RESEARCH LETTER

A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species

Trine Aakvik[1], Kristin Fløgstad Degnes[2], Rannveig Dahlsrud[1], Frank Schmidt[3], Ragnar Dam[3], Lihua Yu[2], Uwe Völker[3], Trond Erling Ellingsen[2] & Svein Valla[1]

[1]Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway; [2]Department of Biotechnology, SINTEF Materials and Chemistry, SINTEF, Trondheim, Norway; and [3]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany

Correspondence: Svein Valla, Department of Biotechnology, Norwegian University of Science and Technology, 7491 Trondheim, Norway. Tel.: +47 735 933 20; fax: +47 73 591 283; e-mail: svein.valla@biotech.ntnu.no

Received 26 March 2009; accepted 27 March 2009.
Final version published online 20 May 2009.

DOI:10.1111/j.1574-6968.2009.01639.x

Editor: Marco Soria

Keywords
metagenome; marine; RK2; vector; broad host range.

Abstract

The majority of microorganisms in natural environments are difficult to cultivate, but their genes can be studied via metagenome libraries. To enhance the chances that these genes become expressed we here report the construction of a broad-host-range plasmid vector (pRS44) for fosmid and bacterial artificial chromosome (BAC) cloning. pRS44 can be efficiently transferred to numerous hosts by conjugation. It replicates in such hosts via the plasmid RK2 origin of replication, while in *Escherichia coli* it replicates via the plasmid F origin. The vector was found to be remarkably stable due to the insertion of an additional stability element (*parDE*). The copy number of pRS44 is adjustable, allowing for easy modifications of gene expression levels. A fosmid metagenomic library consisting of 20 000 clones and BAC clones with insert sizes up to 200 kb were constructed. The 16S rRNA gene analysis of the fosmid library DNA confirmed that it represents a variety of microbial species. The entire fosmid library and the selected BAC clones were transferred to *Pseudomonas fluorescens* and *Xanthomonas campestris* (fosmids only), and heterologous proteins from the fosmid library were confirmed to be expressed in *P. fluorescens*. To our knowledge no other reported vector system has a comparable potential for functional screening across species barriers.

Introduction

Current estimates suggest that the vast majority of the microorganisms in natural environments cannot be cultivated (Torsvik *et al.*, 1990; Amann *et al.*, 1995; Torsvik & Ovreas, 2002). An enormous source of genetic information therefore remains undiscovered even after extensive screening based on standard cultivation methods. To explore these sources, novel approaches have been developed that involve direct isolation and cloning of DNA from environmental samples into suitable vectors, thus creating complex metagenomic libraries (Handelsman *et al.*, 1998). Such libraries have been made from both soil and marine environments (reviewed in Daniel, 2005; DeLong, 2005). In addition, Venter *et al.* (2004) reported in 2004 the first example of the use of the 'whole-genome shotgun sequencing' approach to marine microbial populations collected from the Sargasso Sea.

The vectors most often used for large insert metagenomic studies are the *Escherichia coli* F-factor-based vectors for bacterial artificial chromosome (BAC) and fosmid (packing by λ phage) cloning (Kim *et al.*, 1992; Shizuya *et al.*, 1992). Metagenomic libraries can be analysed for novel genes and pathways with sequence-based techniques or through activity screening involving analyses of expression of novel phenotypic traits in surrogate hosts. The advantage of such functional screening approaches is that they can detect activities originating from genes whose functions cannot be predicted by bioinformatic analyses of DNA or protein sequences. On the other hand, identification of novel activities by functional screening is dependent on successful transcription and translation of the cloned genes. Even though novel activities have been expressed using *E. coli* as host, there is a potential advantage of expanding the range of bacterial hosts to capture additional expression capabilities. This was demonstrated by Gabor *et al.* (2004) in a study showing that only 40% of the genes from the genomes of 32 prokaryotes could be detected when expressed in *E. coli*. The study also revealed significant differences in the predicted

expression modes between distinct taxonomic groups of organisms. Another study by Martinez *et al.* (2004) showed that *E. coli*, *Pseudomonas putida* and *Streptomyces lividans* differed in their abilities to express heterologous gene clusters.

To exploit the potential involved in using many different hosts for expression of genes obtained from environmental samples new biological tools (vectors) are needed. Ideally, it should be possible to efficiently transfer entire libraries with large inserts from *E. coli* (where they are constructed) to many different hosts, and the vectors should be stably maintained in these hosts. BAC vectors that are capable of being transferred to one or a few hosts in addition to *E. coli* have been described (Sosio *et al.*, 2000; Martinez *et al.*, 2004; Hain *et al.*, 2008), but they all share limitations with respect to either efficiency of transfer, lack of autonomous replication or host range. RK2 replicons potentially eliminate these types of limitations, and large-sized poorly characterized cosmid-RK2-derivatives have been used for metagenomic library constructions (Wexler *et al.*, 2005).

Here, we report the development of a relatively small and functionally well-understood combined BAC and fosmid vector based on the RK2 replicon. It is generally known that vectors constructed from this replicon function in numerous Gram-negative bacterial species (Thomas & Helinski, 1989), and have even been transferred to Gram-positive bacteria, yeast and mammalian cells (Poyart & Trieu-Cuot, 1997; Bates *et al.*, 1998; Waters, 2001). An important advantage is that libraries constructed in this vector can be conjugatively transferred from *E. coli* to a number of hosts with an efficiency that ensures no significant loss of representativity relative to the library present in the *E. coli* donor. Another advantage of the RK2 replicons is that their copy numbers can be manipulated across species barriers using well-defined mutants in the essential replication initiation and plasmid copy number control gene *trfA* (Haugan *et al.*, 1995). As expression normally correlates with gene dosage, this feature may represent a very important advantage with respect to detection of activities from genes expressed at low levels.

Materials and methods

Bacterial strains, plasmids and growth media

The bacterial strains and plasmids used in this study are described in Table 1. *Escherichia coli* strains were grown in Luria–Bertani (LB) medium or on LB-agar at 37 °C. *Pseudomonas fluorescens* and *Xanthomonas campestris* strains were grown at 30 °C in LB or on Difco Pseudomonas Isolation Agar (PIA) (*P. fluorescens*) and in Difco Yeast Mold (YM) broth or on YM agar (*X. campestris*). Antibiotics (when relevant) were used at the following concentrations: chloramphenicol, 12.5 μg mL^{-1} (*E. coli*), 30 μg mL^{-1} (*X. campestris*); kanamycin, 50 μg mL^{-1} (*E. coli* and *X. campestris*); tetracycline 10 μg mL^{-1} (*E. coli*), 15 μg mL^{-1} (*X. campestris*) or 25 μg mL^{-1} (*P. fluorescens*). 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside and isopropyl-β-D-thiogalactoside were used for blue–white selection of clones with inserts in *E. coli*. Clones in EPI300 were switched from single copy to high-copy number by L-arabinose induction, using the solution from the Copy Control Fosmid Library Production Kit, Epicentre. Expression of *trfA* from *PmG5* was induced by addition of *m*-toluate at 0.5 mM.

Standard DNA manipulations

Agarose gel electrophoresis, and routine DNA manipulations were performed according to the methods of Sambrook & Russel (2001), or using commercially available kits. DNA sequencing was performed using the Big-Dye Terminator v1.1 Cycle kit (Applied Biosystems) or by MWG-Biotech AG.

Southern hybridization analysis was conducted with the DIG-High Prime DNA Labeling and DIG Nucleic Acid Detection Kit (Roche Diagnostics) according to the supplier's instructions.

Vector constructions, transformation and conjugative matings

pRS44 was constructed by introducing the *parDEoriT* region and a kanamycin resistance gene to pCC1FOS. The entire nucleotide sequence of this vector is known. The suicide transposon vector pRS48 was constructed by cloning *trfA* and the tetracycline resistance gene into pKD20, using *E. coli* S17.1(λ*pir*) as host. For further details, see Table 1.

The transposon in pRS48 was inserted into the chromosomes of electrocompetent *P. fluorescens* NCIMB10525 and *X. campestris* B100-152 by standard electroporation (13 V cm^{-1}, 100Ω, 25 μF), as described for *E. coli* by Sambrook & Russel (2001), and the transformants designated NCIMB10525::Tn*RS48* and B100-152::Tn*RS48* were selected on PIA or YM agar containing tetracycline. pRS44-derived clones were transformed into *E. coli* S17.1 by standard heat-shock transformation (Chung *et al.*, 1989).

Conjugative matings were performed on LB-agar without antibiotic selection overnight at 30 °C. The mixtures were then plated on PIA with *m*-toluate and kanamycin (*P. fluorescens)* or on YM agar with *m*-toluate, tetracycline and chloramphenicol (*X. campestris)*, followed by incubation at 30 °C for 48 and 72 h, respectively.

Plasmids were isolated from cultures of *P. fluorescens* and *X. campestris* by Qiagen Plasmid Midi Kit (Qiagen) and electroporated into EPI300 (13 V cm^{-1}, 100Ω, 25 μF).

Construction of the metagenomic fosmid library

Marine sediment samples were collected from the intertidal zone at the Trondheims fjord. DNA was extracted from five

Table 1. Bacterial strains and plasmids used in this study

Bacterial strains or plasmids	Properties	Sources or references
E. coli		
ElectroMAX DH10B	F-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K ë-*rps*L *nup*G	Invitrogen
EPI300	Phage T1-resistant and *lacZ*$^-$ strain with L-arabinose-induced chromosomally expressed TrfA, (F$^-$ *mcrA* Δ*(mrr-hsdRMS-mcrBC)* φ*80dlacZ*Δ*M15* Δ*lacX74 recA1 endA1 araD139* Δ*(ara, leu)7697 galU galK* λ$^-$ *rpsL nupG trfA tonA dhrf*)	Epicentre
S17.1	Strain with the RK2 *tra* genes for conjugative transfer integrated in the chromosome *(RP4-2-Tc::Mu-Km::Tn7*, *pro, res*$^-$ *mod*$^+$, Tp^r Sm^r)	Simon *et al*. (1983)
S17.1(λ*pir)*	λ*pir* lysogen of strain S17.1	De Lorenzo *et al*. (1993)
Pseudomonas fluorescens		
NCIMB 10525	*P. fluorescens* wild type	NCIMB
NCIMB10525::Tn*RS48*	Derivative of NCIMB 10525 with transposon Tn*RS48* from pRS48 integrated into the chromosome	This work
Xanthomonas campestris		
B100-152	Spontaneous *xanA* exopolysaccharide-negative mutant	Hötte *et al*. (1990)
B100-152::Tn*RS48*	Derivative of B100-152 with transposon Tn*RS48* from pRS48 integrated into the chromosome	This work
Plasmids		
B9	pTA44 BAC clone with a 130-kb DNA fragment originating from the plant *Ipomoea nil* cloned into the HindIII site, Cm^r, Km^r	This work
B19	pTA44 BAC clone with a 130-kb DNA fragment originating from the plant *Ipomoea nil* cloned into the HindIII site, Cm^r, Km^r	This work
Litmus28	*ColE1* replicon, Ap^r, 2.8 kb	New England BioLabs
pCC1FOS	Cloning vector for Copy Control Fosmid library containing *ori2* and *oriV*, *parABC*, *cos* and *loxP*, Cm^r, 8.1 kb	Epicentre
pDrive	Cloning vector for direct ligation of pcr products, *colEI* replicon, Ap^r, Km^r, 3.9 kb	Qiagen
pHH100G5	RK2-based plasmid containing a mutant *Pm* promoter designated *PmG5*, Km^r, 8.8 kb	Gimmestad *et al*. (2004)
pIndigoBAC5	BAC cloning vector containing *ori2*, *parABC*, *cos* and *loxP*, Cm^r, 7.5 kb	Epicentre
pJB321	RK2 minimal replicon containing *parDE*, Ap^r, 5.6 kb	Blatny *et al*. (1997a)
pJB658	RK2 expression vector containing the *Pm* promoter and the gene encoding the regulatory protein XylS and TrfA, Ap^r, 6.8 kb	Blatny *et al*. (1997b)
pKD20	Suicide vector encoding a mini-Tn*5* transposon with NdeI/NotI cloning sites for introduction of genes under *PmG5*/*xylS* control, Ap^r, Km^r, 8.0 kb	Bakkevig *et al*. (2005)
pLitmusTcBam2	Litmus28 derivative containing *tetA* and *tetR*, Ap^r, Tc^r, 5.1 kb	Bakkevig *et al*. (2005)
pRS43	Derivative of pCC1FOS with an insertion of a 1.2-kb PCR fragment encoding *parDEoriT* from pJB321, Cm^r, 9.3 kb	This work
pRS44	Derivative of pRS43 with an insertion of a 1.0-kb PstI/SacI PCR fragment encoding Km^r from pHH100G5 cloned in the HpaI site, Cm^r, Km^r, 10.3 kb	This work
pRS47	*trfA* from pJB658 cloned as a 1.2-kb NdeI/NotI PCR fragment into pKD20, Ap^r, Km^r, 9.1 kb	This work
pRS48	Derivative of pRS47 with a 2.2-kb BamHI fragment containing *tetA* and *tetR* from pLitmusTcBam2 cloned in the DraIII/SacI sites, Ap^r, Tc^r, 10.5 kb	This work
pRS49	pRS44 with Fosmid Control DNA (Epicentre) cloned in the Eco27I-site, Cm^r, Km^r, 46.3 kb	This work
pTA44	Derivative of pRS44 where the HindIII site in the Km^r gene is removed by site specific mutagenesis, Cm^r, Km^r, 10.3 kb	This work
RK2	Ap^r, Km^r, Tc^r, 60.1 kb	Pansegrau *et al*. (1994)

*Ap^r, ampicillin resistance; Cm^r, chloramphenicol resistance; Km^r, kanamycin resistance; Tc^r, tetracycline resistance.

parallels, each containing 5 g of sediment soil according to the sodium dodecyl sulphate (SDS)-based method described by Zhou *et al.* (1996), with the modification that the shaking treatment of the samples was carried out at 125 r.p.m. instead of 225 r.p.m. At the end, a total of about 85 μg of DNA was obtained, and the visualized (agarose gel electrophoresis) size of the DNA seemed to cover a range between 15 and 100 kb with the highest intensity around 50 kb. The isolated DNA was size selected (35–40 kb) and purified through two steps of agarose gel electrophoresis, the latter in a low-melting-point (LMP) gel followed by extraction of the DNA with GELase enzyme according to the recommendations of the manufacturer (Epicentre). The environmental DNA was at no point subjected to ethidium bromide or UV light, and this was handled as described in Rondon *et al.* (2000).

Preparation of pRS44 vector DNA was carried out as follows: four parallels of 4 μg pRS44 vector DNA were digested with 50 U Eco72I in reaction volumes of 1 mL for 5 h at 37 °C. Consecutive dephosphorylation was performed by adding 50 U CIP to each reaction followed by further incubation at 37 °C for 2 h. The DNA was isopropanol precipitated and resuspended in a total volume of 60 μL sterile water. To remove any potential still phosphorylated DNA, the DNA was treated with 1 μL Fast-Link DNA Ligase (from kit, Epicentre) for 2 h and purified on an LMP gel. Linear DNA was recovered from the gel with GELase enzyme, as above (Epicentre).

The fosmid library with environmental insert-DNA was constructed according to the protocol of the Copy Control Fosmid Library Production kit (Epicentre). In the ligation reaction, 4 μg of the insert DNA was mixed with 3 μg of vector DNA in a reaction volume of 100 μL. Ten microlitres of this ligation mixture was further used for each packaging reaction, as described in the Epicentre protocol. Approximately, 20 000 colonies of the obtained fosmid library were robotically picked and gridded into 384 well plates (Genetix QPixII robotic colony picker), and the library was also stored as a pooled freeze stock.

PCR amplification of rRNA fragments and denaturing gradient gel electrophoresis (DGGE)

The 16S rRNA genes were amplified by PCR using the universal bacterial primer Bac341f with GC clamp (Muyzer *et al.*, 1993) and Bac907r (Teske *et al.*, 1996) with DNA extracted from a mixed culture containing the entire fosmid library as template. The PCR amplifications were performed as described by Burr *et al.* (2006).

The PCR products were analysed by DGGE, using the Ingeny phorU 2 × 2 apparatus (GRI Molecular Biology). Samples (10 μL) were loaded onto 8% polyacrylamide/bis (37.5 : 1) gels with denaturing gradients from 35% to 55% in 1 × TAE electrophoresis buffer. Electrophoresis was performed at 100 V, 60 °C for 18 h. Gels were then stained with SYBR Gold (Cambridge BioScience) in 1 × TAE for 1 h at room temperature and visualized under UV illumination.

Amplified 16S rRNA gene was cloned into the pDrive cloning vector, and sequenced with M13 primer pairs.

Proteomic analysis of fosmid-encoded proteins

Pseudomonas fluorescens containing fosmid clones with environmental inserts or the corresponding vector without insert were selected for analysis by two-dimensional gel electrophoresis (2-DE). LB medium was inoculated with 0.2‰ stationary-phase cultures and incubated to an $OD_{600\,nm}$ of 0.8. Twenty-five ODU were harvested by centrifugation (15 min at 9000 ***g*** and 5 °C). The pellets were washed twice with 4 mL HEPES/PMSF buffer (20 mM HEPES, 0.3 mg mL^{-1} PMSF, pH 7.5), followed by centrifugation at 3500 ***g*** for 6 min at 5 °C.

The pellets were resuspended in 8 M urea and 2 M thiourea and disrupted by sonication (5 × 30 s), after which the lysates were cleared by centrifugation at 16 000 ***g*** and 4 °C for 30 min. The supernatants were centrifuged again at 16 000 ***g*** and 4 °C for 10 min. The protein concentrations were determined according to Bradford (1976) and the crude protein extracts were stored at − 80 °C.

Isoelectric focusing was performed with commercially available 24-cm Immobilized pH gradient (IPG) strips in the pH range 4–7 (Amersham Biosciences, Freiburg, Germany). IPG strips were loaded with 100 μg of crude protein extract by rehydration for 24 h in a solution containing 8 M urea (Merck, Darmstadt, Germany), 2 M thiourea (Sigma, St. Louis, MO), 1% w/v CHAPS (Roth, Karlsruhe, Germany), 20 mM DTT (ICN-Biomedicals) and 0.5% v/v pharmalyte 3–10 (Amersham Biosciences, Piscataway, NJ). IEF using the Multiphor II unit (Amersham Biosciences) and SDS-polyacrylamide gel electrophoresis (PAGE) using the Protean Dodeca Cell (BioRad, Munich, Germany) were performed as described previously (Thiele *et al.*, 2007). The resulting 2-DE gels were fixed with 50% v/v ethanol and 12% v/v acetic acid for 1–2 h, subsequently stained with silver nitrate according to Blum *et al.* (1987) and scanned with an HP Scanjet Scanner. 2-DE analyses were conducted in duplicate.

Protein spots exclusively present in both duplicates of individual cultures harbouring insert-containing vectors were identified using the DELTA-2D image analysis software (Decodon GmbH, Greifswald, Germany), manually excised from silver-stained 2-DE gels and destained using 30 μL of 15 mM $K_3[Fe(CN)_6]$ 100 mM $Na_2S_2O_3$ for 30 min at 37 °C. Then, gel slices were washed threefold with Lichrosolv water and once with 100 μL 200 mM NH_4HCO_3/50% acetonitrile (ACN) for 15 min at 37 °C, dehydrated with ACN and

digested in a trypsin solution (10 ng mL^{-1} of 20 mM NH_4HCO_3/50% ACN) overnight at 37 °C.

For peptide extraction, gel pieces were covered with 20 μL 0.1% acetic acid and subsequently 15 μL 0.05% acetic acid/50% ACN and incubated in an ultrasonic water bath for 30 min. The eluate of both extractions was combined and concentrated in a vacuum centrifuge. Finally, the samples were dissolved in 12 μL buffer A for MS.

All experiments were performed on a Proxeon nano-LC system (Odense, Denmark) connected to a linear quadrupole ion trap – Orbitrap (LTQ Orbitrap) mass spectrometer (ThermoElectron, Bremen, Germany) equipped with a nano-ESI source. For liquid chromatography separation we used a PepMap 100 column (C18, 3 μm, 100 Å) (Dionex, Sunnyvale CA) capillary of 15 cm bed length. The flow rate used was 300 nL min^{-1} for the nanocolumn, and the solvent gradient used was 2% solvent B to 50% solvent B in 30 min. Solvent A was 0.1% formic acid, whereas aqueous 100% acetonitrile in 0.1% formic acid was used as solvent B. The mass spectrometer was operated in a data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Raw data were processed using the TPP pipeline (v. 1.4) available online from ISB (Seattle) to generate MASCOT generic files (mgf). Afterwards, a database search was performed by tandem mass spectrometry ion search algorithms from the Mascot house (v2.2.1) by database comparisons with *P. fluorescens* SBW25 FASTA and the SWISSPROT (release 55.5 all entries) sequences with 10 p.p.m. tolerance for the precursor and 0.3 Da for the MS/MS fragments. Moreover, trypsin was selected and variable modifications, such as methionine oxidation and carbamidomethylation of cysteine were allowed. Peptides were considered to be identified by Mascot when probability < 0.05 (probability based ion scores threshold > 41) was achieved. Proteins were considered to be identified if at least two peptides were identified.

Construction of BAC clones

BAC clones with larger inserts than those present in the fosmid library were constructed by the company Bio S&T Inc. (Canada). In this cloning, a slightly modified derivative of pRS44 (denoted pTA44) was used as vector. In pTA44 the HindIII site in the kanamycin resistance gene has been removed by site-specific mutagenesis, facilitating the use of the remaining HindIII site for insertion of heterologous DNA. As cloning material DNA from the plant *Ipomoea nil* was used. Nuclei were prepared from 100 g leaves and embedded in 5 mL 2% (w/v) LMP agarose plugs as described by Zhang *et al.* (1995). After partial digestion with HindIII, the DNA was size selected using pulse field gel electrophoresis (PFGE), and the DNA in the range of 50–350 kb was electro-eluted from the gel (by PFGE). In reactions of 50 μL, 100 ng of this insert DNA was ligated to 20 ng of HindIII-digested pTA44- and pIndigoBAC vector DNA (the latter for comparison of cloning efficiency). Two microlitres of the ligation mixtures were subsequently electro-transformed into 20 μL of ElectroMax DH10B cells, and a selection of the resulting white colonies were picked for further analysis.

Results and discussion

Construction of a broad-host-range metagenome vector system

The broad-host-range fosmid and BAC vector (pRS44, Fig. 1) was constructed using the commercially available pCC1FOS vector as a starting point. pCC1FOS has two origins of replication, *ori2* from the F plasmid and *oriV* from RK2. *ori2* functions in *E. coli* and is active during construction of libraries in this host, while it is not active in most other hosts. In pCC1FOS *oriV* is included to be able to produce large quantities of vector DNA in a particular *E. coli* host. This host carries a chromosomally integrated mutant version (copy-up) of the gene encoding the replication initiation and copy number control protein TrfA that can be activated by expression from the inducible *pBAD* promoter. Instead here we use *oriV* as a tool for replication of metagenome libraries in non-*E. coli* hosts. In addition we have inserted the origin of conjugative transfer (*oriT*) to allow efficient transfer of the libraries from *E. coli* to such alternative hosts. This is a very important feature because it is well known that the efficiency of conjugational transfer is usually much better than transformation of naked DNA. We predicted that vector stability could become an important feature and we therefore also inserted the stabilization element *parDE* from RK2, because it is known that *parDE* stabilizes RK2 vectors across species barriers (Sia *et al.*, 1995; Blatny *et al.*, 1997a). Finally, the kanamycin resistance gene was inserted to provide an alternative selection marker, potentially useful in some hosts. The BamHI and HindIII (pTA44, see legend to Fig. 1) sites are useful for sticky-end BAC cloning, while the Eco72I site is used for blunt-end fosmid cloning. The *lac* system for blue–white screening was kept from pCC1FOS.

To allow replication of pRS44 in new hosts, expression of *trfA* (we used wild type) is needed. However, we decided not to include *trfA* as part of pRS44, as such a vector would contain two active replication systems in *E. coli*, possibly leading to plasmid instability. Instead we constructed a suicide vector (pRS48, Fig. 1), which can be used to insert a derivative of transposon Tn*5* expressing the TrfA protein from the inducible *PmG5* promoter. This promoter is a mutant derivative of wild-type *Pm,* which is known to be active in many hosts (Mermod *et al.*, 1986; Ramos *et al.*, 1988; Keil & Keil, 1992). The inducibility in addition allows

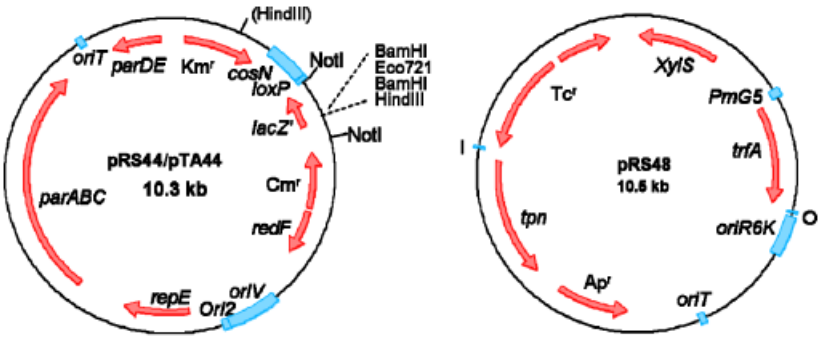

Fig. 1. Plasmid vectors for use in genomic library constructions (pRS44/pTA44) and for support of vector replication in hosts other than *Escherichia coli* (pRS48). In pTA44 the HindIII site in the Km^r gene is removed, making it possible for cloning of HindIII-digested DNA into the remaining HindIII site. pRS44/pTA44 can replicate as a single copy replicon via *ori2* and *repE*, while *oriV* contributes to a medium copy number if its replication initiation protein TrfA is expressed in the same cell. pRS44/pTA44 DNA can easily be prepared in large quantities in the *E. coli* strain EPI300 by expressing a mutant *trfA* gene from an arabinose-induced promoter, as described by Wild *et al*. (2002). *cosN* is the site used for packaging of the environmental DNA library in bacteriophage λ particles, BamHI and Eco72I sites are used for BAC and fosmid cloning respectively, and NotI is suitable for size determination of the inserts. The *trfA*-gene is inserted into the chromosome of hosts of interest by the transposon present in the narrow-host-range plasmid pRS48. The inside and the outside ends of the transposon (designated Tn*RS48*) are marked I and O, respectively. *tpn*, gene encoding the transposase, which is not a part of the transposon; *XylS*, gene encoding activator of *PmG5* transcription in the presence of benzoic acid-type inducers, like *m*-toluate; *OriT*, origin of conjugative transfer. For further details see Table 1 and the text.

for modification of the amount of TrfA produced. pRS48 replicates in the *E. coli* strain S17.1(λ*pir*), which expresses the Pir protein needed for replication initiation of the plasmid R6K origin, *oriR6K* (De Lorenzo *et al*., 1993).

Plasmid stability in the absence of selection

Plasmid stability may potentially become critical for the functioning of the metagenome cloning vector described here, and to quantify this we measured the rate by which it became lost in the absence of antibiotic selection in *E. coli* EPI300 (Fig. 2). As controls in this experiment we used the native RK2 plasmid and pCC1FOS. By the repeated transferring, growth was monitored over about 230 generations, a number which enormously exceeds the number of generations taking place in laboratory-scale batch cultures. The experiments showed that plasmid loss could easily be detected for pCC1FOS, while both pRS44 and a derivative of it containing a 36-kb control DNA insert (pRS49) were remarkably stable, like whole RK2. This experiment therefore clearly confirmed the relevance of introducing *parDE* into the vector.

Construction of a metagenomic library in pRS44

Initial experiments showed that the standard 36 kb insert used as a control for the commercially available vectors was packaged and established in *E. coli* at similar frequencies for pRS44 as for pCC1FOS. However, we wanted to test the vectors with DNA from the environment, because it is well known that it may be difficult or inefficient to clone DNA from such samples (Robe *et al*., 2003), and because we wanted to test the behaviour of the environmental DNA inserts in new hosts.

A fosmid library consisting of about 20 000 clones was constructed using DNA isolated from marine sediments as cloning material. End sequencing and restriction digest analysis of a selection of the clones confirmed that the library consists of plasmids with different environmental inserts of the expected size of around 35 kb. To explore the microbial diversity represented by this library, regions of 16S rRNA genes were amplified and analysed by DGGE. More than 10 different bands could be clearly distinguished, indicating that the DNA in the library originates from many different genotypes, as expected. In addition, sequencing of 24 amplified 16S rRNA gene fragments resulted in identification of 10 different genotypes, none of which were identical to existing sequences in the databases (data not shown). It could therefore be concluded that the library contains environmental DNA of diverse origins, and that pRS44 has the features required for metagenomic studies.

Verification of heterologous expression of environmental proteins in *P. fluorescens* by 2-DE and MS

In order to study if cloned genes in the metagenomic library are expressed in a non-*E. coli* host, crude protein extracts from five *P. fluorescens* library clones and cells harbouring the vector only were subjected to 2-DE in duplicate. Protein patterns of all six crude protein extracts were analysed with the DELTA-2D software in order to identify protein spots present in both duplicates of only one environmental DNA-containing strain but not in any other clone or the

strain containing the vector only. Even though gels were generally remarkably similar, such individual spots could easily be identified, in some cases also in parts of the gel that were not heavily crowded by host proteins. As an example, Fig. 3 shows the 2-DE images of the control containing the vector only (*P. fluorescens* pRS44) and the *P. fluorescens* metagenome library clone P-13. As can be seen there is a particular very clear spot that is present in both technical replicates from the strain with the metagenomic insert, while missing in the vector control strain. The protein in this spot was subjected to LC-MS/MS measurements and the respective spectra were searched against a *P. fluorescens* specific database or the entire SwissProt database, respectively. Whereas clear cut matches to the *P. fluorescens* database were observed for all host proteins tested, no significant match against the *P. fluorescens* or the SwissProt database was obtained for the presumable insert DNA-encoded protein depicted in Fig. 3, even if the search was performed with information-rich spectra. This observation proves the concept by clearly indicating expression of an additional, non-*P. fluorescens* protein encoded by the environmental DNA insert.

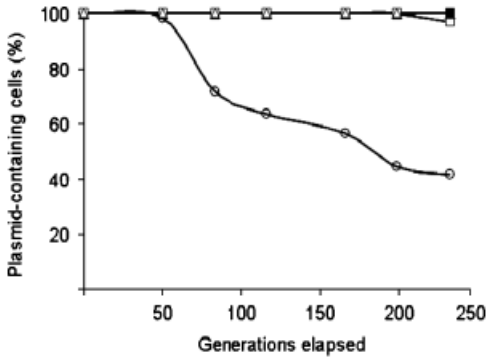

Fig. 2. Plasmid stability of pRS44 and pRS49 in the absence of antibiotic selection. Exponentially growing cells in shake flasks (in the presence of selection) were diluted 10^5 times in medium lacking antibiotics. The cultures were then grown overnight and the dilution procedures were repeated until about 230 generations had elapsed. After each growth step cells were plated on LB-agar lacking antibiotics. From each step, 184 colonies were picked and duplicated into 96-well plates containing media with and without chloramphenicol. ■, pRS44; □, pRS49; Δ, RK2; and ○, pCC1FOS.

Construction of BAC clones with inserts up to around 200 kb

To test the capability of the vector (as a BAC) to hold even larger inserts than those described above high-molecular-weight DNA from plant nuclei was cloned into pTA44. This resulted in BAC clones with inserts up to 200 kb. PFGE was used to determine the size of the inserts, and Fig. 4 shows the result for 22 of the obtained BAC clones. The ligation and transformation efficiencies were similar to what was observed in parallel experiments with the commercially available BAC cloning vector pIndigoBAC5. These experiments confirmed that pRS44/pTA44 has retained the capacity of the parent pCC1FOS to carry and maintain very large inserts in *E. coli*, and the performance of the new vector in alternative hosts could therefore be tested.

Transfer of fosmid library and BAC clones from *E. coli* to *P. fluorescens* and *X. campestris*

The *E. coli* strain EPI300 does not contain the *tra* genes required for mobilized *oriT*-mediated conjugation to new hosts, and for this reason the entire fosmid library and

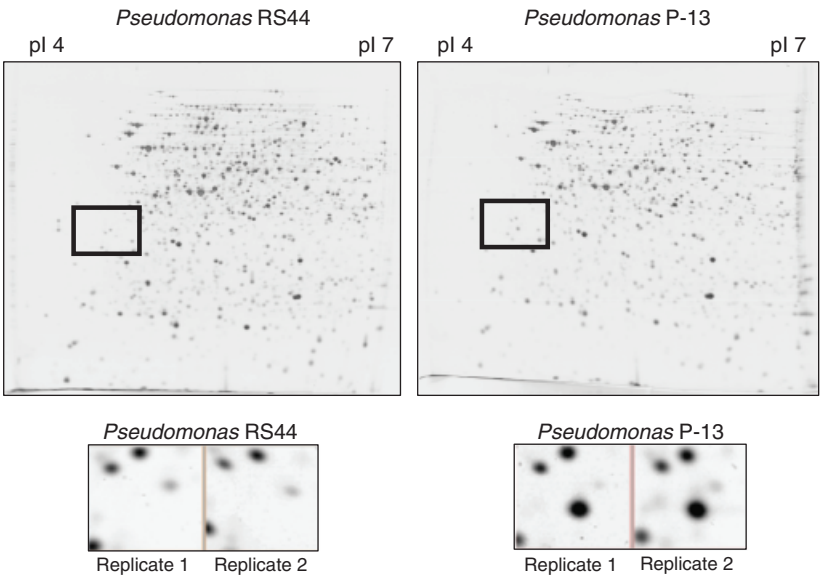

Fig. 3. Comparative analysis of protein patterns of *Pseudomonas fluorescence*-containing the vector pRS44 or the vector with environmental DNA inserts. Crude protein extracts of the control *P. fluorescence* pRS44 and a strain (P-13) containing pRS44 with an environmental DNA insert were separated by 2-DE and protein patterns were compared with the software package DELTA-2D after staining with silver nitrate. In the lower part detailed zoomed regions of the two technical replicates of each strain containing insert-encoded proteins are shown. An insert-encoded protein is located in the centre of the enlarged section.

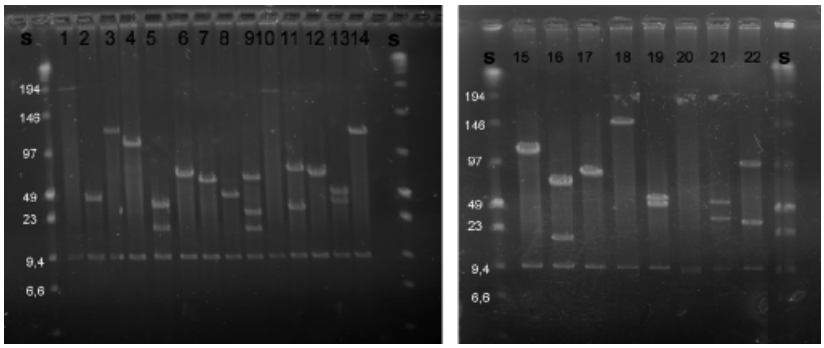

Fig. 4. Pulsed field gel electrophoretic analysis of 22 NotI-digested BAC clones (lanes 1–22). S, molecular weight standard (kb) (New England BioLabs).

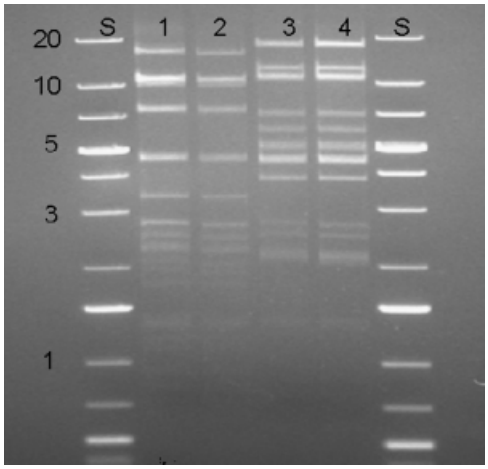

Fig. 5. Agarose gel electrophoretic analysis of two HindIII-digested BAC clones before and after passage through *Pseudomonas fluorescens*::*TnRS48*. Lanes 1 and 2, plasmid B9 before and after transfer, respectively; lanes 3 and 4, plasmid B19 before and after transfer, respectively; S, molecular weight standard (kb) (Fermentas).

selected BACs were first transformed into strain S17.1, which has the RK2 *tra* genes integrated into the chromosome. Direct library construction in S17.1 can also be carried out but is more complicated due to lower frequencies of transformation. However, it was easy to obtain the required number of transformants from the pre-existing library in EPI300; thus, this additional step will apparently not significantly reduce the representativity of the clones in the original libraries.

Before the fosmid and BAC plasmids were transferred to two selected hosts, *P. fluorescens* (fosmid library and selected BAC clones) and *X. campestris* (fosmid clones), the transposon in pRS48 (carrying the *trfA* gene) was inserted into their chromosomes by electroporation. The entire fosmid library and the selected BAC-clones were then conjugatively transferred to the new hosts. In all cases kanamycin-(*P. fluorescens*) and chloramphenicol-(*X. campestris*) resistant clones were obtained at high frequencies, demonstrating that complete fosmid libraries and BAC clones with large inserts could efficiently be transferred.

Plasmids were isolated from independent transconjugants, retransformed into *E. coli* EPI300, digested with HindIII and analysed by agarose gel electrophoresis. Fosmids (30–35-kb inserts) and BACs with inserts up to 130 kb could be recovered in an intact state from the respective hosts (Fig. 5). Plasmids with the largest inserts (around 190 kb) were difficult to retransform into *E. coli*, presumably due to low levels of plasmid recovery and also inefficient transformation. However, Southern hybridization analyses showed that also these plasmids were present in the plasmid state in *P. fluorescens* (data not shown).

Based on all these experiments it could be concluded that pRS44/pTA44 has retained the capacity of the parent plasmid pCC1FOS to hold and maintain very large inserts, that it is more stably maintained than its parent and that it has the capacity to be efficiently transferred to and stably maintained both as a fosmid and a BAC in presumably a large number of non-*E. coli* hosts. To our knowledge, this is the only available vector system that combines all these features, and pRS44 should therefore represent a very useful tool for functional screening across species barriers.

Acknowledgement

This work was supported by a grant from the Norwegian Research Council.

Authors' contribution

T.A. and K.F.D. contributed equally to this work.

References

Amann RI, Ludwig W & Schleifer KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.

Bakkevig K, Sletta H, Gimmestad M, Aune R, Ertesvåg H, Degnes K, Christensen BE, Ellingsen TE & Valla S (2005) Role of the

Pseudomonas fluorescens alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *J Bacteriol* **187**: 8375–8384.

Bates S, Cashmore AM & Wilkins BM (1998) IncP plasmids are unusually effective in mediating conjugation of *Escherichia coli* and *Saccharomyces cerevisiae*: involvement of the Tra2 mating system. *J Bacteriol* **180**: 6538–6543.

Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K & Valla S (1997a) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl Environ Microb* **63**: 370–379.

Blatny JM, Brautaset T, Winther-Larsen HC, Karunakaran P & Valla S (1997b) Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in Gram-negative bacteria. *Plasmid* **38**: 35–51.

Blum H, Beier H & Gross H. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**: 93–99.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**: 248–254.

Burr MD, Clark SJ, Spear CR & Camper AK (2006) Denaturing gradient gel electrophoresis can rapidly display the bacterial diversity contained in 16S rDNA clone libraries. *Microb Ecol* **51**: 479–486.

Chung CT, Niemela S-L & Miller RH (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *P Natl Acad Sci USA* **86**: 2172–2175.

Daniel R (2005) The metagenomics of Soil. *Nature Rev* **3**: 470–478.

DeLong EF (2005) Microbial community genomics in the ocean. *Nature Rev* **3**: 459–469.

De Lorenzo V, Cases I, Herrero M & Timmis KN (1993) Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in *Pseudomonas putida* with *lacZ-tet* bicistronic reporters. *J Bacteriol* **175**: 6902–6907.

Gabor E, Alkema WBL & Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**: 879–886.

Gimmestad M, Sletta H, Karunakaran P, Bakkevig K, Ertesvåg H, Ellingsen TE, Skjåk-Bræk G & Valla S (2004) New mutant strains of *Pseudomonas fluorescens* and variant thereof, methods for their production, and uses thereof in alginate production. Patent WO 2004/011628 A1.

Hain T, Otten S, von Both U, Chatterjee SS, Technow U, Billion A, Ghai R, Mohamed W, Domann E & Chakraborty T (2008) Novel bacterial artificial chromosome vector pUvBBAC for use in studies of the functional genomics of *Listeria* spp. *Appl Environ Microb* **74**: 1892–1901.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245–R249.

Haugan K, Karunakaran P, Tøndervik A & Valla S (1995) The host range of RK2 minimal replicon copy-up mutants is limited by species-specific differences in the maximum tolerable copy-number. *Plasmid* **33**: 27–39.

Hötte B, Rath-Arnold I, Pühler A & Simon R (1990) Cloning and analysis of a 35.3-kilobase DNA region involved in exopolysaccharide production in *Xanthomonas campestris* pv. *campetris*. *J Bacteriol* **172**: 2804–2807.

Keil S & Keil H (1992) Construction of a cassette enabling regulated gene expression in the presence of aromatic hydrocarbons. *Plasmid* **27**: 191–199.

Kim UJ, Shizuya H, de Jong PJ, Birren B & Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucl Acids Res* **20**: 1083–1085.

Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA & Osburne MS (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microb* **70**: 2452–2463.

Mermod N, Ramos JL, Lehrbach PR & Timmis KN (1986) Vector for regulated expression of cloned genes in a wide range of Gram-negative bacteria. *J Bacteriol* **167**: 447–454.

Muyzer G, Dewaal EC & Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for the 16S rRNA. *Appl Environ Microb* **59**: 695–700.

Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Shwab H, Stanisich VA & Thomas CM (1994) Complete nucleotide sequence of Birmingham IncPα plasmids – compilation and comparative analysis. *J Mol Biol* **239**: 623–663.

Poyart C & Trieu-Cuot P (1997) A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to β-galactosidase in Gram-positive bacteria. *FEMS Microbiol Lett* **156**: 193–198.

Ramos JL, Gonzalez-Carrero M & Timmis KN (1988) Broad-host range expression vectors containing manipulated *meta*-cleavage pathway regulatory elements of the TOL plasmid. *FEBS Lett* **226**: 241–246.

Robe P, Nalin R, Capellano C, Vogel TM & Simonet P (2003) Extraction of DNA from soil. *Eur J Soil Biol* **39**: 183–190.

Rondon MR, August PR, Bettermann AD *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microb* **66**: 2541–2547.

Sambrook J & Russel DW (2001) *Molecular Cloning, a Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y & Simon M (1992) Cloning and stable maintance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *P Natl Acad Sci, USA* **89**: 8794–8797.

Sia EA, Roberts RC, Easter C, Helinski DR & Figurski DH (1995) Different relative importances of the *par* operons and the effect of conjugal transfer on the maintenance of intact promiscuous plasmid RK2. *J Bacteriol* **177**: 2789–2797.

Simon R, Priefer U & Pühler A (1983) A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Bio/Technology* **1**: 784–791.

Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM & Donadio S (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nature Biotechnol* **18**: 343–345.

Teske A, Wawer C, Muyzer G & Ramsing NB (1996) Distribution of sulfate-reducing bacteria in a stratified fjord (Mariager Fjord, Denmark) as evaluated by most-probable number counts and denaturing gradient gel electrophoresis of PCR-amplified ribosomal DNA fragments. *Appl Environ Microb* **62**: 1405–1415.

Thiele T, Steil L, Gebhard S, Scharf C, Hammer E, Brigulla M, Lubenow N, Clemetson KJ, Völker U & Greinacher A (2007) Profiling of alterations in platelet proteins during storage of platelet concentrates. *Transfusion* **47**: 1221–1233.

Thomas CM & Helinski DR (1989) Vegetative replication and stable inheritance of IncP plasmids. *Promiscuous Plasmids in Gram-Negative Bacteria* (Thomas CM, ed), pp. 1–25. Academic Press, London.

Torsvik V & Ovreas L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240–245.

Torsvik V, Goksoyr J & Daae FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microb* **56**: 782–787.

Venter JC, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Waters VL (2001) Conjugation between bacterial and mammalian cells. *Nat Genet* **29**: 375–376.

Wexler M, Bond PL, Richardson DJ & Johnston AWB (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* **7**: 1917–1926.

Wild J, Hradecna Z & Szybalski W (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genom Res* **12**: 1434–1444.

Zhang HB, Zhao X, Ding X, Paterson AH & Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. *Plant J* **7**: 175–184.

Zhou J, Bruns MA & Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microb* **62**: 316–322.