JOURNAL OF BACTERIOLOGY, Nov. 2006, p. 7876–7884 Vol. 188, No. 22
0021-9193/06/$08.00+0 doi:10.1128/JB.00942-06

Structural Organization and Functional Properties of Miniature DNA Insertion Sequences in Yersiniae[▽]

Eliana De Gregorio, Giustina Silvestro, Rossella Venditti, Maria Stella Carlomagno, and Pier Paolo Di Nocera*

Dipartimento di Biologia e Patologia Cellulare e Molecolare, Facoltà di Medicina, Università Federico II, Via S. Pansini 5, 80131 Napoli, Italy

Received 29 June 2006/Accepted 14 August 2006

YPALs (*Yersinia pal*indromic sequences) are miniature DNA insertions scattered along the chromosomes of yersiniae. The spread of these intergenic repeats likely occurred via transposition, as suggested by the presence of target site duplications at their termini and the identification of syntenic chromosomal regions which differ in the presence/absence of YPAL DNA among *Yersinia* strains. YPALs tend to be inserted closely downstream from the stop codon of flanking genes, and many YPAL targets overlap rho-independent transcriptional terminator-like sequences. This peculiar pattern of insertion supports the hypothesis that most of these repeats are cotranscribed with upstream sequences into mRNAs. YPAL RNAs fold into stable hairpins which may modulate mRNA decay. Accordingly, we found that YPAL-positive transcripts accumulate in *Yersinia enterocolitica* cells at significantly higher levels than homologous transcripts lacking YPAL sequences in their 3′ untranslated region.

Bacterial insertion sequences (ISs) are mobile genetic elements ranging in size from 800 to 2,500 bp which are widely distributed among bacteria (21, 5). Typically, ISs encode a transposase which mediates their movement and feature terminal inverted repeats (TIRs) 10 to 50 bp long, which serve as recognition sites for the transposase. Most ISs generate short direct repeats (target site duplications [TSDs]) at the point of insertion. For each element, the length of the TSD is fixed and ranges from 2 to 13 bp. Differences in the structural organization, coding capacity, and transposition properties make it possible to sort ISs into approximately 20 major subfamilies (21).

In recent years, it has emerged that small IS-like sequences called MITEs (for *mi*niature *t*ransposable *e*lements) may be a relevant genome component in several eukaryotic species (16, 20). These elements characteristically measure 150 to 400 bp, carry long TIRs, and are flanked by TSDs of variable lengths. MITEs likely represent deletion derivatives of longer, autonomous ISs, which have been mobilized by tranposases encoded by partly related mobile donor elements (20). Several MITE families have also been identified in archaeabacteria (5). Three families of MITEs have been described for eubacteria. RUP (*r*epeat *u*nit of *p*neumococcus) elements are spread in ~100 copies in the genome of the *Streptococcus pneumoniae* 4 strain, and it has been proposed that their mobilization is mediated by the IS*630*-Spn1 element (24). NEMIS (*Ne*isseria *m*iniature *i*nsertion *s*equences) are 108- to 158-bp-long repeats which make up ~2% of the *Neisseria meningitidis* genome. In contrast to RUP elements, which are mostly interspersed with other repeated DNA sequences, NEMIS sequences are frequently located next to *Neisseria* genes and are transcribed into mRNAs. Hairpins formed by the pairing of NEMIS TIRs are targeted by RNase III, and this interaction regulates sets of *N. meningitidis* genes at the posttranscriptional level (12, 13).

A different type of eubacterial MITE is represented by ERIC (*e*nterobacterial *r*epetitive *i*ntergenic *c*onsensus) sequences. These elements, which measure 69 to 127 bp, are moderately abundant (20 to 25 copies) in the genomes of several enterobacteria (19) but are overrepresented in the genomes of yersiniae (15). In *Yersinia* species, most ERIC sequences are inserted immediately downstream from open reading frames (ORFs) and hence are cotranscribed into mRNAs with upstream genes. ERIC RNA may fold into robust RNA hairpins, and changes in their relative position and orientation within the mRNA molecule have been shown to differently influence the processing rate of neighboring RNA segments (15).

In this paper we report on the genomic and functional organization of a novel family of repeated DNA sequences present in yersiniae. The YPAL (*Yersinia pal*indromic elements) sequences share properties with both ERIC and NEMIS sequences and represent a novel example of MITEs which can play a role as RNA elements in posttranscriptional control.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Yersinia enterocolitica* strains used in this work and growth conditions have been described previously (15). The *Escherichia coli* strains W3110 and HT115 (W3110 *rnc-14*::ΔTn*10*) are described in the work of Takiff et al. (28). The strains SK5006 (*thr leu* pDK39 Cm^r *rnb-500*), SK5003 (*thr leu pnp-7 rnb-500*), and SK5695 (*thr leu rne-1*) are described in reference 2.

RNA analyses. Total bacterial RNA was purified on RNeasy columns (QIAGEN). YPAL-positive transcripts were monitored by Northern analyses by using as probes radiolabeled DNA segments 300 to 400 bp in length resulting from the amplification of *Y. enterocolitica* strain Ye161 DNA with pairs of gene-specific oligonucleotides. Processed YPAL RNA species were detected by high-resolution Northern analyses as described previously (14). Reverse transcription (RT)-PCR analyses were carried out by reverse transcribing 200 nanograms of total *Y.*

* Corresponding author. Mailing address: Dipartimento di Biologia e Patologia Cellulare e Molecolare, Facoltà di Medicina, Università Federico II, Via S. Pansini 5, 80131 Naples, Italy. Phone: 0039-81-7462059. E-mail: dinocera@unina.it.

▽ Published ahead of print on 8 September 2006.

TABLE 1. Primers used in this work

Primer function	Name	Sequence (5′ to 3′)
RT-PCR	rRNA for	GCGCTTAACGTGGGAACTGC
	rRNA rev	TCAGTCTTTGTCCAGGGGGC
	1047 for	TATGGTGGCGGTCTTCCTGG
	1047 rev	TTTCGGATCGTTAGGCACGC
	686-A for	ACTTATTCTGTCGGGCGCTG
	686-A rev	TAACCGTCTTACCTTCCGCC
Northern probe	408-A for	CATTTGCCCGTTTCCAGATC
	408-A rev	ACGCGGATCCGATCAAATAC
	686-B for	TGTGGTGACTGGTGAGATGG
	686-B rev	CGGTTTTCTTGCTCTGCTAC
	ypal for	TGAGGTTAATGACAAAGTGCCCGTA
	ypal rev	TCAGACTGCTGACAAACCTCAAGGA
	Ypal45 for	TAATACGACTCACTATAGGGAGAACCCGCGAAATGCGGGTTGAGG
	Ypal45 rev	ATTTAGGTGACACTATAGAATACACCCGCATTTCGCGGGTTCAGA
YPAL antisense RNA template	408-B for	ATTTAGGTGACACTATAGAATACACCTTGTCATTAGATGGGGACCC
	408-B rev	TAATACGACTCACTATAGGGAGAAAAATGATAAGCCGCAACGCTAG
	686-C for	ATTTAGGTGACACTATAGAATACGCGAAGGCCGAGTTGTTGAAGAG
	686-C rev	TAATACGACTCACTATAGGGAGATTAGAGGAGGGCTATCCGGTGGG
YPAL sense RNA template	408-C for	TAATACGACTCACTATAGGGAGACAGTGCCTGGATATACCTTTACC
	408-C rev	ATTTAGGTGACACTATAGAATACCACAGCAAACTGACCAAAGAATA
	408-D for	TAATACGACTCACTATAGGGAGACAGTGCCTGGATATACCTTTACC
	408-D rev	ATTTAGGTGACACTATAGAATACGCCGAAATCCTGTCCTCTTTCGA

enterocolitica RNA by random priming. The resulting cDNAs were amplified by using pairs of gene-specific oligonucleotides. One oligonucleotide within each pair had been previously ^{32}P end labeled at the 5′ terminus with polynucleotide kinase. To adequately monitor gene-specific RNA levels by RT-PCR, cDNAs were amplified under nonsaturating cycling conditions, and low-cycle (14 to 18 cycles) PCR analyses were performed for each set of amplified genes. As an internal control, a 140-nucleotide (nt)-long amplified segment of 16S rRNA from *Y. enterocolitica* was used. rRNA primers were added to the PCRs two to four cycles before the amplification of the ORF-specific mRNAs was completed. Amplified DNA fragments were separated using 6% polyacrylamide-8 M urea gels and quantitated by phosphorimagery.

The uniformly ^{32}P-labeled RNAs used as probes in the RNase protection assays were obtained by transcribing in vitro linear DNA templates as described previously (12). DNA templates were obtained by PCR amplification of DNA derived from the *Y. enterocolitica* strain Ye161. One of the two primers included in its 5′ end region the sequence of the T7 RNA polymerase promoter (see underlined residues in Table 1). Twenty micrograms of total RNA derived from *Y. enterocolitica* cells was mixed with ^{32}P-labeled antisense RNA probes in 30 μl of hybridization buffer (75% formamide, 20 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl, 0.1% sodium dodecyl sulfate). Samples were incubated at 95°C for 5 min, cooled down slowly, and kept at 45°C for 16 h. After a 60-min incubation at 33°C with RNase T1 (2 μg/ml), samples were treated with proteinase K (50 μg/ml) for 15 min at 37°C, extracted once with phenol, precipitated with ethanol, resuspended in 80% formamide, and loaded onto 6% polyacrylamide-8 M urea gels.

To obtain the YPAL-positive RNA control electrophoresed in lane 3 of Fig. 6B, the YPAL 45 element and flanking sequences were amplified by using a primer including the T7 RNA polymerase promoter.

In vitro RNA cleavage assays. Uniformly ^{32}P-labeled RNAs were obtained by transcribing in vitro linear DNA templates with T7 RNA polymerase in the presence of radiolabeled [^{32}P]UTP. Templates were obtained by PCR amplification of *Yersinia* DNA with pairs of 50-mers, one of which included the T7 RNA polymerase promoter in its 5′ end region. The T7 promoter sequence allowed the in vitro synthesis of RNA substrates for processing assays. Degradation assays with whole bacterial cell extracts were carried out essentially as described previously (11).

Oligonucleotides. Sequences of all PCR primers used in this study are reported in Table 1.

Computer analyses. A YPAL element found next to ERIC sequences in the sequenced 8081 strain of *Y. enterocolitica* (www.sanger.ac.uk/Projects/Y_*enterocolitica*) was used as a query in BLAST searches to fetch homologous DNA segments from the genomes of the same strain and the *Yersinia pestis* CO92 strain (26).

Consensus sequences from multiple alignments of YPAL family members were established with the program CONS of the EMBOSS package. Secondary structure modeling was done using the Mulfold program (www.bioinfo.rpi.edu/applications/mfold), which predicts RNA secondary structure by free energy minimization (30).

RESULTS

Organization of YPAL repeats. During analyses aimed at the characterization of the family of ERIC repeats in *Y. enterocolitica*, we fortuitously encountered a palindromic intergenic sequence located next to an ERIC element. As revealed by BLAST searches, the latter was found to be a member of an abundant DNA family spread throughout the chromosomes of both *Y. enterocolitica* and *Y. pestis*. A member of this family had been earlier called YPAL (3), and we will keep the acronym for clarity. The palindromic nature of YPAL repeats is highlighted in Fig. 1. As RNA sequences, i.e., allowing for the formation of GU base pairs, YPALs can fold, regardless of their orientation, into stable, low-free-energy branched hairpins (Fig. 1). YPALs are present in ~100 copies in sequenced *Yersinia* genomes (Fig. 2) and feature a modular organization. Complete repeats include the same external modules (A and H modules in Fig. 2) but a different set of internal modules. Eight different YPAL subfamilies could be distinguished by sequence alignments. The highest numbers of repeats belong to subfamilies 2 and 8. The size of subfamily 2, which includes elements measuring 167 bp, is similar in *Y. enterocolitica* and *Y. pestis*. In contrast, the size of subfamily 8, made by 130-bp-long repeats, is much reduced in *Y. pestis*. The 98-bp-long module I, defining the members of the subfamily 8, comes in two sequence variants (Ia and Ib), neither of which bears significant

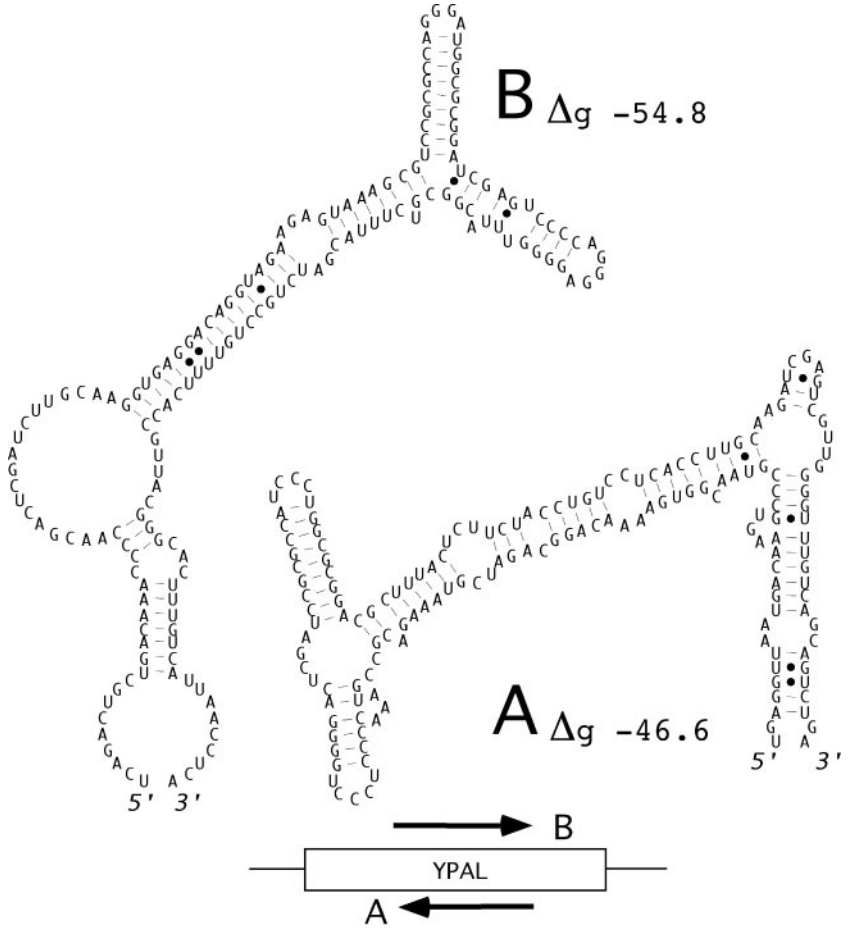

FIG. 1. Secondary structure of YPAL repeats. Hairpin structures formed by a representative YPAL element inserted in the mRNA in the two possible A and B orientations are shown. GU pairing is highlighted by dots. The Gibbs free energy of each hairpin is indicated. Secondary structure modeling was done using the Mulfold program. The element analyzed is the 167-bp-long YPAL 45 element (see the text).

sequence homology to the bodies of other YPALs (Fig. 2). Interestingly, modules Ia and Ib were found in 2/3 and 1/3 of the elements identified in the *Y. enterocolitica* 8081 strain, respectively. In contrast, members of YPAL subfamily 8 identified in the *Y. pestis* CO92 strain all feature module Ia.

In the chromosome of *Y. enterocolitica* strain 8081, six members of subfamily 2 are interrupted at the same position between modules E and F (Fig. 2) by the insertion of 1,376 bp of foreign DNA. Homology searches run at the IS finder site (http://www-is.biotoul.fr) enabled us to establish that the intervening DNA sequence is ISYen*1*, a low-copy-number IS found exclusively in the *Y. enterocolitica* species (27). The two additional members of the ISYen*1* family resident in the chromosome of *Y. enterocolitica* strain 8081 are not associated with YPAL DNA. The insertion of ISYen*1* is not accompanied by the duplication of target site sequences.

YPALs are miniature mobile DNA elements. YPAL repeats show the typical structure of miniature DNA insertion sequences. The hypothesis that their genomic spread occurred via transposition was supported by the presence of TSDs at the termini of most elements (Fig. 3). In contrast to the majority of ISs, whose insertion is accompanied by the generation of TSDs of fixed length, most (85%) YPALs are flanked by TSDs of variable size (Fig. 3). The length of the duplicated segment ranged from 3 to 26 bp, with the most frequently found TSDs measuring 18 or 20 bp (Fig. 3B). YPAL targets do not exhibit sequence homologies. However, it is remarkable that they span, or are located next to, regions of dyad symmetry. Intriguingly, YPAL target sites often correspond to the DNA counterpart of rho-independent transcription terminators, the RNA structures responsible for the detachment of the transcribing RNA polymerase from the DNA template. This conclusion was supported both by the presence of runs of thymidines immediately next to palindromes targeted by YPAL (Fig. 3A) and by the finding that most targets were found located 20 to 50 bp downstream from the stop codon of annotated ORFs in sequenced *Yersinia* genomes.

The notion that YPALs are mobile DNA elements was directly supported by in silico analyses. The chromosomal distribution of YPALs varies significantly between *Y. pestis* and *Y. enterocolitica*, and only a few syntenic YPAL-positive regions are shared by the two species. This is not surprising, since *Y. pestis* and *Y. enterocolitica* are evolutionarily distant. In contrast, *Y. pestis* and *Yersinia pseudotuberculosis* are evolutionarily closer, *Y. pestis* being regarded as a clone that evolved from *Y. pseudotuberculosis* only 1,500 to 20,000 years ago (1). Com-

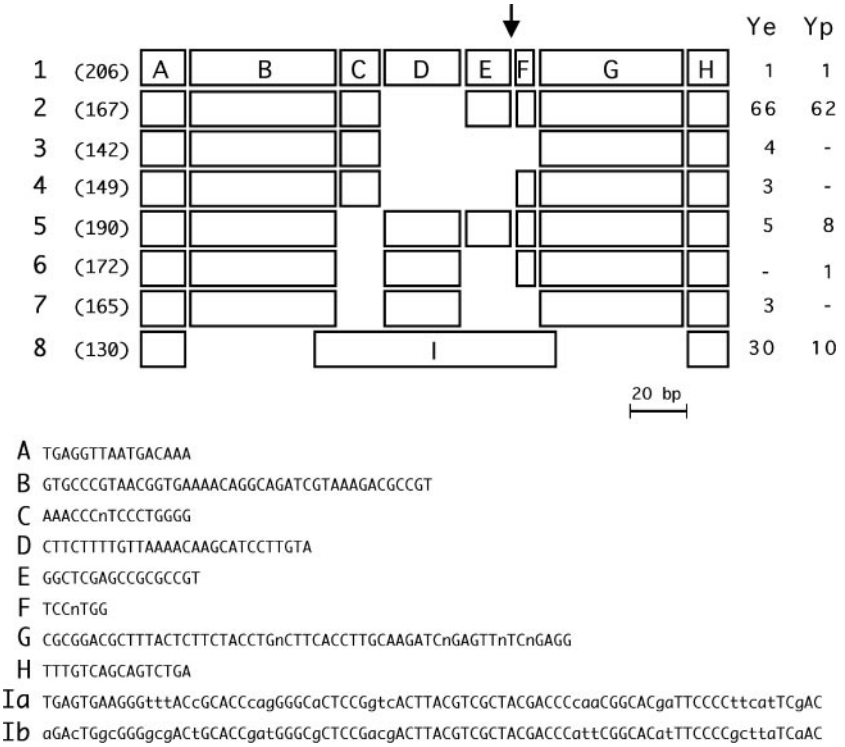

FIG. 2. Structural organization of YPAL elements. The nine modules labeled A to I found in YPAL elements are shown. Numbers to the left (1 to 8) denote YPAL subfamilies. The length in bp of the elements of each subfamily is given in parenthesis. The number of elements within each subfamily found in the *Y. enterocolitica* 8081 (Ye) and the *Y. pestis* CO92 (Yp) strains is reported. The consensus sequence content of the nine modules derived from the comparison of *Y. enterocolitica* YPAL repeats is shown at the bottom. Two versions of the module I (Ia and Ib) are reported. Uppercase residues denote sequence identity. An arrow marks the site of insertion of ISYen*1*.

parisons of the genomes of the *Y. pestis* CO92 and *Y. pseudotuberculosis* IP32953 strains led to the identification of homologous chromosomal regions which carry YPAL DNA only in one of the two species (Fig. 4). The finding of a single copy of the TSD at the empty chromosomal sites (Fig. 4) corroborated the notion that target site duplication is induced upon insertion, ruling out the formal possibility that YPALs could preferentially integrate between preexisting tandem duplications. Empty and filled YPAL chromosomal sites also were subsequently identified among different laboratory strains of *Y. enterocolitica* by DNA sequence analyses (data not shown).

Changes in the distribution of YPAL sequences among *Yersinia* strains might also reflect the loss of YPAL DNA from specific sites, prompted either by specific recombination between the flanking TSDs or by replicational slippage between the TSDs.

YPAL elements transcribed into RNA. All the YPAL sequences found in the *Y. enterocolitica* chromosome are located within intergenic regions. Remarkably, most elements (70 repeats) are inserted close to the stop codons of flanking ORFs. This suggests that YPALs may be cotranscribed along with upstream coding sequences and that their ability to fold into RNA hairpins may have functional significance.

To address this issue, total RNA from *Y. enterocolitica* strain 161 was analyzed by Northern blotting. We monitored two members of YPAL subfamily 2 (repeats 45 and 9) located 89 bp and 56 bp downstream from the stop codon of the IMP dehydrogenase/GMP reductase gene encoded by ORF YE686 and the pyrophosphatase gene encoded by ORF YE408, respectively (Fig. 5A). The ORF YE686 probe detected a major mRNA species measuring ~1,380 nt, which plausibly spans both ORF YE686 and the flanking YPAL 45 element. Two transcripts corresponding to ORF YE408, measuring ~650 and ~850 nt, were detected by the ORF YE408 probe. The sizes of the bands are in accord with the hypothesis that the 850-nt-long transcript spans both ORF YE408 and the flanking YPAL 9 element and that the 650-nt-long RNA species may originate with the removal of YPAL sequences from the transcript (Fig. 5A).

Northern data were complemented by RNase protection experiments (Fig. 5B). When YE686 transcripts were monitored, a predominant RNA species, corresponding to the accumulation in *Y. enterocolitica* cells of transcripts encompassing both ORF YE686 and the flanking YPAL 45 element, was protected (Fig. 5B, lane 3). In contrast, when YE408 transcripts were monitored, two bands of protection were detected (Fig. 5B, lane 6). The size of the protected species is in accord with the hypothesis that two predominant mRNA segments spanning ORF YE408 accumulate within *Y. enterocolitica* cells, one of which encompasses the flanking YPAL 9 element. From the relative intensities of the bands, it could be inferred that the latter RNA species is two to three times more abundant than the 400-nt-long RNA species lacking YPAL sequences.

Processing of YPAL-positive transcripts. Both the Northern and RNase protection experiments suggest that YPAL sequences may be cleaved off from YPAL-positive transcripts in

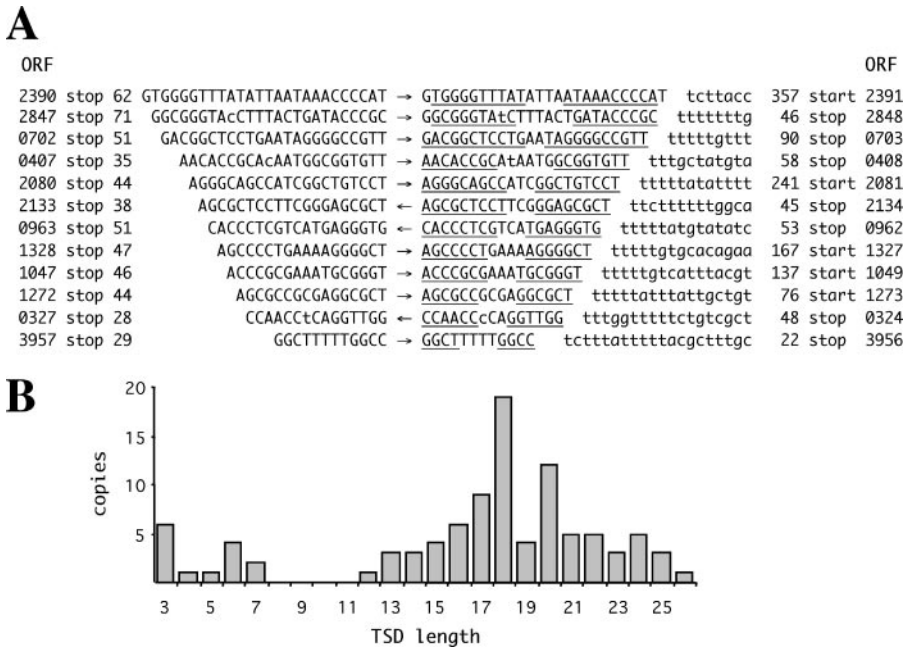

FIG. 3. YPAL target sites. (A) Some of the TSDs induced by the insertion of YPALs in the chromosome of the *Y. enterocolitica* 8081 strain, as well as flanking T-rich segments, are shown. Base changes at TSDs are in lowercase letters. Arrows mark YPAL sequences and their orientations. Regions of dyad symmetry are underlined. ORFs flanking YPAL sequences are indicated. The distance in bp separating YPAL termini from either the start or the stop codon of each ORF is shown. (B) Length variation of TSDs flanking YPAL elements in the *Y. enterocolitica* 8081 strain.

vivo. This hypothesis was confirmed by in vitro RNA degradation assays. A DNA fragment spanning the 3′ end region of ORF YE408 and the flanking YPAL 9, the 167-nt-long element belonging to the abundant subfamily 2 already analyzed in Fig. 5, was used as a template to synthesize in vitro-radiolabeled YPAL-positive RNAs of known length. By using the Mulfold program, we checked that foreign sequences present in the RNA synthesized in vitro did not interfere with the formation of the YPAL RNA hairpins shown in Fig. 1. Challenging the RNA obtained (408-A RNA) with *Y. enterocolitica* whole-cell extracts resulted in the accumulation of three major RNA species (Fig. 6A, lane 2). Bands a to c had the size expected for RNA moieties generated by cleavages occurring at the boundaries of YPAL 9 sequences. Results were corroborated by experiments carried out with a substrate similar to the 408-A RNA but lacking 161 nt at the 3′ end (408-B RNA) (Fig. 6A, lane 4). By using this shortened RNA substrate, we were able to detect the same a and b bands derived from the processing of the 408-A RNA. The faintness of the RNA species labeled d, easily detectable only upon prolonged exposure of the autoradiogram, plausibly reflects its intrinsic instability as an RNA segment. The same cleavage pattern was obtained by challenging the 408-B RNA probe with whole-cell extracts derived from the wild-type *E. coli* W3110 strain (Fig. 6A, lane 5). The 408-B RNA was therefore challenged with whole cellular extracts derived from *E. coli* strains harboring mutant alleles for different ribonucleases.

The *E. coli* W3110 (Fig. 6A, lane 5) and RNase III-negative HT115 (lane 6) strains are isogenic. The strain SK5006 (lane 7), which carries a temperature-sensitive allele of the *rnb* gene, encoding RNase II, was grown at 32°C in order to serve as a control for the isogenic SK5003 strain (lanes 8 and 9), which in addition carries the inactive *pnp*-7 allele of the polynucleotide phosphorylase gene, and the SK5695 strain (lanes 10 and 11),

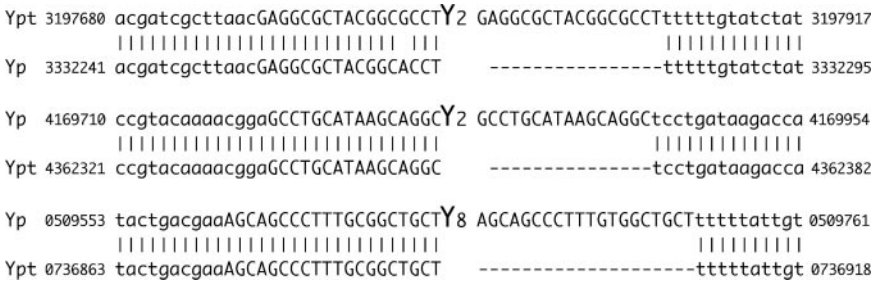

FIG. 4. Filled and empty YPAL sites. Homologous DNA regions from the *Y. pseudotuberculosis* IP32953 (Ypt) (9) and *Y. pestis* CO92 (Yp) (14) strains are aligned. Numbers refer to genome residues. Capital letters denote TSDs induced by the insertion of YPAL DNA. Y2 and Y8 are members of YPAL subfamilies 2 and 8, respectively.

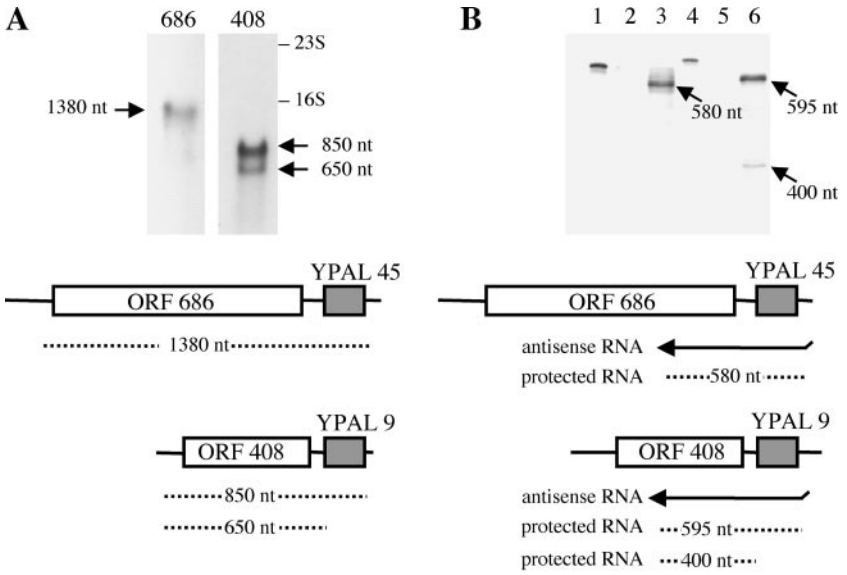

FIG. 5. Analysis of YPAL-positive transcriptional units. (A) Ten micrograms of total RNA from Ye161 cells was separated on a 1% agarose gel, blotted onto nitrocellulose, and probed with amplimers spanning the IMP dehydrogenase/GMP reductase (ORF YE686) and the pyrophosphatase (ORF YE408) genes, respectively. The positions of 16S and 23S rRNA are indicated. Major hybridization bands are indicated by arrows. The mRNA species detected and the corresponding chromosomal regions are drawn at the bottom. (B) RNase protection of YPAL-positive transcripts. In the diagram shown, the RNA probes and protected RNA species are drawn as thick and dotted lines, respectively. T1 RNase-resistant RNA hybrids were electrophoresed on 6% polyacrylamide-8 M urea gels. Protected RNA species are marked by arrows. Lanes: ORF 686 and ORF 408 RNA probes unreacted (lanes 1 and 4) or hybridized to 50 μg of tRNA from *E. coli* (lanes 2 and 5) or to 20 μg of total RNA from the *Y. enterocolitica* Ye161 strain (lanes 3 and 6).

which carries the *rne-1* temperature-sensitive allele of the RNase E gene. Significant alteration of the cleavage pattern was observed only with extracts derived from HT115 cells, suggesting that RNase III may play a role in the processing of YPAL-positive mRNAs.

Short YPAL-specific RNA moieties matching in size the b band in Fig. 6A were identified in vivo (Fig. 6B, lane 1) and may originate from cleavage of YPAL-positive transcripts. RNA species detected in vitro and in vivo measure ~200 nt, thus exceeding in length canonical YPAL repeats, such as YPAL 9, analyzed in Fig. 6A, which measure 167 bp. Highly structured RNA also can run abnormally in denaturing gels. To rule out technical artifacts, an RNA species of known length made by the 167 nt of the YPAL 45 element, the 34 nt of the two flanking TSDs, and an additional 30 nt was used as a control in Northern experiments (Fig. 6B, lanes 2 and 3).

Similarly, by using a 130-bp-long YPAL subfamily 8 repeat as a probe in Northern analyses, a ~160-nt RNA species was detected in vivo (data not shown). Discrepancies between the sizes of YPAL elements and the lengths of the hypothetical RNA cleavage products may suggest that YPAL sequences are not targeted per se and that additional base pairing provided by complementary sequences generated at the site of YPAL insertion is required for endonucleolytic cleavage by RNase III to occur. The hypothesis was reinforced by the finding that in vitro, the cleavage products shown in Fig. 6A were no longer detected when a shorter RNA substrate carrying the 167 residues of the YPAL 9 element but lacking flanking TSDs was used as a substrate (data not shown).

Stability of YPAL-positive RNAs in vivo. As shown by both Northern and RNase protection data, YPALs are transcribed into mRNAs. To investigate the role that these sequences may play as RNA elements in vivo, we first identified, by means of PCR analyses, *Yersinia* strains which lack YPAL sequences at specific loci (not shown). Subsequently, the levels of homologous mRNAs were measured in "filled" and "empty" strains by quantitative RT-PCR analyses. Total RNA from *Y. enterocolitica* cells was reverse transcribed into cDNA and amplified under nonsaturating cycling conditions to ensure that the yield of amplified products was proportional to the amount of cellular RNA targeted by PCR. *Y. enterocolitica* 16S rRNA sequences were used as an internal control. The 16S rRNA primers were added to the PCRs two to four cycles before the amplification of the ORF-specific mRNAs was completed (see Materials and Methods). In strain Ye161, ORFs YE686 and YE1047 are flanked 3′ by YPALs 45 and 41, respectively. In contrast, in strain Ye25, the corresponding intergenic sequences both lack YPAL DNA. The levels of transcripts corresponding to either ORF were found to be ~10-fold higher in strain Ye161 than in strain Ye25 (Fig. 7). As for the YE686 transcripts (Fig. 5B), RNase protection experiments indicated that most transcripts spanning the YE1047 ORF retained in their 3′ end sequences corresponding to YPAL 41 (data not shown).

These data support the hypothesis that the presence of YPAL sequences at the 3′ end may increase mRNA stability. The simplest interpretation is that YPAL RNAs may act as stabilizers by forming RNA hairpins able to counteract the progression of the 3′-5′ exonucleases present in the degradosome (8).

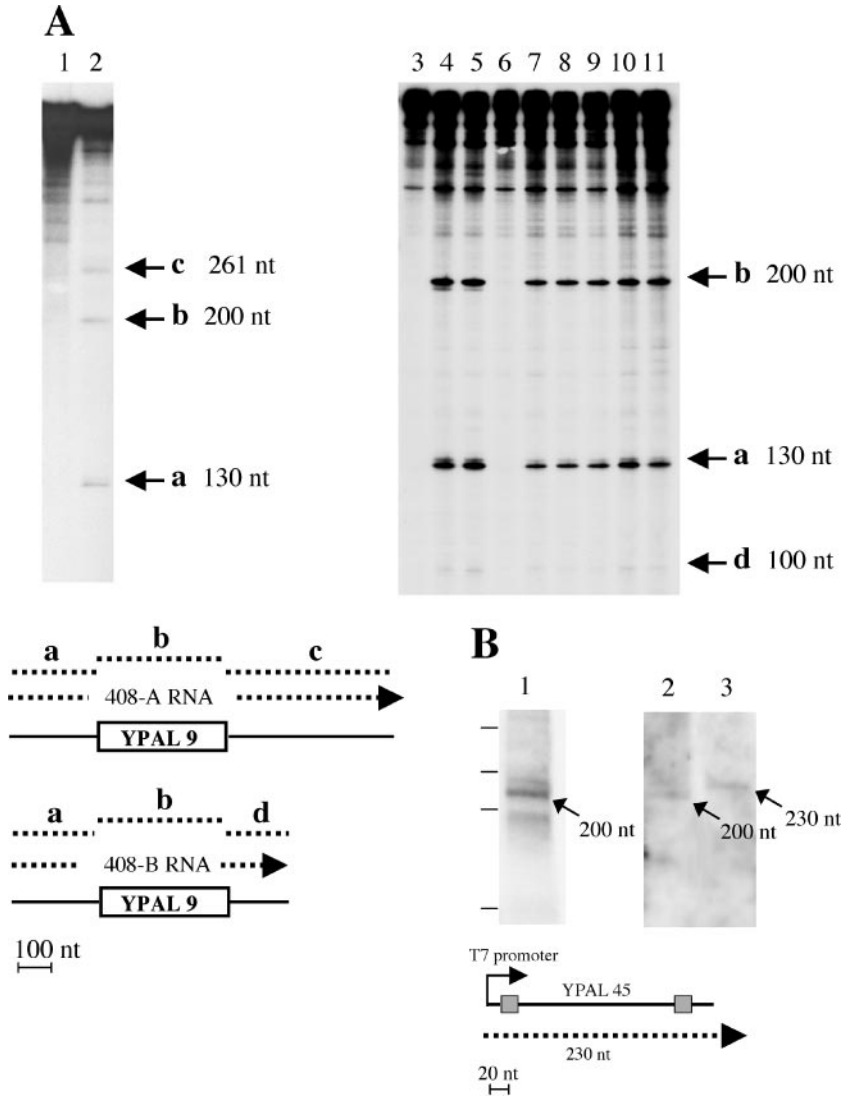

FIG. 6. (A) Processing of YPAL-positive RNA in vitro. The radiolabeled 408-A RNA spanning the element YPAL 9 was incubated for 5 min at 37°C either alone (lane 1) or with 0.5 μg of S100 cellular lysates from the *Y. enterocolitica* strain Ye161 (lane 2). The shorter 408-B RNA was incubated for 5 min at 37°C either alone (lane 3) or with 0.5 μg of S100 cellular lysates from strain Ye161 (lane 4) or *E. coli* strain W3110 (lane 5), HT115 (lane 6), SK5006 (lane 7), SK5003 (lanes 8, 9), or SK5695 (lanes 10, 11). All strains were grown at 32°C; extracts in lanes 9 and 11 were derived from SK5003 and SK5695 cells grown to early logarithmic phase at 32°C and shifted to 44°C for 45 min before harvesting to inactivate RNases II and E, encoded by the *rnb-500* and *rne-1* alleles, respectively. Reaction products were separated on 6% polyacrylamide-8 M urea gels. The RNA substrates and the corresponding processed RNA species a to d are sketched at the bottom. (B) Identification of processed YPAL RNA species in vivo. Total RNA (20 μg) from the Ye161 strain was separated on 6% polyacrylamide-8 M urea gels and transferred to nitrocellulose (lanes 1 and 2). A T7-driven transcript spanning YPAL 45 and flanking TSDs (filled boxes in the diagram at the bottom) and containing in addition 7 nt at the 5′ end and 23 nt at the 3′ end was electrophoresed in lane 3 as a control. The filters were hybridized to a DNA segment spanning the YPAL 9 element. The ~200-nt-long YPAL RNA identified in vivo, the size of the YPAL RNA synthesized in vitro, and the positions of the coelectrophoresed DNA molecular weight markers are shown.

DISCUSSION

We describe in this report the structural and functional characteristics of a relatively abundant set of sequences spread in *Yersinia* genomes. YPAL elements vary in size and sequence composition and can be sorted into several subfamilies. Subfamily 1 includes the largest repeat type, present in one copy in both *Y. enterocolitica* and *Y. pestis*. Members of subfamilies 2 to 7 differ from this prototype sequence in the lack of one or more body segments. Elements belonging to subfamily 8 share the terminal A and H repeats with other YPALs but feature a different body. Taking into account only complete elements, i.e., those including both A and H terminal modules, the YPAL family found in the sequenced 8081 strain of *Y. enterocolitica* is composed of >100 members. The family size is smaller in the sequenced CO92 strain of *Y. pestis*, and this correlates primarily with the reduced size of subfamily 8 in this species. The coexistence of subfamilies is typical of DNA repeats spread in a stepwise manner by bursts of transposition, and YPALs represent indeed a novel class of miniature ISs or MITEs. The mobile nature of YPAL sequences is directly

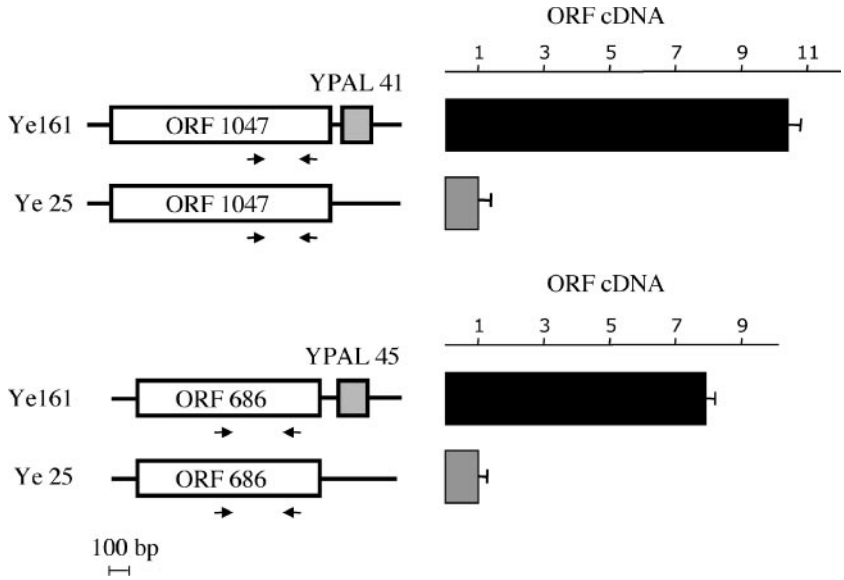

FIG. 7. RT-PCR analyses of YPAL-positive transcripts. Total RNA (200 nanograms) derived from *Y. enterocolitica* strains Ye161 and Ye25 was reverse transcribed, and the cDNA obtained was amplified by PCR with pairs of gene-specific oligonucleotides. Within each pair, one oligonucleotide was ^{32}P end labeled to allow the detection of the amplified segments by autoradiography. Reaction products were run on 6% polyacrylamide-8 M urea gels and quantitated by phosphorimaging. RNA levels were calculated relative to the amount of the coamplified 16S-specific rRNA fragment (see also Materials and Methods) and are expressed in arbitrary units. Black and gray bars denote the amounts of YE686- and YE1047-specific mRNA in the YPAL-positive strain Ye161 and the YPAL-negative strain Ye25, respectively. The ORF YE686 has been described in the legend to Fig. 5. The ORF YE1047 encodes a serine hydroxymethyltransferase. Arrows denote PCR primers.

supported by the finding of homologous chromosomal regions carrying or lacking YPAL DNA in different yersiniae. The identification of empty chromosomal sites in both *Y. pestis* and *Y. pseudotuberculosis* genomes allows us to conclude that transposition of YPAL still occurred after the speciation of *Y. pseudotuberculosis* into *Y. pestis*. As documented by the identification of empty and filled YPAL chromosomal sites among laboratory strains, the mobilization of YPAL sequences is still active in *Y. enterocolitica*.

All MITE families identified so far in eubacteria induce the duplication of the dinucleotide TA upon genomic insertion (15, 22, 24). In contrast, the TSDs which flank YPALs vary in both sequence and size (Fig. 3 and 4). Most YPAL targets overlap or coincide with palindromic sequences. This finding is not novel, since IS*30* (25), IS*1397* (29), IS*903* (18), and IS*621* (10) similarly tend to insert into regions of dyad symmetry. The notion that many YPAL targets may coincide with rho-independent transcriptional terminators also is not unprecedented, since *Y. pestis* IS*1541* (23) and *Mycoplasma fermentans* IS*1630* (6) have also been found inserted at rho-independent transcriptional terminators.

MITEs are mobilized by transposases encoded by ancestral ISs as well as by evolutionarily unrelated elements (5, 16, 20, 24). Distinct ISs may have played a role in the mobilization of YPALs. The conclusion stems primarily from the observation that target sequences were not always found duplicated at YPAL termini. The original targets may have been altered by mutation. However, two IS families are known to insert at transcriptional terminators. IS*1541* does not induce TSDs (23), whereas IS*1630* produces TSDs, ranging from 9 to 22 bp in length (6). It is possible that multiple endonucleases mobilize YPAL insertion and some produce TSDs while others do not.

YPALs are frequently located at the 3′ ends of *Yersinia* genes and often between genes transcribed in a convergent manner. Because some YPAL-positive transcripts accumulate in *Y. enterocolitica* at levels 8 to 10 times higher than those of homologous, YPAL-negative transcripts (Fig. 7), the element may impede the degradosome. In this respect, YPAL may function analogously to the intergenic REP sequences found in enterobacteriaceae (17). YPAL elements might also stimulate degradation of some mRNAs in which they are embedded. Because RNase III cleavage occurs near several YPAL hairpins (Fig. 6), stabilization or degradation of RNA may depend on context (4, 7, 12). Future work is needed to define the role of TSD sequences in RNA cleavage/stabilization.

YPALs are not restricted to *Yersinia*. A dozen 170-bp-long YPAL elements were found by BLAST analyses in the plant pathogen *Erwinia carotovora*. *Yersinia* and *Erwinia* elements are 70% homologous. YPALs may thus be an ancient component of the gamma-proteobacteria that were eliminated from most bacterial species.

Sequence repeats closely resembling YPALs have been identified by in silico surveys for several microorganisms (G. Silvestro and P. P. Di Nocera, unpublished data). Interestingly, many of these sequences may similarly function in posttranscriptional control as *cis*-acting elements according to the pattern of interspersion with coding segments.

ACKNOWLEDGMENTS

We are indebted to Ida Luzzi and Francesca Berlutti for providing us with *Yersinia* strains and Judah Lee Rosner for suggestions and critical reading of the manuscript.

This work was funded by a grant assigned to Pier Paolo Di Nocera by the PRIN 2004 agency of the Italian Ministry of University and Scientific Research.

REFERENCES

1. **Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel.** 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA **96:**14043–14048.
2. **Arraiano, C. M., S. D. Yancey, and S. R. Kushner.** 1988. Stabilization of discrete mRNA breakdown products in *ams pnp rnb* multiple mutants of *Escherichia coli* K-12. J. Bacteriol. **170:**4625–4633.
3. **Bachellier, S., J. M. Clement, and M. Hofnung.** 1999. Short palindromic repetitive DNA elements in enterobacteria, a survey. Res. Microbiol. **150:** 627–639.
4. **Blaszczyk, J., J. Gan, J. E. Tropea, D. L. Court, D. S. Waugh, and X. Ji.** 2004. Noncatalytic assembly of ribonuclease III with double-stranded RNA. Structure **12:**457–466.
5. **Brugger, K., P. Redder, Q. She, F. Confalonieri, Y. Zivanovic, and R. A. Garrett.** 2002. Mobile elements in archaeal genomes. FEMS Microbiol. Lett. **206:**131–141.
6. **Calcutt, M. J., J. L. Lavrrar, and K. S. Wise.** 1999. IS*1630* of *Mycoplasma fermentans*, a novel IS*30*-type insertion element that targets and duplicates inverted repeats of variable length and sequence during insertion. J. Bacteriol. **181:**7597–7607.
7. **Calin-Jageman, I., and A. W. Nicholson.** 2003. RNA structure-dependent uncoupling of substrate recognition and cleavage by Escherichia coli ribonuclease III. Nucleic Acids Res. **31:**2381–2392.
8. **Carpousis, A. J.** 2002. The Escherichia coli RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. Biochem. Soc. Trans. **30:**150–155.
9. **Chain, P. S., E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser, and E. Garcia.** 2004. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA **101:**13826–13831.
10. **Choi, S., S. Ohta, and E. Ohtsubo.** 2003. A novel IS element, IS*621*, of the IS*110*/IS*492* family transposes to a specific site in repetitive extragenic palindromic sequences in *Escherichia coli*. J. Bacteriol. **185:**4891–4900.
11. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2002. The abundant class of nemis repeats provides RNA substrates for ribonuclease III in Neisseriae. Biochim. Biophys. Acta **1576:**39–44.
12. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2003. Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs. Biochem. J. **374:**799–805.
13. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2003. Asymmetrical distribution of *Neisseria* miniature insertion sequence DNA repeats among pathogenic and nonpathogenic *Neisseria* strains. Infect. Immun. **71:**4217–4221.
14. **De Gregorio, E., L. Chiariotti, and P. P. Di Nocera.** 2001. The overlap of Inr and TATA elements sets the use of alternative transcriptional start sites in the mouse galectin-1 gene promoter. Gene **268:**215–223.
15. **De Gregorio, E., G. Silvestro, M. Petrillo, M. S. Carlomagno, and P. P. Di Nocera.** 2005. Enterobacterial repetitive intergenic consensus sequence repeats in yersiniae: genomic organization and functional properties. J. Bacteriol. **187:**7945–7954.
16. **Feschotte, C., N. Jiang, and S. R. Wessler.** 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. **3:**329–341.
17. **Higgins, C. F., R. S. McLaren, and S. F. Newbury.** 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression, evolution by gene conversion? A review. Gene **72:**3–14.
18. **Hu, W. Y., W. Thompson, C. E. Lawrence, and K. M. Derbyshire.** 2001. Anatomy of a preferred target site for the bacterial insertion sequence IS903. J. Mol. Biol. **306:**403–416.
19. **Hulton, C. S. J., C. F. Higgins, and P. M. Sharp.** 1991. ERIC sequences, a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. **5:**825–834.
20. **Jiang, N., C. Feschotte, X. Zhang, and S. R. Wessler.** 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements MITEs. Curr. Opin. Plant Biol. **7:**115–119.
21. **Mahillon, J., C. Leonard, and M. Chandler.** 1999. IS elements as constituents of bacterial genomes. Res. Microbiol. **150:**675–687.
22. **Mazzone, M., E. De Gregorio, A. Lavitola, C. Pagliarulo, P. Alifano, and P. P. Di Nocera.** 2001. Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic neisseriae. Gene **278:**211–222.
23. **Odaert, M., A. Devalckenaere, P. Trieu-Cuot, and M. Simonet.** 1998. Molecular characterization of IS*1541* insertions in the genome of *Yersinia pestis*. J. Bacteriol. **180:**78–181.
24. **Oggioni, M. R., and J. P. Claverys.** 1999. Repeated extragenic sequences in prokaryotic genomes, a proposal for the origin and dynamics of the RUP element in Streptococcus pneumoniae. Microbiology **145:**2647–2653.
25. **Olasz, F., J. Kiss, P. Konig, Z. Buzas, R. Stalder, and W. Arber.** 1998. Target specificity of insertion element IS30. Mol. Microbiol. **28:**691–704.
26. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Genome sequence of Yersinia pestis, the causative agent of plague. Nature **413:**523–527.
27. **Partridge, S. R., and R. M. Hall.** 2003. The IS*1111* family members IS*4321* and IS*5075* have subterminal inverted repeats and target the terminal inverted repeats of Tn*21* family transposons. J. Bacteriol. **185:**6371–6384.
28. **Takiff, H. E., S. M. Chen, and D. L. Court.** 1989. Genetic analysis of the *rnc* operon of *Escherichia coli*. J. Bacteriol. **171:**2581–2590.
29. **Wilde, C., S. Bachellier, M. Hofnung, and J.-M. Clement.** 2001. Transposition of IS*1397* in the family *Enterobacteriaceae* and first characterization of ISKpn*1*, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units. J Bacteriol. **183:**4395–4404.
30. **Zuker, M.** 1989. On finding all suboptimal foldings of an RNA molecule. Science **244:**48–52.