Sintezna genomika

Ponovno zapisovanje (Rewriting)
Preoblikovanje kode (Refactoring)
Transplantacija genoma
Sintezni genomi
Sintezni organizmi

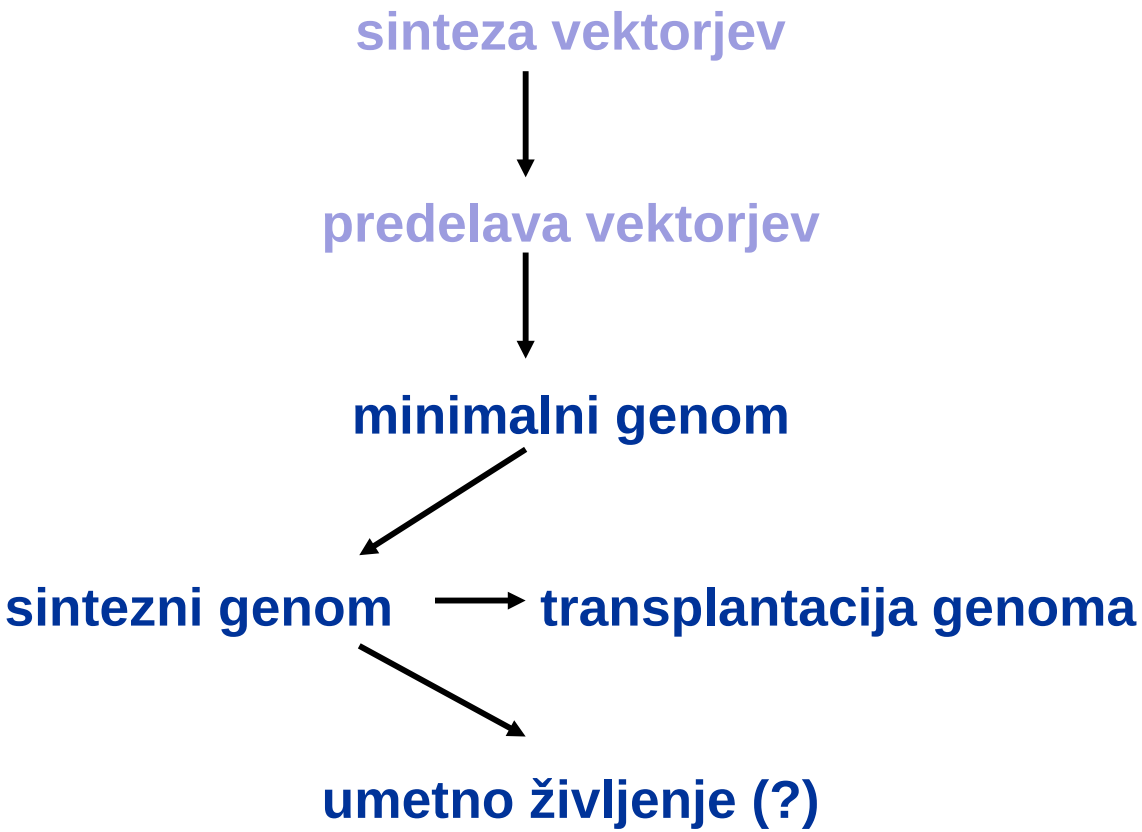
Koncept ‚sinteznega življenja‘

- Celic, kot so se razvile v evoluciji, ne razumemo v celoti, zato bi za inženirske namene bilo smiselno pripraviti (sintezni) celični sistem, ki bi se obnašal v celoti predvidljivo.
- Sintezne celice bi bilo mogoče zasnovati tako, da bi opravljale samo posamezne naloge in bi zato potrebovale manj energije kot naravne celice.
- Do sintezne celice lahko pridemo na dva načina: s poenostavitvijo obstoječih genomov (zmanjšanje genoma; *genome reduction*), ali pa z inženirsko zasnovanim minimalnim genomom, ki bi ga vstavili v celice – govorimo o pristopih ‚od zgoraj navzdol‘ oz. ‚od spodaj navzgor‘.
- Pristop ‚od spodaj navzgor‘ bi lahko temeljil na minimalnih celicah, ki bi bile prilagojene le za življenje v posebnih pogojih, lahko pa bi uporabili brezcelične sisteme / in vitro umetne celice.
- Vprašanje je, kako definiramo življenje.
- Kako obsežen je minimalni genom? Za prostoživeče evkarionte verjetno ~12 Mbp / ~5000 genov (kvasovke), za prokarionte ~1,3 Mbp / 1300 genov (*Pelagibacter ubique*). Od bakterij bolje razumemo mikoplazme, ki običajno živijo parazitsko in imajo genom velik >0,58 Mbp (*M. genitalium*: 583.000 bp, 482 genov), nekatere so sposobne tudi rasti v laboratoriju na posebnih gojiščih (*M. mycoides, M. capricolum*).

Zmanjšanje genoma

- Hamilton Smith (NN 1978 za odkritje restriktaz, 1970 *Hin*dII): poskus določitve minimalnega genoma *M. genitalium*. Ta vrsta bakterij zaradi parazitskega načina življenja nima vseh potrebnih genov za sintezo aminokislin in nukleotidov ter za oksidativni metabolizem.
- Za določitev minimalnega nabora genov lahko uporabimo dva raziskovalna pristopa: s primerjavo z ostalimi genomi ali redukcionistični pristop.
- Pri primerjalnem pristopu analiziramo, kateri geni so skupni praktično vsem organizmom, hkrati pa tudi preverjamo, koliko se med predstavniki različnih vrst neke širše skupine organizmov med seboj razlikujejo (manj ko se, bolj je verjetno njihova funkcija pomembna).
- Samo 60-70 genov je skupnih vsem živim bitjem, a to ne zadošča za življenje, saj so različne skupine organizmov razvile različne tipe molekul za opravljanje nujnih funkcij. Mikoplazme imajo skupnih 200-300 genov, kar je bolj verjetno minimalno število esencialnih genov.
- Pri redukcionističnem pristopu sistematično odstranjujemo iz genoma posamezne gene (knock-out, npr. pri *M. genitalium* z uporabo transpozonov, ki prekinejo zaporedje gena in ga s tem inaktivirajo, pri *B. subtilis* pa z uporabo nereplicirajočih plazmidov).
- Bacillus nujno potrebuje 271 genov od 4800 (verjetno ne bi ostalo pri tem, če bi deletirali hkrati vse neesencialne gene), M. genitalium pa ima 100 neesencialnih genov (torej nujno rabi 370 genov za proteine, a za 100 od teh ne poznajo funkcije v celici).
- Cilj raziskav na mikoplazmah je (bil) razviti celice *M. laboratorium*, delecijske mutante *M. genitalium* z minimalnim naborom genov. Tak genom je mogoče tudi sintetizirati in celico, ki bi vsebovala samo sintezni genom, bi pogojno lahko imenovali ‚sintezna celica‘ (vse sestavine na začetku ne bi bile sintetizirane, pač pa bi kot osnovo uporabili obstoječo naravno celico).

Sintezna genomika



Gene, 164 (1995) 49–53

GENE 09226

Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides

(Gene synthesis; plasmid synthesis; β-lactamase; in vitro recombination; assembly PCR; DNA shuffling; mutagenesis)

Willem P.C. Stemmer[a], Andreas Crameri[a], Kim D. Ha[b], Thomas M. Brennan[b] and Herbert L. Heyneker[b]

[a] *Affymax Research Institute, Palo Alto, CA 94304, USA; and* [b] *ProtoGene Laboratories, Inc., Mountain View, CA 94043, USA*

Received by F. Bolivar: 3 March 1995; Revised/Accepted: 15 June/23 June 1995; Received at publishers: 24 July 1995

SUMMARY

Here, we describe assembly PCR as a method for the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos). The method, which is derived from DNA shuffling [Stemmer, Nature 370 (1994a) 389–391], does not rely on DNA ligase but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process. A 1.1-kb fragment containing the TEM-1 β-lactamase-encoding gene (*bla*) was assembled in a single reaction from a total of 56 oligos, each 40 nucleotides (nt) in length. The synthetic gene was PCR amplified and cloned in a vector containing the tetracycline-resistance gene (Tc^R) as the sole selectable marker. Without relying on ampicillin (Ap) selection, 76% of the Tc^R colonies were Ap^R, making this approach a general method for the rapid and cost-effective synthesis of any gene. We tested the range of assembly PCR by synthesizing, in a single reaction vessel containing 134 oligos, a high-molecular-mass multimeric form of a 2.7-kb plasmid containing the *bla* gene, the α-fragment of the *lacZ* gene and the pUC origin of replication. Digestion with a unique restriction enzyme, followed by ligation and transformation in *Escherichia coli*, yielded the correct plasmid. Assembly PCR is well suited for several in vitro mutagenesis strategies.

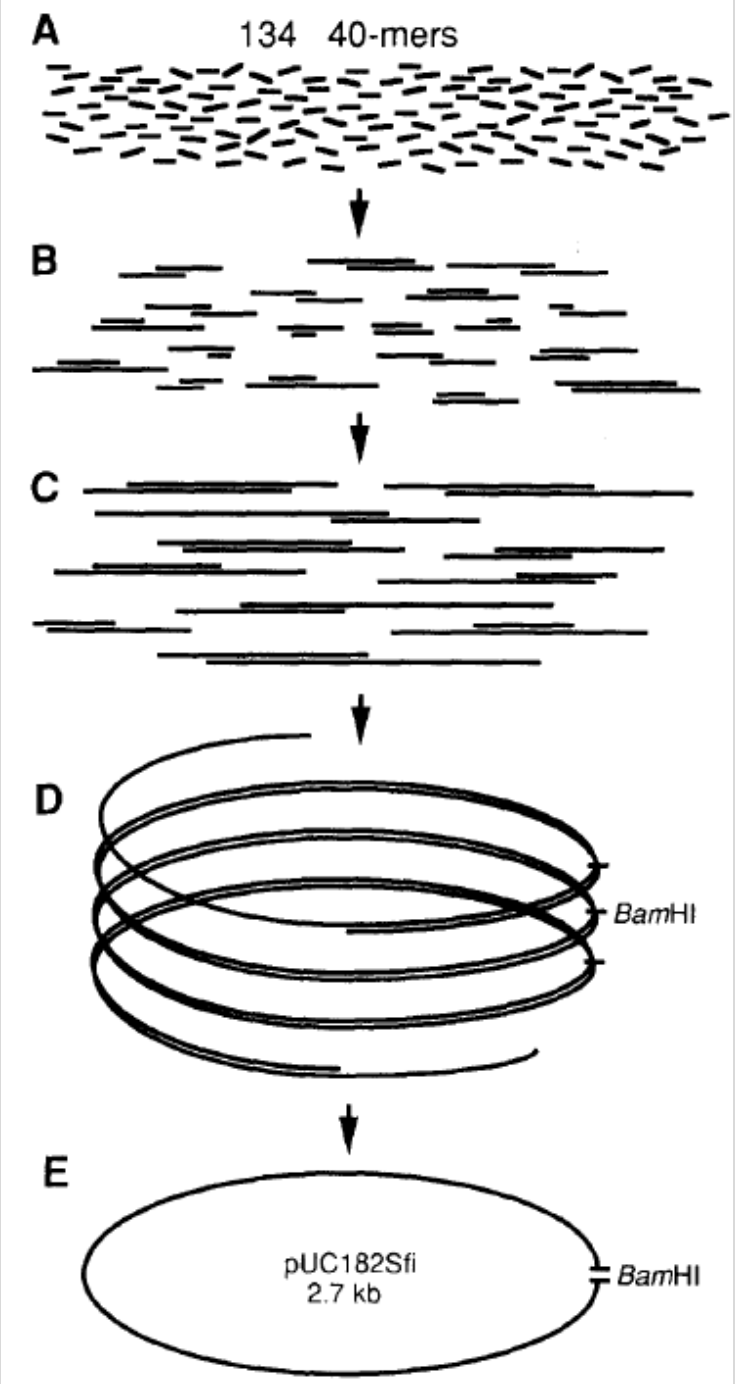


Fig. 3. Protocol for assembly of synthetic plasmid pUC182Sfi. Preparation of oligos was described in the legend to Fig. 1a. Equal volumes drawn from each of the 134 oligo solutions were combined to a final concentration of approx. 250 µM of mixed oligos, prior to a 250-fold dilution in 20 µl of Geneamp XL PCR mix (Perkin Elmer, Branchburg, NJ, USA). The amplification process consisted of three stages. The conditions for the first PCR program, which started with

Generating a synthetic genome by whole genome assembly: ϕX174 bacteriophage from synthetic oligonucleotides

Hamilton O. Smith, Clyde A. Hutchison III[†], Cynthia Pfannkoch, and J. Craig Venter[‡]

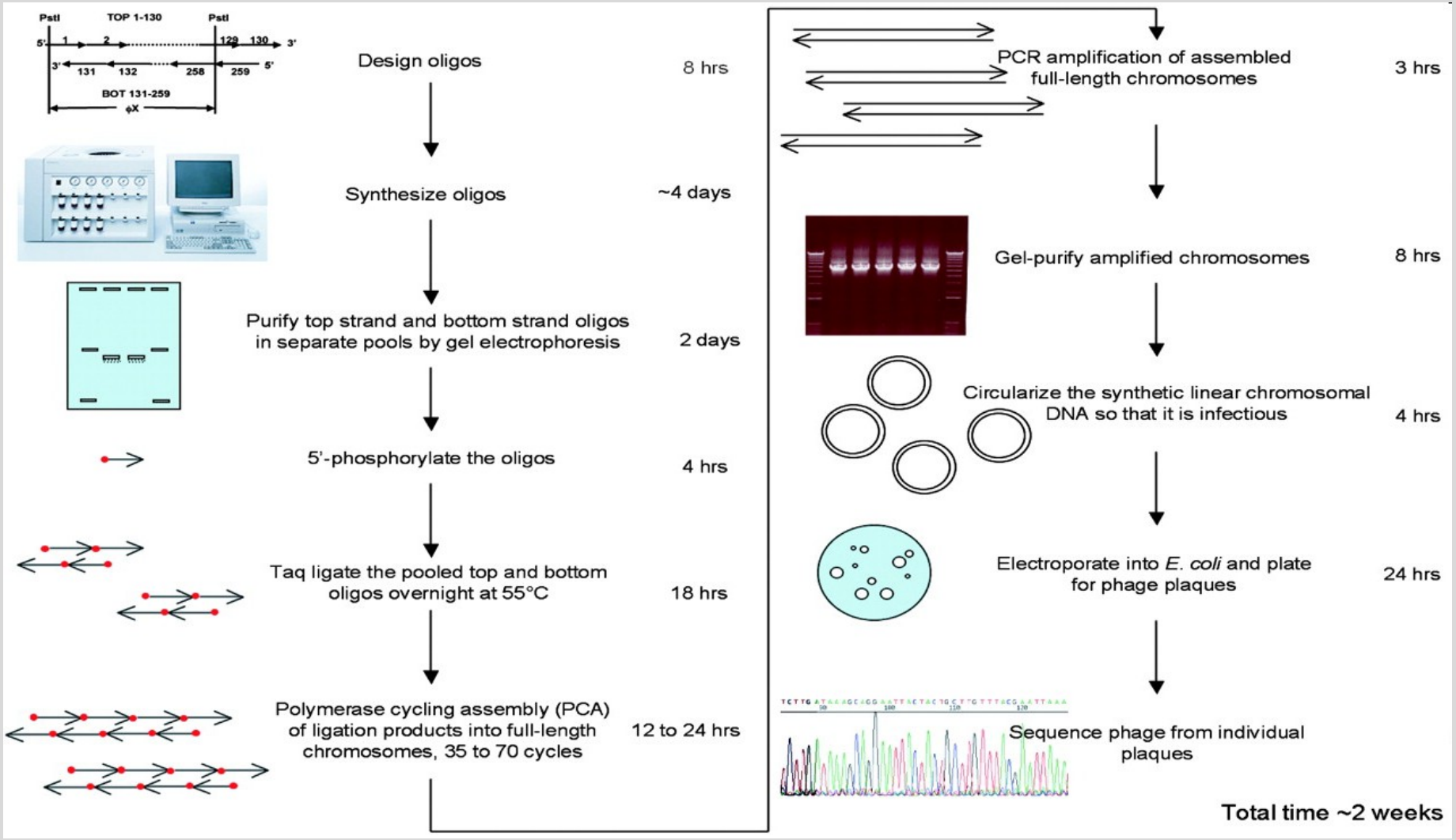
Institute for Biological Energy Alternatives, 1901 Research Boulevard, Suite 600, Rockville, MD 20850

Contributed by J. Craig Venter, November 3, 2003

We have improved upon the methodology and dramatically shortened the time required for accurate assembly of 5- to 6-kb segments of DNA from synthetic oligonucleotides. As a test of this methodology, we have established conditions for the rapid (14-day) assembly of the complete infectious genome of bacteriophage ϕX174 (5,386 bp) from a single pool of chemically synthesized oligonucleotides. The procedure involves three key steps: (*i*) gel purification of pooled oligonucleotides to reduce contamination with molecules of incorrect chain length, (*ii*) ligation of the oligonucleotides under stringent annealing conditions (55°C) to select against annealing of molecules with incorrect sequences, and (*iii*) assembly of ligation products into full-length genomes by polymerase cycling assembly, a nonexponential reaction in which each terminal oligonucleotide can be extended only once to produce a full-length molecule. We observed a discrete band of full-length assemblies upon gel analysis of the polymerase cycling assembly product, without any PCR amplification. PCR amplification was then used to obtain larger amounts of pure full-length genomes for circularization and infectivity measurements. The synthetic DNA had a lower infectivity than natural DNA, indicating approximately one lethal error per 500 bp. However, fully infectious ϕX174 virions were recovered after electroporation into *Escherichia coli*. Sequence analysis of several infectious isolates verified the accuracy of these synthetic genomes. One such isolate had exactly the intended sequence. We propose to assemble larger genomes by joining separately assembled 5- to 6-kb segments; ≈60 such segments would be required for a minimal cellular genome.

truncated species. Although such oligonucleotides are highly useful as primers for PCR amplification and DNA sequencing, only small (a few hundred base pairs) synthetic genes can generally be accurately and directly synthesized without multiple repair/selection steps. For example, the recent report (9) of the assembly of a partially active poliovirus from cloned synthetic segments of DNA from which polio genomic RNA (7,440 bases) could be transcribed was quite complex and took many months to accomplish. First, segments 400–600 bp long were individually assembled and cloned, and 5–15 isolates of each were sequenced to find one that was correct or readily correctable by oligonucleotide mutagenesis. These segments were then assembled into three larger segments of the polio genome, recloned, and finally assembled to produce a full-length product. This slow process would not be practical for synthesizing a 300,000-bp chromosome. We have now improved the methodology for synthesis of multigene segments of a genome as a step toward synthesis of a cellular genome. As a test of this methodology we have established conditions for global assembly of the infectious genome of bacteriophage ϕX174 (5,386 bp) from a single pool of chemically synthesized oligonucleotides. ϕX174 presents no known hazard because it infects only certain enteric bacteria and is not a human, plant, or animal pathogen. Therefore, its choice for synthesis serves to separate safety issues from ethical considerations and other potential risks associated with synthetic genomics.

ϕX174 (ϕX) is the prototypical minute icosahedral bacterio-



Schematic diagram of the steps in the global synthesis of infectious φX174 bacteriophage from synthetic oligonucleotides.

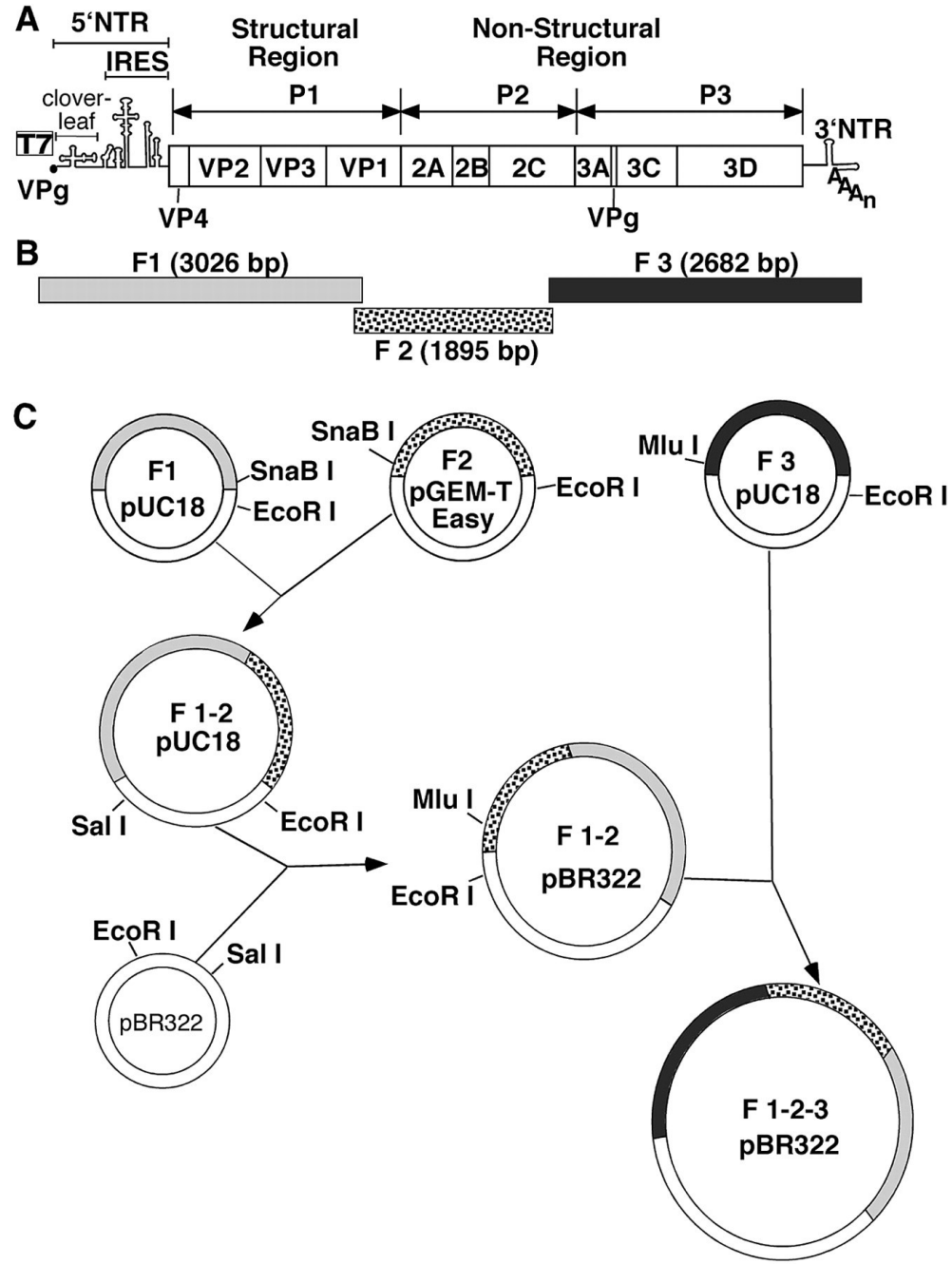
Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template

Jeronimo Cello, Aniko V. Paul, Eckard Wimmer*

Full-length poliovirus complementary DNA (cDNA) was synthesized by assembling oligonucleotides of plus and minus strand polarity. The synthetic poliovirus cDNA was transcribed by RNA polymerase into viral RNA, which translated and replicated in a cell-free extract, resulting in the de novo synthesis of infectious poliovirus. Experiments in tissue culture using neutralizing antibodies and CD155 receptor–specific antibodies and neurovirulence tests in *CD155* transgenic mice confirmed that the synthetic virus had biochemical and pathogenic characteristics of poliovirus. Our results show that it is possible to synthesize an infectious agent by in vitro chemical-biochemical means solely by following instructions from a written sequence.

SCIENCE VOL 297 9 AUGUST 2002

ssRNA, 7741 nt



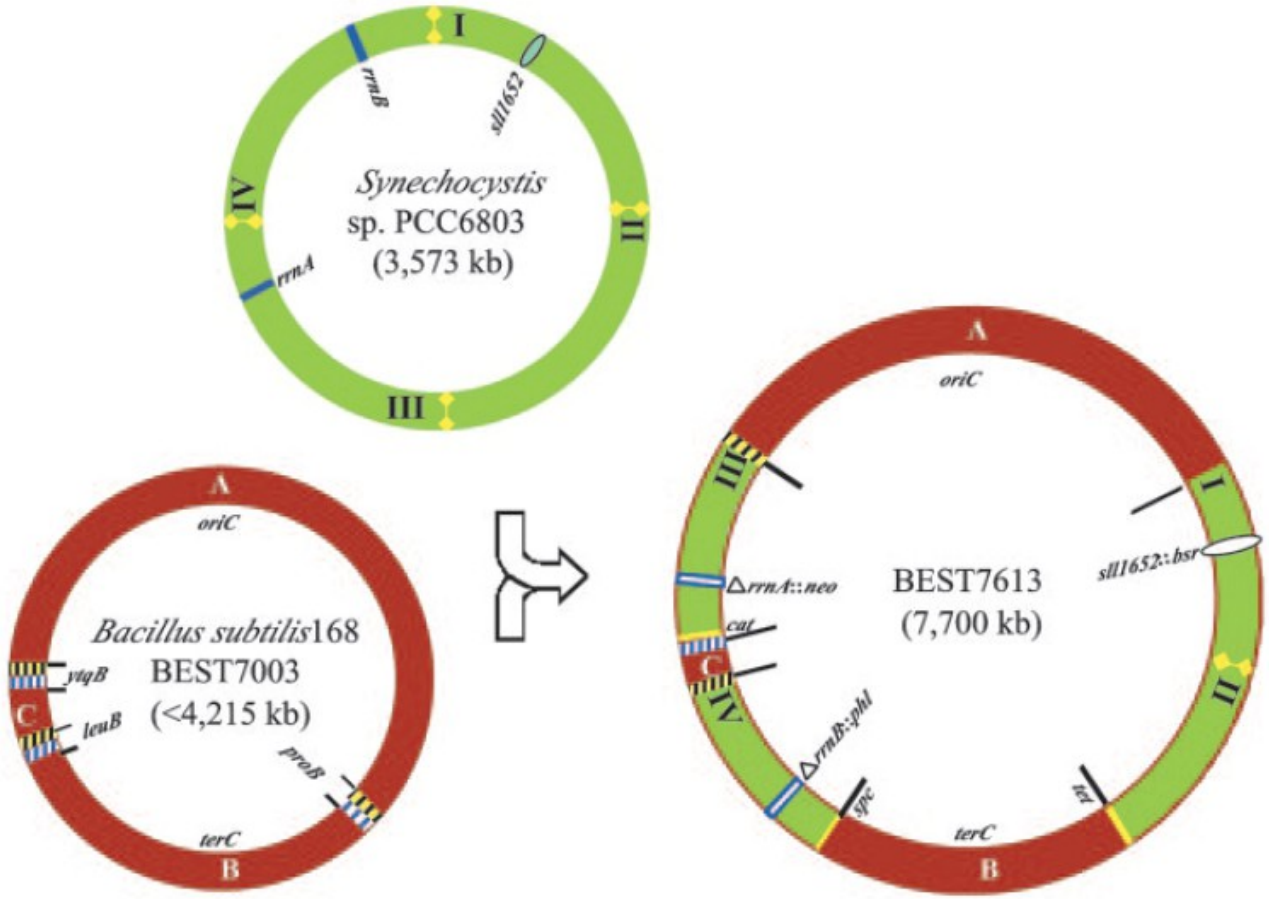
Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome

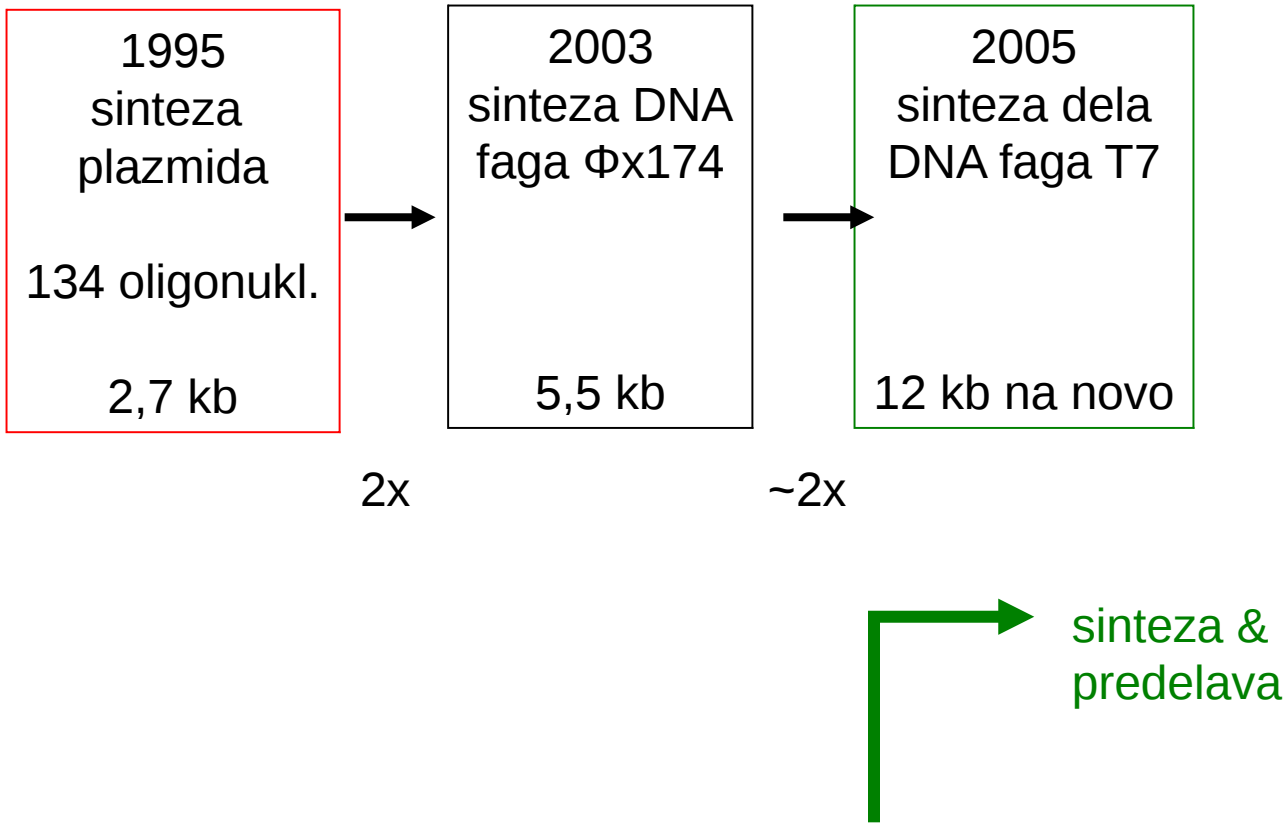
Mitsuhiro Itaya*, Kenji Tsuge, Maki Koizumi, and Kyoko Fujita

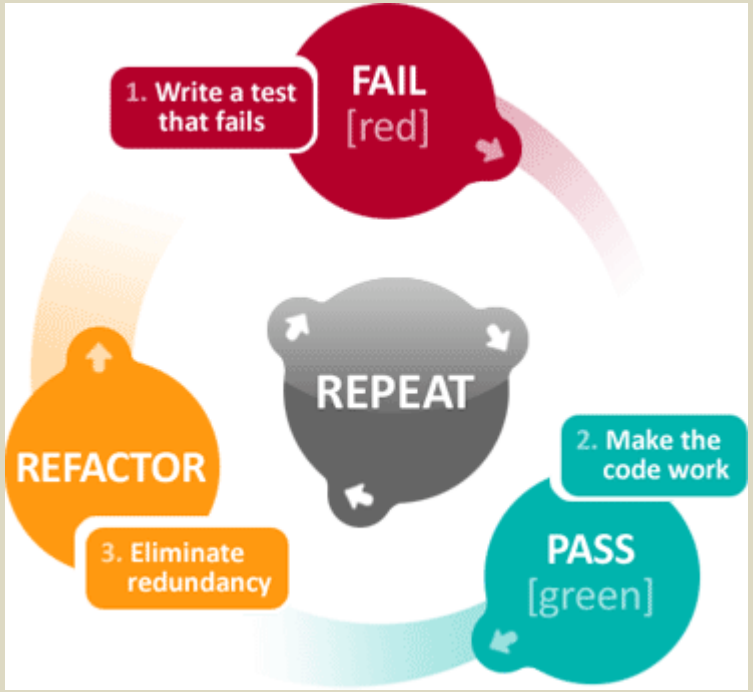
Mitsubishi Kagaku Institute of Life Sciences, 11 Minamiooya, Machida-shi, Tokyo 194-8511, Japan

Edited by J. Craig Venter, The J. Craig Venter Institute, Rockville, MD, and approved September 16, 2005 (received for review May 10, 2005)

Cloning the whole 3.5-megabase (Mb) genome of the photosynthetic bacterium *Synechocystis* PCC6803 into the 4.2-Mb genome of the mesophilic bacterium *Bacillus subtilis* 168 resulted in a 7.7-Mb composite genome. We succeeded in such unprecedented large-size cloning by progressively assembling and editing contiguous DNA regions that cover the entire *Synechocystis* genome. The strain containing the two sets of genome grew only in the *B. subtilis* culture medium where all of the cloning procedures were carried out. The high structural stability of the cloned *Synechocystis* genome was closely associated with the symmetry of the bacterial genome structure of the DNA replication origin (*oriC*) and its termination (*terC*) and the exclusivity of *Synechocystis* ribosomal RNA operon genes (*rrnA* and *rrnB*). Given the significant diversity in genome structure observed upon horizontal DNA transfer in nature, our stable laboratory-generated composite genome raised fundamental questions concerning two complete genomes in one cell. Our megasize DNA cloning method, designated megacloning, may be generally applicable to other genomes or genome loci of free-living organisms.




1995
sinteza
plazmida
134 oligonukl.
2,7 kb
2003
sinteza DNA
faga Φx174
5,5 kb
2005
sinteza dela
DNA faga T7
12 kb na novo
2x
~2x
sinteza &
predelava



http://www.improveit.mx/en/tddcourse.html

TDD = test-driven development

Preoblikovanje kode je v programiranju pristop izboljševanja programja, s čimer izboljšamo njegovo delovanje in podaljšamo življenjsko dobo izdelku. Končna uporabnost izdelka se ne spremeni, izboljšajo se le njegove lastnosti.

Ponovno zapisovanje predstavlja manjšo spremembo v programju kot preoblikovanje kode in z njim samo delno popravimo lastnosti, za katere smo opazili, da ne delujejo optimalno.

Preoblikovanje je ključni sestavni del razvoja, ki temelji na testiranju (TDD).

Molecular Systems Biology (2005) doi:10.1038/msb4100025

www.molecularsystemsbiology.com
Article number: 2005.0018

molecular systems biology

Refactoring bacteriophage T7

Leon Y Chan[1,3], Sriram Kosuri[2,3] and Drew Endy[2,*]

[1] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA and [2] Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] These authors contributed equally to this work
* Corresponding author. Division of Biological Engineering, Massachusetts Institute of Technology, 68-580, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Tel.: +1 617 258 5152; Fax: +1 617 253 5865; E-mail: endy@mit.edu

Received 15.7.05; accepted 23.7.05

Natural biological systems are selected by evolution to continue to exist and evolve. Evolution likely gives rise to complicated systems that are difficult to understand and manipulate. Here, we redesign the genome of a natural biological system, bacteriophage T7, in order to specify an engineered surrogate that, if viable, would be easier to study and extend. Our initial design goals were to physically separate and enable unique manipulation of primary genetic elements. Implicit in our design are the hypotheses that overlapping genetic elements are, in aggregate, nonessential for T7 viability and that our models for the functions encoded by elements are sufficient. To test our initial design, we replaced the left 11 515 base pairs (bp) of the 39 937 bp wild-type genome with 12 179 bp of engineered DNA. The resulting chimeric genome encodes a viable bacteriophage that appears to maintain key features of the original while being simpler to model and easier to manipulate. The viability of our initial design suggests that the genomes encoding natural biological systems can be systematically redesigned and built anew in service of scientific understanding or human intention.
Molecular Systems Biology 13 September 2005; doi:10.1038/msb4100025

Levih 11.515 bp so nadomestili s preurejenih 12.179 bp → fag T7.1; ukinili so prekrivanja genov → lažje modeliranje in manipuliranje.

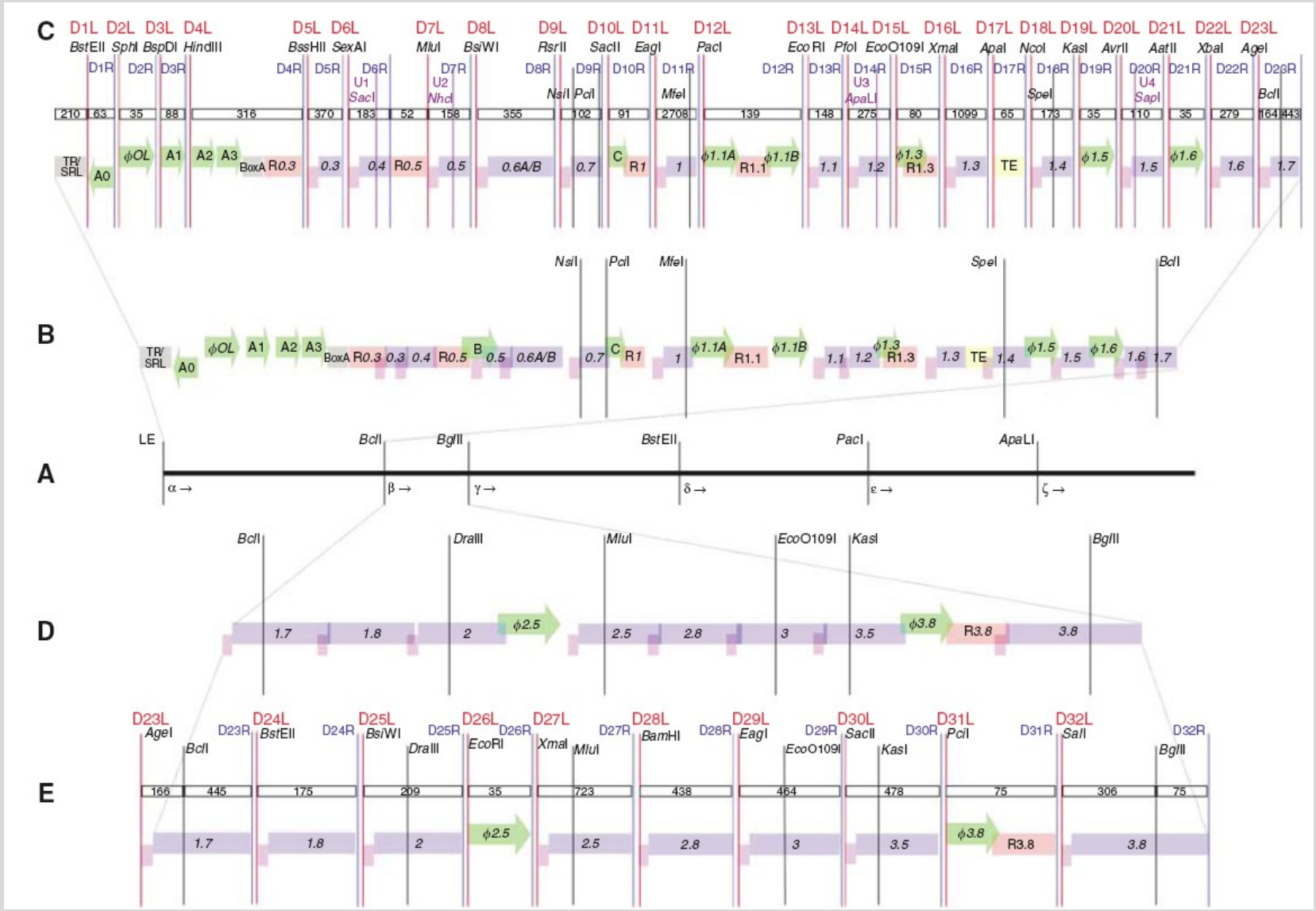


Figure 2 Genome design. (**A**) We split the wild-type T7 genome into six sections, *alpha* through *zêta*, using five restriction sites unique across the natural sequence. (**B**) Wild-type section *alpha* genetic elements: protein coding regions (blue), RBSs (purple), promoters (green), RNaseIII recognition sites (pink), a transcription terminator (yellow), and others (gray). Elements are labeled by convention (Dunn and Studier, 1983). Images are not to scale, but overlapping boundaries indicate elements with shared sequence. The five useful natural restriction sites across section *alpha* are shown (black lines). (**C**) T7.1 section *alpha* parts. Parts are given integer numbers, 1–73, starting at the left end of the genome. Unique restriction site pairs bracket each part (red/blue lines, labeled D[part #]L/R). Added unique restriction sites (purple lines, U[part #]) and part length (# base pairs, open boxes) are shown. We do not know if sequence changes in and around parts 6 and 7 destroy the minor *E. coli* promoter, B. (**D**) Wild-type section *beta* genetic elements. (**E**) T7.1 section *beta* parts. Supplementary Figure S2 depicts the six sections, alpha through *zêta*, which make up the T7.1 genome.

Molecular Systems Biology (2005) doi:10.1038/msb4100025

www.molecularsystemsbiology.com
Article number: 2005.0018

molecular systems biology

Refactoring bacteriophage T7

Leon Y Chan[1,3], Sriram Kosuri[2,3] and Drew Endy[2,*]

[1] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA and [2] Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] These authors contributed equally to this work
* Corresponding author. Division of Biological Engineering, Massachusetts Institute of Technology, 68-580, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Tel.: +1 617 258 5152; Fax: +1 617 253 5865; E-mail: endy@mit.edu

Received 15.7.05; accepted 23.7.05

Natural biological systems are selected by evolution to continue to exist and evolve. Evolution likely gives rise to complicated systems that are difficult to understand and manipulate. Here, we redesign the genome of a natural biological system, bacteriophage T7, in order to specify an engineered surrogate that, if viable, would be easier to study and extend. Our initial design goals were to physically separate and enable unique manipulation of primary genetic elements. Implicit in our design are the hypotheses that overlapping genetic elements are, in aggregate, nonessential for T7 viability and that our models for the functions encoded by elements are sufficient. To test our initial design, we replaced the left 11 515 base pairs (bp) of the 39 937 bp wild-type genome with 12 179 bp of engineered DNA. The resulting chimeric genome encodes a viable bacteriophage that appears to maintain key features of the original while being simpler to model and easier to manipulate. The viability of our initial design suggests that the genomes encoding natural biological systems can be systematically redesigned and built anew in service of scientific understanding or human intention.
Molecular Systems Biology 13 September 2005; doi:10.1038/msb4100025

Levih 11.515 bp so nadomestili s preurejenih 12.179 bp → fag T7.1; ukinili so prekrivanja genov → lažje modeliranje in manipuliranje.

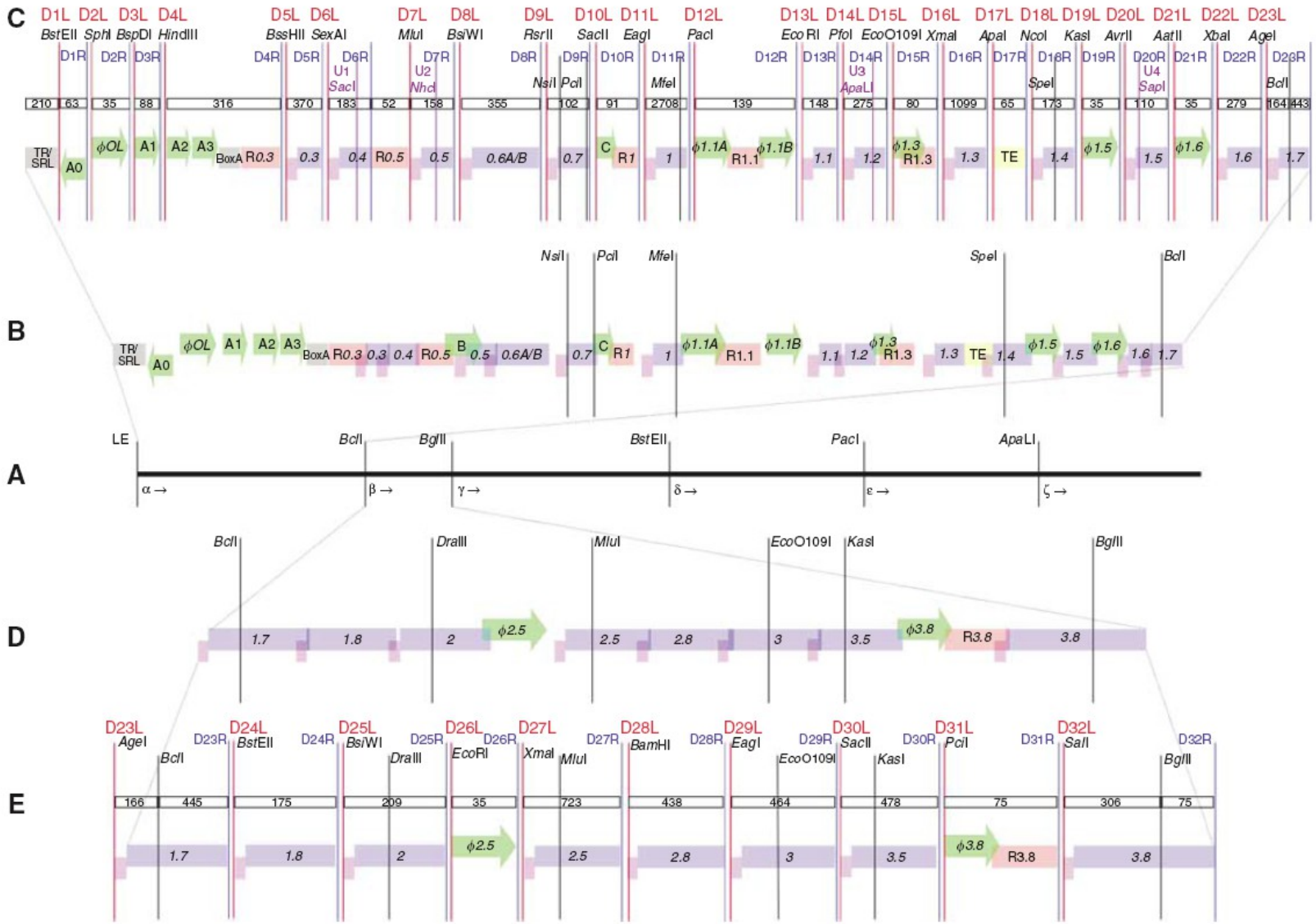


Figure 2 Genome design. (**A**) We split the wild-type T7 genome into six sections, *alpha* through *zêta*, using five restriction sites unique across the natural sequence. (**B**) Wild-type section *alpha* genetic elements: protein coding regions (blue), RBSs (purple), promoters (green), RNaseIII recognition sites (pink), a transcription terminator (yellow), and others (gray). Elements are labeled by convention (Dunn and Studier, 1983). Images are not to scale, but overlapping boundaries indicate elements with shared sequence. The five useful natural restriction sites across section *alpha* are shown (black lines). (**C**) T7.1 section *alpha* parts. Parts are given integer numbers, 1–73, starting at the left end of the genome. Unique restriction site pairs bracket each part (red/blue lines, labeled D[part #]L/R). Added unique restriction sites (purple lines, U[part #]) and part length (# base pairs, open boxes) are shown. We do not know if sequence changes in and around parts 6 and 7 destroy the minor *E. coli* promoter, B. (**D**) Wild-type section *beta* genetic elements. (**E**) T7.1 section *beta* parts. Supplementary Figure S2 depicts the six sections, alpha through *zêta*, which make up the T7.1 genome.

Table 2. Landmarks in the race toward the synthesis of artificial minimal genomes

References	Title	Remarks
Sekiya *et al.* (1979)	Total synthesis of a tyrosine suppressor tRNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA	Khorana's research group presents the last in a series of 16 papers, published along 5 years, describing the chemical synthesis of a 207-bp tRNA gene
Nambiar *et al.* (1984)	Total synthesis and cloning of a gene coding for the ribonuclease S protein	The first protein-coding gene is synthesized in the Sidney Brenner's group. Two people, working full-time during a year and a half, were needed to synthesize this 330-bp DNA fragment (S.A. Brenner, pers. commun.)
Stemmer *et al.* (1995)	Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides	The first report of the synthesis of extended stretches of DNA, using a PCR-based method for the synthesis of a 2.7-kb plasmid from a pool of short, overlapping synthetic oligonucleotides
Cello *et al.* (2002)	Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template	The first chemically synthesized genome. A functional 7740-bp poliovirus genome, with all expected biochemical and pathogenic properties was obtained
Yount *et al.* (2003)	Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus	Viral genome assembled from cDNAs
Smith *et al.* (2003)	Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides	Venter's group synthesizes a 5385-bp bacteriophage genome in 2 weeks by serialized oligonucleotides assembly and amplification
Kodumal *et al.* (2004)	Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster	A couple of months were sufficient to synthesize a genome fragment of 32 kb, the biggest piece of man-made DNA at that time, incorporating several genes needed for the synthesis of a pharmaceutical compound
Tian *et al.* (2004)	Accurate multiplex gene synthesis from programmable DNA microchips	Presents an oligonucleotide synthesis method miniaturized on photo-programmable microfluidic chips, allowing both reduced error frequencies and lower costs. The procedure allowed the synthesis of all 21 genes that encode proteins of the *E. coli* 30S ribosomal subunit and to improve translation efficiency *in vitro* through alteration of codon bias.
Gibson *et al.* (2008)	Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome	The complete synthesis of the 582 970-bp genome for the pathogenic bacterium *M. genitalium* is reported. Venter's group used assembled chemically synthesized oligonucleotides, joined by *in vitro* recombination to produce intermediate assemblies, which were all cloned as bacterial artificial chromosomes in *E. coli*, and transferred into yeast, to assemble the full-length genome

FEMS Microbiol Rev **33** (2009) 225–235

Presaditev genoma
2007

Mycoplasma capricolum → *Mycoplasma mycoides* LC

RESEARCH ARTICLE

Genome Transplantation in Bacteria: Changing One Species to Another

Carole Lartigue, John I. Glass,* Nina Alperovich, Rembert Pieper, Prashanth P. Parmar, Clyde A. Hutchison III, Hamilton O. Smith, J. Craig Venter

As a step toward propagation of synthetic genomes, we completely replaced the genome of a bacterial cell with one from another species by transplanting a whole genome as naked DNA. Intact genomic DNA from *Mycoplasma mycoides* large colony (LC), virtually free of protein, was transplanted into *Mycoplasma capricolum* cells by polyethylene glycol–mediated transformation. Cells selected for tetracycline resistance, carried by the *M. mycoides* LC chromosome, contain the complete donor genome and are free of detectable recipient genomic sequences. These cells that result from genome transplantation are phenotypically identical to the *M. mycoides* LC donor strain as judged by several criteria.

SCIENCE VOL 317 3 AUGUST 2007

from incompatibility between the two genomes (*6*). Transplantation of nuclei as intact organelles into enucleated eggs is a well-established procedure in vertebrates (*7*–*9*). Our choice of the term "genome transplantation" comes from the similarity to eukaryotic nuclear transplantation in which one genome is cleanly replaced by another.

Genome transplantation is a requirement for the establishment of the new field of synthetic genomics. It may facilitate construction of useful microorganisms with the potential to solve pressing societal problems in energy production, environmental stewardship, and medicine. Chemically synthesized chromosomes must eventually be transplanted into a cellular milieu where the encoded instructions can be expressed. We have long been interested in defining a minimal genome that is just sufficient for cellular

1. izolacija genomske DNA iz M.m. (sev, ki je odporen na Tet)
2. 2. transformacija celic M.c. in selekcija na Tet

Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Chuck Merryman, Lei Young, Vladimir N. Noskov, John I. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith*

We have synthesized a 582,970–base pair *Mycoplasma genitalium* genome. This synthetic genome, named *M. genitalium* JCVI-1.0, contains all the genes of wild-type *M. genitalium* G37 except MG408, which was disrupted by an antibiotic marker to block pathogenicity and to allow for selection. To identify the genome as synthetic, we inserted "watermarks" at intergenic sites known to tolerate transposon insertions. Overlapping "cassettes" of 5 to 7 kilobases (kb), assembled from chemically synthesized oligonucleotides, were joined by in vitro recombination to produce intermediate assemblies of approximately 24 kb, 72 kb ("1/8 genome"), and 144 kb ("1/4 genome"), which were all cloned as bacterial artificial chromosomes in *Escherichia coli*. Most of these intermediate clones were sequenced, and clones of all four 1/4 genomes with the correct sequence were identified. The complete synthetic genome was assembled by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*, then isolated and sequenced. A clone with the correct sequence was identified. The methods described here will be generally useful for constructing large DNA molecules from chemically synthesized pieces and also from combinations of natural and synthetic DNA segments.

genome, we needed to establish convenient and reliable methods for the assembly and cloning of much larger synthetic DNA molecules.

Strategy for synthesis and assembly. The native 580,076-bp *M. genitalium* genome sequence (*Mycoplasma genitalium* G37 ATCC 33530 genomic sequence; accession no. L43967) (*3*) was partitioned into 101 cassettes of approximately 5 to 7 kb in length (Fig. 1) that were individually synthesized, verified by sequencing, and then joined together in stages. In general, cassette boundaries were placed between genes so that each cassette contained one or several complete genes. This will simplify the future deletion or manipulation of the genes in individual cassettes. Most cassettes overlapped their adjacent neighbors by 80 bp; however, some segments overlapped by as much as 360 bp. Cassette 101 overlapped cassette 1, thus completing the circle.

Short "watermark" sequences were inserted in cassettes 14, 29, 39, 55 and 61. Watermarks are inserted or substituted sequences used to identify or encode information into DNA. This information can be either in noncoding or coding sequences (*10*–*12*). Most commonly, watermarking has been used to encrypt information within coding sequences without altering the amino acid sequences (*10*, *11*). We opted to insert watermark sequences at

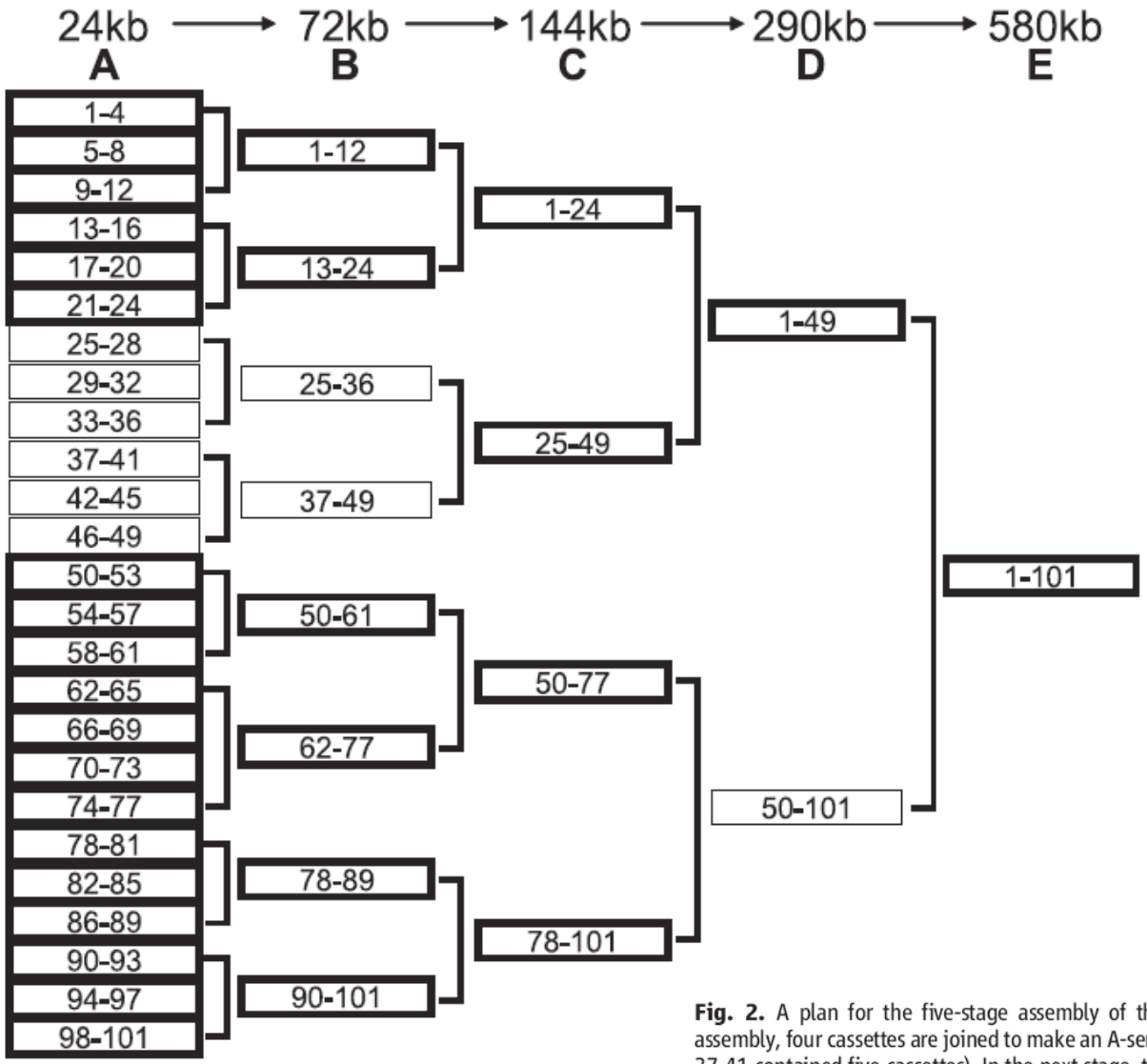


Fig. 2. A plan for the five-stage assembly of the *M. genitalium* chromosome. In the first stage of assembly, four cassettes are joined to make an A-series assembly approximately 24 kb in length (assembly 37-41 contained five cassettes). In the next stage, three A-assemblies are joined together to make a total of eight ~72-kb B-series assemblies (assembly B62-77 contained four A-series assemblies). The eighth-genome B-assemblies are taken two at a time to make quarter-genome C-series assemblies. These assemblies were all made by in vitro recombination (see Fig. 3) and cloned into *E. coli* using BAC vectors. Half-genome and whole-genome assemblies were made by in vivo yeast recombination. Assemblies in bold boxes were sequenced to verify their correctness. For the final molecule, the D-series half molecules were not employed. Rather, we assembled the whole molecule from the four C-series quarter molecules.

S četrtinskimi genomi so hkrati transformirali kvasovke, kjer so popravljalni mehanizmi zlepili prekrivajoče se konce segmentov.

Dobili so krožni genom, ki je imel razen naravnih še nekatera dodatna zaporedja, ki so omogočala selekcijo in replikacijo v kvasovkah.

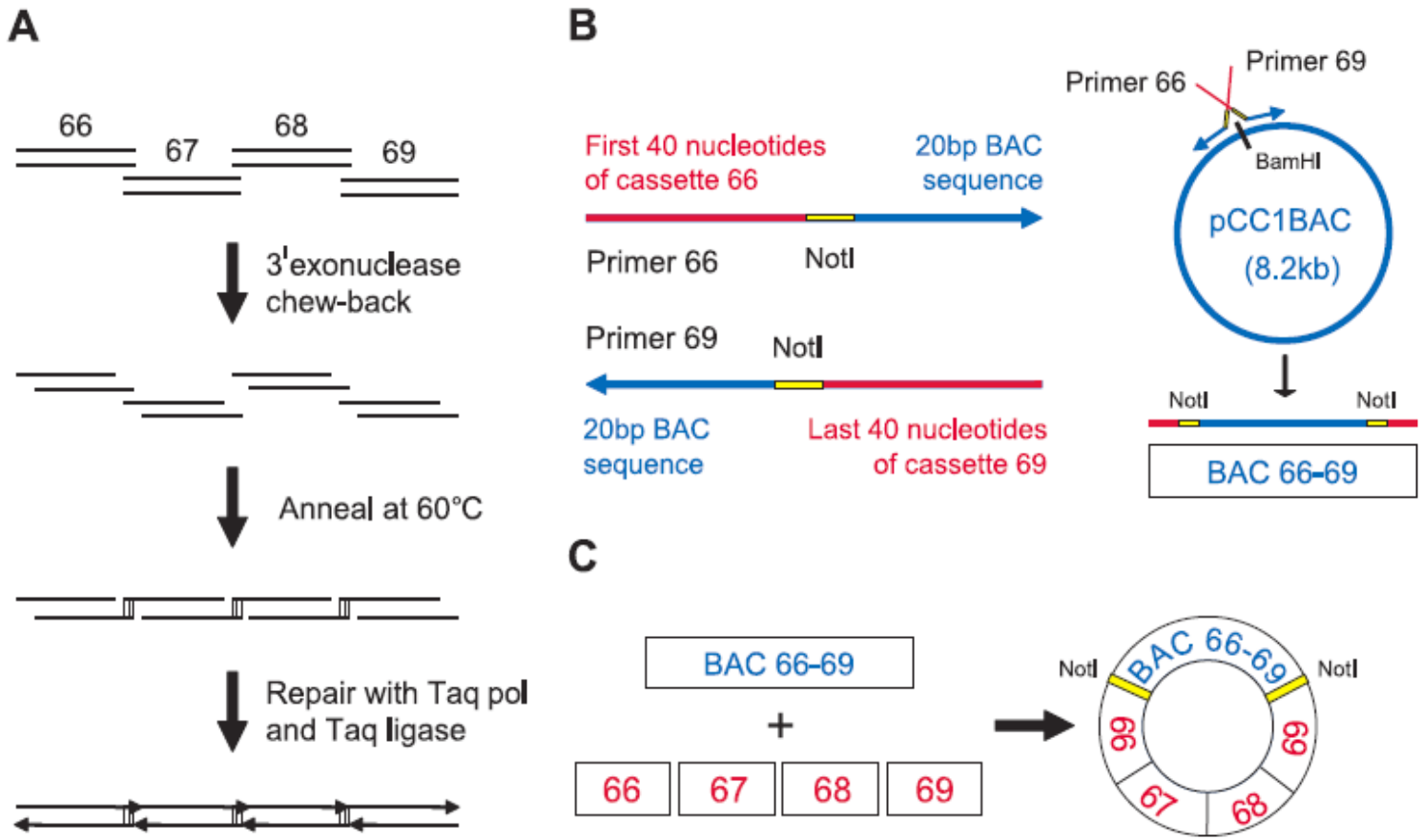


Fig. 3. Assembly of cassettes by in vitro recombination. (**A**) Diagram of steps in the in vitro recombination reaction, using the assembly of cassettes 66 to 69 as an example. (**B**) BAC vector is prepared for the assembly reaction by PCR amplification using primers as illustrated. The linear amplification product, after gel purification, is included in the assembly reaction of (A), such that the desired assembly is circular DNA containing the four cassettes and the BAC DNA as depicted in (**C**).

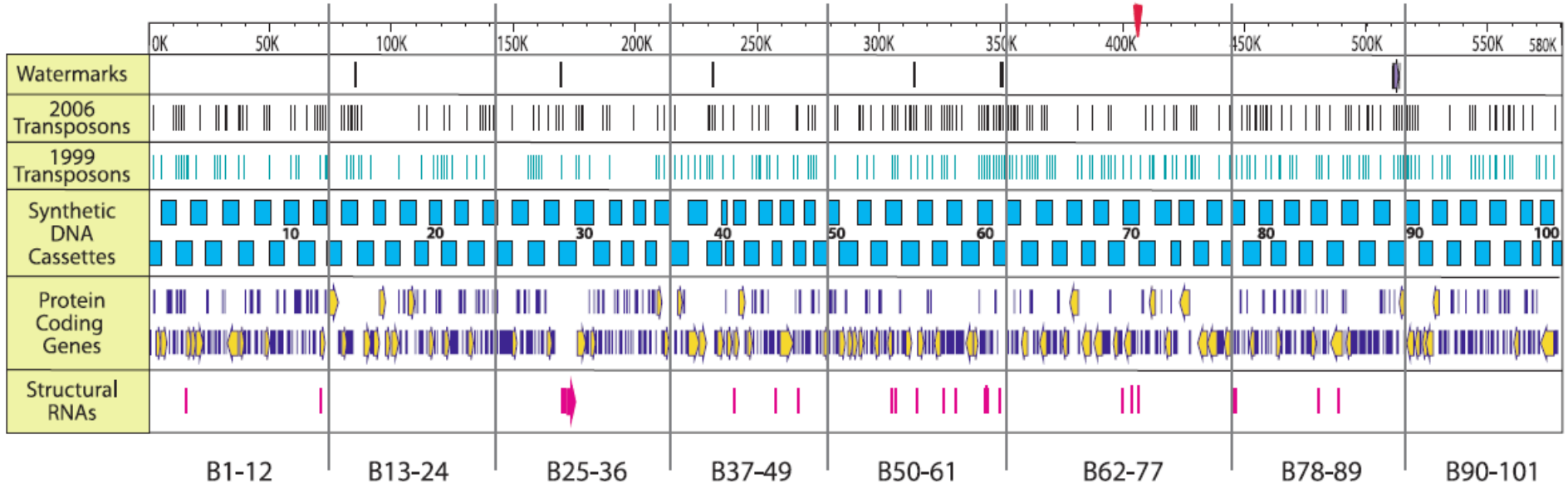
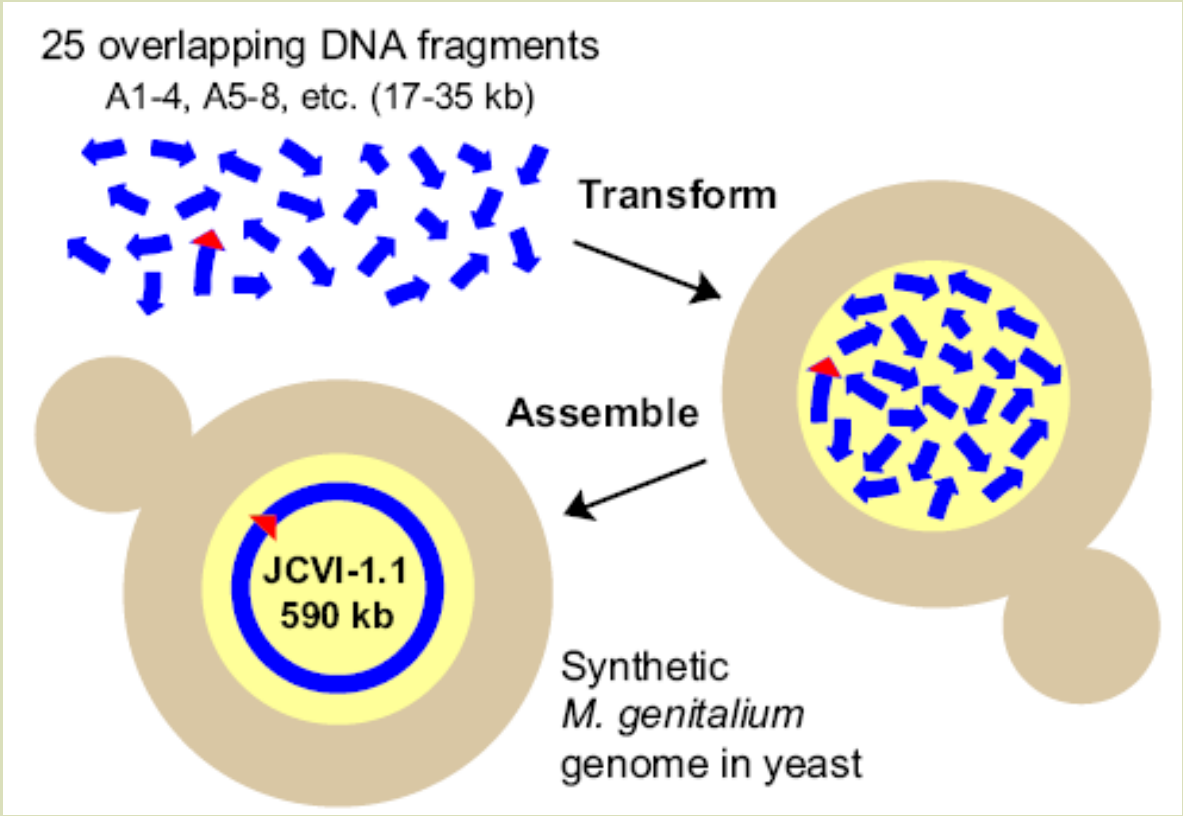


Fig. 1. Linear GenomBench (Invitrogen) representation of the circular 582,970-bp *M. genitalium* JCVI-1.0 genome. Features shown include locations of watermarks and the aminoglycoside resistance marker, viable Tn4001 transposon insertions determined in our 1999 and 2006 studies (*3*, *4*), overlapping synthetic DNA cassettes that comprise the whole genome sequence, 485 *M. genitalium* protein-coding genes, 43 *M. genitalium* rRNA, tRNA, and structural RNA genes, and B-series assemblies (Fig. 2). The red dagger on the genome coordinates line shows the location of the yeast/*E. coli* shuttle vector insertion. Table S1 lists cassette coordinates; table S2 has FASTA files for all 101 cassettes; table S3 lists watermark coordinates; table S4 lists the sequences of the watermarks.

One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome



PNAS | December 23, 2008 | vol. 105 | no. 51

Nucleic Acids Research, 2009, 1–7
doi:10.1093/nar/gkp687

Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides

Daniel G. Gibson*

The J. Craig Venter Institute, Synthetic Biology Group, 9704 Medical Center Drive, Rockville, MD 20850, USA

Received January 1, 2009; Revised August 1, 2009; Accepted August 4, 2009

ABSTRACT

Here it is demonstrated that the yeast *Saccharomyces cerevisiae* can take up and assemble at least 38 overlapping single-stranded oligonucleotides and a linear double-stranded vector in one transformation event. These oligonucleotides can overlap by as few as 20 bp, and can be as long as 200 nucleotides in length. This straightforward scheme for assembling chemically-synthesized oligonucleotides could be a useful tool for building synthetic DNA molecules.

fragments could be assembled by yeast into an entire *Mycoplasma genitalium* genome (17). Subsequently, this process was improved and 25 overlapping fragments, between 17 and 35 kb in length, were assembled at once into this genome (18). Later work showed that six smaller fragments were also acceptable (19).

To turn yeast into a 'factory' able to produce whole genomes and large constructs of any reasonable sequence, what remains is to demonstrate the assembly of chemically synthesized oligonucleotides into appropriate dsDNA molecules in yeast (Figure 1). For this to succeed, a single yeast cell would need to take up all of the necessary oligonucleotides as well as a vector that

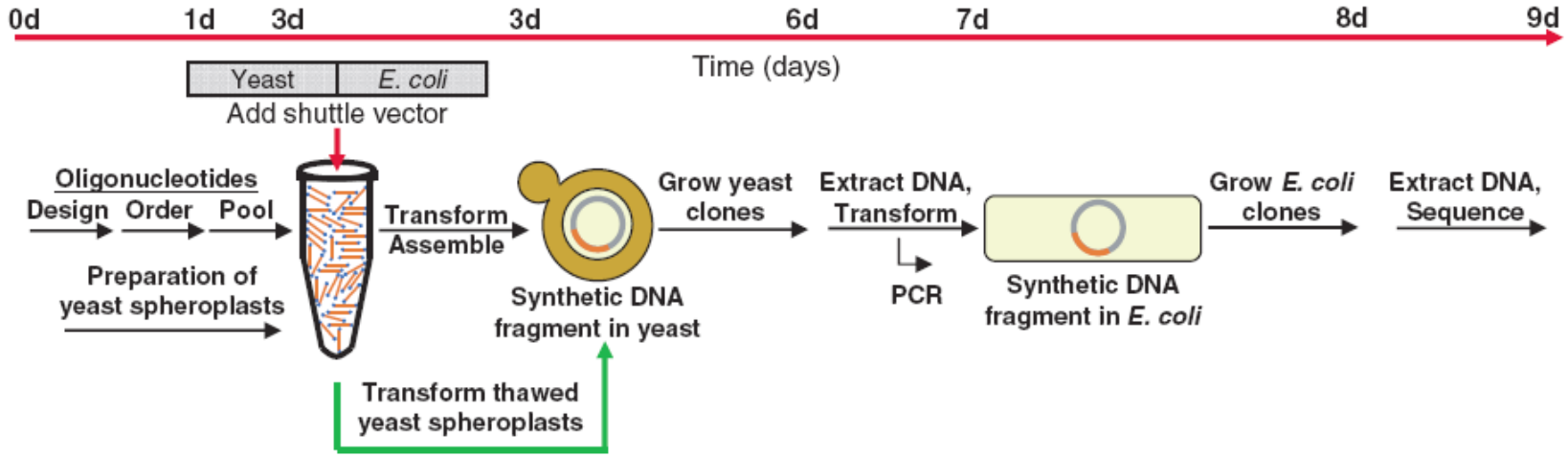
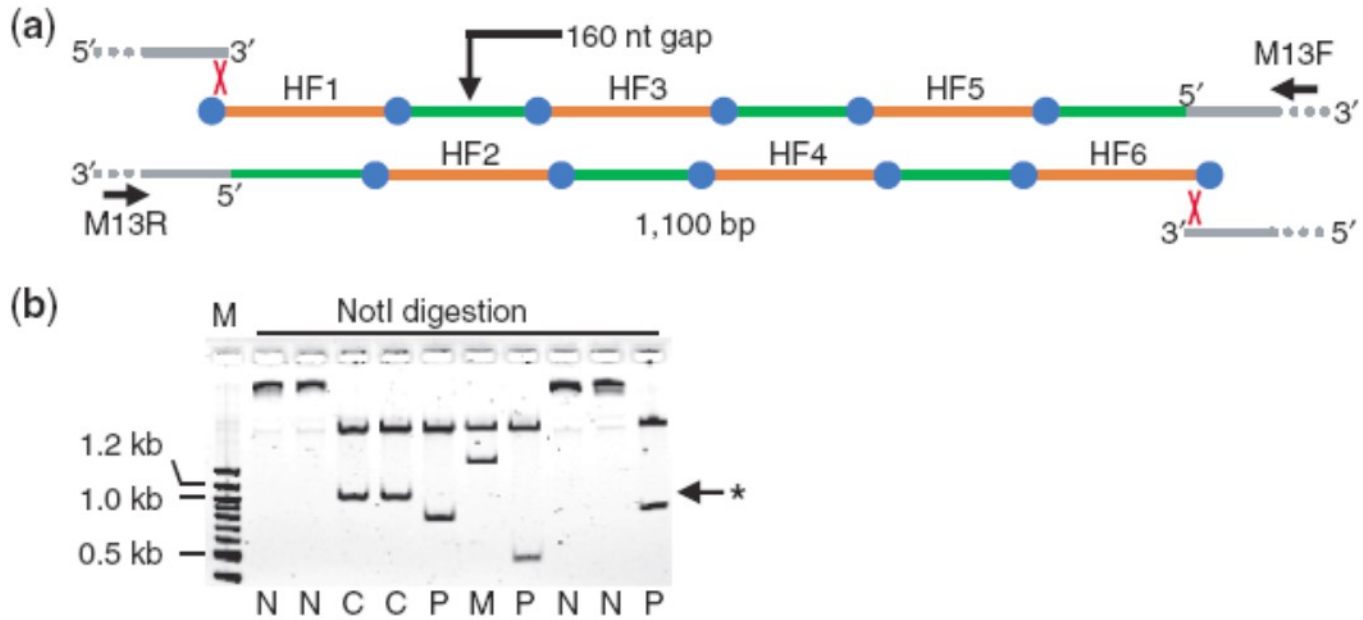
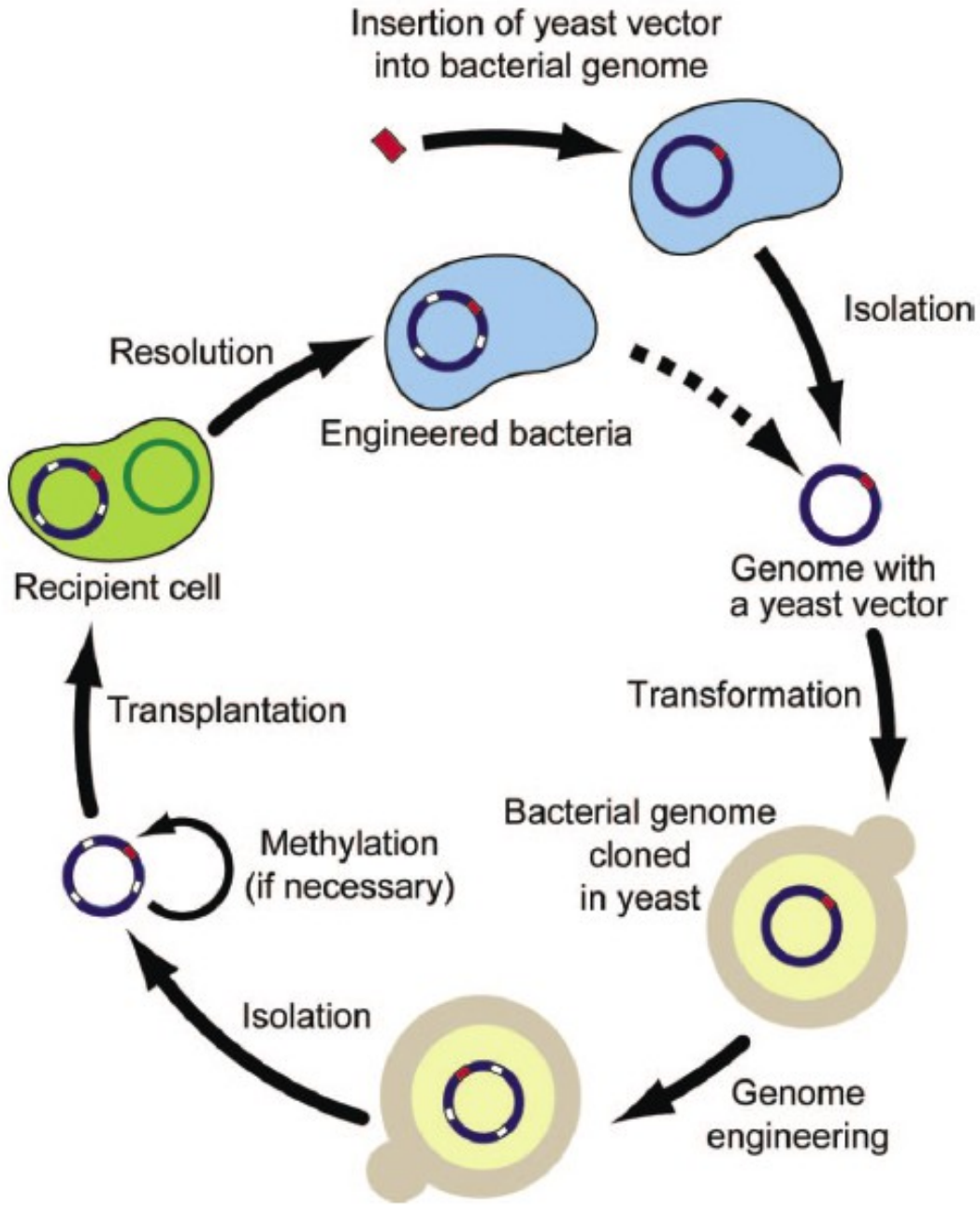


Figure 1. Schematic overview and timeline for the assembly of overlapping ssDNA oligonucleotides (orange lines with blue circles) into a linear dsDNA yeast/*E. coli* shuttle vector (pRS313; grey) within the nucleus of a yeast cell. Following a single transformation event, a synthetic dsDNA fragment (orange) is produced. These fragments are recovered from yeast and then transferred to *E. coli* for more efficient amplification.




Insertion of yeast vector
into bacterial genome
Isolation
Genome with
a yeast vector
Transformation
Bacterial genome
cloned
in yeast
Genome
engineering
Isolation
Methylation
(if necessary)
Transplantation
Recipient cell
Resolution
Engineered bacteria

Programming cells by multiplex genome engineering and accelerated evolution

Harris H. Wang[1,2,3]*, Farren J. Isaacs[1]*, Peter A. Carr[4,5], Zachary Z. Sun[6], George Xu[6], Craig R. Forest[7] & George M. Church[1]

The breadth of genomic diversity found among organisms in nature allows populations to adapt to diverse environments[1,2]. However, genomic diversity is difficult to generate in the laboratory and new phenotypes do not easily arise on practical timescales[3]. Although *in vitro* and directed evolution methods[4–9] have created genetic variants with usefully altered phenotypes, these methods are limited to laborious and serial manipulation of single genes and are not used for parallel and continuous directed evolution of gene networks or genomes. Here, we describe multiplex automated genome engineering (MAGE) for large-scale programming and evolution of cells. MAGE simultaneously targets many locations on the chromosome for modification in a single cell or across a population of cells, thus producing combinatorial genomic diversity. Because the process is cyclical and scalable, we constructed prototype devices that automate the MAGE technology to facilitate rapid and continuous generation of a diverse set of genetic changes (mismatches, insertions, deletions). We applied MAGE to optimize the 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway in *Escherichia coli* to overproduce the industrially important isoprenoid lycopene. Twenty-four genetic components in the DXP pathway were modified simultaneously using a complex pool of synthetic DNA, creating over 4.3 billion combinatorial genomic variants per day. We isolated variants with more than fivefold increase in lycopene production within 3 days, a significant improvement over existing metabolic engineering techniques. Our multiplex approach embraces engineering in the context of evolution by expediting the design and evolution of organisms with new and improved properties.

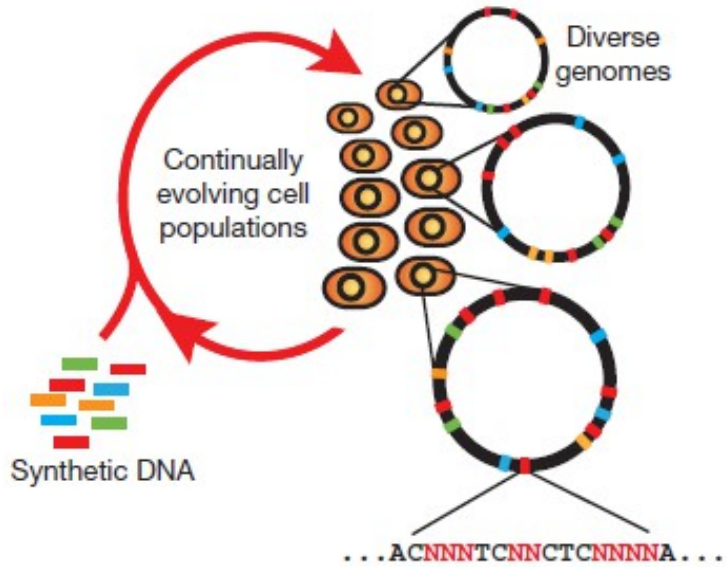


Figure 1 | Multiplex automated genome engineering enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations across a large population of cells through the repeated introduction of synthetic DNA. Each cell contains a different set of mutations, producing a heterogeneous population of rich diversity (denoted by distinct chromosomes in different cells). Degenerate oligo pools that target specific genomic positions enable the generation of a diverse set of sequences at each chromosomal location.

Nature, 2009

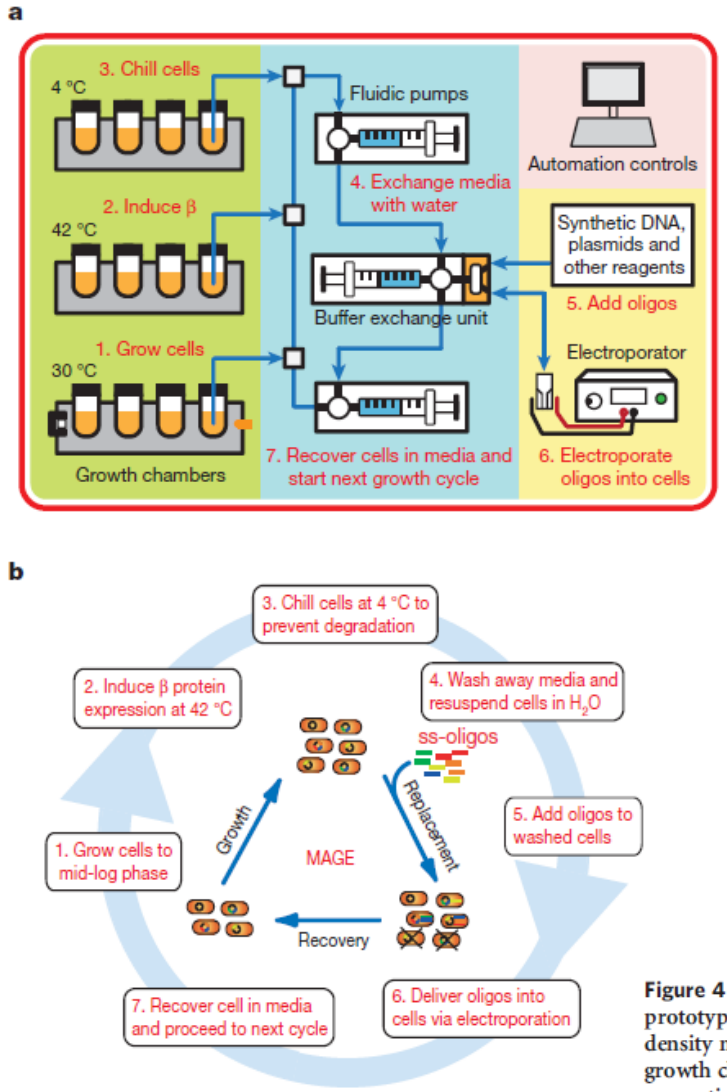


Figure 4 | MAGE automation. **a**, Detailed schematic diagram of MAGE prototype including climate-regulated growth chambers with real-time cell density monitors (green), anti-fouling fluidics for transfer of cells between growth chambers and exchange of media and buffers (blue), and real-time generation of competent cells for transformation with synthetic DNA (yellow). Cultures are carried through different chambers at different temperature regimes (30 °C, 42 °C, 4 °C) depending on the necessary MAGE steps (that is, cell growth, heat-shock, cooling). Cells are made electrocompetent by concentration onto a filter membrane and resuspension with wash buffer. Oligos are delivered into cells by electroporation. **b**, Step-by-step diagram of MAGE cycling steps at a total run time of 2–2.5 h per cycle. Owing to high voltage (18 kV cm^{-1}) electroporation, ~95% of cells are killed at each cycle. Hence, the electroporation event serves to both introduce oligos into cells and to dilute the cell population, cells are then recovered and grown to mid-log phase (7×10^8 cells ml^{-1}) in liquid medium for the subsequent cycle.

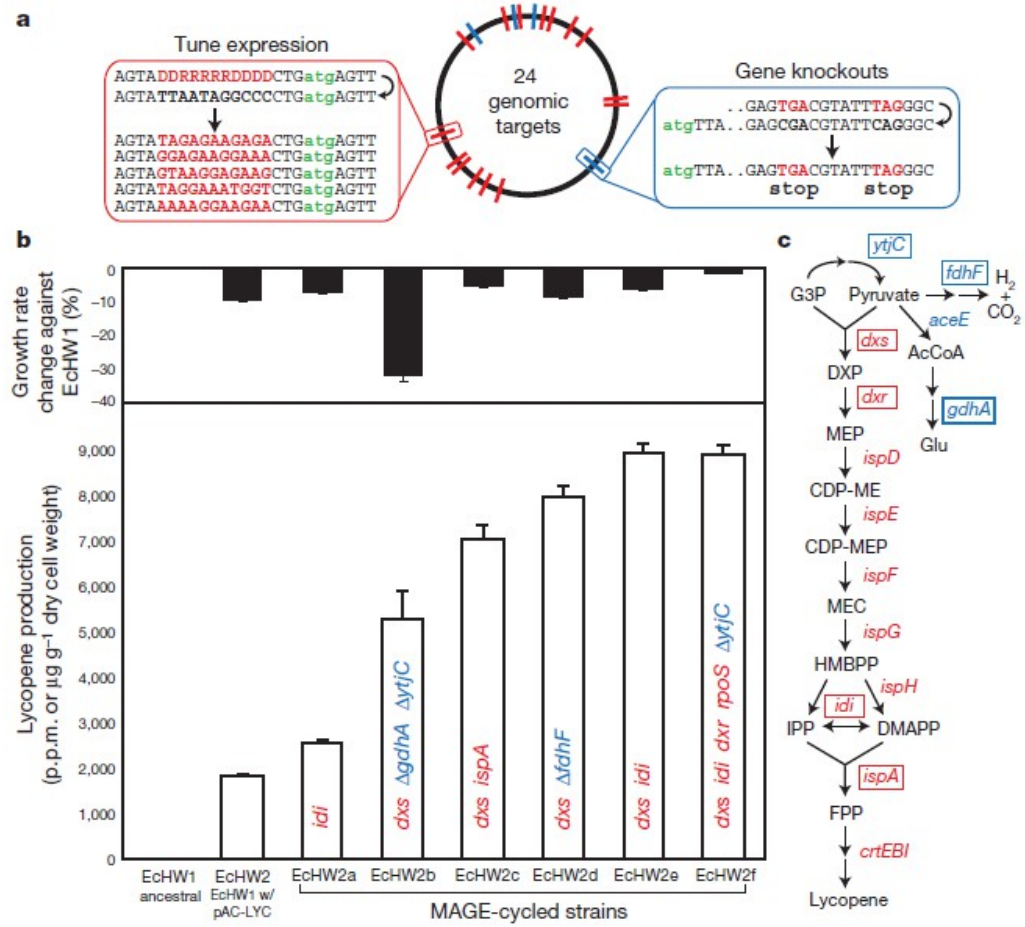
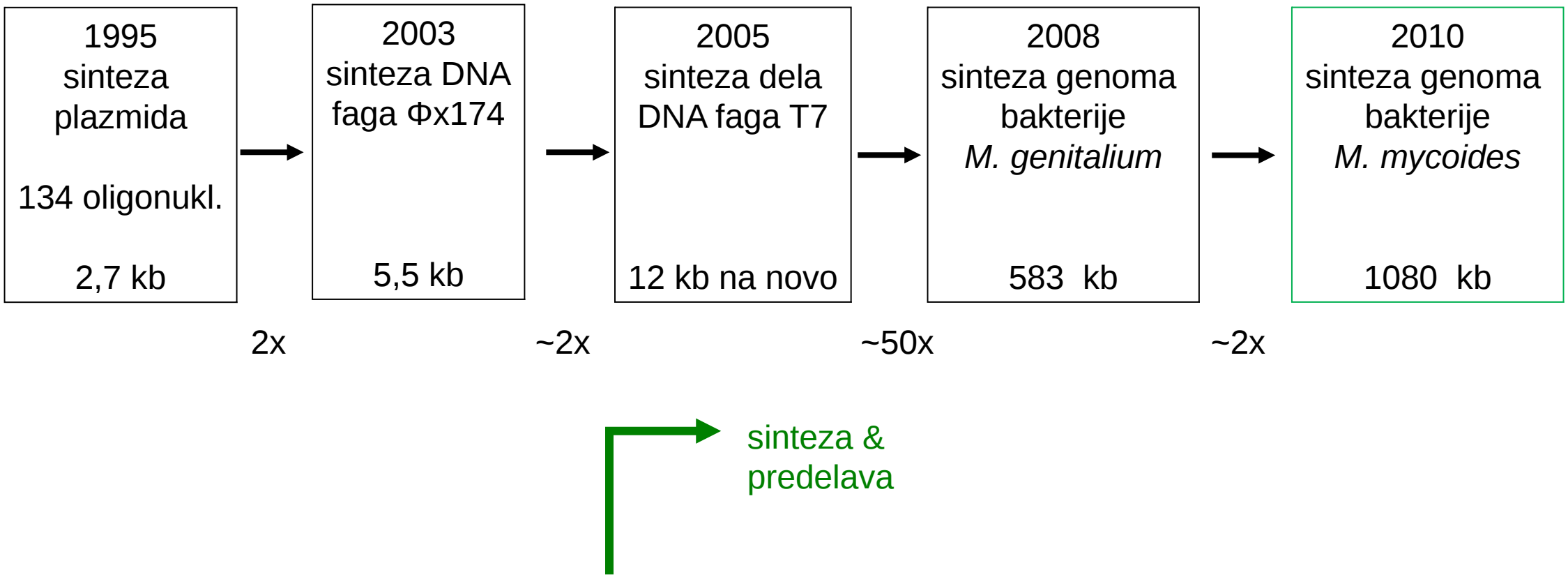


Figure 5 | Optimization of the DXP biosynthesis pathway for lycopene production. **a**, Genomic positions of 24 targeted genes with the RBS optimization strategy on the left (red) and gene knockout strategy on the right (blue). The gene knockout strategy involves the introduction of two nonsense mutations. All 90-mer oligos contain two phosphorothioated bases at the 3′ and 5′ termini. **b**, Black bars represent the growth rate of isolated variants (EcHW2a–f) relative to the ancestral EcHW1 strain. White bars represent lycopene production in p.p.m., which is normalized by dry cell weight in ancestral and mutant strains. Colour-coded labels in each white bar represent genetic modifications found by sequencing. All error

Table 1. The smallest published natural genomes

Metabolic mode	Lifestyle	Organism	Environment	Genome size (kb)	ORFs	G+C content (%)	Cell size, diameter (mm)*
A. Heterotrophy	A.1. Symbionts	*Mycoplasma genitalium*†	Human cell	580	477	31.7	0.3
		Nanoarchaeum equitans‡	*Ignicoccus* sp.	490	536	31.6	0.4
		Buchnera aphidicola BCc§¶	*Cinara cedri*	420	362	20.1	3.0
		Carsonella ruddii§‖	*Pachypsylla venusta*	159	182	16.0	3.0
		*Sulcia muelleri***††	*Homalodisca coagulate*	245	227	22.4	30.0
	A.2. Free living	*Pelagibacter ubique*‡‡	Sea water	1308	1354	29.7	0.3
B. Autotrophy	B.1. Symbionts	*Ruthia magnifica*§	*Calyptogena magnifica*	1200	976	34.0	1.0
		Vesicomyosocius okutanii§	*Calyptogena okutanii*	1000	937	31.6	1.0
	B.2. Free living	*Prochlorococcus marinus*§§	Sea water	1657	1717	30.8	0.6


1995
sinteza
plazmida
134 oligonukl.
2,7 kb
2003
sinteza DNA
faga Φx174
5,5 kb
2005
sinteza dela
DNA faga T7
12 kb na novo
2008
sinteza genoma
bakterije
M. genitalium
583 kb
2010
sinteza genoma
bakterije
M. mycoides
1080 kb
2x
~2x
~50x
~2x
sinteza &
predelava

Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,[1] John I. Glass,[1] Carole Lartigue,[1] Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Mikkel A. Algire,[1] Gwynedd A. Benders,[2] Michael G. Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] Chuck Merryman,[1] Sanjay Vashee,[1] Radha Krishnakumar,[1] Nacyra Assad-Garcia,[1] Cynthia Andrews-Pfannkoch,[1] Evgeniya A. Denisova,[1] Lei Young,[1] Zhi-Qing Qi,[1] Thomas H. Segall-Shapiro,[1] Christopher H. Calvey,[1] Prashanth P. Parmar,[1] Clyde A. Hutchison III,[2] Hamilton O. Smith,[2] J. Craig Venter[1,2]*

We report the design, synthesis, and assembly of the 1.08–mega–base pair *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a *M. capricolum* recipient cell to create new *M. mycoides* cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including "watermark" sequences and other designed gene deletions and polymorphisms, and mutations acquired during the building process. The new cells have expected phenotypic properties and are capable of continuous self-replication.

SCIENCE VOL 329 2 JULY 2010

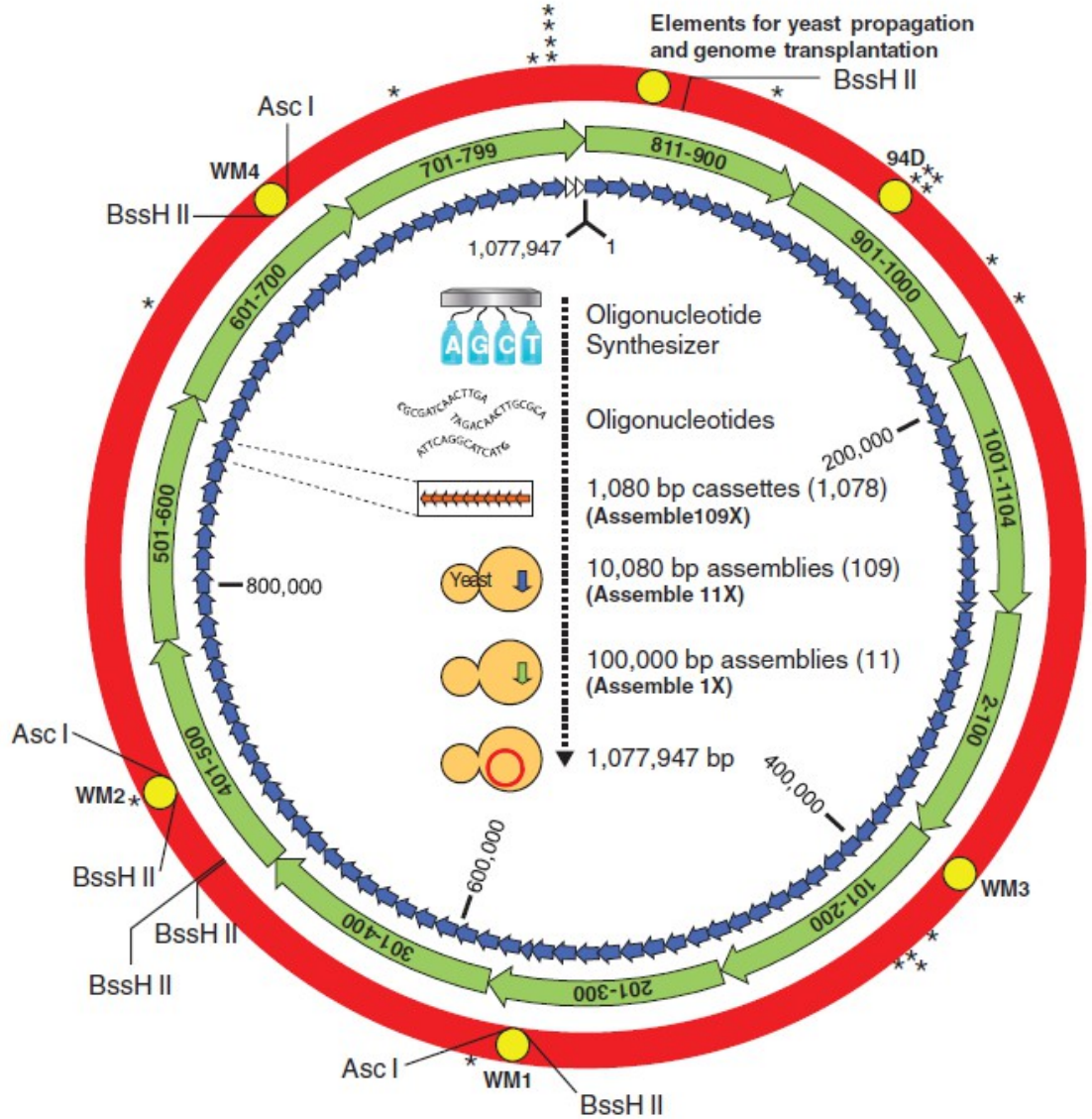


Fig. 1. The assembly of a synthetic *M. mycoides* genome in yeast. A synthetic *M. mycoides* genome was assembled from 1078 overlapping DNA cassettes in three steps. In the first step, 1080-bp cassettes (orange arrows), produced from overlapping synthetic oligonucleotides, were recombined in sets of 10 to produce 109 ~10-kb assemblies (blue arrows). These were then recombined in sets of 10 to produce 11 ~100-kb assemblies (green arrows). In the final stage of assembly, these 11 fragments were recombined into the complete genome (red circle). With the exception of two constructs that were enzymatically pieced together in vitro (*27*) (white arrows), assemblies were carried out by in vivo homologous recombination in yeast. Major variations from the natural genome are shown as yellow circles. These include four watermarked regions (WM1 to WM4), a 4-kb region that was intentionally deleted (94D), and elements for growth in yeast and genome transplantation. In addition, there are 20 locations with nucleotide polymorphisms (asterisks). Coordinates of the genome are relative to the first nucleotide of the natural *M. mycoides* sequence. The designed sequence is 1,077,947 bp. The locations of the Asc I and BssH II restriction sites are shown. Cassettes 1 and 800-810 were unnecessary and removed from the assembly strategy (*11*). Cassette 2 overlaps cassette 1104, and cassette 799 overlaps cassette 811.

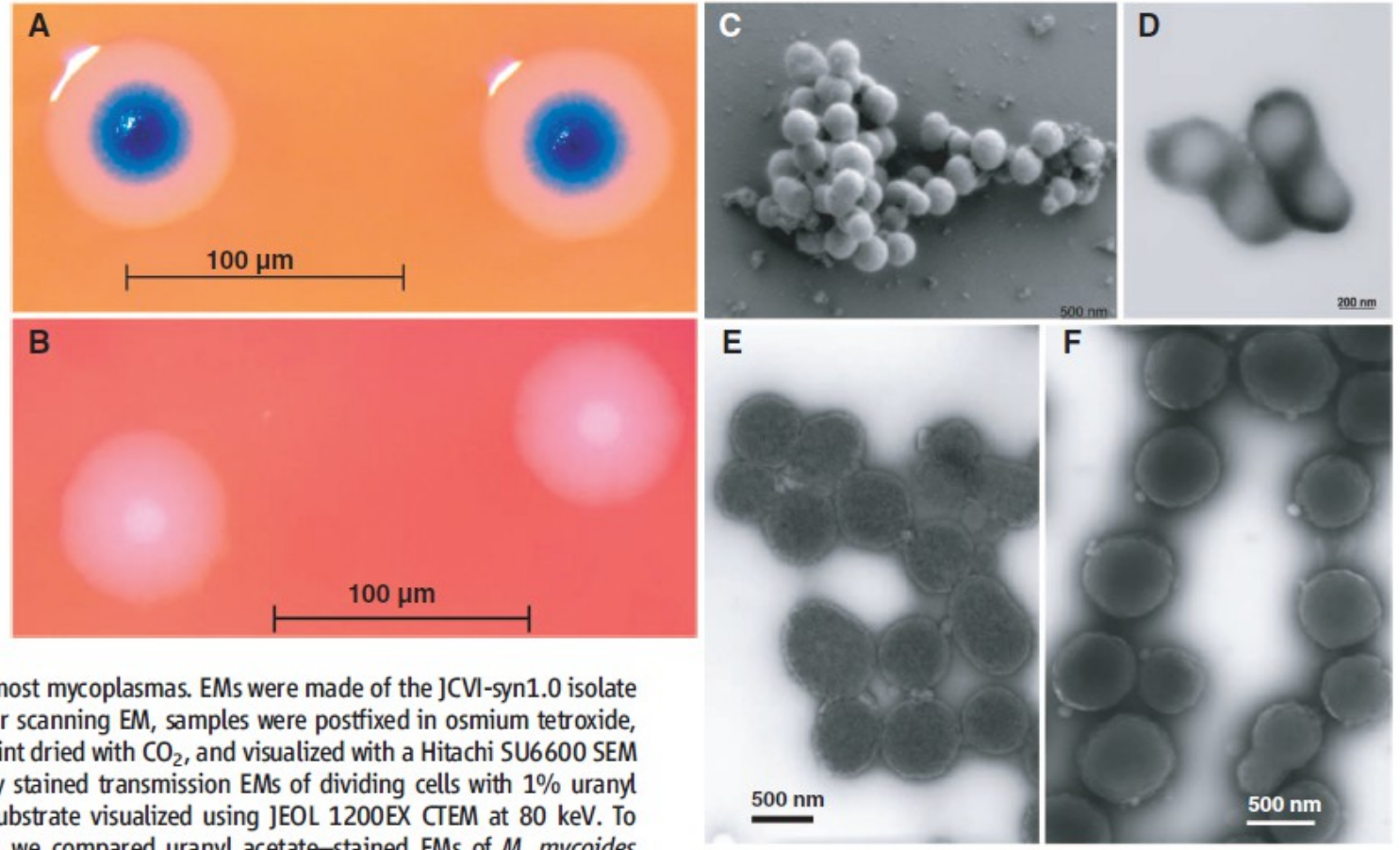


Fig. 5. Images of *M. mycoides* JCVI-syn1.0 and WT *M. mycoides*. To compare the phenotype of the JCVI-syn1.0 and non-YCp WT strains, we examined colony morphology by plating cells on SP4 agar plates containing X-gal. Three days after plating, the JCVI-syn1.0 colonies are blue because the cells contain the *lacZ* gene and express β-galactosidase, which converts the X-gal to a blue compound (**A**). The WT cells do not contain *lacZ* and remain white (**B**). Both cell types have the fried egg colony morphology characteristic of most mycoplasmas. EMs were made of the JCVI-syn1.0 isolate using two methods. (**C**) For scanning EM, samples were postfixed in osmium tetroxide, dehydrated and critical point dried with CO_2, and visualized with a Hitachi SU6600 SEM at 2.0 keV. (**D**) Negatively stained transmission EMs of dividing cells with 1% uranyl acetate on pure carbon substrate visualized using JEOL 1200EX CTEM at 80 keV. To examine cell morphology, we compared uranyl acetate–stained EMs of *M. mycoides* JCVI-syn1.0 cells (**E**) with EMs of WT cells made in 2006 that were stained with ammonium molybdate (**F**). Both cell types show the same ovoid morphology and general appearance. EMs were provided by T. Deerinck and M. Ellisman of the National Center for Microscopy and Imaging Research at the University of California at San Diego.

- TetR, LacZ,… vodni tisk (*watermark*)
- gostiteljske celice se ne odzivajo na metilacijski vzorec
- preverjanje funkcionalnosti segmentov dolžine 100.000 bp (kombinacije z naravnim preostankom genoma