Human-gut phages harbor sporulation genes

Schwartz DA[1], Rodríguez-Ramos J[2], Shaffer M[2], Flynn R[2], Daly R, Wrighton KC[2], Jay T. Lennon[1]

[1] Department of Biology, Indiana University, Bloomington, IN 47405

[2] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80521

* Corresponding authors: danschw@iu.edu and lennonj@iu.edu

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-encoded 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 encoded 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.

SIGNIFICANCE

Phages acquire bacterial genes and use them to alter host metabolism in ways that enhance their 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 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 identify 31 phage-encoded homologs of sporulation genes that are mostly limited to phages infecting spore-forming bacteria. Nearly one-third 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.

40 **MAIN TEXT**

41 Microbiomes in the human gut are made up of a diverse community of bacteria, archaea,
42 and microeukaryotes, as well as well as viruses that infect these microorganisms (1). Members of
43 the phylum Bacillota (formerly Firmicutes) include many spore-forming lineages such as *Bacillus*
44 and *Clostridium*. These spore-forming members are dominant in healthy gut microbiomes, with
45 many strains within this group also considered common intestinal pathogens (2, 3). Sporulation is
46 a complex form of dormancy, involving hundreds of genes, that helps these bacteria contend with
47 spatial and temporal variation in environmental conditions in human guts and facilitate
48 transmission (2, 4).
49
50 Viruses of microbes, such as bacteriophages, play an important role in shaping gut
51 microbiomes (1). Phage fitness is thought to be enhanced through the encoding of bacterial-like
52 auxiliary metabolic genes (AMGs) that can reprogram and sustain host metabolism during
53 infection (5). The acquisition of other, non-metabolic genes may allow phages to alter other aspects
54 of bacterial physiology (6). One of the most important determinants of phage fitness is the
55 metabolic activity of the host cell (7, 8). Bacterial metabolism spans orders of magnitude, ranging
56 from exponential growth to being nearly inert when cells engage in certain types of dormancy such
57 as sporulation (9, 10). By entering a state of reduced metabolic activity, microorganisms can
58 defend themselves against phage attack (11, 12) altering selection in ways that could modify
59 coevolutionary dynamics.
60
61 Previous work has demonstrated that some phage isolates encode sporulation genes (e.g.
62 13, 14, 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 from vegetative cell to and endospore (16). When expressed in a host (*Bacillus subtilis*), the phage-encoded 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 metagenomic assembled genomes (vMAGS) using DRAM-v (17) (Fig. S1 and supplementary text). Specifically, we targeted homologs of well characterized sporulation genes found in *B. subtilis* and *Clostridioides difficile*. We reasoned that phage-encoded genes can only affect sporulation 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 vMAGs assembled from human gut environments for which host predictions had been made (18, 19). To minimize potential for contamination by bacterial sequences, we inspected the annotations of 6,542 gut-derived vMAGs in which sporulation genes were detected, with an average of 117 vMAGs inspected per enriched sporulation gene (Fig. S2).

Phages encode structural genes required for sporulation — Our search identified homologs of 31 phage-encoded homologs of sporulation genes (Table 1). These sporulation genes were enriched in phages that infect spore-forming hosts (Fig. 1). Many of the phage-encoded homologs were structural (i.e., non-regulatory) 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*, *spoIID*) 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 encode 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 the phages control a metabolic process by phage-encoded 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 with sigma factors where the ectopic expression of phage-encoded *sigG* and *sigF* homologs altered the transcriptional program of *B. subtilis* resulting in reduced spore yield (13).

Most notable among phage-encoded 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. S3). 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-encoded *spo0A* may not require the normal host signals to activate or repress the initiation of host sporulation (24). Beside transcriptional regulators, phage genes included other potential post-transcriptional regulators (RNA binding *spoVG*, and translation-related genes *cca* and *smpB*). Taken together, 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 structural genes, this is likely to be a more efficient strategy for altering the course of a complex cellular program.

Phage-encoded sporulation genes occur in diverse environments — The recovery of phage-encoded sporulation genes is not restricted to the human gut. We identified sporulation genes in vMAGs originating from diverse environments (Table S1). Of the 30 sporulation genes identified in gut-derived vMAGs 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 host defense mechanisms (25). Specifically, our study provides genomic and metagenomic evidence that phages encode multiple homologs of 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-encoded sporulation genes remain to be investigated (15). Our 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.

CODE AND DATA AVAILABILITY

All code and data used in this study are available at https://github.com/LennonLab/spore_amg. In addition, prior to publication, all data and code will be made available on Zenodo.

ACKNOWLEGEMENTS

Research was supported by the National Science Foundation (DEB-1934554 JTL and DAS, DBI-2022049 JTL, EAR-1847684 KCW), US Army Research Office Grant (W911NF-14-1-0411 JTL and W911NF-22-1-0014), the National Aeronautics and Space Administration (80NSSC20K0618 JTL).

REFERENCES

1. Shkoporov AN, Hill C. 2019. Bacteriophages of the human gut: the “known unknown” of the microbiome. Cell Host & Microbe 25:195-209.

2. Koopman N, Remijas L, Seppen J, Setlow P, Brul S. 2022. Mechanisms and applications of bacterial sporulation and germination in the entestine. International Journal of Molecular Sciences 23:3405.

3. 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.

4. 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. Journal of Bacteriology:e00079-22.

5. Jacobson TB, Callaghan MM, Amador-Noguez D. 2021. Hostile takeover: How viruses reprogram prokaryotic metabolism. Annual Review of Microbiology 75:515-539.

6. Hargreaves KR, Kropinski AM, Clokie MR. 2014. Bacteriophage behavioral ecology: How phages alter their bacterial host's habits. Bacteriophage 4:e29866.

7. Nabergoj D, Modic P, Podgornik A. 2018. Effect of bacterial growth rate on bacteriophage population growth rate. MicrobiologyOpen 7:e00558.

8. Igler C. 2022. Phenotypic flux: The role of physiology in explaining the conundrum of bacterial persistence amid phage attack. Virus Evolution 8:veac086.

9. Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proceedings of the National Academy of Sciences 118:e2016810118.

10. Ghosh S, Korza G, Maciejewski M, Setlow P. 2015. Analysis of metabolism in dormant spores of *Bacillus* species by ^{31}P nuclear magnetic resonance analysis of low-molecular-weight compounds. Journal of bacteriology 197:992-1001.

11. Meeske AJ, Nakandakari-Higa S, Marraffini LA. 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature 570:241-245.

12. 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. Biology Direct 7:7-40.

13. Schwartz D, Lehmkuhl B, Lennon J. 2022. Phage-encoded sigma factors alter bacterial dormancy. mSphere:e00297-22.

14. 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.

15. Butala M, Dragoš A. 2022. Unique relationships between phages and endospore-forming hosts. Trends in Microbiology doi:https://doi.org/10.1016/j.tim.2022.11.009.

16. Fimlaid KA, Shen A. 2015. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. Current Opinion in Microbiology 24:88-95.

17. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research 48:8883-8900.

18. 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.

19. 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.

20. Sonenshein AL. 1970. Trapping of unreplicated phage DNA into spores of *Bacillus subtilis* and its stabilization against damage by ^{32}P decay. Virology 42:488-95.

21. 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. Molecular Microbiology 101:625-644.

22. 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. Journal of Bacteriology 179:2845-2851.

23. 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*. Journal of Biological Chemistry 269:16977-16982.

24. Ireton K, Rudner DZ, Siranosian KJ, Grossman AD. 1993. Integration of multiple developmental signals in *Bacillus subtilis* through the *Spo0A* transcription factor. Genes & Development 7:283-294.

25. 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.

TABLES

Table 1. Sporulation genes detected in viral genomes and metagenomes. These genes were enriched in phages of spore-forming hosts and were validated to have a viral origin by manual inspection of annotations. KO represents the KEGG ortholog identifier. 'Locus no.' and `Gene` refer to the gene locus number and name of 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). Locus `Type` reflects whether a KO is a regulatory gene (R), structural gene (S), or a hypothetical of uncharacterized function (U). The `Function` column provides a description of the KO from Subtiwiki for *B. subtilis* or KEGG for *C. difficle.*

KO	Locus no.	Gene	Type	Function
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*		RNA polymerase sigma-E factor
	BSU_15330	*sigG*		RNA polymerase sigma-G factor
	BSU_23450	*sigF*		RNA polymerase sigma-F factor
	CD630_07720	*sigF*	R	RNA polymerase sigma-F factor
	CD630_12300	*sigK*		RNA polymerase sigma-K factor
	CD630_26420	*sigG*		RNA polymerase sigma-G factor
	CD630_26430	*sigE*		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	*spoIIID*	R	Stage III sporulation protein D
	CD630_1260	*spoIIID*	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+S	sporulation initiation inhibitor protein Soj
	CD630_36720	*soj*	R+S	Transcriptional regulator, sporulation initiation inhibitor, chromosome partitioning protein
K00390	BSU_10930	*yitB*	S	phosphoadenosine phosphosulfate reductase
K00640	CD630_15950	*cysE*	S	Serine acetyltransferase (SAT)
K00820	CD630_1200	*glmS*	S	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
K00974	BSU_22450	*cca*	S	CCA-adding enzyme
K01142	BSU_40880	*exoA*	S	Exodeoxyribonuclease, repair of oxidative DNA damage in spores
K01449	BSU_02600	*cwlJ*	S	cell wall hydrolase CwlJ
	BSU_22930	*sleB*		spore cortex-lytic enzyme
	CD630_35630	*NA*	S	putative spore cortex-lytic hydrolase
K02049	BSU_30610	*ytlC*	S	ABC transporter ATP-binding protein
K02343	CD630_160	*dnaX*	S	DNA polymerase III subunits gamma and tau
K03466	BSU_16800	*spoIIIE*	S	spore DNA translocase
K03497	BSU_40960	*parB*	S	stage 0 sporulation protein J
	CD630_36710	*spo0J*	S	Stage 0 sporulation protein J, site-specific DNA-binding protein
K03657	CD630_7490	*NA*	S	putative DNA helicase, UvrD/REP type
K03664	BSU_33600	*smpB*	S	SsrA-binding protein
K03698	BSU_9930	*yhaM*	S	3'-5' exoribonuclease yhaM
K06381	BSU_36750	*spoIID*	S	stage II sporulation protein D
	CD630_1240	*spoIID*	S	Stage II sporulation protein D
K06412	BSU_490	*spoVG*	S	septation protein SpoVG
	CD630_35160	*spoVG*	S	Regulator required for spore cortex synthesis
K07171	CD630_34610	*EndoA*	S	Endoribonuclease toxin
K10716	BSU_31322	*yugO*	S	potassium channel protein YugO
K10979	BSU_13410	*ykoV*	S	DNA repair protein YkoV
K01448[7]	BSU_17410	*cwlC*	S	mother cell lysis
	BSU_01530	*cwlD*		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

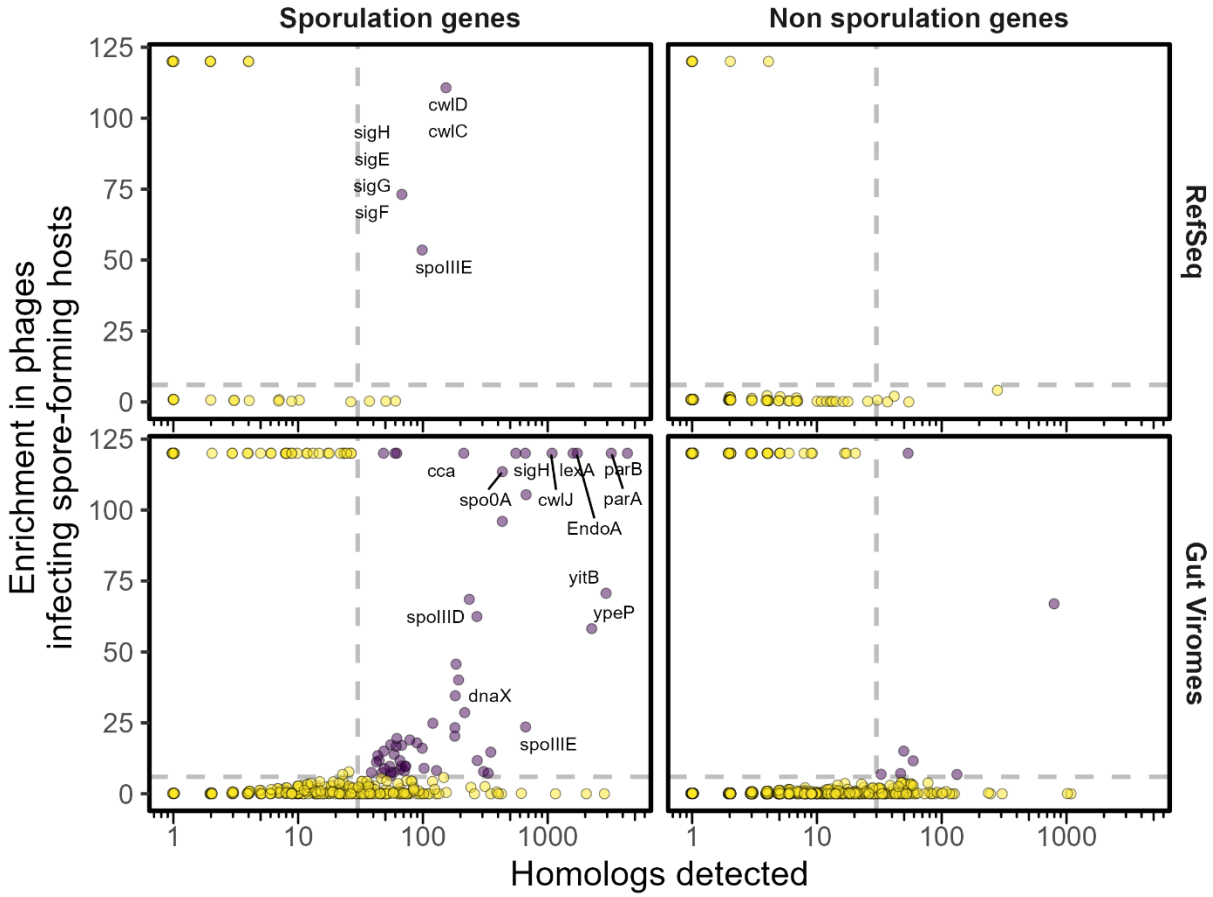
230

231

FIGURE LEGENDS

Fig. 1. Sporulation genes enriched in phages that infect spore-forming bacteria. Homologs of sporulation and non-sporulation genes were identified in RefSeq isolate phage genomes and in published gut viromes (18, 19) using DRAM-v. A hypergeometric enrichment test was performed for each gene to evaluate if it was found in phages that infect spore-forming hosts more than the random expectation given the total number of phages. The enrichment result (–Log_{10} *(hypergeometric P value)* on y-axis) is plotted as a function of the number of homologs detected (x-axis). Purple points represent enriched genes with an adjusted *P*-value < 10^{-6} (horizontal dashed line) and a sample size > 30 (vertical dashes line). Representative names of *B. subtilis* sporulation genes are provided for genes that were enriched and of viral origin.

244 **Fig. 1**



245