S3 Text. Technical considerations regarding the composition of the BASEL collection and the phenotyping of bacterial defense systems

Considerations regarding the composition of the BASEL collection

The choice of isolation host is known to be "perhaps the most critical part of the isolation process" because it pre-determines the properties and range of phages that can be sampled [1]. We therefore specifically designed a highly phage-sensitive host strain, *Escherichia coli* K-12 ΔRM, in order to eliminate as many biases in the isolation process as possible (see *Materials and Methods* as well as *Composition of the BASEL collection* in the main text).

However, one important feature of *E. coli* K-12 ΔRM is that it displays the rough LPS of all K12-lineage laboratory strains and lacks the O-antigen chains that are otherwise fully covering the cell surface of natural enterobacterial isolates (Figs 3A and 12A) [2]. On one hand, the absence of this formidable barrier enabled the isolation of a huge diversity of phages from all families that are, with very few exceptions, largely unable to infect *E. coli* K-12 with restored O16-type O-antigen expression (Figs 3A, 4, and 6-11). However, using a strain with rough LPS as isolation host intrinsically excludes any phages for which the O-antigen is not only an optional primary receptor but an essential part of the host recognition. Two common examples for such phages are iconic *Salmonella* phage P22 and *Gamaleyavirus* G7C, a relative of N4 (see Fig 10E), that both bind to very specific types of O-antigen and then target the glycan chain by enzymatic activities of their tailspikes to generate directional movement towards the cell surface [3, 4]. It is clear that this kind of phages are not highly abundant among *E. coli* phages because, e.g., a study that had isolated fifty phages primarily using strains with smooth LPS found 1) mostly the same taxonomic groups as those in the BASEL collection and reported that 2) many phages isolated on smooth hosts can also infect *E. coli* K-12 [5]. However, the targeting of O-antigen chains as an essential primary

receptor is not uncommon among podoviruses (and some myoviruses) [6, 7], and the isolation of *E. coli* phages with hosts expressing full smooth LPS indeed revealed a bigger diversity of (narrow-host-range) podoviruses than was reported in studies solely relying on *E. coli* K-12 [5]. Given seemingly low abundance of O-antigen specialists and that most realistic applications of the BASEL collection will be based on infecting *E. coli* K-12 strains with rough LPS, we do not feel that the absence of these O-antigen-dependent phages compromises the usefulness of our work. However, we imagine that a future study could generate a dedicated, optional expansion of the BASEL collection containing a diverse and representative selection of these phages that are specialized in the O16-type O-antigen of *E. coli* K-12. Similarly, we excluded all tailless phages like *Microviridae* and *Inoviridae* from the BASEL collection because their biology, evolution, and host interactions are so different from the lytic *Caudovirales* that we feel their investigation is beyond the scope of our current work [8, 9].

Any bacteriophage collection of finite size is inherently unable to include all rare phage groups and cannot comprehensively cover all genera of highly diverse and abundant families such as the *Drexlerviridae* (Fig 4B). Despite this shortcoming, we feel that the BASEL collection provides a reasonably complete snapshot of *E. coli* phage diversity because, e.g., all common and almost all previously described protein receptors of all included phage groups are covered (Figs 4-8). Increasing the number of phage isolates in the BASEL collection might therefore not greatly increase its biological diversity but could jeopardize its usefulness by making the handling more complicated. It also does not appear that sampling most phages from sewage plant inflow is a major limitation, because we did not observe any difference in the phages sampled from sewage and, e.g., river water (S5 Table). Previous work had sampled overall similar sets of phage groups on matter if the phages came from sewage or from infant guts [5, 10-15], possibly because many *E. coli* phages in the environment might derive directly or indirectly from fecal contaminations.

Another limitation of our isolation host strain *E. coli* K-12 ΔRM is that it might introduce additional biases beyond its rough LPS phenotype. As an example, laboratory adaptation might have inactivated also other possible phage receptors on the cell surface. A relevant example for the possible consequences of such a bias is the observation that phage T5 of the T phages is a highly unusual member of the *Markadamsvirinae* subfamily of *Demerecviridae* because it uses the FhuA protein as its terminal receptor (Fig 7) [16]. This peculiar feature of the T phages is easily explained with the *btuB* loss-of-function mutation of many early *E. coli* B strains that were used to sample the T phages but later reverted back to functional BtuB expression, e.g., in the lineage leading to *E. coli* B REL606 that is sensitive to all tested *Markadamsvirinae* (Fig 12B) [16, 17]. Similarly, the T phages do not contain any *Vequintavirinae* phage despite the abundance of this group, and we indeed find that none of our twelve *Vequintavirinae sensu stricto* can lyse *E. coli* B REL606 (Fig 12B). Since this defect is probably due to impaired adsorption, we speculate that the truncated *E. coli* B LPS core and / or changes of the ECA or any other primary receptor might be responsible for this surprising phenotype.

Another possible factor biasing the diversity of isolated bacteriophages could be remaining immunity systems in the sampling strain *E. coli* K-12 ΔRM. While we strain lacks all known restriction systems of *E. coli* K-12 as well as the RexAB and PifA Abi systems (S1 Text), several systems remain that either have been poorly studied or are supposed to have only a narrow target range [18]. As an example, the cryptic prophage *e14* of *E. coli* K-12 encodes the Lit Abi system that cleaves the elongation factor Tu (EF-Tu) when sensing the major capsid protein of *Tevenvirinae* [18, 19]. It is intuitive that the presence of *lit* in *E. coli* K-12 ΔRM might somehow bias the range of *Tevenvirinae* phages that we are sampling. However, although we confirmed by whole-genome sequencing that both the host as well as the T4 phage should carry alleles of *lit* and the major capsid protein gene triggering abortive infection [19], we observe robust plaque formation of T4 on our *E. coli* K-12 strains (see, e.g., in S4B Fig). Similarly, recent work

growing phage T4 on *E. coli* K-12 hosts never reported any problems with abortive infection [5, 20, 21]. We are therefore skeptical if abortive infection by *lit* can significantly affect the isolation of bacteriophages with *E. coli* K-12 ΔRM. Besides Lit, the RnlAB type II toxin-antitoxin system aborts the growth of T4 by activation of the RnlA RNase toxin when the phage lacks the *dmd* antitoxin gene [18, 22]. However, given the broad conservation of *dmd* among *Tevenvirinae* and the absence of any reports that RnlAB could target different phages, we do not think that this system has a significant impact on our bacteriophage isolation experiments. Like for Lit and RnlAB, we see no evidence that the other proposed immunity systems of *E. coli* K-12 will have relevant impact on the diversity of sampled phages. These are mostly ill-characterized or, in case of DicB, were only shown to work when ectopically expressed [23].

Considerations regarding the phenotyping of bacterial immunity systems

Our phage phenotyping experiments with diverse immunity systems were performed as quantitative top agar assays to determine the efficiency of plating (EOP) of each phage on a host carrying a given immunity system to generate robust, quantitative data (see *Materials and Methods*). However, there are a few technical caveats associated with our approach that need to be considered for a comprehensive interpretation of our results. Most importantly, for the phenotyping in the *E. coli* K-12 ΔRM host we cloned the different immunity systems onto plasmids which, though most were cloned with their native promoter, might affect their functionality by change of copy number (see *Materials and Methods* as well as S3 and S4 Tables). As an example, it was shown that phage T4 wildtype is resistant to RexAB but becomes sensitive when this immunity system is overexpressed from a multicopy plasmid [24]. However, the ColE1 and SC101 origins of replications used in this study have rather low copy numbers (of around 40 and 3-4, respectively [25]) and, e.g., our *rexAB* construct with ColE1 origin of replication has no detectable effect on the growth of phage T4 or any relative (Fig 8C), while an *rIIAB* mutant of this phage was sensitive as described previously (S4B and S4C Figs) [18]. The use of two different origins of replication

is a consequence of our initial choice to clone immunity systems onto plasmids isogenic to the EcoRI and EcoRV plasmids of Pleška, Qian et al. [26] which later resulted in problems with toxicity for some constructs. These were then cloned into the lower-copy plasmid with SC101 origin of replication instead (S4 Table). Though it is clear that expression levels could be different between these backbones as already evidenced form the difference in toxicity, we do no feel that this compromises our approach considerably. As an example, the two type III RM systems EcoP1_I and EcoCFT_II have a similar recognition sequence (Fig 3B) that is similarly abundant in the diverse phage genomes (S5 Table), but EcoP_1 was cloned with an SC101 origin of replication while EcoCFT_II was cloned with a ColE1 origin of replication. However, the phenotyping data do not show EcoCFT_II would be more potent compared to EcoP1_I (Figs 4 and 6-11) – sometimes one or the other has the stronger effect on EOP, sometimes they are similar.

Another more transparent caveat is that the graphs displaying the EOP of different phages on hosts with restriction-modification (RM) systems are not directly comparable quantitatively. The reason is that, due to differences in genome size, GC content, and evolution towards restriction site avoidance, the different genomes have vastly different numbers of recognition sites for a given RM system [27] (S5 Table). In extreme cases such as for type II RM systems and many tested podoviruses, recognition sites can be completely absent (S5 Table). One might argue that displaying an EOP (of ca. 1) for phage / RM interactions in the absence of recognition sites is not useful because – just biochemically – a cleavage of the phage chromosome was barely possible. However, given that the number of these sites is strongly under selection and part of the phages' strategy to counter their host's defenses, we find it more appropriate to show all data and highlight for the reader when no recognition sites were present.

References (S3 Text)

1. Hyman P. Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. Pharmaceuticals (Basel). 2019;12(1). Epub 2019/03/14. doi: 10.3390/ph12010035. PubMed PMID: 30862020; PubMed Central PMCID: PMCPMC6469166.
2. Liu D, Reeves PR. *Escherichia* coli K12 regains its O antigen. Microbiology. 1994;140 (Pt 1):49-57. doi: 10.1099/13500872-140-1-49. PubMed PMID: 7512872.
3. Broeker NK, Barbirz S. Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. Molecular microbiology. 2017;105(3):353-7. Epub 2017/06/16. doi: 10.1111/mmi.13729. PubMed PMID: 28618013.
4. Prokhorov NS, Riccio C, Zdorovenko EL, Shneider MM, Browning C, Knirel YA, et al. Function of bacteriophage G7C esterase tailspike in host cell adsorption. Molecular microbiology. 2017;105(3):385-98. Epub 2017/05/18. doi: 10.1111/mmi.13710. PubMed PMID: 28513100.
5. Korf IHE, Meier-Kolthoff JP, Adriaenssens EM, Kropinski AM, Nimtz M, Rohde M, et al. Still Something to Discover: Novel Insights into *Escherichia coli* Phage Diversity and Taxonomy. Viruses. 2019;11(5). doi: 10.3390/v11050454. PubMed PMID: 31109012; PubMed Central PMCID: PMCPMC6563267.
6. Hantke K. Compilation of Escherichia coli K-12 outer membrane phage receptors - their function and some historical remarks. FEMS Microbiol Lett. 2020;367(2). Epub 2020/02/06. doi: 10.1093/femsle/fnaa013. PubMed PMID: 32009155.
7. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363(4). Epub 2016/01/13. doi: 10.1093/femsle/fnw002. PubMed PMID: 26755501.
8. Hay ID, Lithgow T. Filamentous phages: masters of a microbial sharing economy. EMBO reports. 2019;20(6). Epub 2019/04/07. doi: 10.15252/embr.201847427. PubMed PMID: 30952693; PubMed Central PMCID: PMCPMC6549030.
9. Doore SM, Fane BA. The microviridae: Diversity, assembly, and experimental evolution. Virology. 2016;491:45-55. Epub 2016/02/14. doi: 10.1016/j.virol.2016.01.020. PubMed PMID: 26874016.
10. Olsen NS, Forero-Junco L, Kot W, Hansen LH. Exploring the Remarkable Diversity of Culturable *Escherichia coli* Phages in the Danish Wastewater Environment. Viruses. 2020;12(9). Epub 2020/09/10. doi: 10.3390/v12090986. PubMed PMID: 32899836; PubMed Central PMCID: PMCPMC7552041.
11. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. Nat Commun. 2020;11(1):378. Epub 2020/01/19. doi: 10.1038/s41467-019-14042-z. PubMed PMID: 31953385; PubMed Central PMCID: PMCPMC6969025.
12. Sørensen PE, Van Den Broeck W, Kiil K, Jasinskyte D, Moodley A, Garmyn A, et al. New insights into the biodiversity of coliphages in the intestine of poultry. Sci Rep. 2020;10(1):15220. Epub 2020/09/18. doi: 10.1038/s41598-020-72177-2. PubMed PMID: 32939020; PubMed Central PMCID: PMCPMC7494930.
13. Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguilo-Ferretjans M, Christie-Oleza J, et al. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. Environ Microbiol. 2019;21(6):2112-28. Epub 2019/03/19. doi: 10.1111/1462-2920.14590. PubMed PMID: 30884081; PubMed Central PMCID: PMCPMC6563131.
14. Smith R, O'Hara M, Hobman JL, Millard AD. Draft Genome Sequences of 14 *Escherichia coli* Phages Isolated from Cattle Slurry. Genome Announc. 2015;3(6). Epub 2016/01/02. doi: 10.1128/genomeA.01364-15. PubMed PMID: 26722010; PubMed Central PMCID: PMCPMC4698387.

15. Pacifico C, Hilbert M, Sofka D, Dinhopl N, Pap IJ, Aspock C, et al. Natural Occurrence of *Escherichia coli*-Infecting Bacteriophages in Clinical Samples. Front Microbiol. 2019;10:2484. Epub 2019/11/19. doi: 10.3389/fmicb.2019.02484. PubMed PMID: 31736918; PubMed Central PMCID: PMCPMC6834657.
16. Demerec M, Fano U. Bacteriophage-Resistant Mutants in *Escherichia Coli*. Genetics. 1945;30(2):119-36. Epub 1945/03/01. PubMed PMID: 17247150; PubMed Central PMCID: PMCPMC1209279.
17. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. Journal of molecular biology. 2009;394(4):653-80. doi: 10.1016/j.jmb.2009.09.021. PubMed PMID: 19765592.
18. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol. 2020;7(1):371-84. Epub 2020/06/20. doi: 10.1146/annurev-virology-011620-040628. PubMed PMID: 32559405.
19. Yu YT, Snyder L. Translation elongation factor Tu cleaved by a phage-exclusion system. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(2):802-6. Epub 1994/01/18. doi: 10.1073/pnas.91.2.802. PubMed PMID: 8290603; PubMed Central PMCID: PMCPMC43037.
20. Trojet SN, Caumont-Sarcos A, Perrody E, Comeau AM, Krisch HM. The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. Genome Biol Evol. 2011;3:674-86. Epub 2011/07/13. doi: 10.1093/gbe/evr059. PubMed PMID: 21746838; PubMed Central PMCID: PMCPMC3157838.
21. Mutalik VK, Adler BA, Rishi HS, Piya D, Zhong C, Koskella B, et al. High-throughput mapping of the phage resistance landscape in *E. coli*. PLoS biology. 2020;18(10):e3000877. Epub 2020/10/14. doi: 10.1371/journal.pbio.3000877. PubMed PMID: 33048924; PubMed Central PMCID: PMCPMC7553319 following competing interests: VKM, AMD, and APA consult for and hold equity in Felix Biotechnology, Inc.
22. Koga M, Otsuka Y, Lemire S, Yonesaki T. *Escherichia coli rnlA* and *rnlB* compose a novel toxin-antitoxin system. Genetics. 2011;187(1):123-30. Epub 2010/10/29. doi: 10.1534/genetics.110.121798. PubMed PMID: 20980243; PubMed Central PMCID: PMCPmc3018318.
23. Ragunathan PT, Vanderpool CK. Cryptic-Prophage-Encoded Small Protein DicB Protects *Escherichia coli* from Phage Infection by Inhibiting Inner Membrane Receptor Proteins. Journal of bacteriology. 2019;201(23). Epub 2019/09/19. doi: 10.1128/JB.00475-19. PubMed PMID: 31527115; PubMed Central PMCID: PMCPMC6832061.
24. Shinedling S, Parma D, Gold L. Wild-type bacteriophage T4 is restricted by the lambda *rex* genes. Journal of virology. 1987;61(12):3790-4. Epub 1987/12/01. doi: 10.1128/JVI.61.12.3790-3794.1987. PubMed PMID: 2960831; PubMed Central PMCID: PMCPMC255994.
25. Jahn M, Vorpahl C, Hubschmann T, Harms H, Muller S. Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. Microb Cell Fact. 2016;15(1):211. doi: 10.1186/s12934-016-0610-8. PubMed PMID: 27993152; PubMed Central PMCID: PMCPMC5168713.
26. Pleska M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, et al. Bacterial Autoimmunity Due to a Restriction-Modification System. Curr Biol. 2016;26(3):404-9. Epub 2016/01/26. doi: 10.1016/j.cub.2015.12.041. PubMed PMID: 26804559.
27. Rusinov IS, Ershova AS, Karyagina AS, Spirin SA, Alexeevski AV. Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses. BMC Genomics. 2018;19(1):885. Epub 2018/12/12. doi: 10.1186/s12864-018-5324-3. PubMed PMID: 30526500; PubMed Central PMCID: PMCPMC6286503.

204