NOTES

Sequencing Bacillus anthracis Typing Phages Gamma and Cherry Reveals a Common Ancestry

Derrick E. Fouts,1* David A. Rasko,1‡ Regina Z. Cer,1 Lingxia Jiang,1 Nadia B. Fedorova,1 Alla Shvartsbeyn,1 Jessica J. Vamathevan,1 Luke Tallon,1 Ryan Althoff,† Tamara S. Arbogast,1 Douglas W. Fadrosh,1 Timothy D. Read,2 and Steven R. Gill1†

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850,1 and Biological Defense Research Directorate, Naval Medical Research Center, 12300 Washington Avenue, Rockville, Maryland 208512

Received 30 November 2005/Accepted 22 February 2006

The genetic relatedness of the Bacillus anthracis typing phages Gamma and Cherry was determined by nucleotide sequencing and comparative analysis. The genomes of these two phages were identical except at three variable loci, which showed heterogeneity within individual lysates and among Cherry, Wβ, Fah, and four Gamma bacteriophage sequences.

Bacteriophages are useful tools for bacterial species and strain differentiation (7, 20). Gamma phage (Wγ) susceptibility is an initial test for differentiating Bacillus anthracis from closely related Bacillus cereus group species (1, 3). Even though lysis is highly specific for B. anthracis, there are a few B. cereus strains that can be infected by Gamma phage (8, 29). The history of Gamma phage is quite complex. McCloy isolated strains that can be infected by Gamma phage (8, 29). The phage, could infect both plaques on 11950 and found it to be somewhat infecting only a few strains of B. anthracis unable to infect Bacillus and rough (nonencapsulated) properties of being able to infect both smooth (encapsulated) strain W with a lysate of W phage (3). It has the unique capsule (21), limiting their usefulness as typing phages. Wβ formed turbid plaques on B. anthracis and failed to infect the original source, B. cereus strain W; however, a rare clear plaque mutant, called Wα, could infect both B. anthracis and B. cereus strain W. Both Wβ and Wα could infect only B. anthracis specific, infecting only a few strains of B. cereus (21). Wβ formed turbid plaques on B. anthracis and failed to infect the original source, B. cereus strain W; however, a rare clear plaque mutant, called Wα, could infect both B. anthracis and B. cereus strain W. Both Wβ and Wα could infect only B. anthracis strains that lacked a capsule (21), limiting their usefulness as typing phages. Gamma phage was originally isolated by Brown and Cherry in 1955 as a W phage variant formed by reinfesting B. cereus strain W with a lysate of W phage (3). It has the unique properties of being able to infect both smooth (encapsulated) and rough (nonencapsulated) B. anthracis strains and being unable to infect Bacillus strains that are lysogenic for Wβ phage. Since many B. anthracis strains are encapsulated, Wγ became a valuable tool for typing B. anthracis strains. Another B. anthracis phage called Cherry phage has also been used for typing, albeit less frequently (13); however, its relationship to Gamma phage was not known. Gamma and Cherry phages appear identical under the electron microscope, and both belong to the Siphoviridae morphotype (13, 36).

We completed and analyzed the nucleotide sequences of the Gamma and Cherry phages to determine their genetic relatedness. During the course of sequencing, it was determined through restriction enzyme mapping and PCR experiments that the stock phage preparations were heterogeneous, which led to the acquisition and sequencing of a second Gamma phage preparation from USAMRIID. Comparison of the complete genome sequences has revealed the location of three distinct variable genetic loci. These variable loci were also compared with the sequences of Wβ, Wγγ, WγΦ, and Fah (Table 1). Overall, this work provides a striking example of how diagnostic bacteriophages can evolve over several years in different laboratories.

Bacterial strains and phage DNA isolation. Gamma-LSU phage (Wγγ) and Cherry phage (WγΦ) DNA were provided by Pamala R. Coker while at Louisiana State University. The phages were propagated on B. anthracis strain Vollum by plating on Trypticase soy agar with 5% sheep blood (Remel, Kansas) followed by amplification in nutrient broth. Bacterial cells were removed from the lysate by filtration through a 0.22-μm syringe filter prior to isolation of bacteriophage genomic DNA. A stock Gamma-USAMRIID phage (WγΦ) lysate was obtained from John Ezzell, USAMRIID, Fort Detrick, MD, and propagated on B. cereus ATCC 4342. A single isolated plaque was picked after overnight growth from a lawn of B. cereus ATCC 4342 using the agar layer method (2). Bacteriophages from this plaque were propagated on B. cereus ATCC 4342 on agar plates (2). The resulting cell lysate was passed over DE52 cellulose resin to remove unpackaged, contaminating nucleic acids in the lysate (18). The flowthrough was then filtered through a 0.22-μm syringe filter to remove bacterial cells. Wγγ and WγΦ genomic DNA was purified using a QIAGEN Lambda DNA extraction kit (QIAGEN, Germany). The DNA extraction procedure was modified from the QIAGEN Lambda DNA extraction kit (QIAGEN, Germany) by resuspending the polyethylene glycol phage pellet in 215 μl of buffer L4 and 4.3
Phage variant

<table>
<thead>
<tr>
<th>Locus II (bp), form, and source or accession no.</th>
<th>Gamma-Porton (W/H9253)</th>
<th>Gamma-d'Herelle (W/H9253)</th>
<th>Gamma-Cherry (W/H9253)</th>
<th>Gamma-d'Herelle (W/H9253)</th>
<th>Gamma-Cherry (W/H9253)</th>
<th>Gamma-Cherry (W/H9253)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A† 301–336; II;</td>
<td>30512–31874; fosfomycin resistance (HMM)</td>
<td>31279–32641; fosfomycin resistance (HMM)</td>
<td>31150–32512; fosfomycin resistance (HMM)</td>
<td>31825–33199; fosfomycin resistance (HMM)</td>
<td>32641–34015; fosfomycin resistance (HMM)</td>
<td>33199–34573; fosfomycin resistance (HMM)</td>
</tr>
<tr>
<td>A† 29652–29664; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
</tr>
<tr>
<td>A† 29652–29664;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
</tr>
<tr>
<td>A† 29652–29664;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
<td>29664–29700; II;</td>
</tr>
</tbody>
</table>

Missing bp 28798–31971 (Cherry, W/H9253) 2840,864 24748–27392; A* 29818–29857; I; Emboss, VI; 

An asterisk indicates that the form was found in the deposited sequence; a dagger indicates that the form was not found in the deposited sequence but was observed with PCR. n.d., not determined (the locus was not found in the deposited sequence but was observed with PCR).
Gamma and Cherry phages package DNA using a 3’ overhang cos site mechanism. Since the large terminase protein of Gamma and Cherry phages grouped with phages of gram-positive bacteria having known 3’ overhang single-stranded cohesive (cos) ends and we observed no terminal redundancy in the genome sequence, which would have suggested a pac site mechanism of DNA packaging, we hypothesized that the Gamma/Cherry phage packages DNA using a 3’ overhang cos site mechanism. We tested this hypothesis by sequencing the PCR product that was formed after religation of the cos ends (Fig. 1).

FIG. 1. Linear representation of the Gamma/Cherry phage genomes. (A) A consensus molecule is depicted with the 3’ protruding cos ends indicated at the beginning and end of the molecule. The χ² value was determined and graphed below the ORFs. Each ORF is color coded based on predicted function. See the key for definitions. Regions highlighted yellow are areas of heterogeneity and are labeled with roman numerals. GenBank accession numbers are indicated to the right of each linear illustration. (B) A digital photograph of an ethidium bromide-stained 1% Tris-acetate-EDTA gel indicates the sizes of the three different forms observed within variable locus I. A 1-kb ladder was loaded into the leftmost lane, followed by forms A, B, and C. The PCR products were obtained through amplification using primers 10BE and 10AK (red arrows). The sequence of form A was determined by cloning and sequencing the PCR product from primers 10BB and 10AX on WY. Forms B and C were determined from the assemblies of WY and WY, respectively.
Two unique primers, P44087 (TCAATCTGACTAATTCA GCAGC) and P44086 (GATAGAGATAGATCTACGAC CC), were designed to face outward and read the DNA sequence of each end of the linear phage genomic DNA (Fig. 1). By comparing the sequence of the PCR product to the sequence of the ends, the sequence of the cos site, GCAGC CCCC (Fig. 1A), was determined to be 9 nucleotides in length, which is similar to the cos site of related Clostridium perfringens phage φ3626 (GGCGATTGTCGC) and identical to the cos site of B. anthracis bacteriophage Fah (22).

**Regions of heterogeneity.** We identified three heterogeneous loci while comparing the sequences of the Gamma and Cherry phages (Fig. 1A, yellow highlighted areas). We first became aware of heterogeneity near the integrase when performing confirmatory restriction mapping of the WγC genome from a plaque-purified phage preparation grown on B. cereus ATCC 4342 (Table 1, locus I). The map revealed additional DNA that was not included in the Cherry phage assembly (data not shown). Primers 10BB (AATTGTATCATCGAGTATTAAT GA) and 10AX (TGTAAGTATCGATACCTAATCG) were designed to subclone this contesting region using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA), for the production of a microlibrary for sequencing and for primer walking of the PCR product. For diagnostic purposes, primers 10BE (TGTGGTG AGC) and 10AK (TTTGGTATCATCGATACCTAATCG) were designed nested sets of the primers P44705 (TGATTTTCTA CTACACTTCC) to first amplify a 2,165-bp product and then primer P41871 (CCCATACAATCTAAATGGGAG) and P41870 (TGCAATAACGTGTCGTGGT) to obtain high-quality sequence data close to the ambiguity region (Fig. 1). The sequences of these PCR products confirmed that the form II ambiguity region is linked to form A (WγC) and the form I ambiguity region is linked to form C (WγC). Since this study was completed, the sequences of two additional Gamma phage isolates (WγU [28] and WγP [unpublished]), Wβ (28), and Fah (22) have become available for comparison. With the addition of Fah, this locus was expanded to 13 amino acids, with a total of four different variations observed (Table 1).

A third locus of heterogeneity between Gamma and Cherry phages was identified during comparative analysis of the three phage genomes (Table 1 and Fig. 1A, locus III). WγU and WγC main assemblies have identical sequences in this region, while Wγδ and a 7,578-bp variant assembly from WγU (Fig. 1A) share a different sequence. This region in WγU/WγC encodes three proteins (GAMMAUSAM0038/CHERRY0036, GAMMAUSAM0039/CHERRY0037, and GAMMAUSAM0040/ CHERRY0038). Both GAMMAUSAM0038/CHERRY0036 and GAMMAUSAM0039/CHERRY0037 have matches to proteins with no known function from other phages. GAMMAUSAM0040/CHERRY0038 is predicted to encode a fosfomycin resistance protein (Table 1). It is unclear whether GAMMA USAM0040/CHERRY0038 is able to produce a functional protein, because the insertion of a cytosome nucleotide at position 67 caused a frameshift in both Wγδ and WγC; however, a nonframeshifted homolog, gp41 U, in Wγδ, was recently shown to confer fosfomycin resistance (28).

The equivalent region in WγL and a 7,578-bp assembly from WγU (Fig. 1) is larger than the region in the Wγδ and WγU main assemblies, encoding two proteins (GAMMALSU0036/GAMMAUSAM0007 and GAMMALSU0037/GAMMAUSAM 0008). GAMMALSU0036/GAMMAUSAM0007 is predicted to encode a 479-amino-acid protein with 95 copies of a G-X-X repeat that is found in members of the collagen superfamily and proteins that are structural components of the exosporium of B. anthracis (33) and B. cereus (35) spores and form a triple helix. The distribution of repeats has the structure [GXX]3-T-[GXX]37-P-[GXX]7-P-[GXX]137-T-[GXX]37. This open reading frame (ORF) is predicted to belong to the collagen repeat superfamily based on HMM (PF03191) and BLASTP matches. GAMMAL SU0037/GAMMAUSAM0008 is predicted to encode a 193-amino-acid protein that matches HMM PF07883, a cupin domain protein. In bacteria, proteins with one or two cupin domains, which form a beta barrel structure, can have either isomerase or transferase activity that modify cell wall carbohydrates. The best NCBI-BLASTP match is a hypothetical protein, CTC01899 from Clostridium tetani E88.

We propose that the Gamma phage encodes the collagen repeat protein either to function in host recognition or possibly...
FIG. 2. Putative attP of Gamma and Cherry phages. Predicted attP sites are depicted as stem-loop structures when present (A and B). The linear representations of Wy5 (A), Wy1 (B), and Wy/C-Wy+ (C) are color coded as in Fig. 1. The known attP site from φC31 (31) is provided for reference (D). The predicted structure of attB within BA1618 of B. anthracis is indicated (E) (R. Calendar, personal communication). Stem-loop structures aid in visualization of the inverted repeats that flank the core att sequence (red bold). Predicted secondary structures and their free energies are from MFOLD. The stop codon of the phage integrase gene is underscored, while the boldface type denotes the common sequence 5' of the breakpoint of each of the three forms observed.
to make the bacillus spore more stable, ensuring its survival under stress. It is also entirely possible that either the collagen repeat protein or the cupin domain protein or both account for under stress. It is also entirely possible that either the collagen repeat protein or the cupin domain protein or both account for under stress. It is also entirely possible that either the collagen repeat or the cupin domain protein or both account for under stress. It is also entirely possible that either the collagen repeat or a predicted secondary structure for the smallest attP cannot. The Gamma phage has not been shown to form lysogens in B. anthracis, but the allelic variant Wα has been shown to survive within B. anthracis spores (16). This phage-trapping phenomenon has been observed during infection of B. subtilis 3610 by the virulent phage Pβ (32) and by phage PBS1 in B. subtilis SB19 (34).

There is also the question of the origin of fosfomycin resistance and the collagen repeat/cupin domain regions. It is possible that through propagation of these phages on various hosts, in various labs, they acquired these loci via recombination with prophages that existed in the host genome. We have evidence that contradicts this hypothesis, because PCRs on B. cereus strain W and on a mitomycin-induced prophage from strain W (presumably Wβ) gave products for both regions (data not shown). This indicates that these two forms existed in the parental host strain W.

The Gamma/Cherry phage is predicted to encode serine recombinases. The type of recombinase encoded by a bacteriophage determines target site specificity. For example, tyrosine recombinases that have a tropism for tRNA genes typically have what appears to be a target site duplication flanking the ends of the integrated prophage genome, which corresponds to the core sequence of the att site. In contrast, serine recombinases have very small core att sites that are flanked by inverted repeats (31) and may or may not have any recognizable target site duplication. An in silico method to identify the type of recombinase is to use HMMs. GAMMAUSAM0027 of Wγ5, GAMMALSU0027 of Wγ4, and CHERRY0027 of WγC match PUB0255, an HMM model for serine recombinases, above the trusted cutoff. Multiple sequence alignments of the Gamma/Cherry recombinases with members of the serine recombinase family (data not shown) enabled recognition of the catalytic serine residue at amino acid residue 13.

Nucleotide sequence and structure of a putative attP site. Sequence analysis of the three forms (A, B, and C) near the integrase attP region of the Gamma/Cherry phage revealed a conserved breaking point 31 nucleotides downstream of the integrase stop codon (Fig. 2A, yellow ORF). Further inspection of this region revealed inverted repeats with the breakpoint in the center of predicted stem-loop structures (Fig. 2). It is common for serine recombinases to use inverted repeats as the substrate for integration (31). The MFOLD program (37) was used to calculate the structure and free energy of the putative attP region for two of the three forms (Fig. 1 and 2). The largest attP region (Fig. 1B and 2A, form A) was predicted to form the best inverted repeat and the most stable structure (ΔG = −10.6 kcal). The medium-sized attP region (Fig. 1B and 2B, form B) formed a stem-loop with predicted free energy of −5.4 kcal (Fig. 2). We were unable to find an inverted repeat or a predicted secondary structure for the smallest attP region (Fig. 1B and 2C, form C), suggesting that this form may have been derived through illegitimate recombination. Given this data, we hypothesize that form C is unable to integrate into attB, while forms A and B may be functional attP substrates capable of site-specific integration into attB. Form A is the ancestral form, since Wβ has this sequence (Table 1). Form A was shown to serve as a substrate for site-specific recombination in B. anthracis by targeting BA1618 (Fig. 2) (R. Calendar, personal communication). Bacteriophages Fah, WγC, and WγP also have form A (Table 1), suggesting that these phages are also capable of integration/excision reactions. Further studies are necessary to determine whether form B is a functional attP sequence.

We present the complete nucleotide sequences and a comparison of two B. anthracis-specific bacteriophages that are used for typing. To our surprise, we discovered heterogeneity within each of three phage lysate stocks used to make purified phage DNA for whole shotgun sequencing. There was also heterogeneity between Gamma phage stock lysates from two different sources and among Wβ, Wγ4, Wγ3, and Fah sequences. We conclude that the Gamma phage, Cherry phage, and Fah are essentially the same phage, containing variations at three distinct locations within the genome and demonstrating significant heterogeneity within their populations.

Nucleotide sequence accession numbers. The nucleotide sequences of B. anthracis Wγ4, WγU1, and WγC genomes and minor variant assemblies have been deposited at GenBank (http://www.ncbi.nlm.nih.gov/GenBank/) under accession numbers DQ222851 to DQ222855 and DQ294634.

We thank Pamela R. Coker for supplying the Gamma-LSU and Cherry phage lysates and purified genomic DNA preparations and John W. Ezzell for supplying the Gamma-USAMRIID phage lysate. We also thank Robert T. DeBoy and Eric Eisenstein for insightful discussions about the manuscript, Shu-mu Sozhamannan and Karen E. Nelson for critically reviewing the manuscript, and Richard Calendar for communicating unpublished results and for stimulating discussions on the history of the W phages. This work was supported by NSF grant 0242162.

REFERENCES


