INFECTION AND IMMUNITY, Dec. 1998, p. 5731–5742 Vol. 66, No. 12
0019-9567/98/$04.00+0

Complete DNA Sequence and Detailed Analysis of the *Yersinia pestis* KIM5 Plasmid Encoding Murine Toxin and Capsular Antigen

LUTHER E. LINDLER,[1]* GREGORY V. PLANO,[2] VALERIE BURLAND,[3] GEORGE F. MAYHEW,[3] AND FREDERICK R. BLATTNER[3]

Department of Bacterial Diseases, Division of Communicable Diseases and Immunology, Walter Reed Army Institute of Research, Washington, D.C. 20307-5100[1]; Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, Florida 33176[2]; and Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706[3]

Received 23 June 1998/Returned for modification 12 August 1998/Accepted 11 September 1998

***Yersinia pestis*, the causative agent of plague, harbors at least three plasmids necessary for full virulence of the organism, two of which are species specific. One of the *Y. pestis*-specific plasmids, pMT1, is thought to promote deep tissue invasion, resulting in more acute onset of symptoms and death. We determined the entire nucleotide sequence of *Y. pestis* KIM5 pMT1 and identified potential open reading frames (ORFs) encoded by the 100,990-bp molecule. Based on codon usage for known yersinial genes, homology with known proteins in the databases, and potential ribosome binding sites, we determined that 115 of the potential ORFs which we considered could encode polypeptides in *Y. pestis*. Five of these ORFs were genes previously identified as being necessary for production of the classic virulence factors, murine toxin (MT), and the fraction 1 (F1) capsule antigen. The regions of pMT1 encoding MT and F1 were surrounded by remnants of multiple transposition events and bacteriophage, respectively, suggesting horizontal gene transfer of these virulence factors. We identified seven new potential virulence factors that might interact with the mammalian host or flea vector. Forty-three of the remaining 115 putative ORFs did not display any significant homology with proteins in the current databases. Furthermore, DNA sequence analysis allowed the determination of the putative replication and partitioning regions of pMT1. We identified a single 2,450-bp region within pMT1 that could function as the origin of replication, including a RepA-like protein similar to RepFIB, RepHI1B, and P1 and P7 replicons. Plasmid partitioning function was located ca. 36 kb from the putative origin of replication and was most similar to the *parABS* bacteriophage P1 and P7 system. *Y. pestis* pMT1 encoded potential genes with a high degree of similarity to a wide variety of organisms, plasmids, and bacteriophage. Accordingly, our analysis of the pMT1 DNA sequence emphasized the mosaic nature of this large bacterial virulence plasmid and provided implications as to its evolution.**

The facultative intracellular parasite *Yersinia pestis* harbors at least three plasmids, one of which is common to the enteropathogenic species *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* (30, 72). The other two plasmids, designated pMT1 and pPCP1, are unique to *Y. pestis* (10) and are thought to promote the ability of this organism to penetrate deep tissues and to contribute to the acute infection associated with this species. In fact, the *Y. pestis* genome shares much homology with that of *Y. pseudotuberculosis* (2, 63), yet the infection caused by the latter organism is usually mild and self-limiting (15). Accordingly, a logical starting point to understanding the difference in the pathogenesis of *Y. pestis* and *Y. pseudotuberculosis* is to study the genes encoded on the plasmids unique to plague, pMT1 and pPCP1.

The 9.5-kb plasmid pPCP1 encodes a bacteriocin termed pesticin, a pesticin immunity protein, and a plasminogen activator (89). Loss of this plasmid increases the 50% lethal dose of the organism by a factor of 10^5 when the organism is injected subcutaneously in the mouse model of infection (90). The only characterized virulence determinant encoded by pPCP1, the plasminogen activator, has been implicated in deep tissue invasion by *Y. pestis* (11) and functions in the flea vector (58). These facts demonstrate that a plasmid, specifically harbored by *Y. pestis*, encodes a virulence factor necessary for the acute infection caused by the organism and that a single protein can influence the life cycle of the organism at multiple stages.

The largest extrachromosomal element present in *Y. pestis* was commonly called the cryptic plasmid until 1983. Protsenko et al. (73) demonstrated that the capsular protein fraction 1 (F1) and the murine toxin (MT) were both encoded by the ~100-kb element now called pMT1. The genes for each of the proteins have been cloned from *Y. pestis* EV76 and sequenced previously (36, 37, 49). Data addressing the involvement of these proteins in plague pathogenesis are open to interpretation since the effect that mutational loss has on the 50% lethal dose depends on the animal model used in the study as well as the route of infection (8, 9). However, pMT1 does appear to contribute to the acute phase of plague infection, as evidenced by the fact that strains lacking the 100-kb plasmid demonstrate reduced morbidity (27, 80, 96).

Information pertaining to the genetic characterization of the pMT1 molecule is limited. The size of the plasmid has been found to vary from approximately 90 to 288 kb in size (31). Furthermore, pMT1 has been found to integrate at multiple sites into the chromosome of *Y. pestis* at high frequency (74), with speculation that the observed integration of pMT1 into the chromosome may have been due to IS*100* homology between the two molecules. Both F1 and MT gene activation have been characterized in relation to environmental cues such as temperature and calcium (28). F1 capsule synthesis is maximal at 37°C in the absence of extracellular calcium while

* Corresponding author. Mailing address: Department of Bacterial Diseases, WRAIR, Bldg. 40, Room 2105, Washington, DC 20307-5100. Phone: (202) 782-3532. Fax: (202) 782-0748. E-mail: Dr._Luther_Lindler@wrsmtp-ccmail.army.mil.

murine toxin expression is induced at 26°C. F1 expression is therefore maximum under conditions similar to those that induce the expression of one of the major virulence determinants of *Y. pestis* (91–93). In contrast, MT production is induced in an environment similar to that which *Y. pestis* would be expected to encounter in the flea vector. The presence of genes induced under these widely different conditions indicates the presence of at least two networks regulating expression of virulence determinants operating in plague.

DNA-sequencing technology has progressed to the point that large amounts of genetic material can be sequenced in a relatively short time. Several facts make pMT1 a good candidate for large-scale DNA sequencing. First, the plasmid is unique to *Y. pestis*. Second, some derivative of the ~100-kb plasmid is always present in clinical isolates (31). Third, we already know that genes regulated by two different environmental stimuli that mimic different environments encountered during the life cycle of plague are present on this molecule. Here, we report and annotate the entire DNA sequence of the pMT1 plasmid derived from the *Y. pestis* laboratory strain KIM.

MATERIALS AND METHODS

Bacterial strains, media, and plasmid isolation. *Y. pestis* KIM10+ is a strain that contains only pMT1 (71). Plasmid DNA was prepared by growing *Y. pestis* KIM10+ in heart infusion broth (Difco Laboratories, Detroit, Mich.) at 26 to 30°C followed by alkaline lysis and polyethylene glycol precipitation (4, 46). Purified pMT1 was used in preparing DNA libraries as described below.

pMT1 library construction and DNA sequencing. Libraries of pMT1 were prepared by random shearing of plasmid DNA and size fractionation (62) and then cloned into the M13 Janus vector (12). Random phage clones were isolated, and their DNA was purified as described elsewhere (68). The DNA templates were subjected to dye-terminator sequencing by using the Prism cycle sequencing kit and ABI 377 automated sequencers (Applied Biosystems Division of Perkin Elmer, Foster City, Calif.). Sequences were assembled into contigs with the use of the Seqman II program (DNASTAR, Madison, Wis.). Suitable clones were selected for further sequencing from the opposite end to fill in coverage, resolve ambiguities, and close gaps (12). The final coverage was approximately eightfold.

DNA sequence analysis and annotation. Open reading frames (ORFs) that were at least 50 amino acids in length were identified with GeneQuest (DNASTAR). Codon usage was assessed in the program by second- and third-order statistical comparison (6) with a matrix built from all available sequences for *Yersinia* species. Although this matrix was more useful than one built from *Escherichia coli* genes, it was necessarily constructed from a relatively small data set. Generally, the start codon (including GTG and TTG) farthest upstream was used to annotate the ORF (5). For the first pass, amino acid sequences were searched against the current GenPept database by using the BLOSUM62 matrix by the DeCypher II System (TimeLogic, Inc., Incline Village, Nev.). Subsequent searches of the Swiss Protein, *E. coli*, and nonredundant GenBank databases were obtained via the Internet with BLAST software (1) from the National Center for Biotechnology Information homepage (www.ncbi.nlm.gov/BLAST/). Pairwise protein alignments were done with the BLAST algorithm (1). Protein localization was predicted for relevant translated *orf* genes with the use of the PSORT program (66). The prediction of membrane-associated helices was done with the TMpred program (45). Where appropriate, multiple protein sequences were aligned by using the algorithm developed by Lipman et al. (55). These programs can be found as part of Pedros Molecular Biology Tools at Internet site http://zguw.ibb.waw.pl/pedros.htm.

GenBank nucleotide sequence accession number. The annotated sequence was deposited in GenBank under accession no. AF074611.

RESULTS AND DISCUSSION

General overview. The fully assembled circular DNA sequence of pMT1 was 100,990 bp in length. An initial screening of the sequence with the DNASTAR program GeneQuest revealed 145 potential ORFs along the entire length of the plasmid. Each individual ORF was translated into the predicted protein, which then was used to search the various databases (GenBank, Swiss Protein, GenPept, and *E. coli*) for proteins with potentially significant homologies. The pertinent results of our searches are illustrated in Fig. 1 and summarized in Table 1. Several factors were taken into consideration for deciding whether a potential match was significant enough to report. In general, if the pMT1 ORF had significant similarity to known proteins in the database, we assigned the putative protein encoded by that ORF a like function. Homologies were considered to be significant when at least 25% of the amino acids were identical for at least 35% of the protein in the database. We decided on 25% identity to give a reasonable baseline, to which added conservative amino acid substitutions often result in higher similarity scores between protein molecules. However, in specific instances, we designated protein function as similar based on less than 25% identity. These instances are indicated in the text where relevant. The extent of homology with the database protein was set at 35% to allow for the possibility that protein domains might have different functions in different molecular contexts. We lowered our stringency when deciding whether a putative protein might function in pathogenesis. In these cases, when the region of homology included at least 20% amino acid identity with a protein that might interact with or substitute for the action of a host protein, we considered it a potential virulence factor. More weight was given to potential alignments when the homology between the *Y. pestis* ORF and the target protein sequence was located in a domain that had a known function in host physiology. Although these possibilities will require experimental confirmation, we felt it important given the fact that pMT1 is specifically harbored by *Y. pestis* and is thought to promote deep tissue spread of the organism. Finally, when the putative protein did not contain significant similarity to any known proteins, we analyzed the upstream DNA for ribosome binding sites (85) and also considered the known codon usage for *Yersinia* genes. After applying these criteria to the 145 potential ORFs initially identified on pMT1, we were left with 115 putative coding regions. Of these 115 putative ORFs, 38% had no regions of significant homology to any protein in the current databases and 7% had significant homology with previously described hypothetical proteins.

We noted an approximately 24-kb inversion when our pMT1 sequence was compared to the sequence recently submitted by Hu et al. (45a) (accession no. AF053947). The two IS*100* elements which form 1,954-bp inverted repeats (only one base difference) are 24,440 bp apart. Coordinates of the IS*100* elements are bp 46382 and 48337 and bp 72777 and 74730. It was not possible to deduce from the sequence whether this difference was due to misassembly of individual sequence reads at the areas of homology or the result of in vivo recombination in one of the two sequenced plasmids. Therefore, to confirm our physical map with the DNA sequence, we performed *Sph*I and *Hin*dIII restriction digests of our plasmid DNA and compared the fragment sizes with those predicted from the two sequence arrangements. Diagnostic fragments were obtained, confirming that the sequence assembly is correct for the molecule sequenced by us.

Potential virulence factors. An important reason for performing large sequencing projects is to aid the discovery of new virulence factors which might be used as vaccine candidates or as targets for therapeutic drugs. Since *Y. pestis* is a facultative intracellular parasite and pMT1 is thought to enhance deep tissue spread of the organism, we took note of several ORFs that had limited homology with proteins that might function during various stages of the organism's life cycle. These proteins are listed in Table 2. Although many of these homologies do not meet our criteria for general ORF homologies, we felt that a more relaxed standard should be applied to protein homologies in order to aid future research pertaining to plague pathogenesis.

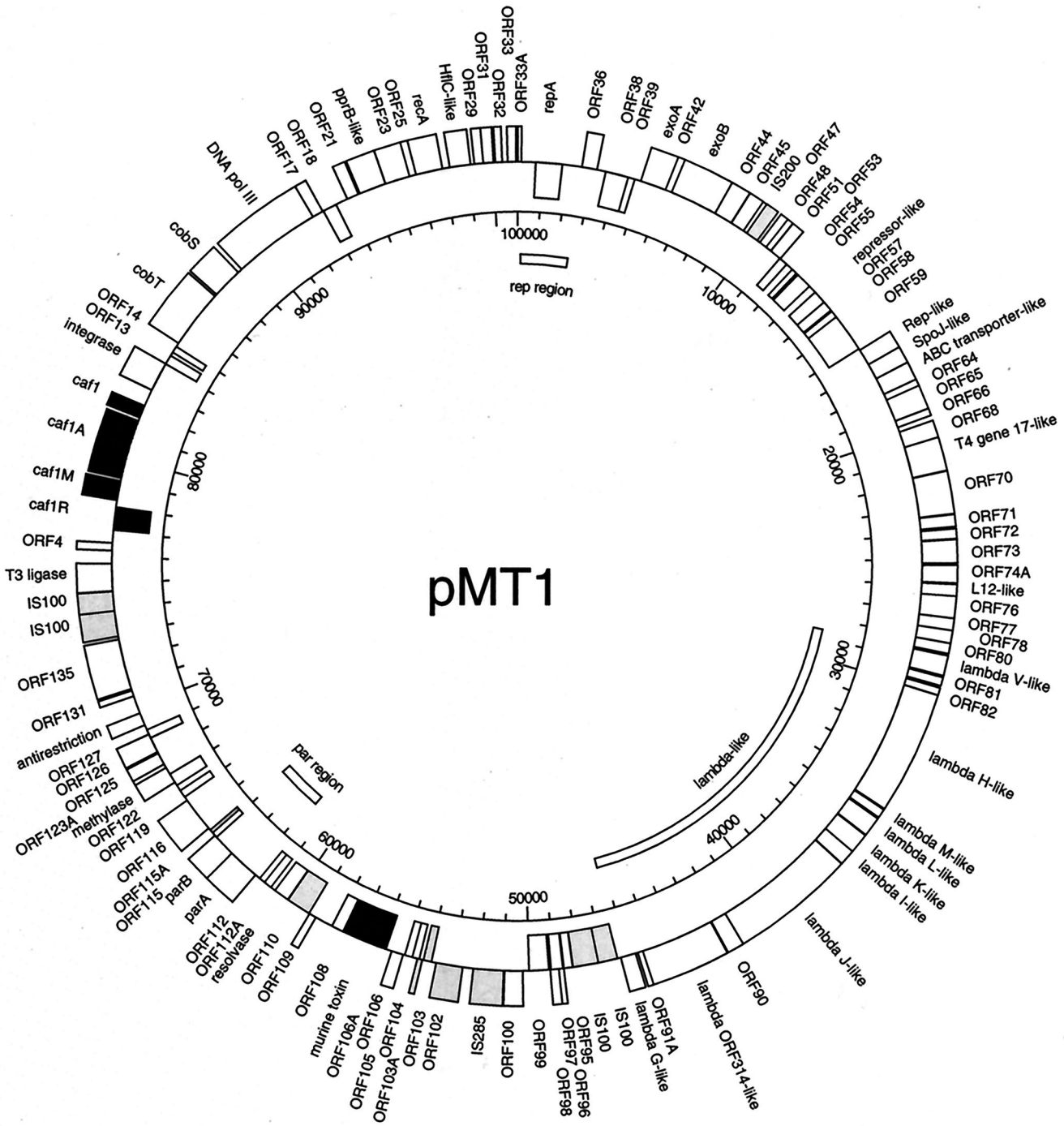

FIG. 1. Map of the whole pMT1 plasmid. The outer circle shows ORFs with their orientation denoted by their positions: outside the ring indicates clockwise, and inside the ring indicates counterclockwise. Known virulence proteins are indicated by filled boxes; insertion-associated ORFs are indicated by shaded boxes. The scale is in basepairs. All ORF designations, except the previously known genes of the capsular antigen operon and the murine toxin (black square), are putative and derived from database matches. The map was derived from the annotated DNA sequence by the computer program GeneScene, under development at DNASTAR, and edited in Adobe Illustrator 7. Nomenclature of the ORFs is as indicated in Table 1.

Potential evolution of the F1 capsule and murine toxin coding regions. The coding region for the F1 capsule protein and accessory factors was located between genes that encode proteins with a high degree of identity to phage integrases and ligases as well as IS*100*. ORF2 was similar to bacteriophage T3 ligase, and ORF12 showed a high degree of identity to a *Vibrio cholera* prophage integrase. The genes for these two phage-like proteins also flank *orf4*, which may encode a potential vasorelaxant (Fig. 1 and Table 2). Thus, the F1 capsule-coding region and one potential virulence factor may have originated by a recombination event initiated through mobile DNA metabolism.

Although the region of pMT1 that encodes the F1 capsular protein has little homology with the *E. coli* genome at the nucleotide level, we found a region from bp 77738 to 77780

TABLE 1. ORFs identified in *Y. pestis* pMT1 DNA sequence by classification[a]

ORF class	Designation	Function or comments	Organism or element (gene if known)	Accession no.	Location (bp)
DNA metabolism	ORF1	IS*100*	*Y. pestis* IS*100* (*orfB*)	U59875	73885–74661
	ORF2	Ligase	Bacteriophage T3	X05031	74680–75777
	ORF12	Integrase	*Vibrio cholera*	U39068	82931–84109
	ORF16	DNA Pol III	*E. coli*	M19334	88955–92479
	ORF26	RecA	*Bacteroides fragilis*	M63029	96910–97986
	ORF34	RepA	*E. coli* plasmid ColV	L01250	Complement, 717–1781
	ORF41	*exoA*	Bacteriophage T4	X01804	4968–6053
	ORF43	*exoB*	Bacteriophage T4	X01804	6271–8199
	ORF46	IS*200*	IS*200*	U22457	9675–10184
	ORF60	Rep-like	*Coxiella burnetti*	L34077	16197–16895
	ORF61	SpoJ-like	*Streptococcus pneumoniae*	AF000658	16862–17563
	ORF69	Gene 17-like	Bacteriophage T4	X52394	20457–21713
	ORF93	IS*100*	*Y. pestis* IS*100* (*orfB*)	U59875	Complement, 46449–47231
	ORF94	IS*100*	*Y. pestis* IS*100* (*orfA*)	U59875	Complement, 47228–48250
	ORF101	IS*285*	*Y. pestis* IS*285* (*orf2*)	X78303	51013–52221
	ORF102	Transposase	*Enterobacter*	U60777	52648–53712
	ORF108	Membrane endonuclease	*E. coli* plasmid-pKM101 (*nuc*)	U09868	Complement, 57629–58117
	ORF111	Resolvase	*Pseudomonas syringae* (*stbA*)	L48985	Complement, 60161–60781
	ORF113	ParA	Bacteriophage P1	X02954	61767–63041
	ORF114	ParB	Bacteriophage P1	K02380	63038–64009
	ORF123	Adenine-specific DNA methylase	*E. coli* pEC156 EcoVIII methylase	U48806	66648–67325
	ORF128	Antirestriction	*E. coli*	Z34467	69208–69714
	ORF135	DNA partitioning	*Rhizobium meliloti* (Orf1 and Orf2 of pRmeGR4a)	X69105	70730–72739
			Shigella sonnei (*psiB*)	U82272	
			Streptococcus pneumoniae (*spoOJ*)	AF000658	
	ORF136	IS*100*	*Y. pestis* IS*100* (*orfA*)	U59875	72863–73882
Protein metabolism	ORF28	HflC-like	*Vibrio parahaemolyticus*	U09005	98281–99111
	ORF63	ABC transporter/ATP binding	*Archaeglobus fulgidus* (AF1064)	AE001029	17500–18198
	ORF75	L12 ribosomal protein L12e	*Haloferax volcanii*	X58924	25927–26361
Gene regulation	ORF5	Caf1R	*Y. pestis* (*caf1R*)	X61996	Complement, 77118–78041
	ORF22	PrpB-like	*Pseudomonas putida*	X80272	94557–95636
	ORF56	Repressor of flagellum synthesis	*Salmonella abony* (*fljA*)	D26167	Complement, 13278–13841
Known virulence	ORF6	Caf1M	*Y. pestis* (*caf1M*)	X61996	78318–79127
	ORF8	Caf1A	*Y. pestis* (*caf1A*)	X61996	79152–81653
	ORF9	Caf1	*Y. pestis* (*caf1*)	X61996	81734–82246
	ORF107	Murine toxin	*Y. pestis* (*ymt*)	X92727	Complement, 55788–57551
Lambda-like	ORF80a	V major tail fiber	Bacteriophage lambda	P03733	28560–29303
		Intimin	*E. coli* O157:H7 (*eae*)	P43261	
	ORF84	H tail fiber	Bacteriophage lambda	AF007380	30041–34618
	ORF85	M minor tail fiber	Bacteriophage lambda	P03737	34660–34995
	ORF86	L minor tail fiber	Bacteriophage lambda	P03738	35052–35783
	ORF87a	K tail assembly	Bacteriophage lambda	P03729	35815–36570
	ORF88	I tail assembly	Bacteriophage lambda	P03730	36561–37148
	ORF89	J host specificity	Bacteriophage lambda	P03749	37164–41801
	ORF91	Hypothetical protein ORF314	Bacteriophage lambda	P03745	42469–45405
	ORF92	Tail fiber assembly	Bacteriophage lambda (*tfa*)	225931	45707–46315
Hypothetical in database[b]	ORF15	CobT	*Pseudomonas denitrificans* (*cobT*)	P29934	85075–87441
	ORF15a	CobS	*Pseudomonas denitrificans* (*cobS*)	P29933	87539–88771
	ORF29	Hypothetical	Bacteriophage P22	X78401	99265–99636
	ORF33a	Hypothetical regulatory	Bacteriophage P1	76816	100922–147
	ORF38	Hypothetical lipoprotein	*Bacillus subtilis* (*orfK yzeA*)	L16808, Z93102	Complement, 3530–4552
	ORF59	Long hypothetical protein	*Pyrococcus horikoshii* (PHBW005)	AB009472	Complement, 14573–16132
	ORF73	SRPI hypothetical protein	*Synechococcus* PCC7942 pANL	Q55032	24271–25146
	ORF104	Hypothetical protein	*E. coli*	U70214	Complement, 54408–54803
	ORF105	Hypothetical protein	*E. coli*	U70214	Complement, 54694–55002
	ORF116	Hypothetical	*Sphingomonas* S88	U51197	64388–65785
	ORF131	Hypothetical	*E. coli*	AE000133	70427–70657
Fragments[c]	ORF23	DNA polymerase	*Lactococcus lactis*	U78771	95646–96641
	ORF33	Type II	*Helicobacter pylori*	AE000647	100590–100925
	ORF99	Hypothetical protein	*Methanobacterium thermo-autotrophicum*	AE000913	Complement, 49210–50004
	ORF103	Hypothetical transposase	*Salmonella typhimurium*	Z29513	Complement, 53911–54234
	ORF103a	IS*600*	*Shigella sonnei*	X05952	54281–54481
	ORF106	Hypothetical	*Shigella flexneri*	U97489	55073–55543
	ORF106a	IS*801*	*Pseudomonas syringae*	X57269	55589–55729
	ORF110	Hypothetical	*Salmonella typhimurium*	Z29513	Complement, 59154–60140
	ORF115a	SamB-like	*Salmonella typhimurium*	D90202	87539–88771

[a] ORFs listed were assigned a putative function according to our criteria outlined in the general overview section of Results and Discussion. Classification then was based on these putative functions.
[b] Homology above our criteria with proteins in the database that have not been assigned a function.
[c] ORFs that appear to be remnants of larger proteins in the database.

TABLE 2. ORFs that may be potential virulence factors

ORF	Location (bp)	Homologous protein (target)	% Homology[a]	Accession no.	Reference(s)
ORF4	76298–76603	C-type natriuretic peptide from *Squalus acanthias*	43/30	P41319	83
ORF17	92476–92919	Delta insecticidal protein from *Bacillus thuringiensis*	40/18	P05628	35
ORF18	Complement, 92949–93512	RTX[b] toxin of *Actinobacillus pleuropneumoniae*	21/11	D16582	32, 65
ORF21	94015–94448	Laminin of *Homo sapiens*	23/5	Q16787	79, 95
		Paramysin-related protein of *Onchocerca gibsoni*	21/18	U20609	25, 99
ORF72	23873–24244	Major myristoylated alanine-rich protein kinase C substrate	24/32	P29966	41
ORF74a	25221–25883	Bacteriophage lambda V protein	40/41	P03733	81
		Citrobacter freundii intimin	30/10	Q07591	82

[a] Percentage of identical amino acids compared with the percentage of the total target protein sequence.
[b] RTX represents repeats in the structural toxin.

within the *caf1R* gene that was 88% identical to the *E. coli afrR* locus (Table 3). The AfrR protein is a pilus expression transcriptional regulator (GeneBank accession no. L08467 [98]). The similarity extended for 42 bp, which would precisely encode the amino acids FYDSQQTFTREFKK. The deduced peptide sequence lies within the region of homology between the Caf1R and AfrR proteins as well as many other transcriptional regulators in the *E. coli* AraC family. These findings support the idea that the *Y. pestis* protein is a member of the AraC family (49) of transcriptional regulators and suggest a conserved block of nucleotides involved in the functional evolution of these *cis*-acting regulatory elements. Furthermore, the fact that the *Y. pestis caf1R* gene encodes a region identical to a small region of an *E. coli* pilus regulatory gene suggests that the plague capsule operon may have originated as an adhesin. This possibility is supported by the fact that the two other genes necessary for F1 expression are similar to pilin chaperones (36) and membrane anchor proteins (49).

The molecular evolution of the region surrounding the other known virulence factor of *Y. pestis*, MT, may have occurred through several abortive or imprecise transposition events from another plasmid or bacteriophage. Several features found within the pMT1 DNA sequence suggest this possibility. First, the MT-coding locus, *ymt*, is flanked by a partial gene duplication event with several incomplete transposon sequences and ORFs (Fig. 2). A striking example of an incomplete ORF was found after ORF110 was compared with ORF103. The ORF110 peptide was predicted to be composed of 312 amino acids and was found to be 78% identical over its entire length to a group of repeated hypothetical proteins in the *E. coli* (13) and *Salmonella typhimurium* (GenBank accession no. Z29513) genomes designated the YadD family. Second, there is a partial repetition of DNA that encodes the *S. typhimurium parA* and *parB* loci (17). The *par* repetition was noted by the fact that nucleotides 54023 through 54219 are an imperfect partial duplication of bp 59329 through 59486 (Fig. 2). To the right of the second copy of the *Salmonella par* region, we identified an ORF that had 38% identity with a resolvase encoded by *Pseudomonas syringae* (40). To the right of the putative resolvase, we located a small 85-bp region (nucleotides 60886 to 60970) that was 89% identical to the *Salmonella* plasmid-partitioning sequences. Third, we located two fragments of transposons in this region. The first remnant was found as a partial ORF from bp 54281 through 54481 and would be predicted to encode amino acids that were 69% similar to residues 75 to 116 of the *Shigella sonnei* IS*600* hypothetical 31-kDa protein (57). The second partial ORF was encoded by nucleotides 55589 to 55729 and was predicted to encode a polypeptide that was 88% similar to IS*801* transposase amino acids 260 to 277 (78). The fourth piece of evidence that the MT gene was acquired by *Y. pestis* through an illegitimate recombination event from another replicon is the presence of DNA homology with other known plasmids (Table 3). Taken together, these observations strongly suggest that *Y. pestis* acquired the MT-coding sequences and possibly other virulence factors through recombination events that originated from mobile genetic elements. However, no one single event can explain the molecular architecture that we observed.

If *Y. pestis* pMT1 is a mosaic of different genetic elements or regions, we reasoned that a base composition analysis of the plasmid might indicate and confirm potential regions acquired by horizontal gene transfer. In agreement with the hypothesis that the regions of the *Y. pestis* plasmid containing the F1 capsule genes and the gene encoding the MT originated by transfer of mobile genetic elements, we noted that the regions of pMT1 which contained these genes had a guanine-plus-cytosine (G+C) content significantly different from those of surrounding regions of DNA (Fig. 3). Specifically, the DNA

TABLE 3. Nucleotide identities between *Y. pestis* pMT1 and DNA sequences in the nonredundant GenBank database[a]

Position (bp) in pMT1	Homologous DNA source (% identity)[b]	GenBank accession no.
77738–77780	*E. coli afrR* (88)	L08467
44273–44300	Bacteriophage K3 gene 37 (93)	X00615
45200–45395	Bacteriophage lambda (87)	J02459
45347–45395	*Salmonella typhimurium pagJ* (88)	AF013776
53973–53997	*E. coli hlyR* (96)	X07565
54023–54219	*Salmonella typhimurium* plasmid partitioning (88.5)	M97752
59329–59486	*Salmonella typhimurium* plasmid partitioning (88)	M97752
59891–60117	*Salmonella typhimurium* plasmid partitioning (79)	M97752
60886–60970	*Salmonella typhimurium* plasmid partitioning (89)	M97752
54411–55559	*E. coli* chromosomal DNA (90)	U70214
54418–55559	*Shigella flexneri* RT (90)	U97489
55347–55399	*Pseudomonas putida* (85)	AB004059
58641–58691	*Salmonella enteritidis virR* (92)	D14491
58641–58691	*Salmonella choleraesuis* pMBA1 (92)	X54148
58641–58691	*Salmonella dublin* pSDL2 (92)	M58505
58675–58745	*Salmonella* pSC101 (84)	X01654
61836–62386	Bacteriophage P7 *par* region (78)	X17529
61661–61934	*E. coli* F plasmid RepF1B (89)	M26308
58627–58662	*Neisseria gonorrhoeae* pFA7 (89)	X01654
58627–58662	*Klebsiella pneumoniae* pLST1000 (89)	X64367
58627–58662	*E. coli* pRK2 (89)	U05773
69212–69714	*E. coli* antirestriction protein (92)	Z34467
72557–72630	*E. coli* F plasmid *psiB* (96)	X12462
72558–72630	*Shigella sonnei* plasmid ColIb-P9 *psiB* (82)	U82272

[a] Identities of 78% or greater for a continuous group of 25 nucleotides were considered significant.
[b] In some cases, nucleotide identity between pMT1 and the target DNA sequence varied over the length of the homologous region. In these instances, the percent identity was averaged for all homologous regions.

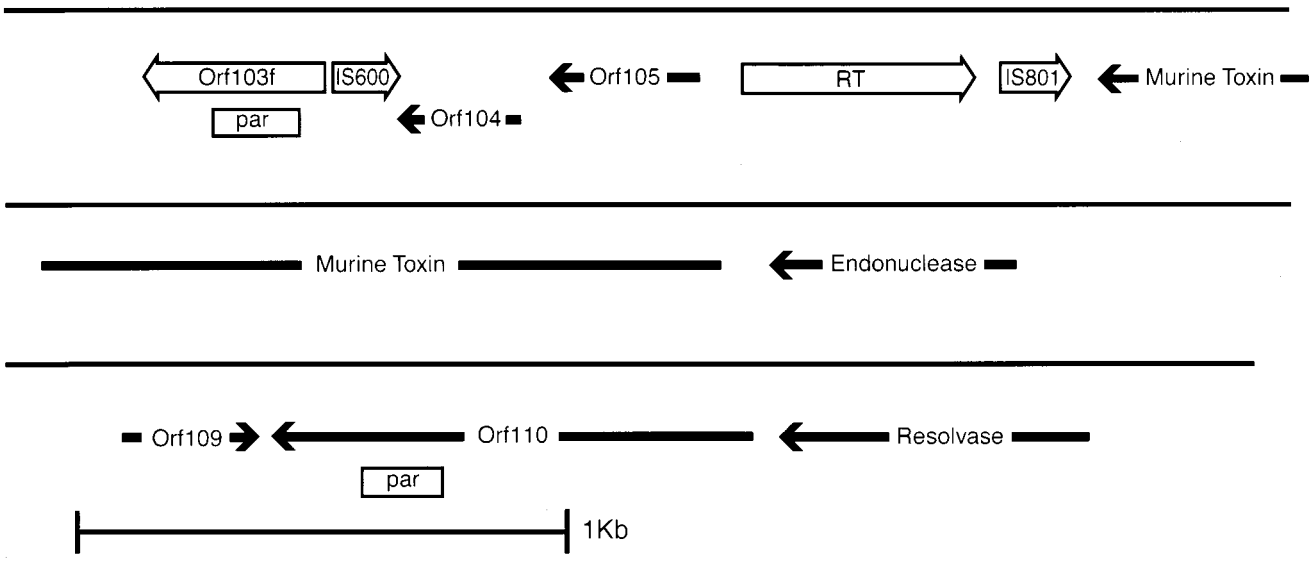

FIG. 2. Map of *Y. pestis* pMT1 region encoding the MT virulence determinant. The region from ca. bp 53500 to 61300 is shown and is represented by the solid line. ORFs that were predicted to be intact are shown as solid arrows. ORFs that appear only partial and were judged to be remnants of intact coding regions are shown as open arrows. The open boxes labeled par designate regions with high nucleotide homology with the *S. typhimurium par* locus (17). RT, reverse transcriptase-like partial ORF.

encoding the *caf1R*, *caf1M*, *caf1A*, and *caf1* genes (bp 3067 through 8195) had a G+C content of 39.2% and the *ymt* locus (bp 82727 through 84490) had a G+C content of 38.1%. These base compositions can be contrasted with the overall 50.1% G+C content of the entire pMT1 molecule. Our observation that the two known virulence factors encoded by pMT1 have G+C contents significantly different from that of the surrounding DNA may be similar to one of the criteria currently used to define pathogenicity islands (39). The differing base composition of these known virulence genes relative to that of surrounding pMT1 DNA does suggest, in a limited way, that plasmid-borne virulence genes may be analogous to chromosomal pathogenicity islands. The observation that virulence genes generally have a base composition different from that of the flanking DNA has been observed for the hemolysin genes present on the *E. coli* virulence plasmid pO157 (14) and within the pathogenicity island of this organism which encodes the locus for enterocyte effacement (70). If the trend of virulence genes, or blocks of genes, to have a G+C content different from that of the surrounding genetic material is genuine, the region located to the right of IS*200* on pMT1, as shown in Fig. 3, may warrant further investigation. Database searches with putative ORF48 through ORF58 translated from pMT1 sequences in this area did not reveal any homology with known

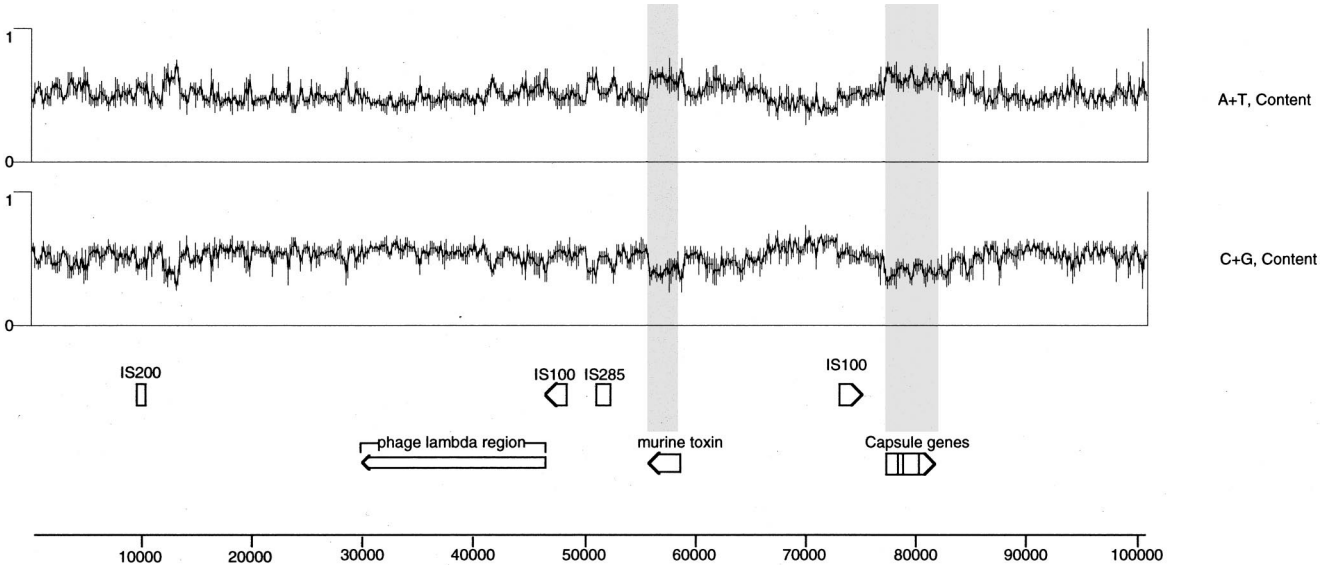

FIG. 3. Base composition of pMT1. The plots showing A+T and G+C content were derived by DNASTAR's GeneQuest program, which also displays selected ORFs and other annotated features to the correct scale. The shaded bars mark the previously known virulence genes to highlight the low G+C composition of these regions. Note a small region to the right of IS*200* that also shows low G+C content (discussed in Results and Discussion). The scale at the left is equivalent to 0 to 100% for each plot. The scale below the figure shows the plasmid genome in base pairs.

or potential virulence factors. In fact, only ORF56 displayed any significant homology with any of the proteins in the current databases.

Insertion sequence elements. The MT plasmid appeared to be a chimera of many types of mobile genetic elements including bacteriophage, plasmids, insertion sequences, and one defective bacterial reverse transcriptase which may be a remnant of a retron (87, 94). Insertion sequences are known to promote illegitimate recombination events and promote genetic plasticity. With this fact in mind, we noted that pMT1 encoded several IS elements as well as what appeared to be nonfunctional remnants of mobile genetic elements. IS elements were considered to be an allele of a previously published mobile element when the DNA sequence was at least 85% identical to entries in the current GenBank database. An element was considered to be an incomplete copy when the identity covered less than 80% of the known length of functional elements in the database. Using these criteria, we noted four apparently complete IS elements: one copy of IS*200* (3, 86) from bp 9576 to 10292, two copies of IS*100* (31, 59) from bp 46383 to 48336 and 72777 to 74730, and one copy of IS*285* (31) from bp 50872 to 55256.

Our analysis revealed one potential new IS element located at bp 52465 to 53758. These coordinates mark the beginning and end of a direct repeat sequence, GATGATAA, that flanks a putative transposase which we designated ORF102. ORF102 had the greatest identity, 40% over 96% of the target protein, with a putative transposase previously found in *Enterobacter aerogenes* (88) as well as 36% identity over 96% of a putative transposase previously described in *Yersinia enterocolitica* (77). We analyzed the region surrounding ORF102 for DNA sequence features previously found in IS*1328* of *Y. enterocolitica*. The only nucleotide feature that we found was an exact match of a GGAGG potential ribosome binding site that was 8 bp upstream of the ORF102 ATG translation initiation codon. We did not find any inverted repeat sequences of 5 bp or longer in contrast to the 6-bp inverted repeats described for *Y. enterocolitica* IS*1328*. This result was not particularly surprising given the amount of sequence divergence at the protein and DNA levels. Accordingly, we decided to identify this potential IS with a new designation, IS*1618*.

The *Y. pestis* MT plasmid included a 1,150-bp region that was 93% identical to the *E. coli* chromosome (5) at the DNA level. We identified a putative protein coding sequence within this homologous region, designated ORF106, that was 91% identical over 32% of a previously described *Shigella flexneri* reverse transcriptase (RT)-like protein (76). The full-length *S. flexneri* RT is 431 amino acids compared to 151 residues for ORF106. Given the high degree of identity of ORF106 with the amino terminus of a bacterial RT, we speculate that the pMT1 ORF may be a remnant of a bacterial retron element (94). However, the pMT1 DNA sequence immediately upstream of ORF106 lacks any inverted repeat sequences or potential RNA structures that have been associated with bacterial retron elements (44, 52, 94). The fact that ORF106 includes such high identity to a bacterial RT but none of the other features associated with retrons suggests that this region of DNA may have been acquired by *Y. pestis* pMT1 through a mechanism other than retrotransposition. Furthermore, the high level of conservation of the DNA sequence in this area of pMT1 suggests that these sequences were acquired recently in the evolution of this plasmid.

Partial lambdoid prophage. ORF80a through ORF92 show a high degree of similarity to lambda (81) and phage BF23 (67) proteins (Fig. 1). The amount of amino acid identity ranged from 31% over 54% of lambda protein J to 61% over 91% of lambda protein G. Further support for the lambdoid origin of this region of pMT1 was found in the gene order of several of the ORFs. Specifically, ORF84, ORF85, ORF86, ORF87a, ORF88, and ORF89 are in the same order as the genes encoding lambda proteins H, M, L, K, I, and J. The gene that encodes one of the lambda-like proteins, ORF80a, appears to have been partially duplicated to generate *orf74a*. The partial duplication event is indicated by our observation that ORF74a is 45% identical over 47% of ORF80a. We analyzed the DNA surrounding ORF80a and ORF74a for repeat elements that may have been involved in this duplication but without success. Accordingly, a mechanism for the partial duplication and rearrangement of these sequences encoded by pMT1 was not evident in the DNA sequence. It is interesting that ORF74a may encode a potential virulence factor (Table 2).

Other mobile genetic elements. Some regions of pMT1 obviously arose from either plasmid or phage gene transfer and recombination events. The DNA sequence data that suggest an origin of known virulence factors in relation to putative mobile genetic events have already been presented. However, in general, it was difficult to determine exactly where the potential prophage or other plasmid molecules began and ended. To the best of our ability, the most that can be determined about the origins of various regions of *Y. pestis* pMT1 is the mosaic nature of the entire molecule. As an example, the DNA sequence between ORF12 and ORF43 includes putative gene products with a high degree of homology to *E. coli* DNA polymerase III (ORF16), *Bacteroides fragilis* RecA (ORF26), and the gene 47 polypeptide of bacteriophage T4 (ORF42). Given a mosaic genetic element, the most reliable means of determining the origin of specific regions of pMT1 is comparison of nucleotide sequences. Table 3 lists all significant nucleotide homologies found between pMT1 and the current nonredundant GenBank databases excluding previously entered *Y. pestis* sequences or known insertion sequences. Two important features were noted from this analysis. First, most of the significant nucleotide homologies occur in a segment containing 30% of the pMT1 sequence from bp 44272 to 72640. The high conservation of DNA sequence in this region could indicate either that these sequences were acquired relatively late in pMT1 evolution or that the functions encoded by these sequences are very important in plasmid maintenance and function. In fact, most of the nucleotide homology was found with genes required for plasmid maintenance, such as partitioning during cell division. Second, the majority of conserved DNA sequence was similar to that of large plasmids known to be involved in the pathogenesis of several enteric organisms. Taken together, these observations indicate an evolutionary linkage among enteric virulence plasmids.

In support of a link between the evolution of plasmids found in enteric organisms, we noted that ORF123, a putative DNA methylase, had 91% identity to ORFL7074 found on *E. coli* pO157 (14). Furthermore, Krause et al. (51) have reported a small region of homology between the *Y. pseudotuberculosis* virulence plasmid and a *Salmonella dublin* virulence plasmid. We found a 51-bp region (nucleotides 58641 to 58691) that was 92% identical to the common sequence found on the *S. dublin* plasmid (51). The 51-bp region of homology encoded by pMT1 was located upstream of *S. dublin* ORF1 (51) and did not include the inverted repeats reported by these investigators. The possible implications of this nucleotide homology remains to be elucidated.

Evolution of pMT1 could have been facilitated by either conjugational (50) or transductional (100) movement of genetic material. The *Y. pestis* murine toxin plasmid has been shown to integrate into the chromosome and to promote trans-

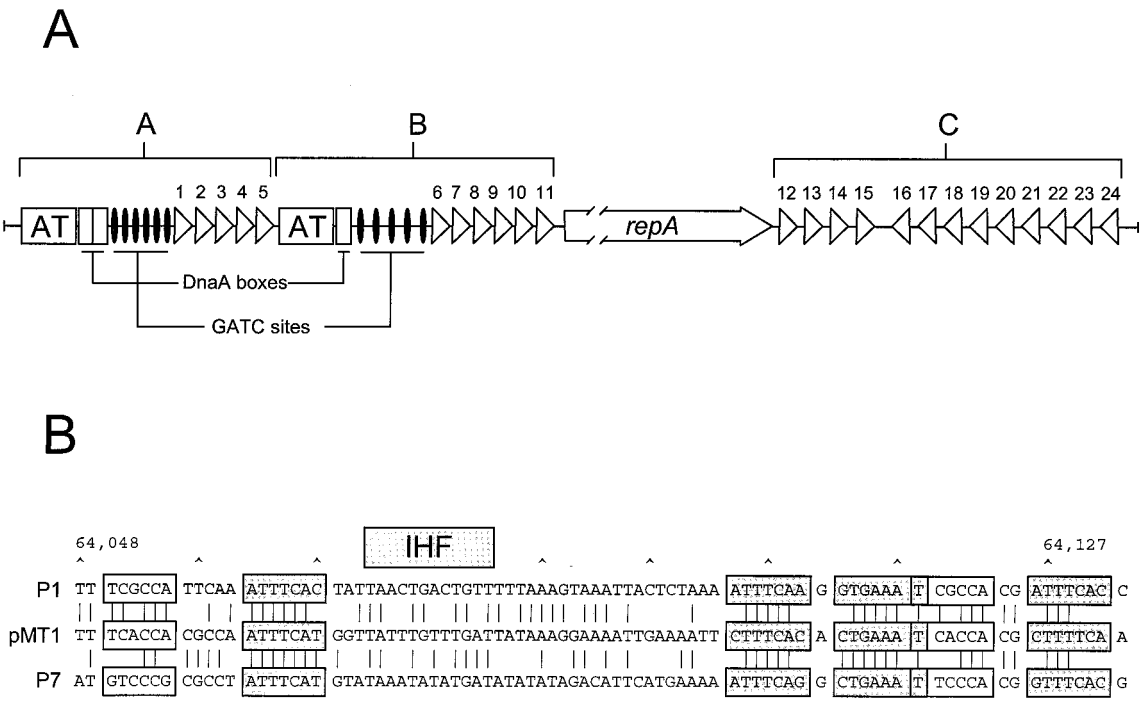

FIG. 4. Organization of the pMT1 replicon. (A) The pMT1 replicon (bp 151 to 2601) consists of a single *repA* gene (open arrow) flanked by sets of repeated elements (iterons) indicated by triangles and numbered 1 to 24. The region upstream from *repA* contains a number of sites present in many bacterial origin sequences including DnaA-binding sites, GATC Dam methylase recognition sites and two A-T-rich regions. The iterons and additional sequence elements are organized into two potentially functional origins of replication (regions A and B). Downstream of *repA* are 13 additional repeats (region C) that by analogy with other iteron-based replicons, are involved in plasmid-copy-number control. (B) Alignment of the pMT1 (bp 64048 to 64113) and P1 and P7 *parS* partition sites. The hexamer motifs (open boxes), heptamer motifs (shaded boxes), and the IHF recognition sites are shown. Base identities are indicated by lines.

fer of genetic markers to a recipient strain (74). With these facts in mind, we looked for genes that might promote or be associated with transfer of pMT1 between bacterial cells. We took particular note of ORF128 since it displayed 93% identity over 95% of the predicted sequence of an *E. coli* antirestriction protein (21). Antirestriction proteins are thought to function during conjugation to inhibit cleavage of the donor DNA before methylation can occur and are usually located near the origin of transfer (*oriT*) of the plasmid. In fact, we searched the area downstream from *orf128* between bp 69714 and 70397 for features found in F-like plasmid *oriT* sequences (33, 38, 53, 97) without success. An expanded search of the entire pMT1 sequence did not reveal any other portion of the plasmid that might function as *oriT*. Since the plasmid has evolved as part of the *Y. pestis* genetic material, it is possible that any *oriT* that may have been present on pMT1 has been lost or has become nonfunctional due to mutation. Further experimentation will be required to determine if the plasmid encodes an *oriT* and where it is located.

Replication and partition functions. DNA sequence analysis revealed a single potential plasmid replication region (bp 151 to 2601) consisting of a structural gene (*repA*) and additional sequence elements characteristic of plasmid replicons that employ an iteron-based replication initiation and control mechanism. The ~32-kDa predicted *repA* gene product (ORF34) showed a high degree of similarity to a number of plasmid replication initiation proteins, including those associated with the RepFIB (62% identity), P1 (47% identity), P7 (47% identity), and RepHI1B (40% identity) replicons. Upstream and downstream of the *repA* locus we found two sets of 19-bp direct DNA repeats (iterons). The location and orientation of these repeats and their orientation relative to *repA* are presented in Fig. 4A. The consensus sequence for these 19-bp repeats (5′-AACCACTGTAGAGAGTAAA-3′) is most similar to the 17-bp direct repeats associated with the RepHI1B replicon (5′-ATCCACTATACCGGGTA-3′), matching at 12 of the 17 possible positions (34). The direct repeats of iteron-containing plasmids have been shown to provide specific multiple binding sites for the plasmid Rep protein, which is an essential step in the initiation of plasmid replication. In addition to their role in plasmid replication, Rep proteins of iteron-carrying plasmids are involved in the regulation of their own synthesis, i.e., autoregulation (19), and in the control of plasmid copy number (18).

Additional sequence elements characteristic of origins of replication are present in the sequence upstream from the *repA* coding region. These include three DnaA binding sites, two A-T-rich regions, and 11 Dam methylase (GATC) recognition sequences (Fig. 4A). These elements are organized into two separate potentially functional origins of replication (regions A and B). Binding of the Rep protein to iterons at the replication origin has been shown to cause bending and/or melting of the adjacent A-T-rich region (18, 64). The Rep protein, in concert with the host DnaA protein, then functions to recruit host replication proteins into the open replication initiation complex (64).

The region downstream of the RepA coding region (region C) contains an additional 13 direct repeats (Fig. 4A). The downstream cluster of repeats in a number of iteron-containing plasmids has been shown to be nonessential for plasmid replication but plays an essential role in plasmid-copy-number control (18, 69). Furthermore, incompatibility of iteron-containing plasmids is primarily a function of the iterons and the binding specificity of the replication initiation protein for these

DNA sequences (18). However, neither the copy number nor the incompatibility characteristics of a plasmid can be empirically determined from the sequence or genetic organization of these repeats.

The partitioning system of pMT1 (bp 61661 to 64161) is located approximately 36-kb (Fig. 1) from the origin of replication and appears to closely resemble the *parABS* system of bacteriophages P1 (18) and P7 (56). The pMT1 *parA* gene is predicted to encode an ~44.9-kDa product (ORF113) that shows 90% identity and 95% similarity to the P7 ParA protein and 57% identity and 75% similarity to the P1 ParA protein. The pMT1 *parB* locus is predicted to encode an ~36.4-kDa product (ORF114) that is 67% identical and 83% similar to the P7 ParB protein and 44% identical and 62% similar to the P1 ParB protein. The P1 and P7 ParA proteins are ATPases whose activities are stimulated in vitro by ParB (23). ParB is a DNA binding protein that recognizes the *cis*-acting *parS* site, which lies immediately downstream of *parB* (22, 43). The genetic organization, spacing, and specific sequence motifs of the pMT1 *parS* site are similar to those found at the P1 and P7 *parS* sites. Figure 4B shows an alignment of the pMT1 *parS* site with both the P1 and P7 *parS* sites. Two repeated elements, a hexamer box (open box) and a heptamer box (shaded box), have been shown to be required for proper ParB binding and partitioning activity (22). The relatively well-conserved heptamer motifs represent individual binding sites for ParB. The hexamer motifs, which differ somewhat between the P1 and P7 *parS* sites, are responsible for the species specificity of the ParB-*parS* interaction (42). Interestingly, the P1 *parS* site, which does not function with P7 Par proteins, can be converted to a completely functional P7 *parS* site by exchanging just 5 bases located within the two hexamer motifs (42, 75) shown in Fig. 4B. Also, in the P1 and P7 systems, ParB interacts cooperatively with the integration host factor (IHF) for *parS* binding (22). We noted a putative IHF binding site within the putative pMT1 *parS* site (Fig. 4B). The amino acid and nucleotide sequence similarities between the pMT1, P1, and P7 *parABS* partitioning components suggest that the pMT1 system functions in a manner analogous to that of the P1 and P7 systems.

Metabolic genes. In order to survive, obviously a pathogen must be able to scavenge or to synthesize required cellular precursors. We found two potential anabolic genes encoded by pMT1. ORF15 is highly similar (28% identity over 42% of the target protein) to the *Pseudomonas denitrificans cobT* gene product, and ORF15a is similar (42% identity over 75% of the target protein) to the *P. denitrificans* CobS protein (16). Interestingly, *cobT* and *cobS* are linked to each other yet are not located with other *cob* gene clusters in *P. denitrificans* (16). Furthermore, the two genes are arranged in reverse order compared to their relative locations in *P. denitrificans*. The products of the *cobT* and *cobS* genes are involved in conversion of flavin mononucleotide to vitamin B_{12}. In *E. coli* and *S. typhimurium* there are at least 28 genes necessary for the synthesis of this vitamin cofactor which are linked on the chromosome (47). To our knowledge, this is the first example of genes necessary for vitamin B_{12} synthesis being located on a plasmid. We find this fact very intriguing given the fact that *Y. pestis* is a facultative intracellular parasite. It remains to be determined if these other loci are present on the *Y. pestis* chromosome or if the organism scavenges the precursors necessary for the CobT and CobS homologs encoded on pMT1 for use as a substrate in vitamin synthesis.

Gene mosaics. Shapiro (84) and Boyd et al. (7) have noted that the mosaic nature of genetic organization can be seen at multiple levels. Our examination of the molecular structure of pMT1 supports the view that evolution occurs by a natural genetic engineering process as well as by classical mechanisms of genetic drift and selection. Specific examples of these processes have already been presented. Briefly, ORF74a may be an example of a bacteriophage protein that has undergone classic gene duplication and has evolved by selection of a more important function in pathogenesis. The highly conserved group of 42 nucleotides found in the *E. coli afrR* locus and the *Y. pestis caf1R* regulatory elements may represent DNA that encodes a functional domain which in nature can generate different regulatory proteins in the *araC* family. The regions of pMT1 that encode groups of genes with significant identity to bacteriophage lambda genes and the regions of highly conserved nucleotide sequences (Table 3) indicate the overall mosaic structure of this molecule.

To extend the theme of mosaic nature of genetic material, we noted an example of a single gene that appears to be composed of domains derived from other individual protein molecules. ORF135 shows greater than 20% identity to at least four different proteins in the current databases. Beginning at the amino terminus, residues 37 through 201 were 26% identical to a *S. pneumoniae* SpoJ-like protein (20). SpoJ is a chromosomal partitioning protein originally identified in *Bacillus subtilis* (54). The next region of homology within ORF135, predicted amino acids 135 to 310, displayed 38% identity to *Rhizobium meliloti* ORF1 (60). This third region of homology, residues 309 through 602, was 23% identical to another *R. meliloti* protein, designated ORF2 (61). The percent identity between ORF2 and ORF135 was below our general cutoff; however, given the high degree of identity to *R. meliloti* ORF1, we decided that the homology was genuine. Both ORF1 and ORF2 of *R. meliloti* are necessary for stable replication of plasmid DNA and are thought to function in partitioning of the replicon into daughter cells. The last region of homology within ORF135 occurred in both the nucleotide and protein sequences. As shown in Table 3, bp 72558 to 72585 was homologous to *E. coli psiB* and bp 72558 to 72630 was homologous to *S. sonnei psiB*. ORF135 residues 575 to 646 were 65% identical to *S. sonnei* PsiB (48). The PsiB-like proteins are thought to be responsible for inhibition of the cellular SOS response during bacterial conjugation (29). Interestingly, the region of protein identity, residues 610 to 634, would encode a polypeptide of only approximately 8-kDa compared to the 15-kDa molecular mass for the *S. sonnei* PsiB protein. Therefore, the PsiB-like region of ORF135 may define a small domain which, in the context of the large protein, could perform the same function as PsiB. Alternatively, the PsiB region of ORF135 may not function in SOS inhibition but rather has evolved to some other function in the current molecular context.

Although it is not obvious from the DNA or protein sequence how ORF135 might have evolved, the homologies strongly suggest that this putative protein is involved in plasmid partitioning. The molecular structure of ORF135 could have arisen either by mutation of an ancestral protein or by domain splicing similar to exon shuffling in eucaryotes (26). The fact that some of the regions of homology overlap each other would argue for mutation of an ancestral protein. Specifically, region 1 of ORF135, which has homology with SpoJ-like proteins, overlaps region 2, which has homology with *R. meliloti* ORF1 by 66 amino acids. Similarly, region 3, which has identity with *R. meliloti* ORF2, overlaps the PsiB region by 27 residues. Only the second and third regions of homology, *R. meliloti* ORF1 and ORF2, respectively, appear to have no appreciable overlap (two amino acids). Taken together, our analysis of ORF135 suggests a combination of mutation and potential splicing or

recombination of independent genes to create the pMT1 putative polypeptide. In this scenario a protein similar to SpoJ-*R. meliloti* ORF1 would have been recombined in frame with a protein similar to *R. meliloti* ORF2-PsiB. The molecular mechanism for such natural genetic engineering in prokaryotes is not clear; however, directly repeated sequences have been suggested as one possibility (24). We found many directly repeated sequences within ORF135, although none could explain the potential joining of the two proteins as proposed above.

Summary. The nucleotide sequence of *Y. pestis* pMT1 has provided a wealth of new information. Our analysis has allowed us to identify several genes to target for further study in order to access their possible roles in pathogenesis. Deciphering the potential roles of these proteins improves our understanding not only of disease but also of host physiology. As more complete virulence plasmid DNA sequences become available we will begin to understand the mosaic nature of these molecules and what new combinations we might expect in the future. Detailed molecular analysis of the structure of virulence plasmids will impact our ability to predict the emergence of bacterial pathogens as well as to detect their presence. Although much of our specific analysis requires confirmation of protein function, it does confirm and expand our knowledge about pathogen evolution.

ACKNOWLEDGMENTS

We thank R. D. Perry for supplying pMT1 purified DNA for nucleotide sequencing. J. Norris-Thomas is thanked for her help in manuscript preparation. We thank N. J. Perna, N. Saunders, and B. Larson for helpful comments during preparation of the manuscript. We also thank G. Plunkett III for constructing the codon usage matrix for *Yersinia* and N. J. Perna for help with database searches. We also thank D. J. Rose and the technical staff of the Wisconsin Genome Project for sequencing and also G. Peyrot for editing the aligned project.

This work was supported by Public Health Service grant P01 HG10428 (F.R.B.) and the United States Army Medical Research and Materiel Command (L.E.L.).

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Bercovier, H., H. H. Mollaret, J. M. Alonso, J. Brault, G. R. Fanning, A. G. Steigerwalt, and D. J. Brenner.** 1980. Intra- and interspecies relatedness of *Yersinia pestis* by deoxyribonucleic acid hybridization and its relationship to *Yersinia pseudotuberculosis*. Curr. Microbiol. **4:**225–229.
3. **Beuzon, C. R., and J. Casadesus.** 1997. Conserved structure of IS*200* elements in Salmonella. Nucleic Acids Res. **25:**1355–1361.
4. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
5. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirpatrick, M. A. Goeden, D. J. Rose, R. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1462.
6. **Borodovsky, M., and J. McIninch.** 1993. Genmark: parallel gene recognition for both DNA strands. Comput. Chem. **17:**123–133.
7. **Boyd, E. F., C. W. Hill, S. M. Rich, and D. L. Hartl.** 1996. Mosaic structure of plasmids from natural populations of *Escherichia coli*. Genetics **143:** 1091–1100.
8. **Brubaker, R. R.** 1972. The genus *Yersinia*: biochemistry and genetics of virulence. Curr. Top. Microbiol. **57:**111–158.
9. **Brubaker, R. R.** 1983. The Vwa^{+} virulence factor of yersinae: the molecular basis of the attendant nutritional requirement for Ca^{++}. Rev. Infect. Dis. **5:** S748–S758.
10. **Brubaker, R. R.** 1991. Factors promoting acute and chronic diseases caused by yersiniae. Clin. Microbiol. Rev. **4:**309–324.
11. **Brubaker, R. R., E. D. Beesley, and M. J. Surgalla.** 1965. *Pasturella pestis*: role of pesticin I and iron in experimental plague. Science **149:**422–424.
12. **Burland, V., D. L. Daniels, G. Plunkett III, and F. R. Blattner.** 1993. Genome sequencing on both strands: the Janus strategy. Nucleic Acids Res. **21:**3385–3390.
13. **Burland, V., G. Plunkett III, H. J. Sofia, D. L. Daniels, and F. R. Blattner.** 1995. Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. **23:**2105–2119.
14. **Burland, V., Y. Shao, N. T. Perna, G. Plunkett III, H. J. Sofia, and F. R. Blattner.** The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Submitted for publication.
15. **Butler, T.** 1983. Plague and other yersinia infections, p. 111–159. *In* W. B. Greenbugh III and T. C. Merigan (ed.), Current topics in infectious disease. Plenum Press, New York, N.Y.
16. **Cameron, B., C. Guilhot, F. Blanche, L. Cauchois, M. C. Rouyez, S. Rigault, S. Levy-Schil, and J. Crouzet.** 1991. Genetic and sequence analysis of the *Pseudomonas denitrificans* DNA fragment containing two *cob* genes. J. Bacteriol. **173:**6058–6065.
17. **Cerin, H., and J. Hackett.** 1993. The parVP region of the Salmonella typhimurium virulence plasmid pSLT contains four loci required for incompatibility and partition. Plasmid **30:**30–38.
18. **Chattoraj, D. K., and T. D. Schneider.** 1997. Replication control of plasmid P1 and its host chromosome: the common ground. Prog. Nucleic Acid Res. Mol. Biol. **57:**145–186.
19. **Chattoraj, D. K., K. M. Snyder, and A. L. Abeles.** 1985. P1 plasmid replication: multiple functions of RepA protein at the origin. Proc. Natl. Acad. Sci. USA **82:**2588–2592.
20. **Cheng, Q., E. A. Campbell, A. M. Naughton, S. Johnson, and H. R. Masure.** 1997. The *com* locus controls genetic transformation in *Streptococcus pneumoniae*. Mol. Microbiol. **23:**683–692.
21. **Chilley, P. M., and B. M. Wilkins.** 1995. Distribution of the *ardA* family of antirestriction genes on conjugative plasmids. Microbiology **141:**2157–2164.
22. **Davis, M. A., and S. J. Austin.** 1988. Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. EMBO J. **7:**1881–1888.
23. **Davis, M. A., K. A. Martin, and S. J. Austin.** 1992. Biochemical activities of the ParA partition protein of the P1 plasmid. Mol. Microbiol. **6:**1141–1147.
24. **De Chateau, M., and L. Bjorck.** 1996. Identification of interdomain sequences promoting the intronless evolution of a bacterial protein family. Proc. Natl. Acad. Sci. USA **93:**8490–8495.
25. **Donelson, J. E., B. O. L. Duke, D. Moser, W. Zeng, N. E. Erondu, R. Lucius, A. Renz, M. Karam, and G. Z. Flores.** 1988. Construction of *Onchocerca volvulus* cDNA libraries and partial characterization of the cDNA for a major antigen. Mol. Biochem. Parasitol. **31:**241–250.
26. **Doolittle, R. F.** 1995. The multiplicity of domains in proteins. Annu. Rev. Biochem. **64:**287–314.
27. **Drozdov, I. G., A. P. Anisimov, S. V. Samoilova, I. N. Yezhov, S. A. Yeremin, A. V. Karlyshev, V. M. Krasilniova, and V. I. Kravchenko.** 1995. Virulent non-capsulate *Yersinia pestis* variants constructed by insertion mutagenesis. J. Med. Microbiol. **42:**264–268.
28. **Du, Y., E. Galyov, and A. Forsberg.** 1995. Genetic analysis of virulence determinants unique to *Yersinia pestis*. Contrib. Microbiol. Immunol. **13:** 321–324.
29. **Dutreix, M., A. Backman, J. Celerier, M. M. Bagdasarian, S. Sommer, A. Bailone, R. Devoret, and M. Bagdasatian.** 1988. Identification of *psiB* genes of plasmids F and R6-5. Molecular basis for *psiB* enhanced expression in plasmid R6-5. Nucleic Acids Res. **16:**10669–10679.
30. **Ferber, D. M., and R. R. Brubaker.** 1981. Plasmids in *Yersinia pestis*. Infect. Immun. **31:**839–841.
31. **Filippov, A. A., N. S. Solodovnikov, L. M. Kookleva, and O. A. Protsenko.** 1990. Plasmid content in *Yersinia pestis* strains of different origin. FEMS Microbiol. Lett. **67:**45–48.
32. **Frey, J., R. Meier, D. Gygi, and J. Nicolet.** 1991. Nucleotide sequence of the hemolysin I gene from *Actinobacillus pleuropneumoniae*. Infect. Immun. **59:** 3026–3032.
33. **Frost, L. S., K. Ippen-Ihler, and R. A. Skurry.** 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. Microbiol. Rev. **58:**162–210.
34. **Gabant, P., A. O. Chahdi, and M. Couturier.** 1994. Nucleotide sequence and replication characteristics of RepHI1B: a replicon specific to the IncHI1 plasmids. Plasmid **31:**111–120.
35. **Galjart, N. J., N. Sivasubramanian, and B. A. Federici.** 1987. Plasmid localization, cloning and sequence of the gene encoding a 27.3-kilodalton cytolytic protein from *Bacillus thuringiensis* subspecies *morrisoni* (PG14). Curr. Microbiol. **16:**171–177.
36. **Galyov, E. E., A. V. Karlyshev, T. V. Chernoskaya, D. A. Dolgikh, O. Y. Smirnov, K. I. Volkovoy, V. M. Abramov, and V. P. Zav'yalov.** 1991. Expression of the envelope antigen F1 of *Yersinia pestis* is mediated by the product of *caf1M* gene having homology with the chaperone protein PapD of *Escherichia coli*. FEBS Lett. **286:**79–82.
37. **Galyov, E. E., O. Y. Smirnov, A. V. Karlishev, K. I. Volkovoy, A. I. Denesyuk, I. V. Nazimov, K. S. Rubtsov, V. M. Dalvadyanz, and V. P. Zav'yalov.** 1990. Nucleotide sequence of the *Yersinia pestis* gene encoding F1 antigen and the primary structure of the protein. Putative T and B cell epitopes. FEBS Lett. **277:**230–232.
38. **Goodrich, J. A., M. L. Schwartz, and W. R. McClure.** 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation

and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18:**4993–5000.
39. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
40. **Hanekamp, T. D. Kobayashi, S. Hayes, and M. M. Stayton.** 1997. Avirulence gene D of *Pseudomonas syringae* pv. tomato may have undergone horizontal gene transfer. FEBS Lett. **415:**40–44.
41. **Harlan, D. M., J. M. Graff, D. J. Stumpo, R. L. Eddy, Jr., T. B. Stows, J. M. Boyle, and P. J. Blackshear.** 1991. The human myristoylated alanine-rich C kinase substrate (MARCKS) gene. Analysis of its gene product, promoter and chromosomal localization. J. Biol. Chem. **266:**14399–14405.
42. **Hayes, F., and S. J. Austin.** 1993. Specificity determinants of the P1 and P7 centromere analogs. Proc. Natl. Acad. Sci. USA **90:**9228–9232.
43. **Hayes, F., and S. J. Austin.** 1994. Topological scanning of the P1 plasmid partition site. J. Mol. Biol. **243:**190–198.
44. **Herzer, P. J., S. Inouye, and M. Inouye.** 1992. Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34bp intergenic sequence. Mol. Microbiol. **6:**345–354.
45. **Hofman, K., and W. Stoffel.** 1993. Tmbase—a database of membrane spanning protein segments. Biol. Chem. **347:**166–172.
45a. **Hu, P., J. Elliott, P. McCready, E. Skowronski, J. Garnes, A. Kobayashi, R. R. Brubaker, and E. Garcia.** 1998. Structural organization of virulence-associated plasmids of *Yersinia pestis.* J. Bacteriol. **180:**5192–5202.
46. **Humphreys, G. O., G. A. Willshaw, and E. S. Anderson.** 1975. A simple method for the preparation of large quantities of pure plasmid DNA. Biochim. Biophys. Acta **383:**457–463.
47. **Jeter, R., J. C. Escalante-Semerena, D. Roof, B. Olivera, and J. Roth.** 1987. Synthesis and use of vitamin B_{12}, p. 551–556. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
48. **Jones, A. J., P. T. Barth, and B. M. Wilkins.** 1992. Zygotic induction of plasmid *ssb* and *psiB* genes following conjugative transfer of IncI1 plasmid ColIb-P9. Mol. Microbiol. **6:**605–613.
49. **Karlyshev, A. V., E. E. Galyov, V. M. Abramov, and V. P. Zav'yalov.** 1992. Caf1R gene and its role in the regulation of capsule formation of *Y. pestis*. FEBS Lett. **305:**37–40.
50. **Koltsova, E. G., Y. G. Suchkov, and S. A. Lebedeva.** 1975. Transmission of a bacteriocinogenic factor in *Pasturella pestis*. Sov. Genet. **7:**507–510.
51. **Krause, M., J. Harwood, J. Fierer, and D. Guiney.** 1991. Genetic analysis of homology between the virulence plasmids of *Salmonella dublin* and *Yersinia pseudotuberculosis*. Infect. Immun. **59:**1860–1863.
52. **Lampson, B. C., S. Inouye, and M. Inouye.** 1991. msDNA of bacteria. Prog. Nucleic Acid Res. Mol. Biol. **40:**1–24.
53. **Lanka, E., and B. M. Wilkins.** 1995. DNA processing reactions in bacterial conjugation. Annu. Rev. Biochem. **64:**141–169.
54. **Lewis, P. J., and J. Errington.** 1997. Direct evidence for active segregation of *oriC* regions on the *Bacillus subtilis* chromosome and co-localization with the SpoOJ partitioning protein. Mol. Microbiol. **25:**945–954.
55. **Lipman, D., S. Altschul, and J. Kececioglu.** 1989. A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. USA **86:**4412–4415.
56. **Ludtke, D. N., and S. J. Austin.** 1987. The plasmid-maintenance functions of the P7 prophage. Plasmid **18:**93–98.
57. **Matsutani, S., H. Ohtsubo, Y. Maeda, and E. Ohtsubo.** 1987. Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196:**445–455.
58. **McDonough, K. A., A. M. Barnes, T. J. Quan, J. Montenieri, and S. Falkow.** 1993. Mutation in the *pla* gene of *Yersinia pestis* alters the course of plague bacillus-flea (Siphonaptera: ceratophyllidae) interaction. J. Med. Entomol. **30:**772–780.
59. **McDonough, K. A., and J. M. Hare.** 1997. Homology with a repeated *Yersinia pestis* DNA sequence IS*100* correlates with pesticin sensitivity in *Yersinia pseudotuberculosis*. J. Bacteriol. **179:**2081–2085.
60. **Mercado-Blanco, J., and J. Olivares.** 1994. A protein involved in stabilization of a large non-symbiotic plasmid of *Rhizobium meliloti* shows homology to eukaryotic cytoskeletal proteins and DNA binding proteins. Gene **139:** 133–134.
61. **Mercado-Blanco, J., and J. Olivares.** 1994. The large non-symbiotic plasmid pRmeGR4a of *Rhizobium meliloti* GR4 encodes a protein involved in replication that has homology with the RepC protein of *Agrobacterium* plasmids. Plasmid **32:**75–79.
62. **Millon, J., H. A. Kirkpatrick, H. L. Kijenski, C. A. Bloch, C. K. Rode, G. F. Mayhew, D. J. Rose, G. Plunkett III, V. Burland, and F. R. Blattner.** Subdivision of the *Escherichia coli* K-12 genome for sequencing: manipulation and DNA sequence of transposable elements introducing unique restriction sites. Submitted for publication.
63. **Moore, R. L., and R. R. Brubaker.** 1975. Hybridization of deoxyribonucleic acid sequences of *Yersinia enterocolitica* and other selected members of *Enterobacteriaceae*. Int. J. Syst. Bacteriol. **25:**336–339.
64. **Mukhopadhyay, G., K. M. Carr, J. M. Kaguni, and D. K. Chattoraj.** 1993. Open-complex formation by the host initiator, DnaA, at the origin of P1 plasmid replication. EMBO J. **12:**4547–4554.
65. **Nagai, S., T. Yagihashi, and A. Ishihama.** 1993. DNA sequence of an allelic variant of the *Actinobacillus pleuropneumoniae*-RTX-toxin I (ApxIA) from serotype 10. Microb. Pathog. **15:**485–495.
66. **Nakai, K., and M. Kanehisa.** 1991. Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins Struct. Funct. Genet. **11:**95–110.
67. **Nakayama, S., T. Kaneko, H. Ishimaru, H. Moriwaki, and K. Mizobuchi.** 1994. Cloning, sequencing and expression of bacteriophage BF23 late genes 24 and 25 encoding tail proteins. J. Bacteriol. **176:**7280–7290.
68. **Olsen, C. H., F. R. Blattner, and D. L. Daniels.** 1991. Simultaneous preparation of up to 768 single-stranded DNAs for use as templates in DNA sequencing. Methods **3:**27–32.
69. **Pal, S. K., R. J. Mason, and D. K. Chattoraj.** 1986. P1 plasmid replication. Role of initiator titration in copy number control. J. Mol. Biol. **20:**275–285.
70. **Perna, N. T., G. F. Mayhew, G. Posfai, S. Elliot, M. S. Donnenberg, J. B. Kaper, and F. R. Blattner.** 1998. Molecular evolution of the pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **66:**3810–3817.
71. **Perry, R. D., M. L. Pendrak, and P. Schuetze.** 1990. Identification and cloning of the hemin storage locus involved in the pigmentation phenotype of *Yersinia pestis*. J. Bacteriol. **172:**5929–5937.
72. **Portnoy, D. A., and R. J. Martinez.** 1985. Role of a plasmid in the pathogenicity of *Yersinia* species. Curr. Top. Microbiol. Immunol. **118:**29–51.
73. **Protsenko, O. A., P. I. Anisimov, O. T. Mosarov, N. P. Donnov, Y. A. Popov, and A. M. Kokushkin.** 1983. Detection and characterization of *Yersinia pestis* plasmids determining pesticin I, fraction I antigen and mouse toxin synthesis. Genetika **19:**1081–1090.
74. **Protsenko, O. A., A. A. Filippov, and V. V. Kutyrev.** 1991. Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into *Yersinia pestis* chromosome. Microb. Pathog. **11:**123–128.
75. **Radnedge, L., M. A. Davis, and S. J. Austin.** 1996. P1 and P7 plasmid partition: ParB protein bound to its partition site makes a separate discriminator contact with the DNA that determines species specificity. EMBO J. **15:**1155–1162.
76. **Rajakumar, K., C. Sasakawa, and B. Adler.** 1997. *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect. Immun. **65:**4606–4614.
77. **Rakin, A., and J. Heesemann.** 1995. Virulence-associated fyuA/irp2 gene cluster of *Yersinia enterocolitica* biotype 1B carries a novel insertion sequence IS*1328*. FEMS Microbiol. Lett. **129:**287–292.
78. **Romantschuk, M., G. Y. Richter, P. Mikhopadhyay, and D. Mills.** 1991. IS*801*, an insertion sequence element isolated from *Pseudomonas syringae* pathovar *phaseolicola*. Mol. Microbiol. **5:**617–622.
79. **Ryan, M. C., R. Tizard, D. R. VanDevanter, and W. G. Carter.** 1994. Cloning of the Lam3 gene encoding the alpha 3 chain of the adhesive ligand epiligrin: expression in wound repair. J. Biol. Chem. **269:**22779–22787.
80. **Samoilova, S. V., L. V. Somoilova, I. N. Yezhov, I. G. Drozdov, and A. P. Anisimov.** 1996. Virulence of pPST+ and pPST− strains of *Yersinia pestis* for guinea pigs. J. Med. Microbiol. **45:**440–444.
81. **Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Peterson.** 1982. Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. **162:** 729–773.
82. **Schauer, D. B., and S. Falkow.** 1993. Attaching and effacing locus of a *Citrobacter freundii* biotype that causes transmissible murine colonic hyperplasia. Infect. Immun. **61:**2486–2492.
83. **Schofield, J. P., D. S. Jones, and J. N. Forrest, Jr.** 1991. Identification of C-type natriuretic peptide in heart of spiny dogfish shark (*Squalus acanthias*). Am. J. Physiol. **261:**F734–F739.
84. **Shapiro, J. A.** 1992. Natural genetic engineering in evolution. Genetika **86:** 99–111.
85. **Shine, J., and L. Dalgarno.** The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
86. **Simonet, M., B. Riot, N. Fortineau, and P. Berche.** 1996. Invasin production by *Yersinia pestis* is abolished by insertion of an IS*200*-like element within the *inv* gene. Infect. Immun. **64:**375–379.
87. **Singer, M. F.** 1995. Unusual reverse transcriptases. J. Biol. Chem. **270:** 24623–24626.
88. **Smith, C. A., M. Pinkney, D. G. Guiney, and C. M. Thomas.** 1993. The ancestral IncP replication system consisted of contiguous *oriV* and *trfA* segments as deduced from a comparison of the nucleotide sequences of diverse IncP plasmids. J. Gen. Microbiol. **139:**1761–1766.
89. **Sodeinde, O. A., and J. D. Goguen.** 1988. Genetic analysis of the 9.5-kilobase virulence plasmid of *Yersinia pestis*. Infect. Immun. **56:**2743–2748.
90. **Sodeinde, O. A., Y. V. B. K. Subrahmanyam, K. Stark, T. Quan, Y. Bao, and J. D. Goguen.** 1992. A surface protease and the invasive character of plague. Science **258:**1004–1007.
91. **Straley, S. C.** 1988. The plasmid-encoded outer-membrane proteins of *Yersinia pestis*. Rev. Infect. Dis. **10:**S323–S326.
92. **Straley, S. C.** 1991. The low-Ca^{+2} response virulence regulon of human-pathogenic yersiniae. Microb. Pathog. **10:**87–91.

93. **Straley, S. C., and R. R. Brubaker.** 1981. Cytoplasmic and membrane proteins of yersiniae cultivated under conditions simulating mammalian intracellular environment. Proc. Natl. Acad. Sci. USA **78:**1224–1228.
94. **Temin, H. M.** 1989. Reverse transcriptases: retrons in bacteria. Nature **339:** 254–255.
95. **Vidal, F., C. Baudoin, C. Miquel, M. F. Galliano, A. M. Christiano, J. Uitto, J. P. Ortonne, and G. Meneguzzi.** 1995. Cloning of the laminin alpha 3 chain gene (LAMA3) and identification of a homozygous deletion in a patient with Herlitz junctional epidermolysis bullosa. Genomics **30:**273–280.
96. **Welkos, S. L., K. M. Davis, L. M. Pitt, P. L. Worsham, and A. M. Friedlander.** 1995. Studies on the contribution of the F1 capsule-associated plasmid pFra to the virulence of *Yersinia pestis*. Contrib. Microbiol. Immunol. **13:**299–305.
97. **Willetts, N., and R. Skurry.** 1987. Structure and function of the F factor and mechanism of conjugation, p. 1,110–1,133. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
98. **Wolf, M. K., and E. C. Boedeker.** 1990. Cloning of the genes for AF/R1 pili from rabbit enteroadherent *Escherichia coli* RDEC-1 and DNA sequence of the major structural subunit. Infect. Immun. **58:**1124–1128.
99. **Zhang, D., and D. J. Miller.** 1995. Characterization of a novel non-muscle myosin-related protein from *Onchocerca gobsoni*. Int. J. Parasitol. **25:**1385–1391.
100. **Zitman, D., and R. Ben-Gurion.** 1972. Transduction of *Pasturella pestis*. Virology **47:**513–516.

Editor: V. A. Fischetti