



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/110735/>

Version: Accepted Version

Article:

Fogg, Paul C M, Haley, Joshua A, Stark, W Marshall et al. (2017) Genome integration and excision by a new Streptomyces bacteriophage, ϕ Joe. Applied and Environmental Microbiology. e02767-16. pp. 1-43. ISSN: 0099-2240

<https://doi.org/10.1128/AEM.02767-16>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

20 **Abstract**

21 Bacteriophages are the source of many valuable tools for molecular biology and genetic
22 manipulation. In *Streptomyces*, most DNA cloning vectors are based on serine integrase site-
23 specific DNA recombination systems derived from phage. Because of their efficiency and
24 simplicity, serine integrases are also used for diverse synthetic biology applications. Here we
25 present the genome of a new *Streptomyces* phage, ϕ Joe, and investigate the conditions for
26 integration and excision of the ϕ Joe genome. ϕ Joe belongs to the largest *Streptomyces* phage
27 cluster (R4-like) and encodes a serine integrase. The *attB* site from *S. venezuelae* was used
28 efficiently by an integrating plasmid, pCMF92, constructed using the ϕ Joe *int/attP* locus. The
29 *attB* site for ϕ Joe integrase was occupied in several *Streptomyces* genomes, including *S.*
30 *coelicolor*, by a mobile element that varies in gene content and size between host species.
31 Serine integrases require a phage-encoded recombination directionality factor (RDF) to
32 activate the excision reaction. The ϕ Joe RDF was identified and its function was confirmed *in*
33 *vivo*. Both the integrase and RDF were active in *in vitro* recombination assays. The ϕ Joe site-
34 specific recombination system is likely to be an important addition to the synthetic biology and
35 genome engineering toolbox.

36 **Importance**

37 *Streptomyces* spp. are prolific producers of secondary metabolites including many clinically
38 useful antibiotics. Bacteriophage-derived integrases are important tools for genetic
39 engineering as they enable integration of heterologous DNA into the *Streptomyces*
40 chromosome with ease and high efficiency. Recently researchers have been applying phage
41 integrases for a variety of applications in synthetic biology, including rapid assembly of novel
42 combinations of genes, biosensors and biocomputing. An important requirement for optimal
43 experimental design and predictability when using integrases, however, is the need for
44 multiple enzymes with different specificities for their integration sites. In order to provide a
45 broad platform of integrases we identified and validated the integrase from a newly isolated
46 *Streptomyces* phage, ϕ Joe. ϕ Joe integrase is active *in vitro* and *in vivo*. The specific

- 47 recognition site for integration is present in a wide range of different Actinobacteria, including
- 48 *Streptomyces venezuelae*, an emerging model bacterium in *Streptomyces* research.

49 **Introduction.**

50 Over the past few decades, serine integrases have become widely established as tools
51 for genome engineering and synthetic biology (1, 2). Serine integrases are phage-encoded,
52 DNA site-specific recombinases that mediate recombination between two short (<50 bp)
53 sequences. The integration reaction occurs during the establishment of lysogeny, during
54 which the integrase causes a single crossover between the *attB* site on the bacterial
55 chromosome and the *attP* site on the circularised phage genome leading to the integrated
56 phage DNA flanked by the recombinant sites, *attL* and *attR* (1, 3). Integrase dimers bind to
57 the two *att* sites and produce double-strand breaks with 2 bp overhangs (3, 4); the cut ends
58 are then exchanged and the DNA backbone is re-ligated to produce the recombinant products
59 (5). The *attL* and *attR* sites each contain reciprocal halves of the *attP* and *attB* sites (6). As
60 integrases are unable to use *attL* and *attR* as substrates without an accessory protein, the
61 recombination directionality factor (RDF), the integrated phage genome is stable until the
62 RDF-encoding gene is expressed during prophage induction (3). Recombination between *attL*
63 and *attR* is the excision reaction and is essentially the reverse of integration, releasing the
64 phage genome and reforming *attP* and *attB*. Whilst only integrase is required to mediate
65 integration, excision requires both integrase and the RDF. Genome engineers have exploited
66 these systems to integrate genes of interest into a specific site on the chromosome, which can
67 either be the endogenous *attB* or an introduced *attB* or *attP* used as a docking site (1). The
68 simplicity of the serine integrase mediated site-specific recombination systems means that
69 they are reliably portable to heterologous hosts where DNA can be integrated stably and in
70 single copy.

71 The simple requirements of serine integrases make them amenable to a wide variety
72 of applications. The earliest examples of this were to integrate an *attP* plasmid into a target
73 genome containing the cognate *attB* (or *vice versa*) (7), allowing stable delivery of genes into
74 diverse species, including bacteria (6, 8–10), mice (11), mosquitos (12) and humans (13).
75 More complex genetic engineering approaches use integrases in *in vitro* ordered assembly of

76 multiple DNA fragments (14, 15). *In vivo* genome manipulations can also be achieved either
77 by iterative rounds of recombination (16, 17) or multiplexing orthogonal integrases/*att* sites
78 (18). Integrase mediated DNA rearrangements can also be used to provide permanent genetic
79 memory in novel types of biosensors (19, 20). Some applications, such as *post factum*
80 modifications (15) or biocomputing (19, 21), need controlled excision and this requires
81 integrase and its cognate RDF. The RDF binds directly to the integrase protein and is thought
82 to induce a conformational change that allows *attL* and *attR* to be used as recombination
83 substrates whilst inhibiting recombination of *attB* and *attP* (22, 23).

84 A limiting factor for the use of serine integrases for complex, multiplexed applications
85 is the number of well-characterized integrases and, perhaps more pressingly, RDFs. Only
86 seven integrase/RDF pairs have been characterized to date (from phages TP901-1 (24), ϕ C31
87 (22), ϕ BT1 (25), Bxb1 (23), ϕ Rv1 (26) and SPBc (27), and from the excisive element of
88 *Anabaena* and *Nostoc* cyanobacteria species (28)), but many more integrases have been
89 studied without their RDFs (1, 2, 29–31). Integrase genes are easily identified by comparative
90 sequence analysis and, when the integrase is prophage encoded, the attachment sites can
91 also be predicted (31). RDFs, however, are far more difficult to predict because known
92 examples share little sequence homology, vary markedly in size and also differ in gene
93 location in phage genomes (1). Expansion of the available arsenal of serine integrases and
94 RDFs is desirable to enable advanced synthetic biology applications.

95 Phages that encode serine integrases are prevalent in Gram-positive bacteria, and in
96 particular in Actinobacteria. Here, we describe a newly isolated *Streptomyces* phage, ϕ Joe,
97 and its serine integrase (Int) that is only distantly related to characterized integrases. ϕ Joe Int
98 is active *in vivo* in *Streptomyces* and *E. coli*, the integrase protein is readily purified and is able
99 to carry out efficient *in vitro* recombination. We also describe the ϕ Joe RDF, a 6.8 kDa protein
100 that is able to promote excisive recombination and inhibit integration.

101

102 **Materials and Methods**

103 **Growth media**

104 *Escherichia coli* strains were generally grown in LB, except where otherwise noted. Antibiotics
105 were added for selection where appropriate (apramycin: 50 µg/ml, chloramphenicol: 50 µg/ml,
106 kanamycin: 50 µg/ml, ampicillin: 100 µg/ml). Preparation of competent cells and
107 transformation of *E. coli* were performed as described in Sambrook *et al.*, 2001 (32).
108 *Streptomyces* strains were grown on Mannitol Soya agar (33) supplemented with 10 mM
109 MgCl₂ for plating conjugation mixtures and antibiotics, where required (apramycin: 50 µg/ml,
110 nalidixic acid: 25 µg/ml).

111 **Phage Isolation.** The procedures for isolation, plating and titre of phage with *Streptomyces*
112 as the isolation host are described in detail in Kieser *et al.*, 2000 (33). Raw soil samples were
113 enriched for environmental phage using *S. coelicolor* M145 as a propagation host (34). Briefly,
114 3 g of soil was added to 9 ml Difco™ nutrient (DN) broth (BD Diagnostics, Oxford, UK)
115 supplemented with 10 mM CaCl₂, 10 mM MgSO₄ and 0.5% glucose. *Streptomyces* spores
116 were added to a concentration of 10⁶ colony forming units/ml (cfu/ml) and incubated at 30°C
117 with agitation for 16 h. Soil and bacteria were removed by centrifugation and filtration through
118 a 0.45 µm filter. A dilution series of the filtrate in SM buffer (100 mM NaCl, 8.5 mM MgSO₄, 50
119 mM Tris-HCl pH 7.5, 0.01% gelatin) was plated with *S. coelicolor* spores to isolate single
120 plaques. Phage were recovered from single, well-isolated plaques by single plaque soak outs
121 in DN broth and re-plated with the host strain for three rounds of plaque purification. A high
122 titre phage preparation was generated from plates inoculated with sufficient plaque forming
123 units (pfu) to generate almost confluent lysis (33). The phage suspensions were filtered,
124 pelleted by ultracentrifugation and resuspended in 0.5 ml SM buffer (35). The concentrated
125 phage were further purified by caesium chloride isopycnic density gradient centrifugation (36).

126 **Next Generation Sequencing.** Phage DNA was extracted by phenol:chloroform purification
127 (32) and the presence of pure phage DNA was confirmed by restriction digest. Phage DNA

128 was sequenced and assembled in collaboration with Dr Darren Smith at NU-OMICS
129 (Northumbria University). DNA was prepared for next generation sequencing on the Illumina
130 MiSeq platform using the Nextera XT library preparation kit (Illumina, Saffron Waldon, UK).
131 Samples were loaded and run using a 2 × 250 cycle V2 kit. DNA samples were diluted to 0.2
132 ng/μl, prior to normalization and pooling. Paired end sequencing reads were provided as
133 FASTQ files (NU-OMICS, Northumbria University, Newcastle, UK) and subjected to
134 downstream analysis. ORF prediction and annotations were assigned using DNA master
135 (Lawrence lab, Pittsburgh, PA), Glimmer (37) and Genemark (38). The annotated genome
136 sequence was submitted to GenBank (accession number: KX815338).

137 **Electron Microscopy.** Purified phage were negatively stained with uranyl acetate (39) and
138 imaged in a FEI Tecnai 12 G2 transmission electron microscope fitted with a CCD camera.

139 **Mass Spectrometry.** Whole phage samples were run into a 7 cm NuPAGE Novex 10% Bis-
140 Tris gel (Life Technologies) at 200 V for 6 mins. The total protein band was excised and
141 digested in-gel with 0.5 μg trypsin, overnight at 37°C. Peptides were extracted, concentrated
142 and loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity
143 Symmetry C₁₈, 5 μm trap (180 μm x 20 mm Waters) and a nanoAcquity HSS T3 1.8 μm C₁₈
144 capillary column (75 μm x 250 mm, Waters). The nanoLC system was interfaced with a maXis
145 HD LC-MS/MS system (Bruker Daltonics) with CaptiveSpray ionisation source (Bruker
146 Daltonics). Positive ESI-MS and MS/MS spectra were acquired using AutoMSMS mode.
147 Instrument control, data acquisition and processing were performed using Compass 1.7
148 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics. The collision energy
149 and isolation width settings were automatically calculated using the AutoMSMS fragmentation
150 table, absolute threshold 200 counts, preferred charge states: 2 – 4, singly charged ions
151 excluded. A single MS/MS spectrum was acquired for each precursor and former target ions
152 were excluded for 0.8 min unless the precursor intensity increased fourfold. Protein
153 identification was performed by searching tandem mass spectra against the NCBI nr database

154 using the Mascot search program. Matches were filtered to accept only peptides with expect
155 scores of 0.05 or better.

156 **Plasmid Construction.** Plasmids used in this study are listed in Table 1 and oligonucleotides
157 in Table 2. General molecular biology techniques including plasmid DNA preparation, genomic
158 DNA preparation, restriction endonuclease digestion and agarose gel electrophoresis were
159 performed as described in Sambrook *et al.*, 2001 (32). In-fusion cloning technology (Clontech)
160 was generally used for construction of plasmids. Polymerase chain reaction (PCR)-amplified
161 DNA was generated using primers with Infusion tags for insertion into plasmid vectors, which
162 had been cut with restriction endonucleases. The ϕ Joe integrating plasmid, pCMF92, was
163 created by Infusion cloning of the ϕ Joe *int* gene and *attP* region, obtained by PCR with Joe
164 Int-attP F/R primers and ϕ Joe genomic DNA as a template, into the 3.1 kbp EcoRI-SphI
165 fragment from pSET152. Plasmid pCMF91 was generated by inserting the amplified *attP* site
166 prepared using ϕ Joe genomic DNA as a template and primers Joe *attP* F/R into EcoRI
167 linearized pSP72. The integration sites in *S. coelicolor* were named *attLsc* and *attRsc* and
168 were amplified from *S. coelicolor* gDNA using Joe *attB1* F/R and Joe *attB2* F/R. The *attB* site
169 from *S. venezuelae* (*attBsv*) was amplified using *S. venezuelae* gDNA with Joe *attB Sv* F/Joe
170 *attB* R primers. All three attachment sites were inserted into EcoRI-linearized pGEM7 to
171 produce pCMF90, 94 and 95, respectively. The reconstituted *S. coelicolor attB* sequence
172 (*attBsc*) was prepared from two complementary oligonucleotides, Joe *attB* Recon F and Joe
173 *attB* Recon R (Ultramers primers, IDT) that were annealed and inserted into EcoRI-linearized
174 pGEM7 to produce pCMF97. pCMF98 contains the ϕ Joe *attLsv* and *attRsv* sites in head-to-
175 tail orientation and was isolated by transformation of an *in vitro* recombination reaction
176 between pCMF91 (containing ϕ Joe *attP*) and pCMF95 (containing *attBsv*) into *E. coli*. The
177 *attLsv* and *attRsv* sites in pCMF98 were confirmed by Sanger sequencing (GATC Biotech Ltd,
178 London, UK). The recombination reporter plasmid pCMF116 was constructed by PCR
179 amplification of *lacZ α* using *E. coli* MG1655 gDNA (40) as a template and Joe BzP forward
180 and reverse primers encoding the ϕ Joe *attBsv* and ϕ Joe *attP*, respectively, resulting in the

181 *attBsv* and *attP* sites flanking the *lacZα* gene in head-to-tail orientation. The amplified DNA
182 was inserted into *Xmn*I-linearized pACYC184. pCMF103 was constructed in the same way
183 as pCMF116 except that Joe LzR F/R primers containing the ϕJoe *attLsv* and *attRsv* sites
184 were used.

185 The integrase expression plasmid for protein purification, pCMF87, was constructed
186 by insertion of a PCR fragment encoding the ϕJoe *int* gene, amplified from ϕJoe gDNA using
187 primers Joe H6-Int F/R, into *Nco*I-linearized pEHISTEV expression vector. ϕJoe *g52*, encoding
188 the RDF, was PCR-amplified from ϕJoe gDNA using primers Joe MBP-*g52* F/R and inserted
189 into pETFPP_2 MBP-tag expression vector linearized by PCR with *Cle*F/R to create pCMF96.
190 For *in vivo* recombination assays the integrase expression plasmid, pCMF107, was
191 constructed by insertion of a PCR fragment encoding the ϕJoe *int* gene, amplified from ϕJoe
192 gDNA using primers Joe pBAD Int F/R, into *Nco*I-linearized pBAD-HisA expression vector. A
193 ϕJoe *gp52* and integrase co-expression plasmid, pCMF108, was created by amplification of
194 each gene using Joe pBAD *gp52* F/R and Joe pBAD Int Co-Ex F/Joe pBAD Int R primers,
195 respectively, and insertion of both PCR products simultaneously into pBAD-HisA. The co-
196 expression insert from pCMF108 was subsequently PCR-amplified using Joe H6-*gp52* F/Joe
197 H6-Int R primers and transferred to *Nco*I-linearized pEHISTEV to produce an alternative
198 expression vector, pCMF117.

199 **Conjugation and integration of plasmids in *Streptomyces*.** Transfer of plasmids into
200 *Streptomyces* strains was performed according to the procedures described by Kieser *et al.*,
201 (2000) (33). Conjugation donors were produced by introduction of plasmids into the non-
202 methylating *E. coli* strain, ET12567, containing an RP4 derivative plasmid (pUZ8002), by
203 transformation. Recipient *Streptomyces* spores were used at a concentration of 10⁸/ml, mixed
204 with the *E. coli* donors, plated onto mannitol soya agar supplemented with 10 mM MgCl₂ with
205 no antibiotic selection and incubated at 30°C overnight. Plates containing the donor cells were
206 overlaid with 1 ml water containing 0.5 mg nalidixic acid (for *E. coli* counterselection) and
207 antibiotic for selection of exconjugants (apramycin) before further incubation of all plates at

208 30°C for three days. Integration efficiency was calculated as the number of apramycin-
209 resistant colonies/10⁸ cfu (8).

210 **Protein Purification.** *E. coli* BL21(DE3) containing the relevant expression plasmid were
211 grown (37°C with agitation) in 500 ml 2YT medium (1.6% w/v tryptone, 1.0% w/v yeast extract,
212 0.5 w/v NaCl) to mid-exponential growth phase. The cultures were rapidly chilled on ice for 15
213 min, IPTG was added (final concentration 0.15 mM) and the cultures were further incubated
214 (17°C, 16 h, with agitation). Cells were harvested by centrifugation, resuspended in 20 ml lysis
215 buffer (1 M NaCl, 75 mM Tris pH 7.75, 0.2 mg/ml lysozyme, 500 U Basemuncher
216 Endonuclease; Expedeon Ltd.) and incubated on ice (30 min). The cells were lysed by
217 sonication and debris was removed by centrifugation (18,000 g, 5 min, 4°C). The supernatant
218 was applied to a 5 ml HisTrap FF crude column that had been pre-equilibrated with binding
219 buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4) on an ÄKTA pure
220 25 chromatography system (GE Healthcare). Bound, his-tagged protein was eluted with a step
221 gradient of binding buffer containing 125 mM and 250 mM imidazole. Imidazole was removed
222 from the eluted fractions by pooling the fractions containing the desired protein and applying
223 the pooled solutions to a HiPrep 26/10 desalting column (GE Healthcare) equilibrated with
224 imidazole-free binding buffer. Finally, the protein extracts were subjected to size exclusion
225 chromatography on a HiLoad 16/60 Superdex column. Purified protein fractions were
226 concentrated in a Vivaspin sample concentrator (GE Healthcare) and quantified by
227 absorbance at 280 nm on a Nanodrop spectrophotometer (Thermo Scientific). Protein analysis
228 was performed by denaturing acrylamide gel electrophoresis using pre-made gels (4-12%
229 gradient acrylamide; Expedeon Ltd.); gels were stained with InstantBlue (Expedeon, Ltd.). For
230 storage, an equal volume of 100% glycerol was added to protein samples before freezing at -
231 80°C.

232 ***In vitro* Assays.** Recombination reactions (final volume of 20 µl) were carried out in ϕC31
233 RxE buffer (10 mM Tris pH 7.5, 100 mM NaCl, 5 mM DTT, 5 mM spermidine, 4.5% glycerol,
234 0.5 mg/ml BSA) (41), Bxb1 RxE buffer (20 mM Tris pH 7.5, 25 mM NaCl, 1 mM DTT, 10 mM

235 spermidine, 10 mM EDTA) (23) or TG1 RxE (as Bxb1 RxE plus 0.1 mg/ml BSA) (42).
236 Integrase and RDF proteins were added at the concentrations indicated for each experiment.
237 Plasmids containing the recombination substrates were used at 100ng per reaction. Reactions
238 were either incubated at 30°C for 2 h (to reach steady state) or for specified times. Reactions
239 were stopped by heat (10 min, 75°C), the buffer was adjusted to be compatible with restriction
240 enzymes and the plasmids were digested with XhoI (NEB). The linearized reaction mixtures
241 were run on a 0.8% agarose gel and the relative band intensities were measured to assess
242 activity. Recombination efficiencies were calculated as intensity of product band(s)/sum
243 intensity of all bands.

244 **Bioinformatics.** The ϕ Joe genome was visualized using DNAplotter (43). The *attB* DNA
245 alignment and logo consensus sequence were created with Jalview (44). Protein sequence
246 alignments for visual presentation were produced using the Clustalw (45) program within the
247 Bioedit suite (46). Protein alignments for phylogenetic analysis were produced using Clustal
248 Omega (47) and maximum likelihood trees were created in Mega6 (48). The BLOSUM62
249 similarity matrix was used for protein alignment and annotation (49). Structural alignment of
250 the small RDF proteins was carried out with Promals3D (50). Band densities for *in vitro* assays
251 were measured using the FIJI GelAnalyzer module (51). Accession numbers for all sequences
252 used here are provide in Table S2.

253 **Results and Discussion.**

254 **Isolation of actinophage ϕ Joe and genome sequence.** Raw soil samples were enriched
255 for environmental phage using *S. coelicolor* strain M145 as a propagation host. The phage
256 chosen for further analysis, ϕ Joe, is a siphovirus with a capsid diameter of 46.5 nm (SD 1.6
257 nm, n=9) and a long flexible tail of 199.5 nm (SD 12 nm, n = 8) with clear striations visible in
258 most images (Fig. 1). ϕ Joe is able to plaque on a broad range of *Streptomyces* hosts,
259 producing lytic infection of seven out of nine species tested (Table 3). *Saccharopolyspora*
260 *erythraea* (formerly *Streptomyces erythraeus*) and *Streptomyces venezuelae* were resistant
261 to infection.

262 Genomic DNA was extracted from high titre ϕ Joe suspensions ($>10^{10}$ pfu/ml) and
263 sequenced on the Illumina MiSeq platform with 2,542x coverage. The phage genome is
264 48,941 bases (Accession: KX815338) with a GC content slightly lower than the host bacteria;
265 65.5% compared to ~72% for most *Streptomyces* species. BLASTn was used to measure
266 nucleotide identity for the closest relatives to ϕ Joe; the generalized transducing phage ϕ CAM
267 (52) and two newly sequenced *Streptomyces* phages, Amela and Verse (Fig. S2), are 73, 76
268 and 76% identical, respectively, in global alignments. The ϕ Joe genome contains 81 predicted
269 open reading frames (Fig. 2), the majority of which have similar amino acid sequences to the
270 three phages above and the well characterized R4 phage (53). Notably, similarity to ϕ Joe
271 integrase (gp53) is absent from each of the closest genome matches but is instead present in
272 several more distantly related phages (Fig. 3), indicative of phage mosaicism (54). Specifically,
273 ϕ Joe integrase is homologous to the uncharacterized integrases from five complete phages -
274 Lannister (78% amino acid identity), Zemyla (74%), Danzina (73%), Lika (73%) and Sujidade
275 (73%) (Fig. 3). Comparison to known integrases suggests that the catalytic serine is likely to
276 be at position 46 in the protein sequence (VRLSVFT).

277 Purified phage particles were submitted for shotgun LC-MS/MS analysis to determine
278 the structural proteome. At least one peptide match was detected from fourteen ϕ Joe gene

279 products, five of which have predicted functions – Portal, Capsid, Tail Tape Measure, Scaffold,
280 Head-Tail Adaptor (Figure 2, Table S1). The remaining nine gene products have no known
281 function but all cluster close to the predicted structural genes within a region of the genome
282 spanning ~21 kbp.

283 **Characterization of ϕ Joe integrase and attachment sites.** For most phage-encoded
284 integration systems, the *attP* site lies adjacent to the *int* gene encoding the integrase. The *attP*
285 sites for serine integrases are characteristically about 45 to 50 bp in length and contain
286 inverted repeat sequences flanking a spacer of approximately 20 bp (3, 55). Examination of
287 the ϕ Joe genome identified a candidate *attP* site located 18 bp upstream of the *int* gene. A
288 plasmid, pCMF92, was constructed by replacing the ϕ C31 *int/attP* locus from the widely used
289 integrating vector pSET152, with the ϕ Joe *int/attP* locus (Fig.S1). Integration of pCMF92 would
290 confirm whether the integrase is functional, the nature of the *attP* site and, by rescuing the
291 DNA flanking the integrated plasmid, the identity of the *attB* site could be deduced (Fig. 4).
292 pCMF92 was introduced into *S. coelicolor* J1929 and *S. lividans* TK24 by conjugation and
293 apramycin resistant colonies were obtained, but the frequencies were low (10^{-4} to 10^{-5}
294 exconjugants/cfu) compared to other integrating vectors (10^{-2} to 10^{-3} exconjugants/cfu) (9,
295 18). To test whether integration was site-specific, four *S. coelicolor*:pCMF92 cell lines were
296 amplified from independent exconjugants and the genomic DNAs were analysed by Southern
297 blotting using a probe derived from the ϕ Joe *int* gene. In the four cell lines pCMF92 had
298 integrated into one of two different integration sites, as revealed by hybridisation of the probe
299 to two different restriction fragments (data not shown).

300 We then sought to characterize the two integration sites for pCMF92 in *S. coelicolor* by
301 rescuing the integrated plasmids along with flanking DNA into *E. coli*. In pCMF92 there is 3.9
302 kbp of DNA between the ϕ Joe *attP* site and the PstI cleavage site that contains the plasmid
303 origin of replication and the apramycin resistance gene (Fig. S1). Genomic DNA from two *S.*
304 *coelicolor*:pCMF92 cell lines, each containing pCMF92 integrated into one of the two different
305 integration sites, was digested with PstI endonuclease, self-ligated and introduced into *E. coli*

306 DH5 α by transformation. The rescued plasmids were sequenced over the recombination sites
307 to validate the nature of the ϕ Joe *attP* site and to identify the chromosomal positions of the
308 two *S. coelicolor* integration sites. The ϕ Joe *attP* site was confirmed to be ≤ 50 bp and the 5'
309 GG dinucleotide at the centre of an imperfect inverted repeat is predicted to be where the
310 crossover occurs (Fig. 4A).

311 The two *S. coelicolor* integration sites for pCMF92 are located 3.9 kbp apart, separated
312 by an apparent mobile genetic element comprising *sco2603*, encoding a putative serine
313 integrase with 68% identity to ϕ Joe integrase and two further genes (Fig. 4B). Its product,
314 SCO2603, is 68% identical to ϕ Joe integrase. We hypothesized that the ϕ Joe integrating
315 plasmid is inefficient in *S. coelicolor* because an ancestral and optimal *attB* site is occupied by
316 the SCO2603-encoding element. The two integration sites for pCMF92 in *S. coelicolor* were
317 therefore called *attLsc* and *attRsc* to reflect the provenance of the sites containing the mobile
318 element. To test this hypothesis, the sequence of the ancestral *attB* site, *attBsc*, was predicted
319 by removing the sequence between *attLsc* and *attRsc*, including the *attP* moieties that would
320 have originated from the inserted mobile element (Fig. 4C). The reconstituted *attBsc* was used
321 to interrogate the GenBank *Streptomyces* database for closely related extant sequences.
322 Three species were chosen from the top ten hits returned (*S. avermitilis*, *S. albus* and *S.*
323 *venezuelae*, Fig. 4D) and assayed for *in vivo* integration efficiency. *S. venezuelae* was the
324 only host to support highly efficient integration after conjugation with pCMF92, 160-fold greater
325 frequency than *S. coelicolor* and 1,600-fold greater than *S. lividans* (Fig. 5A). The integration
326 frequencies for pCMF92 into *S. venezuelae* are similar to those reported for other
327 characterized serine integrases (9, 18) and we demonstrate below that the *attB* site from *S.*
328 *venezuelae*, *attBsv*, is indeed used efficiently by ϕ Joe integrase. Plasmid pCMF92 could
329 therefore be used as a new integrating vector for use in this newly emerging model system for
330 *Streptomyces* research.

331 The *S. venezuelae attBsv* site was used as a BLASTn query to estimate the prevalence
332 of potential ϕ Joe insertion sites in sequenced species. In many instances, each half of the

333 query sequence matched separate locations in the target genome, suggesting that ϕ Joe-like
334 *attB* sites are frequently occupied by either a prophage or a similar mobile element to that
335 observed in *S. coelicolor* J1929. Hits were subsequently filtered for matches of at least 80%
336 coverage with an e-value of $<1 \times 10^{-10}$ and a bit score >75 , which revealed numerous apparently
337 unoccupied ϕ Joe *attB* sites in diverse *Streptomyces*, *Kitasatospora* and *Dermacoccus* species
338 (Fig. 4D). Generally, the *attB* site for ϕ Joe and the SCO2603 integrase-encoding elements is
339 located 74bp from the end of an ORF encoding a SCO2606-like predicted B12 binding
340 domain-containing radical SAM protein. Insertions this close to the end of an ORF may not
341 necessarily cause loss of function of the gene product and this could explain the prolific
342 number of mobile elements that use this locus as an insertion site. Other than the
343 recombination genes, the genetic content of the mobile elements located here varies markedly
344 in different bacterial species (Fig. S2). Some *Streptomyces* strains have an almost identical
345 SCO2603-containing genetic element to *S. coelicolor* J1929 (e.g. WM6391), others have no
346 genes other than the recombination genes (e.g. NRRLF-5123) and some contain up to 40 kbp
347 between the predicted *attL* and *attR* sites (Fig. S2).

348 **ϕ Joe integrase catalyses efficient *in vivo* and *in vitro* integration.** In order for an integrase
349 to have broad appeal as a bioengineering tool it must be functional in heterologous hosts. As
350 a proof of principle, we tested the activity of ϕ Joe integrase in *E. coli* by cloning the integrase
351 gene into an arabinose-inducible expression vector, pBAD-HisA, to produce pCMF107.
352 Meanwhile, we constructed a reporter plasmid, pCMF116, containing the *E. coli lacZ α* gene
353 flanked by ϕ Joe *attB_{sv}* and *attP* sites in head to tail orientation (Fig. S3). Both plasmids were
354 introduced into *E. coli* TOP10 cells (Invitrogen) by co-transformation and plated on selective
355 agar plates containing 0.2% L-arabinose and 80 μ g/ml X-Gal. pBAD-HisA lacking an insert
356 was used as a negative control. All of the transformants were white in the presence of ϕ Joe
357 *int*, indicating efficient recombination between the *attB_{sv}* and *attP* sites leading to loss of the
358 *lacZ α* gene (Fig. 5B & S3). ϕ Joe integrase and its cognate *attB_{sv}* and *attP* sites are, therefore,
359 active in *E. coli*.

360 Another key application for serine integrases is for *in vitro* combinatorial assembly of
361 genes for optimising expression of metabolic pathways (14, 15). In this application different
362 integrases are used to join (by recombination) specific pairs of DNA fragments tagged with
363 their cognate attachment sites. In theory this procedure can be multiplexed to assemble many
364 DNA fragments together using different, orthogonally acting integrases. The aim is to generate
365 artificial operons with defined or random order. To test the suitability of ϕ Joe Int for *in vitro*
366 recombination reactions, the integrase gene was cloned into the His-tag expression vector
367 pEHISTEV and purified after overexpression in *E. coli*. *In vitro* recombination assays were
368 carried out with ϕ Joe *attP* (pCMF91) versus each of *attBsc*, *attLsc*, *attRsc* and *attBsv*
369 (pCMF97, pCMF90, pCMF94 and pCMF95, respectively) and using a range of ϕ Joe integrase
370 concentrations. Successful recombination between attachment sites produces a co-integrant
371 plasmid, which can be distinguished from the substrate plasmids by a restriction digest and
372 agarose gel electrophoresis (Fig. S3). In this assay, recombination was undetectable when
373 *attLsc* (pCMF90) or *attRsc* (pCMF94) were used with *attP* (pCMF91) as substrates. A small
374 amount of recombination was observed ($\leq 2\%$, Fig. S4) when the reconstituted *attBsc*
375 (pCMF97) was used with *attP* (pCMF91). However, consistent with the observations in *E. coli*
376 and in *Streptomyces*, the *S. venezuelae attBsv* site (pCMF95) was a highly efficient substrate
377 for recombination with the ϕ Joe *attP* site. ϕ Joe integrase was effective over a broad range of
378 concentrations (50 – 1000 nM) (Fig. 5C & S4). Using 200 nM integrase, detectable
379 recombination product was produced after ~10-15 min, and after 2 h approximately 70% of
380 the substrate molecules were converted to product (Fig. 5C & D).

381 There are only 6 bp that differ between *attBsc* and *attBsv*, and all the differences are
382 on the left-hand arm of the *attB* sites (Fig. 4C). Previously, a mutational analysis of the ϕ C31
383 *attB* site showed that mutationally sensitive bases occur 2, 15 and 16 bases to either side of
384 the crossover dinucleotide (56). As two of the differences between *attBsc* and *attBsv* are also
385 2 and 16 bases from the putative crossover 5'GG (Fig. 4C), these base pair differences might
386 account for the poor activity of *attBsc* in the *in vitro* assays.

387 **Identification and validation of the ϕ Joe RDF protein, gp52.** Although there are dozens of
388 serine integrases that have been described in the literature, there are only seven published
389 RDFs for serine integrases (ϕ C31 gp3 (22), ϕ BT1 gp3 (25), Bxb1 gp47 (23), TP901 ORF7/Xis
390 (24), Anabaena/Nostoc Xisl (57), SPBc SprB (27), and ϕ Rv1 Rv1584c/Xis (26)). The Bxb1
391 and ϕ C31 RDFs are amongst the largest of these RDF proteins (approximately 27.5 kDa, 250
392 amino acids) and their genes are located in proximity to the phage DNA replication genes.
393 Both RDFs have functions during phage replication in addition to acting as RDFs but they are
394 evolutionarily unrelated (25, 58). The RDFs from ϕ BT1 and another ϕ C31-like phage, TG1,
395 are close relatives of the ϕ C31 RDF at the sequence level (85% and 59% identical,
396 respectively); furthermore, the ϕ BT1-encoded RDF acts on ϕ C31 integrase and *vice versa*
397 (25). The ϕ Rv1 and SPBc RDFs are located within 1 or 2 ORFs of the *int* gene, a feature which
398 is reminiscent of the *xis* genes that act with tyrosine integrases. ϕ Rv1, SPBc, TP901 and
399 Anabaena/Nostoc RDFs are much smaller proteins than ϕ C31 gp3 or Bxb1 gp47 (58 and 110
400 amino acids). Given the variation in RDF size, sequence and genomic location, there are no
401 sound generalizations yet for identifying new RDFs in phage genomes.

402 A list of four candidate genes (*g40*, *43*, *49* and *52*) for the ϕ Joe RDF was drawn up
403 based on comparable size to known, small RDFs and genomic location (i.e. not located
404 amongst the late/structural genes) (Fig. 2). One of the potential RDF genes (*g52*) is adjacent
405 to *int* in the ϕ Joe genome, but it is transcribed divergently, with the *attP* site situated between
406 *int* and *g52* (Fig. 2). Unlike the other candidate RDFs, gp52 homologues are only found in
407 those phages with ϕ Joe-like integrases (Fig. 3), and phylogenetic analysis of gp52 and the
408 integrase indicated that both proteins have followed a parallel evolutionary path (Fig. S5).
409 Pairwise alignment of the 6.8 kDa (62 amino acids) gp52 protein with other known small RDFs
410 revealed homology with ϕ Rv1 RDF (25.7% identity and 35.1% similarity; Fig. 6A). Also,
411 examination of the mobile elements that have inserted into the *attB* sites in *S. coelicolor* and
412 other *Streptomyces* spp, revealed that they also contain a gene encoding a gp52 homologue
413 in a similar genetic context i.e. the *int* and *g52* genes are adjacent to the *attL* and *attR* sites,

414 respectively, and would flank *attP* after excision (Fig. 4B & S2). The predicted secondary
415 structure of ϕ Joe gp52 contains an alpha-helix in the N-terminal region, a beta-sheet in the C-
416 terminal region and an unstructured region in between (Fig. S6). Alignment of the ϕ Joe-like
417 RDFs found in intact phages and the RDFs found in the SCO2603-encoding mobile elements
418 indicated that both of the structured regions are well conserved, particularly the putative alpha-
419 helix, but the centre of the protein is variable (Fig. S6).

420 RDFs are able to influence integrase-catalysed recombination in two ways; they
421 activate the *attL* x *attR* reaction to regenerate *attP* and *attB* (excision) and they inhibit the *attB*
422 x *attP* integration reaction (22, 23). We were unable to produce sufficient soluble gp52 protein
423 for *in vitro* assays when expressed with a simple histidine-tag; however, a maltose-binding
424 protein MBP-gp52 fusion protein was more soluble. We tested the ability of MBP-gp52 to
425 inhibit integration by titrating the protein against a fixed concentration of integrase at MBP-
426 gp52:Int ratios of 1:2 to 22.5:1. When the MBP-gp52 was in excess integration was repressed
427 to less than 10%; however, at less than equimolar concentrations, recombination was
428 equivalent to the control in which no MBP-gp52 was added (Fig. 6B). These results are similar
429 to observations for ϕ C31 and Bxb1 integrases and their cognate RDFs, gp3 and gp47 (22,
430 23).

431 To test the ability of gp52 to activate an excision reaction, a plasmid containing the
432 cognate *attLsv* and *attRsv* sites was produced, pCMF98 (Fig. S3). The MBP-gp52 protein was
433 unable to promote efficient excision under any conditions tested (not shown). Removal of the
434 MBP-tag, using 3c protease, increased excision activity but the reaction was still inefficient
435 after 2 h incubation (Fig. 6C). Longer incubations of 5 – 20 h further increased the amount of
436 substrates converted to product up to 45%, but also led to significant amounts of excision
437 products (10-20%) by the integrase alone. Thus, in comparison to the activity of other RDFs,
438 gp52 has rather poor activity; ϕ C31 gp3 activates approximately 60 to 80% conversion of the
439 *attL* x *attR* substrates to products (22) and similar results are obtained with other RDFs (23,
440 25, 26).

441 To test the excision ability of ϕ Joe gp52 *in vivo*, a *g52* and *int* co-expression operon
442 was designed in which *int* and *g52* were located directly downstream of the T7 promoter and
443 ribosome binding site (RBS) in the expression vector pEHISTEV to produce pCMF117. A
444 reporter plasmid, pCMF103, was produced containing the *lacZ α* gene flanked by ϕ Joe *attLsv*
445 and *attRsv* sites (Fig. S3). pCMF117 and pCMF103 were introduced into *E. coli* BL21(DE3)
446 cells by co-transformation and plated onto LB agar supplemented with 0.5 mM IPTG to induce
447 expression of the *g52-int* operon (30). The reporter plasmid was then extracted from the
448 BL21(DE3) transformants and introduced into *E. coli* DH5 α to determine the percentage of
449 plasmids that had undergone *attLsv* x *attRsv* recombination and had lost the *lacZ α* gene. As
450 controls, plasmids expressing either only integrase (pCMF87) or only gp52 (pCMF100) were
451 also introduced together with the reporter (pCMF103) into BL21(DE3) and the assay was
452 repeated using the same procedure. When ϕ Joe integrase alone was expressed, excision
453 occurred at a frequency of 37.6% (SD=5.1%, n=5) but when co-expressed with gp52 the
454 frequency rose to 96.8% (SD=1.3%, n=5) (Fig. 6D). Expression of gp52 without integrase led
455 to no detectable excision events (Fig. 6D). Although overall recombination *in vivo* was higher
456 than *in vitro*, the relative levels of *attLsv* x *attRsv* recombination by ϕ Joe integrase alone and
457 ϕ Joe integrase with gp52 were comparable. Taken together, the *in vivo* and *in vitro* data
458 indicate that ϕ Joe gp52 has RDF activity.

459 The observation that ϕ Joe integrase has a basal level of excision activity in the absence of its
460 RDF is highly unusual for a phage-encoded integrase and further study may provide novel
461 insights into the mechanism and evolution of the serine integrases. *Streptomyces* phage ϕ BT1
462 integrase was shown to catalyse bidirectional recombination, albeit at extremely low levels
463 (59). The archetypal ϕ C31 integrase is only able to mediate *attL* x *attR* recombination in the
464 absence of gp3 when certain mutations are introduced just upstream or within a motif, the
465 coiled coil motif, required for subunit-subunit interactions during synapsis of DNA substrates
466 (60). The coiled coil motifs are also thought to play a role in inhibiting recombination between
467 *attL* and *attR* in the absence of the RDF; the ϕ C31 IntE449K mutation or its RDF, gp3, relieves

468 this inhibition (55, 60–62). Three independent structural predictions indicate the presence of
469 a coiled coil domain in the ϕ Joe Int C-terminal domain (A395-T453, Fig. S7). The high basal
470 excision activity of ϕ Joe integrase could be due to incomplete inhibition of synapsis by the
471 coiled coil motif when integrase is bound to *attL* and *attR*, reminiscent of the hyperactive ϕ C31
472 mutant IntE449K (60). Natural bidirectional, large serine recombinases include the
473 transposases TnpX (63) and TndX (64) from clostridial integrated conjugative elements
474 (ICEs); ϕ Joe integrase could be an evolutionary intermediate between these bi-directional
475 recombinases and the highly directional recombinases such as ϕ C31 and Bxb1 integrases.
476 Our data show that, under the *in vitro* conditions used, gp52 was highly effective at inhibiting
477 integration by ϕ Joe integrase but only weakly activated excision. It remains to be seen whether
478 this system, with its unusual properties, is sufficiently robust to regulate phage genome
479 integration and excision according to the developmental choices of ϕ Joe.

480 The properties of the ϕ Joe integrase and gp52 are compatible with some of the existing
481 applications for serine integrases, but they could also present opportunities for new
482 applications. ϕ Joe integrase is highly efficient in integration assays *in vivo* and *in vitro*, and *in*
483 *vivo* excision when the RDF is present. In *attB* x *attP* integration assays, the yield of products
484 by ϕ Joe integrase was comparable to well established integrases such as those of ϕ C31 or
485 Bxb1. Furthermore, ϕ Joe integrase is active in buffers compatible with other characterized
486 integrases indicating that it could be used in DNA assembly procedures in combination with
487 other integrases. Although yet to be tested, assemblies generated with ϕ Joe integrase could
488 later be used as substrates for modification by ϕ Joe integrase in a single step. The innate
489 excision activity of ϕ Joe integrase could excise a fragment flanked by *attLsv* or *attRsv* sites
490 and, in the same reaction, replace it via an integration reaction. ϕ Joe integrase could therefore
491 provide a more streamlined tool than the existing requirement for two steps by the more
492 directional integrases such as those from ϕ C31 and Bxb1 (15). Furthermore, given that ϕ Joe
493 Int can mediate basal levels of excision in the absence of RDF, integrating plasmids based on
494 ϕ Joe *int/attP* may display a degree of instability. Selection for the plasmid marker would ensure

495 plasmid maintenance when desired but, if the plasmid is easily lost without selection, this trait
496 could be desirable if there is a need to cure the strain of the plasmid or during studies on
497 synthetic lethality.

498 **Conclusions.** On the basis of sequence and genome organisation, phage Joe is a member
499 of a large cluster of R4-like *Streptomyces* phages. Its closest relatives at the nucleotide level
500 are *Streptomyces* phages Amila and Verse with very high levels of nucleotide identity in the
501 regions encoding essential early and structural genes. However, Joe integrase is more closely
502 related to the integrases from five other R4-like cluster phages - Lannister, Danzina, Zemlya,
503 Lika and Sujidade. At the present time the majority of *Streptomyces* phages belong to the R4-
504 like cluster phages, but there is a continuum of relatedness throughout the cluster; for example
505 R4 is a more distant relative to ϕ Joe than any of the other phages mentioned above.

506 We identified the RDF for Joe integrase on the basis of its gene location, small size and distant
507 similarity to another known RDF, Rv1584c. Although this identification was relatively
508 straightforward, it is not clear yet how general such an approach might be. The activity of ϕ Joe
509 integrase and RDF contributes to the growing number of complete serine integrase site-
510 specific recombination systems that are available for use in synthetic biology applications. The
511 ϕ Joe *int/attP* plasmid, pCMF92, also adds to the number of useful integrating vectors for use
512 in *Streptomyces* species. However, and unusually for a phage integrase, ϕ Joe Int displays a
513 significant level of excisive recombination in the absence of its RDF while still being efficient
514 at mediating integration. This bi-directional property could be applied in new ways in future
515 applications of serine integrases.

516

517 **Acknowledgements**

518 We are grateful to Dr Darren Smith (NU-OMICS, Northumbria University) for phage genome
519 sequencing and assembly, and to the York Biosciences Technology Facility for proteomics

520 and electron microscopy. This research was performed with funding from the Biotechnology
521 and Biological Sciences Research Council (project grant BB/K003356/1) and the Microbiology
522 Society (formerly Society for General Microbiology) with a Harry Smith Vacation Scholarship
523 for JH.

524

525

526 **References.**

- 527 1. **Fogg PCM, Colloms S, Rosser S, Stark M, Smith MCM.** 2014. New applications for
528 phage integrases. *J Mol Biol* **426**:2703–2716.
- 529 2. **Groth AC, Calos MP.** 2004. Phage integrases: Biology and applications. *J Mol Biol*
530 **335**:667–678.
- 531 3. **Smith MCM.** 2015. Phage-encoded Serine Integrases and Other Large Serine
532 Recombinases. *Microbiol Spectr* **3**:1–19.
- 533 4. **Smith MCA, Till R, Smith MCM.** 2004. Switching the polarity of a bacteriophage
534 integration system. *Mol Microbiol* **51**:1719–1728.
- 535 5. **Olorunniji FJ, Buck DE, Colloms SD, McEwan AR, Smith MCM, Stark WM,**
536 **Rosser SJ.** 2012. Gated rotation mechanism of site-specific recombination by ϕ C31
537 integrase. *Proc Natl Acad Sci U S A* **109**:19661–6.
- 538 6. **Thorpe HM, Smith MC.** 1998. In vitro site-specific integration of bacteriophage DNA
539 catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S*
540 *A* **95**:5505–5510.
- 541 7. **Kuhstoss S, Richardson MA, Rao RN.** 1991. Plasmid cloning vectors that integrate
542 site-specifically in *Streptomyces* spp. *Gene* **97**:143–146.
- 543 8. **Fayed B, Ashford DA, Hashem AM, Amin MA, El Gazayerly ON, Gregory MA,**
544 **Smith MCM.** 2015. Multiplexed integrating plasmids for engineering of the
545 erythromycin gene cluster for expression in *Streptomyces* spp. and combinatorial
546 biosynthesis. *Appl Environ Microbiol* **81**:8402–8413.
- 547 9. **Gregory MA, Till R, Smith MCM.** 2003. Integration site for *Streptomyces* phage
548 ϕ BT1 and development of site-specific integrating vectors. *J Bacteriol* **185**:5320–
549 5323.
- 550 10. **Hong Y, Hondalus MK.** 2008. Site-specific integration of *Streptomyces* ϕ C31
551 integrase-based vectors in the chromosome of *Rhodococcus equi*. *FEMS Microbiol*
552 *Lett* **287**:63–68.
- 553 11. **Chavez CL, Keravala A, Chu JN, Farruggio AP, Cuéllar VE, Voorberg J, Calos**
554 **MP.** 2012. Long-Term Expression of Human Coagulation Factor VIII in a Tolerant
555 Mouse Model Using the ϕ C31 Integrase System. *Hum Gene Ther* **23**:390–398.
- 556 12. **Meredith JM, Basu S, Nimmo DD, Larget-Thierry I, Warr EL, Underhill A,**
557 **McArthur CC, Carter V, Hurd H, Bourgouin C, Eggleston P.** 2011. Site-specific
558 integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae*
559 significantly reduces *Plasmodium* infections. *PLoS One* **6**:e14587.
- 560 13. **Groth AC, Olivares EC, Thyagarajan B, Calos MP.** 2000. A phage integrase directs
561 efficient site-specific integration in human cells. *Proc Natl Acad Sci U S A* **97**:5995–
562 6000.
- 563 14. **Zhang L, Zhao G, Ding X.** 2011. Tandem assembly of the epothilone biosynthetic
564 gene cluster by in vitro site-specific recombination. *Sci Rep* **1**:141.
- 565 15. **Colloms SD, Merrick CA, Olorunniji FJ, Stark WM, Smith MCM, Osbourn A,**
566 **Keasling JD, Rosser SJ.** 2014. Rapid metabolic pathway assembly and modification
567 using serine integrase site-specific recombination. *Nucleic Acids Res* **42**.
- 568 16. **Dafhnis-Calas F, Xu Z, Haines S, Malla SK, Smith MCM, Brown WRA.** 2005.
569 Iterative in vivo assembly of large and complex transgenes by combining the activities
570 of ϕ C31 integrase and Cre recombinase. *Nucleic Acids Res* **33**:e189.
- 571 17. **Xu Z, Lee NCO, Dafhnis-Calas F, Malla S, Smith MCM, Brown WRA.** 2008. Site-
572 specific recombination in *Schizosaccharomyces pombe* and systematic assembly of a
573 400kb transgene array in mammalian cells using the integrase of *Streptomyces* phage
574 ϕ BT1. *Nucleic Acids Res* **36**:e9.
- 575 18. **Fayed B, Younger E, Taylor G, Smith MCM.** 2014. A novel *Streptomyces* spp.
576 integration vector derived from the *S. venezuelae* phage, SV1. *BMC Biotechnol* **14**:51.
- 577 19. **Siuti P, Yazbek J, Lu TK.** 2013. Synthetic circuits integrating logic and memory in
578 living cells. *Nat Biotechnol* **31**:448–52.
- 579 20. **Bonnet J, Subsoontorn P, Endy D.** 2012. Rewritable digital data storage in live cells

- 580 via engineered control of recombination directionality. *Proc Natl Acad Sci* **109**:8884–
581 8889.
- 582 21. **Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D.** 2013. Amplifying genetic logic
583 gates. *Science* (80-) **340**:599–603.
- 584 22. **Khaleel T, Younger E, Mcewan AR, Varghese AS, Smith MCM.** 2011. A phage
585 protein that binds ϕ C31 integrase to switch its directionality. *Mol Microbiol* **80**:1450–
586 1463.
- 587 23. **Ghosh P, Wasil LR, Hatfull GF.** 2006. Control of phage Bxb1 excision by a novel
588 recombination directionality factor. *PLoS Biol* **4**:0964–0974.
- 589 24. **Breüner A, Brøndsted L, Hammer K.** 1999. Novel organization of genes involved in
590 prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J*
591 *Bacteriol* **181**:7291–7297.
- 592 25. **Zhang L, Zhu B, Dai R, Zhao G, Ding X.** 2013. Control of directionality in
593 *Streptomyces* phage ϕ BT1 integrase-mediated site-specific recombination. *PLoS One*
594 **8**:e80434.
- 595 26. **Bibb LA, Hancox MI, Hatfull GF.** 2005. Integration and excision by the large serine
596 recombinase ϕ Rv1 integrase. *Mol Microbiol* **55**:1896–1910.
- 597 27. **Abe K, Kawano Y, Iwamoto K, Arai K, Maruyama Y, Eichenberger P, Sato T.**
598 2014. Developmentally-Regulated Excision of the SP β Prophage Reconstitutes a
599 Gene Required for Spore Envelope Maturation in *Bacillus subtilis*. *PLoS Genet*
600 **10**:e1004636.
- 601 28. **Ramaswamy KS, Carrasco CD, Fatma T, Golden JW.** 1997. Cell-type specificity of
602 the *Anabaena* fdxN-element rearrangement requires xisH and xisl. *Mol Microbiol*
603 **23**:1241–1249.
- 604 29. **Xu Z, Brown WRA.** 2016. Comparison and optimization of ten phage encoded serine
605 integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol*
606 **16**:13.
- 607 30. **Xu Z, Thomas L, Davies B, Chalmers R, Smith M, Brown W.** 2013. Accuracy and
608 efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases
609 for the integration of DNA into the human genome. *BMC Biotechnol* **13**:87.
- 610 31. **Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK,**
611 **Voigt CA.** 2014. Permanent genetic memory with >1-byte capacity. *Nat Methods*
612 **11**:1261–1266.
- 613 32. **Sambrook J, Fritsch EF, Maniatis T.** 2001. *Molecular Cloning: A Laboratory Manual.*
614 Cold Spring Harb Lab **3**:2344.
- 615 33. **Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA.** 2000. *Practical*
616 *Streptomyces Genetics.* John Innes Cent Ltd.
- 617 34. **Bentley SD, Chater KF, Cerdeño-Tárraga a-M, Challis GL, Thomson NR, James**
618 **KD, Harris DE, Quail M a, Kieser H, Harper D, Bateman a, Brown S, Chandra G,**
619 **Chen CW, Collins M, Cronin a, Fraser a, Goble a, Hidalgo J, Hornsby T,**
620 **Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O’Neil S,**
621 **Rabbinowitsch E, Rajandream M, Rutherford K, Rutter S, Seeger K, Saunders D,**
622 **Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek a, Woodward J,**
623 **Barrell BG, Parkhill J, Hopwood D a.** 2002. Complete genome sequence of the
624 model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141–147.
- 625 35. **Fogg PCM, Hynes AP, Digby E, Lang AS, Beatty JT.** 2011. Characterization of a
626 newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain
627 SB1003. *Virology*, 2011/10/25 ed. **421**:211–221.
- 628 36. **Clokic MRJ, Kropinski AM.** 2009. *Bacteriophages Methods in molecular biology.*
- 629 37. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial
630 gene identification with GLIMMER. *Nucleic Acids Res* **27**:4636–4641.
- 631 38. **Besemer J, Borodovsky M.** 2005. GeneMark: Web software for gene finding in
632 prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **33**.
- 633 39. **Booth DS, Avila-Sakar A, Cheng Y.** 2011. Visualizing Proteins and Macromolecular
634 Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. *J Vis*

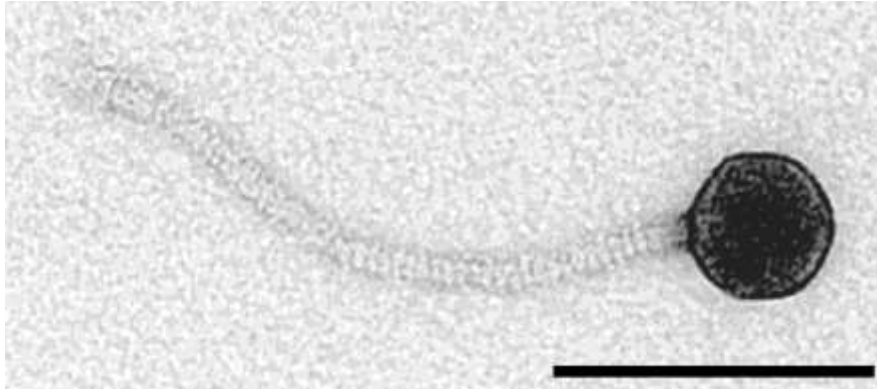
- 635 Exp 1–8.
- 636 40. **Frederick R, Blattner * Guy Plunkett III * Craig A Bloch Nicole T Perna Valerie**
637 **Burland Monica Riley Julio Collado-Vides Jeremy D Glasner Christopher K**
638 **Rode George F Mayhew Jason Gregor Nelson Wayne Davis Heather A**
639 **Kirkpatrick Michael A Goeden Debra J Rose Bob Mau Ying Shao.** 1997. The
640 Complete Genome Sequence of *Escherichia coli* K-12. *Science* (80-) **277**:1453–
641 1462.
- 642 41. **McEwan AR, Raab A, Kelly SM, Feldmann J, Smith MCM.** 2011. Zinc is essential
643 for high-affinity DNA binding and recombinase activity of pc31 integrase. *Nucleic*
644 *Acids Res* **39**:6137–6147.
- 645 42. **Morita K, Morimura K, Fusada N, Komatsu M, Ikeda H, Hirano N, Takahashi H.**
646 2012. Site-specific genome integration in alphaproteobacteria mediated by TG1
647 integrase. *Appl Microbiol Biotechnol* **93**:295–304.
- 648 43. **Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J.** 2009. DNAPlotter:
649 Circular and linear interactive genome visualization. *Bioinformatics* **25**:119–120.
- 650 44. **Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ.** 2009. Jalview
651 Version 2-A multiple sequence alignment editor and analysis workbench.
652 *Bioinformatics* **25**:1189–1191.
- 653 45. **Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H,**
654 **Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T, Higgins D.** 2007.
655 ClustalW and ClustalX version 2. *Bioinformatics* **23**:2947–2948.
- 656 46. **Hall T.** 1999. BioEdit: a user-friendly biological sequence alignment editor and
657 analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**:95–98.
- 658 47. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H,**
659 **Remmert M, Söding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation
660 of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst*
661 *Biol* **7**:539.
- 662 48. **Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S.** 2013. MEGA6: Molecular
663 evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:2725–2729.
- 664 49. **Pearson WR.** 2013. Selecting the right similarity-scoring matrix. *Curr Protoc*
665 *Bioinforma* **43**:3.5.1–9.
- 666 50. **Pei J, Grishin N V.** 2014. PROMALS3D: Multiple protein sequence alignment
667 enhanced with evolutionary and three-dimensional structural information. *Methods*
668 *Mol Biol* **1079**:263–271.
- 669 51. **Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T,**
670 **Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ,**
671 **Hartenstein V, Eliceiri K, Tomancak P, Cardona A.** 2012. Fiji: an open-source
672 platform for biological-image analysis. *Nat Methods* **9**:676–682.
- 673 52. **Monson R, Salmond GP.** 2012. Genome sequence of a new *Streptomyces coelicolor*
674 generalized transducing bacteriophage, PhiCAM. *J Virol* **86**:13860.
- 675 53. **McDonald JE, Smith DL, Fogg PCM, McCarthy AJ, Allison HE.** 2010. High-
676 throughput method for rapid induction of prophages from lysogens and its application
677 in the study of shiga toxin-encoding *Escherichia coli* strains. *Appl Environ Microbiol*,
678 2010/02/09 ed. **76**:2360–2365.
- 679 54. **Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF.** 1999. Evolutionary
680 relationships among diverse bacteriophages and prophages: all the world's a phage.
681 *Proc Natl Acad Sci U S A* **96**:2192–7.
- 682 55. **Rutherford K, Yuan P, Perry K, Sharp R, Van Duyne GD.** 2013. Attachment site
683 recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res*
684 **41**:8341–8356.
- 685 56. **Gupta M, Till R, Smith MCM.** 2007. Sequences in attB that affect the ability of
686 phiC31 integrase to synapse and to activate DNA cleavage. *Nucleic Acids Res*
687 **35**:3407–19.
- 688 57. **Ramaswamy KS, Carrasco CD, Fatma T, Golden JW.** 1997. Cell-type specificity of
689 the *Anabaena* fdxN-element rearrangement requires xisH and xisl. *Mol Microbiol*

- 690 23:1241–1249.
- 691 58. **Savinov A, Pan J, Ghosh P, Hatfull GF.** 2012. The Bxb1 gp47 recombination
692 directionality factor is required not only for prophage excision, but also for phage DNA
693 replication. *Gene* **495**:42–48.
- 694 59. **Zhang L, Ou X, Zhao G, Ding X.** 2008. Highly efficient in vitro site-specific
695 recombination system based on *Streptomyces* phage ϕ BT1 integrase. *J Bacteriol*
696 **190**:6392–6397.
- 697 60. **Rowley PA, Smith MCA, Younger E, Smith MCM.** 2008. A motif in the C-terminal
698 domain of ϕ C31 integrase controls the directionality of recombination. *Nucleic Acids*
699 *Res* **36**:3879–3891.
- 700 61. **Rutherford K, Van Duyn GD.** 2014. The ins and outs of serine integrase site-
701 specific recombination. *Curr Opin Struct Biol* **24**:125–131.
- 702 62. **Hwang WC, Golden JW, Pascual J, Xu D, Cheltsov A, Godzik A.** 2014. Site-
703 specific recombination of nitrogen-fixation genes in cyanobacteria by XisF-XisH-XisI
704 complex: Structures and models. *Proteins Struct Funct Bioinforma* n/a–n/a.
- 705 63. **Lyras D, Adams V, Lucet I, Rood JI.** 2004. The large resolvase TnpX is the only
706 transposon-encoded protein required for transposition of the Tn4451/3 family of
707 integrative mobilizable elements. *Mol Microbiol* **51**:1787–1800.
- 708 64. **Wang H, Mullany P.** 2000. The large resolvase TndX is required and sufficient for
709 integration and excision of derivatives of the novel conjugative transposon Tn5397. *J*
710 *Bacteriol* **182**:6577–6583.
- 711 65. **Wilkinson CJ, Hughes-Thomaz ZA, Martin CJ, Bohm I, Mironenko T, Deacon M,**
712 **Wheatcraft M, Wirtz G, Stanton J, Leadlay PF.** 2002. Increasing the efficiency of
713 heterologous promoter in actinomycetes. *J Mol Microbiol Biotechnol* **4**:417–426.
- 714 66. **Liu H, Naismith JH.** 2009. A simple and efficient expression and purification system
715 using two newly constructed vectors. *Protein Expr Purif* **63**:102–111.
- 716 67. **Fogg MJ, Wilkinson AJ.** 2008. Higher-throughput approaches to crystallization and
717 crystal structure determination. *Biochem Soc Trans* **36**:771–775.
- 718 68. **Coulsox AR, Barrell BG.** 1978. The Nucleotide Sequence of Bacteriophage. *Nucleic*
719 *Acids Res* **16**:355.
- 720 69. **Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ.** 1999. Evidence that the
721 extracytoplasmic function sigma factor σ^E is required for normal cell wall structure in
722 *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**:204–211.
- 723
- 724

725 **Figures**

726

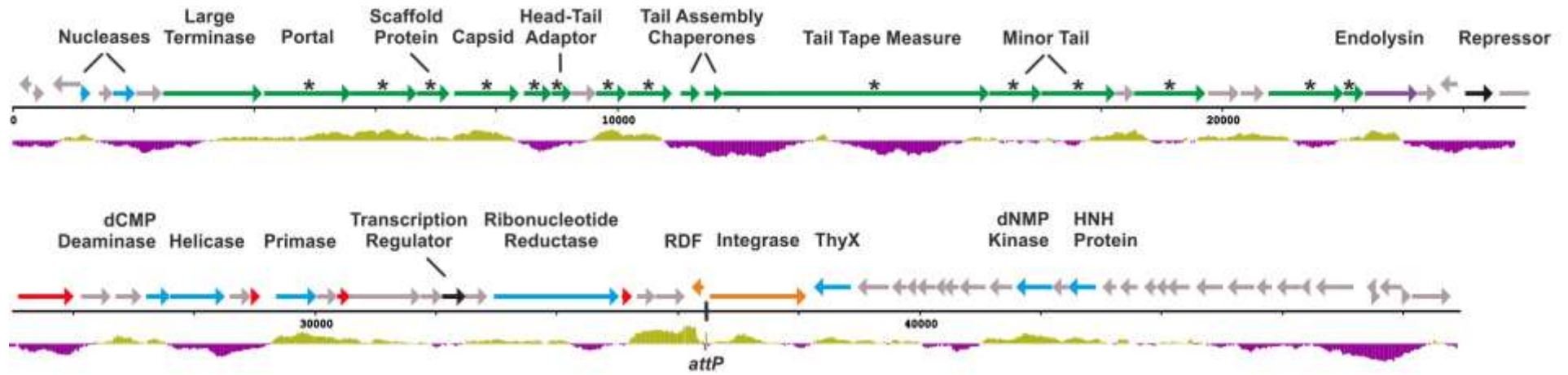
727 **Figure 1: A ϕ Joe virion imaged by transmission electron microscopy.**



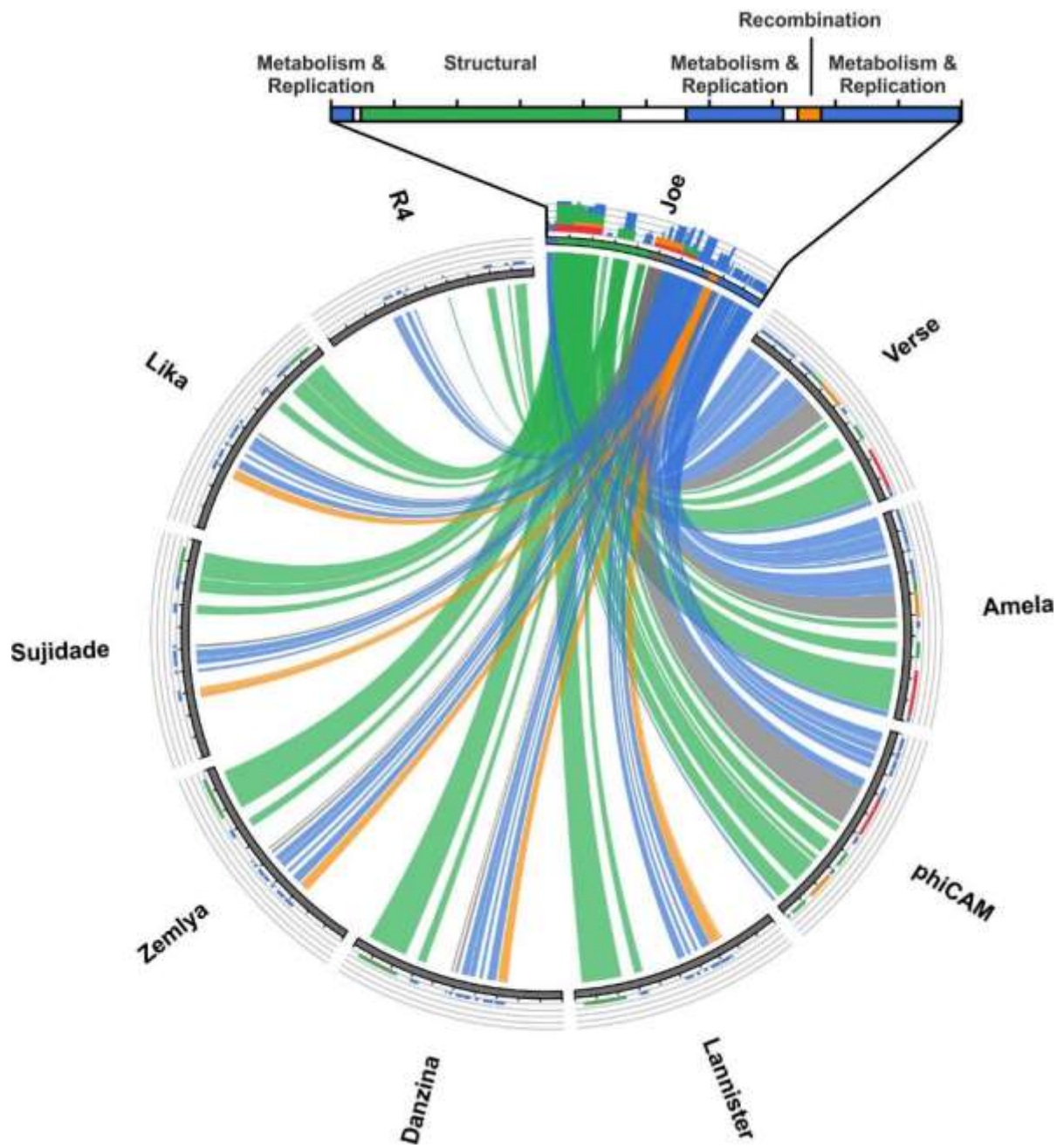
728 **Figure 2: Schematic of the ϕ Joe genome.**

729

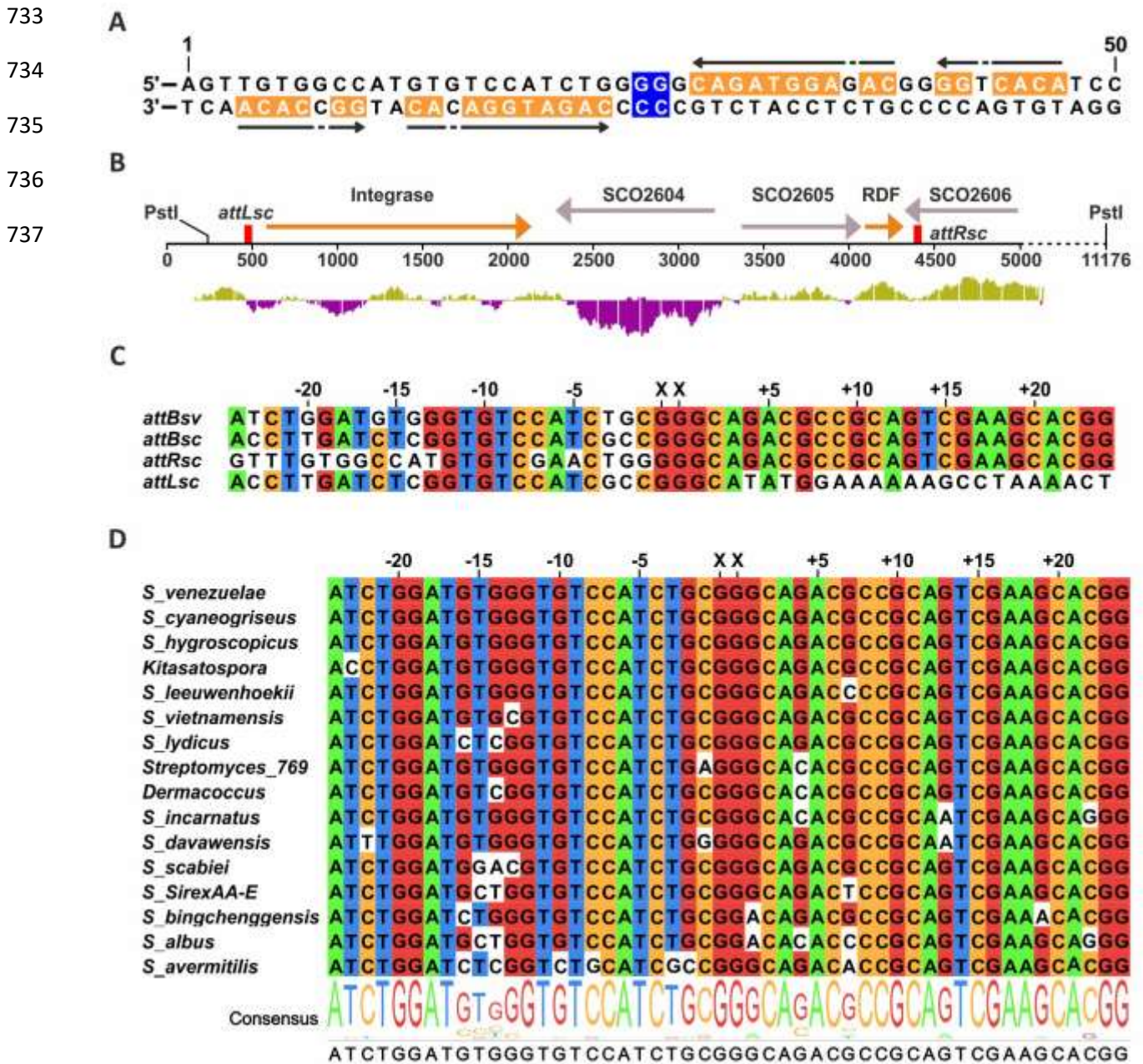
730



731 Figure 3: Circos plot of the ϕ Joe genome versus nine related phages.

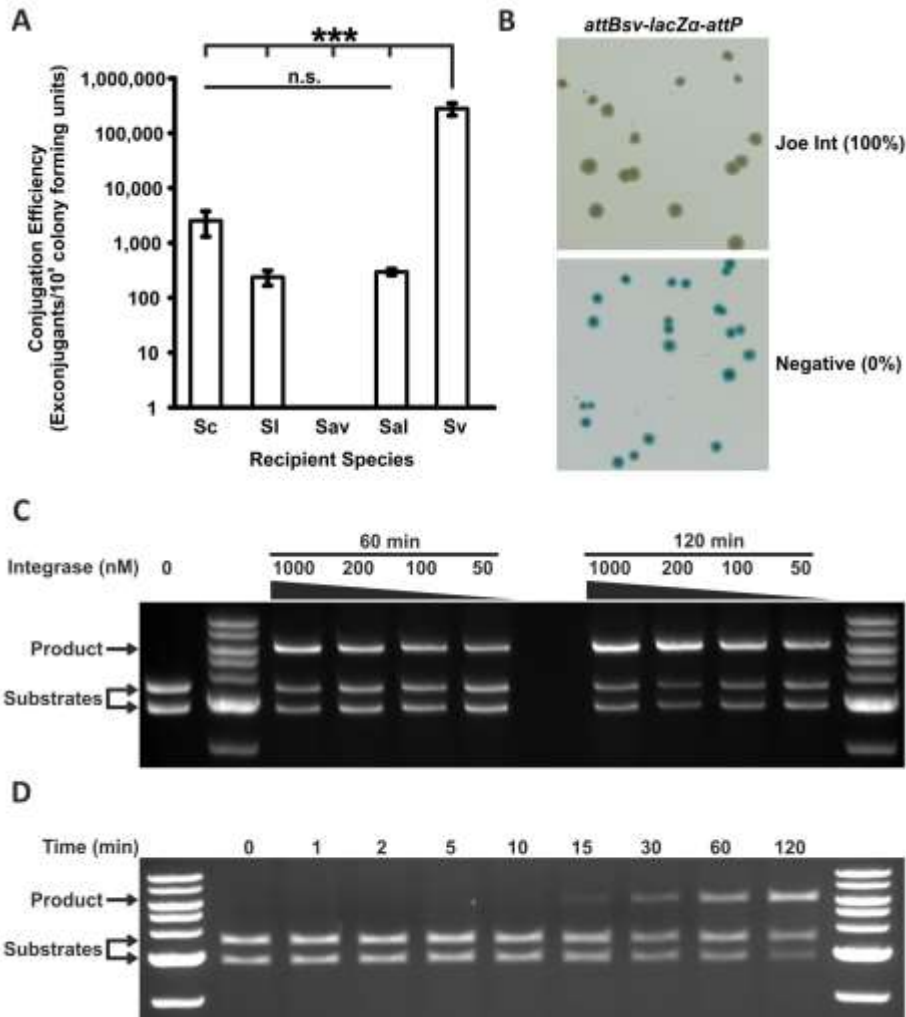


732 **Figure 4: ϕ Joe attachment sites and integration sites.**



738 **Figure 5: Activity of ϕ Joe integrase in vivo and in vitro.**

739



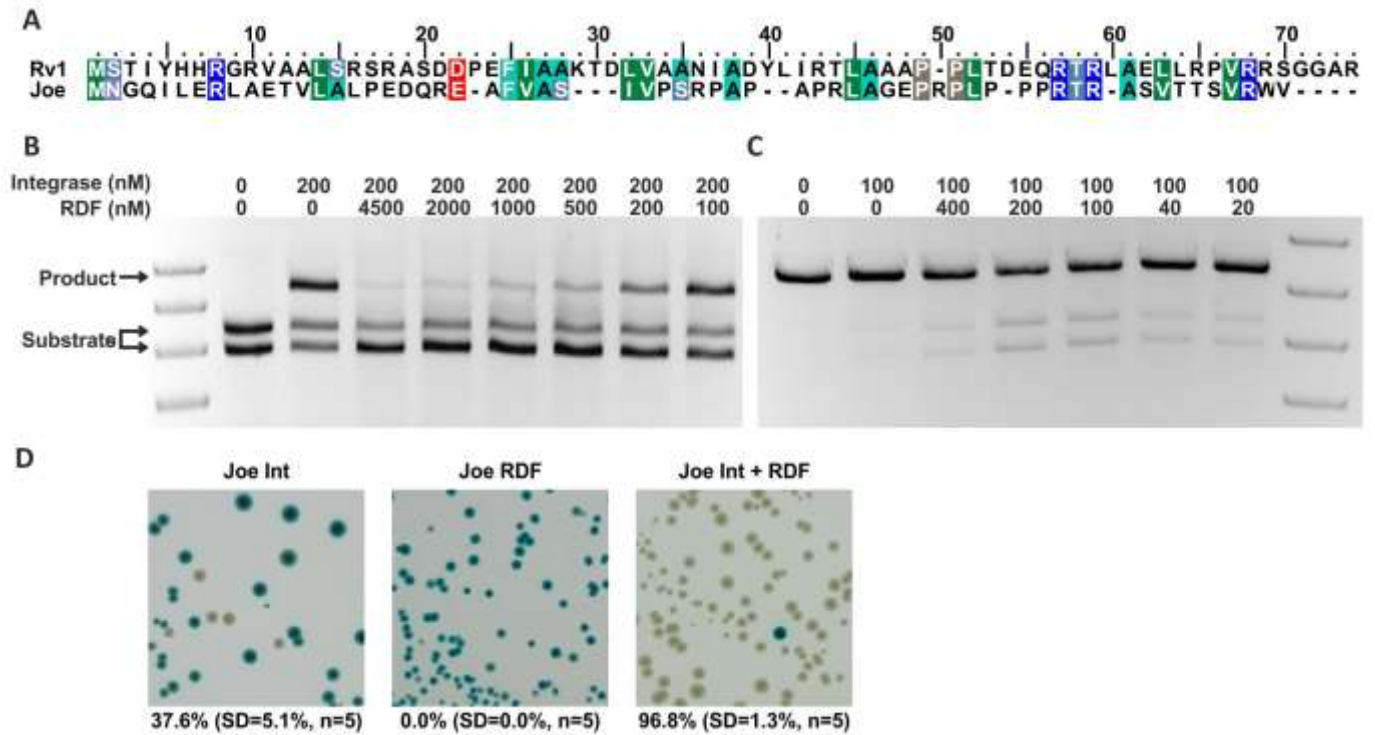
740

741

742 **Figure 6: Identification of the ϕ Joe RDF, gp52.**

743

744



745 **Figure Legends.**

746 **Figure 1: A ϕ Joe virion imaged by transmission electron microscopy.** Viral particles were
747 negatively stained with uranyl acetate and this image was taken at 220,000x magnification.
748 The scale bar represents 100 nm.

749 **Figure 2: Schematic of the ϕ Joe genome.** The genome is 48,941 bp in length. ORFs were
750 predicted using GeneMark and Glimmer then manually curated. ORFs are labelled and colour-
751 coded based on their predicted function. Orange = recombination; cyan = metabolism and
752 DNA processing/replication; green = structural proteins; purple = lysis; black = regulatory; grey
753 = hypothetical proteins with no known function; red = candidate RDF genes. Genes marked
754 with an asterisk encode structural proteins that were detected by tandem MS:MS. The
755 histogram below the genome contains purple bars to indicate below-average GC content
756 (65.5%) and green bars to indicate above-average GC content (1000 nt window size, 20 nt
757 step).

758 **Figure 3: Circos plot of the ϕ Joe genome versus nine related phages.** A Blastn
759 comparison was carried out for ϕ Joe, the five sequenced phage with a ϕ Joe-like integrase,
760 the three closest whole genome matches and the well-characterized R4 phage. The E-value
761 cut-off was set to 1×10^{-100} and the HSPs to 100, ribbons are coloured by genomic regions as
762 defined in Figure 1 and depicted above the Circos plot. The histograms above each genome
763 are coloured to reflect relative homology to the ϕ Joe sequence based on Blast score
764 (Red>Orange>Green>Blue).

765 **Figure 4: ϕ Joe attachment sites and integration sites. A.** Diagram of ϕ Joe *attP* showing
766 the central dinucleotides (Purple) and imperfect inverted repeats (Orange and arrows). **B.**
767 Schematic of the genomic context of the two *S. coelicolor* integration sites (*attLsc* and *attRsc*,
768 red boxes) used by the ϕ Joe integrating plasmid pCMF92. The location of the PstI sites used
769 for identification of the *att* sites are shown. The DNA between the *attLsc* and *attRsc* sites is
770 an apparent mobile genetic element with homologous integrase and RDF genes (orange
771 arrows) to those of ϕ Joe. **C.** Alignment of *S. venezuelae attB (attBsv)* with the two *S. coelicolor*
772 *att* sites (*attRsc* and *attLsc*) and the reconstituted *attB* site (*attBsc*) that would be produced by
773 excision of the DNA between *attRsc* and *attLsc*. **D.** Alignment of closely related *attB* sites
774 identified by a Blastn search against the non-redundant Genbank database. Hits were first
775 filtered for matches of at least 80% and then for an e-value of $<1 \times 10^{10}$ and a bit score >75 .
776 Nucleotide positions in C and D are shown as distance from the crossover dinucleotides (XX).

777 **Figure 5: Activity of ϕ Joe integrase *in vivo* and *in vitro*.** **A.** Conjugation efficiency of an
778 integrating vector, containing ϕ Joe *int* and *attP*, into five recipient species - *Streptomyces*
779 *coelicolor* (**Sc**), *S. lividans* (**Sl**), *S. venezuelae* (**Sv**), *S. albus* (**Sal**) and *S. avermitilis* (**Sav**).
780 Levels of significance for *S. venezuelae* versus all other species in a one-way ANOVA was p
781 $= <0.001$ (3 asterisks), all other comparisons were non-significant (**n.s.**). Error bars are
782 standard deviation (**Sc** n=5, **Sv** and **Sl** n=3, **Sal** and **Sav** n=2). **B.** Representative image of an
783 *in vivo* integration assay to assess *attBsv/attP* recombination by ϕ Joe integrase (pCMF107)
784 and a negative control (pBAD-HisA). Recombination leads to deletion of an intervening *lacZ α*
785 gene and white colonies, inactivity produces blue colonies. Integration efficiency is shown in
786 brackets (n=3). **C.** Representative image of *in vitro* recombination of two substrate plasmids,
787 *attP* (pCMF91) and *attBsv* (pCMF95), to produce the co-integrand plasmid pCMF98. The

788 concentration of ϕ Joe Integrase and incubation time for each reaction is indicated above the
789 gel. **D.** Time-course for the integration reaction shown in part C.

790 **Figure 6: Identification of the ϕ Joe RDF, gp52.** **A.** Alignment of ϕ Joe and RV1 RDFs,
791 coloured using the BLOSUM62 scheme. **B.** Representative agarose gel showing *in vitro*
792 inhibition of integration by ϕ Joe RDF. The concentration of ϕ Joe Integrase and RDF for each
793 reaction is indicated above the image. Reactions were stopped after 2 h and linearized using
794 XhoI. **C.** Representative agarose gel showing *in vitro* excision reactions catalysed by ϕ Joe
795 Integrase and RDF. The concentration of ϕ Joe Integrase and RDF for each reaction is
796 indicated above the image. Reactions were stopped after 2 h and linearized using XhoI. **D.** *In*
797 *vivo* excision assay to assess *attLsv x attRsv* recombination by ϕ Joe integrase alone, ϕ Joe
798 RDF alone and ϕ Joe integrase co-expressed with the RDF. Recombination leads to deletion
799 of an intervening *lacZ α* gene and white colonies, inactivity produces blue colonies. Expression
800 from the T7 promoter successfully achieved almost complete excision activity for ϕ Joe Int +
801 RDF.

802 **Table 1. Plasmids used in this study**

803

804

Plasmid	Description	Resistance	Reference
pSET152	ϕ C31 <i>int</i> + <i>attP</i> integrating vector	Apra	(65)
pEHISTEV	Expression vector, T7 promoter, C-terminal HIS6, TEV cleavage site	Kan	(66)
pETFPP_2	Expression vector; HIS6-MBP-3c Cleavage Site	Kan	(67)
pBAD-HisA	Expression vector, araBAD inducible promoter	Amp	Invitrogen
pCMF87	pEHISTEV + ϕ Joe <i>int</i> (gp53)	Kan	This Study
pCMF90	pGEM7 + <i>S. coelicolor attRsc</i> (274 bp)	Amp	This Study
pCMF91	pSP72 + ϕ Joe <i>attP</i> (354 bp)	Amp	This Study
pCMF92	ϕ Joe <i>int</i> + <i>attP</i> integrating vector; pSET152	Apra	This Study
pCMF94	pGEM7 + <i>S. coelicolor attLsc</i> (419 bp)	Amp	This Study
pCMF95	pGEM7 + <i>S. venezuelae attBsv</i> (462 bp)	Amp	This Study
pCMF96	pETFPP_2 + ϕ Joe MBP-RDF (gp52)	Kan	This Study
pCMF97	pGEM7 + <i>S. coelicolor</i> reconstituted <i>attBsc</i> (152 bp)	Amp	This Study
pCMF98	ϕ Joe <i>attLsv/attRsv</i> ; pCMF91 integrated into pCMF95	Amp	This Study
pCMF100	pEHISTEV + ϕ Joe RDF	Kan	This Study
pCMF103	pACYC184 + ϕ Joe <i>attLsv-lacZα-attRsv</i>	Cm	This Study
pCMF107	pBAD + ϕ Joe <i>int</i>	Amp	This Study
pCMF108	pBAD + ϕ Joe RDF + <i>int</i> co-expression	Amp	This Study
pCMF116	pACYC184 + ϕ Joe <i>attBsv-lacZα-attP</i>	Cm	This Study
pCMF117	pEHISTEV + ϕ Joe RDF + <i>int</i> co-expression	Kan	This Study
pGEM7	General cloning vector	Amp	Promega
pSP72	General cloning vector; Accession X65332	Amp	Promega
pACYC184	General cloning vector; Accession X06403	Cm	(68)
pUZ8002	Conjugation helper plasmid; RK2 derivative with defective oriT	Kan	(69)

805 **Table 2. Primers used in this study**

806

Primer	Sequence (5' – 3')
Joe Int-attP F	CCGTCGACCTGCAGGCATGCCGTTCCCGCAGGTCAGAGC
Joe Int-attP R	ACATGATTACGAATTCTGTGGATCAGAACGTCTCGG
Joe H6-Int F	TTTCAGGGCGCCATGATGAGTAACCGACTACATG
Joe H6-Int R	CCGATATCAGCCATGTCAGAACGTCTCGGCGAAG
Joe attP F	TACCGAGCTCGAATTAAGACCGTCTCAGCCAGG
Joe attP R	TATCATCGATGAATTTTCAGTGAAGACGGACAGG
Joe attB1 F	CCGGGGTACCGAATTTGTGACGTCAGCCACAGC
Joe attB1 R	TAGACTCGAGGAATTTGACAAGGAGTGGCTCTGG
Joe attB2 F	CCGGGGTACCGAATTTGACTGCGTGCCGTCAGCC
Joe attB2 R	TAGACTCGAGGAATTTGCTCGTGTCTGTCTGTGAG
Joe attB Sv F	CCGGGGTACCGAATTACCAGGTGGTGGATGAGC
Joe attB Recon F	TAGACTCGAGGAATTACCTTGATCTCGGTGTCCATCGCCGGGCAGACG CCGCAGTCGAAGCACGG
Joe attB Recon R	CCGGGGTACCGAATTGACAAGGAGTGGCTCTGG
Joe MBP-gp52 F	TCCAGGGACCAGCAATGAACGGACAGATCCTGG
Joe MBP-gp52 R	TGAGGAGAAGGCGCGCTACACCCAGCGCACCGA
CleF	CGCGCCTTCTCCTCACATATGGCTAGC
CleR	TTGCTGGTCCCTGGAACAGAACTTCC
Joe H6-gp52 F	TTTCAGGGCGCCATGAACGGACAGATCCTGGAG
Joe H6-gp52 R	CCGATATCAGCCATGCTACACCCAGCGCACCGA
Joe pBAD Int F	GAGGAATTAACCATGAGTAACCGACTACATG
Joe pBAD Int R	TGAGAACCCCCCATGTCAGAACGTCTCGGCGAAG
Joe pBAD gp52 F	GAGGAATTAACCATGAACGGACAGATCCTGGAG
Joe pBAD Int Co-Ex F	AGTGGTAGGTTCTCGCCATG
Joe pBAD gp52 R	GAGGAACCTACCACTCTACACCCAGCGCACCGA
Joe LzR F	GGGTGTCAGTGAAGTAGTTGTGGCCATGTGTCCATCTGGGGGCAGACG CCGCAGTCGAAGCACGGCGATTTTCGGCCTATTGGT
Joe LzR R	CCTGCCACATGAAGCGGATGTGACCCCGTCTCCATCTGCCCGCAGATG GACACCCACATCCAGATAATACGCAAACCGCCTCT
Joe BzP F	GGGTGTCAGTGAAGTATCTGGATGTGGGTGTCCATCTGCGGGCAGACG CCGCAGTCGAAGCACGGCGATTTTCGGCCTATTGGT
Joe BzP R	CCTGCCACATGAAGCGGATGTGACCCCGTCTCCATCTGCCCCAGATG GACACATGGCCACAATAATACGCAAACCGCCTCT
SPBc H6-sprA F	CCGATATCAGCCATGGAGTTAAAAACATTGTT
SPBc H6-sprA R	TTTCAGGGCGCCATGCTTACTACTTTTCTTAGTGG
SPBc MBP-sprB F	TCCAGGGACCAGCAATGGAACCTTACCAACGT
SPBc MBP-sprB R	TGAGGAGAAGGCGCGAAGCTTACTCTGCCTTCC
SPBc LZR F	GGGTGTCAGTGAAGTAGTGCAGCATGTCATTAATATCAGTACAGATAAA GCTGTATATTAAGATACTTACTACATATCTACGATTTTCGGCCTATTGGT
SPBc LZR R	CCTGCCACATGAAGCTGGCACCCATTGTGTTTACAGGAGATACAGCTTT ATCTGTTTTTTAAGATACTTACTACTTTTCTAATACGCAAACCGCCTCT

807

808 **Table 2. ϕ Joe Host Range.**

Host Species	Lysis (pfu ⁸⁰⁹) [^]
<i>Streptomyces albus</i> J1074	(2x10 ⁹) ⁸¹⁰
<i>Streptomyces avermitilis</i>	(2x10 ⁹) ⁸¹⁰
<i>Streptomyces coelicolor</i> J1929	(2x10 ⁸) ⁸¹¹
<i>Streptomyces coelicolor</i> M145	✓ ⁸¹¹
<i>Streptomyces griseus</i>	(4x10 ⁸) ⁸¹²
<i>Streptomyces lividans</i> TK24	(7x10 ⁷) ⁸¹²
<i>Streptomyces nobilis</i>	(1x10 ⁴) ⁸¹³
<i>Streptomyces scabies</i>	(6x10 ⁷) ⁸¹³
<i>Streptomyces venezuelae</i>	X ⁸¹⁴
<i>Streptomyces venezuelae</i> VL7	X
<i>Streptomyces venezuelae</i> VS1	X ⁸¹⁵
<i>Streptomyces venezuelae</i> 10712	X
<i>Saccharopolyspora erythraea</i>	X ⁸¹⁶

817

818 [^] Pfu/ml values quoted are illustrative of the relative plaquing efficiencies when
 819 challenged with the same phage stock propagated on *S. coelicolor* J1929