Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries.

L. Terrón-González, C. Medina, M. C. Limón-Mortés, E. Santero

SUPPLEMENTARY METHODS

Construction of metagenomic vectors.

We started from a previous modification of the pCC1FOS™ fosmid vector (Epicentre), which contained two *Ceu*I sites flanking the cloning site (*Eco*72I) for the metagenomic DNA (M. Ferrer, personal communication). The different elements were introduced into the fosmid as follows:

(i) *oriT*: A 314 bp fragment bearing *oriT* was amplified from the RP4 plasmid[1] using the OriTHpaIFw and OriTHpaIRev primers, which generate an *Hpa*I restriction site at each end of the amplified DNA fragment, and cloned into the unique *Hpa*I site of pCC1FOS-CeuI, thus generating pMPO561.

(ii) p*sal* promoter followed by the nut_L site (N-utilization leftward site) from the lambda phage: A 238 bp DNA fragment containing these two elements was synthesised by recursive PCR[2], using the primers psalnut1, psalnut2, psalnut3 y psalnut4 and cloned into the *Eco*RV site pBluescript II SK+ (Stratagene). The nut_L site was located 48 bp downstream of the p*sal* transcription initiation. The fragment of interest was obtained by digesting with *Xba*I plus *Hind*III, and cloned into the unique *Nar*I site of the pMPO561, in the appropriate orientation in relation to the metagenomic DNA cloning site, just upstream of the T7 gene 10 promoter already present in pCC1FOS, to yield pMPO571.

(iii) A promoterless *gfp* gene with the ribosome binding site from the T7 gene-10 next to the metagenomic DNA cloning site: This construction required several steps. First, part of the pCC1FOS-CeuI fosmid that contained the chloramphenicol resistance gene and part of *red*F was amplified using the primers HindChlFw and BstZredFRv (the resulting fragment is flanked by the *Hind*III and *Bst*1107I sites), and this fragment was cloned into the *EcoR*V site of pBluescript II SK+. Second, the *gfp* gene has an *Eco*72I site, which was necessary to eliminate. It was mutated to a *Mlu*I site using the primers KpnISDpT77GFP, GFPMluIFwsolap, GFPMluIRvsolap and GFPXbaI-TFB-PCRsolap and the pMPO634 plasmid[3] as template by overlapping PCR[4]. Base substitutions did not change the aminoacid sequence of the Gfp. The 5' end of the primer KpnISDpT77GFP contained the ribosome binding site from the T7 gene-10 sequence in such a way that it was placed during the PCR at the right distance upstream of the amplified *gfp* coding region. The product of the PCR was digested with *Kpn*I and *Xba*I (their restriction sites are at the ends of the fragment), its ends made blunt with T4 DNA polymerase, and cloned between the two *Hind*III sites of the intermediate plasmid next to the chloramphenicol resistance gene from pCC1FOS-CeuI previously cloned. After this second cloning the construction in the pBluescript II SK+ vector contained the T7 Shine-Dalgarno fused to the initiation codon of *gfp* without *Eco*72I site followed by the chloramphenicol resistance gene and part of *red*F. A fragment comprising the T7 Shine-Dalgarno, the *gfp* without *Eco*72I site and part of the chloramphenicol resistance gene was isolated by digesting with *Xho*I and *Sca*I and used to replace the 5' end of the chloramphenicol resistance gene that is deleted by digesting pMPO571 with *Hind*III and *Sca*I. The final fosmid pMPO579 (Fig. 1) has the chloramphenicol resistance gene reconstituted and bears a promoterless *gfp* just downstream of the metagenomic insert site (*Eco*72I).

Construction of specialized strains.

The EPI300™ -T1 *E. coli* strain was used to integrate in its genome the two constructions bearing the viral expression components. These constructions were integrated in the *trg* locus, by replacing its coding sequence through a modification of the method described by Datsenko and Wanner[5].

Construction of the MPO553 strain involved integration of two consecutive DNA fragments. First, the p*lacUV5* promoter was obtained as an *Eco*RI-*Bam*HI fragment from pNK736 plasmid[6], and cloned into pBluescript II SK+ digested with *Not*I and *Bam*HI, to yield pMPO556. The *nasF* attenuator from pMPO27 flanked by *Eco*RI sites[7], was cloned into pMPO556 digested with *Eco*RI, downstream of p*lacUV5*, thus generating pMPO557. The kanamycin resistance gene was amplified from pKD4[5] using the primers Sac-P1 and Sac-P2, and cloned into pMPO557 digested with *Eco*RV, thus generating pMPO558. The first part of the construction in pMPO558 was amplified by PCR using the primers trgEc-P12 and trgEc-BSK2. The 5' ends of these primers are homologous to the limits of the genomic sequence to be replaced by the amplified construction. The PCR product was digested with *Dpn*I, and electroporated into the EPI300-T1^{R} strain containing pKD46[5] to integrate it in the *trg* locus.

To build the second part of the construction, the gene coding for the chloramphenicol resistance from pKD3 was amplified using the Sac-P1 and Sac-P2 primers and cloned into pGP1-2[8] digested with *Bam*HI, donwstream the gene-1 (coding for the T7 RNA polymerase), thus generating pMPO559. The fragment bearing the gene-1 and the chloramphenicol resistance gene was amplified by PCR from pMPO559 using the primers trgEc-P12 and terSacP2-G1, whose 5' ends are homologous to the limits of the genomic sequence to be replaced by the construction amplified. The PCR product was

digested with *Dpn*I and electroporated into the previously constructed EPI300-T1 strain derivative, which also contained pKD46, to integrate the second DNA fragment downstream of the first one in a way that the kanamycin resistance gene from pKD4 was replaced by the T7 gene-1 and the chloramphenicol resistance gene from pKD3[5]. Finally, the chloramphenicol resistance gene was deleted from the genome using pCP20[9]. A schematic of genomic integration in the resulting MPO553 strain, which bears the lacUV5 promoter, the *nasF* attenuator and the gene-1, is shown in Figure 2. The orientation of this construction is the opposite of the *trg* gene transcription.

For the MPO554 strain construction, the fragment containing the *nahR* gene with its own promoter and the divergent *psal* promoter, was obtained from pCNB4-S2[10] by restriction with *Not*I and *Bam*HI, and cloned into pBluescript II KS+ digested with the same restriction enzymes, thus generating pMPO563. The chloramphenicol resistance gene from pKD3 was amplified using the primers Sac-P1 and Sac-P2 and cloned into pMPO563 digested with *Eco*RV, downstream from and in the same orientation as the *nahR* gene, thus generating pMPO564. Then, the gene N from lambda DNA was amplified using the NotN and KspN primers. The PCR product was digested with *Not*I and *Ksp*I and cloned into pMPO564 partially digested with *Ksp*I and completely digested with *Not*I, thereby generating pMPO565, which has the gene N downstream the p*sal*, and in the same orientation. The construction in pMPO565 was amplified by PCR using the primers trgEc-BKS and trgEc-P1, whose 5' ends are homologous to the limits of the genomic sequence to be replaced by the amplified construction. We digested the PCR product with *Dpn*I and electroporated it into the EPI300-T1^{R} strain containing pKD46 to integrate the construction in the *trg* locus. Finally, the chloramphenicol resistance gene was removed from the genome using pCP20. A schematic of the genomic integration in the MPO554 strain, which bears the *nahR* gene

and the *psal* promoter followed by the gene N, is shown in Figure 2. The orientation of p*sal-N* is the opposite of the *trg* gene transcription.

For MPO555 strain construction, the thnB-6-thnC primer was auto-annealed by incubating it for 5 minutes at 85ºC plus 30 minutes at room temperature and the 3' ends of the annealed product filled-in with Klenow to create a 14 bp blunt-ended product. This product was cloned into pMPO565 digested with *Hpa*I to generate a frameshift in the 60th codon of gene N, which resulted in pMPO575. Construction of the MPO555 strain followed the same procedure as that for MPO554 construction, except that the frameshifted plasmid pMPO575 was used instead of pMPO565 (Fig. 2).

REFERENCES

1. Holloway, B.W. Plasmids that mobilize bacterial chromosome. *Plasmid* **2,** 1-19 (1979).

2. Prodromou, C. & Pearl, L.H. Recursive PCR: a novel technique for total gene synthesis. *Protein Eng*. **5**, 827-829 (1992).

3. Tomás-Gallardo, L. *et al*. Molecular and biochemical characterization of the tetralin degradation pathway in *Rhodococcus* sp. strain TFB. *Microbial Biotechnol*. **2**, 262-273 (2009).

4. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59 (1989).

5. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640-6645 (2000).

6. Simons, R.W., Hoopes, B.C., McClure, W.R. & Kleckner, N. Three promoters near the termini of IS10: pIN, pOUT, and pIII. *Cell* **34**, 673-682 (1983).

7. Royo, J.L., Manyani, H., Cebolla, A. & Santero, E. A new generation of vectors with increased induction ratios by overimposing a second regulatory level by attenuation. *Nucleic Acids Res*. **33**, e169 (2005).

8. Tabor, S. & Richardson, C.C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of especific genes. *Proc. Natl. Acad. Sci. USA* **82**, 1074-1078 (1985).

9. Cherepanov, P.P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14 (1995).

10. Cebolla, A., Sousa, C. & de Lorenzo, V. Rational design of a bacterial transcriptional cascade for amplifying gene expression capacity. *Nucleic Acids Res*. 29,759-766 (2001).

Supplementary Table S1. Oligonucleotides used in this work.

Oligonucleotide	Sequence (5' to 3')
OriTHpaIFw	CAAGTTAACCTTGCCCTCATCTGTTACG
OriTHpaIRev	TCGGTTAACCCAGTCGGTAGATATTCCAC
psalnut1	TAAGGCGCCTTATTGCTGGTGCCCGGCCGGGCGCAATATTCATGTTG ATGATTTATTATATATCGAGTGGTGTATTTATCAATATTGTTTGCTCC
psalnut2	GTCACCTTCATGGTGGTCAGTGCGTCCTGCTGATTAATAACGATAAC GGAGCAAACAATATTGATAAATACACC
psalnut3	GCACTGACCACCATGAAGGTGACGCTCTTAAAAATTAAGCCCTGAA GAAGGGCAGCATTCAAAGCAGAAGGCTTTGGGGTGTGTG
psalnut4	TATGGCGCCCCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTATC ACACACCCCAAAGCCTTCTGC
HindChlFw	CAGGCATGCAAGCTTGAG
BstZredFRv	GGTATACCGGCATACAGC
KpnISDpT77GFP	TAGAGGGTACCAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT ATGAGTAAAGGAGAAGAACTTTTC
GFPMluIRvsopal	GGAACTACAAGACGCGTGCTGAAGTCAAG
GFPMluIFwsolap	CTTGACTTCAGCACGCGTCTTGTAGTTCC
GFPXbaI-TFB-PCRsolap	AGGTCTTCTAGATTATTTGTATAGTTCATC
Sac-P1	TATAGAGCTCTGTAGGCTGGAGCTGCTTC
Sac-P2	TATAGAGCTCATATGAATATCCTCCTTAG
trgEc-P12	GGTTTTTTGCATCACATCAGGTTGGTTCCGTTATTTGCCTGCATTCTA GGGTGTAGGCTGGAGCTGCTTC
trgEc-BSK2	CGCGAGGTTCTGCCGACACAGAATGTTTGTGCAGACGGAATACATC CACCCCATGATTACGCCAAGCTCG
terSacP2-G1	CCGGGCGCTTTTTTTTTGCGCGAATTCGATTATAGAGCTCATATGAA TATTTACTAACTGGAAGAGGCAC
NotN	TATGCGGCCGCCCACTGGCGGTGATACTG
KspN	TAACCGCGGAAAGCCAAGGCCAATATC
trgEc-BKS	GGTTTTTTGCATCACATCAGGTTGGTTCCGTTATTTGCCTGCATTCTA GGTAAAACGACGGCCAGTGAGC
trgEc-P1	CGCGAGGTTCTGCCGACACAGAATGTTTGTGCAGACGGAATACATC CACCGTGTAGGCTGGAGCTGCTTC
thnB-6-thnC	GATCATGCAT