JOURNAL OF BACTERIOLOGY, Oct. 2004, p. 6560–6574 Vol. 186, No. 19
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.19.6560–6574.2004


# Complete Genomic Sequence of Bacteriophage B3, a Mu-Like Phage of *Pseudomonas aeruginosa*†

Michael D. Braid,[1] Jennifer L. Silhavy,[1] Christopher L. Kitts,[1] Raul J. Cano,[1] and Martha M. Howe[2]*

*Biological Sciences Department and Environmental Biotechnology Institute, California Polytechnic State University, San Luis Obispo, California,[1] and Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, Tennessee[2]*

Received 24 March 2004/Accepted 9 July 2004

**Bacteriophage B3 is a transposable phage of *Pseudomonas aeruginosa*. In this report, we present the complete DNA sequence and annotation of the B3 genome. DNA sequence analysis revealed that the B3 genome is 38,439 bp long with a G+C content of 63.3%. The genome contains 59 proposed open reading frames (ORFs) organized into at least three operons. Of these ORFs, the predicted proteins from 41 ORFs (68%) display significant similarity to other phage or bacterial proteins. Many of the predicted B3 proteins are homologous to those encoded by the early genes and head genes of Mu and Mu-like prophages found in sequenced bacterial genomes. Only two of the predicted B3 tail proteins are homologous to other well-characterized phage tail proteins; however, several Mu-like prophages and transposable phage D3112 encode approximately 10 highly similar proteins in their predicted tail gene regions. Comparison of the B3 genomic organization with that of Mu revealed evidence of multiple genetic rearrangements, the most notable being the inversion of the proposed B3 immunity/early gene region, the loss of Mu-like tail genes, and an extreme leftward shift of the B3 DNA modification gene cluster. These differences illustrate and support the widely held view that tailed phages are genetic mosaics arising by the exchange of functional modules within a diverse genetic pool.**

Bacteriophage B3 is a temperate, transposable phage of *Pseudomonas aeruginosa* that was isolated from a pathogenic strain of *P. aeruginosa* (strain 3) containing four different prophages: A3, B3, C3, and D3 (38). These four phages are serologically dissimilar and display no prophage cross immunity (38). B3 is not induced by UV irradiation and can transduce genes when using *P. aeruginosa* strain 1 (PAO1) as the donor and recipient strain (38). DNA hybridization and recombination studies showed that B3 is somewhat related to other *P. aeruginosa* transposable phages, such as B39 and D3112, and to the transposable *Escherichia coli* phage Mu (2, 49–51). Phages B3, D3112, and B39 also possess similar serological characteristics (3). Like many other transposable phages, B3 integrates its linear genome essentially randomly into the host chromosome during both lytic and lysogenic development, often causing mutations in host genes (9, 50, 90). In addition, B3 and many other transposable phages carry essentially random host DNA fragments on the ends of the packaged genome and have identical terminal 5′-TG dinucleotides (2, 49, 74, 90). Unlike Mu, B3 adsorption requires bacterial pili and surface growth (74). Electron micrographs of B3 show it is morphologically distinct from Mu (24, 84). Both are tailed phages with a regular polygonal head; however, B3 displays a flexible, noncontractile tail (84), while Mu possesses a rigid, contractile tail structure (24). Prior to this work, little was known about the sequence or organization of B3 genes.

Multiple species of the genus *Pseudomonas* are commonly found in the environment. This subset of the gamma-proteobacteria is metabolically and catabolically diverse, and several species are known to play important roles in elemental cycling. *P. aeruginosa* is commonly found in soil and water and is one of the best-characterized members of the genus. *P. aeruginosa* strain PAO1 has been extensively studied and is used as the model strain for *Pseudomonas* genetics (37, 39, 75). The complete genomic sequence of *P. aeruginosa* strain PAO1 was finished in 2000 (85). Several *Pseudomonas* plasmids have also been characterized, including some isolated by selective growth enrichment and others specifically constructed for the degradation of petroleum hydrocarbons and other toxic chemicals (1, 15, 30, 47, 92). Thus, there is significant potential for use of *P. aeruginosa* for bioremediation of contaminated environments.

In addition to its metabolic capability and environmental versatility, *P. aeruginosa* is also studied because of its ability to cause human disease and its resistance to many antibiotics (7, 12, 27, 52). *P. aeruginosa* is an opportunistic pathogen that causes infections at a number of sites, including the urinary tract, respiratory system, and central nervous system (7, 12). The success of *P. aeruginosa* as a pathogen can be attributed to its large number of virulence factors, including those that confer weakened host defenses, resistance to antibiotics, and production of extracellular enzymes and toxins (18, 45, 67, 73). The ongoing analysis of the factors that contribute to *P. aeruginosa* virulence hold promise for the development of better antibiotics and methods for treatment of such infections (26, 79, 80).

In this investigation, the genome of bacteriophage B3, a transposable phage of *P. aeruginosa*, was sequenced, analyzed,

* Corresponding author. Mailing address: Department of Molecular Sciences, University of Tennessee Health Science Center, 858 Madison Ave., Memphis, TN 38163. Phone: (901) 448-8215. Fax: (901) 448-8462. E-mail: mhowe@utmem.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

and compared to the genomic sequences of Mu (68) and multiple Mu-like prophages (10, 20, 32–34, 48, 70–72, 83).

## MATERIALS AND METHODS

**Plasmids, phage, and bacterial strains.** Bacteriophage B3 (ATCC 15692-B1) and the host bacterium, *P. aeruginosa* strain PAO1 (ATCC 15692), were obtained from the American Type Culture Collection (Manassas, Va.). The cloning vector pPCR-Script Amp, a derivative of pBluescript, and the *E. coli* XL1-Blue transformation recipient (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*/F′ *proAB lacI*$^{q}$ Δ*lacZM15* Tn*10* were obtained from Stratagene (La Jolla, Calif.).

**Phage DNA isolation, cloning, and sequencing.** Phage B3 was propagated for DNA extraction using the *P. aeruginosa* PAO1 host and a plate lysate method in which the phage-infected cells are suspended in a soft agar overlay (66). After the agar overlay was harvested and subjected to centrifugation to remove agar and cell debris, the supernatant was treated with DNase and RNase (Sigma, St. Louis, Mo.) and the phage particles were concentrated by precipitation with polyethylene glycol 8000 (Fisher Biotech) and resuspended in SM buffer (66). Phage DNA was then isolated by phenol extraction and ethanol precipitation (76). The phage DNA was sheared, treated with *Bal* 31 exonuclease (Promega, Madison, Wis.), and fractionated on a 0.7% SeaPlaque agarose gel (FMC Corporation, Newark, Del.), and fragments in the 1.6- to 3-kb range were purified using a Qiaex II gel extraction kit (QIAGEN, Valencia, Calif.) (76) and cloned into pPCR-Script Amp using a PCR-Script cloning kit (Stratagene). After *E. coli* XL1-Blue cells (Stratagene) were transformed and plated on Luria-Bertani agar containing ampicillin (50 μg/ml; Fisher Scientific, Hanover Park, Ill.), X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (80 μg/ml; Fisher Scientific), and IPTG (isopropyl-β-D-thiogalactoside) (20 mM; Fisher Scientific), white colonies were purified, cultures were grown, and plasmid DNA was isolated using an UltraClean plasmid mini-prep kit (Mo Bio Laboratories, Solana Beach, Calif.). Plasmid inserts were sequenced from both ends using standard T3 and T7 primers and Dye Terminator chemistry (PE Applied Biosystems, Foster City, Calif.) on ABI Prism 373 and 377 automated DNA sequencers (PE Applied Biosystems). For direct B3 genome sequencing, custom primers were designed from the assembled clone sequences using PrimerSelect software (DNASTAR, Inc., Madison, Wis.), and the sequence was determined using an ABI Prism 377 automated DNA sequencer (PE Applied Biosystems).

**Sequence assembly and annotation.** Genome analysis was performed using the Lasergene software suite (DNASTAR, Inc.). The processed sequences were aligned and edited using SeqMan II and coding regions were predicted using GeneQuest. Where multiple open reading frames (ORFs) were possible, the presence of ribosome-binding sites, codon bias, and genetic overlap were used to help choose the final ORF presented. Searches of public databases for homologous DNA sequences were performed with BLASTN; whereas searches for protein sequence homologues were performed with BLASTP, BLASTX, PSI-BLAST, and TBLASTX (4, 5; http://www.ncbi.nlm.nih.gov/BLAST).

Variable host DNA sequences were identified by using a BLASTN search against the *P. aeruginosa* genome and the entire nonredundant GenBank sequence database. Searches for helix-turn-helix motifs were performed using GYM (69; http://www.cs.fiu.edu/~giri/bioinf/GYM2/prog.html), and searches for potential regulatory protein-binding sites were performed with BIOPROSPECTOR (54; http://robotics.stanford.edu/~xsliu/BioProspector/).

**Nucleotide sequence accession number.** The DNA sequence of the B3 genome was deposited in the GenBank data bank and has been assigned accession number AF232233.

## RESULTS AND DISCUSSION

**Determination of the B3 genome sequence.** The complete DNA sequence of the B3 genome was determined by using a shotgun sequencing approach. Phage B3 DNA fragments (1.6 to 3 kb long) were ligated into pPCR-Script Amp plasmid vector DNA. After transformation of *E. coli*, the plasmids were recovered and their inserts were sequenced from both ends using vector primers and automated DNA sequencers. The B3 sequences were assembled using SeqMan II software (DNASTAR, Inc.), creating a genome scaffold. Direct B3 genome sequencing with 91 custom primers was used to fill in gaps, resolve ambiguities, and validate the assembled genome; upon completion, direct genome sequence data covered approximately 83% of the B3 genome.

A combined total of 655 plasmid and direct genome sequences were used to assemble the complete B3 genome. The average sequence, after end trimming and vector removal, was 578 bases long. Each nucleotide position was sequenced a minimum of two times, in general at least once from each strand. When a nucleotide position was sequenced twice from a single strand, both cloned DNA and direct genome sequencing were used. Only 1.8% (720 bp) of the complete genome was represented by single-strand coverage. On average, each nucleotide position was sequenced 9.85 times, with 336 sequences representing the upper DNA strand and 319 sequences covering the lower strand. The linear genome is composed of 38,439 bp (GenBank accession number AF232233) with a G+C content of 63.3%. This high G+C content correlates with that of *P. aeruginosa* strain PAO1, which displays a G+C content of 66.6% in most predicted coding regions (85).

**Characterization of the ends of the B3 genome and attached host DNA.** Assembly of multiple right-end sequences from plasmid clones revealed a position-specific divergence in sequence homology that defined the right end of the B3 genome. Since only one clone contained B3 left-end DNA, this approach could not be used to define the precise genome left end. Therefore, direct genome sequencing of B3 DNA isolated from phage particles was performed with a primer walk-out strategy analogous to that used previously for Mu (44). The resulting sequences showed a loss of specific base identification immediately 5′ to terminal 5′-TG dinucleotides at both ends of the B3 genome (data not shown). We interpret this loss of base identity to be due to host DNA attached to the terminal 5′-TG dinucleotides, since all bases would be represented in nearly equal proportions if random host DNA segments were present at the ends of the B3 genome. These results confirmed previous findings for B3 (50, 74).

The presence of *Pseudomonas* DNA on both ends of the packaged B3 genome was confirmed by performing a search for nucleotide sequence homologues by BLASTN analysis (5) of B3 left- and right-end clones. Sequences of the putative host DNA in the one left-end clone and three right-end clones displayed nearly perfect homology with different regions of the PAO1 genome, as expected for a transposable phage. In all cases, host homology ceased precisely at the 5′-TG B3 end. Nine clones lacking B3 homology contained sequences homologous to different regions of the PAO1 genome (data not shown). These clones were most likely derived by cloning of host DNA from the ends of B3 phage particle DNA or from particles containing only transducing DNA.

**BLASTN analysis of the B3 genome.** The B3 genome was screened for nucleotide sequence homology with bacterial, bacteriophage, and viral DNA sequences using BLASTN analysis (5). Significant nucleotide homology was found for only two regions of the B3 genome. The first region reflected strong homology (expect value of 5e-24; 83% identity over 178 bp) between a region of bacteriophage ϕE125 (93) and ORF48 of B3. As will be discussed later, these homologous regions comprise about one-quarter of the *dam* DNA modification genes of both phages.

The second region consisted of the last 111 bp (bp 38328 to 38439) of the B3 genome, which displayed strong homology

(expect value of 2e-41) with a sequence containing the PAO1 cyanide-insensitive oxidase A and B operon (*cioAB*; GenBank accession number Y10528; 19), but no homology with the sequenced PAO1 genome (85). Analysis of this 2,925-bp *cioAB* sequence by BLASTN against the PAO1 genome revealed strong homology with *cioAB* bp 1 to 2812 (corresponding to bp 4403831 to 4406643 in the PAO1 genome; 19, 85). However, the remaining 113 bases of the *cioAB* sequence shared no similarity with the PAO1 genome sequence but displayed nearly perfect homology with the right end of the B3 genome (106 of 113 nucleotides [nt]). This region of B3/*cioAB* homology lies 59 bp downstream of the termination codon of the *cioB* gene, with an orientation such that the right end of the B3 genome is closest to *cioB*. The simplest explanation is that there is a B3 or B3-like prophage or a cryptic B3-like sequence downstream of *cioB* in the PAO1 substrain (PAO6049) from which the *cioAB* sequence was obtained (19).

**Assignment of probable B3 genes.** Putative ORFs within the B3 genome were located using GeneQuest software (DNASTAR, Inc.) and validated by multiple methods. The predicted protein sequences were used for BLASTP analysis (5) to identify ORFs displaying significant homology with other bacteriophage or bacterial protein sequences. Segments of B3 nucleotide sequence were also subjected to BLASTX analysis (5) to screen for possible homologues in other B3 reading frames. Sequence upstream and in the early portion of each predicted ORF was examined manually for the presence of ribosome-binding site sequences (e.g., GGAGG) (57) within 4 to 14 bases upstream of potential start codons. The sequence was also examined for the close proximity of start codons to stop codons, as is frequently observed in prokaryotic operons (57). When multiple possible ORFs were present, codon usage was used to aid the choice of the most likely coding region. As observed for many organisms with high G+C DNA (16), most B3 ORFs show a strong preference for G+C in the first and third base positions, with the second position displaying a preference for A+T (Table 1). Interestingly, the 1.5-kb region from 10.0 to 11.5 kb on the B3 genome exhibits less striking preference and has a lower G+C content, 53.9% instead of 63.3%, which may reflect its relatively recent incorporation into the genome.

The above analysis identified 59 putative B3 protein-coding regions whose salient characteristics are given in Table 1. The ORFs are numbered 1 to 58 starting at the left end of the phage genome (and include a second nested protein-coding region ORF36Z within ORF36). In 13 cases where start site features were poor, alternative possible start sites are also given; they are designated with ′ and ″ symbols in Table 1. Six ORFs are predicted to encode very short polypeptides with fewer than 50 amino acids, and seven others would encode polypeptides only 50 to 90 amino acids long. Four potential translation initiation codons were identified: AUG (43 examples), GUG (10 examples), CUG (4 examples), and UUG (1 example). Many of the non-AUG start codons are located in the putative late gene region (ORF24 to ORF58) and may reflect the use of rare initiation codons as a translational regulatory mechanism for some B3 genes. Since 6 of the 13 short ORFs begin with non-AUG start codons, it is also possible that some of these ORFs may not encode protein products at all but may be nonfunctional gene remnants generated during the evolutionary process.

**Detection of homologues and prediction of gene function.** With the availability of genome sequences for many characterized bacteriophages, one fruitful approach for predicting protein function is the detection among homologous proteins some whose functions are already known. To take advantage of this approach, we performed homology searches with BLASTP and PSI-BLAST algorithms (4, 5) against protein sequences in the protein database at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). BLASTP and the first round of PSI-BLAST perform similar searches and were run with and without a low-complexity filter, respectively. In most cases, the homologues identified and scores obtained were similar (see Table S2 in the supplemental material). The second iteration of PSI-BLAST incorporates sequence information from the strong homologues identified in the first round for a second search, with the consequence that scores for the first homologues usually increase and new homologues are detected (4). The results showed that many B3 genes are similar to genes in Mu and Mu-like prophages (defined as prophages that encode transposase homologues and a substantial number of other Mu protein homologues in the early, lysis, and head gene regions) (see Table S2 in the supplemental material). The identity of the highest scoring homologue and the homologue most useful in predicting the gene's function are shown in Table 1, with their scores presented as expect values, which roughly indicate the likelihood of the observed match occurring at random. For B3 genes with homologues of known function, we were able to assign predicted functions, which are shown in Fig. 1.

**(i) Immunity and early genes.** Predicted ORFs 1 to 15 appear to constitute a single operon which is transcribed leftward from a regulatory region located between ORF15 and ORF16. On the basis of their similarity to proteins involved in transposition, B3 ORF11 and ORF12 are predicted to encode the two transposase subunits analogous to the B and A proteins of Mu, respectively (Table 1 and Fig. 1; see Table S2 in the supplemental material). Likewise, B3 ORF1 and ORF2 are homologues of the Mu regulatory proteins Mor and GemA. These similarities lead us to propose that ORFs 1 to 15 constitute the primary B3 early operon. The only other early operon protein for which a function can be proposed is that of ORF14, which exhibits similarity to two proposed transcriptional regulators (Table 1; see also Table S2 in the supplemental material). On the basis of its location upstream of the transposase genes, ORF14 may encode the B3 lytic phase repressor analogous to the Mu Ner protein responsible for down-regulating early transcription and repressor transcription at late times during the lytic cycle (89). Of the remaining B3 early operon ORFs, five had no detectable homologues (ORF3, -4, -7, -9, and -10) and the remaining five had bacterial (ORF5 and -8) or prophage (ORF5, -6, -13, and -15) homologues of unknown function (see Table S2 in the supplemental material). The early operon of phage Mu also contains a number of relatively small genes of as yet unknown function (Fig. 1).

The untranslated region between ORF15 and ORF16 contains a typical bacterial promoter, with a −35 hexamer (TTG CCA) separated by 16 nt from a −10 hexamer (TTTGTT)

TABLE 1. Proposed B3 coding regions

| Gene | Start position[a] | RBS[b] | Start codon | Size (aa)[c] | ORF codon usage (avg G+C %)[d] | | | Representative homologue(s)[e] | BLASTP expect value[f] | PSI-BLAST2 expect value[g] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Base 1 | Base 2 | Base 3 | | | |
| Left[h] | | | | | | | | | | |
| ORF1 | 486 | GGAGG (5) | ATG | 120 | 65.0 | 35.9 | 70.8 | VioMu transcriptional regulator CV2165, AAQ59838 | 1e-04 | 8e-38 |
| | | | | | | | | Mu transcription activator Mor/E17, VMOR_BPMU | 0.91 | 3e-08 |
| ORF2 | 1052 | GGAGG (6) | ATG | 187 | 70.6 | 39.1 | 82.9 | Pnm2* hypothetical protein NMA1301, CAB84553 | 6e-12 | 2e-48 |
| | | | | | | | | Mu protein Gem/E16/gp16, VG16_BPMU | 0.001 | 2e-19 |
| ORF3 | 1506 | GAGG (9) | ATG | 155 | 71.0 | 47.7 | 75.5 | — | — | — |
| ORF4 | 1966 | GAAG (6) | ATG | 152 | 69.0 | 42.8 | 64.5 | — | — | — |
| ORF4′[i] | 1888 | AGGGGG (6) | GTG | 126 | 72.2 | 34.9 | 60.3 | ND | ND | ND |
| ORF5 | 2534 | AAGAAG (9) | ATG | 214 | 67.3 | 45.8 | 82.8 | Conserved protein RL111 in PAPI-1, AAP84235 | 4e-33 | 1e-78 |
| | | | | | | | | Sti3 hypothetical protein STY1595, AH0683 | 5e-24 | 5e-68 |
| ORF5′ | 2579 | GGGGG (6) | ATG | 229 | 66.4 | 43.6 | 83.0 | ND | ND | ND |
| ORF6 | 3204 | AAGAGG (9) | ATG | 207 | 67.6 | 38.2 | 78.2 | DucMul conserved protein HD0097, AAP95100 | 1e-56 | 9e-90 |
| ORF7 | 3397 | GGAGG (9) | ATG | 66 | 57.6 | 44.0 | 74.3 | — | — | — |
| ORF8 | 3764 | GGAGG (8) | ATG | 124 | 69.4 | 48.4 | 75.0 | Hypothetical protein RL112 in PAPI-1, AAP84236 | 3e-08 | 1e-36 |
| ORF8′ | 3824 | GGAG (4) | GTG | 144 | 70.8 | 51.4 | 71.5 | ND | ND | ND |
| ORF8′′ | 3983 | ATGAG (8) | CTG | 197 | 76.1 | 54.3 | 58.9 | ND | ND | ND |
| ORF9 | 4045 | AGGAG (7) | ATG | 93 | 63.5 | 31.2 | 88.2 | — | — | — |
| ORF10 | 4383 | AGGGGG (7) | ATG | 113 | 70.8 | 46.0 | 75.2 | — | — | — |
| ORF11 | 5557 | AGGAGG (6) | ATG | 390 | 67.7 | 39.0 | 79.2 | VioMu conserved protein CV2157, AAQ59830 | e-119 | e-139 |
| | | | | | | | | General secretion pathway ExeA, GSPA_AERHY | 4e-07 | 1e-71 |
| | | | | | | | | Mu transposase protein B, VPB_BPMU | 0.004 | 0.001 |
| | | | | | | | | Tn*552* transposase subunit TnpB, S11779 | — | 0.019 |
| ORF12 | 7338 | GGAG (6) | ATG | 593 | 66.0 | 44.5 | 79.6 | Sti3 hypothetical protein STY1604, AI0684 | 0.0 | 0.0 |
| | | | | | | | | RadMu transposase DRA0075, E75601 | 1e-05 | 3e-79 |
| | | | | | | | | Tn*552* transposase subunit TnpA, S11780 | — | 2e-20 |
| ORF13 | 8267 | AGGGG (5) | ATG | 308 | 72.1 | 39.3 | 75.0 | BorMu NMB1002-like protein BB3646, CAE35619 | 9e-71 | e-105 |
| ORF14 | 8586 | GGAG (5) | ATG | 102 | 63.8 | 43.2 | 80.4 | BorMu NMB1003-like protein BB3645, CAE35618 | 7e-23 | 1e-31 |
| | | | | | | | | Transcription regulator AGR_L_810, G98181 | 7e-04 | 2e-16 |
| ORF15 | 9064 | GGAG (8) | ATG | 161 | 66.4 | 44.7 | 81.4 | VioMu hypothetical protein CV2153, AAQ59826 | 5e-36 | 2e-77 |
| ORF16 | 9683 | GAGG (9) | ATG | 163 | 74.9 | 56.5 | 64.4 | — | — | — |
| ORF17 | 9904 | AGGAG (7) | ATG | 69 | 65.2 | 43.5 | 68.1 | NeisMu2* hypothetical protein NMB1004, B81133 | 3e-13 | 2e-22 |
| | | | | | | | | DucMu3 hypothetical protein HD1538, AAP96324 | 1e-12 | 2e-18 |
| Right[j] | | | | | | | | | | |
| ORF18 | 9990 | AGGCGG (5) | GTG | 136 | 62.5 | 52.9 | 51.4 | RS6 transcription repressor RSc1907, CAD15609 | 3e-12 | 5e-29 |
| | | | | | | | | Phage P22 c2 repressor, CAA60873 | 0.18 | 3e-08 |
| ORF18′ | 10167 | GAAGG (7) | GTG | 77 | 58.5 | 53.3 | 50.7 | ND | ND | ND |
| *ORF19*[k] | *10030* | *GAGG (6)* | *GTG* | *44* | *56.8* | *52.3* | *70.5* | — | — | — |
| ORF20 | 10406 | TAAGGAG (7) | GTG | 195 | 56.4 | 40.5 | 55.9 | — | — | — |
| ORF21 | 11022 | AGAAGG (6) | ATG | 36 | 44.4 | 44.5 | 66.7 | — | — | — |
| *ORF22* | *11138* | *GTAGG (4)* | *TTG* | *128* | *64.0* | *43.8* | *53.9* | — | — | — |
| ORF22′ | 11144 | GTAGG (10) | TTG | 126 | 64.2 | 44.5 | 53.2 | ND | ND | ND |
| ORF22′′ | 11177 | AGGATG (8) | TTG | 115 | 64.3 | 44.4 | 53.0 | ND | ND | ND |
| ORF23 | 11538 | AGGGAG (5) | ATG | 21 | 66.6 | 42.9 | 71.4 | — | — | — |
| ORF24 | 11682 | GGAG (7) | ATG | 98 | 66.3 | 49.0 | 85.7 | BorMu membrane protein BB3639, CAE35612 | 1e-12 | 2e-20 |
| | | | | | | | | BcepMu holin protein BcepMu21; AAS47861 | 1e-04 | 3e-15 |

*Continued on following page*

TABLE 1—*Continued*

| Gene | Start position[a] | RBS[b] | Start codon | Size (aa)[c] | ORF codon usage (avg G+C %)[d] | | | Representative homologue(s)[e] | BLASTP expect value[f] | PSI-BLAST2 expect value[g] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Base 1 | Base 2 | Base 3 | | | |
| ORF25 | 11918 | GGAGG (10) | CTG | 264 | 68.2 | 53.4 | 76.5 | BorMu lytic transglycosylase BB3638, CAE35611 | 3e-68 | ND |
| ORF26 | 12709 | GAGG (8) | GTG | 70 | 62.9 | 52.8 | 67.2 | — | — | — |
| ORF27 | 12908 | GGAGG (4) | ATG | 203 | 74.4 | 50.7 | 75.9 | Sti3 probable lipoprotein STY1611, AH0685 | 2e-13 | 1e-62 |
| | | | | | | | | BcepMu Rz, Rz1 proteins BcepMu23, -24, AAS47863, -4 | 0.007 | 1e-41 |
| ORF28 | 13523 | AAGGGG (7) | ATG | 129 | 67.4 | 34.9 | 79.1 | Stm7* inner membrane protein STM4216, AAL23040 | 9e-22 | 1e-31 |
| | | | | | | | | Sti3 membrane protein STY1612, AI0685 | 2e-18 | 4e-34 |
| ORF29 | 13902 | GGAG (5) | ATG | 102 | 63.8 | 44.1 | 83.3 | Sti3 hypothetical protein STY1613, AB0686 | 4e-32 | 5e-40 |
| | | | | | | | | Phage Mu gp26, VG26_BPMU | 12 | — |
| ORF30 | 14207 | AAGAAG (4) | GTG | 193 | 70.0 | 47.1 | 84.4 | VioMu conserved protein CV2141, AAQ59814 | 4e-39 | 2e-52 |
| | | | | | | | | Phage Mu gp27, VG27_BPMU | 9e-06 | 3e-09 |
| ORF30′ | 14216 | AAGAAG (14) | ATG | 190 | 70.5 | 46.8 | 84.2 | ND | ND | ND |
| ORF31 | 14788 | GGGG (8) | ATG | 486 | 62.9 | 42.6 | 82.9 | RadMu conserved protein DRA0094, H75603 | 2e-76 | e-140 |
| | | | | | | | | Mu large terminase subunit[l] gp28, VG28_BPMU | 3e-28 | e-148 |
| ORF32 | 16248 | GGAGG (7) | ATG | 488 | 68.3 | 43.9 | 81.4 | BorMu gp29-like protein BB3631, CAE35604 | e-153 | 0.0 |
| | | | | | | | | Phage Mu gp29, VG29_BPMU | 2e-10 | e-137 |
| ORF33 | 17714 | GGAG (7) | ATG | 418 | 68.9 | 46.9 | 78.3 | DucMu3 gp30-like protein HD1557, AAP96341 | 9e-53 | 3e-83 |
| | | | | | | | | Mu F protein gp30, VPF_BPMU | 2e-29 | e-128 |
| *ORF34* | *18980* | *GAG (7)* | *GTG* | *32* | *59.4* | *62.5* | *81.2* | — | — | — |
| ORF35 | 19094 | GGAGG (4) | ATG | 190 | 68.9 | 46.8 | 75.8 | DucMu3 G-like protein HD1558, AAP96342 | 5e-30 | 2e-60 |
| | | | | | | | | Mu G protein, VPG_BPMU | 0.002 | 3e-17 |
| ORF36 | 19857 | GAAAAG (3) | CTG | 423 | 73.0 | 47.7 | 79.7 | VioMu conserved protein CV2135, AAQ59808 | 8e-58 | e-120 |
| | | | | | | | | BcepMu protease protein BcepMu32, AAS47872 | 1e-45 | e-120 |
| ORF36′ | 20097 | GGAGG (11) | ATG | 343 | 72.9 | 47.3 | 78.7 | ND | ND | ND |
| ORF36Z[m] | 20664 | AGCAGG (8) | ATG | 154 | 76.6 | 42.2 | 79.2 | DucMu3 conserved protein HD1559, AAP96343 | 2e-11 | 3e-36 |
| | | | | | | | | BcepMu scaffold protein BcepMu33, AAS91496 | 7e-05 | 5e-26 |
| ORF37 | 21132 | AAGGGAG (11) | ATG | 125 | 68.8 | 57.6 | 79.2 | DucMu3 hypothetical protein HD1560, AAP96344 | 7e-17 | 3e-49 |
| ORF38 | 21521 | AAGGAG (5) | ATG | 309 | 63.7 | 46.3 | 84.2 | DucMu3 hypothetical protein HD1561, AAP96345 | e-105 | e-144 |
| | | | | | | | | BcepMu capsid protein BcepMu34, AAS47873 | 2e-17 | 1e-75 |
| ORF39 | 22497 | GGAG (8) | ATG | 209 | 77.1 | 45.0 | 76.1 | — | — | — |
| ORF40 | 23126 | GGAG (7) | GTG | 172 | 72.1 | 45.9 | 77.4 | DucMu3 hypothetical protein HD1562, AAP96346 | 4e-31 | 7e-68 |
| | | | | | | | | Mu gp36, VG36_BPMU | 0.079 | 2e-15 |
| ORF41 | 23646 | GAAGG (9) | ATG | 154 | 70.8 | 45.4 | 74.1 | DucMu3 hypothetical protein HD1563, AAP96347 | 1e-14 | 5e-44 |
| ORF42 | 24107 | GGAG (5) | ATG | 68 | 61.8 | 41.2 | 83.8 | DucMu3 hypothetical protein HD1564, AAP96348 | 0.17 | ND |
| ORF43 | 24317 | AAGGAG (5) | ATG | 250 | 57.6 | 44.4 | 85.6 | RS1 conserved protein RSc0863, CAD14565 | 5e-36 | 2e-95 |
| | | | | | | | | Sti3 tail sheath protein STY1625, AF0687 | — | 0.34 |
| ORF44 | 25072 | AAGGCCG (7) | ATG | 168 | 72.6 | 43.5 | 77.4 | DucMu1 hypothetical protein HD0137, AAP95133 | 2e-11 | 9e-56 |
| ORF45 | 25596 | ACGGA (11) | CTG | 43 | 72.1 | 28.0 | 67.5 | — | — | — |
| ORF45′ | 25575 | — | ATG | 50 | 70.0 | 28.0 | 68.0 | ND | ND | ND |
| ORF45′′ | 25428 | AAGGA (10) | CTG | 99 | 72.7 | 47.5 | 61.6 | ND | ND | ND |
| *ORF46* | *25741* | *AAGGAAG (8)* | *GTG* | *26* | *73.1* | *46.2* | *84.7* | — | — | — |
| ORF47 | 25863 | AAGGTG (8) | GTG | 61 | 62.3 | 62.3 | 70.5 | XccP1 small conserved protein XCC2968, AAM42240 | 2e-11 | 1e-21 |
| | | | | | | | | Mu Com translational regulator, VCOM_BPMU | 0.009 | 0.028 |
| ORF47′ | 25818 | — | GTG | 76 | 59.2 | 60.5 | 71.0 | ND | ND | ND |

*Continued on following page*

TABLE 1—*Continued*

| Gene | Start position[a] | RBS[b] | Start codon | Size (aa)[c] | ORF codon usage (avg G+C %)[d] | | | Representative homologue(s)[e] | BLASTP expect value[f] | PSI-BLAST2 expect value[g] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Base 1 | Base 2 | Base 3 | | | |
| ORF47″ | 25851 | AGCAGA (9) | GTG | 65 | 61.5 | 61.5 | 70.8 | ND | ND | ND |
| ORF48 | 26018 | GGAG (4) | ATG | 264 | 63.2 | 42.5 | 78.8 | BorMu modification methylase BB3607, CAE35580 | e-122 | ND |
| | | | | | | | | Phage ϕE125 gp27 methylase, AAL40300 | e-114 | ND |
| ORF49 | 26894 | — | ATG | 1,099 | 71.2 | 49.0 | 74.3 | DucMu3 tail-length tape measure HD1569, AAP96352 | 1e-23 | — |
| | | | | | | | | Phage Mu gp42 tail-length tape measure, VG42_BPMU | — | 7e-10 |
| ORF49′ | 26921 | AAGGA (11) | CTG | 1,090 | 71.2 | 49.1 | 74.3 | ND | ND | ND |
| ORF50 | 30136 | GCAGG (7) | CTG | 337 | 62.0 | 50.0 | 80.8 | DucMul hypothetical protein HD0145, AAP95140 | 1e-37 | e-155 |
| ORF50′ | 30193 | GGA (6) | ATG | 318 | 62.0 | 50.0 | 80.8 | ND | ND | ND |
| ORF51 | 31151 | AGGAAG (7) | ATG | 307 | 59.0 | 49.6 | 76.5 | DucMul hypothetical protein HD0146, AAP95141 | 6e-35 | e-137 |
| ORF52 | 32077 | GGTGG (10) | ATG | 568 | 68.2 | 48.3 | 81.5 | DucMul conserved protein HD0148, AAP95143 | 4e-83 | 0.0 |
| ORF52′ | 32497 | AGGAG (11) | GTG | 428 | 69.7 | 48.4 | 81.6 | ND | ND | ND |
| ORF52″ | 32656 | AGGAG (11) | TTG | 375 | 69.9 | 48.8 | 82.1 | ND | ND | ND |
| ORF53 | 33770 | AGGAG (5) | ATG | 272 | 60.7 | 51.1 | 77.2 | DucMul hypothetical protein HD0150, AAP95145 | 4e-35 | 3e-87 |
| ORF54 | 34598 | AGGAG (5) | ATG | 76 | 56.5 | 43.4 | 79.0 | DucMul hypothetical protein HD0151, AAP95146 | 2e-08 | 3e-24 |
| ORF55 | 34600 | AGGAG (5) | GTG | 151 | 74.2 | 57.0 | 59.0 | XfP6* hypothetical protein XF2116, E82599 | 2e-10 | 3e-33 |
| ORF55′ | 34825 | AAGGGG (7) | GTG | 76 | 69.7 | 56.5 | 73.7 | ND | ND | ND |
| ORF56 | 35042 | GGAGG (6) | ATG | 736 | 64.6 | 49.0 | 80.3 | DucMu3 conserved protein HD1580, AAP96361 | e-132 | 0.0 |
| ORF57 | 37249 | GGAGG (9) | ATG | 266 | 60.5 | 53.8 | 70.3 | — | — | — |
| ORF58 | 38128 | AGGGATG (5) | ATG | 74 | 67.5 | 33.8 | 79.7 | — | — | — |

[a] First nucleotide of proposed initiation codon.
[b] Identity of bases between first and last matches to the ribosome-binding site (RBS) sequence UAAGGAGGUGA (57). The number of base pairs between position G in the above RBS sequence and base 1 of the initiation codon is shown in parentheses.
[c] aa, amino acids.
[d] ORF codon usage, measured by the average G+C content (as a percentage), at base positions 1, 2, and 3 of the codon.
[e] In general, the homologues listed include the one with the greatest similarity and the one most useful for predicting the function of the B3 ORF. Homologue descriptions include the name of the phage or phage-related element (Table 2), the protein function if known, the ORF number or name, and an accession number, in that order. Homologue descriptions for proteins that are not phage associated begin with the protein function. —, no homologue with a score better than 5.0 was detected; ND, not determined.
[f] BLASTP values were generated at NCBI (http://www.ncbi.nlm.nih.gov) with the low-complexity filter applied. —, no homologue with a score better than 5.0 was detected; ND, not determined.
[g] PSI-BLAST expect values for iteration 2 were generated with the low-complexity filter turned off and with homologues scoring 0.005 or better in iteration 1. —, no homologue with a score better than 5.0 was detected; ND, not determined.
[h] Genes transcribed leftward on map.
[i] Alternate start sites are indicated by ′ and ″.
[j] Genes transcribed rightward on map.
[k] ORFs deemed least likely to function are shown in italics.
[l] In the sequence databases (accession numbers AF083977 and NC_000929), Mu gp28 is mistakenly identified as a portal protein; gp28 is the large subunit of the Mu terminase (69, 83).
[m] ORF36Z is entirely contained within ORF36 in the same reading frame and has the same function as Mu Z.

(Fig. 2). The high similarity of this promoter to consensus promoter sequences (−35 TTGACA, 16- to 18-nt spacer, and −10 TATAAT [29]) suggests that it will be recognized directly by the host RNA polymerase and efficiently drive leftward transcription of the early operon.

The untranslated region between ORF17 and ORF18 should represent a second important regulatory region controlling both leftward transcription of ORF16 and ORF17 and rightward transcription beginning with ORF18 (Fig. 1). This region contains another good candidate bacterial promoter, with a −35 hexamer (TTGTGT) located 17 nt upstream of a −10 hexamer (AATATT) (Fig. 2). This promoter is oriented to initiate rightward transcription. Examination of this region for a promoter driving leftward transcription revealed only two relatively poor candidates, one containing a −35 sequence (TTCACC) located 14 nt upstream of a −10 sequence (CAT CTT) and a second containing a −35 hexamer (TTGACA) located 20 nt upstream of a −10 hexamer (TATTCA) (Fig. 2). Whereas the hexamer sequences in these candidates are good, the spacer distances are shorter and longer, respectively, than those of a typical good bacterial promoter (29, 65). Nevertheless, direct comparison of the effect of spacer length on promoter activity in *E. coli* and *P. aeruginosa* revealed that longer spacer lengths are tolerated in *P. aeruginosa* (65), making the

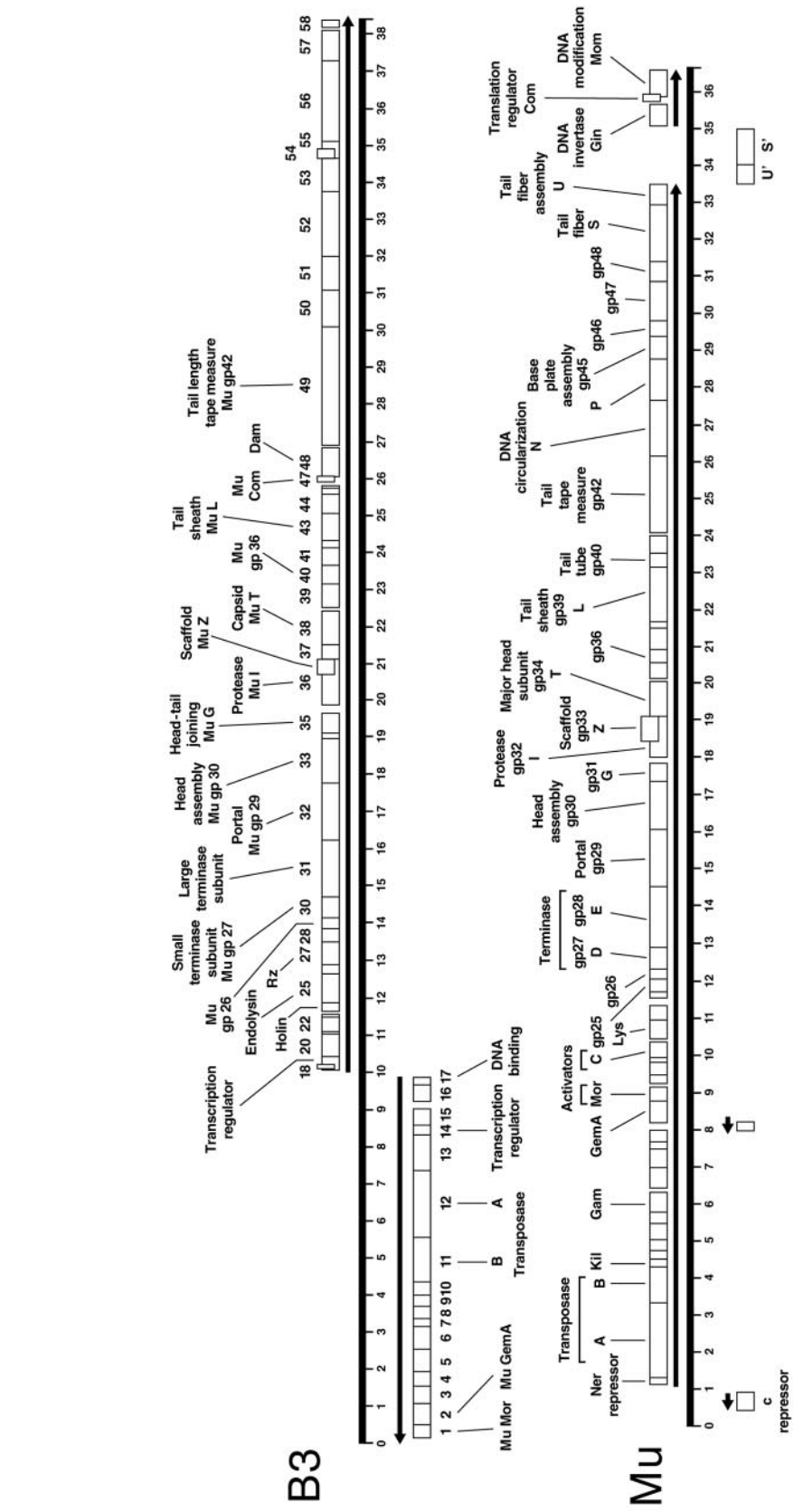


FIG. 1. Comparison of the genome organization for phages B3 and Mu. The genomes of B3 and Mu are shown as thick lines and are drawn approximately to scale, with tick marks at 1-kb intervals. ORFs are shown as boxes; their positions were derived from GenBank accession numbers AF083977 and NC_000929 for the Mu genome and this work for B3 (GenBank accession number AF232233). The direction of transcription is indicated by an arrow; in addition, genes transcribed rightward are shown above the line, and those transcribed leftward are positioned below the line. Relevant Mu genes are identified by their gene names or sequential genome gp numbers, and functions if known. The predicted B3 gene functions shown were deduced from the known functions of homologous genes in Mu or other phages; for B3 genes with Mu homologues, the identity of the homologous Mu gene is also given.


FIG. 2. Potential regulatory elements in untranslated regions. The sequences for both DNA strands in relevant predicted untranslated regions are shown. The beginning or end of the flanking ORFs are shown by the large brackets; the ORFs continue in the directions of the broken lines. Potential −10 and −35 hexamers are indicated by bullets, which are pointed in the direction of transcription.

second promoter a plausible candidate for driving transcription of ORF16 and ORF17.

One would expect these two untranslated regions to contain binding sites for a B3 repressor protein that would keep early transcription turned off in a lysogen. Searches with BIOPROSPECTOR (54) identified multiple related sequences that overlapped all three promoters and, thus, might serve as repressor-binding sites, but we were unable to develop a single convincing consensus sequence from them. Nevertheless, there are two candidate ORFs that could encode the B3 immunity repressor protein, ORF17 and ORF18. The predicted protein encoded by ORF18 exhibits amino acid sequence similarity to the phage P22 c2 repressor and a long list of predicted phage repressor proteins (Table 1; see Table S2 in the supplemental material). The putative ORF17 protein, while not homologous to known regulators, also contains a potential helix-turn-helix DNA-binding motif (data not shown). The genes for both proteins lie close to the ORF17-ORF18 and ORF15-ORF16 intergenic regions where the immunity repressor is predicted to bind, a location consistent with the typically modular organization of phage genes and regulatory elements (8, 35). They are also just downstream of bacterial promoters and, thus, would be transcribed early after infection when repressor protein synthesis is needed to establish lysogeny. The remaining ORFs in this region, ORF13, -15, -16, -20, -21, and -22 have no currently identifiable function. On the basis of their locations in predicted early transcripts, the proteins produced by these ORFs could function as (i) accessory regulators of repressor synthesis, (ii) positive regulators for subsequent phases of transcription, (iii) auxiliary factors regulating or involved in replication and transposition, and (iv) factors stimulating or inhibiting host functions affecting lysogenic or lytic development.

**(ii) Middle and late genes.** It is well-known that phages and other viruses typically possess mechanisms that confer sequential expression of their genes. For example, genes for genome replication are usually expressed early in the lytic cycle, and genes for phage particle morphogenesis and cell lysis are expressed late. Phage Mu transcription occurs in three phases: early, middle, and late (60). Mu middle and late transcripts are activated by the Mor and C proteins, respectively (59, 63).

The identification of typical bacterial promoter sequences in the ORF15-ORF16 and ORF17-ORF18 intergenic regions makes it likely that transcription of ORF1 through ORF15, ORF16 and ORF17, and ORF18 through ORF23 occurs early during the lytic cycle and is catalyzed by the host RNA polymerase. Since the leftward promoter for transcription of ORF16 and ORF17 may be inefficient due to its long 20-nt spacer, it is possible that a positive regulator produced from one of these transcripts might stimulate transcription from that region, making ORF16 and ORF17 the B3 middle genes. It is equally possible that ORF16 and ORF17 are simply early genes and that B3 has only two phases of transcription, early and late.

The first identifiable B3 late genes are ORF24, ORF25, and ORF27. The predicted ORF25 protein has strong amino acid sequence similarity to more than 50 soluble, predominantly bacterial, lytic transglycosylases (see Table S2 in the supplemental material), and therefore, likely encodes a B3 endolysin that participates in host cell lysis. With the recent release of the BcepMu sequence (GenBank accession number AY539836), we can also identify nearby ORF24 and ORF27, which encode potential membrane or membrane-anchored proteins, as homologues of the predicted BcepMu holin and Rz homologues, proteins predicted to play roles in host cell lysis analogous to those of the well-studied auxiliary lysis proteins of phage lambda (6).

Most of the ORFs from ORF28 through ORF38 are the head genes of B3; they display significant amino acid sequence similarity and gene order with the known head genes of Mu (22, 25, 68) (Table 1 and Fig. 1; see Table S2 in the supplemental material). In particular, ORF30 and ORF31 are homologous to the small and large subunits of the Mu terminase (68, 82) and are followed by the portal and head assembly genes ORF32 and ORF33, respectively. In Mu, the head assembly gene is followed by the Mu G gene (gp31), and ORF35 of B3 is a Mu G homologue. Although the precise role of Mu G protein is not known, it is proposed to be involved in head-tail joining (25). In many tailed phages, including Mu, the

overlapping genes for the protease and scaffolding proteins and the gene encoding the major capsid protein follow the above genes (68). On the basis of similarity to BcepMu proteins (86), we can also identify ORF36 as the gene encoding the B3 protease and nested scaffolding protein (designated ORF36Z) and ORF38 as encoding the B3 major head protein, the capsid protein. Thus, the order of B3 head genes roughly parallels that in Mu and other tailed phages.

In many phage genomes, a cluster of tail genes follows the head genes, with the tail-length tape measure gene being preceded by the tail sheath and tail tube genes (14, 68). Phage B3 has the two-gene DNA modification gene cluster located immediately upstream of the predicted B3 tail-length tape measure gene, ORF49; however, ORF43 exhibits some similarity to the tail sheath protein of phage Sti3 (Table 1; see Table S2 in the supplemental material), and other ORF43 homologues are closely related to other phage tail sheath proteins (data not shown). With the exception of ORF43 and ORF49 (and possibly ORF40, a homologue of Mu gp36 of unknown function), there is no significant homology to any characterized phage tail genes. Since B3 has a flexible noncontractile tail morphologically similar to that of phages lambda and A118 (46, 55, 84), one might expect approximately 10 B3 genes to be involved in tail morphogenesis. The right half of the B3 genome contains 15 ORFs of as yet unidentified function (ORF40 through -42, ORF44 through -46, and ORF50 through -58). Excluding the very small ORFs (ORF45 and ORF46) and ORF58, which had no detectable homologue, there are 12 remaining B3 ORFs, and we predict that most of these are involved in B3 tail morphogenesis. After completion of the BLAST analysis, we discovered that genes at the right end of the phage D3112 genome encode proteins with very high similarity to those from B3 ORF41 through ORF57, with the notable exception of ORF45 through ORF48, which include the B3 DNA modification gene cluster. These D3112 genes also have the same order as their B3 homologues (91).

Examining the sequences in untranslated regions and very short ORFs between ORF22 and ORF57 both manually and with the program BIOPROSPECTOR (54), we failed to detect any consistent sequence elements that might serve as promoters or binding sites for regulators of late transcription. We were also unable to detect elements which could form a potential RNA stem-loop structure, followed by a string of T residues, features characteristic of Rho-independent terminators (11). Thus, insight into possible regulatory sequences and mechanisms must await experimental determination of transcript ends within this late gene region. Interestingly, there is a poor, but recognizable, promoter just upstream of ORF58 (Fig. 2), raising the possibility that it is a moron, an autonomous genetic module containing a protein-coding region flanked by a promoter and terminator (41).

**(iii) DNA modification gene cluster.** The Mu DNA modification gene cluster contains two genes, *com* and *mom*, which are located at the extreme right end of the genome (Fig. 1). The Mom protein modifies about 15% of the adenine residues in Mu DNA to acetamidoadenine, protecting it from cleavage by a variety of restriction enzymes (31, 42). In the absence of Com, translation of *mom* mRNA is inefficient due to formation of a stem-loop structure that occludes the ribosome-binding site and start codon for Mom translation (31, 42). The Mu Com protein is a zinc finger protein that binds to an adjacent stem-loop structure, destabilizing the *mom* RNA stem-loop and allowing translation of Mom (31, 42).

The B3 DNA modification gene cluster also contains two genes: ORF47, which encodes a Com homologue containing the four highly conserved zinc finger cysteines, and ORF48 which encodes a DNA adenine methyltransferase homologue (see Table S2 in the supplemental material). Folding of RNA sequences containing the translation initiation region of ORF48 using the Zuker (94) Mfold web server (http://www.bioinfo.rpi.edu/applications/mfold) produced a variety of structures depending on the length of sequence used. In these structures, the ORF48 ribosome-binding site bases GGAG were partially or completely sequestered in double-stranded regions (data not shown), leaving open the possibility that translation of ORF48 is similarly regulated.

**Global organization of the B3 genome: functional coding regions.** Although there is significant amino acid sequence similarity between proteins encoded by B3 and Mu, there are several differences in genome organization, with the most dramatic being the opposite orientations of their early operons (Fig. 1). In B3, the leftmost 9 kb containing the proposed primary early operon is transcribed leftward, whereas in Mu, the early genes are transcribed rightward (Fig. 1).

The proposed B3 late genes are transcribed rightward, as are the Mu late genes, and the order of B3 head genes parallels that of the Mu head genes (Fig. 1). One notable difference between the B3 and Mu genome organization is the location of the B3 *com-dam* DNA modification gene cluster within the tail gene region about 12 kb from the genome right end; the corresponding *com-mom* gene cluster of Mu is located at the extreme right end of the Mu genome (Fig. 1).

**B3 evolution.** Homologues of B3 genes were found in a large array of phages, prophages, and phage-related elements (see Table S2 in the supplemental material). Table 2 lists the names of the phages, prophages, and phage-related elements with significant homology to B3 genes, their hosts, and genome locations and reveals a dramatic increase in the number of Mu-like transposable phage family relatives since the annotation of the Mu genome in 2002 (68). The hosts containing these prophages and elements, as well as their ecological niches, are quite diverse, ranging from nonpathogenic enteric bacteria, such as *E. coli* K-12 (87), to human or animal pathogens, including *E. coli* O157:H7 (32), *Neisseria meningitidis* (48, 71, 88), *Salmonella enterica* (70), *Bordetella bronchiseptica* (72), *Vibrio cholerae* (33), *Haemophilus influenzae* (20), *Burkholderia cenocepacia* (86), and *Haemophilus ducreyi* (GenBank accession number NC_002940), to predominantly nonpathogenic soil organisms, such as *Shewanella oneidensis* (34) and *Chromobacterium violaceum* (10), to plant pathogens, such as *Xylella fastidiosa* (83). This diversity strongly supports the conclusion that tailed phages are genetic mosaics derived by multiple stepwise recombinational exchanges that occur within a single, large gene pool for tailed phages (14, 35, 36).

While B3, Mu, and multiple bacterial genome prophages are clearly closely related, the level of similarity varies greatly across the B3 genome (Fig. 3). The genes with the greatest similarity and highest frequency of homologues are located in the early and head regions. Consistent with the difference in tail morphology, homology with the Mu family was absent from

TABLE 2. Phages, prophages, and phage-related elements with homology to phage B3[a]

| Phage or element[b] | Bacterial host[c] | Gene(s)[d] |
|---|---|---|
| Phage A118 | *Listeria monocytogenes* | ORF1-ORF72 |
| Phage BcepMu | *Burkholderia cenocepacia* | gp1-gp53 |
| BorMu | *Bordetella bronchiseptica* RB50 | BB3606-BB3657 |
| Brs1* | *Brucella suis* 1330 | ?BR0584-BR0599? (ext) |
| Bruc1* | *Brucella melitensis* 16M | BMEI1340-BMEI1350 |
| Cc1* | *Caulobacter crescentus* CB215 | ?CC2776-CC2790 (ext) |
| CP-933T* | *Escherichia coli* O157:H7 EDL933 | Z2966-Z2994 |
| CV1* | *Chromobacterium violaceum* ATCC 12472 | CV0406-CV0432 |
| CV2* | *Chromobacterium violaceum* ATCC 12472 | ?CV1474-CV1479? |
| Phage D108 | *Escherichia coli* | Probably 50 to 60 genes[e] |
| Phage D3112 | *Pseudomonas aeruginosa* | p01-p55[f] |
| DucMu1 | *Haemophilus ducreyi* 35000HP | HD0089-HD0156 |
| DucMu2 | *Haemophilus ducreyi* 35000HP | HD0479-HD0540 |
| DucMu3 | *Haemophilus ducreyi* 35000HP | HD1517-HD1581 |
| FluMu | *Haemophilus influenzae* Rd KW20 | HI1476-HI1523 |
| Phage HF2 | *Haloferax volcanii* | ORF1-ORF121 |
| Phage K139 | *Vibrio cholerae* | ORF1-ORF44 |
| MiniFluMu* | *Haemophilus influenzae* Rd KW20 | HI1568-HI1571 |
| Phage Mu | *Escherichia coli* | gp1-gp55 |
| MuSo1 | *Shewanella oneidensis* MR-1 | SO0641-SO0683 |
| MuSo2 | *Shewanella oneidensis* MR-1 | SO2652-SO2704 |
| NeisMu1* | *Neisseria meningitidis* B MC58 | NMB1078-NMB1121 |
| NeisMu2*[g] | *Neisseria meningitidis* B MC58 | NMB0985-NMB0991 |
| NMB1* | *Neisseria meningitidis* B MC58 | NMB1002-NMB1007 |
| Pnm1 | *Neisseria meningitidis* A Z2491 | NMA1821-NMA1884 |
| Pnm2* | *Neisseria meningitidis* A Z2491 | NMA1281-NMA1330 |
| Pnm3* | *Neisseria meningitidis* A Z2491 | NMA1185-NMA1199? |
| Pnm4/5*[h] | *Neisseria meningitidis* A Z2491 | NMA1208-NMA1231? |
| Phage 03 | *Pseudomonas syringae* DC3000 | PSPTO3385-PSPTO3429 |
| Phage V | *Shigella flexneri* | SfV ORF1-ORF53 |
| Phage 186 | *Escherichia coli* | ORF1-ORF97 |
| Phage P2 | *Escherichia coli* | ORF1-ORF42 |
| Phage ϕP27 | *Escherichia coli* | ORF1-ORF58 |
| Phage ϕE125 | *Burkholderia mallei* | ORF1-ORF70 |
| Phage PP01 | *Pseudomonas putida* KT2440 | PP3849-PP3920 |
| RadMu | *Deinococcus radiodurans* R1 | DRA0074-DRA0119 |
| RS1* | *Ralstonia solanacearum* GMI1000 | ?RSc0839-RSc0892? |
| RS6 | *Ralstonia solanacearum* GMI1000 | RSc1896-RSc1948 |
| SP14* | *Escherichia coli* O157:H7 Sakai | ?ECs2713-ECs2774 |
| SP18 | *Escherichia coli* O157:H7 Sakai | ECs4943-ECs4998 |
| Sti3[i] | *Salmonella* serovar Typhi CT18[j] | STY1591-STY1643 |
| Phage ST64B | *Salmonella* serovar Typhimurium[j] | Sb1-Sb56 |
| Stm7* | *Salmonella* serovar Typhimurium LT2[j] | STM4196-STM4219 |
| Tum2 | *Agrobacterium tumefaciens* C58 | AGR_C_1731-AGR_C_1774 |
| VibMu* | *Vibrio cholerae* N16961 | ?VC1788-VC1803 (ext) |
| VioMu[k] | *Chromobacterium violaceum* ATCC 12472 | ?CV2111-CV2165? |
| XacP1* | *Xanthomonas axonopodis* 903 | ?XAC 2640-XAC2655? (ext) |
| XccP1 | *Xanthomonas campestris* ATCC33913 | XCC2962-3012 |
| XfP1 | *Xylella fastidiosa* 9a5c | XF0678-XF0733 |
| XfP4* | *Xylella fastidiosa* 9a5c | XF1642-XF1711 |
| XfP5* | *Xylella fastidiosa* 9a5c | XF0485-XF0536? |
| XfP6* | *Xylella fastidiosa* 9a5c | XF2110-XF2132 (ext) |

[a] Much of the information given here for prophages and phage-related elements found in bacterial genome sequences was taken from the review by Casjens (13). The genome sequences for *Bordetella bronchiseptica*, *Chromobacterium violaceum*, and *Haemophilus ducreyi* were completed after that review; their B3-homologous prophages and elements are named here using the conventions of Morgan et al. (68) and Casjens (13). Specifically, new Mu-like prophages were named by adding a three-letter prefix to Mu (e.g., VioMu), and short phage-related elements are identified by a strain designator and asterisk (e.g., CV1*). For other phages and phage elements that were named by viral and bacterial genome sequencing teams, we use the names given by those teams.

[b] The names of phages known to be capable of infection and multiplication are designated Phage. Elements that appear to be too short to allow independent multiplication are marked with an asterisk. Newly identified prophages with high homology to Mu are designated BorMu, DucMu (pronounced duke), and VioMu according to the convention of Morgan et al. (68).

[c] For infectious phage, a common host is listed. For prophages and elements, the strain whose genome was sequenced is given.

[d] The genes included in each phage, prophage, or element are listed using the genomic sequential ORF numbering system. A question mark is added when the terminal gene is uncertain. (ext) indicates that the identification of B3-homologous genes has resulted in the extension of the likely outer limits of prophages and phage-related elements beyond those annotated by Casjens (13).

[e] Phage D108 has not been sequenced, but it is very closely related to Mu (21); therefore, it probably has 50 to 60 genes.

[f] The sequence database annotation for D3112 contains 55 genes, whereas the published paper lists only 53 genes (91).

[g] NeisMu2 is also called MuMenB (62).

[h] Pnm4 and Pnm5 were originally annotated as two separate elements (48, 71). Their close proximity and the identification of intervening genes as B3 homologues suggest that they may be a single longer element that we call Pnm4/5.

[i] Sti3 is also called SalMu (86).

[j] The complete names for *Salmonella* species are as follows: *Salmonella* serovar Typhi is *Salmonella enterica* subsp. *enterica* serovar Typhi (strain CT18), and *Salmonella* serovar Typhimurium LT2 is *Salmonella enterica* serovar Typhimurium (strain LT2).

[k] VioMu is also called ChromoMu (86).

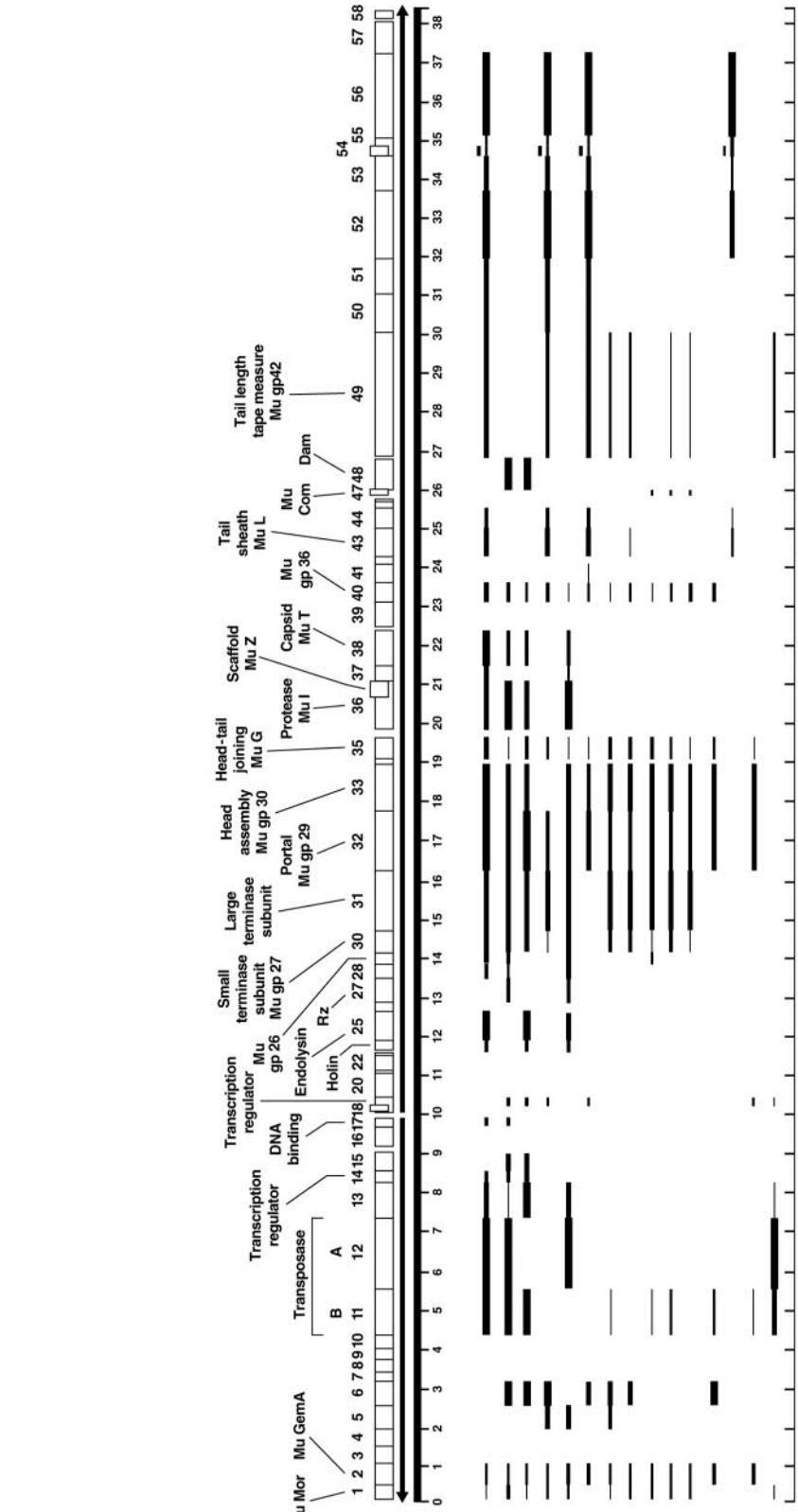


FIG. 3. Diagrammatic representation of the degree of relatedness for B3 genes with homologues in other Mu-like phages and prophages. The B3 map, ORFs, and predicted protein functions are shown as in Fig. 1, except that all ORFs are shown above the kilobase line. Phages and prophages with homology to at least six B3 genes are included, and their names are listed on the left. Horizontal lines drawn below a B3 gene indicate the presence of a homologue in the corresponding Mu-like phage or prophage. The thickness of the line indicates the degree of similarity, achieved by dividing the BLAST expect values in Table S2 in the supplemental material into five groups, with the thickest line representing the greatest similarity (group I) and decreasing thickness indicating decreasing similarity. The expect value boundaries for the five groups are as follows: e-50 or better for group I, e-49 to e-15 for group II, e-14 to e-4 for group III, 0.001 to 5.0 for group IV, and >5.0 for group V on PSI-BLAST iteration 1 but 1.0 or better on PSI-BLAST iteration 2.

the predicted tail region (see Table S2 in the supplemental material), but the three *H. ducreyi* prophages closely related to B3, DucMu1, DucMu2, and DucMu3 (Fig. 3), and D3112 (data not shown) exhibit high similarity across both predicted B3 tail gene regions. Consistent with the conclusion that proteins which function together are often grouped together in the genome, many of the B3-related phages and prophages in Fig. 3 appear to have retained or lost the entire B3 head gene module or tail gene module rather than just a subset of those genes. The most notable exceptions are XfP5, Pnm2, and VibMu, which are known to be cryptic phage remnants lacking many phage genes (Table 2) (13). It remains to be determined whether the other exceptions are also defective or have simply evolved proteins that perform the same function but exhibit little amino acid sequence similarity, as observed previously for other phages, including the lambdoid phages HK022 and HK97 (41, 68).

The Mu-like prophages shown in Fig. 3 all share the global gene organization of Mu, with early genes followed by lysis genes, then head genes, and finally tail genes. In most cases, the order of genes within these functional groups is also conserved, and B3 hypothetical ORFs without homologues are replaced by different hypothetical ORFs in the other phages, although the number of such ORFs often varies. The inversion of the B3 early gene region relative to Mu allows us to divide the Mu-like phages and prophages into two groups. Those with rightward transcription of the early genes include Mu, FluMu, SP18, Pnm1, Pnm2, DucMu1, DucMu2, MuSo1, MuSo2, and D3112. Those which, like B3, have the early operon transcribed leftward include BorMu, VioMu, DucMu3, Sti3, and BcepMu. Genes in prophages BorMu, VioMu, and DucMu3 also exhibited the greatest similarity to B3 genes (Fig. 3).

**B3 transposase subunit evolution.** The BLAST and PSI-BLAST searches with ORF12 identified a number of homologues in genomic Mu-like prophages and multiple annotated transposase proteins, including the primary transposase subunits of RadMu and Tn*552* (Table 1; see also Table S2 in the supplemental material). The Tn*552* transposase TnpA, like the Mu A protein, is a member of the DDE superfamily of transposases that include retroviral integrases and the transposases of the IS*3* family of bacterial insertion sequences (28, 53). This superfamily consists of a relatively heterogeneous group of proteins that share at least a ~200-residue catalytic core and perform closely related strand transfer reactions (28, 53). Despite the lack of similarity to the Mu A transposase subunit, these homologies clearly identified ORF12 as encoding the primary transposase subunit for B3. Its location, in the early operon, also parallels that of the Mu A transposase (Fig. 1).

Curiously, similar searches with ORF11 identified four different groups of proteins. Several of those with the greatest similarity were annotated as homologues of A subunits of the bacterial type II general secretion pathway (GSP) used for the second step of secretion of multiple hydrolytic enzymes and toxins from the periplasm to the extracellular environment (77, 78). The second group contained true GSPA proteins. The third group contained the MshM proteins of the bacterial type IV MSHA (mannose-sensitive hemagglutination antigen) secretion pathway used for synthesis of type IV pili and for transfer of DNA during conjugation and transfer of T-DNA into plant cells by *Agrobacterium tumefaciens* (17, 56, 61). Many of the type IV secretion pathway proteins are closely related to those of the type II pathway, and MshM shares homology with the GSPA component in particular (56). The fourth group, with considerably lower similarity, contained a number of transposase subunits similar to and including the Mu B and Tn*552* TnpB transposase subunits. Strikingly, the proteins with the greatest similarity to B3 ORF 11 were encoded next to the Mu A protein homologue within BcepMu (86) and the closely related Mu-like prophages VioMu, DucMu (DucMu1, -2, and -3), BorMu, and VibMu, a position analogous to that of the B gene in Mu (Fig. 1). In the case of BcepMu, the B3 ORF11 homologue (BcepMu gp8) has been annotated as ExeA and proposed to serve as a potential virulence factor for pathogenesis (86). In contrast, the similarities and locations described above lead us to propose that ORF11 and its phage and prophage homologues encode the second transposase subunit analogous to Mu B, an ATPase subunit that brings the target DNA to the transposase complex (28, 64).

The observed similarity of the transposase subunits to ATP-binding subunits of the type II and type IV secretion systems is real. Beginning a PSI-BLAST homology search with each of the two best-characterized proteins, Mu B (64) and the *Aeromonas hydrophila* type II secretion protein ExeA (also called GSPA) (40, 81), led to recovery of the same three groups of homologous proteins in the second iteration (data not shown). Furthermore, the conserved ~150-amino-acid ATPase domain of the ExeA subunit (COG 3267) is located within the ~300-amino-acid conserved transposase ATPase subunit domain (COG 2842) (58; http://www.ncbi.nlm.nih.gov; data not shown). Whereas similarity of the 500- to 700-amino-acid GSPA secretion proteins to each other usually extended over the entire length of both proteins, similarity of the smaller ~300-amino-acid type IV proteins and the 300- to 400-amino-acid transposase subunits was generally limited to the N-terminal 250 to 300 amino acids of ExeA. This region of ExeA contains the three motifs characteristic of ATPases, the Walker A motif, the $Mg^{2+}$-binding site, and the Walker B motif (81). Thus, it seems likely that it is the related ATPase domains in these groups of proteins that are responsible for their detection as homologues.

Clearly, the BLAST scores show that ORF11 of B3 and its prophage homologues are much more similar to the ATPase domains of the secretion genes than to the Mu B-like transposase subunits (Table 1; see also Table S2 in the supplemental material); yet, like the transposase subunits, they are only ~400 amino acids long. Perhaps the secretion proteins and ORF11 homologues evolved from a common ATPase ancestor, either by deleting the C terminus of a long protein precursor to form the transposase subunit or by adding 100 to 300 amino acids to the ATPase domain to form the secretion protein, with the added region playing a secretion-specific role. Nevertheless, at this point, we cannot rule out the possibility that ORF11 and its prophage homologues participate in both transposition and protein secretion. A test of their ability to complement an *exeA* or *mshM* mutant is clearly warranted, as is a test for transposition of an ORF 11 mutant phage.

**B3 DNA modification gene evolution.** The B3 DNA adenine methylase gene, ORF48, was the only B3 gene with significant nucleotide sequence similarity to any other phage or bacterial gene; it shares similarity with gene 27, the DNA adenine meth-

ylase gene, of the temperate phage ϕE125 of *Burkholderia thailandensis* (93). Not surprisingly, the amino acid sequences of the ORF48 and gene 27 proteins exhibited extremely high similarity over the entire length of both proteins. Very high protein similarity was also observed for the proteins annotated as DNA modification methylases of the BorMu, VioMu, and phage 03 prophages of *Bordetella bronchiseptica*, *Chromobacterium violaceum*, and *Pseudomonas syringae* (see Table S2 in the supplemental material). There were about 70 additional ORF48 homologues; most were bacterial proteins and had considerably poorer scores. This group contained multiple methylase genes associated with restriction-modification systems, e.g., *Mbo*I and *Dpn*II, and multiple cytosine-specific methyltransferases as well. Thus, it remains to be seen whether ORF48 and its prophage homologues perform an adenine methylation or some other type of modification and whether that modification provides a survival advantage to the phage as Mom does for Mu (31, 42).

Curiously, B3 is the only phage in Fig. 3 that encodes both Com and Dam homologues. Phages VioMu and BorMu encode Dam homologues, but no Com homologue (Fig. 3). Phages SP18, FluMu, and Mu encode Com homologues but no Dam homologue (Fig. 3). A BLASTP search with the Mu Mom protein sequence revealed that SP18 encodes a Mom DNA modification protein just downstream of its *com* gene (data not shown), but as observed previously by Morgan et al. (68), FluMu encodes a different non-Mom-like protein at that position, as does Pnm1 (68). A BLASTP search with the FluMu protein sequence identified homologues in Pnm1 and Pnm2 as well as DucMu1, DucMu2, and DucMu3 (data not shown). The lack of similarity between these new genes and Mom or Dam raises the possibility that these genes may perform a DNA modification different from both methylation and acetamidoadenine modification. In Mu, the Com protein is needed to prevent translation of Mom until late in the lytic cycle, because high-level Mom expression is lethal (31, 43). The absence of intact Com homologues in these other phages (Pnm1 contains a mutant *com* gene [68]) suggests that expression of their DNA modification genes may not be lethal, thereby making Com unnecessary.

In Mu and all but one of the above phages, the DNA modification gene(s) is located very close to the right end of the phage genome, whereas in DucMu1 and B3, the DNA modification gene(s) is located internally, approximately 12 kb from the right end (data not shown). Thus, there are striking differences in the gene number, modification protein sequence, and gene locations in this group of closely related phages.

**Summary.** The sequencing of transposable phage B3 particle DNA revealed that B3 has a linear genome, 38,439 bp long, with variable host DNA fragments attached to the genome ends. Comparison of the B3 genetic map to that of Mu revealed evidence of multiple genetic rearrangements and substitutions. The results from homology searches with the 59 predicted B3 ORFs allowed us to predict potential functions for almost half of the genes, defining distinct transposition, regulation, lysis, head, and tail regions. Homologous proteins were found in multiple related phages with a diverse range of bacterial hosts, raising the possibility that genetic manipulations based on transposable phage technology (23, 90) can be applied to this broad spectrum of pathogenic and nonpathogenic bacteria. The sequence also provides much of the essential information needed for the development of B3 vectors for use in bioremediation.

## ACKNOWLEDGMENTS

This work was supported in part by Research and Reimbursement Agreement 5753 from the UNOCAL Corporation to R.J.C. for the Environmental Biotechnology Institute for bioremediation studies at the UNOCAL Guadalupe Oil Fields, by National Science Foundation grant MCB-0318108 to M.M.H., and by a Van Vleet Chair of Excellence Professorship held by M.M.H.

We thank Alice Hamrick for assistance with G+C content analysis of the alternate ORFs and Christy Houde for assistance with the primer walk-out sequencing.

## REFERENCES

1. **Abraham, W. R., B. Nogales, P. N. Golyshin, D. H. Pieper, and K. N. Timmis.** 2002. Polychlorinated biphenyl-degrading microbial communities in soils and sediments. Curr. Opin. Microbiol. **5:**246–253.
2. **Akhverdyan, V. Z., E. A. Khrenova, M. A. Reulets, T. V. Gerasimova, and V. N. Krylov.** 1985. Characterization of transposon phages of *Pseudomonas aeruginosa* belonging to two groups differing in DNA-DNA homology. Genetika **21:**735–746.
3. **Akhverdyan, V. Z., E. A. Khrenova, V. G. Bogush, T. V. Gerasimova, N. B. Kirsanov, and V. N. Krylov.** 1984. Wide distribution of transposable phages in natural populations of *Pseudomonas aeruginosa*. Genetika **20:**1612–1619.
4. **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.
5. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
6. **Blasi, U., and R. Young.** 1996. Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. Mol. Microbiol. **21:**675–682.
7. **Bodey, G. P., R. Bolivar, V. Fainstein, and L. Jadeja.** 1983. Infections caused by *Pseudomonas aeruginosa*. Rev. Infect. Dis. **5:**279–313.
8. **Botstein, D.** 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. **354:**484–490.
9. **Bourkaltseva, M. V., and V. N. Krylov.** 1997. Comparison of D3112 and B3 transposable phage insertions located in the *Pseudomonas aeruginosa* chromosome. Genetika **33:**553–555.
10. **Brazilian National Genome Project Consortium.** 2003. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. Proc. Natl. Acad. Sci. USA **100:**11660–11665.
11. **Brendel, V., G. H. Hamm, and E. N. Trifonov.** 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. **3:**705–723.
12. **Campa, M., M. Bendinelli, and H. Friedman (ed.).** 1993. *Pseudomonas aeruginosa* as an opportunistic pathogen. Plenum Press, New York, N.Y.
13. **Casjens, S.** 2003. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. **49:**277–300.
14. **Casjens, S., G. F. Hatfull, and R. W. Hendrix.** 1992. Evolution of ds-DNA tailed-bacteriophage genomes. Semin. Virol. **3:**383–397.
15. **Chayabutra, C., and L. Ju.** 2000. Degradation of *n*-hexadecane and its metabolites by *Pseudomonas aeruginosa* under microaerobic and anaerobic denitrifying conditions. Appl. Environ. Microbiol. **66:**493–498.
16. **Chen, L. L., and C. T. Zhang.** 2003. Seven GC-rich microbial genomes adopt similar codon usage patterns regardless of their phylogenetic lineages. Biochem. Biophys. Res. Commun. **306:**310–317.
17. **Christie, P. J.** 1997. *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. J. Bacteriol. **179:**3085–3094.
18. **Cowell, B. A., S. S. Twining, J. A. Hobden, M. S. Kwong, and S. M. Fleiszig.** 2003. Mutation of *lasA* and *lasB* reduces *Pseudomonas aeruginosa* invasion of epithelial cells. Microbiology **149:**2291–2299.
19. **Cunningham, L., M. Pitt, and H. D. Williams.** 1997. The *cioAB* genes from *Pseudomonas aeruginosa* code for a novel cyanide-insensitive terminal oxidase related to the cytochrome bd quinol oxidases. Mol. Microbiol. **24:**579–591.
20. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. Fitzhugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.

21. **Gill, G. S., R. C. Hull, and R. Curtiss III.** 1981. Mutator bacteriophage D108 and its DNA: an electron microscopic characterization. J. Virol. **37:**420–430.
22. **Grimaud, R.** 1996. Bacteriophage Mu head assembly. Virology **217:**200–210.
23. **Groisman, E. A.** 1991. *In vivo* genetic engineering with bacteriophage Mu. Methods Enzymol. **204:**180–212.
24. **Grundy, F. J., and M. M. Howe.** 1984. Involvement of the invertible G segment in bacteriophage Mu tail fiber biosynthesis. Virology **134:**296–317.
25. **Grundy, F. J., and M. M. Howe.** 1985. Morphogenetic structures present in lysates of amber mutants of bacteriophage Mu. Virology **143:**485–504.
26. **Ha, U. H., Y. Wang, and S. Jin.** 2003. DsbA of *Pseudomonas aeruginosa* is essential for multiple virulence factors. Infect. Immun. **71:**1590–1595.
27. **Hancock, R. E. W.** 1998. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermentative gram-negative bacteria. Clin. Infect. Dis. **27:**S93-S99.
28. **Haren, L., B. Ton-Hoang, and M. Chandler.** 1999. Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. **53:**245–281.
29. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:**2343–2361.
30. **Haro, M.-A., and V. de Lorenzo.** 2001. Metabolic engineering of bacteria for environmental applications: construction of *Pseudomonas* strains for biodegradation of 2-chlorotoluene. J. Biotechnol. **85:**103–113.
31. **Hattman, S.** 1999. Unusual transcriptional and translational regulation of the bacteriophage Mu *mom* operon. Pharmacol. Ther. **84:**367–388.
32. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
33. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleischmann, W. C. Nierman, and O. White.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477–483.
34. **Heidelberg, J. F., I. Paulsen, K. Nelson, E. Gaidos, W. Nelson, T. Read, J. Eisen, R. Seshadri, N. Ward, B. Methe, R. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. DeBoy, R. Dodson, A. Durkin, D. Haft, J. Kolonay, R. Madupu, J. Peterson, L. Umayam, O. White, A. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. Utterback, L. McDonald, T. Feldblyum, H. Smith, J. Venter, K. Nealson, and C. Fraser.** 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nat. Biotechnol. **20:**1118–1123.
35. **Hendrix, R. W., G. F. Hatfull, and M. C. Smith.** 2003. Bacteriophages with tails: chasing their origins and evolution. Res. Microbiol. **154:**253–257.
36. **Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA **96:**2192–2197.
37. **Holloway, B. W., and A. F. Morgan.** 1986. Genome organization in *Pseudomonas*. Annu. Rev. Microbiol. **40:**79–105.
38. **Holloway, B. W., J. B. Egan, and M. Monk.** 1960. Lysogeny in *Pseudomonas aeruginosa*. Aust. J. Exp. Biol. Med. Sci. **38:**321–330.
39. **Holloway, B. W., V. Krishnapillai, and A. F. Morgan.** 1979. Chromosomal genetics of *Pseudomonas*. Microbiol. Rev. **43:**73–102.
40. **Jahagirdar, R., and S. P. Howard.** 1994. Isolation and characterization of a second *exe* operon required for extracellular protein secretion in *Aeromonas hydrophila*. J. Bacteriol. **176:**6819–6826.
41. **Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix.** 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. **299:**27–51.
42. **Kahmann, R.** 1984. The *mom* gene of bacteriophage Mu. Curr. Top. Microbiol. Immunol. **108:**29–47.
43. **Kahmann, R., A. Seiler, F. G. Wulczyn, and E. Pfaff.** 1985. The *mom* gene of bacteriophage Mu: a unique regulatory scheme to control a lethal function. Gene **39:**61–70.
44. **Kamp, D., and R. Kahmann.** 1981. Two pathways in bacteriophage Mu transposition? Cold Spring Harbor Symp. Quant. Biol. **45:**329–336.
45. **Kang, P. J., A. R. Hauser, G. Apodaca, S. Fleiszig, J. Wiener-Kronish, K. Mostov, and J. N. Engel.** 1997. Identification of *Pseudomonas aeruginosa* genes required for epithelial cell injury. Mol. Microbiol. **24:**1249–1262.
46. **Katsura, I.** 1983. Tail assembly and injection, p. 331–346. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
47. **Kellogg, S. T., D. K. Chatterjee, and A. M. Chakrabarty.** 1981. Plasmid-assisted molecular breeding: new technique for enhanced biodegradation of persistent toxic chemicals. Science **214:**1133–1135.
48. **Klee, S. R., X. Nassif, B. Kusecek, P. Merker, J.-L. Beretti, M. Achtman, and C. R. Tinsley.** 2000. Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. Infect. Immun. **68:**2082–2095.
49. **Krylov, V. N., V. G. Bogush, and J. Shapiro.** 1980. Bacteriophages of *Pseudomonas aeruginosa* with DNA similar in structure to that of phage Mu1. I. General description, localization of sites sensitive to endonucleases in DNA, and structure of homoduplexes of phage D3112. Genetika **16:**824–832.
50. **Krylov, V. N., V. G. Bogush, A. S. Ianenko, and N. B. Kirsanov.** 1980. *Pseudomonas aeruginosa* bacteriophages with DNA structure similar to the DNA structure of Mu1 phage. II. Evidence for similarity between D3112, B3, and B39 bacteriophages: analysis of DNA splits by restriction endonucleases, isolation of D3112 and B3 recombinant phages. Genetika **16:**975–984.
51. **Krylov, V. N., V. Z. Akhverdian, V. G. Bogush, E. A. Khrenova, and M. A. Reulets.** 1985. Modular structure of the genes of phages-transposons of *Pseudomonas aeruginosa*. Genetika **21:**724–734.
52. **Lambert, P. A.** 2002. Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. J. R. Soc. Med. **95**(Suppl. 41)**:**22–26.
53. **Leschziner, A. E., T. J. Griffin IV, and N. D. Grindley.** 1998. Tn552 transposase catalyzes concerted strand transfer *in vitro*. Proc. Natl. Acad. Sci. USA **95:**7345–7350.
54. **Liu, X., D. L. Brutlag, and J. S. Liu.** 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of coexpressed genes. Pac. Symp. Biocomput. **2001:**127–138.
55. **Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar.** 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. **35:**324–340.
56. **Lory, S.** 1998. Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting. Curr. Opin. Microbiol. **1:**27–35.
57. **Ma, J., A. Campbell, and S. Karlin.** 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. **184:**5733–5745.
58. **Marchler-Bauer, A., J. B. Anderson, C. DeWeese-Scott, N. D. Fedorova, L. Y. Geer, S. He, D. I. Hurwitz, J. D. Jackson, A. R. Jacobs, C. J. Lanczycki, C. A. Liebert, C. Liu, T. Madej, G. H. Marchler, R. Mazumder, A. N. Nikolskaya, A. R. Panchenko, B. S. Rao, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, S. Vasudevan, Y. Wang, R. A. Yamashita, J. J. Yin, and S. H. Bryant.** 2003. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res. **31:**383–387.
59. **Margolin, W., G. Rao, and M. M. Howe.** 1989. Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. J. Bacteriol. **171:**2003–2018.
60. **Marrs, C. F., and M. M. Howe.** 1990. Kinetics and regulation of transcription of bacteriophage Mu. Virology **174:**192–203.
61. **Marsh, J. W., and R. K. Taylor.** 1999. Genetic and transcriptional analyses of the *Vibrio cholerae* mannose-sensitive hemagglutinin type 4 pilus gene locus. J. Bacteriol. **181:**1110–1117.
62. **Masignani, V., M. M. Giuliani, H. Tettelin, M. Comanducci, R. Rappuoli, and V. Scarlato.** 2001. Mu-like prophage in serogroup B *Neisseria meningitidis* coding for surface-exposed antigens. Infect. Immun. **69:**2580–2588.
63. **Mathee, K., and M. M. Howe.** 1990. Identification of a positive regulator of the Mu middle operon. J. Bacteriol. **172:**6641–6650.
64. **Maxwell, A., R. Craigie, and K. Mizuuchi.** 1987. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. Proc. Natl. Acad. Sci. USA **84:**699–703.
65. **McLean, B. W., S. L. Wiseman, and A. M. Kropinski.** 1997. Functional analysis of sigma-70 consensus promoters in *Pseudomonas aeruginosa* and *Escherichia coli*. Can. J. Microbiol. **43:**981–985.
66. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
67. **Miyata, S., M. Casey, D. W. Frank, F. M. Ausubel, and E. Drenkard.** 2003. Use of the *Galleria mellonella* caterpillar as a model host to study the role of the type III secretion system in *Pseudomonas aeruginosa* pathogenesis. Infect. Immun. **71:**2404–2413.
68. **Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix.** 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. **317:**337–359.
69. **Narasimhan, G., C. Bu, Y. Gao, X. Wang, N. Xu, and K. Mathee.** 2002. Mining protein sequences for motifs. J. Comput. Biol. **95:**707–720.
70. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. G. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848–852.
71. **Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather,**

**S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt, and B. G. Barrell.** 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature **404:**502–506.

72. **Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. **35:**32–40.
73. **Pukatzki, S., R. H. Kessin, and J. J. Mekalanos.** 2002. The human pathogen *Pseudomonas aeruginosa* utilizes conserved virulence pathways to infect the social amoeba *Dictyostelium discoideum*. Proc. Natl. Acad. Sci. USA **99:** 3159–3164.
74. **Roncero, C., A. Darzins, and M. J. Casadaban.** 1990. *Pseudomonas aeruginosa* transposable bacteriophages D3112 and B3 require pili and surface growth for adsorption. J. Bacteriol. **172:**1899–1904.
75. **Rothmel, R. K., A. M. Chakrabarty, A. Berry, and A. Darzins.** 1991. Genetic systems in *Pseudomonas*. Methods Enzymol. **204:**485–514.
76. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
77. **Sandkvist, M.** 2001. Biology of type II secretion. Mol. Microbiol. **40:**271–283.
78. **Sandkvist, M.** 2001. Type II secretion and pathogenesis. Infect. Immun. **69:**3523–3535.
79. **Sarac, M. S., A. Cameron, and I. Lindberg.** 2002. The furin inhibitor hexa-D-arginine blocks the activation of *Pseudomonas aeruginosa* exotoxin A in vivo. Infect. Immun. **70:**7136–7139.
80. **Schmidtchen, A., I. M. Frick, E. Andersson, H. Tapper, and L. Bjorck.** 2002. Proteinases of common pathogenic bacteria degrade and inactivate the antibacterial peptide LL-37. Mol. Microbiol. **46:**157–168.
81. **Schoenhofen, I. C., C. Stratilo, and S. P. Howard.** 1998. An ExeAB complex in the type II secretion pathway of *Aeromonas hydrophila*: effect of ATP-binding cassette mutations on complex formation and function. Mol. Microbiol. **29:**1237–1247.
82. **Siboo, I. R., F. Sieder, K. Kumar, M. M. Howe, and M. S. DuBow.** 2004. Characterization of P*lys*-proximal morphogenetic genes of transposable bacteriophage Mu. Arch. Virol. **149:**241–259.
83. **Simpson, A. J. G., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. C. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. S. Briones, M. R. P. Bueno, A. A. Camargo, L. E. A. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. R. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J. S. Ferreira, V. C. A. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. S. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. C. Leite, E. G. M. Lemos, M. V. F. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. B. N. Madeira, H. M. F. Madeira, C. L. Marino, M. V. Marques, E. A. L. Martins, E. M. F. Martins, A. Y. Matsukuma, C. F. M. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteiro-Vitorello, D. H. Moon, M. A. Nagai, A. L. T. O. Nascimento, L. E. S. Netto, A. Nhani, Jr., F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. G. Pereira, H. A. Pereira, Jr., J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J. de M. Rosa, V. E. de Rosa, Jr., R. G. de Sa, R. V. Santelli, H. E. Sawasaki, A. C. R. da Silva, F. R. da Silva, A. M. da Silva, W. A. Silva, Jr., J. F. da Silveira, M. L. Z. Silvestri, W. J. Siqueira, A. A. de Souza, A. P. de Souza, M. F. Terenzi, D. Truffi, S. M. Tsai, M. H. Tsuhako, H. Vallada, M. A. Van Sluys, S. Verjovski-Almeida, A. L. Vettore, M. A. Zago, M. Zatz, J. Meidanis, and J. C. Setubal.** 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature **406:**151–157.
84. **Slayter, H. S., B. W. Holloway, and C. E. Hall.** 1964. The structure of *Pseudomonas aeruginosa* phages. J. Ultrastruct. Res. **11:**274–281.
85. **Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. L. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K.-S. Wong, Z. Wu, I. T. Paulsen, J. Reitzer, M. H. Saier, R. E. W. Hancock, S. Lory, and M. V. Olson.** 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature **406:**959–964.
86. **Summer, E. J., C. F. Gonzalez, T. Carlisle, L. M. Mebane, A. M. Cass, C. G. Savva, J. J. LiPuma, and R. Young.** 2004. *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. J. Mol. Biol. **340:**49–65.
87. **Taylor, A. L.** 1963. Bacteriophage-induced mutation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **50:**1043–1051.
88. **Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. Deboy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter.** 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287:**1809–1815.
89. **Van de Putte, P., M. Giphart-Gassler, N. Goosen, T. Goosen, and E. van Leerdam.** 1981. Regulation of integration and replication functions of bacteriophage Mu. Cold Spring Harbor Symp. Quant. Biol. **45:**347–353.
90. **Van Gijsegem, F., A. Toussaint, and M. J. Casadaban.** 1987. Mu as a genetic tool, p. 25–39. *In* N. Symonds, A. Toussaint, P. Van de Putte, and M. M. Howe (ed.), Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
91. **Wang, P. W., L. Chu, and D. S. Guttman.** 2004. Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. J. Bacteriol. **186:**400–410.
92. **Watanabe, K.** 2001. Microorganisms relevant to bioremediation. Curr. Opin. Biotechnol. **12:**237–241.
93. **Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer.** 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. **184:**4003–4017.
94. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.