Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution

Raúl Fernández-López[1], M. Pilar Garcillán-Barcia[1], Carlos Revilla[1], Miguel Lázaro[2], Luis Vielva[2] & Fernando de la Cruz[1]

[1]Departamento de Biología Molecular (Unidad Asociada al C.I.B., C.S.I.C.); and [2]Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain

Correspondence: Fernando de la Cruz, Departamento de Biología Molecular (Unidad Asociada al C.I.B., C.S.I.C.), Universidad de Cantabria, C. Herrera Oria s/n, 39011 Santander, Spain. Tel.: +34 942 2019 42; fax: +34 942 2019 45; e-mail: delacruz@unican.es

Received 31 March 2006; revised 30 June 2006; accepted 30 June 2006.
First published online 5 October 2006.

DOI:10.1111/j.1574-6976.2006.00042.x

Editor: Ramon Diaz Orejas

Abstract

Plasmids cannot be understood as mere tools for genetic exchange: they are themselves subject to the forces of evolution. Their genomic and phylogenetic features have been less studied in this respect. Focusing on the IncW incompatibility group, which includes the smallest known conjugative plasmids, we attempt to unveil some common trends in plasmid evolution. The functional modules of IncW genetic backbone are described, with emphasis on their architecture and relationships to other plasmid groups. Some plasmid regions exhibit strong phylogenetic mosaicism, in striking contrast to others of unusual synteny conservation. The presence of genes of unknown function that are widely distributed in plasmid genomes is also emphasized, exposing the existence of ill-defined yet conserved plasmid functions. Conjugation is an essential hallmark of IncW plasmid biology and special attention is given to the organization and evolution of its transfer modules. Genetic exchange between plasmids and their hosts is analysed by following the evolution of the type IV secretion system. Adaptation of the *trw* conjugative machinery to pathogenicity functions in *Bartonella* is discussed as an example of how plasmids can change their host *modus vivendi*. Starting from the phage paradigm, our analysis articulates novel concepts that apply to plasmid evolution.

Introduction

Plasmids naturally fall into discrete clusters, which can be identified based on analysis of DNA sequence similarity, genetic organization and phenotypic properties. Because members of a given cluster share a mechanism for DNA replication or partition control, they cannot coexist stably in the same host. They are said to be incompatible, so plasmid clusters are traditionally referred to as incompatibility groups. Plasmids belonging to the same incompatibility group are frequently found in bacterial natural isolates. For instance, hundreds of conjugative plasmids carrying genes for antibiotic resistance and able to replicate in *Enterobacteria* were classified in about 20 incompatibility groups (Datta & Hughes, 1983). More recently, the appearance of many plasmid genomes in sequence databases has allowed a new type of molecular analysis *in silico*, and has brought about comparative genomics and evolutionary analysis of the relationships between incompatibility groups (Couturier *et al.*, 1988; Osborn & Boltner, 2002). Such comparative analysis may help our understanding of the general principles that govern the genetic organization, function and evolution of plasmid genomes. The purpose of this review is to analyse one of those clusters, the IncW plasmids. These plasmids have the smallest genome size among natural conjugative plasmids studied so far. Thus, they represent what can be logically considered a minimal conjugative plasmid genome. Their analysis reveals that even these minimal plasmid genomes contain a number of genes that are classically considered as nonessential but, being present in many unrelated plasmids, can imply more fundamental functions than initially anticipated. The IncW prototype, plasmid R388, contains 33 926 bp. IncW plasmids have been found in a wide variety of bacterial species (Table 1), so they are considered to have a broad host range. Small size and broad host range make the IncW backbone a good example for the analysis of the genetic organization and phylogeny of a conjugative plasmid genome.

The genetic organization of mobile DNA elements such as plasmids and bacteriophages differs from the chromosomal

Table 1. Host range of IncW plasmids

Bacterial class	Species	Plasmid	References
Alphaproteobacteria	*Agrobacterium tumefaciens*	pSa,R388,R7K	Loper & Kado (1979), Chernin *et al*. (1984)
	Agrobacterium rhizogenes	pSa	Vlasak & Ondrej (1985)
	Brucella abortus	pSA	Rigby & Fraser (1989)
	Rhizobium leguminosarum	pSa	Tait *et al*. (1982b)
	Rhizobium trifolii	pSa	Tait *et al*. (1982b)
	Zymomonas mobilis	pSa	Strzelecki *et al*. (1987)
	Alcaligenes eutrophus	pSa	Tait *et al*. (1982b)
Betaproteobacteria	*Methylophilus methylotrophus*	pSa	Nagate *et al*. (1978)
Gammaproteobacteria	*Acinetobacter calcoaceticus*	pSa,R388	Hinchliffe & Vivian (1980)
	Aeromonas liquefasciens	RA2, RA3	Hedges & Datta (1971)
	Aeromonas salmonicida	pSa	Aoki *et al*. (1971)
	Citrobacter freundii	pIE1077	Gotz *et al*. (1996)
	Enterobacter	pHH1191,pHH1193,pHH1303, pHH1307	Valentine & Kado (1989)
	Enterobacter cloacae	pIE1080,pIE1081	Gotz *et al*. (1996)
	Erwinia amylovora	pSa	Tait *et al*. (1983)
	Erwinia carotovora	pSa	Shaw & Kado (1986)
	Erwinia rubrifasciens	pSa	Shaw & Kado (1986)
	Escherichia coli	R388	Datta & Hedges (1971)
	Klebsiella sp.	pSa	Datta & Hedges (1972b)
	Klebsiella K9	pHH720	Datta *et al*. (1980)
	Klebsiella pneumoniae	pIE522	Gotz *et al*. (1996)
	Legionella pneumophila	pSa	Dreyfus & Iglewski (1985)
	Proteus rettgeri	pSA,R7K	Coetzee *et al*. (1972)
	Proteus mirabilis	pSa	Datta & Hedges (1972a)
	Providencia stuartii	pSa	Tait *et al*. (1983)
	Pseudomonas aeruginosa	pSa,R388,R7K	Bradley & Cohen (1976)
	Pseudomonas fluorescens	pSa	Tait *et al*. (1983)
	Pseudomonas glumae	pSa	Smith *et al*. (1979)
	Pseudomonas putida	pSa	Tait *et al*. (1983)
	Pseudomonas solanacearum	R388	Morales & Sequeira (1985)
	Pseudomonas syringae	pSa	Shaw & Kado (1986)
	Pseudomonas stutzeri	pSa	Tait *et al*. (1983)
	Salmonella dublin	pIE321, pIE1105, pIE1097, pIE1105	Smalla *et al*. (2001)
	Salmonella enteritidis	pSa	Causey & Brown (1978)
	Salmonella typhimurium	pSa,R388,R7K,pIE306,pIE1056, pIE1095,pIE1098	Watanabe *et al*. (1968), Bradley & Cohen (1976), Sanderson *et al*. (1981), Gotz *et al*. (1996)
	Serratia marcescens	pSa	Tait *et al*. (1983)
	Shigella flexneri	pSa,R388,R7K,pIE774	Bradley & Cohen (1976); Gotz *et al*. (1996)
	Shigella sonnei	pIE384 pIE385 pIE1078	Gotz *et al*. (1996)
	Stenotrophomonas maltophilia	R388	Avison *et al*. (2000)
	Vibrio cholerae	pSa	Hamilton-Miller (1979)
	Xanthomonas campestris	pSa	Shaw & Kado (1986)
	Xanthomonas campestris	pXV2	Wu & Tseng (2000)
	Salmonella ordonez	pSa	Valentine & Kado (1989)
Deltaproteobacteria	*Myxococcus virescens*	pBL1005,pBL1009, pBL1010, pBL1011,pBL1013,pBL1014, pBL1016,pBL1017,pBL1021, pBL1024	Morris *et al*. (1978)
	Myxococcus xanthus	pSa	von Kruger & Parish (1981)
Bacteroidetes	*Rhodothermus marinus*	pRM21	Ernstsson *et al*. (2003)

organization of their hosts. Plasmid organization, reviewed in detail by Thomas (2000), is based on the confluence and clustering of different survival and propagation functions, arranged in functional modules. Modular architecture has well-known functional advantages (Toussaint & Merlin, 2002), and modular organization is not a special hallmark of plasmids nor of mobile DNA elements, but a common feature of living systems (Rainey & Cooper, 2004). Nevertheless, in the case of plasmids and phages, modular architecture is a consequence of evolutionary mosaicism (Brussow & Hendrix, 2002). Mosaicism refers to the quality of being formed by the accretion of parts of different origin, and in phage evolution refers specifically to the different phylogenies of the component parts of a viral genome (Hendrix *et al.*, 2000). These component parts are usually clustered in the genome, and therefore they are considered to constitute both functional and evolutionary modules. Genetic mosaicism has been described in the genomic organization of mycobacteriophages (Pedulla *et al.*, 2003), lambdoid bacteriophages (Juhala *et al.*, 2000) and *Staphylococcus aureus* bacteriophages (Kwan *et al.*, 2005).

Following the phage paradigm, our analysis of the IncW backbone will initially consider the plasmid genome as a mosaic of functional modules with various phylogenetic origins. We will describe each module from a functional as well as from an evolutionary point of view, emphasizing, where possible, the links between phylogeny and function. We will describe in detail both replication and transfer functions, but paying also attention to additional functions encoded in the plasmid genome. Many features found in the molecular archaeology of the IncW backbone will serve as examples of how genetic organization and evolution occurs in the 'hazardous' life of conjugative plasmids.

The IncW family of broad host range plasmids

IncW plasmids were named after T. Watanabe, who described the first member of the group, pSa (called S-a in the original paper), isolated from *Shigella* (Watanabe *et al.*, 1968). Hedges & Datta (1971) isolated R388 from *Escherichia coli* and Kontomichalou isolated R7K from *Providencia rettgeri* (Kontomichalou, 1971; Coetzee *et al.*, 1972). Although many other plasmids belonging to the IncW group were later described (Datta & Hedges, 1972b, 1983; Pattishall *et al.*, 1977; Shinagawa *et al.*, 1982; Dorokhina & Korotiaev, 1984) only these three were physically characterized initially. Heteroduplex analysis by electron microscopy revealed a high degree of DNA sequence similarity between the three, with differences mainly assigned to antibiotic resistance genes (Gorai *et al.*, 1979). Sequence comparisons between pSa (Close & Kado, 1991, 1992; Okumura & Kado, 1992; Belogurov *et al.*, 2000) and R388 (GenBank accession no. BR000038) revealed that they are more than 95% identical at the DNA level. We have sequenced two additional IncW plasmids, pIE321 and pIE522 (Smalla *et al.*, 2001), and their backbone genomes are 95% and 99% identical to R388 (C. Revilla *et al.*, manuscript in preparation). In summary, available data underline the existence of a conserved IncW backbone, while the three characterized examples presumably represent the recent expansion of a successful pre-existing assemblage.

Initial interest in IncW plasmids focused on their small size, wide spectrum of antibiotic resistances, broad host range (Table 1) and their unusual ability to suppress oncogenicity caused by *Agrobacterium tumefaciens*. Early studies on pSa, R388 and R7K by Bradley & Cohen (1976) determined that IncW pili are rigid filaments, 10–12 nm thick, with an average length of 450 nm. In the genera *Escherichia*, *Salmonella*, *Shigella* and *Pseudomonas*, W-pili are present at average frequencies of up to three per cell. IncW plasmid conjugation only occurs when mating takes place on solid surfaces (Bradley *et al.*, 1980). Thus, IncW plasmids are considered 'surface maters', like IncP and IncN plasmids, but unlike IncF and IncI plasmids, which can mate in liquid. Successful transfer and stable inheritance of IncW plasmids have been reported in many bacterial genera (Table 1), most of them belonging to the *Proteobacteria*. However, not all *Proteobacteria* are able to harbour IncW plasmids: in the case of *Thiobacillus ferrooxidans*, plasmid pUFR034 could be transferred from *E. coli* donors, but was not stable in the recipient (Liu *et al.*, 2000).

Genetic structure of plasmid R388

The genome of plasmid R388 (GenBank accession no. BR000038) is 33 926 bp long and contains 43 putative ORFs longer than 50 amino acids (Fig. 1). A function could be assigned to all but five genes, based on database similarity searching and/or available experimental data. The genetic structure of R388 is compact, with a pronounced absence of noncoding regions except for two 558-bp-long direct repeats (LDR1 and LDR2), which are described below. Based on functional assignment, we have divided R388 genes into five functional groups. Genes belonging to these individual groups are localized to discrete regions of the genome, which we call modules. These modules are organized in two major sectors: a sector devoted to general maintenance functions, and a sector devoted to conjugation. General maintenance functions are dedicated to the survival of the plasmid within a certain host and we divided have them in three main modules: replication, stable inheritance and establishment. Conjugation is classically divided into DNA transfer replication (*Dtr*) and mating pore formation (*Mpf*) modules. As we will discuss below, those two sectors display different genomic arrangements and phylogenetic features.

As stated previously, the IncW plasmids originally isolated and characterized have an identical backbone. To

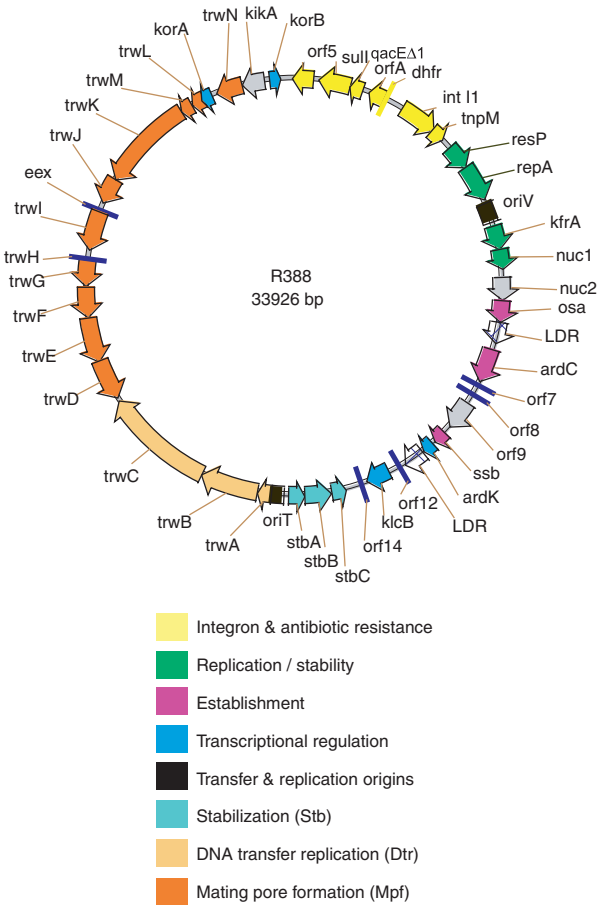

Fig. 1. Genetic map of plasmid R388. The figure shows the genetic organization of the plasmid, as inferred from analysis of the DNA sequence (GenBank accession no. B000038). ORFs and other sequence features are depicted in different colours after assignment to functional modules according to the code at the right (see text for further details).

evaluate the conservation of individual genes and global arrangement, we performed a PSI-BLAST search with all the R388 ORFs. As can be seen from Table 2, 28 out of the 35 genes that comprise the R388 backbone (excluding the integron) have their closest homologues in genes located in other plasmids. To evaluate best the similarity of R388 to other conjugative plasmids, we organized a matrix with the BLAST scores of each R388 ORF with respect to a series of representative plasmid genomes. We included in the analysis plasmids that appeared frequently in Table 2 as well as representatives of well-known incompatibility groups (Fig. 2). Although the BLAST score is not a good estimator of evolutionary distance, the matrix in Fig. 2 serves to illustrate the global relationship of R388 with other plasmid groups. The analysis readily reveals the important notion that the R388 genome is essentially made up of a combination of genes originating in other mobile genetic elements. This result highlights the existence of a specific gene repertoire that belongs to elements involved in horizontal gene transfer (mainly conjugative plasmids), and reinforces the concept of the existence of a horizontal gene pool (Thomas, 2000). A similar situation has been found in the analysis of different conjugative plasmids, and the notion of these elements as mosaic assemblages of genes belonging to the horizontal gene pool appears repeatedly in the literature (Schneiker *et al.*, 2001; Greated *et al.*, 2002; Gonzalez *et al.*, 2003; Batchelor *et al.*, 2004; Heuer *et al.*, 2004). Many of the genes found in the horizontal gene pool are involved in some of the main functions of plasmid physiology (replication, partition, establishment, conjugation and phenotypic traits). However, the appearance of recurrent genes of unknown function, in otherwise unrelated plasmids, suggests also that they may have an adaptive value for plasmid survival in certain environments.

Figure 2 also shows that, although R388 genes belong to the horizontal gene pool, there is no global conservation of plasmid architecture with respect to other incompatibility groups, as no particular plasmid exhibits higher homology with all or a majority of R388 modules. The mating pore formation (*Mpf*) region is the most conserved overall, showing detectable homology with all conjugative plasmids analysed, the IncF cluster being the most distal member. Conservation of synteny is observed within each of the conjugation modules (*Mpf* and *Dtr*) but not in the replication, partition or establishment modules. R388 is therefore composed of a specific repertoire of genes extracted from the horizontal gene pool, but arranged in an architecture unique for the IncW incompatibility group.

General maintenance functions in the IncW backbone

Replication module

Replication is an essential function for plasmid survival, and plasmids have evolved alternative strategies for ensuring their own endurance within their host genomes. Plasmid replication proceeds by one of two basic mechanisms. A protein (either chromosomal or plasmid-encoded Rep protein) recognizes a specific DNA sequence (the origin of replication, *oriV*) and proceeds either by (1) nicking one DNA strand to produce a 3′-OH free end (rolling-circle mechanism) or (2) melting the double-stranded DNA to allow the generation of an RNA replication primer and the assemblage of a replisome (θ mechanism). Because plasmids possess a wide variety of strategies to control the process of replication precisely, a large number of variants of the two basic mechanisms have been observed (Paulsson, 2002; Giraldo, 2003).

Early studies on pSa revealed that IncW replication proceeds bidirectionally from the origin of vegetative replication (*oriV*) and that plasmid copy number is low, between

Table 2. ORFs in the R388 genome

ORF	Gene name	Position (bp)*	G+C content (mol%)†	Protein size (amino acids)	Closest homologue‡	Percentage identity§	*E*-value¶	Accession no.
ORF1	repA	5129–6097	59.2	323	RepA protein pBFp1 plasmid	48% (145/297)	e-75	AAQ94182
ORF2	kfrA	6818–7444	58.8	209	KfrA *Pseudomonas* sp. CT14 plasmid	43% (91/209)	e-34	ABA25964
ORF3	parB	7444–7965	58.7	174	ParB protein RP4 plasmid	60% (102/170)	e-45	AAA26415
ORF4	nuc2	8172–8771	59.6	200	Putative protein pWW0 plasmid	30% (32/106)	0.029	NP_542916
ORF5	osa	8777–9343	56.2	189	FiwA pTB11 plasmid	26% (70/193)	e-20	CAG30876
ORF6	ardC	10079–10969	57.6	297	Antirestriction protein pAT plasmid	48% (138/287)	e-65	AC3172
ORF7	orf7	11108–11305	65.6	66	Hypothetical protein *Polaromonas* sp. JS666	52% (33/63)	e-10	ZP_00506016
ORF8	orf8	11381–11542	51.57	54	Not found			
ORF9	orf9	11613–12386	61.7	258	Hypothetical protein pMOL98 plasmid	67% (171/255)	e-83	CAD58034
ORF10	ssb	12588–12938	58.0	117	SSB (*Xylella fastidiosa*)	35% (39/111)	e-10	NP_299066
ORF11	ardK	12984–13313	58.4	110	KorA2 protein pWW0 plasmid	43% (31/71)	e-05	NP_542828
ORF12	orf12	14008–14349	56.6	143	Plasmid hypothetical protein (*Ralstonia metallidurans*)	37% (33/87)	e-08	ZP_00025237
ORF13	klcB	14368–15003	60.5	212	Klcb protein RK2 plasmid	47% (94/197)	e-41	P52605
ORF14	orf14	15169–15417	59.3	83	Hypothetical protein pWW0 plasmid	51% (38/74)	e-11	NP_542818
ORF15	stbC	15960–15577	66.1	128	Hypothetical protein pXFa51 plasmid	47% (22/46)	0.32	NP_061706
ORF16	stbB	16669–15977	51.4	231	Hypothetical protein pWW0 plasmid	49% (117/235)	e-54	NP_542911
ORF17	stbA	17091–16699	55.2	141	Putative TraD protein pWW0 plasmid	41% (40/97)	e-09	NP_542912
ORF18	trwA	17545–17907	61.3	121	TraK protein R46 plasmid	48% (18/37)	0.012	NP_511203
ORF19	trwB	17912–19432	55.2	507	TraJ protein R46 plasmid	43% (194/442)	e-89	NP_511202
ORF20	trwC	19435–22332	60.1	966	TrwC protein pXAC64 plasmid	51% (485/947)	0.0	NP_644759
ORF21	trwD	23494–22421	56.1	358	Conjugal transfer protein pXFa51 plasmid	41% (143/344)	e-72	NP_061671
ORF22	trwE	24647–23463	60.5	395	TraL protein pSB102 plasmid	40% (154/378)	e-60	NP_361040
ORF23	trwF	25447–24650	55.2	266	TraO protein R46 plasmid	35% (104/291)	e-41	NP_511196
ORF24	trwG	26153–25461	56.5	231	TraE protein R46 plasmid	41% (92/224)	e-39	NP_511195
ORF25	trwH	26293–26153	55.0	47	TraN protein R46 plasmid	47% (20/42)	1.7	NP_511194
ORF26	trwI	27435–26410	59.7	342	Conjugal transfer protein pXFa51 plasmid	32% (99/303)	e-31	NP_061667
ORF27	eex	27679–27452	54.2	76	TraG protein pIPO2T plasmid	35% (29/81)	e-04	NP_444523
ORF28	trwJ	28356–27679	57.7	226	TraF protein pIPO2T plasmid	38% (85/221)	e-25	NP_444522
ORF29	trwK	30833–28356	55.8	823	TraE protein pIPO2T plasmid	42% (353/829)	0.0	NP_444521
ORF30	trwM	31150–30839	57.3	104	Channel protein VirB3 homologue (*Brucella melitensis*)	37% (36/97)	e-10	NP_541004
ORF31	trwL	31501–31166	55.5	112	Not found			
ORF32	korA	31730–31437	47.4	98	KorA protein R46 plasmid	32% (28/85)	e-04	NP_511187
ORF33	trwN	32550–31855	59.0	198	VirB1 protein pXAC64 plasmid	46% (90/194)	e-33	NP_644764
ORF34	kikA	33145–32555	54.7	197	Hypothetical protein (*A. tumefaciens*)	36% (51/140)	e-16	AE008939
ORF35	korB	33282–33566	54.2	95	DNA-binding protein (*Janthinobacterium* sp. J3)	35% (37/105)	e-14	BAC56733
ORF36	orf5	440–33869	65.6	166	Putative acetyltransferase plasmid pFBAOT6	100% (166/166)	e-94	YP_067851
ORF37	sul1	1407–571	61.8	279	Sul1 dihydropteroate synthase R46 plasmid	100% (279/279)	e-143	NP_511228
ORF38	qacE	1748–1404	50.2	115	QacEdelta1 R46 plasmid	100% (115/115)	e-46	NP_511227
ORF39	orfA	2348–1917	53.1	144	Not found			

Table 2. Continued.

ORF	Gene name	Position (bp)*	G+C content (mol%)†	Protein size (amino acids)	Closest homologue‡	Percentage identity§	*E*-value¶	Accession no.
ORF40	dhfr	2639–2406	56.2	78	Dihydrofolate reductase gene (*E. coli*).	100% (78/78)	e-33	P00384
ORF41	intI1	2878–3888	61.2	337	Integrase IntI1 (*Pseudomonas aeruginosa*)	100% (337/337)	0.0	AAK19119
ORF42	tnpM	3836–4210	62.6	194	TnpM (*Shigella flexneri 2a*)	97% (79/81)	e-33	AAL08438
ORF43	resP	4480–5133	57.1	218	Site-specific recombinase, Tn-*21* resolvase family	44% (88/197)	e-36	NP_720486

*Position in base pairs referred to the unique *Sal*I restriction site (bp 1).

†G+C content was calculated using total gene length.

‡The closest homologue to each R388 ORF refers to the most similar gene (with assigned function if available) found in the databases by PSI-BLAST. IncW plasmid pSa and Trw proteins from *Bartonella* sp. were excluded.

§Percentage identity and number of identical residues over the region comprising the detectable homology (in parentheses) are also shown.

¶*E*-values shown correspond to the PSI-BLAST first iteration.

R388	Orf5	SulI	QaCE	OrfA	IntI1	TnpM	ResP	RepA	KfrA	Nuc1	Nuc2	Osa	ArdC	Orf7	Orf8	Orf9	Ssb	ArdK	Orf12	KlcB	Orf14
pKM101							e-5						0.001				0.001	e-5			
pWW0			e-12		0.001		e-10		0.001		0.001		0.001				0.001	e-5			e-11
pXAC64						0.001	e-4	e-25	0.001		0.001							e-6			
pIPO2								e-48	e-11	e-33						e-82					
pSB102							e-6	e-45								e-73					
pXFa51							e-5									e-73	0.001	0.001			
R100	e-94	e-174	e-46		0.00	e-36	e-12			e-25						e-6	0.001				
pTi					e-8		e-7			0.001			e-58								
RP4							e-5			e-75		e-20	e-42							e-32	

R388	StbC	StbB	StbA	TrwA	TrwB	TrwC	TrwD	TrwE	TrwF	TrwG	TrwH	TrwI	Eex	TrwJ	TrwK	TrwM	TrwL	KorA	TrwN	KiKA	KorB
pKM101		e-34		0.001	e-39	0.0	e-60	e-45	e-41	e-34	0.001	e-23	0.001	e-102	e-102	e-5		e-4	e-32		0.001
pWW0		e-54	e-8	0.001	e-37	0.0	e-12	e-18	e-14	0.001		0.001		0.001	e-63	0.001			e-29		
pXAC64	e-5	e-43	e-7		e-83	0.0	e-61	e-33	e-23	e-5		e-6		e-8	e-79	e-4			e-33		
pIPO2					0.001		e-54	e-57	e-9	e-34		e-24	e-4	e-25	0.0	e-8			e-27		
pSB102					0.001		e-55	e-60	e-33	e-37	0.001	e-26	0.001	e-20	0.0	e-8			e-31		
pXFa51	e-4	e-61	e-54	0.001	e-4		e-12	e-34	e-32	e-37		e-31	0.001	e-10	0.0	e-7			e-33		
R100					e-42	e-41									0.001						
pTi					0.001		e-44	e-45	e-23	e-7		e-9			e-105	0.001			e-15		
RP4							e-14	0.001	e-4						e-20						

Fig. 2. Distribution of R388 gene homologues among different conjugative plasmid genomes. The figure represents a matrix of similarity scores for each R388 gene (columns) to a number of representative plasmids (rows). Each cell shows the PSI-BLAST *E*-value for a given R388 gene relative to the indicated plasmid. Dark grey shadowed boxes indicate *E*-values obtained in the first iteration, while light-grey shadowed boxes denote first-iteration *E*-values below the threshold (score < 0.001) but that reached it in subsequent iterations. Representative plasmids from different incompatibility groups (pKM101 for the IncN, R100 for the IncF and RP4 for the IncP groups) or plasmids found to contain close homologues to R388 ORFs according to Table 2 were selected.

two and three copies per cell (Tait *et al.*, 1982a). The replication regions of pSa and R388 are 95% identical at the DNA sequence level. They are composed of *oriV* plus an operon containing two ORFs coupled in translational fusion. The first ORF codes for ResP (218 amino acids), a homologue of Tn*21* resolvase (24% identity). The gene *resP* is transcriptionally coupled to gene *repA*, coding for R388 replication initiator protein (323 amino acids). ResP provides a resolution activity for plasmid dimers produced in vegetative replication. Experimentally, ResP was able to stabilize minireplicons containing the IncW replicon (Okumura & Kado, 1992).

RepA does not belong to any of the canonical replicase families. A PSI-BLAST search using RepA_R388 as a query converged at the second iteration, indicating that IncW RepA and their homologues form a coherent family. A phylogenetic tree was constructed (Fig. 3) that shows the evolutionary spread and wide distribution of IncW-like Rep proteins in bacterial phyla. Although no evolutionary link to other families of replication proteins was obvious, the fact that pSa replication is bidirectional (Tait *et al.*, 1982a) and the presence of an iteron region composed of six 15-bp direct repeats in the adjacent *oriV* suggest a θ mechanism of replication in which RepA would bind these iterons.

Some RepA homologues have been analysed experimentally. pBFp1, a 24-kb partially sequenced plasmid isolated from a marine biofilm, is incompatible with pSa and thus a member of the IncW group (Bergstrom *et al.*, 2004). Its RepA protein is 44.7% identical to R388_RepA and, as in R388, pBFp1 *oriV* is adjacent to RepA and contains five iterons, 75% identical to those in pSa/R388. Plasmid pXV2 from *Xanthomonas campestris* has another characterized

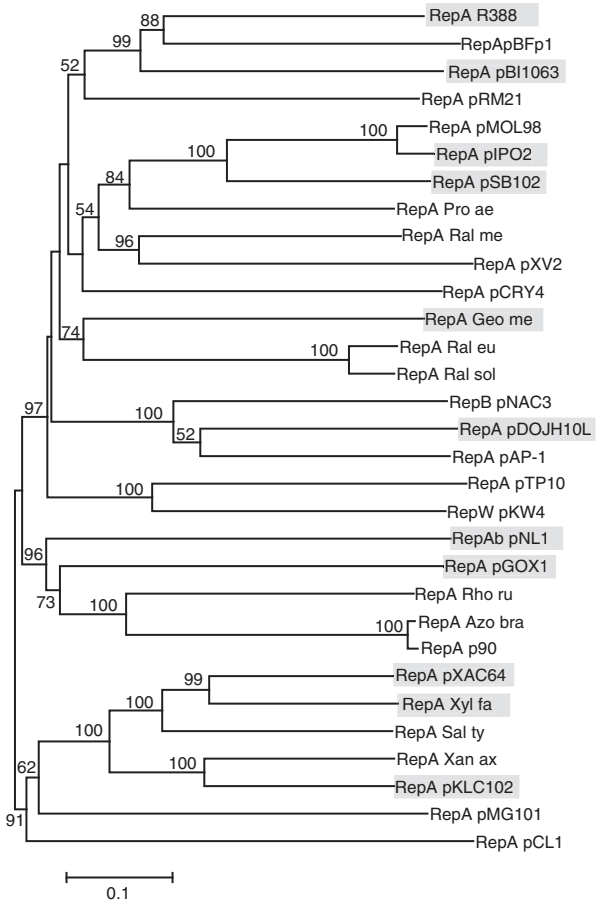

Fig. 3. Phylogenetic tree of IncW replicases. Dendrogram of RepA_R388 homologues found by PSI-BLAST in plasmids and genomic islands. The tree was constructed using a neighbour-joining algorithm with MEGA software. Bootstrap values for 1000 replicates are indicated. Grey boxes indicate Rep proteins belonging to conjugative plasmids, either determined experimentally or assumed by the presence in cis of additional conjugation genes. Plasmid names are indicated except in the cases where no specific name has been given for the mobile element (sequences obtained from genome projects). In these cases an abbreviation indicating the corresponding organism is given. (Pro ae, *Prosthecochloris aestuarii* DSM 271; Ral me, *Ralstonia metallidurans* CH34; Geo me, *Geobacter metallireducens* GS-15; Ral eu, *Ralstonia eutropha JMP134*; Ral sol *Ralstonia solanacearum*; Rho ru, *Rhodospirillum rubrum* ATCC 11170; Azo bra, *Azospirillum brasilense* plasmid of 90 kB; Xyl fa, *Xylella fastidiosa*; Sal ty *Salmonella typhimurium* DT104; Xan ax, *Xanthomonas axonopodis* pv. citri str. 306).

IncW replicon, incompatible with pSa (Wu & Tseng, 2000). pXV2_RepA is 39% identical to pSa/R388_RepA. Its *oriV* is also adjacent to RepA and contains four iterons. These iterons show a lower degree of conservation to pSa/R388 than pBFp1. A core of three bases spaced by two or three nucleotides to a track of three adenines [**ACG**NN(N)**AAA**] is conserved in the three plasmids (invariant nucleotides are shown in bold type). pXV2 minireplicons have narrow host range, being maintained in *Xanthomonas* but not in *Pseudomonas*, *Escherichia*, *Erwinia* or *Rhizobium*. The smallest plasmid with an IncW-like replicase known to date, *Rhodothermus marinus* pRM21 plasmid (2935 bp), is also unable to replicate in *E. coli* (Ernstsson *et al.*, 2003) and its copy number in *Rhodothermus* is 40 copies per cell (Bjornsdottir *et al.*, 2005). pRM21 and pSa/R388 RepA proteins are 34% identical, and pRM21 *oriV* contains three 19-bp iterons with little sequence similarity to those in pSa/R388.

Another group of plasmids isolated from soil (pMOL98, pSB102 and pIPO2) also bears characteristic IncW-like RepA proteins (Fig. 3), which are 29.6%, 21.2% and 22.8% identical to RepA_R388, respectively. Contrary to the examples above, their *oriV*s are located several kilobases away from their cognate *repA* genes. In these *oriV*s, three (pSB102) or four (pIPO2 and pMOL98) iterons can be found. *oriV* iterons from these plasmids are highly similar (94.4%) and eight of the nine first nucleotides are identical to the corresponding nucleotides in pSa/R388 iterons.

There are also RepA_R388 homologues in gram-positive bacteria that include Rep proteins of plasmids from *Corynebacterium* and *Bifidobacterium*. In pNAC3, a 10 224-bp plasmid from *Bifidobacterium longum*, a putative *oriV* is located immediately before the *repA* gene. It contains three 21-bp iterons and an AT-rich region (Corneau *et al.*, 2004). A second representative is *Corynebacterium glutamicum* pCRY4 plasmid. This 48-kb plasmid has a low copy number (two to three plasmids per chromosome) and displays an *oriV* adjacent to *repA* gene with five 22-bp iterons (Tauch *et al.*, 2003). It was suggested, therefore, that replication in these plasmids proceeds by a θ mechanism as in pSa (Tauch *et al.*, 2003).

All the data described above suggest that the IncW-like family of replicons is widely spread among bacteria, including the *Proteobacteria*, *Actinobacteria* and the *Bacteroidetes*/chlorobi group. The classical distinction between gram-negative θ replicons and gram-positive rolling circle replicons is no longer sustained: θ replicons and rolling-circle replicons have been found to function in both groups (del Solar *et al.*, 1998), as the pSa/R388 family of replicases clearly shows. The presence of homologous RepA proteins, adjacent or distant to an iteron-containing *oriV*, suggests a common replication mechanism, although major differences in copy number (two to three copies in pSa, R388, pCRY4 and pMG101, vs. 40 copies in *Rhodothermus* pRM21) and host range (e.g. pXV2 replicates only in *Xanthomonas* whereas pSa/R388 have broad host range). Finally, plasmids from closely related genera do not cluster together (as shown by the RepA phylogenetic tree in Fig. 3), suggesting a history of horizontal gene transfer between distant bacterial species.

Functions for stable inheritance

Plasmids of low copy number encode stability systems devoted to ensuring a correct segregation between daughter cells, while plasmids of high copy number usually rely on stochastic distribution. Faithful plasmid segregation is guaranteed by one or more of three different systems: dimer resolution, partition and genetic addiction (Zielenkiewicz & Ceglowski, 2001). When homologous recombination takes place between two plasmid molecules present in the same cell, a plasmid dimer is produced. Plasmid dimers have two *oriV*s and, therefore, replicate preferentially and accumulate in the cytoplasm. The progressive decrease in copy number produced by the formation and accumulation of plasmid dimers and higher multimers results in a catastrophic decrease of plasmid stability (Summers *et al.*, 1993). For this reason, most plasmids code for a specific resolvase or, more rarely, rely on a host-encoded resolution system. The resolvase ResP accounts for this function in plasmid R388. Faithful segregation of plasmids between daughter cells before bacterial division is accomplished by plasmid partition systems. Among plasmids in gram-negative bacteria, there are two main classes of partition systems, represented by the *par*/*sop* gene family (Walker-type ATPases) and the actin-like ATPase family, represented by the *par* locus of plasmid R1 (Gerdes *et al.*, 2000). Although the mechanism of plasmid partition is not fully understood, both systems seem to produce dynamic filaments that propel plasmid molecules to both sides of the division septum (Ebersbach & Gerdes, 2005). The third mechanism of plasmid stability is genetic addiction. Plasmid addiction systems are responsible for postsegregational killing of plasmid-free daughter cells after cell division. Three main classes of addiction systems have been described: (1) the antisense RNA-based *hok-sok* system of R1 plasmid (Gerdes *et al.*, 1997), (2) protein-based addiction systems such as the *ccd* system of F plasmid (Jaffe *et al.*, 1985) or the *parDE* locus of plasmid RP4 (Roberts *et al.*, 1990) and (3) restriction-modification systems, less studied in relation to plasmid physiology, but nevertheless found in the genomes of some (Nakayama & Kobayashi, 1998). Finally, it should be mentioned that plasmid stability is a complex phenomenon and mutations in various other genes, apart from those described above, show decreased plasmid stability (see, e.g., the *parCBA* operon in RK2 described below).

A PSI-BLAST search of all R388 ORFs of at least 50 amino acids yielded no detectable homology to any of the essential proteins in the stability systems described above, with the only exception being ResP protein, involved in dimer resolution. The R388 region that extends from *oriV* to the first LDR contains, however, a total of four ORFs that show significant similarity to proteins involved in plasmid stable inheritance that do not belong to the classes mentioned above. These are *kfrA*, *nuc1*, *nuc2* and *osa*. They are organized in one operon (Chen & Kado, 1994). KfrA from the IncP plasmid RK2 (43% identity to KfrA_R388) binds DNA and is a repressor of its own synthesis (Jagura-Burdzy & Thomas, 1992). It forms an extended coiled-coil domain able to form multimers suggested to provide a bridge for host factors involved in plasmid partition (Jagura-Burdzy & Thomas, 1992). Adjacent to *kfrA* in R388 is *nuc1*, a homologue of *parB*_RK2 (61% identity). In spite of its name, *parB*_RK2 does not belong to the ParB family of ATPases involved in plasmid partition. Nuc1_R388 and ParB_RK2 belong to the *Staphylococcus aureus* thermonuclease family. The role of Nuc1 in IncW physiology is puzzling. Close & Kado (1992) demonstrated that Nuc1_pSa is secreted and elicits a nuclease activity in the extracellular medium. R388 *nuc1* and RK2 *parB* genes have secretion signals that localize their encoded proteins to the periplasm. ParB_RK2 was purified and showed endo- and exonuclease activity on single-stranded DNA (ssDNA) (Johnson *et al.*, 1999). It was also capable of nicking supercoiled double-stranded DNA nonspecifically. In RK2 and other IncPα plasmids, *parB* belongs to an atypical partition module (the *parCBA* operon). Mutational analysis demonstrated that *parCBA* was involved in plasmid stable inheritance. However, *parC* or *parB* mutants only showed a twofold decrease in plasmid RK2 stability, whereas *parA* mutants showed a much stronger phenotype (Easter *et al.*, 1998). In R388, no *parC* or *parA* homologues could be found and the function of an isolated *parB* homologue is enigmatic. Nevertheless, and given that this gene in isolation appears in many plasmid systems, we presume the function it plays is important for plasmid physiology.

An additional ORF in this region is encoded by the *osa* gene. Protein Osa (oncogenicity suppression agent) inhibits T-DNA transfer to plant cells mediated by Ti plasmids of *Agrobacterium tumefaciens*. Inner membrane protein Osa suppresses Ti oncogenicity by inhibiting VirE and VirF protein transport from the infecting bacteria to the plant cell (Close & Kado, 1991; Chen & Kado, 1994; Lee *et al.*, 1999) and interferes specifically with VirD4 coupling function, blocking T-DNA transfer (Cascales *et al.*, 2005). Osa also inhibits RSF1010 mobilization by the pTi plasmid both between *Agrobacterium* cells and from *Agrobacterium* to plant cells. Osa is 38.2% identical to FiwA, a protein from IncP plasmids shown to exert fertility inhibition on IncW plasmids (Fong & Stanisich, 1989). Thus, Osa belongs to a protein family that impairs conjugation of coresident plasmids (fertility inhibition), and might constitute a factor of fitness (favouring a given conjugative plasmid when competing with others for recipient cells).

Because pSa/R388 plasmids have only two or three copies per cell (Tait *et al.*, 1982a), it seems unlikely that they lack a system to ensure their correct segregation. However, we detected neither any of the ATPases characteristic of the main stability systems nor any evidence of a toxin–antitoxin locus,

but only genes (*kfrA*, *nuc1*) known to play secondary roles in typical partition systems. It is thus possible that these two genes, together with some of the orphan genes situated in their vicinity, constitute a yet unknown system for plasmid stable inheritance. Alternatively, there is an additional locus in the R388 genome that might be implicated in plasmid stability, namely the *stbABC* operon, adjacent to the origin of transfer. As homologues of this system were found associated with the conjugation machinery in many other plasmids, the *stbABC* operon will be described in detail later.

Long direct repeats and establishment module

Two 558-bp-long direct repeats (LDR1 and LDR2; 99.8% identity) are present in the R388 genome (Fig. 1). No ORF above 50 amino acids in size is present inside these elements. Interestingly, two 386-bp LDRs present in the soil plasmid pIPO2 (Repeat Regions 4 and 7) share a conserved core of 62 bp with 90.1% identity to R388 LDRs. pIPO2 LDRs neither code for any protein nor have any assigned function. As could be expected for any repeated DNA segment, the LDRs create some instability in the R388 genome, and R388ΔLDR, an R388 derivative produced by deletion of the intervening sequence between both LDRs, was isolated spontaneously (our unpublished results).

Located between LDR1 and LDR2 are six ORFs. Three of them (*ardC*, *ssb* and *ardK*) have an assignable function based on experimental data or homology searching. Because their function has implications in plasmid establishment after conjugation, the module containing them has been termed accordingly. Gene *ardC* codes for a 297-amino-acid protein. ArdC was reported to show an antirestriction activity that protected plasmid DNA from the action of host restriction barriers (Belogurov *et al.*, 2000). ArdK, a transcriptional regulator for *ardC*, is found about 3 kb downstream in the R388 sequence. ArdK is homologous to ArdK_pKM101, a transcriptional regulator of ArdA and ArdB antirestriction genes in the IncN plasmid pKM101 (Belogurov *et al.*, 1993). ArdA/ArdB systems in pKM101 are active against type I and II restriction endonucleases, while ArdC in R388 protects ssDNA from type III endonucleases. ArdC has no significant homology to ArdA or ArdB proteins, except for the presence of a small motif common to all antirestriction proteins (Belogurov *et al.*, 2000). Although there is no significant homology between ArdA or ArdB in pKM101 and ArdC_R388, the regulatory protein (ArdK) is related in both systems (25.4% identity). Homologues of *ardC* can be found in widely different plasmids and closest relatives to ArdC_R388 occur in pTi and pRi plasmids from *Agrobacterium*. Antirestriction proteins would therefore be factors contributing to plasmid fitness in their natural environments.

The third ORF within this region with an assigned homology is *ssb*, coding for an ssDNA binding protein. SSB proteins bind ssDNA and are involved in many functions in DNA metabolism: enhancement of helix destabilization by DNA helicases, prevention of reannealing of ssDNA, protection from nuclease digestion, organization and stabilization of replication origins, primosome assembly, priming specificity, enhancement of replication fidelity, enhancement of polymerase processivity and promotion of polymerase binding to the template (Meyer & Laine, 1990). SSB proteins are widespread in conjugative plasmids (Moreira, 2000). In the IncI plasmid ColIb-P9, the *ssb* promoter is activated immediately after conjugation along with another gene (*psiB*) that prevents triggering of the SOS response in the recipient cell upon ssDNA entry from the donor (Jones *et al.*, 1992). This unique mechanism of transcriptional induction promotes plasmid establishment in a recipient cell after conjugation, when the conjugating ssDNA enters the cell (Jones *et al.*, 1992).

Two proteins between LDR1 and LDR2 are susceptible to be transported to the recipient cell. Although there is no experimental evidence for this transport, ArdC and SSB proteins show a set of features that makes this possibility very appealing. ArdC has a region of homology to the N-terminal domain of TraC1 primase of plasmid RP4 (Rees & Wilkins, 1990). This domain is responsible for primase active transport to the recipient cell during conjugation (Miele *et al.*, 1991). Thus, ArdC might be actively transported to the recipient cell and thus protect the incoming strand from the action of type III endonucleases (ssDNA endonucleases). In the case of SSB, this protein may cover the transferred strand during transport or might even be actively transported to the recipient cell. ssDNA binding protein VirE2 in *Agrobacterium tumefaciens* has been demonstrated to associate with the transferred strand and to be actively transported to the recipient cell (Christie, 1988; Sundberg, 1996). Whatever the mechanisms, they are not essential for conjugation: R388ΔLDR lacks *ardC* and *ssb* but conjugates at the same frequency as R388 in *E. coli* to *E. coli* mating under laboratory conditions (our unpublished results). In natural environments, where recipient cells have restriction barriers and other defence mechanisms against incoming DNA, this region might be necessary for the establishment of IncW plasmids.

A further five genes, three between LDR1 and LDR2 (*orf7*, *orf8*, *orf9*) and two downstream of LRD2 (*orf12* and *orf14*), code for five R388 orphan proteins. No function could be assigned to them after exhaustive similarity searching. Orf9, however, has unnamed homologues in several conjugative plasmids such as pIPO2, pSB102, pXFa51 and R100 among others. Orf9 proteins are predicted by Pfam (Sonnhammer *et al.*, 1997) to contain a conserved IMP dehydrogenase/GMP reductase domain. This domain is involved in the biosynthesis of guanosine and in maintaining the balance between A- and G-containing nucleotides. Although the possible role of such activity in plasmid physiology is puzzling, the conservation of this family of proteins among different plasmids points to the

existence of conserved genes in plasmid systems of as yet unknown function. It should be noted that the genes adjacent to *orf9* in R388 are not significantly similar to any adjacent gene in the loci of the cited *orf9* homologues.

Genetic mosaicism in general maintenance modules

Phylogenetic analyses of the ORFs corresponding to the replication, stable inheritance and establishment modules unveiled that these parts of the R388 genome display genetic mosaicism. In the replication module, *resP* and *repA* are transcribed under the control of the same promoter, but this arrangement is not conserved among other plasmids with replicases belonging to the IncW family. The association of a Tn*21*-like resolvase and *repA* seems to be a unique feature of R388/pSa. The stable inheritance module contains ORFs from diverse origins. *parB*_R388 is found in one operon together with *osa* and *nuc*, while in RK2 it belongs to the *parABC* operon. In summary, each gene outside the conjugation modules has its own idiosyncratic phylogenetic tree, indicating that they, even if organized in operons, have

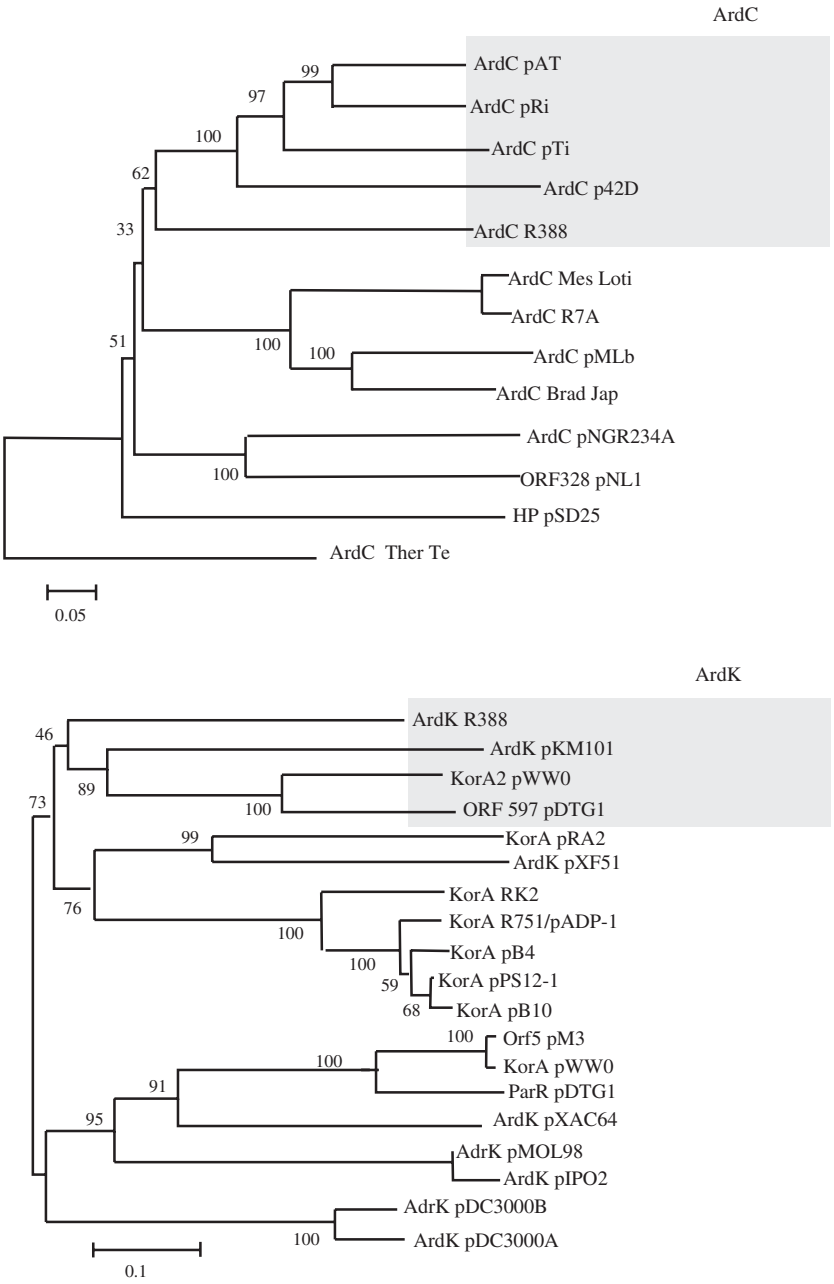

Fig. 4. Genetic mosaicism in the ArdC/ArdK antirestriction system. Phylogenetic dendrograms of ArdC and ArdK proteins were constructed using the neighbour-joining algorithm with MEGA software. Bootstrap values for 1000 replicates are indicated above each branch. R388_ArdC clusters in a monophyletic branch (shadowed) with homologues present in plasmids from *Rhizobium* and *Agrobacterium* from the pTi/pRi plasmid family. R388_ArdK clusters in a monophyletic branch (shadowed) with transcriptional regulators present in IncN plasmid pKM101 and TOL plasmid pWW0. In the case of ArdK_pKM101 it has been shown that ArdK controls the expression of antirestriction systems distantly related to ArdC (see text).

different evolutionary histories. Perhaps the clearest example of modular evolution is provided by the different phylogenies of the antirestriction protein ArdC and its transcriptional regulator, ArdK. Although ArdC is clearly linked to homologues present mainly in symbiotic plasmids from *Rhizobium* and pTi plasmids from *Agrobacterium*, ArdK is homologous to transcriptional regulators of ArdA/B systems in plasmid pKM101 (Fig. 4).

Different phylogenies for adjacent ORFs imply that these genes incorporated into the ancestral IncW genome at different times and from different sources. Recombination might be an essential force driving the evolution of these regions of the R388 genome. The mosaic origin of different genes involved in the same process may indicate that this region of the IncW genome evolved by capture and combination of different genes that enhanced plasmid fitness. A similar process is believed to occur in phage evolution: the 'moron accretion hypothesis' suggests that many functions in the phage genome are gained by capture and rearrangement of individual genes (Hendrix *et al.*, 2000).

Sequence features and regulation network

At least six ORFs in the R388 sequence (*ardK*, *klcB*, *stbA*, *trwA*, *korA* and *korB*) were designated transcriptional regulators, although it is possible that other small ORFs with no assigned function also have a regulatory role. One important question in plasmid biology is if plasmids should be considered as coherent genomes or as a random collection of useful genes. A valuable hint would be to determine if there is evidence of global regulation, if their genomes tend to behave in an integrated way or, alternatively, if each functional module works independently. Previous work unveiled the existence of complex global regulation networks within conjugative plasmid genomes. This is the case of the CUP network in pKM101 (Delver & Belogurov, 1997) and the *kil-kor* regulon in IncP plasmids (Kornacki *et al.*, 1990). The CUP regulation network in pKM101 controls the expression of nine ORFs involved in plasmid establishment. CUP elements consist of eight repeated sequences situated immediately upstream of each of the regulated ORFs. ArdK_pKM101 binds a consensus sequence in the CUP elements, which are thought to have a role both in recombination and incorporation of new genes and in establishment of a coordinate regulation network (Delver & Belogurov, 1997). In the IncP backbone, three global regulators (KorA, KorB and TrbA) exert a coordinate regulation of replication and transfer functions (Zatyka *et al.*, 1994; Macartney *et al.*, 1997). The presence of repeated sequences in several promoters of the plasmid genome is an indicator of the existence of a coordinate regulation network: in RP4, KorB, a homodimeric regulator, binds a 13-bp palindromic sequence repeated 12 times in the plasmid genome (Balzer *et al.*, 1992; Dostal *et al.*, 2003).

Owing to the lack of experimental data available on the regulation network of R388, we performed a sequence analysis following the RP4 and pKM101 paradigm. That is, we searched for the presence of sequence features with potential regulatory role, namely inverted repeats, direct repeats and iterons in the intergenic regions of R388. The results are summarized in Fig. 5. Five families of repeated elements, suggesting potential binding sites, were found in R388 intergenic regions. As was the case in the *kil-kor* regulon of IncP plasmids, repeated sequences were found in promoters from different regions of the plasmid. Sometimes, members of a family appeared in different promoters and on several occasions sequence elements from different families occurred simultaneously in the same operon. Several combinations of potential regulatory sites were found, suggesting the existence of a regulatory network that may operate through the use of a set of regulatory proteins acting on different promoters.

Some of these binding sites could be putatively ascribed to regulators found in the R388 genome. SDR-B repeats present in the *Mpf* promoters are probably binding sites for transcriptional regulators KorA and KorB (Seubert *et al.*, 2003). SDR-A repeats are identical to ArdK_pKM101 binding sites, so they are probable regulatory sites for ArdK_R388. IT-A corresponds to the iterons present in *oriV*. Additionally, the role of TrwA as transcriptional regulator of the *Dtr* module has also been demonstrated (Moncalian *et al.*, 1997).

Although we found sequence features with potential regulatory role over the entire R388 genome, a sharp distinction between their distribution in the conjugation module with respect to other regions of the plasmid (replication, stable inheritance and establishment) should be pointed out. The conjugation modules, divided into *Dtr* and *Mpf* regions (see below), present binding sites for their cognate transcriptional regulators only (TrwA for *Dtr* and KorAB for *Mpf*). We were unable to find shared binding sites between *tra* and other plasmid modules, although there is still the possibility of interactions at the protein–protein level. On the other hand, intergenic regions in the replication, stable inheritance and establishment modules contained combinations of different sequence features. Assuming they are binding sites for transcriptional regulators, these sequence features imply a complex regulatory network at least in these regions of the genome. Current efforts are being made to elucidate experimentally the existence of this network.

General features of modules for plasmid conjugation

R388 conjugation has been extensively studied. A detailed description of the transfer mechanism is not the aim of this

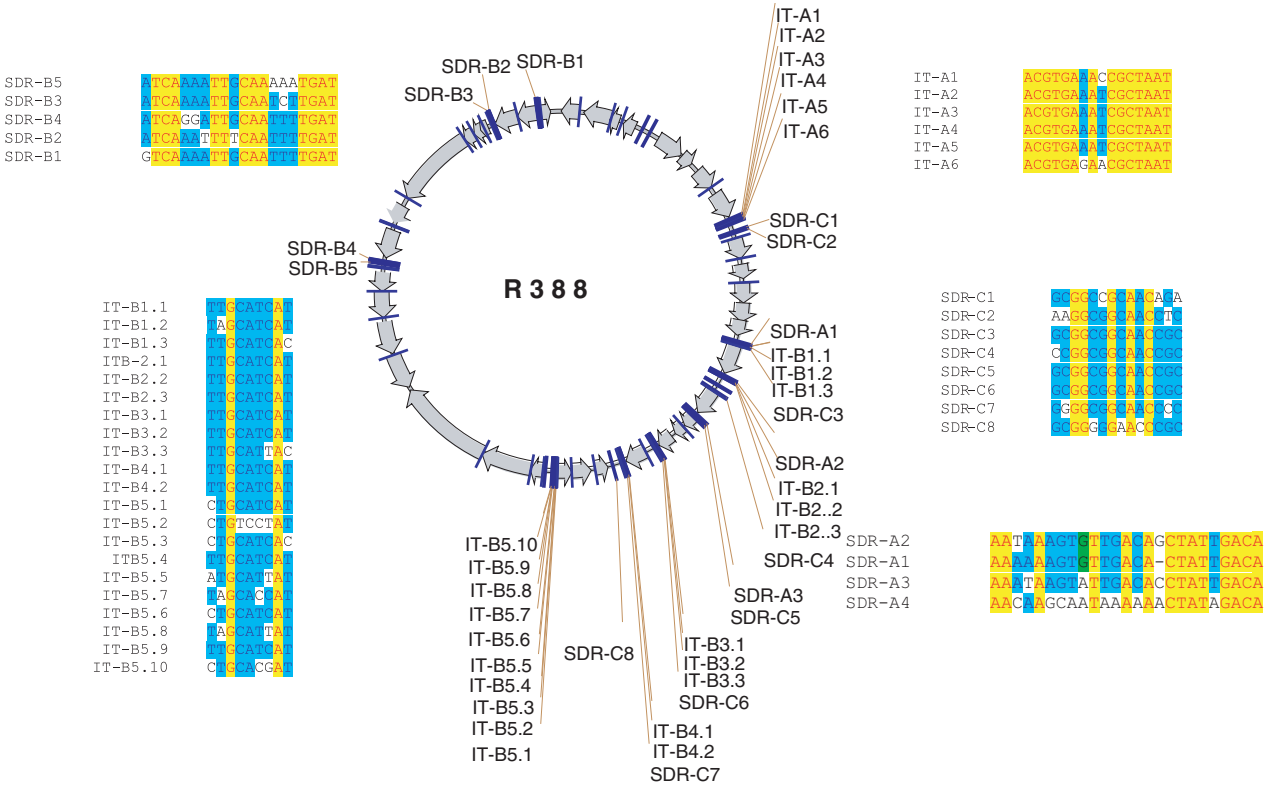

Fig. 5. Short direct repeats and iterons in the R388 genome. The figure shows the R388 genetic map with an indication of the location of a number of repeated sequences. R388 sequence analysis revealed the presence of five major sequence element families. Their position and sequence conservation is indicated. IT-A iterons, located in *oriV*, constitute binding sites for RepA replicase. SDR-A is composed of four 25-bp repeats present in four promoter regions of the establishment region. SDR-B direct repeats are presumed KorA-binding sites present in promoter regions of operons involved in conjugative pilus expression and assembly. SDR-C is composed of eight repeats present in the stability and establishment regions. IT-B is a family of 9-bp iterons spaced by three nucleotides present in variable number (2–10) in the promoter region of the *stbABC* operon and in four promoters of the establishment region. Invariant residues are shown in red on a yellow background, while conserved residues are shown in black on a blue background.

review (for reviews on this topic see Gomis-Ruth *et al.*, 2004; Cabezon & de la Cruz, 2006; Llosa & de la Cruz, 2005). Conjugation can be described as a DNA transfer machinery coupled to a T4SS (Llosa *et al.*, 2002). This functional distinction is reflected in the genetic organization of the R388 conjugation region in two separate modules. Relaxase, coupling protein and a nicking-accessory protein constitute the DNA-Transfer Replication region (*Dtr*) whereas T4SS proteins are part of the Mating Pore Formation region (*Mpf*). The arrangement of R388 conjugation genes and those of three related plasmids are depicted in Fig. 6. IncN plasmid pKM101 is the closest homologue to IncW in the conjugation region, both in *Dtr* and in *Mpf* modules. And pWW0 is the prototype of the *Pseudomonas* IncP-9 group, which includes many plasmids from *Pseudomonas* that carry xenobiotic degradation pathways. pXAC64 belongs to the plant pathogen *Xanthomonas axonopodis* and other plasmids from this genus have an identical genetic arrangement. In pKM101, a homologous regulatory system is conserved, and overall amino acid sequence identity in the conjugation region is about 30%. Taking into account the divergence rate for conserved proteins in prokaryotes (Doolittle *et al.*, 1996), and assuming Dtr proteins are of the conservative type, the closest relative outside the IncW incompatibility group (i.e. the IncN group) would have diverged earlier than *Escherichia* and *Bacillus*. Gene order is an evolutionary feature that is highly variable in bacterial genomes. Initial studies suggested that bacterial genomes have stable architectures, and that gene order is globally conserved between different species (Ochman & Wilson, 1987). But when many bacterial genomes became available, a huge degree of variation was found. Now synteny is considered as an evolutionary trait that is quickly lost in evolution (Mushegian & Koonin, 1996; Itoh *et al.*, 1999). If (1) shuffling of a genome structure is virtually neutral in long-term evolution, (2) shuffling depends extensively on recombination and DNA uptake (Silva *et al.*, 2003) and (3) these two processes are known to occur more often in plasmid genomes than in chromosomes, why is gene order so well conserved in conjugation operons, even when sequence similarity is so

low? A similar situation is found in the genomes of bacteriophages (Brussow & Hendrix, 2002) and no evident answer has yet been reported. Perhaps the assembly of a complex structure such as a conjugative pilus or a viral capsid requires an order in genome arrangement that ensures a temporal expression profile and coordinated levels of gene expression. Genetic mechanisms that may explain this process have been demonstrated for flagellum biosynthesis and metabolic networks (Zaslaver *et al.*, 2004). The genetic arrangement of flagellar genes shows a considerable degree of synteny conservation in eubacteria (Canals, 2006) and archaea (Patenge, 2001). Flagella construction is organized in operons under the control of a hierarchy of promoter activities and temporal regulation, so that the earlier a given gene product participates in flagellum assembly, the sooner is the cognate promoter activated. In this arrangement, clustering of genes that participate in the same step of flagellum development optimizes the design (Kalir & Alon, 2004). Thus, when complex integration of individual protein functions is required, conservation of global arrangement may be as important for fitness as the individual sequence and structure of each of the components.

As discussed above, replication, stability and establishment regions have a mosaic structure. A labile genomic arrangement (in evolutionary terms) may reflect a region devoted to a cellular function able to be accomplished by the sum of individual protein functions. On the other hand, cellular functions that require a higher degree of integration between individual protein activities (i.e. conjugation) show a conserved, ordered gene arrangement. Thus, synteny conservation might be a sign of a genetic region encoding a highly structured molecular machine.

Structure and evolution of the DNA transfer replication region

Genetic organization and function

The region comprised between *stbC* and *korB* (17 990 bp) contains the region involved in conjugative DNA transfer. The synteny of this region is conserved in a large number of plasmids, some of which are shown in Fig. 6. The transfer region contains three functional modules: *Dtr*, involved in DNA processing; Mpf, involved in mating pair formation; and Stb, a potential postconjugation establishment system. In R388, three *Dtr* genes (the relaxase *trwC*, the coupling protein *trwB* and the regulatory protein *trwA*) are organized in a unique operon adjacent to *oriT* (Llosa *et al.*, 1994b). Three additional conserved genes (*stbA*, *stbB* and *stbC*) are arranged in a putative operon located on the opposite side of *oriT*, i.e. in the leading region of the transferred strand. Strikingly, a similar gene arrangement of homologous genes in the immediate vicinity of their respective *oriT*s can be found in a number of other plasmids, especially in plasmids with *Dtr* modules more closely related to that of R388 (Fig. 6). In pKM101 there is experimental evidence that the lack of StbABC results in plasmid instability in a $RecA^+$ background, although the reasons for this instability are not clear

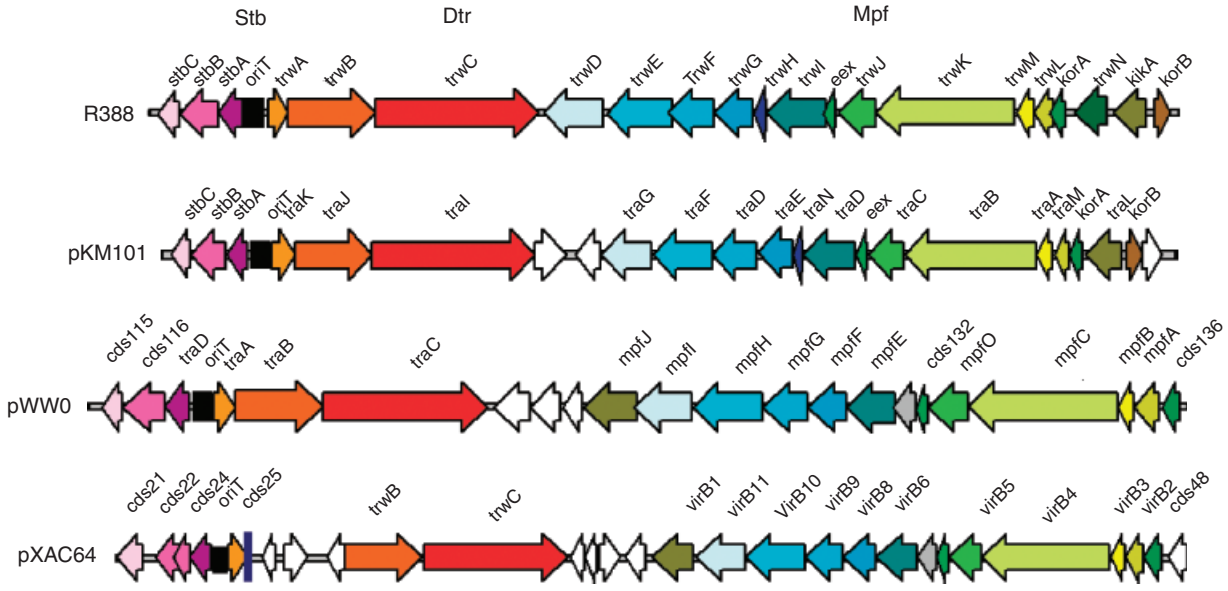

Fig. 6. Synteny conservation in conjugation modules. The figure shows the gene organization of the *Stb*/*Dtr*/*Mpf* modules of plasmids R388, pKM101, pWWO and pXAC64, with homologous genes drawn in the same colour. In R388 the region depicted covers 18 kb. In spite of the conserved genetic architecture, the degree of sequence conservation is always lower than 55% amino acid identity, with a rough median of 25% identity. Coloured boxes indicate the general arrangement for *Stb*, *Dtr* and *Mpf* modules.

(Paterson *et al.*, 1999). The *stb* operon is preceded both in R388 and in pKM101 by a pattern of direct repeats present in the *stbABC* putative promoter regions. In R388 there are 10 iterons of 9 bp separated by 2-bp spacers. Small groups of two to four iterons, also separated by 2-bp spacers, are present in four other promoters of the establishment region (IT-B in Fig. 5). Although the function of these iterons has not been defined, their presence in the promoter region may indicate a regulatory role. Positional conservation of an *stb* locus relative to the *Dtr* region among this series of plasmids might also suggest a role for these proteins in the conjugation process. Both possibilities remain to be investigated. StbB is a remote homologue of ParB ATPases and the absence of canonical partition systems in the genome of R388 makes this locus a possible candidate for such function.

Evolution

The evolutionary history of the *Dtr* module can be traced using relaxase TrwC as its essential element. Relaxases are the only proteins shared by mobilizable and conjugative plasmids, so they have been used to classify conjugation systems in different families (Francia *et al.*, 2004). TrwC has two distinct domains: an N-terminal relaxase and a C-terminal helicase (Llosa *et al.*, 1996). The helicase domain is not conserved in all relaxases, not even within the TrwC family, as shown in Fig. 7. On the other hand, it is widespread among different families of proteins, including replication, recombination and repair proteins. As a consequence, and in order to avoid long-branch attraction phenomena, the phylogeny of TrwC was reconstructed from protein sequence alignments using only the relaxase domain (300 N-terminal amino acids). TrwC relaxase and its homologues constitute the MOB_F family. As shown in the resulting phylogenetic tree (Fig. 7), the MOB_F family has two main branches; the first one groups homologues hosted mainly in bacteria belonging to the *Proteobacteria*, while the second clusters homologues from the *Actinobacteria*.

Plasmids closest to R388 are found in branch (b). They all display a *Dtr* organization equivalent to that in Fig. 6, including the existence of a single nicking-accessory protein. Besides, their *Mpf* systems are homologous to VirB of the Ti plasmids from *A. tumefaciens*. Branch (a) contains the IncF plasmids, whose relaxases and coupling proteins are homologous to the corresponding R388 proteins, but whose *Mpf* system is highly divergent from VirB T4SS (Lawley *et al.*, 2003). Classical F-like plasmids (branch a.1) have two accessory proteins (TraY and TraM), although this is not the case for more distant F homologues (branch a.2) that have only one (in the cases of pYJ016 and pPBPR1) or none (pNL1). F-like and VirB-like *Mpf* systems have differential conjugation properties: F-like plasmids produce long flexible pili, adapted to conjugation in liquid, while W-like or N-like plasmids produce rigid pili, able to conjugate only in solid media. Thus, the existence of phylogenetic branches (a) and (b) is probably a consequence of the adaptation of an ancestral *Dtr* module to different *Mpf* machineries. It should be noted that after this long process, the three-dimensional structures of relaxases from both branches are still nearly identical, as shown by the structures of R388_TrwC and F_TraI (Datta *et al.*, 2003; Guasch *et al.*, 2003), conserving the four catalytic tyrosil residues involved in *nic* cleavage and strand transfer reactions (Grandoso *et al.*, 2000).

Association of TrwC-like relaxases to their cognate coupling proteins is linked to the presence of a full conjugative apparatus, as no mobilizable plasmid with an MOB_F relaxase has yet been discovered. However, there are two instances in which a W-like relaxase and coupling protein appear associated with no trace of a known *Mpf* apparatus: *Cyanobacteria Nostoc* PCC7120 plasmid gamma, mobility of which has not been analysed experimentally, and a contig from a *Shewanella* sp. obtained from a sequencing project of environmental samples from the Sargasso Sea (Venter *et al.*, 2004) (branch b.2 in Fig. 7). In this particular case, *trwC* and *trwB* homologues are surrounded by ORFs of either unknown function or related to bacteriophages Phi or P1 and Lambda. Comparison of the marine *Shewanella* contig to the complete genome of *Shewanella oneidiensis* revealed that *trwBC* are located in a 44-kb genomic island inserted between genome locus SO1154 and TonBR (Acc. No. SO1156). Many ORFs present in that genomic island are related to proteins associated with CISPR loci [repetitive DNA sequences, believed to be mobile (Jansen *et al.*, 2002), and present in many prokaryotic genomes]. This result opens the intriguing possibility that TrwC and TrwB could be involved in other mechanisms of DNA mobilization besides classical conjugation. In this respect, it might be noteworthy to point out that TrwC relaxase catalyses conjugation-independent site-specific recombination and integration reactions acting on *oriT* sites (Llosa *et al.*, 1994a; Draper *et al.*, 2005).

TrwC-like relaxases of plasmids from the *Actinobacteria* are clustered in branch (c). They form two separate clades. The first clade (c.1) is mainly composed of relaxases of rolling-circle replication plasmids from *Corynebacterium* and *Mycobacterium*. The second clade (c.2) includes relaxases of *Streptomyces* linear plasmids. They were neither associated with putative coupling proteins nor with other proteins involved in conjugation. Conjugation in *Streptomyces* has been described to proceed by dsDNA transfer, and only requires one conjugation protein, a septal translocator homologuous to FtsK/SpoIIIE (remote homologue of TrwB-like coupling proteins) (Grohmann *et al.*, 2003).

Homologues to MOB_F relaxases in these plasmids (clade c.2) contain just the relaxase domain, and there is no

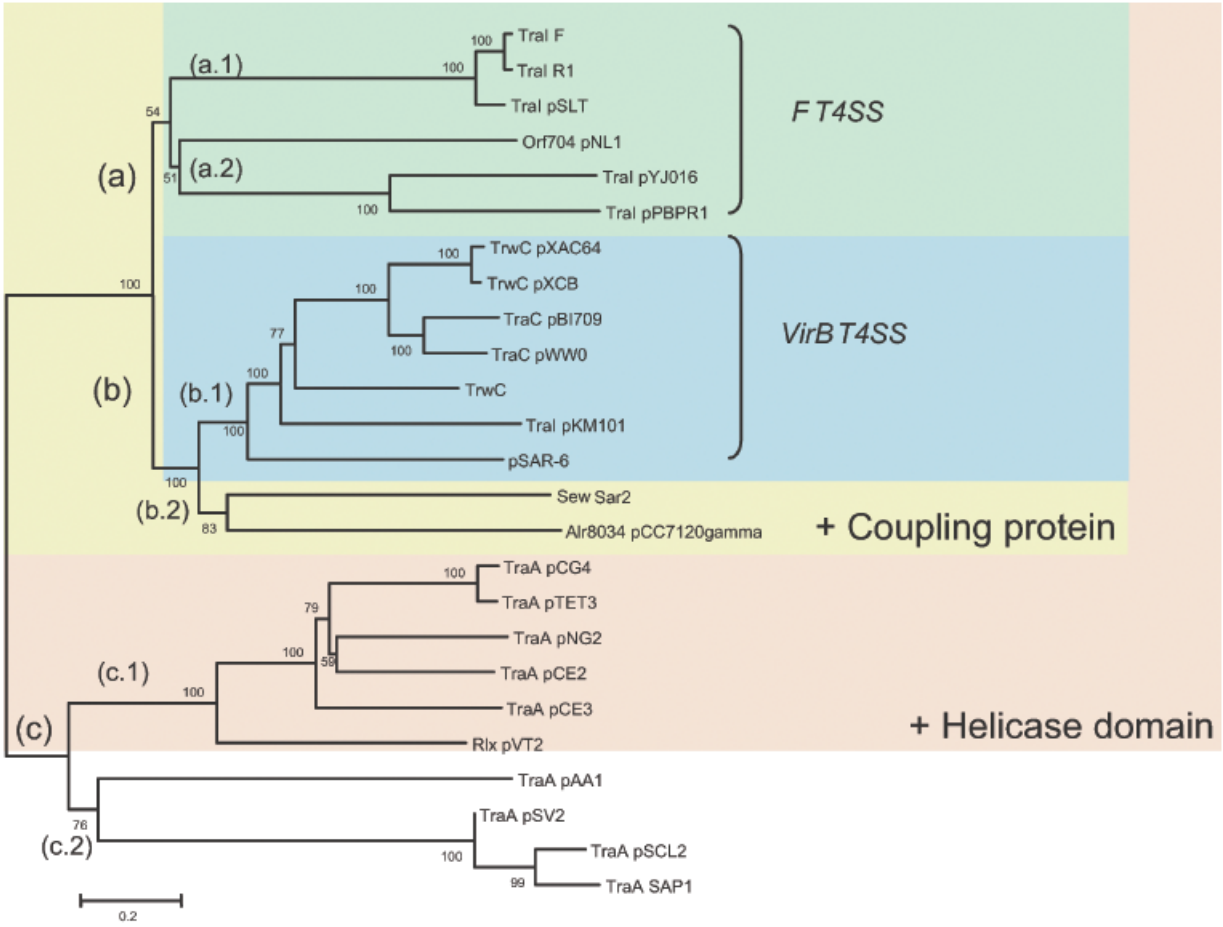

Fig. 7. Evolution of TrwC-like relaxases. Phylogenetic tree (unrooted) of TrwC relaxase domain homologues. The proteins shown were all found by PSI-BLAST, which converged after four iterations. The dendrogram was constructed using a neighbour-joining algorithm. Bootstrap values for 1000 replicates are indicated. Three major branches can be observed in the figure. Branch (a) includes conjugative plasmids with F-like T4SS, further divided into two subbranches: (a.1) clusters classical F-like plasmids with two accessory proteins (TraY and TraM), while (a.2) clusters atypical F-like plasmids with either one (pYJ016, pPBPR1) or no accessory protein (pNL1). Branch (b) is also composed of two subbranches. Subbranch (b.1) contains plasmids with a *trwB*-like coupling protein and a VirB-like T4SS, while (b.2) clusters TrwC-like relaxases with a TrwB-like coupling protein but no T4SS. In this last case, the location of the *Dtr* gene assemblage was either chromosomal (*Shewanella* Sar2) or plasmidic (pCC7210-γ from *Nostoc* sp.). Branch (c) clusters TrwC-homologues found in gram-positive plasmids and it is also composed of two subbranches. Subbranch (c.1) includes relaxases with a helicase domain, present in plasmids of *Corynebacterium* and *Mycobacterium*, while (c.2) includes TrwC-homologues present in species from the genus *Streptomyces* that do not contain an attached helicase domain. The different coloured backgrounds indicate the milestone events associated with TrwC evolution. The red box includes relaxases containing a helicase domain (all branches except c.2). The yellow box indicates relaxases associated in a putative operon with a coupling protein (all branches except c). Blue and green boxes indicate plasmids with a VirB-like T4SS or F-like T4SS, respectively.

experimental evidence for their involvement in conjugation. Corynebacterial homologues (clade c.1) do have relaxase and helicase domains. The backbone of these plasmids is also related to the IncW group at the replication level. Not surprisingly therefore, it was hypothesized that they may serve as vehicles of genetic exchange between *Corynebacteria* and gram-negative bacteria, as they carry homologues of both replicases and relaxases (Tauch *et al.*, 2002). Although their lack of a conjugative pilus-forming apparatus and of a coupling protein makes it unlikely that these plasmids are self-transmissible, the possibility remains that they are mobilizable.

As a consequence of the data described above, and according to the phylogenetic tree shown in Fig. 7, three important milestones (shown by different background colours in the figure) occurred during evolution of TrwC from a common MOB_F ancestral protein. The first event was the accretion of a helicase domain to the relaxase gene, an early event that preceded the split of MOB_F relaxases in a branch comprising plasmids from *Corynebacteria* and another branch that includes plasmids hosted by gram-negative bacteria. The second milestone was the association with a coupling protein gene, a genetic organization present in all plasmids from gram-negative bacteria and in *Nostoc*

PCC7120. Finally, the MOB_F *Dtr* became associated with one of two alternative *Mpf* genetic determinants: either the IncF T4SS, or a VirB-type T4SS. In the distal branch that clusters plasmids harbouring VirB-like *Mpf* regions (b.1), synteny is conserved along the entire conjugation region.

Structure and evolution of the *Mpf*

Structure

The R388 *Mpf* region codes for genes necessary for substrate transport and to assemble the conjugative W-pilus. Eleven genes constitute the T4SS (*trwDEFGHIJKLMN*), one is responsible for surface exclusion (*eex*), two are transcriptional regulators (*korA* and *korB*) and one has an ill-defined function (*kikA*). Sequence analysis and experimental data (our unpublished results) indicate that the 15 genes of this region are organized in four transcriptional units (*korB*, *trwN-kikA*, *korA-trwLMKJ-eex-trwI* and *trwHGFED*).

Three promoter regions are located in the *Mpf* module, and five putative binding sites for KorAB regulators were found by sequence inspection (SDR-B in Fig. 5). The *trwH* promoter has two putative Kor binding sites (SDR-B5 and SDR-B4) and controls expression of operon *trwHGFED*. The *korA* promoter contains also two SDR-B sites (sites 2 and 3) and controls expression of an operon that includes *korA*, *trwJKL*, *eex* and *trwI*. The third control region contains two divergent promoters (*trwN* and *korB* promoters), one controlling *trwN* and *kikA* expression and the other controlling *korB* transcription. Both promoters probably share a common –35 box, while each orientation presents its own putative –10 box. SDR-B1 is found in this region in the vicinity of the common -35 box. The presence of these sequence features may indicate the need for a complex transcriptional regulatory scheme for the *Mpf* apparatus, presumably directed by KorA and KorB (Seubert *et al.*, 2003). A similar regulatory organization is conserved in the *Bartonella trw* locus (see below) and in pKM101 (Seubert *et al.*, 2003).

The Trw proteins in the *Mpf* region are homologous to VirB proteins from *Agrobacterium tumefaciens* pTi (de Paz *et al.*, 2005). As stated before, an important feature of the *Mpf* region is the conservation of synteny among plasmids with little sequence identity. Thus, the first *Mpf* operon in R388, comprising *trwHGFED* (*virB 7-8-9-10-11* homologues), contains the majority of conserved members of T4SS and thus is responsible for the formation of the 'core' structure of the conjugative pilus (de Paz *et al.*, 2005). Along with VirB4, a cluster of homologues to these proteins form the natural competence system in *Helicobacter* (Hofreuter *et al.*, 1998, 2001; Smeets & Kusters, 2002). The same core structure is also present in other pathogens such as *Wolbachia* and *Rickettsia* (Malek *et al.*, 2004). The *trwLMKJI* (VirB2-3-4-5-6) operon encodes less conserved proteins, although it also contains one of the core ATPases (TrwK/VirB4). Yeast two-hybrid analysis of the *Agrobacterium* pTi system (Ward *et al.*, 2002) and the *Bartonella* T4SS (Shamaei-Tousi *et al.*, 2004) determined that most interactions detected between elements of the T4SS involved members of the same operon. VirB4 was the most common protein involved in interactions between both sets of proteins. In conclusion, synteny conservation could be a consequence of the evolutionary constraints imposed by the functional interactions among the gene products in each operon, as explained in the beginning of 'General features of modules for plasmid conjugation'.

Evolution

Sequence conservation in T4SSs varies among the different protein constituents. Nevertheless, phylogeny of the T4SS can be traced based on the phylogenetic trees of two major ATPases: VirB11 and VirB4. Dendrograms showing the evolution of TrwD (VirB11) and TrwK (VirB4) are depicted in Figs 8 and 9. TrwD and TrwK trees are largely congruent, although there is a higher degree of conservation between VirB4 homologues (see below). Congruence in both trees indicates that the main body of T4SS coevolved from its remote origin. Coevolution of the remaining T4SS components is suggested by the phylogenetic trees shown in Fig. 10. Although these trees were constructed from alignments of proteins containing an overall lower degree of homology, and thus contain more uncertainty in the resulting dendrograms, the picture shown in Fig. 10 is descriptive enough to underline overall conservation. With some interesting exceptions, such as the divergence in VirB5 and VirB10 homologues between pKM101 and R388, which can be explained by allelic replacement, most components of T4SS show an impressive congruence in their phylogenetic trees. As a consequence, and using TrwK and TrwD as the most confident dendrograms (Figs 8 and 9) we can attempt to describe some important events in the evolution of the ancestral VirB-like T4SS.

The higher sequence conservation in VirB4 allowed the incorporation of more distant homologues. Thus, it was possible to incorporate the IncF and IncH proteins (TraC and TrhC) in the TrwK tree that, along with some homologues from archaeal plasmids, and the Trb system of plasmid RP4, constitute the root of the tree (Fig. 9). In both the VirB4 and the VirB11 tree, the Trb apparatus from IncP and *Agrobacterium* plasmids appears as a monophyletic branch of ancestral origin, clearly divergent from classical VirB homologues. While the Trb and IncF/IncH branches contain only plasmid-encoded members, the VirB branch is more diverse and contains homologues in conjugative

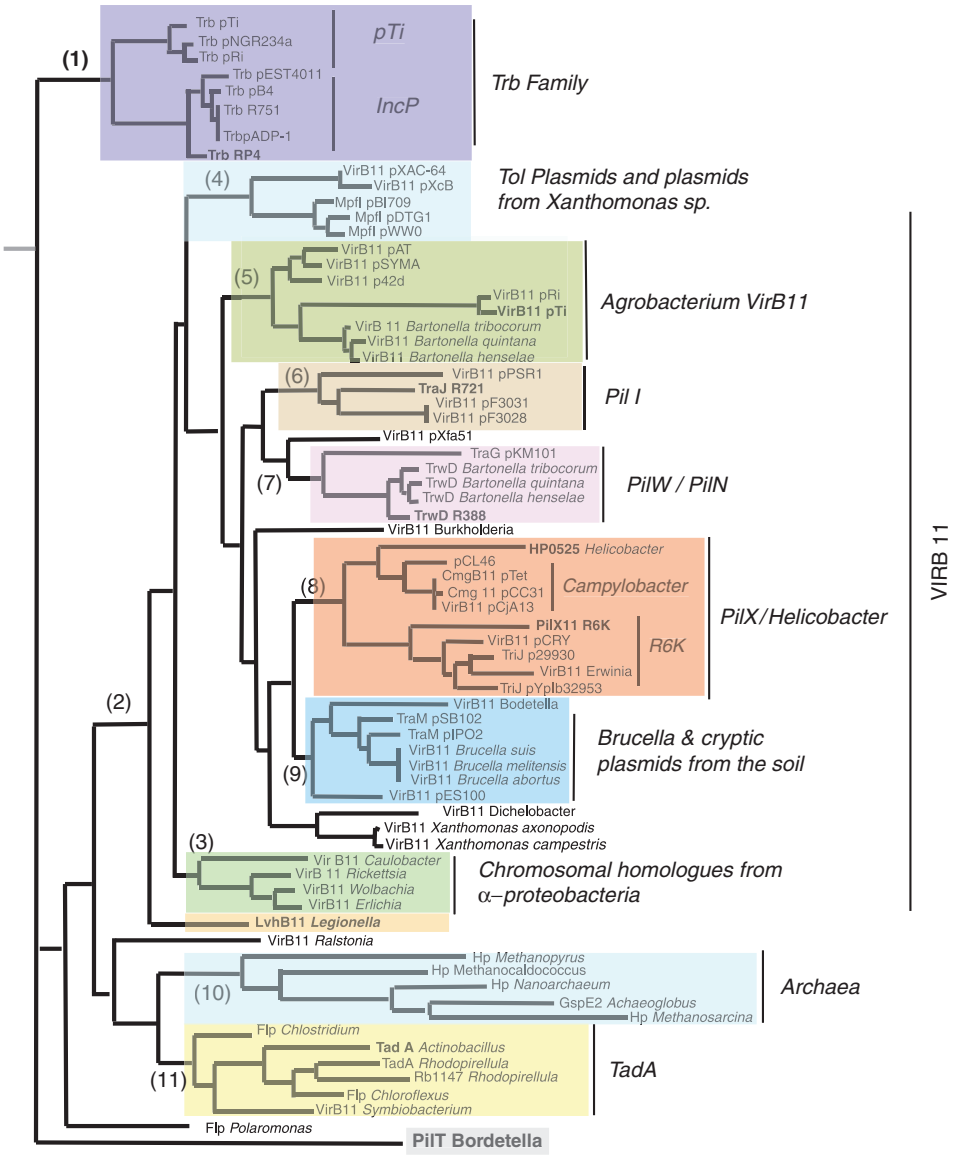

Fig. 8. Phylogenetic tree of TrwD (VirB11 homologues). The tree was constructed using a maximum-likehood algorithm (PHYLIP software, J. Felsestein, http://evolution.genetics.washington.edu/phylip.html) for TrwD and homologues. The tree was rooted using PilT, a remote homologue from a Type II Secretion System. The tree shows two major subdivisions. Branch (1) clusters the Trb protein family, present in IncP and pTi plasmids. These are the most distant homologues of the VirB common trunk and therefore appear as an outgroup. Closer to the VirB common trunk but acting also as outgroups are the monophyletic branches of TadA homologues (10) and *Archaea* homologues (11). The VirB family (2) splits in two branches that include most homologues contained in the databases. Outside these two major subdivisions is the LvhB system of *Legionella*. Branch (3) includes chromosomal homologues from the *Alphaproteobacteria*. Branch (4) includes plasmids from *Xanthomonas* and pWWO-like plasmids from *Pseudomonas*. The classical VirB system from *Agrobacterium* and closely related systems are clustered in branch (5). This branch includes two monophyletic groups: one includes the symbiotic plasmids from *Rhizobium* and *Sinorhizobium*, while the second includes a branch for pTi plasmids from *Agrobacterium* and a branch comprising the chromosomal VirB system from *Bartonella*. The IncI family (prototype R721) is clustered in branch (6), close to branch (7), which includes IncN and IncW homologues, including TrwD. Branch (7) also clusters the *trw* systems from *Bartonella*. Branch (8) groups VirB11 homologues present in *Helicobacter* involved in the delivery of virulence factors with the conjugative systems of plasmids from *Campylobacter* and the IncX group of conjugative plasmids. Branch (9) comprises homologues from cryptic plasmids from soil and the chromosomal counterparts of *Brucella*.

plasmids as well as in chromosomes. IncW plasmids are clustered with IncN plasmids and the *Bartonella trw* locus (see below). Importantly, there is no correlation between the phylogenetic trees obtained with relaxase and coupling protein on one side, and those from *Mpf* components on the other (M.P. Garcillan-Barcia, manuscript in preparation).

The dual role of T4SS in plasmid conjugation and protein secretion has been emphasized (Lessl & Lanka, 1994;

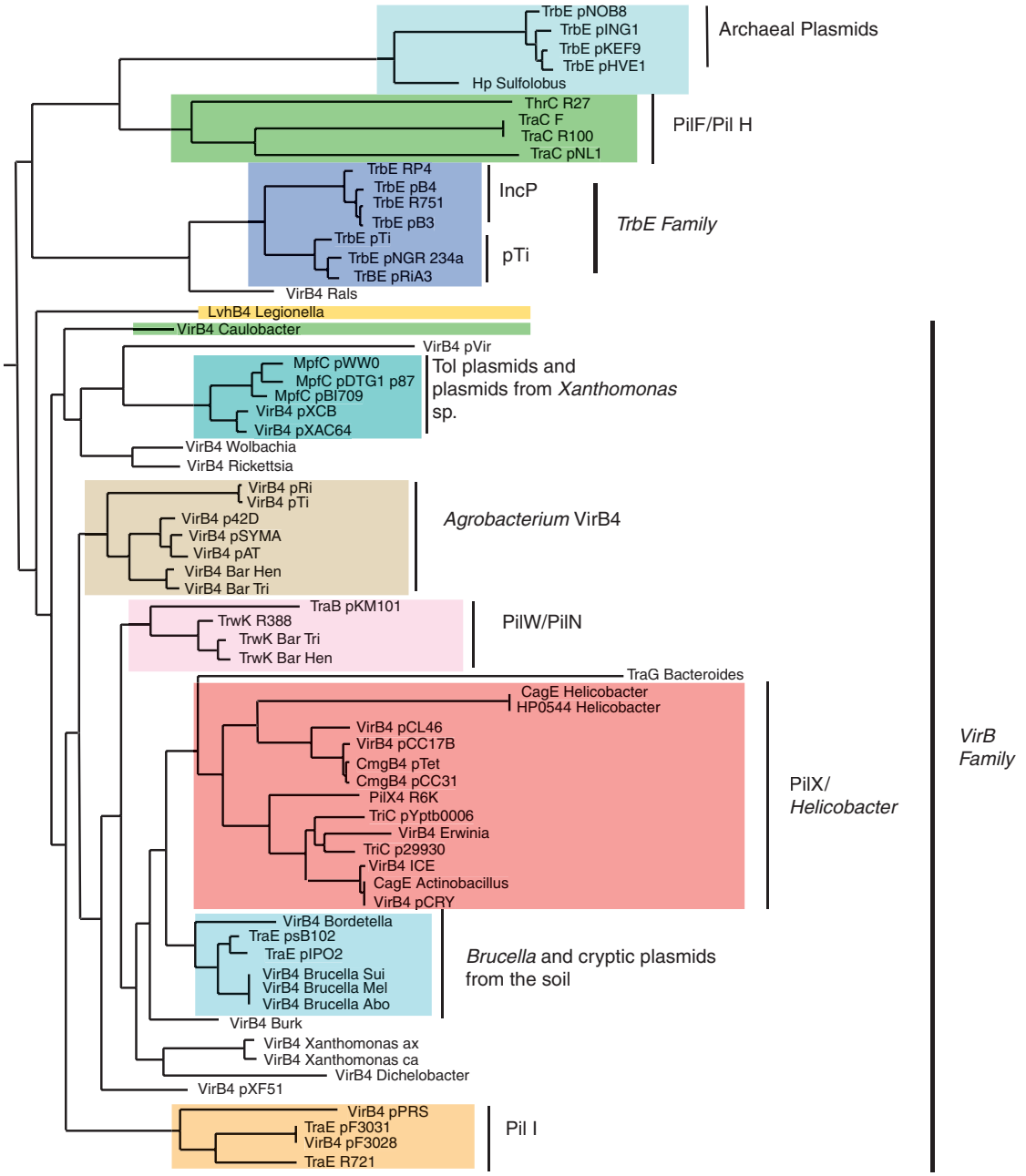

Fig. 9. Phylogenetic tree of TrwK (VirB4 homologues). The tree was constructed as that in Fig. 8, and follows the same conventions. Both trees are largely congruent (denoted by branches with the same colour background) with some exceptions. It should be noted that the higher conservation between VirB4-like proteins allows the inclusion in this tree of the ancient divergence of type F (F-like plasmids) and type H (IncH plasmids) *Mpf*s (dull green branch, second from top) as well as a monophyletic branch of Archaeal plasmids (light blue, top branch). The major discrepancy between VirB11 and VirB4 trees is the divergence between the clustering of type W/N *Mpf*s and type I *Mpf* homologues in the VirB11 tree, and their radical separation in the VirB4 dendrogram (light pink and orange, respectively).

Christie & Vogel, 2000; Llosa & de la Cruz, 2005; Schroder & Lanka, 2005). The finding that some T4SS in pathogens can effectively mediate the transfer of mobilizable plasmids (Vogel *et al.*, 1998) reinforces the idea of a versatile machinery able to perform both functions. The distinction between T4SS involved in protein transport or plasmid conjugation is not sustained by phylogenetic reconstruction (Frank *et al.*, 2005). Instead of an ancestral separation

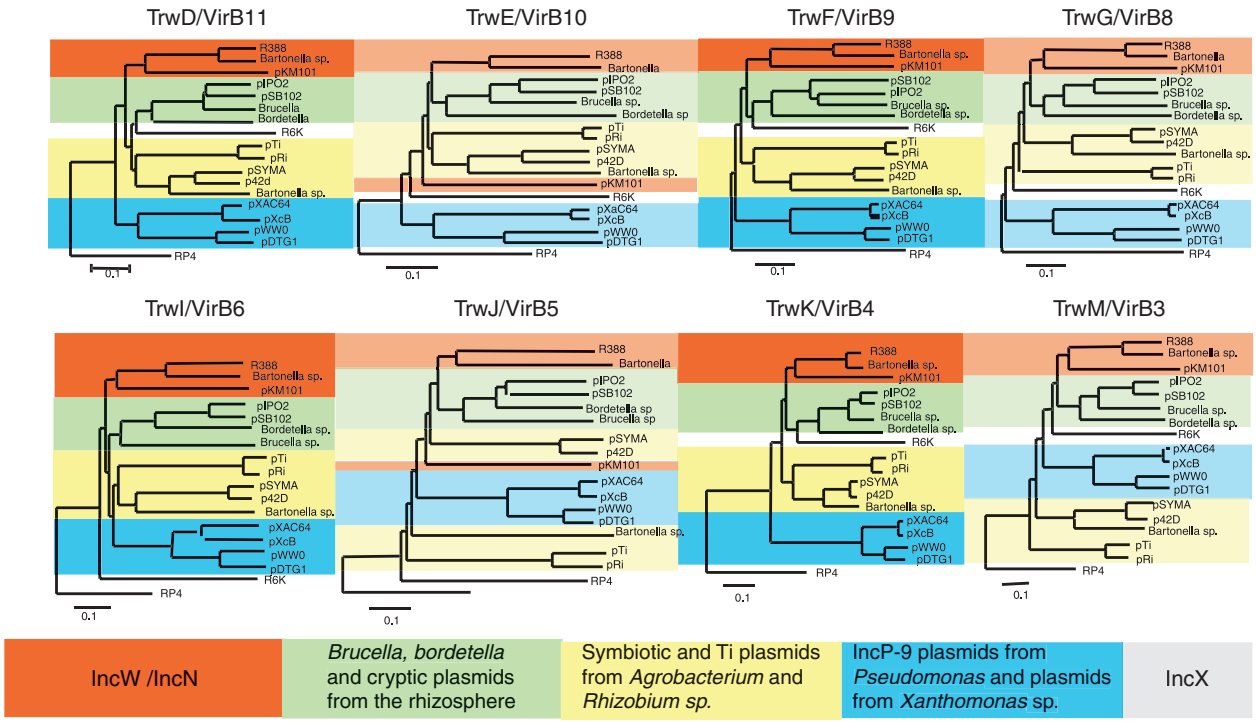

Fig. 10. Coevolution of T4SS components. Phylogenetic trees (unrooted) of selected members of different T4SS components. Dendrograms were constructed using a neighbour-joining algorithm using MEGA software (Kumar *et al.*, 2004). Major branches are highlighted in different colours. Conservation of the topology in different trees (congruence) indicates protein coevolution.

between the two classes of T4SS, TrwD and TrwK trees show multiple branches that include representatives of both classes. One of the clearest examples of how *Mpf* systems devoted to conjugation can be adapted to perform a different role in the physiology of their natural hosts is the case of the *trw* locus in *Bartonella* species (Seubert *et al.*, 2003).

Trw locus in *Bartonella*

Seubert *et al.* (2003) described a T4SS in *Bartonella tribocorum* involved in intraerythrocytic parasitism. This locus was found in other *Bartonella* species and it is strikingly closely related to the IncW *Mpf* system (up to 80% amino acid identity for individual components). The phylogenetic relationship between both systems was clear, as even KorAB regulators and binding sites were conserved. Two major differences were apparent between both Mpf systems: the *Bartonella* locus lacks the two genes not explicitly involved in conjugation (*kikA* and *eex*) while it presents a number of tandem gene duplications (the *trwL* gene is present in seven tandem copies, and a segment containing *trwJIH* is present in five tandem copies). KorAB-dependent regulation is conserved, and even Kor proteins from R388 are able to repress the *Bartonella trw* locus (Seubert *et al.*, 2003). KorAB putative binding sites are also conserved, and have been duplicated in accordance with *trwL* and *trwJIH* tandem duplications. Comparison between the genomes of *Bartonella* and a close relative, *Brucella*, revealed that the *Bartonella trw* locus appears to have been inserted as a pathogenicity island between two conserved genes in *Brucella*. Thus, insertion of the *trw* locus occurred after divergence of *Brucella* and *Bartonella*, but early in *Bartonella* speciation (the *trw* locus is present in all *Bartonella* species). The sequence identity between the *trw* locus from *Bartonella* and the *Mpf* system of R388 indicates that this T4SS was incorporated into the pathogen genome from an IncW plasmid. Extensive sequence conservation since divergence of the Trw systems of *Bartonella* sp. and R388 indicates that T4SS from conjugative plasmids can be incorporated into bacterial genomes and adapted to perform different functions without major changes (de Paz *et al.*, 2005). The dendrograms in Figs 8 and 9 indicate that exchange of T4SSs between plasmids and chromosomes occurred several times in the course of evolution.

Concluding remarks: general trends on plasmid evolution

The genetic programme of a plasmid needs not to be as impermeable in relation to its host as that of a

bacteriophage. On the other hand, it has to be robust enough to adapt to variable genetic as well as environmental backgrounds provided by its different hosts. The genetic constitution of a plasmid should reflect the dynamics provoked by this dichotomy. Here we have analysed the genetic constitution of a group of small conjugative plasmids, the IncW plasmids. The main conclusions we derive from this analysis are given below.

Recombination and mosaicism

In classical phylogenetics, there is a common line of ancestors for a given genome, and it is possible to draw a common tree for the majority of the genes in a genome. Genes found to have different ancestors are commonly attributed to horizontal gene transfer. If recombination were uncommon between plasmids, it would be possible to trace a 'species tree' for plasmid backbone evolution. But when phylogenetic dendrograms for different genes of the same plasmid are constructed, it is commonly found that unrelated proteins result in different evolutionary trees. This is also the case in phage evolution, and therefore the evolution pattern in plasmids and phages is more similar to a web than to a canonical tree (Papke & Doolittle, 2003). As discussed in the Introduction, this is known as genetic mosaicism.

We show in this review that in R388, phylogenetic mosaicism is found not only when considering the genome as a whole, but also in functionally linked modules (*Dtr* and *Mpf*) and even inside a functional module (replication, stable inheritance, establishment). On the other hand, regions with unexpected synteny conservation and congruent phylogenetic trees, as opposed to independent functional modules of strong mosaicism, are also present. This antithetical situation may be a reflection of the profound differences in the way that complex functions are encoded in a genome. Thus, conjugative plasmids might serve as optimal experimental systems to analyse mechanisms of evolution of genome organization.

Genes from the horizontal gene pool

Homology search showed that the IncW backbone is mainly composed of genes coming from the horizontal gene pool. Plasmid genomes appear therefore as a mosaic of genes from this genetic repertoire, specifically from mobile DNA elements. Besides genes encoding housekeeping functions (replication, stability and conjugative transfer), conjugative plasmids contain a number of additional genes (found in many plasmids) with unclear functions. It seems logical to speculate that these genes are responsible for important phenotypes in natural environments, phenotypes that we cannot reproduce experimentally under laboratory conditions. Further research is needed to clarify their role in plasmid physiology.

Owing to the scope of this review, we have not dealt with the structure and evolution of the integron, the most fluid part of the genetic constitution of R388. The R388 integron (shown in yellow in Fig. 1) is a mobile element of recent acquisition, as are transposons containing antibiotic resistance genes in other IncW plasmids. Their insertion in the IncW backbone occurred as a consequence of the selective pressure exerted by the use of antibiotics to fight human pathogens (de la Cruz & Davies, 2000). The changes in genetic structure brought about by integrons belong to a different evolutionary scale and thus are not relevant for this review. For the structure and evolution of integrons the reader is directed to reviews by Rowe-Magnus & Mazel (2002), and Rowe-Magnus *et al*. (2002).

Interrelations between plasmids and their bacterial hosts

A paradigm for a close encounter between mobile DNA elements and their host genomes is the T4SS. This molecular machinery has been extensively exchanged between plasmids and different bacterial chromosomes. The appearance of chromosomal and plasmid-based T4SS in most branches of the trees shown in Figs 8 and 9 is an indicator that this exchange occurred many times in evolution (rather than a single exchange and subsequent divergence). Symbiosis and pathogenesis are closely related to T4SS, indicating that this exchange had an important impact on the lifestyle and even speciation of bacteria. In spite of this, we have shown that most plasmid genes are idiosyncratic, and disseminated among plasmids and other mobile genetic elements in what constitutes the horizontal gene pool.

Plasmids are defined by a conserved genetic structure

Association of functional modules or isolated genes by recombination is a driving force in plasmid evolution. This implies that different gene architectures are continuously arising as a result of extensive recombination events. Gene shuffling will produce a wide variety of different functional combinations, and the virtual repertoire of different genomes is nearly infinite. In the case of bacteriophages, where similar mechanisms operate, no genomically defined phage has been isolated more than once (Pedulla *et al*., 2003), indicating that the diversity of viable and evolutionarily fit varieties is extraordinarily high. This concept has been confirmed by a recent review (Weigel & Seitz, 2006), in which the authors demonstrate that recombination and replication functions in bacteriophages display maximal mosaicism: any theoretical assortment of replication vs.

recombination genes was found in nature. We would like to emphasize that the extreme diversity displayed by bacteriophages seems not to occur in the case of plasmids, where there is a plethora of examples of virtually identical plasmids isolated many times. The isolation of three representatives of the IncW family in a short period of time (1968–1971) in Japan, Greece and Great Britain is a good example of a widely distributed clonal expansion. In that case, the three founding members of IncW family, pSa, R388 and R7K, have a common backbone diverging in the resistance determinants they carry. This backbone has been shown to be unique for IncW plasmids, whereas all of their genes are commonly found in other mobile elements.

Other examples of relative homogeneity in natural plasmids correspond to the IncF (Mulec *et al*., 2002) and IncP families (Pansegrau *et al*., 1994). Whether these observations are the product of a biased sampling (mainly circumscribed to antibiotic-resistant plasmids in the *Enterobacteriaceae*) or they reflect a widespread phenomenon will only be known for certain when more sequences become available. In this regard, the outcome of many bacterial genome projects has enriched the databases with a significant number of plasmid sequences, offering a wider perspective. Efforts are being carried out in order to organize this huge amount of data and determine the extent and content of the horizontal gene pool (Leplae *et al*., 2004). Mobile genetic elements are collected and classified in the ACLAME database (http://aclame.ulb.ac.be) offering a reliable tool to approach the genomics and phylogenetics of these elements.

Plasmid regulatory networks

When a plasmid has been analysed for regulatory circuits, for example the IncP plasmid RK2 or the IncN plasmid pKM101, a complex regulatory network appeared. The existence of a network is anticipated by the abundance of repeated sequences in intergenic regions. R388 and other plasmids also contain such sequences and thus probably a coordinated regulatory network. Coordinated regulation is a consequence of the persistence of stable plasmid genetic backbones. As mentioned before, plasmids seem to be half way between phages and chromosomes in their genetic organization. The regulation networks of conjugative plasmids should be studied in more detail in order to confirm the existence of a global organization in their genomes, as an alternative to a mere juxtaposition of independent modules.

Acknowledgements

This work was supported in part with funds from the Ministerio de Educación y Ciencia of Spain (Grant BFU2005-03477) and the 6th European Union Framework Program (project LSHM-CT-2005_019023) to F.C. M.P.G.-B. and C.R. were supported by contracts from the cited EU project. R.F.-L. is a predoctoral student supported by an FPU fellowship from the Ministerio de Educación y Ciencia of Spain. We wish to thank Matxalen Llosa and Ellen Zechner for helpful discussions and critical reading of the manuscript.

References

Aoki T, Egusa S, Kimura T & Watanabe T (1971) Detection of R factors in naturally occurring *Aeromonas salmonicida* strains. *Appl Microbiol* **22**: 716–717.

Avison MB, von Heldreich CJ, Higgins CS, Bennett PM & Walsh TR (2000) A TEM-2beta-lactamase encoded on an active Tn1-like transposon in the genome of a clinical isolate of *Stenotrophomonas maltophilia*. *J Antimicrob Chemother* **46**: 879–884.

Balzer D, Ziegelin G, Pansegrau W, Kruft V & Lanka E (1992) KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Res* **20**: 1851–1858.

Batchelor RA, Pearson BM, Friis LM, Guerry P & Wells JM (2004) Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology* **150**: 3507–3517.

Belogurov AA, Delver EP & Rodzevich OV (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J Bacteriol* **175**: 4843–4850.

Belogurov AA, Delver EP, Agafonova OV, Belogurova NG, Lee LY & Kado CI (2000) Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. *J Mol Biol* **296**: 969–977.

Bergstrom M, Hermansson M & Dahlberg C (2004) Isolation and sequencing of the replication region of plasmid pBFp1 isolated from a marine biofilm. *Plasmid* **51**: 179–184.

Bjornsdottir SH, Thorbjarnardottir SH & Eggertsson G (2005) Establishment of a gene transfer system for *Rhodothermus marinus*. *Appl Microbiol Biotechnol* **66**: 675–682.

Bradley DE & Cohen DR (1976) Basic characterization of W-pili. *J Gen Microbiol* **97**: 91–103.

Bradley DE, Taylor DE & Cohen DR (1980) Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12. *J Bacteriol* **143**: 1466–1470.

Brussow H & Hendrix RW (2002) Phage genomics: small is beautiful. *Cell* **108**: 13–16.

Cabezon E & de la Cruz F (2006) TrwB: an F(1)-ATPase-like molecular motor involved in DNA transport during bacterial conjugation. *Res Microbiol* **157**: 299–305.

Canals R, Ramirez S, Vilches S, Horsburgh G, Shaw JG, Tomas JM & Merino S (2006) Polar flagellum biogenesis in *Aeromonas hydrophilia*. *J Bacteriol* **188**: 542–555.

Cascales E, Atmakuri K, Liu Z, Binns AN & Christie PJ (2005) *Agrobacterium tumefaciens* oncogenic suppressors inhibit

T-DNA and VirE2 protein substrate binding to the VirD4 coupling protein. *Mol Microbiol* **58**: 565–579.
Causey SC & Brown LR (1978) Transconjugant analysis: limitations on the use of sequence-specific endonucleases for plasmid identification. *J Bacteriol* **135**: 1070–1079.
Chen CY & Kado CI (1994) Inhibition of *Agrobacterium tumefaciens* oncogenicity by the osa gene of pSa. *J Bacteriol* **176**: 5697–5703.
Chernin LS, Lobanok EV, Fomicheva VV & Kartel NA (1984) Crown gall-suppressive Incw R plasmids cause a decrease in auxin production in *Agrobacterium tumefaciens*. *Mol Gen Genet* **195**: 195–199.
Christie PJ & Vogel JP (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* **8**: 354–360.
Christie PJ, Ward JE, Winans SC & Nester EW (1988) The *Agrobacterium tumefaciens* virE2 gene product is a single-stranded-DNA-binding protein that associates with T-DNA. *J Bacteriol* **170**: 2659–2667.
Close SM & Kado CI (1991) The osa gene of pSa encodes a 21.1-kilodalton protein that suppresses *Agrobacterium tumefaciens* oncogenicity. *J Bacteriol* **173**: 5449–5456.
Close SM & Kado CI (1992) A gene near the plasmid pSa origin of replication encodes a nuclease. *Mol Microbiol* **6**: 521–527.
Coetzee JN, Datta N & Hedges RW (1972) R Factors from *Proteus rettgeri*. *J Gen Microbiol* **72**: 543–552.
Corneau N, Emond E & LaPointe G (2004) Molecular characterization of three plasmids from *Bifidobacterium longum*. *Plasmid* **51**: 87–100.
Couturier M, Bex F, Bergquist PL & Maas WK (1988) Identification and classification of bacterial plasmids. *Microbiol Rev* **52**: 375–395.
Datta N & Hedges RW (1971) Compatibility groups among fi–R factors. *Nature* **234**: 222–223.
Datta N & Hedges RW (1972a) Host ranges of R factors. *J Gen Microbiol* **70**: 453–460.
Datta N & Hedges RW (1972b) Trimethoprim resistance conferred by W plasmids in *Enterobacteriaceae*. *J Gen Microbiol* **72**: 349–355.
Datta N & Hughes VM (1983) Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. *Nature* **306**: 616–617.
Datta N, Dacey S, Hughes V, Knight S, Richards H, Williams G, Casewell M & Shannon KP (1980) Distribution of genes for trimethoprim and gentamicin resistance in bacteria and their plasmids in a general hospital. *J Gen Microbiol* **118**: 495–508.
Datta S, Larkin C & Schildbach JF (2003) Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure (Cambridge)* **11**: 1369–1379.
de la Cruz F & Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**: 128–133.
de Paz HD, Sangari FJ, Bolland S, Garcia-Lobo JM, Dehio C, de la Cruz F & Llosa M (2005) Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology-SGM* **151**: 3505–3516.
del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M & Diaz-Orejas R (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**: 434–000.
Delver EP & Belogurov AA (1997) Organization of the leading region of IncN plasmid pKM101 (R46): a regulation controlled by CUP sequence elements. *J Mol Biol* **271**: 13–30.
Doolittle RF, Feng DF, Tsang S, Cho G & Little E (1996) Determining divergence times of the major kingdoms of living organizms with a protein clock. *Science* **271**: 470–477.
Dorokhina OV & Korotiaev AI (1984) Incompatibility properties of pKMR plasmids determining antibiotic resistance and the capacity to produce colonization antigen. *Antibiotiki* **29**: 91–95.
Dostal L, Khare D, Bok J, Heinemann U, Lanka E & Welfle H (2003) RP4 repressor protein KorB binds to the major groove of the operator DNA: a Raman study. *Biochemistry* **42**: 14476–14482.
Draper O, Cesar CE, Machon C, de la Cruz FC & Llosa M (2005) Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc Natl Acad Sci USA* **102**: 16385–16390.
Dreyfus LA & Iglewski BH (1985) Conjugation-mediated genetic exchange in *Legionella pneumophila*. *J Bacteriol* **161**: 80–84.
Easter CL, Schwab H & Helinski DR (1998) Role of the parCBA operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. *J Bacteriol* **180**: 6023–6030.
Ebersbach G & Gerdes K (2005) Plasmid segregation mechanisms. *Annu Rev Genet* **39**: 453–479.
Ernstsson S, Bjornsdottir SH, Jonsson ZO, Thorbjarnardottir SH, Eggertsson G & Palsdottir A (2003) Identification and nucleotide sequence analysis of a cryptic plasmid, pRM21, from *Rhodothermus marinus*. *Plasmid* **49**: 188–191.
Fong ST & Stanisich VA (1989) Location and characterization of 2 functions on Rp1 that inhibit the fertility of the Incw plasmid-R388. *J Gen Microbiol* **135**: 499–502.
Francia MV, Varsaki A, Garcillan-Barcia MP, Latorre A, Drainas C & de la Cruz F (2004) A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev* **28**: 79–100.
Frank AC, Alsmark CM, Thollesson M & Andersson SGE (2005) Functional divergence and horizontal transfer of type IV secretion systems. *Mol Biol Evol* **22**: 1325–1336.
Gerdes K, Jacobsen JS & Franch T (1997) Plasmid stabilization by post-segregational killing. *Genet Eng (NY)* **19**: 49–61.
Gerdes K, Moller-Jensen J & Bugge JR (2000) Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol* **37**: 455–466.
Giraldo R (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev* **26**: 533–554.
Gomis-Ruth FX, Sola M, de la Cruz F & Coll M (2004) Coupling factors in macromolecular type-IV secretion machineries. *Curr Pharm Design* **10**: 1551–1565.

Gonzalez V, Bustos P, Ramirez-Romero MA *et al.* (2003) The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol* **4**: R36.

Gorai AP, Heffron F, Falkow S, Hedges RW & Datta N (1979) Electron microscope heteroduplex studies of sequence relationships among plasmids of the W incompatibility group. *Plasmid* **2**: 485–492.

Gotz A, Pukall R, Smit E, Tietze E, Prager R, Tschape H, van Elsas JD & Smalla K (1996) Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl Environ Microbiol* **62**: 2621–2628.

Grandoso G, Avila P, Cayon A, Hernando MA, Llosa M & de la Cruz F (2000) Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. *J Mol Biol* **295**: 1163–1172.

Greated A, Lambertsen L, Williams PA & Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* **4**: 856–871.

Grohmann E, Muth G & Espinosa M (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* **67**: 277–301.

Guasch A, Lucas M, Moncalian G, Cabezas M, Perez-Luque R, Gomis-Ruth FX, de la CF & Coll M (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* **10**: 1002–1010.

Hamilton-Miller JM (1979) Mechanisms and distribution of bacterial resistance to diaminopyrimidines and sulphonamides. *J Antimicrob Chemother* **5**: 61–73.

Hedges RW & Datta N (1971) fi – R-factors giving cloramphenicol resistance. *Nat New Biol* **234**: 220–221.

Hendrix RW, Lawrence JG, Hatfull GF & Casjens S (2000) The origins and ongoing evolution of viruses. *Trends Microbiol* **8**: 504–508.

Heuer H, Szczepanowski R, Schneiker S, Puhler A, Top EM & Schluter A (2004) The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1 beta group without any accessory genes. *Microbiology-SGM* **150**: 3591–3599.

Hinchliffe E & Vivian A (1980) Restriction mediated by pAV2 affects the transfer of plasmids in *Acinetobacter calcoaceticus*. *J Gen Microbiol* **121**: 419–423.

Hofreuter D, Odenbreit S, Henke G & Haas R (1998) Natural competence for DNA transformation in *Helicobacter pylori*: identification and genetic characterization of the comB locus. *Mol Microbiol* **28**: 1027–1038.

Hofreuter D, Odenbreit S & Haas R (2001) Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol* **41**: 379–391.

Itoh T, Takemoto K, Mori H & Gojobori T (1999) Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol Biol Evol* **16**: 332–346.

Jaffe A, Ogura T & Hiraga S (1985) Effects of the ccd function of the F plasmid on bacterial growth. *J Bacteriol* **163**: 841–849.

Jagura-Burdzy G & Thomas CM (1992) kfrA gene of broad host range plasmid RK2 encodes a novel DNA-binding protein. *J Mol Biol* **225**: 651–660.

Jansen R, Embden JD, Gaastra W & Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565–1575.

Johnson EP, Mincer T, Schwab H, Burgin AB & Helinski DR (1999) Plasmid RK2 ParB protein: purification and nuclease properties. *J Bacteriol* **181**: 6010–6018.

Jones AL, Barth PT & Wilkins BM (1992) Zygotic induction of plasmid ssb and psiB genes following conjugative transfer of Incl1 plasmid Collb-P9. *Mol Microbiol* **6**: 605–613.

Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF & Hendrix RW (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299**: 27–51.

Kalir S & Alon U (2004) Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell* **117**: 713–720.

Kontomichalou P (1971) R-factors controlling resistance to the penicillins. Thesis. University of Athens, Greece.

Kornacki JA, Burlage RS & Figurski DH (1990) The kil-kor regulon of broad-host-range plasmid RK2: nucleotide sequence, polypeptide product, and expression of regulatory gene korC. *J Bacteriol* **172**: 3040–3050.

Kumar S, Tamura K & Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163.

Kwan T, Liu J, Dubow M, Gros P & Pelletier J (2005) The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci USA* **102**: 5174–5179.

Lawley TD, Klimke WA, Gubbins MJ & Frost LS (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**: 1–15.

Lee LY, Gelvin SB & Kado CI (1999) pSa causes oncogenic suppression of *Agrobacterium* by inhibiting VirE2 protein export. *J Bacteriol* **181**: 186–196.

Leplae R, Hebrant A, Wodak SJ & Toussaint A (2004) ACLAME: a classification of mobile genetic elements. *Nucleic Acids Res* **32**: D45–D49.

Lessl M & Lanka E (1994) Common mechanisms in bacterial conjugation and Ti-mediated T-Dna transfer to plant-cells. *Cell* **77**: 321–324.

Liu Z, Guiliani N, Appia-Ayme C, Borne F, Ratouchniak J & Bonnefoy V (2000) Construction and characterization of a recA mutant of *Thiobacillus ferrooxidans* by marker exchange mutagenesis. *J Bacteriol* **182**: 2269–2276.

Llosa M & de la Cruz F (2005) Bacterial conjugation: a potential tool for genomic engineering. *Res Microbiol* **156**: 1–6.

Llosa M, Bolland S & delaCruz F (1994a) Genetic organization of the conjugal DNA processing region of the Incw plasmid-R388. *J Mol Biol* **235**: 448–464.

Llosa M, Bolland S & de la Cruz F (1994b) Conjugation-independent, site-specific recombination at the Orit of the

Incw plasmid R388 mediated by Trwc. *J Bacteriol* **176**: 3210–3217.
Llosa M, Grandoso G, Hernando MA & de la Cruz F (1996) Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J Mol Biol* **264**: 56–67.
Llosa M, Gomis-Ruth FX, Coll M & de la Cruz FF (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol* **45**: 1–8.
Loper JE & Kado CI (1979) Host range conferred by the virulence-specifying plasmid of *Agrobacterium tumefaciens*. *J Bacteriol* **139**: 591–596.
Macartney DP, Williams DR, Stafford T & Thomas CM (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.
Malek JA, Wierzbowski JM, Tao W, Bosak SA, Saranga DJ, Doucette-Stamm L, Smith DR, McEwan PJ & McKernan KJ (2004) Protein interaction mapping on a functional shotgun sequence of *Rickettsia sibirica*. *Nucleic Acids Res* **32**: 1059–1064.
Meyer RR & Laine PS (1990) The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev* **54**: 342–380.
Miele L, Strack B, Kruft V & Lanka E (1991) Gene organization and nucleotide sequence of the primase region of IncP plasmids RP4 and R751. *DNA Seq* **2**: 145–162.
Moncalian G, Grandoso G, Llosa M & de la Cruz F (1997) OriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J Mol Biol* **270**: 188–200.
Morales VM & Sequeira L (1985) Suicide vector for transposon mutagenesis in *Pseudomonas solanacearum*. *J Bacteriol* **163**: 1263–1264.
Moreira D (2000) Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol Microbiol* **35**: 1–5.
Morris DW, Ogden-Swift SR, Virrankoski-Castrodeza V, Ainley K & Parish JH (1978) Transduction of *Myxococcus virescens* by coliphage P1CM: generation of plasmids containing both phage and *Myxococcus* genes. *J Gen Microbiol* **107**: 73–83.
Mulec J, Starcic M & Zgur-Bertok D (2002) F-like plasmid sequences in enteric bacteria of diverse origin, with implication of horizontal transfer and plasmid host range. *Curr Microbiol* **44**: 231–235.
Mushegian AR & Koonin EV (1996) Gene order is not conserved in bacterial evolution. *Trends Genet* **12**: 289–290.
Nagate T, Inoue M, Inoue K & Mitsuhashi S (1978) Plasmid-mediated sulfanilamide resistance. *Microbiol Immunol* **22**: 367–375.
Nakayama Y & Kobayashi I (1998) Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. *Proc Natl Acad Sci USA* **95**: 6442–6447.
Ochman H & Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**: 74–86.
Okumura MS & Kado CI (1992) The region essential for efficient autonomous replication of pSa in *Escherichia coli*. *Mol Gen Genet* **235**: 55–63.
Osborn AM & Boltner D (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**: 202–212.
Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Schwab H, Stanisich VA & Thomas CM (1994) Complete nucleotide-sequence of Birmingham Incp-alpha plasmids – compilation and comparative-analysis. *J Mol Biol* **239**: 623–663.
Patenge N, Berendes A, Engelhardt H, Schuster SC & Oesterhelt D (2001) The *fla* gene cluster is involved in the biogenesis of flagella in *Halobacterium salinarum*. *Mol Microbiol* **41**: 653–663.
Papke RT & Doolittle WF (2003) Phage evolution: new worlds of genomic diversity. *Curr Biol* **13**: R606–R607.
Paterson ES, More MI, Pillay G, Cellini C, Woodgate R, Walker GC, Iyer VN & Winans SC (1999) Genetic analysis of the mobilization and leading regions of the IncN plasmids pKM101 and pCU1. *J Bacteriol* **181**: 2572–2583.
Pattishall KH, Acar J, Burchall JJ, Goldstein FW & Harvey RJ (1977) Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups. *J Biol Chem* **252**: 2319–2323.
Paulsson J (2002) Multileveled selection on plasmid replication. *Genetics* **161**: 1373–1384.
Pedulla ML, Ford ME, Houtz JM *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171–182.
Rainey PB & Cooper TE (2004) Evolution of bacterial diversity and the origins of modularity. *Res Microbiol* **155**: 370–375.
Rees CE & Wilkins BM (1990) Protein transfer into the recipient cell during bacterial conjugation: studies with F and RP4. *Mol Microbiol* **4**: 1199–1205.
Rigby CE & Fraser ADE (1989) Plasmid transfer and plasmid-mediated genetic exchange in *Brucella abortus*. *Can J Vet Res* **53**: 326–330.
Roberts RC, Burioni R & Helinski DR (1990) Genetic characterization of the stabilizing functions of a region of broad-host-range plasmid RK2. *J Bacteriol* **172**: 6204–6216.
Rowe-Magnus DA & Mazel D (2002) The role of integrons in antibiotic resistance gene capture. *Int J Med Microbiol* **292**: 115–125.
Rowe-Magnus DA, Davies J & Mazel D (2002) Impact of integrons and transposons on the evolution of resistance and virulence. *Pathogenicity Islands and the Evolution of Pathogenic Microbes*, *Vol. 2* (Hacker J & Kaper JB, eds), pp. 167–188, Springer Books, Berlin.
Sanderson KE, Janzer J & Head J (1981) Influence of lipopolysaccharide and protein in the cell envelope on recipient capacity in conjugation of *Salmonella typhimurium*. *J Bacteriol* **148**: 283–293.
Schneiker S, Keller M, Droge M, Lanka E, Puhler 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.

Schroder G & Lanka E (2005) The mating pair formation system of conjugative plasmids – A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54**: 1–25.

Seubert A, Hiestand R, de la CF & Dehio C (2003) A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol* **49**: 1253–1266.

Shamaei-Tousi A, Cahill R & Frankel G (2004) Interaction between protein subunits of the type IV secretion system of *Bartonella henselae*. *J Bacteriol* **186**: 4796–4801.

Shaw JJ & Kado CI (1986) Development of a *Vibrio* bioluminescens gene set to monitor citopathogenic bacteria during the ongoing disease process in a non-disruptive manner. *Bio/Technology* **4**: 560–564.

Shinagawa M, Makino S, Hirato T, Ishiguro N & Sato G (1982) Comparison of DNA sequences required for the function of citrate utilization among different citrate utilization plasmids. *J Bacteriol* **151**: 1046–1050.

Silva FJ, Latorre A & Moya A (2003) Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet* **19**: 176–180.

Smalla K, Heuer H, Gotz A, Niemeyer D, Krogerrecklenfort E & Tietze E (2001) Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity of IncQ-like plasmids (Vol. 66, p. 4854, 2000). *Appl Environ Microbiol* **67**: 490.

Smeets LC & Kusters JG (2002) Natural transformation in *Helicobacter pylori*: DNA transport in an unexpected way. *Trends Microbiol* **10**: 159–162.

Smith SL, Stone D, Novak P, Baccanari DP & Burchall JJ (1979) R plasmid dihydrofolate reductase with subunit structure. *J Biol Chem* **254**: 6222–6225.

Sonnhammer ELL, Eddy SR & Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins-Structure Function and Genetics* **28**: 405–420.

Strzelecki AT, Goodman AE & Rogers PL (1987) Behavior of the IncW plasmid Sa in *Zymomonas mobilis*. *Plasmid* **18**: 46–53.

Summers DK, Beton CWH & Withers HL (1993) Multicopy plasmid instability – the dimer catastrophe hypothesis. *Mol Microbiol* **8**: 1031–1038.

Sundberg C, Meek L, Carroll K, Das A & Ream W (1996) VirE1 protein mediates export of the single-stranded DNA-binding protein VirE2 from *Agrobacterium tumefaciens* into plant cells. *J Bacteriol* **178**: 1207–1212.

Tait RC, Close TJ, Rodriguez RL & Kado CI (1982a) Isolation of the origin of replication of the IncW-group plasmid pSa. *Gene* **20**: 39–49.

Tait RC, Lundquist RC & Kado CI (1982b) Genetic map of the crown gall suppressive IncW plasmid pSa. *Mol Gen Genet* **186**: 10–15.

Tait RC, Close TJ, Hagiya M, Rodriguez RL & Kado CI (1983) Construction of cloning vector from the IncW plasmid pSa and the use in the analysis of crown gall tumor formation. *Genetic Engineering in Eukaryotes* (Lurquin PF & Kleinhofs A, eds), pp. 111–123. Plenum Publishing Corporation, New York.

Tauch A, Schneiker S, Selbitschka W *et al.* (2002) The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIPO2 isolated from bacteria of the wheat rhizosphere. *Microbiology-SGM* **148**: 1637–1653.

Tauch A, Puhler A, Kalinowski J & Thierbach G (2003) Plasmids in *Corynebacterium glutamicum* and their molecular classification by comparative genomics. *J Biotechnol* **104**: 27–40.

Thomas CM (2000) Paradigms of plasmid organization. *Mol Microbiol* **37**: 485–491.

Toussaint A & Merlin C (2002) Mobile elements as a combination of functional modules. *Plasmid* **47**: 26–35.

Valentine C & Kado C (1989) Molecular genetics of IncW plasmids. *Promiscuous Plasmids of Gram-Negative Bacteria* (Christopher T, ed), pp. 125–153. Academic Press, London.

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

Vlasak J & Ondrej M (1985) The Effect of Sa and Minisa plasmids on the induction of plant crown galls and hairy roots by *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Folia Microbiologica* **30**: 148–153.

Vogel JP, Andrews HL, Wong SK & Isberg RR (1998) Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* **279**: 873–876.

von Kruger WM & Parish JH (1981) beta-Lactamase activity and resistance to penicillins in *Myxococcus xanthus*. *Arch Microbiol* **130**: 150–154.

Ward DV, Draper O, Zupan JR & Zambryski PC (2002) Peptide linkage mapping of the *Agrobacterium tumefaciens* vir-encoded type IV secretion system reveals protein subassemblies. *Proc Natl Acad Sci USA* **99**: 11493–11500.

Watanabe T, Furuse C & Sakaizumi S (1968) Transduction of various R factors by phage P1 in *Escherichia coli* and by phage P22 in Salmonella typhimurium. *J Bacteriol* **96**: 1791–1795.

Weigel C & Seitz H (2006) Bacteriophage replication modules. *FEMS Microbiol Rev* **30**: 321–381.

Wu LT & Tseng YH (2000) Characterization of the IncW cryptic plasmid pXV2 from *Xanthomonas campestris* pv. *vesicatoria*. *Plasmid* **44**: 163–172.

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG & Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nat Genet* **36**: 486–491.

Zatyka M, Jagura-Burdzy G & Thomas CM (1994) Regulation of transfer genes of promiscuous IncP alpha plasmid RK2: repression of Tra1 region transcription both by relaxosome proteins and by the Tra2 regulator TrbA. *Microbiology* **140**: 2981–2990.

Zielenkiewicz U & Ceglowski P (2001) Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. *Acta Biochim Polon* **48**: 1003–1023.