High-throughput cloning and expression in recalcitrant bacteria

Geertsma, Eric R.; Poolman, Berend

Published in:
Nature Methods

DOI:
10.1038/NMETH1073

Publication date:
2007

High-throughput cloning and expression in recalcitrant bacteria

Eric R Geertsma & Bert Poolman

Supplementary text and figures:

Supplementary Figure 1 Frequency of *Sfi*I sites yielding identical 3′extensions in pro- and eukaryotic genomes.

Supplementary Table 1 The distribution of *Sfi*I sites over pro- and eukaryotic gene transcripts.

Supplementary Table 2 Basic and additional LIC vectors constructed.

Supplementary Methods

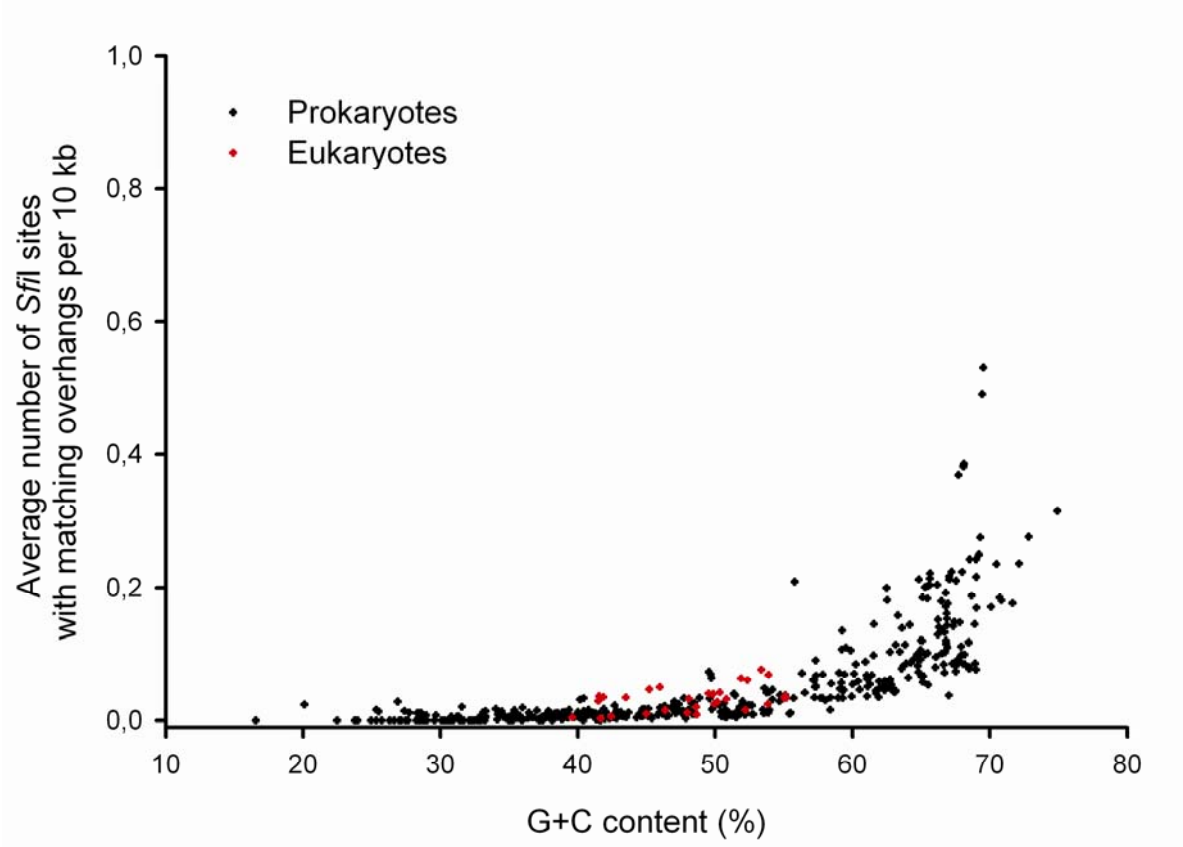
Supplementary Note Detailed overview of the LIC process.

Supplementary Data is available on the Nature Methods webiste.

A step-by-step protocol is available in the Protocols Network on the Nature Protocols website.

Supplementary Fig. 1

Frequency of *Sfi*I sites yielding identical 3'extensions in pro- and eukaryotic genomes.



Analysis of the occurrence of *Sfi*I sites yielding identical overhangs in genomes and combined transcripts as a function of the G+C content of the DNA. For each DNA, the datapoint of the most occurring *Sfi*I site of this type is shown. The full dataset is available in **Supplementary Data**.

Supplementary Table 1

The distribution of *Sfi*I sites over pro- and eukaryotic gene transcripts.

***Sfi*I sites in transcript**	**Percentage of transcripts**	
	prokaryotes	**eukaryotes**
0	**92.18**	**92.06**
1	**6.42**	**6.49**
2	**1.09**	**1.04**
3	**0.22**	**0.23**
4	**0.06**	**0.08**
5	**0.02**	**0.03**
6	**0.01**	**0.02**
7	**0.00**	**0.01**
8	**0.00**	**0.01**
9	**0.00**	**0.01**
>9	**0.00**	**0.02**

For the pro- and eukaryotic datapoints, 1.484.533 and 745.558 gene transcripts were analyzed, respectively.

Supplementary Table 2

Basic and additional LIC vectors constructed.

Vector name	Protein sequence	Protein sequence after TEV protease cleavage	Expression host
pREnLIC	M-His_{10}-G-TEV site-protein	G-protein	*L. lactis* NZ9000
pREcLIC	MGGGFA-protein-TEV site-His_{10}	MGGGFA-protein-ENLYFQ	*L. lactis* NZ9000
pREcLIC-GFP	MGGGFA-protein-TEV site-GFP-His_{10}	MGGGFA-protein-ENLYFQ	*L. lactis* NZ9000
pRE-USP45-nLIC	M-ssUSP45[1)]-His_{10}-G-TEV site-protein	G-protein	*L. lactis* NZ9000
pBADnLIC	M-His_{10}-G-TEV site-protein	G-protein	*E. coli*
pBADcLIC	MGGGFA-protein-TEV site-His_{10}	MGGGFA-protein-ENLYFQ	*E. coli*
pBADcLIC-GFP	MGGGFA-protein-TEV site-GFP-His_{10}	MGGGFA-protein-ENLYFQ	*E. coli*
pBAD-OmpA-nLIC	M-ssOmpA[2)]-His_{10}-G-TEV site-protein	G-protein	*E. coli*

[1)] ssUSP45 indicates the signal sequence of the *L. lactis* USP45 protein. The pRE-USP45-nLIC vector is used if the N-terminus of the membrane protein of interest is predicted to be on the outside of the cytoplasmic membrane, or to replace the signal sequence of the protein of interest with the ssUSP45.

[2)] ssOmpA indicates the signal sequence of the *E. coli* OmpA protein. The pBAD-OmpA-nLIC vector is used if the N-terminus of the membrane protein of interest is predicted to be on the outside of the cytoplasmic membrane, or to replace the signal sequence of the protein of interest with the ssOmpA.

Supplementary Methods

Ligation-Independent Cloning (LIC). We amplified inserts using Phusion DNA polymerase (Finnzymes) and gene specific primers extended at the 5' side with LIC specific tails (Table A). We purified vectors containing a LIC cassette using a plasmid isolation kit (Wizard ® Plus, Promega). DNA was additionally purified by extraction with phenol:chloroform and chloroform to remove tracer amounts of proteins, followed by ethanol precipitation. Vectors (~5 μg of DNA) were digested overnight with 25 units of *Swa*I (Roche) at 25°C. We gel-purified PCR products and digested vectors using a gel purification kit (GFX PCR DNA & Gel Band Purification Kit, GE), and eluted the material in 10 mM Tris-HCl, pH 7.5, 0.2 mM Na-EDTA. Purified material was stored at 4°C.

For a typical reaction, we used 200 ng of *Swa*I-digested vector or equimolar quantities of inserts, and adjusted the volume to 10 μl with milliQ. Next, we added 3 μl 5X buffer (250 mM Tris-HCl, 75 mM $(NH_4)SO_4$, 35 mM $MgCl_2$, 0.5 mM EDTA, 50 mM 2-mercapto-ethanol, 0.1 mg/ml BSA, pH 8.8), 1.5 μl 25 mM dCTP (for the vector) or 1.5 μl 25 mM dGTP (for the insert), and 0.5 μl (1U/μl) T4 DNA polymerase (Roche). The sample was incubated at 20°C for 30 min. Subsequently, we heat inactivated the T4 DNA polymerase (20 min at 75°C). The material, now in a "LIC-ready" state, can be stored at 4°C for prolonged periods (over 6 months). We mixed LIC-ready vector (1 μl) and insert (3 μl) and after a 5 min incubation at RT transformed the material to 75 μl chemically-competent *E. coli* MC1061. Cells were plated on Luria Broth supplemented with ampicillin (100 μg / ml).

Table A. Primer extensions required for Ligation Independent Cloning.

Type of primer	Sequence (5'→3')
nLIC forward	AT GGT GAG AAT TTA TAT TTT CAA GGT + gene specific (no start)
nLIC reverse	T GGG AGG GTG GGA TTT TCA TTA + gene specific (no stop)
cLIC forward	ATG GGT GGT GGA TTT GCT + gene specific (no start)
cLIC reverse	TTG GAA GTA TAA ATT TTC + gene specific (no stop)

Vector Backbone Exchange. Prior to the VBEx procedure, we purified pRExLIC-derived vectors and the pERL vector using a plasmid isolation kit (Wizard ® Plus, Promega). Furthermore, we additionally purified plasmid pERL (but not the pRExLIC-derived vectors) by extraction with phenol:chloroform and chloroform to remove tracer amounts of proteins, followed

by ethanol precipitation. We performed the exchange of the vector backbone of pRExLIC-derived vectors in a small volume (10 μl) in a PCR machine with heated lid to avoid condensation. We mixed ~125 ng of the pERL vector (containing the *L. lactis* origin of replication) and ~125 ng of a pRExLIC-derived vector (containing the *L. lactis cat* gene) and adjusted the volume to 10 μl by adding 1 μl 10 X buffer (100 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 500 mM NaCl, 1 mg/ml BSA), 5 U *Sfi*I (Fermentas) and sufficient milliQ. The sample was incubated for 80 min at 50°C, and 20 min at 80°C to inactivate *Sfi*I. After cooling to RT, we started ligation by the addition of 1.5 μl 8 mM Na_2-ATP, pH 7, and 0.5 U T4 DNA ligase (Roche). The sample was incubated for 1 hr at 20°C and 20 min at 65°C to heat inactivate the T4 DNA ligase. Subsequently, we transformed 2 μl of the sample to 30 μl electrocompetent *L. lactis* NZ9000 (see below) and plated aliquots on M17 plates[1] (Difco) supplemented with 0.5% glucose, 0.5 M sucrose, 5 μg/ml chloramphenicol. Parafilm-sealed plates were incubated at 30°C until colonies appeared (~18 hrs).

Electrotransformation of *L. lactis*. Preparation of electrocompetent *L. lactis* NZ9000 was essentially done as described[2,3], but with some critical modifications. Briefly, we grew cells in M17 supplemented with 0.5% glucose, 0.5 M sucrose and 2% glycine at 30°C to OD_{600} = ~0.5. Cells were harvested by centrifugation at 5000 x *g* for 15 min at 4°C. Following washes with 1 volume ice-cold solution A (0.5 M sucrose and 10% glycerol, prepared in milliQ), 0.5 volume solution A supplemented with 50 mM Na-EDTA, pH 7.5, and 0.25 volume solution A, we resuspended cells in 0.01 volume solution A. Aliquots of 40 μl were flash-frozen in liquid nitrogen and stored at -80°C until use. For electroporation, cells were thawed on ice, combined with plasmid DNA, and transferred to an ice-cooled electroporation cuvet (2 mm gap). We exposed cells to a single electrical pulse with a field strength of 2 kV, 25 μF capacitance and 200Ω resistance. Immediately following discharge, we mixed the cells with 1 ml ice-cold M17 supplemented with 0.5% glucose, 0.5 M sucrose, 20 mM $MgCl_2$ and 2 mM $CaCl_2$, and left them on ice for 10 min. Subsequently, cells were incubated at 30°C for 2 hrs and we plated aliquots on M17 agar supplemented with 0.5% glucose, 0.5 M sucrose and 5 μg/ml chloramphenicol. Plates were sealed and incubated overnight at 30°C.

1. B. E. Terzaghi and W. E. Sandine, *Appl Microbiol* **29** (6), 807 (1975).
2. H. Holo and I. F. Nes, *Appl Environ Microbiol* **55** (12), 3119 (1989).
3. J. M. Wells, P. W. Wilson, and R. W. Le Page, *J Appl Bacteriol* **74** (6), 629 (1993).

Equipment and settings. Immunoblots were visualized with a Fujifilm LAS-3000 imaging system and analyzed using AIDA software (Raytest; Isotopenmessgeräte, GmbH).

Supplementary Note

Detailed overview of the LIC process.

nLIC cassette: A construct is made that contains an N-terminal 10 His-tag, followed by a TEV protease cleavage site and Your Favorite Protein (YFP). Using the 3'→5' exonuclease activity of T4 DNA polymerase and dedicated tail-sequences, long defined overhangs are generated. These overhangs have a sufficiently high annealing temperature that mere mixing of complementary overhangs suffices in generating stable DNA sequences, ready for cell transformation.

The vector holds the nLIC cassette which contains a *SwaI* site. After *SwaI* digestion, the vector is treated with T4 DNA polymerase in the presence of **dCTP**, in order to generate the *nLIC-ready* overhangs (illustrated below).

1. nLIC cassette

```
N  Met His His His His His His His His His His Gly Glu Asn Leu Tyr
5' ATG CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT GGT GAG AAT TTA TAT TTAAATCCCACCCTCCCAG
3' TAC GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA CCA CTC TTA AAT ATA AATTTAGGGTGGGAGGGTC
```

2. After digestion with *Swa*I

```
N  Met His His His His His His His His His His Gly Glu Asn Leu Tyr
5' ATG CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT GGT GAG AAT TTA TAT TT    AAATCCCACCCTCCCAG
3' TAC GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA CCA CTC TTA AAT ATA AA    TTTAGGGTGGGAGGGTC
```

3. After treatment with T4 DNA polymerase + **dCTP**

```
N  Met His His His His His His His His His His Gly Glu Asn Leu Tyr
5' ATG CAT CAT CAC CAT CAT CAC CAT CAC CAT C                           AAATCCCACCCTCCCAG
3' TAC GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA CCA CTC TTA AAT ATA AA                  C
```

The insert is PCRed using primers with dedicated nLIC tails. After removal of primers and nucleotides, the insert is treated with T4 DNA polymerase in the presence of **dGTP**, in order to generate the nLIC-ready overhangs (illustrated below).

1. PCR

```
N  is Gly Glu Asn Leu Tyr Phe Gln Gly Met      X   X
5' AT GGT GAG AAT TTA TAT TTT CAA GGT ATG YFP TAA TGA AAA TCC CAC CCT CCC A
3' TA CCA CTC TTA AAT ATA AAA GTT CCA TAC YFP ATT ACT TTT AGG GTG GGA GGG T
```

2. After treatment with T4 DNA polymerase + **dGTP**

```
N  is Gly Glu Asn Leu Tyr Phe Gln Gly      X   X
5' AT GGT GAG AAT TTA TAT TTT CAA GGT YFP TAA TG
3'                          GTT CCA YFP ATT ACT TTT AGG GTG GGA GGG T
```

Subsequently, the nLIC-ready vector and insert are mixed. The defined overhangs anneal to stable structures with a T_m of approximately 44°C and 58°C for the 5’ and 3’ end of the gene, respectively. Small, 1 basepair gaps remain, which will be filled *in vivo*.

a. 5’ end, before annealing

```
N  Met His His His His His His His His His H                          is Gly Glu Asn Leu Tyr Phe Gln Gly
5′ ATG CAT CAT CAC CAT CAT CAC CAT CAC CAT C                          AT GGT GAG AAT TTA TAT TTT CAA GGT YFP
3′ TAC GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA CCA CTC TTA AAT ATA AA.                            GTT CCA YFP
```

b. 5’ end, after annealing

```
N  Met His His His His His His His His His His Gly Glu Asn Leu Tyr Phe Gln Gly
5′ ATG CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT GGT GAG AAT TTA TAT TTT CAA GGT YFP
3′ TAC GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA CCA CTC TTA AAT ATA AA. GTT CCA YFP
```

1. 3’ end, before annealing

```
N  Met      X   X
5′ ATG YFP TAA TG                          AAA TCC CAC CCT CCC AG
3′ TAC YFP ATT ACT TTT AGG GTG GGA GGG T                       C
```

2. 3’ end, after annealing

```
N  Met      X   X
5′ ATG YFP TAA TG. AAA TCC CAC CCT CCC AG
3′ TAC YFP ATT ACT TTT AGG GTG GGA GGG TC
```

cLIC cassette: A construct is made that contains a relatively small N-terminal modification (MGGGFA), Your Favorite Protein (YFP), and a C-terminal TEV protease cleavage site followed by a 10 His-tag. Using the 3'→5' exonuclease activity of T4 DNA polymerase and dedicated tail-sequences, long defined overhangs are generated. These overhangs have a sufficiently high annealing temperature that mere mixing of complementary overhangs suffices in generating stable DNA sequences,ready for cell transformation.

The vector holds the cLIC cassette which contains a *SwaI* site. After *SwaI* digestion, the vector is treated with T4 DNA polymerase in the presence of **dCTP**, in order to generate the *nLIC-ready* overhangs (illustrated below).

1. cLIC cassette

```
     M   G   G   G   F     N   L   Y   F   Q   G   H   H   H   H   H   H   H   H   H   H   X
5′ C ATG GGT GGT GGA TTT A AAT TTA TAC TTC CAA GGT CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT TAA
3′ G TAC CCA CCA CCT AAA T TTA AAT ATG AAG GTT CCA GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA ATT
```

2. After digestion with SwaI

```
     M   G   G   G   F         N   L   Y   F   Q   G   H   H   H   H   H   H   H   H   H   H   X
5′ C ATG GGT GGT GGA TTT       A AAT TTA TAC TTC CAA GGT CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT TAA
3′ G TAC CCA CCA CCT AAA       T TTA AAT ATG AAG GTT CCA GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA ATT
```

3. After treatment with T4 DNA polymerase + **dCTP**

```
     M   G   G   G   F         N   L   Y   F   Q   G   H   H   H   H   H   H   H   H   H   H   X
5′ C                         A AAT TTA TAC TTC CAA GGT CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT TAA
3′ G TAC CCA CCA CCT AAA                           CCA GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA ATT
```

The insert is PCRed using primers with dedicated cLIC tails. After removal of primers and nucleotides, the insert is treated with T4 DNA polymerase in the presence of **dGTP**, in order to generate the cLIC-ready overhangs (illustrated below).

1. PCR

```
   M   G   G   G   F   A       E   N   L   Y   F   Q
5′ ATG GGT GGT GGA TTT GCT YFP GAA AAT TTA TAC TTC CAA
3′ TAC CCA CCA CCT AAA CGA YFP CTT TTA AAT ATG AAG GTT
```

2. After treatment with T4 DNA polymerase + **dGTP**

```
   M   G   G   G   F   A       E   N   L   Y   F   Q
5′ ATG GGT GGT GGA TTT GCT YFP G
3′                       GA YFP CTT TTA AAT ATG AAG GTT
```

Subsequently, the cLIC-ready vector and insert are mixed. The defined overhangs anneal to stable structures with a T_m of approximately 41°C and 31°C for the 5’ and 3’ end of the gene, respectively. Small, 1 basepair gaps remain, which will be filled *in vivo*.

a. 5’ end, before annealing

```
      M   G   G   G   F  A
5′ C                          ATG GGT GGT GGA TTT GCT YFP
3′ G TAC CCA CCA CCT AAA                        GA YFP
```

b. 5’ end, after annealing

```
      M   G   G   G   F  A
5′ C ATG GGT GGT GGA TTT GCT YFP
3′ G TAC CCA CCA CCT AAA .GA YFP
```

1. 3’ end, before annealing

```
       E   N   L   Y   F   Q         N   L   Y   F   Q   G   H   H   H   H   H   H   H   H   H   X
5′ YFP G                        A AAT TTA TAC TTC CAA GGT CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT TAA
3′ YFP CTT TTA AAT ATG AAG GTT                        CCA GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA ATT
```

2. 3’ end, after annealing

```
       E   N   L   Y   F   Q   G   H   H   H   H   H   H   H   H   H   H   X
5′ YFP G.A AAT TTA TAC TTC CAA GGT CAT CAT CAC CAT CAT CAC CAT CAC CAT CAT TAA
3′ YFP CTT TTA AAT ATG AAG GTT CCA GTA GTA GTG GTA GTA GTG GTA GTG GTA GTA ATT
```

Supplementary Data

Frequency and sequence of 3’ extensions of *Sfi*I sites in pro- and eukaryotic genomes.

suppl_data.xls