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

Peter Norberg¹,², 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.
The ability of prokaryotes to exchange genes by means of horizontal gene transfer (HGT) has far-reaching implications for our understanding of prokaryotic evolution. One of the most important contributors to HGT is conjugal 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 genes.

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 multi-resistance. 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 plasmids and F plasmids 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 bacteria and to yeast and eukaryotic cell lines. A recent study using genomic signatures also suggested a broad host range of the IncP-1 plasmids. Furthermore, they can also harbour a wide spectrum of antibiotic-resistance genes. Five evolutionary clades have hitherto been described for IncP-1 plasmids: ε-clade, β-clade, γ-clade, δ-clade, and ζ-clade. Several previous studies of the evolution of these plasmids focus on differences in MGE incorporated into the backbone. 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, 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. 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 plasmids. 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 hypothesis.

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 biofilm, 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 search. 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 homologs 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, truG and truO, whereas pB4 genes truC2 and truK present the longest genetic distance. Similarly, the pJK5 genes trbB, trbE, trbf, trH, trf, klcB and klkA presented the closest, and the two genes upf30.5 and klcE in the same plasmid presented the longest genetic distance to pMCBF1. Only plasmids pAKD4 and pQKH54 differ and 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 ancestors).

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 A, contains the seven genes trfA, ssh, trbA, trbB, trbC, trbD and trbE. Although the AA sequences for the genes ssh 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 A, contains the seven genes trbF to trbL. Region B contains the 11 genes trae to truO, and region C contains the five genes kfrA, korB, korA, incC and klcE. 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 pIB1 was previously described as a recombinant 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 A, A, B and C of 24 IncP-1 plasmids (plasmid pEST4011 was excluded from the analysis as it lacks the genes in A). The network, which presents a combinatorial genealogy 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 study, the β-clade could be divided into...
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 pIJ1B, 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 ψ-test, which was recently described to yield reliable results for diverged DNA sequences.26 We analysed the complete concatenated segment, as well as three regions separately, to analyse the frequency and location of recombination crossovers (segments A, B, and C 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 possible analyse 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, and ζ clades were represented by single backbone, it was impossible to investigate whether recombination had occurred within these clades. Consequently, the ψ-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 pOKH54, 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

---

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

| Positions | Gene name* | Function* |
|-----------|------------|-----------|
| D 1-315   | trbA       | Mating pair formation (Mpf) regulation |
| D 592-1551| trbB       | Mpf, ATPase, protein kinase |
| D 1564-1986| trbC     | Mpf |
| D 1990-2301| trbD     | Mpf |
| D 2298-4841| trbE     | Mpf |
| D 4868-5620| trbF     | Mpf |
| D 5639-6529| trbG     | Mpf |
| D 6532-7029| trbH     | Mpf |
| D 7035-6405| trbI     | Mpf |
| D 8423-9190| trbJ     | Mpf |
| D 9201-9413| trbK     | Entry exclusion |
| D 9425-11107| trbL    | Mpf, topoisomerase |
| D 11043-11714| trbM    | Mpf |
| D 11172-12349| trbN    | Mpf |
| D 12346-13044| trbP    | Mpf |
| D 12822-13496| up30.5  | Outer membrane protein |
| C 15971-16660| orf 17   | Hypothetical prot. |
| D 16005-17684| trnA    | Transposition |
| D 17867-18595| trnB    | NTP binding |
| D 18685-19809| trnQ  | Transposition of Tn5053 |
| D 19970-20484| trnR    | Resolvase |
| C 20523-21020| orf 22   | Hyp. open reading frame |
| C 20537-20773| mere (orf-1)  | Mercury resistance |
| C 20770-21203| merD    | Regulation |
| C 21152-20374| merA    | Mercury reductase |
| C 22795-23040| merF    | Mercury transporter |
| C 23951-24385| merR    | Regulatory prot. |
| C 23043-23318| merP    | Mercury binding |
| C 23334-23684| merT    | Mercury transport |
| C 23756-24190| merR    | Regulatory prot. |
| D 24321-24917| resA   | Resolvase |
| C 24919-25794| yacC   | Hyp. prot. with exonuclease domain |
| C 25864-30351| traCZ  | DNA primase |
| C 30355-30732| traD   | DNA transfer |
| C 30757-32817| traE   | DNA topoisomerase |
| C 32833-33582| traF   | Maturation peptidase |
| C 33366-35270| traG  | DNA transport during transfer |
| C 35563-35934| traH   | Relaxosome stabilization |
| C 35267-37480| traI   | DNA relaxase |
| C 37518-37889| traJ   | orT binding |
| C 38266-38673| traK  | orT binding |
| C 38673-39398| traL   | Transfer protein, Topoisomerase |
| C 39398-39835| traM   | Transfer protein |
| C 40326-40492| traN  | Muramidase |
| C 40631-40978| up54.8 (traQ?)  | |
| C 41146-43377| orf45  | Transcription regulator, LysR family |
| C 41352-42974| oprKF  | Multi-drug efflux (MDE) outer membrane prot. NodT family |
| C 42946-46113| oqxB mexF | MDE transporter |
| C 46133-47332| mexE   | MDE membrane fusion prot. |
| C 47515-47829| ispS1  | Transposase |
| D 47941-48975| orf50  | Membrane prot. |
| D 48988-50091| ispS2  | |
| D 49811-50263| trnA   | Transposase |
| D 50545-51546| krfA  | Regulation, transcriptional repressor |
| C 51894-52931| korB  | Regulation, transcriptional repressor |
| C 51699-52736| korA  | Regulation, transcriptional repressor |
| C 52733-53824| incC  | Regulation, partition |
| C 54107-54433| kicE  | Stable inheritance |
| C 54596-54814| kicB  | Stable inheritance |
| C 54871-55107| kicA  | Stable inheritance |
| C 55241-55498| korC  | Regulation, transcriptional repressor |
| C 55488-56612| kicB  | Stable inheritance |
| C 56840-57652| istB?  | ATPase |
| C 57642-59135| orf63  | Resolvase |
| C 59386-59850| kicA  | Antirestriction system |
| C 60971-62410| trfA  | DNA binding, replication initiation |
| C 62179-62523| ssb   | Single-stranded DNA binding |

*By similarity to sequences in GenBank, nucleotide.

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

| Positions | Gene name* | Function* |
|-----------|------------|-----------|
| D 16099-17817| korC  | Hypothetical protein |
| C 16311-16790| korD  | Hyp. prot. |
| D 17844-18728| trnB  | NTP binding protein |
| D 18818-19942| trnQ  | Transposition |
| D 20003-20617| trnR  | Resolvase |
| C 20656-20976| trnM  | Modulator of transposition |
| C 20992-21228| merE  | Mercury transport |
| C 21225-21626| merD  | Regulation |
| C 21644-22456| merB  | Organomercurial lyase |
| C 22291-22842| merR2 | Regulation |
| C 22399-22761| merR  | Hyp. prot. |
| C 22542-22724| merR1 | Regulation |
| C 23820-24458| merB1 | Organomercurial lyase |
| C 24439-25192| merG  | Organomercurial resistance |
| C 25228-27084| merA  | Mercury reductase |
| C 27125-27436| merP  | Mercury transport |
| C 27613-28017| merT  | Mercury transport |
| C 27701-28270| merR1 | Regulation |
| C 27827-28189| merP  | Hyp. prot. |
| C 27970-28152| merP  | Hyp. prot. |

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

*By similarity to sequences in GenBank, nucleotide.
| Region A | Region B | Region C |
|----------|----------|----------|
| trlA     | trlB     | trlC     |
| ssb      | trlD     | ktrA     |
| trlE     | trlG     | korB     |
| trlF     | trlT     | korA     |
| trlM     | trlN     | incC     |
| trlN     | trlR     | kdeE     |
| A_1      | B        | C        |

**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 A_1 and A_2), B and C were identified as suitable targets for further phylogenetic and signature analysis as they are present in all plasmids.
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 words[^25]. 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 DNA[^26]. Once a plasmid conjugates to a new host, its signature will ameliorate towards that of the host.

By applying recently developed algorithms[^20,21], 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 genome DNA sequences currently available from GenBank. The genomic signatures in the four segments A, B, C and D 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 statistical significant at the class level, the A- and D-segment in plasmid 1 were present in all plasmids from the $\alpha$- and $\delta$-clades, as well as the B-segment in the plasmids from the $\alpha$-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 $\beta$ plasmids[^23]. 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 $\beta$-1 subclade and a hitherto unknown clade (Fig. 4). An alternative view would thus be that pBP136 is a result of a $\beta$-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

---

[^20]: [Link to reference]
[^21]: [Link to reference]
[^22]: [Link to reference]
[^23]: [Link to reference]
on events such as the mutation, speciation and recombination of the backbone core regions31.

All investigated conjugative plasmids, including IncP-1 plasmids, contain at least one entry exclusion gene32, 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 events33,34. 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 times33,35. 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 generations36, which may allow time for recombination. Recombination can function as a powerful and essential driving force of evolution by deleting deleterious mutations37, collecting beneficial mutations38 and increasing the rate of adaptation39,40. 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 localisation 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 mechanisms31,41–43, but may also be affected by environmental factors44. Given sufficient residence time, plasmid signatures ameliorate towards that of the chromosome44,45,42. 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 origin36,31. The suggested hosts (Fig. 5) are within groups that are known to harbour IncP-1 plasmids32. 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 plasmids44,41. 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, A, 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 dinucleotide46 and trinucleotide signatures47 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 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 transconjugant41. 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.43 recently demonstrated that the IncP-1z plasmid RK2 has a mean G+C content of the backbone of 66.6 mol%,

| Sequence subset | P-value A + B + C | P-value A | P-value B | P-value C |
|-----------------|------------------|-----------|-----------|-----------|
| All             | 0.00             | 0.00      | 0.00      |
| β-1, β-2        | 0.00             | 0.00      | 0.00      |
| β-1             | 1.57×10⁻⁶        | 0.71      | 0.31      |
| β-2, pKJK5      | 0.00             | 0.24      | 0.00      |
| pEST401, pKJK5, pMCBF1, pQKH54 | 4.68×10⁻¹⁴ | 9.57×10⁻⁵ | 2.38×10⁻⁵ | 0.00 |
| pQKH54, pMCBF1, RK2, pTP6 | 0.00 | 0.00 | 0.01 | 8.37×10⁻⁵ |
| pKJK5, pMCBF1, pQKH54, α | 9.31×10⁻³ | 8.29×10⁻⁴ | 3.04×10⁻⁵ | 4.12×10⁻⁶ |

Test for statistical significance of recombination within the concatenated region A + B + C as well as in three subregions A (A + A), 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.
whereas the mean G + C content of pQKH54 (IncP-1γ) 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 histories15. 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 content15,16 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 genes4,40,48.
which has major implications for the treatment of human pathogens. Several studies have demonstrated that IncP-1 plasmids can spread to57,58 and be maintained in59,60 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 leaky61,62 and incompatibility does not immediately segregate two plasmids63, 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 AB6 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 medium64 with 10 g of added NaCl® 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 Smal 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 using M13 forward and reverse primers from the pHUSESCRIPT II SK+ and the ABI BigDye Terminator Cycle Sequencing kit (Applied Biosystems). Sequencing was carried out at KS 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 protist 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 program41. 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 program42, 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-Classifier23,24, 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 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 shemes 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.
available from GenBank. All four segments A, A', B and C in each of the 25 analysed plasmids were then separately matched 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.

**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 A, A', 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.

We applied the Mann–Whitney test, 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 parameter is the aspect of the data that are different. This is precisely what we are interested in: the ranks of various bacterial 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.
1. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. 19, 2226–2238 (2002).

2. 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).

3. Lake, J. A., Jain, R. & Rivera, M. C. Mix and match in the tree of life. Science 283, 2027–2030 (1999).

4. Lawrence, J. G. Gene transfer, speciation, and the evolution of bacterial genomes. Curr. Opin. Microbiol. 2, 519–523 (1999).

5. 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).

6. Fluit, A. C. Towards more virulent and antibiotic-resistant Salmonella? FEMS Immunol. Med. Microbiol. 43, 1–11 (2005).

7. Schlüter, A., Szczepanowski, R., Pühler, A. & Schlüter, A. Sequence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. FEMS Microbiol. Ecol. 31, 449–477 (2001).

8. 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).

9. 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).

10. 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).

11. Musovic, S., Svearegaard, G., Groer, N. & Sorensen, 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).

12. Heienmann, J. A. & Sprague, G. F. Jr. Conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 340, 205–209 (1989).

13. Waters, V. L. Conjugation between bacterial and mammalian cells. Nat. Genet. 17, 376–377 (1997).

14. Suzuki, H., Yano, H., Brown, C. J. & Top, E. M. Predicting plasmid promiscuity based on genomic signature. J. Bacteriol. 192, 6045–6050 (2010).

15. Pansegrau, W. et al. Complete nucleotide sequence of Birmingham IncP plasmids. Compilation and comparative analysis. J. Mol. Biol. 239, 623–663 (1994).

16. Thorsted, P. B. et al. Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncPc backbone. J. Mol. Biol. 282, 969–990 (1998).

17. Vedler, E., Valter, M. & Heinari, A. The completely sequenced plasmid pEST4011 contains a novel IncP1 backbone and a catabolic transposon from the epilithon which mobilize recombinant plasmid pD10. Appl. Environ. Microbiol. 58, 1292–1300 (1992).

18. Bahl, M. I., Hansen, L. H., Goesmann, A. & Sorensen, S. J. The multiple antibiotic resistance IncP-1 plasmid pK12 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).

19. Haines, A. S., Jones, K., Batt, S. M., Koshevka, I. A. & Thomas, C. M. Sequence of plasmid pBS228 and reconstruction of the IncP-1 alpha phylogeny. Plasmid 58, 76–83 (2007).

20. Trefault, N. et al. Genetic organization of the catabolic plasmid pP4 from Ralstonia eutropha [MP134 (pP4)] reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. Environ. Microbiol. 6, 655–668 (2004).

21. 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).

22. Tennstedt, T., Szczepanowski, R., Krahn, I., Pühler, A. & Schlüter, A. Sequence of the 68,869 bp IncP-1 alpha plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. Plasmid 55, 218–238 (2006).

23. Schlüter, A. et al. The 64,508 bp IncP-1beta 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).

24. Dallenberg, C., Linberg, C., Torvik, V. L. & Hermansson, M. Conjugal plasmids isolated from bacteria in marine environments show various degrees of similarity, which are not closely related to well characterized plasmids. Appl. Environ. Microbiol. 63, 4692–4697 (1997).

25. Sen, D. et al. Comparative genomics of pAKD4, the prototype IncP-1 delta plasmid with a complete backbone. Plasmid 63, 98–107 (2010).
and signature analysis), and analysed results. V.I. 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 pMCBF1 and pMCBF6 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/