

# Complete sequences of pIJ101-based *Streptomyces*-*Escherichia coli* shuttle vectors

Katelyn V. Brown and S. Eric Nybo\*

## Abstract

High-copy-number plasmids are indispensable tools for gene overexpression studies in prokaryotes to engineer pathways or probe phenotypes of interest. The development of genetic tools for the industrially relevant Actinobacteria is of special interest, given their utility in producing keratolytic enzymes and biologically active natural products. Within the Actinobacteria, *Streptomyces*-*Escherichia coli* shuttle vectors based on the SCP2\* and pIJ101 incompatibility groups are widely employed for molecular cloning and gene expression studies. Here, the sequences of two commonly used pIJ101-based *Streptomyces*-*E. coli* shuttle vectors, pEM4 and pUWL201, were determined using next-generation sequencing. These plasmids drive the expression of heterologous genes using the constitutive *ermE*\* promoter. pEM4 was found to be 8.3 kbp long, containing a  $\beta$ -lactamase gene, thiostrepton resistance marker, the *lacZa* fragment, a ColE1 origin of replication and the *Streptomyces* pIJ101 origin of replication. pUWL201 was found to be 6.78 kbp long, containing a  $\beta$ -lactamase gene, thiostrepton resistance marker, the *lacZa* fragment, a ColE1 origin of replication and the *Streptomyces* pIJ101 origin of replication. Interestingly, the sequences for both pEM4 and pUWL201 exceed their previously reported size by 1.1 and 0.4 kbp, respectively. This report updates the literature with the corrected sequences for these shuttle vectors, ensuring their compatibility with modern synthetic biology cloning methodologies.

## DATA SUMMARY

The sequences for pEM4 (GenBank Accession No. MN970094) and pUWL201 (GenBank Accession No. MN992950) have been deposited in the National Center for Biotechnology Information database. Supplementary data have been deposited in Figshare DOI: <https://doi.org/10.6084/m9.figshare.27170772.v1>.

## INTRODUCTION

Next-generation sequencing (NGS) techniques, such as RNA-seq and Illumina whole-genome sequencing (WGS), have transformed the biological sciences via deep sequencing of whole-cell transcriptomic profiles and genomes at significantly reduced cost [1, 2]. WGS can also be applied to determine unknown plasmid sequences with great accuracy [3]. As such, WGS is a powerful tool for determining plasmid vector sequences for plasmids that have been in widespread use for several decades but for which no sequence information is available in the literature.

pIJ101 is a high-copy plasmid vector replicated in *Streptomyces lividans* and *Streptomyces coelicolor* at a copy number of ~70–100 copies per cell [4]. Since the first report of its use in the cloning of *Streptomyces* genes in the 1980s, many different derivatives of pIJ101 have been developed, including pWHM3 and pUWL201 [5, 6]. These vectors are commonly used to overexpress genes of interest in a wide variety of *Streptomyces* spp. [7, 8].

In early reports, cloning via restriction digest and ligation into a multiple cloning site (MCS) was the gold-standard technique for studying gene expression in *Streptomyces*. Yet, due to limitations in the cost and scope of Sanger sequencing, the sequences

*Access Microbiology* is an open research platform. Pre-prints, peer review reports, and editorial decisions can be found with the online version of this article. Received 01 August 2024; Accepted 08 October 2024; Published 23 October 2024

**Author affiliations:** <sup>1</sup>Department of Pharmaceutical Sciences, College of Pharmacy, Ferris State University, Big Rapids, MI 49307, USA.

**\*Correspondence:** S. Eric Nybo, [EricNybo@Ferris.edu](mailto:EricNybo@Ferris.edu)

**Keywords:** next-generation sequencing; pIJ101; plasmid; *Streptomyces*; synthetic biology.

**Abbreviations:** ATP, adenosine triphosphate; LA, lysogeny broth agar; LB, lysogeny broth; MCS, multiple cloning site; NGS, next-generation sequencing; ssDNA, single-stranded deoxyribonucleic acid; WGS, whole genome sequencing.

000893.v3 © 2024 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

for pIJ101-based plasmids were unreported during the early decades of actinomycete genetic engineering. However, modern cloning techniques are sequence-specific, including Golden-gate assembly [9], Gibson assembly [10] and even yeast recombination assembly [11], methods for which these plasmid vectors would be incompatible without exact sequencing information. Therefore, in this report, we determined the sequence for the popular pEM4 and pUWL201 shuttle vectors using NGS [7, 8, 12–15].

## Highlights

Two closely related *Escherichia coli*–*Streptomyces* shuttle vectors were sequenced and analysed. Updated sequences and vector maps for pEM4 and pUWL201 facilitate the incorporation of these plasmids into modern synthetic biology workflows.

## METHODS

### Bacterial strains, growth conditions and plasmid extraction

*E. coli* JM109 (Promega) was used to propagate plasmids for sequencing analysis. *E. coli* JM109 was grown in Lysogeny Broth (Miller formulation) or on solid lysogeny broth agar plates. Plasmids were introduced into chemically competent *E. coli* JM109 cells by standard procedures [16]. For *E. coli* strains harbouring plasmids, ampicillin was added at a final 100 µg ml<sup>−1</sup> concentration.

Plasmid DNA was extracted from *E. coli* JM109 cells using the Wizard Plus SV Minipreps DNA Purification System (Promega Catalogue No. A1460) following the manufacturer's instructions. The plasmid was eluted in nuclease-free water. Following the manufacturer's instructions, plasmid concentration was measured via a Cytation BioTek Take3 plate reader (Agilent Technologies). Sufficient plasmid purity for whole plasmid sequencing was determined by a 260/230 absorbance ratio of 2.0–2.2 and a 260/230 absorbance ratio of ≥1.8.

### Next-generation plasmid sequencing

The plasmid DNA samples were diluted to a final concentration of 40–65 ng µl<sup>−1</sup>, and aliquots of ≥35 µl were submitted for complete plasmid sequencing. Whole plasmid verification was performed at the Massachusetts General Hospital Center for Computational and Integrative Biology (MGH-CCIB) DNA Core (Cambridge, MA). Sequencing was carried out on an Illumina MiSeq platform with V2 chemistry. Plasmid sequences were assembled using MGH CCIB's *de novo* assembler UltraCycler v1.0 (Brian Seed and Huajun Wang, unpublished).

## RESULTS AND DISCUSSION

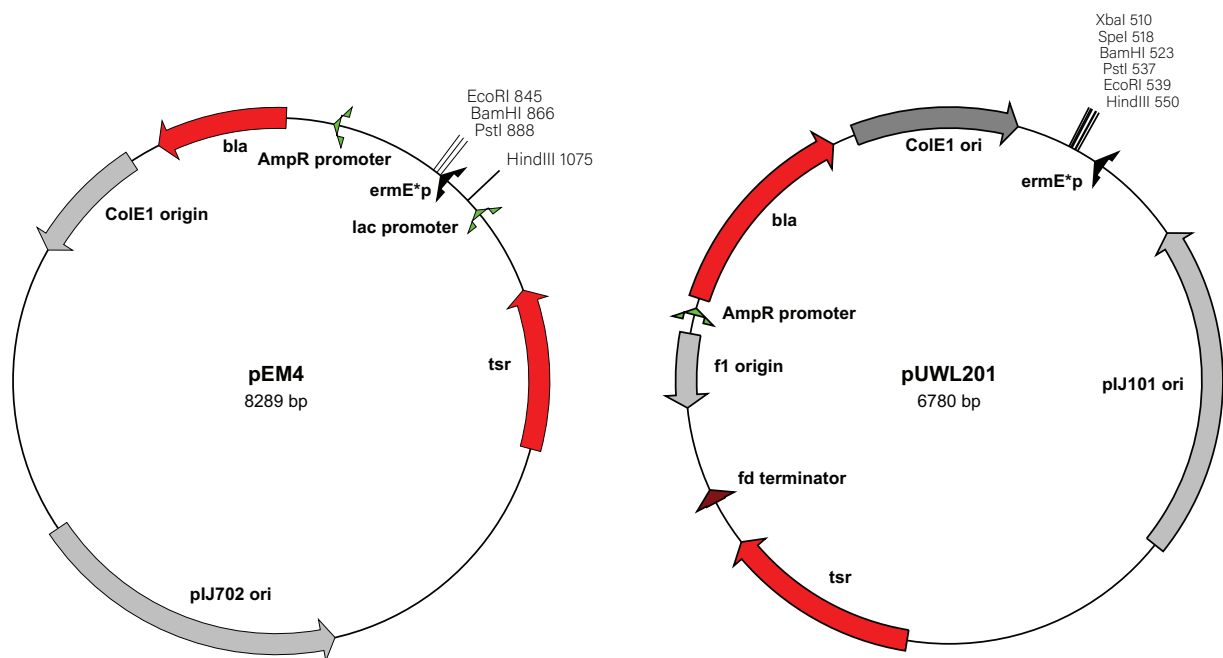
### Sequencing and analysis of pEM4 and pUWL201

Plasmids pEM4 and pUWL201 were sequenced via the Illumina MiSeq NGS platform. pEM4 was sequenced with 209 963 reads, and pUWL201 was sequenced with 12 699 reads. pEM4 is a high-copy number, *E. coli*–*Streptomyces* spp. shuttle vector, which is derived from pWHM4 via insertion of the strong *ermE*\**p* promoter from *Saccharopolyspora erythraea* [17]. The *ermE*\**p* promoter (i.e. the 'ermE-up' promoter) is a mutated version of the *ermE* promoter that features a TGG deletion of the *ermE* P1 promoter [18, 19]. pEM4 (GenBank Accession No. MN970094) was determined to be an 8289 bp plasmid, which features an additional 1089 bp of sequence as compared to the initially reported plasmid size of 7200 bp [6]. The extra bp of the sequence

**Table 1.** Plasmid features of pEM4 and pUWL201

Plasmid features	pEM4	pUWL201
Size (bp)	8289	6780
G+C content (%)	62.6	57.8
ColE1 origin	6901...7583	6383...285
pIJ101 origin	3847...5433	1052...2422
<i>bla</i> resistance	7678...249	5626...6285
<i>tsr</i> resistance	1607...2416	3564...4373
<i>ermE</i> * <i>p</i> promoter	892...933	633...671
Polylinker	846...1073	509...632
LacZ alpha	688...756	4772...4840

Feature is not present in the vector.



**Fig. 1.** Plasmid maps of pEM4 and pUWL201. The *Streptomyces* and *E. coli* origins of replication are coloured silver. The drug resistance markers for *bla<sup>R</sup>* and *tsr<sup>R</sup>* are coloured red.

include the *ermE\***p* promoter and sequence encoding a partial ATP-binding cassette transporter from pIJ486 carried over from the original cloning scheme. pEM4 contains the *bla* gene encoding ampicillin resistance, the *tsr* 23S-rRNA methyltransferase from *Streptomyces azureus* encoding thiostrepton resistance, the high-copy-number ColE1 origin of replication from pUC19 and the pIJ101 origin of replication for *Streptomyces* spp. (Table 1 and Fig. 1).

pUWL201 is another high-copy, *E. coli*–*Streptomyces* shuttle vector derived from pIJ4070 via introducing a 280 bp fragment encoding the *ermE\***p* promoter into the *Kpn*I–*Xba*I sites of the polylinker region [5]. pUWL201 was determined to be 6780 bp in size, which is 380 bp larger than the originally reported plasmid size of 6400 bp. Sequencing of the MCS revealed the presence of one additional *Xba*I restriction site at position 618 in addition to the site in the polylinker. pUWL201 was determined to have the *bla* gene encoding ampicillin resistance, the *tsr* 23S-rRNA methyltransferase from *S. azureus* encoding thiostrepton resistance, the ColE1 origin of replication for *E. coli* and the pIJ101 origin of replication for *Streptomyces* spp. (Table 1 and Fig. 1). The plasmid also features the fd phage terminator from bacteriophage fd [20]. The fd terminator arrests transcription of the upstream *tsr* gene. The bacteriophage f1 origin of DNA replication is from bacteriophage f1 and initiates and terminates viral ssDNA synthesis [21]. Using a helper phage, the f1 origin facilitates the packaging of ssDNA into phage particles [21]. The f1 origin does not interfere with plasmid replication or play any role without a helper phage.

## SUMMARY

The sequences for the broadly used pEM4 and pUWL201 *E. coli*–*Streptomyces* shuttle vectors were determined via NGS and analysed. Both vectors revealed additional nucleotide sequences and were larger in size than previously reported. The characterisation of these vectors reveals that vectors traditionally used for the genetic engineering of prokaryotes may contain carryover sequences not explicitly required for plasmid function. This work should facilitate efforts to generate ‘minimal’ versions of these *E. coli*–*Streptomyces* spp. vectors and will enable their adaptation to synthetic biology.

## AUTHOR NOTES

The raw sequence reads are included as a supplementary file with the online version of the article.

### Funding information

Research reported in this publication was partially supported by the National Science Foundation under grant nos. ENG-2321976 and CHE-2348596 (S.E.N.). Start-up funds and a faculty research grant from Ferris State University (to S.E.N.) also funded this study.

**Acknowledgements**

Dr. Jurgen Rohr and Dr. Steven Van Lanen (University of Kentucky) are gratefully acknowledged for the gift of plasmids pEM4 and pUWL201.

**Author contributions**

Conceptualization: S.E.N. Data curation: K.V.B. Formal analysis: K.V.B. and S.E.N. Investigation: S.E.N. and K.V.B. Funding acquisition: S.E.N. Methodology: S.E.N. Project administration: S.E.N. Visualization: K.V.B. Writing – original draft: S.E.N. Writing – review and editing: K.V.B. and S.E.N.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**References**

- Li R, Zhu H, Ruan J, Qian W, Fang X, *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 2010;20:265–272.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
- Suckling L, McFarlane C, Sawyer C, Chambers SP, Kitney RI, *et al.* Miniaturisation of high-throughput plasmid DNA library preparation for next-generation sequencing using multifactorial optimisation. *Synth Syst Biotechnol* 2019;4:57–66.
- Kieser T, Hopwood DA, Wright HM, Thompson CJ. pJ101, a multi-copy broad host-range *Streptomyces* plasmid: Functional analysis and development of DNA cloning vectors. *Molec Gen Genet* 1982;185:223–238.
- Doumith M, Weingarten P, Wehmeier UF, Salah-Bey K, Benhamou B, *et al.* Analysis of genes involved in 6-deoxyhexose biosynthesis and transfer in *Saccharopolyspora erythraea*. *Mol Gen Genet* 2000;264:477–485.
- Vara J, Lewandowska-Skarbek M, Wang YG, Donadio S, Hutchinson CR. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *J Bacteriol* 1989;171:5872–5881.
- Weiser J, Buriánková K, Kalachová L, Branny P, Pernodet JL. Construction and testing of a bacterial luciferase reporter gene system for in vivo measurement of nonsense suppression in *Streptomyces*. *Folia Microbiol (Praha)* 2006;51:62–64.
- Weitnauer G, Gaisser S, Trefzer A, Stockert S, Westrich L, *et al.* An ATP-binding cassette transporter and two RRNA methyltransferases are involved in resistance to avilamycin in the producer organism *Streptomyces viridochromogenes* Tü57. *Antimicrob Agents Chemother* 2001;45:690–695.
- Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 2008;3:e3647.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;6:343–345.
- Joska TM, Mashruwala A, Boyd JM, Belden WJ. A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. *J Microbiol Methods* 2014;100:46–51.
- Gullón S, Olano C, Abdelfattah MS, Braña AF, Rohr J, *et al.* Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline steffimycin. *Appl Environ Microbiol* 2006;72:4172–4183.
- Schmutz E, Mühlenweg A, Li SM, Heide L. Resistance genes of aminocoumarin producers: two type II topoisomerase genes confer resistance against coumermycin A1 and clorobiocin. *Antimicrob Agents Chemother* 2003;47:869–877.
- Schwartz D, Berger S, Heinzelmann E, Muschko K, Welzel K, *et al.* Biosynthetic gene cluster of the herbicide phosphinothricin tripeptide from *Streptomyces viridochromogenes* Tü494. *Appl Environ Microbiol* 2004;70:7093–7102.
- Temuujin U, Chi WJ, Lee SY, Chang YK, Hong SK. Overexpression and biochemical characterization of DagA from *Streptomyces coelicolor* A3(2): an endo-type  $\beta$ -agarase producing neoagarotetraose and neoagarohexaose. *Appl Microbiol Biotechnol* 2011;92:749–759.
- Sambrook J, W Russell D. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY 999: Cold Spring Harbor Laboratory Press; 2001.
- Quirós LM, Aguirrezabalaga I, Olano C, Méndez C, Salas JA. Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*. *Mol Microbiol* 1998;28:1177–1185.
- Bibb MJ, Janssen GR, Ward JM. Cloning and analysis of the promoter region of the erythromycin-resistance gene (ermE) of *Streptomyces erythraeus*. *Gene* 1986;41:E357–E368.
- Schmitt-John T, Engels JW. Promoter constructions for efficient secretion expression in *Streptomyces lividans*. *Appl Microbiol Biotechnol* 1992;36:493–498.
- Otsuka J, Kunisawa T. Characteristic base sequence patterns of promoter and terminator sites in phi X174 and fd phage DNAs. *J Theor Biol* 1982;97:415–436.
- Dotto GP, Horiuchi K, Zinder ND. The functional origin of bacteriophage f1 DNA replication. Its signals and domains. *J Mol Biol* 1984;172:507–521.

**The Microbiology Society is a membership charity and not-for-profit publisher.**

**Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.**

**Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org)**