

The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIPO2 isolated from bacteria of the wheat rhizosphere

Andreas Tauch,¹ Susanne Schneiker,² Werner Selbitschka,² Alfred Pühler,² Leo S. van Overbeek,³ Kornelia Smalla,⁴ Christopher M. Thomas,⁵ Mark J. Bailey,⁶ Larry J. Forney,⁷ Andrew Weightman,⁸ Piotr Ceglowski,⁹ Tony Pembroke,¹⁰ Erhard Tietze,¹¹ Gunnar Schröder,¹² Erich Lanka¹² and Jan Dirk van Elsas³

Author for correspondence: Christopher M. Thomas. Tel: +44 121 414 5903. Fax: +44 121 414 5925. e-mail: c.m.thomas@bham.ac.uk

- ¹ Zentrum für Genomforschung, Universität Bielefeld, D-33615 Bielefeld, Germany
- ² Lehrstuhl für Genetik, Universität Bielefeld, D-33615 Bielefeld, Germany
- ³ Plant Research International, 6700AA Wageningen, The Netherlands
- ⁴ Biologische Bundesanstalt für Land-und Forstwirtschaft, D-38104 Braunschweig, Germany
- ⁵ School for Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
- ⁶ NERC Institute of Virology and Environmental Microbiology, Oxford OX1 3SR, UK
- ⁷ Department of Biology, University of Idaho, Moscow, ID 83844, USA
- ⁸ Cardiff School of Biosciences, University of Wales, Cardiff CF10 3TL, UK
- ⁹ Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland
- ¹⁰ Department of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland
- ¹¹ Robert Koch-Institut, Bereich Wernigerode, D-38855 Wernigerode, Germany
- ¹² Max-Planck-Institut für Molekulare Genetik, Dahlem, D-14195 Berlin, Germany

Plasmid pIPO2 is a cryptic, conjugative, broad-host-range plasmid isolated from the wheat rhizosphere. It efficiently self-transfers between α , β and γ *Proteobacteria* and has a mobilizing/retromobilizing capacity for IncQ plasmids. The complete nucleotide sequence of pIPO2 is presented on the basis of its mini-Tn5::luxABtet-tagged derivative, pIPO2T. The pIPO2 sequence is 39815 bp long and contains at least 43 complete ORFs. Apart from a suite of ORFs with unknown function, all of the genes carried on pIPO2 are predicted to be involved in plasmid replication, maintenance and conjugative transfer. The overall organization of these genes is different from previously described plasmids, but is similar to the genetic organization seen in pSB102, a conjugative plasmid recently isolated from the bacterial community of the alfalfa rhizosphere. The putative conjugative transfer region of pIPO2 covers 23 kb and contains the genes required for DNA processing (Dtr) and mating pair formation (Mpf). The organization of these transfer genes in pIPO2 is highly similar to the genetic organization seen in the environmental plasmid pSB102 and in pXF51 from the plant pathogen *Xylella fastidiosa*. Plasmids pSB102 and pXF51 have recently been proposed to form a new family of environmental broad-host-range plasmids. Here it is suggested that pIPO2 is a new member of this family. The proposed Mpf system of pIPO2 shares high amino acid sequence similarity with equivalent VirB proteins from the type IV secretion system of *Brucella* spp. Sequence information was used to design primers specific for the detection of pIPO2. Environmental DNA from a range of diverse habitats was screened by PCR with these primers. Consistently positive signals for the presence of pIPO2 were obtained from a range of soil-related habitats, including the rhizospheres of young wheat plants, of field-grown oats and of grass (all gramineous plants), as well as from the rhizosphere of tomato plants. These data add to the growing evidence that plasmids carry advantageous genes with as yet undefined functions in plant-associated communities.

Keywords: bacterial genome, exogenous plasmid isolation, plasmid stability, conjugative transfer, horizontal gene spread

The GenBank accession number for the pIPO2T sequence reported in this paper is AJ297913.

INTRODUCTION

With a basis in the recent sequencing of whole bacterial genomes, it has become increasingly clear that lateral gene transfer processes have played a major role in the evolution of bacteria and in their adaptation to their natural environments (de la Cruz & Davies, 2000; Kurland, 2000; Ochman *et al.*, 2000). It has been estimated that up to 18% of the genome of *Escherichia coli* has been acquired as a result of recombination events following horizontal gene transfers (Lawrence & Ochman, 1998). Evidence for the key role of lateral gene transfer in bacterial evolution is undoubtedly strong, but direct proof for the mechanisms involved, and their ecological impact in natural bacterial communities, is scarce. It has been postulated that, in addition to the primary gene pool, which is represented by the conserved sectors of bacterial chromosomes (the core chromosomal gene pool), bacterial communities maintain a second gene pool (the 'horizontal' gene pool), which is carried by mobile genetic elements such as plasmids (Thomas, 2000a).

In plant-related environments, such as the rhizosphere, conjugation has been shown to be a key mechanism that can facilitate the gene flow between Gram-negative bacteria (van Elsas *et al.*, 1988; Smit *et al.*, 1991, 1993; Pukall *et al.*, 1996; Lilley & Bailey, 1997). Thus, studies with broad-host-range plasmids of incompatibility groups IncP-1, IncN and IncW, as well as studies with the phytosphere-isolated plasmid pQM, showed that rates of transfer between pseudomonads were greater in the rhizospheres of wheat and sugar beet, respectively, than in bulk soil (van Elsas *et al.*, 1988; Lilley *et al.*, 1994). These results indicate that naturally occurring conjugative plasmid transfer may be a frequent process with significance for the adaptive capacity of rhizospheric bacterial communities. It is therefore likely that naturally occurring self-transmissible plasmids carrying the genes necessary for conjugative DNA transfer play a major role in gene-exchange processes in plant-related microbial communities (van Elsas *et al.*, 2000a).

In previous work describing the capacity of plant-associated bacterial populations to effect gene transfer, a set of cryptic plasmids with gene-mobilizing capacity was isolated from bacterial populations of the wheat rhizosphere (van Elsas *et al.*, 1998). These plasmids were isolated by a triparental mating system (triparental exogenous isolation) allowing the isolation of mobile genetic elements by virtue of their capacity to mobilize non-selftransferable plasmids (Hill *et al.*, 1992, 1996; Smalla *et al.*, 2000). The system consisted of an *E. coli* donor strain carrying the mobilizable but non-selftransferable IncQ plasmid pMOL187, and a plasmidless rifampicin-resistant *Ralstonia eutropha* recipient strain (Top *et al.*, 1994). Plasmid pMOL187 contains the *czc* gene cassette, encoding resistances against ions of cadmium, zinc and cobalt, which can only be expressed in the *R. eutropha* recipient strain. Following a multi-partner mating between indigenous wheat rhizosphere

bacteria and the two partners of the mating system, plasmids with broad-host-range mobilizing capacity can be obtained in the recipient via co-transfer with pMOL187 (Top *et al.*, 1994). This procedure, thus, selects naturally occurring plasmids primarily on the basis of their capacity to mobilize and self-transfer, i.e. to efficiently produce conjugation bridges between the mating partners employed.

One of the plasmids obtained from bacterial populations of the wheat rhizosphere, termed pIPO2, had an estimated size of about 38 kb and was not classifiable into any known group by replicon typing via PCR and hybridization with replicon-specific probes (van Elsas *et al.*, 1998). The plasmid could mobilize IncQ plasmids to many different Gram-negative bacterial species and had a broad self-transfer range among the α , β and γ subclasses of the *Proteobacteria* (van Elsas *et al.*, 1998). Plasmid pIPO2 was able to mobilize the IncQ plasmid pIE723 to these hosts at considerable frequencies under rhizosphere conditions in the field (van Elsas *et al.*, 1998). The presence of pIPO2 enabled both *Pseudomonas fluorescens* (van Elsas *et al.*, 1998) and *R. eutropha* (Szpirer *et al.*, 1999) to capture selectable IncQ plasmids from other hosts via retrotransfer.

In this study, we describe the complete nucleotide sequence and genetic organization of plasmid pIPO2, on the basis of the mini-Tn5::luxAB::tet-tagged derivative pIPO2T. This plasmid is proposed to belong to a new family of environmental broad-host-range plasmids including pSB102 and pXF51, each of which are plasmids isolated from plant-associated bacteria carrying a similarly organized set of genes necessary for bacterial conjugation. Plasmid pSB102 was isolated from the bacterial community of the alfalfa rhizosphere (Schneiker *et al.*, 2001), and pXF51 is a plasmid isolated from the plant pathogen *Xylella fastidiosa* (Simpson *et al.*, 2000; Marques *et al.*, 2001). The sequence information of pIPO2 was used to design specific primers that were used for PCR to determine in what ecological niches pIPO2-like plasmids could be found.

METHODS

Bacterial strains and plasmids. Plasmid pIPO2 was originally obtained in *R. eutropha* as a cryptic plasmid that co-transferred with the IncQ plasmid pMOL187 (van Elsas *et al.*, 1998). It was then transferred into *P. fluorescens* R2f (*chr*::Tn5) with pSUP104 as the selectable mobilizable plasmid, and further transferred into a nalidixic-acid- and rifampicin-resistant derivative of strain R2f (strain R33), using plasmid pIE723 as the indicator plasmid (van Elsas *et al.*, 1998). Curing of pIE723 yielded *P. fluorescens* R33(pIPO2), which was routinely used in subsequent transfer experiments. To facilitate selection, pIPO2 was tagged with a mini-Tn5::luxAB::tet transposon (de Lorenzo *et al.*, 1990) using a triparental mating/mobilization strategy (van Elsas *et al.*, 1998), yielding *P. fluorescens* R2f(pIPO2T). All strains were stored at -80°C in 20% (w/v) glycerol. For growth of the latter strain, LB broth (Sambrook *et al.*, 1989) supplemented with tetracycline hydrochloride ($20\ \mu\text{g ml}^{-1}$) was used, with incubation at 27°C with shaking.

Plasmid isolation, shotgun-library construction and DNA sequencing. Plasmid DNA of pIPO2T was isolated from *P. fluorescens* R2f by means of the alkaline-SDS lysis and column purification procedure, according to the NucleoBond Plasmid Purification Protocol (Macherey-Nagel, Düren, Germany). Ultrasonication was used in a common ultrasonicator (at maximal amplitude). A time-course was done (0–2–5–10–20 s) with subsamples of the pIPO2 DNA preparation, and the treated subsamples were run on an agarose gel to determine the mean fragment sizes. The time resulting in mean fragment sizes of around 1000 bp was used to treat the full pIPO2 DNA preparation. The shotgun library was constructed in the vector pBluescript II SK⁻ by using 0.6–1.2 kb size fractions of the pIPO2T DNA. Sequencing was performed in the different laboratories of participating authors and was completed at the University of Bielefeld with the following strategy. Templates for DNA sequencing were prepared by automated alkaline lysis of *E. coli* clones carrying recombinant plasmids by means of the QIAGEN BioRobot 9600. Cycle-sequencing reactions were carried out using purified template DNA, the Prism Ready Reaction Dye Deoxy Termination Kit and fluorescently labelled forward and reverse primers. The products of cycle sequencing were sequenced by running them on a Prism ABI 377 automated DNA sequencer (Applied Biosystems). Computer-assisted assembly of the random shotgun sequences was carried out with the STADEN sequence analysis package (Staden, 1996). Gaps in the pIPO2T sequence were closed by primer walking with oligonucleotides designed to contig ends (Staden, 1996), using pIPO2T DNA as template. Final assembly and editing of the DNA sequence data resulted in a single, circular molecule with a total length of 45 319 bp.

DNA sequence analysis and annotation. The complete DNA sequence of pIPO2T was initially analysed with the automated genome interpretation system GenDB (Center for Genome Research, Bielefeld, Germany). Evaluation of the data collection in the GenDB database files was performed with the STADEN software package (Staden, 1996). In addition, features from searches against the nucleotide database at NCBI with the BLAST-N algorithm, version 2.0 (Altschul *et al.*, 1997), were analysed and incorporated into the sequence annotation. The complete sequence of pIPO2T was searched for σ^{70} -dependent promoters, using the Neural Network Promoter Prediction approach (Reese *et al.*, 1996). Repeat regions within the pIPO2T nucleotide sequence were identified with the REPUTER software (Kurtz & Schleiermacher, 1999).

Amino acid sequence similarities were calculated with the ALIGN computer program (Myers & Miller, 1988). Prediction of signal peptides and transmembrane helical segments was performed with the neural network systems SignalP (Nielsen *et al.*, 1997) and PredictProtein (Rost *et al.*, 1995), respectively. Protein sequence alignment of TraR was done using the CLUSTAL W software, version 1.4, with the BLOSUM series as protein weight matrix; shadings were drawn following the Gonnet PAM250 matrix. The annotated DNA sequence of pIPO2T was re-oriented in such a way that the orientation and order of genes is the same as in the pSB102 sequence (Schneiker *et al.*, 2001).

Distribution of bacteria with pIPO2. To determine the distribution of pIPO2, PCR analysis of environmental total microbial community DNA was performed as described previously (van Elsas *et al.*, 1998). The samples from which DNA was extracted were as follows.

Seven-day-old wheat. *Triticum aestivum* cv. *Sicco* grown in Flevo silt loam soil in microcosms, Wageningen, The Netherlands (van Elsas *et al.*, 1998).

Twenty-three-day-old rice. *Oryza sativa* L. japonica Zhongzao 9037 grown in Flevo silt loam soil in microcosms, Wageningen, The Netherlands (Lin *et al.*, 2000).

Thirty-day-old maize. *Zea mays* grown in Varzea, as well as in Cerrado field soils in Sete Lagoas, Minas Gerais, Brazil (Rosado *et al.*, 1998).

Mature wheat. *Triticum aestivum* cv. *Sicco* grown in Flevo silt loam soil microplots, Wageningen, The Netherlands (van Overbeek *et al.*, 1995).

Mature oats. *Avena sativa* var. Gigant grown in Wildekamp loamy sand (KCl; pH 5.3; 3% organic matter) field soil, Wageningen, The Netherlands.

Tomato. *Lycopersicon esculentum* cv. Moneymaker grown in Ede loamy sand soil in microcosms, Wageningen, The Netherlands.

Potato. *Solanum tuberosum* cv. Desirée grown in Ede loamy sand soil microcosms, Wageningen, The Netherlands.

Cauliflower. *Brassica oleracea* L. grown in Ens loam field soil, Ens, Noordoostpolder, The Netherlands.

Ramanas (white radish). *Raphanus sativus* grown in Ens loam field soil, Ens, Noordoostpolder, The Netherlands.

Flevo silt loam, Wageningen, The Netherlands. Described in van Elsas *et al.* (1998).

Flevo silt loam treated with petroleum/dibenzothiophene. In soil microcosms, Wageningen, The Netherlands (Duarte *et al.*, 2001).

Lovinkhoeve. Silt loam soil (KCl; pH 7.2; 2% organic matter) sampled from arable land at Lovinkhoeve experimental field, Flevopolder, The Netherlands, eight times over one year, on average once every 6 weeks. It should be noted that five time points with positive responses were found in summer and autumn. The samples were provided by Dr E. Smit, RIVM, Bilthoven, The Netherlands.

Ede loamy sand. Obtained from a field microplot (KCl; pH 5.5; 3% organic matter), Wageningen, The Netherlands.

Hellendoorn loamy sand. ILS field soil (KCl; pH 5.2–5.5; 4% organic matter) obtained from Hellendoorn, Overijssel, The Netherlands (van Elsas *et al.*, 2000b).

Westmaas clay. Arable field soil collected at Westmaas, Noordholland, The Netherlands.

A soil. Oil-polluted soil collected in Amsterdam, The Netherlands.

FiOS. High-organic matter forest soil obtained from Viikki, Helsinki, Finland.

Wildekamp soil. See above under mature oats.

Guaira. Soil obtained from arable land with maize in Guaira, Parana, Brazil (Rosado *et al.*, 1998).

Manure. Pig and cow manure obtained from the BBA Braunschweig experimental farm, Germany.

Seawater. Obtained from Fleves island and Eretria, Greece. Courtesy of Dr A. Karagouni, Athens, Greece.

Wastewater. Obtained from Athens wastewater outflow. Courtesy of Dr A. Karagouni, Athens, Greece.

Environmental DNA was obtained using a direct DNA extraction method originally described by Smalla *et al.* (1993) and modified by van Elsas & Smalla (1995). Briefly, the

Table 1. PCR-based detection of pIPO2 in various environmental habitats on the basis of directly extracted environmental DNA

Habitat	Location/type*	Detection of pIPO2 with†	
		PCR system I‡	PCR system II (<i>repA</i>)‡
Rhizosphere soil	7-Day-old wheat	+	+
	23-Day-old rice	+	+
	30-Day-old maize (Varzea, Cerrado)	+	+
	Mature wheat	–	–
	Mature oats	+	+
	Grass (Wildekamp)	+	+
	Tomato	+	+
	Potato	–	–
	Cauliflower	+	ND
	Ramanas	–	–
Bulk soil	Flevo silt loam (FSL)	–	–
	FSL treated with petroleum/dibenzothiophene	–	–
	Lovinkhoeve (three time points)§	–	–
	Lovinkhoeve (five time points)§	+	+
	Ede loamy sand	–	–
	Hellendoorn loamy sand	–	–
	Hellendoorn loamy sand, compost- treated	–	–
	Westmaas clay	–	–
	A soil (Amsterdam)	–	–
	FiOS (Finland)	–	–
Wildekamp loamy sand	+	+	
Guaira (Brazil)	–	–	
Manure	From pigs and cows	–	–
Seawater	Greece	–	–
Wastewater	Greece	–	–

ND, Not done.

* For details of sample source see Methods.

† DNA was produced at least twice and screened for production of an appropriate band by PCR.

‡ +, Indicates that at least some positive samples were found via PCR/hybridization; –, indicates the complete absence of specific product.

§ Lovinkhoeve soil was sampled eight times at 6-week intervals, yielding five positive responses spread across summer and autumn.

protocol consisted of a homogenization step, a cell lysis step on the basis of bead beating, and a few extraction and purification steps, the final ones of which were commonly based on the Wizard DNA clean-up system (Promega). For all habitats investigated (Table 1), this method yielded preparations of DNA amplifiable by PCR. Two PCR primer systems were employed to screen for the presence of pIPO2-specific sequences. The first system (I), based on primers pIPO2f and pIPO2r, amplified a product of 305 bp (from position 44091 to 44396 on the pIPO2 sequence, covering ORF33 and a little of ORF32) and was previously shown to be specific for pIPO2 (van Elsas *et al.*, 1998). The second system (II) was designed on the basis of the sequence of the putative *repA* gene. The primers pIPO2rep-x2 (5'-CGCAACCTTGCCACATCG-3') and pIPOr-y2 (5'-TAGGTAGCTCATGCGATAGG-3') amplified a product of 546 bp. PCR with environmental DNA

was performed using 1 µl (5–20 ng) of each DNA extract. To assess the template quality of the isolated environmental DNA, controls were performed with each extract spiked with serially diluted pIPO2 DNA. The PCR conditions applied in system I were as described previously (van Elsas *et al.*, 1998). Thermal cycling in system II was performed as follows: denaturation at 94 °C for 3 min; 40 cycles at 94 °C for 1 min, 50–55 °C for 90 s and 72 °C for 1 min; final extension at 72 °C for 20 min. PCR products were run on 1.2% (w/v) agarose gels (Sambrook *et al.*, 1989), and ethidium-bromide-stained bands were visualized under UV illumination. Gels were blotted to nylon membrane filters, and blots were hybridized at high stringency according to the protocol supplied by the manufacturer (Roche Diagnostics). Digoxigenin-labelled amplicons produced on plasmid pIPO2 with systems I (probe I) and II (probe II) were used as specific probes. Signals were

detected following exposure of X-ray films, according to standard procedures. PCR results were considered positive and specific if a hybridization signal indicating a product of expected size with the specific probe was observed. All habitats were tested by fresh isolation of DNA followed by PCR a minimum of two times.

RESULTS AND DISCUSSION

DNA sequence annotation of pIPO2

The complete nucleotide sequence of pIPO2 was determined by sequencing of its tagged derivative, pIPO2T, of which a shotgun library was constructed. Assembly of the pIPO2T sequence was carried out with the STADEN sequence analysis package. The DNA sequence was determined on both strands, with a mean coverage of 5.75 per consensus base. In total, 533 reads were assembled, with a mean length of 489.25 bases per read. This gave a single contiguous DNA sequence of 45 319 bp (mean G + C content of 60.3 mol%), consisting of 39 815 bp equivalent to the pIPO2 mother plasmid, 5495 bp of mini-Tn5::luxABtet transposon, and an additional 9 bp inserted by target-site duplication at the integration site. The genome size of pIPO2 and its deduced restriction maps are in line with the previous estimation of size (about 38–39 kb) made on the basis of restriction with EcoRI and SphI (van Elsas *et al.*, 1998; A. C. Keijzer-Wolters & J. D. van Elsas, unpublished data). The pIPO2T sequence contains 46 ORFs, as identified by automated genome interpretation. The names and positions of these ORFs are listed in Table 2. The set of ORFs includes the luxAB and tetA(C) coding regions, which were inserted into orf34 of pIPO2 by the tagging procedure with the mini-Tn5 derivative. Fig. 1 presents a detailed map of the 46 ORFs found on pIPO2T, as well as the positions of large direct and inverted repeat sequences. Data for pIPO2T, such as the G + C content of the coding regions, the length and molecular mass of predicted proteins and the highest similarities to known proteins in databases, are summarized in Table 2. The putative coding regions of pIPO2 are dominated by genes involved in plasmid replication, partitioning/stable maintenance and conjugative transfer, whereas only a few ORFs cannot be attributed to these essential plasmid backbone genes (Table 2; Fig. 1). These observations are in line with the 'cryptic' character of plasmid pIPO2, as reported previously (van Elsas *et al.*, 1998).

An automated search for σ^{70} -dependent promoters with the Neural Network Promoter Prediction program (Reese *et al.*, 1996) resulted in the prediction of a total of 79 putative promoters when using a threshold score of 0.8, which recognizes 60% of all promoters, with a rate of false positive predictions of 0.4%. Following the logic of the gene organization we screened the 79 candidates for those that appeared upstream of the first ORF in each putative operon. Additional promoters upstream of other ORFs were chosen if their position relative to the ribosome-binding site seemed logical. Nine putative pIPO2 promoters were selected on the

basis of these criteria (Table 3). The most notable omission from the list of putative promoters is an obvious candidate for a promoter for *orf1* or *traA*. Promoter cloning and transcript mapping will be needed in the construction of a more solid assessment of genetic organization.

Putative replication and partitioning functions

A putative pIPO2 replication initiator protein, RepA, was found on the basis of its similarity to the replication initiator protein of the self-transmissible, broad-host-range IncW plasmid pSa, originally identified in *Shigella flexneri* (Okumura & Kado, 1992). No adjacent direct-repeat clusters typical for replication origin regions were identified in pIPO2. However, because the putative *rep* gene is embedded in a gene cluster (ORF30–ORF34) that consists of two different putative operons (*orf34* and *orf33*), it is possible that the replication origin is further away and therefore difficult to predict. This constellation is not unique, since the *repABC* replicons of many large plasmids, particularly those found in the *Rhizobiaceae*, have a *rep* gene as the third ORF in a partitioning/replication operon and the location of the replication origin has not been defined in these systems (reviewed by Thomas, 2000b). To identify the replication requirements of plasmid pIPO2, the isolation of a mini replicon will provide more information.

Within the cluster ORF30–ORF34 are two genes whose deduced polypeptide products showed significant similarity to the ParA and ParB proteins of plasmid RP4 (Table 2), which are part of a complex multimer resolution/post-segregational killing system (reviewed by Pansegrau *et al.*, 1994a). The ORF34 protein of pIPO2, deduced from joining the 5' and 3' sequence fragments of pIPO2T (Table 2), contains a potential helix–turn–helix motif at its C-terminal end (position 190–211; Dodd & Egan, 1990). The protein is likely to function in dimer resolution, since its amino acid sequence shows homology to resolvases of the Tn3 family. However, the disruption of *orf34*, caused by the tagging procedure with the mini-Tn5::luxABtet transposon in pIPO2T, did not result in plasmid instability in *P. fluorescens*. The fact that pIPO2T was stably maintained in *P. fluorescens* R2f, even without tetracycline selection, suggests that *orf34* is dispensible for stable plasmid inheritance in at least some bacterial hosts (van Elsas *et al.*, 1998). Conversely, the low rate of transfer from *P. fluorescens* R2f to *E. coli* as well as to *Pseudomonas corrugata*, coupled with the consistently poor growth of the putative *E. coli* transconjugants, indicated poor maintenance of the pIPO2T plasmid in these hosts (van Elsas *et al.*, 1998). An apparent redundancy of *parA*-like auxiliary functions would not be unprecedented: IncP-9 plasmids also possess an apparently cryptic putative resolvase (*res*) gene adjacent to their *rep* gene (Greated *et al.*, 2000).

In addition, pIPO2 carries a putative operon (ORF26–ORF22) whose key deduced proteins show amino acid

Table 2. Location of the coding regions of pIPO2T and the closest relationship of the deduced proteins to previously studied proteins

ORF no.	Position in sequence (nt)	Gene	G + C content (mol %)	Protein		Amino acid identity to pSB102	Closest relationship (amino acid identity/similarity to protein; micro-organism and/or plasmid)	GenBank accession no.
				Length (aa)	Size (kDa)			
ORF31	76–1464	<i>repA</i>	61.4	462	50.7	57% to ORF1	26%/36% to RepA protein; pSa17	U30471
ORF30	1464–1928	<i>orf30</i>	64.7	154	16.9	Not in pSB102	60%/71% to ParB protein; RP4	M59825
ORF29	2714–4063	<i>orf29</i>	61.1	449	49.1	67% to ORF4	39%/43% to hypothetical protein; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF28b	4331–4723	<i>orf28b</i>	65.4	130	14.6	60% to ORF5	–	–
ORF28a	4753–5043	<i>orf28a</i>	65.3	96	10.5	25% to ORF6	–	–
ORF27	5148–6182	<i>orf27</i>	68.5	344	37.0	Not in pSB102	40%/51% to KfrA protein; RP4	L27758
ORF26	7740–8099	<i>ssb</i>	59.9	119	13.2	72% to ORF10	34%/45% to single-stranded DNA binding protein; RP4	X00713
ORF25	8103–8501	<i>orf25</i>	55.4	132	14.6	69% to ORF11	26%/39% to partition gene repressor; <i>Pseudomonas putida</i> pM3	AF078924
ORF24	8498–9289	<i>orf24</i>	56.6	263	28.7	71% to ORF12	36%/54% to small IncC; RP4	X03962
ORF23	9290–10429	<i>orf23</i>	62.0	379	40.0	57% to ORF13	40%/53% to transcriptional repressor protein KorB; RP4	Y00448
ORF22	10426–10902	<i>orf22</i>	58.3	158	16.9	62% to ORF14	35%/49% to conserved hypothetical protein TP0087; <i>Treponema pallidum</i>	AE001193
ORF21	11167–11709	<i>orf21</i>	57.9	180	20.3	39% to ORF15	27%/43% to hypothetical protein; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF20	11706–12800	<i>traR</i>	61.9	364	40.3	62% to ORF16	41%/57% to nickase protein; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF19	13457–12879	<i>traQ</i>	61.3	192	20.7	50% to ORF18	50%/64% to TrbM protein; RP4	M93696
ORF18	14031–13471	<i>traP</i>	59.5	186	20.2	60% to ORF19	40%/51% to ORF12 protein; <i>Brucella suis</i>	AF141604
ORF17	16620–14428	<i>traO</i>	60.4	730	81.0	53% to ORF21	23%/37% to DNA primase; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF16	17330–16632	<i>orf16</i>	55.6	232	25.4	44% to ORF22	–	–
ORF15	19937–17424	<i>traN</i>	58.8	837	91.7	64% to ORF23	35%/46% to conjugal transfer protein; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF14	21552–20485	<i>traM</i>	58.9	355	39.8	74% to ORF25	74%/83% to VirB11; <i>Brucella abortus</i>	AF226278
ORF13	22687–21536	<i>traL</i>	62.1	383	39.1	66% to ORF26	53%/66% to VirB10; <i>Brucella abortus</i>	AF226278
ORF12	23559–22687	<i>traK</i>	60.3	290	31.5	66% to ORF27	64%/74% to VirB9; <i>Brucella suis</i>	AF141604
ORF11	24266–23556	<i>traJ</i>	58.9	236	26.3	71% to ORF28	56%/70% to VirB8; <i>Brucella suis</i>	AF141604
ORF10	24436–24272	<i>tral</i>	57.0	54	5.5	61% to ORF29	46%/67% to VirB7; <i>Brucella suis</i>	AF141604
ORF9	25655–24573	<i>traH</i>	59.8	360	37.9	79% to ORF30	30%/49% to VirB6; <i>Brucella abortus</i>	AF226278
ORF8	25963–25667	<i>traG</i>	59.6	98	10.3	56% to ORF31	34%/45% to conserved plasmid protein; <i>Xylella fastidiosa</i>	AE004030
ORF7	26785–26066	<i>traF</i>	58.3	239	26.1	62% to ORF33	44%/61% to VirB5; <i>Brucella abortus</i>	AF226278
ORF6	29271–26791	<i>traE</i>	55.7	826	94.1	81% to ORF34	72%/82% to VirB4; <i>Brucella suis</i>	AF141604
ORF5	29601–29278	<i>traD</i>	58.6	107	12.1	82% to ORF35	57%/74% to VirB3; <i>Brucella suis</i>	AF141604
ORF4	29919–29611	<i>traC</i>	60.5	102	10.5	52% to ORF36	51%/68% to VirB2; <i>Brucella suis</i>	AF141604
ORF3	32398–30050	<i>traB</i>	61.4	782	36.4	68% to ORF37	42%/53% to DNA topoisomerase I; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF2	33169–32408	<i>traA</i>	63.9	253	26.2	65% to ORF38	57%/64% to VirB1; <i>Brucella suis</i>	AF141604
ORF1	33622–33227	<i>orf1</i>	51.0	135	15.3	46% to ORF39	21%/31% to hypothetical protein; <i>Xylella fastidiosa</i> pXF51	AE003851
ORF45	35529–34471	<i>orf45</i>	62.5	352	37.0	36% to ORF40	–	–
ORF44	35617–35883	<i>orf44</i>	60.6	88	9.8	46% to ORF42	–	–
ORF43	35945–36181	<i>orf43</i>	64.9	78	8.5	80% to ORF41	–	–

Table 2 (Cont.)

ORF no.	Position in sequence (nt)	Gene	G+C content (mol %)	Protein		Amino acid identity to pSB102	Closest relationship (amino acid identity/similarity to protein; micro-organism and/or plasmid)	GenBank accession no.
				Length (aa)	Size (kDa)			
ORF42	36190–36606	<i>orf42</i>	66.6	138	16.0	Not in pSB102	–	–
ORF41	36603–36740	<i>orf41</i>	60.9	45	5.2	Not in pSB102	–	–
ORF40	36761–37027	<i>orf40</i>	65.2	88	9.4	44% to ORF45	–	–
ORF39	37036–37299	<i>orf39</i>	65.2	87	9.3	52% to ORF46	37%/51% to hypothetical protein; <i>Xylella fastidiosa</i>	AE004023
ORF38	37316–37540	<i>orf38</i>	61.7	74	8.0	69% to ORF47	–	–
3' End*	38237–37632	<i>orf34-3'</i>	67.5	201	ND	69% to ORF53*	73%/82% to resolvase ParA; RP4	M59825
ORF37	38419–39486	<i>luxA</i>	44.9	355	40.2	Not in pSB102	100%/100% to LuxA; mini-Tn5:: <i>luxAB::tet</i>	M10961
ORF36	39513–40487	<i>luxB</i>	44.7	324	36.3	Not in pSB102	100%/100% to LuxB; mini-Tn5:: <i>luxAB::tet</i>	M10961
ORF35	42057–43247	<i>tetA(C)</i>	61.4	396	41.5	Not in pSB102	100%/100% to TetA(C); mini-Tn5:: <i>luxAB::tet</i>	J01749
ORF34*	43816–43676	<i>orf34-5'</i>	62.4	46	5.0	69% to ORF53*	46%/57% to resolvase ParA; RP4	M59825
ORF33	44014–44367	<i>orf33</i>	63.3	117	13.1	60% to ORF68	–	–
ORF32	44342–45319	<i>orf32</i>	62.6	325	37.0	59% to ORF69	18%/30% to cell filamentation protein; <i>Xylella fastidiosa</i>	AE003992

ND, not determined.

* The joint *orf34* gene has a G+C content of 67.1 mol% and encodes a protein of 226 aa with a molecular mass of 24.3 kDa. The deduced protein revealed 71% identity and 80% similarity to the ParA protein of RP4 (accession no. M59825).

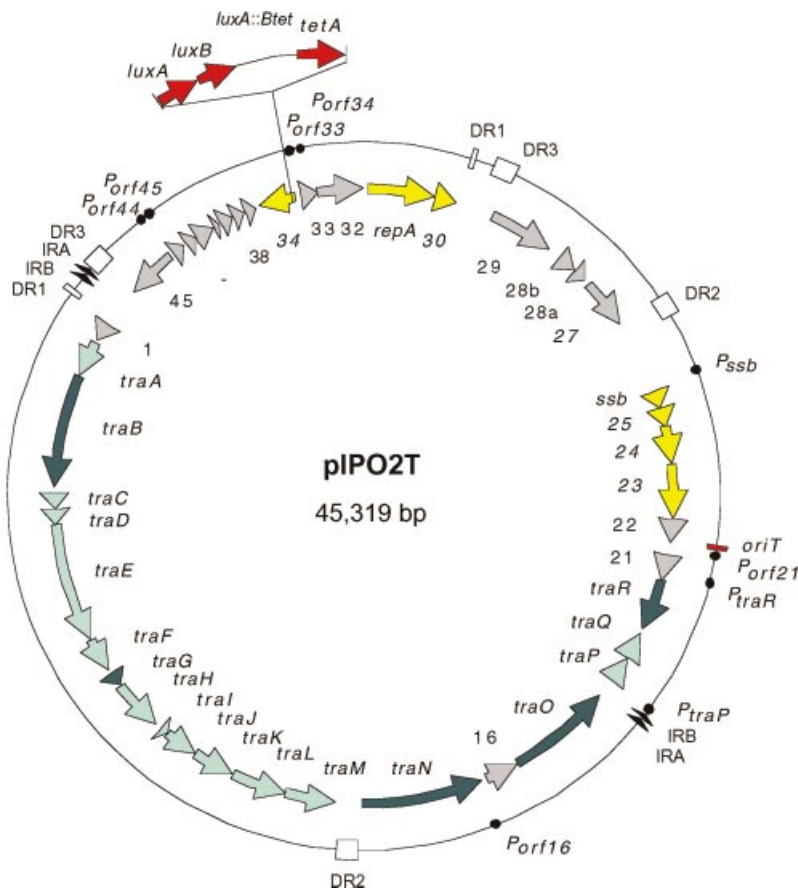


Fig. 1. Genetic map of the 45319 bp pIPO2 derivative, pIPO2T. ORFs deduced from the complete plasmid sequence of pIPO2T are shown by arrows indicating the direction of transcription. Different colours indicate replication and maintenance functions (yellow), mating pair formation functions (light green) and DNA processing functions (dark green). Hypothetical coding regions are shown in grey. The mini-Tn5::*luxAB*::*tet* transposon used for plasmid tagging of pIPO2 is marked in red. The approximate positions of the putative origin of transfer (*oriT*), putative promoter sequences (*P*), long range direct repeats (DR1–DR3) and long range inverted repeats (IRA and IRB) are presented. The lengths of the direct repeats are 96 bp (DR1), 386 bp (DR2) and 403 bp (DR3). The inverted repeats have lengths of 42 bp (IRA) and 38 bp (IRB).

Table 3. Putative promoter sequences identified in the pIPO2T sequence

ORF name	Position in sequence (nt)	Sequence		Directions*	
		–35 region	–10 region		
<i>ssb</i>	7660–7703	agctgt ttgcat	ggtttgttttatgagtgt	tataat gtgct	→
ORF21	11071–11114	ggcta ttgcag	gataggctaatcgccgg	tatagt acgcg	→
<i>traR</i>	11607–11650	cacga ttgagt	caatcgaacggctgcgcg	gatcat cgacg	→
<i>traP</i>	14090–14047	at ttgcc a	agatatataatcttggttc	tacaac aagcc	←
ORF16	17472–17429	cgtg ttgaca	agccccttccttcaatc	aaaat gcgga	←
ORF45	35601–35558	ccaaa ttgcc c	gttgtgtgtttccattata	taccat cttgg	←
ORF44	35539–35582	cgcg tttgt t	gttttcatctaccaagatggta	tataat ggaaa	→
ORF34	43991–43948	acaat ttggc g	tataagcctgcctctat	ttta atgtatg	←
ORF33	43932–43975	gggca ttgac g	atgcaatgccatacat	taaa atagagg	→

* The direction of the arrows indicates the promoter orientation (→, clockwise; ← anticlockwise) relative to the map shown in Fig. 1.

sequence similarity to proteins from the central control region of IncP-1 plasmids (Table 2; Fig. 2). Both the *incC* and *korB* gene products have been shown to be required for the segregational stabilization (partitioning) phenotype, in instances where they have been studied in detail (reviewed by Gerdes *et al.*, 2000; Bignell & Thomas, 2001). *KorB* belongs to the ParB family of DNA-binding proteins that bind the *cis*-acting centromere-like sequence, and *IncC* belongs to the ParA

family of ATPases needed for symmetric distribution of the KorB–DNA complexes. Note that ParA and ParB here should not be confused with ParA and ParB referred to in the previous paragraph. Centromere-like sequences in other systems are normally found upstream or downstream of the *incC* operon, appearing as one or more inverted or direct repeats. The most obvious candidate for a centromere-like region is therefore the cluster of direct repeats upstream of the *ssb* promoter

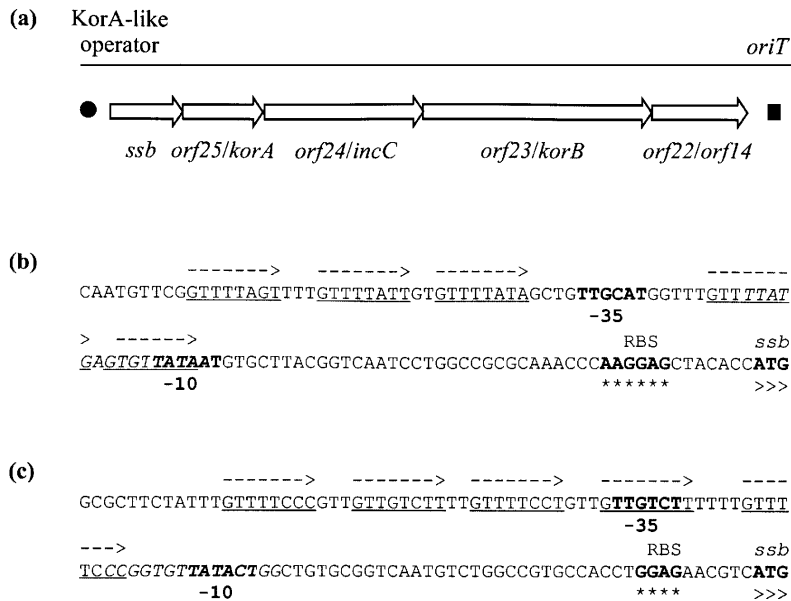


Fig. 2. Putative maintenance region in pIPO2 and pSB102. (a) Genetic organization of the maintenance region in pIPO2 and pSB102. The genes are shown by arrows indicating the direction of transcription. The highly similar coding regions *orf22* and *orf14* are part of the pIPO2 and pSB102 regions, respectively. The approximate position of a putative KorA-like operator and the proposed *oriT* site (see Table 3) are shown. (b) Nucleotide sequence of the *ssb* upstream region of pIPO2. (c) Nucleotide sequence of the *ssb* upstream region of pSB102. Putative -10 and -35 signals and ribosome-binding sites are specifically marked. The ATG start codon of *ssb* is indicated. The inverted repeat sequence of a putative KorA-like operator is typed in italics. Direct repeats are underlined and shown by arrows.

(Fig. 2). An unusual feature of this putative operon is the inclusion of a gene encoding a single-stranded DNA-binding protein (Ssb). *ssb* genes are generally associated with replication or transfer regions rather than with partitioning genes (Golub & Low, 1986; Jovanovic *et al.*, 1992). A comparison between the organization of the *ssb* operon in plasmid pIPO2 and the recently described rhizobial plasmid pSB102 (Schneiker *et al.*, 2001) revealed complete collinearity. In addition, there is considerable sequence conservation at the -35 and -10 sites (Fig. 2), and the repeat sequences upstream of the putative operon were conserved between the two plasmids.

In IncP-1 plasmids, the *incC-korB* (*parA-parB*) operon is autoregulated in part by KorA, as well as by KorB (Pansegrau *et al.*, 1994a). The proposed promoter for this operon in pIPO2 and pSB102 contains an inverted repeat sequence overlapping the -10 region (Fig. 2). This sequence may represent the binding site for the putative KorA-like protein in pIPO2 (*orf25*) and pSB102 (*orf11*). In IncP-1 plasmids, KorA also binds at six other places on the plasmid and serves to coordinate expression of replication and stable inheritance functions, whereas KorB binds at 12 sites (Pansegrau *et al.*, 1994a). A search in the rest of the pIPO2 sequence for additional copies of the proposed target for the KorA-like protein proved negative. This suggests that this protein does not form the centre of a complex regulon of replication, transfer and partitioning genes, as it does in IncP-1 plasmids (Pansegrau *et al.*, 1994a; Thorsted *et al.*, 1998). The adjacent ORF27 encodes a protein with similarity to the KfrA protein of IncP-1 plasmids, which may belong to the broad family of SMC proteins on the basis of its structure (Jagura-Burdzy & Thomas, 1992; Macartney *et al.*, 1997; Cobbe & Heck, 2000). In IncP-1 plasmids, KfrA autoregulates an operon implicated in providing auxiliary maintenance functions, and its role in pIPO2 is likely to be similarly non-essential, since

the closely related plasmid pSB102 does not carry an equivalent gene (Schneiker *et al.*, 2001).

Putative conjugative transfer region

More than half of the plasmid, i.e. approximately 23 kb, is occupied by the putative transfer region, delimited by ORF1/ORF2 (*traA*) and a sequence region proposed to function as the transfer origin (*oriT*) of pIPO2 in between ORF21 and ORF22 (Figs 1 and 3). The proposed *oriT* resembles the T-border sequences (5'-TATCCTGC-3', see Fig. 3) which flank the T-DNA segment of the agrobacterial Ti plasmids, a region that is transferred from bacteria to plant cells by a conjugation-like process. These so-called nick region sequences are part of the IncP-1 *oriT* region and contain eight conserved base pairs, which have been called the *nic*-site (for a review, see Pansegrau & Lanka, 1996). The orientation of the *nic*-site in the lower strand (reverse complementary strand) of pIPO2 predicts that the transfer (*tra*) genes would be transmitted into the recipient cell last, as observed for most other conjugation systems (each of the plasmids listed in Fig. 3). This is the same as the pIPO2-related plasmids pSB102 and pXF51 (Marques *et al.*, 2001; Schneiker *et al.*, 2001), which have significant sequence similarity with pIPO2 at the proposed *nic* region and carry inverted repeat structures (Fig. 3) which might be important for target recognition during DNA processing. Each of the 17 reading frames in between *traA* and *oriT* has significant sequence similarity to a conjugative transfer gene product and/or to a product of a type IV secretion system (Christie & Vogel, 2000; Zechner *et al.*, 2000), suggesting a single and unique transfer region. Potential Tra components of pIPO2 fall into two classes, the DNA processing functions (*orfs* 3, 8, 15, 17 and 20, highlighted in green in Fig. 1; Table 2) and the mating pair formation functions (*orfs* 2, 4-7, 9-14, 18 and 19, highlighted in light green in Fig. 1; Table 2), as outlined below.

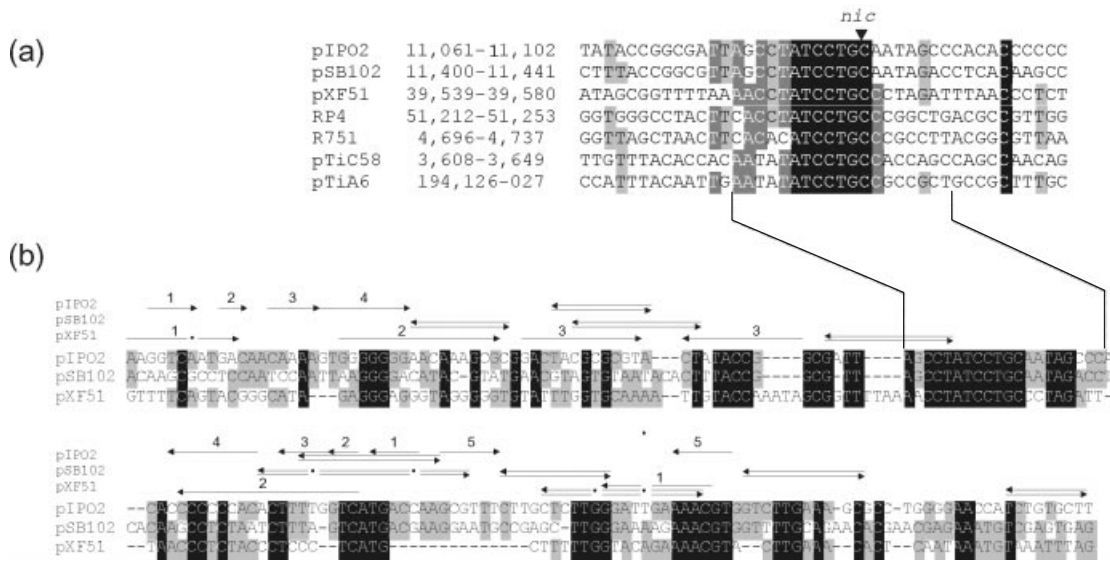


Fig. 3. The proposed *oriT* sequence of pIPO2 is homologous to the *nic* region of IncP-1-like *oriTs*. (a) Sequence alignment showing the putative pIPO2 *nic* region compared to the known *nic* regions of IncP-1 and Ti plasmids. Potential *nic* regions of the closely related plasmids pSB102 and pXF51 are included. The basis for the alignment was the consensus sequence determined for the group of IncP-1-like *oriTs* (Zechner *et al.*, 2000). Nucleotide positions that are conserved throughout are highlighted with black. Dark grey highlighting marks positions that are conserved in >70% of the aligned sequences; light grey highlighting is used to show nucleotides with weak conservation (57%). The *nic*-site as determined for the IncP-1 and Ti transfer systems is indicated. (b) The proposed *oriT* regions of pIPO2, pSB102 and pXF51 share sequence homology and contain common structures. Identical nucleotides conserved throughout are indicated by black highlighting; conservation in two of the three aligned sequences is shown by grey highlighting. Arrows indicate inverted repeat regions in the corresponding sequences, with each pair of repeats designated by identical numbers. Unnumbered parallel arrows in pairs indicate palindromic sequences. The accession numbers for the sequences shown in (a) and (b) are AJ304453 (pSB102), AE003851 (pXF51), M93696 [RP4 (*oriT*)], X54458 [R751 (*oriT*)], AJ237588 [pTiC58 (left border)] and NC_002377 [pTiA6 (border A)].

DNA processing components

Sequence interpretation from sequence alignments allowed us to predict ORF20 to function as a VirD2-like relaxase (Pansegrau *et al.*, 1993), ORF15 as the TraG-like protein VirD4 (Lessl *et al.*, 1992; Balzer *et al.*, 1994; Moncalián *et al.*, 1999), and ORF3 as the topoisomerase III (Top) analogue of TraE of IncP-1 plasmid RP4 (Li *et al.*, 1997). The VirD2-like protein TraR of pIPO2 is predicted to be involved in the transesterification of the plasmid strand destined to be transferred into the recipient, because it contains each of the three defined motifs conserved in other relaxases of IncP-1 and Ti origin (Pansegrau *et al.*, 1994b) (Fig. 4). Whether ORF21 upstream of *traR* encodes a DNA chaperone or a relaxosomal accessory protein as found in IncP-1 plasmids (TraJ/TraK; for a review, see Pansegrau & Lanka, 1996) remains to be determined, since no significant sequence similarity to these IncP-1 functional proteins was detectable. The TraG analogue TraN of pIPO2 may play an intrinsic role directly in DNA transport, as suggested by the three-dimensional structure of the truncated hexameric TraG-like protein TrwB of R388 (Gomis-Rüth *et al.*, 2001) which shares striking similarities with the F₁-ATPase (Egelman, 2001). The sequence alignments suggest that pIPO2 TraB functions as a topoisomerase I (Table 2). The only other topoisomerase found in transfer systems, IncP-1

TraE, is a topoisomerase of type III (Li *et al.*, 1997). ORF17 shares similarity with the conjugative primase PriL of the IncM plasmid pACM1 (Preston *et al.*, 2000). However, the region of similarity matches only part of the PriL molecule, which most likely does not represent the primase moiety but instead represents something else (located towards the C terminus) with as yet unknown function.

Mating pair formation system

The proposed mating pair formation (Mpf) system – responsible for pilus biogenesis and initial contact to the recipient cell – consists of a set of *tra* gene products with pronounced similarity to the corresponding *vir* gene products of *Brucella* spp. (Table 2) and *Bartonella henslelea*. Both of these organisms encode type IV secretion systems that are proposed to function in mammalian pathogenesis (O’Callaghan *et al.*, 1999; Schmiederer & Anderson, 2000; Seira *et al.*, 2000). While this gene block in pIPO2 could conceivably encode a mechanism to deliver proteins to the tissues of the plants in whose rhizosphere the host bacteria grow, it seems more likely that they are devoted to plasmid transfer, since it is known that the plasmid is self-transmissible and many other self-transmissible plasmids from Gram-negative bacteria carry a very similar block of genes, which in many cases has been shown to

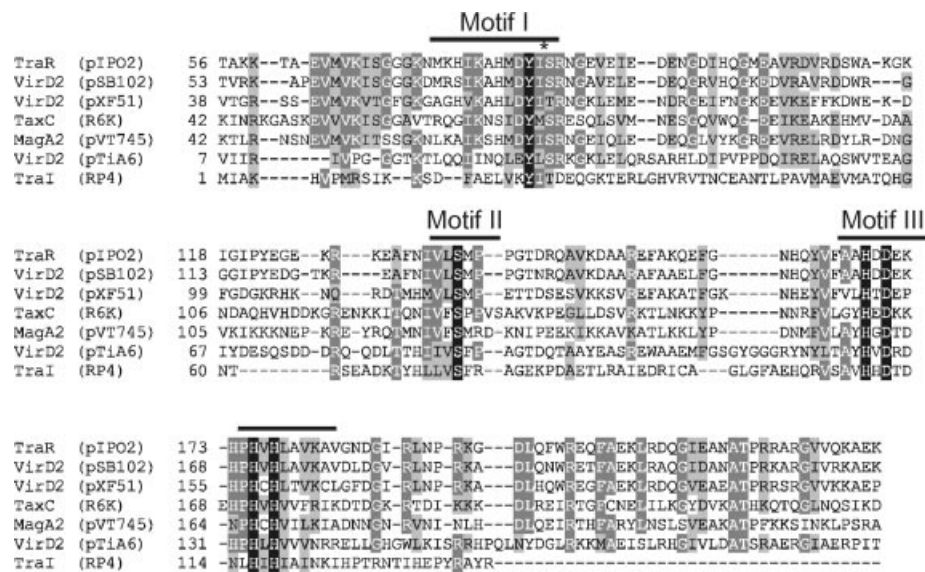


Fig. 4. TraR of pIPO2 possesses three sequence motifs conserved among relaxases. The similarity of the sequences towards TraR of pIPO2 is of decreasing order from top to bottom. Motif I contains the active-site tyrosine (indicated by an asterisk) that is involved in DNA transesterification of RP4 TraI (Pansegrau *et al.*, 1993). Motif II contains residues probably involved in the tight binding of the nick region, whereas motif III contains histidine residues altering the cleavage-joining activity of TraI (Pansegrau *et al.*, 1994b). Amino acid positions that are conserved throughout are indicated by black highlighting. Identical residues present in at least 70% of the sequences are marked by dark grey highlighting; similar residues are indicated by light grey highlighting. The accession numbers for the sequences shown in this figure are Q9PHG1 [VirD2 (pSB102)], AE003851 [VirD2 (pXF51)], S72275 [TaxC (R6K)], NP_067550 [MagA2 (pVT745)], NP_059814 [VirD2 (pTiA6)] and CAA38336 [TraI (RP4)].

TraC	MQTTIDNSGTVKALSIMAVFLGLALMAAEPFAF/ QGGLD	7.2 kDa
TraF	MKMKKAFGLVTLTSLMSSAFA/ QIPVT	23.8 kDa
TraG	MRKLTVLLFVAGTLLSM LAG/ CKEDK	8.3 kDa*
TraI	MKQAFVLLVVAL LAG/ CATNK	3.9 kDa*
TraK	MKRLALILALSAAFCAAQA/ ADVPQ	29.6 kDa
TraN	MEMPKWLKVVLVGVVFIATVGAVWLAGFFFFAFS/ KTNPF	87.7 kDa
TraP	MHKLVLSSIL LAA/ CSSPP	18.7 kDa*
TraQ	MKRLAITSALAAFFTAVPAHA/ NDGLF	18.4 kDa
Orf22	MTMNPAILKGAAGLIVAGALLAFAAGRQQAQA/ ADVAE	13.8 kDa

Fig. 5. Predicted signal peptides and cleavage sites of proteins encoded on plasmid pIPO2. Signal sequences were predicted using the SignalP software (Nielsen *et al.*, 1997). Cleavage sites are marked by a solidus (/). The putative lipoprotein consensus motif (Pugsley, 1993) is in bold and underlined. The molecular mass of the mature proteins is indicated. In the case of the predicted lipoproteins, the molecular mass of the lipid moiety has to be added (marked by an asterisk).

be necessary for the transfer. The collinear arrangement of the *tra* genes of pIPO2, the *virB* genes of the *Agrobacterium tumefaciens* Ti plasmid, the *tra* genes of pKM101 (IncN) and R388 (IncW), the *lhb* genes of *Legionella pneumophila*, the *vir* genes of *Brucella* spp. and of *Bartonella henselae*, and the *ptl* genes of *Bordetella pertussis* is conserved in each of these systems (Christie & Vogel, 2000; Zechner *et al.*, 2000). The VirB/Trb proteins of the Ti-plasmid/RP4 systems are thought to form a multicomponent transport pore spanning the bacterial envelope (Thorstenson *et al.*, 1993; Daugelavičius *et al.*, 1997; Grahn *et al.*, 2000).

Signal sequence prediction (Nielsen *et al.*, 1997) revealed that nine proteins encoded by pIPO2, including TraC, TraE and TraK, possess a signal peptide, like the corresponding proteins of the Ti plasmid (Fig. 5). Three proteins, including those encoded by TraG (the pro-

posed IncN-like entry exclusion function Eex) and TraI, contain signal sequences that resemble the signal peptidase II cleavage site found in lipoprotein precursors (Pugsley, 1993).

TraA of pIPO2 is proposed to encode the VirB1-like transglycosylase function involved in local destruction of the peptidoglycan layer (Bayer *et al.*, 1995). VirB1 analogues appear to act as optimization factors, since in conjugative transfer systems the transglycosylase function is not essential. TraC shows the features of precursors for pilins, i.e. its structure consists of a long signal peptide of more than 30 residues, and two transmembrane helices in the core of approximately 80, mostly hydrophobic, residues (Eisenbrandt *et al.*, 1999, 2000). IncP-1 TrbC and Ti plasmid VirB2 pilin precursor proteins were shown to be processed into cyclic polypeptides containing a head-to-tail covalent peptide bond

TraC	(pIPO2)	83	ILAGLLIGG-AAELAGYLLG
VirB2	(pSB102)	83	ILAGLLIGG-AAELARFLLG
VirB2	(<i>Br. suis</i>)	86	VLGGALVWGA-AAEIASYLLR
PtIA	(<i>Bor. pert.</i>)	83	VVLAGLLIGA-SAEIARYLLT
XF2055	(<i>X. fast.</i>)	115	QNSLSAITGQ-GAEITAFESH
TrbC	(RP4)	104	QNMSTFFGR-GAEIAALGNG
TrbC	(R751)	104	QNMSTFFGR-GAEIAALTDC
TrbC	(pRi1724)	92	ATNIVGLFGATGASIGLTSEQ
TrbC	(pNGR234a)	92	ATTIVGLFGATGASIGLTNDR
TrbC	(pTiC58)	93	ATQIVALFGATGASIGELHSQ
TrbC	(pTi-SAK.)	93	ATQIIALFGATGASIGEMEAR

Fig. 6. TraC of pIPO2 resembles the IncP-1 type pilin precursor protein TrbC. The sequences of TrbC homologues found in both conjugative DNA transfer systems and in type IV secretion systems are compared to the C-terminal part of RP4 TrbC precursor protein, which is the prepropilin being processed into cyclic pilin (processing sites are marked by arrows). Identical residues conserved in >80% of all sequences are indicated by black highlighting. Similar amino acids present in >80% of the sequences are marked dark grey highlighting. Light grey highlighting indicates similar or identical residues conserved within each of the two groups. The accession numbers for the sequences shown in this figure are Q9RPY3 [VirB2 (pSB102)], AAD56612 [VirB2 (*Brucella suis*)], AAB27428 [PtIA (*Bordetella pertussis*)], F82605 [XF2055 (*Xylella fastidiosa*)], AAA26429 [TrbC (RP4)], NP_044241 [TrbC (R751)], NP_066710 [TrbC (pRi1724)], P55396 [TrbC (pNGR234a)], NP_059759 [TrbC (pTiC58)] and NP_053256 [TrbC (pTi-SAKURA)].

between the N and C termini. The processing involves three different enzymes, two of which act on the C-terminal part of TrbC. Similarly, the pIPO2 TraC precursor may function as a substrate for polypeptide cyclization because the recognition site A-E-I/L-A required for this reaction seems to be conserved (Fig. 6). A plasmid-encoded function analogous to the signal peptidase-like function of IncP-1 TraF has not been found among pIPO2 genes but might be provided by a host-encoded component, consistent with the mechanism of Ti plasmid VirB2 processing. The conserved A-E-I/L-A recognition motif of these proteins is also found in several homologues of type IV secretion systems, indicating a close relationship between both IncP-1 pili and potential pili of macromolecular secretion systems (Eisenbrandt *et al.*, 2000). However, the alignments in Fig. 6 clearly indicate that the putative pilin of pIPO2 is not in the same cluster as the Ti plasmid-related TrbC proteins.

Plasmid pIPO2 TraE belongs to the TrbE/VirB4 family (Zechner *et al.*, 2000). The members of this family represent the largest proteins of the mating pair formation components, some of which consist of more than 800 aa. The oligomeric state of the corresponding Ti-encoded protein, VirB4, is at least dimeric (Dang *et al.*, 1999). VirB4 is associated with the cytoplasmic membrane and contains nucleotide-binding motifs of the Walker A and B types (Walker *et al.*, 1982). It was suggested that the protein possesses ATPase activity (Shirasu *et al.*, 1994). Thus, the VirB4 analogues might play an active role in the transport of substrates rather

than just having a structural role in the transport complex. Data on the function of the Ti plasmid-encoded proteins VirB3, VirB5, VirB6, VirB7, VirB8, VirB9 and VirB10, which have pIPO2 relatives, are rather scarce. Exceptions are that VirB5 might function extracellularly as a component associated with the T-pilus (Schmidt-Eisenlohr *et al.*, 1999) and that the Vir proteins B7, B8, B9 and B10 probably interact with each other by forming stoichiometric complexes (Das & Xie, 2000; Kumar *et al.*, 2000).

According to its sequence, the pIPO2 protein TraM belongs to the VirB11 family of traffic/secretion NTPases (Krause *et al.*, 2000a; Planet *et al.*, 2001). A few members of this superfamily [Ti VirB11, R388 TrwD, RP4 TrbB and the *Helicobacter pylori* 0525 (*Cag* α) protein] have been shown to hydrolyse NTPs, to form hexamers of identical subunits in solution, and seem to be associated with the cytoplasmic membrane (Krause *et al.*, 2000a, b). The three-dimensional structure of the *H. pylori* protein *Cag* α shows the form of a six-clawed grapple, which is supposed to open and close upon ATP binding/hydrolysis (Yeo *et al.*, 2000). This property suggests that the hexamers may function as chaperones or have pore character, and that they may be involved in substrate transmission into target cells.

Plasmid pIPO2 belongs to a new family of environmentally important plasmids

The gene organization of the putative conjugative transfer region (Tra region) of pIPO2 has been found to be conserved in pSB102 (Schneiker *et al.*, 2001) and in the *X. fastidiosa* plasmid pXF51 (Marques *et al.*, 2001), indicating that the three Tra systems belong to the same class (Fig. 7). The interruption of the Mpf gene clusters by the proposed DNA processing gene *traB/top* next to *traA/traL* is unique for pIPO2, pSB102 and also for pXF51. The sequence-related *traE* gene has been found in IncP-1 plasmids in the Tra1 region (which encodes DNA processing genes) as a *tra* gene that is non-essential at least for intraspecific *E. coli* matings (Lessl *et al.*, 1993), even in the absence of chromosomally encoded *topA* and *topB* genes (G. Schröder & E. Lanka, unpublished data). Complementation by other chromosome-encoded topoisomerases, for instance, the topoisomerase IV (Deibler *et al.*, 2001), may be the reason for this finding.

The presence of long repetitive sequences in pIPO2, pSB102 and pXF51 is a striking feature of this group and differentiates them from other sequenced broad-host-range plasmids, such as those of the IncP-1 group (Fig. 7). Each of the three plasmids carries at least one pair of unrelated direct repeats spanning several hundreds of nucleotides. The differences in the nucleotide sequences apply for repeats within a plasmid, as well as for repeats among the three plasmids. Each pair of unrelated direct repeats within pSB102 and pIPO2 carries nearly perfect inverted repeats. It is noteworthy that in both plasmids at least one pair of inverted repeats flanks a varying number of ORFs of unknown function. The function of

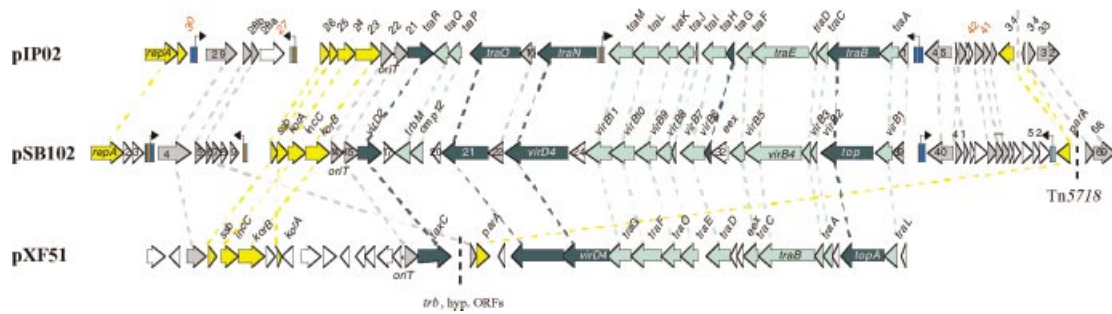


Fig. 7. Comparison of the genetic organization of pIPO2, pSB102 and pXF51, which share common sets of genes involved in plasmid transfer and maintenance. Linear genetic maps of pIPO2 (39815 bp), pSB102 (55578 bp) and pXF51 (51158 bp) are presented. ORFs are shown by arrows indicating the direction of transcription. Homologous ORFs are connected by dashed lines. The colours of the arrows and dashed lines indicate different postulated functions of the deduced proteins: light green, mating pair formation functions; dark green, DNA processing functions; yellow, replication and maintenance. ORFs encoding proteins of unknown function are shown in grey if they are related to any other ORF of plasmids pSB102 or pXF51, or are shown without colour if they are unrelated among the three plasmids. Genes only present in pIPO2 are typed in red. The approximate position of transposon insertions in pIPO2 and pSB102 is shown. Direct repeats within the plasmid sequences are indicated as rectangles, with different shades of blue indicating the respective pair of repeats in individual plasmids. Inverted repeats that are part of different direct repeats are indicated by black arrows.

the direct and inverted repeats remains unknown. These structures might promote rearrangements and thus might have played a role in the evolution of these plasmids. A possibly related phenomenon is observed in at least some of the IncP-1 β plasmids where there are clusters of tandemly repeated inverted repeats that border major sectors of the plasmid genome (Thorsted *et al.*, 1998). It is suggested that these may have played a role in the evolution of the plasmid genome.

The compact transfer regions of the three plasmids pIPO2, pSB102 and pXF51 contain elements related to components of the Ti plasmid T-DNA transfer system and the conjugative transfer system of IncP-1 plasmids. This modular design principle of the combination of functional units of different phylogenetic origin apparently is valid for most of the conjugative transfer systems (Pansegrau & Lanka, 1996; Zechner *et al.*, 2000). Although related to the Ti plasmid VirB proteins, the proposed mating pair formation components of the three plasmids are much more closely related to proteins of the type IV secretion system of *Brucella* spp. (Table 2). The latter are thought to be responsible for the secretion or delivery to mammalian cells of as yet unknown substrates, most likely proteinaceous toxins.

An additional common feature of pIPO2 and pSB102 is the existence of a replication initiator protein, RepA, which is embedded by putative ORFs of unknown function. However, both plasmids, pIPO2 and pSB102, lack iteron-like structures supposed to function as part of the vegetative origin, i.e. as binding sites for RepA. Two regions in pIPO2 and pSB102 contain clusters of predicted ORFs with unknown functions. Strikingly, many of the ORFs of unknown function are related between pIPO2 and pSB102, suggesting a possible involvement of these ORFs in host/plasmid functioning in natural habitats.

In conclusion, plasmid pIPO2 fits well into the new class of environmental plasmids defined by plasmid pSB102 (Schneiker *et al.*, 2001). Interestingly, pIPO2, as well as pSB102 and pXF51, resides in plant-associated bacteria. This further supports the hypothesis that all three plasmids might be variants of an archetypical plasmid class associated with phytosphere bacteria.

Determination of the distribution of plasmid pIPO2

Plasmid pIPO2 was initially isolated by the *E. coli*/R. *eutropha*-based triparental exogenous plasmid isolation system (Top *et al.*, 1994), selecting for plasmids with IncQ plasmid mobilizing capacity. pIPO2 was hereby isolated from extracts of soil associated with wheat seedlings, at just above the detection limit of 10^{-9} per recipient (van Elsas *et al.*, 1998). In contrast, exogenous isolation of plasmids from rhizosphere soil of field-grown mature wheat, maize and sugarbeet repeatedly failed under the same conditions. This indicated a lower incidence of functional pIPO2-like plasmids in the respective soil or rhizospheric bacterial communities compared to the numbers associated with young wheat plants.

In the present work, a direct molecular screen for pIPO2 sequences in total microbial community DNA, obtained from replicate samples from a range of different environmental habitats, was performed (Table 1). The previously developed PCR primer system I shown to be specific for pIPO2 (van Elsas *et al.*, 1998), as well as primer system II (which amplified part of the *repA* gene homologue), was used. Primer system II was, on theoretical grounds, specific for plasmid pIPO2, as BLAST searches against the whole GenBank/EMBL database did not reveal any significant match with any sequence, including those of plasmids pSB102, pXF51, R388, RP4 and R751. In addition, laboratory tests showed that the

broad-host-range plasmids R388, RN3, RP4, R751, RSF1010, pIE723 and pIE639, as well as a range of 10 *E. coli* narrow-host-range plasmids, did not produce amplicons, whereas plasmid pIPO2 DNA consistently produced the expected PCR fragment. pIPO2 DNA added to each environmental DNA extract consistently caused amplification in direct correlation with the concentration of added target DNA, i.e. progressive dilution of added target DNA in the environmental extract yielded progressively fainter signals which ultimately became extinct. This established a detection threshold for each habitat, which was calculated to be about 5×10^3 plasmids per g of soil or per ml of water.

The detection of pIPO2 prevalence in the different habitats was consistent in primer systems I and II (Table 1). Evidence for the occurrence of pIPO2 was obtained for seven of the 10 rhizospheres tested and for two of the 11 bulk soils studied, whereas such evidence was not found for seawater, wastewater, manure and compost (added to soil) (Table 1). Thus, the only habitats with consistent evidence for pIPO2-specific sequences were the rhizospheres of young wheat, rice and maize, and those of oats, grass, tomato and cauliflower, as well as soils previously used for cultivation of crop plants such as the Lovinkhoeve (samples from five time points, but not in three other samples) and Wildekamp arable land (summer 2000 sample, oats grown in the field) samples. On the basis of these results, we postulate that pIPO2 either has a preference for host organisms that thrive in the rhizospheres of a variety of crop plants or confers a property on its host that promotes survival in these environments. Conversely, pIPO2 appears not to survive in hosts that prefer bulk soils or it does not promote survival of its host in bulk soils.

A number of microbial groups are known to be subject to changes in relative dominance in the rhizosphere of crop plants such as wheat, maize, tomato and potato. For instance, Miller *et al.* (1989) showed that fluorescent pseudomonads are avid responders to young roots of wheat and maize, enhancing their abundance. Upon plant ageing, these organisms were shown to decline. This principle can most likely be extended to many crop plants, including the ones used in this study (wheat, maize, tomato and cauliflower). Smalla *et al.* (2001) recently showed that the dynamics of microbial populations associated with field-grown strawberry, oilseed rape and potato was strongly dependent on plant type and on season, i.e. most likely due to plant developmental stage. By and large, and possibly a bit speculatively, one can assume that a range of bacterial groups, which includes the Gram-negative and Gram-positive rhizosphere-strategists (such as pseudomonads, stenotrophomonads, several enterics, bacilli and paenibacilli), are able to respond to the enhanced availability of nutrients in root exudates by enhancing their prevalence. Such a response can be different per bacterial group and per plant, and is most likely highly dependent on the types of compounds present in the root exudates. Thus, we suggest that the detection of plasmids at different stages of seedling development probably re-

flects changes in the composition of the associated microbial community.

Conclusion and prospects for further work

The broad-host-range plasmid type represented by pIPO2 was found to be a compact genetic structure containing mainly conjugation and inheritance functions. A region containing adjacent ORFs that encode small proteins with unknown function was identified (ORF32–ORF45). This region is homologous to a similar region found in plasmid pSB102 (Schneiker *et al.*, 2001). The potential phenotype encoded by the ORFs of this region might be identified by studies that address host/plasmid functioning in the natural habitat, i.e. including the plant as a major driving factor. Since most other pIPO2 functions are related to plasmid replication, maintenance and conjugative transfer functions, the plasmid at present appears to be 'cryptic' (van Elsland *et al.*, 1998). Therefore, expression of transfer-related genes might cause an enhanced metabolic load without a balancing advantage, explaining the apparently reduced survival in bulk soils of bacteria carrying pIPO2.

Given its similarity to the novel broad-host-range conjugative plasmid pSB102 and to the *X. fastidiosa* plasmid pXF51, we propose that pIPO2 is an example of a novel class of plasmids that are prevalent in hosts that associate in symbiotic or in other ways with plants. The direct molecular evidence that pIPO2 is prevalent in the rhizosphere of several crop plants is consistent with this hypothesis and pinpoints the original host of pIPO2 as an organism that associates primarily with plants and takes advantage (currently unknown) of the presence of pIPO2. The most obvious possibility is that the clustered and probably coordinately expressed *orfs* of unknown function (*orfs* 38–45) encode a metabolic pathway or apparatus that confers an advantageous phenotype to its host, which is perhaps specific to plant-associated communities. One aspect of future work will be to define what this phenotype is.

ACKNOWLEDGEMENTS

This study was supported by a Concerted Action grant from the European Community DGXII (MECBAD consortium), as well as by the RESERVOIR project. We further thank Anneke Keijzer-Wolters for preparing the clones, and Rogier Doornbos and Biserka Sveshterova for assisting in some of the experiments.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Balzer, D., Pansegrau, W. & Lanka, E. (1994). Essential motifs of relaxase (TraI) and TraG proteins involved in conjugative transfer of Plasmid RP4. *J Bacteriol* **176**, 4285–4295.
- Bayer, M., Eferle, R., Zellnig, G., Teferle, K., Dijkstra, A., Koraimann, G. & Högenauer, G. (1995). Gene 19 of plasmid R1 is required for

- both efficient conjugative DNA transfer and bacteriophage R17 infection. *J Bacteriol* **177**, 4279–4288.
- Bignell, C. & Thomas, C. M. (2001).** The bacterial ParA–ParB partitioning proteins. *J Biotechnol* **91**, 1–34.
- Christie, P. J. & Vogel, J. P. (2000).** Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* **8**, 354–360.
- Cobbe, N. & Heck, M. M. S. (2000).** SMCs in the world of chromosome biology – from prokaryotes to higher eukaryotes. *J Struct Biol* **129**, 123–143.
- Dang, T. A., Zhou, X. R., Graf, B. & Christie, P. J. (1999).** Dimerization of the *Agrobacterium tumefaciens* VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. *Mol Microbiol* **32**, 1239–1253.
- Das, A. & Xie, Y. H. (2000).** The *Agrobacterium* T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another. *J Bacteriol* **182**, 758–763.
- Daugelavičius, R., Bamford, J. K., Grahn, A. M., Lanka, E. & Bamford, D. H. (1997).** The IncP plasmid-encoded cell-envelope-associated DNA transfer complex increases cell permeability. *J Bacteriol* **179**, 5195–5202.
- Deibler, R. W., Rahmati, S. & Zechiedrich, E. L. (2001).** Topoisomerase IV, alone, unknots DNA in *E. coli*. *Genes Dev* **15**, 748–761.
- de la Cruz, F. & Davies, J. (2000).** Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**, 128–133.
- de Lorenzo, V., Herrero, M., Jakubzik, U. & Timmis, K. N. (1990).** Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in Gram-negative eubacteria. *J Bacteriol* **172**, 6568–6572.
- Dodd, I. B. & Egan, J. B. (1990).** Improved detection of helix–turn–helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* **18**, 5019–5026.
- Duarte, G. F., Rosado, A. S., Seldin, L., De Araujo, W. & van Elsas, J. D. (2001).** Analysis of bacterial community structure in sulfurous-oil-containing soils and detection of species carrying dibenzothiophene desulfurization (*dsz*) genes. *Appl Environ Microbiol* **67**, 1052–1062.
- Egelman, E. H. (2001).** Pumping DNA. *Nature* **409**, 573–575.
- Eisenbrandt, R., Kalkum, M., Lai, E. M., Lurz, R., Kado, C. I. & Lanka, E. (1999).** Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. *J Biol Chem* **274**, 22548–22555.
- Eisenbrandt, R., Kalkum, M., Lurz, R. & Lanka, E. (2000).** Maturation of IncP pilin precursors resembles the catalytic Dyad-like mechanism of leader peptidases. *J Bacteriol* **182**, 6751–6761.
- Gerdes, K., Ayora, S., Canosa, I. & 10 other authors (2000).** In *The Horizontal Gene Pool – Bacterial Plasmids and Gene Spread*, pp. 49–85. Edited by C. M. Thomas. Amsterdam, The Netherlands: Harwood Academic Publishers.
- Golub, E. I. & Low, K. B. (1986).** Unrelated conjugative plasmids have sequences which are homologous to the leading region of the F factor. *J Bacteriol* **166**, 670–672.
- Gomis-Rüth, F. X., Moncalián, G., Perez-Luque, R., Gonzalez, A., Cabezón, E., de la Cruz, F. & Coll, M. (2001).** The bacterial conjugation protein TrwB resembles ring helicases and F₁-ATPase. *Nature* **409**, 637–641.
- Grahn, A. M., Haase, J., Bamford, D. H. & Lanka, E. (2000).** Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: implications for related macromolecule transport systems. *J Bacteriol* **182**, 1564–1574.
- Greated, A., Titiok, M., Krasowiak, R., Fairclough, R. & Thomas, C. M. (2000).** The replication and stable inheritance functions of IncP-9 plasmid pM3. *Microbiology* **146**, 2249–2258.
- Hill, K. E., Weightman, A. J. & Fry, J. C. (1992).** Gene transfer in the aquatic environment: persistence and mobilization of the catabolic recombinant plasmid pD10 in the epilithon. *Appl Environ Microbiol* **58**, 1292–1300.
- Hill, K. E., Marchesi, J. R. & Fry, J. C. (1996).** Conjugation and mobilization in the epilithon. In *Molecular Microbial Ecology Manual*, pp. 1–28. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Jagura-Burdzy, G. & Thomas, C. M. (1992).** *kfrA* gene of broad host range plasmid RK2 encodes a novel DNA-binding protein. *J Mol Biol* **225**, 651–660.
- Jovanovic, O. S., Ayres, E. K. & Figurski, D. H. (1992).** The replication initiator operon of promiscuous plasmid RK2 encodes a gene that complements an *Escherichia coli* mutant defective in single-stranded DNA-binding protein. *J Bacteriol* **174**, 4842–4846.
- Krause, S., Bárcena, M., Pansegrau, W., Lurz, R., Carazo, J. M. & Lanka, E. (2000a).** Sequence-related protein export NTPases encoded by the conjugative transfer region of RP4 and by the *cag* pathogenicity island of *Helicobacter pylori* share similar hexameric ring structures. *Proc Natl Acad Sci USA* **97**, 3067–3072.
- Krause, S., Pansegrau, W., Lurz, R., de la Cruz, F. & Lanka, E. (2000b).** Enzymology of type IV macromolecule secretion systems: the conjugative transfer regions of plasmids RP4 and R388 and the *cag* pathogenicity island of *Helicobacter pylori* encode structurally and functionally related nucleoside triphosphate hydrolases. *J Bacteriol* **182**, 2761–2770.
- Kumar, R. B., Xie, Y. H. & Das, A. (2000).** Subcellular localization of the *Agrobacterium tumefaciens* T-DNA transport pore proteins: VirB8 is essential for the assembly of the transport pore. *Mol Microbiol* **36**, 608–617.
- Kurland, C. G. (2000).** Something for everyone – horizontal gene transfer in evolution. *EMBO Rep* **1**, 92–95.
- Kurtz, S. & Schleiermacher, C. (1999).** REPUTER – fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**, 426–427.
- Lawrence, J. C. & Ochman, H. (1998).** Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95**, 9413–9417.
- Lessl, M., Pansegrau, W. & Lanka, E. (1992).** Relationship of DNA transfer systems: essential transfer factors of plasmid RP4, Ti, and F share common sequences. *Nucleic Acids Res* **20**, 6099–6100.
- Lessl, M., Balzer, D., Weyrauch, K. & Lanka, E. (1993).** The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J Bacteriol* **175**, 6415–6425.
- Li, Z., Hiasa, H., Kumar, U. & DiGate, R. J. (1997).** The *traE* gene of RP4 encodes a homologue of *Escherichia coli* DNA topoisomerase III. *J Biol Chem* **272**, 19582–19587.
- Lilley, A. K. & Bailey, M. J. (1997).** The acquisition of indigenous plasmids by a genetically marked pseudomonad population colonizing the sugar beet phytosphere is related to local environmental conditions. *Appl Environ Microbiol* **63**, 1577–1583.
- Lilley, A. K., Fry, J. C., Day, M. J. & Bailey, M. J. (1994).** *In situ* transfer of an exogenously isolated plasmid between *Pseudomonas* spp. in sugar beet rhizosphere. *Microbiology* **140**, 27–33.

- Lin, M., Smalla, K., Heuer, H. & van Elsas, J. D. (2000). Effect of an *Alcaligenes faecalis* inoculant strain on bacterial communities in flooded soil microcosms planted with rice seedlings. *Appl Soil Ecol* **15**, 211–225.
- Macartney, D. P., Williams, D. R., Stafford, T. & Thomas, C. M. (1997). Divergence and conservation of the partitioning and global regulation functions in the central control region of the IncP plasmids RK2 and R751. *Microbiology* **143**, 2167–2177.
- Marques, M. V., da Silva, A. M. & Gomes, S. L. (2001). Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. *Plasmid* **45**, 184–199.
- Miller, H. J., Henken, G. & van Veen, J. A. (1989). Variation and composition of bacterial populations in the rhizospheres of maize, wheat and grass cultivars. *Can J Microbiol* **35**, 656–660.
- Moncalián, G., Cabezón, E., Alkorta, I., Valle, M., Moro, F., Valpuesta, J. M., Goñi, F. M. & de la Cruz, F. (1999). Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *J Biol Chem* **274**, 36117–36124.
- Myers, E. W. & Miller, W. (1988). Optimal alignments in linear space. *Comput Appl Biosci* **4**, 11–17.
- Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Sys* **8**, 581–599.
- O'Callaghan, D., Cazevieuille, C., Allardet-Servent, A., Boschiroli, M. L., Bourg, G., Foulongne, V., Frutos, P., Kulakov, Y. & Ramuz, M. (1999). A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion systems is essential for intracellular survival of *Brucella suis*. *Mol Microbiol* **33**, 1210–1220.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Okumura, M. S. & Kado, C. I. (1992). The region essential for efficient autonomous replication of pSa in *Escherichia coli*. *Mol Gen Genet* **235**, 55–63.
- Pansegrau, W. & Lanka, E. (1996). Enzymology of DNA transfer by conjugative mechanisms. *Progr Nucleic Acids Res Mol Biol* **54**, 197–251.
- Pansegrau, W., Schoumacher, F., Hohn, B. & Lanka, E. (1993). Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of *Agrobacterium tumefaciens* Ti plasmids: analogy to bacterial conjugation. *Proc Natl Acad Sci USA* **90**, 11538–11542.
- Pansegrau, W., Lanka, E., Barth, P. T. & 7 other authors (1994a). Complete nucleotide sequence of Birmingham IncP α plasmids: compilation and comparative analysis of sequence data. *J Mol Biol* **239**, 623–663.
- Pansegrau, W., Schröder, W. & Lanka, E. (1994b). Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (Tral) of conjugative plasmid RP4. *J Biol Chem* **269**, 2782–2789.
- Planet, P. J., Kachlany, S. C., DeSalle, R. & Figurski, D. H. (2001). Phylogeny of genes for secretion NTPases: identification of the widespread *tadA* subfamily and development of a diagnostic key for gene classification. *Proc Natl Acad Sci USA* **98**, 2503–2508.
- Preston, K. E., Radomski, C. C. A. & Venezia, A. (2000). Nucleotide sequence of a 7-kb fragment of pACM1 encoding an IncM DNA primase and other proteins associated with conjugation. *Plasmid* **44**, 12–23.
- Pugsley, A. P. (1993). The complete general secretory pathway in Gram-negative bacteria. *Microbiol Rev* **57**, 50–108.
- Pukall, R., Tschäpe, H. & Smalla, K. (1996). Monitoring the spread of broad host and narrow host range plasmids in soil microcosms. *FEMS Microbiol Ecol* **20**, 53–66.
- Reese, M. G., Harris, N. L. & Eckman, F. H. (1996). Large scale sequencing specific neural networks for promoter and splice site recognition. In *Biocomputing: Proceedings of the 1996 Pacific Symposium*. Edited by L. Hunter & T. E. Klein. Singapore: World Scientific Publishing.
- Rosado, A. S., Duarte, G. F., Seldin, L. & van Elsas, J. D. (1998). Genetic diversity of *nifH* gene sequences in *Paenibacillus azotofixans* strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR-amplified gene fragments. *Appl Environ Microbiol* **64**, 2770–2779.
- Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995). Prediction of helical transmembrane segments at 95% accuracy. *Protein Sci* **4**, 521–533.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Schmidt-Eisenlohr, H., Domke, N., Angerer, C., Wanner, G., Zambryski, P. C. & Baron, C. (1999). Vir proteins stabilize VirB5 and mediate its association with the T pilus of *Agrobacterium tumefaciens*. *J Bacteriol* **181**, 7485–7492.
- Schmiederer, M. & Anderson, B. (2000). Cloning, sequencing, and expression of three *Bartonella henselae* genes homologous to the *Agrobacterium tumefaciens* VirB region. *DNA Cell Biol* **19**, 141–147.
- Schneiker, S., Keller, M., Dröge, M., Lanka, E., Pühler, A. & Selbitschka, W. (2001). The genetic organization and evolution of the broad-host-range mercury resistance plasmid pSB102 isolated from a microbial population residing in the rhizosphere of alfalfa. *Nucleic Acids Res* **29**, 5169–5181.
- Shirasu, K., Koukolíková-Nicola, Z., Hohn, B. & Kado, C. I. (1994). An inner-membrane-associated virulence protein essential for T-DNA transfer from *Agrobacterium tumefaciens* to plants exhibits ATPase activity and similarities to conjugative transfer genes. *Mol Microbiol* **11**, 581–588.
- Sieira, R., Comerci, D. J., Sanchez, D. O. & Ugalde, R. A. (2000). A homologue of an operon required for DNA transfer in *Agrobacterium* is required in *Brucella abortus* for virulence and intracellular multiplication. *J Bacteriol* **182**, 4849–4855.
- Simpson, A. J. G., Reinach, F. C., Arruda, & 113 other authors (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151–157.
- Smalla, K., Cresswell, N., Mendonca-Hagler, L. C., Wolters, A. & van Elsas, J. D. (1993). Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. *J Appl Bacteriol* **74**, 78–85.
- Smalla, K., Osborn, A. M. & Wellington, E. M. H. (2000). Isolation and characterisation of plasmids. In *The Horizontal Gene Pool – Bacterial Plasmids and Gene Spread*, pp. 207–248. Edited by C. M. Thomas. Amsterdam, The Netherlands: Harwood Academic Publishers.
- Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H. & Berg, G. (2001). Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Appl Environ Microbiol* **67**, 4742–4751.
- Smit, E., van Elsas, J. D., van Veen, J. A. & de Vos, W. M. (1991). Detection of plasmid transfer from *Pseudomonas fluorescens* to indigenous bacteria in soil by using bacteriophage Φ R2f from donor counterselection. *Appl Environ Microbiol* **57**, 3482–3488.

- Smit, E., Venne, D. & van Elsas, J. D. (1993). Mobilization of a recombinant IncQ plasmid between bacteria on agar and in soil via co-transfer or retrotransfer. *Appl Environ Microbiol* **59**, 2257–2263.
- Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**, 233–241.
- Szpirer, C., Top, E., Couturier, M. & Mergeay, M. (1999). Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. *Microbiology* **145**, 3321–3329.
- Thomas, C. M. (editor) (2000a). *The Horizontal Gene Pool – Bacterial Plasmids and Gene Spread*. Amsterdam, The Netherlands: Harwood Academic Publishers.
- Thomas, C. M. (2000b). Paradigms of plasmid organisation. *Mol Microbiol* **37**, 485–491.
- Thorsted, P. B., Macartney, D. P., Akhtar, P. & 9 other authors (1998). Complete sequence of the IncP β plasmid R751: implications for evolution and organization of the IncP backbone. *J Mol Biol* **282**, 969–990.
- Thorstenson, Y. R., Kuldau, G. A. & Zambryski, P. C. (1993). Subcellular localization of seven VirB proteins of *Agrobacterium tumefaciens*: implications for the formation of a T-DNA transport structure. *J Bacteriol* **175**, 5233–5241.
- Top, E., de Smet, I., Verstraete, W., Dijkmans, R. & Mergeay, M. (1994). Exogenous isolation of mobilizing plasmids from polluted soils and sludges. *Appl Environ Microbiol* **60**, 831–839.
- van Elsas, J. D. & Smalla, K. (1995). Extraction of microbial community DNA from soils. In *Molecular Microbial Ecology Manual*, pp. 1.3.3: 1–11. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- van Elsas, J. D., Trevors, J. T. & Starodub, M.-E. (1988). Bacterial conjugation between pseudomonads in the rhizosphere of wheat. *FEMS Microbiol Ecol* **53**, 299–306.
- van Elsas, J. D., McSpadden Gardener, B. B., Wolters, A. C. & Smit, E. (1998). Isolation, characterization, and transfer of cryptic gene-mobilizing plasmids in the wheat rhizosphere. *Appl Environ Microbiol* **64**, 880–889.
- van Elsas, J. D., Fry, J., Hirsch, P. & Molin, S. (2000a). Ecology of plasmid transfer and spread. In *The Horizontal Gene Pool – Bacterial Plasmids and Gene Spread*, pp. 175–206. Edited by C. M. Thomas. Amsterdam, The Netherlands: Harwood Academic Publishers.
- van Elsas, J. D., Kastelein, P., van Bekkum, P., van der Wolf, J. M., de Vries, P. M. & van Overbeek, L. S. (2000b). Survival of *Ralstonia solanacearum* biovar 2, the causative agent of potato brown rot, in field and microcosm soils in temperate climates. *Phytopathology* **90**, 1358–1366.
- van Overbeek, L. S., van Veen, J. A. & van Elsas, J. D. (1995). Induced reporter gene activity, enhanced stress resistance, and competitive ability of a genetically modified *Pseudomonas fluorescens* strain released into a field plot planted with wheat. *Appl Environ Microbiol* **63**, 1965–1973.
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945–951.
- Yeo, H. J., Savvides, S. N., Herr, A. B., Lanka, E. & Waksman, G. (2000). Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol Cell* **6**, 1461–1472.
- Zechner, E. L., de la Cruz, F., Eisenbrandt, R. & 8 other authors (2000). Conjugative transfer processes. In *The Horizontal Gene Pool – Bacterial Plasmids and Gene Spread*, pp. 87–174. Edited by C. M. Thomas. Amsterdam, The Netherlands: Harwood Academic Publishers.

Received 12 November 2001; revised 18 December 2001; accepted 21 January 2002.