JOURNAL OF BACTERIOLOGY, Feb. 2006, p. 1184–1187 Vol. 188, No. 3
0021-9193/06/$08.00+0 doi:10.1128/JB.188.3.1184–1187.2006

Comparative Genomic Analysis of 18 *Pseudomonas aeruginosa* Bacteriophages†

Tony Kwan,[1] Jing Liu,[1] Michael DuBow,[2] Philippe Gros,[3] and Jerry Pelletier[3]*

Targanta Therapeutics, Inc., 7170 Frederick Banting, 2nd Floor, Ville Saint Laurent, Quebec, Canada H4S 2A1[1]; Institut de Génétique et Microbiologie, Université Paris Sud, Bâtiment 409, 91405 Orsay, France[2]; and Department of Biochemistry and McGill Cancer Center, McIntyre Medical Sciences Building, McGill University, Montréal, Québec, Canada H3G 1Y6[3]

Received 15 August 2005/Accepted 6 November 2005

A genomic analysis of 18 *P. aeruginosa* phages, including nine newly sequenced DNA genomes, indicates a tremendous reservoir of proteome diversity, with 55% of open reading frames (ORFs) being novel. Comparative sequence analysis and ORF map organization revealed that most of the phages analyzed displayed little relationship to each other.

Recent studies of *Staphylococcus aureus* and *Mycobacterium tuberculosis* bacteriophages indicate these to be rich reservoirs of novel protein information that remain to be mined and analyzed (5, 9). Indeed, ~25% of the predicted open reading frames (ORFs) of *S. aureus* phages have no homology to predicted proteins in GenBank, with a much larger proportion (65%) of ORFs lacking functional annotation (5). Similar studies of mycobacterial phages revealed that between 50 to 75% of ORFs contain no match in GenBank (9).

To extend this information base, we sequenced nine *Pseudomonas aeruginosa* bacteriophages: F10, PA73, 119X, M6, F8, PA7, PA16, and SD1-M (obtained from H.-W. Ackermann, Felix d'Hérelle Reference Center for Bacterial Viruses, Québec City, Canada) and PA11 (obtained from the American Type Culture Collection, Manassas, VA). A comparative analysis was performed with the sequences of seven published *P. aeruginosa* double-stranded DNA bacteriophages (phiCTX, D3112, B3, phiKMV, D3, F116, and phiKZ) (1, 2, 4, 6–8, 12) and two unpublished bacteriophages (PaP2 and PaP3; NCBI accessions NC_005884 and NC_004466, respectively). With the exception of phages SD1-M and phiKZ (~281 kbp each), the remaining 16 phage genomes span a narrow-size range from ~35 kbp to ~72 kbp (Table 1).

Combined sequence analysis of the 18 phage genomes revealed a G+C content of 54.4%, significantly lower than the *P. aeruginosa* PA01 host genome (66.6%) (Table 1). This is surprising and a contrast to the similarity in G+C content between viral and host genomes observed in other studies with *S. aureus* (33.7% [phage] versus 32.9% [host]) (5), *M. tuberculosis* (63.6% [phage] versus 65.6% [host]) (9), and *Streptococcus pneumoniae* (39.8% [phage] versus 39.7% [host]) (data not shown). The G+C content was higher in the predicted coding regions (54.7%) than in the noncoding regions (50.5%) but still did not approach the 66.6% G+C content of the host genome. To determine the distribution of the lower G+C content in the phage genome, an analysis was performed in which the G+C content within a sliding window of 500 bp was determined for the entire length of each genome. The results obtained with phages F10, phiCTX, B3, and D3 revealed specific regions of high A+T content that significantly differ from other segments of their genomes (Fig. 1). This discrepancy between the phage and host G+C content may be due to (i) lateral gene transfer occurring from other hosts and/or phages with lower G+C contents, (ii) a recent invasion of these phages into *P. aeruginosa* from hosts with lower G+C contents, or (iii) a characteristic feature of these phages that has been stably maintained along their evolutionary history.

Bacteriophage ORFs were predicted by scanning all reading frames for the presence of a start codon (AUG, UUG, CUG, GUG, or AUA) and terminating at a minimum of 33 codons downstream. Putative ORFs were also scored for the presence of a Shine-Dalgarno sequence (5′GAAACC3′) centered 8 to 12 bp upstream of the start codon and defined from 500 known protein coding sequences (30 bp upstream of the AUG initiation codon) in *P. aeruginosa* by using the CONSENSUS program (3, 11). For the overall gene organization for each phage genome, see Fig. S1 in the supplemental material. A total of 1,894 ORFs are predicted from the genomes of the 18 phages (Table 1). The gene maps demonstrate that the coding regions are tightly packed, with very few intergenic spaces between them (see Fig. S1 in the supplemental material). The average gene coding potential of each phage genome is 93.0% (Table 1), with approximately 1.5 genes per kbp, slightly lower than what has been reported for staphylococcal phages (1.67 genes/kbp) (5) and *M. tuberculosis* phages (1.69 genes/kbp) (9).

Several important points emerged when the predicted proteins were examined for similarity to known bacterial and bacteriophage sequences that were deposited in public databases (see Table S1 in the supplemental material). First, the phage proteomes are rich sources of untapped protein sequence diversity. The biological function of a large proportion of predicted proteins (1,562 genes; 82% of the proteome) cannot be determined by comparison to current entries in genome databases (Table 1). Only 332 predicted proteins (18%) can be structurally or functionally annotated (Table 1). Second, 449

* Corresponding author. Mailing address: McIntyre Medical Sciences Building, Room 810, 3655 Promenade Sir William Osler, McGill University, Montreal, Quebec, Canada H3G 1Y6. Phone: (514) 398-2323. Fax: (514) 398-7384. E-mail: jerry.pelletier@mcgill.ca.

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

TABLE 1. Bacteriophage genomics summary[a]

Phage	Genome size (bp)	% GC	Predicted genes	% Coding	Gene density (genes/kbp)	No. of genes with:			Avg gene size (nt)
						Putative function	Unknown function	NDM	
F10	39,199	62.1	80	96.2	2.04	25	55	49	193
PA73	42,999	53.6	75	96.9	1.74	12	63	54	226
119X	43,365	44.9	56	94.3	1.29	3	53	3	258
PA11	49,639	44.8	71	90.7	1.43	18	53	38	222
M6	59,446	64.5	110	96.3	1.85	13	97	80	261
F8	66,015	54.9	109	93.4	1.65	17	92	74	226
PA7	72,009	55	116	95	1.61	15	101	87	246
PA16	72,010	55	101	94.4	1.4	16	85	73	246
SD1-M	281,083	36.9	367	91.1	1.31	32	335	66	233
phiCTX	35,580	62.6	47	91.3	1.32	29	18	12	233
D3112	37,611	64.3	55	93.8	1.46	16	39	17	215
B3	38,439	63.2	59	97.1	1.53	21	38	18	216
phiKMV	42,519	62.3	48	90.1	1.13	13	35	27	269
PaP2	43,783	45.4	58	92.4	1.32	2	56	49	249
PaP3	45,503	52.2	71	91.6	1.56	18	53	71	197
D3	56,425	57.8	95	89.1	1.68	29	66	40	179
F116	65,195	63.2	70	92.9	1.07	22	48	32	289
phiKZ	280,334	36.8	306	87.1	1.09	31	275	260	266
Total	1,371,154		1,894			332	1,562	1,050	
Average		54.4		93.0	1.47				235

[a] NDM, no database match; nt, nucleotides.

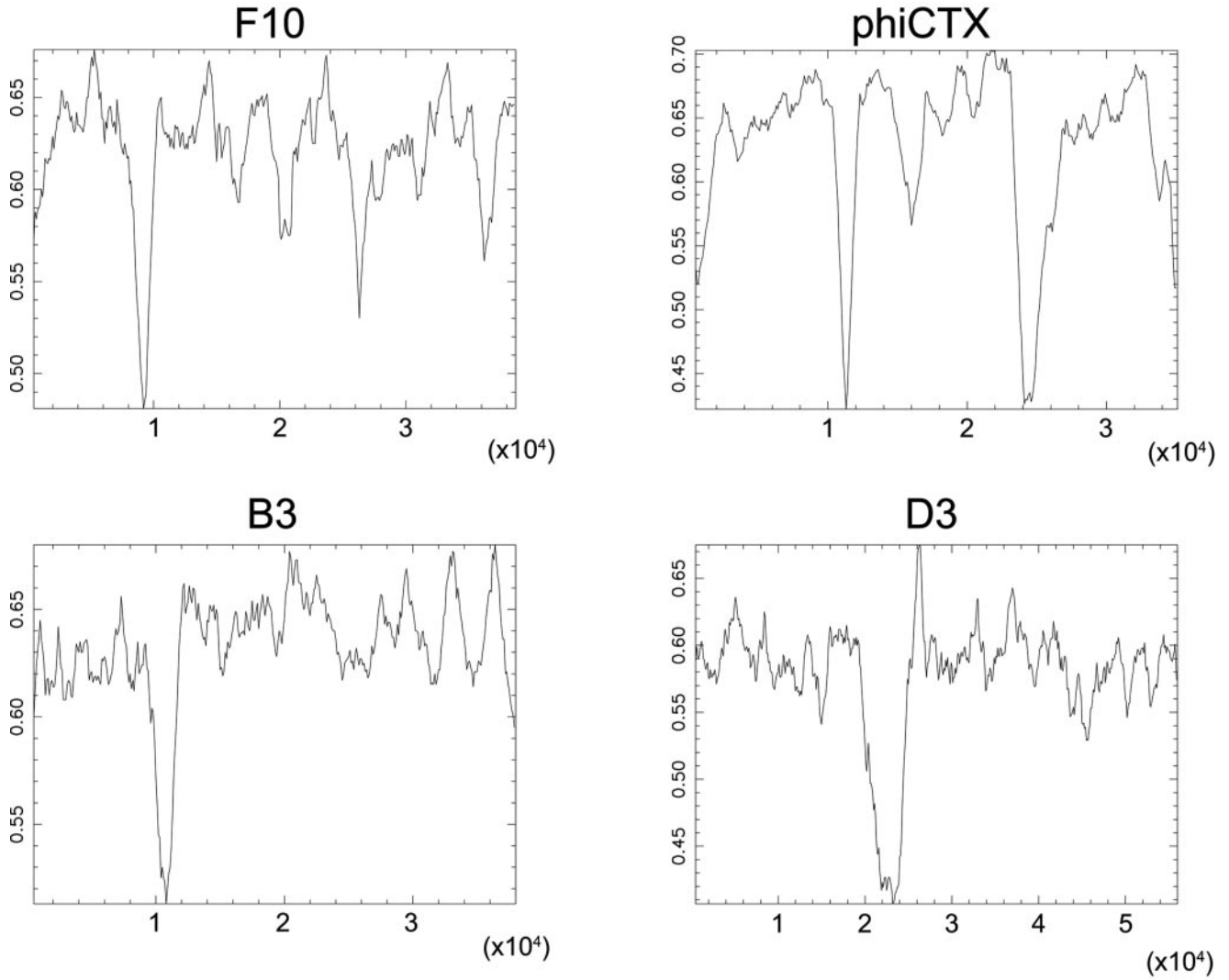

FIG. 1. Compositional analysis of four *P. aeruginosa* bacteriophage genomes. The G+C contents (vertical axes) are plotted over the entire length of the sequence (horizontal axes) for phages F10, phiCTX, B3, and D3. The plots were made with the Isochore program of the EMBOSS package with a sliding window size of 500 bp.

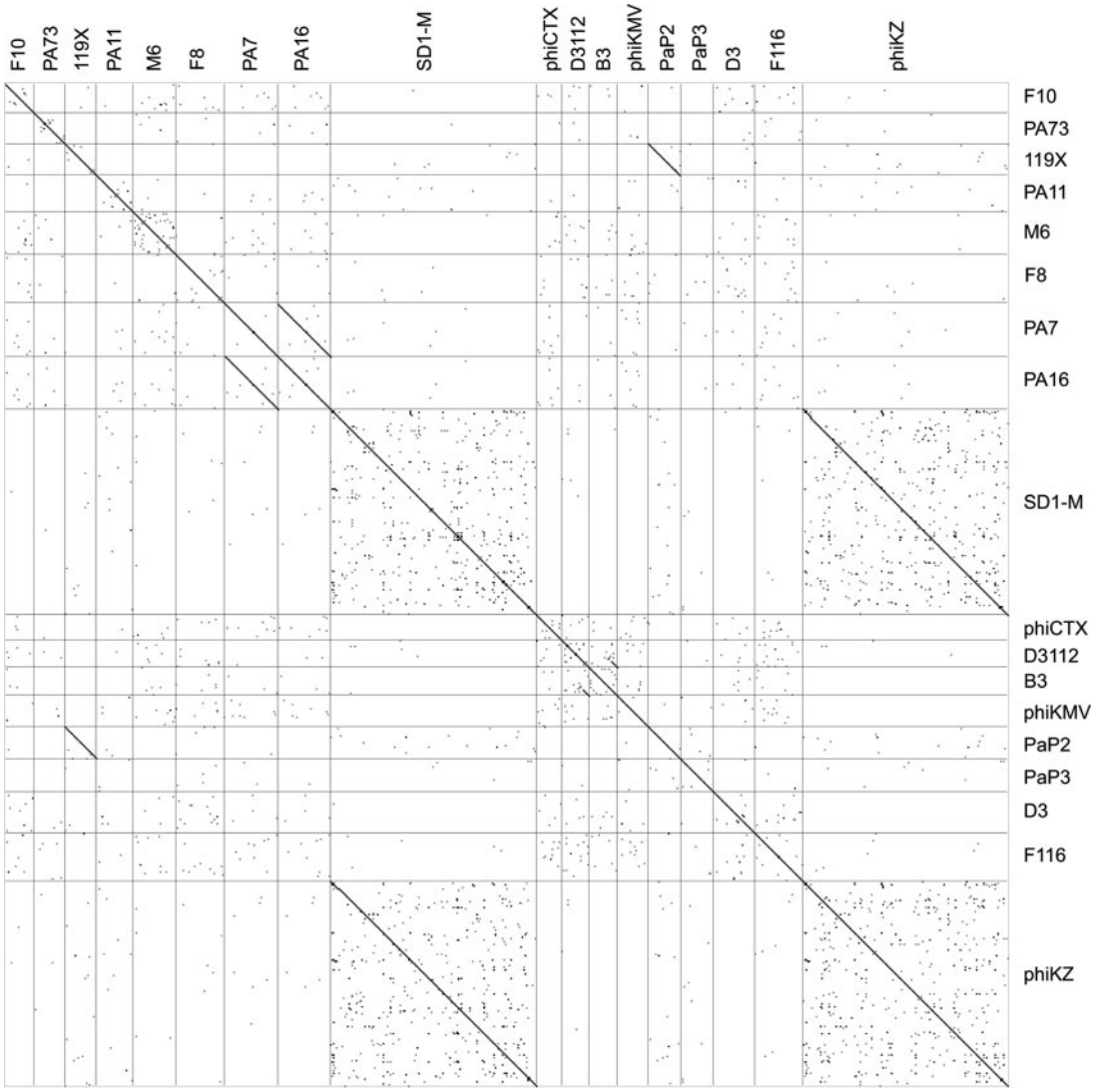

FIG. 2. Comparative nucleotide sequence analysis of *P. aeruginosa* bacteriophage genomes. Dot matrices comparing the relatedness of the nucleotide sequences of each phage genome were generated with the software program Dotter (10) using a sliding window of 25 bp.

ORFs (24%) show sequence similarity to only other ORFs encoded by *P. aeruginosa* phage genomes (data not shown). This number is higher than the proportion of phage genes with identifiable homologs among other phages (295 genes; 16% of the proteome) or within the *P. aeruginosa* host genome (18 genes; 1% of the proteome), indicating that gene transfer among *P. aeruginosa* phages may be more predominant than recombination between *P. aeruginosa* phages and phages of other species or between *P. aeruginosa* phages and their host. Third, a significant proportion of predicted ORFs (1,050 genes; 55% of the proteome) are unique to the *P. aeruginosa* phage proteome reported herein and show no database match (Table 1) to any publicly available prokaryotic sequence.

A pairwise comparison of the nucleotide sequences (Fig. 2) and of the proteomes (Table 2) of the 18 phages reveals that with three exceptions, the *P. aeruginosa* phages show little sequence relatedness to each other (Fig. 2) These three exceptions are PA7 and PA16 (90% identity), 119X and PaP2 (93% identity), and SD1-M and phiKZ (99% identity) (Fig. 2). Phages SD1-M and phiKZ share almost 100% identity over their entire sequence, with the exception of two insertions (a 1,095-bp segment in SD1-M and a 345-bp segment in phiKZ) that show no homology to any known bacterial or phage nucleotide sequences and do not encode any proteins of known function.

By providing primary sequence information from a large group of *P. aeruginosa* phages, this report not only provides a compendium of novel protein sequences, but also sets the stage for future studies aimed at better understanding virus/host relationships.

Nucleotide sequence accession numbers. The *P. aeruginosa* bacteriophage genomes have been deposited into the NCBI and assigned the following GenBank accession numbers: DQ163912 (F10), DQ163913 (PA73), DQ163914 (119X), DQ163915 (PA11), DQ163916 (M6), DQ163917 (F8), DQ163918 (PA7), DQ163919 (PA16), and DQ163920 (SD1-M).

TABLE 2. Shared *P. aeruginosa* bacteriophage genes[a]

Source	Phage	No. of shared genes from different pairs of indicated bacteriophages																	
		In house									GenBank								
		F10	PA73	119X	PA11	M6	F8	PA7	PA16	SD1-M	phiCTX	D3112	B3	phiKMV	PaP2	PaP3	D3	F116	phiKZ
GenBank	F10		1			1	1	3	3		1	3					8	2	
	PA73	**1**				19	1					5	5						
	119X					2	1	1	1			1			54				
	PA11					2								4		11			
	M6	**1**	**9**				6					9	9		2	1			
	F8	**1**				**3**									1		1		
	PA7								103			2		2	1				
	PA16							**98**				2		2	1				
	SD1-M										1						1		340[b]
In house	phiCTX																2		1
	D3112	**1**	**1**			**4**							18		1		2	2	
	B3		**2**			**4**						**11**		1				1	
	phiKMV				**1**											4			
	PaP2			**53**															
	PaP3				**5**									**1**					
	D3	**5**										**1**						5	1
	F116		**1**														**1**		
	phiKZ									**309**[b]									

[a] The numbers in roman type and bold type are derived from BLAST cutoff E values of 10^{-4} and 10^{-20} respectively. The phages are grouped according to in-house-sequenced double-stranded DNA phages and publicly available double-stranded DNA phages from GenBank.

[b] Although phiKZ has only 306 predicted genes, the numbers of shared genes between phiKZ and SD1-M at BLAST cutoff E values of 10^{-4} and 10^{-20} are 340 and 309, respectively, which is higher than the actual number of genes within phiKZ. This is due to the fact that a number of phiKZ genes have BLAST hits to multiple SD1-M genes and are therefore counted multiple times.

We thank the anonymous reviewers for their helpful suggestions in improving the manuscript. We thank H.-W. Ackermann (Felix d'Hérelle Reference Center for Bacterial Viruses, Québec City, Canada) for kindly providing many of the *Pseudomonas* phages used in this study.

We thank the National Research Council (Canada) Industrial Research Assistance Program for partial support of our research program. T.K. and J.L. are recipients of Natural Sciences and Engineering Research Council of Canada Industrial Research Fellowships.

We are grateful to the scientific and support personnel of Targanta Therapeutics for their contributions to this project.

This work was performed while M.D., P.G., and J.P. were scientific consultants.

REFERENCES

1. **Braid, M. D., J. L. Silhavy, C. L. Kitts, R. J. Cano, and M. M. Howe.** 2004. Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. J. Bacteriol. **186:**6560–6574.
2. **Byrne, M., and A. M. Kropinski.** 2005. The genome of the Pseudomonas aeruginosa generalized transducing bacteriophage F116. Gene **346:**187–194.
3. **Hertz, G. Z., G. W. Hartzell III, and G. D. Stormo.** 1990. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. Comput. Appl. Biosci. **6:**81–92.
4. **Kropinski, A. M.** 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J. Bacteriol. **182:** 6066–6074.
5. **Kwan, T., J. Liu, M. DuBow, P. Gros, and J. Pelletier.** 2005. The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. Proc. Natl. Acad. Sci. USA **102:**5174–5179.
6. **Lavigne, R., M. V. Burkal'tseva, J. Robben, N. N. Sykilinda, L. P. Kurochkina, B. Grymonprez, B. Jonckx, V. N. Krylov, V. V. Mesyanzhinov, and G. Volckaert.** 2003. The genome of bacteriophage phiKMV, a T7-like virus infecting Pseudomonas aeruginosa. Virology **312:**49–59.
7. **Mesyanzhinov, V. V., J. Robben, B. Grymonprez, V. A. Kostyuchenko, M. V. Bourkaltseva, N. N. Sykilinda, V. N. Krylov, and G. Volckaert.** 2002. The genome of bacteriophage phiKZ of Pseudomonas aeruginosa. J. Mol. Biol. **317:**1–19.
8. **Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi.** 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of Pseudomonas aeruginosa: implications for phage evolution and horizontal gene transfer via bacteriophages. Mol. Microbiol. **31:**399–419.
9. **Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:**171–182.
10. **Sonnhammer, E. L., and R. Durbin.** 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:**GC1–GC10.
11. **Stormo, G. D., and G. W. Hartzell III.** 1989. Identifying protein-binding sites from unaligned DNA fragments. Proc. Natl. Acad. Sci. USA **86:**1183–1187.
12. **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.