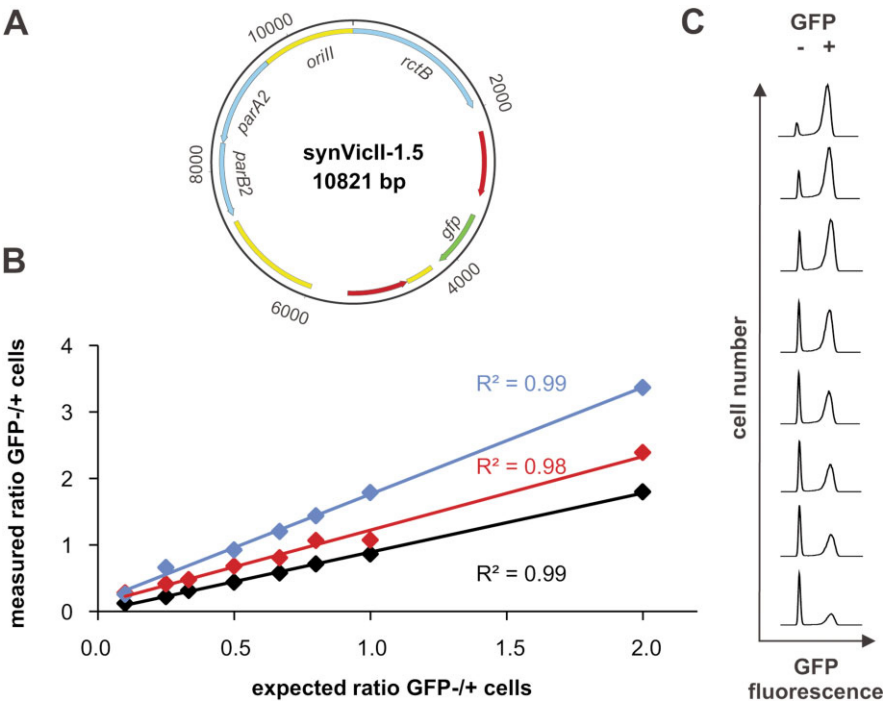


Figure 2. Quantification of GFP$^{+/-}$ cells by flow cytometry. (**A**) The gene encoding GFPmut3* was integrated into the *I-Sce*I site of synVicII-1.0 (compare to Fig. 1C). (**B**) Quantification of GFP$^{+/-}$ in defined mixtures. *E. coli* strains NZ23 (GFP$^-$) and SMS20 (GFP$^+$) were grown in 50 mL LB medium at 37°C. 0.2 mM IPTG was added at an OD_{600} of around 0.15. After 1 h ($OD_{600} \approx 0.6$) 10 mL cells were prepared for flow cytometry analysis as described in the method section. Cells were diluted in 5 mL TBS and GFP$^-$ and GFP$^+$ cells were mixed in different ratios as given on the x-axis. Samples were stained with 5 mL of Hoechst 33258 (9 µg/mL TBS) and analyzed by flow cytometry. Measured ratios of GFP$^{+/-}$ cells were plotted against expected ratios for three independent experiments (colored diamonds and regression lines). (**C**) Exemplary flow cytometry histograms for data in B (black line).

stability assay a GFP, which is rapidly degraded, is required. Since the GFP-AAV variant is rapidly degraded and is initially measured in 85% of the cells (the GFP-LVA variant was only detected in 78% of the cells) we chose the GFP-AAV variant for the replicon stability assay.

To test the assay we designed and constructed an inducible replicon loss system. It is based on the finding that trimethoprim inhibits an enzyme which recycles an essential cofactor depleted by ThyA [38]. Thus, trimethoprim treatment leads to a lack of the cofactor, which is toxic for the cells. Under conditions where thymine is provided in the medium the cell does not need ThyA. Consequently, the cell could eliminate the *thyA* gene. This elimination is supported if the counter-selection agent trimethoprim is added. To be able to simulate replicon loss we inserted *thyA* into synVicII-1.3 and transformed it into a strain with a deletion of the chromosomal *thyA* copy. The strain cultivated in medium supplemented with thymine was treated with trimethoprim to allow growth of replicon-free cells only. As expected, the percentage of GFP$^+$ cells remained constantly high for the control without trimethoprim treatment for the analyzed time period, indicating stable maintenance of the replicon (Fig. 3A). In contrast, the number of GFP$^+$ cells decreased dramatically 2.6 h after trimethoprim addition.

Therefore, we conclude that the new assay based on GFP could be used as a suitable measure of replicon stability.

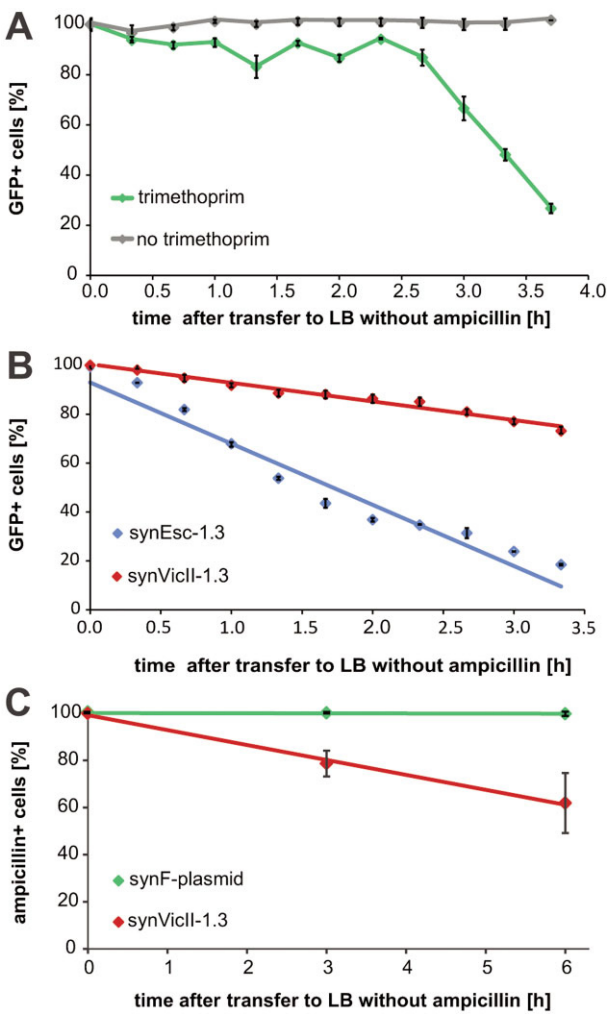
3.3 Stability of the *oriII*-based replicon synVicII-1.3

The assay was then applied to the two synthetic secondary chromosomes synEsc-1.3 and synVicII-1.3 to measure their stability. The respective *E. coli* strains SMS18 and SMS28 were first cultivated in LB medium with ampicillin. In early exponential growth phase the selection pressure was eliminated by transferring the cells into medium without ampicillin. The ratio of GFP$^{+/-}$ cells was measured temporally by flow cytometry. After 3.3 h, synEsc-1.3 was only present in around 12% of the cells (Fig. 3B). In contrast, synVicII-1.3 was maintained in 75% of the cells at the same time point. The slopes of regression lines indicate a replicon loss rate of 8% per hour for synVicII-1.3, which is considerably smaller than the value of 25% for synEsc-1.3.

As a second control, we compared the stability of synVicII-1.3 to synF-plasmid. This replicon carries the replication origin of the F plasmid and its own segregation system consisting of *sopABC*. We also used this comparison to further evaluate our flow cytometry-based assay by assessing replicon stability by the conventional plating method (Fig. 3C). Indeed, the value of 6% replicon loss rate per hour for synVicII correlates well with the 8% determined with the flow cytometry assay above (Fig. 3C). Consistent with previous findings, the F plasmid replication and segregation system results in very high stability with 100% of the cells containing synF-plasmid after the time measured here [39]. In summary, our data show that the stability of synVicII is in-between the replicons based on the *E. coli* chromosomal origin and the F plasmid investigated here.

3.4 Copy number analysis suggests that synVicII replicates similar to chromosome II in *V. cholerae*

Beside stability, the copy number is an important characteristic of replicons. While many plasmids have copy numbers higher than the primary chromosome a secondary chromosome has a copy number similar to the primary chromosome. In rich medium, *E. coli* is capable of overlapping replication cycles [40, 41]. To analyze the details of DNA replication we performed cell cycle analysis of the *E. coli* strain SMS18, carrying synVicII-1.3. To this end we determined primary chromosome copy numbers in cells after blocking replication initiation by rifampicin and cephalexin treatment. In agreement with previous studies, we found that replication is started at eight origins of the primary chromosome under the respective growth conditions (Fig. 4) [42, 43]. The detection of one 8- and one 16-chromosome peak and no intermediates indicate regular control of the primary chromosome replication and thus a lack of perturbation by synVicII (Fig. 4B).



As a critical parameter for cell cycle analysis the *oriC*/*ter* ratio was determined by qPCR to be 3.2 (±0.7). The replication pattern should subsequently appear as shown in Fig. 4 (for details of cell cycle analysis see Section 2). To estimate the copy number of synVicII we determined the *oriII*/*oriC* ratio. This ratio should be 1 if the synthetic chromosome is replicated at the same time point as the primary chromosome and 0.31 if synVicII is replicated when the primary chromosome terminates (considering the above determined *oriC*/*ter* ratio of 3.2). However, the *oriII*/*oriC* ratio was found to be 0.54 (±0.06) indicating that synVicII initiates later than *oriC*, but before replication of the primary chromosome terminates. It is important to note that the method details are critical for proper analysis of secondary replicon copy numbers. For two DNA-isolation methods with or without initial DNA fragmentation it was not possible to detect synVicII by qPCR (data not shown). Our data were subsequently generated by qPCR based on whole cell lysates (see Section 2 for details). Using a logarithmic interpolation of the copy number distribution of primary chromosome loci we determined the potential time point of synVicII replication (Fig. 4, see Section 2 for details). The copy number could result from replication of synVicII starting after 53% of the primary chromosome is replicated assumed that replication of synVicII takes place at a defined time of the cell cycle. This time point is similar to the replication start of the ChrII replication in *V. cholerae*, which initiates replication

Figure 3. Stability of synthetic secondary chromosomes in *E. coli* MG1655. (**A**) Proof of principle for the stability assay with the inducible replicon-loss system based on the trimethoprim counterselection (see text for details). Strain SMS32 was grown in LB supplemented with ampicillin, 0.2 mM IPTG and 200 µg/mL thymine at 37°C to an OD_{600} of about 0.15. The cells were centrifuged at 3000 × *g* for 3 min at RT and transferred to pre-warmed LB supplemented with 200 µg/mL thymine and 0.2 mM IPTG but no ampicillin. Whereas one culture was additionally treated with 20 µg/mL trimethoprim (diluted in DMSO) to induce the loss of synVicII-1.4, the other culture was treated with the corresponding amount of DMSO. Samples were taken every 20 min and analyzed by flow cytometry as described in the method section. The first value was set to 100% and the other values related to this value. (**B**) Stability of synEsc-1.3 (blue) and synVicII-1.3 (red) in *E. coli* MG1655 determined by flow cytometry. Strains SMS18 and SMS28 were grown in LB supplemented with ampicillin and 0.2 mM IPTG at 37°C to an OD_{600} of about 0.15. Cells were transferred to ampicillin free medium as above. Cells were cultivated to an OD_{600} of around 1.5, diluted 1:10 into fresh pre-warmed medium to allow continued exponential growth. Samples were taken every 20 min and analyzed by flow cytometry. (**C**) Stability of synVicII-1.3 (red) compared to synF-plasmid (green) in *E. coli* MG1655 measured by the plate count method. Strains SMS18 and SMS62 were cultivated in LB with ampicillin at 37°C to an OD_{600} of around 0.15 and transferred to pre-warmed LB without antibiotic. Cells were cultivated for 6 h at 37°C. Three and six hours after the transfer cells were plated on non-selective plates. After overnight cultivation colonies were re-streaked on ampicillin and non-selective LB agar plates. Values are given as percentage of ampicillin resistant cells. Results are obtained from three independent experiments including at least 250 colonies per time point and strain.

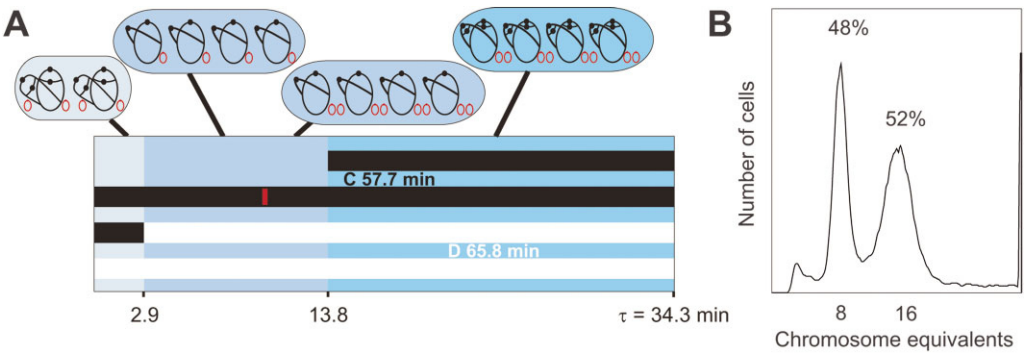


Figure 4. Cell cycle analysis of the *E. coli* MG1655 strain SMS18 carrying synVicII-1.3. (**A**) Scheme demonstrating the replication pattern of cells in different intervals of the cell cycle. Blue rods represent cells with chromosomes (black lines), origins (black dots), and secondary chromosomes (red lines) in four stages of the cell cycle. One generation time, found to be 34.3 min, is shown by the width of the rectangle with the different shades of blue representing the periods from cell birth to termination (0–2.9 min), from termination to initiation (2.9–13.8 min) and from initiation to cell division (13.8–34.3 min). The black line represents the C period (time for one complete round of replication) and the white line the D period (time from termination to cell division). The *oriC*/*ter* ratio of 3.2 (±0.7) was determined by qPCR. Initiation of the synVicII replication is indicated by the red bar. Based on the *oriII*/*oriC* ratio of 0.54 (± 0.06) this happens 30.6 min after initiation of the primary chromosome. (**B**) Histogram of one exemplary rifampicin–cephalexin experiment based on flow cytometry analysis for the calculation of cell cycle parameters. The fraction of cells containing 8 chromosomes was 48% and the fraction of cells containing 16 chromosomes was 52% (for determination of 8 and 16 chromosome peaks see Supporting information, Fig. S1). The average of three biological replicates was 49 and 51%, respectively.

after about 54–66% of ChrI is replicated according to a previous study [33].

To verify this finding we performed comparative genomic hybridization (CGH) using a custom microarray (see Section 2 for details). Probes on the microarray match the *E. coli* chromosome as well as synVicII. A common hybridization control for bacterial CGH is DNA from stationary phase cultures where chromosomes are not replicating, resulting in a relative copy number of 1 for all genetic loci. The problem in our case is that nothing is known about the copy number of synVicII in stationary phase. To generate a suitable control DNA we integrated a stretch of DNA matching the three probes for synVicII into the *E. coli* chromosome (giving strain FSK18; see Section 2). Stationary phase cells of the respective strain should have a relative copy number of 1 for the synVicII probes compared to the *E. coli* chromosome. As described above the DNA isolation is critical. Since we could not apply the whole cell approach used for the qPCR, we used a protocol shown to be suitable before [44]. The isolated DNA was first tested by qPCR and gave similar results to the whole cell approach showing that the DNA isolation procedure could be used (data not shown). DNA from strain SMS18 growing exponentially in LB medium was hybridized against the hybridization control. Respective fluorescence ratios were plotted relative to the respective chromosomal position (Fig. 5). The relative abundance of chromosomal loci diminishes exponentially with increasing distance from the origin for exponentially growing populations [45]. Exponential curves were therefore fitted to each chromosome arm separately. The finding that the two curves meet at *oriC* indicates good data quality (Fig. 5). We carried out three biological replicates one of which is shown in Fig. 5. Fitted curves were used to calculate average *oriC/ter* ratios of 1.8 ± 0.1. This is considerably less than measured by qPCR, a phenomenon observed before [46]. The copy number ratio of synVicII to *oriC* was 0.69 ± 0.04. These ratios fit to synVicII replication at a constant time point of the cell cycle after 67% ± 15 of the *E. coli* primary chromosome is replicated. This value correlates well with the qPCR results above and could indicate that synVicII-replication timing is similar to that of ChrII in *V. cholerae* relative to the respective primary chromosome.

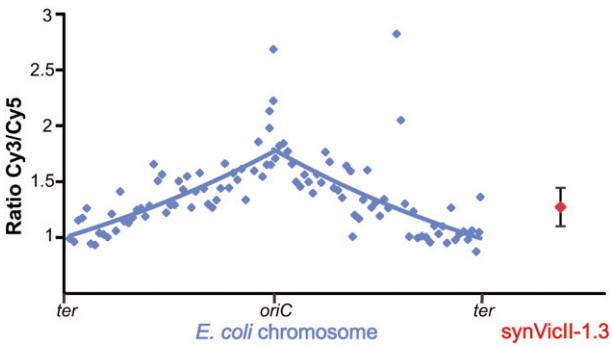


Figure 5. Comparative genomic hybridization of *E. coli* strain SMS18 harbouring synVicII-1.3. DNA of exponentially grown SMS18 was hybridized against DNA from stationary phase cells of strain FSK18 (see Section 2 for details). Respective ratios are plotted against their genomic position. Blue dots represent the 104 probes for the *E. coli* chromosome, blue lines the exponential curve fitted to chromosome arms. The red dot represents the mean of the three probes of synVicII with the corresponding SD.

4 Discussion

Here, we present the design, construction, and initial characterization of secondary synthetic chromosomes in *E. coli*. Important goals in synthetic biology are the standardization and simplification of genetic tools [47]. The design of our synthetic secondary chromosome synVicII is inspired by these goals. A specialized integration site allows the parallel integration of replication origins from natural resources, refactoring processes, or design. The second integration site allows iterative integration of large DNA fragments. This feature is supported by the rare *I-Sce*I site and the yeast origin and selection marker. The gene encoding an unstable GFP allows constant monitoring of the replicon on single-cell level if needed. Genetic engineering of bacteria for biotechnological applications is traditionally based on either the use of plasmids or manipulation of the primary chromosome. While plasmids have the advantage of allowing convenient in vitro construction, the integration of new genetic features to the primary chromosome guarantees high stability of the same. Synthetic secondary chromosomes could potentially combine both advantages. The handling of the secondary chromosomes we constructed is very similar to that of plasmids. Even with extension of their size by more than an order of magnitude the secondary chromosomes can be treated like plasmids since transformation of replicons up to 240 kb has been shown for *E. coli* [48]. The question if secondary chromosomes are as stably inherited as primary chromosomes and thus more stable than plasmids leads to two questions: first, what determines replicon stability and second, what distinguishes a secondary chromosome from a plasmid. In the following, we will address a set of criteria over these questions to discuss the results of our current study.

An easy answer to the first question would be that replicon stability describes how frequent a replicon is lost in a bacterial culture. However, any replicon could be easily stabilized by using selection markers or insertion of addiction systems [49]. These approaches kill the cells that have lost the replicon and therefore reduce the rate of biomass increase. Although this reduction will usually not be severe, we consider it important to focus on the intrinsic replicon stability, i.e., positive selection. By this, we mean mechanisms that lead to stable inheritance of a replicon without negative selection against cells that have lost the replicon. The intrinsic stability could, e.g., be increased by adding a functional partitioning system to the replicon [50]. While many bacterial chromosomes and plasmids encode such systems the *E. coli* chromosome does not [51, 52]. In *Vibrio* species, both chromosomes harbor distinct partitioning systems [53]. We included the respective *parAB2* genes in our synthetic secondary chromosome synVicII (Fig. 1A and B). This might explain its higher stability compared to synEsc, which lacks such a system (Fig. 4B). An alternative explanation for the lower stability of synEsc is that it contains only the minimal *oriC* sequence and lacks *mioC*, which has been shown to stabilize *oriC*-based replicons [50, 54]. ParAB2 have been shown previously to stabilize plasmids that carry a respective centromere-like parS2 site [53]. However, whether the partitioning system of our synthetic secondary chromosome synVicII is fully functional remains to be proven. This is because it encodes only one *parS2* site compared to nine sites encoded on *V. cholerae* ChrII. Notably the very stable replicon based on the F plasmid contains 12 centromere-like sequences (Fig. 3C) [55]. We conclude that there is potential to further increase synVicII stability by increasing the number of parS2 sites. It has to be noted that the partitioning system of *V. cholerae* ChrII is not as modular as other partitioning systems. This is first because the centromere like *parS2* sites are distributed over the whole ChrII and not clustered beside the partitioning genes as for example in the F plasmid and many other plasmids [53]. Secondly, ParAB2 have been shown to be involved in regulation of DNA replication in addition to their role in partitioning [23, 56]. Another way to increase synVicII-stability could be to delete sequences like the yeast 2μ ori, the URA-3 and the oriR6K after the final assembly to decrease the metabolic cost of the replicon. As mentioned above one could also stabilize the *oriC*-based replicon for example by extending the origin to the neighboring *mioC* gene. Potentially, such a replicon could even be more stable than synVicII. However, we consider this route unfavorable because an additional copy of *oriC* will lead to competition of both, the primary and secondary chromosome. This leads to integration of *oriC* of the secondary replicon into the primary chromosome [36, 57]. Such a phenomenon has not been observed for *V. cholerae oriII*-derived replicons in *E. coli*.

We suggest that yet another factor determining the intrinsic replicon stability is the cell cycle dependent replication. While plasmids usually replicate at random time points during the cell cycle, the initiation of chromosomal origins is synchronized to a specific time point of the cell cycle [14, 15, 33, 41]. We consider this aspect of synchronization between cell division and chromosome replication to be an important characteristic to distinguish plasmids from chromosomes. In this respect the secondary chromosome of *V. cholerae* can be classified as a real chromosome because its replication is coordinated with the cell cycle [14]. However, it has to be mentioned that cell-cycle specific replication might also apply to plasmids, although contradicting results have been found for example for the F plasmid [58–60]. It is important to note that although we assume cell cycle dependent replication of synVicII for the interpretation of copy numbers (Figs. 4 and 5) it remains to be proven that this is really the case for *V. cholerae oriII* dependent replication in *E. coli*. The method of choice here would be a Meselson Stahl density shift experiment that we plan for future approaches [61]. The copy numbers we determined could alterna-

tively be interpreted as average of synVicII replication at many different time points of the cell cycle.

One other characteristic of chromosomes is that they encode essential genes. This is also true for ChrII of *V. cholerae* which harbors about 40 essential genes according to the most recent study [62]. Bacteria are generally regarded as haploid, with one copy of the chromosomes per cell. This would make a copy number of one an additional characteristic of a chromosome while plasmids mostly have higher copy numbers. However, polyploidy might actually be more than an exception in bacteria [63]. The *E. coli* cell also contains multiple chromosomes per cell under conditions supporting rapid growth (Fig. 4). However, this overlapping replication is a specialty of some fast growing bacteria and is only to guarantee that in the end every descendent cell gets precisely one chromosome copy. For a secondary chromosome we would suggest that the copy number should be the same as the primary chromosome or lower, as it is the case for the natural secondary chromosome in *V. cholerae* as well as the synthetic secondary chromosome synVicII [16, 33] (Figs. 4 and 5). Our interpretation that *oriII*-driven replication in synVicII is timed similar to ChrII in *V. cholerae* relative to the primary chromosome would suggest that all parts needed for the timing are included on synVicII. This is, however, contradicting a recent publication showing that a special site on ChrI of *V. cholerae* contributes to the regulation of ChrII replication by binding RctB [64]. It would be conclusive to analyze how or if replication of synVicII is changed in an engineered *E. coli* encoding the respective regulatory site from *V. cholerae* ChrI on the primary chromosome.

We have developed a flow-cytometry-based assay to measure secondary chromosome stability. While GFP encoded on secondary replicons was used before to detect respective copy numbers we use it here to measure replicon stability [36]. A very recent study describes a similar approach [65]. One challenge they experienced is the separation of GFP^- cells from background caused by small particles in the used buffer. The solution we present here is the staining of cellular DNA using Hoechst. Since only cells will be stained and not small particles this allows correct quantification of GFP^- cells. The second challenge described in the published study is the overestimation of replicon stability caused by continued fluorescence of the stable GFP after replicon loss [65]. The obvious solution to this problem we present here is the use of unstable GFP variants.

Our current work is intended as a foundation for the use of synthetic secondary chromosomes for biotechnological as well as basic research applications. We show that *V. cholerae* ChrII serves as valuable template and the use of its replication origin in synthetic replicons is superior to the *oriC*-based construct used here. Importantly, others reported similar results. Liang and colleagues split the *E. coli* chromosome in two linear replicons [66]. Interestingly, attempts to include a copy of the *E. coli oriC* on each replicon failed, whereas the two-replicon *E. coli* could be supported with one *E. coli oriC* and the *oriII* from *V. cholerae* on the second replicon. As the technical capacity for the handling and assembling large DNA fragments increases, we expect a growing need for knowledge of chromosome-like replicons. This study, as well as future characterization, optimization and extension of synthetic secondary chromosomes such as synVicII, first introduced here, will help to meet that need.

We gratefully thank all members of the Waldminghaus group for helpful discussions and Matthew McIntosh for critical reading of the manuscript. We thank Nadine Schallopp for excellent technical assistance. Melanie Blokesch, Mark K. Chee, Ulrich Kück, Søren Molin, Minou Nowrousian, and Joseph T. Wade are acknowledged for providing strains and plasmids. We thank Anke Becker for printing the microarrays and Lotte Søgaard-Andersen for sharing the microarray scanner. This work was supported within the LOEWE program of the State of Hesse.

The authors declare no commercial or financial conflict of interest.

5 References

[1] Liu, T., Khosla, C., Genetic engineering of *Escherichia coli* for biofuel production. *Annu. Rev. Genet.* 2010, *44*, 53–69.
[2] Huang, C. J., Lin, H., Yang, X., Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* 2012, *39*, 383–399.
[3] Wendisch, V. F., Bott, M., Eikmanns, B. J., Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr. Opin. Microbiol.* 2006, *9*, 268–274.
[4] Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A. et al., Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 2008, *319*, 1215–1220.
[5] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N. et al., Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 2010, *329*, 52–56.
[6] Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S. et al., Total synthesis of a functional designer eukaryotic chromosome. *Science* 2014, *344*, 55–58.
[7] Suwanto, A., Kaplan, S., Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: Presence of two unique circular chromosomes. *J. Bacteriol.* 1989, *171*, 5850–5859.
[8] Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L., Ramuz, M., Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. *J. Bacteriol.* 1993, *175*, 7869–7874.
[9] Pohlmann, A., Fricke, W. F., Reinecke, F., Kusian, B. et al., Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nat. Biotechnol.* 2006, *24*, 1257–1262.
[10] Trucksis, M., Michalski, J., Deng, Y. K., Kaper, J. B., The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14464–14469.

[11] Ren, S. X., Fu, G., Jiang, X. G., Zeng, R. et al., Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 2003, *422*, 888–893.
[12] White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K. et al., Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* 1999, *286*, 1571–1577.
[13] Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A. et al., DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 2000, *406*, 477–483.
[14] Egan, E. S., Lobner-Olesen, A., Waldor, M. K., Synchronous replication initiation of the two *Vibrio cholerae* chromosomes. *Curr. Biol.* 2004, *14*, R501–502.
[15] Rasmussen, T., Jensen, R. B., Skovgaard, O., The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. *EMBO J.* 2007, *26*, 3124–3131.
[16] Duigou, S., Knudsen, K. G., Skovgaard, O., Egan, E. S. et al., Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. *J. Bacteriol.* 2006, *188*, 6419–6424.
[17] Egan, E. S., Waldor, M. K., Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* 2003, *114*, 521–530.
[18] Pal, D., Venkova-Canova, T., Srivastava, P., Chattoraj, D. K., Multipartite regulation of *rctB*, the replication initiator gene of *Vibrio cholerae* chromosome II. *J. Bacteriol.* 2005, *187*, 7167–7175.
[19] Venkova-Canova, T., Chattoraj, D. K., Transition from a plasmid to a chromosomal mode of replication entails additional regulators. *Proc. Natl. Acad. Sci. USA* 2011, *108*, 6199–6204.
[20] Venkova-Canova, T., Saha, A., Chattoraj, D. K., A 29-mer site regulates transcription of the initiator gene as well as function of the replication origin of *Vibrio cholerae* chromosome II. *Plasmid* 2012, *67*, 102–110.
[21] Venkova-Canova, T., Srivastava, P., Chattoraj, D. K., Transcriptional inactivation of a regulatory site for replication of *Vibrio cholerae* chromosome II. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 12051–12056.
[22] Yamaichi, Y., Fogel, M. A., Waldor, M. K., *par* genes and the pathology of chromosome loss in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 630–635.
[23] Venkova-Canova, T., Baek, J. H., Fitzgerald, P. C., Blokesch, M., Chattoraj, D. K., Evidence for two different regulatory mechanisms linking replication and segregation of *Vibrio cholerae* chromosome II. *PLoS Genet.* 2013, *9*, e1003579.
[24] Duigou, S., Yamaichi, Y., Waldor, M. K., ATP negatively regulates the initiator protein of *Vibrio cholerae* chromosome II replication. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 10577–10582.
[25] Koch, B., Ma, X., Lobner-Olesen, A., *rctB* mutations that increase copy number of *Vibrio cholerae oriCII* in *Escherichia coli*. *Plasmid* 2012, *68*, 159–169.
[26] Yamaichi, Y., Gerding, M. A., Davis, B. M., Waldor, M. K., Regulatory cross-talk links *Vibrio cholerae* chromosome II replication and segregation. *PLoS Genet.* 2011, *7*, e1002189.
[27] Clark, D. J., Maaloe, O., DNA replication and division cycle in *Escherichia coli*. *J. Mol. Biol.* 1967, *23*, 99–112.
[28] Gietz, R. D., Schiestl, R. H., Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2007, *2*, 1–4.
[29] Colot, H. V., Park, G., Turner, G. E., Ringelberg, C. et al., A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 10352–10357.
[30] Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P. et al., New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* 1998, *64*, 2240–2246.
[31] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C. et al., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 2009, *6*, 343–345.
[32] Datsenko, K. A., Wanner, B. L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 6640–6645.
[33] Stokke, C., Waldminghaus, T., Skarstad, K., Replication patterns and organization of replication forks in *Vibrio cholerae*. *Microbiology* 2011, *157*, 695–708.
[34] Serrania, J., Vorholter, F. J., Niehaus, K., Puhler, A., Becker, A., Identification of *Xanthomonas campestris pv. campestris* galactose utilization genes from transcriptome data. *J. Biotechnol.* 2008, *135*, 309–317.
[35] Waldminghaus, T., Skarstad, K., ChIP on Chip: Surprising results are often artifacts. *BMC Genomics* 2010, *11*, 414.
[36] Lobner-Olesen, A., Distribution of minichromosomes in individual *Escherichia coli* cells: Implications for replication control. *EMBO J.* 1999, *18*, 1712–1721.
[37] Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E. et al., Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* 1995, *20*, 448–455.
[38] Stringer, A. M., Singh, N., Yermakova, A., Petrone, B. L. et al., FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. *PLoS One* 2012, *7*, e44841.
[39] Niki, H., Ichinose, C., Ogura, T., Mori, H. et al., Chromosomal genes essential for stable maintenance of the mini-F plasmid in *Escherichia coli*. *J. Bacteriol.* 1988, *170*, 5272–5278.
[40] Cooper, S., Helmstetter, C. E., Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* 1968, *31*, 519–540.
[41] Skarstad, K., Boye, E., Steen, H. B., Timing of initiation of chromosome replication in individual *Escherichia coli* cells. *EMBO J.* 1986, *5*, 1711–1717.
[42] Waldminghaus, T., Weigel, C., Skarstad, K., Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. *Nucleic Acids Res.* 2012, *40*, 5465–5476.
[43] Stokke, C., Flatten, I., Skarstad, K., An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS One* 2012, *7*, e30981.
[44] Lobner-Olesen, A., von Freiesleben, U., Chromosomal replication incompatibility in Dam methyltransferase deficient *Escherichia coli* cells. *EMBO J.* 1996, *15*, 5999–6008.
[45] Sueoka, N., Yoshikawa, H., The chromosome of *Bacillus subtilis*. I. Theory of marker frequency analysis. *Genetics* 1965, *52*, 747–757.
[46] Johnsen, L., Flatten, I., Morigen, Dalhus, B. et al., The G157C mutation in the *Escherichia coli* sliding clamp specifically affects initiation of replication. *Mol Microbiol* 2011, *79*, 433–446.
[47] Martinez-Garcia, E., Calles, B., Arevalo-Rodriguez, M., de Lorenzo, V., pBAM1: An all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* 2011, *11*, 38.
[48] Sheng, Y., Mancino, V., Birren, B., Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res* 1995, *23*, 1990–1996.
[49] Kroll, J., Klinter, S., Schneider, C., Voss, I., Steinbuchel, A., Plasmid addiction systems: Perspectives and applications in biotechnology. *Microb. Biotechnol.* 2010, *3*, 634–657.
[50] Lobner-Olesen, A., Atlung, T., Rasmussen, K. V., Stability and replication control of *Escherichia coli* minichromosomes. *J. Bacteriol.* 1987, *169*, 2835–2842.
[51] Niki, H., Hiraga, S., Subcellular localization of plasmids containing the *oriC* region of the *Escherichia coli* chromosome, with or without the *sopABC* partitioning system. *Mol. Microbiol.* 1999, *34*, 498–503.
[52] Livny, J., Yamaichi, Y., Waldor, M. K., Distribution of centromere-like *parS* sites in bacteria: insights from comparative genomics. *J. Bacteriol.* 2007, *189*, 8693–8703.
[53] Yamaichi, Y., Fogel, M. A., McLeod, S. M., Hui, M. P., Waldor, M. K., Distinct centromere-like *parS* sites on the two chromosomes of *Vibrio* spp. *J. Bacteriol.* 2007, *189*, 5314–5324.

[54] Lobner-Olesen, A., Boye, E., Different effects of *mioC* transcription on initiation of chromosomal and minichromosomal replication in *Escherichia coli*. *Nucleic Acids Res.* 1992, *20*, 3029–3036.
[55] Mori, H., Kondo, A., Ohshima, A., Ogura, T., Hiraga, S., Structure and function of the F plasmid genes essential for partitioning. *J. Mol. Biol.* 1986, *192*, 1–15.
[56] Kadoya, R., Baek, J. H., Sarker, A., Chattoraj, D. K., Participation of chromosome segregation protein ParAI of *Vibrio cholerae* in chromosome replication. *J. Bacteriol.* 2011, *193*, 1504–1514.
[57] Skarstad, K., Lobner-Olesen, A., Stable co-existence of separate replicons in *Escherichia coli* is dependent on once-per-cell-cycle initiation. *EMBO J.* 2003, *22*, 140–150.
[58] Keasling, J. D., Palsson, B. O., Cooper, S., Replication of mini-F plasmids during the bacterial division cycle. *Res. Microbiol.* 1992, *143*, 541–548.
[59] Cooper, S., Keasling, J. D., Cycle-specific replication of chromosomal and F-plasmid origins. *FEMS Microbiol. Lett.* 1998, *163*, 217–222.
[60] Helmstetter, C. E., Thornton, M., Zhou, P., Bogan, J. A. et al., Replication and segregation of a miniF plasmid during the division cycle of *Escherichia coli*. *J. Bacteriol.* 1997, *179*, 1393–1399.
[61] Meselson, M., Stahl, F. W., Vinograd, J., Equilibrium sedimentation of macromolecules in density gradients. *Proc. Natl. Acad. Sci. USA* 1957, *43*, 581–588.
[62] Kamp, H. D., Patimalla-Dipali, B., Lazinski, D. W., Wallace-Gadsden, F., Camilli, A., Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathog.* 2013, *9*, e1003800.
[63] Pecoraro, V., Zerulla, K., Lange, C., Soppa, J., Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS One* 2011, *6*, e16392.
[64] Baek, J. H., Chattoraj, D. K., Chromosome I controls chromosome II replication in *Vibrio cholerae*. *PLoS Genet.* 2014, *10*, e1004184.
[65] Loftie-Eaton, W., Tucker, A., Norton, A., Top, E. M., Flow cytometry and real-time quantitative PCR as tools for assessing plasmid persistence. *Appl. Environ. Microbiol.* 2014, *80*, 5439–5446.
[66] Liang, X., Baek, C. H., Katzen, F., *Escherichia coli* with two linear chromosomes. *ACS Synth. Biol.* 2013, *2*, 734–740.

Biotechnology Journal

Supporting Information for DOI 10.1002/biot.201400031

Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*

Sonja J. Messerschmidt, Franziska S. Kemter, Daniel Schindler and Torsten Waldminghaus

Supporting information

Figure S1

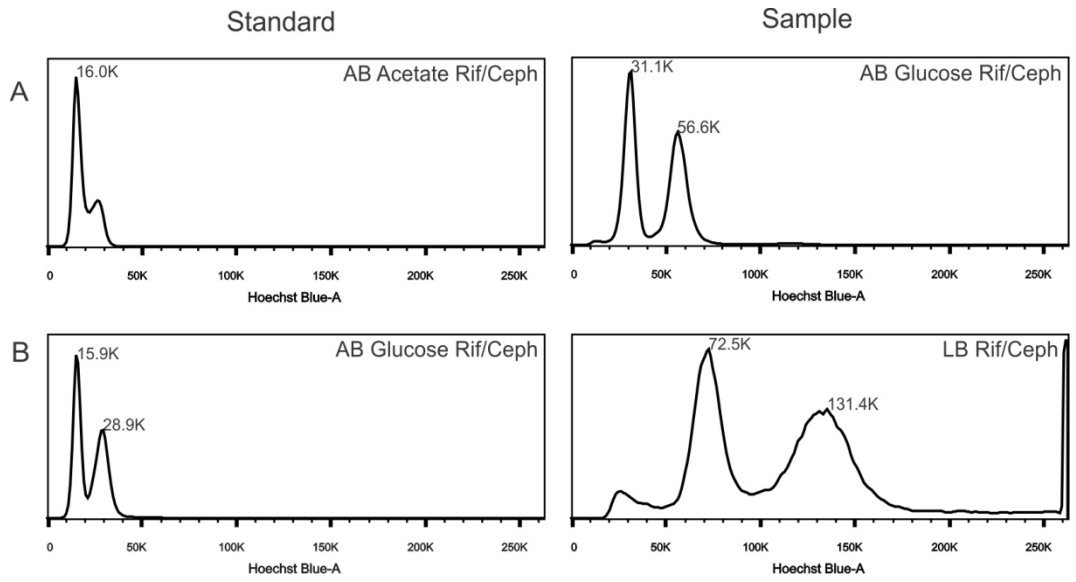


Figure S1 Determination of chromosome copy numbers in flow cytometry histograms of Hoechst stained cells. **A** *E. coli* strain SMS18 grown in AB Glucose medium was treated with rifampicin and cephalexin at an OD_{450} of 0.15 (Sample). *E. coli* MG1655 cells grown in AB Acetate was used as standard. Both cell types where stained with Hoechst and the standard in addition with FITC. The samples where than mixed and measured together by flow cytometry. FITC staining allowed separation of sample and standard population via the FITC channel. The two peaks from the slow growing AB Acetate cells are known to represent cells containing 1 and 2 chromosomes. Mean values of the peak populations where calculated and indicated in the histograms near the respective peak. Numbers indicate that cells in AB Glucose cells contain two or four chromosomes as seen before [1]. **B** SMS18 cells grown in AB Glucose shown in A were used as standard for cells grown in LB and treated with rifampicin and cephalexin at an OD_{600} of 0.15. The 2- and 4-chromosome peaks of the AB Glucose cells result in a Hoechst signal of 16 and 29K, respectively. Hence, the peaks of 73 and 131K of the LB sample represent 8 and 16 chromosomes.

Figure S2

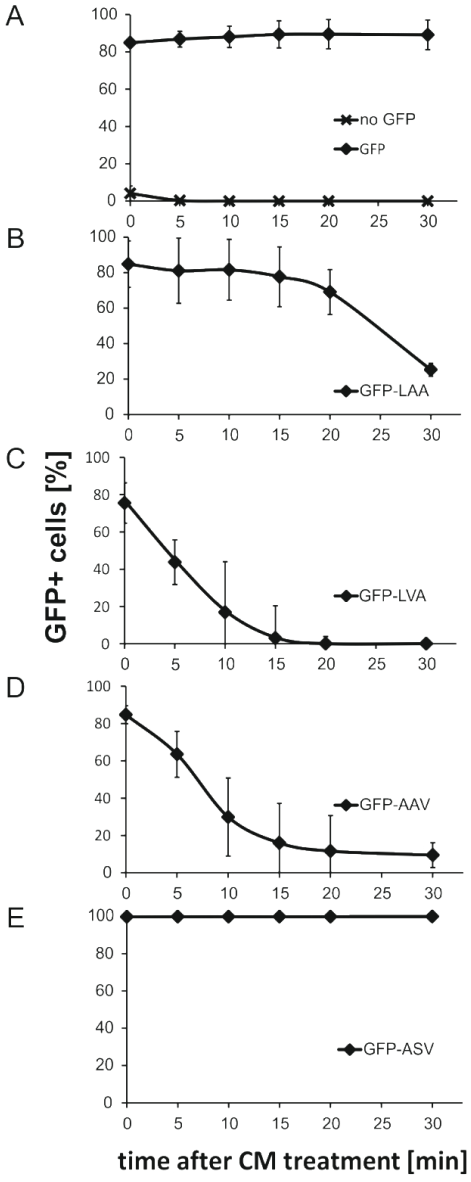


Figure S2 Degradation of different GFP variants encoded on synthetic chromosomes in *E. coli* MG1655. Strains NZ23 and SMS16-20 were grown in LB medium with ampicillin. 0.2 mM IPTG was added at an OD_{600} of 0.2. Translation was inhibited 1 h after induction by addition of chloramphenicol (100 µg/ml). GFP fluorescence was measured in 5 min intervals by flow cytometry. Degradation rates are dependent on the last three amino acids of the ssrA tag indicated in the diagrams. Results shown are mean values of three independent experiments with the indicated standard deviation. **A** SMS20 with *gfp*mut3* containing no degradation tag and NZ23 (*wt*, no GFP) were used as controls. **B-E** Degradation of the different GFP variants in SMS16-19.

Table S1: Strains and plasmids used in this work

Strain	Relevant genotype and characteristics[a)]	Reference
E. coli DH5α λ *pir*	*supE*44, Δ*lacU*169 (Φ*lacZ*ΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λ *pir* phage lysogen	[2]
E. coli XL1 Blue	*supE*44, *hsdR*17, *recA*1, e*ndA*1, *gyrA*96, *thi1, relA*1, *lac*$^-$ [F' *proAB, lac*^{q}*Z*M15, Tn10(tet^r)]	Stratagene, Agilent Technologies, Germany
E. coli MG1655	wild type	[3]
V. cholerae El Tor N16961	StrR	[4]
S. cerevisiae PJ69-4a	MATa, *trp*1-901, *leu2*-3, *leu2*-112, *URA*-3-52, *his*3-200, *gal4*Δ, *gal80*Δ, GAL2-ADE2, LYS2::GAL1HIS3, *met*2::GAL7-	[5]
TB13	*E. coli* DY330 with CamGATC cluster, Cam	[6]
JW3718	BW25113 Δ*gidB*	[7]
NZ23	*E. coli* MG1655 synVicII-1.0	This work
SMS16	*E. coli* MG1655 synVicII-1.1	This work
SMS17	*E. coli* MG1655 synVicII-1.2	This work
SMS18	*E. coli* MG1655 synVicII-1.3	This work
SMS19	*E. coli* MG1655 synVicII-1.4	This work
SMS20	*E. coli* MG1655 synVicII-1.5	This work
SMS28	*E. coli* MG1655 synEsc-1.3	This work
SMS32	*E. coli* MG1655 synVicII-1.6	This work
SMS62	*E. coli* MG1655 synF-plasmid	This work
FSK9	JW3718 pMA407	This work
FSK18	*E. coli* MG1655 with 3 microarray probes for synVicII (in AmpR, URA-3, 2µ *ori*), KanR	This work
Plasmid		
pUC57kan	Cloning vector, KanR	GenScript, Piscataway Township, NJ, USA
pNEB193	Donor of AmpR	NEB, Ipswich, MA, USA
pPS11	Donor of R6K*oriγ,* CMR, KanR	[8]
pRSII426	Donor of 2µ *ori*, *URA-3*	[9]
pAMD135	Donor of *thyA*	[10]
pJBA27	AmpR; pUC18Not-$P_{A1/04/03}$-RBSII-*gfp*mut3*-T_0-T_1	[11]
pJBA110	AmpR; pUC18Not-$P_{A1/04/03}$-RBSII-*gfp*(LAA)-T_0-T_1	[11]
pJBA111	AmpR; pUC18Not-$P_{A1/04/03}$-RBSII-*gfp*(LVA)-T_0-T_1	[11]

pJBA112	AmpR; pUC18Not-$P_{A1/04/03}$-RBSII-*gfp*(AAV)-T_0-T_1	[11]
pJBA113	AmpR; pUC18Not-$P_{A1/04/03}$-RBSII-*gfp*(ASV)-T_0-T_1	[11]
pBR-FRT-kan-FRT	KanR	[12]
synF-plasmid	synX-0.3 with F *origin*, *sopABC*	This work
synX-0.1	R6K*oriγ*; 2µ *ori*, *URA-3*, AmpR	This work
synX-0.3	synX-0.1 with *gfp*-AAV	This work
synVicII-0.1	*rctB* (coord. 1134–3110), *parAB* (coord. 1070018-1072220), R6K*oriγ*; 2µ *ori*, *URA-3*, AmpR	This work
synVicII-0.2	synVicII 0.1 + *oriII* (coord. 1072221-750)	This work
synVicII-1.0	synVicII 0.1 + *oriII* (coord. 1072221-1133)	This work
synVicII-1.1	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*(LAA)	This work
synVicII-1.2	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*(LVA)	This work
synVicII-1.3	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*(AAV)	This work
synVicII-1.4	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*(ASV)	This work
synVicII-1.5	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*mut3*	This work
synEsc-1.0	synX-0.1 + *oriC* (coord. 3923616-3924064)	This work
synEsc-1.3	synEsc-1.0+ $P_{A1/04/03}$-RBSII-*gfp*(AAV)	This work
synVicII-1.6	synVicII-1.3 + *thyA*	This work
pMA407	pUC57kan with the cloned qPCR templates for AmpR, *oriII* (coord. 1070232-1070355), *oriC* (coord. 3921529-3921685)	This work
pMA411	pBR-FRT-kan-FRT with Amp*R*, URA-3, 2µ *ori* microarray probes for synVicII, KanR	This work

a) Genomic positions are indicated according to the genome annotation NC_002506.1 for ChrII of *V. cholerae* and NC_ 000913.2 for *E. coli* MG1655.

Table S2: Oligonucleotides used in this work

Name[a)]	Sequence from 5' to 3'
1	GGGGCCAAGCGGCCGCATGAGCTCAGAAGAAAAACGATTGATC
2	GGGTAGGGGTTCCGCGCGGGATAGAAAGCACTGAGTCAGG
3	GGGGCTTTCTATCCCGCGCGGAACCCCTATTTGTTTATTTTTC
4	GGGTTATTACCCTGTTATCCCTAGTGGAACGAAAACTCACGTTAAGGG
5	GGGTTCCACTAGGGATAACAGGGTAATAACCTGTTGATAGTACGTACTAAGCTC
6	GGGCCAGCAAAACTAACCATGTCAGCCGTTAAGTGTTCCTG
7	GGGACGGCTGACATGGTTAGTTTTGCTGGCCGCATCTTCTC
8	GGGCTTAGGCGCTGCCAACGAAGCATCTGTGCTTCATTTTG
9	GGGCAGATGCTTCGTTGGCAGCGCCTAAGAAACCAATAAGG
10	GGGCTTCTGAGCTCATGCGGCCGCTTGGCAATGAAAAGAGAACAAACGATAG
14	CAATCTCAATTCGATCGGCCTGCACT
15	CAGTCAATCGCGGCTTCAGAGACTTC
16	GGTTAGATCCGTATCACACTTACCGT
17	CCATCTTTGTGAGTTCTTGGCAATTTGATCAATCGTTTTTCTTCTGAGCTCATGCGGCCGCGTCGTTGTATCTCCTTCCTCTCGTAC
26	CTTAACGTGAGTTTTCGTTCCACTAGGGATAACAGGGTAATTTTATCAAAAAGAGTGTTGACTTGTGAG
27	GTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAGGTTCCAAGCTAGCTTGGATTCTCAC
117	AGAGCGGCCGCCGCGGAACCCCTATTTGTTTATTTT
118	GCGGCGGCCGCAACGAAGCATCTGTGCTTCATTTTG
124	GCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTCCGCCAATGATGATGACGTC
125	GAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCTGGGATCGTGGGTTAATTTAC
126	CATTGTTATCCGCTCACAAGTCAACACTCTTTTTGATAAACCATGGCCGCGGGATTTAGAC
127	CTTAACGTGAGTTTTCGTTCCACTAGGGATAACAGGGTAATTTAGATAGCCACCGGCGCTTTAATG
111	GCGCCCGGGAGGAGTGGCAGCATATAGAACAGCTAAAGGGTAGTGCTGAAGGAAGCATACGATACCCCGCGCGGAACCCCTATTTGTTTATTTTTC
112	GCGTCTAGAACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAGTGGAACGAAAACTCACGTTAAGGG
113	GCGTCTAGACCTTGAGCTTGAGATTGCTG
114	ATACCCGGGGCCGCCCTACTATCGTTAAA
352	CATGTTTGACAGCTTATCATCGATAAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCACTGACCCAATGCGTCTCCCTT
353	TAACTGTGATAAACTACCGCATTAACGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCTGGGGCCGCCCTACTATCGTT
354	GGCGGATTTTCTCCAGCTTCTGTATTGGTAAGTAACCGCGCTTACGAAGCAAAGATCAGTTGGGTGCACGA
355	CATTTTGCTGCGATCCAGGCCTTTTTCCTGCTCGCGTTTTTTAATCAGTAAGAGCGCTTTTGAAGCTGGG
361	CGACGTTGTAAAACGACGGCCAGTGCAATGCTATCGAGAATATGGCGTACCAG
362	ATGCATTCGCGAGGTACCGAGCTCGTTACTGCTGAAAGACGCAGGTATTTCGCTT
608	CAAAATGAAGCACAGATGCTTCGTTCATGGAGCGGCGTAACCGTC
609	CGGCATCTCTGATAGCCTGAG
610	CCGGCGCTGGAGAATAGGTG
611	GAAAAATAAACAAATAGGGGTTCCGCGCGTCGACAGCGACACACTTG
3921366 fw	GAGAATATGGCGTACCAGCA
3921366 rv	AAGACGCAGGTATTTCGCTT
3921366 p	CAACCTGACTTCGGTCCGCG 5'Fam-3'Tamra
ter-fw	TCCTCGCTGTTTGTCATCTT
ter-rv	GGTCTTGCTCGAATCCCTT
ter-p	CATCAGCACCCACGCAGCAA 5'Fam-3'Tamra
ori2fw	CCTTGAGCTTGAGATTGCTG

ori2rv	GCCGCCCTACTATCGTTAAA
ori2probe	TGCTCACGCTGAGCCTCATTCA 5'Fam-3'Tamra
AmpR fv	TAACACTGCGGCCAACTTAC
AmpR rv	AAGGCGAGTTACATGATCCC
AmpR Sonde	TCGGTCCTCCGATCGTTGTCA 5'Fam-3'Tamra
MA_0001	CCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGTAACGGTG
MA_0002	GAAGAGATGATGATCAAACTGTTTGGCGAACCGGTGGCTTTCGATGAAGAGCAGCCGCAG
MA_0003	CGAGTCTACGCAACTATCGGCAGCTACGGCATTTATCGCCCACTGTCGATTACCAAAGTT
MA_0004	TTTCAGATCCAGCGCCGAGATATAACCCCATGCTGGCTGTTTACATGGCAGACCAAATGG
MA_0005	GTTAATCCGCCCTGCTCAACAAACCACTGATAACCATCATCGGCCAACATTTGCGTCCAC
MA_0006	CGGCTTTCTGCCCCTTCGATCACCTGATTATCATTAATACGGAAAGAGAGCAGCACCGCT
MA_0007	TGAAGCTTCTGTACTGGTTACGCGCGCTTCCGATCACAATCCGCTACTCGTTGAATTCAG
MA_0008	CCAGCATCACCACGATCAGAATGGAACCGTAGGCAATCGCTACACCGTATTCGCCATCTT
MA_0009	TCTGATCGGTTCCTGCGTTAGTTCACATCACGACTCATTTTTTCGCTCTCACCGGCATCC
MA_0010	GTCTGGGGGCGCAACGACCGCTTTGTGCCGATGGATGCGGGTCTGCGTCTGCTGTCCGGC
MA_0011	ATCTGGAAATTCGCGTCATGCCTTATACCCACAAACAGTTGCTGATGGTGGCGCGTGATG
MA_0012	CTGGAGCCGGAAGGGCAGGTGGTAGCGGATGCGCTTATTACGTTTGTTTGTATTGATCTT
MA_0013	ACAGATGCGTCATAACCCGCTTGTTTGATGGTTTCGATCAGCTGTTCTGCACTGGCAGTC
MA_0014	AACGGCAGTTGAACCAGAATGCCATCGATGGTGTTGTCGGCATTCAGCGTATCGATAAGC
MA_0015	TTTAAATACCAACATCAGCACGGCAGCCATTGCACCAAGACCGCTGATCAGAATCGCTGG
MA_0016	TTCGGAATCCAGCGCGTTGACACGCACAATGGTTTCAATATCGCGATACAGCGGATGTTG
MA_0017	TTGGTGGCTACGACGGGATTCGAACCTGTGACCCCATCATTATGAGTGATGTGCTCTAAC
MA_0018	TGATATCAACCTGTATGGCTGGCATTTGCCGCTTGACGCACATCCTGAGCTGGGCAATAA
MA_0019	ATCGATCAGCAAGGCATGATCTTTCTTGCCGAGCGCGGAAATTAGCTGATCCATGATCCC
MA_0020	TTCTTTAGCGATATCGGCGGGCGGGCGCTACTAATGCATGGTACTGAAGGTGAAGTGTAT
MA_0021	AATCGATGGGTTTGAAAATGGCGTTGCTGGCGTGAAATTCACCACCGCAAAGAAACGCGC
MA_0022	GTCAGTAACATAATGCCGCCCAGCGTAAAGGCGAGGGTGAAATCAAGGAAACTTAACCCG
MA_0023	TGGTGTTCGATCTGCAGAGCAAACTCGATCGTATTATCAGTTGGGGCCAGCAATCCATCG
MA_0024	GGGCACAATTGAATACCGATCATTGTTTGGCAGTTACAGCCTGACCGTTGACGACACGGT
MA_0025	TTGCAACCGGTGCAGTTTGCGTTTTACCGCCTGCTTTTTCAAATAATTGCGGCGGCATGG
MA_0026	CAGCAGTACAGAAAAGATCAGCGCTGTAATCAGGATCTTTTCGGGTCCGATTCGATCGCC
MA_0027	ATAAAGACCTTGTTGATCTCTTTATCGTTGTACGCCCAGTCAGGTGCGGTACTGCGTGCG
MA_0028	ACGATTTCATGGACGTGACAAACGATGATTACATCCGCCTGTTATCGGCACTGTTGCCGC
MA_0029	GGTGGCTGTATGTAAGCTGGGTTTGCGGCTCTATATAGAACCCATAACCAGTAGGGGACA
MA_0030	CCCGCAGGCGTCACACTGGCGGAAAAACAAACACTGGTACGTAACAATGGTTCAGAAGTT
MA_0031	TGATAACCGCTTTCAGCAGACGGTGGTTGATCATGTCGCCATATTTACGGATCGGCGAAG
MA_0032	GGACGAAACCCGGAAATCTGATCGGCAATGGCGCTTATGTTCTTAAAGAGCGCGTAGTCA
MA_0033	AGAATCGTGTCATCGGCCAGCGGGACAATTTTATTGATTTCGTGGGACGTGCGGAAACCA
MA_0034	AATGGATCAACAAACTGCCGCCCGCACGTCGTGAGGATGAAGACGTGAAAGAGATCCGTT
MA_0035	GGAGCGGGCTTTCGTAATCTTGCCCTGGGCATTATTGTCGTATAAACGCTTTACCGGACA
MA_0036	TCGGCAACCATATTTACGCAACGCAGGTCGATGGCTGCGGATTGCGGATATTCTTCTTTG
MA_0037	CAGTGAATGAGAGCGTACCCGCGCTTTCCCAAAGAATAACGGTAAAGGTGATCAGCATGC
MA_0038	TAACGGCAGGACACTGTTTATTGACACCTCCAAAGGGTAAAGCCGATAAAGCAGTGGCGC
MA_0039	TTTCGGTGCAGGTTGCGAGAAATACGGGTCGCTGAGCATATTTTGCAGCAGTGGGAGAAT
MA_0040	AGGTGAAAGATTGCTCTGGTCACGGCGTAAAAATTCAATGGCCAGCTGGGTGGGATCAAG
MA_0041	ATCTGAGCGTTCCGCAGCTGGTCTTGTCCTTTATCCCGAAAAACCCGTTTGCCGATCTTA
MA_0042	ACCGTCTGGATAGTGACTTCGCGATTGATTTTGAGGAAATTGATTGCCGAATCGCCGGGT

MA_0043 ACAGCGTGCTAATGGCGCAGAGATCGACGACGGTGCTATCAACGGTATCAAAGTCGGTTT
MA_0044 AGTTCAGAGGCCAGCTTTAACGCGTCCGCACGATCCGGATCGAAAAATACCTGCACGATA
MA_0045 CCCGACGTGCTGACGCAACTGGCGAATCAGAGATTTTTGCTGACTTAATGAACGTGCAGT
MA_0046 CCGTTGCGTGAGGTTGCTGGCTGGGAACACGATCTTCATGCTGCAATGAACAACATTCAG
MA_0047 AAGGATCGGGGCATGATCTTCAACGGCTTTCATCATTGTTGCCGACAAATTCTGACGCGC
MA_0048 ATACGGATCGGGGATTTCACATTCGTTATCCCAGTGACCAAACAGCATCACTTTGCCGCG
MA_0049 TTCTTTCAACGAGTATTCGAAGATGTTGTTGATCTGCGTGCCGTAACCGCGCGAACGCGT
MA_0050 AATAAAACGACATATCCAGAAAAATATACACTAAGTGAATGATATCTTCCGATTTATCTT
MA_0051 GTCGATCTCTGGGGACAACGCTTTATTGAAGTATGGTCGGGGATGCCGACGCTGTTTTTG
MA_0052 GGAACCGGGCTTGGTCTGGCGTTAAGTCAACGCATCATTAATGCCCATCAGGGTGATATT
MA_0053 ATTCCGGTTTATGCGCCGGATAACGTTAATCATCCGCTGTGGGTGGAACGCATTGCCCAA
MA_0054 TCAACAAAGAAGGTACCCGTCCTGCGGTGGTTATCCCAACCAACGAAGAACTGGTTATCG
MA_0055 AAACAAAATTCACCCGAATCCATGAGTGCGCCACCTCCAAATTTTGCCAGCTGGATCGCG
MA_0056 AAAGCTGAAGGCAACAAAACCTGCGGCTTTGTTGATCGGGCCACCAAGGTCGATTGCTGT
MA_0057 CCCCGTTACAAAAAGTGGTGCGACGAATACTTCTACCTCAAACATCGCAACGAACAGCGC
MA_0058 AACATTATCTGGTGATCACTGCGTTGGGTGCCGATCGCCCTGGAATTGTGAACACCATCA
MA_0059 CGGTTTGAGCACTGATCCTGGCGATCGGATCGAATTGACCATATCAACATGACCAAAGCG
MA_0060 TAATCACAAAACGGTCACCGGAAACCGCCTGCCCTACATAATGGACACAATCAGCAAGCC
MA_0061 GCAAAAGTAGCAAAGTGGCGCAGTGCCTGAATAAACGCCATATTCTGATCGTGCTCGACG
MA_0062 AAATTATGATGGGTCCACGCGTGTCGGCGGTGAGGCGTAGCTTAATAAAGGTTGCTCTAC
MA_0063 GATCAGGCGCTGAAAACGATCCTGGTGACCAATCCCACGCTGCAATTCGATGGCGATATT
MA_0064 CCAACAACCAGATGGCTAATCTGCCTCGTAAGCGCGGAGGTACATTTTCAGTGACCACGA
MA_0065 ATGACCAATCAGATTTTCCTGCACCCGACGGAAGTCGTCTTCACTCCAGGTTTGCAGCAA
MA_0066 GGGTGTTCCAGTGATTACGTCATGGCGACCAAAGATGGCCGTATGATTTTGACCGATGGA
MA_0067 AAAGGTTGGGATGAATATAAGCAACGCTGGCACAAACCAGCCGGTTCTGGTTCACGCCAT
MA_0068 GATCTTATGCAGGCGTTAAGCGATCTCAACCGGCCGGAAATTCGCTGTATCATTTTGCGC
MA_0069 ATCTTCGGGGCGATGAACGAAGGGATGACCGGCGATCCGAGTATTTTAATCGCCAAGTCA
MA_0070 TGTGATAAACCAGATCGAACTTCATCCGCTGATGCAACAACGCCAGCTACACGCCTGGAA
MA_0071 ATGCGGCGTAAACGCCTTATCCGGCTACATGTCAACGACAGTTGTAGGCCTGATAAGACG
MA_0072 AAGAGTTGAGTAAGATTCGTAGCAAAGCGGAGCAGGCACTGAAACAGAGCCGTTATCGCC
MA_0073 CCAGCATGATCAACCGTTTCGAAACGGAGCGGCCCAGCTTACCTGCCATTGCACTAAATA
MA_0074 AGAGCTGAAAGCGTTCAATGTTCGTGACGGTTATGGCATTCGTTGTGCGCTCACCTCTGA
MA_0075 TGATCCTCGAGTAATAAACTCACCGGTATAGACGTTGAGGTTGGTCACCCAGCCATCTGC
MA_0076 GAAAAAACAACGTAATTTACGTAGCATGGCGGCCCAGGCCGTTGAACAAGTCGTCGAGCA
MA_0077 CGCTGATCCTGATCTCTCACGACCGCGACTTCCTCGATCCGATCGTCGATAAAATTATTC
MA_0078 TTATTCTGGATCCCGCTGGATATGCCGCTGAAATTCACTCTCTCATGGATGAAAGGCGCG
MA_0079 AATTACGGAAGATGTGTTGAGGATCTTGTTGGTTTCCGCCTGCCCTTTCCACTCTTCGCG
MA_0080 GTTACGGTCAGTTGCTGAGCTTCACCGACTGTTCCGGTTATGTAACCCGTTATGACCATG
MA_0081 ACTCCTTTATGCAGGCATGTGGGCTGGTGAATGATCATGTGGTTGGCTGCTGTTGCTATC
MA_0082 TTTTACCAACAATAATTCCTGCCTGCGGCCCGCCTAACAACTTGTCGCCGGAGAAACTCA
MA_0083 GCGCGTACCCACATACCGCTGTTGTTAAAACAGGTAACCCAGCAGTGATCGCTTCTAGAA
MA_0084 GGCCAGAATATGGCGGATTTGCGTGCTTTGGCTTAAGGCATCGTAAACAGGCAGGATAGA
MA_0085 GCTCGCGGTAGTTATCAATAGTGAGCTTCTGGATATTCTCTTCTTTGATCCCCGCAGGGG
MA_0086 GGGCAATTCGCGCTTCGAACTGTAAGCCTGTATGGGCATCCTGATTTTGATACAAGCGTC
MA_0087 AGCCTATTCCTGGGAATACCCGAACCCACGTCTGCTGGCGAAAGATATCAAACAGCGTTT
MA_0088 TTTCCCACTTAACCATGACTTTGGGACCTTAGCTGGCGGTCTGGGTTGTTTCCCTCTTCA
MA_0089 TACGCGTAAAGCGAGGATCGATCACAATCAGCTTCGCGCCGTTGTGAATTTTGGCTTCCA

MA_0090	TACTGGTGCGCGTATTGTCACCAGCCACGATGATGTCTGTTTGATTGAGTTTCAGGTGCC
MA_0091	GCTGTTAACGACAAAAGGTAAAGCTACCGTTGCTTTTGCCCGTGAAGCGCGTACCGAAGT
MA_0092	AAATGAACGTATTCTGGTGCTGGACGGCGGTATGGGCACCATGATCCAGAGTTATCGACT
MA_0093	TCTCTGACACTTAATGAGAACTGCGATCCCACCGTACGCCACGACATGGAGCGTTTTTTC
MA_0094	TTTCTGCGTCTCTTCGATGGCGTTGATCATCTTCTCCAGATTCTCCAGCCGATCGCCGCC
MA_0095	TTGCGTTTCGTCCAGCATCAGCTGACGACGCCAGACATCTATCTGTTGGCTAAGATAACG
MA_0096	TCAGTATCAGGAAAACGATCTGCCGGATCTGATCGCTTCGCTCGATCAACCGTTCCTGCT
MA_0097	AATGTGTATCCGCCGATGAGCACCTTACTTGAGAAGCAAGGCATTGAGCTGATTCAGGGG
MA_0098	TTACTTCTCAGCGTCAGCCTGCTCCGTTGCATCGTCTGTTTTATCTCCCTGTATCCCAGT
MA_0099	CAGCGCACTTCCCGTTACAGGTCAGAGCATTGACAGTAAATGGCGGAGCCACTCAGGGAA
MA_0100	GCCAGCGGGGATCATGATCAGGCAGGTTGAGAAAGTGCTGAAGTGAATGTGTTTCGAGCA
MA_0101	CCACGATCCCAGCCATTCTTCTGCCGGATCTTCCGGAATGTCGTGATCAAGAATGTTGAT
MA_0102	GGATCGCACTGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGGAAAGG
MA_0103	AAGATCCTGCAAAACGATCGGGACCGCGGATCATAGCCTAAACTGCGCAAGAGATCTTCT
MA_0104	GATAGCGATCCGAGGTGAGAATGATCTGTTGATTACCTTCCAGCAGGGCGTTGAAGGTGT
MA_0105	AGGAGTGGCAGCATATAGAACAGCTAAAGGGTAGTGCTGAAGGAAGCATACGATACCCCG
MA_0106	GACCCAATGCGTCTCCCTTGTCATCTAAACCCACACCGGGTGTCATAATCAACCAATCGT
MA_0107	AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCC

[a)] Oligos named MA… are microarray probes with 5´ C6-amino linker

Supporting References

[1] Stokke, C., Flatten, I., Skarstad, K., An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS One* 2012, *7*, e30981.
[2] Miller, V. L., Mekalanos, J. J., A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J Bacteriol* 1988, *170*, 2575-2583.
[3] Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T.*, et al.*, The complete genome sequence of *Escherichia coli* K-12. *Science* 1997, *277*, 1453-1462.
[4] Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A.*, et al.*, DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 2000, *406*, 477-483.
[5] James, P., Halladay, J., Craig, E. A., Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* 1996, *144*, 1425-1436.
[6] Bach, T., Skarstad, K., An *oriC*-like distribution of GATC sites mediates full sequestration of non-origin sequences in *Escherichia coli*. *Journal of Molecular Biology* 2005, *350*, 7-11.
[7] Baba, T., Ara, T., Hasegawa, M., Takai, Y.*, et al.*, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006, *2*, 2006 0008.
[8] Srivastava, P., Fekete, R. A., Chattoraj, D. K., Segregation of the replication terminus of the two *Vibrio cholerae* chromosomes. *J Bacteriol* 2006, *188*, 1060-1070.
[9] Chee, M. K., Haase, S. B., New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 2012, *2*, 515-526.
[10] Stringer, A. M., Singh, N., Yermakova, A., Petrone, B. L.*, et al.*, FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. *PLoS One* 2012, *7*, e44841.
[11] Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P.*, et al.*, New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* 1998, *64*, 2240-2246.
[12] Blokesch, M., TransFLP--a method to genetically modify *Vibrio cholerae* based on natural transformation and FLP-recombination. *J Vis Exp* 2012.

3.2 Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*

In Kapitel 3.1 wurde synVicII erfolgreich in *E. coli* etabliert. Darauf aufbauend wurde synVicII optimiert und tiefgreifender charakterisiert. Die daraus resultierende synVicII-2.0 Version ermöglicht nun deutlich umfassendere Anwendungen.

Dieses Kapitel zeigt, dass synVicII gegenüber synthetischen sekundären Chromosomen basierend auf *oriC,* eine deutlich höhere genomische Integrität besitzt. Synthetische sekundäre Chromosomen mit einem *oriC* können bei andauernder Kultivierung in das Chromosom von *E. coli* integrieren, wohingegen synVicII eigenständig bleibt, sogar, wenn homologe Bereiche zum *E. coli* Chromosom in synVicII eingefügt werden. Es stellt sich die Frage, ob weitere synthetische sekundäre Chromosomen basierend auf anderen *Vibrionaceae* Spezies ähnliche Eigenschaften besitzen. In dem Kapitel kann gezeigt werden, dass auf neun weiteren Arten basierende synthetische sekundäre Chromosomen in *E. coli* repliziert werden können. Eine Kombination mehrerer der hergestellten synthetischen sekundären Chromosomen in einem *E. coli* Stamm ist vermutlich aufgrund einer vorliegenden Inkompatibilität nicht möglich.

synVicII bedarf weiterer Optimierungen für die Anwendung in der Biotechnologie. Dafür ist es wichtig die Replikonstabilität zu erhöhen, zudem müssen DNA-Fragmente möglichst effizient in synVicII kloniert und hergestellte synVicII-Varianten leicht in Rezipienten transferiert werden können. Deshalb wurden in diesem Kapitel mehrere Optimierungen an synVicII durchgeführt. Durch evolutive Experimente wurde versucht Mutationen in synVicII zu identifizieren, die einen positiven Einfluss auf die Replikonstabilität von synVicII aufweisen. Es wurde ein *origin of transfer* in synVicII integriert, um eine Konjugation von synVicII zu ermöglichen. Des Weiteren wurden Typ IIS-Erkennungssequenzen mutiert, wodurch ein synVicII MoClo-Set etabliert werden konnte. Zudem wurden die synVicII Elemente, die lediglich für die Konstruktion benötigt werden, von FRT-Sequenzen flankiert, um diese abschließend nicht mehr notwendigen Elemente durch eine Flp-Rekombination entfernen zu können.

Sonja Messerschmidt hat in Rücksprache mit Torsten Waldminghaus die Studie konzipiert und durchgeführt. Daniel Schindler hat die MoClo kompatiblen Minichromosomen konzipiert, hergestellt und charakterisiert. Franziska Kemter hat in Rücksprache mit Torsten Waldminghaus und Sonja Messerschmidt die FRT Rekombination etabliert und in Rücksprache mit Torsten Waldminghaus die Kopienzahl bestimmt. Celine Zumkeller hat unter Anleitung von Torsten Waldminghaus und Sonja Messerschmidt die Konjugation von synVicII etabliert. Das Manuskript wurde von Sonja Messerschmidt und Torsten Waldminghaus verfasst unter Beteiligung von Franziska Kemter und Daniel Schindler.

Manuscript under revision

Frontiers in Bioengineering and Biotechnology,

section Synthetic Biology

Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*

*Sonja Messerschmidt, Daniel Schindler, Celine Mara Zumkeller, Franziska Kemter, Nadine Schallopp and Torsten Waldminghaus**

LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-University Marburg, Hans-Meerwein-Str. 6, D-35043 Marburg, Germany

**corresponding author*

E-mail: torsten.waldminghaus@synmikro.uni-marburg.de

Abstract

Learning by building is one of the core ideas of synthetic biology research. Consequently the way to fully understand what a chromosome is one needs to build more and more chromosomes including failures to be expected and accepted. The last years have seen exciting synthetic-chromosome studies and there is certainly more to come. One approach to apply synthetic genomics to chromosome biology is the design, construction and testing of secondary chromosomes. We had previously introduces the synthetic secondary chromosome synVicII in *E. coli*. It is based on the replication mechanism of the secondary chromosome in *Vibrio cholerae*. Here we present a detailed analysis of its genetic characteristics and a directed evolution approach to optimize replicon stability. We probe the origin diversity of secondary chromosomes from *Vibrionaceae* by construction of several new respective replicons. Finally we present a synVicII version 2.0 with several innovations including its full complementarity to the popular Molecular Cloning (MoClo) assembly system. The presented work extends the basis to use secondary chromosomes in *E. coli* to answer basic research questions and for biotechnological applications.

Introduction

New DNA-assembly methods have been developed in recent years and costs of DNA synthesis are constantly decreasing (Chao *et al.* 2014; Kosuri and Church 2014). These two factors are the main driving force for an increasing number of synthetic chromosome projects (Schindler and Waldminghaus 2015; Lee *et al.* 2013; Gibson *et al.* 2008; Annaluru *et al.* 2014). This development was started by Venter and co-workers who constructed a whole *Mycoplasma genitalium* chromosome with a size of 583 kb from scratch (Gibson *et al.* 2008). Two years later, a synthetic chromosome was introduced into bacterial cells replacing the natural chromosome (Gibson *et al.* 2010). The two synthetic chromosomes had in common that they were basically copies of natural genome sequences. The next thing to do would consequently be using the new methodologies to engineer on a chromosome wide scale. The stepwise replacement of chromosome III in *Saccharomyces cerevisiae* with a designed synthetic chromosome synIII was a step in this direction (Annaluru *et al.* 2014). In addition, genome-wide recoding of codons is now possible (Lajoie *et al.* 2013; Ostrov *et al.* 2016). A recent project aiming at a minimal genome with respect to gene content showed that our knowledge about a functional genetic setup is still limited (Hutchison *et al.* 2016). Initially designed chromosomes based on the current state of understanding were not functional and cycles of design, synthesis, and testing of engineered chromosomes were needed. One possibility for such design-build-test cycles to improve our knowledge on chromosome biology is the introduction of extra replicons (Birchler 2015; Messerschmidt *et al.* 2015; Yu *et al.* 2007; Nasuda *et al.* 2005). An extra replicon facilitates more severe engineering and respective testing because cell viability can be maintained by the original genetic setup. Cai and coworkers are currently constructing a so-called neochromosome which relocates all the tRNA genes of *S. cerevisiae* (Pennisi 2014). The sequence of those genes can result in stalling of the DNA-replication machinery leading to hotspots of recombination and retrotransposon insertion. To avoid stalled DNA replication on the main chromosomes, the tRNA genes are transferred to the neochromosome and then added to the synthetic genome of *S. cerevisiae* Sc2.0, an ongoing project in which all 16 yeast chromosomes are exchanged by synthetic ones (Dymond *et al.* 2011). A prerequisite for such neochromosomes is a replication mechanism that ensures chromosome-like replication. In bacteria this is a challenge because the genetic content is stored on a single chromosome replicated from a single replication origin. A secondary copy of this replication origin as driver of an extra replicon has been shown to cause several problems probably due to competition with the native replication origin (Lobner-Olesen 1999; Skarstad and Lobner-Olesen 2003). One interesting alternative is the replication origin of the secondary chromosome of *Vibrio cholerae*. This origin has been shown to replicate in *E. coli* and was used in several respective genome engineering projects (Liang *et al.* 2013; Milbredt *et al.* 2016; Messerschmidt *et al.* 2015; Egan and Waldor 2003; Zhou *et al.* 2016).

V. cholerae is a model system for multi-chromosome bacteria. Its primary chromosome (ChrI) has a size of 2.96 Mbp and the secondary chromosome (ChrII) a size of 1.07 Mbp (Heidelberg *et al.* 2000). While ChrI is replicated from the DnaA-controlled replication origin I (*oriI*), similar to *E. coli*, chrII is replicated from the RctB-controlled origin II (*oriII*) (Duigou *et al.* 2006; Egan and Waldor 2003). Both chromosomes encode their own segregation systems (*parAB1* and *parAB2*) (Yamaichi *et al.* 2007). The core *oriII* region is flanked by the *parAB2* and the *rctB* gene. ParB2 seems not only to participate in segregation but also in the regulation of DNA replication of ChrII (Venkova-Canova *et al.* 2013). The regulation of the replication timing in this two-chromosome system has been extensively studied over the last years (Egan *et al.* 2004; Rasmussen *et al.* 2007). It was shown that ChrI initiates DNA replication first followed by initiation at *oriII* after about two thirds of the primary chromosome is replicated (Rasmussen *et al.* 2007; Stokke *et al.* 2011; Val *et al.* 2016).

On the basis of *oriII* from *V. cholerae* we previously constructed a prototype of the synthetic secondary chromosome synVicII in *E. coli* (Messerschmidt *et al.* 2015). Here, we present a thorough characterization and introduce several innovations leading to a new version of synVicII to satisfy the need for well understood and easy-to-use replication systems for bioengineering and synthetic biology applications.

Results and Discussion

Genetic integrity of synVicII

Genetic circuits for biotechnological applications might be integrated into the primary chromosome of a production strain or alternatively be placed on a secondary synthetic chromosome or plasmid. However, full control of the genetic setup is mandatory. Integration of a secondary replicon into another replicon would for example destroy its genetic context and attributes (Haldimann and Wanner 2001). Notably, the use of an additional copy of the primary chromosome origin to drive secondary chromosome replication is known to result in frequent integration into the primary chromosome (Lobner-Olesen 1999; Skarstad and Lobner-Olesen 2003). To test if the synthetic secondary chromosome synVicII is also prone to integration into the primary *E. coli* chromosome, we measured the degree of integration after an extended cultivation of respective cultures for 1 or 3 days by Southern blot analysis (Fig. 1). *E. coli* strain SMS65, carrying the *E. coli oriC*-based replicon synEsc was used as control. Total DNA from respective strains was digested with *Nco*I leading to linearization of non-integrated replicons. A potential integration would lead to a band shift of the detected DNA fragment. Such a shift was seen for a portion of cells carrying synEsc while the synVicII DNA fragment was unchanged even after three days of continuous cultivation (Fig. 1 A).

The conditions of cell growth tested here might not fully reflect the diversity of conditions that a production strain might face. To simulate more challenging conditions regarding genetic integrity we transferred synVicII to a strain with a DNA replication defect caused by a deletion of the SeqA protein (Waldminghaus and Skarstad 2009; Lu *et al.* 1994). This strain background has been shown to increase the frequency of replicon integration into the primary chromosome (Skarstad and Lobner-Olesen 2003). In fact, the control replicon synEsc was integrated throughout the population after only one day of cultivation (Fig. 1 A). In contrast, synVicII remained a separate replicon even after three days of continuous cultivation (Fig. 1 A).

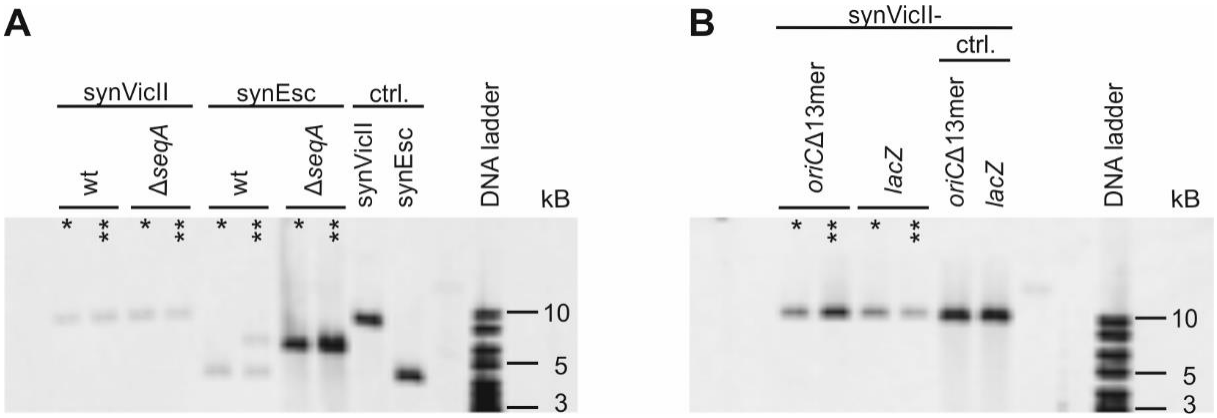


Figure 1 synVicII does not integrate into the *E. coli* chromosome. **A** *E. coli* wildtype strains carrying either synVicII- 1.3 (SMS18) or synEsc-1.31 (SMS65) and Δ*seqA* carrying synVicII-1.3 (SMS66) or synEsc1.31 (SMS67) were grown exponentially over three days in LB medium at 37°C. Genomic DNA was extracted after one (one asterisk) or three days (two asterisks) and blotted as described in Material and Methods after digestion with *Nco*I. 1 µg of DNA used for strains SMS18, SMS65 and SMS66; 500 ng from SMS 67 and 25 ng of synVicII synEsc. *Nco*I digestion linearizes synVicII and yields a 9134 bp fragment and 4378 bp for synEsc. **B** Sequences with homology to the *E. coli* chromosome were inserted into synVicII, either an inactive *oriC* (synVicII-1.301) or a part of *lacZ* (synVicII-1.302). Respective strains SMS72 and SMS74 were analyzed as above. Non-integrated linearized fragments are 9891 bp fragment for synVicII-1.301 and a 10199 bp fragment for synVicII-1.302 as seen for the control with purified replicons.

A notable difference between the two replicons compared here is that synEsc contains sequences with homology (*oriC*) to the *E. coli* primary chromosome while synVicII shares no homology with the *E. coli* chromosome. This raises the question if the different integration behavior of the replicons is simply caused by their ability to function in a homologous recombination reaction. As backbone for future biotechnology applications, synVicII would potentially carry large amounts of genetic content. This would certainly increase the chance of sequences with homology to the *E. coli* chromosome. To test if the synVicII-integration frequency is dependent on homologous sequences on the replicon, we inserted two different genetic regions that also occur on the *E. coli* chromosome. First, a synVicII version was constructed carrying 1065 bp of *lacZ* (synVicII-1.302). Second, a copy of *oriC* was inserted similar to synEsc but made inactive by deletion of 44 bp in the initiation region (synVicII-1.301).

Although these two versions of synVicII carried sequences with homology to the *E. coli* chromosome and thus being a potential target for homologous recombination, the replicons did not integrate but remained as separate entity during extended cultivation (Fig. 1 B).

Directed evolution results in stabilized versions of synVicII

Different applications of a synthetic secondary chromosome might require different characteristics. If the replicon is for example used to analyze the stabilizing effect of different genetic elements, it would be important to use a replicon which is lost from a cell population over time under non-selective conditions. On the other hand, a replicon and its genetic content would need to be stably transmitted from cell to cell if it is used in biotechnological processes. Notably, the prototype of synVicII showed a certain degree of instability (Messerschmidt *et al.* 2015). A potential way of increasing the genetic stability could be the use of directed evolution. As in natural evolution, variation and selection is a central part of directed evolution (Chatterjee and Yuan 2006; Mills *et al.* 1967). It has been applied to optimize many different cellular features as enzymes, regulatory RNA or cell characteristics (Wang *et al.* 2016; Derkx *et al.* 2014; Waldminghaus *et al.* 2008). Directed evolution is also an important tool within synthetic biology (Cobb *et al.* 2013; Kang *et al.* 2015). It has previously been used to generate plasmids with increased copy numbers (Tao *et al.* 2005). The setup for a directed evolution system to generate stabilized versions of synVicII is shown in figure 2 A. The basic idea is that *E. coli* cells carrying synVicII are cultivated without antibiotic selection. After some generations, a proportion of the population will have lost the replicon and others will still carry a synVicII copy. Versions of synVicII with a stabilizing mutation will belong to the later ones and to select for them the cells are transferred to growth medium with antibiotic selection. This alternating cultivation is carried out several times and finally individual clones are analyzed further. Replication characteristics of secondary replicons can change due to mutations on the primary chromosome (Ederth *et al.* 2002; Lopilato *et al.* 1986). Since we were interested in mutations of synVicII itself, the replicons of individual clones were isolated and retransformed into a “clean” genetic background (Fig. 2 A). Stabilizing mutations might target different mechanisms of replicon maintenance. One possibility are mutations leading to an increased copy number. A higher replicon copy number leads to increased stability because just by chance it is more likely for each daughter cell to get at least one replicon copy. In fact, amino acid changes in replication initiator proteins are frequently found to increase the replicon copy number (Fang *et al.* 1993; Wadood *et al.* 1997). This is also true for replicons based on *oriII* of *V. cholerae* similar to synVicII (Koch *et al.* 2012). Such copy-up mutations are not desirable for synVicII because one of its main features is its low copy number comparable to the primary chromosome (Messerschmidt *et al.* 2015). We developed a simple screen for copy-up

mutations by growing candidate clones in medium with different concentrations of ampicillin. The respective logic would be that a higher copy number of the replicon correlates with a higher copy number of the β-lactamase gene and consequently its higher expression. Cells carrying a replicon with a higher copy number should therefore tolerate higher amounts of the β-lactam antibiotic ampicillin. Measuring the growth of cells with synVicII or one of four different evolved versions showed very similar growth in the standard ampicillin concentration of 100 µg/ml (Fig. 2 B). In contrast, only one strain grew at an elevated ampicillin concentration of 1,500 µg/ml, suggesting that this strain carries a copy-up mutation (Fig. 2 C). Increased stability of the remaining candidates compared to the original synVicII was measured by the number of colony-forming units after 6 hours of exponential growth without selection pressure (Fig. 2 D). To verify that our reasoning of the ampicillin-growth test was correct and to further characterize the evolved synVicII versions we performed copy-number measurements by comparative genomic hybridization (CGH) on a custom made microarray. Probes on the array match the *E. coli* chromosome as well as synVicII. DNAs from exponentially growing strains carrying the potentially copy-up mutation (candidate 3) or a non-copy-up version (candidate 4) were hybridized against the hybridization control of non-replicating *E. coli* strain FSK18 as described previously (Messerschmidt *et al.* 2015). Respective fluorescence ratios were plotted relative to the chromosomal position (Fig. 2 E, F). The relative abundance of chromosomal loci diminishes exponentially with increasing distance from the origin for exponentially growing populations as seen for the primary chromosome (Fig. 2 E, F; (Sueoka and Yoshikawa 1965)). Fitted curves were used to calculate average *oriC*/*ter* ratios which were very similar for the two strains and within biological replicates (Suppl. table S5 and figure S1). In contrast, the copy number of the predicted copy-up version of synVicII was 9.5 relative to the terminus of the primary chromosome – almost three times higher than *oriC*. The predicted non-copy-up version of synVicII had a copy number of 3.5 very similar to the *oriC* copy number. We conclude that (i) the directed evolution approach introduced here is able to produce both, copy-up and non-copy-up versions of synVicII that are stabilized; (ii) the growth test with different ampicillin concentrations is a valid and simple measure of replicon copy number. Notably, the directed evolution setup introduced here should in principle be suitable for any other secondary replicon optimization.

We hypothesized that the copy-up phenotype of candidate 3 was caused by a change in the amino acid sequence of the initiator protein RctB as found previously (Koch *et al.* 2012). By sequencing, we indeed found a point mutation leading to an exchange of a serine to tyrosine at position 555. Because the position of stabilizing mutations would be more difficult to predict for the non-copy-up candidate 4, we sequenced the entire replicon. Two single point mutations were found, one in the replication origin *oriII* and one in the transcriptional terminator of the encoded *gfp* gene. To derive the contribution of each of these mutations, we introduced them individually to an otherwise

unchanged synVicII and measured respective replicon stabilities (Fig. S2). While the mutation at *gfp* slightly increased the stability compared to synVicII, the origin mutation actually decreased it. It appears, that both mutations act synergistically to increase replicon stability as found in candidate 4. Certainly, further analyses are needed to understand the molecular basics of this finding.

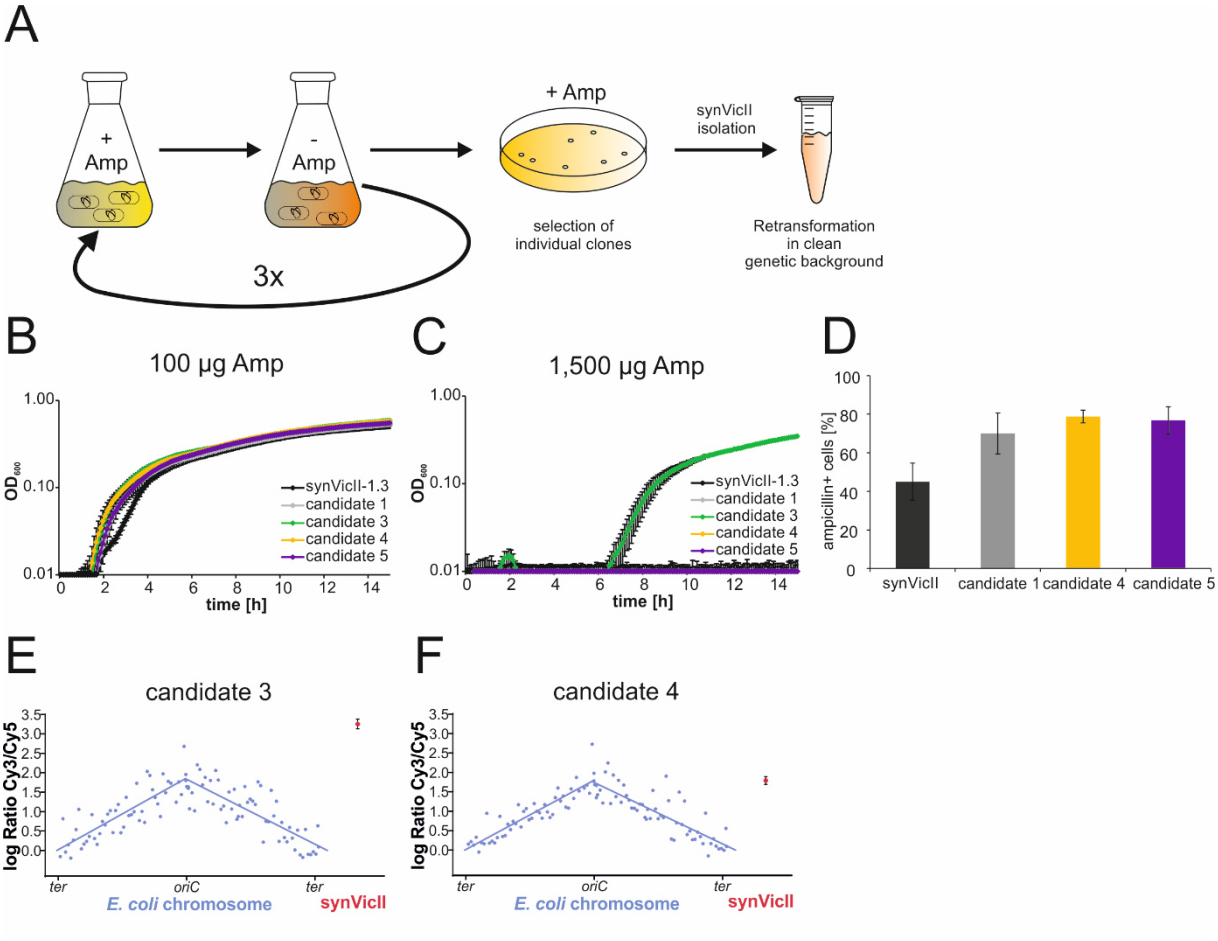


Figure 2 Directed evolution experiment reveal stabilized versions of synVicII-1.3. **A** Work-flow scheme for the identification of stabilized synVicII versions. *E. coli* strain SMS18 carrying synVicII-1.3 was grown in LB medium with ampicillin and then shifted to medium without antibiotic (see Material and Methods for details). The process was repeated for three times before cells were plated onto LB agar with ampicillin. synVicII was isolated from these candidate clones and retransformed into *E. coli* MG1655. **B and C** Growth test to distinguish low-copy and high-copy versions of evolved synVicII versions. Candidate strains were grown in LB medium with low and high ampicillin concentrations as indicated in a 96-well plate at 37°C. **D** Stability of evolved candidate versions of synVicII in comparison to synVicII-1.3. Given numbers are mean values of ampicillin resistant cells after 6 h of cultivation without antibiotic selection from three biological replicates (Messerschmidt *et al.* 2015). **E and F** Comparative genomic hybridization of *E. coli* strain SMS81 harboring synVicII-candidate 3 and SMS79 harboring synVicII-candidate 4, respectively. DNA of exponentially grown cells was hybridized against DNA of a hybridization control (see Material and Methods for details). Respective logarithmic values of ratios are plotted against their genomic position. Blue dots represent the 104 probes for the *E. coli* chromosome, blue lines the linear curve fitted to chromosome arms. The red dot represents the mean of the three probes of synVicII with the corresponding standard deviation.

Probing the origin diversity of secondary chromosomes from *Vibrionaceae*

If it is an attractive idea to have a synthetic secondary chromosome for biotechnology applications and basic research, the question occurs if it might also be interesting to have tertiary, quaternary or octonary chromosomes in addition. Spreading the genetic information to multiple replicons might actually have considerable benefits (Milbredt *et al.* 2016; Liang *et al.* 2013; Schindler and Waldminghaus 2015). Since using the replication origin of the *V. cholerae* secondary chromosome as basis for a synthetic secondary chromosome in *E. coli* has proven a suitable approach we set out to probe the origin diversity of the *Vibrio* genus for its potential as tertiary chromosome. To this end we constructed new replicons based on eight replication origins derived from different *Vibrio* species and one *Photobacterium* (Fig. 3 A). The backbone was a newer version of synVicII including an origin of transfer to allow conjugational replicon transfer (see following chapter and Material and Methods). A prerequisite of having two replicons in a cell in addition to the primary chromosome is that they are not incompatible. Incompatibility is a long known phenomenon describing the observation that a plasmid is not kept in a cell which harbors a plasmid of the same ancestry (Scaife and Gross 1962; Bouet *et al.* 2007). The molecular mechanisms underlying incompatibility can be different but are mostly related to replicon segregation and replication (Bouet *et al.* 2007). To measure the compatibility of replicons we constructed additional versions with an alternative selection marker (kanamycin instead of ampicillin). All replicons were able to replicate in *E. coli* and the growth of respective strains was relatively similar except variations of the lag-phase duration (Fig. 3 B). We performed crosswise conjugations in all possible combinations with replicons based on an F plasmid origin as positive control. Except this control, none of the pairwise combination produced a significant amount of transconjugants (data not shown). We conclude that replicons based on replication origins from secondary chromosomes of the *Vibrio* genus and *Photobacterium* all belong to the same incompatibility group and are not suited for combination in one host cell. However, they all replicate within the heterologous host *E. coli* and could be used as alternative to synVicII in principle.

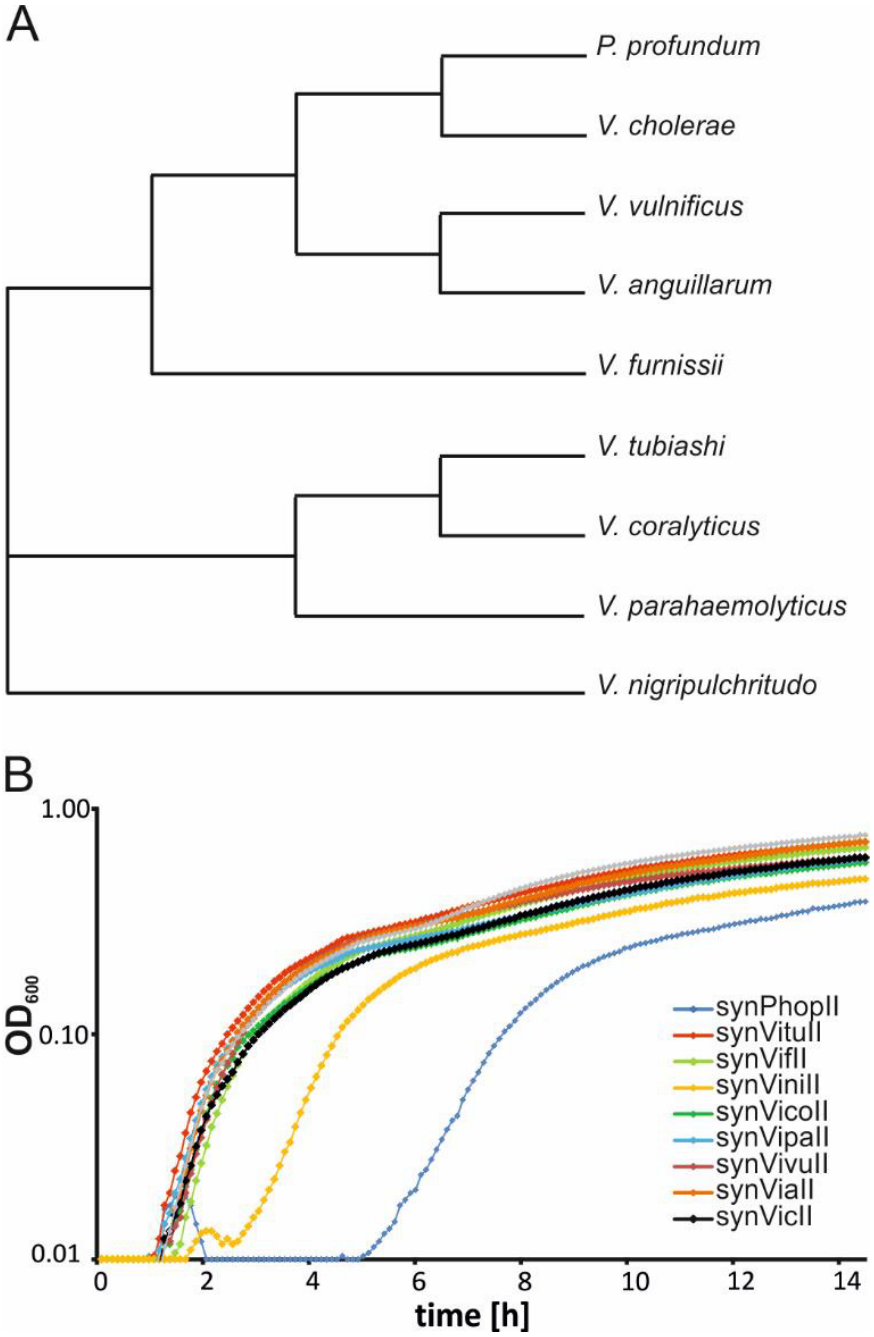


Figure 3 Probing the origin diversity of secondary chromosomes from diverse *Vibrio*. **A** Phylogenetic tree of analyzed *Vibrionaceae* species based on 16s rRNA sequences. The respective alignment was calculated with EMBOS needle (Rice *et al.* 2000). 16s rRNA sequences (the one nearest the origin) were derived from the following genomes: NC_002505.1 for *V. cholerae,* NC_006370.1 for *P. profundum*, NC_015633.1 for *V. anguillarum*, NC_016602.1 for *V. furnissii*, NC_004603.1 for *V. parahaemolyticus*; NC_005139.1 for *V. vulnificus*, NC_022528.1 for *V. nigripulchritudo*, NZ_CP009354.1 for *V. tubiashi* and NZ_CP009264.1 for *V. coralliilyticus*. **B** Growth curves of *E. coli* MG1655 strains carrying synthetic secondary chromosomes based on different *Vibionaceae* replication origins. Strains used are: SMS121 (synPhopII(ampR)), SMS101 (synVituII(ampR)), SMS102 (synVifII(ampR)), SMS106 (synViniII(ampR)), SMS107 (synVicoII(ampR)), SMS108 (synVipaII(ampR)), SMS110 (synVivuII(ampR)), SMS134 (synViaII(amp)) and NZ72 (synVicII-1.352). Cells were grown in LB medium with ampicillin OD_{600} was measured in 5 min intervals in a Victor X3 microplate reader.

New version of synVicII

Well characterized replicons are a prerequisite for solid genetic work in basic research and biotechnology. We had previously developed a prototype of the synthetic secondary chromosome synVicII (Messerschmidt *et al.* 2015). Meanwhile we have introduced several innovations as summarized in figure 4 A into a new version synVicII-2.0. A first change to the previous synVicII is the introduction of an origin of transfer (*oriT*) to allow transfer of the replicon via conjugation. This feature is especially important for larger replicons because efficiencies of isolation and transformation drop with replicon size (Gowland and Hardman 1986; Sheng *et al.* 1995). We have successfully tested the transfer of synVicII-2.0 versions with different inserts from a donor strain carrying the conjugation machinery to wildtype *E. coli* cells (data not shown). A second new feature of synVicII-2.0 is the possibility to excise a region of the replicon which is needed only for the construction process. This region includes the conditional replication origin *ori*R6K, the yeast marker and replication origin and *oriT* (Fig. 4 A). The excision is mediated by two flanking FRT recombination sites and a simple readout of successful loss of this region is possible through an inserted *mCherry* reporter gene (Fig. 4 B). Removing this construction region will limit interference with the genetic content of interest. The third change to synVicII was its conversion into a MoClo-compatible replicon. MoClo is an assembly framework based on type IIs restriction enzyme and was developed by Sylvestre Marillonnet and colleagues (Weber *et al.* 2011; Werner *et al.* 2012). The MoClo system is now widely used with still increasing popularity (Schindler *et al.* 2016; Engler *et al.* 2014; Kakui *et al.* 2015). It consists of vector sets (Level 0, level 1, level M, level P) with the 4 bp overhang of each vector matching the overhangs of the preceding and following vector, respectively. Assembling multiple fragments into an acceptor vector is possible because the resistance markers as well as the type IIs restriction sites are alternating. Assemblies of different numbers of fragments are facilitated by a set of specific endlinkers. To make the benefits of the MoClo system accessible for synVicII engineering we removed all 12 *Bpi*I and *Bsa*I restriction sites by a two-step multi fragment assembly in yeast (Fig. 4 A; see Material and Methods for details). In addition, we introduced level M or level P MoClo cassettes consisting of the suicide gene *ccdB* and the reporter *lacZ* flanked by either *Bpi*I or *Bsa*I sites (Fig. 4 A; (Schindler *et al.* 2016)). This resulted in 14 synVicII backbones with full compatibility to the MoClo system (Weber *et al.* 2011; Werner *et al.* 2012). Insertion of genetic content of interest will remove the *ccdB-lacZ* cassette generating viable white colonies. Because the synVicII backbone as well as the endlinker plasmids possess an ampicillin resistance marker, the marker of all 14 Level M and P endlinker plasmids was changed to chloramphenicol (Table S2).

To test if the new version of synVicII retains its previous characteristics we constructed an “empty” replicon by performing a MoClo reaction with synVicII-2.11 and the respective endlinker pMA657 only since the cloning cassette would permit viable wildtype cells (= synVicII-2.111, suppl. table S2).

This replicon showed a very similar stability within *E. coli* cells compared to the original synVicII-1.3 as measured by flow cytometry and a colony counting approach as previously described (Fig. 5 A and B;(Messerschmidt *et al.* 2015)). CGH analysis showed that the synVicII-2.111 copy number lays between the copy number of the replication origin and terminus of the primary chromosome as shown for the original synVicII-1.3 (Messerschmidt *et al.* 2015). We conclude that the new version of synVicII remains the previously established genetic characteristics despite the introduced changes.

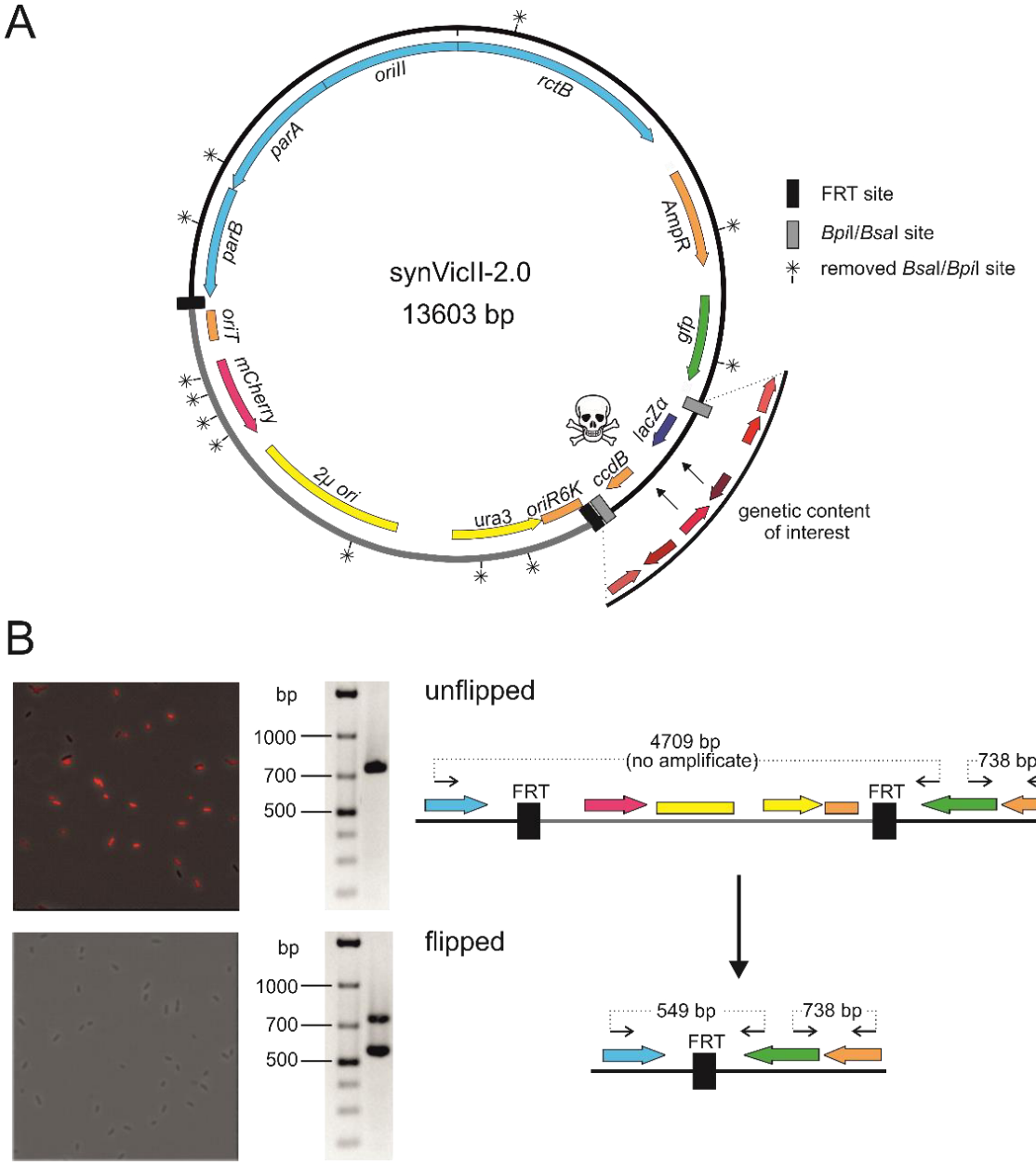


Figure 4 New features of synVicII-2.0 **A** Scheme synVicII-2.0. Genes are indicated by colored arrows and origins as blocks. In comparison to its precursor synVicII-1.3, the *I-Sce*I recognition site was displaced at a new locus between *gfp* and *oriR6K* (for construction see materials and methods). *Bsa*I/*Bpi*I recognition sites were removed rendering it suitable for Modular Cloning (MoClo). Insertion of *oriT* allows conjugal transfer of synVicII. For extension of the synVicII backbone, a *lacZα-ccdB* cassette flanked by *Bsa*I or *Bpi*I recognition sites was inserted to use blue/white screening and *ccdB* toxicity in standard *E. coli* strains for efficient detection of recombinant DNA (Bernard and Couturier 1992). Flanking of the region only needed for construction purposes by FRT sites allows removal by flippase-based site specific recombination (Cherepanov and Wackernagel 1995). **B** The red fluorescence reporter under the control of the P_{lac} promoter allows easy readout of successful recombinations as shown by fluorescence microscopy and PCR analyis of unflipped (top pannel) and flipped (bottom pannel) as illustrated in the right pannel. Data shown are for strain NZ67 carrying synVicII-1.34 transformed with pcp20 at 30°C which activates the FRT recombination. Upon heat induction, the heat sensitive replicon pCP20 got lost. Colony PCR with primers (back arrows in right pannel) 716/24 and 858/25 also confirmed successful flipping. Elongation time was short enough to allow amplification of max. 800 bp and expected fragment sizes are indicated in the right pannel.

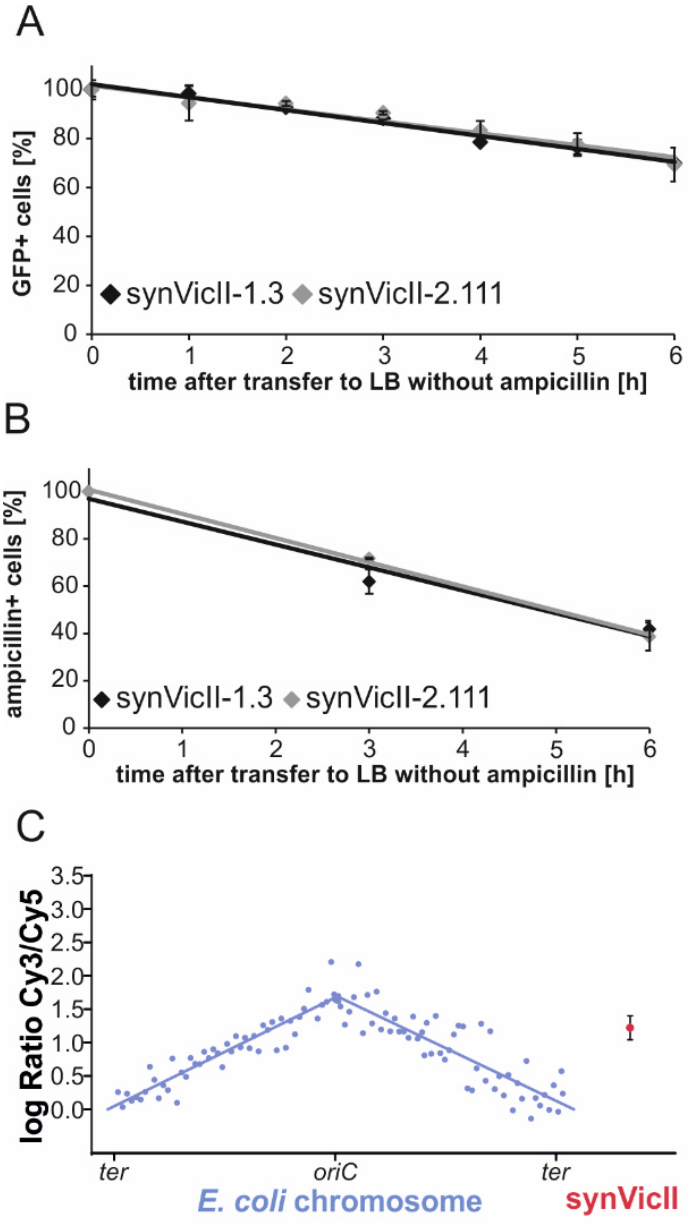


Figure 5 Conserved genetic characteristics of new synVicII. **A** Stability of synVicII-1.3 (back) and synVicII-2.111 (grey) in *E. coli* MG1655 determined by flow cytometry measured as described (Messerschmidt *et al.* 2015). Strains carrying synVicII-1.3 (SMS18) or synVicII-2.111 (DS292) were grown in LB supplemented with ampicillin and 0.2 mM IPTG at 37°C to an OD_{600} of about 0.15, transferred to ampicillin-free medium and cultured for 6 h. Cultures were kept in exponential phase by diluting into fresh pre-warmed medium at an OD_{600} higher than 1.2. Samples were taken every 1 h and *gfp* fluorescence as proxy for replicon resence measured by flow cytometry. **B** Stability of synVicII-1.3 (black) compared to synVicII-2.111 (grey) in *E. coli* MG1655 measured by counting ampicilin resistant colonies after transfer to medium without antibiotic selection at indicated time points (Messerschmidt *et al.* 2015). Results are from three biological replicates with a total of 300 colonies per time point and strain. **C.** Comparative genomic hybridization of *E. coli* strain DS292 harboring synVicII-2.111. DNA of exponentially grown DS292 was hybridized against a control with fully replicated chromosomes (see Material and Methods for details). Respective logarithmic values of ratios are plotted against their genomic position. Blue dots represent the 104 probes for the *E. coli* chromosome, blue lines the linear curve fitted to chromosome arms. The red dot represents the mean of the three probes of synVicII with the corresponding standard deviation.

Materials and Methods

Bacterial strains, plasmids, oligonucleotides, and culture conditions

Strains, plasmids and oligonucleotides are listed in table S1-S3. Pre-cultures of *E. coli* were grown in 3 ml LB medium. Antibiotics and inductors were used with the following concentrations if not indicated otherwise: ampicillin (100 µg/µl), kanamycin (100 µg/µl) and IPTG (100 µg/µl). Cultures of *S. cerevisiae* were as described previously (Messerschmidt *et al.* 2015).

Construction of secondary chromosomes and plasmids

All replicons were constructed by Gibson assembly and reactions transformed into *E. coli* XL-1 Blue or *E. coli* DH5*α λ pir* if not indicated otherwise (Gibson *et al.* 2009). The previously published synVicII-1.3 was changed stepwise towards synVicII-2.0 as follows. A *Xho*I recognition site was inserted by Gibson assembly of a PCR product with primers 327/328 from synVicII-1.5 (Messerschmidt *et al.* 2015) and *Nru*I-digested synVicII-1.3, resulting in synVicII-1.31. For the construction of synVicII-1.32, the *I-Sce*I site in synVicII-1.31 was replaced with a *Pvu*II site. For that purpose *I-Sce*I digested synVicII-1.31 and oligonucleotide 814 were assembled by homologous recombination in *S. cerevisiae* strain pJ69-4a as described (Colot *et al.* 2006; Gietz and Schiestl 2007). For the construction of synVicII-1.311, the RhaT promotor was amplified with primers 329/330 from pWBT5 (Schluter *et al.* 2015). The forward primer has an overhang adding one FRT site. The fragment was integrated into XhoI-cut synVicII-1.31. To generate synVicII-1.312, *rfp* was amplified with primers 331 and 597 from pSB1C3 J04450 (iGEM). The reverse primer has an overhang adding one FRT site. The PCR product was then assembled with *Nhe*I-digested synVicII-1.311.

To generate synVicII-1.33, the FRT site with *Sma*I and the FRT site with *I-Sce*I recognition site were amplified with primers 817/818 and 815/816 from synVicII-1.312 and integrated into I-SceI-cut synVicII-1.32. synVicII-1.34 was made by amplification of *rfp* with primers 819/820 from pSB1C3J04450 (iGEM) and integration into *Sma*I-digested synVicII-1.33. synVicII-1.35 was constructed by combining PCR-amplified *oriT* (primers 874 and 875 from pUC18-R6KT-egfp) and *Sma*I-digested synVicII-1.34. To construct synVicII-1.36, *lacZ* and *ccdB* were amplified with oligonucleotides 1002 and 1005 from pMA58 (Schindler *et al.* 2016). Genes were integrated into *I-Sce*I digested synVicII-1.35.

The mutation of *Bpi*I and *Bsa*I recognition sites within synVicII-1.36 was made by cutting the replicon with one of the enzymes and transforming the fragments into yeast together with bridging DNA fragments changing the respective sites. Bridging DNA for the mutation of four *Bpi*I and two *Bsa*I recognition sites was generated by designing pairs of 60 bp oligonucleotides with 20 bp annealing region (primer pairs: 1628-1635 & 1638-1641). The resulting 100 bp DNA fragments were generated by a 3 cycle PCR with respective primer pairs. Additional four *Bpi*I sites were deleted by replacing *rfp* with an optimized *mCherry* amplified with primer pair 1636 & 1637 from pMA17. pMA17 was generated by a MoClo Reaction into pICH41276 using two PCR products to remove a recognition site with primer pairs 69 & 70 and 71 & 72 from template pWBT5^{mCherry}. One *Bsa*I site was mutated by amplifying *bla* from pMA53 with primer 214 & 215. synVicII-1.36 was cut with *Bpi*I or *Bsa*I and transformed together with the corresponding DNA parts into *S. cerevisiae* VL6-48N to produce either *Bpi*I recognition site free synVicII-1.361 or *Bsa*I recognition site free synVicII-1.362 by *in vivo* homologous recombination. For each construct *S. cerevisiae* colonies were pooled, cultivated in

50 ml SD-ura and plasmid DNA extracted. The plasmid DNA was digested with *Bpi*I or *Bsa*I to remove false positives and subsequently transformed into *E. coli* DB3.1 λpir. Positive clones were verified by restriction analysis. DNA of synVicII-1.361 and synVicII-1.362 was pooled in equimolar concentration, digested with *Bpi*I and *Bsa*I and transformed into *S. cerevisiae* to generate *Bpi*I and *Bsa*I recognition site free synVicII-1.37. In order to generate MoClo compatible level M and level P backbones synVicII-1.37 was cut with *Not*I and transformed in 14 reactions with the respective 7 Level M and 7 Level P MoClo cassettes to generate synVicII-2.01 to synVicII-2.07 and synVicII-2.11 to synVicII-2.17, respectively. Corresponding Level M and P MoClo cassettes were amplified using primer 1029 & 1030 and templates pMA60-pMA66 respectively primer 1031 & 1032 and templates pMA67-pMA73 (Schindler *et al.* 2016).

The existing MoClo endlinker of the Marillonnet group possess *bla* and interfere with the MoClo synVic2.0 backbones. Therefore, the *bla* gene was exchanged with *cat*. To this end, Level M and P endlinker plasmids were amplified with primer pair 582 and 1099 and the *cat* gene with primer pair 581 and 1100 from pMA44 (Daniel Schindler, unpublished) resulting in plasmids pMA667-680.

synVicII-1.301 was constructed by inserting PCR-amplified *oriC* without one of the 13mers from gDNA of strain SMS18 into *I-Sce*I digested synVicII-1.3. To generate synVicII-1.302, part of *lacZ* was amplified with primers 876 and 877 from gDNA of strain SMS18 and integrated into *I-Sce*I digested synVicII-1.3.

synVicII-0.11 was constructed by assembling *gfp*-AAV amplified with primers 28/29 from synVicII-1.3 with *I-Sce*I-digested synVic-0.1 by homologous recombination in yeast. For construction of synVicII-1.313, the *oriII* of the directed evolution candidate 4 (synVicII-1.8) was amplified with primers 14/16 and assembled with *Not*I-digested synVicII-0.11 by homologous recombination in yeast. The *gfp* from synVicII-1.8 was amplified with primers 26/27 and integrated into *I-Sce*I-digested synVicII-1.0 by Gibson assembly resulting in synVicII-1.314.

For construction of synthetic secondary chromosomes based on different Vibrio genomes, *oriII*s with *parAB* and *rctB* were amplified from gDNA of the respective strain. gDNA was isolated with the phenol-chloroform method as described in (Schindler *et al.* 2016). To facilitate origin cloning, the *oriII* in synVicII-1.35 was replaced with *lacZα*. For this construction *lacZα* was amplified with primers 1132/1133 and assembled with *Eco*RI-*Sal*I-digested synVicII-1.35 by yeast homologous recombination. All oligonucleotides for *oriII* cloning have fitting overhangs to the neighboring fragment in the backbone synVicII-1.351 (at least 25 bp) allowing the construction with Gibson assembly and add an *Asc*I site to allow *oriII* release.

To construct synPhopII(AmpR), the *oriII* region was amplified with primers 1148/1149 from gDNA of *P. profundum* and assembled with *Asc*I-digested synVicII-1.351. synVituII(AmpR), synVifII(AmpR), synViniII(AmpR), synVicoII(AmpR), synVipaII(AmpR), synVivuII(AmpR), and synViaII(AmpR) were constructed accordingly with respective primers and templates listed in supplementary table S4.

To exchange the ampicillin resistance marker in synVicII-1.3 with a kanamycin resistance marker, *kan* was amplified with primer 30/31 from pUC57 and assembled with *Bgl*I-digested synVicII-1.3 by homologous recombination in yeast to generate synVicII-1.7.

For the construction of other kanamycin resistant replicons, the backbone synVicII-0.3 with was generated by relegation of *Asc*I-digested synVicII-1.351 (= synVicII-1.3511) followed by cutting with *Bgl*I. This linearized fragment was assembled with the kanamycin cassette amplified with primers

1435/31 from synVicII-1.7 by homologous recombination in yeast as described above. For synPhopII(*kan*), *Asc*I-digested synVicII-0.3 was ligated with the *Asc*I-digested *oriII* part of synPhopII(AmpR), the corresponding ampicillin resistant replicon. synVituII(*kan*), synVifII(*kan*), synViniII(*kan*), synVicoII(*kan*), synVipaII(*kan*), synVivuII(*kan*), and synViaII(*kan*) were constructed accordingly. synF-2.0 was constructed the same way with F *ori* amplification from synF-plasmid with primers 1487/1488.

Directed evolution experiments

E. coli strain SMS18 carrying synVicII-1.3 was grown overnight in LB medium with ampicillin and was then 1:1,000 diluted in LB medium without antibiotics. After 8 h of growth cells were transferred 1:10,000 into LB with ampicillin and grown overnight. The procedure was repeated for three days and finally 100 µl of culture plated on selective plates. Replicons of individual clones were isolated and retransformed into *E. coli* MG1655. Replicon stability was measured as before (Messerschmidt *et al.* 2015). To select for copy-up and non-copy-up mutants, candidates were grown in LB medium with either 100 or 1,500 µg/ml ampicillin in a 96-well plate in a microplate reader (Victor X3 Multilabel Plate Reader, PerkinElmer) at 37°C. The 150 µl of main culture was inoculated 1:1,000, covered with 70 µl of mineral oil and growth curves recorded for 14.5 hrs.

Comparative genomic hybridization

Microarray construction, sample preparation, hybridization and data processing were essentially performed as described (Messerschmidt *et al.* 2015). Instead of harvesting cells in stationary phase for a hybridization control, exponentially growing cells of strain FSK18 were treated with 150 µg/ml Rifampicin for 2 h. Lysed cells were treated with 60 µg/ml RNaseA for 1 h at 65°C before DNA isolation with phenol-chloroform and ethanol precipitation.

Southern Blot experiments

For Southern hybridization, genomic DNA was extracted from 1.5 ml culture at an OD_{600} of 0.3 as described with the following minor changes (Skarstad and Lobner-Olesen 2003). Treatment in the DNA-isolation buffer was performed at 4°C and the RNaseA incubation was for 1 h. After phenol-chloroform extraction DNA was precipitated with ethanol and Na-acetate. For blotting, usually 1 µg of *Nco*I-digested chromosomal DNA was separated on 1 % agarose gels and transferred by vacuum blotting to an Amersham Hybond-N membrane (GE Healthcare, Chalfont St Giles). Exceptions with other amount of DNA are mentioned in the figure legends (SMS67 and the replicon controls synVicII-1.3/synEsc-1.3). DNA was detected with a DIG labelled AmpR probe (PCR DIG Probe Synthesis Kit, Roche, Penzberg) as PCR product from primers 793/794 with synVicII-1.3 as a template.

Acknowledgements

We thank all members of the Waldminghaus lab for help and fruitful discussions. We are grateful to Vladimir Larionov, Sylvestre Marillonnet, Michael L. Kahn and Matthew McIntosh for providing strains or plasmids and we thank the Flow Cytometry and Genomics Core Facility (ZTI, Marburg) for providing respective devices. Jörg Overmann and Elke Lang are acknowledged for sharing *Vibrio* strains and help with growing them. This work was supported within the LOEWE program of the State of Hesse and a grant of the Deutsche Forschungsgemeinschaft (Grant No. WA 2713/4-1).

References

Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, Yeluru A, Murugan S, Kandavelou K, Agier N, Fischer G, Yang K, Martin JA, Bilgel M, Bohutski P, Boulier KM, Capaldo BJ, Chang J, Charoen K, Choi WJ, Deng P, DiCarlo JE, Doong J, Dunn J, Feinberg JI, Fernandez C, Floria CE, Gladowski D, Hadidi P, Ishizuka I, Jabbari J, Lau CY, Lee PA, Li S, Lin D, Linder ME, Ling J, Liu J, London M, Ma H, Mao J, McDade JE, McMillan A, Moore AM, Oh WC, Ouyang Y, Patel R, Paul M, Paulsen LC, Qiu J, Rhee A, Rubashkin MG, Soh IY, Sotuyo NE, Srinivas V, Suarez A, Wong A, Wong R, Xie WR, Xu Y, Yu AT, Koszul R, Bader JS, Boeke JD, Chandrasegaran S (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344 (6179):55-58. doi:10.1126/science.1249252

Bernard P, Couturier M (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol 226 (3):735-745

Birchler JA (2015) Promises and pitfalls of synthetic chromosomes in plants. Trends Biotechnol 33 (3):189-194. doi:10.1016/j.tibtech.2014.12.010

Bouet JY, Nordstrom K, Lane D (2007) Plasmid partition and incompatibility--the focus shifts. Mol Microbiol 65 (6):1405-1414. doi:10.1111/j.1365-2958.2007.05882.x

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

Chatterjee R, Yuan L (2006) Directed evolution of metabolic pathways. Trends Biotechnol 24 (1):28-38. doi:10.1016/j.tibtech.2005.11.002

Cherepanov PP, Wackernagel W (1995) Gene disruption in *Escherichia coli:* TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158 (1):9-14

Cobb RE, Sun N, Zhao H (2013) Directed evolution as a powerful synthetic biology tool. Methods 60 (1):81-90. doi:10.1016/j.ymeth.2012.03.009

Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A 103 (27):10352-10357. doi:10.1073/pnas.0601456103

Derkx PM, Janzen T, Sorensen KI, Christensen JE, Stuer-Lauridsen B, Johansen E (2014) The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. Microb Cell Fact 13 Suppl 1:S5. doi:10.1186/1475-2859-13-s1-s5

Duigou S, Knudsen KG, Skovgaard O, Egan ES, Lobner-Olesen A, Waldor MK (2006) Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. J Bacteriol 188 (17):6419-6424. doi:10.1128/JB.00565-06

Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S, Bader JS, Boeke JD

(2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477 (7365):471-476. doi:10.1038/nature10403

Ederth J, Isaksson LA, Abdulkarim F (2002) Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the *Escherichia coli* RNA polymerase. Mol Genet Genomics 267 (5):587-592. doi:10.1007/s00438-002-0689-y

Egan ES, Lobner-Olesen A, Waldor MK (2004) Synchronous replication initiation of the two *Vibrio cholerae* chromosomes. Curr Biol 14 (13):R501-502. doi:10.1016/j.cub.2004.06.036

Egan ES, Waldor MK (2003) Distinct replication requirements for the two *Vibrio cholerae* chromosomes. Cell 114 (4):521-530

Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JD, Patron NJ, Marillonnet S (2014) A golden gate modular cloning toolbox for plants. ACS Synth Biol 3 (11):839-843. doi:10.1021/sb4001504

Fang FC, Durland RH, Helinski DR (1993) Mutations in the gene encoding the replication-initiation protein of plasmid RK2 produce elevated copy numbers of RK2 derivatives in *Escherichia coli* and distantly related bacteria. Gene 133 (1):1-8

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, 3rd, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319 (5867):1215-1220. doi:10.1126/science.1151721

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329 (5987):52-56. doi:10.1126/science.1190719

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6 (5):343-345. doi:10.1038/nmeth.1318

Gietz RD, Schiestl RH (2007) Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2 (1):1-4. doi:10.1038/nprot.2007.17

Gowland PC, Hardman DJ (1986) Methods for isolating large bacterial plasmids. Microbiol Sci 3 (8):252-254

Haldimann A, Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol 183 (21):6384-6393. doi:10.1128/jb.183.21.6384-6393.2001

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406 (6795):477-483. doi:10.1038/35020000

Hutchison CA, 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC (2016) Design and synthesis of a minimal bacterial genome. Science 351 (6280):aad6253. doi:10.1126/science.aad6253

Kakui Y, Sunaga T, Arai K, Dodgson J, Ji L, Csikasz-Nagy A, Carazo-Salas R, Sato M (2015) Module-based construction of plasmids for chromosomal integration of the fission yeast *Schizosaccharomyces pombe*. Open Biol 5 (6):150054

Kang Z, Zhang J, Jin P, Yang S (2015) Directed evolution combined with synthetic biology strategies expedite semi-rational engineering of genes and genomes. Bioengineered 6 (3):136-140. doi:10.1080/21655979.2015.1011029

Koch B, Ma X, Lobner-Olesen A (2012) *rctB* mutations that increase copy number of *Vibrio cholerae oriCII* in *Escherichia coli*. Plasmid 68 (3):159-169. doi:10.1016/j.plasmid.2012.03.003

Kosuri S, Church GM (2014) Large-scale de novo DNA synthesis: technologies and applications. Nat Methods 11 (5):499-507. doi:10.1038/nmeth.2918

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342 (6156):357-360. doi:10.1126/science.1241459

Lee BR, Cho S, Song Y, Kim SC, Cho BK (2013) Emerging tools for synthetic genome design. Mol Cells 35 (5):359-370. doi:10.1007/s10059-013-0127-5

Liang X, Baek CH, Katzen F (2013) *Escherichia coli* with Two Linear Chromosomes. ACS Synth Biol. doi:10.1021/sb400079u

Lobner-Olesen A (1999) Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control. EMBO J 18 (6):1712-1721. doi:10.1093/emboj/18.6.1712

Lopilato J, Bortner S, Beckwith J (1986) Mutations in a new chromosomal gene of *Escherichia coli* K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives. Mol Gen Genet 205 (2):285-290

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in *E. coli*. Cell 77 (3):413-426

Messerschmidt SJ, Kemter FS, Schindler D, Waldminghaus T (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. Biotechnol J 10 (2):302-314. doi:10.1002/biot.201400031

Milbredt S, Farmani N, Sobetzko P, Waldminghaus T (2016) DNA Replication in Engineered *Escherichia coli* Genomes with Extra Replication Origins. ACS Synth Biol. doi:10.1021/acssynbio.6b00064

Mills DR, Peterson RL, Spiegelman S (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc Natl Acad Sci U S A 58 (1):217-224

Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR (2005) Stable barley chromosomes without centromeric repeats. Proc Natl Acad Sci U S A 102 (28):9842-9847. doi:10.1073/pnas.0504235102

Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G, Gonzales A, Wanner BL, Norville JE, Lajoie MJ, Church GM (2016) Design, synthesis, and testing toward a 57-codon genome. Science 353 (6301):819-822. doi:10.1126/science.aaf3639

Pennisi E (2014) Building the ultimate yeast genome. Science 343 (6178):1426-1429. doi:10.1126/science.343.6178.1426

Rasmussen T, Jensen RB, Skovgaard O (2007) The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. EMBO J 26 (13):3124-3131. doi:10.1038/sj.emboj.7601747

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16 (6):276-277

Scaife J, Gross JD (1962) Inhibition of multiplication of an Flac factor in Hfr cells of *Escherichia coli* K-12. Biochem Biophys Res Commun 7:403-407

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synth Biol. doi:10.1021/acssynbio.6b00089

Schindler D, Waldminghaus T (2015) Synthetic chromosomes. FEMS Microbiol Rev 39 (6):871-891. doi:10.1093/femsre/fuv030

Schluter JP, Czuppon P, Schauer O, Pfaffelhuber P, McIntosh M, Becker A (2015) Classification of phenotypic subpopulations in isogenic bacterial cultures by triple promoter probing at single cell level. J Biotechnol 198:3-14

Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res 23 (11):1990-1996

Skarstad K, Lobner-Olesen A (2003) Stable co-existence of separate replicons in *Escherichia coli* is dependent on once-per-cell-cycle initiation. EMBO J 22 (1):140-150. doi:10.1093/emboj/cdg003

Stokke C, Waldminghaus T, Skarstad K (2011) Replication patterns and organization of replication forks in *Vibrio cholerae*. Microbiology 157 (Pt 3):695-708. doi:10.1099/mic.0.045112-0

Sueoka N, Yoshikawa H (1965) The chromosome of *Bacillus subtilis.* I. Theory of marker frequency analysis. Genetics 52 (4):747-757

Tao L, Jackson RE, Cheng Q (2005) Directed evolution of copy number of a broad host range plasmid for metabolic engineering. Metab Eng 7 (1):10-17

Val ME, Marbouty M, de Lemos Martins F, Kennedy SP, Kemble H, Bland MJ, Possoz C, Koszul R, Skovgaard O, Mazel D (2016) A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci Adv 2 (4):e1501914. doi:10.1126/sciadv.1501914

Venkova-Canova T, Baek JH, Fitzgerald PC, Blokesch M, Chattoraj DK (2013) Evidence for two different regulatory mechanisms linking replication and segregation of *Vibrio cholerae* chromosome II. PLoS Genet 9 (6):e1003579. doi:10.1371/journal.pgen.1003579

Wadood A, Dohmoto M, Sugiura S, Yamaguchi K (1997) Characterization of copy number mutants of plasmid pSC101. J Gen Appl Microbiol 43 (6):309-316

Waldminghaus T, Kortmann J, Gesing S, Narberhaus F (2008) Generation of synthetic RNA-based thermosensors. Biol Chem 389 (10):1319-1326. doi:10.1515/BC.2008.150

Waldminghaus T, Skarstad K (2009) The *Escherichia coli* SeqA protein. Plasmid 61 (3):141-150. doi:10.1016/j.plasmid.2009.02.004

Wang X, Wang G, Li X, Fu J, Chen T, Wang Z, Zhao X (2016) Directed evolution of adenylosuccinate synthetase from *Bacillus subtilis* and its application in metabolic engineering. J Biotechnol 231:115-121. doi:10.1016/j.jbiotec.2016.05.032

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6 (2):e16765. doi:10.1371/journal.pone.0016765

Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioeng Bugs 3 (1):38-43. doi:10.4161/bbug.3.1.18223

Yamaichi Y, Fogel MA, Waldor MK (2007) *par* genes and the pathology of chromosome loss in *Vibrio cholerae*. Proc Natl Acad Sci U S A 104 (2):630-635. doi:10.1073/pnas.0608341104

Yu W, Han F, Gao Z, Vega JM, Birchler JA (2007) Construction and behavior of engineered minichromosomes in maize. Proc Natl Acad Sci U S A 104 (21):8924-8929. doi:10.1073/pnas.0700932104

Zhou J, Wu R, Xue X, Qin Z (2016) CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. Nucleic Acids Res 44 (14):e124. doi:10.1093/nar/gkw475

Supporting Information

Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*

Sonja Messerschmidt, Daniel Schindler, Celine Mara Zumkeller, Franziska Kemter, Nadine Schallopp and Torsten Waldminghaus

Figure S1

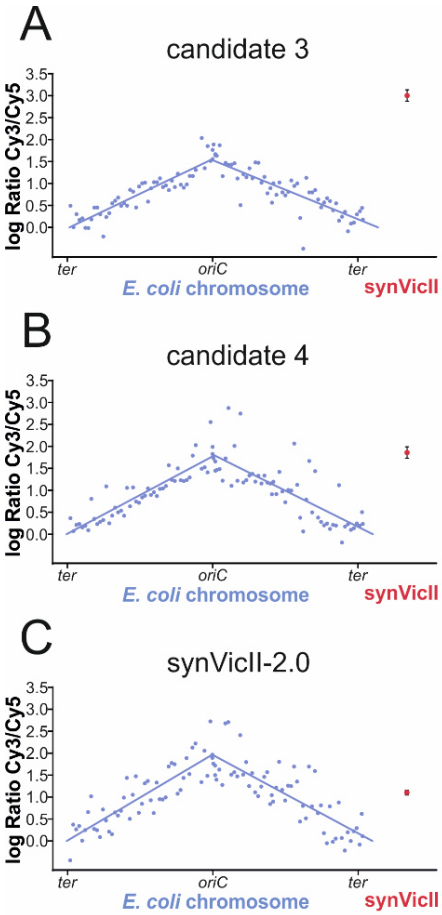


Figure S1 Biological replicates of CGH analyses of *E. coli* strains harboring different versions of synVicII. DNA of exponentially grown cells was hybridized against DNA from rifampicin-runout cells of strain FSK18 (see Materials and Methods for details). Respective logarithmic values of ratios are plotted against their genomic position. Blue dots represent the 104 probes for the *E. coli* chromosome, blue lines the linear curve fitted to chromosome arms. The red dot represents the mean of the three probes of synVicII with the corresponding standard deviation. **A** Strain SMS81 harboring synVicII candidate 3 as in Fig. 2E. **B** Strain SMS79 harboring synVicII candidate 4 as in Fig. 2F. **C** Strain DS292 harboring synVicII-2.111 as in Fig. 5C.

Figure S2

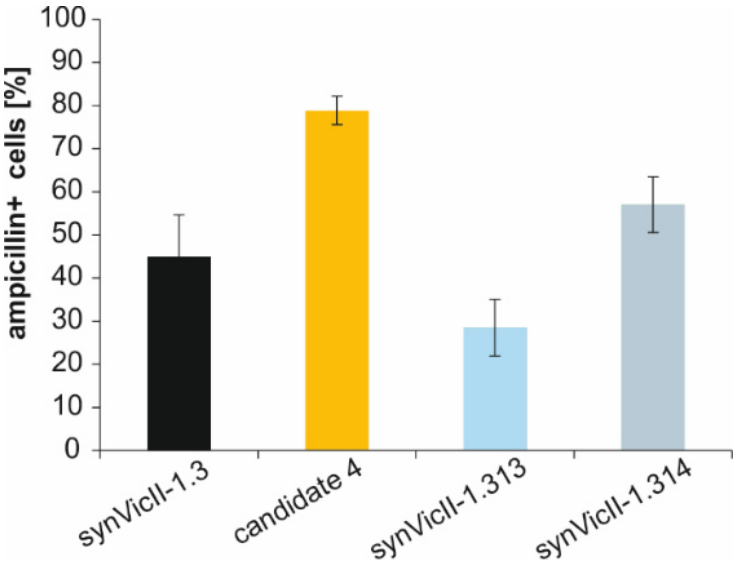


Figure S2 Stability of synVicII versions carrying one of the point mutations found in evolution candidate 4. Given numbers are mean values of ampicillin resistant cells after 6 h of cultivation without antibiotic selection from three biological replicates. Values for synVicII-1.3 and candidate 4 are taken from Fig. 2D. synVicII-1.313 carries a mutation in the replication origin and synVicII-1.314 in the transcriptional terminator of the *gfp* gene.

Table S1: Strains used in this work

Strain	Relevant genotype	Reference
E. coli DH5*α λ pir*	*supE*44, Δ*lacU*169 (Φ*lacZ*ΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λ *pir* phage lysogen	(Miller and Mekalanos 1988)
E. coli DB3.1 λpir	F- *gyrA*462 *endA*1 *glnV*44 Δ(sr1-*recA*) *mcrB mrr* *hsdS*20(rB-, mB-) *ara*14 *galK*2 *lacY*1 *proA*2 *rpsL*20 *xyl*5 Δ*leu mtl*1, λ *pir* phage lysogen	(House *et al.* 2004)
E. coli XL1 Blue	*supE*44, *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi1, relA*1, *lac*⁻ [F' *proAB, lac*^{q}ZM15, Tn10(tet^{r})]	Stratagene, Agilent Technologies, Germany
E. coli MG1655	wild type	(Blattner *et al.* 1997)
E. coli WM3064	Donor strain for conjugation: *thrB1004 pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)*567* Δ*dapA1341*::[*erm pir*(wt)]	William Metcalf
V. cholerae El Tor N16961	StrR	(Heidelberg *et al.* 2000)
V. tubiashii		DSMZ Braunschweig No. 19142
V. nigripulchritudo		DSMZ No. Braunschweig 21607
V. corallilyticus		DSMZ No. Braunschweig 19607
V. anguillarum		DSMZ Braunschweig No.

Strain	Relevant genotype	Reference
V. parahaemolyticus		21597
V. furnissii		DSMZ Braunschweig No. 10027
V. vulnificus		DSMZ Braunschweig No. 14383
Photobacterium profundum		DSMZ Braunschweig No. 10143
S. cerevisiae PJ69-4a	MATa, *trp*1-901, *leu*2-3, *leu*2-112, *ura*-3-52, *his*3-200, *gal4*Δ, *gal80*Δ, *gal*2-*ade*2, *lys*2::*gal*1*his3*, *met*2::*gal*7-	DSMZ Braunschweig No. 21095
S. cerevisiae VL6-48N	MATα *trp*1-Δ1 *ura*3-Δ1 *ade*2-101 *his*3-Δ200 l*ys*2 *met*14 cir°	(James *et al.* 1996)
DS292	*E. coli* MG1655 synVicII-2.111	(Kouprina *et al.* 1998)
SMS18	*E. coli* MG1655 synVicII-1.3	This work
SMS65	*E. coli* MG1655 synEsc-1.31	(Messerschmidt *et al.* 2015)
SMS66	*E. coli* MG1655 Δ*seqA* synVicII-1.3	This work
SMS67	*E. coli* MG1655 Δ*seqA* synEsc-1.31	This work
SMS72	*E. coli* MG1655 synVicII-1.301	This work
SMS74	*E. coli* MG1655 synVicII-1.302	This work
SMS121	*E. coli* MG1655 synPhopII(amp)	This work
SMS101	*E. coli* MG1655 synVituII(amp)	This work
SMS102	*E. coli* MG1655 synVifII(amp)	This work
SMS106	*E. coli* MG1655 synViniII(amp)	This work
SMS107	*E. coli* MG1655 synVicoII(amp)	This work
SMS108	*E. coli* MG1655 synVipaII(amp)	This work
SMS110	*E. coli* MG1655 synVivuII(amp)	This work
SMS134	*E. coli* MG1655 synViaII(amp)	This work
NZ72	*E. coli* MG1655 synVicII-1.352	This work

Table S2: Plasmids used in this work

Plasmid	Relevant characteristics[a)]	Reference
pICH41276	MoClo Level 0 Plasmid	(Weber *et al.* 2011)
pICH50872	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50881	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50892	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50900	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50914	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50927	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH50932	MoClo Level M endlinker	(Weber *et al.* 2011)
pICH79255	MoClo Level P endlinker	(Weber *et al.* 2011)
pICH79264	MoClo Level P endlinker	(Weber *et al.* 2011)
pICH79277	MoClo Level P endlinker	(Weber *et al.* 2011)
pICH79289	MoClo Level P endlinker	(Weber *et al.* 2011)

Plasmid	Relevant characteristics[a)]	Reference
pICH79290	MoClo Level P endlinker	(Weber *et al.* 2011)
pICH79300	MoClo Level P endlinker	(Weber *et al.* 2011)
pICH79311	MoClo Level P endlinker	(Weber *et al.* 2011)
pMA17	pICH41276 *mCherry*	This work
pMA53	Level 1 MoClo plasmid	(Schindler *et al.* 2016)
pMA58	Level 1 MoClo Plasmid	(Schindler *et al.* 2016)
pMA60	MoClo Level M 1	(Schindler *et al.* 2016)
pMA61	MoClo Level M 2	(Schindler *et al.* 2016)
pMA62	MoClo Level M 3	(Schindler *et al.* 2016)
pMA63	MoClo Level M 4	(Schindler *et al.* 2016)
pMA64	MoClo Level M 5	(Schindler *et al.* 2016)
pMA65	MoClo Level M 6	(Schindler *et al.* 2016)
pMA66	MoClo Level M 7	(Schindler *et al.* 2016)
pMA67	MoClo Level P 1	(Schindler *et al.* 2016)
pMA68	MoClo Level P 2	(Schindler *et al.* 2016)
pMA69	MoClo Level P 3	(Schindler *et al.* 2016)
pMA70	MoClo Level P 4	(Schindler *et al.* 2016)
pMA71	MoClo Level P 5	(Schindler *et al.* 2016)
pMA72	MoClo Level P 6	(Schindler *et al.* 2016)
pMA73	MoClo Level P 7	(Schindler *et al.* 2016)
pMA667	MoClo Level M endlinker	This work
pMA668	MoClo Level M endlinker	This work
pMA669	MoClo Level M endlinker	This work
pMA670	MoClo Level M endlinker	This work
pMA671	MoClo Level M endlinker	This work
pMA672	MoClo Level M endlinker	This work
pMA673	MoClo Level M endlinker	This work
pMA674	MoClo Level P endlinker	This work
pMA675	MoClo Level P endlinker	This work
pMA676	MoClo Level P endlinker	This work
pMA677	MoClo Level P endlinker	This work
pMA678	MoClo Level P endlinker	This work
pMA679	MoClo Level P endlinker	This work
pMA680	MoClo Level P endlinker	This work
pUC57kan	Cloning vector, *kan*	GenScript, Piscataway Township, NJ, USA
pWBT5^{mCherry}	P_{rhaT}	(Schlüter *et al.* 2015)
synEsc-1.3	synX-0.1 + *oriC* (coord. 3923616-3924064)	(Messerschmidt *et al.* 2015)
synEsc-1.31	synEsc-1.3 with *mioC*	This work
synF-2.0	synVicII-0.3+ F $origin_{synF\text{-}plasmid}$	This work
synF-plasmid	synX-0.3 with F *origin*, *sopABC*	(Messerschmidt *et al.* 2015)
synPhop(AmpR)	synVicII-1.351 + $oriII_{P.\ profundum}$ (coord. 2234799-2736)	This work
synPhop(*kan*)	synVicII-0.3+ $oriII_{P.\ profundum}$	This work
synViaII(AmpR)	synVicII-1.351 + $oriII_{V.\ anguillarum}$ (coord.	This work
synViaII(*kan*)	synVicII-1.351 + $oriII_{V.\ anguillarum}$	This work
synVicII-0.1	*rctB* (coord. 1134–3110), *parAB* (coord. 1070018-1072220), R6K*orig*; *2µ ori*, *ura3*, AmpR	(Messerschmidt *et al.* 2015)

Plasmid	Relevant characteristics[a)]	Reference
synVicII-0.11	synVicII-0.1 + *gfp*(AAV)	This work
synVicII-0.3	synVicII-1.351 AmpR::*kan*	This work
synVicII-1.0	synVicII 0.1 + *oriII* (coord. 1072221-1133)	(Messerschmidt *et al.* 2015)
synVicII-1.3	synVicII-1.0 + $P_{A1/04/03}$-RBSII-*gfp*(AAV)	(Messerschmidt *et al.* 2015)
synVicII-1.301	synVicII-1.3 + *oriC* w/o 13mer (coord. 3923811-3924568)	This work
synVicII-1.302	synVicII-1.3 + *lacZ* (coord. 363541-364605)	This work
synVicII-1.31	synVicII-1.3 + *Xho*I between *2µ ori* and *parB*	This work
synVicII-1.311	synVicII-1.31 + P_{rhaT} in *Xho*I	This work
synVicII-1.312	synVicII-1.311 + *rfp* in *Nhe*I	This work
synVicII-1.313	synVicII-0.11 + $oriII_{\text{candidate 4 evolution}}$	This work
synVicII-1.314	synVicII-1.0 + $gfp_{\text{candidate 4 evolution}}$	This work
synVicII-1.32	synVicII-1.31 Δ*I-Sce*I, +*Pvu*II	This work
synVicII-1.33	synVicII-1.32 + FRT site *Sma*I + FRT site *I-Sce*I	This work
synVicII-1.34	synVicII-1.33 + *rfp*	This work
synVicII-1.35	synVicII-1.34 + *oriT*	This work
synVicII-1.351	synVicII-1.35, $oriII_{V.cholerae}$*::lacZα*	This work
synVicII-1.3511	synVicII-1.351 Δla*cZα*	This work
synVicII-1.36	synVicII-1.35 + *lacZα* + *ccdB*	This work
synVicII-1.36	synVicII-1.35 + *lacZα* + *ccdB*	This work
synVicII-1.361	synVicII-1.36 w/o *Bpi*I recognition sites	This work
synVicII-1.362	synVicII-1.36 w/o *Bsa*I recognition sites	This work
synVicII-1.37	synVicII-1.36 w/o *Bpi*I and *Bsa*I recognition sites	This work
synVicII-2.01	synVicII-1.37 MoClo Level M 1 (pMA60)	This work
synVicII-2.02	synVicII-1.37 MoClo Level M 2 (pMA61)	This work
synVicII-2.03	synVicII-1.37 MoClo Level M 3 (pMA62)	This work
synVicII-2.04	synVicII-1.37 MoClo Level M 4 (pMA63)	This work
synVicII-2.05	synVicII-1.37 MoClo Level M 5 (pMA64)	This work
synVicII-2.06	synVicII-1.37 MoClo Level M 6 (pMA65)	This work
synVicII-2.07	synVicII-1.37 MoClo Level M 7 (pMA66)	This work
synVicII-2.11	synVicII-1.37 MoClo Level P 1 (pMA67)	This work
synVicII-2.111	synVicII-2.11 and pMA678 endlinker	This work
synVicII-2.12	synVicII-1.37 MoClo Level P 2 (pMA68)	This work
synVicII-2.13	synVicII-1.37 MoClo Level P 3 (pMA69)	This work
synVicII-2.14	synVicII-1.37 MoClo Level P 4 (pMA70)	This work
synVicII-2.15	synVicII-1.37 MoClo Level P 5 (pMA71)	This work
synVicII-2.16	synVicII-1.37 MoClo Level P 6 (pMA72)	This work
synVicII-2.17	synVicII-1.37 MoClo Level P 7 (pMA73)	This work
synVicoII(AmpR)	synVicII-1.351 + $oriII_{V.\ coralyticus}$ (coord. 1636573-1642380)	This work
synVicoII(*kan*)pMA893	synVicII-0.3+ $oriII_{V.\ coralyticus}$	This work
synVifII(AmpR)	synVicII-1.351 + $oriII_{V.\ furnissii}$ (coord. 1027513-1033467)	This work
synVifII(*kan*)	synVicII-0.3+ $oriII_{V.\ furnissii}$	This work
synVinII(AmpR)	synVicII-1.351 + $oriII_{V.\ nigripulchritudo}$ (coord.2212141-5725)	This work
synVinII(*kan*)	synVicII-0.3+ $oriII_{V.\ nigripulchritudo}$	This work

Plasmid	Relevant characteristics[a)]	Reference
synVipaII(AmpR)	synVicII-1.351 + *oriII*$_{V.\ parahaemolyticus}$ (coord. 1874837-3408)	This work
synVipaII(*kan*)	synVicII-0.3+ *oriII*$_{V.\ parahaemolyticus}$	This work
synVituII(AmpR)	synVicII-1.351 + *oriII*$_{V.tubiashi}$ (coord. 1764492-3576)	This work
synVituII(*kan*)	synVicII-0.3+ *oriII*$_{V.tubiashi}$	This work
synVivuII(AmpR)	synVicII-1.351 + *oriII*$_{V.\ vulnificus}$ (coord.1853576-2515)	This work
synVivuII(*kan*)	synVicII-0.3+ *oriII*$_{V.\ vulnificus}$	This work

[a)] Genomic positions are indicated according to the following genome annotations: NC_002506.1 for ChrII of *V. cholerae,* NC_ 000913.2 for *E. coli* MG1655, for ChrII of *P. profundum* NC_00637.1, *V. anguillarum* NC_015637.1, *V. furnissii* NC_016628.1, *V. parahaemolyticus* NC_004605.1; *V. vulnificus* NC_005140.1, *V. nigripulchritudo* NC_022543.1, *V. tubiashi* NZ_CP009355.1 and *V. coralliilyticus* NZ_CP009265.1

Table S3: Oligonucleotides used in this work

Name	Sequence from 5' to 3'
14	CAATCTCAATTCGATCGGCCTGCACT
16	GGTTAGATCCGTATCACACTTACCGT
26	CTTAACGTGAGTTTTCGTTCCACTAGGGATAACAGGGTAATTTTATCAAAAAGAGTGTTGACTTGTGAG
27	GTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAGGTTCCAAGCTAGCTTGGATTCTCAC
28	GATAGGTGCCTCACTGATTAAG
29	GGCTCTAAGGGCTTCTCAGT
30	GATTTGGCAAAATCCTGACTCAGTGCTTTCTATCCCGTGCAGCTCTGGCCCGTGTCT
31	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAGAAAAACTCATCGAGCATCAAATGAAAC
69	TGAAGACATGCTTGATGAGGATCTGCAGAGGAG
70	TGAAGACTAGGTTTTCTTCTGCATTACGGGGC
71	TGAAGACATAACCATGGGCTGGGAGGCCT
72	TGAAGACATAGCGACGGCCAGTGCCAAGCT
214	CGCTCATGAG ACAATAACCC
215	GCCTTCTTGA CGAGTTCTTC TGACTGTCAG ACCAAGTTTA CTCATATA
327	CGTTGGCAGCGCCTAAGAAACCAATAAGGCTAAGCCCTCGAGCCCTAAAACGCACAAAGCCC
328	TCAATGCTGAGCACGCTAAGTTT
329	GATGCGGGCTTTGTGCGTTTTAGGGCTCGAGGTACTTTTCGTAAGGGTATGG
330	CTAAGAAACCAATAAGGCTAAGCCCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCCTCGAGCATCTCCGACG AGATGAGT
331	GTACGTACTATCAACAGGTTCCAAGGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGCTAGCAGAGGAGAAAT TAAGCATGGCTTCCTCCGAAGACGT
581	GAGTAAACTTGGTCTGACAGTCATCGCAGTACTGTTGTATTCATTAAGC
582	CTGTCAGACCAAGTTTACTCATATATACTTTAGATTG
597	CGTTTTTTATTGGTGAGAATCCAAGGCTAGCTTAAGCACCGGTGGAGTGAC
793	GTTACCCAGGTCGATTTCAG
794	AAATCGTCCGCTCTATGCAG
814	TCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACAATGAGTAACAGCTGTAGTTTATCAAAAAGAGTGTT GACTTGTGAGCGGATAACAATG
815	ATGCGGGCTTTGTGCGTTTTAGGGCGTCGGAGATGCTCGAGGAAG
816	CTAAGAAACCAATAAGGCTAAGCCCGGGTGCATGACAGGAAGTTCCTATACTTTCTAG

Name	Sequence from 5’ to 3’
817	GTACGTACTATCAACAGGTTC
818	CGTTTTTTATTGGTGAGAATCCAAGTAGGGATAACAGGGTAATATAGGTCTGCGAAGTTCCTATACTTTCTAG
819	CCTAAGAAACCAATAAGGCTAAGCCCTATAAACGCAGAAAGGCCCAC
820	TATAGGAACTTCCTGTCATGCACCCGGGCAATACGCAAACCGCCTCTC
874	GCGGGGAGAGGCGGTTTGCGTATTGCCCGCTAGCCCTTAAGGTATACTTTCCGCTGC
875	AAGTATAGGAACTTCCTGTCATGCACCCGGGAGCTTATCGGCCAGCCTCGC
876	CCCTTAACGTGAGTTTTCGTTCCACTAGGGATAACAGGGTAATCGGCTTACCATCCAGCGCCAC
877	ACAAGTCAACACTCTTTTTGATAAACTCTATCGTGCGGTGGTTGAAC
1002	TTGGTGAGAATCCAAGTAGGGATAATTGCGGCCGCAGCTGGCACGACAGGTTTGCCG
1004	AACTTCGCAGACCTATATTACCCTGATGCGGCCGCGACTATGCGGCATCAGAGC
1005	AACTTCGCAGACCTATATTACCCTGTTGCGGCCGCGTCACAGCTTGTCTGTAAGCGGATGCC
1029	CGGGCGTTTTTTATTGGTGAGAATCCAAGTAGGGATAATTTCCTGCACTCTGTGGTCTCA
1030	CGGGCGTTTTTTATTGGTGAGAATCCAAGTAGGGATAATTCCTGCACTCTGTGAAGACAA
1031	TAGAAAGTATAGGAACTTCGCAGACCTATATTACCCTGTTGCCCGGCCACTTCGTGTCCC
1099	CACATTTCCCCGAAAAGTGC
1100	GCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTC
1132	GAATAGGAACTTCCTCGAGCATCTCCGACGGCGCGCCAGCTGGCACGACAGGTTTG
1133	GAAAAATAAACAAATAGGGGTTCCGCGGGCGCGCCCTATGCGGCATCAGAGCAGATTG
1148	GAACTTCCTCGAGCATCTCCGACGGCGCGCCGATATACCCAAGCTAGACCG
1149	GAAAAATAAACAAATAGGGGTTCCGCGGGCGCGCCCTAGCAAGGATGATCAGAGAC
1435	CCTCGAGCATCTCCGACGGCGCGCCTGCAGCTCTGGCCCGTGTCT
1487	GAGCATCTCCGACGGCGCGCCCATGGAGCGGCGTAACCGTC
1488	GGCCAGAGCTGCAGGCGCGCCCGTCGACAGCGACACACTTG
1489	GAGCATCTCCGACGGCGCGCCCCGCCAATGATGATGACGTC
1490	GGCCAGAGCTGCAGGCGCGCCCTGGGATCGTGGGTTAATTTAC
1628	TGGTTTAAATACACCGCCAGCCATGAAAGATGAGGCTGATCAGTTTGTCGTCTCACCGAC
1629	GGGTGGCGCTAACGCTTGATCAATGATCCCTGCCCAGTTAGTCGGTGAGACGACAAACTG
1630	CTTTGCAAATAGTCCTCTTCCAACAATAATAATGTCAGATCCTGTAGACACCACATCATC
1631	TGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTGTCTACAGG
1632	CCCGCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCCTCGAAG
1633	AGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAGGACAGAAAATTTG
1634	GAAAGTATAGGAACTTCAGAGCGCTTTTGAAAACCAAAAGCGCTCTGATGACGCACTTTC
1635	TTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTCATCAGAGCG
1636	TAACAATTTCACACATACTAGAGAAAGAGGAGAAATACTAGATGGTGAGCAAGGGCGAGG
1637	CTGAGCCTTTCGTTTTATTTGATGCCTGGCTCTAGTATTACTTGTACAGCTCGTCCATGC
1638	ACTGACCAATGCCTCGCTGCACTGGAACATTTTGGCAAGCGCTTTTTGGTCCTCCGCTTC
1639	TGAGAAAGGTAAAGAGATGCAAGCCAAGCTCGATAGTGGTGAAGCGGAGGACCAAAAAGC
1640	TGCGTTCAGTTCAGAAGTTTTAATTGCCATAAGTTACTTACCCCTGATTCAGTGATGACC
1641	AAAAGTGCTTTAGAGCTAGAGCGCGTACTTCATTCACATTGGTCATCACTGAATCAGGGG

Table S4: Primers and templates for the construction of different *Vibrio* replicons

Replicon	gDNA as *oriII* template[a)]	Oligonucleotides
synVituII(amp)	*V. tubiashi*	1166+1167
synVifII(amp)	*V. furnissi*	1156+1157
synViniII(amp)	*V. nigripulchritudo*	1160+1161
synVicoII(amp)	*V. coralliilyticus*	1152+1153
synVipaII(amp)	*V. parahaemolyticus*	1162+1163
synVivuII(amp)	*V. vulnificus*	1168+1169
synViaII(amp)	*V. anguillarum*	1411+1223; 1222+1412

[a)]Strain designations are given in table S1.

Table S5: Relative origin copy numbers derived from CGH experiments[a)]

synVicII-	**oriC/ter ratio**	**oriII/oriC**	**oriII/ter**
candidate 3	3,6	2,6	9,5
	2,9	2,8	8,0
candidate 4	3,4	1,0	3,5
	3,5	1,0	3,6
2.0	3,2	0,7	2,3
	3,9	0,5	2,2

[a)]Numbers are calculated for experiments shown in figures 2, 5 and S1 as indicated.

Supplementary references

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331):1453-1462

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406 (6795):477-483. doi:10.1038/35020000

House BL, Mortimer MW, Kahn ML (2004) New recombination methods for *Sinorhizobium meliloti* genetics. Appl Environ Microbiol 70 (5):2806-2815

James P, Halladay J, Craig EA (1996) Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 144 (4):1425-1436

Kouprina N, Annab L, Graves J, Afshari C, Barrett JC, Resnick MA, Larionov V (1998) Functional copies of a human gene can be directly isolated by transformation-associated recombination cloning with a small 3' end target sequence. Proc Natl Acad Sci U S A 95 (8):4469-4474

Messerschmidt SJ, Kemter FS, Schindler D, Waldminghaus T (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. Biotechnol J 10 (2):302-314. doi:10.1002/biot.201400031

Miller VL, Mekalanos JJ (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J Bacteriol 170 (6):2575-2583

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synth Biol. doi:10.1021/acssynbio.6b00089

Schlüter JP, Czuppon P, Schauer O, Pfaffelhuber P, McIntosh M, Becker A (2015) Classification of phenotypic subpopulations in isogenic bacterial cultures by triple promoter probing at single cell level. J Biotechnol 198:3-14. doi:10.1016/j.jbiotec.2015.01.021

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6 (2):e16765. doi:10.1371/journal.pone.0016765

3.3 Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors

Das Designen und die Assemblierung von DNA-Sequenzen ist einer der grundlegenden Prozesse in der synthetischen Biologie. Das MoClo-System hat sich etabliert, um damit effizient und modularisiert DNA-Fragmente zu assemblieren. MoClo basiert auf Typ IIS-Endonukleasen, die gerichtet außerhalb ihrer Erkennungssequenz schneiden, wodurch die jeweilige Erkennungssequenz bei entsprechendem Design des Experiments verloren geht. Es ist wichtig, dass die DNA-Fragmente, neben den passenden, durch Typ IIS-Endonukleasen entstehenden DNA-Überhängen, frei von weiteren Erkennungssequenzen sind. Deshalb ist eine sorgfältige Planung notwendig und wird durch Computerprogramme deutlich erleichtert.

In diesem Kapitel wird das Computer Programm MARSeG (*Motif Avoiding Randomized Sequence Generator*) beschrieben. MARSeG ermöglicht die Generierung von degenerierten, zufälligen DNA-Sequenzen die zeitgleich definierte DNA-Sequenzmotive ausschließen. MARSeG wurde verwendet, um ein Fluoreszenz-Repressor-Operator System (FROS) mit variablen Sequenzen zwischen den Operator-Bindestellen zu generieren. Für die DNA-Assemblierung der MARSeG generierten FROS-Sequenzen wurden MoClo-Vektoren dahingehend optimiert, dass ein Arbeiten mit Sequenzbibliotheken möglich ist. Dies konnte erreicht werden indem ein negativer Selektionsmarker in die Vektoren integriert wurde. Zusätzlich wurde das Set der MoClo-Vektoren um Vektoren mit einem konditionalen Replikationsursprung erweitert, damit bei Experimenten zur chromosomalen Integration Zellen die ein Plasmid aufgenommen haben nicht wachsen können. Durch MARSeG und die modifizierten MoClo-Vektoren konnte die FROS-Sequenz erfolgreich assembliert, in das Chromosom von *E. coli* integriert und die Funktionalität gezeigt werden. Des Weiteren konnte belegt werden, dass die assemblierte FROS-Sequenz mit variablen Sequenzen zwischen den Operatorsequenzen gegenüber einer FROS-Sequenz ohne variable Sequenzen keine rekombinationsbedingte Verkleinerung der DNA-Sequenz aufweist.

Daniel Schindler hat in Rücksprache mit Torsten Waldminghaus die MoClo Plasmide optimiert, die oligonukleotidbasierte Sequenzgenerierung entwickelt, sowie die DNA-Assemblierung der Sequenzen etabliert. Theodor Sperlea hat unter Anleitung von Torsten Waldminghaus und Daniel Schindler das Programm MARSeG programmiert und die variablen Sequenzen des generierten FROS mit Torsten Waldminghaus analysiert. Sarah Milbredt hat in Rücksprache mit Torsten Waldminghaus und Daniel Schindler die FROS-Sequenz assembliert. Fluoreszenzmikroskopische Aufnahmen wurden von Sarah Milbredt durchgeführt und die Daten mit Torsten Waldminghaus zusammen ausgewertet. Alle Autoren haben gleichwertig zum verfassen des Manuskripts beigetragen.

ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

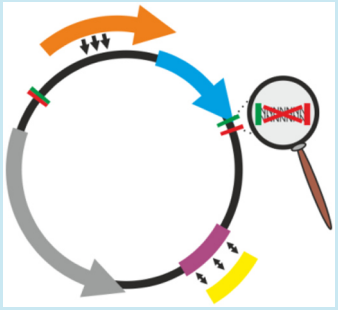
Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors

Daniel Schindler,[†] Sarah Milbredt,[†] Theodor Sperlea,[†] and Torsten Waldminghaus*

Chromosome Biology Group, LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-Universität Marburg, Hans-Meerwein-Str. 6, D-35043 Marburg, Germany

Supporting Information

ABSTRACT: Efficient assembly of large DNA constructs is a key technology in synthetic biology. One of the most popular assembly systems is the MoClo standard in which restriction and ligation of multiple fragments occurs in a one-pot reaction. The system is based on a smart vector design and type IIs restriction enzymes, which cut outside their recognition site. While the initial MoClo vectors had been developed for the assembly of multiple transcription units of plants, some derivatives of the vectors have been developed over the last years. Here we present a new set of MoClo vectors for the assembly of fragment libraries and insertion of constructs into bacterial chromosomes. The vectors are accompanied by a computer program that generates a degenerate synthetic DNA sequence that excludes "forbidden" DNA motifs. We demonstrate the usability of the new approach by construction of a stable fluorescence repressor operator system (FROS).



KEYWORDS: *genome engineering, chromosome, software, Escherichia coli, sequence design, synthetic biology*

Biotechnology as well as basic research in biology often includes changing the organism of interest. In some cases, one might want to teach microorganisms to produce some valuable chemical, in other cases one wants to see the effect of additional factors or how cells compete without a certain component. Thus, the ability to introduce changes in an efficient way is a key for future life science developments. Alterations of organisms will, in most cases, be made on the DNA level from which the phenotypic characteristics are derived. The development of genetic modification started in the 1970s with the first recombinant DNA being used to transform cells and has since been extended enormously. Especially the research field of synthetic biology came along with a multitude of new techniques for DNA manipulation and assembly.[1−5] These new DNA assembly approaches were developed to overcome certain limitations of traditional cloning strategies. One major issue is that cloning based on DNA ligase and regular restriction endonucleases often leaves the respective cut sites as scar in the assembled product. However, there are at least four DNA assembly approaches for scar-free assembly of DNA fragments.[1−3] First, the Gibson assembly is based on homologous ends of DNA fragments, which are fused in an *in vitro* reaction including an exonuclease, a DNA polymerase and a DNA ligase.[5] This is similar to the second approach where the homologous ends are fused *in vivo* by the highly efficient recombination system of the yeast *Saccharomyces cerevisiae*.[6] In a third approach, the ligase cycling reaction (LCR), the homology is not mediated by the DNA fragment ends but by a bridging oligonucleotide.[7] A fourth approach makes use of type IIs restriction enzymes.[8] These enzymes are distinct from other restriction enzymes in that they cut outside their recognition site. They are directional and the positioning of the recognition site allows determination where the DNA is cut. Notably, the actual cut site can be freely chosen allowing the design of scar-less assemblies.

An important benefit of the four described methods as compared to traditional cloning is their suitability for fast, single reaction multifragment assembly. The first three approaches are dependent on homologous regions of about 20−40 bps which will determine the position of fragments in a multipart assembly. With type IIs restriction sites the required homology is limited to only 4 bps. This fact was used to develop hierarchical assembly systems based on vectors with defined 4 bp sequences to fit one another.[8−10] Such a system allows the efficient assembly of many fragments into a destination vector independent of the actual subfragment sequence or size. Probably the most popular type IIs-based assembly framework is the MoClo system developed by Sylvestre Marillonnet and colleagues.[9,11] It consists of sets of seven vectors with the 4 bp overhang ends of each vector matching the overhangs of the preceding and following vector, respectively. Assembling fragments from one vector set (one level) into the next is possible because the resistance markers as well as the type IIs restriction enzymes and sites are alternating. A set of endlinkers is used to generate matching ends for assembly of different numbers of fragments into one acceptor vector.[9,11]

One important benefit of the MoClo approach is that it is based on mixing complete plasmids eliminating the need for PCR or fragment isolation. Recently, the MoClo system was adapted to or optimized for special purposes as transcription unit assembly in plants, mammals, fungi or bacteria.[12−15] Here

Received: March 18, 2016

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

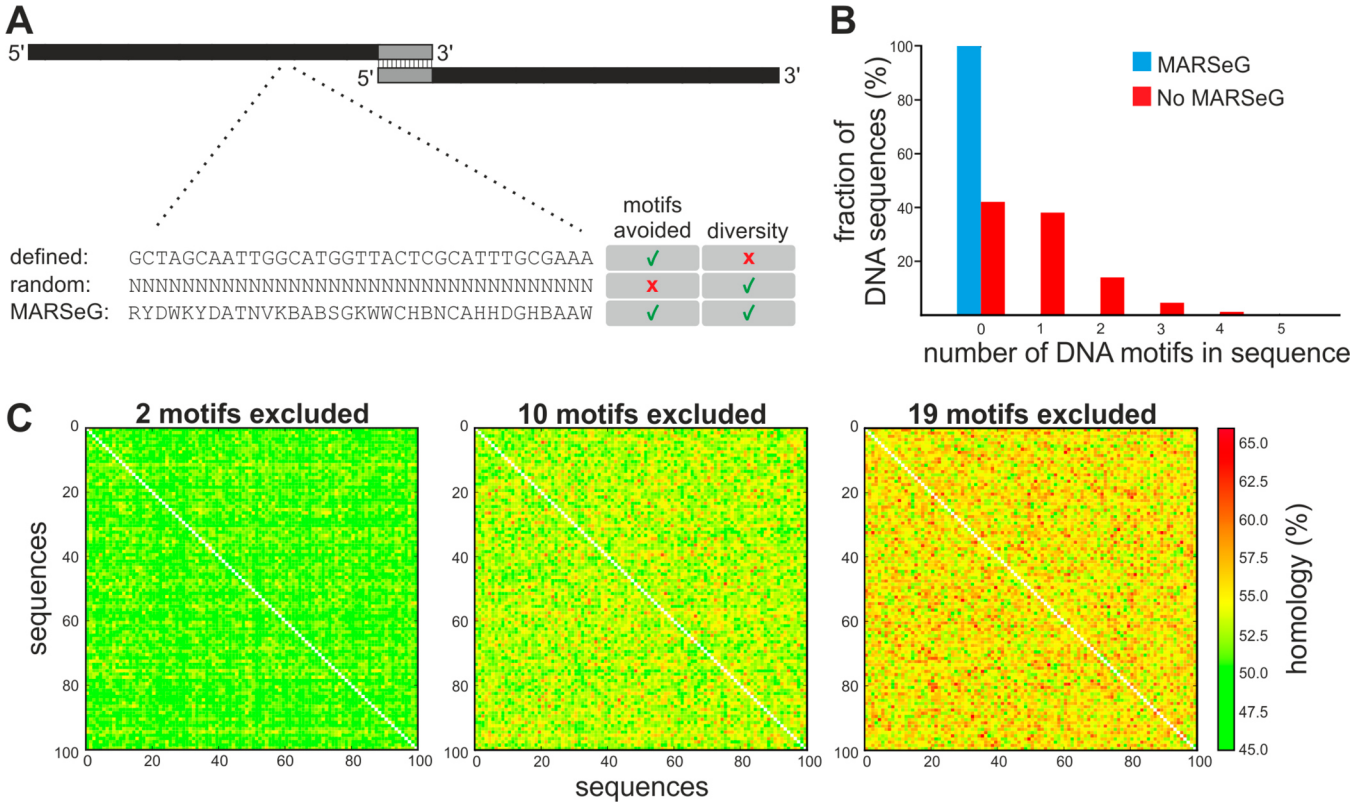


Figure 1. Generation of diverse DNA sequences that exclude of a list of motifs by MARSeG. (A) Double-stranded DNA fragment libraries are generated by annealing and extension of two single-stranded oligonucleotides (black lines) with partial overlap (gray boxes). Three possible designs are shown below. A fully defined sequence (first line) could exclude a list of motifs but does not confer diversity; a fully randomized sequence (second line) confers diversity but might lead to sequences including unwanted motifs. Sequences generated with MARSeG (third line) confer diversity while excluding unwanted motifs. (B) Sequences of 200 bps were generated completely random (red) or with MARSeG (blue) and the number of motifs from a list of 19 "forbidden" restriction enzyme recognition sites (list III in Table S1) was counted for 500 derived sequences. (C) Trade-off between the amount of excluded motifs and diversity in MARSeG generated sequences. After generating degenerate sequences using MARSeG with three motif lists as indicated (Table S1), 100 sequences were defined from each respective template. Pairwise sequence homology values were calculated using a Smith–Waterman algorithm. The degree of homology is color coded as indicated.

we present modifications of the MoClo system for efficient cloning of sequence libraries and for the construction of fragments to be inserted into the *E. coli* chromosome. We introduce a computer tool for sequence design and show the feasibility of our approach by designing and assembling a FROS array (fluorescence repressor operator system).

FROS is a widely used tool for spatial and temporal visualization of genetic loci *in vivo* and has been applied to various different organisms.[16−18] Fluorescently labeled DNA binding proteins are used to highlight specific binding sites, which are integrated at a gene locus of interest by homologous recombination. It was initially applied with tandem repeats of the *lac* operator and a *gfp* fused Lac repressor in yeast and CHO-cells.[16] Also a *tet* operator-based FROS system was generated and used in yeast.[19] As the transfer of FROS to bacteria was not very successful due to instability caused by large homologous regions; arrays were optimized by insertion of random spacers in between the operator repeats to decrease homology.[20,21] As further improvement the number of binding sites can be reduced from 250 to 64 to limit interference with the replication machinery.[22] FROS was subsequently applied successfully in various bacteria to gain new insights into the localization, replication and segregation of chromosomes.[17,23,24] Nevertheless, the design and generation of DNA sequences with many repetitive elements remains challenging. In this paper we present a new set of MoClo vectors that allowed generation of a FROS array with 64 binding sites of two different operators in just 4 cloning steps based on a single pair of degenerate oligonucleotides and its subsequent integration into the chromosome of *E. coli*.

Efficient assembly of DNA fragments is critical for modern molecular biology approaches. It was predicted that software tools will have an increasing importance for DNA assembly approaches.[2] Often, sequences are needed that have specific DNA motifs at defined sites but not at others. Other DNA motifs, as for example restriction sites, need to be excluded throughout the whole construct. It might be straightforward to design a single exact sequence with these characteristics based on extension of two DNA oligonucleotides with an overlap region at one end (Figure 1A). However, efficient cloning strategies should allow working with libraries generated from mixtures of DNA oligonucleotides to lower the overall costs. Here we present the computer program MARSeG (Motif Avoiding Randomized Sequence Generator) that generates degenerated sequences with a high degree of diversity while excluding a list of DNA motifs provided by the user (Figure 1A). An example for its application could be the design of 20 spacer sequences with a length of 200 bps each, that are used to separate transcription units within a large scale gene circuit assembly. Notably, these spacer sequences should not harbor recognition sites for a list of restriction enzymes. Instead of designing and buying 20 individual sequences one could just

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

order a fully randomized sequence with 200 Ns and receive an oligonucleotide mix to be cloned into a vector backbone. However, a certain amount of these sequences will have at least one of the "forbidden" DNA motifs. We tested this assumption by comparing random sequences with sequences generated with MARSeG (Figure 1B). Almost 60% of completely random sequences with a length of 200 bps contain at least one motif from a list of 19 restriction enzyme recognition sites (motif list III in Table S1, Figure 1B). The computer tool MARSeG reduces the diversity of sequences in such a way that "forbidden" motifs are excluded while maintaining high sequence diversity. This leads to sequence collections without any appearance of the "forbidden" motifs (Figure 1B). The degree of MARSeG library diversity will depend on the number and type of DNA motifs to be excluded as shown by analysis of the overall sequence homology for 100 example sequences for three different lists with two, ten or 19 "forbidden" DNA motifs, respectively (Table S1, Figure 1C). An alternative approach would be to generate many sequences of the desired length and exclude all sequences that do contain one or more of the "forbidden" motifs or other undesired characteristics.[25] However, this approach will only generate individual sequences and no sequence libraries as MARSeG does. MARSeG is open source and available, including a detailed user manual, through the Web site (http://www.synmikro.com/marseg).

The MoClo vectors are widely used and some specialized derivatives or part libraries have been developed.[12,13,26] We changed the existing vectors to facilitate library cloning, multifragment assembly and insertion of constructs into bacterial chromosomes *via* homologous recombination techniques. An overview of the modifications is depicted in Figure 2A and a list of new vector sets is given in Table S2. The starting point for our modifications was a set of MoClo vectors kindly provided by Sylvestre Marillonnet. The respective Level 1 vectors have been described previously and the level M and P vectors differ from previous vectors by the fact that they do not contain T-DNA borders for agrobacterium delivery.[9] Working with libraries instead of individual sequences poses special requirements on the DNA assembly system. Most importantly, the percentage of positive clones should be near 100% because clones are not selected individually. To suppress vectors still containing the *lacZ* cassette instead of the desired fragment, we added *ccdB* gene in such a way that it is lost with the *lacZ* gene upon successful cloning (*ccdB*+ vectors). The *ccdB* gene product is a small cytotoxin that kills *E. coli* cells that are not engineered to express the antitoxin CcdA or possess a mutated gyrase.[27] As expected, cloning with the *ccdB*+ vector led to elimination of the blue colonies still harboring the *lacZ-ccdB* MoClo cassette (Figure 2B).

A second change to previous MoClo vectors is a size reduction of the sequence remaining between level 1 fragments in higher level assemblies. Respective sequences where placed in the original level 1 vectors between the *Bpi*I and *Bsa*I sites and contain restriction sites to facilitate the analysis of assembled transcription sites. For this purpose they were certainly helpful but could be deleterious in other cases for example as potential recombination sites if occurring to frequently. To keep this short sequences remaining between the assembled fragments as small as possible we deleted the 12 bp between the *Bsa*I and the *Bpi*I cut sites for the whole level 1 vector set.

Very often it is desired to introduce constructed gene circuits or pathways into the host chromosome as the genomic stability

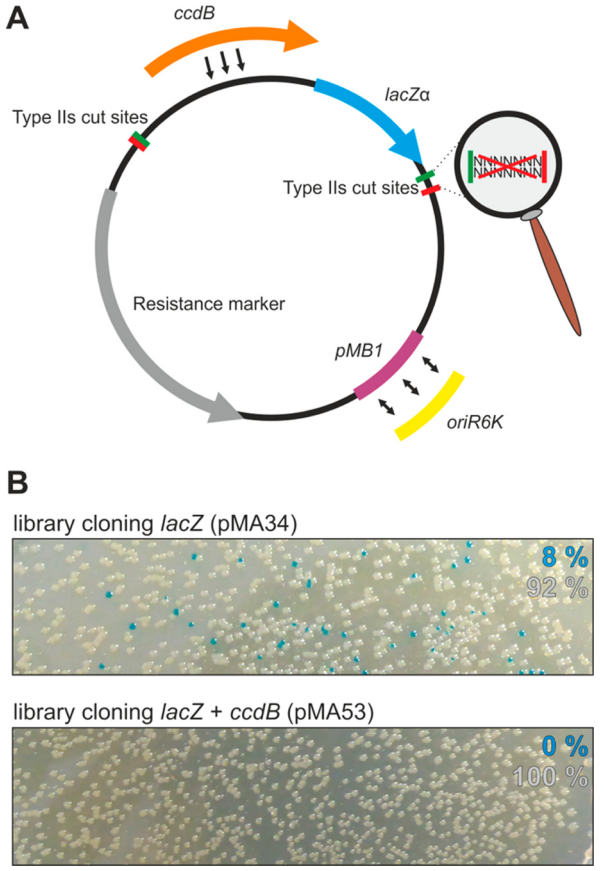


Figure 2. Optimization of MoClo vectors for library cloning and chromosomal insertions. (A) Schematic drawing of changes to existing MoClo vectors: arrows = insertion, double arrows = exchange, red cross = deletion. (B) Cloning into standard MoClo vectors produces some background consisting of original vectors, indicated by blue colonies (top panel). New vectors including *ccdB* lead to white colonies only (bottom panel). Percentage of colonies is given in the respective color.

is higher compared to plasmid based expression.[28] In addition, the cell to cell variability of plasmid copy numbers makes it difficult to derive quantitative data for exact measurements of expression phenomena.[29] Chromosomal insertions into the *E. coli* chromosome are straightforward with the phage lambda based recombination system.[30] However, a frequent problem are false-positives originating from transferred plasmids even if those just served as PCR template or were supposed to be cut by restriction enzymes. To eliminate this problem we exchanged the original *pMB1* replication origin with *oriR6K*. This conditional replication origin does only replicate in *E. coli* strains expressing the lambda *pir* gene and thus, replicons based on *oriR6K* are not able to replicate in wildtype *E. coli*. As a proof of principle we cloned building blocks for a chromosomal insertion into level 1 vectors, including homologous regions targeting the *lac* locus, a chloramphenicol resistance marker flanked by FRT sites to remove the cassette after successful integration *via* "flipping" and a fluorescence gene with a constitutive promoter. After a one-step assembly of all four parts into one of our new vectors the assembled construct could readily be inserted into the *E. coli* chromosome by recombineering to generate red fluorescent cells (data not shown). All vectors as well as the parts described here and below (72 plasmids in total) are available through a request

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

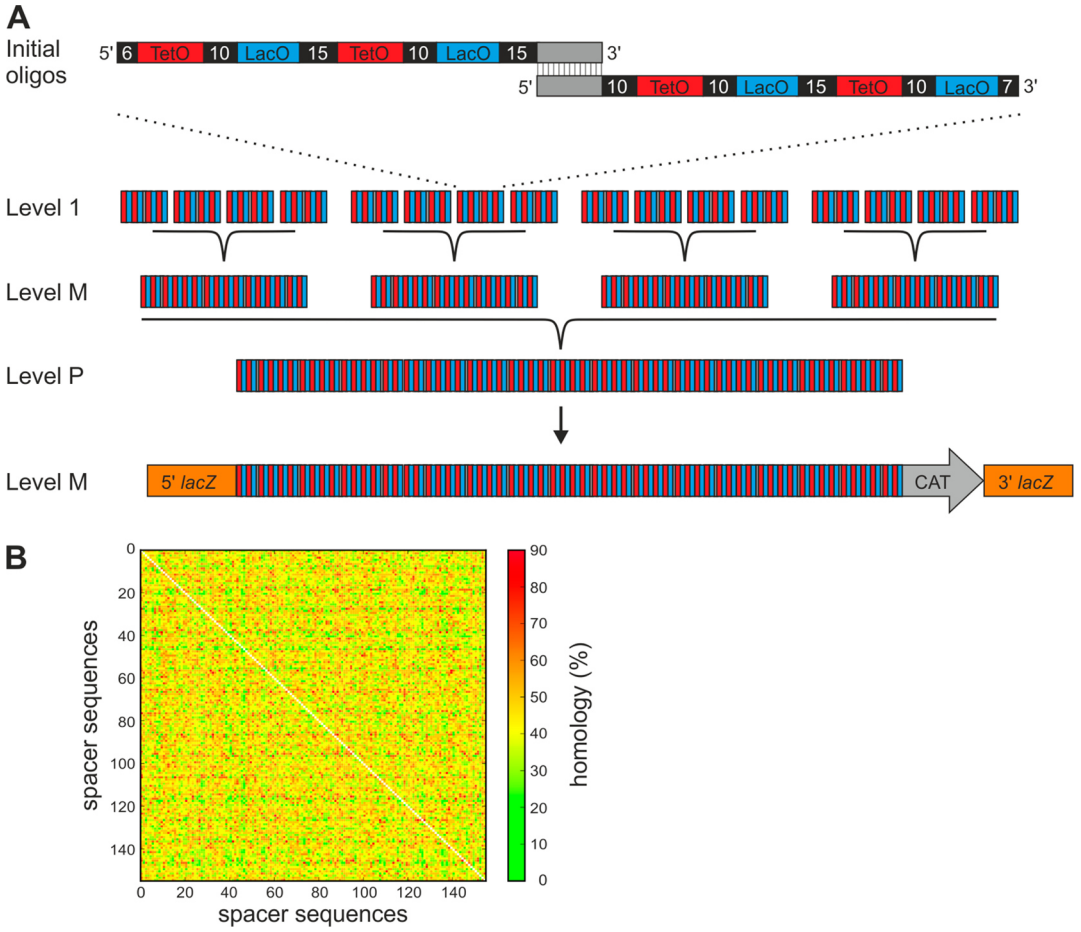


Figure 3. Assembly of the LacO/TetO operator array. (A) DNA oligonucleotides were annealed (gray boxes), elongated and enriched *via* PCR. Linker lengths (in bps) are shown as white numbers. The resulting library was cloned into seven level 1 vectors. Sets of four level 1 vector libraries were assembled into level M acceptor vectors and four resulting individual vectors were combined into level P to gain the final array. For integration into *E. coli lacZ*, flanking regions and a chloramphenicol cassette (flanked by FRT sites) were assembled together with the final array into level M. (B) Spacer sequence homology matrix. The sequenced FROS array assembly parts (pMA281−284, see Table S4) were disassembled and the pairwise homologies of spacer sequences were calculated using a Smith−Waterman algorithm. The respective homology is color coded as indicated.

form on our homepage (http://www.synmikro.com/plasmidrequest). An overview of the new MoClo vectors and their position within the MoClo hierarchy is shown in Figure S1.

To test the usability of the MARSeG program and the new MoClo vectors we applied these tools to a more challenging assembly, namely the construction of a FROS system. Such systems consist of an array of operator sites which are bound by a fluorescence marker fused to the respective repressor protein to visualize a specific genomic region by microscopy. These arrays are difficult to assemble because the operator sequences are homologous to one another. Such repetitive sequences have been shown to be especially difficult to assemble with methods relying on larger homology parts as Gibson assembly.[8,31] The array we designed contains *tet* as well as *lac* operators to allow more flexibility in the choice of binding proteins. Construction of a FROS array of 128 operators (64 TetO plus 64 LacO) was based on building blocks with 8 alternating operator sequences separated by variable linker sequences to reduce homology between building blocks (Figure 3A). The basic building blocks were generated by elongation of two overlapping DNA oligonucleotides designed with MARSeG (Figure 3A, see Methods section for details). Fragment libraries were applied to a MoClo reaction with seven level 1 vectors (Figure 3A). After transformation of the cloning reaction the generated plasmid libraries were directly purified from the liquid *E. coli* culture and used for a four part assembly into level M vectors (Figure 3A). Each of these parts should contain 32 operators with each of the operators being separated by a different spacer sequence of different length which was designed by MARSeG to be diverse on one hand but to not contain recognition sites of the type IIs endonucleases used (Figure 3A). To test this diversity, we sequenced the operator array regions after the second assembly step (pMA281−284, see Table S4) and aligned all spacer sequences with one another. The respective homology matrix is shown in Figure 3B. Notably, none of the 155 spacer sequences appeared more than once in the array. Homologies ranged between 0 and 90%, clearly showing that the design and cloning approach presented here is able to produce a suitable amount of sequence diversity.

To further test the functionality of the constructed array it was integrated into the *E. coli* chromosome *via* the new vector system as described above. Cells carrying this integration were transformed with a plasmid allowing inducible expression of the fluorescence protein mVenus fused to the TetR repressor. Fluorescence microscopy showed clear formation of foci in cells with the constructed FROS array insertion as expected (Figure 4A). In contrast, only diffuse fluorescence was seen in cells

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

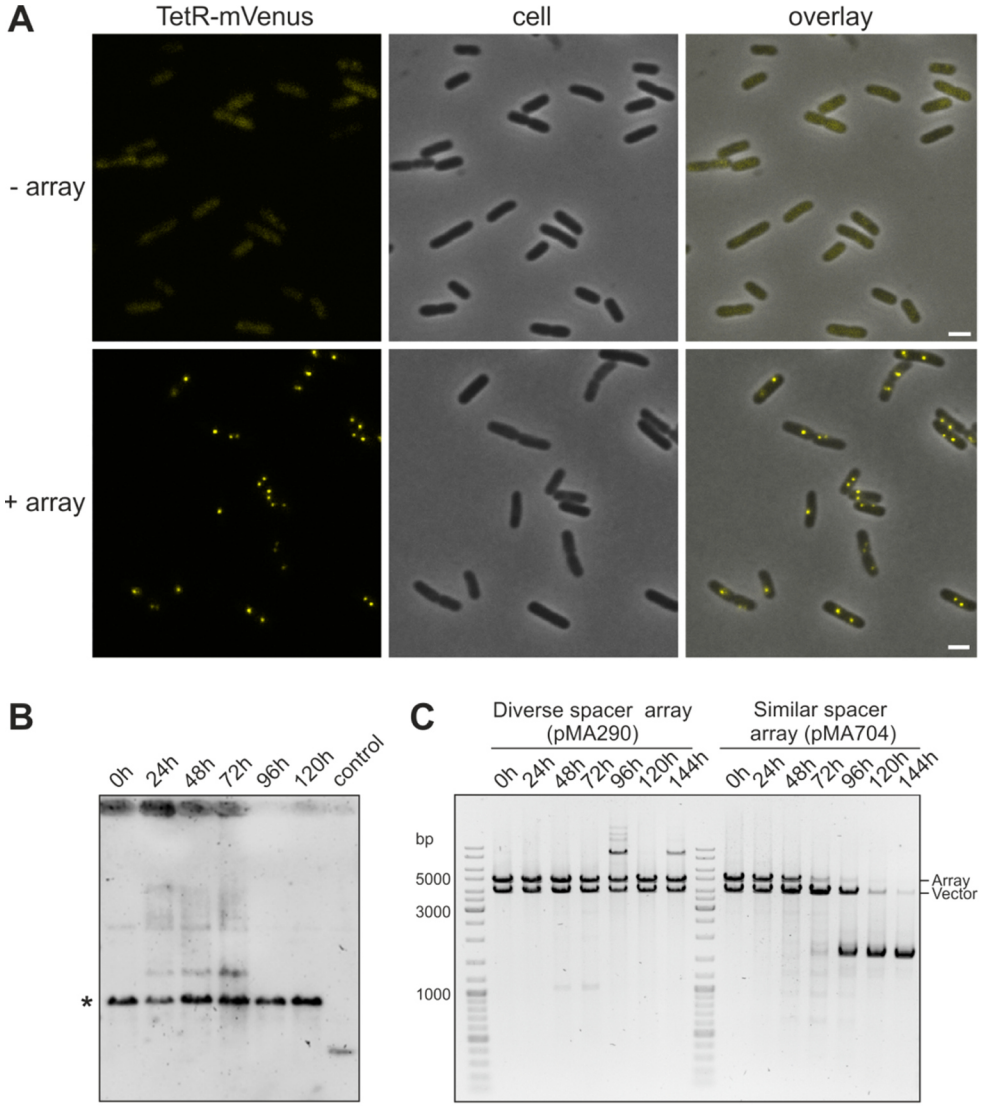


Figure 4. *In vivo* functionality of the constructed FROS array. (A) Fluorescence microscopy of *E. coli* cells harboring a plasmid encoding a TetR-mVenus fusion and either no FROS array (top panel; strain SM100) or a chromosomal integration of the new FROS array (bottom panel; strain SM112). The scale bar is 2 μm. (B) Southern Blot analysis to test stability of the LacO/TetO array during extended cultivation. Chromosomal DNA was isolated from strain SM93 after different time points of cultivation as indicated and cut with *Nde*I. DNA was plotted on a membrane after separation on an agarose gel and the array detected with a probe directed against *lacI*. Black asterisk highlights the size of the array (11247 bps). As control we used DNA from wildtype *E. coli* MG1655 without FROS integration resulting in a fragment of 7520 bps. (C) Genetic stability of a FROS array with MARSeG-designed variable spacer sequences (left) compared to an equivalent array with each spacer sequence being similar to one another. *E. coli* strains DS366 and DS367 carrying plasmids pMA290 and pMA704 respectively were cultivated for the indicated time periods. Plasmid DNA was isolated and cut with *Bpi*I to release the array (4817 bp) and the vector backbone (3968 bp) as indicated.

lacking the FROS array (Figure 4A). These results demonstrate the functionality of the constructed FROS array. A common problem with FROS arrays in which the spacer sequences between the operators have the same sequence and are not diverse as in our case is their genetic instability caused by homologous recombination events. This can lead to undesired size reduction of the respective FROS array. To test if the FROS array presented here is resistant to such recombination events we cultured the *E. coli* strain carrying the array for an extended period of 120 h (see Material section for details). After 24 h periods we measured the array size by Southern Blotting (Figure 4B). No fragments smaller than the expected 11247 bps were detected over the entire test period supporting genetic stability of the constructed FROS array (Figure 4B). To further test if the genetic stability of the FROS array with MARSeG-based design outperforms that of an array with the same operator setup but similar instead of diverse spacer sequences we constructed such a "bad-design-array" with 128 operator sites as above. We cultivated the respective plasmid pMA704 in *E. coli* MG1655 continuously for several days in parallel to cells carrying a similar plasmid with the MARSeG-designed spacers (pMA290). The plasmid DNA was isolated after 24 h intervals and cut with *Bpi*I to release the 4817 bp FROS array. A respective band can be seen at all analyzed time points for the FROS array with MARSeG design (Figure 4C, left). In contrast the FROS array band becomes weaker starting at 48 h of cultivation in case of the similar spacer sequences (Figure 4C, right). In addition, smaller bands occur on the agarose gel at later time points, clearly indicating plasmid size reduction through homologous recombination. We conclude

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

that the FROS array designed using MARSeG has a higher genetic stability as an array with all spacer sequences being similar. It is important to note that stable FROS arrays with variable linkers have been constructed before.[17] However, the assembly approach presented here presents a 3-fold improvement to the earlier work. First, the MARSeG design excludes unwanted restriction sites to omit unwanted cloning of erroneously cut subfragments instead of full fragments. Second, the previous assembly approach required laborious purification of DNA fragments for cloning instead of the plasmid-based MoClo approach used here. Third, the previous approach included a doubling of operator sites in each assembly step while the MoClo hierarchy used here generates a 4-fold increase of sites in each step. This reduces the number of cloning steps which will be more important the bigger the assembly of interest is.

Chromosomal insertions into the *lac* operon are a common approach but are limited to *E. coli* strains that actually carry this gene region. To allow more flexibility and potentially target multiple chromosomal sites we have designed and constructed flanking regions for five additional chromosomal loci in the new MoClo vectors (Table S4, Figure S2). We have used respective vectors to assemble a cassette targeting *tnaA* and could successfully use it for insertion of the FROS array[32] (Table S3 and S4). As for the integration into *lacZ* we observed fluorescence foci showing functional chromosomal integration (data not shown).

The ability to efficiently assemble DNA constructs and integrate them into a host genome is still a main bottleneck in basic and applied molecular biology research. New methods have been developed over the last years allowing multifragment assembly based on different principles. The next step must be the adaptation and optimization of these new approaches to specific systems. Here we present tools for the design and efficient multifragment assembly of genetic constructs for chromosomal insertion. Our new MoClo vectors are fully compatible with previously published MoClo kits of the Marillonnet group. Although we focus on manipulation of the *E. coli* chromosome our approach should be applicable in many bacteria that allow genetic modification *via* homologous recombination. We expect the approach presented here to be especially valuable for the design and construction of synthetic chromosomes which is now technically possible.[1,4,5,33]

MATERIAL AND METHODS

A detailed description of the materials and methods is provided in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00089.

Figures S1, S2; Supplementary Methods; Tables S1−S5 (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: torsten.waldminghaus@synmikro.uni-marburg.de.

Author Contributions

†DS, SM and TS contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully thank Sylvestre Marillonnet (Halle, Germany) for providing MoClo vectors and helpful discussions. Johan Elf (Uppsala, Sweden), Federico Katzen and Xiquan Liang (Thermo Fisher Scientific, Carlsbad, USA), Alexander Böhm (Marburg, Germany), William Margolin (Houston, USA) and Michael L. Kahn (Washington, USA) are acknowledged for providing strains and/or plasmids. We thank Julian Sohl, Joel Eichmann and Patrick Sobetzko from the Waldminghaus lab for helping with experiments and data analysis as well as Nadine Schallopp for excellent technical assistance and the whole working group for fruitful discussions. We are grateful to Manuel Seip for help with setting up the web pages. This work was supported within the LOEWE program of the State of Hesse.

REFERENCES

(1) Schindler, D., and Waldminghaus, T. (2015) Synthetic chromosomes. *FEMS microbiology reviews 39*, 871−891.

(2) 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.

(3) Karas, B. J., Suzuki, Y., and Weyman, P. D. (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Res. 23*, 57−68.

(4) Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., Cai, Y., Zeller, K., Agmon, N., Han, J. S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., London, V., Yeluru, A., Murugan, S., Kandavelou, K., Agier, N., Fischer, G., Yang, K., Martin, J. A., Bilgel, M., Bohutski, P., Boulier, K. M., Capaldo, B. J., Chang, J., Charoen, K., Choi, W. J., Deng, P., DiCarlo, J. E., Doong, J., Dunn, J., Feinberg, J. I., Fernandez, C., Floria, C. E., Gladowski, D., Hadidi, P., Ishizuka, I., Jabbari, J., Lau, C. Y., Lee, P. A., Li, S., Lin, D., Linder, M. E., Ling, J., Liu, J., London, M., Ma, H., Mao, J., McDade, J. E., McMillan, A., Moore, A. M., Oh, W. C., Ouyang, Y., Patel, R., Paul, M., Paulsen, L. C., Qiu, J., Rhee, A., Rubashkin, M. G., Soh, I. Y., Sotuyo, N. E., Srinivas, V., Suarez, A., Wong, A., Wong, R., Xie, W. R., Xu, Y., Yu, A. T., Koszul, R., Bader, J. S., Boeke, J. D., and Chandrasegaran, S. (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science 344*, 55−58.

(5) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O., 3rd, and Venter, J. C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 52−56.

(6) Ma, H., Kunes, S., Schatz, P. J., and Botstein, D. (1987) Plasmid construction by homologous recombination in yeast. *Gene 58*, 201−216.

(7) de Kok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt, D., Shapland, E. B., Serber, Z., Dean, J., Newman, J. D., and Chandran, S. S. (2014) Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol. 3*, 97−106.

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

(9) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One 6*, e16765.

(10) Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T., and Baldwin, G. S. (2015) BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synth. Biol. 4*, 781−787.

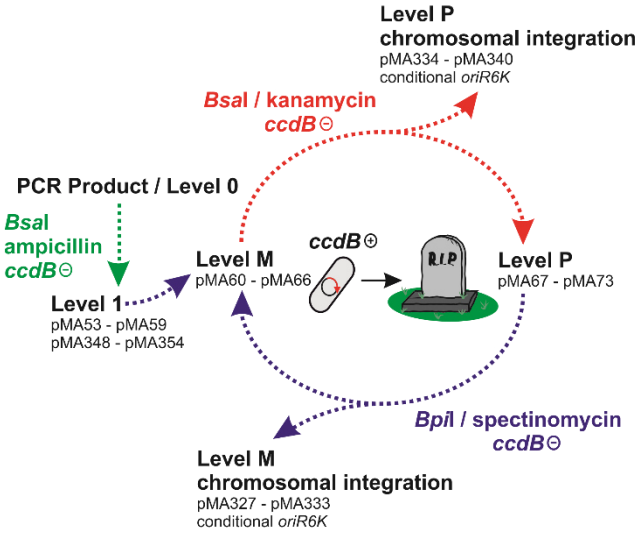
(11) Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012) Fast track assembly of multigene constructs

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

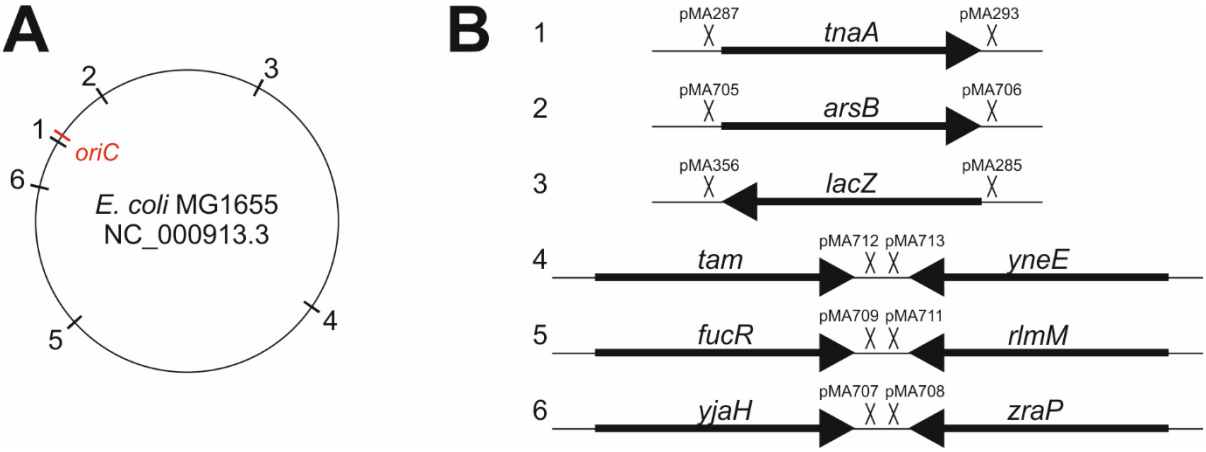
using Golden Gate cloning and the MoClo system. *Bioengineered bugs 3*, 38−43.
(12) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol. 4*, 975−986.
(13) Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014) A golden gate modular cloning toolbox for plants. *ACS Synth. Biol. 3*, 839−843.
(14) Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res. 42*, 13440−13451.
(15) Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One 6*, e19722.
(16) Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A. S. (1996) *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using *lac* operator/repressor recognition. *J. Cell Biol. 135*, 1685−1700.
(17) Lau, I. F., Filipe, S. R., Soballe, B., Okstad, O. A., Barre, F. X., and Sherratt, D. J. (2003) Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol. 49*, 731−743.
(18) Matzke, A. J., Huettel, B., van der Winden, J., and Matzke, M. (2005) Use of two-color fluorescence-tagged transgenes to study interphase chromosomes in living plants. *Plant Physiol. 139*, 1586−1596.
(19) Michaelis, C., Ciosk, R., and Nasmyth, K. (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell 91*, 35−45.
(20) Gordon, G. S., Sitnikov, D., Webb, C. D., Teleman, A., Straight, A., Losick, R., Murray, A. W., and Wright, A. (1997) Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell 90*, 1113−1121.
(21) Dworkin, J., and Losick, R. (2002) Does RNA polymerase help drive chromosome segregation in bacteria? *Proc. Natl. Acad. Sci. U. S. A. 99*, 14089−14094.
(22) Mettrick, K. A., and Grainge, I. (2016) Stability of blocked replication forks *in vivo*. *Nucleic Acids Res. 44*, 657−668.
(23) Thanbichler, M., and Shapiro, L. (2006) MipZ, a spatial regulator coordinating chromosome segregation with cell division in *Caulobacter*. *Cell 126*, 147−162.
(24) Wang, X., Montero Llopis, P., and Rudner, D. Z. (2014) *Bacillus subtilis* chromosome organization oscillates between two distinct patterns. *Proc. Natl. Acad. Sci. U. S. A. 111*, 12877−12882.
(25) Casini, A., Christodoulou, G., Freemont, P. S., Baldwin, G. S., Ellis, T., and MacDonald, J. T. (2014) R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol. 3*, 525−528.
(26) Iverson, S. V., Haddock, T. L., Beal, J., and Densmore, D. M. (2016) CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol. 5*, 99−103.
(27) Bernard, P. (1996) Positive selection of recombinant DNA by CcdB. *Biotechniques 21*, 320−323.
(28) Santos, C. N., Regitsky, D. D., and Yoshikuni, Y. (2013) Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat. Commun. 4*, 2503.
(29) Bentley, W. E., and Quiroga, O. E. (1993) Investigation of subpopulation heterogeneity and plasmid stability in recombinant *Escherichia coli* via a simple segregated model. *Biotechnol. Bioeng. 42*, 222−234.
(30) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640−6645.
(31) Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res. 39*, e82.
(32) Waldminghaus, T., Weigel, C., and Skarstad, K. (2012) Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. *Nucleic Acids Res. 40*, 5465−5476.
(33) Messerschmidt, S. J., Kemter, F. S., Schindler, D., and Waldminghaus, T. (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. *Biotechnol. J. 10*, 302−314.

DOI: 10.1021/acssynbio.6b00089
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Supplementary Figures



Supplementary Figure S1: New vectors and their hierarchy within the MoClo system. Level 1 plasmids are generated by MoClo reactions of level 0 plasmids or PCR products, indicated by green reaction. In level 1 forward or reverse orientation of the same input DNA can be generated based on the destination vector. Level 1 vectors can be combined to level M plasmids with high yields (pMA60-pMA66; *pMB1* origin) or for low background chromosomal integration (pMA327-pMA333; conditional *oriR6K*), indicated by blue reaction. Combination of level M plasmids result in level P, destination plasmids can be chosen for high yields (pMA67-pMA73; *pMB1* origin) or low background chromosomal integration (pMA334-pMA340; conditional *oriR6K*), indicated by red reaction. Level P plasmids can be combined to level M. Level M and level P vectors create an infinite cloning circle. For level M and level P MoClo reactions additional endlinkers are used for the assembly (*1*).



Supplementary Figure S2: Insertion sites on the *E. coli* chromosome to be targeted with the indicated MoClo vectors. A Overview of insertion-site positions on the *E. coli* chromosome. **B** Genetic context of insertion sites. Insertions occur either into non-essential genes (1-3) or in intergenic regions (4-6). Plasmid names of respective MoClo parts are indicated (compare suppl. table S4).

Supplementary Methods

Bacterial strains, plasmids, oligonucleotides, and culture conditions

All strains, plasmids and oligonucleotides used in this study are listed in Table S3-S5. Bacteria were cultured in LB medium under vigorous shaking or on LB plates at 37°C with the desired antibiotic if not indicated otherwise. For microscopic analysis, cultures were grown in AB glucose casamino acids (CAA) (*2*). Antibiotic selection and other ingredients were used at the following concentrations: ampicillin (amp) 100 µg/ml, chloramphenicol (cat) 30 µg/ml, kanamycin (kan) 35 µg/ml, spectionmycin (spec) 100 µg/ml, IPTG = 100 µg/ml and X-Gal = 20 µg/ml. pMA289 was generated by Gibson assembly (*3*) of the *Spe*I linearized plasmid pKG110 with TetR (amplified with primers 1070 + 1065, from pLAU53) and mVenus (amplified with primers 935 + 1064 from pBAD24-LacI-venus).

MoClo assembly

MoClo reactions were carried out according to (*1*) in 25 µl reactions with 40 fmol of each DNA part (100 ng per 4 kb DNA) in T4 Ligase buffer (Promega) with 1 µl T4 Ligase standard or highly concentrated and 1 µl of the desired type IIs endonuclease *Bsa*I (NEB) or *Bpi*I (Thermo Scientific)/*Bbs*I (NEB). Reactions were performed in 150 µl tubes in a PCR cycler using the following program: 300 min 37°C, 20 min 50°C, 10 min 80°C. 5-25 µl of the reaction were transformed into *E. coli* Top10 by heat shock and plated on LB plates with the respective antibiotic.

Amplicons for level 1 MoClo Assembly were generated by standard PCR. The respective primer design was as follows:

Fw-Primer (5´-3´) TTTTAGGAAGGTCTCGGGAG-specific amplicon sequence

Rv-Primer (5´-3´) TTAATCCTTGGTCTCCAGCG-specific amplicon sequence

Details on the construction of the LacO/TetO FROS array and MARSeG are provided in the supplementary methods.

Fluorescence microscopy and data evaluation

Cells were grown in AB glucose CAA to OD_{450} ~ 0.15. 1 ml of the culture was harvested by centrifugation and cells re-suspended in 25 µl fresh AB glucose CAA. 2 µl of the cells were transferred to 1 % agarose pads containing 1 % TAE. Fluorescence microscopy was performed with a Nikon Eclipse Ti-E microscope with a phase-contrast Plan Apo l oil objective (100; numerical aperture, 1.45) with the AHF HC filter set F36-528 YFP (excitation band pass [ex bp] 500/24-nm, beam splitter [bs] 515-nm, and emission [em] bp 535/30-nm filters) and an Argon Ion Laser (Melles Griot). Images were acquired with an Andor iXon3 885 electron-multiplying charge-coupled device (EMCCD) camera. For quantification of fluorescence foci, 20 images were taken for every strain and the first 700 cells were used for further analyses. Images were analyzed by Fiji using the MircobeJ plugin (*4*).

FROS array stability assay

Strain SM93 was incubated in 100 ml LB chloramphenicol for 120h. At an OD_{600} of around 2, a new culture was inoculated 1:10.000 repeatedly. After 24, 48, 72, 96 and 120h samples were taken for Southern blot analysis. Chromosomal DNA was isolated and cut with *Nde*I. The DIG-labeled probe, which specifically binds to *lacI* was amplified with primers 893 and 894 using the PCR DIG probe synthesis kit (Roche).

Construction of the LacO/TetO FROS array

The basic oligonucleotides (791 and 792) were designed by MARSeG and synthesized by IDT (Leuven, Belgium). Annealing and elongation was done in 12 parallel reactions with a temperature gradient (initial denaturation: 95°C 3min; 5 cycles: 95°C 3 min, gradient of 57-62°C 20 sec, 72°C 20 sec; final elongation 72°C 10 min, ∞ 10°C). PCR products were purified and the library enriched by PCR with oligonucleotides 466 and 467. The PCR product with a size of about 322 bp was extracted from an agarose gel. This fragment library was used for seven *BsaI*-based MoClo reactions with vectors pMA53-pMA59, respectively. The reactions were transformed into *E. coli* Top 10 cells and 50 µl were plated on LB media with amp, IPTG and X-Gal. The rest was used to inoculate LB-amp medium for plasmid library purification. In the next step, four libraries of level 1 were combined together in one level M vector. Therefore 4 different *Bpi*I-based MoClo reactions were prepared: 1. pMA61 + level 1 libraries 2,3,4,5 + end linker pICH50914; 2. pMA62 + level 1 libraries 3,4,5,6 + end linker pICH50927; 3. pMA65 + level 1 libraries 6,7,1,2 + end linker pICH50881; 4. pMA66 + level 1 libraries 7,1,2,3 + end

linker pICH50892. Single colonies were analyzed and assemblies named pMA283, pMA281, pMA284 and pMA282, respectively. These 4 fragments were combined to one part into level P in a *Bsa*I-based MoClo reaction with pMA68 and endlinker pICH79277. A positive clone was identified via *BpiI* restriction digestion and named pMA290. Plasmid pMA704 was constructed analog to pMA290 with oligonucleotides 1351 and 1352. Instead of using libraries, individual level 1 vectors were used for the assembly. Respective inserts were verified by sequencing.

For the integration of the array from plasmid pMA290 into the *E. coli* chromosome the 3' flanking region of *lacZ* was amplified with primers 687 + 688 and 689 + 690 and assembled into pMA352 via a *BsaI* based Moclo reaction, resulting in pMA285. The 5' flanking region of *lacZ* was amplified with primers 845 and 846 and assembled into pMA348 and named pMA356**.** Other flanking regions as outlined in figure S2 were cloned accordingly using primers 1059, 1060, 1095, 1096, 1351, 1352, 1419-1434 (Table S5). The chloramphenicol gene with flanking FRT sites was amplified with primer 703 and 704 and assembled into pMA56 which resulted in pMA298. In the final step the following components were assembled in a *Bpi*I-based MoClo reaction: pMA327, endlinker pICH50914, pMA290, pMA289, pMA356, pMA285. The resulting plasmid was named pMA292. Integration into *E. coli* was carried out as described by Datsenko and Wanner (*5*). pMA292 was cut with *BsaI* and the reaction was transformed into *E. coli* AB330. Positive clones were identified by chloramphenicol resistance, spectinomycin sensitivity and white appearance. Integration was verified by southern blot analysis (*6*)(data not shown). Further the Integration into *lacZ* was transferred into strains #1 and #16 of Liang *et al*. (*7*) via P1 transduction and verified by PCR. FRT recombination was carried out via pCP20 based recombination (*8*).

MARSeG

The computer program MARSeG (Motif Avoiding Randomized Sequence Generator) creates a list of degenerated sequences from a template sequence and a list of motifs. Internally, sequences are handled in a 4 bit per DNA character data format, where every bit indicates the possibility of the occurrence of one of the four defined DNA bases in the given DNA character. For example, adenine, cytosine, guanine and thymine are represented by 0001_2, 0010_2, 0100_2 and 1000_2, respectively, while a position where either cytosine or guanine are possible (commonly represented by a S) is represented by 0110_2 and a position, where all of the nucleobases are possible (N) is represented by a 1111_2. This way, DNA characters are represented as sets of possible outcomes of defining this possibly degenerate character. Such an encoding enables set-specific computations such as NAND and AND to assess whether one DNA characters possibilities overlap with those of another. For every generated output sequence and every motif in the list, the program removes all possible occurrences of the motif and removes it by defining one of the degenerated DNA characters, which is randomly

chosen in such a way that the corresponding DNA base in the motif is removed from the set of possible defined DNA bases using an XOR operator. MARSeG will also evaluate all output sequences by defining sets of defined sequences and calculating their pairwise homology and LCS values. Both, the template sequence as well as the motifs can contain degenerated DNA characters. MARSeG is implemented in Java and R, runs on MS Windows OS and is available through our homepage including a detailed manual.

Supplementary Tables

Table S1: DNA motifs to be excluded in sequences designed by MARSeG

DNA sequence	Restriction enzyme	Motif lists		
		I	II	III
GTGCAC	*Apa*LI / *Alw*44I	X	X	X
TCTAGA	*Xba*I	X	X	X
CTCGAG	*Xho*I		X	X
GAAGAC	*Bpi*I		X	X
GGTCTC	*Bsa*I		X	X
GTCTTC	*Bpi*I (rv)		X	X
GAGACC	*Bsa*I (rv)		X	X
GATATC	*Eco*RV		X	X
GAATTC	*Eco*RI		X	X
AAGCTT	*Hind*III		X	X
CCATGG	*Nco*I			X
GCGGGC	*Cac*8I			X
CTGCAG	*Pst*I			X
CAGCTG	*Pvu*II			X
GTCGAC	*Sal*I			X
CCCGGG	*Sma*I / *Xma*I			X
TTTAAA	*Dra*I			X
GCCNNNNNGGC	*Bgl*I			X
CCTNNNNNAGG	*Xag*I / *Eco*NI			X

Table S2: Overview of optimized MoClo plasmids and their respective changes.

Parental vector	Level	Reduced scar	*ccdB* +	*oriR6K*
pICH47802	1	pMA34	pMA53	-
pICH47811	1	pMA35	pMA54	-
pICH47822	1	pMA36	pMA55	-
pICH47831	1	pMA37	pMA56	-
pICH47841	1	pMA38	pMA57	-
pICH47852	1	pMA39	pMA58	-
pICH47861	1	pMA40	pMA59	-
pICH47732	1	pMA341	pMA348	-
pICH47742	1	pMA342	pMA349	-
pICH47751	1	pMA343	pMA350	-
pICH47761	1	pMA344	pMA351	-
pICH47772	1	pMA345	pMA352	-
pICH47781	1	pMA346	pMA353	-
pICH47791	1	pMA347	pMA354	-
pAGM6323	M	-	pMA60	pMA327
pAGM6401	M	-	pMA61	pMA328
pAGM6413	M	-	pMA62	pMA329
pAGM6425	M	-	pMA63	pMA330
pAGM6437	M	-	pMA64	pMA331
pAGM6449	M	-	pMA65	pMA332
pAGM6451	M	-	pMA66	pMA333
pAGM6311	P	-	pMA67	pMA334
pAGM6463	P	-	pMA68	pMA335
pAGM6475	P	-	pMA69	pMA336
pAGM6487	P	-	pMA70	pMA337
pAGM6499	P	-	pMA71	pMA338
pAGM6501	P	-	pMA72	pMA339
pAGM6323	P	-	pMA73	pMA340

Table S3: Strains used in this study.

Strain	Characteristics	Resistance	Reference
#1	*oriC*(wt), *tos1*	-	(*7*)
AB330	*cf*. DY330, *lacZ*+, *gal*+	-	Alexander Böhm
DB3.1	F- *gyrA462 endA1* Δ(*sr1-recA*) *mcrB mrr hsdS20*(rB-, mB-) *supE44 ara-14 galK2 lacY1 proA2 rpsL20*(SmR) *xyl-5* λ- *leu mtl1*	streptomycin	(*9*)
DB3.1λpir	*cf*. DB3.1, λ+	streptomycin	(*10*)
DH5αλpir	F- Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (rK-, mK+) *phoA supE44 thi-1 gyrA96 relA1* λ+	nalidixic acid	(*11*)
DS366	*E. coli* MG1655 with pMA290	kanamycin	this study
DS367	*E. coli* MG1655 with pMA704	kanamycin	this study
SM93	*cf*. #1; *lacZ::*TetO/LacO-array-FRT-cat-FRT	chloramphenicol	this study
SM100	*cf*. #1; + pMA289	chloramphenicol	this study
SM112	*cf*. #1; *lacZ::*TetO/LacO-array-FRT + pMA289	chloramphenicol	this study
TOP10	F- *mcr*A Δ(*mrr-hsdRMS-mcrBC*) Φ80*lac*ZΔM15 Δ*lacX74*	streptomycin	Invitrogen

Strain	Characteristics	Resistance	Reference
	recA1 araD139 Δ(*ara leu*) 7697 *galU galK rpsL* (StrR) *end*A1 *nupG*		
XL1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*[F′ *proAB lacIqZ*ΔM15 *Tn10* (TetR)]	tetracycline, nalidixic acid	Stratagene

Table S4: Plasmids used in this study.

Plasmid	Characteristics	Resistance	Reference
pBad24-LacI-venus	*venus* fluorophor	ampicillin	(*12*)
pCP20	FLP^+, λ $cI857^+$, λ p_R Rep^{ts}	ampicillin, chloramphenicol	(*8*)
pDendra2b	FRT-*cat*-FRT	ampicillin, chloramphenicol	Johan Elf
pKG110	P*nahR*	chloramphenicol	(*13*)
pLAU53	*tetR*	ampicillin	(*14*)
pAGM6311	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6323	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6401	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6413	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6425	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6437	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6449	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6451	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6463	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6475	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6487	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6499	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pAGM6501	Level M MoClo plasmid	spectinomycin	Sylvestre Marillonett
pAGM6513	Level P MoClo plasmid	kanamycin	Sylvestre Marillonett
pDONR201	Gateway Donor vector	kanamycin	Invitrogen
pICH47732	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47742	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47751	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47761	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47772	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47781	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47791	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47802	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47811	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47822	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47831	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47841	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47852	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH47861	Level 1 MoClo plasmid	ampicillin	(*1*)
pICH50881	Level M MoClo endlinker	ampicillin	(*1*)
pICH50892	Level M MoClo endlinker	ampicillin	(*1*)
pICH50914	Level M MoClo endlinker	ampicillin	(*1*)
pICH50927	Level M MoClo endlinker	ampicillin	(*1*)
pICH79277	Level P MoClo endlinker	ampicillin	(*1*)

Plasmid	Characteristics	Resistance	Reference
pMA34	pICH47802 reduced scar	ampicillin	this study
pMA35	pICH47811 reduced scar	ampicillin	this study
pMA36	pICH47822 reduced scar	ampicillin	this study
pMA37	pICH47831 reduced scar	ampicillin	this study
pMA38	pICH47841 reduced scar	ampicillin	this study
pMA39	pICH47852 reduced scar	ampicillin	this study
pMA40	pICH47861 reduced scar	ampicillin	this study
pMA53	pMA34 + *ccdB*	ampicillin	this study
pMA54	pMA35 + *ccdB*	ampicillin	this study
pMA55	pMA36 + *ccdB*	ampicillin	this study
pMA56	pMA37 + *ccdB*	ampicillin	this study
pMA57	pMA38 + *ccdB*	ampicillin	this study
pMA58	pMA39 + *ccdB*	ampicillin	this study
pMA59	pMA40 + *ccdB*	ampicillin	this study
pMA60	pAGM6323 + *ccdB*	spectinomycin	this study
pMA61	pAGM6401 + *ccdB*	spectinomycin	this study
pMA62	pAGM6413 + *ccdB*	spectinomycin	this study
pMA63	pAGM6425 + *ccdB*	spectinomycin	this study
pMA64	pAGM6437 + *ccdB*	spectinomycin	this study
pMA65	pAGM6449 + *ccdB*	spectinomycin	this study
pMA66	pAGM6451 + *ccdB*	spectinomycin	this study
pMA67	pAGM6311 + *ccdB*	kanamycin	this study
pMA68	pAGM6463 + *ccdB*	kanamycin	this study
pMA69	pAGM6475 + *ccdB*	kanamycin	this study
pMA70	pAGM6487 + *ccdB*	kanamycin	this study
pMA71	pAGM6499 + *ccdB*	kanamycin	this study
pMA72	pAGM6501 + *ccdB*	kanamycin	this study
pMA73	pAGM6513 + *ccdB*	kanamycin	this study
pMA281	MoClo assembly of 4 parts into pMA62. Level 1 libraries: pMA55, pMA56, pMA57, pMA58, endlinker: pICH50927	spectinomycin	this study
pMA282	MoClo assembly of 4 parts into pMA66. Level 1 libraries: pMA59, pMA53, pMA54, pMA55, endlinker: pICH50892	spectinomycin	this study
pMA283	MoClo assembly of 4 parts into pMA61. Level 1 libraries: pMA54, pMA55, pMA56, pMA57, endlinker: pICH50914	spectinomycin	this study
pMA284	MoClo assembly of 4 parts into pMA65. Level 1 libraries: pMA58, pMA59, pMA53, pMA54, endlinker: pICH50881	spectinomycin	this study
pMA285	500 bp *lacZ* 3´flank PCR fragment in pMA352	ampicillin	this study
pMA287	500 bp tnaA 5´ flank PCR fragment in pMA348	ampicillin	this study
pMA289	pKG110 *tetR-mVenus*	chloramphenicol	this study
pMA290	MoClo assembly of 4 parts into	kanamycin	this study

Plasmid	Characteristics	Resistance	Reference
	pMA68: pMA281-pMA284 endlinker: pICH79277		
pMA292	MoClo assembly of 4 parts into pMA327: pMA356, pMA290, pMA298, pMA285 endlinker: pICH50914	chloramphenicol, spectinomycin	this study
pMA293	500 bp tnaA 3´ flank PCR fragment in pMA352	ampicillin	this study
pMA298	FRT-*cat*-FRT PCR fragment in pMA56	ampicillin, chloramphenicol	this study
pMA299	MoClo assembly of 4 parts into pMA327: pMA287, pMA290, pMA293, pMA298 endlinker pICH50914	chloramphenicol, spectinomycin	this study
pMA327	pMA60 *oriR6K*	spectinomycin	this study
pMA328	pMA61 *oriR6K*	spectinomycin	this study
pMA329	pMA62 *oriR6K*	spectinomycin	this study
pMA330	pMA63 *oriR6K*	spectinomycin	this study
pMA331	pMA64 *oriR6K*	spectinomycin	this study
pMA332	pMA65 *oriR6K*	spectinomycin	this study
pMA333	pMA66 *oriR6K*	spectinomycin	this study
pMA334	pMA67 *oriR6K*	kanamycin	this study
pMA335	pMA68 *oriR6K*	kanamycin	this study
pMA336	pMA69 *oriR6K*	kanamycin	this study
pMA337	pMA70 *oriR6K*	kanamycin	this study
pMA338	pMA71 *oriR6K*	kanamycin	this study
pMA339	pMA72 *oriR6K*	kanamycin	this study
pMA340	pMA73 *oriR6K*	kanamycin	this study
pMA341	pICH47732 reduced scar	ampicillin	this study
pMA342	pICH47742 reduced scar	ampicillin	this study
pMA343	pICH47751 reduced scar	ampicillin	this study
pMA344	pICH47761 reduced scar	ampicillin	this study
pMA345	pICH47772 reduced scar	ampicillin	this study
pMA346	pICH47781 reduced scar	ampicillin	this study
pMA347	pICH47791 reduced scar	ampicillin	this study
pMA348	pMA341 + *ccdB*	ampicillin	this study
pMA349	pMA342 + *ccdB*	ampicillin	this study
pMA350	pMA343 + *ccdB*	ampicillin	this study
pMA351	pMA344 + *ccdB*	ampicillin	this study
pMA352	pMA345 + *ccdB*	ampicillin	this study
pMA353	pMA346 + *ccdB*	ampicillin	this study
pMA354	pMA347 + *ccdB*	ampicillin	this study
pMA356	500 bp *lacZ* 5´flank PCR fragment in pMA349	ampicillin	this study
pMA704	*cf.* pMA290 with similar spacer sequences	kanamycin	this study
pMA705	500 bp *arsB* 5´flank PCR fragment in pMA349	ampicillin	this study
pMA706	500 bp *arsB* 3´ flank PCR fragment in pMA352	ampicillin	this study

Plasmid	Characteristics	Resistance	Reference
pMA707	540 bp *yjaH-zraP* 5´flank PCR fragment in pMA349	ampicillin	this study
pMA708	394 bp *yjaH-zraP* 3´ flank PCR fragment in pMA352	ampicillin	this study
pMA709	395 bp *fucR-rlmM* 5´flank PCR fragment in pMA349	ampicillin	this study
pMA711	335 bp *fucR-rlmM* 3´ flank PCR fragment in pMA352	ampicillin	this study
pMA712	467 bp tam-yneE 5´flank PCR fragment in pMA349	ampicillin	this study
pMA713	465 bp tam-yneE 3´ flank PCR fragment in pMA352	ampicillin	this study
pUC18-R6KT-miniTn7T-egfp	*oriR6K*	ampicillin, chloramphenicol	(*15*)

Table S5: Oligonucleotides used in this study.

Name	Sequence (5´ → 3´)
161	CGTAATATCCAGCTGAACGG
260	CATTAATGCAGCTGGCACGAC
411	TTGTCTTCACAGAGTGGGGCCGCGGAACCCCTATTTGTTT
412	TTGTCTTCTGCACGAAGTGGGCCAGGAACCGTAAAAAGG
413	CCACTTCGTGCAGAAGACAA
414	GCCCCACTCTGTGAAGACAATGCCAGCGTGAGACCGCAGCTGGC
415	GCCCCACTCTGTGAAGACAAGCAAAGCGTGAGACCGCAGCTGGC
416	GCCCCACTCTGTGAAGACAAACTAAGCGTGAGACCGCAGCTGGC
417	GCCCCACTCTGTGAAGACAATTACAGCGTGAGACCGCAGCTGGC
418	GCCCCACTCTGTGAAGACAACAGAAGCGTGAGACCGCAGCTGGC
419	GCCCCACTCTGTGAAGACAATGTGAGCGTGAGACCGCAGCTGGC
420	GCCCCACTCTGTGAAGACAAGAGCAGCGTGAGACCGCAGCTGGC
466	TTTTAGGAAGGTCTCGGGAG
467	TTAATCCTTGGTCTCCAGCG
573	TATAGTCCTGTCGGGTTTCG
574	ATTCCACCCGCGTGAAGAAG
580	AATATTGAAAAAGGAAGAGTATGGAGAAAAAAATCACTGGATATACCACC
581	GAGTAAACTTGGTCTGACAGTCATCGCAGTACTGTTGTATTCATTAAGC
582	CTGTCAGACCAAGTTTACTCATATATACTTTAGATTG
583	ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
687	AAGGTCTCGGGAGGTCTGCGTCTGGCTGGCTG
688	GCGGTCTCTGTCCTCGGTATCGTCGTATCCCAC
689	CCGGTCTCAGGACAGCTCATGTTATATCCCGCC
690	TTGGTCTCCAGCGTCACTGCCCGCTTTCCAGTC
703	AAGGTCTCGGGAGGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTCATTTAAATGG
704	TTGGTCTCCAGCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGGCGCGCCTACCTGTGACGG
722	CTATGCGGCATCAGAGCAGATTG
723	TTAAGCCAGCCCCGACACC
724	CAATCTGCTCTGATGCCGCATAGTCGCGTGGATCCGGCTTACTAAAAGC
725	TAACGGAGACAGGCACACTGGCCATATCGG
726	CCAGTGTGCCTGTCTCCGTTATCGG
727	GGTGTCGGGGCTGGCTTAATTATATTCCCCAGAACATCAGG
747	TTCTCGCGGTATCATTGCAGC
748	GCTGCAATGATACCGCGAGAACCACGCTCACCGGCTC

Name	Sequence (5´ → 3´)
785	TTTACGGTTCCTGCACTCTG
786	TTCAGCAGCCCGGCCACTTC
787	GAAGTGGCCGGGCTGCTGAATGAGCGTCGCAAAGGCCGAG
788	ACACTTAACGGCTGACATGGGATCTTTTCTACGGGGTCTG
789	CCATGTCAGCCGTTAAGTG
790	CAGAGTGCAGGAACCGTAAAAACCTGTTGATAGTACGTACTAAGC
791	TTTTAGGAAGGTCTCGGGAGNWNCVNTCCCTATCAGTGATAGAGADNTNCNNNNNAATTGTGAGCGGAT AACAATTNNNNATNNNNATNNNTCCCTATCAGTGATAGAGADCHNNNNNGNAATTGTGAGCGGATAACA ATTNNNNNTANNTDNNRNTCGTGGGTAVGANTCAAHGSTVNAATTAGT
792	TTAATCCTTGGTCTCCAGCGNNGSNNNAATTGTTATCCGCTCACAATTNNNNGNANYNTCTCTATCACT GATAGGGANNNWANKNDYGNNNNAATTGTTATCCGCTCACAATTHNNNNNGNANTCTCTATCACTGATA GGGAYNNNNYANNNACTAATTNBASCDTTGANTCBTACCCACGA
811	CGAAATACGGGCAGACATGG
838	GCCCCACTCTGTGAAGACAATGCCGGAGTGAGACCGCAGCTGGC
839	GCCCCACTCTGTGAAGACAAGCAAGGAGTGAGACCGCAGCTGGC
840	GCCCCACTCTGTGAAGACAAACTAGGAGTGAGACCGCAGCTGGC
841	GCCCCACTCTGTGAAGACAATTACGGAGTGAGACCGCAGCTGGC
842	GCCCCACTCTGTGAAGACAACAGAGGAGTGAGACCGCAGCTGGC
843	GCCCCACTCTGTGAAGACAATGTGGGAGTGAGACCGCAGCTGGC
844	GCCCCACTCTGTGAAGACAAGAGCGGAGTGAGACCGCAGCTGGC
845	AAGGTCTCGGGAGCCAACACAGCCAAACATCCG
846	TTGGTCTCCAGCGTAATAACCGGGCAGGCCATG
893	CGCTAACAGCGCGATTTGCTGGTG
894	GGCCAGCCACGTTTCTGCGAAAAC
1059	AAGGTCTCGTGCCTAGCCATCACCAGAGCCAAACCG
1060	TTGGTCTCCTTGCGAATGGTGTATTGATAACCAAAG
1095	AAGGTCTCGGGAGCGGCAGACCAGTTCCCGGCACAG
1096	TTGGTCTCCAGCGACCGGCAAGATCAACAGGTAAAGC
1351	TTTTAGGAAGGTCTCGGGAGTGTTGCTCCCTATCAGTGATAGAGATCAAGTCCATAATTGTGAGCGGAT AACAATTTCAAGTCCATGTTGCTCCCTATCAGTGATAGAGATCAAGTCCATAATTGTGAGCGGATAACA ATTTCAAGTCCATGTTGCTCGTGGGTATGACTCAACGCTGTAATTAGT
1352	TTAATCCTTGGTCTCCAGCGGACTTGAAATTGTTATCCGCTCACAATTATGGACTTGATCTCTATCACT GATAGGGAGCAACATGGACTTGAAATTGTTATCCGCTCACAATTATGGACTTGATCTCTATCACTGATA GGGAATGGACTTGAACTAATTACAGCGTTGAGTCATACCCACGA
1419	AAGGTCTCGGGAGCCTCTGCACTTACACATTCG
1420	TTGGTCTCCAGCGAATGCCTCCCGGATAAAACAC
1421	AAGGTCTCGGGAGTGAGATACTGATATGAGCAACATTACC
1422	TTGGTCTCCAGCGGCACTTTTCTAACAACCTGTGG
1423	AAGGTCTCGGGAGGCATCTGGCGAAAAGACAGC
1424	TTGGTCTCCAGCGATCAGAACGTTTTCTCGTTGGG
1425	AAGGTCTCGGGAGTTACCAGTGGCCCATACCC
1426	TTGGTCTCCAGCGGATGGCGCTTTCAGCAATGG
1427	AAGGTCTCGGGAGACCAATAGCCATCCGATTTGCC
1428	TTGGTCTCCAGCGTCTCCGGCCTGCTACCCTTT
1429	AAGGTCTCGGGAGCGACGCCCCGGCCTTGCCTG
1430	TTGGTCTCCAGCGAAATTCCGTCCGACGCGCAG
1431	AAGGTCTCGGGAGGCTAATGCCTCACTGCAATGG
1432	TTGGTCTCCAGCGATTTACTCCATACGCCGGGC
1433	AAGGTCTCGGGAGATCACGTCAGCTGGTAATGAC
1434	TTGGTCTCCAGCGCTTCGAACTCTCCAGCTAACC

Supplementary References

1. Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs, *PLoS One 6*, e16765.
2. Messerschmidt, S. J., Kemter, F. S., Schindler, D., and Waldminghaus, T. (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*, *Biotechnol J 10*, 302-314.
3. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature methods 6*, 343-345.
4. Jiang, C., Brown, P. J., Ducret, A., and Brun, Y. V. (2014) Sequential evolution of bacterial morphology by co-option of a developmental regulator, *Nature 506*, 489-493.
5. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc Natl Acad Sci U S A 97*, 6640-6645.
6. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J Mol Biol 98*, 503-517.
7. Liang, X., Baek, C. H., and Katzen, F. (2013) *Escherichia coli* with two linear chromosomes, *ACS synthetic biology 2*, 734-740.
8. Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant, *Gene 158*, 9-14.
9. Bernard, P., Kezdy, K. E., Van Melderen, L., Steyaert, J., Wyns, L., Pato, M. L., Higgins, P. N., and Couturier, M. (1993) The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase, *J Mol Biol 234*, 534-541.
10. House, B. L., Mortimer, M. W., and Kahn, M. L. (2004) New recombination methods for *Sinorhizobium meliloti* genetics, *Appl Environ Microbiol 70*, 2806-2815.
11. Miller, V. L., and Mekalanos, J. J. (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*, *J Bacteriol 170*, 2575-2583.
12. Hammar, P., Wallden, M., Fange, D., Persson, F., Baltekin, O., Ullman, G., Leroy, P., and Elf, J. (2014) Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation, *Nat Genet 46*, 405-408.
13. Pazos, M., Natale, P., Margolin, W., and Vicente, M. (2013) Interactions among the early *Escherichia coli* divisome proteins revealed by bimolecular fluorescence complementation, *Environ Microbiol 15*, 3282-3291.
14. Lau, I. F., Filipe, S. R., Soballe, B., Okstad, O. A., Barre, F. X., and Sherratt, D. J. (2003) Spatial and temporal organization of replicating *Escherichia coli* chromosomes, *Mol Microbiol 49*, 731-743.
15. Godeke, J., Paul, K., Lassak, J., and Thormann, K. M. (2011) Phage-induced lysis enhances biofilm formation in *Shewanella oneidensis* MR-1, *ISME J 5*, 613-626.

3.4 Using experimental chromosome construction to study functional interactions between segregation and DNA mismatch repair in *Escherichia coli*

Chromosome Maintenance Systeme werden traditionell durch *in vitro* Studien und die Verteilung von Motiven auf dem Chromosom analysiert. *In vivo* Analysen gestalten sich schwierig, da Sequenzen, wie beispielsweise das GATC-Sequenzmotiv, sehr häufig auf dem Chromosom vorkommen und somit nicht global deletiert werden können. Das GATC-Sequenzmotiv wird durch die Dam-Methyltransferase methyliert. Hemi-methylierte GATC-Sequenzen, die durch die DNA-Replikation entstehen, sind zum einen Erkennungssequenz für SeqA, welches eine Re-methylierung blockiert, und zum anderen kann MutH, ein *DNA mismatch* Reparatur Protein, an hemi-methylierten GATCs parentale von neusynthetisierter DNA unterscheiden. Ob SeqA und MutH miteinander interagieren, konnte bisher nicht gezeigt werden.

Dieses Kapitel kombiniert das in dieser Arbeit in Kapitel 3.1 und 3.2 etablierte synVicII mit der Herstellung degenerierter, Motiv-freier DNA-Sequenzen und deren DNA-Assemblierung aus Kapitel 3.3, um systematisch die Interaktion von SeqA und der *DNA mismatch* Reparatur zu analysieren. Dazu wurden drei sich in der Verteilung von GATCs unterscheidende, synthetische sekundäre Chromosomen mit einer Größe von fast 100 kb durch das MoClo-System assembliert, um anhand dieser die Mutationsraten der einzelnen synVicIIs vergleichend zu analysieren. synVicII ohne GATCs dient als Kontrolle, SeqA und MutH können die DNA-Sequenz außerhalb des synVicII-Rückgrat nicht binden. Die DNA-Sequenz des synVicII mit weiten Abständen zwischen den einzelnen GATCs kann lediglich MutH binden. Wohingegen an den GATC-Clustern des dritten synVicII sowohl MutH als auch SeqA binden können. Die Sequenzen der synVicIIs wurde mittels SMRT Sequenzierungsdaten *de novo* assembliert und die erhaltenen kinetischen Daten belegen, dass die GATCs methyliert sind. SeqA *ChIP-Seq* Daten belegen, dass ein SeqA-Binden *in vivo* nur in dem synVicII mit GATC-Clustern erfolgt. Vergleichende Analyse von DNA-Punktmutationen der drei synVicII-Varianten ermöglichen es dadurch erstmals *in vivo* eine mögliche Interaktion von SeqA und der *DNA mismatch* Reparatur zu belegen. Erste Analysen konnten eine erhöhte Mutationsfrequenz für das synVicII ohne GATCs zeigen. Im Vergleich dazu war die Anzahl an Mutationen in den synVicIIs mit GATCs deutlich geringer, aber zwischen den Replikons sehr ähnlich. Zukünftige vergleichende Analysen der synVicIIs werden zeigen, ob SeqA und die *DNA mismatch* Reparatur funktionell interagieren oder nicht.

Daniel Schindler hat unter Rücksprache mit Torsten Waldminghaus die Studie konzipiert, die Experimente durchgeführt und ausgewertet. Das Manuskript wurde von Torsten Waldminghaus und Daniel Schindler verfasst.

Manuscript in preparation

Using experimental chromosome construction to study functional interactions between segregation and DNA mismatch repair in *Escherichia coli*

Daniel Schindler and Torsten Waldminghaus

Abstract

For many centuries, genes or promoters have been studied by functional characterization of respective mutant versions. Such an approach would certainly help to study chromosome biology. Emerging DNA-assembly techniques now make the construction of larger replicons feasible and open the door for experimental chromosome construction. As a proof of principle, we present here the design and assembly of three synthetic secondary chromosomes of about 100 kbp. The rationale design generates differential binding of the two proteins SeqA and MutH, which share the common target sequence GATC, to the three different chromosomes. The two proteins are involved in chromosome segregation and mismatch repair, respectively. Their functional interaction remains largely unexplored. The three synthetic secondary chromosomes were designed to allow binding of (i) none of both proteins, (ii) only MutH or (iii) SeqA and MutH. Measuring the mutation rates on the three replicons showed that SeqA and MutH appear to act independent of one another and suggest spatial separation of the underlying mechanisms.

Introduction

Chromosomes carry the genetic information in cells of all domains of life and deficiency in maintaining their integrity result in severe effects such as cancer. Chromosomes need to be replicated and segregated within the cell cycle and occurring errors need to be repaired (Badrinarayanan *et al.* 2015). Respective chromosome maintenance systems have evolved (Touzain *et al.* 2011). However, their complete understanding still lacks important insights. In the model organism *Escherichia coli* two important chromosome maintenance systems are based on the DNA sequence motif GATC (Lobner-Olesen *et al.* 2005; Marinus and Lobner-Olesen 2014). This motif is unique as it is methylated at the N^6 position of both adenines in the palindromic sequence by the Dam methyltransferase (Marinus and Lobner-Olesen 2014). Since only un-modified nucleotides are incorporated during the semiconservative process of DNA replication, GATCs will be hemi-methylated (only the parental strand is methylated) for a short time after passage of the replication fork. These hemi-methylated GATC sites are specific binding targets for the MutH protein, a key player in the DNA mismatch repair (MMR) system (Marinus 2012). An occurring mismatch is first bound by a dimer of the MutS protein which then recruits MutL (Su and Modrich 1986; Acharya *et al.* 2003). These two proteins form a complex with MutH at the nearest hemi-methylated GATC (Welsh *et al.* 1987; Cooper *et al.* 1993; Grilley *et al.* 1993). MutH detects the methylated DNA strand as template and cuts specifically the un-methylated newly synthesized strand. UvrD helicase unwinds the DNA and the nascent strand is degraded by an exonuclease (Matson 1986; Viswanathan and Lovett 1998; Yamaguchi *et al.* 1998). Resynthesis is mediated by DNA polymerase III holoenzyme and a DNA ligase finally seals the remaining nick (Lahue *et al.* 1989; Lehman 1974; Nandakumar *et al.* 2007).

Besides MutH, a second protein binds specifically to hemi-methylated GATC sites, namely SeqA (Waldminghaus and Skarstad 2009). This protein was found to sequester the replication origin *oriC* to inhibit early re-initiation (Lu *et al.* 1994; Slater *et al.* 1995). Sequestration is mediated by a high number of GATC sites within *oriC*. SeqA binds and oligomerizes on the hemi-methylated GATCs occurring after initiation of DNA replication and remains bound for about one third of the cell cycle (Bach *et al.* 2003; Lu *et al.* 1994; Slater *et al.* 1995). SeqA blocks the initiator protein DnaA from binding to DnaA boxes within *oriC* (Nievera *et al.* 2006; Waldminghaus and Skarstad 2009). Binding of SeqA is not limited to the replication origin but occurs throughout the genome (Sanchez-Romero *et al.* 2010; Waldminghaus *et al.* 2012). SeqA binds as dimer to pairs of hemi-methylated GATCs not to far away from one another (Slater *et al.* 1995; Brendler and Austin 1999). Preferred binding is on GATC pairs with distances placing them on the same phase of the double helix (Brendler and Austin 1999). The specificity of SeqA to

hemi-methylated GATCs makes its binding dynamically, tracking the replication forks (Waldminghaus *et al.* 2012; Joshi *et al.* 2013). At the replication fork, SeqA is thought to contribute to organization and segregation of the new DNA and to sister chromosome cohesion (Joshi *et al.* 2013).

With two proteins, SeqA and MutH, having the same target DNA motif the question arises if these two factors compete with one another or interact at the newly synthesized DNA. However, our knowledge is limited about the functional relation between chromosome organization and mismatch repair mediated by SeqA and MutH, respectively. Interesting observations in this context are that overproduction of SeqA leads to an inhibition of MMR and it is thought that *seqA* deletion strains have an increased mutation rate as it mimics a *dam* overexpression (Yang *et al.* 2004; Herman and Modrich 1981). However, the multiple functions of SeqA in chromosome replication timing and segregation result in pleiotropic effects of its depletion or overexpression preventing specific conclusions. Here we present an experimental chromosome construction approach to study the functional interaction of SeqA and MutH in *E. coli*. A set of three synthetic secondary chromosomes was designed to allow binding of (i) none of both proteins, (ii) only MutH or (iii) SeqA and MutH. Functional characterization of strains carrying the respective chromosomes led to new insights into chromosome biology in bacteria.

Results

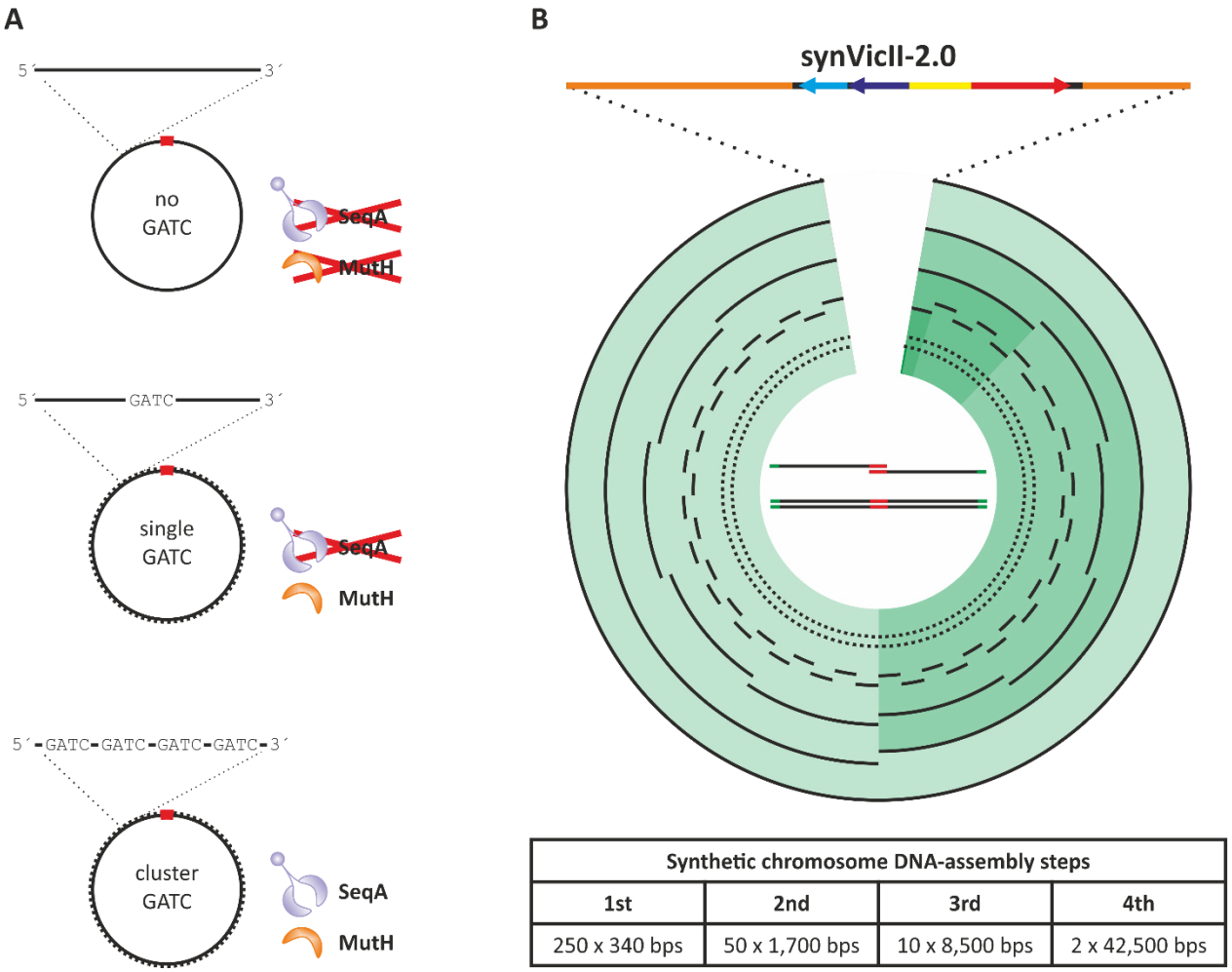
Design and construction of three synthetic secondary chromosomes

Although much is known about the role of SeqA in DNA replication and chromosome organization and MutH in DNA mismatch repair, little is known about their functional interrelation (Waldminghaus and Skarstad 2009; Marinus 2012). A functional interaction of some kind is expected because both proteins share hemi-methylated GATCs as common DNA binding site. The goal of this study was the construction of synthetic secondary chromosomes for which the design determines the binding of either (i) none of both proteins, (ii) only MutH or (iii) SeqA and MutH. We considered the fact that MutH can bind to individual GATCs while SeqA binds as dimer and consequently needs two GATC sites in near proximity at the same DNA strand. The absence of GATCs should abolish binding of both proteins to the respective DNA. Secondary chromosome designs were generated accordingly (i) without GATC sites, (ii) with distantly separated single GATCs or (iii) with clusters of four closely spaced GATCs (Fig. 1A). As backbone for the synthetic secondary chromosome we chose synVicII. This replicon is based on the replication origin of the secondary chromosome of *Vibrio cholerae* and has been carefully characterized to replicate chromosome-like in *E. coli* (Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision). As size of

the new synthetic secondary chromosomes we aimed at roughly 100,000 bps. The chromosome designs exclude genes anywhere except in the synVicII backbone because effects on gene function could result in unwanted phenotypic effects. Thus, the actual chromosome sequence should be random except the GATC sites. We have developed the computer program MARSEG previously for the design of fully synthetic sequence libraries (Schindler *et al.* 2016). The program was used to design three different basic chromosome-building blocks of 340 bps with either no, one or four GATC sites (Fig. S1). Respective DNA oligonucleotide libraries where cloned into entry vectors of the MoClo DNA assembly system (Fig. S1, (Schindler *et al.* 2016)). Systematic arrangement of these building blocks results in the final design as outlined in Figure S1 and S2. Sets of five building blocks were assembled into destination vectors starting with 250 fragments of about 340 bps (Fig. 1B). Further assembly was hierarchical with a final assembly of two chromosome halves and a MoClo compatible version of synVicII (Fig. 1B, (Messerschmidt *et al.* under revision)). The first two assembly steps (340 and 1,700 bps) were carried out with libraries of plasmids while later steps were based on individual fragments. A portion of the synVicII backbone only needed for construction purposes was removed after transfer of the three synthetic secondary chromosomes into a wildtype *E. coli* MG1655 strain as described previously (Messerschmidt *et al.* under revision).

Successful assemblies were confirmed by restriction enzyme cutting and agarose gel electrophoresis (data not shown). Single molecule DNA sequencing (SMRT) was applied to determine the full sequence of the three synthetic secondary chromosomes (Fig. 2). The first synthetic chromosome, named synVicII-noMo (for no motif), is indeed completely free of GATC sites except in the replicon backbone region (Fig. 2A and Fig S3). The second synthetic chromosome, named synVicII-oneGATC, carries single GATC sites with a highly regular spacing of about 700 bps (Fig. 2B and Fig. S2). The third chromosome, named synVicII-fourGATC, comprises regularly spaced clusters of mostly four and sometimes three GATC sites (Fig. 2C and Fig. S2). SMRT data analysis indicates that the Adenines of the GATC sequence in the synthetic secondary chromosomes are methylated by dam as expected (data not shown).

Sequence diversity is a critical factor for replicon integrity. The above design and assembly scheme should result in building blocks with diverse sequences to inhibit homologous recombination. To study this further, we reassembled the individual building blocks of the synthetic chromosomes computationally and calculated the pairwise homology and the longest common substring (Fig. S3). Overall, the building blocks are considerably diverse and sufficient to block homologous recombination. One exception is a 1.7 kbp fragment occurring in synVicII-noMo which probably occurred by chance during library-based assembly.



Synthetic chromosome DNA-assembly steps			
1st	**2nd**	**3rd**	**4th**
250 x 340 bps	50 x 1,700 bps	10 x 8,500 bps	2 x 42,500 bps

Figure 1: Conception and construction of synthetic secondary chromosomes to study chromosome maintenance systems. **A** The three synthetic chromosome designs including the design of the GATC distribution within the sequences respectively. GATC scattering of the two synthetic chromosomes is indicated by black bars. In addition, binding properties of SeqA and MutH for each chromosome is illustrated (red cross = no binding). **B** The chromosomes are constructed from libraries generated by three different pairs of oligonucleotides symbolized by inner single stranded, annealed oligonucleotides and resulting double stranded DNA-fragments (see Material and Methods and Fig S1 for details (Schindler *et al.* 2016)). Consecutive, ordered assemblies are performed in five rounds (inner to outer circle) to generate the synthetic secondary chromosomes, respective steps are shaded in green colors and synVicII-2.0 genes are highlighted in different colors (for details see (Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision; Schindler *et al.* 2016)). The total number of DNA-fragments for each assembly step is represented in the table below the scheme.

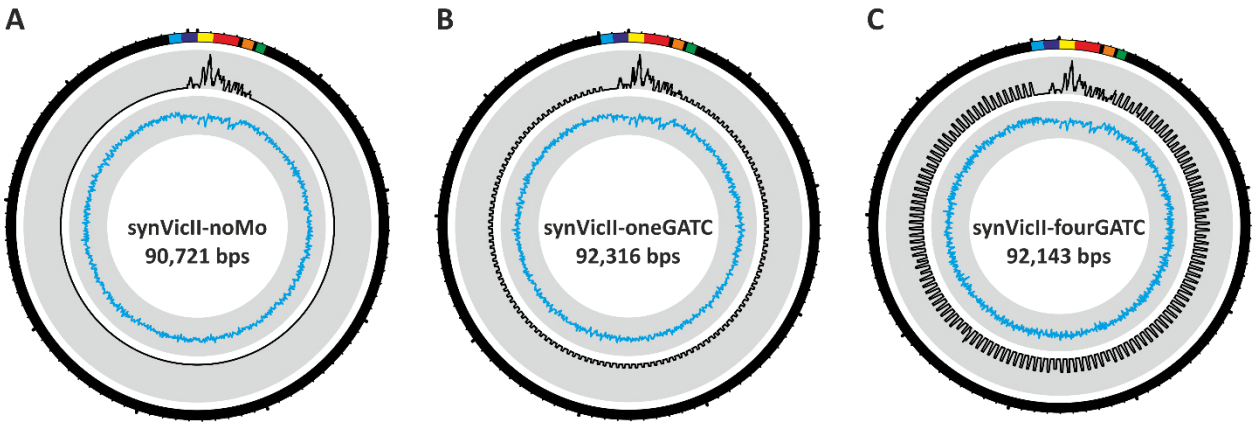


Figure 2: Sequence property illustration of the synthetic secondary chromosome set. The three synthetic chromosomes are illustrated by the outer circle (black) with **A** zero GATCs, **B** single GATCs and **C** clusters of GATCs. Sizes and names of the respective final assembly are indicated. The genes of synVicII-2.0 minimal replicon are highlighted in color (for details see (Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision)). Sequence properties of each finally sequenced replicon is visualized by inner rings. The 250 bp GATC moving window distribution (black) in 50 bp steps and mean GC-content (blue) in a 100 bp window is presented by inner circles. Size of chromosomes is indicated by black lines 10 kb lines accentuated bolt. Chromosomes were plotted using circos (Krzywinski *et al.* 2009).

SeqA binding to the synthetic secondary chromosomes

The rational of the three-different synthetic DNA sequence designs above was that only one (synVicII-fourGATC) is able to bind the SeqA protein and the others don't because they either lack any GATC site (synVicII-noMo) or have only far separated GATCs (synVicII-oneGATC) to prohibit binding of a dimeric SeqA. To test if our assumptions were correct we performed ChIP-Seq experiments with *E. coli* MG1655 strains carrying one of the synthetic secondary chromosomes respectively. As expected, no SeqA binding was detected on synVicII-noMo outside the backbone region where some GATC sites occur (Fig. 3A). A similar ChIP-signal pattern was seen for synVicII-oneGATC, indicating that indeed SeqA is not able to bind to this synthetic DNA sequence despite the occurrence of GATC sites in the synVicII backbone (Fig. 3B). The synthetic secondary chromosome synVicII-fourGATC showed a clearly different ChIP signal (Fig. 3C). Strong binding of SeqA occurred at each GATC cluster as predicted for the dimeric SeqA protein. As reference for SeqA binding, the ChIP signal was compared to known target sites on the primary chromosome (Fig. S4). The results were very similar in biological replicates (compare Fig. S4 and S5). In conclusion, SeqA binding to the three synthetic secondary chromosomes differed as predicted proving that it can be directed by sequence design.

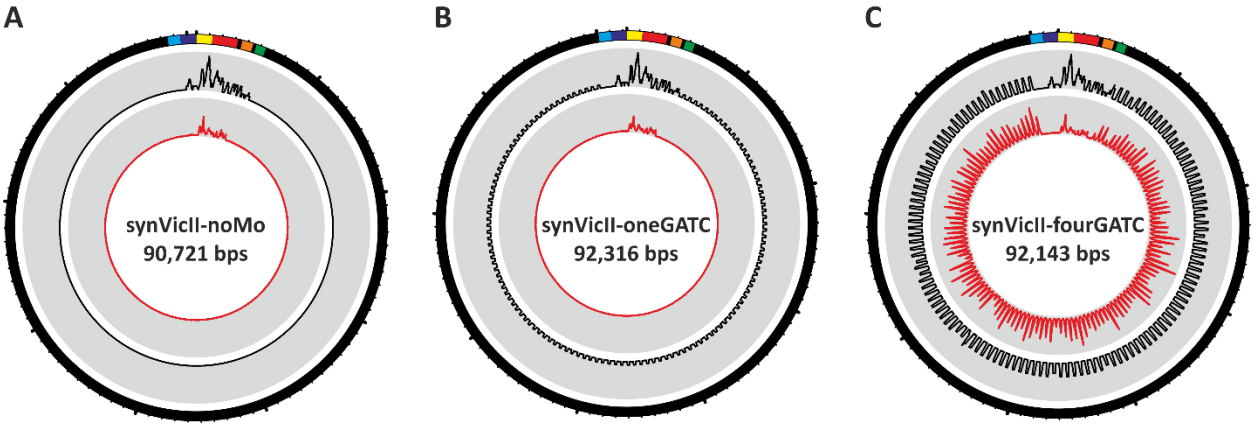
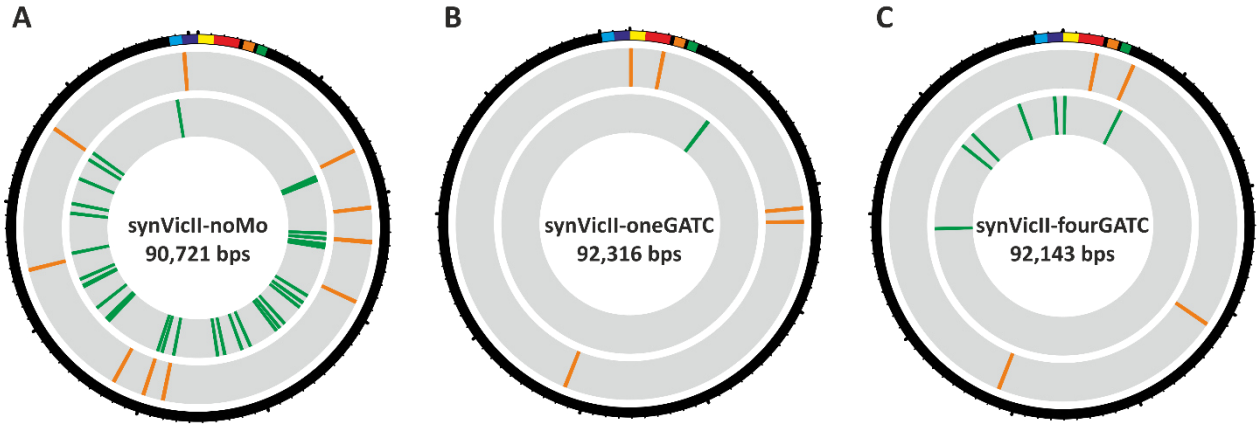


Figure 3: *In vivo* SeqA binding throughout the synthetic secondary chromosome set. **A to C** shows the three synthetic chromosomes. GATC-moving window distribution (black *cf*. Fig. 2) and SeqA ChIP-Seq signals (red) of the chromosomes are visualized by inner circles respectively. ChIP-Seq signals are normalized to the mean synVicII minimal replicon ChIP-Seq signal for comparative illustration purposes. It is clearly indicated that SeqA binds only the GATC clusters in the synVicII-four GATC chromosome beside the synVicII backbone in all three experiments. See Material and Methods for ChIP-Seq details. Size of chromosomes is indicated by black lines 10 kb lines accentuated bolt and plots were generated by circos (Krzywinski *et al.* 2009).

Differential mutation rates on the synthetic secondary chromosomes

To investigate a potential role of SeqA in the DNA mismatch repair the mutation rate on the three synthetic secondary chromosomes had to be measured. To this end, we cultivated three *E. coli* strains carrying one of the replicons for many generations (25 days) in medium containing the mutagenic substance ethyl methanesulfonate (EMS) to induce mismatches (see Material and Method section for details). Total DNA was isolated before and after this extended cultivation and single-nucleotide polymorphisms (SNPs) were determined by next-generation sequencing. The mutation rates on the primary chromosome were very similar in all three strains (Fig. 4D). However, the synthetic secondary chromosome synVicII-noMo had clearly accumulated more mutations (16) compared to synVicII-oneGATC (5) and synVicII-fourGATC (4)(Fig. 4). The high mutation rate on synVicII-noMo was expected because the MutH-dependent mismatch repair relies on hemi-methylated GATCs. However, the maximal *in vivo* distance between a GATC site and a mismatch to be repaired efficiently is unknown. Interestingly, an extended region with almost no SNPs was detected surrounding the chromosome backbone of synVicII-noMo where some GATCs occur (Fig. 4A). This region might provide an estimate of the maximal distance of GATC to mismatch allowing a respective repair reaction.

As an alternative approach for verification of this data we used the regulated expression of a mutant version of the epsilon subunit of DNA polymerase III (*dnaQ926*) with deficiency in proofreading (Badran and Liu 2015). To this end we transformed plasmid MP2 into cells carrying one of the three synthetic secondary chromosomes and induced *dnaQ926* expression with 25 mM of L-arabinose. Since mutation of the plasmid itself could lead to reduced mutagenesis we alternated *dnaQ926* plasmids MP2 and pMA715 every 5 days. After 20 days of cultivation, genomic mutations were detected by next-generation sequencing as above. As in the previous experiment, synVicII-noMo accumulated most mutations compared to the other two synthetic replicons (Fig. 4A-C and E). The synVicII-four GATCs had a slightly raised mutation rate which results from of a point mutation in the primary chromosome causing an amino acid exchange in domain five of MutS. Taking this into account and the similarity of synVicII-oneGATC and synVicII-fourGATC in both experimental approaches with the significantly lower mutation frequency than on synVicII-noMo indicates that SeqA plays a minor role in the DNA mismatch repair process.



D

Strain	*E. coli* chromosome		Synthetic chromosome		Comparison
	SNPs	Frequency	SNPs	Frequency	Ratio
DS341 (synVicII-noMo)	170	$3.7*E^{-5}$	16	$17.6*E^{-5}$	4.82
DS333 (synVicII-oneGATC)	166	$3.6*E^{-5}$	5	$5.4*E^{-5}$	1.51
DS337 (synVicII-fourGATC)	168	$3.6*E^{-5}$	4	$4.3*E^{-5}$	1.20

E

Strain	*E. coli* chromosome		Synthetic chromosome		Comparison
	SNPs	Frequency	SNPs	Frequency	Ratio
DS341 (synVicII-noMo)	98	$2.1*E^{-5}$	33	$36.4*E^{-5}$	17.24
DS333 (synVicII-oneGATC)	217	$4.7*E^{-5}$	1	$1.1*E^{-5}$	0.23
DS337 (synVicII-fourGATC)	287	$6.2*E^{-5}$	7	$7.8*E^{-5}$	1.26

Figure 4: Comparative random mutagenesis results for synthetic secondary chromosome strains. **A to C** Synthetic chromosomes are visualized by black circle, SNPs in EMS experiment (orange) and MP2 (green) experiment are visualized by respective circle. **D and E** Overview of SNPs and the resulting mutation frequency of primary chromosome compared to respective synthetic secondary chromosome for **D** EMS (orange) and **E** MP2 (green) mutagenesis. For comparison of the mutation frequency ratio of primary *E. coli* chromosome divided by the synthetic secondary chromosome normalized to numbers of bps is shown. Details for mutagenesis and sequencing experiments are given in the Material and Methods section. Size of chromosomes is indicated by black lines 10 kb lines accentuated bolt. Visualization of chromosomes was achieved by circos (Krzywinski *et al.* 2009).

Discussion

Synthetic secondary chromosomes – new tools to study chromosome biology

The study presented here is a proof of concept for the use of synthetic secondary chromosomes to study chromosome maintenance. In bacteria, the mechanisms underlying chromosome maintenance often consist of a DNA motif and a respective protein (or more than one) binding specifically to this DNA motif (Touzain *et al.* 2011). Besides the GATC motif introduced above, examples include (i) the *matS* site bound by MatP to organize the chromosomal terminus macrodomain, (ii) KOP sites which direct the FtsK DNA translocase to find the *dif* site for chromosome dimer resolution, (iii) Chi sites, involved in homologous recombination or (iv) SlmA binding sites which are involved in nucleoid occlusion (Tonthat *et al.* 2011; Cho *et al.* 2011; Mercier *et al.* 2008; Bigot *et al.* 2005; Touzain *et al.* 2011; Smith 1988; Messerschmidt and Waldminghaus 2014). All these systems have been studied with various approaches over the last years. However, one important component of their functionality is largely unexplored - the distribution of the respective DNA motifs on the chromosome. This distribution is not random but shows clear biases probably related to the function of the individual systems (Touzain *et al.* 2011; Schindler and Waldminghaus 2015). The main reason of lacking studies on the functional relevance of these motif distributions is a lack of suitable experimental approaches. One option would be deletion or insertion of motifs on the primary chromosome and respective functional analysis of the mutated strains. This approach is problematic because the high number of DNA motifs would require enormous resources to manipulate them. In addition, many such DNA motifs lay within coding sequences and their mutation could have unwanted side effects on protein expression (Sobetzko *et al.* 2016; Bryant *et al.* 2014; Roymondal *et al.* 2009). Using secondary synthetic chromosomes as introduced here omits the problem of interference by using nonsense sequence beside the DNA motif of interest. Of course, it can't be excluded that such nonsense sequences have some sort of unknown effect on cell physiology since "biological neutral DNA sequences" might be an illusion. This fact emphasizes the importance of good biological control experiments as characteristic for biological experiments (Schindler and Waldminghaus 2013; Waldminghaus and Skarstad 2010; Kidder *et al.* 2011). A conceptual advantage of using secondary chromosomes instead of changing the primary chromosome to study maintenance systems is that the later are required for cell viability prohibiting modifications leading to non-functionality. In contrast, an experimental arrangement of DNA motifs rendering a secondary chromosome non-functional could be detected and provide insight into functionality. This is comparable to deleterious mutations of essential

proteins that provide important information but need to be carried out within systems that make them non-essential (Kogoma and von Meyenburg 1983).

One important prerequisite of using synthetic secondary chromosomes to study chromosome maintenance is that their assembly is efficient. If construction takes too much time and money, synthetic secondary chromosomes will have no chance to be widely used as research tool. In the study presented here the three chromosomes were based on only six synthetic DNA oligonucleotids resulting in three MoClo library sets (total synthesis cost of about 450 Euro). This was possible because sequence design of libraries was based on the computer tool MARSeG (Schindler *et al.* 2016). The cost for future synthetic secondary chromosome construction will depend on the diversity of fragments needed. Importantly, the parts used for construction within this study can be reused to speed up future assemblies. We imagine a library of chromosome building blocks which expands over the next years and are working on solutions for efficient sharing and documentation.

SeqA and the mismatch repair

SeqA and MutH share the common DNA target site GATC in its hemi-methylated state. At least two different mechanisms have been suggested on how SeqA might affect MutH mediated mismatch repair. The first idea is based on the finding that SeqA prolongs the hemi-methylated state of DNA by blocking methylation by Dam methyltransferase (Bach *et al.* 2003). MutH distinguishes the old and new DNA strand by their differential methylation right after replication and cuts the unmethylated new strand to initiate the repair process. Consequently, an overproduction of Dam leads to an increased mutation rate probably caused by a reduced period of hemi-methylation of GATC (Herman and Modrich 1981; Yang *et al.* 2004). The MutLS complex might just not find any MutH-bound GATC because MutH can't bind and can't act on fully-methylated GATCs. Mutation rates are also increased in cells lacking Dam (Marinus 2010; Boye *et al.* 1988; Marinus *et al.* 1984). Under such conditions MutH might bind and nick the DNA but will be unable to distinguish the strands and might falsely "repair" the template strand. If the methylation state is a critical parameter for mismatch repair efficiency it is reasonable to assume that changing SeqA levels in the cell will affect mismatch repair via its influence on the genomic methylation state. *seqA* deletion strains might have an increased mutation rate as it mimics conditions of *dam* overexpression. However, if SeqA is important to hold GATCs in a hemi-methylated state to facilitate MutH dependent mismatch repair, the mutation frequency on synVicII-oneGATC which is not binding

SeqA should have been significantly higher compared to synVicII-fourGATC which harbors SeqA bound GATC sites. This was not the case (Fig. 4).

A second idea on SeqAs role in DNA mismatch repair is that it blocks the intrinsic nuclease activity of MutH so that it cleaves a daughter strand only when a mismatch is detected by MutS and MutL (Lee *et al.* 2005). This hypothesis is based on the finding that SeqA overproduction inhibits mismatch repair (Yang *et al.* 2004). It could be that physiological levels of SeqA inhibit MutH from binding hemi-methylated GATCs and upon mismatch recognition, MutS and MutL enable MutH to overcome the inhibitory effect of SeqA (Lee *et al.* 2005). Consequently, the balance of inhibition and activation of MutH would ensure mismatch-repair specific nicking of the hemi-methylated GATC (Lee *et al.* 2005). If this hypothesis is true one would expect such a balance of SeqA dependent inhibition and activation of MutH to be absent on the synthetic secondary chromosome synVicII-oneGATC because it is not binding SeqA (Fig. 3). However, we could not detect a significant difference in mutation rates compared to synVicII-fourGATC were SeqA could potentially play the proposed role of MutH inhibition.

Taken together, our results suggest that MutH mediated mismatch repair occurs largely independent on SeqA. This is surprising if SeqA and MutH compete for the same binding sites. One explanation would be that such a competition is prohibited by special and/or temporal constrains. It could for example be that MutH is somehow associated with the replication fork to increase the chance of MutH binding right after replication before SeqA. Interestingly, the mismatch repair proteins MutS and MutL have both been shown to interact with the replication fork associated β clamp in *E. coli* and *Bacillus subtilis* (Simmons *et al.* 2008; Lopez de Saro *et al.* 2006). Maybe, loading of the third mismatch repair protein MutH on hemi-methylated GATCs is replication fork associated in a similar way. This would imply that in a region right behind the replication fork all GATCs are bound by MutH while SeqA binds later at regions further away from the fork. This model is supported by a recent finding of SeqA structures being spatially separated from the replication fork (Helgesen *et al.* 2015). This finding concludes the hypothesis that a stretch of DNA between SeqA and the replisome presents a preferred site for mismatch repair processes (Helgesen *et al.* 2015).

Material and Methods

Bacterial strains, plasmids, oligonucleotides and culture conditions

All strains, plasmids and oligonucleotides are listed in table S1 to S3. Precultures of *E. coli* were grown in 5 ml LB medium. Antibiotics and other ingredients were used with the following concentrations if not indicated otherwise: ampicillin (100 µg/ml), spectinomycin (100 µg/ml), chloramphenicol (30 µg/ml), kanamycine (35 µg/ml) and diaminopimelic acid (57 µg/ml).

Construction of secondary chromosomes

MARSeG was used to generate random sequences with 0, 1 or 4 GATCs by excluding the motifs shown in table S4. Degenerated oligonucleotides were ordered as IDT Ultramers. MoClo library construction and MoClo reactions were carried out according, to Schindler *et al.* 2016. Five Level one libraries were combined with respective Level M vector (pMA60-pMA66) and endlinker (pICH50872-pICH50932) to produce Level M libraries. For the constructs with single GATCs and GATC clusters, libraries without and with respective GATC(s) were assembled alternating whereas each library was generated starting with a GATC or GATC-free library respectively. Five Level M libraries were combined with respective Level P vector (pMA67-pMA73) and endlinker (pICH79255-pICH79311) to produce Level P vectors. Positive Level P vectors were identified from single colonies by restriction analysis with BsaI and DpnI respectively. Five Level P vectors were combined with the respective MoClo Level M vector (pMA60, pMA67 or pMA333) and endlinker (pICH50927 or pICH50900) to produce the two semi-final Level M constructs for each chromosome with a size of approximately 42 kb. Plasmid DNA > 40 kb was isolated by a previously described protocol (Rondon *et al.* 1999) and verified by restriction analysis with BpiI and DpnI in 0.4 % agarose gels. Final synthetic chromosomes were assembled by combining the corresponding two 42 kb DNA-fragments, synVicII-2.0 (pMA657) and the respective endlinker (pMA678) by a MoClo reaction in 0.5 ml PCR reaction tubes. RbCl competent Top10 cells were transferred into the reaction tube and heat shocked. Positive clones were verified as described for semifinal constructs. For conjugation purposes verified synVicII2.0 variants were transferred into *E. coli* WM3064 strain by chemical transformation and conjugation was carried out as described previously (Messerschmidt *et al.* under revision). Flp/FRT recombination was performed in final strains with a pCP20 derivate (pMA900) with a disrupted *bla* gene to remove construction based elements of synVicII-2.0 backbone.

Random mutagenesis experiments

Cells were cultivated in 1 ml LB with corresponding selection marker on an Eppendorf Thermomixer at 37 °C and 1000 rpm in 2 ml reaction tubes. The reaction tubes were prepared

with two holes in the cap to prevent anaerobic growth. Cultures were inoculated 1:1000 every morning and evening into fresh media. Cultures were stored as glycerol stocks if experiment was interrupted and restarted by 1:100 inoculation. Chemical mutagenesis was carried out using EMS (ethyl methanesulfonate) in a final 1:1000 dilution. For mutagenesis by overexpressing a *dnaQ*-variant under control of the P_{BAD} promoter, MP2 or pMA715 plasmid was used. 25 mM L-arabinose was applied for induction. For a confident constant mutagenesis MP2 and pMA715 were alternated every 5 days. Change of plasmid was performed by chemical transformation into verified single colonies and switch of the corresponding antibiotic. pMA715 was generated by replacing chloramphenicol resistance of MP2 by a spectinomycin resistance via Gibson Assembly. MP2 was amplified with primers 1518/1519 and spectinomycin resistance gene was amplified from pMA60 with primers 1520/1521 to finally obtain pMA715. MP2 and pMA715 cultures were cultivated in presence of 0.4 M glucose if repression of P_{Bad}-dnaQ926 was necessary.

DNA Sequencing

Whole Genome Sequencing of the synVicII chromosome set was performed on a PacBio RSII (Pacific Bioscience). DNA of exponential growing cells (OD = 0.15) was extracted using a previously described protocol (Rondon *et al.* 1999). Subsequent RNase A treatment and standard Phenol/Choloroform extraction was performed prior DNA library preparation. DNA was sheared to approximately 500 to 1000 bp DNA-fragments in a Bioruptor (Diagenode) using 5 cycles of 30 seconds high-sonication and 30 seconds cooling. Library preparation and DNA Sequencing of synVicII chromosomes was performed according, to the manufacture guidelines using SMRTbell™ Template Prep Kit 1.0 and DNA Sequencing Reagent 4.0 v2 Kit (Pacific Bioscience). *De novo* Assembly of the synVicII chromosome set was performed using Genious R9.

For SNP analysis DNA of stationary phase cells was extracted by standard Phenol/Chloroform extraction. SeqA ChIP DNA of two independent cultures was prepared as described previously (Waldminghaus and Skarstad 2010; Waldminghaus *et al.* 2012). Sequencing libraries for Illumina Sequencing of EMS mutagenesis strains (t = 0 d) and MP2 mutagenesis strains (t = 20 d) were prepared by Nextera DNA Library Prep Kit (Illumina) according, to manufacture guidelines. DNA was sequenced on a MiSeq using the MiSeq V2 Reagent Kit with 2 x 250 bp paired end reads (MS-102-2003). Sequencing libraries for Illumina Sequencing of EMS mutagenesis strains (t = 25 d) and SeqA ChiP-Seq samples were prepared as described previously by Ethan Ford (Ford *et al.* 2014). DNA was sequenced on a MiSeq using the MiSeq V3 Reagent Kit with 2 x 75 bp paired end reads (MS-102-3001). Mapping of sequencing reads, SNP-detection and ChIP-signal intensity determination was performed using Genious R9.

Acknowledgements

We thank all members of the Waldminghaus lab for help and fruitful discussions. We are grateful to Sylvestre Marillonnet and William Metcalf for providing strains and/or plasmids. We thank Claudia Quedenau and Wei Chen (MDC Berlin) for PacBio RS Sequencing and Bernadette Boomers, Javier Serrania and Anke Becker (SYNMIKRO Marburg) for help concerning Illumina Sequencing and providing the MiSeq-plattform. This work was supported within the LOEWE program of the State of Hesse and a grant of the Deutsche Forschungsgemeinschaft (Grant No. WA 2713/4-1).

References

Acharya S, Foster PL, Brooks P, Fishel R (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. Mol Cell 12 (1):233-246

Bach T, Krekling MA, Skarstad K (2003) Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. EMBO J 22 (2):315-323. doi:10.1093/emboj/cdg020

Badran AH, Liu DR (2015) Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. Nat Commun 6:8425. doi:10.1038/ncomms9425

Badrinarayanan A, Le TB, Laub MT (2015) Bacterial chromosome organization and segregation. Annu Rev Cell Dev Biol 31:171-199. doi:10.1146/annurev-cellbio-100814-125211

Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, Dennis C, Grigoriev M, Allemand JF, Barre FX, Cornet F (2005) KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. EMBO J 24 (21):3770-3780. doi:10.1038/sj.emboj.7600835

Boye E, Lobner-Olesen A, Skarstad K (1988) Timing of chromosomal replication in *Escherichia coli*. Biochim Biophys Acta 951 (2-3):359-364

Brendler T, Austin S (1999) Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. EMBO J 18 (8):2304-2310. doi:10.1093/emboj/18.8.2304

Bryant JA, Sellars LE, Busby SJ, Lee DJ (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. Nucleic Acids Res 42 (18):11383-11392. doi:10.1093/nar/gku828

Cho H, McManus HR, Dove SL, Bernhardt TG (2011) Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. Proc Natl Acad Sci U S A 108 (9):3773-3778. doi:10.1073/pnas.1018674108

Cooper DL, Lahue RS, Modrich P (1993) Methyl-directed mismatch repair is bidirectional. J Biol Chem 268 (16):11823-11829

Ford E, Nikopoulou C, Kokkalis A, Thanos D (2014) A method for generating highly multiplexed ChIP-seq libraries. BMC Res Notes 7:312. doi:10.1186/1756-0500-7-312

Grilley M, Griffith J, Modrich P (1993) Bidirectional excision in methyl-directed mismatch repair. J Biol Chem 268 (16):11830-11837

Helgesen E, Fossum-Raunehaug S, Saetre F, Schink KO, Skarstad K (2015) Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. Nucleic Acids Res 43 (5):2730-2743. doi:10.1093/nar/gkv146

Herman GE, Modrich P (1981) *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. J Bacteriol 145 (1):644-646

Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D (2013) Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. PLoS Genet 9 (8):e1003673. doi:10.1371/journal.pgen.1003673

Kidder BL, Hu G, Zhao K (2011) ChIP-Seq: technical considerations for obtaining high-quality data. Nat Immunol 12 (10):918-922. doi:10.1038/ni.2117

Kogoma T, von Meyenburg K (1983) The origin of replication, *oriC*, and the DnaA protein are dispensable in stable DNA replication (*sdrA*) mutants of *Escherichia coli* K-12. EMBO J 2 (3):463-468

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19 (9):1639-1645. doi:10.1101/gr.092759.109

Lahue RS, Au KG, Modrich P (1989) DNA mismatch correction in a defined system. Science 245 (4914):160-164

Lee JY, Chang J, Joseph N, Ghirlando R, Rao DN, Yang W (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. Mol Cell 20 (1):155-166. doi:10.1016/j.molcel.2005.08.019

Lehman IR (1974) DNA ligase: structure, mechanism, and function. Science 186 (4166):790-797

Lobner-Olesen A, Skovgaard O, Marinus MG (2005) Dam methylation: coordinating cellular processes. Curr Opin Microbiol 8 (2):154-160. doi:10.1016/j.mib.2005.02.009

Lopez de Saro FJ, Marinus MG, Modrich P, O'Donnell M (2006) The beta sliding clamp binds to multiple sites within MutL and MutS. J Biol Chem 281 (20):14340-14349. doi:10.1074/jbc.M601264200

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in *E. coli*. Cell 77 (3):413-426

Marinus MG (2010) DNA methylation and mutator genes in *Escherichia coli* K-12. Mutat Res 705 (2):71-76. doi:10.1016/j.mrrev.2010.05.001

Marinus MG (2012) DNA Mismatch Repair. EcoSal Plus 5 (1). doi:10.1128/ecosalplus.7.2.5

Marinus MG, Lobner-Olesen A (2014) DNA Methylation. EcoSal Plus 6 (1). doi:10.1128/ecosalplus.ESP-0003-2013

Marinus MG, Poteete A, Arraj JA (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene 28 (1):123-125

Matson SW (1986) *Escherichia coli* helicase II (*urvD* gene product) translocates unidirectionally in a 3' to 5' direction. J Biol Chem 261 (22):10169-10175

Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, Boccard F, Espeli O (2008) The MatP/*matS* site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. Cell 135 (3):475-485. doi:10.1016/j.cell.2008.08.031

Messerschmidt S, Schindler D, Zumkeller CM, Kemter F, Zimmer N, Waldminghaus T (under revision) Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*. Front Bioeng Biotechnol

Messerschmidt SJ, Kemter FS, Schindler D, Waldminghaus T (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. Biotechnol J 10 (2):302-314. doi:10.1002/biot.201400031

Messerschmidt SJ, Waldminghaus T (2014) Dynamic organization: chromosome domains in *Escherichia coli*. J Mol Microbiol Biotechnol 24 (5-6):301-315. doi:10.1159/000369098

Nandakumar J, Nair PA, Shuman S (2007) Last stop on the road to repair: structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. Mol Cell 26 (2):257-271. doi:10.1016/j.molcel.2007.02.026

Nievera C, Torgue JJ, Grimwade JE, Leonard AC (2006) SeqA blocking of DnaA-*oriC* interactions ensures staged assembly of the *E. coli* pre-RC. Mol Cell 24 (4):581-592. doi:10.1016/j.molcel.2006.09.016

Rondon MR, Raffel SJ, Goodman RM, Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. Proc Natl Acad Sci U S A 96 (11):6451-6455

Roymondal U, Das S, Sahoo S (2009) Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome. DNA Res 16 (1):13-30. doi:10.1093/dnares/dsn029

Sanchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, Grainger DC (2010) Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12. MBio 1 (1). doi:10.1128/mBio.00012-10

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synth Biol. doi:10.1021/acssynbio.6b00089

Schindler D, Waldminghaus T (2013) "Non-canonical protein-DNA interactions identified by ChIP are not artifacts": response. BMC Genomics 14:638. doi:10.1186/1471-2164-14-638

Schindler D, Waldminghaus T (2015) Synthetic chromosomes. FEMS Microbiol Rev 39 (6):871-891. doi:10.1093/femsre/fuv030

Simmons LA, Davies BW, Grossman AD, Walker GC (2008) Beta clamp directs localization of mismatch repair in *Bacillus subtilis*. Mol Cell 29 (3):291-301. doi:10.1016/j.molcel.2007.10.036

Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N (1995) *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell 82 (6):927-936

Smith GR (1988) Homologous recombination in procaryotes. Microbiol Rev 52 (1):1-28

Sobetzko P, Jelonek L, Strickert M, Han W, Goesmann A, Waldminghaus T (2016) DistAMo: A Web-Based Tool to Characterize DNA-Motif Distribution on Bacterial Chromosomes. Front Microbiol 7:283. doi:10.3389/fmicb.2016.00283

Su SS, Modrich P (1986) *Escherichia coli mutS*-encoded protein binds to mismatched DNA base pairs. Proc Natl Acad Sci U S A 83 (14):5057-5061

Tonthat NK, Arold ST, Pickering BF, Van Dyke MW, Liang S, Lu Y, Beuria TK, Margolin W, Schumacher MA (2011) Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. EMBO J 30 (1):154-164. doi:10.1038/emboj.2010.288

Touzain F, Petit MA, Schbath S, El Karoui M (2011) DNA motifs that sculpt the bacterial chromosome. Nat Rev Microbiol 9 (1):15-26. doi:10.1038/nrmicro2477

Viswanathan M, Lovett ST (1998) Single-strand DNA-specific exonucleases in *Escherichia coli*. Roles in repair and mutation avoidance. Genetics 149 (1):7-16

Waldminghaus T, Skarstad K (2009) The *Escherichia coli* SeqA protein. Plasmid 61 (3):141-150. doi:10.1016/j.plasmid.2009.02.004

Waldminghaus T, Skarstad K (2010) ChIP on Chip: surprising results are often artifacts. BMC Genomics 11:414. doi:10.1186/1471-2164-11-414

Waldminghaus T, Weigel C, Skarstad K (2012) Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. Nucleic Acids Res 40 (12):5465-5476. doi:10.1093/nar/gks187

Welsh KM, Lu AL, Clark S, Modrich P (1987) Isolation and characterization of the *Escherichia coli mutH* gene product. J Biol Chem 262 (32):15624-15629

Yamaguchi M, Dao V, Modrich P (1998) MutS and MutL activate DNA helicase II in a mismatch-dependent manner. J Biol Chem 273 (15):9197-9201

Yang H, Wolff E, Kim M, Diep A, Miller JH (2004) Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. Mol Microbiol 53 (1):283-295. doi:10.1111/j.1365-2958.2004.04125.x

Supporting Information

Using experimental chromosome construction to study functional interactions between segregation and DNA mismatch repair in *Escherichia coli*

Daniel Schindler and Torsten Waldminghaus

Figure S1

5´-*TTTTAGGAAGGTCTCG*GGAGTNDNGNGGCCNNCNDTGNCRNCNCNTWTNNGYBGHNGCNGNGGCNNCNRCANTNNCANHBGCCDTHNCCHHWANTNDCNGNGCCNNCNCDNCNTWGKCNTANC
NVCNCVTNCNTHNGCHNCNNCNTACDNGCNNCNYGTWNNGNDGCNHNTCGTGGGTAVGANTCAAHGSTVNAATTAGT-3´
5´-AGCACCCATBCTNAGTTDCSABNTTAATCAADNCTCNDWTNCNCNGCNGTNGCNCNCNRCNGTNKCCNTCNMCNNC
GNNGTDNCNNCGNTGNNNTNAYGNNGNNWAACNNCNHCTACNNCTWCNWCNCNCNDCNWCGANNCNCCNNWTDCNGNYGGSNCDNTTNCGHNNCNCRCCNNCCNGCGA*CCTCTGGTTCCTAATT*-3´

Single GATC basic library:

5´-*TTTTAGGAAGGTCTCG*GGAGNNGCKNNGCNVCDTNGYNGNCNRCHNCTNTNNCNCDNTGNNGCCNNCATNNSGGCNNCGDCANTNGNNGGCDTNGNGNNGCHNNCNTTANTTYNGNCCNCGNN
GYNCNNCNTWDNCHCNNCTNTNGNCGVCNTNTCNNGTNGCNGNGNGNACGGCTCWTBTAYCGCAWANTYNTGCTGTT-3´
5´-TGCCGAGWAVATRGCGTWTNARNACGACAAKNNGYTNCNTNHACNGNTNTNACNNCNCNAHCNGNTRCNBTCANCC
NNHAMCNGGNNTNACNNCGNNNTCTAGGCNGNYGNTGKNNTTNCNNWTNCDGNHAACNNAACNYGNGNCDNTNATNCNNHACCYGNNNAHTNCNCNNCGNNGNBGCGA*CCTCTGGTTCCTAATT*-3´

Cluster GATC basic library:

5´-*TTTTAGGAAGGTCTCG*GGAGTNNNGYNGCNNCWNTNGGCNNCCNCGNTGNCGNNGCNNCATNNGCCGNNGCDCNGNNGYNGNCDANGCNCNTNTMNCNGTNGNYGNTGACNNCNCNTWDCCNC
GNCCNNCNACWNTNTANTNNCNNAGCCMHNTHNTWNTNNCCGNGNNWACCTAATHTNATDCGGTWADTKMCCTCTGT-3´
5´-TGGATTADANTAHGCCAWTHAMKGGAGACAANNCNNCGNNNTNATNCNGNYGNNGNNAHTNAHTTNWCANCNNCNA
TNACDNCGNCTAGTCNNCNCRCTAGGMNDCRVCCTAGATVCNGCNCTAGCDNCGNCNCGNNDCGHNNCGNGCCNNCCWGTNNCNNCACVNCNGWTNCNGNYGNNGCGA*CCTCTGGTTCCTAATT*-3´

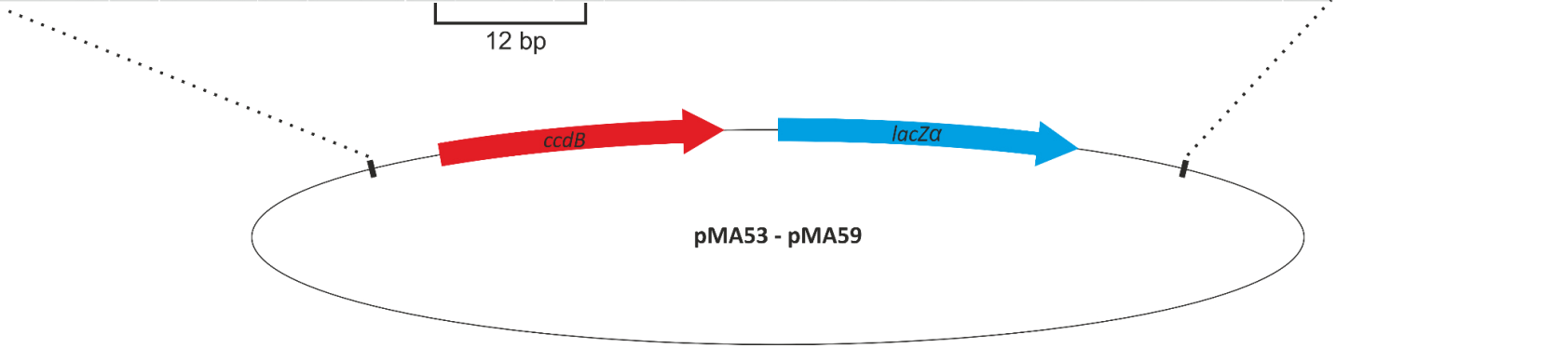


Fig S1: Level 1 MoClo libraries for constructing the set of synthetic secondary chromosomes with varied GATC distribution. Sequences and design of Level 1 MoClo libraries is indicated. Construction based sequences are italic including a BsaI recognition site (red). BsaI generated overhangs are bold and underlined respectively. Annealing region of oligonucleotides (green) and GATCs (red) of respective library are indicated. The final library contend in Level 1 vectors is shaded in grey. Standard IUPAC ambiguity code for nucleotides is used. The spacing within the cluster GATC library is 12 bps between the methylated bases. For details see Material and Method section.

Figure S2

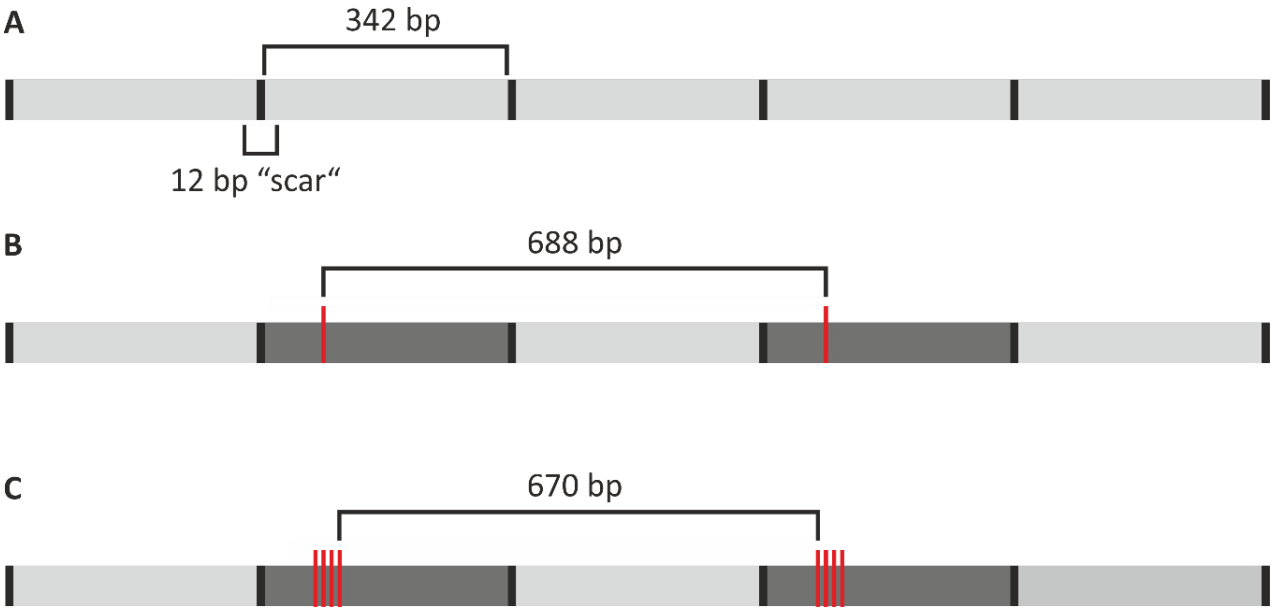


Fig S2: Level M MoClo library composition. Resulting Level M DNA-fragments generated by the combination of five Level 1 libraries, respective Level M vector and corresponding endlinker is shown. **A** Level M libraries consist of five Level 1 library fragments with an average size of 340 bps separated by a construction based 12 bp scar-sequences. **B** Single GATCs (red) have an average distance of 688 bps and **C** GATC cluster have an average spacing of 670 bps, within the cluster the spacing from one Adenine to the next is 12 bps (*cf*. Fig S1).

Figure S3

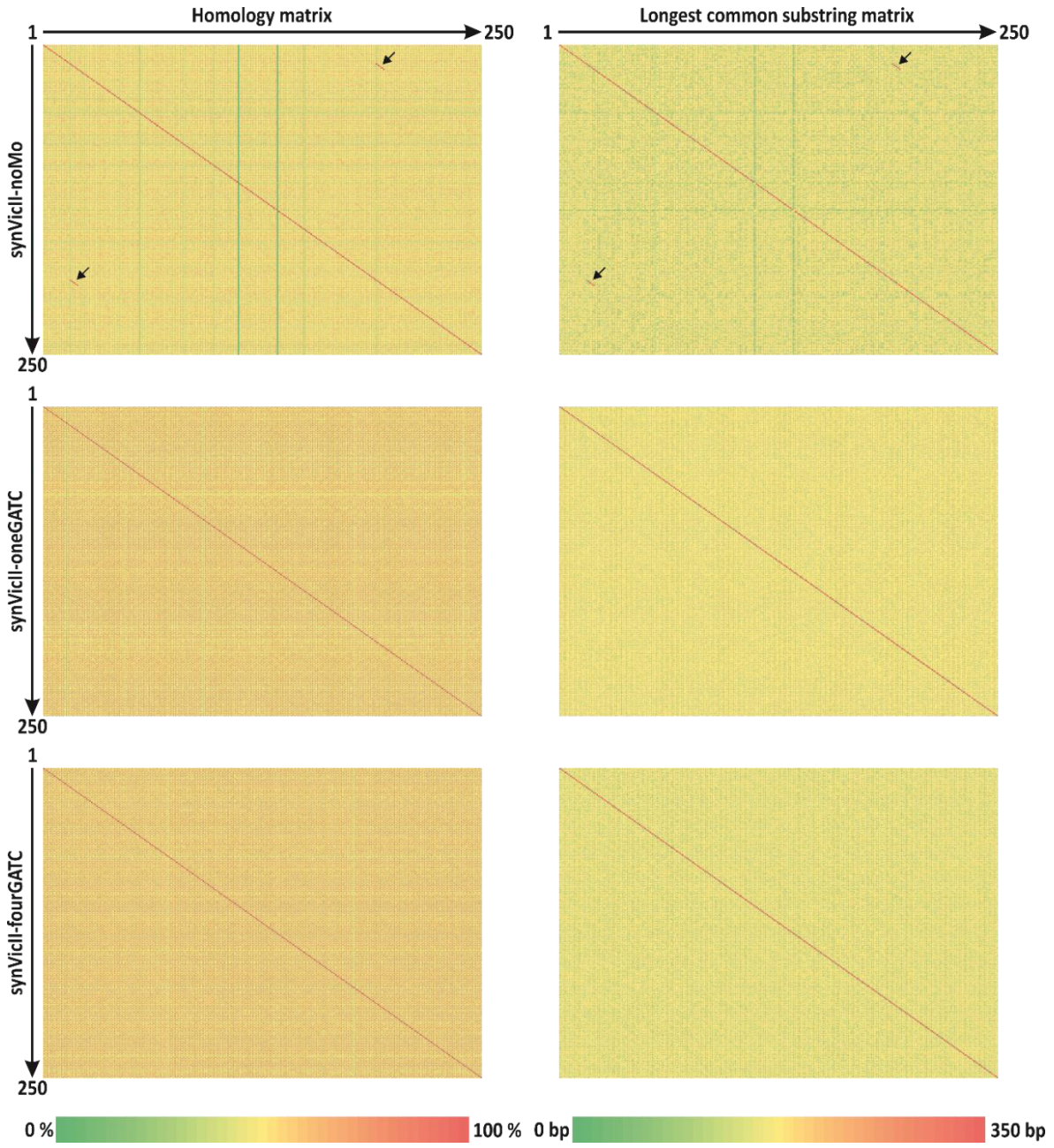


Fig S3: Homology and longest common substring matrices of synthetic chromosomes. Shown are the matrices for synVicII-noMo, synVicII-oneGATC and synVicII-four GATC from top to down. The homology is shown on the left and the longest common substring on the right panel. Chromosome sequences were dissected into the Level 1 MoClo sequences and all sequences of one chromosome were compared to each other by *Needleman-Wunsch* algorithm (homology) or the longest common substring was determined. The heat score for each matrices set is indicated by color (green to red). There is only one perfect sequence match for the synVicII-noMo which occurred by chance during library-based assembly in Level M (black arrows). The other sequences show a moderate homology and a low longest common substring.

Figure S4

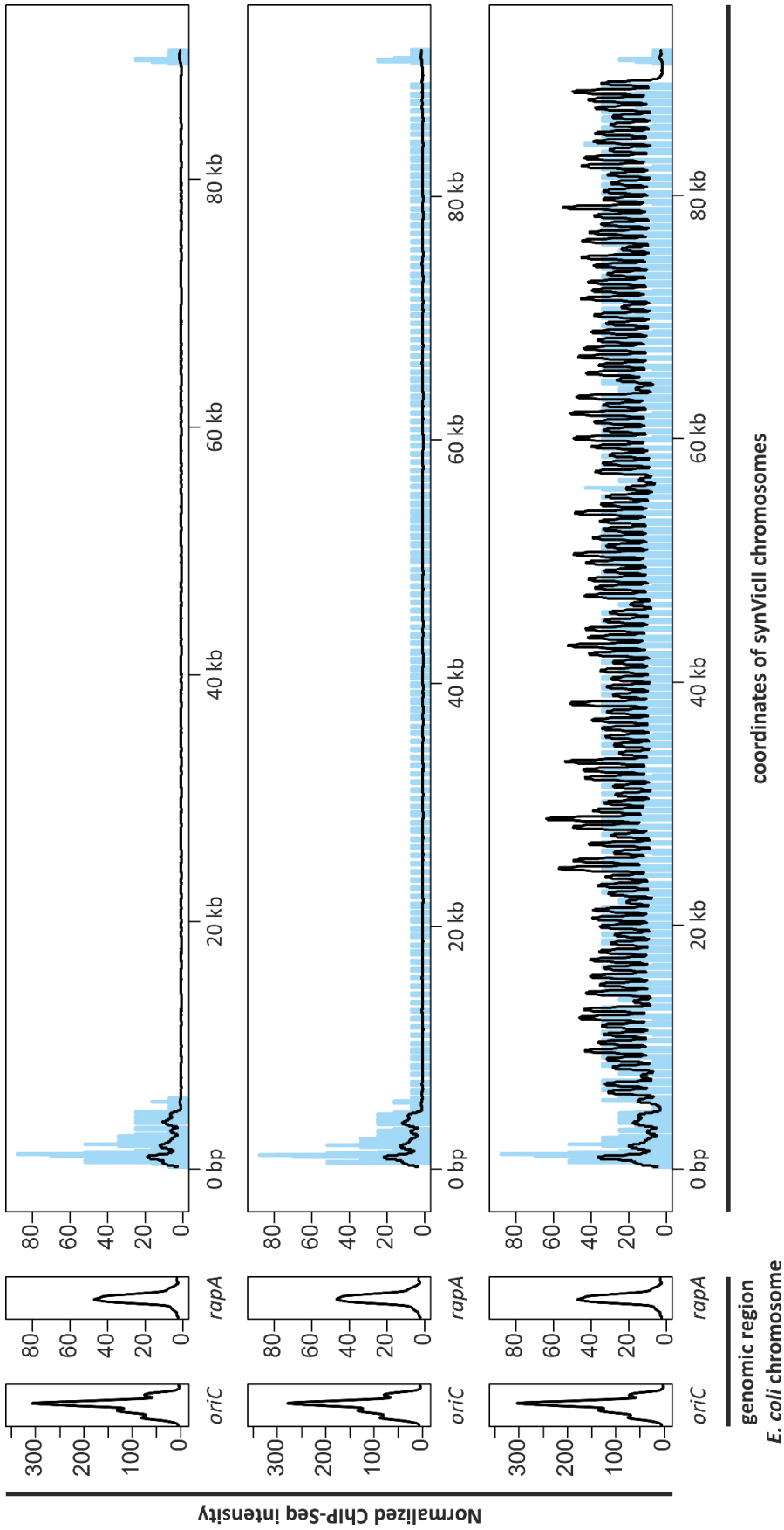


Fig S4: SeqA ChIP-Seq signals for synthetic secondary chromosomes in comparison to *oriC*-region and *rapA* – replicate one. SeqA ChIP-Seq signals are shown for *oriC*, *rapA* and synthetic secondary chromosome for each of the three strains in a 400 bp moving window with 25 bp steps respectively. The GATC distribution (blue) for synthetic secondary chromosomes indicated by a 250 bp moving average window with 50 bp steps. ChIP-signal intensities are normalized to *oriC*. For details see Material and Method section.

Figure S5

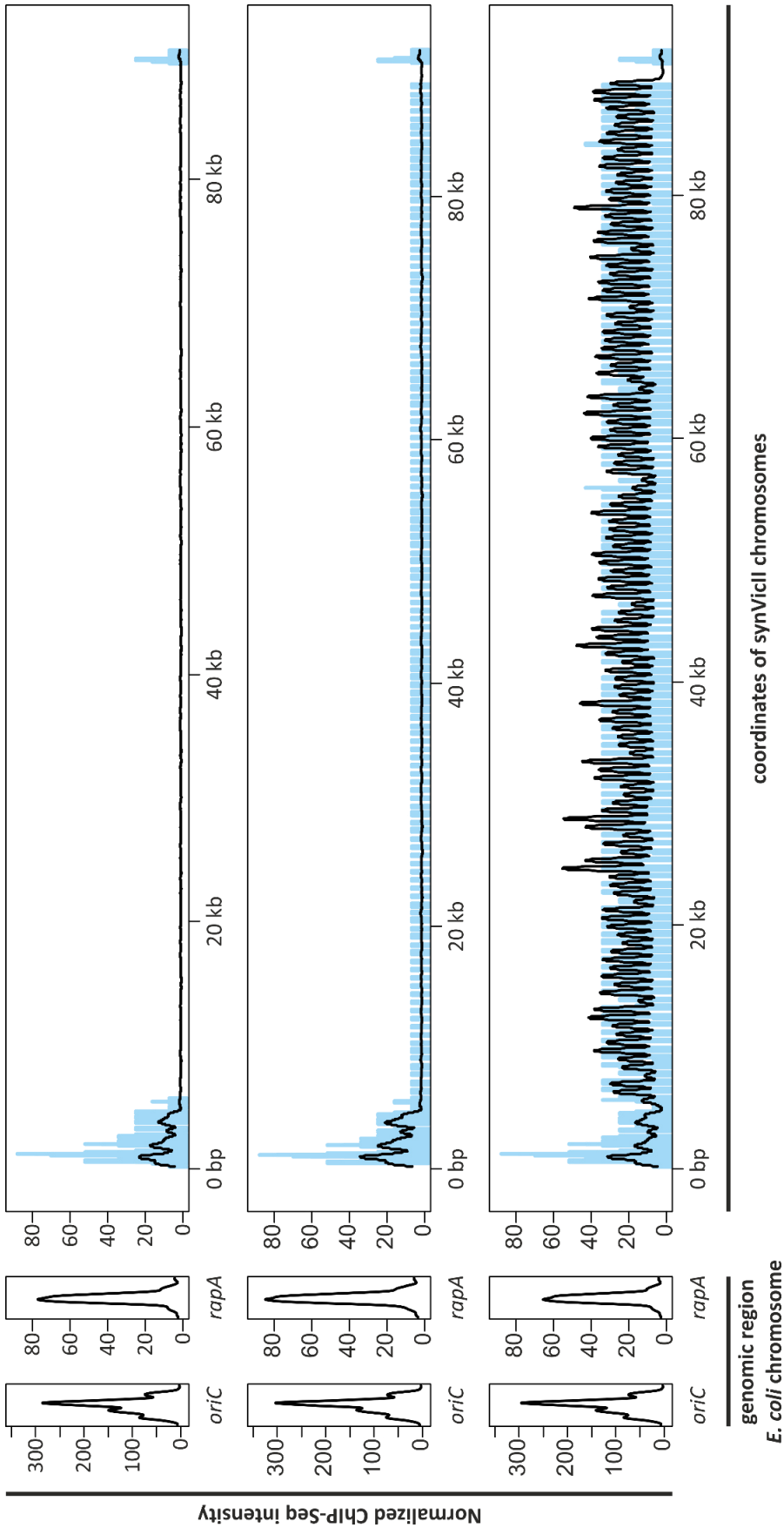


Fig S5: SeqA ChIP-Seq signals for synthetic chromosomes in comparison to *oriC*-region and *rapA* – replicate two. SeqA ChIP-Seq signals are shown for *oriC*, *rapA* and synthetic secondary chromosome for each of the three strains in a 400 bp moving window with 25 bp steps respectively. The GATC distribution (blue) for synthetic secondary chromosomes indicated by a 250 bp moving average window with 50 bp steps. ChIP-signal intensities are normalized to *oriC*. For details see Material and Method section.

Table S1: Strains used in this study.

Strain	Characteristics	Resistance	Reference
E. coli DH5αλpir	*supE*44, Δ*lacU*169 (Φ*lacZ*ΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λ *pir* phage lysogen	nalidixic acid	(Miller and Mekalanos 1988)
E. coli MG1655	*E. coli* wild type	-	(Blattner *et al.* 1997)
E. coli TOP10	F- *mcr*A Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74 recA1 araD139* Δ(*ara leu*) 7697 *galU galK rpsL* (StrR) *end*A1 *nupG*	streptomycin	Invitrogen
E. coli WM3064	Donor strain for conjugation: *thrB1004 pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)*567* Δ*dapA1341*::[*erm pir*(wt)]	-	William Metcalf
DS292	*E. coli* MG1655 pMA682	ampicillin	Messerschmidt *et al.* under revision
DS301	*E. coli* MG1655 pMA683	ampicillin	This work
DS304	*E. coli* MG1655 pMA684	ampicillin	This work
DS320	*E. coli* MG1655 pMA685	ampicillin	This work
DS330	*E. coli* MG1655 pMA689	ampicillin	This work
DS333	*E. coli* MG1655 pMA690	ampicillin	This work
DS337	*E. coli* MG1655 pMA691	ampicillin	This work
DS341	*E. coli* MG1655 pMA692	ampicillin	This work
DS369	*E. coli* MG1655 pMA690 MP2	ampicillin, chloramphenicol	This work
DS372	*E. coli* MG1655 pMA691 MP2	ampicillin, chloramphenicol	This work
DS375	*E. coli* MG1655 pMA692 MP2	ampicillin, chloramphenicol	This work

Table S2: Plasmids used in this study.

Plasmid	Characteristics	Resistance	Reference
MP2	P_{BAD}-*dnaQ926*	chloramphenicol	(Badran and Liu 2015)
pICH50872	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50881	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50892	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50900	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50914	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50927	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH50932	Level M endlinker	ampicillin	(Weber *et al.* 2011)
pICH79255	Level P endlinker	ampicillin	(Weber *et al.* 2011)

Plasmid	Characteristics	Resistance	Reference
pICH79264	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pICH79277	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pICH79289	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pICH79290	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pICH79300	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pICH79311	Level P endlinker	ampicillin	(Weber *et al.* 2011)
pMA53	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA54	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA55	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA56	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA57	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA58	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA59	MoClo Level 1 vector	ampicillin	(Schindler *et al.* 2016)
pMA60	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA61	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA62	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA63	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA64	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA65	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA66	MoClo Level M vector	spectinomycin	(Schindler *et al.* 2016)
pMA67	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA68	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA69	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA70	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA71	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA72	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA73	MoClo Level P vector	kanamycin	(Schindler *et al.* 2016)
pMA333	MoClo Level M *oriR6K* vector	spectinomycin	(Schindler *et al.* 2016)
pMA492	5 respective no GATCs Level M libraries in pMA67, endlinker pICH79289	kanamycin	This work
pMA493	5 respective no GATCs Level M libraries in pMA68, endlinker pICH79290	kanamycin	This work
pMA494	5 respective no GATCs Level M libraries in pMA69, endlinker pICH79300	kanamycin	This work

Plasmid	Characteristics	Resistance	Reference
pMA495	5 respective no GATCs Level M libraries in pMA71, endlinker pICH79255	kanamycin	This work
pMA496	5 respective no GATCs Level M libraries in pMA72, endlinker pICH79264	kanamycin	This work
pMA497	5 respective no GATCs Level M libraries in pMA67, endlinker pICH79289	kanamycin	This work
pMA498	5 respective no GATCs Level M libraries in pMA68, endlinker: pICH79290	kanamycin	This work
pMA500	5 respective no GATCs Level M libraries in pMA71, endlinker: pICH79255	kanamycin	This work
pMA507	40 kb Fragment Level M (no GATCs): pMA492-496 in pMA60, endlinker: pICH50927	spectinomycin	This work
pMA510	5 respective single GATCs Level M libraries in pMA67, endlinker: pICH79289	kanamycin	This work
pMA511	5 respective single GATCs Level M libraries in pMA68, endlinker: pICH79290	kanamycin	This work
pMA512	5 respective single GATCs Level M libraries in pMA69, endlinker: pICH79300	kanamycin	This work
pMA513	5 respective single GATCs Level M libraries in pMA71, endlinker: pICH79255	kanamycin	This work
pMA514	5 respective single GATCs Level M libraries in pMA72, endlinker: pICH79264	kanamycin	This work
pMA515	5 respective single GATCs Level M libraries in pMA67, endlinker: pICH79289	kanamycin	This work
pMA516	5 respective single GATCs Level M libraries in pMA68, endlinker: pICH79290	kanamycin	This work
pMA517	5 respective single GATCs Level M libraries in pMA70, endlinker: pICH79311	kanamycin	This work
pMA518	5 respective single GATCs Level M libraries in pMA71, endlinker: pICH79255	kanamycin	This work
pMA519	5 respective single GATCs Level M libraries in pMA73, endlinker: pICH79277	kanamycin	This work
pMA525	40 kb Fragment Level M (single GATCs): pMA510-514 in pMA60, endlinker: pICH50927	spectinomycin	This work
pMA526	40 kb Fragment Level M (single GATCs): pMA515-519 in pMA66, endlinker: pICH50900	spectinomycin	This work
pMA528	5 respective GATC cluster Level M libraries in pMA67, endlinker: pICH79289	kanamycin	This work
pMA529	5 respective GATC cluster Level M libraries in pMA68, endlinker: pICH79290	kanamycin	This work
pMA530	5 respective GATC cluster Level M libraries in pMA69, endlinker: pICH79300	kanamycin	This work
pMA531	5 respective GATC cluster Level M libraries in pMA71, endlinker: pICH79255	kanamycin	This work

Plasmid	Characteristics	Resistance	Reference
pMA532	5 respective GATC cluster Level M libraries in pMA72, endlinker: pICH79264	kanamycin	This work
pMA533	5 respective GATC cluster Level M libraries in pMA67, endlinker: pICH79289	kanamycin	This work
pMA534	5 respective GATC cluster Level M libraries in pMA68, endlinker: pICH79290	kanamycin	This work
pMA535	5 respective GATC cluster Level M libraries in pMA70, endlinker: pICH79311	kanamycin	This work
pMA536	5 respective GATC cluster Level M libraries in pMA71, endlinker: pICH79255	kanamycin	This work
pMA537	5 respective GATC cluster Level M libraries in pMA73, endlinker: pICH79277	kanamycin	This work
pMA543	40 kb Fragment Level M (GATC cluster): pMA528-532 in pMA60, endlinker: pICH50927	spectinomycin	This work
pMA544	40 kb Fragment Level M (GATC cluster): pMA533-537 in pMA66, endlinker: pICH50900	spectinomycin	This work
pMA561	5 respective no GATCs Level M libraries in pMA73, endlinker: pICH79277	kanamycin	This work
pMA562	5 respective no GATCs Level M libraries in pMA70, endlinker: pICH79311	kanamycin	This work
pMA657	synVicII-2.0	ampicillin	(Messerschmidt *et al.* under revision)
pMA678	Level P endlinker	chloramphenicol	(Messerschmidt *et al.* under revision)
pMA682	synVicII-2.0	ampicillin	(Messerschmidt *et al.* under revision)
pMA683	synVicII-oneGATC pMA525 pMA526 in pMA657, endlinker: pMA678	ampicillin	This work
pMA684	synVicII-fourGATC pMA543 & pMA544 in pMA657, endlinker: pMA678	ampicillin	This work
pMA685	synVicII-noMo pMA507 & pMA687 in pMA657, endlinker: pMA678	ampicillin	This work
pMA687	40 kb Fragment Level M (no GATCs): pMA497, pMA498, pMA500, pMA561, pMA562 in pMA333, endlinker: pICH50900	spectinomycin	This work
pMA689	synVicII-2.0 after Flp/FRT recombination	ampicillin	This work
pMA690	synVicII-oneGATC after Flp/FRT recombination	ampicillin	This work
pMA691	synVicII-fourGATC Flp/FRT recombined	ampicillin	This work
pMA692	synVicII-noMo Flp/FRT recombined	ampicillin	This work
pMA715	*cf*. MP2 *cat*::*spec*	spectinomycin	This work
pMA900	pCP20 with disrupted *bla* gene	chloramphenicol	(Messerschmidt *et al.* unpublished)

Table S3: Oligonucleotides used in this study.

Name	Sequence (5´ -> 3´)
640	TTTTAGGAAGGTCTCGGGAGTNDNGNGGCCNNCNDTGNCRNCNCNTWTNNGYBGHNGCNGNGGCNNCNR CANTNNCANHBGCCDTHNCCHHWANTNDCNGNGCCNNCNCDNCNTWGKCNTANCNVCNCVTNCNTHNGC HNCNNCNTACDNGCNNCNYGTWNNGNDGCNHNTCGTGGGTAVGANTCAAHGSTVNAATTAGT
641	TTAATCCTTGGTCTCCAGCGNCCNNCCRCNCNNHGCNTTNDCNSGGYNGNCDTWNNCCNCNNAGCWNCD NCNCNCWNCWTCNNCATCHNCNNCAAWNNGNNGYANTNNNGTNGCNNCNDTGNNGCNNCMNCTNCCKNT GNCRNCNCNCGNTGNCGNCNCNTWDNCTCNDAACTAATTNBASCDTTGANTCBTACCCACGA
642	TTTTAGGAAGGTCTCGGGAGNNGCKNNGCNVCDTNGYNGNCNRCHNCTNTNNCNCDNTGNNGCCNNCAT NNSGGCNNCGDCANTNGNNGGCDTNGNGNNGCHNNCNTTANTTYNGNCCNCGNNGYNCNNCNTWDNCHC NNCTNTNGNCGVCNTNTCNNGTNGCNGNGNGNACGGCTCWTBTAYCGCAWANTYNTGCTGTT
643	TTAATCCTTGGTCTCCAGCGBNGNNGCNNCNCNTHANNNGYCCAHNNCNTANTNDCNGNGYNCAANNCA AHNGDCNTWNNCNTTNNKGTNGYNGNCGGATCTNNNGCNNCANTNNGGNCMAHNNCCNACTBNCRTNGN CHANCNCNNCANTNTNGNCAHNTNCNTYGNNKAACAGCANRANTWTGCGRTAVAWGAGCCGT
644	TTTTAGGAAGGTCTCGGGAGTNNNGYNGCNNCWNTNGGCNNCCNCGNTGNCGNNGCNNCATNNGCCGNN GCDCNGNNGYNGNCDANGCNCNTNTMNCNGTNGNYGNTGACNNCNCNTWDCCNCGNCCNNCNACWNTNT ANTNNCNNAGCCMHNTHNTWNTNNCCGNGNNWACCTAATHTNATDCGGTWADTKMCCTCTGT
645	TTAATCCTTGGTCTCCAGCGNNGYNGNCNTWGNCNVCACNNCNNTGWCCNNCCGNGCNNHGCDNNGCNC NGCNDCGATCNCGNCVTAGATCCVRCDNMGGATCRCNCNNCTGATCNGCNDCANTANCNNCNACWNTTH ANTHANNGNNGYNGNCNTANTNNNGCNNCNNAACAGAGGKMAHTWACCGHATNADATTAGGT
466	TTTTAGGAAggtctcGGGAG
467	TTAATCCTTggtctcCAGCG
1518	ttctggaccagttgcgtgagcgcattttagcttccttagctcctg
1519	gatcaccaaggtagtcggcaaataaacgccatgggcatgtagtcaaaagc
1520	ttatttgccgactaccttggtgatctc
1521	atgcgctcacgcaactggtc

Table S4: MARSeG excluded DNA motifs for Level 1 libraries.

Name	Konsensussequence (5´ -> 3´)	Reference
GATC	GATC	(Marinus and Lobner-Olesen 2014)
BsaI recognition site	GGTCTC	
BpiI recognition site	GAAGAC	
DnaA box	YYHTMCRGM	(Schaefer and Messer 1991)
chi site	GCTNGTGG	(Cheng and Smith 1984, 1987)
dif site	DBBBCSBATAATRTAYATTATGTHAANT	(Hendrickson and Lawrence 2007)
KOPS	GGGNAGGG	(Bigot *et al.* 2005)
mat site	GTGACRNYGTCAC	(Mercier *et al.* 2008)
SlmA site	GTNANYNNWNACT	(Tonthat *et al.* 2011; Cho *et al.* 2011)
ter site	GNRNGTTGTAAYK	(Coskun-Ari and Hill 1997)
IHF	WATCAANNNNTTR	(Hales *et al.* 1994)
Fis	GNTYAAWWWTTRANC	(Finkel and Johnson 1993)

Name	**Konsensussequence (5´ -> 3´)**	**Reference**
V. cholerae par site chrI	NGTTNCACGTGAAACN	(Yamaichi *et al.* 2007)
V. cholerae par site chrII	NTTTACANTGTAAAN	(Yamaichi *et al.* 2007)
mig site	ANTTTTGCNGNNNNNCNGCAAAANT	(Yamaichi and Niki 2004)
Dcm site	CCWGG	(Marinus and Lobner-Olesen 2014)
TidL site	GTTGACGTCAGC	(Thiel *et al.* 2012)
TidR site	GCTGACGTCAGC	(Thiel *et al.* 2012)
RctB site	WTGATCAW	(Venkova-Canova *et al.* 2006)

Supplementary references

Badran AH, Liu DR (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. Nat Commun 6:8425. doi:10.1038/ncomms9425

Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, Dennis C, Grigoriev M, Allemand JF, Barre FX, Cornet F (2005) KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. EMBO J 24 (21):3770-3780. doi:10.1038/sj.emboj.7600835

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331):1453-1462

Cheng KC, Smith GR (1984) Recombinational hotspot activity of Chi-like sequences. J Mol Biol 180 (2):371-377

Cheng KC, Smith GR (1987) Cutting of chi-like sequences by the RecBCD enzyme of *Escherichia coli*. J Mol Biol 194 (4):747-750

Cho H, McManus HR, Dove SL, Bernhardt TG (2011) Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. Proc Natl Acad Sci U S A 108 (9):3773-3778. doi:10.1073/pnas.1018674108

Coskun-Ari FF, Hill TM (1997) Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in *Escherichia coli*. J Biol Chem 272 (42):26448-26456

Finkel SE, Johnson RC (1993) The Fis protein: it's not just for DNA inversion anymore. Mol Microbiol 7 (6):1023

Hales LM, Gumport RI, Gardner JF (1994) Determining the DNA sequence elements required for binding integration host factor to two different target sites. J Bacteriol 176 (10):2999-3006

Hendrickson H, Lawrence JG (2007) Mutational bias suggests that replication termination occurs near the *dif* site, not at *ter* sites. Mol Microbiol 64 (1):42-56. doi:10.1111/j.1365-2958.2007.05596.x

Marinus MG, Lobner-Olesen A (2014) DNA Methylation. EcoSal Plus 6 (1). doi:10.1128/ecosalplus.ESP-0003-2013

Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, Boccard F, Espeli O (2008) The MatP/*matS* site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. Cell 135 (3):475-485. doi:10.1016/j.cell.2008.08.031

Messerschmidt S, Schindler D, Zumkeller CM, Kemter F, Zimmer N, Waldminghaus T (under revision) Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*. Front Bioeng Biotechnol

Miller VL, Mekalanos JJ (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J Bacteriol 170 (6):2575-2583

Schaefer C, Messer W (1991) DnaA protein/DNA interaction. Modulation of the recognition sequence. Mol Gen Genet 226 (1-2):34-40

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synth Biol. doi:10.1021/acssynbio.6b00089

Thiel A, Valens M, Vallet-Gely I, Espeli O, Boccard F (2012) Long-range chromosome organization in *E. coli*: a site-specific system isolates the Ter macrodomain. PLoS Genet 8 (4):e1002672. doi:10.1371/journal.pgen.1002672

Tonthat NK, Arold ST, Pickering BF, Van Dyke MW, Liang S, Lu Y, Beuria TK, Margolin W, Schumacher MA (2011) Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. EMBO J 30 (1):154-164. doi:10.1038/emboj.2010.288

Venkova-Canova T, Srivastava P, Chattoraj DK (2006) Transcriptional inactivation of a regulatory site for replication of *Vibrio cholerae* chromosome II. Proc Natl Acad Sci U S A 103 (32):12051-12056. doi:10.1073/pnas.0605120103

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6 (2):e16765. doi:10.1371/journal.pone.0016765

Yamaichi Y, Fogel MA, McLeod SM, Hui MP, Waldor MK (2007) Distinct centromere-like *parS* sites on the two chromosomes of *Vibrio spp*. J Bacteriol 189 (14):5314-5324. doi:10.1128/JB.00416-07

Yamaichi Y, Niki H (2004) *migS*, a cis-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome. EMBO J 23 (1):221-233. doi:10.1038/sj.emboj.7600028

3.5 SeqA complexes in *Escherichia coli* exchange proteins rapidly and vary depending on replication patterns

Über die Lokalisation und die generelle Funktion des SeqA-Proteins ist bereits viel bekannt, die Funktionsweise von SeqA an den Replikationsgabeln und am Replikationsursprung sind im Detail jedoch nicht vollständig verstanden. Es fehlen insbesondere vergleichende, quantitative Analysen unter verschiedenen Wachstumsbedingungen, um das Verständnis der Funktionsweise von SeqA zu erweitern und daraus ein Modell von SeqA an den Replikationsgabeln zu generieren.

In dem vorliegenden Kapitel wurde deshalb zum einen analysiert, ob SeqA eine Zellzyklusregulation aufweist. Die Daten zeigen, dass SeqA konstant produziert wird und eine erneute Initiation der DNA-Replikation oder eine Duplikation von *seqA* keine Änderung der SeqA Produktion bewirkt. Zum anderen wurde die Menge an SeqA-Molekülen pro Zelle bestimmt, die unter verschiedenen Wachstumsbedingungen vorliegt, und Analysen des Zellzyklus durchgeführt. Des Weiteren wurde die Fraktion an gebundenen SeqA-Molekülen durch Microfluidics-gekoppelte Fluoreszenzmikroskopie analysiert und die SeqA-Dynamik mittels eines FRAP-Experiments belegt. Außerdem konnte gezeigt werden, dass zwischen der Dam-Methyltransferase und dem SeqA-Protein unter allen Wachstumsbedingungen ein ähnliches Mengenverhältnis vorliegt, was auf ein sensitives Gleichgewicht zwischen den beiden Proteinen hinweist. Die Daten aus den verschiedenen experimentellen Ansätzen wurden kombiniert, um ein abschließendes Modell von SeqA an den Replikationsgabeln zu erstellen. Unter der Annahme, dass das Gleichgewicht zwischen ungebundenem und gebundenem SeqA unter allen Wachstumsbedingungen gleich ist, die Größe der SeqA-Strukturen an den Replikationsgabeln jedoch variieren kann, konnte ein Modell für die Menge an gebundenem SeqA pro Replikationsgabel und frei diffundierendem SeqA über den Zellzyklus generiert werden.

Daniel Schindler hat unter Rücksprache mit Torsten Waldminghaus die Studie konzipiert und durchgeführt. Daniel Schindler und Matthias Bruhn haben unter Rücksprache mit Torsten Waldminghaus die Zellzyklusabhängigkeit von SeqA analysiert sowie die SeqA und Dam-Mengenverhältnisse bestimmt. Microfluidic Experimente und deren Auswertung wurden in der Arbeitsgruppe von Johan Elf durch Ebba Gregorsson Lundius und Johan Elf unter Rücksprache mit Daniel Schindler und Torsten Waldminghaus durchgeführt. Daniel Schindler war für zwei Wochen Gast im Labor von Johan Elf, um das Projekt zu koordinieren und die Microfluidic Technik zu erlernen. Modellierungen von SeqA wurden von Sean Murray unter Rücksprache mit Torsten Waldminghaus durchgeführt. Das Manuskript wurde von Torsten Waldminghaus und Daniel Schindler verfasst unter Beteiligung von Ebba Gregorsson Lundius und Johan Elf.

Manuscript in preparation

SeqA complexes in *Escherichia coli* exchange proteins rapidly and vary depending on replication patterns

Daniel Schindler, Ebba Gregorsson Lundius*, Matthias Bruhn, Sean Murray, Johan Elf, Torsten Waldminghaus*

** equal contribution*

Abstract

Bacterial chromosomes are continuously segregated with simultaneous replication. In fast growing bacteria overlapping replication cycles lead to cells with multiple chromosomes. In *Escherichia coli*, the main player in organization of newly synthesized DNA is the SeqA protein. It binds specifically to hemi-methylated GATCs which occur temporary at newly synthesized DNA. The dynamic association of SeqA with the replication forks is thought to originate from competitive binding of SeqA and re-methylation of GATCs by the Dam methyltransferase. Here we show by quantitative SeqA and Dam analysis and mathematical modeling that SeqA structures differ dependent on the replication pattern. Using high resolution microscopy, we found that SeqA does not bind to fully-methylated replication origins as proposed by others. Fluorescence Recovery After Photobleaching (FRAP) experiments show that SeqA exchanges within seconds between distantly separated SeqA complexes.

Introduction

In eukaryotic cells, DNA replication and segregation are two separated cell cycle processes (Ghosh *et al.* 2006; McIntosh *et al.* 2002). In contrast, bacterial chromosomes are continuously segregated with simultaneous replication (Lobner-Olesen and Kuempel 1992; Kuzminov 2014; Youngren *et al.* 2014). This coordinated process has to be well organized to maintain stability of the genetic material. A protein which might fulfill the task of organizing and segregating the sister chromosomes intuitively needs two important characteristics. First, it has to bind the new DNA specifically. Second, it needs to form a structure to hold different parts of the DNA together or apart from each other. The protein matching these characteristics in *E. coli* and related bacteria is SeqA (Waldminghaus and Skarstad 2009; Joshi *et al.* 2013; Lu *et al.* 1994). Structural analysis of SeqA showed two distinct domains connected by a flexible linker: the C-terminal DNA-binding domain and a N-terminal dimerization domain which is able to multimerise several dimers to build a higher-order structure (Guarne *et al.* 2002; Guarne *et al.* 2005; Chung *et al.* 2009; Waldminghaus and Skarstad 2009). SeqA was discovered as negative regulator of DNA replication initiation in a screen for *oriC* sequestration factors (Lu *et al.* 1994). The origin sequestration by SeqA is mediated by its binding to GATC sites which occur at high frequency within *oriC* (Campbell and Kleckner 1990). GATCs are methylated at the adenine by the Dam methyltransferase (Geier and Modrich 1979; Marinus and Morris 1973; Marinus and Lobner-Olesen 2014). Right after replication the GATCs are methylated on the old strand only (Marinus 1987). SeqA specifically binds this newly synthesized, hemi-methylated DNA resulting in a block of re-initiation by the initiation factor DnaA at *oriC* (von Freiesleben *et al.* 1994; Slater *et al.* 1995). Consequently, deletion of *seqA* or *dam* lead to dysregulated, asynchronous DNA replication initiation (Lu *et al.* 1994; Boye and Lobner-Olesen 1990). Beside SeqAs role at *oriC* it was also found to bind hemi-methylated GATCs throughout the chromosome occurring behind the replication forks (Waldminghaus *et al.* 2012; Waldminghaus and Skarstad 2010; Sanchez-Romero *et al.* 2010). SeqA at the replication forks is thought to contribute to chromosome segregation (Joshi *et al.* 2013; Han *et al.* 2004; Brendler *et al.* 2000; Stokke *et al.* 2011). In this context, SeqA was found to mediate a prolonged sister chromosome cohesion by blocking topoisomerase IV-dependent decatenation (Joshi *et al.* 2013). Recent findings suggest that the SeqA structure is 200 nm away from the replisome but the two sister strands are kept close together with a distance of less than 30 nm (Helgesen *et al.* 2015).

As indicated above, SeqA binding is tightly connected to the methylation action of the Dam methyltransferase. The hemi-methylated GATCs produced by passage of the replication fork are targets for both proteins, Dam and SeqA. Binding of SeqA blocks Dam from re-methylating the respective GATC (Kang *et al.* 1999; Katayama *et al.* 1997; Taghbalout *et al.* 2000). On the other hand,

Dam will switch GATCs into non-targets for SeqA by adding a methyl group to the un-methylated strand. The dynamic interplay between these two processes is thought to generate a dynamic association of SeqA with the replication fork in a treadmilling-like way (Waldminghaus *et al.* 2012; Joshi *et al.* 2013). *In vitro* data indicate that Dam acts as a monomer in a processive manner (Urig *et al.* 2002; Horton *et al.* 2005). During one binding event, Dam scans around 3000 target sites in a random walk, that on average leads to re-methylation of 55 target sites (Urig *et al.* 2002). This indicates that low amounts of Dam are sufficient to re-methylate the whole chromosome (Boye *et al.* 1992; Li *et al.* 2014; Szyf *et al.* 1984). *In vivo* data show the hemi-methylated state to last for 1-2 minutes corresponding to a stretch of 60-120 kb of DNA behind the replication fork (Campbell and Kleckner 1990; Ogden *et al.* 1988). A competition of Dam and SeqA for GATCs is supported by the shortening of the hemi-methylation period and to a hypermutable phenotype under Dam overproduction conditions (Herman and Modrich 1981; Yang *et al.* 2004). Notably, excess SeqA leads to strong deficiencies in chromosome segregation (Saint-Dic *et al.* 2008; Bach *et al.* 2003). A sensitive equilibrium of Dam and SeqA appears to be necessary for maintenance of the cell cycle and genomic stability.

Here we investigate the mechanism of Dam and SeqA action by quantification of SeqA relative to Dam in the context of different replication patterns. We find that SeqA complexes in the cell are not isolated entities. Indeed, there is exchange of SeqA molecules between complexes within seconds. These findings could explain why SeqA binds equally well to new and old replication forks in cases of overlapping replication. Contradicting results had been found for SeqA binding to fully-methylated *oriC* regions (Waldminghaus *et al.* 2012; Helgesen *et al.* 2015; Slater *et al.* 1995; Taghbalout *et al.* 2000). Our high-resolution microscopy data suggest an extended cell-cycle period without SeqA binding to the fully-methylated *oriC*.

Results

SeqA levels increase gradually during the cell cycle

One goal of this study was a quantitative understanding of the proposed SeqA-complex treadmilling process. Experiments were carried out with varied growth conditions leading to different patterns of DNA replication ranging from simple replication in cells grown with acetate as carbon source to complex overlapping replication in LB grown cells. Initiation of DNA replication happens at a fixed time point of the bacterial cell cycle under balanced growth. SeqAs involvement in this process might suggest that its expression could be cell cycle regulated. To test this hypothesis, we analyzed the relative amount of a SeqA-YFP fusion in correlation with the cell size as a proxy for the cell cycle stage

by fluorescence microscopy (Fig. 1). The SeqA-YFP fusion was inserted at the native SeqA locus and verified to be functional as reported previously (Babic *et al.* 2008). Cells grown in four different media were grouped according to cell size (Fig. 1). Respective fluorescence intensities showed a linear increase of the SeqA amount for all four growth conditions with no major steps that would indicate cell cycle dependent regulation. We conclude that SeqA expression is not cell cycle regulated and that SeqA gradually increases from cell birth to cell division. Results were similar in biological replicates (Fig. S1).

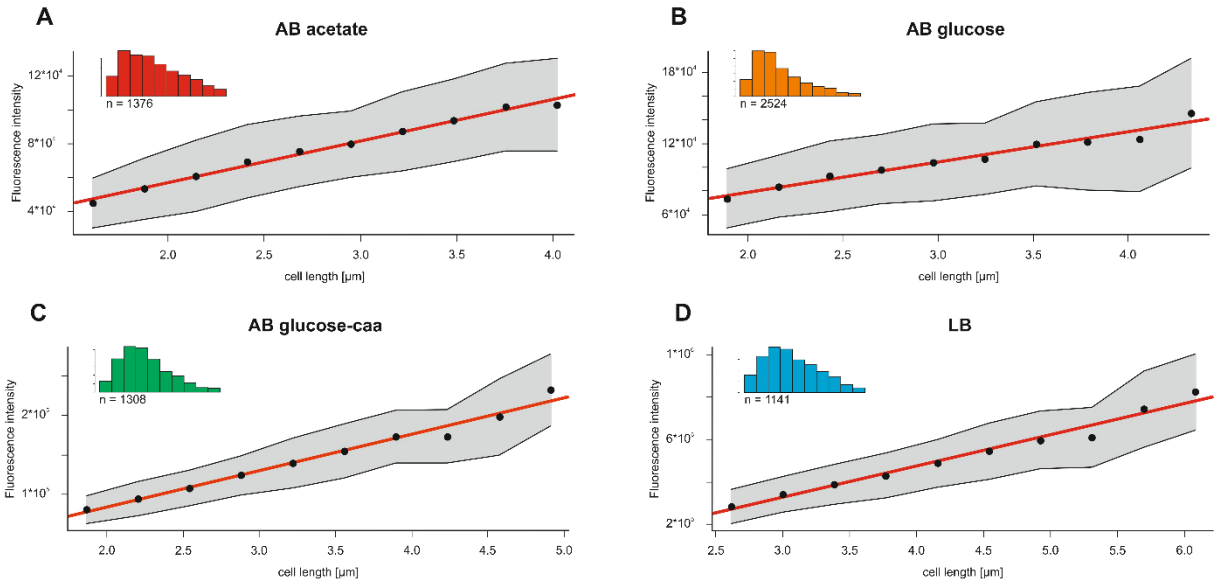


Figure 1: SeqA levels increase linearly during the *E. coli* cell cycle. SeqA-YFP intensities of exponential growing cultures (DS116) were measured by fluorescence microscopy and quantified as described in the Material and Methods section. Four different growth media were used with increasing growth rate from **A** AB acetate to **D** LB. Intensity values of individual cells where grouped according to cell size as proxy for cell cycle stage (black dots) with the respective linear regression (red line) and the standard deviation (grey area). Each medium is color coded by the histogram which gives the cell size distribution with numbers of analyzed cells below the histogram. A replicate data set is shown in supplementary figure S2.

Numbers and concentration of SeqA molecules vary depending on replication patterns

Independent studies report a number of about 1,000 SeqA molecules per cell in minimal medium with glucose as sole carbon source (Li *et al.* 2014; Slater *et al.* 1995; Schmidt *et al.* 2016). However, a systematic quantification of SeqA amounts under varied growth conditions is missing. To this end we performed quantitative Western Blot analysis using a SeqA antiserum and cell counting (Fig. S2, see Method section for details). Normalizing to the 1,000 SeqA molecules in glucose medium we found large differences for the analyzed growth conditions ranging from about 800 in AB acetate to about 3,000 in AB glucose-caa medium and about 5,000 molecules in LB medium (Fig. 2). Beside the

absolute number of SeqA molecules per cell the actual SeqA concentration is a critical parameter. To calculate relative SeqA concentrations we measured cell sizes for the four different growth conditions by flow cytometry. While absolute SeqA numbers varied more than six-fold for the four different growth conditions, the concentration did vary maximally two-fold (Fig. 2). With increasing growth rate the number of replication forks and origins as targets for SeqA increase. The absolute SeqA number but not the SeqA concentration appears to increase gradually with increasing growth rate (Fig. 2). This result was confirmed by an alternative method where SeqA was measured by fluorescence microscopy normalized to HU (Fig. S3, see below).

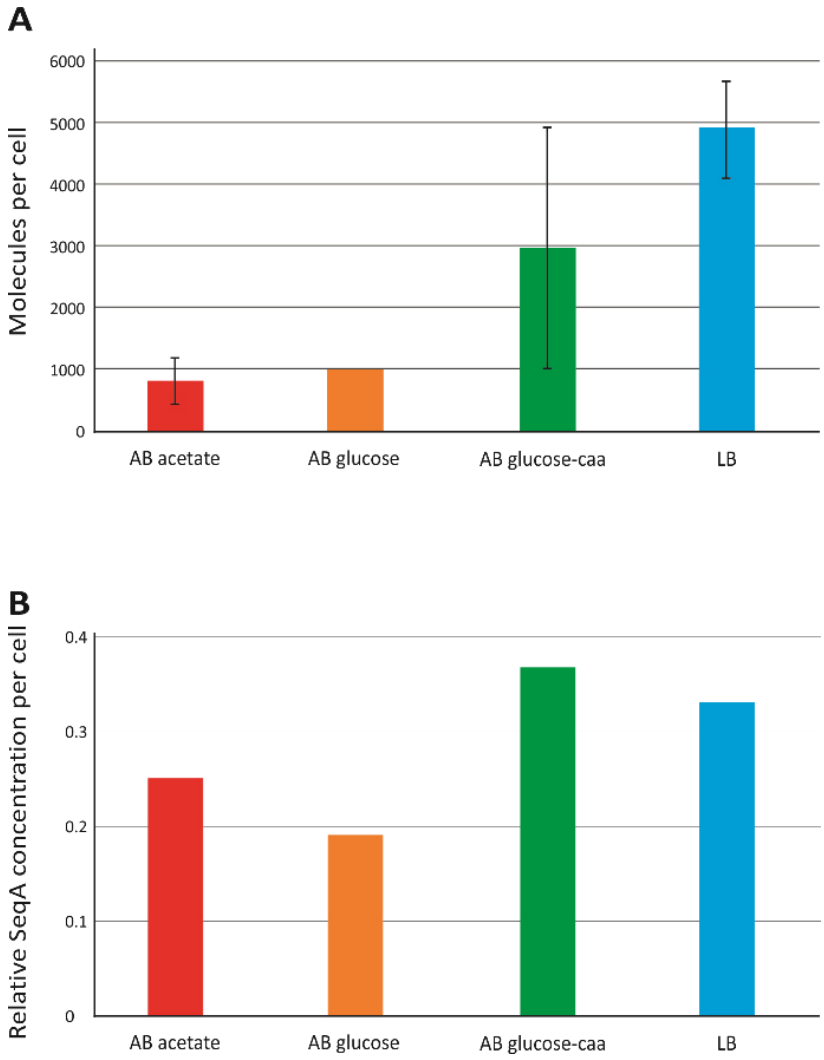


Figure 2: SeqA concentration in exponential growing *E. coli* MG1655 cells at different growth conditions. *E. coli* cells were grown in batch cultures at 37 °C in four different media as indicated. SeqA was detected by an anti-SeqA serum and the intensity was normalized to the cell number per sample (see figure S2). **A** amount of SeqA in AB glucose medium was set to 1,000 molecules per cell as measured with different methods previously (Li *et al.* 2014; Schmidt *et al.* 2016; Slater *et al.* 1995). Other molecule numbers were calculated relative to the glucose medium. SeqA molecules per cell increases with complexity of DNA replication (AB acetate 806 +/- 392; AB glucose-caa 2972 +/- 1949; LB 4915 +/- 776 molecules per cell). **B** cell sizes were determined by flow cytometry to calculate the relative SeqA concentration (See Material and Methods for details).

Constant ratios of Dam methyl-transferase to SeqA for varied replication patterns

SeqA binding is directly dependent on the activity of the Dam methyl-transferase which methylates hemi-methylated GATCs and thus changes them to a non-target for SeqA. The Dam-to-SeqA ratio is therefore a critical parameter for modelling of SeqA binding events. To measure the Ratio of Dam to SeqA we performed quantitative fluorescence microscopy using a HupB-mCherry reporter fusion for normalization. Strain DS126 carries the functional SeqA-YFP fusion as above combined with a HupB-mCherry fusion; strain DS181 encodes a functional Dam-YFP fusion together with HupB-mCherry. Ratios of the YFP signal relative to the mCherry signal was measured by fluorescence microscopy for the four different growth conditions (Fig. 3 A-B). Dividing the respective values by one another gives a relative measure of Dam to SeqA ratios (Fig. 3 C). In medium with glucose as sole carbon source the number of Dam molecules is about 5 % of SeqA molecules correlating well with previous estimates (Boye *et al.* 1992). This Dam-to-SeqA ratio remains relatively unchanged for the four tested growth conditions (Fig. 3 C). To verify our result, an *E. coli* strain was constructed encoding a SeqA-YFP and a Dam-YFP fusion. Comparative western blotting was used with an anti-GFP antibody to detect and quantify both proteins (Fig. S4). For cells grown in LB medium the Dam-to-SeqA ratio was measure to be 0.06 +/- 0.017 verifying the results from relative fluorescence microscopy.

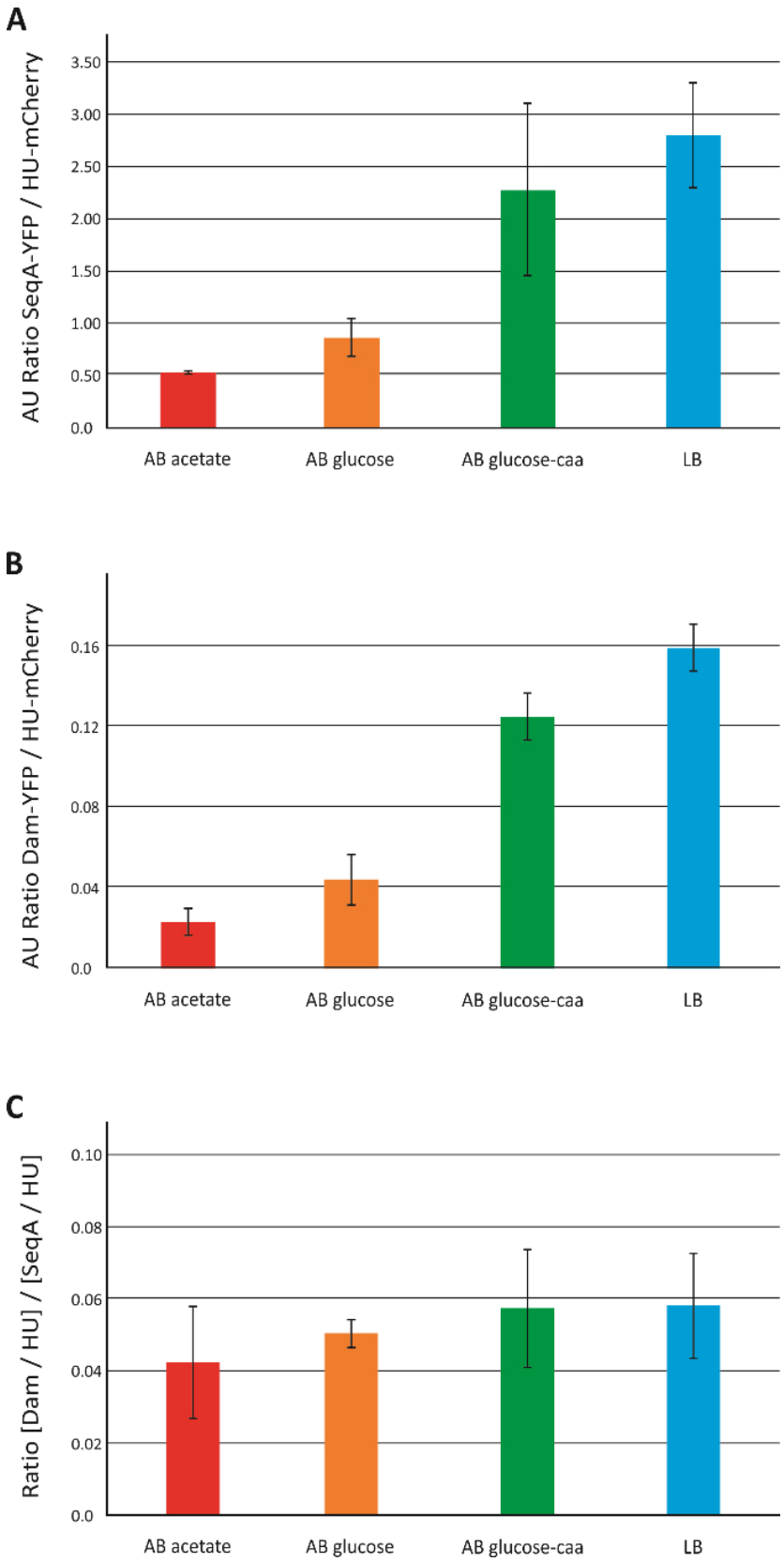


Figure 3: Dam-to-SeqA ratios in exponential growing *E. coli* cells at different growth conditions. Using strains (DS126 and DS183) with a HU-mCherry fusion and **A** SeqA-YFP or **B** Dam-YFP fusion the ratio of SeqA/HU and Dam/HU was determined on single cell level by fluorescence microscopy. **C** respective ratios where used to generate the relative Dam concentration in dependence of SeqA.

SeqA binding to the chromosome

To study SeqA binding to the *E. coli* chromosome we performed high-resolution fluorescence microscopy in microfluidic chips (Ullman *et al.* 2013). The general procedure is outlined in supplementary figure S5 and details are given in the method section and previous publications (Wallden *et al.* 2016). Plotting the SeqA signal of cells sorted according to their volume for slow growing cells (M9 acetate medium) shows a clear accumulation near the cell middle (Fig. 4 A). Under slow growth conditions *E. coli* replicates with a simple cycle where new born cells have one fully replicated chromosome, then initiate replication and go through a D period without replication after termination (Zaritsky and Woldringh 2015; Stokke *et al.* 2012). Since SeqA binding is specific for hemi-methylated DNA which only appears in the C period, SeqA foci should only be seen during C period. Accordingly, the smallest and largest cells lack SeqA foci (Fig. 4 A). This result is contradicting to previous findings where SeqA foci were also seen in cells of B and D period (Helgesen *et al.* 2015). To get a deeper insight the probability for a cell to have a certain number of SeqA foci was calculated relative to cell length (Fig. 4 B). The analysis supports that SeqA molecules are not bound to DNA in most cells within a period before and after cell division. In addition, some cells in C period carry two foci, which is not evident in the aggregated plot (Fig. 4 A). The distance of foci was measured as outlined in figure 4C for the long and short cell axis (Fig. 4 D & E). Distributions of distances between foci appear to be constant throughout replication and are in a similar range for the long and short cell axis. Notably, SeqA clusters are rarely separated by half of the cell length or more. Analysis of individual cells shows that the SeqA clusters split and join frequently but remain at the cell center and its vicinity (Fig. S5).

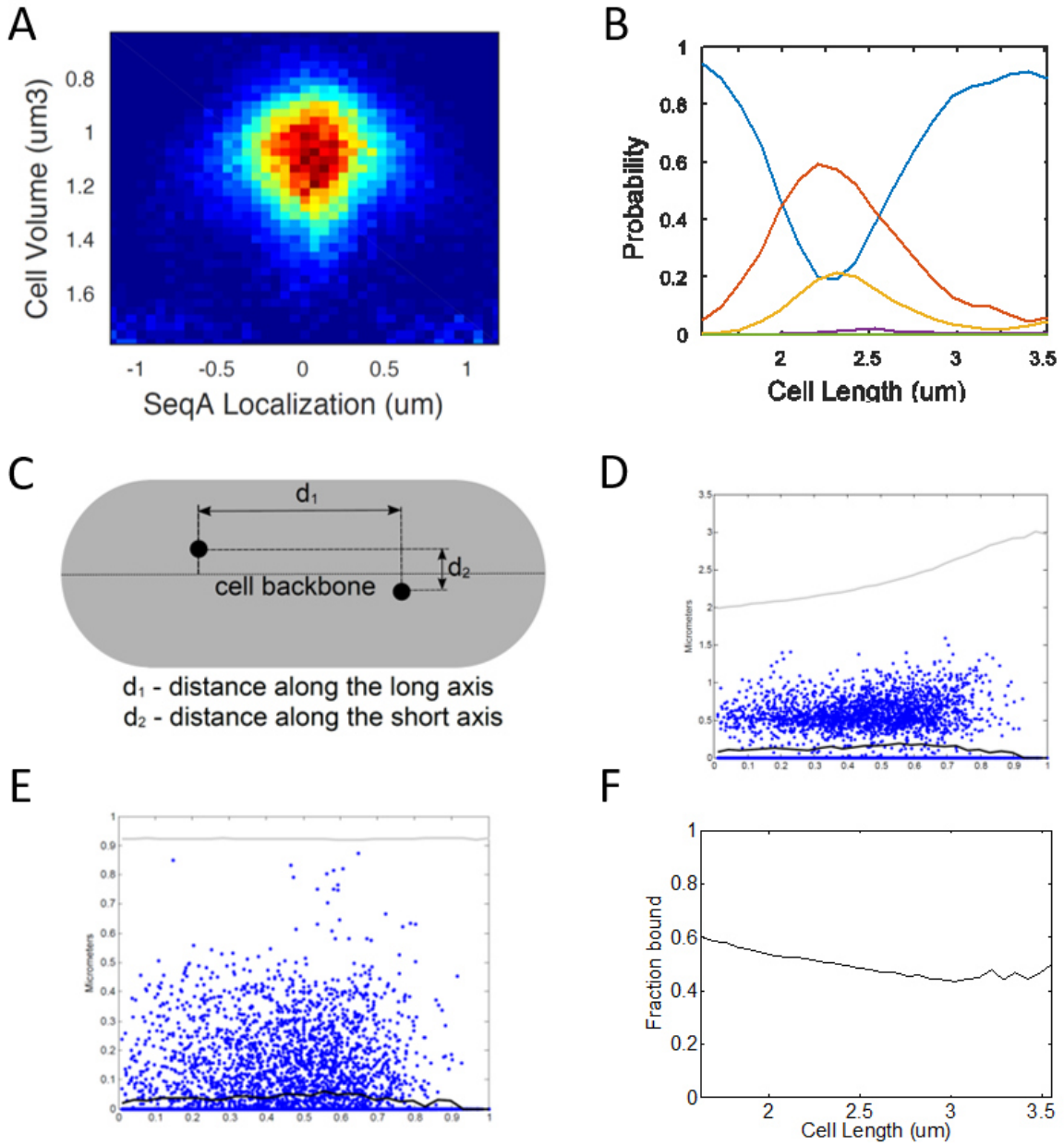


Figure 4: SeqA binding in slow growing *E. coli* MG1655 cells. **A** DS116 was grown in microfluidic devices at 37 °C as described previously (Ullman *et al.* 2013) with acetate as sole carbon source and SeqA-YFP was monitored during the cell cycle. The heatmap shows SeqA spots are only detectable in a certain period of the cell cycle within an *E. coli* population. **B** shows the probability of a cell to possess zero (blue), one (orange), two (yellow), three (purple) or four (green) SeqA-YFP foci depending on the cell size. **C** explains the measurements of distances between distinct spots presented in **D** and **E**. **D** distances of two SeqA clusters along the long axis during the cell cycle in dependence on to the cell middle in individual cells. **E** distance along the short axis the short axis between two SeqA cluster within individual cells. **F** Determination of the SeqA fraction bound to the DNA within the period of the cell cycle where spots appear (*cf.* **A** and **B**).

Individual replication forks bind about 100 SeqA dimers on average

Different estimates of SeqA molecules per replication fork have been made but never experimentally studied systematically (Li *et al.* 2014; Schmidt *et al.* 2016; Slater *et al.* 1995). As first step to calculate respective numbers we determined the fraction of bound SeqA molecules throughout the cell cycle based on the fluorescence microscopy data (Fig. 4 F, see Material and Methods for details). The fraction of SeqA bound to DNA is relatively constant with an average of about 50 %. As second step, we determined the cell cycle parameters by flow cytometry and growth studies (Suppl. table S1; see Material and Methods for details). The derived numbers of forks, the fraction of bound SeqA and the SeqA molecules per cell where used to calculate the number of SeqA bound to individual replication forks during the cell cycle for four different growth rates (Fig. 5). Calculations were based on two alternative assumptions. The first assumption was that under all conditions 50 % of the cellular SeqA is bound to the replication forks as found for the slow growing cells (Fig. 5 A). The second assumption is a simple binding model where a binding constant K is calculated based on the 50 % of SeqA bound to the forks under slow growth conditions and the fraction of SeqA bound at other growth conditions is n/(n+K), where n is the number of forks. Both calculation approaches give an average number of about 100 SeqA dimers per replication fork. Notably, the amount of SeqA bound to individual replication forks can vary up to three-fold within the cell cycle in fast growing cells (Fig. 5).

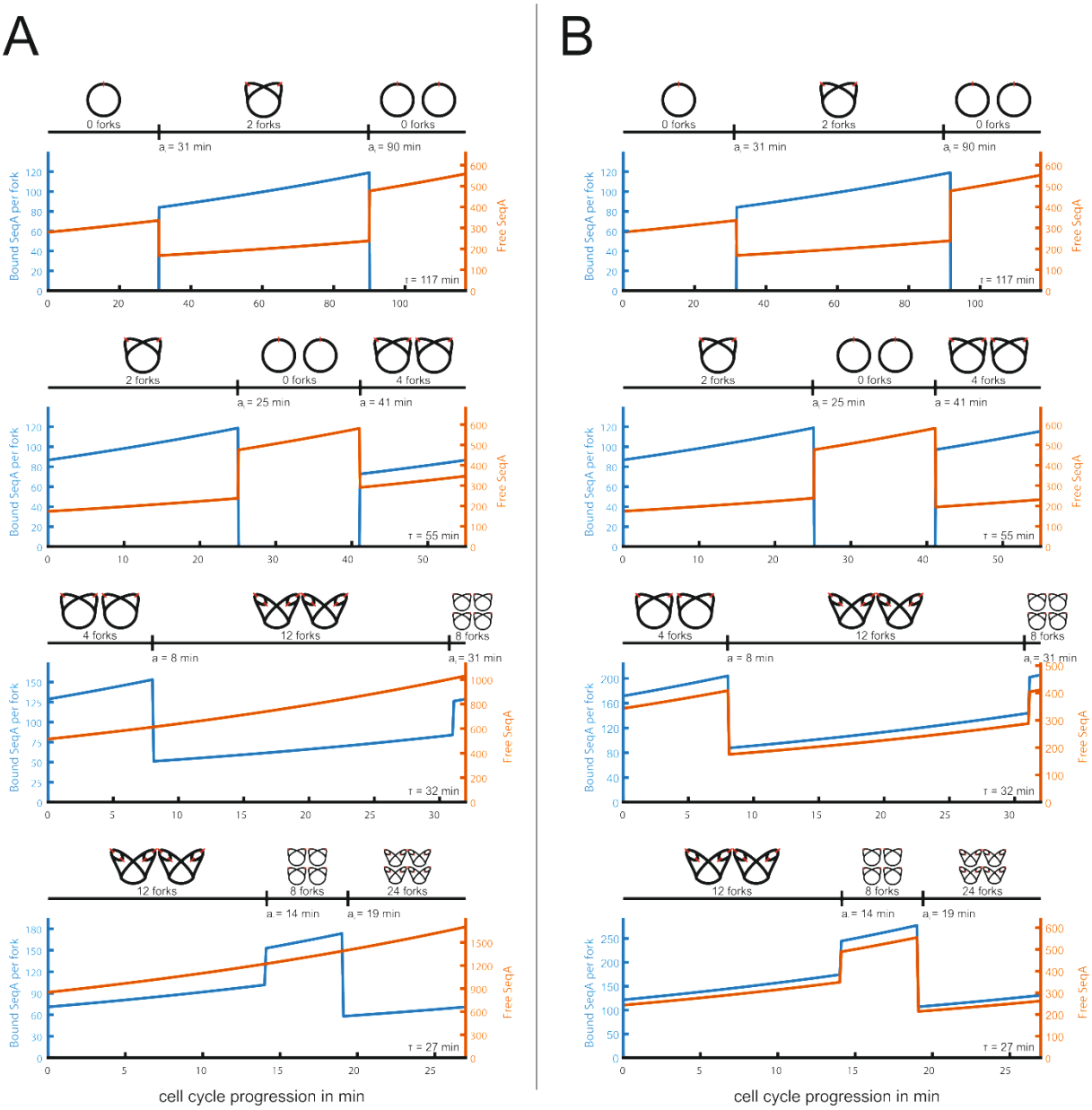


Figure 5: SeqA binding to replication forks. All experimental results are combined to produce two models each with alternative assumptions. Growth conditions are ordered from simple to higher complexity from top to bottom. Above each individual plot chromosomal replication pattern and the number of forks at a certain period of the division cycle is indicated. Additionally, the timepoint of initiation (a_i) and termination (a_t) is given. **A** one model consists of the assumption that 50 % of SeqA is bound (blue) and unbound (orange) to the DNA of replication forks respectively (*cf*. Fig. 4). **B** the second model calculates a binding constant K based on 50 % of SeqA bound at slow growth conditions and the fraction of SeqA bound at other growth conditions is n/(n+K), where n is the number of forks.

Rapid exchange between SeqA clusters within cells

SeqA binding was analyzed in fast growing cells by fluorescence microscopy in microfluidic chambers as above (Fig. 6 A & B). Small cells showed most of the SeqA signal in the cell center while bigger cells comprised two main regions of bound SeqA at the quarter and three quarter position in accordance to previous findings (Fig. 6A (Kuwada *et al.* 2015; Onogi *et al.* 1999; Helgesen *et al.* 2015)). The treadmilling model for SeqA clusters tracking the replication fork would suggest that individual SeqA molecules remain close to one replisome with constant rebinding after falling of the DNA. To test this hypothesis, we performed Fluorescence Recovery After Photobleaching (FRAP) experiments (Fig. 6 C & D). One fluorescence spot within a cell was bleached while the other was not. However, within seconds the bleached fluorescence recovered, indicating a fast exchange of SeqA molecules between distant SeqA structures in the cell.

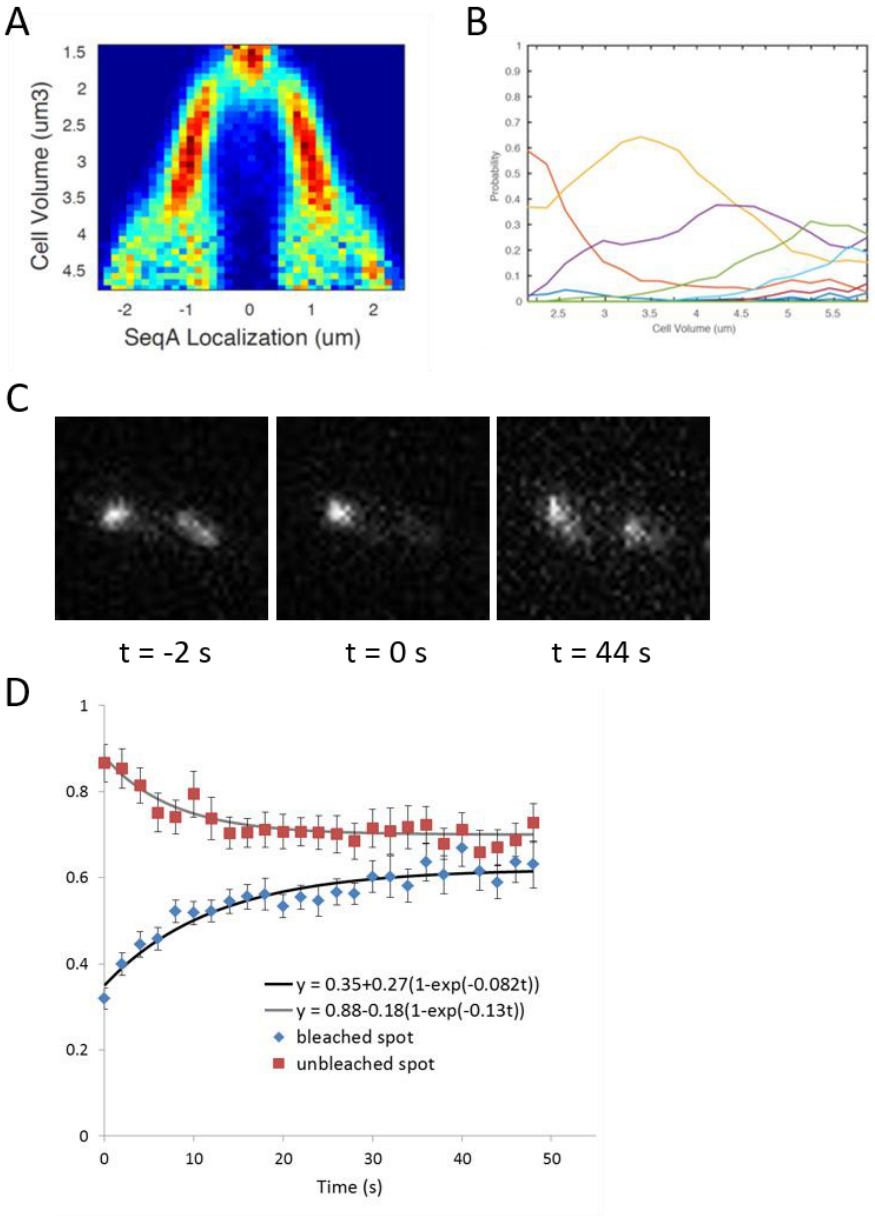


Figure 6: SeqA exchange rate between replisomes after bleaching. **A** DS116 was grown in microfluidic devices at 37 °C as described previously (Ullman *et al.* 2013) in M9 medium with glucose as carbon source supplemented with RPMI. SeqA-YFP was monitored during the cell cycle. The heatmap shows that SeqA spots are detectable during the whole cell cycle. **B** shows the probability of a cell within a population to possess zero (dark blue), one (orange), two (yellow), three (purple), four (green), five (light blue) or six (red) SeqA-YFP foci depending on the cell size. **C** Two SeqA bound replisomes in the same cell just before bleaching (t = -2 s), just after bleaching (t = 0 s) and at equilibrium (t = 44 s). **D** FRAP in fast growing cells with simultaneous replication at two replisomes (n = 20 cells). Fluorescence in one replisome was bleached and fluorescence intensity of both the bleached (blue) and the unbleached (red) replisome was quantified. Recovery curves from the bleached spot and the unbleached spot are fitted to exponential equations. Error bars show the standard deviation of each data point.

Discussion

SeqA structures at the replication forks vary in size and are highly interconnected

SeqA molecules behind the replication fork have been viewed as dynamic filament potentially forming a hyperstructure and moving in a treadmilling-like fashion (Waldminghaus *et al.* 2012; Norris *et al.* 2000). Various assumptions have been made about the length of the SeqA-covered region behind the replication fork, ranging from 30 to 400 kbps (Martina *et al.* 2012; Helgesen *et al.* 2015; Brendler *et al.* 2000; Joshi *et al.* 2013). Here we present the first quantitative approach to analyze the SeqA structures systematically. Previous estimations of SeqA molecules per fork where all based on the assumption that SeqA binding correlates with the period of hemi-methylation. Based on measurements of Campbell and Kleckner it was assumed that GATC sites are hemimethylated for an average time in the range of minutes (Brendler *et al.* 2000; Campbell and Kleckner 1990). Considering a replication time of 1,000 bps per second would for example give 120 kbps of hemi-methylated DNA trailing the replication fork. The *E. coli* genome contains 19124 GATC sites corresponding to one every 243 bps and about 500 GATCs per 120 kbps. If each GATC is bound by one SeqA, about 1,000 molecules would be tracking behind each replication fork (500 at each strand). In *E. coli* cells replicating with up to 24 replication forks this number seems incompatible with any measurement of SeqA molecule numbers (Li *et al.* 2014; Slater *et al.* 1995; Schmidt *et al.* 2016). However, SeqA might actually not bind to each one of the GATC sites. It has been shown that SeqA binds as dimer to a set of two GATCs which are not too far away from each other on the DNA sequence (Brendler *et al.* 2000). Based on the respective biochemical data there should be about 1,750 SeqA dimer binding sites on the *E. coli* chromosome, one every 2,650 bps on average. This would correspond to 90 SeqA molecules behind each replication fork which is very close to our measurements (Fig. 5). This indicates that SeqA binding does not bridge single GATCs on different sister chromosome strands although its contribution to sister chromosome cohesion might suggest so (Joshi *et al.* 2013). In any case, a strict treadmilling model with SeqA molecules binding to the newest hemi-methylated GATC and the SeqA furthest away from the replisome leaving the DNA does not match the fast recovery after photobleaching within seconds as measured here (Fig. 6). Notably, SeqA molecules are highly dynamic not only regarding binding and unbinding behind individual replication forks but also from stretches of newly replicated DNA in different cell halves (Fig. 6). The fast binding-unbinding cycles of SeqA might actually be critical for its main function – the sequestration of the replication origin *oriC* to inhibit early re-initiations. A strict treadmilling at the replication fork including spatial constrains would hold SeqA at the replication forks also when new rounds of replication are initiated before the old forks terminate (multi-fork replication). The high cellular mobility und high turn-over of SeqA observed here might guarantee that some SeqA is available for the important process of origin

sequestration. However, fast binding-unbinding of SeqA might not agree with our current view of *oriC* sequestration where SeqA binds to the replication origin for one third of the cell cycle and then releases it for the initiation protein DnaA to bind (Lu *et al.* 1994; Boye 1991; Campbell and Kleckner 1990; Ogden *et al.* 1988). It remains to be determined how the on-off rate of SeqA binding varies depending on GATC density that might be a main factor in SeqA binding cooperativity (Chung *et al.* 2009). Notably, ectopic GATC clusters have been shown to enhance SeqA binding and genes neighboring *oriC* are significantly enriched in GATCs (Sobetzko *et al.* 2016; Waldminghaus *et al.* 2012).

SeqA does not bind to fully-methylated oriCs

Biochemical data have shown many years ago that SeqA binding is specific for hemi-methylated GATCs but some binding was also observed for fully methylated *oriC* fragments (Slater *et al.* 1995). More recently Helgesen and colleagues find SeqA foci in 90 % of all cells grown slowly in acetate medium (Helgesen *et al.* 2015). This is surprising because there should be an extended B and D period without ongoing replication. Their interpretation is that fully methylated *oriC* is bound by SeqA in accordance with the biochemical experiments (Slater *et al.* 1995). Our findings clearly show a big portion of the cells in acetate grown cells to lack SeqA foci corresponding to cells in B and D period. This is in agreement with ChIP-Chip experiments with synchronized cell cultures which showed SeqA not to bind to fully-methylated replication origins or elsewhere on the chromosome before initiation (Waldminghaus *et al.* 2012). One difference of the Helgesen study compared to our study is they used the *E. coli* strain AB1157 while strain MG1655 was used in this study. Differences between these two strains with regard to segregation have been reported before especially for slow growth conditions (Mercier *et al.* 2008).

Material and Methods

Bacterial strains, plasmids, oligonucleotides, growth conditions and strain construction

All bacterial strains, plasmids and oligonucleotides used in this study are listed in Tables S2 to S4. Cells were grown in LB medium or AB media (Clark and Maaløe 1967; Jensen 1993) supplemented with 10 µg/mL thiamin, 100 µg/mL uridine and either 0.2 % glucose and 0.5 % casamino acids or 0.4 % sodium acetate at 37 °C. Microfluidic experiments were performed in M9 Media with respective carbon source, instead of casamino acids RPMI 1640 amino acids (R7131, Sigma-Aldrich) was used. OD measurements in LB and AB were performed at λ = 600 nm and λ = 450 nm respectively. Antibiotic selection was used at the following concentrations: chloramphenicol 30 µg/mL; ampicillin 100 µg/mL. For chromosomal integration of the Dam-YFP in frame fusion *yfp*-FRT-*cat*-FRT was amplified from pSeqA-C with the primers 750/751. PCR amplicon was

transformed into AB330 utilizing the λ red recombineering system as described previously (Datsenko and Wanner 2000). The genomic insertion was verified by PCR and Sanger sequencing. Transfer of genomic fusions into strains was done by P1 transduction. Removal of selection markers were achieved by Flp/FRT recombination (De Souza Silva and Blokesch 2010).

General microscopic analysis

For analyzing SeqA-YFP protein abundance during cell cycle, cells were grown in corresponding media to detect SeqA-YFP intensity by snapshot imaging of exponential growing cells (OD = 0.15). Cells were imaged on 1 % agarose pads supplemented with respective media except for LB cells, PBS was used. Phase contrast and YFP channel images were acquired using a Nikon Eclipse Ti-E microscope with a phase-contrast Plan Apo l oil objective (100; numerical aperture, 1.45) with the AHF YFP HC Filterset F36-528 (excitation band pass [ex bp] 500/24 nm, beam splitter [bs] 520 nm and emission [em] bp 542/27 nm filters) and Nikon C-HGFIE Intensilight, by an Andor iXon3 885 electron-multiplying charge-coupled device (EMCCD) camera. Detection of single cells and data acquisition was performed using Fiji (Schindelin *et al.* 2012). Data was further analyzed on single cell level in R. *E. coli* MG1655 cells grown in the respective media were used for background subtraction on single cell level. To eliminate false detected cells (e.g. multiple cells detected as single cell), all cells which cell area/cell length ratio differs more than 20 % from the mean as well as cells with negative fluorescence signals after background subtraction were excluded from further analysis. 90 ±5 % of all detected cells were subjected to final analysis. Cells were grouped into 10 subgroups according to the cell length. Mean and corresponding standard deviation were calculated for each group and visualized with the resulting regression line.

Comparative fluorescence microscopy of Dam-YFP and SeqA-YFP normalized to HU-mCherry was performed as described above. In addition, mCherry signal was acquired using the AHF TxRed HC Filterset F36-504 (ex bp 562/40 nm, bs 593 nm and em bp 624/46 nm). Strains were cultivated and analyzed as tandem pairs for each condition. The exposure time was accurately set at each experiment to be confident that no saturation occurs and at the same time mCherry and YFP signals in both strains are detected. Data was analyzed as described before. The mean single cell YFP/mCherry ratio was used to compare the different strains and generate finally a SeqA/Dam ratio for each condition.

Microfluidic sample management, imaging conditions and data evaluation

The preparation and operation of the microfluidic devices used were performed as described in (Ullman *et al.* 2013). The trap depth used was 800 nm. All microscopy experiments were performed using an inverted microscope (Nikon Ti-E) with 100× oil-immersion objectives (either an Apo TIRF 1.49 na or a 100× Plan Apo λ 1.45 na). For phase-contrast imaging, a CFW-1312M (Scion), a DMK 23U274 (the Imaging Source) or an Infinity 2-5M (Lumenera) camera was used. Fluorescence and bright-field images were recorded on Andor Ixon EMCCD cameras. The Andor cameras were equipped with an additional 2× (Diagnostic instruments DD20NLT) or 2.5× lens (Nikon Instruments).

Imaging: phase-contrast images were acquired with a 125 ms exposure. For fluorescence imaging, a 514 nm laser (Coherent Genesis CX STM) was used.

The microscope was controlled using µ-Manager (Edelstein *et al.* 2014), and automated acquisitions were performed using in-house micro-manager plugin. Time-lapsed acquisitions were performed in parallel at multiple microfluidic trap regions, one of which was not exposed to laser. The duration of the acquisition varied from 2–24 hr. In all cases, cells were grown in the microfluidic devices for at least 24 hr prior to imaging to ensure steady-state exponential growth before the start of image acquisition. The temperature of the microfluidic device was maintained using a cage incubator (either OKO lab or Haison) encapsulating the microscope stage.

A custom-written, fully automated analysis pipeline written in MATLAB was used to analyze the time-lapsed microscopy data. Cells in each phase-contrast image were segmented using the method described in (Sadanandan *et al.* 2016). An active contour model based on (Sliusarenko *et al.* 2011) was developed, and a contour was computed for each segmented object. Cells were tracked between frames using the method described in (Magnusson *et al.* 2015). The determination of length, areas, volumes, and widths was based on the contour model as in (Sliusarenko *et al.* 2011).

FRAP – Fluorescence Recovery after Photobleaching

Photobleaching was performed in cells with two distinct SeqA-YFP cluster. One of the clusters was focused with an argon ion laser and bleached by an exposure of ≈25 ms. Images were acquired before and after photobleaching as described before. Measurements of fluorescence in regions of interest as well as calculation of the half-time was performed with a custom-written, fully automated analysis pipeline written in MATLAB.

Western Blotting

Cells of 25 ml culture were harvested in early exponential phase (OD = 0.15), resuspend in 800 µl TE supplemented with 200 µl 5 x loading dye and boiled for 10 min at 95 °C. For normalization cell numbers were determined by Neubauer cell counting chamber. 20 µl of each sample were run on a 15 % acrylamide gel chamber for 2.5 hours at 120 V and transferred by wet blot technique (Hofer® SE300 miniVE Integrated Vertical Electrophoresis and Blotting Unit) for 1 hour at 25 V to a PVDF membrane (ThermoFischer Scientific). Primary anti-SeqA antibody (1:5,000)(kind gift of Kirsten Skarstad) and secondary anti-rabbit IgG HRP-linked Antibody (1:10,000)(ThermoFischer Scientific; Catalog#: 32460) were used for detection. Bio-Rad ChemiDoc™ MP System was applied for signal detection using SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFischer Scientific). Data extraction was performed using the Bio-Rad Image Lab Software. Signal intensity was normalized according to the number of cells and finally means and corresponding standard deviations were calculated.

Comparative Dam and SeqA quantification was performed using an anti-GFP IgG HRP-linked Antibody (1:2,500)(ThermoFischer Scientific; Catalog#: A10260). Growth conditions and analysis were performed as described before, however only in LB Dam-YFP could be detected above background noise. All Western Blot experiments were carried out as technical and biological replicates.

Flow cytometry and cell cycle analysis

Cell cycle analysis of *E. coli* MG1655 was performed as described previously (Waldminghaus *et al.* 2012). SeqA-YFP Fusion was analyzed for functionality by testing strains for synchronous replication via flow cytometry by rifampicin, cephalexin runout experiments as described previously (Milbredt *et al.* 2016). Additionally, SeqA-YFP fusions were verified by detecting distinct foci during fluorescence microscopy and respective bands on western blots.

Acknowledgements

We thank all members of the Waldminghaus and Elf lab for help and fruitful discussions. We are grateful to Alexander Böhm† and Miroslav Radman for providing strains and/or plasmids and we thank the Flow Cytometry and Genomics Core Facility (ZTI, Marburg) for providing respective devices. This work was supported within the LOEWE program of the State of Hesse and a grant of the Deutsche Forschungsgemeinschaft (Grant No. WA 2713/4-1).

References

Babic A, Lindner AB, Vulic M, Stewart EJ, Radman M (2008) Direct visualization of horizontal gene transfer. Science 319 (5869):1533-1536. doi:10.1126/science.1153498

Bach T, Krekling MA, Skarstad K (2003) Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. EMBO J 22 (2):315-323. doi:10.1093/emboj/cdg020

Boye E (1991) The hemimethylated replication origin of *Escherichia coli* can be initiated *in vitro*. J Bacteriol 173 (14):4537-4539

Boye E, Lobner-Olesen A (1990) The role of *dam* methyltransferase in the control of DNA replication in *E. coli*. Cell 62 (5):981-989

Boye E, Marinus MG, Lobner-Olesen A (1992) Quantitation of Dam methyltransferase in *Escherichia coli*. J Bacteriol 174 (5):1682-1685

Brendler T, Sawitzke J, Sergueev K, Austin S (2000) A case for sliding SeqA tracts at anchored replication forks during *Escherichia coli* chromosome replication and segregation. EMBO J 19 (22):6249-6258. doi:10.1093/emboj/19.22.6249

Campbell JL, Kleckner N (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from Dam methyltransferase following the passage of the chromosomal replication fork. Cell 62 (5):967-979

Chung YS, Brendler T, Austin S, Guarne A (2009) Structural insights into the cooperative binding of SeqA to a tandem GATC repeat. Nucleic Acids Res 37 (10):3143-3152. doi:10.1093/nar/gkp151

Clark DJ, Maaløe O (1967) DNA replication and the division cycle in *Escherichia coli*. J Mol Biol 23 (1):99-112. doi:10.1016/s0022-2836(67)80070-6

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97 (12):6640-6645. doi:10.1073/pnas.120163297

De Souza Silva O, Blokesch M (2010) Genetic manipulation of *Vibrio cholerae* by combining natural transformation with FLP recombination. Plasmid 64 (3):186-195. doi:10.1016/j.plasmid.2010.08.001

Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N (2014) Advanced methods of microscope control using μManager software. J Biol Methods 1 (2). doi:10.14440/jbm.2014.36

Geier GE, Modrich P (1979) Recognition sequence of the Dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease. J Biol Chem 254 (4):1408-1413

Ghosh SK, Hajra S, Paek A, Jayaram M (2006) Mechanisms for chromosome and plasmid segregation. Annu Rev Biochem 75:211-241. doi:10.1146/annurev.biochem.75.101304.124037

Guarne A, Brendler T, Zhao Q, Ghirlando R, Austin S, Yang W (2005) Crystal structure of a SeqA-N filament: implications for DNA replication and chromosome organization. EMBO J 24 (8):1502-1511. doi:10.1038/sj.emboj.7600634

Guarne A, Zhao Q, Ghirlando R, Yang W (2002) Insights into negative modulation of *E. coli* replication initiation from the structure of SeqA-hemimethylated DNA complex. Nat Struct Biol 9 (11):839-843. doi:10.1038/nsb857

Han JS, Kang S, Kim SH, Ko MJ, Hwang DS (2004) Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. J Biol Chem 279 (29):30236-30243. doi:10.1074/jbc.M402612200

Helgesen E, Fossum-Raunehaug S, Saetre F, Schink KO, Skarstad K (2015) Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. Nucleic Acids Res 43 (5):2730-2743. doi:10.1093/nar/gkv146

Herman GE, Modrich P (1981) *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. J Bacteriol 145 (1):644-646

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. Cell 121 (3):349-361. doi:10.1016/j.cell.2005.02.021

Jensen KF (1993) The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. J Bacteriol 175 (11):3401-3407

Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D (2013) Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. PLoS Genet 9 (8):e1003673. doi:10.1371/journal.pgen.1003673

Kang S, Lee H, Han JS, Hwang DS (1999) Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. J Biol Chem 274 (17):11463-11468

Katayama T, Akimitsu N, Mizushima T, Miki T, Sekimizu K (1997) Overinitiation of chromosome replication in the *Escherichia coli dnaAcos* mutant depends on activation of *oriC* function by the *dam* gene product. Mol Microbiol 25 (4):661-670

Kuwada NJ, Traxler B, Wiggins PA (2015) Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. Mol Microbiol 95 (1):64-79. doi:10.1111/mmi.12841

Kuzminov A (2014) The precarious prokaryotic chromosome. J Bacteriol 196 (10):1793-1806. doi:10.1128/JB.00022-14

Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157 (3):624-635. doi:10.1016/j.cell.2014.02.033

Lobner-Olesen A, Kuempel PL (1992) Chromosome partitioning in *Escherichia coli*. J Bacteriol 174 (24):7883-7889

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in *E. coli*. Cell 77 (3):413-426

Magnusson KE, Jalden J, Gilbert PM, Blau HM (2015) Global linking of cell tracks using the Viterbi algorithm. IEEE Trans Med Imaging 34 (4):911-929. doi:10.1109/TMI.2014.2370951

Marinus MG (1987) DNA methylation in *Escherichia coli*. Annu Rev Genet 21:113-131. doi:10.1146/annurev.ge.21.120187.000553

Marinus MG, Lobner-Olesen A (2014) DNA Methylation. EcoSal Plus 6 (1). doi:10.1128/ecosalplus.ESP-0003-2013

Marinus MG, Morris NR (1973) Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. J Bacteriol 114 (3):1143-1150

Martina MA, Correa EM, Argarana CE, Barra JL (2012) *Escherichia coli* frameshift mutation rate depends on the chromosomal context but not on the GATC content near the mutation site. PLoS One 7 (3):e33701. doi:10.1371/journal.pone.0033701

McIntosh JR, Grishchuk EL, West RR (2002) Chromosome-microtubule interactions during mitosis. Annu Rev Cell Dev Biol 18:193-219. doi:10.1146/annurev.cellbio.18.032002.132412

Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, Boccard F, Espeli O (2008) The MatP/*matS* site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. Cell 135 (3):475-485. doi:10.1016/j.cell.2008.08.031

Milbredt S, Farmani N, Sobetzko P, Waldminghaus T (2016) DNA Replication in Engineered *Escherichia coli* Genomes with Extra Replication Origins. ACS Synth Biol 5 (10):1167-1176. doi:10.1021/acssynbio.6b00064

Norris V, Fralick J, Danchin A (2000) A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. Mol Microbiol 37 (4):696-702

Ogden GB, Pratt MJ, Schaechter M (1988) The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. Cell 54 (1):127-135

Onogi T, Niki H, Yamazoe M, Hiraga S (1999) The assembly and migration of SeqA-Gfp fusion in living cells of *Escherichia coli*. Mol Microbiol 31 (6):1775-1782

Sadanandan SK, Baltekin O, Magnusson KEG, Boucharin A, Ranefall P, Jalden J, Elf J, Wahlby C (2016) Segmentation and Track-Analysis in Time-Lapse Imaging of Bacteria. IEEE Journal of Selected Topics in Signal Processing 10 (1):174-184. doi:10.1109/jstsp.2015.2491304

Saint-Dic D, Kehrl J, Frushour B, Kahng LS (2008) Excess SeqA leads to replication arrest and a cell division defect in *Vibrio cholerae*. J Bacteriol 190 (17):5870-5878. doi:10.1128/JB.00479-08

Sanchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, Grainger DC (2010) Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12. MBio 1 (1). doi:10.1128/mBio.00012-10

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9 (7):676-682. doi:10.1038/nmeth.2019

Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M (2016) The quantitative and condition-dependent *Escherichia coli* proteome. Nat Biotechnol 34 (1):104-110. doi:10.1038/nbt.3418

Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N (1995) *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell 82 (6):927-936

Sliusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Mol Microbiol 80 (3):612-627. doi:10.1111/j.1365-2958.2011.07579.x

Sobetzko P, Jelonek L, Strickert M, Han W, Goesmann A, Waldminghaus T (2016) DistAMo: A Web-Based Tool to Characterize DNA-Motif Distribution on Bacterial Chromosomes. Front Microbiol 7:283. doi:10.3389/fmicb.2016.00283

Stokke C, Flatten I, Skarstad K (2012) An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. PLoS One 7 (2):e30981. doi:10.1371/journal.pone.0030981

Stokke C, Waldminghaus T, Skarstad K (2011) Replication patterns and organization of replication forks in *Vibrio cholerae*. Microbiology 157 (Pt 3):695-708. doi:10.1099/mic.0.045112-0

Szyf M, Avraham-Haetzni K, Reifman A, Shlomai J, Kaplan F, Oppenheim A, Razin A (1984) DNA methylation pattern is determined by the intracellular level of the methylase. Proc Natl Acad Sci U S A 81 (11):3278-3282

Taghbalout A, Landoulsi A, Kern R, Yamazoe M, Hiraga S, Holland B, Kohiyama M, Malki A (2000) Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E. coli in vitro*. Genes Cells 5 (11):873-884

Ullman G, Wallden M, Marklund EG, Mahmutovic A, Razinkov I, Elf J (2013) High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. Philos Trans R Soc Lond B Biol Sci 368 (1611):20120025. doi:10.1098/rstb.2012.0025

Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A, Jeltsch A (2002) The *Escherichia coli* Dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol 319 (5):1085-1096. doi:10.1016/S0022-2836(02)00371-6

von Freiesleben U, Rasmussen KV, Schaechter M (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. Mol Microbiol 14 (4):763-772

Waldminghaus T, Skarstad K (2009) The *Escherichia coli* SeqA protein. Plasmid 61 (3):141-150. doi:10.1016/j.plasmid.2009.02.004

Waldminghaus T, Skarstad K (2010) ChIP on Chip: surprising results are often artifacts. BMC Genomics 11:414. doi:10.1186/1471-2164-11-414

Waldminghaus T, Weigel C, Skarstad K (2012) Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. Nucleic Acids Res 40 (12):5465-5476. doi:10.1093/nar/gks187

Wallden M, Fange D, Lundius EG, Baltekin O, Elf J (2016) The Synchronization of Replication and Division Cycles in Individual *E. coli* Cells. Cell 166 (3):729-739. doi:10.1016/j.cell.2016.06.052

Yang H, Wolff E, Kim M, Diep A, Miller JH (2004) Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. Mol Microbiol 53 (1):283-295. doi:10.1111/j.1365-2958.2004.04125.x

Youngren B, Nielsen HJ, Jun S, Austin S (2014) The multifork *Escherichia coli* chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. Genes Dev 28 (1):71-84. doi:10.1101/gad.231050.113

Zaritsky A, Woldringh CL (2015) Chromosome replication, cell growth, division and shape: a personal perspective. Front Microbiol 6:756. doi:10.3389/fmicb.2015.00756

Supporting Information

SeqA complexes in *Escherichia coli* exchange proteins rapidly and vary depending on replication patterns

Daniel Schindler, Ebba Gregorsson Lundius*, Matthias Bruhn, Sean Murray, Johan Elf, Torsten Waldminghaus*

** equal contribution*

Figure S1

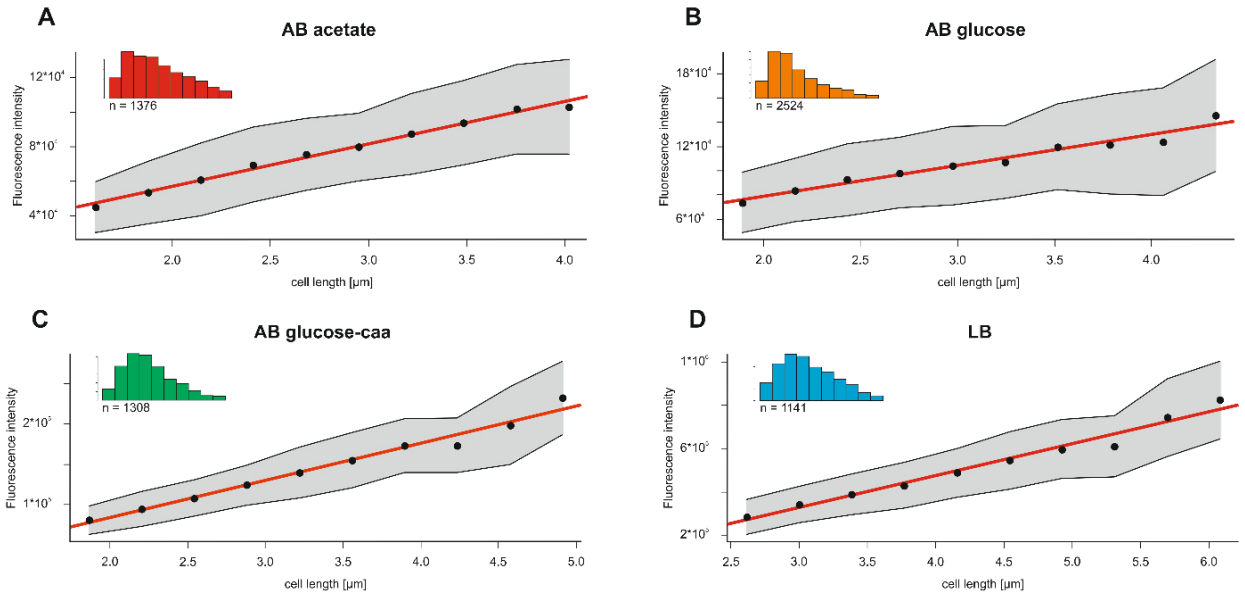


Figure S1: Biological replicate of experiments shown in figure 1. SeqA-YFP intensities of exponential growing cultures (DS116) were measured by fluorescence microscopy and quantified as described in the Material and Methods section. **A to D** Four different growth media were used with increasing growth rate from AB acetate to LB. Intensity values of individual cells where grouped according, to cell size as proxy for cell cycle stage (black dots) with the respective linear regression (red line) and the standard deviation (grey area). Each medium is color coded by the histogram which gives the cell size distribution with numbers of analyzed cells below the histogram.

Figure S2

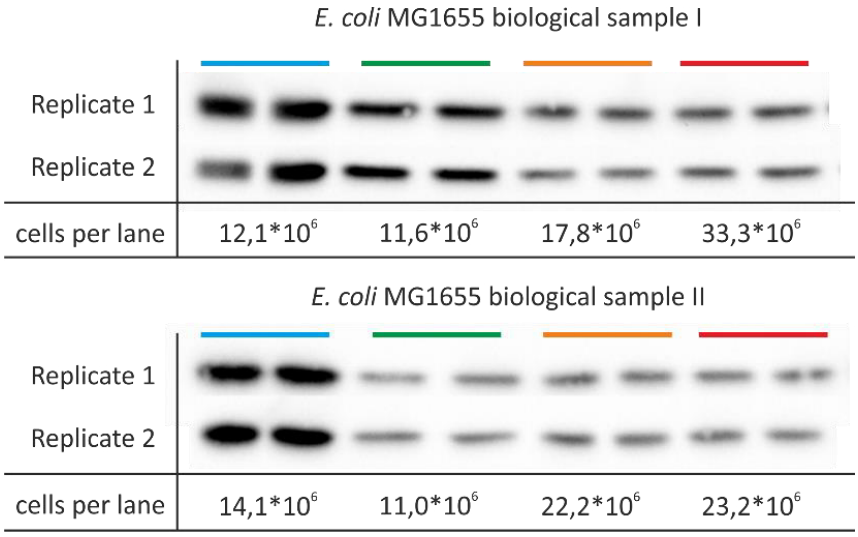


Figure S2: SeqA quantitative western blotting of exponential growing *E. coli* cells in four different media. Figure shows western blots used for quantifications in figure 2. Two biological replicates were performed in two technical replicates as indicated. Cell numbers were determined by counting in Neubauer chamber. Different growth rates are indicated color-coded as in figure S1/S3. For details see the Material and Methods section.

Figure S3

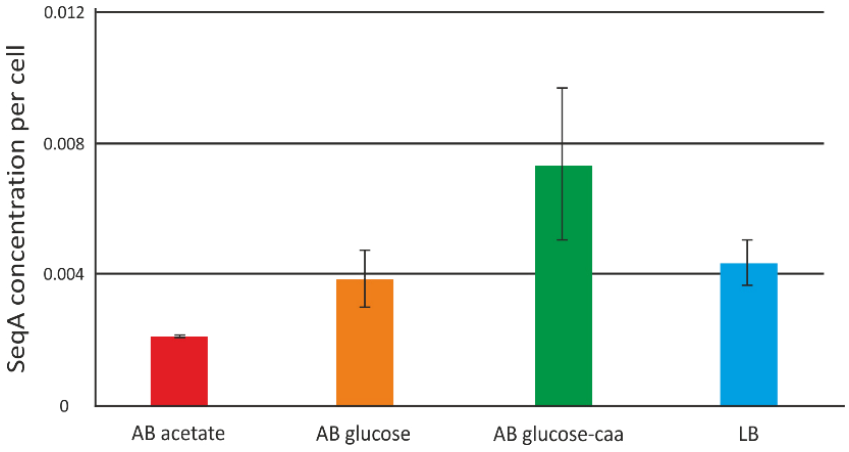


Figure S3: SeqA concentration within cells of strain DS126. Strain DS126 carries *seqA-yfp* and *hupB-mCherry* fusions at the endogenous locus respectively. The fluorescence signals were acquired by fluorescence microscopy and data was extracted with Fiji. Data was used to calculate the concentration of SeqA relative to HU at single cell level and values where divided by the cell area to obtain SeqA concentration within each cell. Shown is the mean of all cells with the respective standard deviation. Results are similar to SeqA concentration determined by flow cytometry (Fig. 2). See Material and Methods for details.

Figure S4

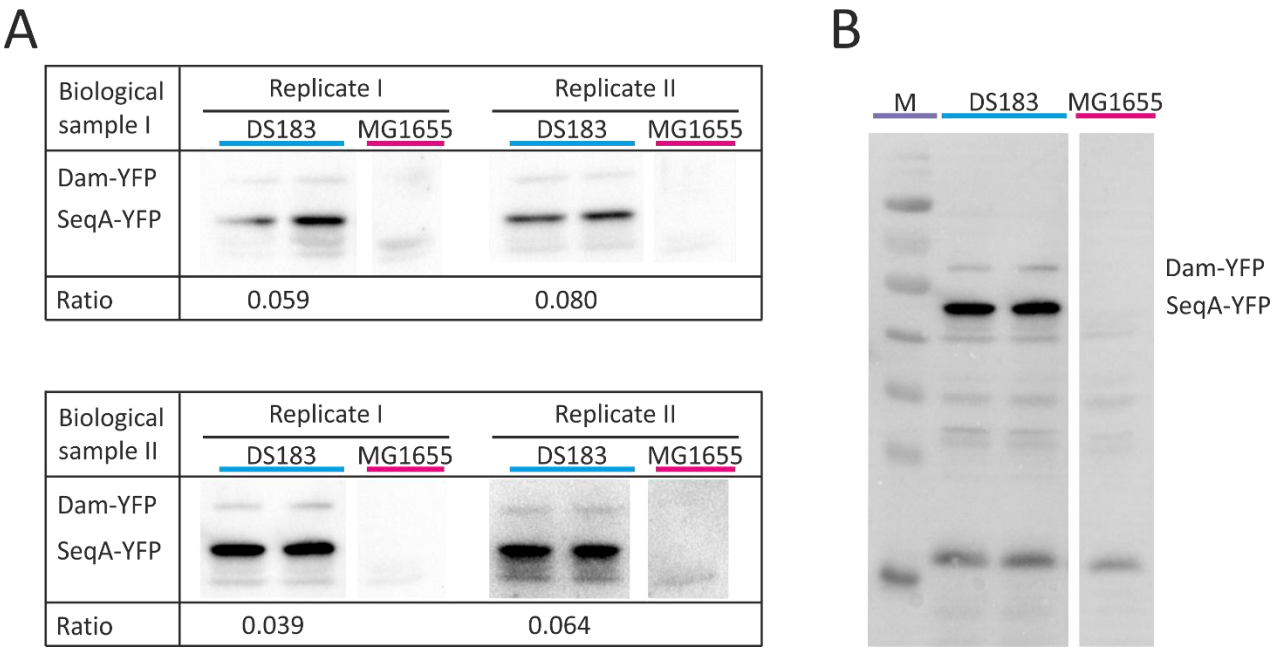


Figure S4: Determination of Dam-YFP/SeqA-YFP ratio by western Blotting. Strain DS183 and *E. coli* M1655 were grown exponentially, harvested and protein extracts were used for PAGE followed by Western Blotting and detection. See Materials and Methods for details. **A** shows the biological and technical replicates. *E. coli* MG1655 served as control and no respective bands were detected. Anti-GFP signals were detected and quantified to calculate indicated ratios for each experiment. **B** shows a respective western blot and *E. coli* MG1655 control clearly indicates that the analyzed bands are specific.

Figure S5

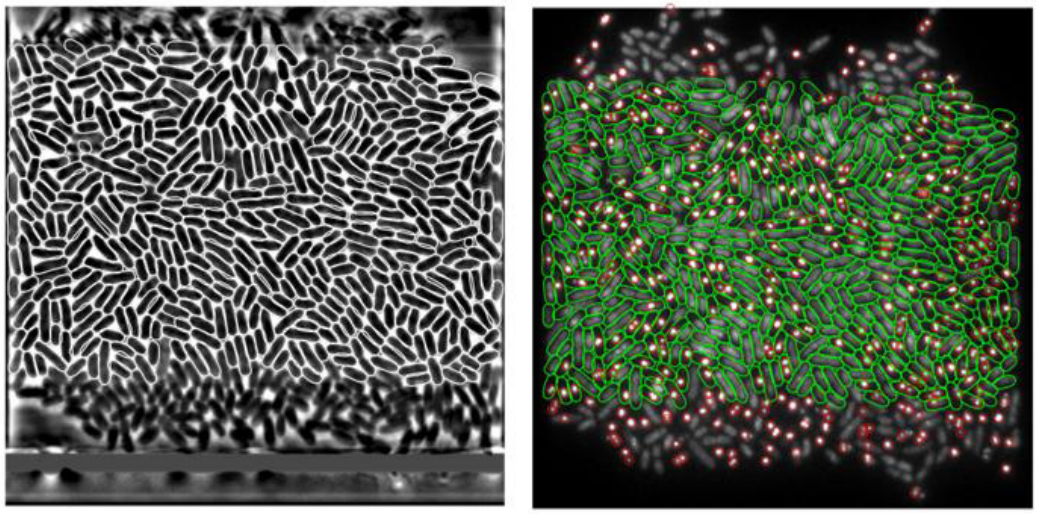


Figure S5: Microscopy and data analysis in microfluidic devices. Phase contrast, brightfield and fluorescence images of the same trap in the microfluidic chip were acquired with minimal time delays in between. The left panel shows an example of a phase contrast image with the contours of segmented cells in white. These contours are then registered onto the fluorescence image via co-registration with a brightfield image taken with the same camera as the fluorescence image. The right panel shows the co-registered cell contours (green) as well as identified fluorescent dots (red circles) in the corresponding fluorescence image.

Table S1: Cell cycle parameters were calculated as described (Waldminghaus *et al.* 2012).

Media	τ[*1]	F[*2]	a_i[*3]	a_t[*5]	C period[*4]
AB acetate	117 +/- 7.8	0.34 +/- 0.01	31	90	59
AB glucose	55 +/- 1.2	0.8 +/- 0.05	41	25	39
AB glucose-caa	32 +/- 1	0.31 +/- 0.02	8	31	55
LB	27 +/- 0.6	0.77 +/- 0.02	19	14	49

[*1] The generation time τ is the mean of three biological replicates.

[*2] F is the fraction of cells that had not initiated as determined by flow cytometric analysis of a rifampicin/cephalexin runout experiment.

[*3] The initiation age a_i was calculated with the formula $a_i = \tau - log(2 - F) * \tau$.

[*4] The C period was calculated by multiplication of relative C period values from Stokke *et al.* 2012 for the respective growth conditions with the generation time.

[*5] The termination age a_t was calculated from a_i, the C period and the number of generations spanned by the replication cycle.

Table 2: Strains used in this study.

Strain	Characteristics	Resistance	Reference
E. coli MG1655	*E. coli* wild type	-	(Blattner *et al.* 1997)
E. coli AB330	*cf*. DY330 (Yu *et al*. 2000), *lacZ*+, *gal*+	-	Alexander Böhm
E. coli seqA-yfp	E. coli *seqA-yfp*	chloramphenicol	(Babic *et al.* 2008)
DS116	*E. coli* MG1655 *seqA-yfp*	chloramphenicol	This study
DS126	*E. coli* TB28 *hupB-mCherry, seqA-yfp*	chloramphenicol	This study
DS179	*E. coli* MG1655 *dam-yfp*	chloramphenicol	This study
DS181	*E. coli* TB28 *dam-yfp, hupB-mCherry*	chloramphenicol	This study
DS183	*E. coli* MG1655 *dam-yfp, seqA-yfp*	chloramphenicol	This study
TB28 HU-mCherry	*E. coli hupB-mCherry*	-	(Jia *et al.* 2014)

Table 3: Plasmids used in this study.

Plasmid	Characteristics	Resistance	Reference
pBR-flp	FLP^+, λ $cI857^+$, λ p_R	ampicillin, tetracycline	(De Souza Silva and Blokesch 2010)
pSeqA-C	*seqA-yfp*-FRT-*cat*-FRT	ampicillin, chloramphenicol	(Babic *et al.* 2008)

Table 4: Oligonucleotides used in this study.

Name	Sequence (5´ -> 3´)
750	GGCGGCACACGTAAAAAGGTGGACGAACTGCTGGCTTTGTACAAACCAGGAGTCGTTTCACCCGCGAA AAAAGGCGGCGGCAGCGCTAGCAAAGG
751	CAGGCGGGCAAAATCAGCCGACAGAATTGAGGGGGCAATCAAATACTGTTTCATCCGCTT CTCCTTGA GAAAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGGCGCG

Supplementary references

Babic A, Lindner AB, Vulic M, Stewart EJ, Radman M (2008) Direct visualization of horizontal gene transfer. Science 319 (5869):1533-1536. doi:10.1126/science.1153498

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331):1453-1462

De Souza Silva O, Blokesch M (2010) Genetic manipulation of *Vibrio cholerae* by combining natural transformation with FLP recombination. Plasmid 64 (3):186-195. doi:10.1016/j.plasmid.2010.08.001

Jia S, Keilberg D, Hot E, Thanbichler M, Sogaard-Andersen L, Lenz P (2014) Effect of the Min system on timing of cell division in *Escherichia coli*. PLoS One 9 (8):e103863. doi:10.1371/journal.pone.0103863

Stokke C, Flatten I, Skarstad K (2012) An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. PLoS One 7 (2):e30981. doi:10.1371/journal.pone.0030981

Waldminghaus T, Weigel C, Skarstad K (2012) Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. Nucleic Acids Res 40 (12):5465-5476. doi:10.1093/nar/gks187

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A 97 (11):5978-5983. doi:10.1073/pnas.100127597

4 Diskussion

Anhand von Methoden der Synthetischen Biologie konnte 2010 erstmals ein Organismus generiert werden, dessen Genom auf einem synthetisch hergestellten Chromosom basiert (Gibson *et al.* 2010). In den letzten Jahren hat die Synthetische Biologie die moderne Molekularbiologie um viele neue Methoden bereichert (Montague *et al.* 2012; Padilla-Vaca *et al.* 2015; Breitling and Takano 2016; MacDonald and Deans 2016; Haellman and Fussenegger 2016). Wichtige Beispiele dafür sind DNA-Synthese und DNA-Assemblierungstechniken, die heute in einer nie dagewesenen Größenordnung möglich sind (Kosuri and Church 2014; Cobb *et al.* 2014; Chao *et al.* 2014; Ellis *et al.* 2011; Carr and Church 2009). Prestigeprojekte könnten antreibende Kräfte für technologische Innovation sein. Beispielsweise sind durch die Sequenzierung des humanen Genoms die Sequenzierungskosten so stark gesunken, dass diese Technik als Serviceleistung heute von fast jedem Labor in Anspruch genommen werden kann (Sboner *et al.* 2011; van Nimwegen *et al.* 2016; Service 2006; Mardis 2006). Insbesondere die Sequenzierung von mikrobiellen Genomen ist technisch und wirtschaftlich ohne Probleme umzusetzen, weshalb immer mehr mikrobielle Genome sequenziert und die generierten Sequenzen in Datenbanken hinterlegt werden(Cochrane *et al.* 2016; Land *et al.* 2015). Wissenschaftler sequenzieren nicht mehr nur einzelne Stämme, sondern auch Umweltproben und damit ganze Lebensgemeinschaften, denn ein Großteil der Mikroorganismen ist nicht unter Laborbedingungen zu kultivieren (Caporaso *et al.* 2012; Venter *et al.* 2004; Human Microbiome Project Consortium 2012; Wilkins *et al.* 2013; Amann *et al.* 1995). Folglich werden die Datenbanken nicht nur mit weiteren, komplett sequenzierten Genomen, sondern auch mit einer großen Anzahl von einzelnen proteinkodierenden Sequenzen gefüllt. In diesen Datenbanken schlummert ein enormes Potential für biotechnologische Anwendungen, denn heute ist es möglich große Gengruppen, die ganze Stoffwechselwege kodieren zu synthetisieren und heterolog zu exprimieren (Awan *et al.* 2016; Keasling 2012; Ro *et al.* 2006; Becker and Wittmann 2016). Dabei müssen nicht alle Gene aus einem Organismus stammen, sondern es können Biosynthesewege bestehend aus den Genen verschiedener Organismen erzeugt werden (Bloch and Schmidt-Dannert 2014; Nielsen 2011). Noch sind diese Biosynthesewege auf eine geringe Anzahl an Genen beschränkt (Bloch and Schmidt-Dannert 2014). Die Synthese von ganzen Chromosomen ist keine Science-Fiction mehr. In Zukunft sollte es möglich sein, Designerorganismen zu generieren, die auf Genen verschiedener Organismen basieren, um somit die optimalen Eigenschaften für beispielsweise eine spezifische biotechnologische Anwendung zu kombinieren. Wird es in Zukunft tatsächlich möglich sein solche Designerorganismen herzustellen und wo liegen die derzeitigen Herausforderungen zur Synthese von Chromosomen?

4.1 Was sind die Herausforderungen bei der Synthese von Chromosomen?

Neben den beachtlichen Kosten, die bei der Synthese ganzer Chromosomen anfallen, ist auch die Komponente der zeitlichen Umsetzung nicht zu unterschätzen. Der erste synthetisch hergestellte Organismus *Mycoplasma mycoides JCVI-syn1.0* hat etwa 40.000.000 US$ gekostet und eine Arbeitszeit von 200 Jahres-Vollzeitäquivalenten aufgebraucht (Sleator 2010; Pennisi 2010; Schindler and Waldminghaus 2015). Studien in diesem Größenmaßstab sind (noch) nicht von einzelnen Laboren zu bewältigen, sondern werden durch große Institutionen oder Kollaborationen durchgeführt. Was ist jedoch neben den wirtschaftlichen Faktoren zu berücksichtigen? Welche Voraussetzungen sind bereits gegeben und welche Herausforderungen müssen bewältigt werden, um solche Projekte umsetzen zu können? Die vorliegende Arbeit hat sich mit der Etablierung von synthetischen sekundären Chromosomen in *Escherichia coli* auseinandergesetzt (Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision; Schindler and Waldminghaus in preparation). Die synthetischen sekundären Chromosomen wurden genutzt, um eine effiziente, vergleichende Analyse von *Chromosome Maintenance* Systemen zu ermöglichen, wodurch in Zukunft generelle Regeln für das Design von synthetischen Chromosomen generiert werden können (Schindler and Waldminghaus in preparation). Im Rahmen der vorliegenden Arbeit sind Herausforderungen für das Design, die Synthese und den Transfer der synthetischen sekundären Chromosomen aufgetreten, die bewältigt werden konnten (Schindler and Waldminghaus in preparation).

4.1.1 Design von synthetischen Chromosomen

Das Design ist die Grundlage eines jeden synthetischen Chromosoms. Das Anwendungsgebiet oder eine Fragestellung sind stets die elementaren Vorgaben für das Design eines synthetischen Chromosoms. Eine Fragestellung in der Synthetischen Biologie ist beispielsweise wie das Design eines Minimalorganismus aussehen würde. Aus diesem Grund haben sich mehrere Studien damit beschäftigt die essentiellen Gene von Bakterien zu identifizieren und auf dem Wissen basierend das Design eines Minimalorganismus zu erschaffen (Gil *et al.* 2004; Jewett and Forster 2010; Xavier *et al.* 2014).

> "*Perfection is finally attained not when there is no longer anything to add but when there is no longer anything to take away.*" Antoine de Saint-Exupery (1900-1944)

Gene, deren Fehlen letale Folgen haben, sind essentiell. Beispielsweise kann das Gen des Initiatorproteins DnaA nicht in natürlich vorkommenden Bakterien deletiert werden, weshalb DnaA zwangsläufig als essentiell bezeichnet werden muss (Fuller *et al.* 1984; Tomizawa and Selzer 1979). Verschiedene Studien postulieren, dass „Leben" auf ein minimales Set von etwa 250 bis 400 essentiellen Genen reduziert werden kann (Hutchison *et al.* 1999; Glass *et al.* 2006; Koonin 2003, 2000; Juhas *et al.*

2014). Es wird sogar behauptet, dass bakterielles Leben mit einem Chromosom von etwa 113 kb Größe und nur 151 Genen möglich sein könnte (Forster and Church 2006). Ein bakterieller Minimalorganismus würde der Grundlagenforschung einen tiefen Einblick in die Funktionsweise des Lebens geben. Allerdings muss ein solcher Minimalorganismus für die biotechnologische Anwendungen um weitere Gene ergänzt werden, denn unter anderem sind die Gene der *DNA mismatch* Reparatur keine essentiellen Gene (Baba *et al.* 2006). Allerdings hat das Fehlen eines oder mehrerer an der DNA-Reparatur beteiligter Gene zur Folge, dass es zu einer genomischen Instabilität kommt. Die Mutationsrate wäre stark erhöht, was eine Verringerung der Fitness zur Folge hätte (Lenhart *et al.* 2016; Drake 1991; Sniegowski *et al.* 1997). Ein minimales Design muss folglich neben den essentiellen Genen um konditionale Gene erweitert werden, die die Stabilität des Designerorganismus garantieren. Selbst wenn ein einheitlicher Konsens über die Auswahl essentieller Gene herrschen würde, stellt sich jedoch immer noch die Frage, wie ein solcher Organismus konstruiert und hergestellt werden kann.

Minimalorganismen – Konstruktionsweisen und gegenwärtiger Forschungsstand

Es gibt zwei verschiedene Herangehensweisen einen Minimalorganismus zu generieren: das *top-down* oder das *bottom-up* Verfahren (Jewett and Forster 2010; Forster and Church 2006, 2007). Beim *top-down* Verfahren wird ein natürlich vorkommender Organismus sukzessive reduziert, wodurch lebensfähige Organismen als Intermediate entstehen, deren Chromosom weiter reduziert werden kann. Ist der Punkt erreich an dem keine weitere Reduktion des Chromosoms mehr möglich ist, sind ausschließlich die essentiellen Gene übrig und der Minimalorganismus wurde konstruiert. Bei dieser Vorgehensweise ist die Abfolge der Gene im finalen reduzierten Chromosom die gleiche wie im ursprünglichen Chromosom. Im Gegensatz zum *top-down* Verfahren wird beim *bottom-up* Verfahren ausschließlich von erworbenem Wissen über essentielle Gene ausgegangen und diese werden zu einem minimalen Chromosom kombiniert. Das *bottom-up* Verfahren stellt die riskantere Herangehensweise dar und ist bisher nicht praktikabel, da essentielle Gene fehlen können und keine lebensfähigen Intermediate wie beim *top-down* Verfahren entstehen. Es ermöglicht die Abfolge der Gene unabhängig von der natürlich vorkommenden Anordnung von Genen zu konstruieren. Somit können durch das *bottom-up* und das *top-down* verschiedene Chromosomen als Endprodukte entstehen. Das Fehlen eines vollständig auf einem *bottom-up* Design basierenden Minimalorganismus und die gescheiterten Versuche einen minimalen *Mycoplasma mycoides* durch das *bottom-up* Verfahren im Rahmen des *Mycoplasma mycoides* JCVI-syn3.0 Projekts herzustellen, zeigen, dass es (noch) nicht möglich ist durch das *bottom-up* Verfahren einen Minimalorganismus zu konstruieren (Hutchison *et al.* 2016).

Vertreter der Gattung *Mycoplasma* haben sich als Modellsystem für ein minimales Design etabliert, der synthetisch hergestellte *Mycoplasma* JCVI-syn3.0 ist der Organismus, der einem Minimalorganismus bisher am nächsten kommt (Hutchison *et al.* 2016; Callaway 2016; Sleator 2016; Service 2016). JCVI-syn3.0 basiert auf JCVI-syn1.0, dessen Chromosom über mehrere „Design – Synthese – Test-Zyklen" um etwa 50 % der Gene reduziert wurde, wodurch die Größe des Chromosoms ebenfalls halbiert werden konnte. Jede weitere Verringerung des Chromosoms scheint starke Auswirkungen auf die Fitness zu haben (Gibson 2014; Hutchison *et al.* 2016). Höchst interessant ist, dass die Funktion von 149 der 473 Gene von JCVI-syn3.0 gänzlich unbekannt ist. Mit JCVI-3.0 besteht die Möglichkeit „Perfektion" (vgl. Zitat Antoine de Saint-Exupery) in Hinblick auf einen auf *Mycoplasma* basierenden Minimalorganismus zu erlangen, wobei weitere Reduktionen zu Lasten der Fitness gehen würden (Hutchison *et al.* 2016). Aufgrund dieser Studie könnte es in Zukunft möglich sein, minimale Designerorganismen mit dem *bottom-up* basierten Verfahren zu erzeugen. Es ist im Bereich des Möglichen sich hierbei nicht auf die Gene eines einzelnen Organismus zu beschränken, sondern eine Kombination von Eigenschaften verschiedener Organismen zu nutzen.

Möglichkeiten und Folgen der Rekodierung ganzer Chromosomen

Ein weiteres Problem beim Design synthetischer Chromosomen ist, dass Chromosomen mehr als die Abfolge ausgewählter Gene sind und ihr Design an die entsprechende Fragestellung bzw. Anwendung angepasst sein muss. Die Komplexität und die Vielschichtigkeit selbst von minimalen Organismen verhindert ein *de novo* Design von Hand. Selbst eine Rekodierung einer vorhandenen Genomsequenz benötigt umfassende Computerprogramme, um ein Design zu generieren. Hier ist das Projekt zur Konstruktion eines um sieben Codons reduzierten *E. coli* Stammes (*rE. coli*-57) zu nennen (Ostrov *et al.* 2016). Das Design wurde auf Grundlage der DNA-Sequenz des *E. coli* Stammes MDS42, der bereits eine Genomreduktion um 14,3 % gegenüber *E. coli* MG1655 aufweist, durch ein Computerprogramm generiert (Ostrov *et al.* 2016; Posfai *et al.* 2006). Dieses Programm beschränkt sich auf sieben Schritte, um das Design zu generieren. Diese Schritte beinhalten unter anderem das Ersetzen der sieben Codons durch alternative Codons, die die gleiche Aminosäure kodieren aber von einer anderen tRNA erkannt werden. Des Weiteren rekodiert das Computerprogramm die DNA-Sequenz, dahingehend, dass Homopolymere einzelner DNA-Basen reduziert werden und die Erkennungssequenzen von drei Typ IIS-Endonukleasen (AarI, BsaI und BsmBI) entfernt werden, um eine optimale DNA-Synthese und Assemblierung zu ermöglichen. Die synthetisierten DNA-Fragmente werden zuerst zu etwa 50 kb großen Teilsegmenten assembliert und anschließend *in vivo* auf Funktionalität getestet.

Die Studie zeigt dabei, dass einige der synthetisierten und assemblierten Teilsegmente extensive Optimierungsschritte benötigen, um funktionell zu sein. Das Design ist folglich nicht fehlerfrei und die Konzeption eines solchen Organismus durch ein Computerprogramm bedeutet nicht, dass dieser Organismus final lebensfähig ist, obwohl keines der im Chromosom kodierten Gene entfernt wurde. Vergleichende Analysen der DNA-Sequenzen des Ausgangsstammes und des entworfenen *rE. coli*-57 im Rahmen dieser Arbeit belegen zudem, dass bei dem Design die Ebene der *Chromosome Maintenance* Systeme außer Acht gelassen wurde. *rE. coli*-57 besitzt durch die automatische Sequenzveränderung 18,5 % (157) weniger *Chi*- und 7,7 % (1302) mehr GATC-Sequenzen als der Ursprungsstamm *E. coli* MDS42, die jedoch für das *Chromosome Maintenance* wesentlich sind. In der vorliegenden Arbeit konnte gezeigt werden, dass *Chromosome Maintenance* Systeme essentielle Funktionen für die Genomstabilität besitzen und somit für das Design von synthetischen Chromosomen berücksichtigt werden müssen (Schindler and Waldminghaus in preparation).

Ausblick auf zukünftige mögliche Chromosomen Designs

Vorausblickend wäre es denkbar in Zukunft einen „Genom Design Standard“ zu etablieren, ähnlich der Standards die für die Assemblierung von Transkriptionseinheiten vorgeschlagen wurden (Weber *et al.* 2011; Rokke *et al.* 2014; Moore *et al.* 2016; Agmon *et al.* 2015). Ein solcher Standard sollte voraussetzen, dass das Design einen Replikationsursprung und eine Terminusregion besitzt. Die Transkriptionseinheiten der Gene sollten auf dem Chromosom ausgehend vom Replikationsursprung zur Terminusregion orientiert werden. Durch eine solche Anordnung wäre eine Kollision von DNA- und RNA-Polymerasen minimiert und DNA-Replikation und Transkription wären bestmöglich innerhalb der Zelle koordiniert (Liu and Alberts 1995). Zusätzlich wäre es innovativ sämtliche Transkriptionseinheiten von *Chromosom Maintenance* Sequenzen zu bereinigen und diese zwischen den Transkriptionseinheiten gezielt und systematisch zu organisieren. Dadurch würden beispielsweise keine DNA-Bindeproteine die Transkription der einzelnen Gene beeinträchtigen. Zudem würde es definierte Bereiche in einem Chromosom geben durch die das Chromosom im dreidimensionalen Raum organisiert und strukturiert werden könnte.

Bisher wurde in der Diskussion der vorliegenden Arbeit, davon ausgegangen, für die Organisation der Gene ein singuläres Chromosom zu verwenden. Das ist aber nicht zwingend notwendig. Die meisten Bakterien haben nur ein einzelnes Chromosom, aber ist gibt auch Arten die mehrere Chromosomen besitzen (Okada *et al.* 2005; Mackenzie *et al.* 1999). Es wäre möglich in Zukunft Designs mit mehreren Chromosomen in einer bakteriellen Zelle zu etablieren (Schindler and Waldminghaus 2015; Liang *et al.* 2013; Milbredt *et al.* 2016). Dabei wäre es denkbar das *E. coli* Chromosom auf mehrere Chromosomen

mit einer Größe von etwa 200 kb aufzuteilen, die daraus resultierenden 23 Chromosomen wären mit molekularbiologischen Methoden *in vitro* zu manipulieren und könnten anschließend wieder in *E. coli* Zellen eingebracht werden. Mit dem 4,6 mb großen *E. coli* Chromosom wäre dies nicht möglich. Bakterien mit zwei oder mehr Chromosomen sind eher die Ausnahme, wohingegen die Organisation des Genoms auf mehrere Chromosomen in Eukaryoten eher die Regel ist (Egan *et al.* 2005). Die Anzahl an Chromosomen bei Eukaryoten schwankt von eins bis zu 16.000 Chromosomen im haploiden Chromosomensatz (Crosland and Crozier 1986; Swart *et al.* 2013). Interessanterweise variiert die Größe der etwa 16.000 Chromosomen von *Oxytricha trifallax* von 469 bp bis 66 kb, wobei 90 % der Chromosomen ein proteinkodierendes Gen und nur 10 % der Chromosomen zwei bis maximal acht proteinkodierende Gene aufweisen (Swart *et al.* 2013). An dieser Stelle sollte ganz klar erwähnt werden, dass *Oxytricha trifallax* eine Ausnahme darstellt, da der Chromosomensatz zusätzlich polyploid ist und die Chromosomen durchschnittlich mit etwa 2.000 Kopien vorliegen. Das Konstruieren eines *E. coli* Stammes mit 4.288 Chromosomen, bei dem jedes Chromosom ein einzelnes proteinkodierendes Gen aufweist, wird auch in Zukunft utopisch sein (Blattner *et al.* 1997).

Ist ein Design für ein synthetisches Chromosom entstanden und soll dieses hergestellt werden, muss dieses entsprechend der Designvorgaben umgesetzt werden. Hier stellt sich zuerst die Frage, ob die gewünschten Änderungen noch durch Änderungen eines vorhandenen Chromosoms mittels Methoden des *Genome Engineering* umgesetzt werden können oder ob eine Neusynthese des Chromosoms notwendig ist (Schindler and Waldminghaus 2015). Doch wie kann die Herausforderung, ein ganzes Chromosom zu synthetisieren, bewerkstelligt werden?

4.1.2 Assemblierung von synthetischen Chromosomen

Durch die Synthetische Biologie ist eine Vielzahl von DNA-Assemblierungsmethoden etabliert worden, doch nur wenige eignen sich, um ganze Chromosomen zu assemblieren (Chao *et al.* 2014; Cobb *et al.* 2014; Ellis *et al.* 2011). Wenn ein entsprechendes Sequenzdesign generiert wurde, wird dieses synthetisiert. Hierbei muss bereits im Design entsprechend bedacht werden, welche DNA-Assemblierungsmethode(n) verwendet werden soll(en). DNA-Fragmente können nur bis zu einer Länge von mehreren hundert Basenpaaren synthetisiert werden, dadurch ist es technisch nicht möglich ein Chromosom als ein zusammenhängendes DNA-Fragment zu synthetisieren. Stattdessen werden synthetische Chromosomen basierend auf vielen kurzen Oligonukleotiden, anhand derer längere doppelsträngige DNA-Fragmente generiert werden, hierarchisch zu einem Chromosom assembliert (Gibson 2012; Gibson *et al.* 2010; Gibson *et al.* 2008a). Es konnte gezeigt werden, dass es effizient ist

eine Kombination aus *in vitro* und *in vivo* Methoden für die DNA-Assemblierung von synthetischen Chromosomen zu verwenden (Zhou *et al.* 2016; Gibson *et al.* 2008a).

Vor- und Nachteile der in vitro DNA-Assemblierung synthetischer Chromosomen

Der Vorteil von *in vitro* DNA-Assemblierungsmethoden wie beispielsweise bei der für JCVI-syn1.0 verwendeten *Gibson Assembly* oder dem in der vorliegenden Arbeit verwendeten MoClo-System ist, dass diese deutlich zeiteffizienter sind als *in vivo* DNA-Assemblierungen in Hefe. Das MoClo-System ist der *Gibson Assembly* aufgrund der geringen Anzahl an Arbeitsschritten überlegen. Allerdings haben Typ IIS-basierte DNA-Assemblierungssysteme spezifische Anforderung an das Design, denn es dürfen keine zusätzlichen Erkennungssequenzen vorliegen bzw. vorliegende Erkennungssequenzen müssen eliminiert werden. Für die *Gibson Assembly* ist es ebenfalls wichtig, dass assemblierte DNA-Fragmente aus den Vektoren in denen sie assembliert wurden, durch eine selten schneidende Endonuklease herausgeschnitten werden können. Dabei ist zu beachten, dass diese Endonuklease nicht innerhalb der assemblierten DNA-Fragmente schneiden darf. Zudem hat die *Gibson Assembly* den Nachteil, dass homologe Bereiche an den Enden der zu assemblierenden DNA-Sequenzen vorhanden sein müssen. In der vorliegenden Arbeit konnte gezeigt werden, dass durch das MoClo-System effizient synthetische sekundäre Chromosomen bis zu einer Größe von 100 kb assembliert werden können (Schindler and Waldminghaus in preparation). Es konnte gezeigt werden, dass es möglich ist das ganze 583 kb große *Mycoplasma genitalium* Genom *in vitro* anhand der *Gibson Assembly* zu assemblieren (Gibson *et al.* 2009). Allerdings sind i*n vitro* Methoden mit zunehmender Größe der DNA-Assemblierung nicht mehr praktikabel, was unter anderem auf Scherkräfte zurückzuführen ist, die die DNA fragmentieren können. Als Lösung für dieses Problem hat sich die *in vivo* DNA-Assemblierung herausgestellt.

Vor- und Nachteile der in vivo DNA-Assemblierung synthetischer Chromosomen

Die *in vivo* DNA-Assemblierung wird häufig in einem heterologen System durchgeführt. Dies hat den Vorteil, dass dadurch mögliche toxische Effekte durch eine erhöhte Anzahl an Genkopien und die damit veränderte Expression unterbunden werden. Die Hefe *Saccharomyces cerevisiae* hat sich derzeit als System der Wahl etabliert (Benders *et al.* 2010; Tagwerker *et al.* 2012; Karas *et al.* 2012; Lartigue *et al.* 2009). Studien konnten eindrucksvoll zeigen, dass in der Hefe synthetische Chromosomen assembliert werden können oder aber ganze, intakte, mikrobielle Chromosomen durch Zellfusion aufgenommen werden können (Gibson *et al.* 2010; Hutchison *et al.* 2016; Karas *et al.* 2013a; Karas *et al.* 2014). In Hefe ist es anschließend möglich die aufgenommenen DNAs durch die Vielzahl an etablierten molekularbiologischen Methoden effizient zu modifizieren (Duina *et al.* 2014; Tsarmpopoulos *et al.* 2016;

Chandran *et al.* 2014; Noskov *et al.* 2010; Lartigue *et al.* 2009). Zudem ist es möglich *S. cerevisiae* ohne großen technischen und wirtschaftlichen Aufwand zu kultiviert.

In Hefe können große DNAs und ganze Chromosomen als episomale Vektoren stabil aufrechterhalten werden, was ein klarer Vorteil gegenüber einer *in vivo* DNA-Assemblierung, wie sie in *Bacillus subtilis* durchgeführt wird, ist (Itaya 1995; Ohtani *et al.* 2012; Itaya *et al.* 2003). Dieses Bakterium kann effizient einzelne, lineare DNA-Fragmente aufnehmen und ins Chromosom integrieren, so dass eine schrittweise DNA-Assemblierung möglich ist. Allerdings ist die Insertion in das Chromosom auch gleichzeitig die Limitierung dieser Methode, denn die Isolation der assemblierten DNAs in Form von episomalen Vektoren und deren Transfer aus dem Chromosom von *B. subtilis* ist bisher nur beschränkt möglich (Itaya and Tanaka 1997; Tanaka and Ogura 1998; Kaneko *et al.* 2005). Es ist derzeit technisch nicht möglich ganze in *B. subtilis* assemblierte Chromosomen zu isolieren.

Ein Vorteil von *S. cerevisiae*, der jedoch gleichzeitig ein Nachteil sein kann, ist die effiziente homologe Rekombination, denn diese benötigt lediglich 20 bp an homologer Sequenz für eine erfolgreiche Rekombination (Gibson 2009). Die homologe Rekombination in Hefe wird genutzt, um wie bereits beschrieben *in vivo* DNAs zu assemblieren, jedoch können dadurch assemblierte DNA-Konstrukte mit bereits kurzen homologen Bereichen instabil werden (Resnick and Nilsson-Tillgren 1990). Zudem ist es schwierig in Hefe Chromosomen mit einem hohen GC-Gehalt zu assemblieren und stabil aufrecht zu erhalten (Noskov *et al.* 2012; Karas *et al.* 2013b). Die natürlichen Chromosomen der Hefe haben nicht einen einzelnen Replikationsursprung, sondern viele in unregelmäßigen Abständen auftretende, autonom replizierende Sequenzen (ARS) (Dhar *et al.* 2012; Musialek and Rybaczek 2015). ARS sind AT-reiche Sequenzabschnitte, was bei einem Design von synthetischen Chromosomen mit hohem GC-Gehalt zwingend zu berücksichtigen ist. Es konnte gezeigt werden, dass es etwa alle 100 kb zusätzlich eingefügte ARS erst möglich machen GC-reiche Chromosomen in Hefe zu assemblieren und als Replikon aufrecht zu erhalten (Noskov *et al.* 2012; Karas *et al.* 2013b). Dies ist nicht notwendig bei Chromosomen mit einem GC-Gehalt ähnlich dem der Hefe (38 %), da die Wahrscheinlichkeit groß ist, dass ARS zufällig vorhanden sind (Karas *et al.* 2013b). ARS besitzen eine 11 bp lange degenerierte DNA-Kernsequenz 5´-WTTTAYRTTTW-3´, welche beispielsweise im 1.084 kb Chromosom von *M. mycoides* 303-mal vorkommt, was durchschnittlich einer ARS pro 3,6 kb entspricht. Dadurch ist die Möglichkeit gegeben, dass einige der DNA-Sequenzen in der Hefe als ARS funktionieren können und eine Insertion von zusätzlichen Hefe-ARS nicht notwendig ist (Karas *et al.* 2013b).

Ist Sinorhizobium meliloti in Zukunft eine Alternative zu S. cerevisiae?

In der Synthetischen Biologie werden immer weitgreifendere Veränderungen an Genomen durchgeführt und neben *E. coli*, *M. mycoides* und *S. cerevisiae* werden viele weitere Modellorganismen bzw. Chassis für diese Zwecke erforscht und etabliert (Adams 2016). Ein interessanter Organismus in diesem Bezug ist *Sinorhizobium meliloti*, da dieser im Begriff ist als Chassis für die Synthetische Biologie etablieren zu werden (Döhlemann *et al.* 2016; Döhlemann *et al.* submitted). *S. meliloti* ist in der Lage homologe Rekombination durchzuführen, allerdings benötigt *S. meliloti* dafür mindestens 200 bp homologer Sequenzen (Becker *et al.* 2009). Dies könnte ein Vorteil für die Assemblierung von DNA-Sequenzen mit kleineren homologen DNA-Sequenzen sein, da die Chance eines Rekombinationsvorgangs geringer ist als in *S. cerevisiae*. Des Weiteren werden keine zusätzlichen ARS benötigt, denn die DNA-Replikation des gesamten Replikons kann, wie in Bakterien üblich, ausgehend von einem Replikationsursprung bewerkstelligt werden. In *S. meliloti* wurde ein auf repABC-Plasmiden basierendes System etabliert, um effizient DNAs zu assemblieren (Döhlemann *et al.* submitted). Dabei handelt es sich um Plasmide, die in einer einfachen Kopienzahl in *S. meliloti* vorliegen und eine hohe Replikonstabilität aufweisen (Döhlemann *et al.* submitted). Dieses könnte vielleicht in Zukunft genutzt werden, um ganze Chromosomen zu assemblieren, allerdings fehlen dazu noch Projektstudien, ob *S. meliloti* ähnlich große episomale DNAs stabil aufrechterhalten kann wie *S. cerevisiae*.

Das Potential von *S. meliloti* ist allerdings enorm, da dieses Bakterium neben dem primären Chromosom von 3,65 mb, zwei Megaplasmide pSymA (1,35 mb) und pSymB (1,68 mb) besitzt, wobei nur pSymB zwei essentielle Gene kodiert (Barnett *et al.* 2001; Finan *et al.* 2001; Capela *et al.* 2001; Galibert *et al.* 2001; diCenzo *et al.* 2013). Wenn die essentiellen Gene von pSymB in das Chromosom von *S. meliloti* transferiert werden, ist es möglich, dass lebensfähige Zellen ohne die beiden Megaplasmide entstehen können (diCenzo and Finan 2015). Angenommen die entstandene Genomreduktion kann für die Assemblierung heterologer DNA-Sequenzen genutzt werden, dann könnte dies bedeuten, dass DNA-Sequenzen mit mehr als 3 mb assembliert und in *S. meliloti* als extra Replikon aufrechterhalten werden könnten. Für *B. subtilis* konnte gezeigt werden, dass dieses Bakterium mehr als 3 mb zusätzlich in das Chromosom integrieren kann (Itaya *et al.* 2005; Watanabe *et al.* 2012). Sollte dies in *S. meliloti* ebenfalls möglich sein, könnte vielleicht ein ganzes *E. coli* Chromosom (4,6 mb) in *S. meliloti* als eigenständiges Replikon aufrechterhalten werden.

Durch Methoden der Synthetischen Biologie ist es möglich ganze Chromosomen zu designen und zu assemblieren. Eine effiziente DNA-Assemblierung basiert meist auf der Kombination von *in vitro* und *in vivo* Techniken mit einer finalen, heterologen *in vivo* DNA-Assemblierung. Finale Assemblierungen

müssen aus diesem Grund isoliert und transplantiert werden, um Zellen zu generieren deren Genom ausschließlich auf dem synthetischen Chromosom basiert. Doch wie können synthetische Chromosomen transplantiert und somit synthetische Zellen generiert werden?

4.1.3 Transplantation von synthetischen Chromosomen

Die Transplantation von synthetischen Chromosomen ist derzeit die größte Hürde zur Herstellung synthetischer Organismen. Große DNAs können assembliert und anschließend isoliert werden (Lartigue *et al.* 2009; Noskov *et al.* 2011; Karas *et al.* 2015). Es ist wichtig, dass die DNA-Extraktion sehr mild durchgeführt wird, um ein Scheren der DNA zu unterbinden (Lartigue *et al.* 2007; Noskov *et al.* 2011). Bisher konnten die einzigen erfolgreichen Transplantationen ganzer Chromosomen in Bakterien nur mit *Mycoplasma*-Arten durchgeführt werden (Lartigue *et al.* 2007; Labroussaa *et al.* 2016). Im Prinzip handelt es sich bei dieser Methode um eine DNA-Transformation, wobei der entscheidende Vorteil das Fehlen einer Zellwand in der Gattung *Mycoplasma* sein dürfte (Razin *et al.* 1998; Dybvig and Voelker 1996).

Limitierende Faktoren bei Genomtransplantationen und wie sie möglicherweise zu umgehen wären

Es wäre denkbar, dass *Mycoplasma* sich in Zukunft als Zellsystem etabliert, um hergestellte synthetische Chromosomen ähnlich einer Computersoftware in einer Zelle zu installieren und zu starten. Der technische Aufwand für eine Genomtransplantation ist immens und die Transformationseffizienz sehr gering (Lartigue *et al.* 2007). Zudem konnte gezeigt werden, dass der Verwandtschaftsgrad von Donor und Rezipient einen Einfluss auf die Effizienz der Transplantation besitzt (Labroussaa *et al.* 2016). Dies bedeutet, dass diese Technik noch viel Optimierung bedarf, um beispielsweise ein *E. coli* Chromosom in *Mycoplasma* zu transplantieren. Sollten *Mycoplasma*-Zellen sich als Rezipienten für Genomtransplantationen etablieren, würde das zwingend notwendige Designregeln für die kodierenden und regulativen DNA-Sequenzen der synthetischen Chromosomen mit sich bringen. Das transplantierte Chromosom wäre abhängig von den im *Mycoplasma*-Rezipienten vorliegenden Proteinen und von der in *Mycoplasma* verwendeten Kodierung von Genen (Inamine *et al.* 1990; Simoneau *et al.* 1993). Würden diese Designanforderungen an das synthetische Chromosom nicht umgesetzt werden, könnten keine lebensfähigen Zellen entstehen.

Ein Projekt, das versucht ein minimales *E. coli* Chromosom (MGE-syn1.0) bestehend aus 449 essentiellen Genen und 267 konditionalen Genen zu assemblieren und dieses anschließend aus der Hefe in *E. coli* zu transplantieren, scheiterte bisher an einer erfolgreichen Transplantation (Zhou *et al.* 2016). Dies kann an einer ineffizienten Transformation liegen oder an einer unzureichenden Qualität der zu transformierenden DNA. In der Studie konnte gezeigt werden, dass bereits eine Transformation der

Teilassemblierung von 308 kb als episomaler Vektor in *E. coli* nicht möglich ist (Zhou *et al.* 2016). In der vorliegenden Arbeit konnte ebenfalls festgestellt werden, dass eine Transformation in *E. coli* mit synthetischen sekundären Chromosomen mit fast 100 kb Größe bereits ineffizient ist. Dies stimmt mit anderen Studien überein, eine Transformation von episomalen DNA-Konstrukten in *E. coli* ist scheinbar größenlimitiert (Sheng *et al.* 1995; Gibson *et al.* 2009). Organismen, die keine Zellwand besitzen, wie *Mycoplasma*, scheinen eine der wenigen Ausnahmen zu sein, in der die Transformation ganzer Chromosomen praktikabel ist.

Um MGE-syn1.0 zu transplantieren, wurde alternativ versucht die Protoplastenfusion von Hefe und *E. coli* zu etablieren, allerdings war dieser Ansatz bisher ebenfalls erfolglos. Andere Studien zeigen, dass eine Protoplastenfusion zur Aufnahme von ganzen Chromosomen in Hefe möglich ist (Karas *et al.* 2013a; Karas *et al.* 2014). Eine Protoplastenfusion zur Generierung von Bakterien, die Hefe-DNA aufgenommen haben, konnte jedoch bisher nicht gezeigt werden. MGE-syn1.0 ist ein *de novo* Design und keine sukzessive Reduktion, weshalb an dieser Stelle nicht außer Acht gelassen werden darf, dass eine Funktionalität des Designs bisher noch nicht gezeigt werden konnte. Obwohl alle einzelnen Komponenten getestet wurden, ist es möglich, dass dem finalen Design essentielle Komponenten fehlen (Xue *et al.* 2015). Die MGE-syn1.0 Studie zeigt damit auf, dass dringend neue Methoden benötigt werden, um synthetische Chromosomen zu transplantieren. Zudem darf eine effiziente Transplantationsmethode nicht auf wenige Arten beschränkt sein, wie beispielsweise die Genomtransplantation in *Mycoplasma*.

Ist die Konjugation die Methode der Zukunft für Genomtransplantationen?

Eine Methode, die sich in Zukunft für Genomtransplantationen anbieten könnte, wäre die Konjugation (Curtiss 1969; Frost 1992). In der vorliegenden Arbeit konnte erfolgreich ein synthetisches sekundäres Chromosom etabliert und charakterisiert werden, welches durch einen *origin of transfer* (*oriT*) konjugiert werden kann (Messerschmidt *et al.* under revision). Es konnte gezeigt werden, dass die erstellten, synthetischen sekundären Chromosomen mit einer Größe von nahezu 100 kb dadurch erfolgreich konjugiert werden können, ohne aufwendiges Isolieren und Transformieren der synVicII-Varianten (Schindler and Waldminghaus in preparation). Eine Konjugation von Bakterien in die anderen Domänen des Lebens ist ebenfalls möglich. Bereits 1989 konnte gezeigt werden, dass beispielsweise ein DNA-Transfer von Bakterien in *S. cerevisiae* durch Konjugation möglich ist (Heinemann and Sprague 1989). Allerdings ist es nicht möglich von *S. cerevisiae* in Bakterien zu konjugieren.

Vielleicht wäre es möglich durch Techniken der Synthetischen Biologie in Zukunft einen Hefe-Stamm zu generieren, der in der Lage ist DNA in Bakterien zu konjugieren. Die Herausforderung hier sollte die unterschiedliche Beschaffenheit der Membranen und Zellwände von Bakterien und Eukaryoten sein, denn der bakterielle Konjugationskomplex besteht zum einen aus einem Transmembrankomplex von etwa 10 verschiedenen Proteinen und zum anderen aus einem Pilus, der den finalen Zellkontakt herstellt (Cabezon *et al.* 2015; Ilangovan *et al.* 2015; Llosa and de la Cruz 2005). Der Konjugationskomplex wird vermutlich nicht durch heterologe Expression der bakteriellen Gene in Hefe zu einem funktionellen Konjugationsprozess führen. Sollte dieser Prozess eines Tages möglich sein, wäre dies eine sehr effiziente und robuste Methode, um ganze, in Hefe assemblierte oder modifizierte Genome wieder in Bakterien einzubringen. Eine Möglichkeit, diese Limitierung der Hefe zu umgehen, wäre es, nach alternativen Chassis zu *S. cerevisiae* zu suchen. Eine solche mögliche Alternative könnte das bereits erwähnte Bakterium *S. meliloti* sein. Wenn sich die DNA-Assemblierung ganzer Genome als möglich erweisen sollte, könnten diese anschließend durch die in *S. meliloti* etablierte Konjugation übertragen werden (Simon *et al.* 1986; Krol and Becker 2014; Simon *et al.* 1983).

Die derzeitige Limitierung der Genom-Transplantation ist verantwortlich dafür, dass Projekte wie das MGE-syn1.0 Projekt nicht abgeschlossen werden können. Dabei ist es fraglich, ob die Transplantation die Limitierung ist oder vielleicht ein fehlerhaftes Design vorliegt. Nichtsdestotrotz werden neue Methoden benötigt, um effizient Transplantationen von ganzen synthetischen Chromosomen zu ermöglichen.

4.2 Potentiale und Anwendungen von synthetischen sekundären Chromosomen

Das Design, die Synthese und Transplantation von ganzen Chromosomen ist heutzutage möglich, bedarf aber noch großer wirtschaftlicher Ressourcen und es gibt keine Erfolgsgarantie. Sowohl in biotechnologischen Anwendungen als auch in der Grundlagenforschung ist es wichtig, gut charakterisierte Systeme für das heterologe Einbringen von DNA-Sequenzen zu haben. Traditionell wird dies durch Plasmide bewerkstelligt, allerdings wird das Verwenden von Plasmiden durch eine Größenlimitierung in Zukunft beschränkt sein, denn die zu kodierenden Informationen werden immer größer und Plasmide mit hoher Kopienzahl sind ihrer Größe limitiert (Fong *et al.* 2007). Des Weiteren verhalten sich Plasmide nicht wie Chromosomen, denn sie besitzen keine zellzyklusabhängige Regulation, wodurch sie für Fragestellungen im Bereich der DNA-Replikation nicht verwendet werden sollten. In der vorliegenden Arbeit konnte in zwei aufeinander aufbauenden Studien zunächst ein synthetisches sekundäres Chromosom (synVicII) in *E. coli* etabliert, charakterisiert und optimiert werden

(Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision). In einer dritten Studie wurde eine Vorgehensweise etabliert, mit der große, variable DNA-Sequenzen effizient und kostengünstig assembliert werden können (Schindler *et al.* 2016). Die vierte Studie dieser Arbeit wendet synVicII an, um das Verständnis von *Chromosome Maintenance* Systemen zu erweitern und grundlegende Regeln für das Design von synthetischen Chromosomen zu generieren (Schindler and Waldminghaus in preparation). Das Potential und die Anwendungsmöglichkeiten synthetischer sekundärer Chromosomen sind jedoch deutlich breiter zu fassen.

4.2.1 Synthetische sekundäre Chromosomen: Konzepte, Designs und erste Erkenntnisse

Die generellen, der DNA-Replikation zu Grunde liegenden Prozesse sind verstanden (O'Donnell *et al.* 2013; Murray 2016; Skarstad and Katayama 2013). Allerdings sind für jede beantwortete Fragestellung neue Fragen aufgekommen. Plasmide und *oriC*-basierte Minichromosomen haben in *E. coli* viel zum Verständnis der Regulation der DNA-Replikation beigetragen (Hiraga 1976; Leonard and Helmstetter 1986; Helmstetter *et al.* 1997; Woelker and Messer 1993). Durch *oriC*-basierte Minichromosomen konnte herausgefunden werden, dass es einen Sequestrierungsfaktor geben muss, der dafür sorgt, dass die DNA-Replikation nur einmal pro Zellzyklus initiiert wird (Russell and Zinder 1987; Campbell and Kleckner 1990). Dieser postulierte Sequestrierungsfaktor konnte in einer nachfolgenden Studie durch ein Screening, in dem ein *oriC*-basiertes Minichromosom verwendet wurde, identifiziert werden und wurde SeqA genannt (Lu *et al.* 1994). *oriC*-basierte Minichromosomen haben jedoch ebenso anwendungsbezogene Nachteile, denn sie haben eine erhöhte Kopienzahl im Vergleich zum Chromosom und konkurrieren um die Regulationsfaktoren der DNA-Replikation (Skarstad and Lobner-Olesen 2003; Lobner-Olesen 1999). Zudem haben *oriC*-basierte Minichromosomen homologe Bereiche zum Chromosom, wodurch diese in das Chromosom integriert werden und als freireplizierende Replikons verloren gehen können (Lobner-Olesen 1999; Skarstad and Lobner-Olesen 2003; Messerschmidt *et al.* under revision). Aus diesem Grund ist es wichtig alternative Systeme für wissenschaftliche Fragestellungen und anwendungsbezogene Prozesse zu entwickeln, die als zweites Chromosom in *E. coli* etabliert werden können. Es konnte gezeigt werden, dass sich Minichromosomen auf Basis des zweiten Chromosoms von *V. cholerae* als alternatives System dazu eignen (Egan and Waldor 2003; Liang *et al.* 2013). In der vorliegenden Arbeit wurde deshalb synVicII, basierend auf dem vorhandenen Wissen zur DNA-Replikation des zweiten Chromosoms in *V. cholerae* als synthetisches sekundäres Chromosom in *E. coli* etabliert (Messerschmidt *et al.* 2015; Messerschmidt *et al.* under revision).

synVicII: ein etabliertes System für vergleichende Analysen von Chromosome Maintenance Systemen

Bisher wurden *Chromosome Maintenance* Systeme aufgrund von der Verteilung von DNA-Sequenzmotiven auf dem Chromosom identifiziert, wobei die meisten Systeme durch computergestützte Analysen von bereits sequenzierten Genomen identifiziert werden konnten (Bigot *et al.* 2005; Sourice *et al.* 1998; Touzain *et al.* 2011; Halpern *et al.* 2007; Mercier *et al.* 2008). Die Funktionalität wurde, sofern möglich, durch Deletion eines oder mehrerer Gene eines *Chromosome Maintenance* Systems bzw. der korrespondierenden Erkennungssequenzen auf dem Chromosom verifiziert oder *in vivo* durch Plasmide analysiert (Yamaichi and Niki 2004; Touzain *et al.* 2011). Alternativ wurden *in vitro* Analysen durchgeführt, durch die ein generelles Verständnis von Protein-Wechselwirkungen mit dem DNA-Sequenzmotiv erlangt werden konnten (Han *et al.* 2004; Brendler and Austin 1999; Slater *et al.* 1995; Mercier *et al.* 2008). Häufig vorkommende DNA-Motive wie beispielsweise die 19.120 GATC-Sequenzen können nicht im gesamten *E. coli* Chromosom deletiert werden. synVicII stellt ein ideales System dar, um eine solche Limitierungen zu umgehen, denn es können sekundäre Chromosomen mit verschiedenen DNA-Sequenzanordnungen vergleichend analysiert werden (Schindler and Waldminghaus in preparation). synVicII ist ein *in vivo* System, das sich wie ein Chromosom verhält. Die vorliegende Arbeit stellt in diesem Bereich erstmals eine erfolgreiche Studie vor, in der durch diese Vorgehensweise systematisch GATC-Sequenzen auf synVicII angeordnet werden und damit die Analyse einer möglichen Interaktion zwischen SeqA und der *DNA mismatch* Reparatur ermöglichen (Schindler and Waldminghaus in preparation). Die Praktikabilität des Ansatzes kann in der Studie dadurch belegt werden, dass eine effiziente DNA-Assemblierung und Generierung der synthetischen sekundären Chromosomen möglich ist. Zudem gehen die Chromosomen auch über eine verlängerte Kultivierungsdauer in *E. coli* nicht verloren. Außerdem belegen die SeqA-*ChIP* Sequenzierungsdaten, dass das experimentelle Design der Studie funktioniert, denn SeqA kann die DNA-Sequenz außerhalb des synVicII-Rückgrats nur binden, wenn GATC-Cluster vorliegen, wodurch *in vitro* Daten zum Bindeverhalten von SeqA bestätigt werden. Die Sequenzierungen der Stämme mit den verschiedenen synthetischen sekundären Chromosomen nach 25 Tagen permanenter Kultivierung und die anschließende Analyse der Einzelnukleotidvarianten innerhalb der Stämme belegt, dass eine funktionierende *DNA mismatch* Reparatur essentiell für die Integrität der DNA-Sequenz ist.

Ist die Interaktion von FtsK mit KOPS hinreichend für die Segregation von Chromosomen?

Die in dieser Arbeit gezeigte Studie belegt, dass synVicII ein etabliertes System zur systematischen Analyse von *Chromosome Maintenance* Systemen ist (Schindler and Waldminghaus in preparation). Das Wissen über *Chromosome Maintenance* Systeme kann nun anhand von synVicII systematisch erweitert

werden. Ein interessantes *Chromosome Maintenance* System, das in Zukunft analysiert werden sollte ist die Interaktion der Translokase FtsK mit *KOPS*, welches in Bakterien konserviert und weit verbreitet ist (Nolivos *et al.* 2012). Wirkt sich *KOPS* auf die Replikonstabilität des synthetischen sekundären Chromosoms aus und ist dieses *Chromosome Maintenance* System in *E. coli* vielleicht sogar hauptverantwortlich für die Segregation der Chromosomen? In *E. coli* ist kein explizites Segregationssystem bekannt und es gibt mehrere Hypothesen darüber, wie die Segregation der Chromosomen funktionieren könnte. Vermutet wird, dass mehrere zelluläre Prozesse für die Segregation des *E. coli* Chromosoms verantwortlich sind und nicht ein einzelner Segregationsmechanismus (Reyes-Lamothe *et al.* 2012; Kleckner *et al.* 2014; Stouf *et al.* 2013). Aber wie können Segregationsmechanismen, die DNA-Motive benötigen, welche über das ganze Chromosom verteilt sind, *in vivo* analysiert werden? Mit der in dieser Arbeit etablierten MoClo-Strategie in Kombination mit MARSeG könnte ein Set von synthetischen sekundären Chromosomen geplant werden, um den Einfluss von *KOPS* auf die Replikonstabilität von synthetischen sekundären Chromosomen in *E. coli* zu analysieren (Schindler *et al.* 2016). Dabei können die synthetischen sekundären Chromosom mit den verschiedenen GATC-Anordnungen bereits als Referenz genommen werden, da dieses frei von *KOPS* sind (Schindler and Waldminghaus in preparation). Die weiteren zu generierenden synthetischen sekundären Chromosomen sollten eine uniforme, eine alternierende und eine zur Terminusregion orientierte Anordnung von *KOPS* aufweisen.

Mit den in der vorliegenden Arbeit etablierten Methoden zur Analyse der Stabilität könnte dadurch der Einfluss von *KOPS* auf die Segregation von Chromosomen analysiert werden. Es sollte zu erwarten sein, dass die Replikonstabilität mit zur Terminusregion orientierten *KOPS* größer ist als bei den anderen synVicII-Varianten, was einen Einfluss auf die Segregation durch FtsK und *KOPS in vivo* bestätigen würde. Bisherige Daten basieren auf Fluoreszenzmikroskopie, bisher konnte damit noch nicht gezeigt werden, ob eine definierte Anordnung von *KOPS* über ein ganzes Chromosom positive oder negative Auswirkungen auf die Segregation eines Replikons hat. Vielleicht kann durch dieses Experiment sogar gezeigt werden, dass die Interaktion von FtsK und *KOPS* in *E. coli* hinreichend ist für die Segregation von Chromosomen. Gleichzeitig könnte der Einfluss von Chromosomen-Catenanen auf die Replikonstabilität analysiert werden, indem das synVicII mit den zur Terminusregion orientierten *KOPS* in zwei Varianten hergestellt wird: mit und ohne *dif*-Sequenz (Val *et al.* 2008; Hendrickson and Lawrence 2007). In Zukunft können systematische, vergleichende, chromosomenweite Analysen von einzelnen oder einer Kombination mehrerer *Chromosome Maintenance* Systeme *in vivo* anhand von synVicII in *E. coli* durchgeführt werden.

4.2.2 Anwendungsmöglichkeiten von sekundären synthetischen Chromosomen

Synthetische sekundäre Chromosomen haben ein enormes Potential für biotechnologische Anwendungen. In Zukunft könnte ein Kapazitätslimit von Plasmiden erreicht werden oder aber eine Feineinstellung der Expression einzelner Gene durch die Kopienzahl der Plasmide nicht gewährleistet sein (Jones *et al.* 2000b). Synthetische sekundäre Chromosomen bieten sich als zellzyklusregulierte Replikons an, um aufwändige chromosomale Integrationen zu umgehen. synVicII ermöglicht die *in vitro* DNA-Assemblierung und das stabile Einbringen von mindestens 100 kb DNA-Sequenz in *E. coli* (Schindler and Waldminghaus in preparation). Darüber hinaus dürfte es möglich sein, synVicII in *E. coli* nachträglich zu manipulieren, zum Beispiel durch das Integrieren von zusätzlichen DNA-Sequenzen. synVicII stellt also eine herausragende Alternative zu Plasmiden und genomischen Integrationen für biotechnologische Anwendungen dar.

Anwendung von synVicII zur Optimierung von Biosynthesewegen

In der Biotechnologie ist die Optimierung von Prozessen, um eine erhöhte Produktion eines Produkts zu bewirken, enorm wichtig (Welch *et al.* 2009; Basler *et al.* 2012). Doch wie wird eine Optimierung durchgeführt? Um Gene eines Biosyntheseweges zu optimieren, werden häufig Zufallsmutageneseprotokolle verwendet (Arnold 1993; Sauer 2001). Dazu werden Kulturen mutagenen Substanzen oder mutagenen physikalischen Einflüssen ausgesetzt. Der Nachteil dabei ist, dass das gesamte Genom zufällig mutiert und Effekte nicht zwingend auf Mutationen der zu optimierenden DNA-Sequenz basieren. Die Zufallsmutagenese kann durch das Verwenden von DNA-Reparatur defizienten Stämme beschleunigt werden, aber die Mutagenese findet trotzdem genomweit statt (Greener *et al.* 1997; Fabret *et al.* 2000; Wright 2004; Lu *et al.* 2001).

Die vorliegende Arbeit zeigt, dass in synthetischen sekundären Chromosomen mit null GATCs eine deutlich erhöhte Mutationsrate vorliegt (Schindler and Waldminghaus in preparation). Würden in synVicII Gene für Biosynthesewege ohne GATCs eingebracht werden, würden diese Sequenzen deutlich stärker mutieren als die übrigen DNA-Sequenzen. Dadurch könnte lokal eine erhöhte Mutationsrate erzielt werden, mit der ausgewählte Sequenzen gezielt durch Zufallsmutagenese optimiert werden könnten. Dies wäre ein deutlicher Vorteil gegenüber klassischen Mutageneseprotokollen oder dem Verwenden von reparaturdefizienten Stämmen, bei denen eine globale Zufallsmutagenese vorliegen würde. Nach einer Mutagenese wäre es einfach zu testen, ob die Optimierung auf der Mutagenese der synVicII DNA-Sequenz oder auf Mutationen des Genoms beruht, indem synVicII mittels Konjugation in den Ursprungsstamm zurücktransferiert wird. Vergleichbare Ansätze verwenden Polymerase I (Pol I) abhängige Plasmide in Kombination mit der Überexpression einer fehlerhaften Pol I (Fabret *et al.* 2000;

Camps *et al.* 2003; Alexander *et al.* 2014). Der Nachteil ist, dass die Mutagenese auf kurze DNA-Sequenzen beschränkt ist, da Pol I nur die DNA-Replikation der ersten etwa 700 bp nach der Initiation der Plasmid-Replikation katalysiert und anschließend Polymerase III die restliche Sequenz des Plasmids repliziert (Camps *et al.* 2003; Itoh and Tomizawa 1979). Der Mutageneseansatz mit synVicII wäre der Methode der Pol I Mutagenese überlegen, da in synVicII mindestens 100 kb DNA-Sequenz zufällig mutiert werden könnten. Dazu bedarf es allerdings noch einer Optimierung des Mutageneseverfahrens, um die Mutationsrate zu erhöhen.

Herstellung eines Minimalorganismus anhand von synVicII

Eine weitere interessante Anwendung von synVicII in der Grundlagenforschung wäre es, nach und nach essentielle Gene vom *E. coli* Chromosom auf synVicII zu übertragen. An dem Punkt, an dem alle essentiellen Gene auf synVicII kodiert wären, könnte das *E. coli* Chromosom verloren gehen und ein Minimal-*E. coli* mit dem *oriII* von *V. cholerae* wäre entstanden. Die Replikation des gesamten *E. coli* Chromosoms nur mit *oriII* von *V. cholerae* ist möglich und somit könnten die bei MGE-syn1.0 vorliegenden Probleme bezüglich Transplantation und Funktionalität umgangen werden (Liang *et al.* 2013; Milbredt *et al.* 2016). Informationen über die essentiellen Gene in *E. coli* liegen aufgrund umfassender Studien vor, die Herstellung eines minimalen *E. coli* Stammes ist folglich nur eine Frage der Zeit (Baba *et al.* 2006; Gerdes *et al.* 2003; Zhou *et al.* 2016; Xue *et al.* 2015). Das beschriebene Vorgehen, ein minimales *E. coli* Chromosom auf Basis von synVicII zu konstruieren, könnte in Zukunft durch das Sinken der DNA-Synthesekosten tatsächlich verwirklicht werden (Kosuri and Church 2014; Carr and Church 2009).

Für biotechnologische Anwendungen ist es wichtig, dass Expressionssysteme äußerst robust und stabil sind (Kroll *et al.* 2010; Sahdev *et al.* 2008). Im Hinblick auf die Grundlagenforschung im Bereich der DNA-Replikation ist ein Replikon, das verloren gehen kann, interessant. Durch ein solches Replikon können Analysen von stabilisierenden und destabilisierenden Elementen ermöglicht werden. Für synVicII ist zumindest in der Grundlagenforschung das Ziel eines etablierten Replikons, um die DNA-Replikation zu analysieren, erreicht, denn die aktuelle MoClo kompatible Version von synVicII erlaubt es stabilisierende und destabilisierende Effekte gleichermaßen zu analysieren. synVicII ermöglicht umfassende Studien im Bereich der *Chromosome Maintenance* Systeme und zur Analyse von Faktoren, die synVicII im Hinblick auf eine biotechnologische Anwendung stabilisieren. Doch wie könnten zukünftige Projekte dazu aussehen?

4.2.3 Zukünftige Optimierungen von synVicII

Für biotechnologische Anwendungen ist es enorm wichtig, dass Expressionsplattformen in den Zellen nicht verloren gehen und die genetische Integrität gewahrt ist. Im Idealfall bräuchte kein Selektionsdruck verwendet werden, denn die Zugabe eines Selektionsdrucks (z.B. ein Antibiotikum) in das Medium im industriellen Produktionsmaßstab sind nicht zu unterschätzende Kosten. Aber wie kann auf einen solchen Selektionsdruck verzichtet werden? Für synVicII kann dies erreicht werden, indem ein essentielles Gen anstelle einer Antibiotikaresistenz verwendet wird. In der vorliegenden Arbeit wurde das essentielle Gen *thyA* auf dem Chromosom von *E. coli* deletiert und anstelle des *bla* Gens in synVicII integriert, was zu einer Erhöhung der Replikonstabilität führte (Messerschmidt *et al.* 2015). Folglich ist es möglich die Replikonstabilität durch den Transfer von essentiellen Genen vom primären *E. coli* Chromosom auf synVicII zu erhöhen und somit auf einen Selektionsdruck (z.B. ein Antibiotikum) zu verzichten. Für ein besseres Verständnis der DNA-Replikation ist es jedoch interessanter welche nicht essentiellen Faktoren eine Erhöhung der Replikonstabilität zur Folge haben.

Durch die Verwendung von synVicII, um die Funktionsweise von *Chromosome Maint*enance Systemen zu erforschen, ist es möglich anhand der erhaltenen Ergebnisse synVicII zu optimieren. Eine erhöhte Replikonstabilität kann zwar durch das Transferieren essentieller Gene vom primären Chromosom auf das sekundäre Chromosom erreicht werden, aber es kann nicht ausgeschlossen werden, dass sich dadurch vielleicht die Kopienzahl von synVicII verändert. Eine verbesserte Replikonstabilität könnte alternativ durch das Optimieren der Segregationseigenschaften von synVicII in *E. coli* erreicht werden. In *Vibrio cholerae* werden die Chromosomen durch ein Partitionierungssystem bestehend aus den Proteinen ParA, ParB und der ParB DNA-Bindesequenz (*parS*) segregiert, wobei jedes Chromosom ein replikonspezifisches ParAB System aufweist (Schumacher and Funnell 2005; Yamaichi *et al.* 2007). Das sekundäre Chromosom weißt dabei neun *parS* über das Chromosom verteilt auf (Yamaichi *et al.* 2007). In synVicII liegt hingegen nur eine *parS* vor, eine Erhöhung der Anzahl an *parS* sollte eine verbesserte Segregation und damit eine erhöhte Replikonstabilität zur Folge haben (Yamaichi *et al.* 2007). Des Weiteren sollte das Einfügen einer *dif*-Sequenz das Auflösen von Chromosomen-Catenanen erlauben und damit ebenfalls zur Replikonstabilität durch eine optimierte Segregation beitragen. In Kombination mit der bereits diskutierten Anordnung von *KOPS*, die möglicherweise einen aktiven Einfluss auf die Segregation der Chromosomen besitzen, könnte die Replikonstabilität des ursprünglichen synVicII vermutlich erhöht werden. Aber es ist fraglich, ob dies bereits ausreichen würde, um eine perfekte Replikonstabilität zu erhalten.

Optimierung von synVicII durch Modifikation des primären Chromosoms

Optimierungen in Bezug auf die Stabilität von synVicII müssen nicht zwingend an synVicII selbst durchgeführt werden. Eine kurz vor dem Einreichen der vorliegenden Arbeit erschienene Studie hat eine bis dahin unbekannte Sequenz (*crtS*) auf Chromosom I von *V. cholerae* charakterisiert, die für die Koordination der Initiation der DNA-Replikation von Chromosom II essentiell ist (Val *et al.* 2016). Es sollte in weiterführenden Studien in Betracht gezogen werden, diese *crtS* in das Chromosom von *E. coli* zu integrieren. synVicII repliziert in *E. coli* und liegt in einfacher Kopienzahl vor, allerdings ist es jedoch nicht eindeutig, ob die Initiation der DNA-Replikation von synVicII in Abhängigkeit des primären Chromosoms koordiniert ist (Messerschmidt A und Kröte). Sollte die zeitliche Koordination des Chromosoms II in *V. cholerae* lediglich von dem Vorhandensein der *crtS* abhängen, würde diese Sequenz auch in *E. coli* einen Einfluss auf die Koordination der Initiation der DNA-Replikation besitzen.

Das Entdecken von *crtS* lässt Spekulationen über weitere, bisher unbekannte Faktoren zu, die für die Stabilität von Chromosom II in *V. cholerae* verantwortlich sind. Doch wie lassen sich solche Faktoren identifizieren? Ein Verlust von Chromosom II hat für *V. cholerae* letale Folgen, weshalb es schwer ist Faktoren zu identifizieren, die die Stabilität von Chromosom II beeinflussen. Eine Studie konnte zeigen, dass es möglich ist die beiden Chromosomen von *V. cholerae* zu fusionieren und dass die Replikation lediglich über den Replikationsursprung von Chromosom I bewerkstelligt werden kann (Val *et al.* 2012). Interessanterweise scheint die Fusion von Chromosomen auch unter natürlichen Bedingungen vorzukommen (Val *et al.* 2014a; Johnson *et al.* 2015). Es konnten *V. cholerae* Stämme aus Patienten isoliert werden, deren sekundäres Chromosom in das primäre Chromosom integriert ist (Johnson *et al.* 2015). Diese Stämme mit nur einem Chromosom würden sich eignen, um vergleichende, globale Transposonmutagenesen durchzuführen. Vergleichende Analysen von monochromosomen *Vibrio cholerae* Stämmen und dem Wildtyp würden es ermöglichen Faktoren zu identifizieren, die einen Einfluss auf die Stabilität von Chromosom II besitzen. Wenn solche Faktoren identifiziert werden würden, könnten diese verwendet werden, um zum einen ein detaillierteres Verständnis über die DNA-Replikation in *V. cholerae* zu erlangen und zum anderen, um synVicII in *E. coli* weiter zu stabilisieren.

4.3 Designerorganismen heute und in Zukunft

Designerorganismen sind nicht natürlich vorkommende Organismen deren Genom auf synthetischen Chromosomen oder auf einer großen Anzahl an DNA-Sequenzveränderungen durch Methoden des *Genome Engineering* basiert. Die vorliegende Arbeit konzentriert sich zwar auf bakterielle synthetische Chromosomen, jedoch ist die Synthese von ganzen Chromosomen nicht auf Bakterien beschränkt (Annaluru *et al.* 2014; Pennisi 2014; Boeke *et al.* 2016). In diesem Kapitel soll übergreifend diskutiert und dargestellt werden, was im Bereich der Designerorganismen möglich ist und möglich sein wird.

4.3.1 Konzepte für zukünftige Designerorganismen: Von Design bis Transplantation

Neben den bereits erwähnten Designerorganismen, die auf das „minimale“ Set an Genen reduziert wurden oder eine reduzierte Anzahl an Codons besitzen, gibt es auch noch weitere interessante Konzepte für mögliche Designerorganismen. Dabei wäre es denkbar beispielsweise Organismen zu generieren, die deutlich schneller wachsen und somit eine höhere Auslastung von industriellen Fermentern erlauben würden. In diesem Bereich gibt es keine öffentlich postulierten Projekte, aber es wurde ein natürlich vorkommender Organismus, *Vibrio natriegens* für molekularbiologische Anwendungen beschrieben. Dieser besitzt eine Verdopplungszeit von unter 10 min und somit einer nur halb so langen Verdopplungszeit verglichen mit *E. coli* (Weinstock *et al.* 2016; Lee *et al.* 2016). Dieser Organismus ist dadurch für die industrielle Anwendung äußerst interessant, da eine größere Auslastung von industriellen Anlagen erreicht werden könnte. Wäre es möglich herauszufinden warum *V. natriegens* so schnell wächst und wäre es möglich diese Eigenschaft auf *E. coli* zu übertragen? Ein somit äußerst schnell wachsender *E. coli* Designerstamm hätte einige Vorteile, denn das breite Spektrum an Methoden und Anwendungen könnte beibehalten werden und müsste nicht für *V. natriegens* etabliert werden. Es ist allerdings fraglich ob die Produktivität eines auf schnelles Wachstum ausgelegten *E. coli* Stammes mit denen der konventionellen *E. coli* Produktionsstämme vergleichbar ist. Hervorzuheben ist, dass dieses konzeptionelle Design eines schnell wachsenden *E. coli* Stammes lediglich ein Beispiel für eine Vielzahl an möglichen Designerorganismen wäre.

Schnelles Wachstum ist nicht der wichtigste Punkt für eine Etablierung eines Organismus in der industriellen Anwendung. Ein Organismus für die biotechnologische Anwendung sollte mit möglichst geringem Aufwand und möglichst kostengünstig kultiviert werden können. Photosynthese betreibende Organismen sind deshalb äußerst interessant, da diese lediglich CO_2 als Kohlenstoffquelle und Licht benötigen (Abed *et al.* 2009; Sarsekeyeva *et al.* 2015). Allerdings haben diese Organismen häufig lange Generationszeiten, der Zellaufschluss ist komplizierter und die Ausbeuten von Produkten sind geringer

(Mori *et al.* 1996; Bernstein *et al.* 2014). Wie wäre es, das Chromosom eines *E. coli* oder *B. subtilis* Produktionsstammes als Chassis zu nutzen und das Chromosom eines autotrophen Bakteriums zu integrieren und somit eine Chimäre als Designerorganismus herzustellen? Eine Chimäre ist ein Fabelwesen aus der griechischen Mythologie, aber sind Chimären nur ein Mythos? Ist es möglich durch Methoden der Synthetischen Biologie chimäre Produktionsstämme herzustellen, die lediglich CO_2, Licht und Mineralien für die Kultivierung benötigen? Die Vorteile wären, dass aufgrund des verwendeten Chassis die bereits etablierten molekularbiologischen Methoden weiter anwendbar wären. Dadurch müssten beispielsweise Zellaufschlussprotokolle nicht verändert werden, so dass bestehende Produktionsabläufe beibehalten werden könnten. Solche Chimären würden die Eigenschaften des schnellen Wachstums mit der Autotrophie und dem etablierten Spektrum an molekularbiologischen Methoden verbinden. Folgen wären eine Verringerung der Kosten für Nährlösungen. Die Produkte könnten entsprechend aufgearbeitet werden und zusätzlich könnte in Zukunft die anfallende Biomasse für die Produktion von beispielsweise Bioethanol verwendet werden (Dong *et al.* 2016; Ghasemi Naghdi *et al.* 2016; Sathish *et al.* 2015). Somit wäre es möglich Ressourcen zu schonen und gleichzeitig Fermentationsprozess weitergehend zu optimieren.

Die Generierung eines solchen chimären Organismus wurde in *B. subtilis* bereits versucht, indem das gesamte Chromosom eines Cyanobakteriums (*Synechocystis* PCC6803) in *B. subtilis* assembliert wurde (Watanabe *et al.* 2012; Itaya *et al.* 2005). Die Studie zeigt eindrucksvoll, dass die Integration von DNAs nicht der limitierende Faktor ist, um eine *B. subtilis* Chimäre zu erschaffen. Allerdings können nur wenige der Genprodukte aus dem *Synechocystis* PCC6803 Chromosom in *B. subtilis* nachgewiesen werden. Es ist naiv zu erwarten, dass durch die Kombination beider Genome eine funktionale Chimäre entsteht, denn es liegt kein enger Verwandtschaftsgrad der Organismen vor, weshalb sich unter anderem die Regulation der Genexpression unterscheidet. Um einen solchen Organismus zu generieren, bedarf es globaler Rekodierung, sowohl der Promotoren als auch der kodierenden DNA-Sequenz des zu integrierenden Genoms, so dass die Expression der Gene durch die Transkriptionsmaschinerie der Chimäre durchgeführt werden kann, was in Zukunft machbar sein sollte (Haimovich *et al.* 2015). Durch die Rekodierung der DNA-Sequenz des zu integrierenden Organismus ist es nicht notwendig das gesamte Genom zu integrieren da ansonsten Genprodukte in doppelter Ausführung vorliegen. Beispielsweise könnte auf die Gene für die DNA-Replikation, Transkription und Translation verzichtet werden. Es fehlen allerdings noch Pilotstudien, um die Funktionalität einer solchen Chimäre zu bestätigen, die heterologe Expression einzelner Gene und ganzer Biosynthesewege ist jedoch schon lange machbar (Harris and Emtage 1986; Frommer and Ninnemann 1995; Malpartida and Hopwood 1984; Arsenault *et al.* 2008).

Synthetische eukaryotische Chromosomen

Die meisten Prokaryoten sind von der genomischen Struktur her auf Effizienz aufgebaut, haben selten lange nichtkodierenden Sequenzbereiche und besitzen in der Regel ein zirkuläres Chromosom. Eukaryoten hingegen besitzen mehrere lineare Chromosomen mit langen nichtkodierenden Bereichen, wofür der Begriff „Junk-DNA“ verwendet wurde (Ohno 1972). Neben den unterschiedlichen Genomstrukturen sind eukaryotische Genome in der Regel deutlich größer als bakterielle Genome und haben eine erhöhte Anzahl von repetitiven Sequenzen. Zudem besitzen sie wie Bakterien transposable Elemente, die zu genomischer Instabilität führen (Boeke *et al.* 2016). Das ideale, eukaryotische Modellsystem zur Erforschung von biologischen Funktionen sollte jedoch möglichst stabile genomische Integrität aufweisen. Dies lässt die Frage zu, ob es möglich ist, eukaryotische Designerorganismen herzustellen, die im Hinblick auf nichtkodierende und destabilisierende Elemente optimiert sind?

Die Hefe *S. cerevisiae* hat sich als eukaryotisches Modellsystem etabliert, durch das gegründete *Synthetic Yeast 2.0 Project* soll erstmals ein synthetischer eukaryotischer Organismus konstruiert werden. 2014 konnte gezeigt werden, dass die generelle Synthese ganzer eukaryotischer Chromosomen möglich ist (Annaluru *et al.* 2014; Gibson and Venter 2014). Dabei werden nicht ganze Chromosomen synthetisiert und transplaniert, sondern die einzelnen Chromosomen werden nach und nach sukzessiv ersetzt. Dies hat den Vorteil, dass direkt ein Rückschluss auf die Funktionalität der veränderten Sequenz gefolgert werden kann. In den kommenden Jahren sollen weitere Chromosomen fertiggestellt und auch sogenannte Neochromosomen etabliert werden (Pennisi 2014).

Neochromosomen sind Chromosomen, die nicht nach einem natürlichen Vorbild geplant sind und zusätzlich zu den Hefechromosomen, ähnlich synthetischen sekundären Chromosomen in Bakterien, in *S. cerevisiae* etabliert werden. Ein Konzept für ein solches Neochromosom ist es, alle tRNA kodierenden Sequenzen auf einem Replikon isoliert zu kodieren, denn tRNA kodierende Sequenzen sind mitverantwortlich für genomische Instabilität (Mularoni *et al.* 2012; Nguyen *et al.* 2010; Pennisi 2014). Zudem ist es interessant festzustellen, wie viele tRNA Gene generell benötigt werden, denn es gibt 64 Codons, folglich müsste es höchstens 64 tRNAs geben, doch in *S. cerevisiae* sind 295 tRNA Gene aufgrund der Genomsequenz vorhergesagt (Chan and Lowe 2016). In Zukunft soll anhand des Neochromosoms mithilfe der im Rahmen des *Synthetic Yeast 2.0* etablierten SCRaMbLE (*Synthetic Chromosome Recombination and Modification by LoxP-mediated Evolution*) Methode evaluiert werden, wie viele tRNA Gene tatsächlich in der *S. cerevisiae* zum Überleben benötigt werden (Dymond and Boeke 2012; Jovicevic *et al.* 2014; Pennisi 2014). SCRaMbLe basiert dabei auf einer *in vivo* Rekombination, wobei jedes nicht essentielle Gen von Erkennungssequenzen (*loxP*) einer Rekombinase (Cre) flankiert ist. Dieses Verfahren

wurde bei dem Design aller Hefechromosomen sowie des tRNA Neochromosoms angewandt, um durch die Induktion von Cre/*lox* Rekombinationsereignisse zu induzieren und so zufällig Duplikationen, Deletionen und Umstrukturierungen innerhalb der Chromosomen zu erzeugen (Hoess *et al.* 1986; Dymond and Boeke 2012; Shen *et al.* 2016). Die vergleichende Auswertung verschiedener, durch SCRaMbLE generierter Stämme, dürfte einen Einblick in die Evolution der Genomstruktur von *S. cerevisiae* geben und vielleicht die Fragestellung beantworten, wie viele tRNA Gene in *S. cerevisiae* tatsächlich benötigt werden (Dymond *et al.* 2011; Shen *et al.* 2016). Das *Synthetic Yeast 2.0 Project* könnte in Zukunft viele interessante Entdeckungen im Bereich der eukaryotischen Chromosomenbiologie machen und Konzepte und Regeln für zukünftige eukaryotische Designerorganismen liefern.

Ein visionäres Projekt hat die Synthese des gesamten humanen Genoms zum Ziel (Boeke *et al.* 2016). Hierbei ist es nicht Ziel synthetische Menschen herzustellen, sondern durch umfassende genetische Modifikationen und Veränderungen der Genomsequenz tiefere Einblicke in die Funktion und Organisation humaner Zellen zu erlangen. Zudem sollen stabile Zelllinien für Forschung und Entwicklung entstehen, ähnlich wie es für die synthetischen *S. cerevisiae* postuliert wird. Es sollen, ähnlich wie im *Synthetic Yeast 2.0 Project*, repetitive und transposable DNA-Sequenzen entfernt werden, was zu einer gesteigerten genomischen Integrität führen könnte. Zudem ist es angedacht die Kopien von Tumorsuppressor-Genen, wie *p53*, im Genom zu erhöhen und zusätzlich durch Rekodierung der kodierenden Sequenzen die Anzahl an CpG-Inseln zu reduzieren was zu einer Reduktion an Mutationen führt. Des Weiteren sollen Gene, die möglicherweise einen negativen Einfluss haben können, entfernt werden insofern sie nicht essentiell sind, ein Beispiel dafür wären Gene von Prionen (Prusiner 1982). Das Projekt soll nicht mit der Fertigstellung eines einzelnen synthetischen Genoms abgeschlossen werden, vielmehr sollen viele verschiedene Genome nach dem Prinzip „*learning by doing*“ konstruiert werden.

Ein solches Projekt mag an dieser Stelle utopisch klingen, aber 1991 wurde das Projekt zur Sequenzierung des humanen Genoms gestartet. Dieses Projekt wurde anfangs sehr kontrovers diskutiert, jedoch konnte durch die Durchführung des Humanen Genom Projekts neben dem Erkenntnisgewinn die Sequenzierungstechnologie extrem weiterentwickelt werden (Leder 1990; Fuller *et al.* 2009; Mardis 2008). Sequenzierungsmethoden haben sich rasant durch Genomsequenzierungsprojekte entwickelt und eine ähnliche Entwicklung wäre für DNA-Synthesetechnologien durch das Projekt zur Synthese eines humanen Genoms denkbar (Kosuri and Church 2014; Carr and Church 2009; Boeke *et al.* 2016). Dementsprechend könnte auch die Synthese des kompletten humanen Genoms bald möglich sein, auch wenn es heute noch utopisch klingt. Rückblickend würde vermutlich kein Wissenschaftler das Projekt der

Sequenzierung des humanen Genoms als utopisch bezeichnen. Zudem ist heute moderne Molekularbiologie ohne das Vorhandensein der Genomsequenz des Modellorganismus fast undenkbar und falls ein neuer Modellorganismus etabliert werden soll, ist einer der ersten Schritte die Sequenzierung des Genoms.

4.3.2 Wirtschaftliche Anwendungen von Designerorganismen

In der Biotechnologie werden Organismen unter anderem dazu genutzt, um in Fermentern im industriellen Maßstab Produkte zu erzeugen. Hierbei ist es enorm wichtig, dass zum einen ein Maximum an Produkt erzielt werden kann und zum anderen die Produktionsstämme eine genomische Integrität aufweisen und resistent gegenüber Infektionen sind. Bakterien können von bakteriellen Viren, den Bakteriophagen, infiziert werden. Wird eine Kultur in einem industriellen Produktionsmaßstab infiziert, ist der wirtschaftliche Schaden beträchtlich (Jones *et al.* 2000a; Los *et al.* 2004). Die Studie zu *rE. coli*-57 zeigt auf, dass die Reduktion von 64 auf 57 Codons eine Resistenz gegenüber Bakteriophagen bewirken würde, denn infizierende Bakteriophagen sind abhängig von 64 Codons (Ostrov *et al.* 2016). Da Bakteriophagen den Translationsapparat der Wirtszelle nutzen und dieser auf 57 Codons basiert, können keine intakten Phagenproteine gebildet werden. Dies würde eine Phagenvermehrung und somit eine weitere Infektion innerhalb einer Kultur unterbinden. Ein solcher Produktionsstamm wäre für die Industrie von sehr wertvoll.

Ein weiteres, anwendungsbezogenes Beispiel ist es geeignete, sichere Designerorganismen für die Produktion im industriellen Maßstab herzustellen. Für die Biotechnologie ist die Zulassung eines Organismus als GRAS (*generally recognized as safe*) Organismus essentiell (Wessels *et al.* 2004). Um den GRAS Status zu erlangen, dürfen Organismen nicht pathogen sein und es dürfen keine toxischen oder antibiotischen Stoffe produziert werden. Designerorganismen könnten in Zukunft entsprechend der GRAS Richtlinien entworfen und hergestellt werden (Liu *et al.* 2015; Taguchi *et al.* 2015). Es wäre zudem möglich durch Rekodierung der Designerorganismen eine weitere Regulationsebene mit einzubeziehen, denn durch Veränderung der Codon-Verwendung in kodierenden DNA-Sequenzen könnten Organismen abhängig von einer nicht proteinogenen Aminosäure gemacht werden. Diese müsste als zusätzliche Komponente in das Medium hinzugegeben werden und würde bewirken, dass Organismen, die versehentlich in die Umwelt gelangt sind, dort nicht überlebensfähig wären (Simon and Ellington 2016; Rovner *et al.* 2015).

Die Synthetische Biologie ist der Übergang von einer erforschenden zu einer produzierenden (synthetisierenden) naturwissenschaftlichen Disziplin. Die vorliegende Arbeit zeigt auf, dass durch die Synthetische Biologie bereits vieles erreicht werden konnte, wie beispielsweise das erste Bakterium, das von einem synthetisch hergestellten Chromosom „kontrolliert“ wird (Gibson *et al.* 2010). Es konnte zudem ein minimaler Organismus basierend auf dem synthetisch hergestellten Organismus konstruiert werden, zudem ist es mittlerweile möglich ganze eukaryotische Chromosomen zu synthetisieren (Hutchison *et al.* 2016; Annaluru *et al.* 2014). In Zukunft werden weitere synthetische Chromosomen hergestellt werden und das dadurch entstehende Wissen wird es vielleicht in Zukunft sogar möglich machen individuell hergestellte synthetische Mikroorganismen für einen bestimmten Zweck zu generieren.

5 Zusammenfassung

Alle Funktionen einer jeden Zelle sind im Genom kodiert, dieses wird in jeder Zellteilung – egal ob ein oder mehrere Chromosomen – gleich auf die Tochterzellen verteilt. Die Integrität des Genoms ist für das Überleben eines jeden Organismus essentiell. Durch die Methoden der Synthetischen Biologie werden umfangreiche Veränderungen an Genomen durchgeführt bzw. ganze Chromosomen synthetisiert, wobei der Fokus meist auf den kodierenden Sequenzen liegt. Chromosomen sind aber mehr als eine Aneinanderreihung von Genen. Chromosomen benötigen Systeme zur Replikation, Segregation, Organisation und Reparatur, was häufig über Wechselwirkungen von Proteinen mit DNA-Sequenzmotiven geschieht. Die vorliegende Arbeit studiert solche, als *Chromosome Maintenance* System bezeichnete Prozesse anhand von synthetischen sekundären Chromosomen in *E. coli*. Können die resultierenden Ergebnisse zum Verständnis der DNA-Replikation in Bakterien beitragen?

Das im Rahmen dieser Arbeit etablierte synthetische sekundäre Chromosom (synVicII) repliziert in *E. coli* wie das sekundäre Chromosom in *V. cholerae*, auf dem es basiert. Das Design wurde nach einer initialen Charakterisierung weiter optimiert. Ein entscheidender Schritt war die Herstellung einer Kompatibilität von synVicII mit dem hierarchischen DNA-Assemblierungssystem MoClo. Parallel wurde eine Vorgehensweise etabliert, um hochvariable, lange DNA-Sequenzen zu generieren, in denen benutzerdefinierte DNA-Sequenzen ausgeschlossen werden können. Die Insertion dieser DNA-Sequenzen in synVicII ermöglicht es, synthetische sekundäre Chromosomen mit einer Größe von 100 kb zu konstruieren. Dadurch konnte im Rahmen der vorliegenden Arbeit erstmals die Interaktion der DNA-Segregation und *DNA mismatch* Reparatur *in vivo* durch ein Set von drei synthetischen sekundären Chromosomen analysiert werden. Beide Prozesse sind auf das Vorhandensein hemi-methylierter GATC-Sequenzen angewiesen und die Arbeit zeigt, dass durch eine strukturierte GATC-Anordnung ein differenzielles Binden der beiden Proteine SeqA und MutH erreicht werden kann.

Da die Funktionsweise von SeqA noch nicht vollständig verstanden ist, wurde außerdem das quantitative Verständnis von SeqA experimentell verbessert. Anhand der Daten wurde ein Modell der SeqA-Strukturen an den Replikationsgabeln generiert. Durch FRAP-Experimente konnte belegt werden, dass SeqA ein dynamisches Protein ist, welches zwischen zwei Bindeereignissen frei in der Zelle diffundiert. SeqA und Dam konkurrieren um die hemi-methylierten GATCs. Es konnte gezeigt werden, dass beide Proteine in einem konstanten Mengenverhältnis vorliegen. Dies könnte ein möglicher Aspekt zur Regulation der Re-methylierung der GATC-Sequenzen in *E. coli* sein. Die Ergebnisse der vorliegenden Arbeit tragen dadurch signifikant zum Verständnis der DNA-Replikation in Bakterien bei.

6 Summary

All functions of each cell are encoded in the genome, which is distributed equally to the daughter cells during each cell division. This applies to chromosomes, regardless the number of chromosomes. The integrity of the genome is essential for the survival of any organism. Using synthetic biology methods, extensive alterations of genomes or entire chromosomes can be synthesized, although the focus is mostly on the coding sequences. However, chromosomes are more than merely a sequential arrangement of genes. Chromosomes need systems for replication, segregation, organization, and repair, which is often done by interactions of proteins with DNA sequence motifs. The present study investigates such so-called chromosome maintenance systems using synthetic secondary chromosomes in *E. coli*. Can the results generate a better understanding of DNA replication in bacteria?

The synthetic secondary chromosome (synVicII) established in this work replicates in *E. coli* similarly to the secondary chromosome in *V. cholerae* on which it is based. After an initial characterization, the design of synVicII was further optimized. A crucial step was to generate compatibility of synVicII with the hierarchical DNA assembly system MoClo. In parallel, an approach was established to generate highly variable, long DNA sequences in which user-defined DNA motifs can be excluded. The insertion of these DNA sequences into synVicII allowed the construction of synthetic secondary chromosomes with a size of 100 kb. These chromosomes make it feasible for the first time to analyze the interaction of DNA segregation and DNA mismatch repair *in vivo* by a set of three synthetic secondary chromosomes. Both processes are dependent on the presence of hemi-methylated GATC sequences in *E. coli*. This work shows that a differentiated binding of the two responsible proteins, for the above processes, SeqA and MutH, can be achieved by an ordered GATC arrangement to allow comparative analysis.

Since the functionality of SeqA has not yet been fully understood, the quantitative understanding of SeqA was experimentally improved. Based on this data a model for SeqA structure variants at the replication fork was generated. It was demonstrated by FRAP experiments that SeqA is a dynamic protein that diffuses freely between two binding events within the cell. SeqA and Dam compete for the hemi-methylated GATCs. It was shown that both proteins are in a similar ratio to each other. This could be an aspect for how GATC re-methylation is regulated. The results of the present work contribute significantly to the understanding of DNA replication in bacteria.

7 Referenzen

Die hier aufgeführten Referenzen sind alle in der Arbeit verwendeten Quellen mit Ausnahme der Ergebniskapitel, diese besitzen am Ende des jeweiligen Kapitels und falls vorhanden im jeweiligen Appendix eine gesonderte Auflistung aller verwendeten Referenzen.

Abed RM, Dobretsov S, Sudesh K (2009) Applications of cyanobacteria in biotechnology. J Appl Microbiol 106 (1):1-12. doi:10.1111/j.1365-2672.2008.03918.x

Acharya S, Foster PL, Brooks P, Fishel R (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. Mol Cell 12 (1):233-246

Adams BL (2016) The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. ACS Synth Biol. doi:10.1021/acssynbio.6b00256

Agmon N, Mitchell LA, Cai Y, Ikushima S, Chuang J, Zheng A, Choi WJ, Martin JA, Caravelli K, Stracquadanio G, Boeke JD (2015) Yeast Golden Gate (yGG) for the Efficient Assembly of *S. cerevisiae* Transcription Units. ACS Synth Biol 4 (7):853-859. doi:10.1021/sb500372z

Alexander DL, Lilly J, Hernandez J, Romsdahl J, Troll CJ, Camps M (2014) Random mutagenesis by error-prone *pol* plasmid replication in *Escherichia coli*. Methods Mol Biol 1179:31-44. doi:10.1007/978-1-4939-1053-3_3

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol Rev 59 (1):143-169

Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, Yeluru A, Murugan S, Kandavelou K, Agier N, Fischer G, Yang K, Martin JA, Bilgel M, Bohutski P, Boulier KM, Capaldo BJ, Chang J, Charoen K, Choi WJ, Deng P, DiCarlo JE, Doong J, Dunn J, Feinberg JI, Fernandez C, Floria CE, Gladowski D, Hadidi P, Ishizuka I, Jabbari J, Lau CY, Lee PA, Li S, Lin D, Linder ME, Ling J, Liu J, London M, Ma H, Mao J, McDade JE, McMillan A, Moore AM, Oh WC, Ouyang Y, Patel R, Paul M, Paulsen LC, Qiu J, Rhee A, Rubashkin MG, Soh IY, Sotuyo NE, Srinivas V, Suarez A, Wong A, Wong R, Xie WR, Xu Y, Yu AT, Koszul R, Bader JS, Boeke JD, Chandrasegaran S (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344 (6179):55-58. doi:10.1126/science.1249252

Arnold FH (1993) Engineering proteins for nonnatural environments. FASEB J 7 (9):744-749

Arsenault PR, Wobbe KK, Weathers PJ (2008) Recent advances in artemisinin production through heterologous expression. Curr Med Chem 15 (27):2886-2896

Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D (2002) FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. Cell 108 (2):195-205

Avery OT, Macleod CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type Iii. J Exp Med 79 (2):137-158

Awan AR, Shaw WM, Ellis T (2016) Biosynthesis of therapeutic natural products using synthetic biology. Adv Drug Deliv Rev 105 (Pt A):96-106. doi:10.1016/j.addr.2016.04.010

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006 0008. doi:10.1038/msb4100050

Bach T, Krekling MA, Skarstad K (2003) Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. EMBO J 22 (2):315-323. doi:10.1093/emboj/cdg020

Bankier AT (2001) Shotgun DNA sequencing. Methods Mol Biol 167:89-100. doi:10.1385/1-59259-113-2:089

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh KC, Davis RW, Federspiel NA, Long SR (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc Natl Acad Sci U S A 98 (17):9883-9888. doi:10.1073/pnas.161294798

Barras F, Marinus MG (1988) Arrangement of Dam methylation sites (GATC) in the *Escherichia coli* chromosome. Nucleic Acids Res 16 (20):9821-9838

Barski A, Zhao K (2009) Genomic location analysis by ChIP-Seq. J Cell Biochem 107 (1):11-18. doi:10.1002/jcb.22077

Basler G, Grimbs S, Nikoloski Z (2012) Optimizing metabolic pathways by screening for feasible synthetic reactions. Biosystems 109 (2):186-191. doi:10.1016/j.biosystems.2012.04.007

Bates DB, Asai T, Cao Y, Chambers MW, Cadwell GW, Boye E, Kogoma T (1995) The DnaA box R4 in the minimal *oriC* is dispensable for initiation of *Escherichia coli* chromosome replication. Nucleic Acids Res 23 (16):3119-3125

Becker A, Barnett MJ, Capela D, Dondrup M, Kamp PB, Krol E, Linke B, Ruberg S, Runte K, Schroeder BK, Weidner S, Yurgel SN, Batut J, Long SR, Puhler A, Goesmann A (2009) A portal for rhizobial genomes: RhizoGATE integrates a *Sinorhizobium meliloti* genome annotation update with postgenome data. J Biotechnol 140 (1-2):45-50. doi:10.1016/j.jbiotec.2008.11.006

Becker J, Wittmann C (2016) Systems metabolic engineering of *Escherichia coli* for the heterologous production of high value molecules-a veteran at new shores. Curr Opin Biotechnol 42:178-188. doi:10.1016/j.copbio.2016.05.004

Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, Assad-Garcia N, Chuang RY, Carrera W, Moodie M, Algire MA, Phan Q, Alperovich N, Vashee S, Merryman C, Venter JC, Smith HO, Glass JI, Hutchison CA, 3rd (2010) Cloning whole bacterial genomes in yeast. Nucleic Acids Res 38 (8):2558-2569. doi:10.1093/nar/gkq119

Bennett ST, Barnes C, Cox A, Davies L, Brown C (2005) Toward the 1,000 dollars human genome. Pharmacogenomics 6 (4):373-382. doi:10.1517/14622416.6.4.373

Bernstein HC, Konopka A, Melnicki MR, Hill EA, Kucek LA, Zhang S, Shen G, Bryant DA, Beliaev AS (2014) Effect of mono- and dichromatic light quality on growth rates and photosynthetic performance of *Synechococcus sp*. PCC 7002. Front Microbiol 5:488. doi:10.3389/fmicb.2014.00488

Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, Dennis C, Grigoriev M, Allemand JF, Barre FX, Cornet F (2005) KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. EMBO J 24 (21):3770-3780. doi:10.1038/sj.emboj.7600835

Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. Cell 75 (2):351-361

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331):1453-1462

Bloch SE, Schmidt-Dannert C (2014) Construction of a chimeric biosynthetic pathway for the *de novo* biosynthesis of rosmarinic acid in *Escherichia coli*. Chembiochem 15 (16):2393-2401. doi:10.1002/cbic.201402275

Boeke JD, Church G, Hessel A, Kelley NJ, Arkin A, Cai Y, Carlson R, Chakravarti A, Cornish VW, Holt L, Isaacs FJ, Kuiken T, Lajoie M, Lessor T, Lunshof J, Maurano MT, Mitchell LA, Rine J, Rosser S,

Sanjana NE, Silver PA, Valle D, Wang H, Way JC, Yang L (2016) GENOME ENGINEERING. The Genome Project-Write. Science 353 (6295):126-127. doi:10.1126/science.aaf6850

Bolivar F, Rodriguez RL, Betlach MC, Boyer HW (1977) Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. Gene 2 (2):75-93

Boye E, Lobner-Olesen A, Skarstad K (1988) Timing of chromosomal replication in *Escherichia coli*. Biochim Biophys Acta 951 (2-3):359-364

Boye E, Lobner-Olesen A, Skarstad K (2000) Limiting DNA replication to once and only once. EMBO Rep 1 (6):479-483. doi:10.1093/embo-reports/kvd116

Boye E, Marinus MG, Lobner-Olesen A (1992) Quantitation of Dam methyltransferase in *Escherichia coli*. J Bacteriol 174 (5):1682-1685

Boye E, Stokke T, Kleckner N, Skarstad K (1996) Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. Proc Natl Acad Sci U S A 93 (22):12206-12211

Breitling R, Takano E (2016) Synthetic Biology of Natural Products. Cold Spring Harb Perspect Biol 8 (10). doi:10.1101/cshperspect.a023994

Brendler T, Austin S (1999) Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. EMBO J 18 (8):2304-2310. doi:10.1093/emboj/18.8.2304

Brezellec P, Hoebeke M, Hiet MS, Pasek S, Ferat JL (2006) DomainSieve: a protein domain-based screen that led to the identification of dam-associated genes with potential link to DNA maintenance. Bioinformatics 22 (16):1935-1941. doi:10.1093/bioinformatics/btl336

Buermans HP, den Dunnen JT (2014) Next generation sequencing technology: Advances and applications. Biochim Biophys Acta 1842 (10):1932-1941. doi:10.1016/j.bbadis.2014.06.015

Cabezon E, Ripoll-Rozada J, Pena A, de la Cruz F, Arechaga I (2015) Towards an integrated model of bacterial conjugation. FEMS Microbiol Rev 39 (1):81-95. doi:10.1111/1574-6976.12085

Callaway E (2016) 'Minimal' cell raises stakes in race to harness synthetic life. Nature 531 (7596):557-558. doi:10.1038/531557a

Campbell JL, Kleckner N (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from Dam methyltransferase following the passage of the chromosomal replication fork. Cell 62 (5):967-979

Camps M, Naukkarinen J, Johnson BP, Loeb LA (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. Proc Natl Acad Sci U S A 100 (17):9727-9732. doi:10.1073/pnas.1333928100

Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, Boistard P, Becker A, Boutry M, Cadieu E, Dreano S, Gloux S, Godrie T, Goffeau A, Kahn D, Kiss E, Lelaure V, Masuy D, Pohl T, Portetelle D, Puhler A, Purnelle B, Ramsperger U, Renard C, Thebault P, Vandenbol M, Weidner S, Galibert F (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. Proc Natl Acad Sci U S A 98 (17):9877-9882. doi:10.1073/pnas.161294398

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6 (8):1621-1624. doi:10.1038/ismej.2012.8

Carr PA, Church GM (2009) Genome engineering. Nat Biotechnol 27 (12):1151-1162. doi:10.1038/nbt.1590

Chan PP, Lowe TM (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res 44 (D1):D184-189. doi:10.1093/nar/gkv1309

Chandran S, Noskov VN, Segall-Shapiro TH, Ma L, Whiteis C, Lartigue C, Jores J, Vashee S, Chuang RY (2014) TREC-IN: gene knock-in genetic tool for genomes cloned in yeast. BMC Genomics 15:1180. doi:10.1186/1471-2164-15-1180

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

Cheng KC, Smith GR (1984) Recombinational hotspot activity of Chi-like sequences. J Mol Biol 180 (2):371-377

Cheng KC, Smith GR (1987) Cutting of chi-like sequences by the RecBCD enzyme of *Escherichia coli*. J Mol Biol 194 (4):747-750

Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP (1996) S-Adenosylmethionine and methylation. FASEB J 10 (4):471-480

Chung YS, Brendler T, Austin S, Guarne A (2009) Structural insights into the cooperative binding of SeqA to a tandem GATC repeat. Nucleic Acids Res 37 (10):3143-3152. doi:10.1093/nar/gkp151

Cobb RE, Ning JC, Zhao H (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. J Ind Microbiol Biotechnol 41 (2):469-477. doi:10.1007/s10295-013-1358-3

Cochrane G, Karsch-Mizrachi I, Takagi T (2016) The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res 44 (D1):D48-50. doi:10.1093/nar/gkv1323

Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids *in vitro*. Proc Natl Acad Sci U S A 70 (11):3240-3244

Cohen SN, Chang AC, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc Natl Acad Sci U S A 69 (8):2110-2114

Cooper DL, Lahue RS, Modrich P (1993) Methyl-directed mismatch repair is bidirectional. J Biol Chem 268 (16):11823-11829

Cooper EM, Muller H, Chandrasegaran S, Bader JS, Boeke JD (2012) The Build-a-Genome course. Methods Mol Biol 852:273-283. doi:10.1007/978-1-61779-564-0_20

Cressey D (2015) DNA repair sleuths win chemistry Nobel. Nature 526 (7573):307-308. doi:10.1038/nature.2015.18515

Crosland MW, Crozier RH (1986) *Myrmecia pilosula*, an Ant with Only One Pair of Chromosomes. Science 231 (4743):1278. doi:10.1126/science.231.4743.1278

Curtiss R, 3rd (1969) Bacterial conjugation. Annu Rev Microbiol 23:69-136. doi:10.1146/annurev.mi.23.100169.000441

Daghfous D, Chatti A, Hammami R, Landoulsi A (2009) Modeling of the full-length *Escherichia coli* SeqA protein, in complex with DNA. Pathol Biol (Paris) 57 (3):e61-66. doi:10.1016/j.patbio.2008.03.013

de Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, Platt D, Shapland EB, Serber Z, Dean J, Newman JD, Chandran SS (2014) Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 3 (2):97-106. doi:10.1021/sb4001992

Dhar MK, Sehgal S, Kaul S (2012) Structure, replication efficiency and fragility of yeast ARS elements. Res Microbiol 163 (4):243-253. doi:10.1016/j.resmic.2012.03.003

diCenzo G, Milunovic B, Cheng J, Finan TM (2013) The $tRNA^{arg}$ gene and *engA* are essential genes on the 1.7-Mb pSymB megaplasmid of *Sinorhizobium meliloti* and were translocated together from the chromosome in an ancestral strain. J Bacteriol 195 (2):202-212. doi:10.1128/JB.01758-12

diCenzo GC, Finan TM (2015) Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome. Mol Genet Genomics 290 (4):1345-1356. doi:10.1007/s00438-015-0998-6

Döhlemann J, Wagner M, Happel C, Sobetzko P, Becker A (submitted) A family of single copy repABC-type shuttle vectors showing chromosome-like maintenance in the alpha-proteobacterium *Sinorhizobium meliloti*. ACS Synth Biol

Döhlemann J, Brennecke M, Becker A (2016) Cloning-free genome engineering in *Sinorhizobium meliloti* advances applications of Cre/*loxP* site-specific recombination. J Biotechnol 233:160-170. doi:10.1016/j.jbiotec.2016.06.033

Dong T, Knoshaug EP, Pienkos PT, Laurens LML (2016) Lipid recovery from wet oleaginous microbial biomass for biofuel production: A critical review. Applied Energy 177:879-895. doi:10.1016/j.apenergy.2016.06.002

Drake JW (1991) Spontaneous mutation. Annu Rev Genet 25:125-146. doi:10.1146/annurev.ge.25.120191.001013

Duggin IG, Bell SD (2009) Termination structures in the *Escherichia coli* chromosome replication fork trap. J Mol Biol 387 (3):532-539. doi:10.1016/j.jmb.2009.02.027

Duggin IG, Wake RG, Bell SD, Hill TM (2008) The replication fork trap and termination of chromosome replication. Mol Microbiol 70 (6):1323-1333. doi:10.1111/j.1365-2958.2008.06500.x

Duigou S, Knudsen KG, Skovgaard O, Egan ES, Lobner-Olesen A, Waldor MK (2006) Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. J Bacteriol 188 (17):6419-6424. doi:10.1128/JB.00565-06

Duina AA, Miller ME, Keeney JB (2014) Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system. Genetics 197 (1):33-48. doi:10.1534/genetics.114.163188

Dybvig K, Voelker LL (1996) Molecular biology of mycoplasmas. Annu Rev Microbiol 50:25-57. doi:10.1146/annurev.micro.50.1.25

Dymond J, Boeke J (2012) The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. Bioeng Bugs 3 (3):168-171. doi:10.4161/bbug.19543

Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S, Bader JS, Boeke JD (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477 (7365):471-476. doi:10.1038/nature10403

Egan ES, Fogel MA, Waldor MK (2005) Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. Mol Microbiol 56 (5):1129-1138. doi:10.1111/j.1365-2958.2005.04622.x

Egan ES, Lobner-Olesen A, Waldor MK (2004) Synchronous replication initiation of the two *Vibrio cholerae* chromosomes. Curr Biol 14 (13):R501-502. doi:10.1016/j.cub.2004.06.036

Egan ES, Waldor MK (2003) Distinct replication requirements for the two *Vibrio cholerae* chromosomes. Cell 114 (4):521-530

El Karoui M, Biaudet V, Schbath S, Gruss A (1999) Characteristics of Chi distribution on different bacterial genomes. Res Microbiol 150 (9-10):579-587

Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 3 (2):109-118. doi:10.1039/c0ib00070a

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3 (11):e3647. doi:10.1371/journal.pone.0003647

Erdeniz N, Nguyen M, Deschenes SM, Liskay RM (2007) Mutations affecting a putative MutLalpha endonuclease motif impact multiple mismatch repair functions. DNA Repair (Amst) 6 (10):1463-1470. doi:10.1016/j.dnarep.2007.04.013

Errington J, Bath J, Wu LJ (2001) DNA transport in bacteria. Nat Rev Mol Cell Biol 2 (7):538-545. doi:10.1038/35080005

European Commission (2005) Synthetic Biology: Applying Engineering to Biology. Report of a NEST High-Level Expert Group. EUR 21796 Luxembourg: Office for Official Publications of the European Communities

Fabret C, Poncet S, Danielsen S, Borchert TV, Ehrlich SD, Janniere L (2000) Efficient gene targeted random mutagenesis in genetically stable *Escherichia coli* strains. Nucleic Acids Res 28 (21):E95

Finan TM, Weidner S, Wong K, Buhrmester J, Chain P, Vorholter FJ, Hernandez-Lucas I, Becker A, Cowie A, Gouzy J, Golding B, Puhler A (2001) The complete sequence of the 1,683-kb pSymB

megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. Proc Natl Acad Sci U S A 98 (17):9889-9894. doi:10.1073/pnas.161294698

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7 (6):461-465. doi:10.1038/nmeth.1459

Fong R, Vroom JA, Hu Z, Hutchinson CR, Huang J, Cohen SN, Kao CM (2007) Characterization of a large, stable, high-copy-number *Streptomyces* plasmid that requires stability and transfer functions for heterologous polyketide overproduction. Appl Environ Microbiol 73 (4):1296-1307. doi:10.1128/AEM.01888-06

Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2:45. doi:10.1038/msb4100090

Forster AC, Church GM (2007) Synthetic biology projects *in vitro*. Genome Res 17 (1):1-6. doi:10.1101/gr.5776007

Foster PL, Lee H, Popodi E, Townes JP, Tang H (2015) Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. Proc Natl Acad Sci U S A 112 (44):E5990-5999. doi:10.1073/pnas.1512136112

Frommer WB, Ninnemann O (1995) Heterologous Expression of Genes in Bacterial, Fungal, Animal, and Plant Cells. Annual Review of Plant Physiology and Plant Molecular Biology 46 (1):419-444. doi:10.1146/annurev.pp.46.060195.002223

Frost LS (1992) Bacterial conjugation: everybody's doin' it. Can J Microbiol 38 (11):1091-1096

Fu C, Donovan WP, Shikapwashya-Hasser O, Ye X, Cole RH (2014) Hot Fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. PLoS One 9 (12):e115318. doi:10.1371/journal.pone.0115318

Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss JA, Schwartz DC, Vezenov DV (2009) The challenges of sequencing by synthesis. Nat Biotechnol 27 (11):1013-1023. doi:10.1038/nbt.1585

Fuller RS, Funnell BE, Kornberg A (1984) The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. Cell 38 (3):889-900

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13 (12):840-852. doi:10.1038/nrg3306

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293 (5530):668-672. doi:10.1126/science.1060966

Geier GE, Modrich P (1979) Recognition sequence of the Dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease. J Biol Chem 254 (4):1408-1413

Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN, Osterman AL (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol 185 (19):5673-5684

Gerding MA, Chao MC, Davis BM, Waldor MK (2015) Molecular Dissection of the Essential Features of the Origin of Replication of the Second *Vibrio cholerae* Chromosome. MBio 6 (4):e00973. doi:10.1128/mBio.00973-15

Ghasemi Naghdi F, González González LM, Chan W, Schenk PM (2016) Progress on lipid extraction from wet algal biomass for biodiesel production. Microbial Biotechnology 9 (6):718-726. doi:10.1111/1751-7915.12360

Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res 37 (20):6984-6990. doi:10.1093/nar/gkp687

Gibson DG (2012) Oligonucleotide assembly in yeast to produce synthetic DNA fragments. Methods Mol Biol 852:11-21. doi:10.1007/978-1-61779-564-0_2

Gibson DG (2014) Programming biological operating systems: genome design, assembly and activation. Nat Methods 11 (5):521-526. doi:10.1038/nmeth.2894

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, 3rd, Smith HO (2008a) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319 (5867):1215-1220. doi:10.1126/science.1151721

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA, 3rd (2008b) 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 (51):20404-20409. doi:10.1073/pnas.0811011106

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329 (5987):52-56. doi:10.1126/science.1190719

Gibson DG, Venter JC (2014) Synthetic biology: Construction of a yeast chromosome. Nature 509 (7499):168-169. doi:10.1038/509168a

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6 (5):343-345. doi:10.1038/nmeth.1318

Gil R, Silva FJ, Pereto J, Moya A (2004) Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68 (3):518-537, table of contents. doi:10.1128/MMBR.68.3.518-537.2004

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 103 (2):425-430. doi:10.1073/pnas.0510013103

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17 (6):333-351. doi:10.1038/nrg.2016.49

Grainge I, Bregu M, Vazquez M, Sivanathan V, Ip SC, Sherratt DJ (2007) Unlinking chromosome catenanes *in vivo* by site-specific recombination. EMBO J 26 (19):4228-4238. doi:10.1038/sj.emboj.7601849

Greener A, Callahan M, Jerpseth B (1997) An efficient random mutagenesis technique using an *E. coli* mutator strain. Mol Biotechnol 7 (2):189-195. doi:10.1007/BF02761755

Grilley M, Griffith J, Modrich P (1993) Bidirectional excision in methyl-directed mismatch repair. J Biol Chem 268 (16):11830-11837

Guarne A, Brendler T, Zhao Q, Ghirlando R, Austin S, Yang W (2005) Crystal structure of a SeqA-N filament: implications for DNA replication and chromosome organization. EMBO J 24 (8):1502-1511. doi:10.1038/sj.emboj.7600634

Guarne A, Zhao Q, Ghirlando R, Yang W (2002) Insights into negative modulation of *E. coli* replication initiation from the structure of SeqA-hemimethylated DNA complex. Nat Struct Biol 9 (11):839-843. doi:10.1038/nsb857

Haellman V, Fussenegger M (2016) Synthetic Biology--Toward Therapeutic Solutions. J Mol Biol 428 (5 Pt B):945-962. doi:10.1016/j.jmb.2015.08.020

Haimovich AD, Muir P, Isaacs FJ (2015) Genomes by design. Nat Rev Genet 16 (9):501-516. doi:10.1038/nrg3956

Halpern D, Chiapello H, Schbath S, Robin S, Hennequet-Antier C, Gruss A, El Karoui M (2007) Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. PLoS Genet 3 (9):1614-1621. doi:10.1371/journal.pgen.0030153

Han JS, Kang S, Kim SH, Ko MJ, Hwang DS (2004) Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. J Biol Chem 279 (29):30236-30243. doi:10.1074/jbc.M402612200

Harris TJ, Emtage JS (1986) Expression of heterologous genes in *E. coli*. Microbiol Sci 3 (1):28-31

Harrison PW, Lower RP, Kim NK, Young JP (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol 18 (4):141-148. doi:10.1016/j.tim.2009.12.010

Hasan AM, Leach DR (2015) Chromosomal directionality of DNA mismatch repair in *Escherichia coli*. Proc Natl Acad Sci U S A 112 (30):9388-9393. doi:10.1073/pnas.1505370112

Hattman S, Brooks JE, Masurekar M (1978) Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco Dam) controlled by the *Escherichia coli dam* gene. J Mol Biol 126 (3):367-380

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406 (6795):477-483. doi:10.1038/35020000

Heinemann JA, Sprague GF, Jr. (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 340 (6230):205-209. doi:10.1038/340205a0

Helgesen E, Fossum-Raunehaug S, Saetre F, Schink KO, Skarstad K (2015) Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. Nucleic Acids Res 43 (5):2730-2743. doi:10.1093/nar/gkv146

Helmstetter CE, Thornton M, Zhou P, Bogan JA, Leonard AC, Grimwade JE (1997) Replication and segregation of a miniF plasmid during the division cycle of *Escherichia coli*. J Bacteriol 179 (4):1393-1399

Hendrickson H, Lawrence JG (2007) Mutational bias suggests that replication termination occurs near the *dif* site, not at *ter* sites. Mol Microbiol 64 (1):42-56. doi:10.1111/j.1365-2958.2007.05596.x

Herman GE, Modrich P (1981) *Escherichia coli* K-12 clones that overproduce Dam methylase are hypermutable. J Bacteriol 145 (1):644-646

Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 36 (1):39-56

Hill TM, Henson JM, Kuempel PL (1987) The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. Proc Natl Acad Sci U S A 84 (7):1754-1758

Hiraga S (1976) Novel F prime factors able to replicate in *Escherichia coli* Hfr strains. Proc Natl Acad Sci U S A 73 (1):198-202

Hoess RH, Wierzbicki A, Abremski K (1986) The role of the *loxP* spacer region in P1 site-specific recombination. Nucleic Acids Res 14 (5):2287-2300

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. Cell 121 (3):349-361. doi:10.1016/j.cell.2005.02.021

Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486 (7402):207-214. doi:10.1038/nature11234

Hutchison CA, 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC (2016) Design and synthesis of a minimal bacterial genome. Science 351 (6280):aad6253. doi:10.1126/science.aad6253

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. Science 286 (5447):2165-2169

Ilangovan A, Connery S, Waksman G (2015) Structural biology of the Gram-negative bacterial conjugation systems. Trends Microbiol 23 (5):301-310. doi:10.1016/j.tim.2015.02.012

Inamine JM, Ho KC, Loechel S, Hu PC (1990) Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. J Bacteriol 172 (1):504-506

Ip SC, Bregu M, Barre FX, Sherratt DJ (2003) Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. EMBO J 22 (23):6399-6407. doi:10.1093/emboj/cdg589

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. Science 333 (6040):348-353. doi:10.1126/science.1205822

Itaya M (1995) Toward a bacterial genome technology: integration of the *Escherichia coli* prophage lambda genome into the *Bacillus subtilis* 168 chromosome. Mol Gen Genet 248 (1):9-16

Itaya M, Fujita K, Ikeuchi M, Koizumi M, Tsuge K (2003) Stable positional cloning of long continuous DNA in the *Bacillus subtilis* genome vector. J Biochem 134 (4):513-519

Itaya M, Tanaka T (1997) Experimental surgery to create subgenomes of *Bacillus subtilis* 168. Proc Natl Acad Sci U S A 94 (10):5378-5382

Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. Proc Natl Acad Sci U S A 102 (44):15971-15976. doi:10.1073/pnas.0503868102

Itoh T, Tomizawa J (1979) Initiation of replication of plasmid ColE1 DNA by RNA polymerase, ribonuclease H, and DNA polymerase I. Cold Spring Harb Symp Quant Biol 43 Pt 1:409-417

Jewett MC, Forster AC (2010) Update on designing and building minimal cells. Curr Opin Biotechnol 21 (5):697-703. doi:10.1016/j.copbio.2010.06.008

Jha JK, Baek JH, Venkova-Canova T, Chattoraj DK (2012) Chromosome dynamics in multichromosome bacteria. Biochim Biophys Acta 1819 (7):826-829. doi:10.1016/j.bbagrm.2012.01.012

Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74:283-315. doi:10.1146/annurev.biochem.73.011303.073859

Johnson SL, Khiani A, Bishop-Lilly KA, Chapman C, Patel M, Verratti K, Teshima H, Munk AC, Bruce DC, Han CS, Xie G, Davenport KW, Chain P, Sozhamannan S (2015) Complete Genome Assemblies for Two Single-Chromosome *Vibrio cholerae* Isolates, Strains 1154-74 (Serogroup O49) and 10432-62 (Serogroup O27). Genome Announc 3 (3). doi:10.1128/genomeA.00462-15

Jones DT, Shirley M, Wu X, Keis S (2000a) Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. J Mol Microbiol Biotechnol 2 (1):21-26

Jones KL, Kim SW, Keasling JD (2000b) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab Eng 2 (4):328-338. doi:10.1006/mben.2000.0161

Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D (2013) Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. PLoS Genet 9 (8):e1003673. doi:10.1371/journal.pgen.1003673

Joska TM, Mashruwala A, Boyd JM, Belden WJ (2014) A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. J Microbiol Methods 100:46-51. doi:10.1016/j.mimet.2013.11.013

Jovicevic D, Blount BA, Ellis T (2014) Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. Bioessays 36 (9):855-860. doi:10.1002/bies.201400086

Juhas M, Reuss DR, Zhu B, Commichau FM (2014) *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. Microbiology 160 (Pt 11):2341-2351. doi:10.1099/mic.0.079376-0

Kadyrov FA, Dzantiev L, Constantin N, Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. Cell 126 (2):297-308. doi:10.1016/j.cell.2006.05.039

Kadyrov FA, Holmes SF, Arana ME, Lukianova OA, O'Donnell M, Kunkel TA, Modrich P (2007) *Saccharomyces cerevisiae* MutLalpha is a mismatch repair endonuclease. J Biol Chem 282 (51):37181-37190. doi:10.1074/jbc.M707617200

Kamada K, Horiuchi T, Ohsumi K, Shimamoto N, Morikawa K (1996) Structure of a replication-terminator protein complexed with DNA. Nature 383 (6601):598-603. doi:10.1038/383598a0

Kaneko S, Akioka M, Tsuge K, Itaya M (2005) DNA shuttling between plasmid vectors and a genome vector: systematic conversion and preservation of DNA libraries using the *Bacillus subtilis* genome (BGM) vector. J Mol Biol 349 (5):1036-1044. doi:10.1016/j.jmb.2005.04.041

Karas BJ, Jablanovic J, Irvine E, Sun L, Ma L, Weyman PD, Gibson DG, Glass JI, Venter JC, Hutchison CA, 3rd, Smith HO, Suzuki Y (2014) Transferring whole genomes from bacteria to yeast spheroplasts using entire bacterial cells to reduce DNA shearing. Nat Protoc 9 (4):743-750. doi:10.1038/nprot.2014.045

Karas BJ, Jablanovic J, Sun L, Ma L, Goldgof GM, Stam J, Ramon A, Manary MJ, Winzeler EA, Venter JC, Weyman PD, Gibson DG, Glass JI, Hutchison CA, 3rd, Smith HO, Suzuki Y (2013a) Direct transfer of whole genomes from bacteria to yeast. Nat Methods 10 (5):410-412. doi:10.1038/nmeth.2433

Karas BJ, Molparia B, Jablanovic J, Hermann WJ, Lin YC, Dupont CL, Tagwerker C, Yonemoto IT, Noskov VN, Chuang RY, Allen AE, Glass JI, Hutchison CA, 3rd, Smith HO, Venter JC, Weyman PD (2013b) Assembly of eukaryotic algal chromosomes in yeast. J Biol Eng 7 (1):30. doi:10.1186/1754-1611-7-30

Karas BJ, Suzuki Y, Weyman PD (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. Chromosome Res 23 (1):57-68. doi:10.1007/s10577-014-9455-3

Karas BJ, Tagwerker C, Yonemoto IT, Hutchison CA, 3rd, Smith HO (2012) Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. ACS Synth Biol 1 (1):22-28. doi:10.1021/sb200013j

Keasling JD (2012) Synthetic biology and the development of tools for metabolic engineering. Metab Eng 14 (3):189-195. doi:10.1016/j.ymben.2012.01.004

Kleckner N, Fisher JK, Stouf M, White MA, Bates D, Witz G (2014) The bacterial nucleoid: nature, dynamics and sister segregation. Curr Opin Microbiol 22:127-137. doi:10.1016/j.mib.2014.10.001

Kolodner R (1996) Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev 10 (12):1433-1442

Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet 1:99-116. doi:10.1146/annurev.genom.1.1.99

Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1 (2):127-136. doi:10.1038/nrmicro751

Korlach J, Turner SW (2012) Going beyond five bases in DNA sequencing. Curr Opin Struct Biol 22 (3):251-261. doi:10.1016/j.sbi.2012.04.002

Kosinski J, Plotz G, Guarne A, Bujnicki JM, Friedhoff P (2008) The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. J Mol Biol 382 (3):610-627. doi:10.1016/j.jmb.2008.06.056

Kosuri S, Church GM (2014) Large-scale *de novo* DNA synthesis: technologies and applications. Nat Methods 11 (5):499-507. doi:10.1038/nmeth.2918

Krawiec S, Riley M (1990) Organization of the bacterial chromosome. Microbiol Rev 54 (4):502-539

Krol E, Becker A (2014) Rhizobial homologs of the fatty acid transporter FadL facilitate perception of long-chain acyl-homoserine lactone signals. Proc Natl Acad Sci U S A 111 (29):10702-10707. doi:10.1073/pnas.1404929111

Kroll J, Klinter S, Schneider C, Voss I, Steinbuchel A (2010) Plasmid addiction systems: perspectives and applications in biotechnology. Microb Biotechnol 3 (6):634-657. doi:10.1111/j.1751-7915.2010.00170.x

Krueger F, Kreck B, Franke A, Andrews SR (2012) DNA methylome analysis using short bisulfite sequencing data. Nat Methods 9 (2):145-151. doi:10.1038/nmeth.1828

Kuempel PL, Henson JM, Dircks L, Tecklenburg M, Lim DF (1991) *dif*, a *recA*-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. New Biol 3 (8):799-811

Kunkel TA, Erie DA (2005) DNA mismatch repair. Annu Rev Biochem 74:681-710. doi:10.1146/annurev.biochem.74.082803.133243

Labroussaa F, Lebaudy A, Baby V, Gourgues G, Matteau D, Vashee S, Sirand-Pugnet P, Rodrigue S, Lartigue C (2016) Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. Nucleic Acids Res 44 (17):8501-8511. doi:10.1093/nar/gkw688

Lahue RS, Au KG, Modrich P (1989) DNA mismatch correction in a defined system. Science 245 (4914):160-164

Lahue RS, Su SS, Modrich P (1987) Requirement for d(GATC) sequences in *Escherichia coli* mutHLS mismatch correction. Proc Natl Acad Sci U S A 84 (6):1482-1486

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342 (6156):357-360. doi:10.1126/science.1241459

Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW (2015) Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15 (2):141-161. doi:10.1007/s10142-015-0433-4

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L,

Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409 (6822):860-921. doi:10.1038/35057062

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. Science 317 (5838):632-638. doi:10.1126/science.1144622

Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, Noskov VN, Denisova EA, Gibson DG, Assad-Garcia N, Alperovich N, Thomas DW, Merryman C, Hutchison CA, 3rd, Smith HO, Venter JC, Glass JI (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 325 (5948):1693-1696. doi:10.1126/science.1173759

Leder P (1990) Can the human genome project be saved from its critics... and itself? Cell 63 (1):1-3

Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A 109 (41):E2774-2783. doi:10.1073/pnas.1210309109

Lee HH, Ostrov N, Wong BG, Gold MA, Khalil A, Church GM (2016). doi:10.1101/058487

Lehman IR (1974) DNA ligase: structure, mechanism, and function. Science 186 (4166):790-797

Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA (2016) Mismatch repair in Gram-positive bacteria. Res Microbiol 167 (1):4-12. doi:10.1016/j.resmic.2015.08.006

Leonard AC, Helmstetter CE (1986) Cell cycle-specific replication of *Escherichia coli* minichromosomes. Proc Natl Acad Sci U S A 83 (14):5101-5105

Levy O, Ptacin JL, Pease PJ, Gore J, Eisen MB, Bustamante C, Cozzarelli NR (2005) Identification of oligonucleotide sequences that direct the movement of the *Escherichia coli* FtsK translocase. Proc Natl Acad Sci U S A 102 (49):17618-17623. doi:10.1073/pnas.0508932102

Li GM (2008) Mechanisms and functions of DNA mismatch repair. Cell Res 18 (1):85-98. doi:10.1038/cr.2007.115

Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157 (3):624-635. doi:10.1016/j.cell.2014.02.033

Liang X, Baek CH, Katzen F (2013) *Escherichia coli* with two linear chromosomes. ACS Synth Biol 2 (12):734-740. doi:10.1021/sb400079u

Liu B, Alberts BM (1995) Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science 267 (5201):1131-1137

Liu L, Guan N, Li J, Shin HD, Du G, Chen J (2015) Development of GRAS strains for nutraceutical production using systems and synthetic biology approaches: advances and prospects. Crit Rev Biotechnol:1-12. doi:10.3109/07388551.2015.1121461

Llosa M, de la Cruz F (2005) Bacterial conjugation: a potential tool for genomic engineering. Res Microbiol 156 (1):1-6. doi:10.1016/j.resmic.2004.07.008

Lobner-Olesen A (1999) Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control. EMBO J 18 (6):1712-1721. doi:10.1093/emboj/18.6.1712

Los M, Czyz A, Sell E, Wegrzyn A, Neubauer P, Wegrzyn G (2004) Bacteriophage contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? J Appl Genet 45 (1):111-120

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in *E. coli*. Cell 77 (3):413-426

Lu X, Hirata H, Yamaji Y, Ugaki M, Namba S (2001) Random mutagenesis in a plant viral genome using a DNA repair-deficient mutator *Escherichia coli* strain. J Virol Methods 94 (1-2):37-43

MacDonald IC, Deans TL (2016) Tools and applications in synthetic biology. Adv Drug Deliv Rev 105 (Pt A):20-34. doi:10.1016/j.addr.2016.08.008

Mackenzie C, Simmons AE, Kaplan S (1999) Multiple chromosomes in bacteria. The yin and yang of *trp* gene localization in *Rhodobacter sphaeroides* 2.4.1. Genetics 153 (2):525-538

Malpartida F, Hopwood DA (1984) Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature 309 (5967):462-464

Mardis ER (2006) Anticipating the 1,000 dollar genome. Genome Biol 7 (7):112. doi:10.1186/gb-2006-7-7-112

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387-402. doi:10.1146/annurev.genom.9.081307.164359

Marinus MG (1987) DNA methylation in *Escherichia coli*. Annu Rev Genet 21:113-131. doi:10.1146/annurev.ge.21.120187.000553

Marinus MG (2010) DNA methylation and mutator genes in *Escherichia coli* K-12. Mutat Res 705 (2):71-76. doi:10.1016/j.mrrev.2010.05.001

Marinus MG, Lobner-Olesen A (2014) DNA Methylation. EcoSal Plus 6 (1). doi:10.1128/ecosalplus.ESP-0003-2013

Marinus MG, Morris NR (1973) Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. J Bacteriol 114 (3):1143-1150

Marinus MG, Poteete A, Arraj JA (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene 28 (1):123-125

Matson SW (1986) *Escherichia coli* helicase II (*urvD* gene product) translocates unidirectionally in a 3' to 5' direction. J Biol Chem 261 (22):10169-10175

McCarthy D, Minner C, Bernstein H, Bernstein C (1976) DNA elongation rates and growing point distributions of wild-type phage T4 and a DNA-delay amber mutant. J Mol Biol 106 (4):963-981

Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, Boccard F, Espeli O (2008) The MatP/*matS* site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. Cell 135 (3):475-485. doi:10.1016/j.cell.2008.08.031

Messer W (2002) The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. FEMS Microbiol Rev 26 (4):355-374

Messerschmidt S, Schindler D, Zumkeller CM, Kemter F, Zimmer N, Waldminghaus T (under revision) Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli.* Front Bioeng Biotech

Messerschmidt SJ, Kemter FS, Schindler D, Waldminghaus T (2015) Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in *Vibrio cholerae*. Biotechnol J 10 (2):302-314. doi:10.1002/biot.201400031

Messerschmidt SJ, Waldminghaus T (2014) Dynamic organization: chromosome domains in *Escherichia coli*. J Mol Microbiol Biotechnol 24 (5-6):301-315. doi:10.1159/000369098

Milbredt S, Farmani N, Sobetzko P, Waldminghaus T (2016) DNA Replication in Engineered *Escherichia coli* Genomes with Extra Replication Origins. ACS Synth Biol 5 (10):1167-1176. doi:10.1021/acssynbio.6b00064

Montague MG, Lartigue C, Vashee S (2012) Synthetic genomics: potential and limitations. Curr Opin Biotechnol 23 (5):659-665. doi:10.1016/j.copbio.2012.01.014

Moore SJ, Lai HE, Kelwick RJ, Chee SM, Bell DJ, Polizzi KM, Freemont PS (2016) EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. ACS Synth Biol 5 (10):1059-1069. doi:10.1021/acssynbio.6b00031

Mori T, Binder B, Johnson CH (1996) Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. Proc Natl Acad Sci U S A 93 (19):10183-10188

Mularoni L, Zhou Y, Bowen T, Gangadharan S, Wheelan SJ, Boeke JD (2012) Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. Genome Res 22 (4):693-703. doi:10.1101/gr.129460.111

Murray H (2016) Connecting chromosome replication with cell growth in bacteria. Curr Opin Microbiol 34:13-17. doi:10.1016/j.mib.2016.07.013

Musialek MW, Rybaczek D (2015) Behavior of replication origins in Eukaryota - spatio-temporal dynamics of licensing and firing. Cell Cycle 14 (14):2251-2264. doi:10.1080/15384101.2015.1056421

Nandakumar J, Nair PA, Shuman S (2007) Last stop on the road to repair: structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. Mol Cell 26 (2):257-271. doi:10.1016/j.molcel.2007.02.026

Nguyen VC, Clelland BW, Hockman DJ, Kujat-Choy SL, Mewhort HE, Schultz MC (2010) Replication stress checkpoint signaling controls tRNA gene transcription. Nat Struct Mol Biol 17 (8):976-981. doi:10.1038/nsmb.1857

Nielsen J (2011) Biofuels: chimeric synthetic pathways. Nat Chem Biol 7 (4):195-196. doi:10.1038/nchembio.548

Nolivos S, Touzain F, Pages C, Coddeville M, Rousseau P, El Karoui M, Le Bourgeois P, Cornet F (2012) Co-evolution of segregation guide DNA motifs and the FtsK translocase in bacteria: identification of the atypical *Lactococcus lactis* KOPS motif. Nucleic Acids Res 40 (12):5535-5545. doi:10.1093/nar/gks171

Noskov VN, Chuang RY, Gibson DG, Leem SH, Larionov V, Kouprina N (2011) Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. Nat Protoc 6 (1):89-96. doi:10.1038/nprot.2010.174

Noskov VN, Karas BJ, Young L, Chuang RY, Gibson DG, Lin YC, Stam J, Yonemoto IT, Suzuki Y, Andrews-Pfannkoch C, Glass JI, Smith HO, Hutchison CA, 3rd, Venter JC, Weyman PD (2012) Assembly of large, high G+C bacterial DNA fragments in yeast. ACS Synth Biol 1 (7):267-273. doi:10.1021/sb3000194

Noskov VN, Segall-Shapiro TH, Chuang RY (2010) Tandem repeat coupled with endonuclease cleavage (TREC): a seamless modification tool for genome engineering in yeast. Nucleic Acids Res 38 (8):2570-2576. doi:10.1093/nar/gkq099

O'Donnell M, Langston L, Stillman B (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harb Perspect Biol 5 (7). doi:10.1101/cshperspect.a010108

Ochman H (2002) Bacterial evolution: chromosome arithmetic and geometry. Curr Biol 12 (12):R427-428

Odsbu I, Klungsoyr HK, Fossum S, Skarstad K (2005) Specific N-terminal interactions of the *Escherichia coli* SeqA protein are required to form multimers that restrain negative supercoils and form foci. Genes Cells 10 (11):1039-1049. doi:10.1111/j.1365-2443.2005.00898.x

Ogden GB, Pratt MJ, Schaechter M (1988) The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. Cell 54 (1):127-135

Ohno S (1972) So much "junk" DNA in our genome. Brookhaven Symp Biol 23:366-370

Ohtani N, Hasegawa M, Sato M, Tomita M, Kaneko S, Itaya M (2012) Serial assembly of *Thermus* megaplasmid DNA in the genome of *Bacillus subtilis* 168: a BAC-based domino method applied to DNA with a high GC content. Biotechnol J 7 (7):867-876. doi:10.1002/biot.201100396

Okada K, Iida T, Kita-Tsukamoto K, Honda T (2005) *Vibrios* commonly possess two chromosomes. J Bacteriol 187 (2):752-757. doi:10.1128/JB.187.2.752-757.2005

Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. Nucleic Acids Res 25 (2):451-452

Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G, Gonzales A, Wanner BL, Norville JE, Lajoie MJ, Church GM (2016) Design, synthesis, and testing toward a 57-codon genome. Science 353 (6301):819-822. doi:10.1126/science.aaf3639

Ozaki S, Katayama T (2009) DnaA structure, function, and dynamics in the initiation at the chromosomal origin. Plasmid 62 (2):71-82. doi:10.1016/j.plasmid.2009.06.003

Padilla-Vaca F, Anaya-Velazquez F, Franco B (2015) Synthetic biology: Novel approaches for microbiology. Int Microbiol 18 (2):71-84. doi:10.2436/20.1501.01.236

Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10 (10):669-680. doi:10.1038/nrg2641

Paszkiewicz K, Studholme DJ (2010) *De novo* assembly of short sequence reads. Brief Bioinform 11 (5):457-472. doi:10.1093/bib/bbq020

Pennisi E (2010) Genomics. Synthetic genome brings new life to bacterium. Science 328 (5981):958-959. doi:10.1126/science.328.5981.958

Pennisi E (2014) Building the ultimate yeast genome. Science 343 (6178):1426-1429. doi:10.1126/science.343.6178.1426

Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. Science 312 (5776):1044-1046. doi:10.1126/science.1126439

Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. Science 216 (4542):136-144

Radman M (2016) Mismatch repair earns Nobel Prize in Chemistry 2015 to Paul Modrich for a biochemical tour de force. DNA Repair (Amst) 37:A22-28. doi:10.1016/j.dnarep.2015.12.004

Rasmussen T, Jensen RB, Skovgaard O (2007) The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. EMBO J 26 (13):3124-3131. doi:10.1038/sj.emboj.7601747

Raymond CK, Pownder TA, Sexson SL (1999) General method for plasmid construction using homologous recombination. Biotechniques 26 (1):134-138, 140-131

Razin S, Yogev D, Naot Y (1998) Molecular biology and pathogenicity of mycoplasmas. Microbiol Mol Biol Rev 62 (4):1094-1156

Reinert K, Langmead B, Weese D, Evers DJ (2015) Alignment of Next-Generation Sequencing Reads. Annu Rev Genomics Hum Genet 16:133-151. doi:10.1146/annurev-genom-090413-025358

Renkawitz J, Lademann CA, Jentsch S (2014) Mechanisms and principles of homology search during recombination. Nat Rev Mol Cell Biol 15 (6):369-383. doi:10.1038/nrm3805

Resnick MA, Nilsson-Tillgren T (1990) DNA homology and chromosome stability: a sensitive yeast genetic system for identifying double-stranded DNA damage. Prog Clin Biol Res 340B:363-369

Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58 (4):586-597. doi:10.1016/j.molcel.2015.05.004

Reyes-Lamothe R, Nicolas E, Sherratt DJ (2012) Chromosome replication and segregation in bacteria. Annu Rev Genet 46:121-143. doi:10.1146/annurev-genet-110711-155421

Riva A, Delorme MO, Chevalier T, Guilhot N, Henaut C, Henaut A (2004) Characterization of the GATC regulatory network in *E. coli*. BMC Genomics 5 (1):48. doi:10.1186/1471-2164-5-48

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440 (7086):940-943. doi:10.1038/nature04640

Rokke G, Korvald E, Pahr J, Oyas O, Lale R (2014) BioBrick assembly standards and techniques and associated software tools. Methods Mol Biol 1116:1-24. doi:10.1007/978-1-62703-764-8_1

Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, Isaacs FJ (2015) Recoded organisms engineered to depend on synthetic amino acids. Nature 518 (7537):89-93. doi:10.1038/nature14095

Russell DW, Zinder ND (1987) Hemimethylation prevents DNA replication in *E. coli*. Cell 50 (7):1071-1079

Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307 (1-2):249-264. doi:10.1007/s11010-007-9603-6

Saint-Dic D, Kehrl J, Frushour B, Kahng LS (2008) Excess SeqA leads to replication arrest and a cell division defect in *Vibrio cholerae*. J Bacteriol 190 (17):5870-5878. doi:10.1128/JB.00479-08

Sanchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, Grainger DC (2010) Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12. MBio 1 (1). doi:10.1128/mBio.00012-10

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74 (12):5463-5467

Sarsekeyeva F, Zayadan BK, Usserbaeva A, Bedbenov VS, Sinetova MA, Los DA (2015) Cyanofuels: biofuels from cyanobacteria. Reality and perspectives. Photosynth Res 125 (1-2):329-340. doi:10.1007/s11120-015-0103-3

Sathish A, Marlar T, Sims RC (2015) Optimization of a wet microalgal lipid extraction procedure for improved lipid recovery for biofuel and bioproduct production. Bioresour Technol 193:15-24. doi:10.1016/j.biortech.2015.06.052

Sauer U (2001) Evolutionary engineering of industrially important microbial phenotypes. Adv Biochem Eng Biotechnol 73:129-169

Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB (2011) The real cost of sequencing: higher than you think! Genome Biol 12 (8):125. doi:10.1186/gb-2011-12-8-125

Schaaper RM (1993) Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. J Biol Chem 268 (32):23762-23765

Schaaper RM, Dunn RL (1991) Spontaneous mutation in the *Escherichia coli lacI* gene. Genetics 129 (2):317-326

Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synth Biol. doi:10.1021/acssynbio.6b00089

Schindler D, Waldminghaus T (2015) Synthetic chromosomes. FEMS Microbiol Rev 39 (6):871-891. doi:10.1093/femsre/fuv030

Schindler D, Waldminghaus T (in preparation) Using experimental chromosome construction to study functional interactions between segregation and DNA mismatch repair in Escherichia coli.

Schoolnik GK, Yildiz FH (2000) The complete genome sequence of *Vibrio cholerae*: a tale of two chromosomes and of two lifestyles. Genome Biol 1 (3):REVIEWS1016. doi:10.1186/gb-2000-1-3-reviews1016

Schumacher MA, Funnell BE (2005) Structures of ParB bound to DNA reveal mechanism of partition complex formation. Nature 438 (7067):516-519. doi:10.1038/nature04149

Serrano L (2007) Synthetic biology: promises and challenges. Mol Syst Biol 3:158. doi:10.1038/msb4100202

Service RF (2006) Gene sequencing. The race for the $1000 genome. Science 311 (5767):1544-1546. doi:10.1126/science.311.5767.1544

Service RF (2016) SYNTHETIC BIOLOGY. Synthetic microbe has fewest genes, but many mysteries. Science 351 (6280):1380-1381. doi:10.1126/science.351.6280.1380

Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, Xue Y, Cai Y, Chen T, Dymond JS, Kang K, Gong J, Zeng X, Zhang Y, Li Y, Feng Q, Xu X, Wang J, Yang H, Boeke JD, Bader JS (2016) SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. Genome Res 26 (1):36-49. doi:10.1101/gr.193433.115

Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res 23 (11):1990-1996

Simon AJ, Ellington AD (2016) Recent advances in synthetic biosafety. F1000Res 5. doi:10.12688/f1000research.8365.1

Simon R, O'Connell M, Labes M, Puhler A (1986) Plasmid vectors for the genetic analysis and manipulation of rhizobia and other gram-negative bacteria. Methods Enzymol 118:640-659

Simon R, Priefer U, Pühler A (1983) A Broad Host Range Mobilization System for *In Vivo* Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Bio/Technology 1 (9):784-791. doi:10.1038/nbt1183-784

Simoneau P, Li CM, Loechel S, Wenzel R, Herrmann R, Hu PC (1993) Codon reading scheme in *Mycoplasma pneumoniae* revealed by the analysis of the complete set of tRNA genes. Nucleic Acids Res 21 (21):4967-4974

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15 (2):121-132. doi:10.1038/nrg3642

Skarstad K, Katayama T (2013) Regulating DNA replication in bacteria. Cold Spring Harb Perspect Biol 5 (4):a012922. doi:10.1101/cshperspect.a012922

Skarstad K, Lobner-Olesen A (2003) Stable co-existence of separate replicons in *Escherichia coli* is dependent on once-per-cell-cycle initiation. EMBO J 22 (1):140-150. doi:10.1093/emboj/cdg003

Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N (1995) *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell 82 (6):927-936

Sleator RD (2010) The story of *Mycoplasma mycoides* JCVI-syn1.0: the forty million dollar microbe. Bioeng Bugs 1 (4):229-230. doi:10.4161/bbug.1.4.12465

Sleator RD (2016) JCVI-syn3.0 - A synthetic genome stripped bare! Bioengineered 7 (2):53-56. doi:10.1080/21655979.2016.1175847

Sliva A, Yang H, Boeke JD, Mathews DJ (2015) Freedom and Responsibility in Synthetic Genomics: The Synthetic Yeast Project. Genetics 200 (4):1021-1028. doi:10.1534/genetics.115.176370

Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL (1981) Structure of chi hotspots of generalized recombination. Cell 24 (2):429-436

Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of E. coli. Nature 387 (6634):703-705. doi:10.1038/42701

Sobetzko P, Jelonek L, Strickert M, Han W, Goesmann A, Waldminghaus T (2016) DistAMo: A Web-Based Tool to Characterize DNA-Motif Distribution on Bacterial Chromosomes. Front Microbiol 7:283. doi:10.3389/fmicb.2016.00283

Sourice S, Biaudet V, El Karoui M, Ehrlich SD, Gruss A (1998) Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site. Mol Microbiol 27 (5):1021-1029

Stokke C, Flatten I, Skarstad K (2012) An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. PLoS One 7 (2):e30981. doi:10.1371/journal.pone.0030981

Stouf M, Meile JC, Cornet F (2013) FtsK actively segregates sister chromosomes in *Escherichia coli*. Proc Natl Acad Sci U S A 110 (27):11157-11162. doi:10.1073/pnas.1304080110

Stranneheim H, Lundeberg J (2012) Stepping stones in DNA sequencing. Biotechnol J 7 (9):1063-1073. doi:10.1002/biot.201200153

Strzelczyk B, Slominska-Wojewodzka M, Wegrzyn G, Wegrzyn A (2003) Non-random distribution of GATC sequences in regions of promoters stimulated by the SeqA protein of *Escherichia coli*. Acta Biochim Pol 50 (4):941-945. doi:035004941

Su SS, Modrich P (1986) *Escherichia coli mutS*-encoded protein binds to mismatched DNA base pairs. Proc Natl Acad Sci U S A 83 (14):5057-5061

Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K, Jung S, Fulton RS, Ly A, McGrath S, Haub K, Wiggins JL, Storton D, Matese JC, Parsons L, Chang WJ, Bowen MS, Stover NA, Jones TA, Eddy SR, Herrick GA, Doak TG, Wilson RK, Mardis ER, Landweber LF (2013) The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol 11 (1):e1001473. doi:10.1371/journal.pbio.1001473

Szybalski W, Kim SC, Hasan N, Podhajska AJ (1991) Class-IIS restriction enzymes--a review. Gene 100:13-26

Szyf M, Avraham-Haetzni K, Reifman A, Shlomai J, Kaplan F, Oppenheim A, Razin A (1984) DNA methylation pattern is determined by the intracellular level of the methylase. Proc Natl Acad Sci U S A 81 (11):3278-3282

Taghbalout A, Landoulsi A, Kern R, Yamazoe M, Hiraga S, Holland B, Kohiyama M, Malki A (2000) Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E. coli in vitro*. Genes Cells 5 (11):873-884

Taguchi S, Ooi T, Mizuno K, Matsusaki H (2015) Advances and needs for endotoxin-free production strains. Appl Microbiol Biotechnol 99 (22):9349-9360. doi:10.1007/s00253-015-6947-9

Tagwerker C, Dupont CL, Karas BJ, Ma L, Chuang RY, Benders GA, Ramon A, Novotny M, Montague MG, Venepally P, Brami D, Schwartz A, Andrews-Pfannkoch C, Gibson DG, Glass JI, Smith HO, Venter JC, Hutchison CA, 3rd (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. Nucleic Acids Res 40 (20):10375-10383. doi:10.1093/nar/gks823

Tanaka T, Ogura M (1998) A novel Bacillus natto plasmid pLS32 capable of replication in *Bacillus subtilis*. FEBS Lett 422 (2):243-246

Taylor AF, Schultz DW, Ponticelli AS, Smith GR (1985) RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. Cell 41 (1):153-163

Tomizawa J, Selzer G (1979) Initiation of DNA synthesis in *Escherichia coli*. Annu Rev Biochem 48:999-1034. doi:10.1146/annurev.bi.48.070179.005031

Touzain F, Petit MA, Schbath S, El Karoui M (2011) DNA motifs that sculpt the bacterial chromosome. Nat Rev Microbiol 9 (1):15-26. doi:10.1038/nrmicro2477

Tsarmpopoulos I, Gourgues G, Blanchard A, Vashee S, Jores J, Lartigue C, Sirand-Pugnet P (2016) In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9. ACS Synth Biol 5 (1):104-109. doi:10.1021/acssynbio.5b00196

Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A, Jeltsch A (2002) The *Escherichia coli* Dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol 319 (5):1085-1096. doi:10.1016/S0022-2836(02)00371-6

Val ME, Kennedy SP, El Karoui M, Bonne L, Chevalier F, Barre FX (2008) FtsK-dependent dimer resolution on multiple chromosomes in the pathogen *Vibrio cholerae*. PLoS Genet 4 (9):e1000201. doi:10.1371/journal.pgen.1000201

Val ME, Kennedy SP, Soler-Bistue AJ, Barbe V, Bouchier C, Ducos-Galand M, Skovgaard O, Mazel D (2014a) Fuse or die: how to survive the loss of Dam in *Vibrio cholerae*. Mol Microbiol 91 (4):665-678. doi:10.1111/mmi.12483

Val ME, Marbouty M, de Lemos Martins F, Kennedy SP, Kemble H, Bland MJ, Possoz C, Koszul R, Skovgaard O, Mazel D (2016) A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci Adv 2 (4):e1501914. doi:10.1126/sciadv.1501914

Val ME, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D (2012) Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues. PLoS Genet 8 (1):e1002472. doi:10.1371/journal.pgen.1002472

Val ME, Soler-Bistue A, Bland MJ, Mazel D (2014b) Management of multipartite genomes: the *Vibrio cholerae* model. Curr Opin Microbiol 22:120-126. doi:10.1016/j.mib.2014.10.003

van Leeuwen J, Andrews B, Boone C, Tan G (2015a) Construction of Multifragment Plasmids by Homologous Recombination in Yeast. Cold Spring Harb Protoc 2015 (9):pdb top084111. doi:10.1101/pdb.top084111

van Leeuwen J, Andrews B, Boone C, Tan G (2015b) Rapid and Efficient Plasmid Construction by Homologous Recombination in Yeast. Cold Spring Harb Protoc 2015 (9):pdb prot085100. doi:10.1101/pdb.prot085100

van Nimwegen KJ, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LE, Grutters JP (2016) Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. Clin Chem. doi:10.1373/clinchem.2016.258632

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J,

Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291 (5507):1304-1351. doi:10.1126/science.1058040

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304 (5667):66-74. doi:10.1126/science.1093857

Viswanathan M, Lovett ST (1998) Single-strand DNA-specific exonucleases in *Escherichia coli*. Roles in repair and mutation avoidance. Genetics 149 (1):7-16

Waldminghaus T, Skarstad K (2009) The *Escherichia coli* SeqA protein. Plasmid 61 (3):141-150. doi:10.1016/j.plasmid.2009.02.004

Waldminghaus T, Weigel C, Skarstad K (2012) Replication fork movement and methylation govern SeqA binding to the *Escherichia coli* chromosome. Nucleic Acids Res 40 (12):5465-5476. doi:10.1093/nar/gks187

Wallden M, Fange D, Lundius EG, Baltekin O, Elf J (2016) The Synchronization of Replication and Division Cycles in Individual *E. coli* Cells. Cell 166 (3):729-739. doi:10.1016/j.cell.2016.06.052

Wang X, Liu X, Possoz C, Sherratt DJ (2006) The two *Escherichia coli* chromosome arms locate to separate cell halves. Genes Dev 20 (13):1727-1731. doi:10.1101/gad.388406

Watanabe S, Shiwa Y, Itaya M, Yoshikawa H (2012) Complete sequence of the first chimera genome constructed by cloning the whole genome of *Synechocystis strain* PCC6803 into the *Bacillus subtilis* 168 genome. J Bacteriol 194 (24):7007. doi:10.1128/JB.01798-12

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171 (4356):737-738

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6 (2):e16765. doi:10.1371/journal.pone.0016765

Weinstock MT, Hesek ED, Wilson CM, Gibson DG (2016) *Vibrio natriegens* as a fast-growing host for molecular biology. Nat Methods 13 (10):849-851. doi:10.1038/nmeth.3970

Welch M, Villalobos A, Gustafsson C, Minshull J (2009) You're one in a googol: optimizing genes for protein expression. J R Soc Interface 6 Suppl 4:S467-476. doi:10.1098/rsif.2008.0520.focus

Welsh KM, Lu AL, Clark S, Modrich P (1987) Isolation and characterization of the *Escherichia coli mutH* gene product. J Biol Chem 262 (32):15624-15629

Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioeng Bugs 3 (1):38-43. doi:10.4161/bbug.3.1.18223

Wessels S, Axelsson L, Bech Hansen E, De Vuyst L, Laulund S, Lähteenmäki L, Lindgren S, Mollet B, Salminen S, von Wright A (2004) The lactic acid bacteria, the food chain, and their regulation. Trends in Food Science & Technology 15 (10):498-505. doi:10.1016/j.tifs.2004.03.003

Wilkins D, van Sebille E, Rintoul SR, Lauro FM, Cavicchioli R (2013) Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. Nat Commun 4:2457. doi:10.1038/ncomms3457

Woelker B, Messer W (1993) The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. Nucleic Acids Res 21 (22):5025-5033

Wolanski M, Donczew R, Zawilak-Pawlik A, Zakrzewska-Czerwinska J (2014) *oriC*-encoded instructions for the initiation of bacterial chromosome replication. Front Microbiol 5:735. doi:10.3389/fmicb.2014.00735

Wright BE (2004) Stress-directed adaptive mutations and evolution. Mol Microbiol 52 (3):643-650. doi:10.1111/j.1365-2958.2004.04012.x

Xavier JC, Patil KR, Rocha I (2014) Systems biology perspectives on minimal and simpler cells. Microbiol Mol Biol Rev 78 (3):487-509. doi:10.1128/MMBR.00050-13

Xue X, Wang T, Jiang P, Shao Y, Zhou M, Zhong L, Wu R, Zhou J, Xia H, Zhao G, Qin Z (2015) MEGA (Multiple Essential Genes Assembling) deletion and replacement method for genome reduction in *Escherichia coli*. ACS Synth Biol 4 (6):700-706. doi:10.1021/sb500324p

Yamaguchi M, Dao V, Modrich P (1998) MutS and MutL activate DNA helicase II in a mismatch-dependent manner. J Biol Chem 273 (15):9197-9201

Yamaichi Y, Fogel MA, McLeod SM, Hui MP, Waldor MK (2007) Distinct centromere-like *parS* sites on the two chromosomes of *Vibrio spp*. J Bacteriol 189 (14):5314-5324. doi:10.1128/JB.00416-07

Yamaichi Y, Niki H (2004) *migS*, a cis-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome. EMBO J 23 (1):221-233. doi:10.1038/sj.emboj.7600028

Yang H, Wolff E, Kim M, Diep A, Miller JH (2004) Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. Mol Microbiol 53 (1):283-295. doi:10.1111/j.1365-2958.2004.04125.x

Zechiedrich EL, Khodursky AB, Cozzarelli NR (1997) Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. Genes Dev 11 (19):2580-2592

Zhou J, Wu R, Xue X, Qin Z (2016) CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. Nucleic Acids Res 44 (14):e124. doi:10.1093/nar/gkw475

Erklärung der selbstständigen Erarbeitung der Dissertation

Hiermit erkläre ich, dass ich die vorliegende Dissertation:

„Konstruktion synthetischer sekundärer Chromosomen zur Charakterisierung von DNA-Reparatur und Segregation in *Escherichia coli*“

selbstständig und ohne unerlaubte Hilfsmittel angefertigt habe. Es wurden keine anderen, als der von mir ausdrücklich angegebenen Hilfsmittel verwendet. Die Dissertation wurde in der jetzigen oder in einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht.

Es handelt sich bei den heute von mir eingereichten Exemplaren um in Wort und Bild völlig übereinstimmende Exemplare.

Des Weiteren erkläre ich, dass die digitalen Abbildungen ausschließlich originale Daten enthalten und in keinem Fall inhaltsverändernde Bildbearbeitung vorgenommen wurde.

Marburg, den 09.11.2016

Daniel Schindler