Evolution of the korA–oriV segment of promiscuous IncP plasmids

Christopher M. Thomas, Christopher A. Smith,† John P. Ibbotson,‡
Lynda Johnston§ and Naijin Wang‖

Author for correspondence: Christopher M. Thomas. Tel: +44 121 414 5903. Fax: +44 121 414 5925.

INTRODUCTION

Considerable attention has been focused on plasmids belonging to Escherichia coli incompatibility group P because of their ability to transfer between, and maintain themselves stably in, diverse species of Gram-negative bacteria (for a review see Thomas & Smith, 1987). Most plasmids of this group fall into one of two subfamilies, IncPα or IncPβ. The best-studied IncPα plasmids form an indistinguishable group of plasmids, RK2, RP1, RP4, R68 and R18 (Pansegrau et al., 1994) (Fig. 1). In this paper, the names RK2 and RP4 will be used because the IncPα segments studied have been derived from both isolates. Detailed molecular genetic analysis of RK2 and RP4 has defined the blocks of genes required for vegetative replication and conjugal transfer as well as the genes involved in segregational stability (Thomas & Helinski, 1989; Guiney & Lanka, 1989; Pansegrau et al., 1994). For vegetative replication the plasmid requires oriV, the replication origin (Figurski & Helinski, 1979; Thomas et al., 1980; Stalker et al., 1981), and trfA, which encodes the initiator protein, TrfA (Figurski & Helinski, 1979; Thomas et al., 1980; Shingler & Thomas, 1984a; Smith & Thomas, 1984). For conjugal transfer the

Keywords: broad-host-range plasmid, plasmid R751, genome evolution, plasmid maintenance, antirestriction
plasmid requires oriT, the transfer origin (Guiney & Helinski, 1979; Guiney & Jakobson, 1983; Pansegrau et al., 1990), as well as proteins encoded in the tra–traA operon (Lessa et al., 1993), the traK–traM operon (Guiney et al., 1989) and the trb operon (Lessa et al., 1992).

A unique feature of the IncP plasmid organization is the coordinated expression of these genes for replication and transfer that is mediated by regulatory proteins encoded in the central control operon (korABF) (Mottallebi-Veshareh et al., 1992; Thomson et al., 1993) (Fig. 1). The product of korB is known to repress transcription of both the trfA and trb operons (Shingler & Thomas, 1984b; Motallebi-Veshareh et al., 1992; Thomson et al., 1993; Williams et al., 1993), while multiple binding sites for KorB are found in both the tra and trb regions (Balzer et al., 1992), although whether KorB modulates gene expression when it binds at these points is unknown.

The korA gene product also acts at many points, but all of these are known promoter regions. Some of these regions coincide with promoters repressed by KorB (korAp, kfrAp, kilAp and trfAp) (Shingler & Thomas, 1984b; Smith et al., 1984; Schreiner et al., 1985; Young et al., 1985, 1987; Theophilus et al., 1985, 1989). At the face-to-face divergent promoters trfAp and trbAp, KorA acts both negatively and positively, repressing trfAp and simultaneously derepressing trbAp (Jagura-Burdzky & Thomas, 1994). The other promoters under KorA control are repressed by a third repressor, KorC, but not KorB (korAp, korBp and kilCp) (Thomas et al., 1988; Kornacki et al., 1990, 1993). It has been suggested that the products of the operons which are co-regulated with replication and transfer functions, but do not encode genes essential for either of these processes, are likely to play an auxiliary role in plasmid maintenance and propagation (Figurski et al., 1982; Larsen et al., 1994). Although no clear role has yet been established for stability systems, the IncP plasmids appear to have systems which have not yet been mapped to specific loci and may turn out to be encoded by these ‘cryptic’ operons (Harayama et al., 1980; Jovanovic et al., 1994).

One way of testing this hypothesis is to exploit the fact that the IncP family has two branches which have undergone considerable divergence from their common ancestor but have retained many fundamental replication, transfer and control functions intact (Lanka et al., 1985; Meyer & Shapiro, 1980; Figurski et al., 1982; Villaroel et al., 1983; Chikami et al., 1985; Yakobson & Guiney, 1983). The best-studied IncPβ plasmid is R751 (Johanputra & Datta, 1974) (Fig. 1). Sequence comparisons have already shown that considerable drift has occurred where the sequences are under less constraint (Smith & Thomas, 1985; Furste et al., 1989; Miele et al., 1991; Waters et al., 1992; Ziegelin et al., 1991). Therefore, conservation of function between IncPα and IncPβ plasmids indicates that the gene provides a selective advantage for the plasmid. In this paper we present a comparison of the plasmid sector between korA and oriV. The results allow identification of the ‘core’ functions in this region which are present in both IncP subfamilies.

METHODS

DNA manipulations and sequencing. Plasmid DNA was isolated from E. coli C600 by the method of Birnboim & Doly (1979) for small-scale preparation and by a modified version of this method for large-scale preparation (Smith & Thomas, 1983). Manipulation and analysis of DNA was essentially as described by Maniatis et al. (1982). DNA sequencing was carried out by the method of Maxam & Gilbert (1980), with minor modifications (Smith & Thomas, 1987), and by chain-termination (Sanget et al., 1977), using either a Sequenase 2 Kit from USB or ABI Kit with dye-terminators, followed by analysis on an ABI 373A automated DNA sequencer. Computer manipulations of the DNA sequence were carried out using the University of Wisconsin GCG package (Devereux et al., 1984).

RESULTS AND DISCUSSION

The korA–oriV segment of R751 is shorter than that of RK2

We previously cloned the 9.8 kb SalI fragment of R751 running from coordinates 0 to 10 on the standard map (Fig. 1). We subcloned orfV from this and determined its DNA sequence (Smith & Thomas, 1985). Subsequently we have sequenced almost all of this SalI fragment. The segment which is the subject of this paper (GenBank accession number U08908) was located by identifying the position of the regulatory gene korA by searching for homology to the RK2 korA gene. The assumption that there would be a korA gene was based on the observation that R751 is phenotypically KorA+ (Figurski et al., 1982). korA was located easily using the program gap (Devereux et al., 1984). By using orfV and korA as reference points we were able to determine that the distance between them is 4.3 kb. This contrasts with a distance of 12.7 kb for RK2 or 7.7 kb when Tn1 is excluded. Clearly there are additional sequences in this segment of RK2 which are not present in R751.

Alignment of R751 and RK2 sequences between korA and oriV

To compare R751 with RK2, the orfV–korA region of RK2 excluding Tn1 was sequenced. Our sequencing of parts of this region has been reported previously (Thomas et al., 1988; Walter et al., 1991), while the intervening regions have been deposited in the databases (orfV–Tn1, L13392; kleA–klaA, L13287), except for the Tn1–korC region, which had been published by others (M32794; Kornacki et al., 1990). The kleA–kleF, klaA–klaC and kla regions have also been sequenced (Kornacki et al., 1993; Goncharoff et al., 1991; Larsen & Figurski, 1994). Inspection of the RK2 DNA sequence on either side of the Tn1 insertion reveals a 5 bp duplication (TTTTA), which strongly indicates that there is a simple insertion with no associated deletions or rearrangements. This is confirmed by the comparison with R751, whereby the R751 sequences which line up with the RK2 sequence on either side of the Tn1 insertion are contiguous and show no evidence of loss of DNA from RK2 (data not shown). Comparison of the orfV–korA regions by a dotplot allowed us to locate regions of high conservation between
R751 and RK2 (Fig. 2). It is clear from these results that kleA, kleB, korC, kleA (korA1), kleB (korA2), kleE, kleF and korA are highly conserved genes. kleA and kleB are in the operon responsible for the KilC' phenotype (Figurski et al., 1982; Larsen & Figurski, 1994), which is in the absence of korC (KilC override). kle genes belong to the region associated with the KilE' phenotype (Young et al., 1987). Analysis of the DNA sequence with the program CODONPREFERENCE showed that all the new ORFs identified in this segment of R751 conformed to the pattern of codon usage expected from known functional ORFs in RK2/RP4.

**Major deletions/insertions**

A number of gaps or insertions are revealed by the alignment of RK2 and R751 sequences in the oriV-korA region. First, there is an extra 600 bp upstream of kleA in R751 which has no significant sequence identity to the equivalent region of RK2 (Fig. 2a). Secondly, sequence similarity between R751 and RK2 disappears immediately after the end of kleA and does not reappear until about 400 bp later, relative to the RK2 sequence. On R751 the gap is only about 100 bp. In both RK2 and R751 there appears to be a long ORF which occupies the whole kleA-korA gap. However, this would have been disrupted in RK2 by the insertion of Tn1, which might have removed the constraint to maintain the integrity of this gene, allowing the drift which is observed in RK2 (Fig. 2a). Thirdly, there is a break in the diagonal where kleC and kleD are in RK2 (Fig. 2b), consistent with the proposal that the kleA/B and kleC/D operons in RK2 arose by duplication (Thomas et al., 1988). There is a second diagonal displaced to the right that shows kleA of R751 aligning with kleC of RK2, which was expected from the relatedness of RK2 kleA and kleC noted previously by Thomas et al. (1988). Comparison of kleA from RK2 and R751 with kleC from RK2 as well as the predicted polypeptide products (RK2 KleA:R751 KleA, 68%; RK2 KleA:RK2 KleC, 38%; R751 KleA:RK2 KleC, 36%) showed that the two kleA genes are more closely related than either is with kleC. Comparison of codon usage of conserved amino acids showed a much higher proportion of completely conserved codons between the two kleA genes than with the kleA–kleC pairs. One would expect that if the higher degree of conservation were due to functional constraints then drift would have allowed greater divergence in the third position of the codons for conserved amino acids. This strongly suggests that the separation of kleA and kleC preceded the separation of the α and β kleA genes. We therefore propose that this duplication was possessed by the common ancestor of the IncP plasmids, and that it has subsequently been lost by the IncPβ subfamily. A fourth nonconserved region is described in the next section.

**The klaAklkBklac (kilaAtel1Atel2) region is completely absent from R751**

In RK2, the korA gene is preceded by an operon which is regulated by KorA and KorB; the operon is associated with cryptic tellurite resistance and IncW plasmid fertility inhibition (Goncharoff et al., 1991; Walter et al., 1991; Yusoff & Stanisch, 1984). This is completely absent from R751 (Fig. 2b), which is consistent with the lack of cryptic tellurite resistance (Bradley, 1985). Downstream of kleF,
RK2 and R751 diverge so that there is no trace of the klaA promoter/operator region. Sequence identity only returns about 300 bp further on with respect to the R751 sequence, at the korA promoter/operator region (Fig. 3a). These genes therefore have either been deleted from R751 or gained by RK2. The degree of similarity between the korA and kilA promoters of RK2 versus the degree of divergence between the korA promoters of RK2 and R751 (Fig. 4b) suggests that the two promoters must have arisen by a duplication event more recently than the divergence of the IncPα and IncPβ subfamilies. However, there are also considerable similarities between korA and trfA promoters within the IncPα subfamily despite the clear existence of both promoters in the common IncP ancestor, suggesting that selective pressure or gene conversion may maintain promoter similarities within a plasmid genome. This is borne out for the KorC-regulated promoters as described below. Therefore, we cannot make any deductions about whether the common ancestor contained the kla operon. However, we can conclude that the klaA-klaB-klaC operon is not essential for efficient survival of IncP plasmids.
The gap between kleF and korA in R751 contains an additional ORF of 89 codons. We have tentatively called this kleG, but we do not know whether it is expressed as a polypeptide. Database searches revealed no unambiguous similarities, but there was alignment with a variety of bacterial integrases, raising the intriguing possibility that in the IncPα ancestor this region could have promoted acquisition of a foreign DNA segment but may then have been lost in rearrangements to bring the new genes under the control of a promoter derived by duplication of the korA promoter.

Conservation of oriV-related sequences

We have previously reported that oriV sequences diverge significantly (about 35% mismatches) between RK2 and R751 (Smith & Thomas, 1985). Fig. 2(a) shows a grid-like pattern due to alignment of the iterons to which TrfA binds (Pinkney et al., 1987; Kittell & Helinski, 1991), which are repeated nine times in the RK2 oriV region and eight times in the R751 oriV region (the ninth repeat in R751 is located on the other side of Tn4321). These repeats are followed by conserved A+T-rich and G+C-rich regions and then by a region in which we previously identified additional TrfA iterons in direct and in inverted orientation (Smith & Thomas, 1985, 1989). Each of the sequence features shows conservation with RK2. A very closely related sequence also appears at the end of kleA (TGACTGATGAGGGC in R751 compared to TGACAAGAGAGGGC in RK2) and R751 oriV, giving rise to the short repeated diagonals aligning between the end of kilC of R751 and oriV of RK2. The significance of these extra sites is not understood. It is possible that they modulate initiation negatively as proposed for the sites upstream of the minimal origin (Thomas et al., 1984). Alternatively, they may assist in the activation of the replication origin by forming a TrfA-mediated loop as was proposed for RSF1010 (Haring & Scherzinger, 1989).

It was previously found that the HaeII oriV was not able to promote the multiplication of an M13-derivative which was defective in the lagging-strand origin, implying that a primosome assembly site, if it exists, must lie further downstream (D. G. Guiney, personal communication). Unfortunately, the sequence alignments did not reveal an obvious candidate for such a lagging-strand origin or primosome assembly site outside the HaeII oriV fragment.

Transcriptional terminators are not conserved

In RK2 there are very strong putative rho-independent transcriptional terminators located upstream of korA (Thomas & Smith, 1986) and kleA (korA) (Thomas et al., 1988); these define the ends of kleABC and kleAkorC operons, respectively, and appear to protect the downstream operons from read-through transcription. Neither of these putative terminators is conserved in R751 and alternatives do not appear to exist. For korAp, the next upstream promoter is only 1.3 kb away. This gap includes the highly conserved hairpin between kleB (korA2) and kleE, but this is unlikely to be a transcriptional terminator.
C. Figrn

Fig. 3 shows alignment of the RK2 and R751 korC genes (Fig. 3a) and predicted KorA proteins (Fig. 3c): at both the DNA and the polypeptide levels there is 79% identity. Many of the DNA changes affect only the degenerate third position of the korA codons. By contrast, the conservation of the overlapping incC ORF is lower (62% identity), suggesting that KorA has been under more constraint than this part of IncC. Similarly, codon-preference identifies korA as the ORF which conforms to standard IncP codon usage. The proposed helix-turn-helix motif in KorA is reasonably conserved, with four changes in the 21 amino acids of this region. Nevertheless, the sequence identified as the KorA operator is completely conserved between R751 and RK2 despite the high degree of divergence that has occurred between the two plasmids in this region. This suggests that any changes in KorA have not altered the basic operator recognition although they may have changed the strength of interaction. Outside the coding region the degree to which drift has occurred is quite remarkable (Fig. 3a). There is apparently no identity except in regions which should be functionally conserved, such as the −10, −35, O_A and O_B regions, unless a number of gaps are introduced. This suggests that drift has occurred both by base deletion/insertion and by base substitution.

Conservation and drift of korC

The korC genes are also highly conserved. However, the least conserved region is the proposed helix-turn-helix motif (Kornacki et al., 1990). Although there are only two KorC-regulated promoters in R751, it is interesting to note that the putative KorC operators are identical to each other but different by two bases from the RK2 KorC operators (Fig. 5). These changes are symmetrical and preserve the perfect inverted-repeat structure. The changed operator sequence may correlate with changes in the KorC protein. This identity of KorC operators within each plasmid subfamily suggests that there is strong selective pressure to maintain similar repressor-operator interactions at all regulatory points on a single genome. Gene conversion may provide the means to transfer any mutational changes from one control region to another.

The klc operon of R751

It has been proposed that the ancestral IncP plasmid contained an operon consisting of klcA (klcC), klcB (prekorC or ORFX) and korC (Thomas et al., 1988; Larsen & Figurski, 1994). R751 appears to contain this operon intact. The high degree of conservation of the klcA ORF suggests that it still plays an important role in the survival of these plasmids (Fig. 6). It has been noted (Larsen & Figurski, 1994) that KlcA shows 31% sequence identity with the Ardb protein of the IncN plasmid pKM101 (Belogurov et al., 1993), suggesting that KlcA may form part of an antirestriction system which could help RK2 to overcome restriction barriers during establishment after conjugative transfer. Fig. 5 includes the Ardb protein in the alignment of KlcA proteins. The degree of identity with the R751 KlcA (32%) is very similar to that with RK2 KlcA. It is clear that the regions of identity with Ardb lie within the regions of highest identity between the two KlcA proteins, which is consistent with the idea that the sequence conservation represents selection for maintenance of a common structure or function. However, the motif that Belogurov et al. (1993) found to be conserved between different antirestriction proteins (ColIb Ardb, pKM101 Ardb, pKM101 Ardb and T7 0.3 gene product), LLREYVNTL in Ardb, termed the Ardb Box, is not conserved between Ardb and the IncP proteins. Nor is the equivalent region conserved in the KlcA proteins.

Immediately after the end of klcA, the plasmids diverge quite dramatically. However, in both plasmids a second ORF preceded by a Shine–Dalgarno sequence follows the klcA ORF with a G+C-rich hairpin lying between the two. The predicted products of the ORFs show 56.5%
R751 has an extra KorB binding site upstream of klcA

The location, organization and spacing of regulatory sequences are almost identical between the two plasmids. KorA operators are located upstream of −35 regions for klcA and R751Ap and overlapping the −10 for korAp. Putative KorC operators overlap −10 regions for klcA and R751Ap. A KorB operator is found downstream of korC in both plasmids, although it is unclear whether binding of KorB in this region plays a role in regulating either korC or klcA.

The only difference between the two plasmids is in the presence of an extra KorB operator directly upstream of the KorA operator in the klcA promoter. We would expect KorB to bind to this site but the spacing between the KorB and KorA operators is much closer here than observed elsewhere. The possibility that binding of these two regulatory proteins is mutually exclusive at these sites should be investigated since this may represent a novel aspect of modulation of gene expression by products of the central control operon.

Conservation of other proteins

The other conserved ORFs encode KleA, KleB, KleE and KleF. The degree of sequence identity varies from 53% (KleB) to 68% (KleA). Comparison to proteins in the databases with known activities have failed to identify likely functions for these gene products. KleA has a high predicted level of α-helical content and a rather regular arrangement of hydrophobic residues interspersed with hydrophilic residues, which could suggest the formation of a coiled-coil structure. However, this feature is less obvious in R751 than it is in RK2. KleB shows a helix-turn-helix motif. This is flanked by four conserved cysteines with a spacing which aligns well with a variety of zinc-finger motifs. These cysteines could coordinate to Zn⁴⁺ to provide a means of holding the helix-turn-helix motif rigid as an alternative to a larger protein scaffold to which helix-turn-helix motifs are normally attached. Although the gene regulation phenotype, KorE⁺, is associated with this region, the mapping carried out by Kornacki et al. (1993) is not consistent with the KleB protein being responsible for this phenotype. Neither KleE nor KleF display features which clearly indicate structure or function. KleG shows about 30% identity with an internal segment of a fusaric acid resistance protein from Pseudomonas cepacia as well as low-level similarity to a variety of integrases.

Concluding remarks

The sequence comparisons described in this paper allow us to define more closely the IncP backbone which is conserved between α and β subfamilies. This provides a simpler picture than is presented by the IncPx plasmids alone. Studies on other regions of the plasmids have assigned many of the genes which are co-regulated by the central control operon to specific functions such as replication, transfer and stability (Shingler et al., 1984a, b; Smith et al., 1984; Motalett-Veshareh et al., 1990, 1992;
Thomson et al., 1993). This left the korA-oriV region with the largest number of unassigned genes. The combination of the comparison described here (summarized in Fig. 7), which shows that only the klc and kle operons belong to the conserved core, and the suggestion by Larsen & Figurski (1994) that the klc operon may provide an antirestriction system, reduces the number of apparently redundant genes. We hope that this comparison will aid future work to focus on the most important genes and the conserved motifs within these genes, as well as helping to simplify the picture of these uniquely integrated plasmid genomes.

ACKNOWLEDGEMENTS

We thank Grazyna Jagura-Burdzy and Peter Thorsted for help in preparing this manuscript. This work was supported by MRC project grants G8224213CB, G8309838CB and G919550CB awarded to C.M.T. Project students involved in various stages of this work were Zaid Marham, Jane Birch and Allison Williams. L.J. was supported by a studentship from the Northern Ireland Education Authority. N.W. was supported by a Chinese Government Scholarship. Oligonucleotides were synthesized by AltaBioscience, University of Birmingham. Some autoradiography was carried out with a Molecular Dynamics Phosphorimager purchased with shared equipment grants from the Wellcome Trust (038654) and the Medical Research Council (G9216078MB). Some of the sequencing was performed on an ABI Automated Sequencer, run by AltaBioscience and purchased with a shared equipment grant from the Wellcome Trust (038654).

REFERENCES


Received 20 October 1994; revised 9 January 1995; accepted 18 January 1995.