

An exogenous chloroplast genome for complex sequence manipulation in algae

Bryan M. O'Neill^{1,*}, Kari L. Mikkelsen¹, Noel M. Gutierrez¹, Jennifer L. Cunningham¹, Kari L. Wolff¹, Shawn J. Szyjka¹, Christopher B. Yohn¹, Kevin E. Redding² and Michael J. Mendez¹

¹Sapphire Energy, Inc., San Diego, CA, 92121 and ²Department Chemistry & Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA

Received July 26, 2011; Revised October 19, 2011; Accepted October 20, 2011

ABSTRACT

We demonstrate a system for cloning and modifying the chloroplast genome from the green alga, *Chlamydomonas reinhardtii*. Through extensive use of sequence stabilization strategies, the *ex vivo* genome is assembled in yeast from a collection of overlapping fragments. The assembled genome is then moved into bacteria for large-scale preparations and transformed into *C. reinhardtii* cells. This system also allows for the generation of simultaneous, systematic and complex genetic modifications at multiple loci *in vivo*. We use this system to substitute genes encoding core subunits of the photosynthetic apparatus with orthologs from a related alga, *Scenedesmus obliquus*. Once transformed into algae, the substituted genome recombines with the endogenous genome, resulting in a hybrid plastome comprising modifications in disparate loci. The *in vivo* function of the genomes described herein demonstrates that simultaneous engineering of multiple sites within the chloroplast genome is now possible. This work represents the first steps toward a novel approach for creating genetic diversity in any or all regions of a chloroplast genome.

INTRODUCTION

A promise of synthetic biology is the ability to rationally alter metabolic processes in ways that would be impossible, or at least prohibitively difficult, through traditional approaches like chemical mutagenesis, breeding, or expression of even a few heterologous genes. While the field remains ill-defined, it generally includes leveraging advanced methods for synthesis and cloning of DNA

molecules to obtain novel sequences with desired functional properties (1,2). Some of synthetic biology's most notable achievements involve the assembly of genetic material into large DNA molecules that resemble chromosomal fragments or even whole genomes (3,4). Such large contigs may then be used to deliver all desired sequences into a target host in a single transformation step (5). However, as exogenous DNA molecules grow in size, so do the challenges of designing and maintaining correct sequences (6,7).

Chloroplast genomes present a unique opportunity for the field of synthetic biology. In a single, relatively small molecule, they encode the most important genes of photosynthesis, nature's principle method for converting sunlight into chemical energy and the progenitor of countless metabolites, and only a few other coding regions that support gene expression in the organelle (8). Most chloroplast genomes range between 150- and 205-kb, and many genomes representing diverse taxa have been sequenced and are publicly available (8). In addition, chloroplast transformation is a well-established technology in both plants and algae (9,10). Thus, these naturally minimized, manipulable genomes, which are of great interest for metabolic engineering for foods, fuels and myriad bio-products, are an ideally suited target for synthetic biology.

Photosynthesis is among the best-understood processes in biology. Studies conducted during the last few decades have revealed the architecture and mechanism of action for every component in the photosynthetic apparatus, including Photosystem II (PSII), a multi-subunit complex responsible for utilizing light energy in oxidoreduction reactions to extract electrons from water and produce oxygen (11). At the core of the PSII complex are four highly conserved proteins: D1, D2, CP43 and CP47. Photochemistry takes place in the D1/D2 heterodimer while CP43 and CP47 bind additional

*To whom correspondence should be addressed. Tel: +1 858 768 4766; Fax: +1 888 501 8353; Email: journal.publications@sapphireenergy.com, bryan.oneill@sapphireenergy.com

pigments to increase the absorption cross-section of the complex (12). These subunits interact extensively with one another via polypeptide sequences that are among the most conserved in all of biology (13). D1, CP47, CP43 and D2 are encoded by *psbA*, *psbB*, *psbC* and *psbD*, respectively, which exist at disparate locations in the chloroplast genome of all photosynthetic eukaryotes (8).

We sought to utilize the power of synthetic biology methods to directly manipulate the core genetics of photosynthesis by cloning a chloroplast genome from algae *ex vivo* using a yeast-bacteria hybrid system (14). This cloning system exploits yeast for its ability to stably maintain large DNA molecules and to support homologous recombination for sequence assembly and modification, and exploits bacteria for its ability to produce large quantities of specific DNA molecules, which are required for biolistic chloroplast transformation (14,15). Here, we demonstrate exogenous assembly and modification of the entire *Chlamydomonas reinhardtii* chloroplast genome, followed by transformation into algae cells and simultaneous alteration of at least six independent sites, including some that encode the core subunits of PSII.

MATERIALS AND METHODS

PCR primers

Primers used in this study are listed in Supplementary Table S1.

Vectors

Hybrid vector elements from pED-R2D2-ADE/URA [including a yeast centromere, yeast autonomously replication sequence, yeast selection marker (*TRP1*), bacterial replication origins (P1 rep and P1 lytic), and bacterial selection marker (Kan^r) (14)] were combined with algae-specific sequences, enabling maintenance of the exogenous algae chloroplast genome in yeast and bacteria. The combination of these DNA sequences is described here. The vector pDOCI was first generated to manipulate pED-R2D2-ADE/URA. Two portions of pED-R2D2-ADE/URA were amplified using PCR primer pairs (462 and 465, and 469 and 473) that anneal to sites surrounding the region encompassing TEL, ADE2 and URA3, assembled into a single DNA fragment by PCR assembly using a single primer pair (462 and 473), digested with NotI and ligated to a NotI-digested variant of pUC19, forming pDOCI. Portions of the *C. reinhardtii* chloroplast genome were then PCR-amplified using two primer pairs specific for adjacent regions near the *psbD* locus (791 and 792, and 793 and 794), digested with NotI and I-SceI and three-way ligated to I-SceI-digested pDOCI to form pDOCI-10. Next, a sequence conferring kanamycin resistance in algae (16) was cloned into pDOCI-10 using SnaBI to form pDOCI-10-Kan. The hybrid vectors for cloning chloroplast DNA, pTRP-10 and pTRP-10-Kan, were constructed using recombination in yeast. pDOCI-10 and pDOCI-10-Kan were digested with PacI and AscI to liberate cassettes that introduce chloroplast

genome-specific elements into the hybrid cloning vector. This cassette was transformed along with pED-R2D2-ADE/URA into yeast. Transformants that correctly integrated each cassette were isolated based on growth on CSM-Trp agar media containing 5-fluoroorotic acid (5-FOA) and by red color. Plasmid DNA was isolated from yeast clones that were grown in CSM-Trp liquid media, transformed into bacteria, isolated and verified by restriction mapping and DNA sequencing.

Often, large pieces of heterologous DNA are unstable in host organisms such as yeast or bacteria. To promote DNA sequence stability, a collection of vectors with yeast and bacterial selection markers was obtained. In one case, a fragment of DNA in pED-R2D2-ADE/URA encompassing the *ADE2* and *URA3* genes was liberated by digestion with SfiI and treated with Klenow fragment to create blunt ends. p322-Strep-ADE/URA was created by cloning this fragment into the PmlI site of a variant of p322 (17) modified to contain a streptomycin resistance marker regulated by the same elements as the kanamycin marker (16). In other cases, pRS416 (18) was used as a source for PCR amplification of *URA3*, pRS415 (18) was used as a source for PCR amplification of *LEU2*, and pRS413 (18) was used as a source for PCR amplification of *HIS3*. In a final case, *LYS2* from *Saccharomyces cerevisiae* genomic DNA and sequences that facilitate insertion between *psbA* and *atpE* in the *C. reinhardtii* chloroplast genome were cloned into pBeloBAC-11 (19).

The exogenous *C. reinhardtii* chloroplast genome was assembled in yeast from a collection of six overlapping fragments that together comprise the entire genome sequence (Figure 1A). Four of these vectors (pSC000, pSC002, pSC004 and pSC005) were obtained from a BAC library generated using total *C. reinhardtii* genomic DNA (clones 09L05, 11A06, 19G12 and 10K17, respectively; CRCCBa library obtained from Clemson University Genomics Institute, Clemson, SC). To allow maintenance of the assembled chloroplast genome in both cloning hosts, hybrid vector elements were introduced into pSC000 to form pSC006 (Supplementary Figure S1A). Briefly, pTrp-10 was digested with NotI to linearize the vector with algae-specific sequences at the termini and co-transformed with pSC000 into yeast. pSC006 plasmid DNA was isolated and verified by restriction mapping (Supplementary Figure S1A) and DNA sequencing. The remaining two vectors, pSC001 and pSC003, were cloned from genomic DNA and plasmids available from the Chlamydomonas Center collection at Duke University, respectively. To obtain pSC001, p322-Strep-ADE/URA and NotI-linearized pTRP-10-Kan were first co-transformed into *C. reinhardtii*. *Chlamydomonas reinhardtii* chloroplast DNA from clones that were successfully transformed with both constructs was used to transform bacteria for evaluation, which yielded one 125-kb DNA molecule wherein a recombination event occurred between the identical 3'-UTRs of the *C. reinhardtii* kanamycin and streptomycin resistance cassettes (details are available upon request). For pSC003, DNA fragments PCR-amplified from plasmids P-19, P-78, P-585 and P-586 (Chlamydomonas Center; Duke University) were assembled into a single DNA molecule

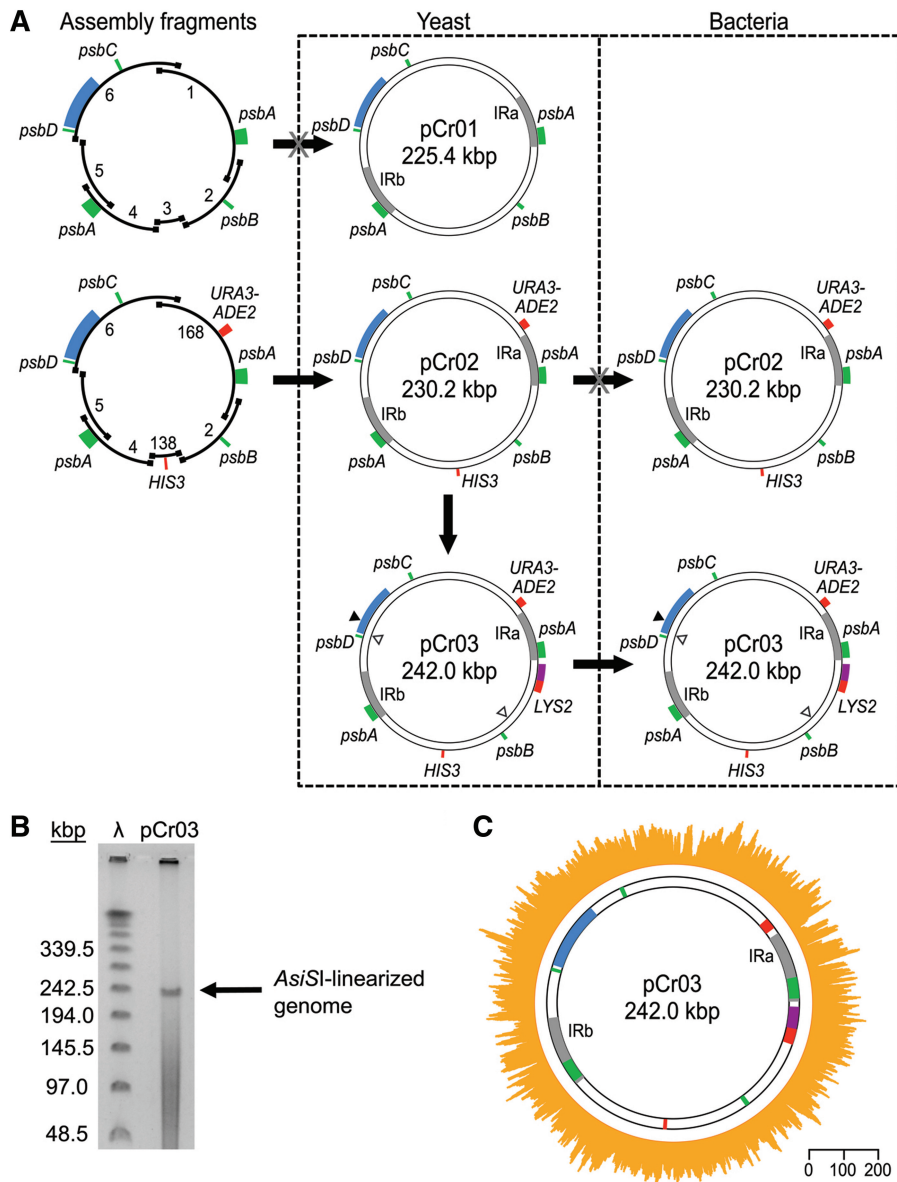


Figure 1. Cloning of the *C. reinhardtii* chloroplast genome. (A) Flow diagram of genome assembly and maintenance in yeast and bacteria. Each arrow represents a transformation step to assemble or modify the genome in yeast or to transfer the genome from yeast to bacteria. Arrows with a gray 'X' indicate transformations that did not yield the target genome in the desired cloning host. In all cases, green boxes indicate PSII-encoding genes, red boxes indicate yeast selection markers, blue boxes indicate the hybrid vector elements from pTrp-10, purple boxes indicate bacterial F-factor replication elements, and gray boxes indicate the large, inverted repeats. Numbers adjacent to each assembly fragment correspond to the pSC vector from which they were liberated. The solid triangles indicate the unique AsiSI restriction site and the open triangles indicate the RsrII restriction sites. (B and C) Analysis of the cloned chloroplast genome. (B) pCr03 was digested with AsiSI and analyzed by pulsed-field gel electrophoresis on a 1% agarose gel in 0.5× TBE. λ indicates the lambda ladder (NEB). (C) Sequence coverage of pCr03 from 454 FLX Titanium pyrosequencing. Features are identical to those in panel (A). Scale bar indicates fold coverage.

by recombination in yeast (details are available upon request).

Gene-disruption cassettes were synthesized and provided as inserts in vendor-provided cloning vectors (DNA2.0; Menlo Park, CA; sequences provided upon request). Each cassette contains a single copy of chloroplast-codon biased *aphA6* (for kanamycin resistance) that is flanked by the same 400–500 bp sequences that exist immediately 5' and 3' of the target gene in

C. reinhardtii. NdeI and SpeI restriction sites were placed at the *aphA6* start codon and immediately downstream of the stop codon, respectively; and NotI restriction sites were placed at both ends of each cassette to allow for linearization.

Gene replacement in the assembly fragment vectors occurred in two steps (Supplementary Figure S2). First, the endogenous PSII core complex genes were displaced using recombination in yeast. This step required the

displacement of five coding regions in four vectors with yeast selection marker cassettes that contain genes for positive selection (*URA3* and *ADE2* or *LEU2*), negative selection (*URA3*), and are targeted using 42-bp tails (Supplementary Figure S2; 5865 and 5866 for *psbAΔ* with *URA3/ADE2*, 5867 and 5868 for *psbBΔ* with *URA3/LEU2*, 5869 and 5870 for *psbCΔ* with *URA3/ADE2*, and 5871 and 5872 for *psbDΔ* with *URA3/LEU2*). For replacements using pSC002 and pSC005, yeast DNA replication and selection elements were added simultaneously with the displacement cassette (details are available upon request). Candidate gene-displaced clones were identified in yeast by PCR and transformed into *Escherichia coli*. Restriction analysis of vector DNA prepared from bacteria shows that all four gene-displaced vectors were generated (Supplementary Figure S2). The second step involved gene replacement. Gene-replacement cassettes were obtained by synthesizing the *(So)psbA* cDNA sequence (DNA2.0; Menlo Park, CA) or by PCR amplifying the PSII gene from *Scenedesmus obliquus* genomic DNA (primers 8316 and 8317 for *psbB*, 8318 and 8319 for *psbC*, and 8320 and 8321 for *psbD*), digesting with NdeI and SpeI, and cloning into its corresponding NdeI and SpeI-digested disruption vector. Linear replacement fragments were liberated by digestion with NotI, mixed with its corresponding gene-displaced assembly fragment vector and transformed into yeast. Candidate gene-replaced clones were isolated by negative selection using 5-FOA for the *URA3* gene, identified in yeast by PCR, and transformed into *E. coli*. Restriction analysis of vector DNA prepared from bacteria shows that all four gene-replaced vectors were generated (Supplementary Figure S2).

Yeast

Experiments were carried out at 30°C in rich media supplemented with 35 mg of adenine unless otherwise indicated. YSE004 (AB1380 *MATa ade2-1 can1-100 lys2-1 trp1 ura3 HIS5+ leu2Δ::lox his3Δ::lox-KanMX-lox*) was transformed using the spheroplast (20) or lithium acetate methods (21). DNA was prepared from yeast cells for bacterial transformation using the standard yeast miniprep method (22) modified to include alkaline lysis.

Bacteria

Experiments were carried out at 37°C in Luria-Bertani broth supplemented with antibiotics. 50 μl of DH10B cells (Life Technologies; Carlsbad, CA) were transformed with 2 μl of yeast miniprep DNA by electroporation in 0.2-cm cuvettes using a BioRad MicroPulser Electroporator (Bio-Rad; Hercules, CA). DNA was isolated from bacterial cells using the Qiagen Large-Construct Kit (Qiagen; Valencia, CA).

Algae

All algae strains used in this study were derived from either *C. reinhardtii* CC1690 wild-type mt+ 21gr or CC137c wild-type mt+ (Chlamydomonas Center; Duke

University). Cells were grown in either TAP or HSM medium and in liquid or on agar plates (23,24). Strains were grown at 28°C, in a 5% CO₂ chamber, with 150 μE of light.

For transformation, algae cells were grown to 10⁶ cells/ml in TAP, collected by centrifugation, and resuspended to 10⁸ cells/ml in TAP. A 500 μl aliquot of concentrated culture was spread on either TAP agar plates supplemented with 100 μg/ml of Kanamycin or on HSM agar plates. DNA was bound to 1 μm gold particles (Seashell Technology; La Jolla, CA, USA) according to the manufacturer's protocol and prepared and bombarded by a Helios Gene Gun (Bio-Rad; Hercules, CA, USA) according to the manufacturer's instructions. Gold particles were shot using 300 psi of Helium at a distance of 6 cm from the plate.

Photosynthetically deficient strains were created by transforming individual, synthetic gene disruption vectors into wild-type algae cells. Homoplasmic transformants were identified by PCR screening and maintained on TAP in the dark.

PCR screening

For yeast, templates for PCR analysis were prepared by suspending cells in lysis buffer (6 mM KHPO₄, pH = 7.5, 6 mM NaCl, 3% glycerol, 3U zymolyase 100T (MP Biomedicals; Solon, OH, USA), incubating at 37°C for 30 min, 95°C for 10 min, and then cooling to 10°C. 2 μl of lysate were used for each PCR reaction. For bacteria, cells were added directly to the PCR reaction. For algae, 10⁶ cells were suspended in lysis buffer (10 mM EDTA, pH = 8), heated to 95°C for 10 min, and cooled to 10°C. 2 μl of lysate were used for each PCR reaction.

Southern blots

To isolate genomic DNA from algae, total DNA was prepared from cells grown on HSM agar plates. Approximately 10⁷ cells were collected, suspended in lysis buffer (150 mM Tris-HCl, pH = 7.5, 200 mM NaCl, 20 mM EDTA and 1% SDS), incubated for 1 h at 37°C, extracted once with phenol/chloroform (1:1) and twice with chloroform, ethanol precipitated and resuspended in TE buffer (10 mM Tris-HCl, pH = 7.4, 1 mM EDTA, 50 μg/ml RNase). For digestions, 10 μg of total algae genomic DNA or 250 ng of the exogenous chloroplast genome DNA isolated from bacteria were incubated with EcoRI-HF or NdeI (New England Biolabs; Ipswich, MA, USA) for 3 h at 37°C in a total volume of 50 μl. Digestion products were separated on a 0.7% agarose gel run in TAE. Gel-embedded samples were then incubated in depurination buffer (0.25 N HCl) for 8 min, denaturation buffer (0.5 N NaOH and 1.5 M NaCl) for 18 min, neutralization buffer (1M Tris-HCl, 1.5 M NaCl, pH = 7.4) for 20 min and transferred to a Hybond-N+ membrane (GE Healthcare; Piscataway, NJ, USA). The DIG High Prime DNA Labeling and Detection Starter Kit II (Roche; Indianapolis, IN, USA) was used to probe Southern blots according to the manufacturer's protocol. Probes specific for sequences adjacent to integration sites for M1, M2, M3, and M4, were generated using

primers 22907 and 22908, 22911 and 22912, 22919 and 22920, and 22905 and 22906, respectively.

PSII quantum yield measurements

Cells grown to mid-log growth in TAP were placed directly into a 1-cm cuvette and immediately put in a JTS-10 LED spectrometer (Bio-Logic; Claix, France). Every measurement was preceded by a dark adaption of 90–120 s. Fluorescence emission was measured by using a 10- μ s pulse of light from a 520-nm LED as the excitation source, and a 670-nm high-pass filter was placed in front of the sample detector to measure fluorescence emission from Chl *a*. A BG39 filter was placed in front of the reference detector to normalize fluorescence from each pulse to the excitation intensity. The initial fluorescence value (F_0) was measured five times every 1 s (and averaged), followed immediately by a 80-ms saturating pulse to fully oxidize the PQ pool, and then fluorescence was measured 200 μ s after the pulse was over to obtain the maximal fluorescence (F_M). Variable fluorescence (F_V/F_M) was calculated as $(F_V - F_0)/F_M$. Measurements were performed three times and averaged.

RESULTS

Cloning the chloroplast genome from *C. reinhardtii*

The fully sequenced 204-kb *C. reinhardtii* chloroplast genome exists as both circular and linear molecules *in vivo* (25,26). This genome is 66% A–T and possesses two copies of a 22-kb inverted repeat sequence (IRA and IRb) separated by two, nearly equally sized unique regions of ~80 kb. Additional short, dispersed repeats make the overall DNA composition >20% repetitive. We chose to assemble the plastid genome as a circular molecule in yeast from a collection of six overlapping fragments that, together, comprise the entire genome sequence (Figure 1A).

Initial attempts to assemble and isolate the *C. reinhardtii* chloroplast genome from just the six fragments were unsuccessful (Figure 1A). It is likely that this strategy failed because of the large inverted- and/or short dispersed repeats of the algal plastome; therefore, we inserted yeast positive selection markers at two additional positions in the overlapping fragments (*URA3-ADE2* into pSC001 to produce pSC168 and *HIS3* into pSC003 to produce pSC138, Figure 1A and Supplementary Figure S1 and ‘Materials and Methods’ section) to facilitate isolation of correct clones and promote sequence stability. Linearized, overlapping fragments were mixed in equimolar amounts and introduced into yeast cells by spheroplast transformation. When selected by growth on media lacking tryptophan, we obtained approximately 100 transformants/ μ g of DNA. The vast majority of these were pink, which indicated that the fragment containing the *ADE2* gene was absent in these clones and that they had likely arisen from undesired recombination events. When selected on media lacking tryptophan, uracil and histidine, we regularly obtained only four transformants per μ g of DNA, but all isolated clones were white, which indicated that they possessed all four yeast selection

markers. Results from PCR-based screens indicated that 10% of the transformants selected on the triple-dropout media (3 of 30) contained a properly assembled genome, a molecule we call pCr02 (Figure 1A and Supplementary Figure S3). These results suggest that the additional yeast markers enabled selection of recombination products that contained the complete chloroplast genome.

Attempts to transform pCr02 from yeast into *E. coli* were unsuccessful (Figure 1A), indicating that the assembled genome may be unstable in bacteria. To promote plastome stability in bacteria, we attempted to integrate the *E. coli* F-factor replicon from pBeloBAC-11 (19) into the candidate pCr02-containing yeast clones (to produce pCr03) at a position opposite the original bacterial origin (Figure 1A and ‘Materials and Methods’ section). DNA was prepared from candidate pCr03-containing yeast clones, transformed into bacteria and analyzed by restriction mapping and DNA sequencing. Analysis of one of the pCr03 clones indicated that the isolated genome was the expected size of 242-kb (Figure 1B). Consistently, alignment of data obtained from 454 FLX Titanium pyrosequencing (done by Engencore; Columbia, SC) to the predicted reference produced a contig with 83-fold average coverage (Figure 1C). The reference sequence in this case was based on an assembly of the plastid DNA trace files generated by the *C. reinhardtii* nuclear genome sequencing project (26), which aligned much better to the 454 sequence data than did the canonical sequence (25). Over 500 1- or 2-nt indels were observed in homopolymer runs of A or T (data not shown); however, since many of these differences were also observed in data obtained from 454 sequencing of the assembly fragments and read-length errors in homopolymer runs are a known artifact in 454 sequencing (27), we disregarded the small indels. We also observed a few larger insertions and deletions (Supplementary Table S2), but these differences were also found in DNA isolated from wild-type algae. Thus, we conclude that no significant mutations were introduced during the multi-step cloning procedure.

In vivo function of the cloned *C. reinhardtii* chloroplast genome

To characterize *in vivo* function of the cloned *C. reinhardtii* chloroplast genome, we used a strategy that relies on the restoration of photosynthesis as the primary selection (Figure 2). We first created a collection of photosynthetically deficient algae strains into which the genome could be transformed (Supplementary Figure S4 and ‘Materials and Methods’ section). Each strain in the collection contains a knockout in the coding region of *psbA*, *psbB*, *psbC* or *psbD*; these modifications render them unable to grow without the addition of organic carbon to compensate for their inability to perform photosynthesis (Supplementary Figure S4). Each strain in the collection is also complemented by transformation with its corresponding wild-type locus (Supplementary Figure S5), demonstrating that they are all transformable. Because its complementation should be tightly linked with the presence of the hybrid cloning elements (Figure 2),

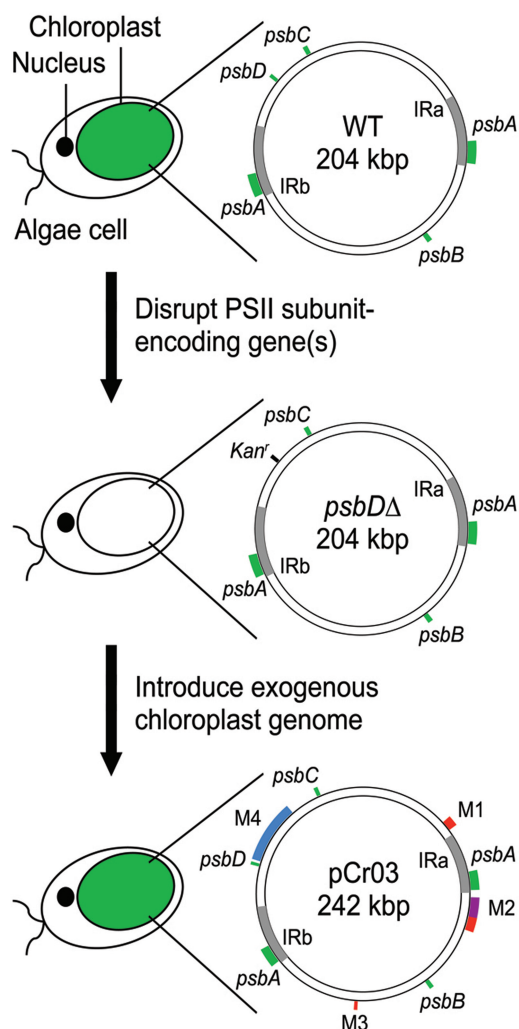


Figure 2. Exogenous chloroplast genome transformation strategy. Color of the chloroplast indicates whether it is able to carry out photosynthesis (green) or not (white). Green boxes indicate PSII-encoding genes, red boxes indicate yeast selection markers, blue boxes indicate the hybrid vector elements from pTrp-10, and purple boxes indicate bacterial F-factor replication elements. Unique markers for screening are labeled M1–M4.

we chose to use the *psbD* knockout strain for initial genome transformations.

The *psbD* knockout strain was transformed with pCr03 (Figure 2) using standard biolistic particle bombardment (see ‘Materials and Methods’ section) and routinely gave rise to more than 100 colonies per transformation. As expected, almost all of the primary transformants (98.4%) screened positive for the hybrid vector backbone (*URA3-ADE2*, F-factor replicon, *HIS3* and hybrid vector backbone, are hereafter called M1, M2, M3 and M4, respectively). Out of 1023 primary pCr03 transformants, 53 (5.2%) screened positive by PCR for the four elements used to clone and stabilize the chloroplast genome (Figure 3A and Table 1). And since transformation of the *psbD* knockout strain with pCr03 digested with RsrII (which cuts pCr03 at two positions; Figure 1A) did not give rise to any clones that were

confirmed to contain M1–M4 (Table 1), the linkage of all four genetic markers observed in the 53 isolates appears to be dependent on pCr03 initially being an intact circular molecule. To examine their genetic stability, 12 of the 53 candidate clones were struck once for single colonies and grown photoautotrophically. PCR screens of these secondary clones indicated that most had lost one or more unique markers, especially M1 and/or M3 (Table 1). However, we were able to obtain clones that retained all four markers, and a few of these were homoplasmic by Southern blot analysis at the M1, M2 and M3 loci (Figure 3B–D and Supplementary Figure S6). We were unable, though, to identify any clones that were homoplasmic at the M4 locus (Figure 4E and Supplementary Figure S6), which may indicate that M4 is not in a viable insertion site or that this piece of DNA is incompatible in some way with the chloroplast genome. These results suggest that the exogenous *C. reinhardtii* chloroplast genome is able to replicate *in vivo*.

Modifying genes encoding the PSII core complex

To explore the potential of the exogenous *C. reinhardtii* chloroplast genome for simultaneous modification of photosynthetic apparatus-encoding genes, we sought to replace the genes encoding the PSII core complex with both synthetic and traditionally cloned orthologs from *S. obliquus* (28). A synthetic cDNA sequence was used for (*So*)*psbA* (which naturally contains introns) and (*So*)*psbB*, (*So*)*psbC* and (*So*)*psbD* were cloned from natural templates (see ‘Materials and Methods’ section). Although their DNA sequences are less well conserved, these genes encode proteins are at least 98% similar and 93% identical to their *C. reinhardtii* ortholog. Each of these genes complemented basic photoautotrophic growth in their respective knockout strain (Supplementary Figure S5). While (*So*)*psbA*, (*So*)*psbC* and (*So*)*psbD* support growth near the level of the same gene from *C. reinhardtii*, (*So*)*psbB* imparts a growth defect (Supplementary Figure S5). Analysis of the quantum efficiency of PSII in each of the complemented strains indicates that while (*So*)*psbB* supports basic PSII function, it does not fully restore PSII function (Figure 4A). However, because of their ability to support basic PSII function, all four genes from *S. obliquus* were carried forward to create a modified exogenous chloroplast.

Gene replacement is a well established technique in yeast molecular biology and can be accomplished without leaving behind a selection marker, enabling unlimited cycles of modification (Figure 4B). In this case, we chose to perform all gene replacements in parallel in the appropriate assembly fragments and assemble the variant genomes *de novo* (Figure 4C and Supplementary Figure S2). The pCr05 DNA molecule was assembled using a strategy similar to that used for pCr03 (Figure 4C). The strategies differ only in the length of one of the assembly fragments (fragment from pSC004), which was truncated (forming pSC093) to remove a redundant copy of *psbA* and reduce the number of gene replacements required. As expected, candidate clones were obtained in both yeast

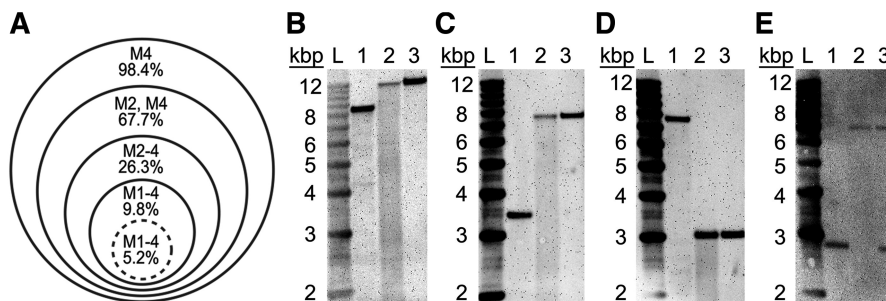


Figure 3. Characterization of the cloned *C. reinhardtii* chloroplast genome *in vivo*. (A) A nested set representing the presence of increasing numbers of markers in primary transformants of pCr03 into a *psbD* knockout strain as determined by PCR (Table 2; primers used as follows: M1, 11 606 and 11 607; M2, 5512 and 5513; M3, 11 456 and 11 457; and M4, 14 067 and 14 068.). The broken circle shows the subset of transformants with M1, M2, M3 and M4 that gave rise to the same genotype upon rescreening. (B–E) Southern blot analysis of EcoRI (B, C and E) or NdeI (D) digests (see ‘Materials and Methods’ section). Probes were specific for sequences adjacent to integration sites for M1 (B), M2 (C), M3 (D) and M4 (E). All samples are arranged as follows: Lane L, 1 kb DNA ladder (Invitrogen; Carlsbad, CA); lane 1, wild-type; lane 2, purified pCr03; and lane 3, a representative algae clone containing all unique markers. A single band in lane 3 indicates homoplasmic integration of the marker, while two bands indicate heteroplasmy with the wild-type locus.

Table 1. Genotypic analysis of *psbD*-deficient algae transformed with pCr03

Marker(s)	Primary transformants (%) ^a			Secondary clones (%) ^b											RsrII-cut primary transformants (%) ^c		
	Total	Full genotype	Confirmed genotype ^d	A1	A6	A7	B1	B11	C5	D1	D9	E1	F6	G1	H1	Full genotype	Confirmed Genotype ^e
None	–	0.6	–	10	1	0	2	0	33	3	6	11	1	2	0	2.4	–
M1	16.1	0	–	0	0	0	0	0	0	0	0	0	0	0	1	0	–
M2	68.6	0.6	–	8	35	5	45	39	6	12	5	0	0	2	5	0	–
M3	29.2	0.1	–	1	0	0	0	0	0	0	0	0	0	0	0	0	–
M4	98.4	27.3	–	58	45	19	27	11	39	30	8	63	16	66	1	50.0	–
M1, M2	–	0	–	0	0	6	0	3	0	0	9	0	0	0	63	0	–
M1, M3	–	0	–	0	0	0	0	0	0	0	1	0	0	0	0	0	–
M1, M4	–	0.9	–	0	1	0	2	0	0	0	0	12	0	0	0	0.3	–
M2, M3	–	0.2	–	0	0	0	0	0	0	1	2	0	0	0	0	0	–
M2, M4	–	36.2	–	23	16	48	18	32	18	42	22	12	51	28	12	31.2	–
M3, M4	–	2.4	–	1	0	1	0	0	0	0	6	0	2	1	0	2.4	–
M2, M3, M4	–	16.5	–	0	0	1	0	0	3	12	16	0	9	1	3	8.9	–
M1, M3, M4	–	0.1	–	0	0	0	0	0	0	0	2	0	0	0	0	0	–
M1, M2, M4	–	5.3	–	0	1	18	0	13	0	0	3	2	16	0	6	1.3	–
M1, M2, M3	–	0.1	–	0	0	0	0	0	0	0	0	0	0	0	0	0	–
M1–M4	–	9.8	5.2	0	0	0	0	2	0	0	19	0	0	0	8	2.2	0

^aReported as a fraction of 1023 primary transformants.

^bResult from 93 isolates for each clone.

^cReported as a fraction of 372 primary transformants.

^dThe 100 isolates that initially screened positive for M1–M4 were re-screened for M1–M4.

^eThe eight isolates that initially screened positive for M1–M4 were re-screened for M1–M4.

and bacteria, and a representative pCr05 clone was selected based on its DNA size (pCr05 is predicted to be 231-kb; Figure 4D) and data generated from 454 FLX Titanium pyrosequencing (Supplementary Figure S7). These data demonstrate that the exogenous *C. reinhardtii* chloroplast genome provides a cloning platform for simultaneous modification of many genes, including those that encode multi-subunit protein complexes.

***In vivo* function of a cloned *C. reinhardtii* chloroplast genome with modified PSII subunit-encoding genes**

Like pCr03, pCr05 was transformed into the *psbD* knockout strain (Figure 2). This selection strategy

ensured that, at a minimum, the endogenous *psbD* coding sequence would be substituted with the ortholog from *S. obliquus*; the ability of pCr03 to simultaneously modify four independent loci suggested that the other PSII-encoding loci could be replaced without direct selection. We first screened for M1, the least frequently observed marker from pCr03. Out of 837 primary pCr05 transformants, 73 (8.7%) screened positive by PCR for this marker (Table 2). Subsequent screening identified five of these isolates as containing all four of the unique markers and (*So*)*psbC* and (*So*)*psbD* (Table 2). Further analysis demonstrated that we were able to obtain clones that were homoplasmic for the *psbC* and *psbD* orthologs (Supplementary Figure S8 and Table 2). (*So*)*psbA* was

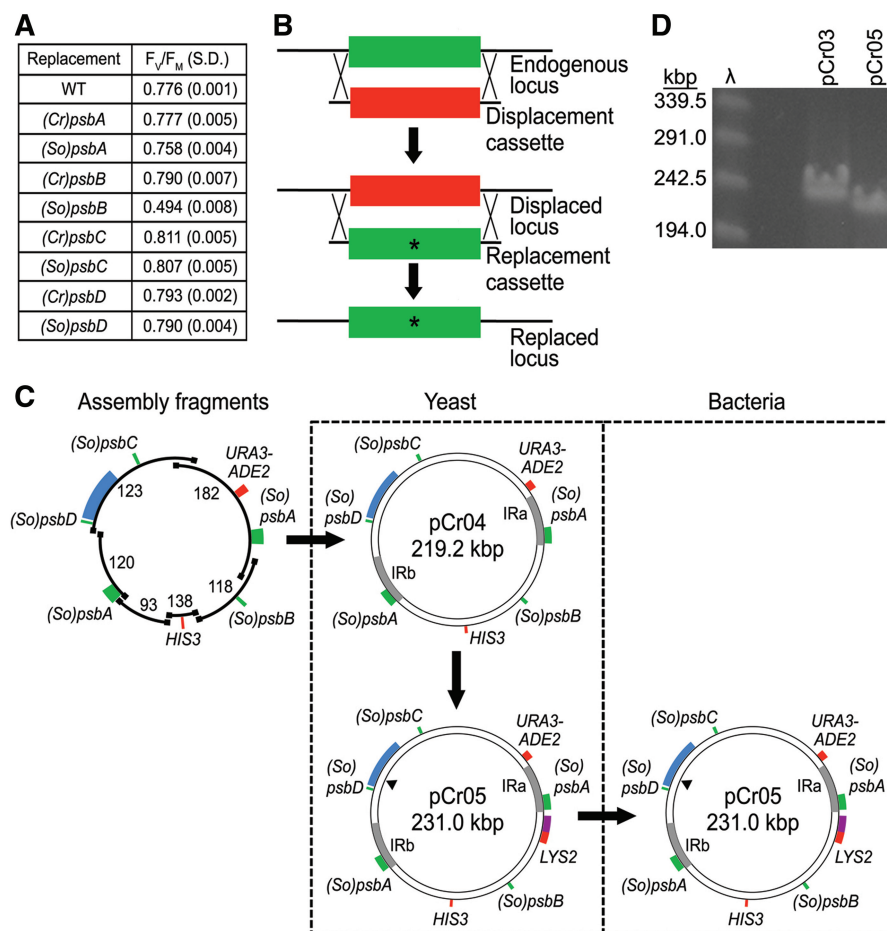


Figure 4. Gene replacement in an exogenous chloroplast genome. (A) Quantum efficiency of PSII in strains with a single PSII gene replacement. Values are reported as the average and standard deviation of three analytical replicates of a representative transformant. (B) The two steps of gene replacement. The endogenous locus (green box) is first displaced by a cassette containing one positive and one negative selection marker (red box) using recombination in yeast. The displacement cassette is then replaced with a new sequence (green box with 'Asterisk') targeted by the same flanking regions. (C) Flow diagram of genome assembly and maintenance in yeast and bacteria. Each arrow represents a transformation step to assemble or modify the genome in yeast or to transfer the genome from yeast to bacteria. In all cases, green boxes indicate PSII-encoding genes, red boxes indicate yeast selection markers, blue boxes indicate the hybrid vector elements from pTrp-34, purple boxes indicate bacterial F-factor replication elements, and gray boxes indicate the large, inverted repeats. Numbers adjacent to each assembly fragment correspond to the pSC vector from which they were liberated. The black triangle in pCr05 indicates the unique AsiSI restriction site. (D) DNA isolated from bacteria was digested with AsiSI and analyzed by pulsed-field gel electrophoresis on a 1% agarose gel in $0.5\times$ TBE. λ indicates the lambda ladder.

not detected in any of the 73 candidates and while (*So*)*psbB* was detected in a few of the candidates, it was lost during cell passage (Table 2). The loss of (*So*)*psbB* was consistent with the hypothesis that a selection pressure is exerted against the gene because it confers a decreased quantum yield of PSII. It is unclear why (*So*)*psbA* was not detected; one possible explanation is that the endogenous *psbA* loci are recalcitrant to modification when not directly selected for.

Taken together, all of the *in vivo* data demonstrate that the cloned *C. reinhardtii* chloroplast genome provides a platform for simultaneous multi-site modification of the photosynthetic apparatus. Replaced sequences can be native or orthologous genes, traditionally cloned or synthetically created. These results also show that this approach can be used to simultaneously modify at least six disparate regions in the chloroplast genome *in vivo*.

DISCUSSION

We have demonstrated the capacity for a yeast-bacteria hybrid cloning system to create genetic diversity in the chloroplast genome at a scale and complexity that would be otherwise impractical using traditional methods. The system may also prove useful for sequence deletions and/or more complicated manipulations like genome rearrangement and minimization. The utility of genome-scale cloning and modification systems has been questioned in other organisms (29); however, in the chloroplast, the need is clear. It has already been demonstrated that, despite tremendous effort and interest, traditional mutagenesis techniques, especially those targeted at single proteins, have failed to yield significantly improved rates of photosynthesis (30,31). Key photosynthetic enzymes are macromolecular assemblies composed of many individually coded elements. In order

Table 2. Genotypic analysis of *psbD*-deficient algae transformed with pCr05

Marker(s)	Primary transformants (%) ^a			Secondary clones (%) ^b											
	Primary genotype	Partial genotype ^c	Full Genotype ^d	A3 ^e	A7 ^f	A12 ^g	B9 ^f	B10 ^h	C4 ^h	C8 ^f	C11 ^f	D2 ^g	D5 ^h	D8 ^f	D12 ^h
None	–	–	–	0	0	0	19	0	66	6	44	0	0	1	0
M1	8.7	2.8	0.1	0	0	0	4	10	9	4	0	0	0	1	0
M2	–	–	–	5	1	1	0	0	5	0	0	0	0	0	0
M3	–	–	–	0	0	3	0	15	0	0	0	0	0	0	0
M4	–	–	–	8	4	11	20	0	12	12	54	0	0	67	0
M1, M2	–	–	–	3	0	0	0	0	0	0	0	0	4	0	0
M1, M3	–	6.0	0	0	0	0	0	24	0	0	0	0	0	0	0
M1, M4	–	–	–	6	6	0	55	1	1	26	1	0	1	8	0
M2, M3	–	–	–	0	1	0	0	32	0	0	0	0	0	0	0
M2, M4	–	–	–	55	47	11	1	2	8	4	1	8	4	8	0
M3, M4	–	–	–	2	3	66	0	2	0	0	0	4	0	0	0
M2, M3, M4	–	–	–	4	25	8	0	2	0	2	0	34	0	14	0
M1, M3, M4	–	–	2.4	0	0	0	0	4	0	0	0	0	1	0	0
M1, M2, M4	–	–	0	16	10	0	0	1	0	19	0	1	85	0	1
M1, M2, M3	–	–	0	0	0	0	0	0	0	0	0	2	0	0	0
M1–M4	–	–	1.4	0	2	0	0	6	0	26	0	51	4	2	99
(<i>So</i>) <i>psbB</i>	–	–	–	4	–	0	–	–	–	–	–	3 ⁱ	–	–	–
(<i>So</i>) <i>psbC</i>	–	–	–	–	–	100	–	90	0	–	–	53 ⁱ	97	–	100

^aReported as a fraction of 837 primary transformants.

^bResult from 93 isolates for each clone.

^cResults from 73 isolates that initially screened positive for M1 and were screened for M3.

^dResults from 50 isolates that initially screened positive for M1 and M3 and were screened for M1–M4.

^ePCR screening of primary isolates indicated that (*So*)*psbB* and (*So*)*psbD* were present.

^fPCR screening of primary isolates indicated that only (*So*)*psbD* was present.

^gPCR screening of primary isolates indicated that (*So*)*psbB*, (*So*)*psbC* and (*So*)*psbD* were present.

^hPCR screening of primary isolates indicated that (*So*)*psbC* and (*So*)*psbD* were present.

ⁱPositive PCR results were non-overlapping.

to explore the states of advanced photosynthetic function, it is likely necessary to evaluate a large genetic state space over a number of genes, simultaneously. The exogenous chloroplast system shown here allows for rapid, systematic, and combinatorial generation of multi-gene photosynthetic mutants. Techniques such as these will be necessary to realize the promise of rationally altering photosynthetic metabolism.

Cloning of the *C. reinhardtii* plastome in both yeast and bacteria required several approaches to combat DNA sequence instability. Larger DNA molecules have been stably cloned in yeast and bacteria without the need for the innovations described here (5,32); however, the highly repetitive nature of the chloroplast genome is an unmet challenge (25). In yeast, we used positive selection markers to promote correct genome assembly by homologous recombination. In bacteria, we used two origins of DNA replication placed on opposite sides of the circular genome to allow for stable maintenance of the plasmid. It remains unclear whether this approach is required because of the genome's size, repetitiveness, or some combination of these and other factors. Regardless, the use of two or more single-copy origins of DNA replication may enable the stable passage of other large and complex molecules through bacteria. This has the obvious benefit of supporting large-scale DNA preparations, but also allows for the use of bacterial genetics and recombination, as well as strains that produce DNA with various methylation patterns. Chloroplast genomes from higher

plants may not have similar challenges with sequence stability in the cloning hosts because they are generally smaller (~160 kb). In fact, the 152-kb soybean plastid genome was cloned as a single bacterial artificial chromosome (33), demonstrating that it is already stable in bacteria.

The *in vivo* function of the genomes described herein demonstrates that simultaneous engineering of multiple sites within the chloroplast genome is now possible. However, to realize the goal of exploring extensive sequence diversity in any or all photosynthetic apparatus-encoding subunits, at least three improvements to this system are required. First, better strategies are needed to combat the recalcitrance of some loci to modification. This may include increased transformation efficiencies, improved algae transformation strains (i.e. multi-gene knockouts), or relocating the hybrid vector elements in M4 to a site that allows for homoplasmic modification. Second, more efficient genome assembly methods are necessary to rapidly create libraries of genome-scale diversity. Pre-integrating the F-factor elements into the fragment(s) derived from pSC001 (Figure 1A) could eliminate the need for an intermediate cloning step in yeast before passage to bacteria. Third, applying a synthetic diversity creation technique able to systematically saturate any region of a subunit's coding sequence with genetic alterations (34) will enable the practical exploration of extensive, non-natural genetic diversity.

Future work is needed to explore the interactions that exist within and between the multi-subunit complexes of the photosynthetic apparatus. That these are among the most conserved sequences in all of biology points to the limited evolutionary diversity that natural mutagenesis or horizontal gene transfer has been able to explore in these essential processes (13,35). We anticipate that the ability to change most or all chloroplast-encoded components at once—for example, to swap all four PSII core subunits with heterologous or modified proteins—may be necessary to spring the photosystem out of an evolutionarily frozen well. As the gatekeeper between the inorganic and organic worlds, the chloroplast is a metabolic engineering target of great importance. An improved photosynthetic apparatus may lead to the crop yield increases necessary to feed our expanding world population, further develop a sustainable green chemistry industry and move closer to commercially-viable algal biofuel.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–8.

ACKNOWLEDGEMENTS

The authors thank K. Botsch for technical assistance and J. Pyle, Y. Poon, S. Briggs, S. Mayfield and N. David for critical reading of the manuscript.

FUNDING

Sapphire Energy, Inc. Funding for open access charge: Sapphire Energy, Inc.

Conflict of interest statement. Bryan O'Neill and Mike Mendez have financial interests in Sapphire Energy, Inc.

REFERENCES

- Carr, P.A. and Church, G.M. (2009) Genome engineering. *Nat. Biotechnol.*, **27**, 1151–1162.
- Khalil, A.S. and Collins, J.J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.*, **11**, 367–379.
- Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science*, **319**, 1215–1220.
- Itaya, M., Tsuge, K., Koizumi, M. and Fujita, K. (2005) Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl Acad. Sci. USA*, **102**, 15971–15976.
- Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M. *et al.* (2010) Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, **329**, 52–56.
- Kouprina, N., Eldarov, M., Moyzis, R., Resnick, M. and Larionov, V. (1994) A model system to assess the integrity of mammalian YACs during transformation and propagation in yeast. *Genomics*, **21**, 7–17.
- Larionov, V., Kouprina, N., Eldarov, M., Perkins, E., Porter, G. and Resnick, M. (1994) Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast*, **10**, 93–104.
- Gao, L., Su, Y.-J. and Wang, T. (2010) Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects. *J. Sys. Evol.*, **48**, 77–93.
- Maliga, P. and Bock, R. (2011) Plastid biotechnology: food, fuel, and medicine for the 21st century. *Plant Physiol.*, **155**, 1501–1510.
- Specht, E., Miyake-Stoner, S. and Mayfield, S. (2010) Micro-algae come of age as a platform for recombinant protein production. *Biotech. Lett.*, **32**, 1373–1383.
- Nelson, N. and Ben-Shem, A. (2004) The complex architecture of oxygenic photosynthesis. *Nature Rev. Mol. Cell. Biol.*, **5**, 971–982.
- Zouni, A., Witt, H.T., Kern, J., Fromme, P., Krauss, N., Saenger, W. and Orth, P. (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature*, **409**, 739–743.
- Shi, T., Bibby, T.S., Jiang, L., Irwin, A.J. and Falkowski, P.G. (2005) Protein interactions limit the rate of evolution of photosynthetic genes in Cyanobacteria. *Mol. Biol. Evol.*, **22**, 2179–2189.
- Hokanson, C.A., Dora, E., Donahue, B.A., Rivkin, M., Finer, M. and Mendez, M.J. (2003) Hybrid yeast-bacteria cloning system used to capture and modify adenoviral and nonviral genomes. *Human Gene Ther.*, **14**, 329–339.
- Boynton, J., Gillham, N., Harris, E., Hosler, J., Johnson, A., Jones, A., Randolph-Anderson, B., Robertson, D., Klein, T., Shark, K. *et al.* (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science*, **240**, 1534–1538.
- Rasala, B.A., Muto, M., Lee, P.A., Jager, M., Cardoso, R.M.F., Behnke, C.A., Kirk, P., Hokanson, C.A., Crea, R., Mendez, M. *et al.* (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.*, **8**, 719–733.
- Franklin, S., Ngo, B., Efu, E. and Mayfield, S.P. (2002) Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J.*, **30**, 733–744.
- Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19–27.
- Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I. and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics*, **34**, 213–218.
- Noskov, V.N., Kouprina, N., Leem, S.H., Ouspenski, I., Barrett, J.C. and Larionov, V. (2003) A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics*, **4**, 16–26.
- Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87–96.
- Hoffman, C.S. and Winston, F. (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene*, **57**, 267–272.
- Gorman, D. and Levine, R. (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA*, **54**, 1665–1669.
- Harris, E. (1989) *The Chlamydomonas Sourcebook: A Comprehensive Guide to Biology and Laboratory Use*. Academic Press, San Diego, CA.
- Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H. and Stern, D.B. (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *Plant Cell*, **14**, 2659–2679.
- Smith, D. and Lee, R. (2009) Nucleotide diversity of the *Chlamydomonas reinhardtii* plastid genome: addressing the mutational-hazard hypothesis. *BMC Evol. Biol.*, **9**, 120.
- Balzer, S., Malde, K., Lanzen, A., Sharma, A. and Jonassen, I. (2010) Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim. *Bioinformatics*, **26**, i420–i425.
- de Cambiaire, J.-C., Otis, C., Lemieux, C. and Turmel, M. (2006) The complete chloroplast genome sequence of the chlorophyceae

- green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. *BMC Evol. Biol.*, **6**, 37.
29. Noireaux, V., Maeda, Y. and Libchaber, A. (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl Acad. Sci. USA*, **108**, 3473–3480.
30. Parikh, M., Greene, D., Woods, K. and Matsumura, I. (2006) Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. *Prot. Engineer. Design Selection*, **19**, 113–119.
31. Smith, S. and Tabita, F. (2003) Positive and Negative Selection of Mutant Forms of Prokaryotic (Cyanobacterial) Ribulose-1,5-bisphosphate Carboxylase/Oxygenase. *J. Mol. Biol.*, **331**, 557–569.
32. Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.
33. Sasaki, C., Lee, S.B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.G. and Jansen, R.K. (2005) Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.*, **59**, 309–322.
34. Rajpal, A., Beyaz, N., Haber, L., Cappuccilli, G., Yee, H., Bhatt, R.R., Takeuchi, T., Lerner, R.A. and Crea, R. (2005) A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc. Natl Acad. Sci. USA*, **102**, 8436–8471.
35. Shi, T. and Falkowski, P.G. (2008) Genome evolution in cyanobacteria: The stable core and the variable shell. *Proc. Natl Acad. Sci. USA*, **105**, 2510–2515.