University of Birmingham
Research at Birmingham

Annotation of Plasmid Genes

Thomas, Christopher; Thomson, Nicholas; Cerdeño-Tárraga, Ana; Brown, Celeste; Top, Eva; Frost, Laura

DOI:
10.1016/j.plasmid.2017.03.006

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Thomas, C, Thomson, N, Cerdeño-Tárraga, A, Brown, C, Top, E & Frost, L 2017, 'Annotation of Plasmid Genes', *Plasmid*, vol. 91, pp. 61-67. https://doi.org/10.1016/j.plasmid.2017.03.006

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 19. Apr. 2024

Table S1. Annotation of the R6K plasmid genome according to the principles in this letter. Note that protein sequences have been removed for brevity. The sequence and annotation have been submitted to ENA at EBI. The accession number is LT827129.

```
R6K
ID   U00000
FH   Key             Location/Qualifiers
FH
FT   CDS             join(39401..39872,1..1637,6585..7367,)
FT                   /gene="tnpA"
FT                   /gene="R6K0001"
FT                   /product="TnpA protein, Transposase Tn3"
FT                   /note="disrupted by an intact Tn3 transposon; a gene in
FT                   Tn5393"
FT   repeat_unit     1632..1637
FT                   /note="duplicated sequence following insertion of IS3
FT                   transposon. Duplicated at position 6585..6591"
FT   repeat_unit     1637..1674
FT                   /note="left inverted repeat (IR-L) for Tn3"
FT   misc_feature    complement(1637..6584)
FT                   /note="Complete transposon Tn3 region; 100% id to Tn3
FT                   from E. coli plasmid pAPEC-02-R (4948bp); 9bp
FT                   (ccatgaaag) deletion compare to ISTN3X (V00613.1)"
FT   CDS             complement(1784..2644)
FT                   /gene="bla"
FT                   /gene="R6K0002"
FT                   /product="putative TEM-1 Beta-lactamase"
FT                   /note="the first 23 amino acids (a signal peptide) do
FT                   not appear in mature, secreted protein"
FT   CDS             complement(2827..3384)
FT                   /gene="tnpR"
FT                   /gene="R6K0003c"
FT                   /product="Transposon Tn3 resolvase"
FT   CDS             3513..6551
FT                   /gene="tnpA"
FT                   /gene="R6K0004"
FT                   /product="transposase for transposon Tn3"
FT                   /note="Start codon=GTG; 100% id with SWALL:Q6EMC8_ECOLI
FT                   (fasta); 99.6% id with SWALL:TNP3_ECOLI - 3 amino acids
FT                   (FHG) deletion from R6K0004 at codon 730. The deleted
FT                   sequence forms part of a perfect duplication in the
FT                   SWALL:TNP3_ECOLI sequence"
FT   repeat_unit     complement(6547..6584)
FT                   /note="right inverted repeat (IR-R) for Tn3"
FT   repeat_unit     6585..6590
FT                   /note="duplicated sequence following insertion of Tn3
FT                   transposon. Duplicated at position 1632..1637"
FT   repeat_unit     complement(7320..7400)
FT                   /note="transposon Tn5393 5' inverted terminal repeat;
FT                   putative"
FT   repeat_unit     7401..7406
FT                   /rpt_type=DIRECT
FT   CDS             complement(7468..7800)
FT                   /gene="R6K0006"
FT                   /product="putative membrane lipoprotein"
FT                   /note="SWALL:Q46691_ECOLI: killing in klebsiellas
FT                   (KikA) from IncN group plasmid pCU1 in E. coli"
FT   CDS             complement(7803..8102)
FT                   /gene="R6K0007"
FT                   /product="hypothetical protein"
FT                   /note="Fasta: Q9L007_STRCO Q9L007 Putative ABC
FT                   transporter integral (238 aa)”
FT   CDS             complement(8195..10030)
FT                   /gene="cplX1"
FT                   /gene="R6K0008"
FT                   /product="CplX1 conjugation coupling protein”
FT                   /note="formerly TaxB; CplX1 is related to IncP TraG-like proteins,
FT                    and is a member of the VirD4 protein superfamily
FT                    involved in coupling the relaxosome to the DNA-transport apparatus"
FT   CDS              complement(10035..11072)
FT                   /gene="tivB11"
```

```
FT                   /gene="R6K0009"
FT                   /product="TivB11 protein, intracellular transport protein”
FT                   /note="formerly PilX11; belongs to Escherichia coli IncX plasmid R6K
FT                    pilus synthesis operon, EMBL:ECO6342; related to P-loop NTPase
FT                    superfamily; VirB11"
FT   CDS             complement(11090..12304)
FT                   /gene="tivB10"
FT                   /gene="R6K0010"
FT                   /product="TivB10 conjugation transfer protein"
FT                   /note="formerly PilX10; involved in unidirectional conjugation,
FT                   belongs to Escherichia coli IncX plasmid R6K pilus synthesis operon,
FT                   EMBL:ECO6342; related to TrbI superfamily (VirB10)"
FT   CDS             complement(12301..13230)
FT                   /gene="tivB9"
FT                   /gene="R6K0011"
FT                   /product="TivB9 conjugation transfer protein"
FT                   /note="formerly PilX9; belongs to Escherichia coli IncX plasmid R6K
FT                   pilus synthesis operon, EMBL:ECO6342; related to VirB9_CagX_TrbG
FT                   superfamily” “The VirB7/VirB9 heterodimer
FT                   localizes at the outer membrane and plays a critical role
FT                   in stabilizing other VirB proteins during assembly of the
FT                   transfer machine.”
FT   CDS             complement(13236..13958)
FT                   /gene="tivB8"
FT                   /gene="R6K0012"
FT                   /product="TivB8 conjugation transfer protein"
FT                   /note="formerly PilX8; belongs to Escherichia coli IncX plasmid R6K
FT                   pilus synthesis operon, EMBL:ECO6342; related to VirB8 superfamily”
FT                   “putative primary constituent of a DNA transporter "
FT   CDS             complement(13951..14076)
FT                   /gene="tivB7"
FT                   /gene="R6K0012a"
FT                   /product="TivB7 conjugation transfer protein; related to VirB7
FT                   superfamily”
FT                   /note=”formerly PilX7, a truncated hypothetical protein in AJ006342”
FT   CDS             complement(14148..15203)
FT                   /gene="tivB6"
FT                   /gene="R6K0013"
FT                   /product="TivB6 conjugation transfer protein”
FT                   /note= “formerly PilX6; belongs to Escherichia coli IncX plasmid R6K
FT                   pilus synthesis operon, EMBL:ECO6342; related to TrbL superfamily
FT                   (VirB6)”
FT   CDS             complement(15215..15478)
FT                   /gene="eex"
FT                   /gene="R6K0014"
FT                   /product="EexX1”
FT                   /note=”putative membrane lipoprotein involved in entry exclusion"
FT   CDS             complement(15485..16222)
FT                   /gene="tivB5"
FT                   /gene="R6K0015"
FT                   /product="TivB5 conjugation transfer protein”
FT                   /note=”formerly PilX5; related to VirB5_like superfamily"
FT   CDS             complement(16224..18983)
FT                   /gene="tivB3-4"
FT                   /product="TivB3-4"
FT                   /note="formerly PilX3 and PilX4; due to a 2bp insertion at nt 2270
FT                   in the EMBL:ECO6342 entry, this gene has been split into 2
FT                    proteins: Q9EUF9_ECOLI PilX3 protein (VirB3 superfamily) and
FT                    Q9EUF8_ECOLI PilX4 protein (VirB4_CagE superfamily); compare to
FT                    sequence at AJ006342"
FT   CDS             complement(19007..19297)
FT                   /gene="tivB2"
FT                   /gene="R6K0017"
FT                   /product="TivB2 conjugation transfer protein”
FT                   /note=”formerly PilX2; putative propilin gene related to TrbC
FT                    superfamily (VirB2)"
FT   CDS             complement(19281..19925)
FT                   /gene="sltX1"
FT                   /gene="R6K0018"
FT                   /product="SltX1 soluble transglycosylase involved in conjugation;
FT                    related to lysozyme-like superfamily (VirB1)"
```

```
FT   CDS             complement(20150..20641)
FT                   /gene="actX"
FT                   /gene="R6K0019"
FT                   /product="ActX transcriptional elongation regulator"
FT   CDS             complement(21002..22159)
FT                   /gene="rlxX1"
FT                   /gene="R6K0020"
FT                   /product="RlxX1,formerly TaxC; R6K oriT-nickase"
FT                   /note="belongs to VirD2 nickase family; Avila P, Nunez B,
FT                   de la Cruz F, 1996 J Mol Biol. 261, 135-43."
FT   CDS             complement(22162..22710)
FT                   /gene="dtrX1"
FT                   /gene="R6K0021"
FT                   /product="DtrX1, formerly TaxA; putative DNA binding protein of ArcA
FT                    superfamily"
FT                   /note="similar to TraY of F-like plasmids; Nunez B, Avila P,
FT                   de la Cruz F, 1997 Mol Microbiol. 24, 135-43."
FT   misc_feature    22479..23018
FT                   /note="E. coli plasmid R6K alpha origin region; blastn 96%
FT                   id to ECR6KALP"
FT   repeat_unit     22713..22809
FT                   /note="complementary inverted sequence in the beta region"
FT   misc_feature    22788..22806
FT                   /note="taxC nic site in oriT-alpha; Avila, Journal of
FT                   Molecular Biology, Volume 261, 16 August 1996, Pages 135-143"
FT   misc_feature    22861..23127
FT                   /note="Escherichia coli plasmid R6K single-strand
FT                   initiation sequence ssiB; direct RNA primer synthesis in
FT                   R6K replication; Nomura, Gene 108(1):15 (1991) Pubmed:1761225"
FT   CDS             complement(23047..23304)
FT                   /gene="R6K0022"
FT                   /product="hypothetical protein"
FT   CDS             complement(23320..23469)
FT                   /gene="R6K0023"
FT                   /product="hypothetical protein"
FT   CDS             complement(23546..23890)
FT                   /gene="R6K0024"
FT                   /product="hypothetical protein"
FT   CDS             complement(23986..24204)
FT                   /gene="R6K0025"
FT                   /product="putative DNA-binding and ATPase protein"
FT   CDS             complement(24197..24460)
FT                   /gene="R6K0026"
FT                   /product="hypothetical protein"
FT   CDS             complement(24505..25062)
FT                   /gene="R6K0027"
FT                   /product="hypothetical protein"
FT   CDS             complement(25104..25622)
FT                   /gene="R6K0028"
FT                   /product="conserved hypothetical protein"
FT                   /note="HMMPfam hit to PF00226, DnaJ domain, score 4.2e-12"
FT   CDS             complement(25694..25837)
FT                   /gene="R6K0029"
FT                   /product="hypothetical protein"
FT   misc_feature    25855..25968
FT                   /note="Escherichia coli plasmid R6K single-strand
FT                   initiation sequence ssiA; direct RNA primer synthesis in
FT                   R6K replication; Nomura, Gene 108(1):15 (1991) Pubmed:1761225"
FT   rep_origin      26142..26423
FT                   /note="Plasmid R6K gamma origin of replication; ref: Proc.
FT                   Natl. Acad. Sci. U.S.A. 82(5):1480 (1985), Pubmed:3883361"
FT   repeat_unit     26236..26257
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43"
FT   repeat_unit     26258..26279
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43, actually a 21bp tandem
```

```
FT                   repeat"
FT   repeat_unit     26280..26301
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43, actually a 21bp tandem
FT                   repeat"
FT   repeat_unit     26302..26323
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43, actually a 21bp tandem
FT                   repeat"
FT   repeat_unit     26324..26345
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43; actually a 21bp tandem
FT                   repeat"
FT   repeat_unit     26346..26367
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43"
FT   repeat_unit     26368..26389
FT                   /note="22bp direct repeat, in the gamma origin of
FT                   replication, ref: David M. Stalker, Roberto Kolter and
FT                   Donald R. Helinski, Journal of Molecular Biology, Volume
FT                   161, 15 October 1982, Pages 33-43, actually a 21bp tandem
FT                   repeat"
FT   misc_feature    26427..26512
FT                   /note="Plasmid R6K silencer of replication origin gamma;
FT                   I.R.Patel, Cell 47(5):785 (1986), Pubmed:2430721"
FT   misc_binding    join(26533..26544,26560..26566)
FT                   /note="RNA Polymerase recognition and binding sequences, ref:  David
FT                   M. Stalker, Roberto Kolter and Donald R. Helinski,
FT                   Journal of Molecular Biology, Volume 161, 15
FT                   October 1982, Pages 33-43"
FT   RBS             26593..26598
FT                   /note="RBS of pi gene, ref:  David M. Stalker, Roberto
FT                   Kolter and Donald R. Helinski, Journal of Molecular
FT                   Biology, Volume 161, 15 October 1982, Pages 33-43"
FT   CDS             26606..27523
FT                   /gene="pir"
FT                   /gene="R6K0030"
FT                   /product="Pi replication initiation protein"
FT                   /function="initiation for plasmid R6K DNA replication"
FT                   /note="Check sequence: conflict in UNIPROT:PIR_ECOLI
FT                   (P03067): a.a. P in PIR_ECOLI is reported N in ref.2 of
FT                   the same entry and in our sequence"
FT   CDS             27527..27982
FT                   /gene="R6K0031"
FT                   /product="hypothetical transmembrane protein"
FT   misc_feature    28020..28552
FT                   /note="E. coli plasmid R6K beta origin region; 532/533bp
FT                   identical with ECR6KBET (blastn); Mol. Gen. Genet.
FT                   208(1-2):263-270(1987). Pubmed:3302610"
FT   repeat_unit     28053..28149
FT                   /note="complementary inverted sequence in the alpha region"
FT   misc_feature    complement(28057..28075)
FT                   /note="RlxX1 (TaxC) nic site in oriT-beta; Avila, Journal of
FT                   Molecular Biology, Volume 261, 16 August 1996, Pages 135-143"
FT   CDS             complement(28884..29168)
FT                   /gene="R6K0033"
FT                   /product="putative ParB centromere binding protein"
FT   CDS             complement(29248..29910)
FT                   /gene="R6K0034"
FT                   /product="Hypothetical protein"
FT                   /note="putative ParA Walker ATPase family protein"
FT   CDS             30185..30823
FT                   /gene="R6K0035"
```

```
FT                   /product="putative resolvase protein"
FT                   /note="maybe a partial resolvase gene (N-terminal) based
FT                   on fasta results"
FT   misc_feature    30827..30848
FT                   /note="Escherichia coli plasmid R6K replication terminus
FT                   site terR2; T.Horiuchi, Cell 54(4):515 (1988), Pubmed:3042153"
FT   misc_feature    complement(30922..30943)
FT                   /note="Escherichia coli plasmid R6K replication terminus
FT                   site terR1, T.Horiuchi, Cell 54(4):515 (1988), Pubmed:3042153"
FT   CDS             complement(31191..31595)
FT                   /gene="R6K0036"
FT                   /product="hypothetical protein"
FT   CDS             complement(31658..32107)
FT                   /gene="R6K0037"
FT                   /product="putative DNA-binding (H-NS histone-like) protein"
FT   CDS             complement(32119..34443)
FT                   /gene="R6K0038"
FT                   /product="putative DNA topoisomerase"
FT   CDS             complement(34463..34708)
FT                   /gene="R6K0039"
FT                   /product="hypothetical protein"
FT   CDS             complement(34755..34919)
FT                   /gene="R6K0040"
FT                   /product="hypothetical protein"
FT   CDS             complement(34959..35108)
FT                   /gene="R6K0041"
FT                   /product="hypothetical protein"
FT   CDS             complement(35123..35593)
FT                   /gene="R6K0042"
FT                   /product="putative nuclease"
FT                   /note="thermonuclease signature, Staphylococcal nuclease
FT                   homologue, secreted (signal peptide)"
FT   repeat_unit     35609..35614
FT                   /rpt_type=DIRECT
FT   repeat_unit     35615..35695
FT                   /note="3' inverted terminal repeat of transposon Tn5393;
FT                   putative"
FT   CDS             complement(35721..36557)
FT                   /gene="strB"
FT                   /gene="R6K0043"
FT                   /product="streptomycin resistance protein B (streptomycin
FT                   phosphotransferase)"
FT   CDS             complement(36557..37360)
FT                   /gene="strA"
FT                   /gene="R6K0044"
FT                   /product="streptomycin resistance protein a (streptomycin
FT                   phosphotransferase)"
FT   repeat_unit     37426..37428
FT                   /rpt_type=DIRECT
FT   repeat_unit     37429..37455
FT                   /note="inverted repeat of IS1133; 3' inverted terminal
FT                   repeat of insertion sequence IS1133; putative"
FT   misc_feature    37429..38660
FT                   /note="IS1133, from Erwinia amylovora"
FT   CDS             complement(37467..38324)
FT                   /gene="R6K0045"
FT                   /product="putative transposase"
FT                   /note="IS3 family element, transposase with integrase
FT                   catalytic domain (Pfam PF00665; rve)"
FT   CDS             complement(38321..38596)
FT                   /gene="R6K0046"
FT                   /product="putative transposase"
FT                   /note="IS3 family element, PFAM: [PF01527] Transposase_8"
FT   repeat_unit     complement(38634..38660)
FT                   /note="5' inverted terminal repeat of insertion sequence
FT                   IS1133; inverted repeat of IS1133; putative"
FT   CDS             complement(38661..39275)
FT                   /gene="R6K0047"
FT                   /product="putative resolvase"
FT   repeat_unit     38661..38663
FT                   /rpt_type=DIRECT
```