Species-specific repetitive extragenic palindromic (REP) sequences in Pseudomonas putida

Isabel Aranda-Olmedo, Raquel Tobes, Maximino Manzanera, Juan L. Ramos and Silvia Marqués*

Consejo Superior de Investigaciones Científicas, Estación Experimental del Zaidín, Departamento de Bioquímica y Biología Molecular y Celular de Plantas, Apdo. de correos 419, E-18080 Granada, Spain

Received November 9, 2001; Revised and Accepted February 22, 2002

ABSTRACT

Pseudomonas putida KT2440 is a soil bacterium that effectively colonises the roots of many plants and degrades a variety of toxic aromatic compounds. Its genome has recently been sequenced. We describe that a 35 bp sequence with the structure of an imperfect palindrome, originally found repeated three times downstream of the rpoH gene terminator, is detected more than 800 times in the chromosome of this strain. The structure of this DNA segment is analogous to that of the so-called enterobacteriaceae repetitive extragenic palindromic (REP) sequences, although its sequence is different. Computer-assisted analysis of the presence and distribution of this repeated sequence in the P.putida chromosome revealed that in at least 80% of the cases the sequence is extragenic, and in 82% of the cases the distance of this extragenic element to the end of one of the neighbouring genes was <100 bp. This 35 bp element can be found either as a single element, as pairs of elements, or sometimes forming clusters of up to five elements in which they alternate orientation. PCR scanning of chromosomes from different isolates of Pseudomonas sp. strains using oligonucleotides complementary to the most conserved region of this sequence shows that it is only present in isolates of the species P.putida. For this reason we suggest that the P.putida 35 bp element is a distinctive REP sequence in P.putida. This is the first time that REP sequences have been described and characterised in a group of non-enterobacteriaceae.

INTRODUCTION

Pseudomonads are able to metabolise an enormous range of natural and synthetic organic compounds. They play a crucial role in the process of mineralisation and recycling of organic matter in nature (1). Their high adaptation capacity and their catabolic potential have instigated biochemical and genetic studies of different species of the genus Pseudomonas over the years (2). Among them, the soil bacterium Pseudomonas putida has deserved special attention due to its potential applications in biotechnology for the control of environmental pollution, promotion of plant growth and control of pathogens (3).

Pseudomonas putida mt-2 (ATCC 33015) was first described by Stanier and co-workers (1) and was later shown to carry two different pathways for benzoate metabolism (4). Pseudomonas putida KT2440 (5) is a plasmid cured spontaneous rmo derivative of P.putida mt-2, isolated to allow genetic analysis and manipulation. This strain has been extensively characterised both physiologically, biochemically and genetically, so that it is currently considered a representative strain of the species P.putida. Its genome was the subject of an early sequencing programme, through which the sequence of its chromosome has recently been determined and is currently in the process of annotation (www.tigr.org).

Replicative extragenic palindromic (REP) elements were first described in Escherichia coli as 35 bp sequences composed of a highly conserved inverted repeat with the potential of forming a stem–loop structure (6,7). The sequences have been extensively characterised in E.coli and Salmonella typhimurium where they are present in the chromosome more than 500 times either as single independent units or as part of different types of clusters, i.e. bacterial interspersed mosaic elements (BIMEs) (8). Similar sequences were found in Klebsiella pneumoniae and other enterobacteria (9). To date, REP sequences have not been described in other prokaryotic microorganisms.

We previously characterised in detail the rpoH gene of P.putida, which is convergent with the mtgA gene. Analysis of the 198 bp intergenic gene region revealed the presence of a 35 bp inverted repeat element located 13 bp downstream of the rho-independent terminator of the rpoH gene. This 35 bp sequence was repeated three times in this intergenic region. In the present study, we have carried out BLAST scans of the available genome sequence of the strain and we have found that this element is repeated more than 800 times in the chromosome, with a high degree of sequence conservation. We have experimentally determined that this specific 35 bp DNA sequence is restricted to strains of the species P.putida. The possible function of this type of sequences is discussed.

*To whom correspondence should be addressed. Tel: +34 958 121011; Fax: +34 958 129600; Email: silvia@eez.csic.es

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors
MATERIALS AND METHODS

Bacterial strains and growth conditions

The following *Pseudomonas* sp. strains were used in this work: *P. putida* KT2440 (5), *P. putida* SMO116 (10), *P. putida* MTB5 (11), *P. putida* MB6 (11), *P. putida* DOT-T1E (12), *P. putida* F1 (13), *Pseudomonas* sp. JLR11 (14), *Pseudomonas aeruginosa* PAO1162 (15), *P. aeruginosa* SSS1 (15), *P. aeruginosa* 7NSK2 (15), *Pseudomonas fluorescens* (Pseudomonas Reference Culture Collection, CSIC, Granada, Spain), *P. fluorescens* PF11 (16), *P. putida* syringae pv syringae (17), *P. putida* stutzeri (18), *P. stutzeri* EEZ29 (19), *Pseudomonas mendocina* KR1 (20), *Pseudomonas vesicularis* CECT-327 (Spanish Type Culture Collection, University of Valencia, Spain). The *E. coli* strains used in this work: DH5α (21) and ET8000 (22). Escherichia coli strains were routinely grown at 37°C in Luria–Bertani (LB) medium (23). *Pseudomonas* strains were grown at 30°C in LB medium or in minimal medium (basal M9 medium supplemented with Fe-citrate, MgSO₄ and trace metals) with glucose (10 mM) as the sole carbon source (24).

When required, antibiotics were used at the following final concentrations (μg/ml): tetracycline (10), kanamycin (50), rifampicin (20), chloramphenicol (30) and streptomycin (50 or 100).

Plasmids

Plasmids pSH27, pMPR and pMPRH have been described previously (25). The following plasmids were constructed in this study: pA1D is a pMPR derivative that carries the 5′-GCCGCCTCTTTCGGGTTTACCCG-3′ REPa1 sequence cloned in the PsI site between λ5′ and λ3′ in the same orientation as it is found down-stream of *ropH*, pA12 is similar to pA1D but with the REPa1 sequence cloned in the opposite orientation. Plasmid pMPRH0 is a pMPR derivative carrying the *ropH* terminator and the REPa1 sequence obtained as a PCR product from pSH27 using the primers 5′-CTGCAGCAACCGCATCCCC-3′ and 5′-CTGCAGAAAGCCTCTACAGGCCGCTCCTACAGGG-3′, the amplified product was cloned in the PsI site between λ5′ and λ3′. Plasmid pA12 is a pMPR derivative carrying the complete REPa sequence, which includes REPa1 and REPa2 (the sequence of the linker is 5′-GCCGCCTCTTTCGGGTTTACCCGCTCCTACAGGG-3′, cloned in the same orientation as it is found downstream of *ropH*, in the PsI site between λ5′ and λ3′). Plasmid pA12 is similar to pA1D, but with the complete REPa sequence cloned in the opposite orientation. Plasmid pMPRA is a pMPR derivative carrying a random sequence of the same length as REPa1 (5′-GCCGCCTCTTTCGGGTTTACCCGCTCCTACAGGG-3′) cloned between λ5′ and λ3′. Plasmids pMPRA2 and pMPRA3 are equivalent to pMPRA1 but carry two copies of the random sequence cloned in direct (pMPRA2) or inverted (pMPRA3) orientation. All constructions were confirmed by sequencing, and plasmids were introduced into *P. putida* by electroporation and into *E. coli* ET8000 by transformation.

DNA manipulations

Total chromosome or plasmid DNA preparations, digestion of DNA with restriction enzymes, ligations, transformations and agarose or polyacrylamide gel electrophoresis were done according to standard methods (23). Plasmid DNA was isolated by the alkaline lysis method with the QIAprep spin plasmid miniprep kit (Qiagen, Hilden, Germany). DNA cloned in plasmids was sequenced on both strands by using appropriate primers and the dideoxy sequencing termination method with dRhodorhodamine (Perkin Elmer, Madrid, Spain). DNA amplification reactions were carried out in a GeneAmp PCR system 2400 with the Expand High Fidelity PCR system (Roche, Barcelona, Spain). For amplification of *Pseudomonas* genomic DNA (0.1 μg/25 μl reaction), the oligonucleotide REPs 5′-GTAGGAGCGGGTTTACCCGCG-3′ (3 μM) was used. The cycling conditions were as follows: 94°C for 3 min, followed by 25 cycles at 94°C for 1 min, 69°C (stringent) or 50°C (relaxed) for 1 min and 70°C for 10 min, followed by a final 10 min step at 70°C.

β-Galactosidase assays

Overnight cultures grown on LB supplemented with the appropriate antibiotics were diluted 1:100 in fresh medium and grown until the exponential phase was reached. Samples were taken to determine β-galactosidase activity as described previously (26). Data are the mean of three independent assays; SD values were <15% of the given values.

Computer-assisted detection of *P. putida* KT2440 palindromic sequences

Stand-alone BLAST was downloaded from the National Center for Biotechnology Information (NCBI) for local use. A database with all the contigs of *P. putida* [The Institute for Genomic Research (TIGR), released November 7, 2000] was screened against the 35 nt sequence 5′-CCGGCCTCTTTCGGGTTTACCCG-3′, which is found down-stream of the *P. putida* *rpoH* gene. We have developed specific computer programs for assisting in the detection and analysis of REP sequences in the genome of *P. putida*. A set of programs deals with the transformation of the BLAST generated output to an appropriate format for the following. (i) Selection of REP sequences by *E* and length. (ii) Generation of a multiple alignment to derive a consensus sequence for the *P. putida* REP sequence. To this end, all selected REP sequences were transformed to a multifasta format. The multiple sequence alignment software used was Multalin (27). (iii) Study of the distance between two REP sequences and their orientation with respect to each other.

The gene-finding program used to annotate the *P. putida* genome was based on a hidden Markov model (HMM) with modules for coding and non-coding regions, START and STOP codons as well as for ribosome binding sites. The HMM is trained on open reading frames (ORFs) with Swissprot homologues and is then used to identify putative genes which score significantly higher than random ORFs (T.S.Larsen and A.Krogh, manuscript in preparation).

Other series of programs analyse REP sequences with respect to the annotated genome in order to detect (i) extra-genic or intragenic location of the REP sequences; (ii) distance between REP sequences restricted to the same intergenic space; (iii) clusters in the same intergenic space; (iv) the size of the intergenic spaces in the *P. putida* genome; (v) the different types of intergenic spaces defined by the different orientation of the limiting genes (genes of equal sense, convergent genes and divergent genes); (vi) the presence of REP sequences in
each type of intergenic space; and (vii) distances between the REP sequence pairs and the end of the convergent genes limiting an intergenic space. The programs are written in BASIC and will be made available by the authors upon request. A version written in C is currently being developed.

RESULTS

Identification of a repetitive element in the P.putida KT2440 chromosome

In the genome of P.putida KT2440 the rpoH gene is followed by a 34 bp inverted repeat sequence that forms a hairpin and has rho-independent terminator activity (25). This sequence is within the 198 bp-long intergenic sequence that separates rpoH from its neighbouring convergent mtgA gene. Detailed analysis of this 198 bp intergenic region revealed that in addition to the hairpin structure, three copies of a well conserved 35 bp sequence located at 13, 57 and 102 bp from the terminator sequence was present; these were organised as a direct unit followed by two copies in the opposite orientation (Fig. 1A). The sequence within each unit was partially palindromic, containing an internal 6 bp inverted repeat. An initial search for homology with the intergenic rpoH/mtgA region against DNA sequences deposited in GenBank revealed almost 30 sequences with significant hits in which the conserved sequence corresponded to the stretch of 35 bp. All these sequences belonged to strains of the species P.putida or to unidentified Pseudomonas strains. The alignment of these sequences revealed a conserved internal inverted repeat 5'-GCGGGN4CCCGC-3'. The element seemed to be widespread in sequences deposited in GenBank for genes belonging to bacteria of the genus Pseudomonas. Consequently, we decided to analyse in detail the presence of this sequence in the recently completed genome of P.putida KT2440 (www.tigr.com).

We selected the 35-base sequence 5'-CCGGGCTCTTCGGGGTAAAGCCGGCCTCCTACAGGG-3' located downstream of the rpoH gene as a query sequence for a BLAST search in the whole P.putida KT2440. The BLAST available at TIGR had limited the number of sequences with hits in the output. Hence, we used a locally executable BLAST without this limitation. A preliminary analysis of the detected sequences allowed us to establish their main features: (i) a length of 35 nt; (ii) the presence of the central palindromic motif GCGGGmnnnCCCGC; and (iii) a dispersed similarity along the 35 bases.

The BLAST parameter ‘W’ defining the size of word for the initial matching was 11, which is the default value. We realised that it was necessary to set this parameter to 7 to detect a pattern with these characteristics. As only the central motif was totally conserved, the parameter penalising the mismatch (q) was also changed from -3 to -2, and since the central palindromic motif did not allow gaps, the BLAST search was done without gaps. From this BLAST search with these parameters in the complete genome of P.putida we obtained a set of around 1300 sequences.

Multiple alignment of the sequences was carried out and we established a filter of significance (E) and length for this set. We selected sequences longer than 19 bases with an E-value <1. This reduced the number of sequences to 804 (Fig. 2 shows the alignment of the best 30 sequences matching the query) and allowed for definition of the following consensus sequence: 5'-c-ggcctcTTCGGGGTaaACCCGGCCTACAAGGG-3' (upercase letters indicate the presence of this base in this position in 90% of the aligned sequences and lowercase letters in 50% of the sequences) with a central palindromic motif (underlined). The central dyad symmetry was maintained in most of the sequences through compensatory base changes. In 98% of the cases, the change GGGi to GGGa was counterbalanced by the change aCCC to tCCC (data not shown). We evaluated the relevance of this palindromic structure with the rationale that,
if the secondary structure was maintained through compensatory mutations, these mutations would appear at a higher frequency than expected. When we analysed the 804 REP sequences, we found 85 base changes on the left side of the central palindrome GCGGGN4CCCGC, and 106 changes on the right side. Among these changes, 21 were symmetrical and complementary. Given a mutation on the left side of the palindrome, we estimated the probability of finding a compensatory mutation on the other side of the palindrome (pcm) as a factor of the probability of finding a mutation in the symmetrical position (106/804 \times 5), and the probability of the change being complementary (1/3). Therefore, the expected frequency of complementary mutations would be 85 \times pcm = 0.75. However, the observed frequency of complementary mutation was 21, 28-fold of what was expected. These data reinforce the importance of conserving the palindromic structure, and allow us to suggest that the secondary structure is conserved through the selection of compensatory mutations. Because of the structure, abundance, degree of conservation and parallelism of these sequences with the REP sequences previously described in enterobacteriaceae, we believe we have found for the first time REP sequences in pseudomonadaceae (6,7).

**Pseudomonas putida** REP sequences are favoured in extragenic spaces between convergent genes

To detect the extragenic or intragenic allocation of the sequences, the distance between two adjacent sequences and the distance to the flanking genes, we developed computer-assisted methods to analyse these sequences in the context of the ORFs in the genome of *P.putida* KT2440. We found that ~80% of these REP sequences were found in extragenic regions. In 89% of cases, the extragenic region was <500 bases in length (Fig. 3A). Taking into account that only 11% of the genome bases of *P.putida* are extragenic (R. Tobe, unpublished data), this location of REP sequences in extragenic spaces is not likely to be random. We performed a chi-square analysis to test the randomness of this distribution. We concluded that for each base, the intra-REP location and the extragenic location were related features with a probability close to 1 (since the P-value of no relation is 0.000000), thus suggesting that the presence of REP sequences in extragenic spaces is not random. This could reflect a selection against the appearance of these elements within coding regions, or a positive selection for the presence of REP sequences in extragenic spaces.

We also analysed the distance between REP sequences within the same intergenic space. In all cases where more than one REP sequence was present, they were separated by 51 bases and always appeared in opposite orientation. When there are several REP sequences within an intergenic region, these sequences form a cluster. We detected 225 isolated REP sequences, 372 REP sequences forming pairs, 36 forming clusters of three sequences, 12 in clusters of four sequences and one cluster of five sequences (Fig. 3B). This mode of organisation has also been described in the *E.coli* chromosome where some REP sequences are organised as complex elements called BIMEs (28). In *P.putida*, when clustered as pairs, REP sequences were always found in opposite orientation separated by short stretches of sequence, opening the possibility of...
secondary structure formation. This would involve the co-evolution of the sequences in a pair, and would imply that the similarity between two elements in a pair should be higher than the similarity between randomly selected pairs of REP sequences. To test this hypothesis, we performed the following analysis. We selected all the pairs of REP sequences (a pair is defined as two inverted REP sequences located in the same intergenic space and <100 bp apart from each other) and ran a BLAST (Bi2seq) between both elements of each pair. The average E-value obtained was 0.00278. Considering the group of 804 REP sequences, there are 322 806 different possible pairs of REP sequences. We randomly selected 120 967 REP pairs of the 322 806 possible different pairs of the 804 REP sequences and ran a BLAST (Bi2seq) as before. The average E-value obtained in this case was 0.01179, which is a value 4.241 times higher than the average E-value between REP sequence pairs. This suggests a selective pressure favouring similarity between REP sequences in a pair.

With the aim of searching for a relationship between the presence of intergenic REP sequences and the orientation of the genes limiting the intergenic regions, we defined four types of intergenic spaces: (i) between convergent genes; (ii) between divergent genes; (iii) between genes positively oriented (coded by the main DNA chain); and (iv) between genes negatively oriented (coded by the complementary DNA chain) (Table 1). If the REP sequences were randomly allocated in the different types of intergenic spaces, their distribution would depend on the number of intergenic spaces of each type. However, the number of REP sequences between convergent genes is strikingly 2-fold higher than the expected value. Moreover, the presence of REP sequences between divergent genes is less than half of the expected value (Table 1). It is worth noting that very similar results were obtained when only REP sequences that appear as isolated elements were analysed. These data suggest that the REP sequences are preferentially localised in intergenic regions limited by convergent genes, their presence between divergent genes being avoided. Given that this feature seemed to us essential in P. putida REP sequences because of putative functional implications, we wondered if other known REP sequences, i.e. E. coli REP sequences, would share this characteristic. With this aim, we performed a similar analysis of E. coli REP sequences based on data available at www.pasteur.fr/recherche/unites/pmtg/repet/tableauBIMEcoli.html and the E. coli genome annotation available at www.ncbi.nlm.nih.gov/genbank/ genomes/Bacteria/Escherichia_coli_K12/U00096.gbk. Strikingly, we also found that for E. coli the observed frequency of REP sequences between convergent genes was 3-fold the expected value (Table 2).

To search for a putative relationship between every intergenic REP sequence and its two flanking genes, we analysed the distances of each REP sequence to its neighbouring gene considering their START (5′-terminus) or STOP (3′-terminus) codon. Since the frequency of REP sequences between convergent genes was the highest, we could define 829 distances to a START codon and 424 distances to a STOP codon. Figure 4 shows a graphic presentation of the frequency of each distance, ordered by sizes. Clearly, in the set of distances to a STOP codon, the values were grouped around a median value of 56 bp, where the most frequent value (mode) was 15 bp. Of the values, >80% were <92 bp (Fig. 4). In contrast, when we analysed the distances to a START codon, the median was

---

Table 1. Statistical analysis of the organisation of the REP sequences that appear in intergenic spaces

<table>
<thead>
<tr>
<th>Intergenic space types</th>
<th>NSb</th>
<th>RFb</th>
<th>EFc</th>
<th>RF/EFc</th>
</tr>
</thead>
<tbody>
<tr>
<td>→←← convergent</td>
<td>1281</td>
<td>209</td>
<td>219.75</td>
<td>0.95</td>
</tr>
<tr>
<td>←←← convergent</td>
<td>1115</td>
<td>155</td>
<td>190.59</td>
<td>0.81</td>
</tr>
<tr>
<td>←→← divergent</td>
<td>719</td>
<td>236</td>
<td>123.34</td>
<td>1.91</td>
</tr>
<tr>
<td>←←→ divergent</td>
<td>678</td>
<td>50</td>
<td>116.31</td>
<td>0.42</td>
</tr>
<tr>
<td>Total</td>
<td>3789</td>
<td>650</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aTypes of intergenic spaces depending on gene orientation, which is schematised by arrows.
bNumber of intergenic spaces of each type in the genome of P. putida.
cREP frequencies: number of REP sequences in each type of intergenic space.
dExpected frequencies of REP sequences in each type of intergenic space.
eRatio between REP frequencies and expected frequencies.

Table 2. Statistical analysis of the organisation of the E. coli BIME sequences that appear in intergenic spaces

<table>
<thead>
<tr>
<th>Intergenic space types</th>
<th>NSb</th>
<th>Bfb</th>
<th>EFc</th>
<th>Bf/EFc</th>
</tr>
</thead>
<tbody>
<tr>
<td>→←← convergent</td>
<td>1113</td>
<td>48</td>
<td>74.91</td>
<td>0.64</td>
</tr>
<tr>
<td>←←← convergent</td>
<td>1195</td>
<td>56</td>
<td>80.43</td>
<td>0.69</td>
</tr>
<tr>
<td>←→← divergent</td>
<td>610</td>
<td>126</td>
<td>41.05</td>
<td>3.06</td>
</tr>
<tr>
<td>←←→ divergent</td>
<td>528</td>
<td>2</td>
<td>35.54</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td>3447</td>
<td>232</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aTypes of intergenic spaces depending on gene orientation, which is schematised by arrows.
bNumber of intergenic spaces of each type in the genome of E. coli.
cBIME frequencies: number of BIME sequences in each type of intergenic space.
dExpected frequencies of BIME sequences in each type of intergenic space.
eRatio between BIME frequencies in each type of intergenic space.

---

Figure 4. Distance of the intergenic REP sequences to the beginning and to the end of a gene. Each intergenic space is limited by two genes. The orientation of these genes determines a START codon or a STOP codon as the limit of each extreme of the intergenic space. For each intergenic REP sequence we analysed its distance to the limits of the intergenic space to generate two sets of data: (i) distances of REP sequences to STOP codons limiting its intergenic space (829 distances); and (ii) distances of REP sequences to START codons limiting its intergenic space (424 distances). The frequency distribution of both data sets is presented.
REP sequences are not specific transcription terminators

Because of the partial palindromic structure and the location of these sequences at the end of ORFs, we decided to test whether they played a role in transcription termination. We cloned the 35 bp element and all possible combinations found downstream of rpoH in different orientations between the $\lambda$ pR promoter and $\lambda$ lacZ, in the plasmids detailed in Materials and Methods, to determine whether they exhibited transcriptional terminal activity. We found that while the true rho-independent terminator of rpoH lowered the activity to ~5% (pMPRH), the presence of the 35 bp sequence in either orientation downstream had little effect on the $\beta$-galactosidase activity, and the two sequences in inverted orientation only reduced 45% of the $\beta$-galactosidase with respect to the construction without the insert (Fig. 1C). To test whether this low termination activity was sequence specific (i.e. exclusive of complementary REP pairs) or due to a putative secondary structure formed between two inverted sequences of this length, we constructed pMPRA2 and pMPRA3, similar to pA12 and pA21, but with two identical random sequences in direct or inverted orientation, instead of the REP sequences. Figure 1C shows that two random sequences in inverted orientation produce the same effect as two REP sequences. These results suggest that the 35 bp element by itself is probably not a terminator, and that the ability of the REP sequence pairs to reduce expression of the downstream gene does not depend on their sequence, but rather on their inverted orientation. In any case, the effect produced is very low compared with the termination activity of a true rho-independent terminator.

REP sequences specifically identify P.putida

Escherichia coli REP sequences have been extensively used in taxonomic studies to determine the diversity of bacterial populations (29). Repetitive element sequence-based PCR (rep-PCR) enables the generation of DNA fingerprint patterns to discriminate bacterial species and strains. The primers more frequently used for rep-PCR-based fingerprinting analysis are REP, ERIC and BOX sequences (30). We decided to test whether the REP elements of P.putida KT2440 could be used in a more specific manner to specifically detect and genotype P.putida. We selected several strains belonging to eight species of the genus Pseudomonas and seven strains of P.putida available in our laboratory collection. We isolated chromosomal DNA from each strain and performed PCR using an oligonucleotide with the P.putida REP consensus sequence. Figure 5 shows that only P.putida strains gave PCR products under these conditions, whereas none of the non-P.putida strains gave any signal, even when using low stringency annealing conditions.

These results suggest that the REP sequence of P.putida KT2440 identified here allows the identification of P.putida strains. The band pattern obtained with all P.putida strains revealed common bands, and several strain-specific bands. This probably reflects genome reorganisations, and allows a very specific genotyping of strains. A genotyping method based on species-specific REP sequences would only require the detection of the presence or absence of PCR products, rather than the more complex analysis of band patterns. Such a method would be suitable for the automation of genotyping procedures.

DISCUSSION

The analysis presented in this work shows the presence of highly repetitive extragenic palindromic sequences in the chromosome of P.putida KT2440. The structure of the REP sequence detected in P.putida is very similar to that described in E.coli and several enterobacteriaceae such as Salmonella sp. and Klebsiella sp., although the DNA sequence is species specific and is dissimilar. In fact, PCR chromosome amplification with an oligonucleotide complementary to the REP element has revealed that this particular sequence can only be detected in strains of the P.putida species. The development of methods and conditions for the optimal use of this sequence in the detection, identification and typing of bacterial strains will make it a valuable tool for taxonomic and population analysis, both in the laboratory and in natural environments. In addition, the use of a single primer in the PCR analysis notably increases the specificity of the reaction, which requires the close presence of two similar sequences in opposite orientation.

REP sequences of different enterobacteriaceae described so far share an internal structure: a 35 base sequence with a central palindromic motif and characteristic positions defining a head and a tail. In P.putida REP sequences, the most conserved region is 29 bp long and contains an internal 14 bp dyad symmetry whose presence tends to be maintained through compensatory changes. This internal palindrome is shorter than the imperfect dyad symmetry of E.coli REP sequences, which is 24 bp long.
In our search, 80% of the REP sequences were found outside ORFs, therefore they did not interfere with gene sequences. This proportion could be even higher since, for the sake of simplicity in our analysis, we have considered as intragenic those REP sequences that overlapped with gene ORFs by only 1 bp. When we analysed the presence of REP sequences close to homologous genes in E.coli and P.putida, we obtained no significant similarity in the position of REP sequences. This suggests that the function of REP sequences is not related to specific gene functions.

In E.coli, only 83 out of 500 REP sequences are present as single units, while the rest are part of more complex mosaic elements, called BIMEs (31). The situation in P.putida seems to be different: only eight BIMEs were found and most of the REP sequences were found as single elements or forming pairs. Although lower than in E.coli, the presence of REP sequence pairs in P.putida is very significant, and strikingly, in these pairs the two REP elements were always found in opposite orientation. This suggests a specific functional role for the pairs (32). The disposition as inverted repeats opens the possibility of stem–loop structure formation although they do not seem to be involved in the transcription termination as determined experimentally in this study. Recently, a role in transcription attenuation has been suggested for E.coli BIMEs (33). Espéli et al. (33) found a similar level (~50%) of decrease in expression of genes located downstream of a pair of inverted REP sequences. Unfortunately, they did not test the putative effect of a random DNA sequence with a similar structure.

Our finding of REP sequences in the non-enterobacteriaceae P.putida supports the hypothesis that these sequences are a more general phenomenon in bacteria. Therefore, it is tempting to speculate that this could mean that they are involved in an important bacterial function(s), though not yet identified. The presence of these sequences in enterobacteriaceae has been related to several functions, such as stabilisation of mRNA (34–36), organisation of the chromosome, insertion of genetic elements, binding site for proteins such as IHF (37), DNA polymerase I (38) and DNA gyrase (39). Thorough studies carried out with this last protein have shown that it is able to bind and cleave REP sequences (40). In addition, HU stimulates high affinity binding of gyrase, but inhibits cleavage of the sequence (41). Yang and Ferro-Luzzi Ames (41) suggested that REP sequences could be the target for gyrase action on the chromosome to maintain the appropriate level of negative supercoiling, and to anchor the chromosome supercoiled domain loops to each other. However, this hypothesis cannot explain why most REP sequences are located at the 3′-terminus of the genes. Changes in DNA supercoiling mediated by DNA gyrase could play a role in the induction of expression of some genes and could provide a general way of increasing expression of genes in the cell in response to environmental stress (42). Initiation of transcription for many genes is sensitive to DNA supercoiling. On the other hand, the situation is especially interesting in the intergenic spaces between convergent genes, where most of the P.putida REP pairs are located. In those cases, the distance of the REP pair to the end of the flanking genes is usually very short (<30 bp), thus suggesting that their function is probably related to those genes. It is known that two neighbouring convergent genes simultaneously transcribed generate positive supercoiling in the DNA ahead of the two RNA polymerases moving towards each other, i.e. the intergenic region (43,44). DNA gyrase, which has been shown to bind REP sequences in E.coli, is known to relax positively supercoiled domains (44). Therefore, it is tempting to suggest that one main role of REP sequences would actually be to allow DNA gyrase to bind and relax DNA when excessive positive supercoiling is generated, especially between two convergent genes. This would explain our finding that most REP pairs are located between convergent genes in P.putida, close to their STOP codon, and why this is also the case in E.coli, as shown in Table 2.

Our results show that the REP sequences appear at a very low frequency between divergent genes. The DNA region between the beginning of two genes is a functionally compromised region, because the promoters of at least two genes are allocated in these regions, where the transcriptional machinery needs to bind. If REP sequences were the target of cellular protein, i.e. DNA gyrase, etc., their presence in this region would interfere with the expression of divergent genes. Also if REP sequences were targets for insertion of genetic elements or for DNA rearrangement, as suggested (45,46), they would be expected to be deleterious if found in promoter regions.

In summary, the finding of REP sequences in non-enterobacteriaceae suggests a more general and crucial role for these sequences. Their significant presence as extragenic, in pairs as inverted repeats, at the end of genes and between convergent genes can give clues towards revealing their function. Finally, the expected broad presence of REP sequences in bacteria together with their species specificity opens new ways for easy and precise bacterial genotyping.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Ana Hurtado for the preparation of chromosomal DNAs and T. S. Larsen and I. Cases for preliminary annotation and suggestions on the handling of the Pseudomonas genome sequence available at www.tigr.org. This study was supported by a grant from the European Commission (BIO 2000-3126-CE).

REFERENCES


