

Corrected Sequence of the Bacteriophage P22 Genome

Marisa L. Pedulla,^{1,2} Michael E. Ford,^{1,2} Tharun Karthikeyan,^{1,2} Jennifer M. Houtz,^{1,2}
Roger W. Hendrix,^{1,2} Graham F. Hatfull,^{1,2} Anthony R. Poteete,^{1,3}
Eddie B. Gilcrease,⁴ Danella A. Winn-Stapley,⁴
and Sherwood R. Casjens^{1,4*}

*Pittsburgh Bacteriophage Institute*¹ and *Department of Biological Sciences*,² *University of Pittsburgh, Pittsburgh, Pennsylvania 15260*; *Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts 01655*³; and *Department of Pathology, University of Utah Medical School, Salt Lake City, Utah 84132*⁴

Received 30 September 2002/Accepted 21 November 2002

We report the first accurate genome sequence for bacteriophage P22, correcting a 0.14% error rate in previously determined sequences. DNA sequencing technology is now good enough that genomes of important model systems like P22 can be sequenced with essentially 100% accuracy with minimal investment of time and resources.

Since its discovery 50 years ago, bacteriophage P22, a double-stranded DNA tailed phage of *Salmonella enterica* serovar Typhimurium (32), has been a prominent model system used in investigations of numerous facets of molecular biology (5, 16, 28). Because of its importance as an experimental system, its genome was originally sequenced in 27 different fragments by many different laboratories starting 20 years ago, but mostly with early technology that was less accurate than current methods (1–4, 6, 8–13, 15, 17–27, 29–31; M. Kroeger and G. Hobom, unpublished data [GenBank accession no. X78401]; M. Sranko and M. Susskind, personal communication). Because of the continuing relevance of both past results and ongoing investigations, such as comparative genomic studies, evolutionary studies, structural studies, etc., we believe that it is important to know the P22 genome sequence accurately. Vander Byl and Kropinski (30) resequenced a few parts of the P22 genome and reported an updated P22 sequence and the resolution of 17 ambiguities among the previously reported sequences. Our independent comparison of the original 27 sequence fragments revealed 28 sequence discrepancies. All 28 were resolved by reanalyzing the original data from our laboratories and by designing oligonucleotide primers to program sequencing reactions across each of the discrepancies by using wild-type P22 virion DNA as the template. Our wild-type P22 came directly from David Botstein's strain collection.

These 28 sequence discrepancies came from disagreements in areas where the individual published sequences overlapped. Since these overlap regions in aggregate covered only 21% of the genome, it seemed likely that there would be additional errors in the remainder of the published sequence. Thus, when we accidentally obtained sequence information from an extremely close relative of P22 (P22-pbi [see below]) which contaminated a preparation of another *Salmonella* phage, we decided to continue collecting sequence information for this

phage until its genome sequence was complete. Shotgun sequencing, performed as previously described (14) with an ABI 3100 capillary sequencer, was continued to 10.2-fold coverage of the whole genome, with complete coverage on both strands, yielding a circular sequence 41,724 bp long. This sequence had 48 differences from the “resolved” P22 genome sequence (Table 1). Our previous experience with contemporary sequencing methodology led us to expect that this newly determined sequence of P22-pbi should be essentially error free as a result of the inherent advantages of shotgun sequencing and the ease of collecting data to high redundancy. The fortuitous availability of two independent determinations of the same genome sequence provided an efficient way to test that assumption, since it allowed attention to be focused specifically on the quality of data at the sites of disagreement. Careful reexamination of the P22-pbi data at each of the 48 sites of disagreement showed that the new sequence was unambiguous in each case.

Phages are so extremely diverse that individuals this similar are very rarely, if ever, independently isolated, so we suspect that P22-pbi is almost certainly a feral version of the original P22 that escaped into the laboratory. If this is true, then some of the differences between P22-pbi and wild-type P22 might be the result of errors in the previously reported sequence. We therefore designed oligonucleotide primers to program sequencing reactions across each of the above-mentioned 48 differences and several other regions of particular interest to us and determined the sequences in these regions by using authentic wild-type P22 virion DNA as the template. In this way, we obtained unambiguous sequence for about 59% of the wild-type P22 genome (mostly outside of the previous overlaps) and found that 44 of the 48 apparent differences between the sequences of our resolved wild-type P22 and P22-pbi were in fact not actual differences at all. These 44 differences are therefore errors (which could be due to sequencing errors, strain differences, or cloning artifacts) in the previously reported P22 sequence. The four authentic differences between P22-pbi and wild-type P22, at bp 13508, 15038, 21030, and 31519 (Table 1), appear to have arisen since P22-pbi's escape.

* Corresponding author. Mailing address: Department of Pathology, Room 5C240, University of Utah Medical School, 30 N. 1900 East, Salt Lake City, UT 84132-2501. Phone: (801) 581-5980. Fax: (801) 581-3607. E-mail: sherwood.casjens@path.utah.edu.

TABLE 1. Corrections to the bacteriophage P22 nucleotide sequence

No.	Resolved P22 ^a		P22-pbi		P22+ Base	Altered gene
	Base	Amino acid	Base	Amino acid		
3378	A	L	G	L	G ^d	1
7484	A	S	G	G	G	10
8714	C	L	G	V	G	26
9226	G	A	C	R	C	14
9227	C	A	G	R	G	14
9228	G	A	C	R	C	14
9231	G	D	C	H	C	14
9375	C	H	T	Y	T	7
11894	A	G	C	G	C	16
12670	C	A	A	D	A	16
12675	T	S	G	A	G	16
13508	C	R	T	K	C	<i>sieA</i>
13918	— ^b	X ^c	G	X	G	Between <i>sieA</i> and <i>orf59a</i>
13927/8	C	X	—	X	—	Between <i>sieA</i> and <i>orf59a</i>
13899/900	G	X	—	X	—	Between <i>sieA</i> and <i>orf59a</i>
14009	—	X	C	X	C	Between <i>sieA</i> and <i>orf59a</i>
15038	T	X	C	X	T	Between <i>arc</i> and <i>ant</i>
18400	C	E	G	Q	G	<i>gtrC</i>
18401	G	I	C	M	C	<i>gtrC</i>
21030	G	X	A	X	G	Between <i>gtrA</i> and <i>int</i>
22784	G	M	C	I	C	<i>eaC</i>
22785	C	P	G	A	G	<i>eaC</i>
30839/40	—	X	G	X	G	Between 24 and <i>c2</i>
30984	A	X	—	X	—	Between 24 and <i>c2</i>
30998	T	X	—	X	—	Between 24 and <i>c2</i>
31519	T	N	C	S	T	<i>c2</i>
34903	C	A	T	A	T	<i>ninB</i> (<i>orf145</i>)
35024	C	H	A	N	A	<i>ninB</i> (<i>orf145</i>)
35053	T	T	C	T	C	<i>ninB</i> (<i>orf145</i>)
35195	T	C	G	C	G	<i>ninD</i> (<i>orf57</i>)
35629	T	V	A	V	A	<i>ninX</i> (<i>orf112</i>)
35941	G	G	C	A	C	<i>ninF</i> (<i>orf58b</i>)
35942	C	G	G	A	G	<i>ninF</i> (<i>orf58b</i>)
35943	G	G	T	C	T	<i>ninF</i> (<i>orf58b</i>)
36199	G	E	A	K	A	<i>ninG</i> (<i>orf203</i>)
36441	T	L	G	L	G	<i>ninG</i> (<i>orf203</i>)
37519	T	A	C	A	C	23
37634	T	F	G	V	G	23
37641	G	C	A	Y	A	23
37650	C	A	T	V	T	23
37663	A	A	C	A	C	23
37671	T	L	G	R	G	23
37673	T	S	G	A	G	23
37676	G	V	A	I	A	23
37892	A	X	—	X	—	Between 23 and 13
37965	C	X	T	X	T	Between 23 and 13
38073/4	—	X	G	X	G	Between 23 and 13
38111	T	X	A	X	A	Between 23 and 13

^a The numbering system is that used for the sequence assigned GenBank accession no. TPA:BK000583 and is based on that suggested by Eppler et al. (9). See the text for the definition of "Resolved P22."

^b A dash (—) indicates that no nucleotide is present at a position where one of the other sequences has one.

^c X, no amino acid is encoded by this region.

^d The G at 3378, a genuine A in the published sequence, is apparently the result of a plasmid cloning artifact.

Are there other errors in the 41% of the wild-type P22 sequence that we did not resequence? We do not think so, since there is no sequence difference between wild-type P22 and P22-pbi in this 41% of the genome and we found no additional discrepancies in the 59% of the wild-type genome that we resequenced. Any other actual error in the previously reported P22 sequence is prohibitively unlikely to have mutated by chance to the same nucleotide in P22-pbi. It is therefore extremely likely that the identity between P22 and P22-pbi represents the true wild-type P22 sequence in these regions

and that the sequence we report here is the correct wild-type P22 sequence. This sequence accurately predicts all 258 experimentally mapped cleavage sites for 46 different restriction enzymes (reference 7 and references therein; S. Casjens, unpublished data); the one previous discrepancy, the absence of an experimentally observed *Xmn*I site at bp 35025, is resolved by the creation of this site in the corrected sequence. Thirteen of the 44 corrections to the wild-type P22 sequence are between genes, and 31 are within genes; among the latter, 23 are changes to nonsynonymous codons and therefore alter the

predicted amino acid sequences of the encoded proteins. This amounts to corrections to the amino acid sequences of 15% of the proteins encoded by the P22 genome. It seems likely that most of the differences between the wild-type P22 sequence reported here and the original P22 sequence are the result of errors in the original sequence rather than strain differences, since a majority of the corrections result in changes to nonsynonymous codons. If coding sequence differences among P22 strains are like those among other closely related organisms, then most would be synonymous. For example, the DNA sequences of gene 23 of P22 (where a number of the differences between P22 and P22-pbi lie) and gene *Q* of phage λ differ by 30 nucleotides, but the encoded protein sequences differ by only six amino acids. We have submitted this corrected bacteriophage P22 sequence to GenBank as an updated and fully annotated complete genome sequence.

Nucleotide sequence accession numbers. The corrected bacteriophage P22 sequence has been assigned GenBank accession no. TPA:BK000583, the 41,724-bp circular sequence of phage P22-pbi has been assigned GenBank accession no. AF527608, and the unambiguous sequence for approximately 59% of the wild-type P22 genome has been assigned GenBank accession no. AY121859, AY121860, AY121861, AY121862, AY121863, and AY121864.

We thank David Botstein for wild-type P22 and Miriam Susskind for access to unpublished information.

This work was supported by NSF grant 990526 to S.R.C. and NIH grants GM51975 to R.W.H. and G.F.H. and GM51609 to A.R.P.

REFERENCES

- Adhikari, P., and P. B. Berget. 1993. Sequence of a DNA injection gene from *Salmonella typhimurium* phage P22. *Nucleic Acids Res.* **21**:1499.
- Backhaus, H. 1985. DNA packaging initiation of *Salmonella* bacteriophage P22: determination of cut sites within the DNA sequence coding for gene 3. *J. Virol.* **55**:458–465.
- Backhaus, H., and J. B. Petri. 1984. Sequence analysis of a region from the early right operon in phage P22 including the replication genes 18 and 12. *Gene* **32**:289–303.
- Berget, P. B., A. R. Poteete, and R. T. Sauer. 1983. Control of phage P22 tail protein expression by transcription termination. *J. Mol. Biol.* **164**:561–572.
- Casjens, S. 1989. Bacteriophage P22 DNA packaging, p. 241–261. In K. Adolph (ed.), *Chromosomes: eukaryotic, prokaryotic and viral*. CRC Press, Boca Raton, Fla.
- Casjens, S., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene 19 to 3 region: identification of a new gene required for lysis. *Virology* **171**:588–598.
- Casjens, S., M. Hayden, E. Jackson, and R. Deans. 1983. Additional restriction endonuclease cleavage sites on the bacteriophage P22 genome. *J. Virol.* **45**:864–867.
- Conlin, C. A., E. R. Vimr, and C. G. Miller. 1992. Oligopeptidase A is required for normal phage P22 development. *J. Bacteriol.* **174**:5869–5880.
- Eppler, K., E. Wyckoff, J. Goates, R. Parr, and S. Casjens. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* **183**:519–538.
- Franklin, N. C. 1985. Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages λ , ϕ 21 and P22. *J. Mol. Biol.* **181**:75–84.
- Hofer, B., M. Ruge, and B. Dreiseikelmann. 1995. The superinfection exclusion gene (*sieA*) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. *J. Bacteriol.* **177**:3080–3086.
- Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1985. The ϕ 80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. *J. Biol. Chem.* **260**:4468–4477.
- Leong, J. M., S. E. Nunes-Duby, A. B. Oser, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1986. Structural and regulatory divergence among site-specific recombination genes of lambdoid phage. *J. Mol. Biol.* **189**:603–616.
- Morgan, G., G. Hatfull, S. Casjens, and R. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J. Mol. Biol.* **317**:337–360.
- Murphy, K. C., A. C. Fenton, and A. R. Poteete. 1987. Sequence of the bacteriophage P22 anti-recBCD (abc) genes and properties of P22 abc region deletion mutants. *Virology* **160**:456–464.
- Poteete, A. R. 1988. Bacteriophage P22, p. 647–682. In R. Calendar (ed.), *The bacteriophages*, vol. II. Plenum Press, New York, N.Y.
- Poteete, A. R. 1982. Location and sequence of the *erf* gene of phage P22. *Virology* **119**:422–429.
- Poteete, A. R., K. Hehir, and R. T. Sauer. 1986. Bacteriophage P22 Cro protein: sequence, purification, and properties. *Biochemistry* **25**:251–256.
- Poteete, A. R., M. Ptashne, M. Ballivet, and H. Eisen. 1980. Operator sequences of bacteriophages P22 and λ . *J. Mol. Biol.* **137**:81–91.
- Poteete, A. R., and T. M. Roberts. 1981. Construction of plasmids that produce phage P22 repressor. *Gene* **13**:153–161.
- Ranade, K., and A. R. Poteete. 1993. Superinfection exclusion (*sieB*) genes of bacteriophages P22 and λ . *J. Bacteriol.* **175**:4712–4718.
- Rennell, D., and A. R. Poteete. 1985. Phage P22 lysis genes: nucleotide sequences and functional relationships with T4 and λ genes. *Virology* **143**:280–289.
- Sampson, L., and S. Casjens. 1993. Nucleotide sequence of *Salmonella* bacteriophage P22 head completion genes 10 and 26. *Nucleic Acids Res.* **21**:3326.
- Sauer, R. T., W. Krovatin, J. DeAnda, P. Youderian, and M. M. Susskind. 1983. Primary structure of the *imm1* immunity region of bacteriophage P22. *J. Mol. Biol.* **168**:699–713.
- Sauer, R. T., W. Krovatin, A. R. Poteete, and P. B. Berget. 1982. Phage P22 tail protein: gene and amino acid sequence. *Biochemistry* **21**:5811–5815.
- Sauer, R. T., J. Pan, P. Hopper, K. Hehir, J. Brown, and A. R. Poteete. 1981. Primary structure of the phage P22 repressor and its gene *c2*. *Biochemistry* **20**:3591–3598.
- Semerjian, A. V., D. C. Malloy, and A. R. Poteete. 1989. Genetic structure of the bacteriophage P22 P_L operon. *J. Mol. Biol.* **207**:1–13.
- Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. *Microbiol. Rev.* **42**:385–413.
- Umlauf, B., and B. Dreiseikelmann. 1992. Cloning, sequencing, and overexpression of gene 16 of *Salmonella* bacteriophage P22. *Virology* **188**:495–501.
- Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. *J. Bacteriol.* **182**:6472–6481.
- Wulff, D. L., Y. S. Ho, S. Powers, and M. Rosenberg. 1993. The *int* genes of bacteriophages P22 and λ are regulated by different mechanisms. *Mol. Microbiol.* **9**:261–271.
- Zinder, N., and J. Lederberg. 1952. Genetic exchange in *Salmonella*. *J. Bacteriol.* **64**:679–699.