PNAS
www.pnas.org

Supplementary Information for

Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality

Jonathan E. Venetz, Luca Del Medico, Alexander Wölfle, Philipp Schächle, Yves Bucher, Donat Appert, Flavia Tschan, Carlos E. Flores-Tinoco, Mariëlle van Kooten, Rym Guennoun, Samuel Deutsch, Matthias Christen and Beat Christen

Corresponding Authors: Matthias Christen and Beat Christen.
E-mail: matthias.christen@imsb.biol.ethz.ch; beat.christen@imsb.biol.ethz.ch

This PDF file includes:

Supplementary text
Figs. S1 to S5
Tables S1 to S6
Captions for Databases S1 to S3
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Databases S1 to S3

Supporting Information Text

Materials and Methods

A. Media and supplements.

Microbial growth conditions. Standard culturing conditions were used to grow microbial strains. *E. coli* strains were cultured in liquid or solid Luria-Bertani (LB) medium at 37°C while *C. crescentus* (*Caulobacter* thereafter) strains were cultured in liquid or solid peptone-yeast extract (PYE) medium at 30°C respectively. *S. cerevisiae* strains were cultured in standard YPD medium supplemented with Adenine hemisulfate (80 $mg{\cdot}l^{-1}$) or synthetic defined medium (SD) at 30°C. Unless otherwise indicated, antibiotics and other supplements were used at the following concentrations: (i) *E. coli*: 10 $\mu g{\cdot}ml^{-1}$ gentamicin (Gm), 50 $\mu g{\cdot}ml^{-1}$ kanamycin (Km), 100 $\mu g{\cdot}ml^{-1}$ ampicillin (Amp); (ii) *Caulobacter*: 20 $\mu g{\cdot}ml^{-1}$ Km, 20 $\mu g{\cdot}ml^{-1}$, nalidixic acid (Na), 10 $\mu g{\cdot}ml^{-1}$ Gm; (iii) *S. cerevisiae*: adenine hemisulfate (80 $mg{\cdot}l^{-1}$), D-glucose (20 $g{\cdot}l^{-1}$)

B. Design of *C. eth-1.0* genome and sequence rewriting into *C. eth-2.0*.

Compilation of the C. eth-1.0 genome design. The comprehensive list of DNA sequences used for the design of *C. eth-1.0* comprised essential DNA loci according to a previously reported essential genome data set from *Caulobacter* NA1000 obtained for growth under rich-medium conditions (1). Cumulatively, *C. eth-1.0* encompasses 575 essential and semi-essential protein-coding sequences, 54 non-coding sequences as well as 1,015 intergenic sequences including gene regulatory features such as promoters and terminators. In addition, 15 redundant genes involved in cellular core metabolism and DNA recombination pathways were also included. Furthermore, 86 non-essential genes were added to the *C. eth-1.0* genome design as control genes for subsequent functionality assessment. Essential and semi-essential DNA sequences were compiled into the 785,701 base-pair (bp) *C. eth-1.0* genome design (2) with order and orientation maintained according to the native *Caulobacter* genome (NCBI Accession: NC_011916.1). To select for faithful assembly and permit stable maintenance in *S. cerevisiae*, auxotrophic marker genes (TRP1, HIS3, MET14, LEU2, ADE2) and a set of 10 ARS sequences (ARS_Max2 (3), ARS416, ARS1018, ARS1113, ARS1213, ARS516, ARS_HI (3), ARS727, ARS4, ARS1323) were seeded across the genome design. Furthermore, the pMR10Y (4) shuttle-vector sequence permitting replication in *S. cerevisiae*, *E. coli* and *Caulobacter* was inserted between the genes CETH_00003 and CETH_03878. The pMR10Y shuttle vector sequence consists of a broad-host range RK2 replicon, kanamycin resistance marker gene, the URA3 gene, CEN, ARS209 sequences and conjugational transfer origin (oriT) (4).

Synthesis optimization and sequence rewriting into the C. eth-2.0 genome design. The *C. eth-1.0* design was rewritten to implement DNA synthesis sequence optimization using the previously reported Genome Calligrapher algorithm and sequence design pipeline (4). Synthesis constraints and disallowed sequences impeding large-scale *de novo* DNA synthesis were refactored by neutral rewriting of protein-coding sequences (synonymous codon replacement) and applying desired base-substitutions within intergenic sequences. Cumulatively, sequence refactoring resulted in removal of a total of 6,713 synthesis and DNA assembly constraints, including removal of 4,342 (100%) high-GC regions (GC content larger than 0.8 within a 99 bp window), 87.2% (767 out of 880) of all direct repeats and 76.9% (466 out of 606) of all hairpin structures with repeat size equal or larger than 8bp, 66.9% (93 out of 139) homopolymeric sequence stretches (larger than six G, eight C, nine A or T) and all 1,045 instances of endonuclease sites for AarI, BsaI, BspQI, PacI, PmeI, I-CeuI and I-SceI enzymes from the *C. eth-1.0* genome design. The GC content was lowered from 66.2% of the *C. eth-1.0* design to 57.4% in the new *C. eth-2.0* design, by requiring GC content within protein-coding sequences not to exceed 0.7 for a 99bp window and not to exceed 0.85 within a 21bp window. Similarly, the AT content was set not to exceed 0.7 within a 99bp window and not to exceed 0.85 within a 21bp window.

Rewriting of protein-coding sequences. Synonymous rewriting of protein-coding sequences was applied in addition to synthesis optimization. The average codon rewriting probability was set to 0.56, resulting in introduction of 133,313 base substitutions across the 785,701 bp *C. eth-2.0* genome design. A subset of segments (segment 7-11 and 29-31) was rewritten in gradual increments with rewriting rates from 12.5% up to 100% implemented. The first four amino acid codons of protein-coding sequences were excluded from rewriting to preserve potential translational initiation signals. Furthermore, rare codons AGT, ATA, AGA, GTA and AGG were set as immutable codons and were neither replaced or introduced upon rewriting, except when required for removal of type IIS restriction sites. The two leucine codons TTA and TTG were erased throughout the *C. eth-1.0* genome design. Furthermore, 163 out of 173 instances of the amber stop codon (TAG) were removed. Cumulatively, sequence refactoring resulted in 123,562 synonymous codon substitutions out of a total of 220,263 codons present across the genome design. To distinguish rewritten genes from the native *Caulobacter* counterparts, suffixes of gene IDs were changes in *C. eth-2.0* from CCNA_ to CETH_ while maintaining the gene ID number (i.e. CCNA_00008 to CETH_00008).

Impact of rewriting on alternative ORFs and CDS internal motifs. While the massive scale of synonymous rewriting ($>$ 56% of all codons replaced) applied to rewrite the *C. eth-1.0* genome maintains protein-coding sequences of a given annotated CDS, while additional overlapping open-reading frame (ORFs) were altered. Similarly, synonymous rewriting erased features and motifs internal to CDS including transcriptional and translational control signals. To assess the impact of synonymous rewriting on alternative genome features and CDS internal motifs, a detailed sequence comparison between the *C. eth-1.0* genome design prior and after rewriting into the *C. eth-2.0* genome design was carried out using BioPython (5). The complete set of alternative ORFs was identified using the Python regex module with the regular expression search pattern [' ((ATG)|(GTG)|(TTG)) ([ATGC] {3,3}) + ? ((TAA)|(TAG)|(TGA))'] across the forward and reverse strand of the *C. eth-1.0* genome sequence prior and

after rewriting. ORFs smaller than 50 bp were discarded. To detect alternative ORFs to the current genome annotation, ORFs were grouped according to shared stop codon positions (ORF groups). Within ORF groups, smaller ORFs sharing identical stop codon positions as the annotated CDS were discarded. This procedure detected in addition to the 676 annotated CDS 3,229 alternative ORFs. For each CDS and alternative ORF, protein identity prior and after rewriting was calculated using BioPython string comparison. From a total of 3,229 alternative ORFs 2,822 (87.4%) showed non-synonymous mutation rates exceeding 20% and were classified as erased upon rewriting. Furthermore, in 2,113 ORFs, (74.9%) a pre-matured stop codon (non-sense mutation) was introduced upon rewriting leading to a truncated protein product. To analyse the effect of rewriting on CDS internal ribosome binding and stalling sites, sequences matching to the bacterial Shine-Dalgarno (SD) consensus sequences were identified across the native *C. eth-1.0* genome design and compared against the corresponding rewritten sequences of the *C. eth-2.0* genome. SD consensus sequences were defined by requiring a match to the hepta-nucleotide sequence 'AAGGAGG' with less than one mismatch or no more than two A to G substitutions. CDS internal ribosome binding sites were classified as erased from the *C. eth-2.0* genome design if rewritten sequences no longer matched the SD consensus sequence according to the above criteria. Similarly, CDS internal transcriptional start sites (TSS) according to a previous reported experimental dataset (6) were classified as erased if, within a 50 bp window upstream of the TSS site, more than 10% of all nucleotides were substituted upon rewriting.

Retro-synthetic partitioning of the C. eth-2.0 genome design. To define the optimal retro-synthetic assembly route, the synthesis optimized *C. eth-2.0* genome design was partitioned using the previously published Genome Partitioner algorithm (2). The partitioning strategy was designed as a four-tier hierarchical assembly process comprised of 236 blocks each 3-4 kb in size (assembly level 1) that were assembled into 37 approximately 20 kb genome segments (assembly level 2) that were further assembled into 16 mega-segments of 40-60 kb in size (assembly level 3) and further assembled into the final 785 kb genome construct. Mega-segment boundaries were set to split 5 auxotrophic markers (TRP1, HIS3, MET14, LEU2, ADE2) seeded at equidistance into the *C. eth-2.0* genome design to serve as click-markers to direct faithful assembly of the whole genome in yeast. Furthermore, segments were required to encompass only intact gene sets to prevent splitting of individual genes and allow for functionality testing at the level of individual segments. Terminal homology overlaps of each assembly level were optimized by the Genome Partitioner algorithm (2) using standard settings (maximal repeat size set to 8 bp, non-specific overlap set to 8 bp). Standard flanking 3' prefix and 5' suffix adapter sequences were previously reported (2).

C. Synthesis and hierarchical assembly of the *C. eth-2.0* genome.

Low-cost DNA synthesis by commercial suppliers. The 236 partitioned 3-4kb assembly blocks used for hierarchical assembly of the *C. eth-2.0* genome design were ordered from two commercial suppliers of low-cost *de novo* DNA synthesis as sequence verified plasmid cloned constructs (Gen9 Inc and GeneArt, Thermo Fisher Scientific, Regensburg Germany). A total of 236 assembly blocks were ordered from the former DNA synthesis supplier. Thereof, 182 were manufactured by low-cost synthesis (77% synthesis success rate). From a second commercial provider (GeneArt, Thermo Fisher Scientific, Regensburg Germany) 54 out of 55 assembly blocks were manufactured by low-cost DNA synthesis (98% synthesis success rate), while for a single remaining assembly block synthesis was accomplished by custom gene synthesis. Constructs were delivered as plasmid cloned constructs in the commercial maintenance vectors pG9m, GeneArt pMS-RQ and pMA-RQ.

BspQI-mediated release of assembly blocks from maintenance vector. Assembly blocks were released from the maintenance vector via a BspQI type IIS restriction digestion. Digestion reactions to release individual 3-4 kb assembly blocks consisted of a 40 µl endonuclease digestion reaction composed of 10 µl (> 5 µg) purified maintenance plasmids containing the appropriate assembly block, 1 µl (10 U) BspQI type IIS restriction enzyme (NEB, USA), 4 µl 10x NEBuffer 3.1 (NEB, USA), and 25 µl nuclease-free water (Promega, USA). The digestion reactions were incubated at 50°C for 1.5 h prior heat-inactivation of the BspQI enzyme at 80°C for 20 min. The restriction digested assembly blocks were purified using the NucleoSpin® Gel and PCR clean up Kit (Macherey-Nagel, Switzerland) and quantified on a Nanodrop ND-1000 spectrometer (Thermo Fisher Scientific, Carlsbad, USA).

Yeast assembly of segment and mega-segments. Column-purified 3-4kb DNA blocks were used for assembly of the 37 segments into the pMR10Y (4) shuttle-vector (pMR10::CEN/ARS::URA3) plasmid. From a mid-log phase *S. cerevisiae* culture (VL6-48N (7), OD_{600} of 0.7) 2 ml of cells were collected by centrifugation and washed in 1 ml 0.9% NaCl-solution. To the washed cells, 100 ng salmon sperm DNA (single stranded from salmon testes, D7656, Sigma-Aldrich, USA), 540 ng linearized pMR10Y (digested with PacI and PmeI) and 300 ng of each DNA block was added. The pellet was resuspended in 500 µl transformation mixture (400 µl 50% PEG solution, 50 µl 1 M Lithium acetate, 50 µl distilled water) and 57 µl DMSO was added prior incubated at RT for 15 min, followed by incubation for 15 min at 42°C. Cells were harvested by centrifugation, resuspended and plated onto selective yeast synthetic defined medium (supplemented with glucose (10 $g{\cdot}L^{-1}$) and adenine (80 $mg{\cdot}L^{-1}$) but lacking uracil) and incubated at 30°C for two days. For the assembly of mega-segments from segments a similar Lithium acetate transformation protocol (8) was used with following modification. The segments were released from the pMR10Y vector by PacI and/or PmeI restriction digestion. The first segment composing the assembled mega-segment was digested using PmeI. The last segment of the mega-segment was digested using PacI. For mega-segments composed of three segments, the centre segment was released by a PacI and PmeI restriction digest. The 20 µl endonuclease digestion reaction consisted of 1 µg segment DNA, 0.5 µl (5 U) PacI / PmeI, 2 µl 10x CutSmart Buffer (NEB, USA) and water to fill up to 20 µl. The restriction digest was conducted for 1 h

at 37°C prior heat-inactivation of the restriction enzyme at 65°C for 20 min. Of each restriction digest, 4 µl (200 ng) of the released segments were transformed into *S. cerevisiae* using the protocol described above.

Assessment of autonomously replicating sequence functionality. Initial attempts to assemble the *C. eth-2.0* chromosome in yeast resulted in isolation of partially assembled chromosome constructs bearing mega-segments 1 to 6 as determined by whole genome sequencing. However, no clones with larger constructs were detected. Close inspection of initial partial assembly products revealed that assemblies failed at the 5'-end of the mega-segment VI, which contains the His3 click marker and the ARS416 at the 3'-end. The Met14 marker and the ARS1213 are located 100 kb downstream of mega-segment VI. We reasoned that either one or multiple ARS sequences were non-functional to promote replication of the GC-rich *C. eth-2.0* chromosome sequence in *S. cerevisiae*. We benchmarked the functionality of ARS elements via assembly of adjacent segments containing a click marker and an ARS element with two different linker sequences. The first linker contained a Cen6 and an ARSH4 site, while the second linker contained a Cen6 site only. The first linker was used as a positive control since the replication of the assembled segments is maintained by the ARSH4 site. The second linker was used to test the functionality of the ARS elements seeded in the design, as the linker itself does not permit for DNA replication. This experiment showed that ARS416 and ARS1213, located at the mega-segment junctions 5 and 6 and 8 and 9 are non-functional while the ARS_Max2 and ARS_HI located at the mega-segment junctions 2 and 3, and 10 and 11 respectively were functional. To repair the initial design, five additional ARS elements (ARS1018, ARS1113, ARS516, ARS727, and ARS1323) were integrated at the 5'-end of the different mega-segments 5,6,9,12 and 15.

Addition of ARS to mega-segments. ARS elements for the repair of the *C. eth-2.0* chromosome design were selected according to following criteria. ARS had to be shorter than 250bp and not contain any of the restriction sites needed for the synthesis of the construct (AarI, BsaI, BspQI, PacI, PmeI). Selected ARS elements satisfying these criteria were flanked by 70bp overhangs to the 3'-end of the mega-segment and the pMR10Y vector and were ordered as G-Blocks (IDT, Skokie, IL, USA). The mega-segments were PmeI restriction digested as described above. To assemble ARS sequences into mega-segments, the following protocol was used. The *S. cerevisiae* (VL6-48N) pre-culture was grown in 5 ml YPD to high density, diluted 1:25 in 50 ml YPD and grown for 4 h prior collection of cells by centrifugation at 1,000 rcf for 5 min, the supernatant was replaced by 25 ml MQ. The cells were centrifuged at 3,000 rcf for 5 min, the supernatant was discarded. The pellet was dissolved in 100 µl Lithium acetate-mix (0.1 M Lithium acetate, 0.01 M Tris-HCl pH 7.5, 0.001 M EDTA pH 8.0) per transformation. To the LiAc-cell mix, 10 µl salmon-sperm DNA (1% w/v salmon-sperm DNA (ssDNA), 0.01 M Tris-HCl pH 7.5, 0.001 M EDTA pH 8.0) and 600 µl PEG-mix (40% w/v Poly(ethylene glycol) 3015-3685 g/mol, 0.01 M Tris-HCl pH 7.5, 0.001 M EDTA pH 8.0) were added per transformation. Of this master mix, 710 µl were aliquoted into 1.5 ml tubes, to which 200 ng of the digested mega-segment and 15 ng of the G-Block was added. The samples were incubated at RT for 30 min on a shaker (Titramax100, Heidolph, Schwabach, DE) at 300 rpm. Following incubation, 70 µl DMSO was added and the samples were incubated at 42°C for 15 min. The cells were centrifuged at 1,000 rcf for 2 min, the supernatant was replaced by 300 µl MQ, 150 µl of the transformants was plated onto SD plates lacking Uracil. The plates were incubated at 30°C, for two days. The correct integration of the ARS was confirmed by diagnostic PCR and Sanger sequencing (Microsynth, Balgach, CH).

Verification of higher-order assemblies in S. cerevisiae using diagnostic PCR. Correct higher-order assemblies for segment and mega-segment assemblies were verified by performing diagnostic PCR reactions across assembly junctions. *S. cerevisiae* colonies were resuspended in 3 µl 0.02 M sodium hydroxide, and incubated for 10 min at 99°C in a thermocycler instrument. The diagnostic PCR reaction contained 5 µl 5 M betaine, 12.5 µl BioRed 2x Mastermix (Bioline, London, UK), 0.5 µl of each diagnostic primer (100 nM) specific for each assembly junction present in the higher-order assembly, 3.5 µl water as well as 3 µl of boiled *S. cerevisiae* cells as template. PCR reactions were cycled in a thermocycler instrument (C1000 touch, BioRad, Cressier, Switzerland) according to following protocol: (1) 5 min at 96°C, (2) 30 s at 96°C, (3) 30 s at 55°C, (4) 1 min at 72°C, (5) repeat steps 2 – 4, 30 times, (6) final elongation 10 min at 72°C. PCR products were analysed on 1.5% agarose gels by electrophoresis to identify clones bearing full-length higher-order assemblies.

Isolation of higher-order constructs from S. cerevisiae. The yeast assembled *C. eth-2.0* chromosome segments and mega-segments cloned into the pMR10Y shuttle vector were extracted from *S. cerevisiae* and transformed into *E. coli* according to the following procedure. *S. cerevisiae* strains bearing segment or mega-segment constructs were inoculated in 5 ml SD medium lacking Uracil and grown overnight. Cells were harvested by centrifugation and resuspended in 250 µl Zymolyase solution (0.143 M 2-mercaptoethanol, 0.01 M Tris-HCl pH 8.0, 1.2 M Sorbitol, 0.005 M $CaCl_2$, 8 mg/mL Zymolyase 20T from Arthrobacter luteus (AMS Biotechnoloy Europe, UK) followed by incubation at 37°C for 1 h. To each digest, 12 µl 1 M Tris-HCl pH 8.0 and 6 µl 0.5 M EDTA pH 8.0 solution was added followed by addition of 250 µl of lysis buffer (GeneJET Plasmid Miniprep Kit, Thermo Scientific, USA) and incubated for 5 to 7 min. To terminate the lysis process, 350 µl neutralization buffer (GeneJET Plasmid Miniprep Kit, Thermo Scientific, USA) was added, mixed and the cell debris was pelleted by centrifugation for 10 min at 15,000 rcf. The clear supernatant (800 µl) was transferred and DNA was precipitated by addition of 640 µl 2-propanol and pelleted by centrifugation in a microfuge (10 min, 15,000 rcf). The resulting pellet was washed with 500 µl 70% ethanol and centrifuged for 5 min, dried and resuspended in 50 µl 0.1 M Tris-HCl buffer, pH 8.0.

Transformation of higher-order constructs into E. coli and isolation. To transform higher-order constructs into *E. coli*, 2 µl of the isolated pMR10Y plasmid DNA was electroporated into *E. coli* (DH5α / DH10B, (90 µl aliquots, OD 25) at 1.35 kV, 200 Ω, and 25 µF using 0.1 cm electrode gap Gene Pulser® cuvettes (Bio-Rad Laboratories, USA). The pulse was applied at time

constants around 4.5 ms. Immediately after the electroporation, transformed *E. coli* were rescued in 1 ml SOC medium and incubated at 37°C for 1 h. 100 µl of each rescued electroporation cell sample was plated onto selective LB medium supplemented with kanamycin (20 µg·ml^{-1}) and incubated at 37°C overnight. Segments and mega-segments were isolated from *E. coli* using the NucleoBond Xtra Midi kit (Macherey-Nagel, Switzerland) from 50 ml *E. coli* cultures grown overnight in LB medium supplemented with kanamycin. The purity of extracted plasmid DNA was assessed by agarose gel-electrophoresis and quantified on a Nanodrop ND-1000 spectrometer (Thermo Fisher Scientific, Carlsbad, USA).

Size confirmation of segments and mega-segments by electrophoresis. Gel electrophoresis was used to assess DNA assemblies of segments. Each assembled segment was released from pMR10Y by restriction digestion as described above and 100 ng were loaded in a well of a 0.5% UltraPure Agarose (Invitrogen) in 1x TAE buffer gel containing 1x GelRed (Biotium, Fremont, CA, USA). The electrophoresis was conducted at 7.1 V/cm for 40 min, using 1x TAE as a buffer. A pulsed field gel electrophoresis CHEF-DR III variable angle system (BioRad, Cressier, Switzerland) was used to assess DNA assemblies of mega-segments. The running gel consisted of 1% UltraPure Agarose (Invitrogen) in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8.3). The temperature of the running buffer (0.5x TBE) was set to 14°C. PacI / PmeI restriction digested DNA samples, 10 µl DNA 35 ng/µl, plus 2 µl 6x Purple Loading Dye (NEB, USA) were loaded into the wells while the pump was turned off. After 10 min of electrophoresis, the pump was turned on again. The gel was run for 15 h at a voltage of 6 V/cm and a pulsed field switch time of 1 – 8 ms. Agarose gels were stained with 1x GelRed (Biotium, Fremont, CA, USA) in 0.5x TBE buffer for 20 min at RT. The gels for both the segments and the mega-segments were imaged using an AlphaImager (ProteinSimple, San Jose, CA, USA).

Sequence verification of segments and mega-segments by next-generation sequencing. To track the assembly progress, sequence verification was performed at the segment and mega-segment level. Chromosome segments and mega-segments cloned into the maintenance plasmid pMR10Y(4) were isolated using the NucleoBond Xtra Midi kit (Macherey-Nagel, Switzerland) from 50 ml *E. coli* cultures grown overnight in LB medium supplemented with kanamycin. DNA yield was quantified on a Nanodrop ND-1000 spectrometer (Thermo Fisher Scientific, Carlsbad, USA) and the quality of isolated plasmid DNA was assessed by gel-electrophoresis using 0.5 % agarose gel against a super-coiled plasmid standard. DNA samples were subjected to tagmentation (9) and barcoded using the Illumina DNA library preparation protocols and sequenced on a MiSeq instrument using standard paired end sequencing. The resulting reads were demultiplexed, filtered to remove low-quality reads and trimmed to remove adapter sequences followed by read alignment to the reference *C. eth-2.0* genome using bwa. The resulting sam file was sorted and converted into the bam format using samtool (10) prior construction of read pileup (mpileup) and calculation of read coverage. The output bcf file was converted into a vcf file using bcftools (11) for SNP calling and assignment.

Release of mega-segments from pMR10Y:ETH plasmids. All mega-segments except mega-segment 1 and 16 were pooled in equimolar ratios (96 ng/40kb mega-segment, 153 ng/60kb mega-segment, 1.6 µg total) in a 1.5 mL tube and digested using 45 units of PacI and PmeI for 8 hours at 37°C. Similarly, mega-segments 1 and 16 were restriction digested using only PacI or PmeI respectively. To remove the excised pMR10Y backbone from the mega-segment inserts, the digests were loaded onto a 0.5% low gelling temperature agarose (Sigma-Aldrich, USA) in 1X TAE buffer gel and subjected to electrophoresis at 11.4 V/cm for 1 h. After the electrophoresis run, the bands corresponding to the mega-segments were recovered from the gel and incubated in 500 µl 1x β -Agarase I Buffer (NEB, USA) for 30 min on ice. This washing step was repeated once. After equilibrating the gel slice, the agarose plug was melted at 65°C for 10 min, followed by an equilibration step at 42°C for 15 min prior addition of 1.5 µl β -Agarase I (NEB, USA) and further incubated at 42°C for 1 h. The digests were transferred into 1.5 mL screw-cap tubes, extracted with one volume buffer-saturated phenol and centrifuged at 17,000 rcf for 10 min. The upper layer was transferred into a new 1.5 ml tube and an equal volume of 2-propanol supplemented with 1µl/100µl Glycogen Blue (Invitrogen) was added prior centrifugation at 17,000 rcf for 20 min. The DNA pellet was washed with 70% ethanol and the dried pellet was dissolved in 20 µl 0.01M Tris-HCl buffer pH 8.0 and incubated overnight at 4°C.

One-step assembly of C. eth-2.0 from 16 mega-segments in yeast. Whole genome assembly was performed in YJV04, a CRISPR engineered *S. cerevisiae* strain based on YPH857 of Philip Hieter (12). In YJV04 the native auxotrophic marker genes URA3, TRP1, HIS3, MET14, LEU2 and ADE2 have been deleted. The whole genome assembly was conducted according to the following procedure (13). A pre-culture of YJV04 (OD600 = 1.67) was diluted 222-fold in 100 mL YPD medium and incubated at 30°C overnight until an OD600 between 0.3 and 0.5 was reached. Cells were harvested by centrifugation at 1,000 rcf for 5 min, washed in 30 ml sterile water and again washed in 20 ml 1 M Sorbitol followed by dissolving the pellet in 20 ml SPE solution (1 M Sorbitol, 0.01 M sodium phosphate, 0.01 M EDTA pH 8.0) supplemented with 40 µl 2-mercaptoethanol and 20 µl Zymolyase solution (10 mg/ml Zymolyase 20T, 0.05 M Tris-HCl pH 7.5, 25% glycerol). The mixture was incubated at 30°C and progression of spheroplast formation was monitored after 20 min by mixing 100 µl of the digested cells with either 900 µl 1 M Sorbitol or 900 µl 2% SDS respectively. The OD600 ratio in absence or presence of SDS was determined on a spectrometer. The Zymolyase digestion was continued until the sample treated with 2% SDS showed a 3 to 5-fold lower OD600 than the control sample. Spheroplasts were harvested by centrifugation at 300 rcf for 10 min and gently dissolved in 50 ml 1 M sorbitol. This washing step was repeated twice again prior carful resuspension of spheroplasts in 2 ml STC solution (1 M sorbitol, 0.01 M Tris-HCl pH 7.5, 0.01 M calcium chloride). For whole genome assembly reactions, 1.8 µg of digested and purified mega-segments in a volume of 20 µl, as well as 2-3 µg ssDNA (Sigma-Aldrich, USA) were pipetted into a sterile 1.5 ml tube and 200 µl of freshly prepared spheroplasts were added and incubated at room temperature for 10 min. To each sample, 800 µl PEG solution (20% PEG 8000, 0.01 M calcium chloride, 0.01 M Tris-HCl pH 7.5) was added, inverted carefully several

times and incubated at room temperature for 10 min. Spheroplasts were collected by centrifugation at 300 rcf for 10 min, resuspended in 800 µl SOS solution (1 M sorbitol, 0.0065 M calcium chloride, 0.25% yeast extract; Difco, BE, 0.5% peptone; Difco) and incubated at 30°C for 40 min. The spheroplast solution was mixed with 7 ml SDSORB-TOP (SD agar containing 1 M Sorbitol and 2.5% agar pre-tempered at 50°C), inverted several times and poured onto SDSORB (SD agar containing 1 M Sorbitol and 2% agar) plates selecting for Trp and Leu. After hardening, the plates were incubated at 30°C for 5-7 days before seeing colonies.

Diagnostic PCR to verify C. eth-2.0 assembly. To assess the correct assembly of the *C. eth-2.0* genome, a PCR was conducted using the yeast cells as template. After growing on a plate, the colony of interest was picked and dissolved in 3 µl 0.02 M sodium hydroxide in a PCR tube. The dissolved colonies were incubated in a thermocycler at 99°C for 10 min. Each PCR reaction contained 5 µl 5 M betaine, 12.5 µl BioRed 2x Mastermix (Bioline, London, UK), 0.5 µl diagnostic primers (100 nM, annealing to each of the 15 assembly junctions to be tested) and 3.5 µl water. The PCR reaction mix was added to the lysed yeast cells and put into the thermocycler to run the following PCR protocol. (1) initial denaturation 5 min at 96°C, (2) denaturation 30 s at 96°C, (3) primer annealing 30 s at 60°C, (4) elongation 1 min at 72°C, (5) repeat steps 2 – 4, 30 times, (6) final elongation 10 min at 72°C. PCR products were analysed by agarose gel-electrophoresis to confirm the correct amplicon size.

Size confirmation of the C. eth-2.0 genome construct by pulsed-field electrophoresis. A pulsed-field gel electrophoresis CHEF-DR III variable angle system (BioRad, Cressier, Switzerland) was used to confirm complete DNA assembly of the *C. eth-2.0* genome. Prior to the electrophoresis step, intact chromosomal DNA was extracted from the *S. cerevisiae* cell. A YJV04 *C. eth-2.0* pre-culture was grown in SD medium selecting for the auxotrophic markers. The pre-culture was diluted 1:15 in SD medium and grown overnight to mid to late log-phase. Cells corresponding to 1.5 ml of the culture were harvested and washed using 1 ml 50 mM EDTA, pH 8.0 and resuspended in 100 µl 50 mM EDTA, pH 8.0. To the washed cells, 50 µl Zymolyase solution (1 mL SCE solution (1 M sorbitol, 0.1 M sodium citrate, 60 mM EDTA, pH 7.0), 9 mg Zymolyase 20T, 50 µL 2-mercaptoethanol) was added prior addition of 250 µl of a 1% low melting point agarose (LMP) solution in 0.125 M EDTA, pH 8.0 equilibrated at 50°C. A total of 80 µl of the yeast cell containing agarose solution was pipetted into CHEF Mapper XA System 50-Well Plug Molds (BioRad, Cressier, Switzerland) and incubated on ice, until the agarose was solidified. The plugs were collected in 2 ml tubes containing 1.5 ml ETB solution (9 ml 0.5 M EDTA pH 8.0, 1 ml 1 M Tris-HCl pH 8.0, 0.5 ml 2-mercaptoethanol) and were incubated at 37°C for 4 hours. The plugs were washed twice with 1.5 ml 50 mM EDTA pH 8.0 prior to adding 2 mL ProteinaseK solution (9 mL 0.5 M EDTA pH 8.0, 1 ml 10% N-Lauroylsarcosine, 10 mg ProteinaseK, 1 mg RNase) followed by an overnight incubation at 37°C. Following the proteinase K digest, the plugs were washed three times with 1.5 ml 50 mM EDTA, pH 8.0. Following an 1 h incubation of the plugs in TE buffer at 37°C, the plugs were incubated in 500 µl CutSmart buffer for 30 min on ice. After removing the buffer, 160 µl restriction enzyme mix (40 µL 5x CutSmart buffer, 111 µl MQ, 3 µl (30 U) PmeI / PacI) was added and incubated at 37°C for 5 h prior to exchanging the restriction enzyme mix to further digest the plugs overnight at 37°C. Prior to loading the plugs into the 1% UltraPure Agarose (Invitrogen) in 0.5x TBE buffer gel, the plugs were incubated in TE buffer for 1 h at room temperature. For the PFGE run, 0.5x TBE buffer at 14°C was used. The switch time was ramped over the course of 24 hours from 60 s to 120 s. The set voltage was 6 V/cm at an angle of 120°. Agarose gels were stained with 1x GelRed (Biotium, Fremont, CA, USA) in 0.5x TBE buffer for 20 min at RT and imaged using an AlphaImager (ProteinSimple, San Jose, CA, USA).

Sequence confirmation of C. eth-2.0 genome construct. To verify the whole genome assembly of *C. eth-2.0* in *S. cerevisiae* and assess chromosome maintenance and stability upon serial propagation of the host strain, total genomic DNA was extracted according to the following procedure. The initially isolated *S. cerevisiae* strain bearing the *C. eth-2.0* genome according to PFGE and diagnostic PCR assessment as well as isolates propagated for twenty, forty and sixty generations were grown overnight to full density. Cells corresponding to 1.5 ml of the overnight culture were harvested at 13,000 rcf for 1 min. The supernatant was discarded and cells were resuspended in 200 µl lysis buffer (1% SDS, 0.1 M sodium chloride, 2% Triton X-100, 0.01 M Tris-HCl pH 8.0, 0.001 M EDTA pH 8.0). To each sample 200 µl acid washed glass beads (400-600 µm Sigma-Aldrich) and 200 µl phenol/chloroform/isoamylalcohol was added. The mix was placed on a FastPrep24 5G (MP Biomedicals, Santa Ana, CA, USA) and homogenized twice at full speed for 15 s with incubation on ice for 2 min in between. The cell debris was pelleted by centrifugation at 13,000 rcf for 5 min and the top layer was transferred into a new 1.5 ml tube supplemented with 100 µl chloroform and vortexed at full speed for 30 s and phase-separated by centrifugation at 13,000 rcf for 2 min. The top layer (150 µl) was transferred into a new 1.5 ml tube and 1 µl RNase solution (0.01 M sodium acetate, 10 mg/mL RNase A (Roche, BS, CH), 0.1 M Tris-HCl pH 7.5) was added and incubated at room temperature for 1 hour. Genomic DNA was precipitated by adding 2.5 times the sample volume precooled (-20°C) 100% ethanol and incubated for 30 min at -20°C. The precipitated DNA was spun collected by centrifugation at 14,000 rcf for 5 min at 4°C and the supernatant was discarded followed by a washing step using 70% ethanol at room temperature. The pellet was air dried and resuspended in 50 µl Tris-HCl pH 8.0 The genomic DNA extract was stored at 4°C. DNA concentrations of each sample was quantified on a Qbit spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). To obtain DNA libraries for next-generation sequencing, DNA samples were processed according to the NEBNext ultra II DNA Library Prep Kit for Illumina (NEB, USA). The sequencing was conducted on an Illumina NextSeq500 sequencer using a NextSeq 500/550 Mid Output v2 kit (FC-404-2001, Illumina, San Diego, CA, USA) at the Functional Genomics Center Zürich. The resulting sequencing output data was demultiplexed and reads were quality filtered, subjected to adapter trimming and mapped to the *C. eth-2.0* genome using bwa. Similar to the sequence analysis for sequence verification of chromosome segments and mega-segments, the resulting sam files were processed using samtools and bcftool to generate coverage data and determined SNP frequencies across the *C. eth-2.0* genome design.

D. Construction of mero-synthetic *Caulobacter* test strains.

Conjugational transfer rates of C. eth-2.0 chromosome segments into Caulobacter. Sequence verified pMR10Y plasmids harbouring a different *C. eth-2.0* chromosome segment were conjugated from *E. coli* S17-1 donor strains into wildtype *Caulobacter* NA1000 to generate a panel of 37 mero-synthetic test-strains. To identify *C. eth-2.0* chromosome segments that elicit toxicity upon boot-up in *Caulobacter*, conjugational transfer frequencies were quantified by counting the occurrence of kanamycin resistant colonies per 108 recipient cells using the conjugation assay and compared against a pMR10Y plasmid control lacking any chromosome segment. *C. eth-2.0* chromosome segments that elicit toxicity in *Caulobacter* were identified using a cut-off of showing more than a 1000-fold reduction in kanamycin resistant trans-conjugants as compared to the empty pMR10Y plasmid control. A total of 12 out of 37 *C. eth-2.0* chromosome segments harboured one or more toxic *C. eth-2.0* genes. Assuming random distribution, the occurrence of toxic *C. eth-2.0* genes was estimated with a Poisson rate constant based on the fraction of genome segments that did not elicit toxicity according to:

$$\lambda = -\ln \frac{N}{S} \quad [1]$$

with $N = 25$ non-toxic segments out of a total of $S = 37$ genome segments assayed yielding a rate constant λ of 0.392. Based on a Poisson distribution with rate constant of 0.392, 25 non-toxic segments, 10 segments with a single toxic gene and 2 segments harbouring two toxic genes are expected. The cumulative size of toxic DNA sequences within individual segments and across the entire *C. eth-2.0* genome design was estimated based on the observed conjugational transfer frequencies for each toxic chromosome segment and an approximated mutation rate of 10^{-8} per base and assuming a phenotypic lag of 2-3 cell divisions prior ceasing of cell division upon conjugational transfer of a toxic chromosome segment.

In vivo adaptation of C. eth-2.0 chromosome segments at the population level. To assay the adaptive dynamics upon *in vivo* boot-up of individual *C. eth-2.0* chromosome segments in *Caulobacter*, populations of 103 up to 105 transconjugants were pooled and pMR10Y plasmids were isolated using the NucleoBond Xtra Midi kit (Macherey-Nagel, Switzerland). Plasmid DNA was quantified by agarose gel-electrophoresis. Tagmentation and barcoding of individual DNA samples was performed using Illumina Nextera XT Index Kit v2 barcodes prior paired-end sequencing on an Illumina MiSeq instrument according to standard sequencing chemistry and protocols. Raw reads were demultiplexed using the indexing read information according to the Illumina workflow. Reads were filtered for high-quality reads and adapter sequences were trimmed prior alignment to the *C. eth-2.0* reference genome sequence using bwa. Read coverage for each *C. eth-2.0* chromosome segment was normalized against total reads mapping to the pMR10Y backbone sequence and compiled into a single data set for plotting the sequence coverage across the entire genome design.

Isolation of evolved C. eth-2.0 chromosome segments and analysis of self-repair. To identify toxic genes and isolate mero-synthetic *Caulobacter* strains evolved for stable maintenance of *C. eth-2.0* chromosome segments, transconjugants were isolated and PCR screened for presence of intact chromosome segments. For PCR screening, sets of multiplex-PCR primers were used that annealed to each block junction present within chromosome segments. Strains that maintained the full-length or the largest portion of a formerly toxic chromosome segments were isolated and corresponding plasmids were isolated using the NucleoBond Xtra Midi kit (Macherey-Nagel, Switzerland). To detect suppressor mutations responsible for tolerance of toxic chromosome segments, individual plasmids bearing intact chromosome segments were sequenced on a MiSeq Illumina instrument using DNA library preparation and indexing according to the Illumina workflow. Read coverage data for each *C. eth-2.0* chromosome segment was calculated and normalized against the pMR10Y backbone. Single base substitutions, indels and larger deletions occurring upon *in vivo* evolution of chromosome segments were mapped using SNP calling by samtools and bcftool. Furthermore, to detect insertion events into *C. eth-2.0* genes by IS288 and IS544 sequences the output sequencing read dataset was searched for reads mapping to the end sequences of native *Caulobacter* insertion sequences (IS511, IS298, ISCc1, ISCc2, ISCc3, ISCc4 and ISCc5) and reading into the *C. eth-2.0* genome design. For *C. eth-2.0* chromosome segments 1, no plasmid born suppressor mutations were identified. To identify chromosomal suppressor mutations that permit tolerance of the formerly toxic *C. eth-2.0* chromosome segment 1, sequencing reads from chromosomal DNA present within the plasmid prepped DNA were aligned onto the native *Caulobacter* reference sequence and SNPs occurring in native copies of mero-diploid genes were identified using samtool and bcftool. To fine-map precise locations of deletions occurring within *C. eth-2.0* chromosome segments upon *in vivo* adaptation of mero-synthetic *Caulobacter* test strains, the sequence coverage data was analysed to detect regions devoid of any sequencing reads. A total of 7 deletions spanning 367bp up to 12,659 bp in size were detected. The precise genome coordinates of deletions were assigned by searching for reads covering both end-points of each deletion.

E. Fault diagnosis of *C. eth-2.0* by transposon sequencing.

Transposon mutagenesis, PCR amplification and sequencing of Tn5 junctions. To generate hyper-saturated transposon libraries, a conjugation based mutagenesis procedure was used as previously described (1). In brief, a Tn*5* transposon derivatives bearing transposon internal barcode sequences were conjugated using the filter mating procedure from an *E. coli* S17-1 donor strain into the mero-diploid *Caulobacter* crescentus test strain bearing the different 20kb long rewritten DNA segments on pMR10Y plasmids. Transposon internal barcodes consisting of a 14 base-pair random sequence adjacent to the transposon end sequences were used to tag independent samples and provide multiplexing capabilities. After overnight incubation on PYE plates supplemented with xylose, cells from each mating filter were harvested, resuspended in 800 µl PYE and 100 µl aliquots were

plated onto PYE plates containing gentamycin, nalidixic acid, kanamycin and xylose. Plates were incubated for three days at 30°C and transposon mutant libraries from each plate were pooled and stored at -80°C as individual pools of 10,000 insertion mutants in 96 deep well plates (Eppendorf AG, Germany). Reference transposon mutant pools were constructed in a similar manner using test-strains bearing only the empty pMR10Y plasmid to subsequently generate a control TnSeq data set.

Parallel amplification of transposon junctions by semi-arbitrary PCR. Amplification of transposon junctions carrying terminal adapters compatible with Illumina sequencing was performed as previously reported (1). In brief, a two-step arbitrary PCR strategy was carried out in 384 well plate format to amplify transposon junctions in parallel from each mutant pool using a reaction volume of 10 µl using 1 µl of a freeze-thawed transposon mutant pool (OD_{600} 1.0) as template and Taq polymerase mix (GoTaq, Promega, Dübendorf, Switzerland) in a thermocycler instrument (C1000 touch, BioRad, Cressier, Switzerland). Primers used for first round amplification consisted of a transposon specific primer (M13 universal) and one of four arbitrary PCR primers as previously described(1). In a second round of PCR, 1.5 µl of the first round PCR products were further amplified using Illumina paired-end primers PE1.0 and PE2.0. The first PCR amplification was performed according to following PCR program: (1) 94°C for 3min, (2) 94°C for 30s, (3) 42°C for 30s, slope -1°C/cycle, (4) 72°C for 1min, (5) repeat step 2-4, 6 times, (6) 94°C for 30s, (7) 58°C for 30s, (8) 72°C for 1min, (9) repeat step 6-8, 25 times, (10) 72°C for 3min, (11) 12°C hold. Product of first round of PCR was further amplified in a second nested PCR reaction using following thermocycling conditions: (1) 94°C for 3min, (2) 94°C for 30s, (3) 64°C for 30s, (4) 72°C for 1min, (5) repeat step 2-4, 30 times, (6) 72°C for 3min, (7) 12°C hold.

DNA library purification and next-generation sequencing of transposon junctions. Amplicons in the 200-700 bp range corresponding to transposon junctions were size selected by gel electrophoresis (2.0% agarose) and purified over a silica column (Machery-Nagel, Switzerland). The DNA concentration of each sample was quantified on a Nanodrop spectrometer Nanodrop ND-1000 spectrometer (Thermo Fisher Scientific, Carlsbad, USA) and samples from different TnSeq libraries were combined. Transposon junction libraries were paired end sequenced (2x150bp) on HiSeq and MiSeq Illumina platforms using standard cluster generation protocols, sequencing chemistry and primers PE 1.0 and 2.0 with a 5% phiX spike-in for calibration of base-calling. Sequencing was performed at the Functional Genomics Center Zürich, Zürich, Switzerland.

Transposon sequencing read processing and mapping of transposon insertion sites. Raw sequencing data processing, read alignment and sample demultiplexing was performed using a previously described analysis pipeline based on Python, Biopython, bwa, and Matlab routines (14). Read filtering criteria included i) removal of low quality reads ii) requiring at least a 15 bp long perfect match to the Tn*5* transposon end sequence [GTGTATAAGAGACAG] and iii) requiring a genomic insert size of at least 15 bp as inferred by testing for overlapping read sequences between mate reads. Sequences corresponding to the Illumina adapters, transposon end sequence and arbitrary PCR primers sequences were trimmed and processed reads were aligned to the *Caulobacter* NA1000 and *C. eth-2.0* reference genomes using bwa (15). Transposon insertion sites were assigned requiring read pairs with correct orientation, perfect matches within the first 15 bases of both reads to a genomic location, unambiguous mapping to either the native genome or the *C. eth-2.0* genome sequence and a genomic insert size smaller than 500 bases. Nucleotide substitutions introduced upon rewriting and sequence optimization of the *C. eth-2.0* genome allowed to unambiguously assign transposon insertions to the native *Caulobacter crescentus* NA1000 reference genome NC_011916 and chromosome segments of *C. eth-2.0*. Demultiplexing into the different TnSeq dataset was performed according to the 14 bp long internal transposon barcode sequences adjacent to the transposon end. Upon insertion into a target locus, Tn*5* transposition generates a nine-base pair long duplication, which was taken into account for subsequent insertion site analysis. Transposon insertion sites were defined according to the first reference base detected immediately after reading outwards of the transposon I-end, which contains an outward pointing Pxyl promoter (1). Global insertion sites statistics, insertion occurrence and distribution across all annotated features of the reference genomes were calculated with Matlab according to previously described metrics and routines (1, 14).

Functionality analysis of C. eth-2.0 genes. Transposon insertion numbers across protein and RNA-coding genes were normalized to account for differences in the global number of transposon insertion sites mapped between the different TnSeq measurements. Reference transposon insertion frequencies for native gene copies residing on the *Caulobacter* chromosome were determined for each gene present on the *C. eth-2.0* genome design using following routine. Essential and semi-essential genes tolerate a low frequency of transposon insertions at sites that are non-disruptive and still permit expression of a functional gene product (5' and 3' portion of essential genes). Transposon hit rates for native genes from a panel of TnSeq measurements across different mero-diploid test strains were linearly fitted against the total number of unique insertions recovered from each individual data set. Based on the linear fit-function, the mean baseline insertion frequency and standard deviation under non-complementing conditions was determined for each *Caulobacter* gene. For the linear fitting procedure, only transposon insertions with unique mapping positions were considered. A Z-score metric was used to quantify the level of functionality of each rewritten *C. eth-2.0* gene (functionality Z-score) and diagnose faulty genes in the genome design according to the following formula:

$$z = \frac{Tn_{compl} - Tn_{baseline}}{\sigma} \quad [2]$$

with Tn_{compl} specifying the normalized number of transposon insertions within a native *Caulobacter* gene under complementing conditions (in presence of a *C. eth-2.0* chromosome segment that carries rewritten gene copies), $Tn_{baseline}$ specifying the normalized number of transposon insertions within the native *Caulobacter* gene according to base-line experiments in absence of complementation and σ specifying the experimentally determined standard deviation of transposon hits for a given *Caulobacter*

gene under non-complementing conditions. The functionality of rewritten *C. eth-2.0* genes was assessed for each gene present on the evolved *C. eth-2.0* genome design with genes that were deleted or partially deleted excluded from the analysis. A total of 42,218 kb of the genome design was deleted upon *in vivo* adaptation (identified deletions within segments, 6, 12, 14, 21, 22, 25, 38). In addition, we discovered that segment 4 containing test-strains were no longer susceptible for conjugation-based transposon mutagenesis. To carry out transposon mutagenesis and assess functionality, a test-strain bearing a partially deleted segment 4 (9.4 kb deletion) was used. Deleted regions of the *C. eth-2.0* genome were enlarged to exclude partially deleted genes and operons from subsequent functionality assessment. According to these criteria, 89.2% (697,348 bp out of 781,541bp excluding the pMR10Y backbone) of the *C. eth-2.0* genome design was subjected to the TnSeq-based functionality analysis. Genes were classified into functional and faulty *C. eth-2.0* genes according to following criteria. Non-essential and redundant *C. eth-2.0* genes were excluded from the functionality analysis. Essential and semi-essential *C. eth-2.0* genes with functionality Z-score equal or larger than 1.65 corresponding to a p value smaller than 0.05 were classified as functional *C. eth-2.0* genes whereas genes with a functionality Z-score smaller than 1.65 were deemed non-functional.

Functionality analysis of C. eth-2.0 genes across gene categories. *C. eth-2.0* genes were grouped according to the clusters of orthologous groups (COG) classification into Metabolism (COG classifier, G, E, F, H, I, P), Cellular processes (COG classifier, D, M, N, O), Transcription (COG classifier, K), DNA replication (COG classifier, L), Energy production (COG classifier, C), Translation (COG classifier, J including tRNAs) and Hypothetical proteins (COG classifier, S). P values for enrichment and de-enrichment in functionality of *C. eth-2.0* genes were calculated based on a cumulative binomial distribution for the above gene categories assuming uniform gene functionality across a given gene set.

Fault diagnosis and classification of error cause. Fault diagnosis of individual faulty *C. eth-2.0* genes was performed according to following classification scheme. Fault causes were grouped into four fault categories, i) mis-annotation, ii) sequence rewriting, iii) synthesis optimization and iv) mutation. Faulty genes that are shorter than annotated, non-existent or lack a functional promoter according to transcriptional start site measurements (6) were assigned to the fault category 'mis-annotation'. Faulty genes with internal promoters embedded within upstream genes that were erased upon rewriting and faulty genes located downstream of functional genes within operons were assigned to the fault category 'sequence rewriting'. Faulty genes with synthesis constraints such as type IIS endonuclease sites, hairpins and direct repeats removed within their promoter regions were assigned to the fault cause 'synthesis optimization'. Faulty genes that harboured sequence changes occurring upon synthesis, assembly and transfer to *Caulobacter* were assigned to the fault category 'mutation'.

Repair of faulty cell division genes and β-galactosidase reporter assay. DNA sequences covering the promoter region of *ftsZ* (CCNA_02623, CETH_02623), *ftsQ* (CCNA_02625, CETH_02625), *murC* (CCNA_02629, CETH_02626) and *murG* (CCNA_02634, CETH_02634) from the native chromosome or the *C. eth-2.0* genome including the first four amino acids of the corresponding protein coding regions were PCR amplified and cloned via BglII and SpeI sites into the *lacZ* reporter plasmid pR9TT using isothermal assembly (16). Reporter plasmids were conjugated into *Caulobacter* NA1000 cells and β-galactosidase-activity of the resulting translational lacZ-reporter constructs were assayed in cells using standard ONPG based assays. The β-galactosidase activities reported represent the average of at least three independent measurements derived from mid-log phase cultures.

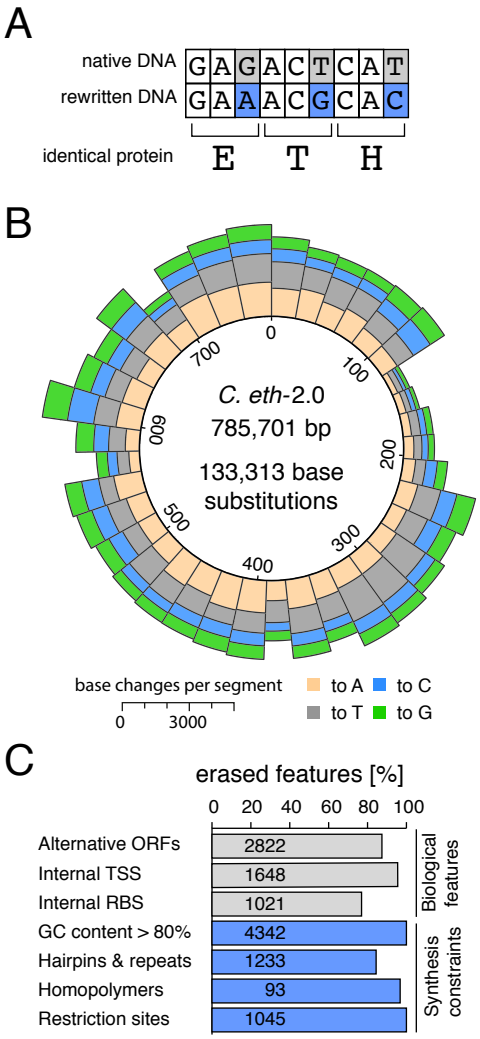


Fig. S1. Rewriting of the bacterial designer genome *C. eth-2.0*. (*A*) Synonymous rewriting maintains identical protein sequences upon genome rewriting. (*B*) Genome map of *C. eth-2.0* illustrating the number of base substitutions introduced per genome segment (radial bars), with base changes to A (beige), T (grey), C (blue) and G (green) colour coded. (*C*) Bar chart summarizing the numbers and categories of genome features and DNA synthesis constraints that have been erased upon genome rewriting.

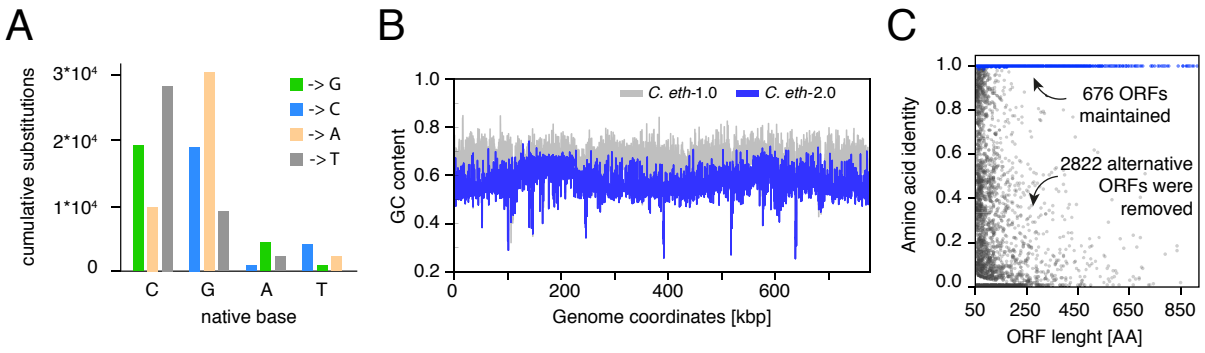
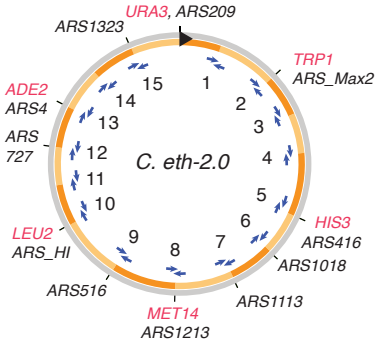


Fig. S2. Massive sequence rewriting of the *C. eth-1.0* genome. (*A*) Bar graph summarizing cumulative base substitutions between the *C. eth-1.0* genome (design based on native sequences) and the rewritten (synthesis streamlined) *C. eth-2.0* genome design. (*B*) GC content plot between the difficult to synthesize (*C. eth-1.0* genome, grey) and the synthesis optimization genome design (*C. eth-2.0* genome, blue). Regions with low GC content pinpoint to yeast ARS sequences introduced to facilitate chromosome maintenance and replication in *S. cerevisiae* strain YJV04. (*C*) Dot plot of the amino acid sequence identity versus ORF length between *C. eth-1.0* and *C. eth-2.0.* A total of 676 ORFs (blue) were maintained in *C. eth-2.0* while 2,822 putative alternative ORFs (grey) were erased upon rewriting.

A

Amplicon	Size [bp]	Amplicon	Size [bp]
1	320	9	861
2	786	10	169
3	400	11	407
4	600	12	897
5	935	13	207
6	601	14	454
7	844	15	975
8	233		

B



C

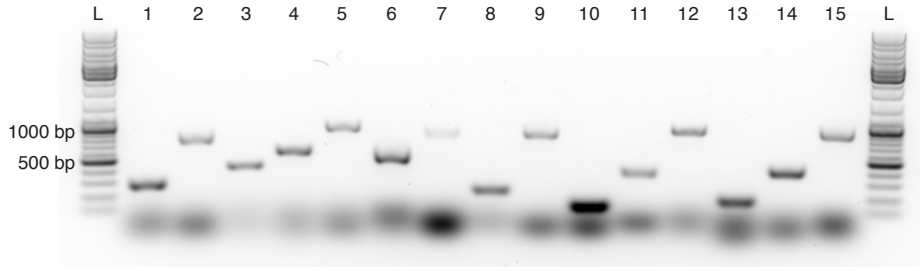


Fig. S3. Confirmation of complete *C. eth-2.0* assembly in yeast. (*A*) Diagnostic PCR was used to verify chromosome assembly. (*B*) Fifteen amplicons covering assembly junctions between mega-segments (light and dark orange boxes) were designed. Location of annealing sites for diagnostic primers (blue arrows) are shown. The pMR10Y vector sequence is represented by a black triangle. (*C*) Size distribution of the PCR amplicons derived for the isolated YJV04 clone 2 bearing the complete *C. eth-2.0* chromosome are shown against a size standard (L).

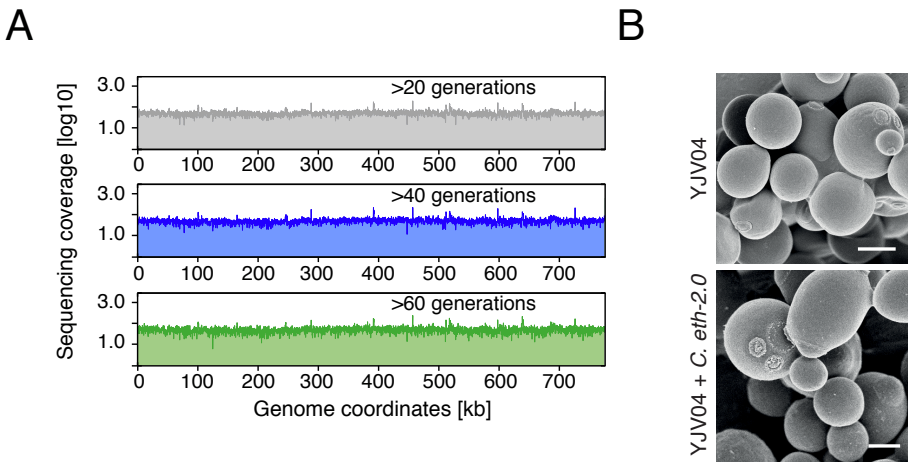


Fig. S4. Cell morphology phenotype and genome stability of yeast cloned *C. eth-2.0*. (*A*) Sequencing coverage retrieved of upon serial outgrowth indicates stable chromosome maintenance of *C. eth-2.0*. Sequencing coverage is shown for *C. eth-2.0* after serial passage for 20, 40 and 60 generations in yeast. (*B*) Phenotypic analysis of YJV04 (upper panel) and YJV04 bearing *C. eth-2.0* (lower panel) by scanning electron microscopy (SEM). Normal yeast cell morphologies and no phenotypic influence of the *C. eth-2.0* chromosome on cellular replication of *S. cerevisiae* is detected.

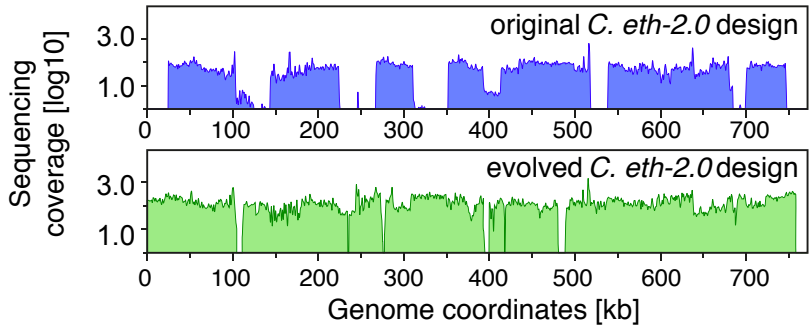


Fig. S5. Stability of *C. eth-2.0* chromosome segments upon conjugation into *Caulobacter*. Sequence coverage plots are shown across the *C. eth-2.0* chromosome upon conjugation of individual segments into *Caulobacter*. Coverage data obtained from sequencing pools of trans-conjugants of the original *C. eth-2.0* genome design (upper panel) and from single isolates (lower panel) evolved for stable maintenance of full length chromosome segments is shown.

Table S1. Codon frequency table of the rewritten *C. eth-2.0* genome

1st base	2nd base: U			2nd base: C			2nd base: A			2nd base: G			3rd base
U	**F**	3,231	(6,924)	**S**	1,925	(1,690)	**Y**	1,787	(2,481)	**C**	552	(1,343)	C
		4,417	(724)		1,192	(285)		2,499	(1,805)		956	(166)	U
	L	0	(42)[a]		208	(207)	*	236	(191)	*	321	(201)	A
		2	(1,105)[a]		3,644	(4,940)	*	8	(173)[a]	**W**	2,526	(2,526)	G
C	**L**	6,042	(3,438)	**P**	3,194	(3,708)	**H**	1,388	(2,920)	**R**	4,205	(9,722)	C
		5,207	(1,040)		3,002	(601)		2,357	(824)		5,265	(2,028)	U
		1,290	(202)		1,709	(289)	**Q**	4,245	(882)		2,287	(428)	A
		8,482	(15,196)		3,662	(6,969)		2,535	(5,898)		2,930	(2,506)	G
A	**I**	4,579	(9,822)	**T**	4,253	(7,047)	**N**	4,119	(2,166)	**S**	3,124	(2,971)	C
		5,863	(621)		1,111	(242)		919	(2,871)		164	(164)[b]	U
		107	(107)[b]		1,095	(168)	**K**	518	(5,275)	**R**	139	(109)[b]	A
	M	4,561	(4,560)		4,275	(3,276)		8,857	(4,100)		192	(222)[b]	G
G	**V**	6,406	(9,013)	**A**	9,173	(17,959)	**D**	4,610	(10,183)	**G**	5,795	(14,411)	C
		4,126	(1,173)		8,215	(1,845)		8,204	(2,630)		6,793	(2,170)	U
		195	(194)[b]		3,060	(480)	**E**	8,885	(4,360)		3,334	(625)	A
		6,309	(6,656)		8,798	(8,963)		4,655	(9,180)		3,076	(1,795)	G

[a] Codons erased upon rewriting.

[b] Rare codons maintained upon rewriting except for removal of forbidden type IIS endonuclease sites.

Table S2. Codon substitutions per *C. eth-2.0* chromosome segments

Segment ID	Coordinates	Size [bp]	Codon substitutions[a]	Rewriting %
seg_1	1..22276	22,276	3,861 (6,387)	60.45
seg_2	22157..41386	19,230	3,451 (5,351)	64.49
seg_3	41267..60572	19,306	3,475 (5,570)	62.39
seg_4	60451..80089	19,639	3,484 (5,747)	60.62
seg_5	79966..101065	21,100	3,550 (5,946)	59.70
seg_6	100946..122064	21,119	3,560 (5,883)	60.51
seg_7	121944..142404	20,461	927 (5,585)	16.60
seg_8	142293..161370	19,078	1,039 (5,121)	20.29
seg_9	161367..182611	21,245	1,441 (5,731)	25.14
seg_10	182491..202737	20,247	1,472 (5,086)	28.94
seg_11	202617..223202	20,586	2,176 (6,265)	34.73
seg_12	223083..245967	22,885	3,766 (6,364)	59.18
seg_13	245848..266880	21,033	3,449 (5,892)	58.54
seg_14	266762..288129	21,368	3,816 (6,211)	61.44
seg_15	288009..309977	21,969	4,127 (6,830)	60.42
seg_16	309856..332605	22,750	4,205 (6,935)	60.63
seg_17	332486..351749	19,264	3,449 (5,706)	60.45
seg_18	351627..374062	22,436	3,897 (6,621)	58.86
seg_19	373942..391434	17,493	3,005 (5,025)	59.80
seg_21	391316..413535	22,220	3,933 (6,873)	57.22
seg_22	413414..434554	21,141	3,731 (6,650)	56.11
seg_23	434431..456204	21,774	3,657 (6,244)	58.57
seg_24	456085..476452	20,368	3,597 (5,813)	61.88
seg_25	476332..496786	20,455	3,494 (5,693)	61.37
seg_26	496667..518097	21,431	3,344 (5,465)	61.19
seg_27	517978..539588	21,611	3,536 (5,898)	59.95
seg_28	539466..559225	19,760	3,842 (6,188)	62.09
seg_29	559105..578007	18,903	2,066 (4,755)	43.45
seg_30	577887..597167	19,281	3,072 (5,322)	57.72
seg_31	597046..617173	20,128	4,754 (5,535)	85.89
seg_32	617051..638739	21,689	3,672 (6,079)	60.40
seg_33	638620..659188	20,569	3,143 (5,268)	59.66
seg_34	659068..681646	22,579	4,122 (6,711)	61.42
seg_35	681526..702398	20,873	2,775 (4,475)	62.01
seg_36	702278..725152	22,875	4,001 (6,693)	59.78
seg_37	725032..748644	23,613	4,310 (7,103)	60.68
seg_38	748524..773851	25,328	4,573 (7,723)	59.21
Total codons substituted in *C. eth-2.0*			123,562 (220,350)	56.08

[a] Total number of codons are indicated in brackets

Table S3. Non-synonymous mutations introduced upon the build process

Segment	Mutation[a]	Stage[b]	Gene annotation, SNP effect[c]
seg_7	128065 G to C	assembly	CETH_00531, LSU ribosomal protein L12P, K127N
seg_7	135815 G to A	synthesis	CETH_00537, RNAP beta' chain, E1030K
seg_8	149437 G to T	synthesis	CETH_00735, beta-barrel assembly protein BamF, terminator
seg_8	150347 T to C	synthesis	CETH_00736, lipoprotein signal peptidase, I109V
seg_8	152937 T to C	synthesis	CETH_00737, isoleucyl-tRNA synthetase, D264G
seg_8	156211 C to T	synthesis	CETH_00749, ferrous iron transport protein B, T258M
seg_8	161242 G to A	synthesis	CETH_R0015, tRNA-Val, 63g>a
seg_9	175114 C to T	synthesis	CETH_00892, phosphoenolpyruvate-protein, S276P
seg_10	185224 C to T	synthesis	CETH_00944; flagellar hook length protein, P337S
seg_10	190540 G to A	synthesis	CETH_01001, ribokinase, R22C
seg_10	193501 C to T	synthesis	CETH_01060, type I secretion protein RsaD, A486P
seg_10	201452 C to T	synthesis	CETH_01104, glycosyltransferase, A100T
seg_11	208036 T to C	synthesis	CETH_01210, nucleotidyltransferase protein, P7S
seg_11	209021 G to A	synthesis	CETH_01211, MobA-like NTP transferase, V265P
seg_11	212793 A to G	synthesis	CETH_01214, YjgP/YjgQ membrane permease, L120P
seg_24	470251 6 bp Δ	synthesis	CETH_02177, hypothetical protein, terminator
seg_29	577864 C to T	synthesis	CETH_02934, conserved hypothetical protein, terminator
seg_30	589432 T to C	synthesis	CETH_03026, two-component regulator PetR, V182T
seg_31	616412 C to T	assembly	CETH_03305, SSU ribosomal protein S7P, D15N
seg_33	649119 G to T	assembly	CETH_03469, arginyl-tRNA synthetase, P597H
seg_38	752493 C to T	assembly	CETH_03833, acyl-carrier protein FabI, M215I

[a] Position of sequence alterations are shown together with nucleotide change.
[b] Stage of the built process where nucleotide change occurred.
[c] Gene ID are provided together with the functional annotation and amino-acid changes.

Table S4. Conjugational transfer frequency of *C. eth-2.0* chromosome segments

Segment ID	Genomecoordinates	Size [bp]	Frequency [log10][a]	Size of toxic DNA part [bp][b]
seg_1	1..22276	22,276	-4.72 ± 0.05	948 ± 105
seg_2	22157..41386	19,230	-1.05 ± 0.05	NA
seg_3	41267..60572	19,306	-2.06 ± 0.16	NA
seg_4	60451..80089	19,639	-1.68 ± 0.1	NA
seg_5	79966..101065	21,100	-2.33 ± 0.08	NA
seg_6	100946..122064	21,119	-4.63 ± 0.11	1,173 ± 262
seg_7	121944..142404	20,461	-3.95 ± 0.09	5,576 ± 1050
seg_8	142293..161370	19,078	-0.83 ± 0.07	NA
seg_9	161367..182611	21,245	-0.95 ± 0.01	NA
seg_10	182491..202737	20,247	-1.12 ± 0.08	NA
seg_11	202617..223202	20,586	-0.96 ± 0.08	NA
seg_12	223083..245967	22,885	-4.95 ± 0.03	564 ± 33
seg_13	245848..266880	21,033	-4.72 ± 0.07	965 ± 152
seg_14	266762..288129	21,368	-1.96 ± 0.06	NA
seg_15	288009..309977	21,969	-1.83 ± 0.03	NA
seg_16	309856..332605	22,750	-4.56 ± 0.07	1,394 ± 195
seg_17	332486..351749	19,264	-4.67 ± 0.02	1,061 ± 53
seg_18	351627..374062	22,436	-2.03 ± 0.14	NA
seg_19	373942..391434	17,493	-4.40 ± 0.32	2,006 ± 1043
seg_21	391316..413535	22,220	-4.64 ± 0.02	1,138 ± 60
seg_22	413414..434554	21,141	-1.55 ± 0.10	NA
seg_23	434431..456204	21,774	-1.41 ± 0.08	NA
seg_24	456085..476452	20,368	-1.35 ± 0.10	NA
seg_25	476332..496786	20,455	-1.90 ± 0.07	NA
seg_26	496667..518097	21,431	-1.75 ± 0.05	NA
seg_27	517978..539588	21,611	-4.57 ± 0.01	1,341 ± 37
seg_28	539466..559225	19,760	-1.39 ± 0.12	NA
seg_29	559105..578007	18,903	-0.77 ± 0.02	NA
seg_30	577887..597167	19,281	-0.72 ± 0.06	NA
seg_31	597046..617173	20,128	-2.00 ± 0.10	NA
seg_32	617051..638739	21,689	-1.17 ± 0.16	NA
seg_33	638620..659188	20,569	-1.39 ± 0.07	NA
seg_34	659068..681646	22,579	-1.28 ± 0.12	NA
seg_35	681526..702398	20,873	-4.41 ± 0.16	1,968 ± 602
seg_36	702278..725152	22,875	-1.43 ± 0.05	NA
seg_37	725032..748644	23,613	-1.93 ± 0.02	NA
seg_38	748524..773851	25,328	-4.82 ± 0.03	767 ± 57
pMR10Y (vector control)			-0.70 ± 0.07	NA
		Cumulative size of toxic DNA parts		18,903 ± 3,650
		Estimated number of toxic DNA parts[c]		14

[a] Log10 transformed conjugational transfer frequency of a given segment relative to the transfer rate of the empty pMR10Y plasmid.
[b] Estimated size of toxic DNA parts based on an effective mutation rate of 10^{-7}.
[c] Estimated total number of toxic DNA parts.

Table S5. List of identified toxic *C. eth-2.0* gene

Segment	Gene_ID	Annotation	Mutation type
seg_1	CETH_00005	DnaQ DNA pol III ϵ subunit	Chrom: Val99 to Val (GTC to GTA)
seg_6	CETH_00390	WaaF, LPS heptosyltransferase [a]	Δ 105919:113454
	CETH_00459	NhaA, Na+/H+ antiporter [a]	
seg_7	CETH_00537	RpoC, RNApol β' [b]	Partial reversion to native sequence
seg_12	CETH_01304 – 1323	Ribosomal protein operon [a]	Δ 235844:236985
seg_13	CETH_01342	RarA, recombination protein [c]	His331 to Arg (CAC to CGC)
seg_16	CETH_01737	DnaB replicative DNA helicase [d]	Val266 to Gly (GTG to GGG)
seg_17	CETH_01760	LptD, LPS-assembly protein [a]	Gln202 to Met (CTG to ATG)
seg_19	CETH_01961	AccC, Biotin carboxylase [e]	ISCc3 insertion 383675: 383676
seg_21	CETH_01986 – 1992	LPS synthesis gene cluster [a]	Δ 395165:403283
seg_27	CETH_02535	TopA, DNA topoisomerase I [a]	Lys70 to Lys (AAA to AAG)
seg_35	CETH_03650	MurU, Mannose-1-P guanylyltransferase	Leu55 to Gly (GTC to GGC)
seg_38	CETH_03835	*parS*[f], FabB, oxoacyl synthase	Gly298 to Gly, (GGT to GGC)

[a] Toxicity demonstrated in *E. coli* ((17), PanDaTox database(18)).
[b] Change in RNApol β and β' stochiometry results in growth arrest in *E.coli* (19).
[c] Toxicity demonstrated in *E. coli* (20).
[d] Toxicity demonstrated in *E. coli* due to disbalance in DnaB:DnaC stochiometry (21).
[e] Toxicity demonstrated in *E. coli* due to disbalance in AccC:AccB stochiometry (22).
[f] Toxicity of extra-copy of centromeric *parS* site has been reported for *Caulobacter* (23).

Table S6. Deletions within *C. eth-2.0* chromosome segments in *Caulobacter*

Segment ID	Deletion	
	Coordinates	Size [bp]
seg_6	105919:113454	7,537
seg_12	235844:236985	1,143
seg_14	276227:279173	2,948
seg_21	395165:403283	8,120
seg_22	419856:420221	367
seg_25	482128:491570	9,444
seg_38	761193:773851	12,659
Cumulative size of deletions		42,218

Additional data table S1 (Data_S1.xlsx)
Functionality analysis of *C. eth-2.0* genome. Raw transposon insertion data, functionality values (z-score), gene annotation and functionality classification are listed for each rewritten *C. eth-2.0* gene.

Additional data table S2 (Data_S2.xlsx)
Fault diagnosis. Data set listing the identified error causes for faulty *C. eth-2.0* genes.

Additional data table S3 (Data_S3.xlsx)
Functionality of rewritten operon genes. Data set listing gene composition of identified functional *C. eth-2.0* operons.

References

1. Christen B, et al. (2011) The essential genome of a bacterium. *Mol. Syst. Biol.* 7(1):528–528.
2. Christen M, Del Medico L, Christen H, Christen B (2017) Genome Partitioner: A web tool for multi-level partitioning of large-scale DNA constructs for synthetic biology applications. *PloS one* 12(5):e0177234.
3. Liachko I, Youngblood RA, Keich U, Dunham MJ (2013) High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. *Genome Res.* 23(4):698–704.
4. Christen M, Deutsch S, Christen B (2015) Genome Calligrapher: A Web Tool for Refactoring Bacterial Genome Sequences for de Novo DNA Synthesis. *ACS synthetic biology* 4(8):927–934.
5. Cock PJA, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)* 25(11):1422–1423.
6. Zhou B, et al. (2015) The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLoS genetics* 11(1):e1004831.
7. Kouprina N, Noskov VN, Larionov V (2006) Selective isolation of large chromosomal regions by transformation-associated recombination cloning for structural and functional analysis of mammalian genomes. *Methods Mol. Biol.* 349:85–101.
8. Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* 2(1):31–34.
9. Adey A, et al. (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology* 11(12):R119.
10. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25(16):2078–2079.
11. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)* 27(21):2987–2993.
12. Spencer F, et al. (1994) Yeast kar1 mutants provide an effective method for YAC transfer to new hosts. *Genomics* 22(1):118–126.
13. Kouprina N, Larionov V (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast saccharomyces cerevisiae. *Nature protocols* 3(3):371.
14. Christen M, et al. (2015) Quantitative Selection Analysis of Bacteriophage φCbK Susceptibility in Caulobacter crescentus. *Journal of molecular biology*.
15. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 26(5):589–595.
16. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6(5):343–345.
17. Sorek R, et al. (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318(5855):1449–1452.
18. Amitai G, Sorek R (2012) Pandatox: a tool for accelerated metabolic engineering. *Bioengineered* 3(4):218–221.
19. Izard J, et al. (2015) A synthetic growth switch based on controlled expression of rna polymerase. *Molecular systems biology* 11(11):840.
20. Shibata T, et al. (2005) Functional overlap between reca and mgsa (rara) in the rescue of stalled replication forks in escherichia coli. *Genes to Cells* 10(3):181–191.
21. Allen G, Kornberg A (1991) Fine balance in the regulation of dnab helicase by dnac protein in replication in escherichia coli. *Journal of Biological Chemistry* 266(33):22096–22101.
22. Choi-Rhee E, Cronan JE (2003) The biotin carboxylase-biotin carboxyl carrier protein complex of escherichia coli acetyl-coa carboxylase. *Journal of Biological Chemistry* 278(33):30806–30812.
23. Toro E, Hong SH, McAdams HH, Shapiro L (2008) Caulobacter requires a dedicated mechanism to initiate chromosome segregation. *Proceedings of the National Academy of Sciences*.