

Video Article

CAPRESI: Chimera Assembly by Plasmid Recovery and Restriction Enzyme Site Insertion

Orlando Santillán¹, Miguel A. Ramírez-Romero², Guillermo Dávila³¹Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México²Oncomedic Incorporation, Ciudad de México³Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de MéxicoCorrespondence to: Orlando Santillán at osantillan@icg.unam.mxURL: <https://www.jove.com/video/55526>DOI: [doi:10.3791/55526](https://doi.org/10.3791/55526)

Keywords: Molecular Biology, Issue 124, Chimeric protein coding gene assembly synonym DNA sequence insertion, restriction enzyme site insertion, plasmid recovery, pUC18/19 vector, antibiotic resistance gene disruption, genetic engineering, fusion protein, primary sigma factor, RpoD, SigA

Date Published: 6/25/2017

Citation: Santillán, O., Ramírez-Romero, M.A., Dávila, G. CAPRESI: Chimera Assembly by Plasmid Recovery and Restriction Enzyme Site Insertion. *J. Vis. Exp.* (124), e55526, doi:10.3791/55526 (2017).

Abstract

Here, we present chimera assembly by plasmid recovery and restriction enzyme site insertion (CAPRESI). CAPRESI benefits from many strengths of the original plasmid recovery method and introduces restriction enzyme digestion to ease DNA ligation reactions (required for chimera assembly). For this protocol, users clone wildtype genes into the same plasmid (pUC18 or pUC19). After the *in silico* selection of amino acid sequence regions where chimeras should be assembled, users obtain all the synonym DNA sequences that encode them. *Ad hoc* Perl scripts enable users to determine all synonym DNA sequences. After this step, another Perl script searches for restriction enzyme sites on all synonym DNA sequences. This *in silico* analysis is also performed using the ampicillin resistance gene (*ampR*) found on pUC18/19 plasmids. Users design oligonucleotides inside synonym regions to disrupt wildtype and *ampR* genes by PCR. After obtaining and purifying complementary DNA fragments, restriction enzyme digestion is accomplished. Chimera assembly is achieved by ligating appropriate complementary DNA fragments. pUC18/19 vectors are selected for CAPRESI because they offer technical advantages, such as small size (2,686 base pairs), high copy number, advantageous sequencing reaction features, and commercial availability. The usage of restriction enzymes for chimera assembly eliminates the need for DNA polymerases yielding blunt-ended products. CAPRESI is a fast and low-cost method for fusing protein-coding genes.

Video Link

The video component of this article can be found at <https://www.jove.com/video/55526/>

Introduction

Chimeric gene assembly has been widely used in molecular biology to elucidate protein function and/or for biotechnological purposes. Different methods exist for fusing genes, such as overlapping PCR product amplification¹, plasmid recovery², homologous recombination³, CRISPR-Cas9 systems⁴, site-directed recombination⁵, and Gibson assembly⁶. Each of these offers different technical advantages; for example, the flexibility of overlapping PCR design, the *in vivo* selection of constructions during plasmid recovery, or the high efficiency of CRISPR-Cas9 and Gibson systems. On the other hand, some difficulties can arise while performing some of these methods; for example, the first two approaches rely on blunt-ended DNA fragments, and ligation of these types of products could be technically challenging compared to sticky-ended ligation. Site-directed recombination can leave traces of extra DNA sequences (scars) on the original, like in the Cre-*loxP* system⁵. CRISPR-Cas9 can sometimes modify other genome regions in addition to the target site⁴.

Here, we introduce chimera assembly by plasmid recovery and restriction enzyme site insertion (CAPRESI), a protocol for fusing protein-coding genes that combines the plasmid recovery method (PRM) with the insertion of restriction enzyme sites on synonym DNA sequences, enhancing ligation efficiency. To ensure amino acid sequence integrity, restriction enzyme sites are inserted on synonym DNA sequence stretches. Among the benefits of CAPRESI are that it can be performed using ordinary laboratory reagents/tools (e.g., enzymes, competent cells, solutions, and thermocycler) and that it can give quick results (when the appropriate enzymes are used). Relying on restriction enzyme sites that emerge from synonym DNA sequences can limit the selection of the exact fusion points inside the proteins of interest. In such cases, target genes should be fused using overlapping oligonucleotides, and restriction enzyme sites should be inserted onto the resistance gene of the vector.

CAPRESI consists of seven simple steps (**Figure 1**): 1) selection of the cloning vector, pUC18 or pUC19⁶; 2) *in silico* analysis of the wildtype sequences to be fused; 3) selection of breaking regions for chimera assembly and plasmid disruption; 4) *in silico* generation of synonym DNA sequences containing restriction enzyme sites; 5) independent cloning of the wildtype genes into the selected plasmid; 6) plasmid disruption

by PCR, followed by restriction enzyme digestion; and 7) plasmid recovery using DNA ligation and bacterial transformation. Chimeric genes produced by this technique should be verified with sequencing.

The pUC18/19 vectors offer technical advantages for cloning and chimera assembly, such as small size (*i.e.*, 2,686 base pairs), high copy number, advantageous sequencing reaction features, and commercial availability⁷. Here, an *Escherichia coli* host was used to assemble and handle the chimeras because bacterial cultures are cheap and grow fast. Given this, subsequent cloning of the fusion fragments into the final target plasmids will be needed (*e.g.*, expression vectors as pRK415 in bacteria or pCMV in mammalian cells).

CAPRESI was tested for fusing two primary sigma factor genes: *E. coli* *rpoD* and *Rhizobium etlisigA*. Primary sigma factors are RNA polymerase subunits responsible for transcription initiation, and they consist of four domains (*i.e.*, $\sigma 1$, $\sigma 2$, $\sigma 3$, and $\sigma 4$)⁸. The amino acid sequence length of proteins encoded by *rpoD* and *sigA* are 613 and 685, respectively. RpoD and SigA share 48% identity (98% coverage). These primary sigma factors were split into two complementary fragments between regions $\sigma 2$ and $\sigma 3$. Two chimeric genes were assembled according to this design: chimera 01 (RpoD $\sigma 1$ - $\sigma 2$ + SigA $\sigma 3$ - $\sigma 4$) and chimera 02 (SigA $\sigma 1$ - $\sigma 2$ + RpoD $\sigma 3$ - $\sigma 4$). DNA fusion products were verified by sequencing.

Protocol

1. CAPRESI Protocol

NOTE: **Figure 1** represents the overall CAPRESI protocol. This technique is based on an *in silico* design and the subsequent construction of the desired chimeras.

1. Selection of the cloning plasmid, pUC18 or pUC19.

1. Select the pUC vector that best fits technical demands.

NOTE: Both vectors have the same sequence, except for the orientation of the multiple cloning site. This is important for the design of oligonucleotides and the PCR plasmid disruption.

2. *In silico* analysis of the wildtype genes and pUC18/19 sequences.

1. Obtain the DNA sequences of the wildtype genes and save them in a FASTA format file (*e.g.*, genes.fas).

NOTE: The sequences of pUC18/19 vectors are contained in the pUC.fas file (**Figure 1**, step 1).

2. Determine which restriction enzymes do not cut the wildtype genes and pUC18/19 sequences by running the REsearch.pl script. Alternatively, perform a restriction enzyme site search with an online tool (**Figure 1**, step 2).

1. Type the following into a terminal:

```
perl REsearch.pl genes.fas
```

```
perl REsearch.pl pUC.fas
```

NOTE: These commands will create the following output files: genes_lin.fas, genes_re.fas, pUC_lin.fas, and pUC_re.fas.

2. Select the appropriate restriction enzymes for cloning the wildtype genes into the multiple cloning site of the chosen pUC18/19 vector by comparing the genes_re.fas and pUC.fas files. Choose the orientation of the insert relative to the ampicillin resistance gene (*ampR*) of the pUC18/19 vector.
3. Design forward and reverse oligonucleotides to amplify the coding region of the wildtype σ genes (external oligonucleotides). Include the proper restriction enzyme site at each 5' end of the oligonucleotides; this will define the insert orientation.
 1. Include a functional ribosome binding site in the forward oligonucleotides.
 2. Insert both wildtype genes in the same orientation inside the vector.

3. Selection of the breaking regions for chimera assembly and plasmid disruption.

1. Obtain the amino acid sequence of the wildtype and the *ampR* genes by running the translate.pl script (**Figure 1**, step 3).

1. Type the following into a terminal:

```
perl translate.pl genes.fas
```

```
perl translate.pl ampR.fas
```

NOTE: This script will create the following output files: genes_aa.fas and ampR_aa.fas.

2. Globally align the two wildtype amino acid sequences with an appropriate software (*e.g.*, MUSCLE)⁹.
3. Select the desired break regions (3-6 amino acids) for chimera assembly based on the sequence alignment. Save them into a file in FASTA format (*e.g.*, regions.fas).
4. Locate the DNA sequences that code for the selected amino acid stretches on both wildtype genes.

4. *In silico* generation of synonym DNA sequences containing restriction enzyme sites.

1. Obtain all the synonym DNA sequences that code for each amino acid sequence stretch selected as breaking regions (**Figure 1**, step 4); these sequences were stored in the regions.fas file.

1. Type the following into a terminal:

```
perl synonym.pl regions.fas
```

NOTE: This script will create the regions_syn.fas output file.

2. Search for restriction enzyme sites found on synonym DNA sequences.

1. Type the following into a terminal: `perl RESynonym.pl regions_syn.fas`

NOTE: This script will create the regions_syn-re.fas output file.

3. Choose one restriction enzyme that is shared between synonym sequences of both wildtype genes; this site will be used to assemble chimeras via restriction enzyme digestion. Verify that the chosen restriction enzyme cut neither the wildtype genes nor the pUC18/19 vector sequences.
 1. Compare the restriction enzymes found on the genes_re.fas, pUC_re.fas, and regions_syn-re.fas files.
4. *In silico* substitute the originals for their synonym DNA sequences at the corresponding loci on both wildtype genes. Append synonym DNA sequences into the wildtype sequence file (genes.fas).
5. Get the amino acid sequence of genes contained on the genes.fas file (*perl translate.pl genes.fas*).
6. Align the amino acid sequences translated from wildtype and synonym DNA sequences. Verify that both sequences are the same.
7. Repeat all the steps of this section with the *ampR* gene present on the pUC18/19 vector.
NOTE: The goal is to disrupt the *ampR* gene (file ampR.fas) into two complementary parts. The restriction enzyme selected to disrupt the *ampR* gene should be different from the one used for chimera assembly.
5. **Independent cloning of the two wildtype genes into the selected pUC18/19 vector (Figure 1, step 3).**
 1. Purify the chosen pUC18/19 plasmid DNA using a plasmid DNA purification kit.
 1. For example, transform *E. coli* DH5 α with pUC18 plasmid and grow the transformants overnight at 37 °C on solid LB-0.3 mg/mL ampicillin (Amp). Pick a transformant colony, inoculate a new LB-0.3 mg/mL Amp plate, and grow it overnight at 37 °C.
 2. Take part of the previous culture and grow it in liquid LB-0.3 mg/mL Amp for 6-8 h at 37 °C. Extract the plasmid DNA using a kit.
 2. PCR amplify wildtype genes from the total genomic DNA using the appropriate external oligonucleotides. Use a high-fidelity DNA polymerase if possible. Select the DNA polymerase that maximizes the yield. Perform PCR according to the manufacturer's guidelines and the melting temperature of the oligonucleotides.
 1. Run PCR cycles, depending upon the DNA polymerase, amplicon size, template, and oligonucleotides used. For example, to amplify *rpoD* DNA, prepare a 50 μ L reaction: 5 μ L of 10x buffer, 2 μ L of 50 mM MgSO $_4$, 1 μ L of 10 mM dNTP mix, 2 μ L of each 10- μ M oligonucleotide (Table 2), 2 μ L of template DNA (*E. coli* DH5 α total DNA), 35.8 μ L of ultra-pure water, and 0.2 μ L of high-fidelity DNA polymerase. Run the following PCR cycles: 1 min at 94 °C for the initial denaturation followed by 30 cycles of amplification (30 s at 94 °C to denature, 30 s at 60 °C to anneal the oligonucleotides, and 2 min at 68 °C to extend).
 2. Dissolve 1.2 g of agarose in 100 mL of double-distilled water by heating them in the microwave. Assemble a gel tray and comb and put them into the horizontal gel caster. Fill the gel tray with the melted agarose solution and let it solidify. Remove the comb.
 3. Place the gel into the electrophoresis chamber. Fill chamber with 1x Tris-acetate-EDTA buffer. Load 50 μ L of PCR products into the gel wells. Electrophorese the samples (e.g., at 110 V for 1 h).
 4. After electrophoresis, stain the gel in 100 mL of double-distilled water containing 100 μ L of 1 mg/mL ethidium bromide solution for 10 min with slow shaking. Wash the gel in double-distilled water for another 10 min in slow shaking; use a horizontal shaker. Caution: Ethidium bromide is a toxic compound; wear protective gloves and a cotton laboratory coat.
 5. Purify wildtype gene DNA from the corresponding bands of the gel using a kit. Visualize the bands by placing the stained gel in a UV chamber (recommended wavelength: 300 nm). Keep the exposure of the gel to a minimum. Use a DNA purification kit and follow the manufacturer's instructions.
Caution: UV radiation is dangerous; wear a protective shield, glasses, and a cotton laboratory coat.
 3. Digest purified wildtype genes and pUC18/19 plasmid DNA with the restriction enzymes according to the manufacturer's instructions. Use fast digestion enzymes when possible. For example, digest 6 μ L of pUC18 DNA (300 ng/ μ L) in 10 μ L of ultrapure water, 2 μ L of 10X buffer, 1 μ L of *KpnI* restriction enzyme, and 1 μ L of *XbaI* restriction enzyme. Leave the reaction at 37 °C for 30 min.
 1. Inactivate the restriction enzymes according to the manufacturer's guidelines. For example, inactivate the double-digestion reaction *KpnI-XbaI* at 80 °C for 5 min.
 4. Mix insert:vector DNA at a volumetric ratio of 3:1. Follow the manufacturer's instructions for a ligation reaction. Use fast ligation enzymes when possible (leave the reaction at 25 °C for 10 min).
NOTE: Typically, the DNA concentration after purification using the kits is good enough for digestion and ligation reactions. For example, add 6 μ L of *rpoD* DNA (*KpnI-XbaI*, 150 ng/ μ L), 3 μ L of pUC18 DNA (*KpnI-XbaI*, 150 ng/ μ L), 10 μ L of 2x buffer, 1 μ L of ultrapure water, and 1 μ L of T4 DNA ligase. Leave ligation reaction at 25 °C for 10 min.
 5. Transform the *E. coli* DH5 α with 2-4 μ L of the ligation reaction, as described in the Supplementary Materials. Plate the transformed cells into solid LB/Amp/X-gal/IPTG plates. Leave the plates overnight at 37 °C.
 6. Select white-colored transformant *E. coli* colonies and streak them on a new LB/Amp plate. Leave the plates overnight at 37 °C.
 7. Perform a colony PCR to choose candidates for sequencing the reaction, as described in the Supplementary Materials. Extract plasmid DNA from candidates. Alternatively, digest candidate plasmid DNA with the same restriction enzymes used for cloning the inserts.
 8. Sequence candidate constructions with a sequencing service provider¹⁰. Assemble the sequence *in silico* using a DNA assembler¹¹ (Figure 1, step 5).
6. **Plasmid disruption by PCR followed by restriction enzyme digestion.**
 1. Design *in silico* forward and reverse oligonucleotides at break regions for wildtype and *ampR* genes, respectively. Substitute *in silico* the wildtype fragment for its corresponding synonym DNA sequence (obtained in step 1.4).
NOTE: This synonym DNA sequence contains a restriction enzyme site. Forward and reverse oligonucleotides overlap at the restriction enzyme site. Oligonucleotides should range from 21-27 nucleotides, end with a cytosine or guanine, and contain a restriction enzyme site. A forward oligonucleotide has the same sequence as its target region on the template DNA. A reverse oligonucleotide is the reverse complementary sequence of the target region on the template DNA.
 2. Use the two wildtype gene constructions as the DNA template for PCR reactions (e.g., pUC18*rpoD* and pUC18*sigA*). Obtain two complementary parts of each construction (pUC18*rpoD* σ 1- σ 2, pUC18*rpoD* σ 3- σ 4, pUC18*sigA* σ 1- σ 2, and pUC18*sigA* σ 3- σ 4) (Figure 1, step 6).
 3. Load DNA samples into an agarose gel 1-1.2% [w/v]. Separate the PCR products from the DNA template by electrophoresis (e.g., 110 V for 1 h). Alternatively, digest the PCR reactions with *DpnI* to break the DNA template.
NOTE: *DpnI* enzyme recognizes only methylated DNA sequences (5'-GATC-3').

1. Purify the samples using a DNA purification kit. Follow the manufacturer's guidelines.
4. Double-digest all the complementary DNA fragments with the appropriate restriction enzymes according to the manufacturer's directions. Use fast-digest restriction enzymes when possible. For example, double-digest the pUC18 ρ D σ 1- σ 2 DNA fragment with *Afl*II and *Spe*I using the manufacturer's protocol. Leave the digestion reaction at 37 °C for 15 min.
5. Inactivate the restriction enzymes according to the manufacturer's guidelines. For example, inactivate *Afl*II-*Spe*I at 80 °C for 20 min.
7. **Plasmid recovery using DNA ligation and bacterial transformation.**
 1. Mix the proper DNA fragments in a volumetric 1:1 ratio to assemble the desired chimeric gene (**Figure 1**, step 7).
NOTE: In this way, the integrity of the *ampR* gene is restored, producing a functional protein.
 1. For example, mix 4 μ L of pUC18 ρ D σ 1- σ 2 DNA (digested with *Afl*II-*Spe*I, 150 ng/ μ L), 4 μ L of pUC18 ρ sigA σ 3- σ 4 DNA (*Afl*II-*Spe*I, 150 ng/ μ L), 10 μ L of 2x buffer, 2 μ L of ultrapure water, and 1 μ L of T4 DNA ligase; the DNA concentration after kit purification is good enough for digestion and subsequent ligation. Leave the ligation reaction at 25 °C for 10 min.
 2. Transform *E. coli* DH5 α with 3-5 μ L of the chimera assembly ligation reaction (Supplementary Materials). Grow the transformant cells in LB/Amp/X-gal/IPTG plates overnight at 37 °C overnight.
 3. Select white-colored transformant *E. coli* colonies and streak them on a new LB/Amp plate. Leave the plates overnight at 37 °C.
 4. Perform a colony PCR to choose candidates for a sequencing reaction, as described in the Supplementary Materials. Visualize bands in a 1-1.2% (w/v) agarose gel by performing electrophoresis (described in step 2.3.2.1).
 5. Grow cultures from positive candidates. Extract plasmid DNA from them using a plasmid DNA purification kit.

2. Sequence Candidate Chimeric Constructions

1. Make sure the quality of the plasmid DNA is good enough for the sequencing reaction by following guidelines from the sequence service provider. Extract DNA using purification kits. For example, on the internet page of the sequence service provider, pay special attention to the recommended DNA concentration for the samples. Obtain the concentration of samples using a DNA quantification instrument.
2. Sequence candidate chimeric constructions with a sequencing service provider¹⁰.
3. Assemble sequencing reads with an *in silico* DNA assembler¹¹.
4. Align assembled versus *in silico*-designed chimeric sequences using sequence alignment tools^{9,12}. Verify that the chimeric gene was fused successfully.

3. Making Preparations

NOTE: All steps involving living cells should be performed in a clean laminar flow hood with the Bunsen burner on.

1. **Producing *E. coli* DH5 α -competent cells.**
 1. To produce *E. coli*-competent cells, follow published protocols¹³ or see the **Supplementary Materials**. Alternatively, use commercially available competent cells.
2. **Transforming *E. coli* DH5 α -competent cells.**
 1. For the chemical transformation of *E. coli* cultures, follow published protocols¹³ or see the **Supplementary Materials**. Alternatively, use electroporation for bacterial transformation. If commercially available competent cells are used, follow the manufacturer's guidelines.
3. **Purifying genomic and plasmid DNA from *E. coli* DH5 α .**
 1. Perform nucleic acid extraction, as stated in the **Supplementary Materials**, using commercially available kits or following published protocols¹³.
4. **Perform colony PCR.**
 1. Use this technique to identify candidate transformant colonies for subsequent sequencing reaction.
NOTE: This method does not represent a final verification of the integrity of the sequences. A detailed description of this technique is available in the **Supplementary Materials**.
5. **Preparing the solutions.**
 1. See **Table 1** for more information about solution preparation.
6. **Download the scripts and sequence files required for the CAPRESI protocol.**
 1. Download gene bank files containing the complete genome sequence of the desired species, extract the DNA sequence of the chosen gene using a genome browser, and save it in fasta file format. Download the Perl scripts required for the CAPRESI protocol. Store the scripts and the sequence files in the same directory.

Representative Results

Figure 1 depicts CAPRESI. Using this method, two chimeric genes were assembled by exchanging the domains of two bacterial primary sigma factors (*i.e.*, *E. coli* RpoD and *R. etli* SigA). The DNA sequences of the *rpoD* and *sigA* genes were obtained using the Artemis Genome Browser¹⁴ from GenBank genome files NC_000913 and NC_007761, respectively. The DNA sequence of the pUC18 vector was obtained from the nucleotide database of the NCBI server. *In silico* analyses of restriction enzyme sites were done on DNA sequences by running the REsearch.pl Perl script (see step 1.2.2). These results were used for oligonucleotide design (step 1.2.3). External oligonucleotides included recognition sites for *Xba*I and *Kpn*I restriction enzymes to clone wildtype genes into the pUC18 multiple cloning site. The forward oligonucleotide also included a ribosome binding site (**Table 2**). Genomic DNA was extracted from *E. coli* DH5 α cells grown in LB/Nal broth (37 °C, 220-250 rpm) and *R. etli* CFN42 cells grown in PY/Nal/CaCl₂ broth (30 °C, 220-250 rpm). These samples were used as DNA templates for wildtype gene amplification (step 1.5.2).

Wildtype genes were amplified using a *Taq* high-fidelity DNA polymerase, which was purified from agarose gels (DNA purification kit), double-digested with *Kpn*I-*Xba*I restriction enzymes, and cloned into the pUC18 vector. Chemically competent *E. coli* DH5 α cells were transformed with 5 μ L of the ligation reaction corresponding to wildtype constructions pUC18*rpoD* and pUC18*sigA* (step 1.5.5). Transformant cells were selected on solid LB/Amp/X-gal/IPTG plates grown at 37 °C. Wildtype constructions were verified by Sanger sequencing¹⁰. Sanger sequence reads were *in silico* assembled using MIRA¹¹ (step 1.5.8).

Amino acid sequences from wildtype genes were obtained by running the translate.pl script (step 1.3.1). After aligning wildtype genes with Clustal¹², the amino acid regions TLV and NLR were selected on the ampicillin resistance gene (*ampR*) and wildtype primary sigma factors, respectively (step 1.3.3). These amino acid regions were used for calculating all the possible DNA synonym sequences that encode them by running the synonym.pl script (step 1.4.1). The RESynonym.pl script searched for restriction enzyme sites found on the previously generated synonym DNA sequences (step 1.4.2). Those synonym regions that introduced the lowest number of changes as compared to the wildtype sequences were selected. For the primary sigma factor genes, wildtype sequences of RpoD (AACTTACGT) and SigA (AACCTTCGC) were exchanged for AACTTAAGG. The latter included an *Afl*II site (bold). For the *ampR* gene, the wildtype sequence ACGCTGGTG was exchanged for its synonym, A**CACTAGTG**. The synonym DNA sequence inserted into the *ampR* gene contained a *Spe*I site (bold). The *in silico* substitution of the wildtype for the corresponding synonym sequences were done to verify the sequence integrity and the oligonucleotide design (steps 1.2-1.4).

Synonym sequence changes were introduced by PCR using the appropriate oligonucleotides (**Table 2**) and wildtype constructions as DNA templates (step 1.5.2). Amplification reactions produced four complementary parts: pUC18*rpoD* σ 1- σ 2, pUC18*rpoD* σ 3- σ 4, pUC18*sigA* σ 1- σ 2, and pUC18*sigA* σ 3- σ 4. The complementary parts disrupted not only sigma factors, but also the *ampR* genes. Complementary parts were purified using a DNA purification kit (step 1.6.3).

These complementary DNA fragments were double-digested with *Afl*II and *Spe*I restriction enzymes (step 1.6.4). According to CAPRESI design, DNA fragments pUC18*rpoD* σ 1- σ 2 was ligated with pUC18*sigA* σ 3- σ 4 to obtain chimera 01, and pUC18*sigA* σ 1- σ 2 was ligated with pUC18*rpoD* σ 3- σ 4 to assemble chimera 02 (steps 1.7.1-1.7.3). Chemically competent *E. coli* DH5 α cells were transformed with 5 μ L of the ligation reaction of pUC18*chim01* and pUC18*chim02* (step 1.7.4). Transformant cells were selected on solid LB/Amp/X-gal/IPTG plates grown at 37 °C (step 1.7.5). Chimeric constructions were verified by Sanger sequencing¹⁰ (step 2). Sanger sequence reads were assembled using MIRA¹¹, and its resulting files are listed in **Table 3**. **Figure 2** shows the alignments (performed using MUSCLE)⁹ of assembled sequences of wildtype and chimeric genes at breaking regions.

Solution	Components		Instructions
	Quantity	Compound	
Luria-Bertani (LB) broth	5 g	yeast extract	Dissolve all components in water. Autoclave. Store at room temperature until use. For solid LB broth, add 15 g of bacteriological agar to the recipe.
	10 g	casein peptone	
	10 g	NaCl	
	up to 1 L	double distilled water	
LB/Amp/X-gal/IPTG plates	100 mL	melted solid LB broth	Add all components to temperate LB broth. Fill Petri dishes and wait until they solidify. Perform steps in a clean laminar flow hood with the Bunsen burner on.
	100 μ L	ampicillin (Amp) [100 mg/mL]	
	100 μ L	X-gal [20 mg/ml]	
	50 μ L	1 M IPTG solution	
peptone yeast (PY) broth	3 g	yeast extract	Dissolve all components in water. Autoclave. Store at room temperature until use. For solid PY broth, add 15 g of bacteriological agar to the recipe. For <i>R. etli</i> culture, add 700 μ L of 1 M CaCl_2 solution prior to use.
	5 g	casein peptone	
	up to 1 L	double distilled water	
PY/ CaCl_2 /NaI plates	100 mL	melted solid PY broth	Add all components to temperate PY broth. Fill the Petri dishes and wait until they solidify. Perform steps in a clean laminar flow hood with the Bunsen burner on.
	700 μ L	1 M CaCl_2 solution	
	100 μ L	nalidixic acid (NaI) [20 mg/mL]	
1 M calcium chloride (CaCl_2) solution	11.1 g	CaCl_2	Dissolve CaCl_2 in water. Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
Tris-EDTA (TE) 10:1 pH 8.0	1 mL	1 M Tris solution	Mix all components. Autoclave. Store at room temperature.
	400 μ L	0.25 M EDTA solution	
	up to 100 mL	double distilled water	
TE 50:20 pH 8.0	5 mL	1 M Tris solution	Mix all components. Autoclave. Store at room temperature.
	8 mL	0.25 M EDTA solution	
	up to 100 mL	double distilled water	
TE 10:1 10 mM NaCl solution	58.4 mg	NaCl	Dissolve NaCl in TE 10:1. Autoclave. Store at room temperature.
	up to 100 mL	TE 10:1 pH 8.0 solution	
lysis solution I	80 mg	lysozyme	Dissolve and mix all components in water. Filter sterilize using a 0.22 μ m filter. Store at room temperature.
	2 mL	0.5 M glucose solution	
	400 μ L	0.5 M EDTA solution	
	500 μ L	1 M Tris solution	
	up to 20 mL	sterilized double distilled water	
lysis solution II	1.2 mL	5 M sodium hydroxide (NaOH) solution	Dissolve all components in water. Filter sterilize using a 0.22 μ m filter. Store at room temperature.
	3 mL	10% sodium dodecyl sulfate (SDS) solution	
	up to 30 mL	sterilized double distilled water	
lysis solution III	29.4 g	potassium acetate (CH_3COOK)	Dissolve and mix all components in water. Filter sterilize using a 0.22 μ m filter. Store at room temperature.
	11.5 mL	absolute acetic acid (CH_3COOH)	
	up to 100 mL	sterilized double distilled water	
1 M magnesium chloride (MgCl_2) solution	9.5 g	MgCl_2	Dissolve MgCl_2 in water. Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
3 M sodium acetate (CH_3COONa) solution	24.6 g	CH_3COONa	

	up to 100 mL	double distilled water	Dissolve CH ₃ COONa in water. Autoclave. Store at room temperature
10% SDS solution	10 g	SDS	Dissolve SDS in water with a magnetic stirring bar at low velocity. Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
3 M CH ₃ COOK solution	29.4 g	CH ₃ COOK	Dissolve CH ₃ COOK in water. Filter sterilize using a 0.22 µm filter. Store at room temperature.
	up to 100 mL	sterilized double distilled water	
proteinase K solution	5 mg	proteinase K	Dissolve proteinase K in TE 50:20 solution. Heat at 37 °C for 1 h. Store at -20 °C.
	up to 1 mL	TE 50:20 pH 8.0 solution	
RNase solution	10 mg	RNase	Dissolve RNase in water. Heat at 95 °C for 10 min. Store at -20 °C.
	up to 1 mL	sterilized ultrapure water	
1 M isopropyl-β-D-thiogalactoside (IPTG) solution	238.3 mg	IPTG	Dissolve IPTG in water. Filter sterilize using a 0.22 µm filter. Store at -20 °C.
	up to 1 mL	sterilized ultrapure water	
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) solution	20 mg	X-gal	Dissolve X-gal in water. Filter sterilize using a 0.22 µm filter. Store at -20 °C.
	up to 1 mL	sterilized ultrapure water	
phenol:chloroform:isoamyl alcohol 24:24:1 solution	24 mL	phenol	Mix all components. Store in an amber capped bottle at 4 °C. CAUTION! Hazardous material. Wear protective glasses, gloves and cotton laboratory gown. Use an extraction hood. DO NOT turn on the Bunsen burner or any other source of fire during handling. Alternative: use commercial phenol:chloroform:isoamyl alcohol 25:24:1 solution
	24 mL	chloroform	
	1 mL	isoamyl alcohol	
0.5 M glucose solution	9 g	glucose	Dissolve glucose in water. Filter sterilize using a 0.22 µm filter. Store at room temperature.
	up to 100 mL	sterilized ultrapure water	
0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8.0 solution	14.6 g	EDTA	Dissolve EDTA in water. Adjust pH with 5 M NaOH solution. Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
0.25 M EDTA pH 8.0 solution	7.3 g	EDTA	Dissolve EDTA in water. Adjust pH with 5 M NaOH solution. Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
1 M Tris pH 8.0 solution	12.1 g	Tris	Dissolve Tris in water. Adjust pH with absolute hydrochloric acid (HCl). Autoclave. Store at room temperature.
	up to 100 mL	double distilled water	
5 M sodium hydroxide (NaOH) solution	19.9 g	NaOH	Dissolve NaOH in water. Store at room temperature. CAUTION! Caustic compound.
	up to 100 mL	sterilized double distilled water	
0.1 M NaOH solution	399 mg	NaOH	Dissolve NaOH in water. Store at room temperature.
	up to 100 mL	sterilized double distilled water	
nalidixic acid (Nal) stock solution	60 mg	nalidixic acid	Dissolve Nal in water. Filter sterilize using a 0.22 µm filter. Store at 4 °C.
	up to 3 mL	0.1 M NaOH solution	
ampicillin (Amp) stock solution	300 mg	ampicillin	Dissolve Amp in water. Filter sterilize using a 0.22 µm filter. Store at 4 °C.
	up to 3 mL	sterilized double distilled water	

10x tris-acetate-EDTA buffer	48.4 g	Tris	Dissolve all components in water. Autoclave. Store at room temperature.
	20 ml	0.5 M EDTA solution	
	11.44 ml	glacial acetic acid	
	up to 1 L	double distilled water	

Table 1: Solution preparation. Instructions for solution preparation.

No.	Code	Sequence	Length (nt)	Features
1	RpoD-FWD	<u>GCTCTAGAGA</u> AAGGAGAT TATCATATGGAGCAAACCCGCAGTCA	43	<i>Xba</i> I site; wild type gene amplification and vector cloning
2	RoD-REV	GGGGT <u>ACCTTAATCGTCCAGGAAGCTACG</u>	29	<i>Kpn</i> I site; wild type gene amplification and vector cloning
3	SigA-FWD	<u>GCTCTAGAGA</u> AAGGAGAT TATCATATGGCAACCAAGGTCAAAGAG	43	<i>Xba</i> I site; wild type gene amplification and vector cloning
4	SigA-REV	GGGGT <u>ACCTTAGCTGTCCAGAAAGCTTCT</u>	29	<i>Kpn</i> I site; wild type gene amplification and vector cloning
3	AmpR-FWD	AC <u>ACTAGT</u> GAAAGTAAAAGAT	21	<i>Spe</i> I site inserted, synonym mutation. Disruption of <i>ampR</i> gene
4	AmpR-REV	T <u>ACTAGT</u> GTTTCTGGGTGAG	21	<i>Spe</i> I site inserted, synonym mutation. Disruption of <i>ampR</i> gene
5	σ 2RpoD-FWD	AA <u>CTTAAGGCTGGT</u> TATTTCTATCGCT	27	<i>Afl</i> III site inserted, synonym mutation. Construction of chim01-02
6	σ 2RpoD-REV	GC <u>CTTAAGTTCGCTTCA</u> ACCATCTCTT	27	<i>Afl</i> III site inserted, synonym mutation. Construction of chim01-02
7	σ 2SigA-FWD	AA <u>CTTAAGGCTCGT</u> CATTTCAATCGCC	27	<i>Afl</i> III site inserted, synonym mutation. Construction of chim01-02
8	σ 2SigA-REV	GC <u>CTTAAGTTCGCTT</u> CGACCATTTCCT	27	<i>Afl</i> III site inserted, synonym mutation. Construction of chim01-02

Table 2: Oligonucleotide sequences. Restriction enzyme sites appear underlined. Ribosome binding site: bold.

Construction	Sequence file	Quality file
chimera01	chim01_out.unpadded.fasta	chim01_out.unpadded.fasta.qual
		chim01_A3/B3/E2/F2/G2/H2.pdf
chimera02	chim02_out.unpadded.fasta	chim02_out.unpadded.fasta.qual
		chim02_C3/D3/E3/F3/G3/H3.pdf
rpoD	rpod_out.unpadded.fasta	rpod_out.unpadded.fasta.qual
sigA	sigA_out.unpadded.fasta	sigA_out.unpadded.fasta.qual

Table 3: Sequence files obtained with the MIRA assembler and DNA sequencer. Sanger sequencing reads from chimeric and wildtype constructions were assembled with MIRA¹¹ software using the mapping option. Chromatograms of chimera sequencing appear as PDF files.

Discussion

CAPRESI was designed as an alternative to the PRM². The original PRM is a powerful technique; it allows for the fusion of DNA sequences along any part of the selected genes. For PRM, wildtype genes should be cloned into the same plasmid. After that, oligonucleotides are designed inside wildtype and antibiotics resistance genes found on the plasmid. Plasmid disruption is achieved by PCR using blunt-ended, high-fidelity DNA polymerases and previously designed oligonucleotides. The ligation of complementary PCR products assembles chimeric genes and restores the antibiotics resistance cassette, resulting in functional plasmid recovery. PRM enables the construction of chimeric gene libraries on the same plasmid background. Unfortunately, the ligation of blunt-ended DNA sequences could be technically arduous. Chimera assembly by overlapping PCR fragments¹ also requires that DNA polymerases yield blunt-ended products and DNA purification for each reaction. The chimeric genes assembled by this technique are linear DNA products. The cloning of each construction into the target vector could delay further analysis.

CAPRESI benefits from many of the strengths of PRM and also simplifies the ligation of complementary DNA fragments by using restriction enzyme sites on synonym DNA sequence regions. For these reasons, CAPRESI does not require blunt-ended DNA polymerases. The use of synonym DNA sequences restrains fusion sites for chimera assembly but enhances ligation efficiency. Another important modification introduced in CAPRESI is that chimeras are assembled and handled in pUC18/19 vectors. These vectors possess several advantageous features, as stated previously. Chimeric DNA could be obtained by propagating the construction vector in *E. coli* hosts. Subsequent cloning of chimeric genes into the final plasmid (e.g., an expression vector) is sometimes required.

CAPRESI also relies on the *in silico* handling of DNA and amino acid sequences. *Ad hoc* Perl scripts enable the selection of the appropriate restriction enzymes to clone wildtype genes, the calculation of all synonym DNA sequences corresponding to break regions (for chimera assembly), and the identification of restriction enzymes for simplifying plasmid recovery. Given these conditions, CAPRESI is an alternative for fusing protein-coding genes. Critical steps of the CAPRESI protocol are: 1) the selection of breaking regions and 2) the purification of PCR products. Breaking regions are stretches where synonym sequences are calculated, producing restriction enzyme sites that are not found on wildtype sequences. The use of restriction enzyme sites for chimera assembly eliminates the need for DNA polymerases that yield blunt-ended products and eases the ligation reaction. Not all regions will have usable restriction enzyme sites; for this reason, it is recommended to move breaking regions by one amino acid at a time until the best sequence stretch is found. If the *in silico* design does not allow for the movement of the breaking regions or the sequence stretches lack synonym DNA sequences with suitable restriction enzyme sites, it is recommended to fuse target genes using overlapping oligonucleotides and to introduce synonym DNA sequences into the ampicillin resistance gene (found on the vector) only. In this way, the genes can be fused at any part and the plasmid recovery relies on the ligation of one unique site (digested with only one restriction enzyme). The flexibility of CAPRESI allows for it to be performed in combination with other techniques.

The purification of PCR products is performed to eliminate template DNA, decreasing the chances of parental plasmid contamination after the ligation of complementary DNA fragments. Fulfilling these two critical steps enhances bacterial transformation efficiency (the number of correctly assembled constructions), allowing CAPRESI to be performed with either chemically competent (CaCl₂) or electrocompetent *E. coli* cells. The first type of competent cells represents the cheapest source of *E. coli* cultures, employed for bacterial transformation in most laboratories. Because of all the features shown previously, CAPRESI represents a useful option for assembling chimeras.

Moreover, CAPRESI could work in combination with overlapping PCR fragments or plasmid recovery techniques. For example, instead of inserting two synonym sequences per complementary DNA portion, the desired breaking region (on the target wildtype genes) could use overlapping oligonucleotides to fuse intervening sequences. The introduction of synonym sequences could be restricted to the *ampR* gene of pUC18/19 vectors, leading to the use of only one restriction enzyme to restore plasmid integrity. These future applications could strengthen CAPRESI by allowing for the fusion of any type of DNA sequences (i.e., not only protein-coding genes), without compromising ligation efficiency.

In order to test CAPRESI, two chimeric sigma factors were assembled by exchanging regions of two wildtype primary sigma factor genes: *rpoD* (*E. coli*) and *sigA* (*R. etli*). All known primary sigma factors hold four domains, σ_1 to σ_4 ⁸. Chimera 01 consists of RpoD σ_1 - σ_2 and SigA σ_3 - σ_4 , while chimera 02 has SigA σ_1 - σ_2 and RpoD σ_3 - σ_4 . The integrity of the chimeric constructions was verified by Sanger sequencing¹⁰ (Figure 2).

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

This work was supported by Consejo Nacional de Ciencia y Tecnología, CONACYT, México (grant number 154833) and Universidad Nacional Autónoma de México. The authors wish to thank Víctor González, Rosa I. Santamaría, Patricia Bustos, and Soledad Juárez for their administrative and technical advice.

References

1. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., & Pease, L.R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*. **77** (1), 61-68 (1989).
2. Vos, M.J., & Kampinga, H.H. A PCR amplification strategy for unrestricted generation of chimeric genes. *Anal. Biochem.* **380** (2), 338-340 (2008).
3. Hawkins, N.C., Garriga, G., & Beh, C.T. Creating Precise GFP Fusions in Plasmids Using Yeast Homologous Recombination. *Biotechniques*. **34** (1), 1-5 (2003).
4. Sander, J.D., & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol.* **32** (4), 347-355 (2014).
5. Nagy, A. Cre recombinase: The universal reagent for genome tailoring. *Genesis*. **26** (2), 99-109 (2000).
6. Gibson, D. G., *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. **6** (5), 343 - 345 (2009).
7. Norrander, J., Kempe, T., & Messing, J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*. **26** (1), 101-106 (1983).
8. Gruber, T. M., & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol.* **57**, 441-466 (2003).
9. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32** (5), 1792-1797 (2004).
10. *Macrogen Inc.*, Seoul, Rep. Of Korea: <http://dna.macrogen.com/eng/index.jsp> (2016).
11. Chevreux, B. *MIRA: an automated genome and EST assembler*. 1-161, <http://www.chevreux.org/thesis/index.html> (2005).
12. Thompson, J. D., Gibson, T. J., Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. **Chapter 2** (Unit 2.3), 1-22 (2002).
13. Sambrook, J. and Russell, D. W. *Molecular Cloning: A laboratory manual*. CSHLP. (2001).
14. Rutherford, K., *et al.* Artemis: sequence visualization and annotation. *Bioinformatics*,. **16** (10), 944-945 (2000).