The Wonderful World of Bacteriophage Genomics

Graham F. Hatfull

Department of Biological Sciences
University of Pittsburgh

Three topics

1. Phage genomes are pervasively mosaic

2. Grouping of phage genomes into Clusters and Subclusters

3. Little picture, big picture.

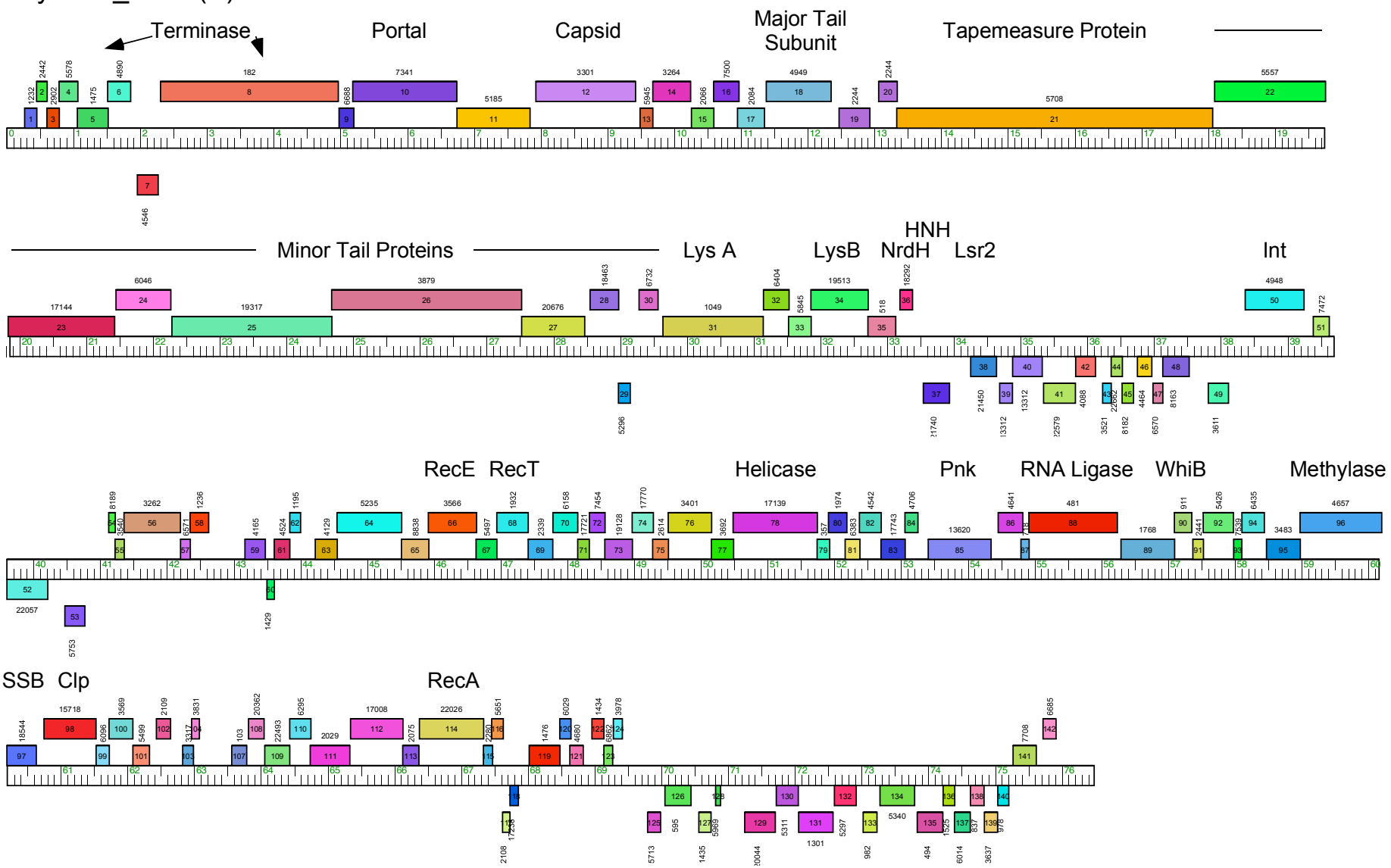
Three topics

1. Phage genomes are pervasively mosaic

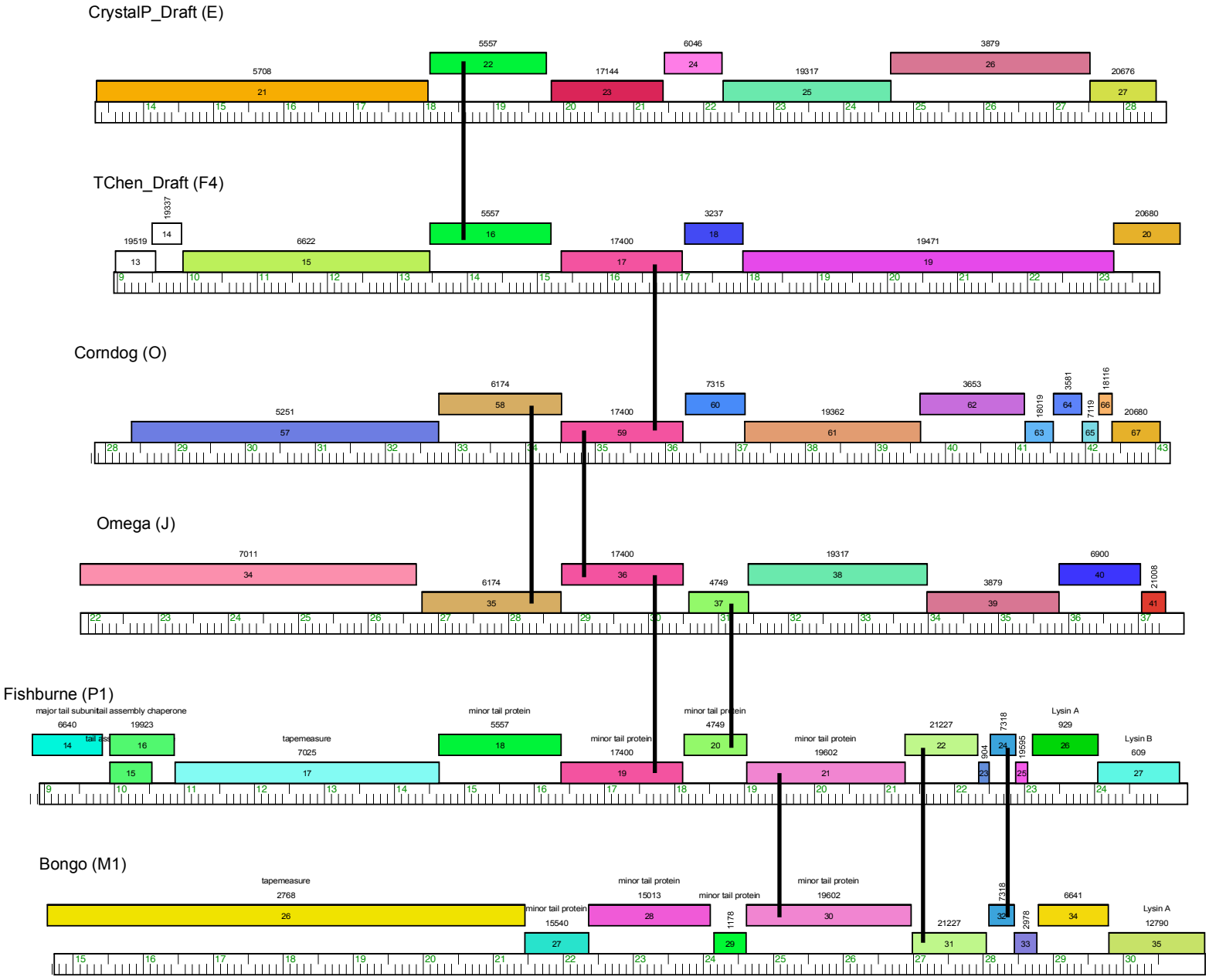
2. Grouping of phage genomes into Clusters and Subclusters

3. Little picture, big picture.

CrystalP_Draft (E)



Genome Mosaics



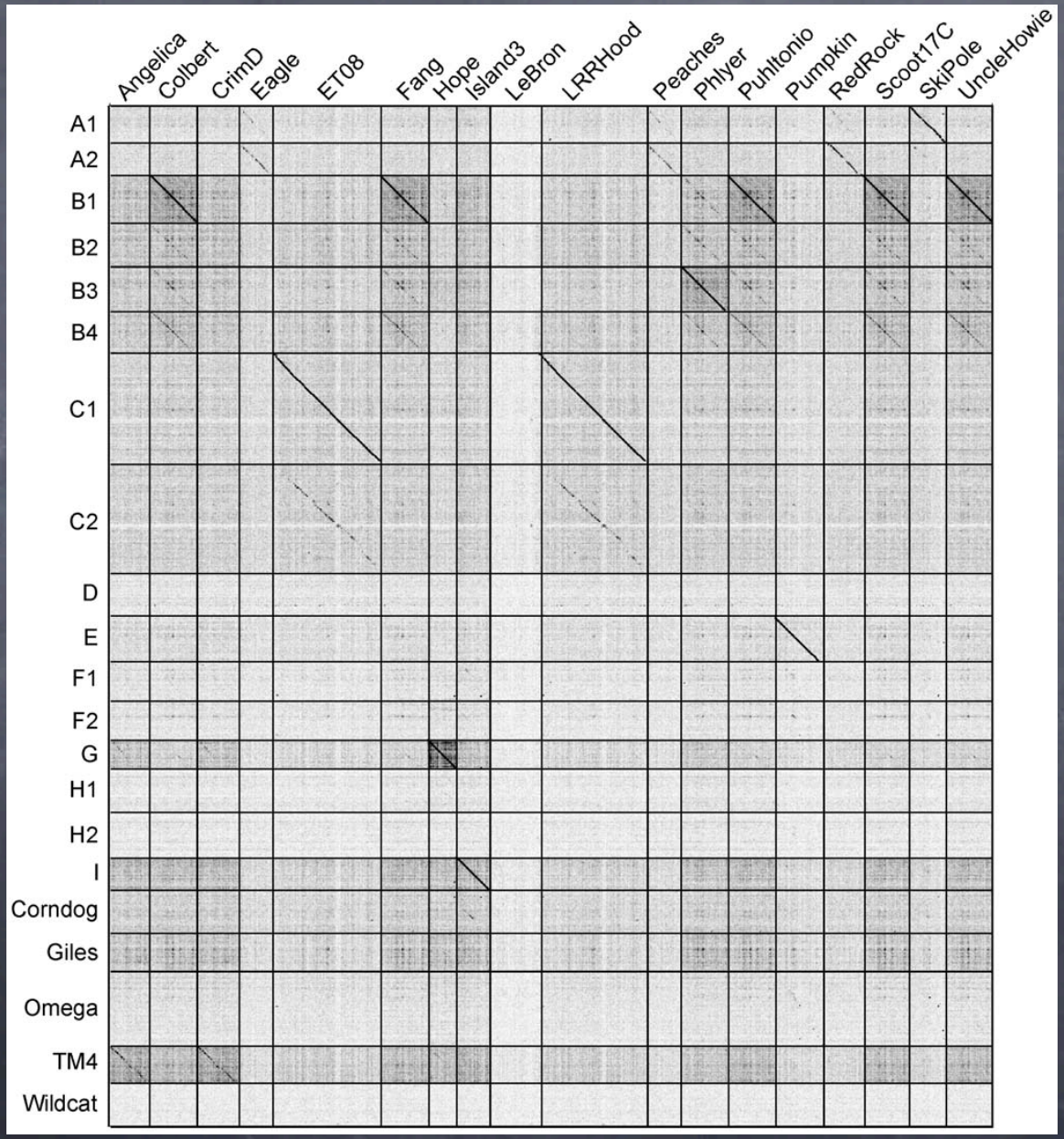
Three topics

1. Phage genomes are pervasively mosaic

2. Grouping of phage genomes into Clusters and Subclusters

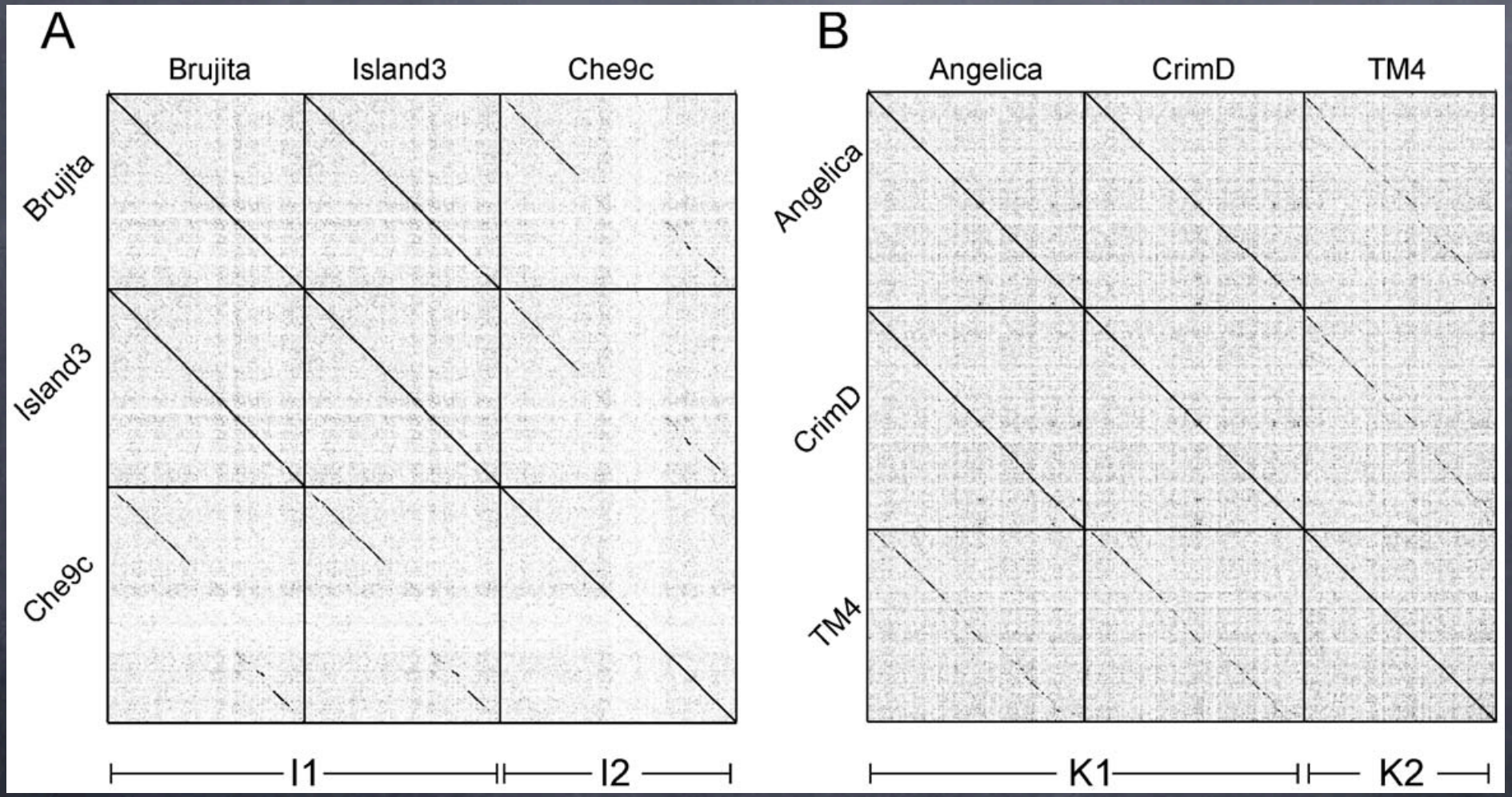
3. Little picture, big picture.

Dotplot showing genome relationships



Pope et al., '11

Clusters and Subclusters

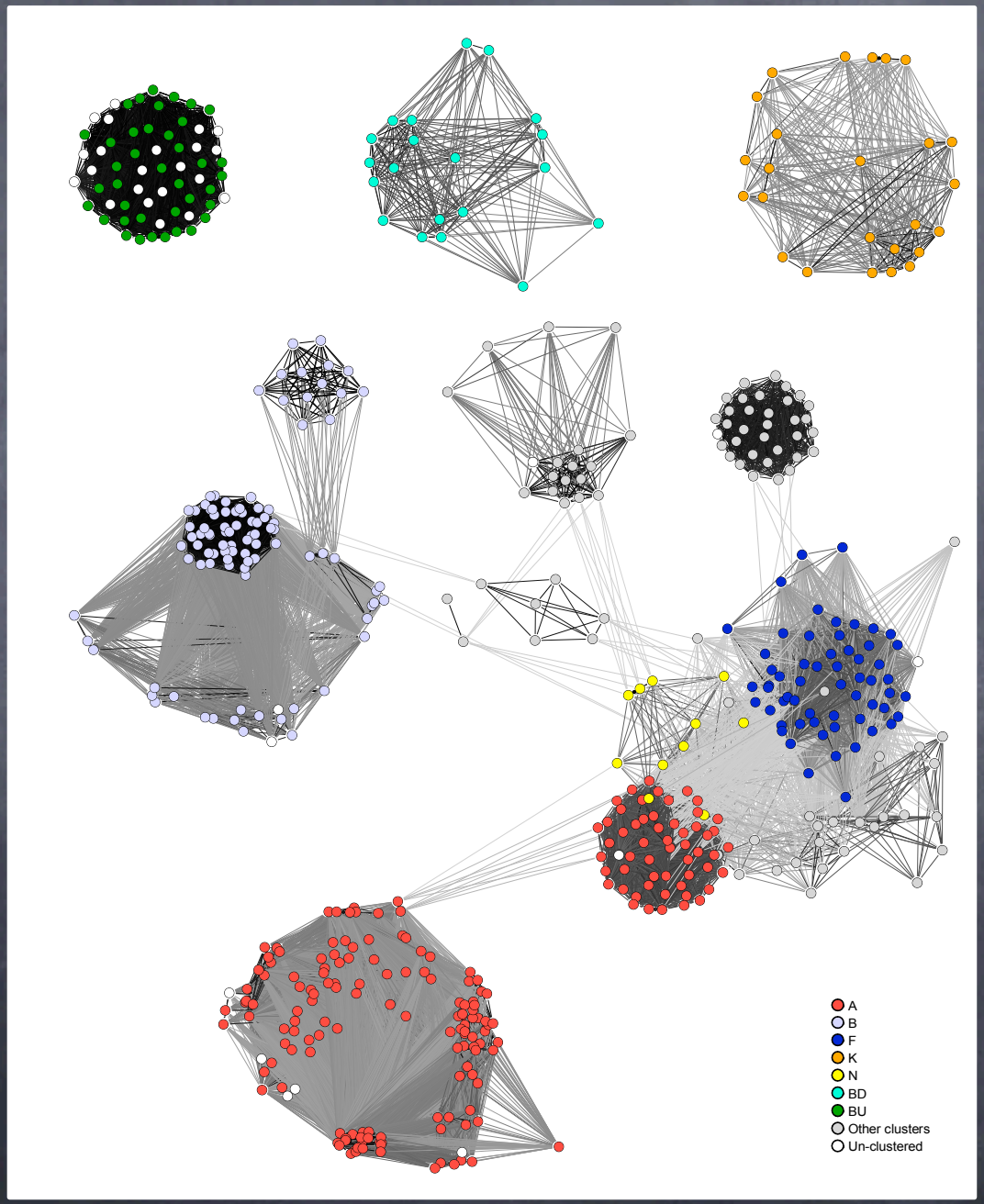


Pope et al., '11

Mycobacteriophage Cluster List

A	18	459	51,609	63.3	2016	Mycobacterium Gordonia
B	7	200	68,729	67.1	2016	Mycobacterium
C	2	81	155,647	64.7	2016	Mycobacterium
D	2	11	64,932	59.5	2016	Mycobacterium
E	--	77	75,492	63.0	2016	Mycobacterium
F	4	112	57,419	61.5	2016	Mycobacterium
G	4	36	42,374	66.9	2016	Mycobacterium
H	2	5	69,469	57.3	2014	Mycobacterium
I	2	5	51,129	66.3	2016	Mycobacterium
J	--	23	110,691	60.9	2016	Mycobacterium
K	6	83	60,089	66.9	2016	Mycobacterium
L	4	28	75,302	58.9	2016	Mycobacterium
M	3	10	81,555	61.3	2016	Mycobacterium
N	--	17	43,076	66.2	2016	Mycobacterium
O	--	7	70,618	65.4	2016	Mycobacterium
P	2	21	48,018	67.0	2016	Mycobacterium
Q	--	6	53,753	67.4	2016	Mycobacterium
R	--	5	71,424	56.0	2014	Mycobacterium
S	--	3	65,193	63.4	2014	Mycobacterium
Singleton	--	55	63,260	62.9	2016	Mycobacterium Rhodococcus Arthrobacter Streptomyces Gordonia Corynebacterium Actinoplanes Tetrasphaera Tsukamurella Microbacterium Rothia Brevibacterium
T	--	4	42,961	66.2	2014	Mycobacterium
U	--	2	69,942	50.4	2014	Mycobacterium
V	--	3	78,354	57.0	2015	Mycobacterium
W	--	3	60,820	67.5	2015	Mycobacterium
X	--	2	88,037	56.7	2015	Mycobacterium
Y	--	2	76,760	66.7	2015	Mycobacterium
Z	--	2	50,807	66.0	2014	Mycobacterium

Cluster isolation and connectivity



Three topics

1. Phage genomes are pervasively mosaic

2. Grouping of phage genomes into Clusters and Subclusters

3. Little picture, big picture.

There's lots to learn from individual genomes

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Genome Sequence of *Gordonia* Phage BetterKatz

Welkin H. Pope, Emily N. Berryman, Kaitlyn M. Forrest, Lilliana McHale, Anthony T. Wertz, Zenas Zhuang, Naomi S. Kasturiarachi, Catherine A. Pressimone, Johnathon G. Schiebel, Emily C. Furbee, Sarah R. Grubb, Marcie H. Warner, Matthew T. Montgomery, Rebecca A. Garlena, Daniel A. Russell, Deborah Jacobs-Sera, Graham F. Hatfull

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

BetterKatz is a bacteriophage isolated from a soil sample collected in Pittsburgh, Pennsylvania using the host *Gordonia terrae* 3612. BetterKatz's genome is 50,636 bp long and contains 75 predicted protein-coding genes, 35 of which have been assigned putative functions. BetterKatz is not closely related to other sequenced *Gordonia* phages.

Received 9 May 2016 **Accepted** 21 June 2016 **Published** 11 August 2016

Citation Pope WH, Berryman EN, Forrest KM, McHale L, Wertz AT, Zhuang Z, Kasturiarachi NS, Pressimone CA, Schiebel JG, Furbee EC, Grubb SR, Warner MH, Montgomery MT, Garlena RA, Russell DA, Jacobs-Sera D, Hatfull GF. 2016. Genome sequence of *Gordonia* phage BetterKatz. Genome Announc 4(4):e00590-16. doi:10.1128/genomeA.00590-16.

Address correspondence to Welkin H. Pope, welkin@pitt.edu.

Gordonia spp. are common environmental inhabitants (1) and several bacteriophages have been isolated using *Gordonia* hosts (2–6). To understand the genetic diversity of *Gordonia* phages, the integrated research-education Science Education Alliance-Phage Hunters Advancing Genomic and Evolutionary Science (SEA-PHAGES) program is using *Gordonia terrae* 3612 to isolate and genomically characterize bacteriophages (7). Phage BetterKatz was recovered from a soil sample from Pittsburgh, PA by direct plating of filtered soil extract on a lawn of *G. terrae*; it was plaque purified, amplified, and viral dsDNA was extracted. BetterKatz virions have a siphoviral morphology with an isometric head, and a flexible tail 220 nm in length.

BetterKatz was sequenced using the Illumina MiSeq platform using 140 bp single-end reads and assembled using Newbler to yield a single major contig of 50,636 bp with an average coverage of 247-fold. The genome has defined ends with 10 base 3′ single stranded DNA extensions (5′-TGCCGCGGTA) and is 67.1% G+C, similar to its host (67.8%). BetterKatz does not share extensive nucleotide sequence similarity to other sequenced phages or prophages, although there are two segments spanning approximately 10 kbp—corresponding to virion structural genes—with similarity to a putative prophage in *Gordonia* sp. KTR9 (8) that is integrated at an *attB* site overlapping a $tRNA^{ala}$ gene (KTR9_RS07590).

Seventy-five BetterKatz protein-coding genes were predicted using Glimmer and Genemark (9, 10) and putative functions were assigned using BLASTP, HHpred, and Phamerator (11, 12); no tRNA genes are predicted using Aragorn (13). All are transcribed rightwards with the exception of five genes—including a tyrosine-integrase and the immunity repressor—near the center of the genome. The attachment site (*attP*) is located immediately downstream of *int* (*39*) and BetterKatz is predicted to integrate into the same *attB* site overlapping a $tRNA^{ala}$ gene, where prophages lie in both *Gordonia* sp. KTR9 and *Gordonia bronchialis* DSM 43427 (14). The genes in the left arm are predominantly virion structure and assembly genes, and several genes in the right arm encode putative DNA metabolism functions including a DNA primase, a DNA methylase, and an exonuclease. We note that 31 of the predicted genes have no amino acid sequence similarity to other actinobacteriophage-encoded proteins in a data set of over 150,000 genes.

The lysis cassette in BetterKatz is located immediately downstream of the virion tail genes and there are two genes with predicted endolysin functions, gp29 that encodes a cysteine protease-like protein and gp30 encoding a glycoside hydrolase. The product of gene *31* has three predicted transmembrane domains and is the likely holin, although gp32 is also a putative membrane protein with four transmembrane domains and may also play a role in lysis. Immediately to the right of the lysis cassette is a leftwards-transcribed HicAB-like toxin-antitoxin system (15). A putative transcription promoter is located upstream of the toxin gene (*35*) and a region of dyad symmetry overlaps the putative −10 motif to which the antitoxin (gp34) may bind to regulate TA transcription.

Accession number(s). The BetterKatz genome sequence is available from GenBank under accession number KU963261.

FUNDING INFORMATION

This work, including the efforts of Graham F. Hatfull, was funded by Howard Hughes Medical Institute (HHMI) (54308198).

REFERENCES

1. **Arenskotter M, Broker D, Steinbuchel A.** 2004. Biology of the metabolically diverse genus *Gordonia*. Appl Environ Microbiol 70:3195–3204. http://dx.doi.org/10.1128/AEM.70.6.3195-3204.2004.
2. **Petrovski S, Seviour RJ, Tillett D.** 2011. Prevention of *Gordonia* and *Nocardia* stabilized foam formation by using bacteriophage GTE7. Appl Environ Microbiol 77:7864–7867. http://dx.doi.org/10.1128/AEM.05692-11.
3. **Petrovski S, Seviour RJ, Tillett D.** 2011. Characterization of the genome of the polyvalent lytic bacteriophage GTE2, which has potential for biocontrol of *Gordonia*-, *Rhodococcus*-, and *Nocardia*-stabilized foams in activated sludge plants. Appl Environ Microbiol 77:3923–3929. http://dx.doi.org/10.1128/AEM.00025-11.
4. **Petrovski S, Tillett D, Seviour RJ.** 2012. Genome sequences and characterization of the related *Gordonia* phages GTE5 and GRU1 and their use as potential biocontrol agents. Appl Environ Microbiol 78:42–47. http://dx.doi.org/10.1128/AEM.05584-11.
5. **Dyson ZA, Tucci J, Seviour RJ, Petrovski S.** 2015. Lysis to kill: evaluation of the lytic abilities, and genomics of nine bacteriophages infective for

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Genome Sequences of *Gordonia* Phages Bowser and Schwabeltier

Matthew T. Montgomery, Welkin H. Pope, Zachary M. Arnold, Aleksandra Basina, Ankitha M. Iyer, Ty H. Stoner, Naomi S. Kasturiarachi, Catherine A. Pressimone, Johnathon G. Schiebel, Emily C. Furbee, Sarah R. Grubb, Marcie H. Warner, Rebecca A. Garlena, Daniel A. Russell, Deborah Jacobs-Sera, Graham F. Hatfull

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

***Gordonia* phages Bowser and Schwabeltier are newly isolated phages infecting *Gordonia terrae* 3612. Bowser and Schwabeltier have similar siphoviral morphologies and their genomes are related to each other, but not to other phages. Their lysis cassettes are atypically situated among virion tail genes, and Bowser encodes two tyrosine integrases.**

Received 9 May 2016 **Accepted** 21 June 2016 **Published** 11 August 2016

Citation Montgomery MT, Pope WH, Arnold ZM, Basina A, Iyer AM, Stoner TH, Kasturiarachi NS, Pressimone CA, Schiebel JG, Furbee EC, Grubb SR, Warner MH, Garlena RA, Russell DA, Jacobs-Sera D, Hatfull GF. 2016. Genome sequences of *Gordonia* phages Bowser and Schwabeltier. Genome Announc 4(4):e00596-16. doi:10.1128/genomeA.00596-16.

Address correspondence to Welkin H. Pope, welkin@pitt.edu.

Gordonia spp. are implicated in foaming of activated sludge in wastewater treatment as well as in opportunistic infections of catheters (1–4). Previously, 17 bacteriophages of *Gordonia* have been isolated, sequenced, and annotated (5–9). Isolation and genomic analysis of bacteriophages using *Gordonia terrae* 3612 as a host within the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) will expand our understanding of the genetic diversity of these viruses (10).

Gordonia phages Bowser and Schwabeltier were isolated by direct plating of filtered soil extracts from Pittsburgh, PA on *G. terrae* 3612. Following plaque purification and amplification, DNA was extracted and sequenced using an Illumina MiSeq with 140 bp single-end reads. Reads were assembled using Newbler into single major contigs of 46,570 bp and 46,895 bp for Bowser and Schwabeltier, respectively, with 1,945-fold and 1,555-fold coverage. The genomes are 67% G+C and have discrete ends with 10 base 3′ single stranded extensions (5′-CGCCGCGGTA). Bowser and Schwabeltier share segments of similarity spanning 60% of their genome lengths with nucleotide identity ranging from 82% to 94%, and are grouped together in Cluster DB using previously described parameters (11). Bowser and Schwabeltier are not closely related to other phages or prophages, although a 1.3 kbp segment is related (75% nucleotide identity) to minor tail protein genes of a putative prophage in *Gordonia* sp. KTR9 (12).

Using GeneMark (13), Glimmer (14), Phamerator (15), and DNA Master (http://cobamide2.bio.pitt.edu), we identified 71 and 72 protein encoding genes in Bowser and Schwabeltier, respectively, approximately 40% of which we could assign putative functions using BLAST (16) and HHpred (17, 18). Neither genome contains tRNA genes. Protein functional assignments include virion structure and assembly proteins, tyrosine integrases, immunity repressors, FtsK-like proteins, an acetyltransferase (Schwabeltier gp30), and several HNH endonucleases.

The lysis cassette in Bowser and Schwabeltier is unusually located within the minor tail protein genes, and includes endolysin and lysin B genes flanking four smaller open reading frames, three of which (e.g., Schwabeliter gp22, gp23, and gp25) are strongly predicted to be membrane proteins. The three putative membrane proteins may all be associated with lysis although it is unclear which plays the holin role. Both phages have leftwards-transcribed tyrosine integrase and immunity repressor genes near the centers of their genomes, and these have the characteristics of the previously described integration-dependent immunity systems (19) in that both the integrases and repressors have C-terminal protein degradation tags. Although the *attP* site is expected to be located within the repressor-coding region in these systems, BLASTn searches failed to identify a corresponding *attB* site in any sequenced *Gordonia* genome. Curiously, Bowser encodes a second rightwards-transcribed tyrosine integrase, but we have also been unable to identify its corresponding *attP* site. It thus remains unclear whether either Bowser or Schwabeltier form lysogens with integrated prophages in *G. terrae* 3612 or any other bacterial host.

Accession number(s). The Bowser and Schwabeltier genomes are available from GenBank under accession numbers KU998235 and KU963252.

FUNDING INFORMATION

This work, including the efforts of Graham F. Hatfull, was funded by Howard Hughes Medical Institute (HHMI) (54308198).

REFERENCES

1. **Blaschke AJ, Bender J, Byington CL, Korgenski K, Daly J, Petti CA, Pavia AT, Ampofo K.** 2007. *Gordonia* species: emerging pathogens in pediatric patients that are identified by 16S ribosomal RNA gene sequencing. Clin Infect Dis 45:483–486. http://dx.doi.org/10.1086/520018.
2. **De los Reyes FL, III, Raskin L.** 2002. Role of filamentous microorganisms in activated sludge foaming: relationship of mycolata levels to foaming initiation and stability. Water Res 36:445–459. http://dx.doi.org/10.1016/S0043-1354(01)00227-5.
3. **De los Reyes FL, III, Rothauszky D, Raskin L.** 2002. Microbial community structures in foaming and nonfoaming full-scale wastewater treatment plants. Water Environ Res 74:437–449. http://dx.doi.org/10.2175/106143002X140233.
4. **Grisold AJ, Roll P, Hoenigl M, Feierl G, Vicenzi-Moser R, Marth E.** 2007. Isolation of *Gordonia terrae* from a patient with catheter-related bacteraemia. J Med Microbiol 56:1687–1688. http://dx.doi.org/10.1099/jmm.0.47388-0.
5. **Dyson ZA, Tucci J, Seviour RJ, Petrovski S.** 2015. Lysis to kill: evaluation

Modes of Phage Evolution