

ARTICLE

Received 25 Jan 2011 | Accepted 8 Mar 2011 | Published 5 Apr 2011

DOI: 10.1038/ncomms1267

The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination

Peter Norberg^{1,2}, Maria Bergström¹, Vinay Jethava³, Devdatt Dubhashi³ & Malte Hermansson¹

Plasmids are important members of the bacterial mobile gene pool, and are among the most important contributors to horizontal gene transfer between bacteria. They typically harbour a wide spectrum of host beneficial traits, such as antibiotic resistance, inserted into their backbones. Although these inserted elements have drawn considerable interest, evolutionary information about the plasmid backbones, which encode plasmid related traits, is sparse. Here we analyse 25 complete backbone genomes from the broad-host-range IncP-1 plasmid family. Phylogenetic analysis reveals seven clades, in which two plasmids that we isolated from a marine biofilm represent a novel clade. We also found that homologous recombination is a prominent feature of the plasmid backbone evolution. Analysis of genomic signatures indicates that the plasmids have adapted to different host bacterial species. Globally circulating IncP-1 plasmids hence contain mosaic structures of segments derived from several parental plasmids that have evolved in, and adapted to, different, phylogenetically very distant host bacterial species.

¹ Department of Cell and Molecular Biology, Microbiology, University of Gothenburg, Box 462, SE 413 46, Gothenburg, Sweden. ² The Institute of Biomedicine, Department of Infectious Diseases, University of Gothenburg, SE 405 30, Gothenburg, Sweden. ³ Department of Computer Science and Engineering, Computing Science, Chalmers University of Technology and University of Gothenburg, SE 412 96, Gothenburg, Sweden. Correspondence and requests for materials should be addressed to P.N. (email: peter.norberg@gu.se).

The ability of prokaryotes to exchange genes by means of horizontal gene transfer (HGT) has far-reaching implications for our understanding of prokaryotic evolution^{1–4}. One of the most important contributors to HGT is conjugative plasmids, which are self-replicating extra-chromosomal units that code for their own cell-to-cell conjugal transfer systems. The plasmid backbone, which contains genes encoding plasmid-related traits, such as replication control and conjugation functions, is usually loaded with accessory genes, such as antibiotic-resistance and heavy-metal-resistance genes. These are themselves often part of other mobile genetic elements (MGEs), such as transposons and integrons. Plasmids are important in bacterial evolution and in adaptation to environmental changes, because they may carry genes that are useful to the host bacterium. The resulting fitness of a plasmid can therefore be thought of as the sum of a ‘selfish’ component, including conjugative transfer, replication and various maintenance functions, and a component that confers advantages on the host cell, exemplified by antibiotic-resistance genes⁵.

The development of antibiotic resistance in pathogenic bacteria is a serious and growing health concern. One particularly problematic development is the emergence of multiresistance; that is, bacteria becoming resistant to many, if not all, medically used antibiotics. Plasmids have an important role in the spread of antibiotic-resistance genes between bacteria and in the development of multiresistance^{6–8}. Knowledge of the manner in which plasmids evolve is thus important if we are to better understand the fundamentals of prokaryotic evolution and the principles underlying the accumulation and spread of antibiotic resistance in bacterial communities.

Research into IncW plasmids⁹ and F plasmids¹⁰ has suggested recombination, and that rare recombination events may be a driving force behind the creation of new plasmid families. The IncP-1 plasmid group has a broad host range and can be stably maintained in almost all Gram-negative bacteria. IncP-1 plasmids have also been demonstrated to conjugate to Gram-positive bacteria¹¹ and to yeast and eukaryotic cell lines^{12,13}. A recent study using genomic signatures also suggested a broad host range of the IncP-1 plasmids¹⁴. Furthermore, they can also harbour a wide spectrum of antibiotic-resistance genes⁷. Five evolutionary clades have hitherto been described for IncP-1 plasmids: α -clade¹⁵, β -clade¹⁶, γ -clade^{17,18}, δ -clade¹⁷ and ϵ -clade¹⁹. Several previous studies of the evolution of these plasmids focus on differences in MGE incorporated into the backbone^{20–22}. Incorporation and expelling of such elements occur more frequently than do changes in the core backbone, exemplified by plasmids with similar backbones, harbouring different transposons^(15,20,23; and the present report), thus providing information on the relatively recent evolution of the plasmids. Long-term evolution, however, should preferably be based on ‘deep characters’, and analysis of the plasmid backbone may reveal important information on how these plasmids evolve and adapt to their hosts.

Information about recombination of the IncP-1 plasmid backbone has hitherto been sparse, except in a few studies in which occasional recent recombination events were suggested^{19,24}. It has been suggested that recent human activities, such as the use of wastewater treatment plants that mix bacteria from a large number of sources, would increase contacts between bacteria and therefore increase recombination between plasmids⁷. Furthermore, the increased mobility of people and goods would be expected to increase the worldwide spread of these plasmids. Isolation of similar plasmid backbone sequences from different parts of the world seems to support this hypothesis¹⁹.

Here we analysed the complete backbone genomes of 25 IncP-1 plasmids, including two novel plasmids from the marine environment. We demonstrate that recombination is not only a recent phenomenon induced by human interference but also has been a continuous and prominent feature of the IncP-1 backbone evolution. Considering recombination, we describe a consensus

phylogeny of the IncP-1 plasmids presenting a divergence into seven distinct clades. We also analysed plasmid DNA signatures and suggest that the IncP-1 plasmids have different host species histories, and that the plasmids have been temporarily isolated in different host bacteria for sufficiently long times for their genomic signatures to have been influenced.

Results

Plasmid backbone analysis. We analysed the complete backbone DNA sequences of two novel IncP-1 plasmids, designated as pMCBF1 and pMCBF6, isolated from a marine biofilm²⁵, and compared them with 23 previously described IncP-1 plasmids retrieved from GenBank (found through BLAST and literature searches). These include the IncP-1 plasmids that resulted from a recent thorough plasmid search¹⁴. Plasmids pMCBF1 (62,689 bp) and pMCBF6 (66,729 bp) presented identical backbones and differed only in their mercury-resistance transposons, the common backbone will hereafter be referred to as pMCBF1. Putative gene functions are shown in Tables 1 and 2.

The genetic distance between the amino-acid (AA) sequence of each backbone gene in pMCBF1 and their corresponding genes in the 23 previously described IncP-1 plasmids was estimated by a maximum likelihood approach. The backbone gene content in the 25 plasmids differs significantly and only 24 homologues of the 41 backbone genes in pMCBF1 were present in all analysed plasmids (Fig. 1). The AA similarity differed also widely with *trbD* being the most conserved gene. Among all 23 plasmids, plasmid pB4 presents the closest genetic distance to pMCBF1 in genes *trbK*, *trbL*, *traG* and *traO*, whereas pB4 genes *traC2* and *traK* present the longest genetic distance. Similarly, the pJKK5 genes *trbB*, *trbE*, *trbJ*, *traH*, *traJ*, *klcB* and *klcA* presented the closest, and the two genes *upf30.5* and *kleB* in the same plasmid presented the longest genetic distance to pMCBF1. Only plasmids pAKD4 and pQKH54 did not have any gene with the closest genetic distance to pMCBF1. Such alterations of relative genetic distances may be explained either by unequal nucleotide substitution rates or by an evolutionary history including homologous recombination (that is, the fact that the different genes in each plasmid backbone have different ancestries).

To reconstruct their evolutionary history, it was necessary to base the phylogenetic analysis on backbone regions, which are conserved and present in all 25 plasmids. Three such relatively large regions were identified and here referred to as regions A, B and C (Fig. 1). Region A was further divided into subregions A₁ and A₂ to decrease its size. Region A₁ contains the seven genes *trfA*, *ssb*, *trbA*, *trbB*, *trbC*, *trbD* and *trbE*. Although the AA sequences for the genes *ssb* and *trbE* in plasmid pEST4011 and pBS228, respectively, was not available because of ‘truncation by insertion’, the counterpart of the genes was still present, allowing it to be included for analysis. Region A₂ contains the seven genes *trbF* to *trbL*. Region B contains the 11 genes *traE* to *traO*, and region C contains the five genes *kfrA*, *korB*, *korA*, *incC* and *kleE*. The DNA sequences were aligned and gap regions were excluded before further analyses. The four regions were also concatenated and analysed as one large (~19,000 nucleotides) segment. Plasmid pIJB1 was previously described as a recombinant²⁶ with a duplication of the genes *trfA* to *trbE*. In this study, we included the second duplicate in the analysis to analyse an intact A region.

Phylogenetic analysis of the IncP-1 backbone. A splits network (Fig. 2a) was initially constructed for 1,000 bootstrap replicates of the concatenated segments A₁, A₂, B and C of 24 IncP-1 plasmids (plasmid pEST4011 was excluded from the analysis as it lacks the genes in A₂). The network, which presents a combinatorial generalization of phylogenetic trees, presented a star-like topology with seven main clades. pMCBF1 formed a novel clade, hereafter called ζ . As visible in a previous study²⁶, the β -clade¹⁶ could be divided into

Table 1 | Location and putative function of the predicted coding regions of pMCBF1.

Positions	Gene name*	Function*
D 1-315	<i>trbA</i>	Mating pair formation (Mpf) regulation
D 592-1551	<i>trbB</i>	Mpf, ATPase, protein kinase
D 1564-1986	<i>trbC</i>	Mpf
D 1990-2301	<i>trbD</i>	Mpf
D 2298-4841	<i>trbE</i>	Mpf
D 4868-5620	<i>trbF</i>	Mpf
D 5639-6529	<i>trbG</i>	Mpf
D 6532-7029	<i>trbH</i>	Mpf
D 7035-6405	<i>trbI</i>	Mpf
D 8423-9190	<i>trbJ</i>	Mpf
D 9201-9413	<i>trbK</i>	Entry exclusion
D 9425-11107	<i>trbL</i>	Mpf, topoisomerase
D 11043-11714	<i>trbM</i>	Mpf
D 11732-12349	<i>trbN</i>	Mpf,
D 12346-13044	<i>trbP</i>	Mpf
D 12822-13496	<i>upf30.5</i>	Outer membrane protein
C 15971-16660	<i>orf 17</i>	Hypothetical prot.
D 16005-17684	<i>tniA</i>	Transposition
D 17867-18595	<i>tniB</i>	NTP binding
D 18685-19809	<i>tniQ</i>	Transposition of Tn5053
D 19870-20484	<i>tniR</i>	Resolvase
C 20523-21020	<i>orf 22</i>	Hyp. open reading frame
C 20537-20773	<i>mere (urf-1)</i>	Mercury resistance
C 20770-21235	<i>merD</i>	Regulation
C 21152-23074	<i>merA</i>	Mercury reductase
C 22795-23040	<i>merF</i>	Mercury transporter
D 23951-24385	<i>merR</i>	Regulatory prot.
C 23043-23318	<i>merP</i>	Mercury binding
C 23334-23684	<i>merT</i>	Mercury transport
C 23756-24190	<i>merR</i>	Regulatory prot.
D 24321-24917	<i>resA</i>	Resolvase
C 24919-25794	<i>yacC</i>	Hyp. prot. with exonuclease domain
C 25864-30351	<i>traC2</i>	DNA primase
C 30355-30732	<i>traD</i>	DNA transfer
C 30757-32817	<i>traE</i>	DNA topoisomerase
C 32833-33582	<i>traF</i>	Maturation peptidase
C 33366-35270	<i>traG</i>	DNA transport during transfer
C 35563-35934	<i>traH</i>	Relaxosome stabilization
C 35267-37480	<i>traI</i>	DNA relaxase
C 37518-37889	<i>traJ</i>	<i>oriT</i> binding
D 38266-38673	<i>traK</i>	<i>oriT</i> binding
D 38673-39398	<i>traL</i>	Transfer protein, Topoisomerase
D 39398-39835	<i>traM</i>	Transfer protein
C 39881-40492	<i>traN</i>	Muraminidase
C 40631-40978	<i>upf54.8 (traO?)</i>	
C 41146-41337	<i>orf 45</i>	Transcription regulator, LysR family
C 41352-42974	<i>oprN?</i>	Multi-drug efflux (MDE) outer membrane prot. NodT family
C 42946-46113	<i>oqxB mexF</i>	MDE transporter
C 46133-47332	<i>mexE</i>	MDE membrane fusion prot.
C 47515-47829	<i>ispS1</i>	
D 47941-48975	<i>orf 50</i>	Membrane prot.
D 48988-50091	<i>ispS1</i>	Transposase
D 49811-50263	<i>tnpA</i>	Transposase
C 50545-51546	<i>krfA</i>	Regulation, transcriptional repressor
C 51894-52931	<i>korB</i>	Regulation, transcriptional repressor
C 51699-52736	<i>korA</i>	Regulation, transcriptional repressor
C 52733-53824	<i>incC</i>	Regulation, partition
C 54107-54433	<i>kleE</i>	Stable inheritance
C 54596-54814	<i>kleB</i>	Stable inheritance
C 54871-55107	<i>kleA</i>	Stable inheritance
C 55241-55498	<i>korC</i>	Regulation, transcriptional repressor
C 55488-56612	<i>kclB</i>	Stable inheritance
C 56840-57652	<i>istB ?</i>	ATPase
C 57642-59135	<i>orf 63</i>	Resolvase
C 59386-59850	<i>klcA</i>	Antirestriction system
C 60971-62410	<i>trfA</i>	DNA binding, replication initiation
C 62179-62523	<i>ssb</i>	Single-stranded DNA binding

Hyp., hypothetical; NTP, nucleoside 5'-triphosphate; prot., protein.

*By similarity to sequences in GeneBank, nucleotide.

Table 2 | Location and putative function of the predicted coding regions of transposon Tn5058 in pMCBF6.

Positions	Gene name*	Function*
D 16099-17817		Hypothetical protein
C 16311-16790		Hyp. prot.
D 17844-18728	<i>tniB</i>	NTP binding protein
D 18818-19942	<i>tniQ</i>	Transposition
D 20003-20617	<i>tniR</i>	Resolvase
C 20656-20976	<i>tniM</i>	Modulator of transposition
C 20992-21228	<i>merE</i>	Mercury transport
C 21225-21626	<i>merD</i>	Regulation
C 21644-22456	<i>merB</i>	Organomercurial lyase
D 22291-22842	<i>merR2</i>	Regulation
C 22399-22761		Hyp. prot.
C 22542-22724	<i>merR</i>	Regulation
C 23820-24458	<i>merB1</i>	Organomercurial lyase
C 24439-25192	<i>merG</i>	Organomercury resistance
C 25228-27084	<i>merA</i>	Mercury reductase
C 27125-27436	<i>merP</i>	Mercury transport
C 27613-28017	<i>merT</i>	Mercury transport
D 27701-28270	<i>MerR1</i>	Regulation
C 27827-28189		Hyp. prot.
C 27970-28152		Hyp. prot.

Hyp., hypothetical; NTP, nucleoside 5'-triphosphate; prot., protein.

*By similarity to sequences in GeneBank, nucleotide.

two subclades, β -1 and β -2. Parallel edges in the phylogenetic network indicated, however, conflicting phylogenetic signals, possibly resulting from homologous recombination. In particular, in addition to plasmid pIJB1, plasmid pAOVO02 was a putative recombinant, not clustering to any of the above-described clades. A second network, excluding these two plasmids, was therefore constructed for comparison (Fig. 2b).

Recombination analysis. To investigate whether the conflicting phylogenetic signals are caused by homologous recombination or homoplasy, we initially used a statistical test, the ϕ -test, which was recently described to yield reliable results for diverged DNA sequences²⁷. We analysed the complete concatenated segment, as well as three regions separately, to analyse the frequency and location of recombination crossovers (segments A₁ and A₂ were analysed as one segment A to decrease bias of multiple testing). To estimate the frequency of recombinant plasmids, we also divided the data set into six representative subgroups. These subgroups were selected on the basis of clade identity to analyse possible recombination events within the β -1 subclade, which harbour enough members to perform such analysis, and between the different clades. Because all three α -clade plasmids have identical backbone sequences, and because the ϵ , γ , δ and ζ clades were represented by single backbones, it was impossible to investigate whether recombination had occurred within these clades. Consequently, the ϕ -test was applied on 28 data sets. After a Bonferroni correction for multiple tests, the significance level was set to $P = 0.05/28 = 0.002$. The results (Table 3) indicated strong statistical significance ($P < 0.002$) for recombination in the vast majority of the data sets. There was no statistically significant support for recombination crossovers within the three separate segments of the β -1 subclade plasmids or for the A-segment of the data set containing plasmids within subclade β -2 and pKJK5 or for the B-segment of the data set containing pQKH54, pMCBF1, RK2 and pTP6. However, there was high statistically significant support for recombination when the three concatenated segments were analysed, indicating that recombination crossovers are located between, but not necessarily within, the three investigated regions.

To further explore and visualize putative recombination crossovers, we used the Bootscan method, which uses a sliding-window

		β										α			δ		γ	ϵ								
		$\beta-1$					$\beta-2$					α	α	δ	γ	ϵ										
		pB3	pB10	R751	pBP136	pJP4	pB8	pUO1	pTP6	pADP-1	pAMMD-1	pAOVO02 R	pA1	pA81	pB4	pCNB	RIK2	pTB11	pBS228	pEST4011	pAKD4	pIJB1 R	pQKH54	pKJK5		
Region A	A ₁	trfA	0.55	0.55	0.61	0.56	0.55	0.55	0.55	0.55	0.56	0.56	0.89	0.58	0.56	0.56	0.65	0.65	0.65	0.61	0.41	0.40	0.41	0.86	0.60	
		ssb	0.28	0.28	0.28	0.27	0.28	0.28	0.28	0.28	0.28	0.33	0.32	0.32	0.32	0.32	0.61	0.61	0.61	*	—	0.28	0.76	0.29	0.29	
		trbA	0.23	0.26	0.25	0.23	0.26	0.25	0.26	0.25	0.26	0.25	0.23	0.25	0.25	0.25	0.42	0.42	0.42	0.50	0.50	0.50	0.53	0.25	0.25	
		trbB	0.24	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.26	0.27	0.26	0.26	0.30	0.30	0.30	0.25	0.25	0.25	0.32	0.20	0.20	
		trbC	0.32	0.27	0.29	0.27	0.27	0.27	0.27	0.27	0.27	0.32	0.30	0.30	0.30	0.30	0.55	0.55	0.55	0.39	0.39	0.39	0.41	0.33	0.33	
		trbD	0.05	0.05	0.08	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.09	0.09	0.09	0.09	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.18	0.10	
		trbE	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.16	0.15	0.15	0.14	0.14	*	0.18	0.72	0.18	0.23	0.13	0.13	
		trbF	0.30	0.29	0.30	0.30	0.29	0.29	0.29	0.29	0.29	0.30	0.24	0.30	0.30	0.29	0.32	0.32	0.32	—	0.33	0.29	0.54	0.32	0.32	
		trbG	0.26	0.26	0.27	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.22	0.23	0.23	0.23	0.31	0.31	0.31	—	0.37	0.26	0.41	0.23	0.23	
		trbH	0.62	0.60	0.61	0.55	0.60	0.60	0.60	0.60	0.60	0.62	0.57	0.60	0.58	0.58	0.85	0.85	0.85	—	0.87	0.60	1.11	0.60	0.60	
	A ₂	trbI	0.35	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.34	0.36	0.35	0.36	0.50	0.50	0.50	—	0.42	0.36	0.54	0.35	0.35	
		trbJ	0.33	0.33	0.33	0.28	0.33	0.33	0.33	0.33	0.33	0.30	0.24	0.24	0.25	0.23	0.38	0.38	0.38	—	0.37	0.33	0.42	0.22	0.22	
		trbK	0.48	0.46	0.48	0.46	0.46	0.46	0.46	0.46	0.46	1.30	0.45	0.44	0.36	0.44	1.75	1.75	1.75	—	0.53	0.46	1.34	1.34	1.34	
		trbL	0.52	0.51	0.51	0.61	0.51	0.51	0.51	0.51	0.51	0.52	0.44	0.44	0.44	0.44	0.66	0.66	0.66	—	0.81	0.51	0.98	0.83	0.83	
Region B		trbM	0.22	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.26	0.32	0.31	0.32	0.34	0.46	0.46	0.46	—	0.40	0.21	—	0.38	0.38	
		trbN	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.36	0.36	0.38	0.36	0.38	0.60	0.60	0.60	—	0.41	0.35	—	0.36	0.36	
		trbP	0.42	0.43	0.44	0.43	0.43	0.43	0.44	0.43	0.43	0.43	0.40	0.43	0.41	0.41	0.56	0.57	0.57	—	0.62	0.43	0.67	0.47	0.47	
		upf30.5	0.54	0.53	0.52	0.53	0.53	0.53	—	—	—	0.51	0.53	0.53	0.53	—	—	—	—	—	—	—	—	—	0.56	
		Tn	traC2	0.75	0.75	0.75	0.59	0.75	0.75	0.75	0.75	0.68	0.68	0.66	0.81	0.66	0.42	0.42	0.42	0.29	0.50	—	0.55	0.40	0.40	
			traD	0.58	0.58	0.58	0.57	0.58	0.58	0.58	0.58	0.58	0.83	0.75	0.73	0.83	0.73	0.97	0.97	0.97	0.61	0.61	(8.27)	0.70	0.78	
			traE	0.19	0.19	0.38	0.19	0.19	0.19	0.19	0.19	0.19	0.20	0.21	0.20	0.20	0.20	1.08	1.08	1.08	1.05	1.05	1.05	0.44	0.22	
			traF	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.24	0.24	0.28	0.25	0.26	0.42	0.42	0.42	0.35	0.34	0.35	0.37	0.32	
			traG	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.22	0.22	0.22	0.25	0.25	0.26	0.26	0.18	
			traH	0.47	0.48	0.48	0.48	0.48	0.48	0.48	0.48	*	*	0.53	0.55	0.53	*	0.64	0.64	0.64	0.54	0.62	*	0.97	0.41	
Region C		traI	0.45	0.45	0.45	0.54	0.45	0.45	0.45	0.45	0.45	0.43	0.41	0.42	0.43	0.42	0.57	0.57	0.57	0.59	0.58	0.59	0.81	0.45		
		traJ	0.42	0.40	0.40	0.41	0.43	0.40	0.40	0.40	0.40	0.48	0.49	0.48	0.48	0.48	0.61	0.61	0.61	0.73	0.72	0.73	1.05	0.35		
		traK	0.53	0.53	0.53	0.54	0.53	0.53	0.53	0.53	0.53	0.91	0.91	0.95	1.31	0.94	0.93	0.93	0.93	0.64	0.62	*	1.24	1.18		
		traL	0.09	0.10	0.10	0.09	0.10	0.09	0.10	0.09	0.09	0.16	0.16	0.18	0.16	0.16	0.17	0.17	0.17	0.20	0.20	0.20	0.21	0.16		
		traM	0.32	0.32	0.32	0.31	0.32	0.32	0.32	0.32	0.32	0.38	0.38	0.38	0.37	0.37	0.46	0.46	0.46	0.59	0.59	0.59	0.49	0.45		
		traN	0.50	0.49	0.55	0.49	0.49	0.49	0.49	0.49	0.49	0.56	0.33	0.56	0.61	0.56	0.57	0.57	0.57	1.00	0.95	1.00	0.74	0.59		
		traO	0.23	0.23	0.22	0.24	0.23	0.23	0.23	0.23	0.23	0.18	0.19	0.18	0.14	0.18	0.19	0.19	0.19	0.45	0.45	0.45	0.61	0.22		
		Tn	kfrA	0.37	0.37	0.38	0.56	0.37	0.37	0.37	0.37	0.37	0.39	0.39	0.44	0.49	0.37	1.23	1.23	1.23	1.30	1.32	1.30	2.18	1.20	
			korB	0.23	0.23	0.23	0.24	0.23	0.23	0.23	0.23	0.24	0.23	0.27	0.26	0.25	0.26	0.25	0.37	0.37	0.37	0.34	0.34	0.34	0.54	0.31
			incC	0.23	0.25	0.23	0.24	0.25	0.23	0.23	0.23	0.23	*	*	0.25	0.25	0.27	0.25	0.27	0.27	0.27	0.32	0.25	0.32	0.45	0.23
		kleE	0.22	0.23	0.21	0.22	0.23	0.21	0.21	0.21	0.21	0.24	0.24	0.24	0.25	0.24	0.59	0.59	0.59	0.65	0.66	0.65	0.77	0.25		
		kleB	0.73	0.68	0.72	0.61	0.68	0.72	0.72	0.70	0.72	0.67	0.67	0.69	0.66	0.66	0.69	0.70	0.69	0.65	0.63	0.65	1.14	0.66		
		kleA	0.85	—	0.88	—	—	0.88	0.88	0.88	0.88	*	—	—	—	—	0.60	0.60	0.60	—	—	—	—	0.99		
		korC	0.49	0.60	0.47	0.92	0.60	0.47	0.47	0.47	0.47	0.58	0.60	0.58	0.61	0.58	0.53	0.53	0.53	0.79	0.79	0.79	—	0.65		
		kicB	0.27	0.30	0.28	0.27	0.30	0.28	0.28	0.28	0.39	0.26	0.30	0.40	0.29	0.31	0.31	0.31	0.31	0.33	0.33	0.36	0.75	0.27		
		kicA	0.75	0.78	0.75	0.76	0.78	0.75	0.75	0.75	1.01	0.75	0.77	0.77	0.92	0.80	0.79	0.84	0.84	—	0.83	0.83	0.83	—	0.59	
			0.50	0.52	0.50	0.52	0.52	0.50	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.52	0.49	0.49	0.49	0.46	0.46	0.46	0.60	0.41		

Figure 1 | Genetic distances between pMCBF1 and other fully sequenced IncP-1 plasmids. Genetic distances between each gene in pMCBF1 and the corresponding genes in the other 23 analysed plasmids. The plasmid(s) with the longest distance to pMCBF1 is marked in red and the plasmid(s) with the shortest distance is marked in blue for respective gene. Genes not present in specific plasmids are marked with '-' and genes that are at least partially present but not expressed as proteins, or proteins not annotated in GenBank are marked with '*'. Three genomic regions, A (further divided into two subregions A₁ and A₂), B and C were identified as suitable targets for further phylogenetic and signature analysis as they were present in all plasmids.

approach, in which a window of a fixed size is moved step-by-step through the sequence alignment. In each step a phylogenetic tree with bootstrap values for each clade is created. The putative recombinant is selected as the query, and the bootstrap support for each of the other plasmids being the one that clusters closest to the query is plotted. Recombination crossovers are indicated as sudden changes in bootstrap supports. Similarity plots were also constructed using a similar sliding-window approach, illustrating the DNA sequence similarity between the query and the other sequences.

The Bootscan and similarity plots support recombination. One example is pAOVO02, which showed a pattern consistent with recombination between the putative parental plasmids R751, pA1 and pKJK5 (Fig. 3a). These were also supported as parental

plasmids by the similarity plot, except for pKJK5, which showed a lesser similarity to pAOVO02 than the other two. Another example is pB3, which generally presented the closest evolutionary relationship to R751 (Fig. 3b) and a close sequence similarity (>95% on average). In a specific pB3 region, however, the Bootscan plot indicated a closer evolutionary relationship to pKJK5, even though the sequence similarity was only 68–88%. A similar alteration in bootstrap support was seen for pB10 (Fig. 3c), which mostly showed the closest relationship to R751 except in one region that was more related to plasmid pA1, supporting a previous suggestion about recombination in pB10 (ref. 24). The SimPlot also indicated a generally high similarity of >95% to R751 and a high similarity to pA1 in the specific region. Finally, additional SimPlot analyses were

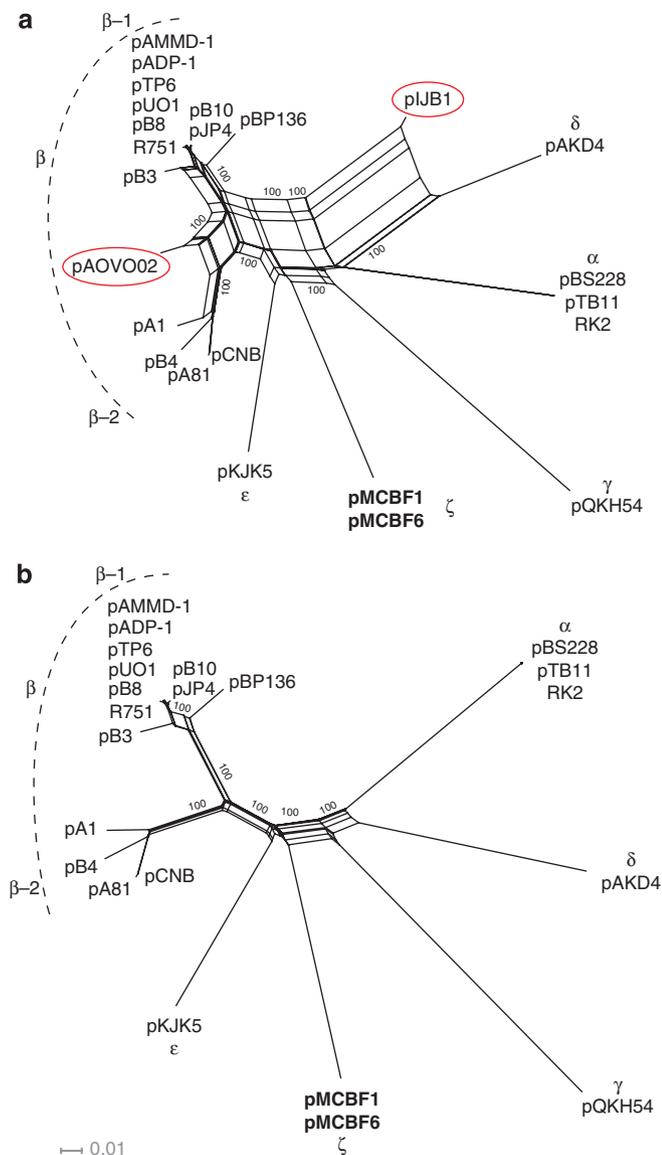


Figure 2 | Phylogenetic analysis of the IncP-1 plasmid backbone.

(a) Phylogenetic network based on the concatenated backbone regions A, B and C of 25 IncP-1 plasmids. The network displays seven main clades, including a novel clade containing the two newly sequenced plasmids pMCBF1 (in bold) and two sub-clades, β -1 and β -2, of the previously described β -clade. The putative recent recombinant plasmids pIJB1 and pAOVO02 are marked with red ellipses. (b) Phylogenetic network excluding the putative recent recombinant plasmids pIJB1 and pAOVO02.

performed to investigate the ancestry of specific recombination fragments. For example, plasmids pB3 and pBP136 shared almost identical sequences with plasmid R751, except in a few regions in which the sequence similarity was significantly less (Fig. 4a). When pBP136 (Fig. 4b) and pB3 (Fig. 4c) were compared with all other plasmids studied here, none of them presented high similarities in these regions for plasmid pBP136 and only plasmid pAOVO02 showed a high similarity in the specific region of pB3. A BLAST search identified no sequence with close similarity to the three regions in pBP136. In summary, we find that the ϕ -test supports recombination between IncP-1 plasmids and Bootscan, and similarity plots further illustrate the recombination crossovers.

Analysis of genomic signatures. Species specificity of a bacterium can be determined by examining its genomic signature (nucleotide

patterns found in its DNA) using different approaches. One such approach is the study of genomic compositions of oligomers of different lengths, so-called DNA words²⁸. The basis for a particular word frequency rests on a multitude of physicochemical properties, such as base stacking energy, propeller twist angle, bendability, position preference and protein deformability, but is also influenced by the codon usage and GC contents of the DNA²⁹. Once a plasmid conjugates to a new host, its signature will ameliorate towards that of the host.

By applying recently developed algorithms^{30,31}, we analysed the genomic signatures in the plasmid backbones to identify putative bacterial hosts. We first created a genomic profile for each of all 1,047 bacterial complete genomic DNA sequences currently available from GenBank. The genomic signatures in the four segments A₁, A₂, B and C for each of the 25 plasmids were then matched against these profiles. To test for statistical significance, we started by investigating whether any of the bacterial species within the genus, which contained the best match, had a high probability of being the host. If no significance was found on the genus level, we stepped up one taxonomic level, testing all members in that specific family. If statistical significance was still not detected, this procedure was repeated until we reached the class level. Thus, the *P*-value indicates whether the signature in a plasmid segment is significantly similar to the signatures of the species in that specific genus, family, order or class (Fig. 5).

The majority of the plasmids presented genomic signatures that were most similar to those of species within the phylum *Proteobacteria* (Fig. 5). Most of these matches were also statistically significant already on the genus or family level. Interestingly, all plasmids had at least two regions with signatures matching species from at least different orders, supporting recombination. In addition, although only statistically significant at the class level, the A₁ segment in plasmid pB3 and all plasmids from the α - and δ -clades, as well as the B-segment in the plasmids from the α -clade, presented a genomic signature most similar to that of species from the *Coriobacteriales* order of the distantly related Gram-positive phylum *Actinobacteria*. To further demonstrate recombination, a statistical test for a cross-region comparison was also performed. In this test, only the best match for a specific segment was compared with the best match for the other segments in that plasmid. The results demonstrate statistically different signatures between all segments that had a best hit on the genus or family level in the above test, which further supports recombination between plasmids from different hosts.

Discussion

We analysed the complete backbone genomes of 25 IncP-1 plasmids and demonstrated a divergence into seven distinct phylogenetic clades, that recombination is a common feature of the plasmid backbone evolution, and an adaptation to different hosts. Evolutionary studies of IncP-1 plasmids are often based on gains and losses of transposons and other MGEs^{20–22}. In particular, the lack of inserted elements was considered to be a sign of ancestry, as in plasmid pBP136, which has been suggested to represent the ancient ancestor of all IncP-1 β plasmids²². However, as MGE are found among plasmids in all described clades, the absence of these may be a poor indicator of ancestry of the IncP-1 group. On the other hand, we demonstrate that plasmid pBP136 is likely to be a recombinant involved in recent recombination events, including parental plasmids from the β -1 subclade and a hitherto unknown clade (Fig. 4). An alternative view would thus be that pBP136 is a result of a β -1 subclade plasmid that has recombined, and exchanged regions, with an ancestral plasmid lacking insertions. Whether there exist such plasmids without insertions or whether insertions can be entirely excised is not yet clear. In any case, frequent insertions and deletions of MGE indicate the recent evolution of plasmids, but the older trajectory of plasmid macroevolution must, as here, be based

Table 3 | Statistical significance of recombination using the ϕ -statistics.

Sequence subset	P-value A + B + C	P-value A	P-value B	P-value C
All	0.00	0.00	0.00	0.00
β -1, β -2	0.00	0.00	0.00	0.00
β -1	1.57$\times 10^{-6}$	0.71	0.31	0.11
β -2, pKJK5	0.00	0.24	0.00	0.00
pEST401, pKJK5, pMCBF1, pQKH54	4.68$\times 10^{-14}$	9.57$\times 10^{-5}$	2.38$\times 10^{-5}$	0.00
pQKH54, pMCBF1, RK2, pTP6	0.00	0.00	0.01	8.37$\times 10^{-5}$
pKJK5, pMCBF1, pQKH54, α	9.31$\times 10^{-13}$	8.29$\times 10^{-4}$	3.04$\times 10^{-5}$	4.12$\times 10^{-6}$

Test for statistical significance of recombination within the concatenated region A + B + C as well as in three subregions A (A_1 , A_2), B and C for all and six subgroups of sequences. Results indicating statistical significance ($P < 0.002$ after a Bonferroni correction for multiple tests) for recombination appear in bold; all other results appear in normal text.

on events such as the mutation, speciation and recombination of the backbone core regions³².

All investigated conjugative plasmids, including IncP-1 plasmids, contain at least one entry exclusion gene³³, which prohibits other plasmids in the same incompatibility family from conjugating to that cell. This exclusion system is believed to confer an evolutionary advantage to the plasmid as it frees the plasmid from competition at segregation during cell division, and protects the plasmid-bearing cell from too many conjugation events^{33,34}. Laboratory experiments suggest that surface exclusion systems in F-plasmids reduce the conjugation rate 100–300 times, and in IncP-1 plasmids this reduction is 10–15 times^{7,33}. As our results indicate frequent recombination of IncP-1 plasmids, which requires the presence of two plasmids in one cell, the experimental results indicating that surface exclusion is leaky are supported by this retrospective study. Furthermore, an early study indicates that different IncP-1 plasmids can coexist in one cell for at least 50 generations³⁵, which may allow time for recombination. Recombination can function as a powerful and essential driving force of evolution by deleting deleterious mutations³⁶, collecting beneficial mutations³⁷ and increasing the rate of adaptation^{38,39}. It is tempting to speculate that there is an optimal balance between saving the plasmid from competition by incompatible plasmids and, on the other hand, allowing sporadic mobility and recombination with plasmids evolved in other host bacteria.

The three backbone regions in pBP136, identified in the similarity plots, did not present a close similarity to any of the other plasmids included in this study (Fig. 4). A BLAST search, which did not find any sequences with a high similarity with these three regions, suggests that previously undescribed IncP-1 plasmid clades exist. It is therefore likely that we have yet seen only a fraction of the IncP-1 plasmid diversity.

No correlation between clade identity and the geographic location of the plasmids was detected by simply comparing isolation site with clade identity. For example, the plasmids of the β -1 subclade were isolated from a hospital (London, UK), a wastewater treatment plant (Braunschweig, Germany), a herbicide spill (Minnesota, USA), industrial sewage (Japan), a mercury-contaminated river (Kazakhstan), Australia and a hospital (Japan)⁴⁰. However, in addition to this apparent worldwide spread, our DNA signature analysis indicates historic isolation of IncP-1 plasmids in specific host bacteria (Fig. 5). Genomic signatures are species specific and likely formed by host replication and repair mechanisms^{31,41–43}, but may also be affected by environmental factors⁴⁴. Given sufficient residence time, plasmid signatures ameliorate towards that of the chromosome^{14,28,42}. We analysed the putative plasmid–host history by using newly developed algorithms based on DNA words of five nucleotides, which were demonstrated to be superior to G + C or dinucleotide signals for classifying a sequence according to its origin^{30,31}. The suggested hosts (Fig. 5) are within groups that are known to harbour IncP-1 plasmids⁷. All plasmids, except pMCBF1, had at least one segment with a genomic signature most similar to those of the *Burkholderiales* order of the *Betaproteobacteria* class

(Fig. 5), signifying the importance of this group as a natural host for IncP-1 plasmids^{14,41}. The finding that all plasmids had segments that clustered with different hosts was also supported by the cross-region analysis, which further supports recombination. Thus, IncP-1 plasmids are recombinants containing regions in their backbones descending from parental plasmids, which have evolved in different hosts and/or under different selection pressures for sufficient time for these unique genomic signatures to evolve. It is noteworthy that with some exceptions the suggested hosts of each segment A_1 , A_2 , B and C are similar for most members within each clade, indicating that recombination happened early in the clade history and that amelioration towards a common DNA signature is slow. In most cases, the best signature match of a segment was statistically significant on the genus or family level, indicating specific adaptation to a host within that genus or family (Fig. 5). On the other hand, in some examples, the signature of the best match was statistically significant only on the order or class level. The cross-region analysis was also unable to demonstrate a statistically significant difference for these regions. Part of the explanation for this low statistical significance might be that the latter regions have resided in several different hosts and have acquired a mixture of signatures. Further development of bioinformatics tools to analyse mixtures of signatures may provide interesting information about the host history of these plasmids that show low statistically significant match to one specific host.

Overall, mean plasmid dinucleotide⁴¹ and trinucleotide signatures¹⁴ were used to suggest plasmid hosts. The latter study showed that the evolutionary host range of the IncP-1 plasmids was broader than the narrow host range of the IncF and IncI plasmids. The hosts suggested in this study, for at least one of the segments in each plasmid, were often close to one of the top five host matches suggested for the overall, whole plasmid analyses by Suzuki *et al.*¹⁴. However, in this study we also demonstrate the significance of homologous recombination in the evolution of IncP-1 plasmids. Segment-wise analyses demonstrated that the combination of a broad host range and recombination leads to the emergence of recombinant IncP-1 backbones that contain segments of significantly different host origins. For example, for six plasmids, the A_1 and B segment signatures showed a similarity to bacteria within Gram-positive *Actinobacteria* (Fig. 5). Interestingly, a recent report showed that the IncP-1 plasmid pKJK5 can transfer to the Gram-positive *Arthrobacter* sp. strain 108 (also class *Actinobacteria*) in soil rhizosphere experiments; this Gram-positive bacterium was in fact the most frequent pKJK5 transconjugant¹¹. The manner in which conjugation was detected showed that the plasmid entered the Gram-positive cell and expressed its fluorescence *gfp* marker gene, but the independent replication of the IncP-1 plasmids was not assessed. It cannot be excluded that IncP-1 plasmids were incorporated into the Gram-positive chromosome and ameliorated, and later recombined to contribute to the present plasmids.

Haines *et al.*⁴⁵ recently demonstrated that the IncP-1 α plasmid RK2 has a mean G + C content of the backbone of 66.6 mol%,

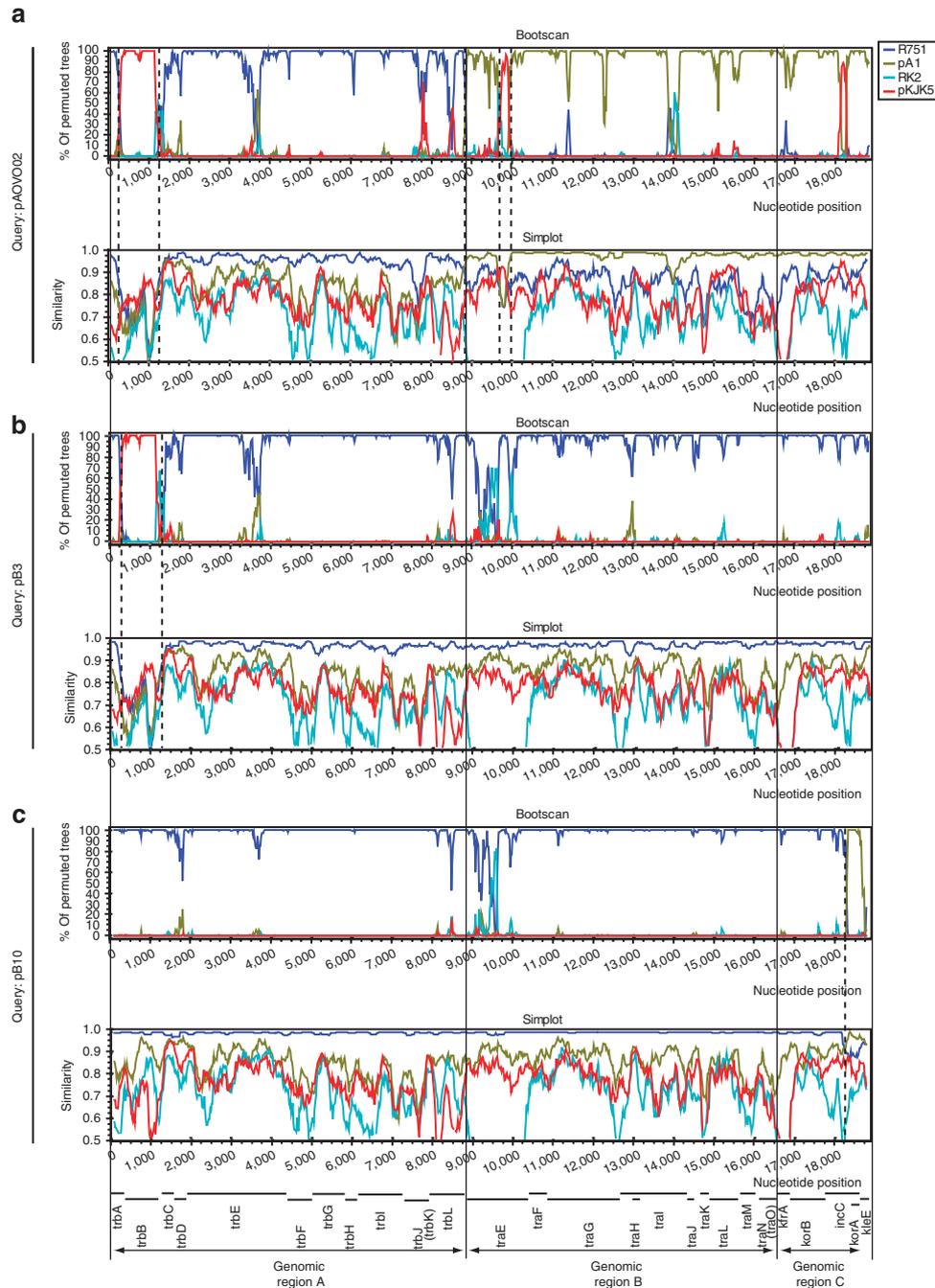


Figure 3 | BootsCan and SimPlot analysis. Analysis of the backbones of plasmids pAOVO02 (**a**), pB3 (**b**) and pB10 (**c**). Each coloured plot corresponds to a specific plasmid depicted in the colour schemes to the right. The bootsCan plot demonstrates phylogenetic relationship to the reference strain, and the SimPlot demonstrates the genetic distances to the reference strain in different parts of the genome. Sudden alterations in bootstrap support, illustrated in the BootsCan plots, indicates recombination. Sequence similarity to the reference strains is represented in the similarity plots beneath the BootsCan plots. Obvious recombination crossovers are highlighted as dotted lines. High sequence similarity indicates recent recombination events. Low sequence similarity indicates ancient recombination events, alternatively recent recombination events involving unanalysed plasmids.

whereas the mean G + C content of pQKH54 (IncP-1 γ) is only 56.6 mol%, and suggested that pQKH54 has resided in a host species with a lower G + C content than that of RK2. The mean G + C content for our suggested hosts for RK2 is 63% whereas the mean G + C for the pQKH54 hosts is 57%, which fits well with the plasmid G + C. Moreover, the pKJK5 backbone genes had a 6.3% lower G + C ratio than that of R751, and these two plasmids were also suggested to have had different host histories¹⁹. The mean G + C content of

our suggested hosts of pKJK5 and R751 is 60 and 65%, respectively. Thus, earlier speculations on plasmid relationships based on G + C content^{19,45} can be substantiated by the DNA signature analysis, which has more predictive power than the G + C content and we can now point to possible hosts.

Perhaps the most important aspects of the evolution and adaptation of the IncP-1 backbone to its different bacterial hosts are the role of these plasmids in HGT and transportation of AB^B genes^{7,40,46},

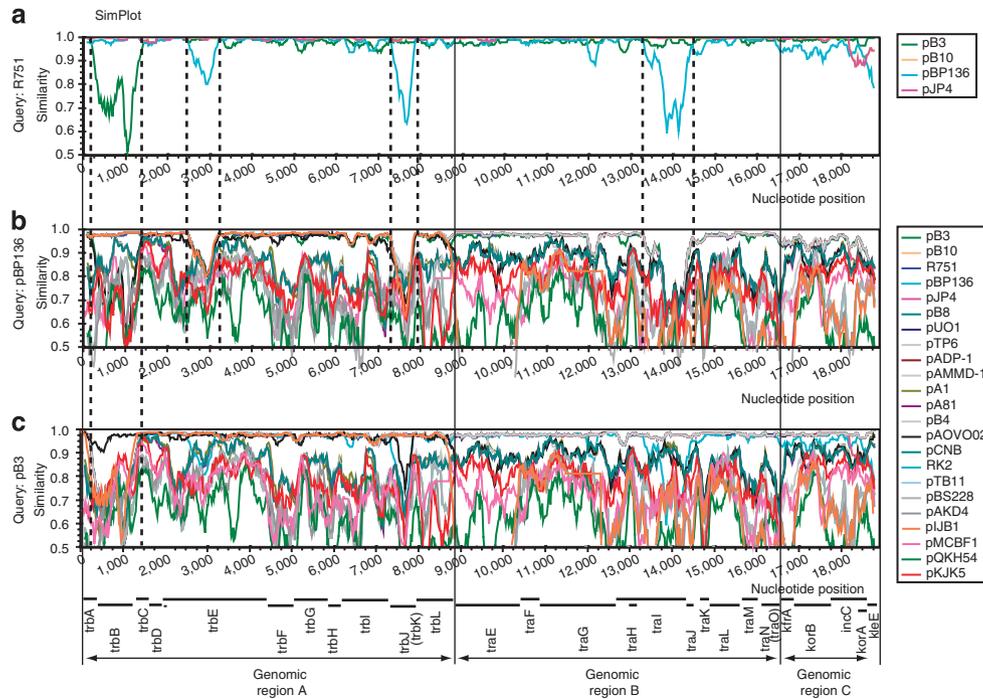


Figure 4 | SimPlot analysis. Similarity plots with plasmids R751, pBP136 and pB3 as reference plasmids. Each coloured plot corresponds to a specific plasmid depicted in the colour schemes to the right and demonstrates the genetic distances from each plasmid to the reference strain in different parts of the genome. The similarity plot of R751 (**a**) highlights one putative recombination event in plasmid pB3 and three putative recombination events in plasmid pBP136. The similarity plots of these two plasmids (**b**, **c**) demonstrate that none of the plasmids included in this study are donors of the recombinant regions in pBP136. Instead, other plasmids from clades that were not previously described were probably involved in these recombination events.

which has major implications for the treatment of human pathogens. Several studies have demonstrated that IncP-1 plasmids can spread to^{47,48} and be maintained in^{40,49} many different bacteria. Our DNA signature analysis demonstrates that the IncP-1 plasmids have been isolated in, and adapted to, different hosts and/or the specific environments the host cells experienced over evolutionary time scales, implying a plasmid/host coevolution. Although surface exclusion has been known to be leaky³³ and incompatibility does not immediately segregate two plasmids³⁵, the extent of direct contact between plasmids in the IncP family is unclear. The frequent pattern of recombination presented here indicates that interactions between IncP-1 plasmid backbones could be direct and not limited to interactions with a third-party MGE. This might be one explanation of the high AB^R mobility in the IncP-1 family, strongly supporting the suggestion of Schlüter *et al.*⁷ that IncP-1 plasmids may be viewed as one of the most potent vehicles for the spread and accumulation of multiantibiotic resistance within and between different bacterial communities.

Methods

Bacterial strains and plasmids and growth conditions. *Pseudomonas putida* UWC1 containing the previously exogenous isolated plasmids pMCBF1 and pMCBF6 (ref. 25) were grown overnight at 26 °C in Luria-Bertani medium⁵⁰ with 10 g of added NaCl l⁻¹ and supplemented with 17 mg l⁻¹ of HgCl₂. *Escherichia coli* were grown overnight at 37 °C in the same medium but supplemented with 50 mg l⁻¹ of ampicillin.

Molecular techniques. Plasmid DNA was obtained using QIAGEN MIDI preps, according to the manufacturer's recommendations (QIAGEN). Shearing of DNA to create a plasmid library was carried out by sonication for 30 s (Branson 1510 sonicator). Sticky ends were filled with Klenow fragments according to the manufacturer's recommendations (MBI Fermenta). Sheared plasmid DNA was subcloned into the *Sma*I site of pBluescript II SK+ (Stratagene) by blunt-end

ligation, and transformed by heat shock (42 °C, 2 min 30 s) into *E. coli* XL-1 Blue (Stratagene). Transformants were picked by blue–white selection; plasmid vectors were isolated and screened for inserts by cutting with restriction enzymes, and analysed on standard agarose gels. Vectors with positive inserts were used as templates in sequencing reactions.

Sequencing. The DNA sequences from the inserts were obtained by using M13 forward and reverse primers from the pBluescript II SK+ and the ABI BigDye Terminator Cycle Sequencing kit (Applied Biosystems). Sequencing was carried out at KI Seq, CGR Sweden, on an ABI 373 automated DNA sequencer (Perkin-Elmer Applied Biosystems). DNA sequences were compiled using Contig Express from the Vector NTI Suite 6.0 (Informax). To close gaps in the sequence, internal custom primers (Invitrogen) were designed. To close gaps and confirm the sequence of the two plasmids, pMCBF1 and pMCBF6 were also sequenced by MWG Biotech AG (Ebersberg; www.mwg-biotech.com) in a 'publication quality' DNA sequencing project, as described by MWG (both strands sequenced and a final data accuracy of >99.995%). Sequences of pMCBF1 and pMCBF6 were deposited in GenBank; Nucleotide Core (accession # AY950444 and EF107516).

DNA and AA sequence analysis. DNA and AA sequences were aligned by using ClustalW included in the BioX program. Genetic distances were calculated using the protdist program included in the phylip package (phylip 3.66), using the Jones–Taylor–Thornton matrix. Gap regions were not eliminated before this analysis as the program itself drops those regions in affected comparisons. All gap regions were, however, removed from the DNA sequence alignment before the phylogenetic analysis. Phylogenetic network analysis and the ϕ -statistics were carried out using the SplitsTree program⁵¹. The splits network (neighbour net) was constructed using the uncorrected *P* character transformation, which computes the proportion of positions at which two sequences differ, and the bootstrap values were derived from 1,000 bootstrap replicates. The SimPlot and Bootscan analyses were performed by using the SimPlot program⁵², with a window size of 200 and 20 bp steps.

All analyses of genomic signatures were based on single intact genomic segments (that is, without alignment and truncation of gap regions). The analysis was carried out by using the program PSTK-Classifier^{30,31}, with a fixed-order Markov model of order 4 (that is, using a word size of five nucleotides). Profiles were first constructed for each of all 1,047 bacterial complete genome sequences currently

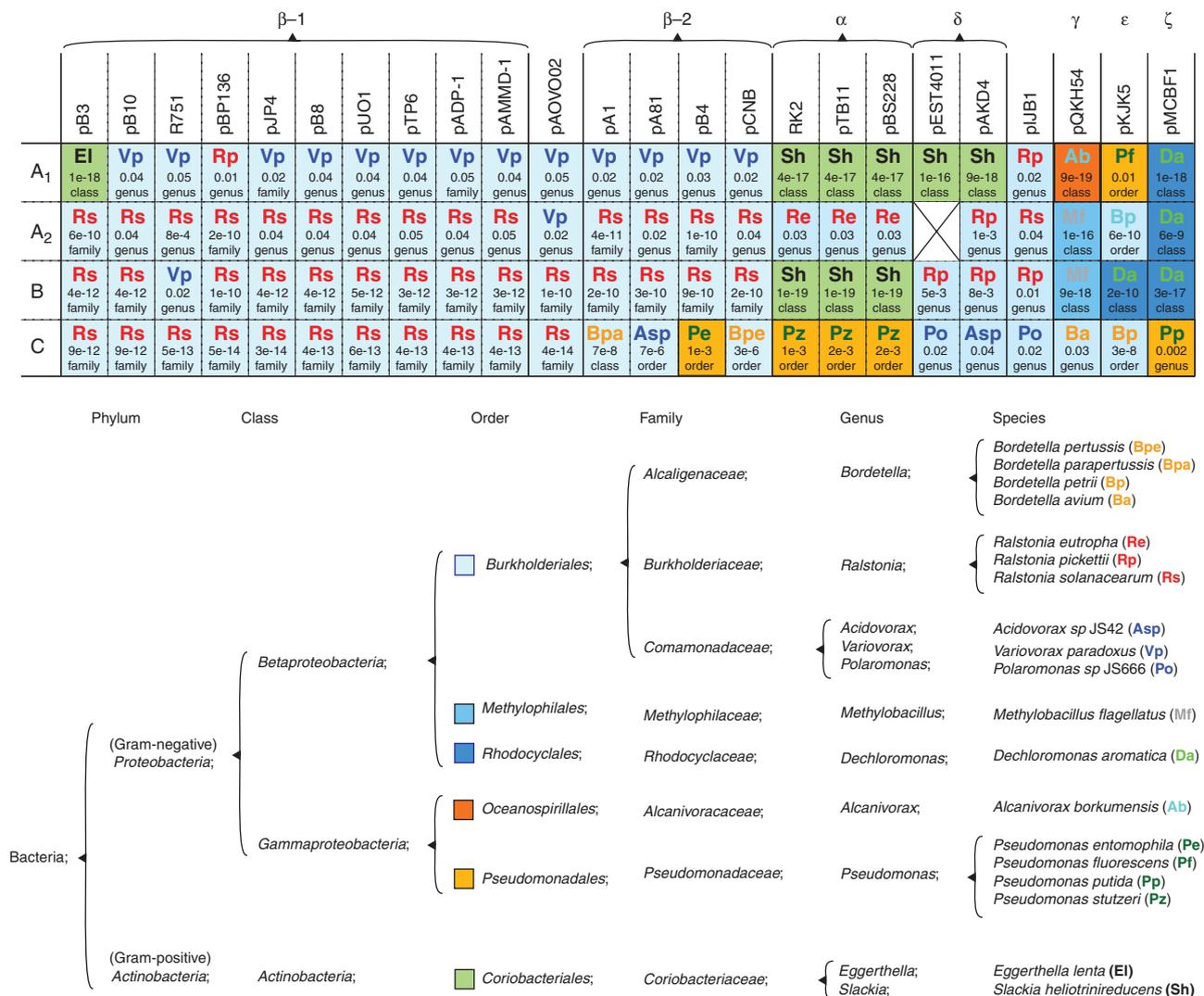


Figure 5 | Analysis of genomic signatures to identify putative hosts. A signature profile was created according to the word frequency for each of the available 1,047 complete genomic bacterial DNA sequences. Further, segments A₁, A₂, B and C of each plasmid were tested independently against these profiles. A P-value indicating the statistical significance was also calculated and indicated for each best match together with the taxonomic level for which the significance was achieved. The background colours in the table demonstrate the order that the putative hosts belong to, and the specific host species are denoted as colour-coded abbreviations.

available from GenBank. All four segments A₁, A₂, B and C in each of the 25 analysed plasmids were then separately matched against these profiles. The Markov classifier determines a score for a bacterium to be the host for a given plasmid. In this way, we can rank various putative host bacteria for a given plasmid. We apply statistical techniques for assessing confidence in our predictions that the top-ranked candidate is the most likely host bacterium: First, we form a list A of the bacteria that are within 5% of the top score. Next, we form a list B of the top-ranked candidate and its closely related neighbours in the Entrez taxonomy database (<http://www.ncbi.nlm.nih.gov/taxonomy>). For this, we traverse the taxonomy up a fixed number of levels and collect all the bacteria that appear below that level. Next we remove from A, those bacteria that also appear in B. Now, our question can be precisely reformulated as follows: Is there a significant difference in scores between the putative hosts in the lists A and B? The null hypothesis is that there is no significant difference, the alternative hypothesis is that there are significantly higher scores in list B. Note that this kind of analysis does not apply to a single putative host but to distinguish two sets of potential hosts. This is required to gain statistical power. In particular, it would assign significance to one taxonomically closely related group of bacteria as being the host as against all the others. We start our analysis on the genus level; that is, we analyse whether the best match is significantly different from the top 5% matches to host bacterial species outside the genus to which the best match belong. If no statistical significance was achieved on the genus level, we moved up one level at a time until the class level was reached.

We applied the Mann-Whitney test⁵³, a powerful non-parametric statistical test to identify whether two samples of observations have equally large values. It computes a test statistic based on the ranks of the elements in a joint series constructed from the two series. The Mann-Whitney test yields a P-value corresponding to observing a result as extreme as observed series under the null hypothesis. There are several reasons to prefer the Mann-Whitney test in our application to other well-known tests, such as the Student's *t*-test: First, it is non-parametric, so it does not assume a fixed underlying distribution such as the Normal distribution, which parametric tests such as the Student's *t*-test do. It is also tailored for ordinal values; that is, the important aspect is the relative order of the data, not their absolute values. This is precisely what we are interested in: the ranks of various bacteria as putative hosts. Furthermore, it is more robust to outliers and hence less likely to assign spurious significance to such data. Finally, it is significantly more efficient than the Student's *t*-test, especially when the underlying distribution is far away from normal.

Another question of interest is whether homologous recombination has created plasmids containing genomic segments, which have evolved in, and adapted to, different host bacterial species. As a complement to the test described above, we performed a cross-region comparison. We compare the best match obtained for each region, and its related neighbours in the hierarchy, against how it compares against the other regions. The null hypothesis is that two regions in a plasmid have evolved in the same host. The alternative hypothesis is that different regions have evolved in different hosts. This test is similar to the test described above with the difference that here we test the best matches against each other irrespective of the top 5% matches.

References

- Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
- Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
- Lake, J. A., Jain, R. & Rivera, M. C. Mix and match in the tree of life. *Science* **283**, 2027–2028 (1999).
- Lawrence, J. G. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* **2**, 519–523 (1999).
- Slater, F. R., Bailey, M. J., Tett, A. J. & Turner, S. L. Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiol. Ecol.* **66**, 3–13 (2008).
- Fluit, A. C. Towards more virulent and antibiotic-resistant *Salmonella*? *FEMS Immunol. Med. Microbiol.* **43**, 1–11 (2005).
- Schlüter, A., Szczepanowski, R., Pühler, A. & Top, E. M. Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol. Rev.* **31**, 449–477 (2007).
- Tennstedt, T., Szczepanowski, R., Braun, S., Pühler, A. & Schlüter, A. Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *FEMS Microbiol. Ecol.* **45**, 239–252 (2003).
- Fernández-López, R. *et al.* Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol. Rev.* **30**, 942–966 (2006).
- Boyd, E. F., Hill, C. W., Rich, S. M. & Hartl, D. L. Mosaic structure of plasmids from natural populations of *Escherichia coli*. *Genetics* **143**, 1091–1100 (1996).
- Musovic, S., Oregaard, G., Kroer, N. & Sørensen, S. J. Cultivation-independent examination of horizontal transfer and host range of an IncP-1 plasmid among Gram-positive and Gram-negative bacteria indigenous to the barley rhizosphere. *Appl. Environ. Microbiol.* **72**, 6687–6692 (2006).
- Heinemann, J. A. & Sprague, G. F. Jr. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).
- Waters, V. L. Conjugation between bacterial and mammalian cells. *Nat. Genet.* **29**, 375–376 (2001).
- Suzuki, H., Yano, H., Brown, C. J. & Top, E. M. Predicting plasmid promiscuity based on genomic signature. *J. Bacteriol.* **192**, 6045–6055 (2010).
- Pansegrau, W. *et al.* Complete nucleotide sequence of Birmingham Inc.P plasmids. Compilation and comparative analysis. *J. Mol. Biol.* **239**, 623–663 (1994).
- Thorsted, P. B. *et al.* Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncP backbone. *J. Mol. Biol.* **282**, 969–990 (1998).
- Vedler, E., Vahter, M. & Heinaru, A. The completely sequenced plasmid pEST4011 contains a novel IncP1 backbone and a catabolic transposon harboring *tfd* genes for 2,4-dichlorophenoxyacetic acid degradation. *J. Bacteriol.* **186**, 7161–7174 (2004).
- Hill, K. E., Weightman, A. J. & Fry, J. C. Isolation and screening of plasmids from the epilithon which mobilize recombinant plasmid pD10. *Appl. Environ. Microbiol.* **58**, 1292–1300 (1992).
- Bahl, M. I., Hansen, L. H., Goesmann, A. & Sørensen, S. J. The multiple antibiotic resistance IncP-1 plasmid pJKK5 isolated from a soil environment is phylogenetically divergent from members of the previously established alpha, beta and delta sub-groups. *Plasmid* **58**, 31–43 (2007).
- Haines, A. S., Jones, K., Batt, S. M., Kosheleva, I. A. & Thomas, C. M. Sequence of plasmid pBS228 and reconstruction of the IncP-1alpha phylogeny. *Plasmid* **58**, 76–83 (2007).
- Trefault, N. *et al.* Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. *Environ. Microbiol.* **6**, 655–668 (2004).
- Kamachi, K. *et al.* Plasmid pBP136 from *Bordetella pertussis* represents an ancestral form of IncP-1beta plasmids without accessory mobile elements. *Microbiology* **152**, 3477–3484 (2006).
- Tennstedt, T., Szczepanowski, R., Krahn, I., Pühler, A. & Schlüter, A. Sequence of the 68,869 bp IncP-1alpha plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. *Plasmid* **53**, 218–238 (2005).
- Schlüter, A. *et al.* The 64 508 bp IncP-1beta antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1beta group. *Microbiology* **149**, 3139–3153 (2003).
- Dahlberg, C., Linberg, C., Torsvik, V. L. & Hermansson, M. Conjugative plasmids isolated from bacteria in marine environments show various degrees of homology to each other and are not closely related to well characterized plasmids. *Appl. Environ. Microbiol.* **63**, 4692–4697 (1997).
- Sen, D. *et al.* Comparative genomics of pAKD4, the prototype IncP-1delta plasmid with a complete backbone. *Plasmid* **63**, 98–107 (2010).
- Bruen, T. C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006).
- Campbell, A., Mrazek, J. & Karlin, S. Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *PNAS* **96**, 9184–9189 (1999).
- Reva, O. N. & Tümmler, B. Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* **6**, 251 (2005).
- Dalevi, D., Dubhashi, D. & Hermansson, M. A new order estimator for fixed and variable length Markov models with applications to DNA sequence similarity. *Stat. Appl. Gen. Mol. Biol.* **5** (2006).
- Dalevi, D., Dubhashi, D. & Hermansson, M. Bayesian classifiers for detecting hgt using fixed and variable order Markov models of genomic signatures. *Bioinformatics* **5**, 517–522 (2006).
- Baquero, F. Environmental stress and evolvability in microbial systems. *Clin. Microbiol. Infect.* **15**, 5–10 (2009).
- Garcillán-Barcia, M. P. & De La Cruz, F. Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid* **60**, 1–18 (2008).
- Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between Bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
- Chikami, G. K., Guiney, D. G., Schmidhauser, T. J. & Helinski, D. R. Comparison of 10 IncP plasmids: homology in the regions involved in plasmid replication. *J. Bacteriol.* **162**, 656–660 (1985).
- Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
- Felsenstein, J. Y. S. The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845–859 (1976).
- Edwards, A. W. The fundamental theorem of natural selection. *Biol. Rev. Camb. Philos. Soc.* **69**, 443–474 (1994).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
- Bahl, M. I., Burmölle, M., Meisner, A., Hansen, L. H. & Sørensen, S. J. All IncP-1 plasmid subgroups, including the novel epsilon subgroup, are prevalent in the influent of a Danish wastewater treatment plant. *Plasmid* **62**, 134–139 (2009).
- Suzuki, H., Sota, M., Brown, C. J. & Top, E. M. Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes. *Nucleic Acids Res.* **36**, e147 (2008).
- Karlin, S. & Burge, C. B. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet.* **11**, 283–290 (1995).
- Mrázek, J. Phylogenetic signals in DNA composition: limitations and prospects. *Mol. Biol. Evol.* **26**, 1163–1169 (2009).
- Kirzhner, V., Paz, A., Volkovich, Z., Nevo, E. & Korol, A. Different clustering of genomes across life using the A-T-C-G and degenerate R-Y alphabets: early and late signaling on genome evolution? *J. Mol. Evol.* **64**, 448–456 (2007).
- Haines, A. S. *et al.* Plasmids from freshwater environments capable of IncQ retrotransfer are diverse and include pQKH54, a new IncP-1 subgroup archetype. *Microbiology* **152**, 2689–2701 (2006).
- Szczepanowski, R. *et al.* Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology* **155**, 2306–2319 (2009).
- Dahlberg, C. *et al.* Interspecies bacterial conjugation by plasmids from marine environments visualized by *gfp* expression. *Mol. Biol. Evol.* **15**, 385–390 (1998).
- Dahlberg, C., Bergström, M. & Hermansson, M. *In situ* detection of high levels of horizontal plasmid transfer in marine bacterial communities. *Appl. Environ. Microbiol.* **64**, 2670–2675 (1998).
- Thomas, C. M. *The Horizontal Gene Pool* (Harwood Academic Publ., 2000).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. S. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory Press, 1982).
- Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
- Lole, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**, 152–160 (1999).
- Ewens, W. J. & Grant, G. R. *Statistical Methods in Bioinformatics: An Introduction* 2nd edn (Springer, 2005).

Acknowledgments

We thank Daniel Dalevi for valuable discussions about the analysis of genomic signatures, and Björn Norberg for artwork. This work was supported by the Swedish Research Council (grant no. 621-2006-2774); the University of Gothenburg; Socialstyrelsen and Svenska Läkaresällskapet foundation; Magnus Bergvalls Foundation; and Wilhelm and Martina Lundgrens Scientific Foundation 1.

Author contributions

M.H. initiated the project. M.B. and M.H. sequenced and annotated the pMCBF1/6 plasmids. P.N. performed the evolutionary analysis (that is, phylogenetic, recombination

and signature analysis), and analysed results. V.J. and D.D. designed and performed the statistical test on genomic signatures. M.H. and P.N. interpreted the results and wrote the manuscript.

Additional information

Accession codes: Sequences of pMCFB1 and pMCFB6 have been deposited in GenBank's Nucleotide Core under accession codes AY950444 and EF107516.

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Norberg, P. *et al.* The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nat. Commun.* 2:268 doi: 10.1038/ncomms1267 (2011).

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>