

OBSERVATION



Human-Gut Phages Harbor Sporulation Genes

[®]Daniel A. Schwartz,^a [®]Josué A. Rodríguez-Ramos,^b Michael Shaffer,^b Rory M. Flynn,^b [®]Rebecca A. Daly,^b [®]Kelly C. Wrighton,^b [®]Jay T. Lennon^a

^aDepartment of Biology, Indiana University, Bloomington, Indiana, USA ^bDepartment of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA

ABSTRACT Spore-forming bacteria are prevalent in mammalian guts and have implications for host health and nutrition. The production of dormant spores is thought to play an important role in the colonization, persistence, and transmission of these bacteria. Spore formation also modifies interactions among microorganisms such as infection by phages. Recent studies suggest that phages may counter dormancy-mediated defense through the expression of phage-carried sporulation genes during infection, which can alter the transitions between active and inactive states. By mining genomes and gut-derived metagenomes, we identified sporulation genes that are preferentially carried by phages that infect spore-forming bacteria. These included genes involved in chromosome partitioning, DNA damage repair, and cell wall-associated functions. In addition, phages contained homologs of sporulation-specific transcription factors, notably spo0A, the master regulator of sporulation, which could allow phages to control the complex genetic network responsible for spore development. Our findings suggest that phages could influence the formation of bacterial spores with implications for the health of the human gut microbiome, as well as bacterial communities in other environments.

IMPORTANCE Phages acquire bacterial genes and use them to alter host metabolism in ways that enhance phage fitness. To date, most auxiliary genes replace or modulate enzymes that are used by the host for nutrition or energy production. However, phage fitness is affected by all aspects of host physiology, including decisions that reduce the metabolic activity of the cell. Here, we focus on endosporulation, a complex and ancient form of dormancy found among the *Bacillota* that involves hundreds of genes. By coupling homology searches with host classification, we identified 31 phage-carried homologs of sporulation genes that are mostly limited to phages infecting spore-forming bacteria. Nearly one-third of the homologs recovered were regulatory genes, suggesting that phages may manipulate host genetic networks by tapping into their control elements. Our findings also suggest a mechanism by which phages can overcome the defensive strategy of dormancy, which may be involved in coevolutionary dynamics of spore-forming bacteria.

KEYWORDS dormancy, evolution, genomics, guts, metagenomes, microbiome, phage, sporulation

Microbiomes in the human gut are made up of a diverse community of bacteria, archaea, and microeukaryotes, as well as viruses that infect these microorganisms (1). Members of the phylum *Bacillota* (formerly *Firmicutes*) include many spore-forming lineages such as *Bacillus* and *Clostridium*. While spore-forming taxa are indicators of healthy gut microbiomes, some strains within this group are common intestinal pathogens (2, 3). Sporulation is a complex form of dormancy, involving hundreds of genes, that helps these bacteria contend with spatial and temporal variation in environmental conditions in human guts and facilitate transmission (2, 4).

Editor Alison Buchan, University of Tennessee at Knoxville

Copyright © 2023 Schwartz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Daniel A. Schwartz, da.schwartz@outlook.com, or Jay T. Lennon, lennonj@iu.edu.

The authors declare no conflict of interest.

Received 19 January 2023 Accepted 23 March 2023 Published 12 April 2023 Viruses of microbes, such as bacteriophages, play an important role in shaping gut microbiomes (1). Phage fitness is thought to be enhanced through the encoding of bacterial-like auxiliary metabolic genes (AMGs) that can reprogram and sustain host metabolism during infection (5). The acquisition of other, nonmetabolic genes may allow phages to alter other aspects of bacterial physiology (6). One of the most important determinants of phage fitness is the metabolic activity of the host cell (7, 8). Rates of bacterial metabolism span orders of magnitude, ranging from exponential growth to being nearly inert when cells engage in certain types of dormancy, such as sporulation (9, 10). By entering a state of reduced metabolic activity, microorganisms can defend themselves against phage attack (11, 12), altering selection in ways that could modify coevolutionary dynamics.

Previous work has demonstrated that some phage genomes possess sporulation genes (13–15). In one example, homologs of sporulation-specific sigma factors (*sigG* and *sigF*) were identified in both lytic and lysogenic phages (13). These sigma factors are essential for the developmental transition of a vegetative cell to an endospore (16). When expressed in a host (*Bacillus subtilis*), the phage-carried sigma factors activate sporulation transcriptional pathways and depress spore yield by up to 99% (13). To date, there has not been any systematic analysis of the prevalence and distribution of sporulation genes in phages. Thus, it remains unknown whether modification of host sporulation is a common phage strategy. In this study, we search for homologs of sporulation genes in genomic and metagenomic data to determine whether phages employ this strategy in human gut microbiomes.

Identifying sporulation homologs in viral genomes and metagenomes. We identified sporulation genes in viral genomes and uncultivated viral genomes (UViGs) using DRAM-v (17) (see Fig. S1 and Text S1 in the supplemental material). Specifically, we targeted homologs of well-characterized sporulation genes found in *B. subtilis* and *Clostridioides difficile*. We reasoned that phage-carried genes can affect sporulation only if they are in phages that infect a spore-forming host. We therefore designed an enrichment test to identify homologs of sporulation genes that were preferentially found in phages that infect spore-forming hosts. We first evaluated our search strategy by looking for sporulation genes in genomes of phage isolates for which the host was known. Next, we applied the same approach to UViGs assembled from human-gut environments, and for which host predictions had been made in the original studies (18, 19). To minimize the potential for contamination by bacterial sequences, we inspected the annotations of 6,542 gut-derived UViGs in which sporulation genes were detected (Fig. S2), with an average of 117 UViGs inspected per enriched sporulation gene (Fig. S3).

Phages possess nonregulatory genes required for sporulation. Our search identified 31 phage-carried homologs of sporulation genes (Table 1). These sporulation genes were enriched in phages that infect spore-forming hosts (Fig. 1). Many of the phage-carried homologs were nonregulatory genes involved in an assortment of sporulation-related processes such as chromosome partitioning, DNA damage repair, and cell wall-associated functions (Table 1). The acquisition of these genes might allow phages to promote or impede specific steps of spore development or its eventual germination. For example, phages may use chromosome segregation genes to increase the probability of entrapment (and survival) of the phage genome in the spore during the asymmetric division separating the developing spore from the mother cell (20). Alternatively, it is possible that some of these genes are used by phages for functions other than sporulation. Chromosome segregation genes are known to be used by phages that establish extrachromosomal, plasmid-like lysogeny (21). Likewise, cell wall hydrolases used by the host to restructure the cell during sporulation (cwlJ, sleB, spollD) could be repurposed as endolysins to burst the host cell at the completion of the phage lytic cycle (22). Further experimental investigation will be required to establish the phage functions of the sporulation gene homologs that we have catalogued in this work.

Phages also possess genes that regulate sporulation. Nearly one-third of the sporulation homologs (n = 9) identified in phage genomes and metagenomes are transcriptional regulators (Table 1). This finding is different than most examples of AMGs, where

TABLE 1 Sporulation genes detected in viral genomes and metagenomes^a

KO [♭]	Locus no. ^c	Gene ^c	Type ^d	Function ^e
K01356	BSU_17850	lexA	R	lexA repressor
K03086	BSU_25200	sigA	R	RNA polymerase sigma factor RpoD
K03091	BSU_00980	SigH	R	RNA polymerase sigma-H factor
	BSU_15320	SigE	R	RNA polymerase sigma-E factor
	BSU_15330	SigG	R	RNA polymerase sigma-G factor
	BSU_23450	sigF	R	RNA polymerase sigma-F factor
	CD630_07720	SigF	R	RNA polymerase sigma-F factor
	CD630_12300	SigK	R	RNA polymerase sigma-K factor
	CD630_26420	SigG	R	RNA polymerase sigma-G factor
	CD630_26430	sigE	R	RNA polymerase sigma-E factor
K04769	BSU_560	spoVT	R	Stage V sporulation protein T
	CD630_34990	spoVT	R	Stage V sporulation protein T
K06283	BSU_36420	spolIID	R	Stage III sporulation protein D
	CD630_1260	spolIID	R	Stage III sporulation protein D
K06284	BSU_370	abrB	R	Transition state regulatory protein AbrB
K07699	BSU_24220	spo0A	R	Stage 0 sporulation protein A
	CD630_12140	spo0A	R	Stage 0 sporulation protein A
K07738	CD630_26400	nrdR	R	Transcriptional regulator, repressor NrdR family
K03496	BSU_40970	parA	R+NR	Sporulation initiation inhibitor protein Soj
	CD630 36720	soj	R+NR	Transcriptional regulator, sporulation initiation inhibitor, chromosome partitioning protein
K00390	BSU_10930	yitB	NR	Phosphoadenosine phosphosulfate reductase
K00640	CD630 15950	cysE	NR	Serine acetyltransferase (SAT)
K00820	CD630 1200	almS	NR	Glucosamine-fructose-6-phosphate aminotransferase (isomerizing)
K00974	BSU 22450	сса	NR	CCA-adding enzyme
K01142	BSU_40880	exoA	NR	Exodeoxyribonuclease, repair of oxidative DNA damage in spores
K01449	BSU 02600	CwlJ	Nr	Cell wall hvdrolase CwlJ
	BSU 22930	sleB	NR	Spore cortex-lytic enzyme
	CD630 35630	NA	NR	Putative spore cortex-lytic hydrolase
K02049	BSU 30610	vtlC	NR	ABC transporter ATP-binding protein
K02343	CD630 160	dnaX	NR	DNA polymerase III subunits gamma and tau
K03466	BSU 16800	spollIE	NR	Spore DNA translocase
K03497	BSU 40960	parB	NR	Stage 0 sporulation protein J
	CD630 36710	spo0J	NR	Stage 0 sporulation protein J, site-specific DNA-binding protein
K03657	CD630 7490	NA	NR	Putative DNA helicase, UvrD/REP type
K03664	BSU 33600	sтрВ	NR	SsrA-binding protein
K03698	BSU 9930	yhaM	NR	3'-5' exoribonuclease yhaM
K06381	BSU_36750	spollD	NR	Stage II sporulation protein D
	CD630 1240	, spollD	NR	Stage II sporulation protein D
K06412	BSU 490	, spoVG	NR	Septation protein SpoVG
	CD630 35160	spoVG	NR	Regulator required for spore cortex synthesis
K07171	CD630 34610	EndoA	NR	Endoribonuclease toxin
K10716	BSU 31322	yuqO	NR	Potassium channel protein YugO
K10979	BSU 13410	ykoV	NR	DNA repair protein YkoV
K014487	BSU 17410	cwlC	NR	Mother cell lysis
	BSU_01530	cwID	NR	Spore cortex peptidoglycan synthesis
K02647	BSU_28670	ysfB	U	Hypothetical protein; similar to carbohydrate diacid transcriptional activator
K03469	BSU_21970	, ypeP	U	Hypothetical protein; similar to RNase HI
K07175	BSU_14810	ylaK	U	Hypothetical protein, similar to PhoH

These genes were enriched in phages infecting spore-forming hosts and were validated by manual inspection of annotations to have a viral origin.

^bKO, KEGG ortholog identifier.

c"Locus no." and "gene" refer to the gene locus number and name of the sporulation gene(s) associated with a KO. BSU loci are from Bacillus subtilis (KEGG taxon T00010), and CD630 loci are from Clostridioides difficile (KEGG taxon T00487).

^dLocus type reflects whether a KO is a regulatory gene (R), a nonregulatory gene (NR), or a hypothetical or uncharacterized function (U).

"The "Function" column provides a description of the KO from SubtiWiki for B. subtilis or KEGG for C. difficle.

the phages control a metabolic process by phage-carried enzymes or by expression of modulators of host enzyme activity (5). It may be that phages manipulate host sporulation by interfering with the tightly regulated transcriptional program that is essential for this complex developmental process (2). Such findings are consistent with recent experimental findings regarding sporulation-specific sigma factors, where the ectopic expression of phage-carried sigG and sigF homologs altered the transcriptional program of B. subtilis, resulting in reduced spore yield (13).



FIG 1 Sporulation genes are enriched in phages that infect spore-forming bacteria. Homologs of both sporulation and nonsporulation genes were identified in RefSeq isolate phage genomes and in published gut viromes (18, 19) using DRAM-v. However, genes enriched among phages infecting spore-forming hosts (purple points) were mostly sporulation-related genes. The number of homologs detected for each gene is plotted on the *x* axis. For each of the genes, a hypergeometric enrichment test evaluated if it was found in phages that infect spore-forming hosts more than the random expectation given the number of phages infecting spore-forming hosts among all phages surveyed. The enrichment result $[-\log_{10}$ (hypergeometric *P* value)] is plotted on the *y* axis. Purple points represent enriched genes with an adjusted *P* value of $<10^{-6}$ (horizontal dashed line) and a sample size of >30 (vertical dashed line). Representative names of *B. subtilis* sporulation genes are provided for genes that were enriched and of viral origin.

Most notable among phage-carried regulators are homologs of *spo0A*, the master regulator of sporulation initiation that is conserved among all spore-forming bacteria (4). Interestingly, the homologs found in phages are truncated versions of *spo0A* that contain the DNA-binding effector domain, but not the receiver domain (Fig. 2). The latter is responsible for modifying the DNA-binding activity in response to environmental and physiological signals received via the phosphorelay signal-transduction system (23). The truncation suggests that phage-carried *spo0A* may not require the normal host signals to activate or repress the initiation of host sporulation (24). In addition to transcriptional regulators, phage genes included other potential posttranscriptional regulators (RNA binding *spoVG*, and translation-related genes *cca* and *smpB*). Taken together, the bioinformatic findings here and laboratory results (13) suggest some phages may overcome dormancy defenses by targeting the regulation of sporulation. Compared to the use of nonregulatory genes, this is likely to be a more efficient strategy for altering the course of a complex cellular program.

Phage-carried sporulation genes occur in diverse environments. The recovery of phage-carried sporulation genes is not restricted to the human gut. We identified sporulation genes in UViGs originating from diverse environments (Table S1). Of the 30 sporulation genes identified in gut-derived UViGs, we found 23 that also occur in phages from terrestrial and aquatic environments (Fig. S4). Thus, phage manipulation of sporulation may be a common phenomenon in environments where spore-forming bacteria are found.

Implications and future directions. Sporulation is an ancient, complex, and important trait that contributes to the persistence and transmission of beneficial and pathogenic members of the mammalian gut microbiome. While sporulation can reduce virus infection, our analysis supports the view that phages may use host-like genes to overcome this host defense mechanism (25). Specifically, our study provides genomic and metagenomic evidence that phages carry homologs of multiple sporulation genes, which may influence the transition of bacteria between active and dormant states in host-associated and environmental ecosystems. The evolutionary drivers and ecological consequences of phage-carried sporulation genes remain to be investigated (15). Our



FIG 2 Multiple sequence alignment of protein sequences of *spo0A* homologs from viromes (n = 29) and diverse bacteria (n = 57; from COG5801). (a) Summary of alignment coverage showing the percentage of nongap characters at each position, separated by the source of the sequence. *Spo0A* from viromes align to the C-terminal effector domain of the bacterial genes. (b) Focus on effector domain, showing all virome sequences aligned with *spo0A* of model spore-forming bacteria *Clostridioides difficile* and *Bacillus subtilis*. The alpha helices of the effector domain are indicated above the sequences. Sequence colors correspond to physicochemical properties of amino acids, using the Zappo coloring scheme. Information on functional and structural domains is from reference 26.

work demonstrates how partitioning phages by a specific host trait (e.g., sporulation) can be used to identify genes used by phages to influence the same host trait.

Data availability. The data products of this study, including sequence data for sporulation genes identified in UViGs, are available at https://doi.org/10.5281/zenodo .7742410. All code used in this study is available at https://github.com/LennonLab/ spore_amg and in the Zenodo repository linked above.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only. TEXT S1, DOCX file, 0.03 MB. FIG S1, TIF file, 2 MB. FIG S2, TIF file, 6.2 MB. FIG S3, TIF file, 6.4 MB. FIG S4, TIF file, 1.9 MB. TABLE S1, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (DEB-1934554 to J.T.L. and D.A.S., DBI-2022049 to J.T.L., EAR-1847684 to K.C.W.), U.S. Army Research Office grants (W911NF-14-1-0411 to J.T.L., W911NF-22-1-0014 to J.T.L., and W911NF-22-S-0008 to J.T.L.), and the National Aeronautics and Space Administration (80NSSC20K0618 to J.T.L.).

REFERENCES

- 1. Shkoporov AN, Hill C. 2019. Bacteriophages of the human gut: the "known unknown" of the microbiome. Cell Host Microbe 25:195–209. https://doi.org/10.1016/j.chom.2019.01.017.
- Koopman N, Remijas L, Seppen J, Setlow P, Brul S. 2022. Mechanisms and applications of bacterial sporulation and germination in the intestine. Int J Mol Sci 23:3405. https://doi.org/10.3390/ijms23063405.
- Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533:543–546. https://doi.org/10.1038/nature17645.
- Galperin MY, Yutin N, Wolf YI, Vera Alvarez R, Koonin EV. 2022. Conservation and evolution of the sporulation gene set in diverse members of the Firmicutes. J Bacteriol 204:e00079-22. https://doi.org/10.1128/jb.00079-22.
- Jacobson TB, Callaghan MM, Amador-Noguez D. 2021. Hostile takeover: how viruses reprogram prokaryotic metabolism. Annu Rev Microbiol 75: 515–539. https://doi.org/10.1146/annurev-micro-060621-043448.
- Hargreaves KR, Kropinski AM, Clokie MR. 2014. Bacteriophage behavioral ecology: how phages alter their bacterial host's habits. Bacteriophage 4: e29866. https://doi.org/10.4161/bact.29866.
- Nabergoj D, Modic P, Podgornik A. 2018. Effect of bacterial growth rate on bacteriophage population growth rate. MicrobiologyOpen 7:e00558. https://doi.org/10.1002/mbo3.558.
- Igler C. 2022. Phenotypic flux: the role of physiology in explaining the conundrum of bacterial persistence amid phage attack. Virus Evol 8:veac086. https://doi.org/10.1093/ve/veac086.
- Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci U S A 118:e2016810118. https://doi .org/10.1073/pnas.2016810118.
- Ghosh S, Korza G, Maciejewski M, Setlow P. 2015. Analysis of metabolism in dormant spores of *Bacillus* species by ³¹P nuclear magnetic resonance analysis of low-molecular-weight compounds. J Bacteriol 197:992–1001. https://doi.org/10.1128/JB.02520-14.
- Meeske AJ, Nakandakari-Higa S, Marraffini LA. 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature 570:241–245. https://doi.org/10.1038/s41586-019-1257-5.
- Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. Biol Direct 7:40. https://doi.org/10.1186/1745-6150 -7-40.
- Schwartz D, Lehmkuhl B, Lennon J. 2022. Phage-encoded sigma factors alter bacterial dormancy. mSphere 7:e00297-22. https://doi.org/10.1128/ msphere.00297-22.
- Van Goethem MW, Swenson TL, Trubl G, Roux S, Northen TR. 2019. Characteristics of wetting-induced bacteriophage blooms in biological soil crust. mBio 10:e02287-19. https://doi.org/10.1128/mBio.02287-19.

- Butala M, Dragoš A. 2022. Unique relationships between phages and endospore-forming hosts. Trends Microbiol 17:S0966-842X(22)00317-1. https:// doi.org/10.1016/j.tim.2022.11.009.
- Fimlaid KA, Shen A. 2015. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. Curr Opin Microbiol 24:88–95. https://doi .org/10.1016/j.mib.2015.01.006.
- Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900. https://doi.org/10.1093/nar/ gkaa621.
- Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. 2021. Massive expansion of human gut bacteriophage diversity. Cell 184: 1098–1109.e9. https://doi.org/10.1016/j.cell.2021.01.029.
- Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. 2020. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe 28:724–740.e8. https://doi .org/10.1016/j.chom.2020.08.003.
- Sonenshein AL. 1970. Trapping of unreplicated phage DNA into spores of Bacillus subtilis and its stabilization against damage by ³²P decay. Virology 42:488–495. https://doi.org/10.1016/0042-6822(70)90292-8.
- Dedrick RM, Mavrich TN, Ng WL, Cervantes Reyes JC, Olm MR, Rush RE, Jacobs-Sera D, Russell DA, Hatfull GF. 2016. Function, expression, specificity, diversity and incompatibility of actinobacteriophage parABS systems. Mol Microbiol 101:625–644. https://doi.org/10.1111/mmi.13414.
- Loessner MJ, Maier SK, Daubek-Puza H, Wendlinger G, Scherer S. 1997. Three *Bacillus cereus* bacteriophage endolysins are unrelated but reveal high homology to cell wall hydrolases from different bacilli. J Bacteriol 179:2845–2851.https://doi.org/10.1128/jb.179.9.2845-2851.1997.
- Grimsley JK, Tjalkens RB, Strauch MA, Bird TH, Spiegelman GB, Hostomsky Z, Whiteley JM, Hoch JA. 1994. Subunit composition and domain structure of the Spo0A sporulation transcription factor of *Bacillus subtilis*. J Biol Chem 269:16977–16982. https://doi.org/10.1016/S0021-9258(19)89485-1.
- Ireton K, Rudner DZ, Siranosian KJ, Grossman AD. 1993. Integration of multiple developmental signals in *Bacillus subtilis* through the SpoOA transcription factor. Genes Dev 7:283–294. https://doi.org/10.1101/gad.7 .2.283.
- Seed KD, Lazinski DW, Calderwood SB, Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494:489–491. https://doi.org/10.1038/nature11927.
- Lewis RJ, Krzywda S, Brannigan JA, Turkenburg JP, Muchová K, Dodson EJ, Barák I, Wilkinson AJ. 2000. The trans-activation domain of the sporulation response regulator Spo0A revealed by X-ray crystallography. Mol Microbiol 38:198–212. https://doi.org/10.1046/j.1365-2958.2000.02134.x.