



OPEN

Systematic analysis of putative phage-phage interactions on minimum-sized phage cocktails

Felipe Molina¹✉, Manuel Menor-Flores², Lucía Fernández³, Miguel A. Vega-Rodríguez² & Pilar García³

The application of bacteriophages as antibacterial agents has many benefits in the “post-antibiotic age”. To increase the number of successfully targeted bacterial strains, phage cocktails, instead of a single phage, are commonly formulated. Nevertheless, there is currently no consensus pipeline for phage cocktail development. Thus, although large cocktails increase the spectrum of activity, they could produce side effects such as the mobilization of virulence or antibiotic resistance genes. On the other hand, coinfection (simultaneous infection of one host cell by several phages) might reduce the potential for bacteria to evolve phage resistance, but some antagonistic interactions amongst phages might be detrimental for the outcome of phage cocktail application. With this in mind, we introduce here a new method, which considers the host range and each individual phage-host interaction, to design the phage mixtures that best suppress the target bacteria while minimizing the number of phages to restrict manufacturing costs. Additionally, putative phage-phage interactions in cocktails and phage-bacteria networks are compared as the understanding of the complex interactions amongst bacteriophages could be critical in the development of realistic phage therapy models in the future.

Bacteriophages, or phages, are the viruses of bacteria. They infect and replicate within their host and, subsequently, lyse and kill the bacterial cell to release new phage offspring. This “life” cycle confers phages with an intrinsic ability to act as antimicrobials, making them an interesting approach to overcome the current global antibiotic resistance crisis brought about by widespread antibiotic misuse. Indeed, the application of bacteriophages as therapeutics and biocontrol agents is increasingly gaining popularity, especially as a tool to fight against multiresistant pathogenic bacteria in several fields such as human health, agriculture, and the food industry^{1,2}.

Besides their lytic ability, another distinctive feature of phages is their high host specificity. This property would prevent disturbance of the surrounding microbiota when phages are used to remove pathogenic bacteria, which is a significant advantage over antibiotic therapy³. Nevertheless, in terms of efficiency, very specific phages (those infecting only a small number of strains) represent a challenge, making it necessary to use mixtures containing different phages with complementing host ranges, the so-called phage cocktails, to increase the number of target strains and become useful for a wider range of applications/patients⁴. Alternatively, a single phage with a broad host range or a polyvalent phage⁵ can be used to target multiple strains from a species or even multiple species in a community. However, polyvalent phages frequently show differences in infection efficiency (EOP) on different hosts due to, for instance, differential receptor-phage binding affinity.

From a logistical standpoint, therapeutic regimes involving only one phage would be simpler and potentially less costly. Indeed, phages that are part of a cocktail must be propagated individually, purified, and mixed prior to use, which might result in higher manufacturing costs. On the other hand, cocktails improve the commercial applicability of phage formulations partly due to their wider range of bacterial targets. Thanks to this ability, cocktails have the potential to be used without prior identification of the pathogens involved. Furthermore, phage cocktails may reduce or almost eliminate the chance of selecting phage-resistant bacteria⁶. To achieve this goal, phages in the cocktail should preferably belong to different families, so that strains becoming resistant to one phage can be potentially targeted by another phage in the mixture, as superinfection immunity is not possible between unrelated phages.

¹Genética, Facultad de Ciencias, Universidad de Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain. ²Escuela Politécnica, Universidad de Extremadura, Avda. de la Universidad s/n, 10003 Cáceres, Spain. ³Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Paseo Río Linares s/n, 33300 Villaviciosa, Asturias, Spain. ✉email: fmolina@unex.es

To date, there are no validated guidelines for the design of individual or mixed phage preparations targeting a given pathogen⁷. Therefore, in order to assemble a phage cocktail with optimal therapeutic potential, it is first necessary to identify candidate phages, and then determine their infection profiles against a diverse set of strains spanning the major lineages of the target bacterium. Nonetheless, bacteriophage cocktails should not be random mixtures of different bacterial viruses. For instance, Merabishvili et al.⁸ recommended some features to be considered for the selection of phages, which include being strictly virulent, not carrying virulence genes, and exhibiting high killing of the target pathogen, good stability and no transducing ability. In some cases, cocktail design would also involve the choice of bacteriophages with different bacterial cell wall receptor recognition sites^{9,10}, thereby minimizing the emergence of resistance mutants. Additionally, phage candidates should exhibit the broadest host range possible against the target bacterium. When necessary, the host range might be broadened by serial passages on different bacterial strains, thus allowing adaptation of the bacteriophages by coevolution with their host¹¹. Also, there are specific methods available to isolate phages with broad host range¹² or to broaden the host range through site-directed mutagenesis of the host-range-determining regions (HRDRs) in the phage tail fiber protein¹³. Likewise, synthetic biology can be used to modify the phage host range by engineering phage genomes through modular swapping of phage tail components¹⁴.

In some cases, it is necessary to target several bacterial species with one formulation. This application would entail the design of even more complex cocktails, containing mixtures of bacteriophages active against each bacterial species. In this case, cocktail design is mainly based on the microbial composition and the frequency of occurrence of different bacterial species in the target infection site. Examples of complex cocktails include the commercial products Pyo- and Intestibacteriophage¹⁵.

Phage cocktail design must take into account not only interactions between phages and their hosts but also phage-phage interactions. The most well-known example of a phage-phage antagonistic relationship is the superinfection immunity mediated by temperate phages. Indeed, bacteria carrying a prophage are typically resistant to infection by related phages¹⁶. Moreover, phage satellites or pathogenicity islands can limit propagation of their helper phage, which confers some level of phage resistance to host bacteria¹⁷. Because of these and other interactions, the host range of a phage cocktail could be different from that predicted theoretically by combining the host range of each individual phage. Indeed, some phages might interact upon coinfection¹⁸ leading to synergetic or antagonistic effects¹⁹. For example, some cocktails show a narrower host range than the sum of the host ranges of all individual phages independently²⁰. This could be due to host infection competition between phages, i.e., coinfection of the same host cell by multiple phages may lead to increased competition for limited cellular resources, thus reducing the fitness of an individual phage²¹. Antagonistic interactions might also occur when phages share the same receptor sites, or due to abortive infection mechanisms²². In turn, synergistic interactions between phages may be related to the rate of infection, the production of progeny, or the time between infection and progeny release²³. More recently, a phage-phage communication system, arbitrium, that allows cross talk among close phages to regulate the lytic/lysogenic life cycles was described by Stokar-Avihail et al.²⁴. Interestingly, this system may allow interactions with more genetically distant phages.

Overall, it seems clear that optimizing the number and combination of phages will be critical for designing effective therapeutic and biocontrol strategies. In a previous work²⁵, the information provided by host range matrices was analyzed by building phage-bacteria infection networks (PBINs) and calculating the nestedness temperature (host range hierarchy of the phages). Subsequently, an estimator of phage cocktail size (Φ) was proposed considering some global properties of the host range matrices such as fill (fraction of successful infections), nestedness temperature, and number of bacteria. Here, we explore a new method, which considers the host range and each individual phage-host interaction, to design the phage mixtures that best suppress the target pathogens while minimizing the required number of phages.

Results and discussion

Designing phage cocktails from phage-bacteria infection networks (PBINs). In a recent work²⁵, a pipeline for designing phage cocktails based on global properties (fill, temperature, and number of bacteria) of the PBINs was proposed. However, host range matrices needed to be processed by two different sorting algorithms (BinMatNest and Ward) and the phage cocktail was not automatically generated (Fig. 1a). Here, we have developed a new approach (see Methods) that considers individually the host range of every phage, designs cocktails using only one sorting algorithm and automatically provides the Minimum Cocktail Size (MCS) (Fig. 1b).

A total of 50 datasets that include bacteria and phages from different sources, such as seafood, plants, live-stock, dairy, sewage, clinical isolates and laboratory collection strains (Table 1) were used to build the PBINs and calculate the Expected Importance (EI) of every bacterium and phage (see Methods) in order to design the corresponding phage cocktails. Whereas most datasets (42 out of 50) comprised a single bacterial species, matrices harboring 2, 5 and 15 species were also surveyed (Table 1). The most frequent species was *Escherichia coli*, which appeared on 18 matrices, followed by *Pseudomonas aeruginosa* and *Staphylococcus aureus*, which appeared on four and three datasets, respectively. However, most bacterial species (43 out of 52) were present in a single dataset. In total, 2,877 bacterial strains, 899 phages and 52,688 phage-host interactions were evaluated. Subsequent processing of the PBINs led to the construction of subnetworks named Phage Cocktail Networks (PCNs), harboring all bacterial strains susceptible to phage infection as well as the phages contained in a hypothetical cocktail (Fig. 1b). Therefore, a processed network displays two types of source nodes (unselected vs. cocktail phages), two types of target nodes (resistant vs. susceptible bacteria) and two types of edges (unselected phages-bacteria vs. cocktail phages-bacteria interactions). Selection of the phages constituting the candidate cocktails was performed by applying both heuristic (Network Metrics) and exhaustive (Exhaustive Search) algorithms from the Cytoscape application PhageCocktail²⁶ (freely available at <https://apps.cytoscape.org/apps/phagecocktail>), seeking the PCN that maximized the expected number of lysed bacteria while keeping a MCS.

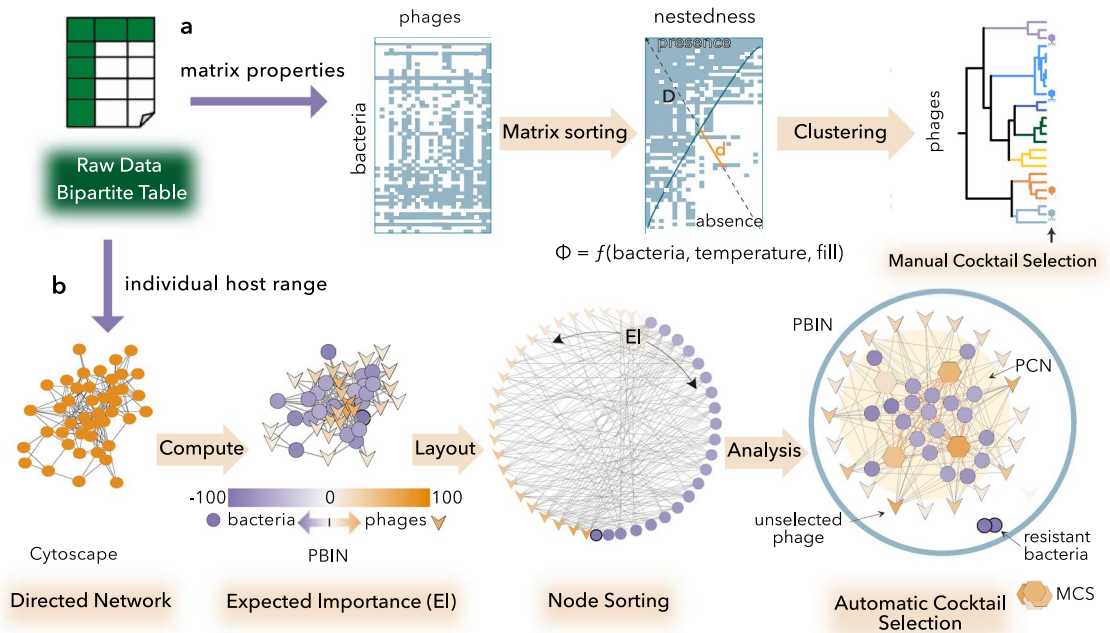


Figure 1. Alternative pipelines for designing phage cocktails. **(A)** Global properties of Phage Bacteria Infection Networks (PBINs). Nestedness algorithms reorder host range data and estimate the deviation (temperature) from a perfectly nested matrix by computing the relative distances (d/D) to the isocline of perfect order (blue line). The metric Φ considers global properties of the networks to estimate phage cocktail size²⁵. Agglomerative hierarchical clustering is used prior to manual selection of the phages constituting each cocktail. **(B)** Automatic determination of the Minimum Cocktail Size (MCS). Bipartite phage-bacteria interaction matrices are imported into Cytoscape as directed networks, and the expected importance (EI) is measured for each node (see Methods). Nodes are colored and sorted by their EI and cocktails are designed, both heuristically and exhaustively, using the app PhageCocktail. The subnetwork (Phage Cocktail Network) harboring all susceptible bacterial strains and phages corresponding to the MCS is selected for each PBIN.

The heuristic algorithm employs EI values and considers phages with a broad host range (high EI_p values) as better candidates for the cocktail, and bacteria susceptible to be infected by fewer phages (low EI_b values) as more problematic. Contrastingly, the exhaustive algorithm evaluates all the possible phage combinations, which warrants the identification of the smallest theoretical cocktail. However, this algorithm imposes a computational runtime penalty and may take several hours to execute²⁶ when the number of phages is large. For instance, one matrix (named Korf et al. in Table 1) needed to be processed overnight. It must be highlighted that, even though alternative algorithms, such as agglomerative hierarchical clustering of phages by their host range, are faster, the exhaustive algorithm employed in this work warrants the best results (Díaz-Galián et al., submitted), i.e., the MCS infecting the highest number of bacteria.

Since Exhaustive Search outperforms Network Metrics, yielding smaller cocktails for the same number of infected bacteria (data not shown), we selected the PCNs obtained with the latter algorithm for further evaluation (Fig. 2). To compare both algorithms the expected importance (EI) of each node, determined by the Network Metrics algorithm (see Methods), is also shown. It is noteworthy that the Exhaustive Search preferably selected generalist phages displaying high EI values (Fig. 2). Although the current implementation of Exhaustive Search in PhageCocktail 1.1 is limited to combinations comprising up to 12 phages, this did not represent a problem in our search given that the size of the largest candidate phage cocktail predicted by the algorithm was 11.

The complexity of a bipartite network depends on its size, symmetry, and fill (fraction of the maximum number of edges). For instance, square matrices result in more complex networks than similarly sized asymmetric networks because the maximum number of edges (possible phage-host interactions) increases. Additionally, the maximum algorithmic complexity²⁷ of a PBIN is found for intermediate fill values because broad-host-range phages simplify cocktail design and any phage with an EI_p of 100 might be able to infect every host in a dataset. With this in mind, we sought to further characterize the PBINs by comparing the number of bacteria, phages and fill of each matrix (Fig. 3a). Most of the PBINs (80%) comprised more bacteria than phages, and the number of nodes (phages plus bacteria) ranged between 10 and 906. Large matrices were lightly filled suggesting a weak negative correlation between matrix size and fill (see below). Two outgroups harbored a significantly higher than average number of phages (166 vs. 18) and bacteria (896 vs. 58). These results showed the lack of consensus regarding the number of bacteria needed for host range determination, which coincides with the data obtained from a recent survey²⁸ that found very little agreement in this regard between researchers (most answers stretched between 20 and 100 but one reached 800). Similarly, bias in data gathering, such as not considering narrow-host-range phages, might skew the fill of the PBINs.

References	Matrix size	Hosts	Phages	Source	
Hong et al. (2013) ^a	21	<i>Salmonella gallinarum</i>	7	3	Sewage, laboratory
Shende et al. (2017) ^a	40	<i>Escherichia coli</i> , <i>Bacillus subtilis</i>	8	5	Manure, sewage water
Mizuno et al. (2020) ^b	64	<i>Citrobacter rodentium</i>	32	2	Sewage treatment plant
Liao et al. (2019) ^a	68	<i>Escherichia coli</i>	17	4	Non-fecal compost, laboratory
Krasowska et al. (2015) ^a	76	<i>Bacillus subtilis</i>	19	4	Soil, laboratory
VHR1 ^b , James L. Van Etten lab	78	<i>Chlorella variabilis</i>	6	13 ^c	Laboratory
Gutiérrez et al. (2015) ^b	90	<i>Staphylococcus</i> , <i>Macrocooccus caseolyticus</i>	45	2	Sewage treatment plant
Hwang et al. (2009) ^a	96	<i>Campylobacter jejuni</i>	16	6	Poultry, sewage, soil, laboratory
Kwiatk et al. (2015) ^a	100	<i>Pseudomonas aeruginosa</i>	20	5	Sewage, clinical
Hammerl et al. (2016) ^a	108	<i>Brucella inopinata</i>	36	3	Laboratory
Xie et al. (2016) ^a	120	<i>Salmonella enterica</i>	12	10	Manure, water, soil, laboratory, cattle feedlots
Magaré et al. (2017) ^a	125	<i>Bacillus novalis</i> , <i>Pseudomonas aeruginosa</i> , <i>Staphylococcus epidermidis</i> , <i>Bacillus cereus</i> , <i>Staphylococcus aureus</i>	5	25	Air
Álvarez et al. (2019) ^a	126	<i>Ralstonia solanacearum</i>	42	3	River water, potatoes, laboratory
Pereira et al. (2016) ^a	126	<i>Salmonella typhimurium</i>	42	3	Sewage, food, water, laboratory
Yu et al. (2016) ^a	155	<i>Pseudomonas syringae</i> pv. <i>actinidiae</i>	31	5	Soil, kiwifruit orchards
VHR5 ^b , PMID: 22,936,928, 26,884,161, 10,430,569, 24,433,295, 22,834,906, 14,592,760	156	<i>Sulfolobus</i> strains, <i>Pyrobaculum arsenaticum</i> , <i>Pyrobaculum oguniense</i> , <i>Sulfolobus islandicus</i> , <i>Acidanus</i> strains	12	13	Acidic hot springs
Denou et al. (2009) ^b	156	<i>Escherichia coli</i>	26	6	Human feces
Schouler et al. (2021) ^b	168	<i>Escherichia coli</i>	56	3	Chicken fecal, recombinant phages
Gutierrez et al. (2010) ^b	195	<i>Staphylococcus epidermidis</i>	65	3	Women's breast milk
Dias et al. (2013) ^a	200	<i>Staphylococcus aureus</i>	20	10	Livestock, sewage
Maura et al. (2012) ^b	219	<i>Escherichia coli</i>	73	3	Human feces
Galtier et al. (2017) ^b	219	<i>Escherichia coli</i> strain	73	3	Feces homogenates from murine gut samples
Alić et al. (2017) ^a	220	<i>Dickeya</i>	55	4	Orchid, wastewater
Molina et al. (2021) (3C) ^a	260	<i>Citrobacter freundii</i> , <i>Citrobacter youngae</i> , <i>Escherichia coli</i> , <i>Hafnia alvei</i> , <i>Klebsiella pneumoniae</i> , <i>Lactobacillus acidophilus</i> , <i>Lactobacillus casei</i> , <i>Lactococcus lactis</i> ssp. <i>Lactis</i> , <i>Salmonella typhimurium</i> , <i>Enterobacter aerogenes</i> , <i>Serratia marcescens</i> , <i>Shigella boydii</i> , <i>Shigella flexneri</i> , <i>Shigella sonnei</i> , <i>Yersinia enterocolitica</i>	26	10	Manure, sewage, laboratory, dairy
Salifu et al. (2013) ^a	270	<i>Rhodococcus equi</i>	27	10	Soil, equine
Arachchi et al. (2014) ^a	300	<i>Listeria monocytogenes</i>	50	6	Laboratory, seafood
Oh et al. (2017) ^a	324	<i>Bacillus cereus</i>	27	12	Laboratory, fermented food, soil
Wandro et al. (2019) ^a	330	<i>Enterococcus faecium</i>	15	22	Sewage human feces
Gunathilaka et al. (2017) ^a	348	<i>Escherichia coli</i>	12	29	Wastewater, laboratory
Jurczak-Kurek et al. (2016) ^a	360	<i>Escherichia coli</i> , <i>Pseudomonas aeruginosa</i> , <i>Salmonella enterica</i> , <i>Staphylococcus sciuri</i> , <i>Enterococcus faecalis</i>	60	6	Clinical, urban sewage
Litt and Jaroni (2017) ^a	378	<i>Escherichia coli</i>	54	7	Clinical, cattle feces
Romero-Suarez et al. (2012) ^a	416	<i>Xanthomonas arboricola</i> pv. <i>juglandis</i>	16	26	Walnut orchards
Wang et al. (2015) ^a	451	<i>Escherichia coli</i>	41	11	Cattle feces, human
Murphy et al. (2013) ^a	480	<i>Lactococcus lactis</i>	20	24	Dairy, Gouda-type cheese-producing plants Lactococcal phages
Sajben-Nagy et al. (2012) ^a	544	<i>Pseudomonas tolaasii</i>	34	16	Laboratory, mushroom
Sekulovic et al. (2014) ^b	555	<i>Clostridium difficile</i>	37	15	Animal and human fecal
Mangieri et al. (2020) ⁵⁰	630	<i>Escherichia coli</i>	30	21	Cattle and sheep feces, bedding material, sewage
Galtier et al. (2016) ^b	876	<i>Escherichia coli</i>	73	12	Sewage
Vu et al. (2019) ^a	1209	<i>Listeria</i>	31	39	Vegetable, seafood, livestock, foods and food processing environments
VHR14 ^b , Mathieu et al. (2020)	1344	<i>Escherichia coli</i>	84	16	Fecal samples of 1-year-old children
Molina et al. (2021) (3A) ^a	1456	<i>Escherichia coli</i>	56	26	Livestock feces, dairy, laboratory
Petsong et al. (2019) ^a	1692	<i>Salmonella enteritidis</i> , <i>Salmonella typhimurium</i>	47	36	Livestock
Jäckel et al. (2017) ^a	2147	<i>Ochrobactrum</i> , <i>Brucella inopinata</i>	113	19	Laboratory
Brady et al. (2017) ^a	2280	<i>Paenibacillus larvae</i>	40	57	Beehive
Lourenço et al. (2020) ^b	2744	<i>Escherichia coli</i>	98	28	Sewage water, laboratory
Continued					

References	Matrix size	Hosts	Phages	Source	
Fong et al. (2019) ⁵¹	2806	<i>Salmonella enterica</i>	61	46	Sediment, cattle feces, sewage effluent, irrigation water, water tanks from an aquaculture facility
Gencay et al. (2019) ^a	2952	<i>Salmonella</i>	72	41	Laboratory, pork meat, environmental and wastewater samples
Korf et al. (2019) ^a	3200	<i>Escherichia coli</i>	64	50	Poultry, sewage, manure, clinical
Saussereau et al. (2014) ^b	8960	<i>Pseudomonas aeruginosa</i>	896	10	Cystic fibrosis isolates, laboratory
Mathieu et al. (2020) ^a	12,450	<i>Escherichia coli</i>	75	166	Fecal
Total	52,688		2877	899	

Table 1. Experimental host range matrices used to generate Phage-Bacteria Infection Networks. ^aFull reference available at Molina et al. (2021) <https://doi.org/10.3389/fmicb.2021.564532>. ^bDownloaded from <https://viralhostrangedb.pasteur.cloud/data-source/>. ^cThe original matrix was trimmed to remove gaps.

Analysis of minimum-sized phage cocktail networks (PCNs). The distribution of the MCS in the different datasets (Fig. 3b) shows that most of the PCNs were indeed single phage formulations, followed by 2- and 3-phage cocktails, whereas larger (4, 5, 7, 8 and 11) cocktail sizes were less common. On the other hand, the number of bacteria in PCNs with a low MCS tended to be smaller. Thus, the minimum number of bacteria in large cocktails (containing four or more phages) was 10, whereas 19% of the PCNs corresponding to smaller cocktails contained nine or fewer bacterial strains.

In a previous work²⁵, the size of candidate phage cocktails was estimated using a metric (Φ) that considers properties of PBINs such as nestedness, temperature, fill and the number of target bacteria. However, Φ does neither evaluate individual phage-host interactions nor provide the minimum possible cocktail size. Comparison of both estimators, MCS and Φ , revealed (Fig. 3c) that Φ was larger than MCS for half (49%) of the PBINs, while being identical for most of the other half (37% of the total). Strikingly, Φ was smaller than MCS for the remaining PBINs (14%) most of which were fairly large networks. Remarkably, Φ takes into account the ecological and evolutionary information provided by the structure of PBINs, but it does not intend to achieve the minimum possible cocktail size. Also, Φ can be even larger than the number of phages assayed²⁵ suggesting that additional phages should be isolated to perform effective biocontrol of the bacterial population. Conversely, the Exhaustive Search algorithm used to determine MCS examines individual phage-host interactions, thereby providing the optimal theoretical cocktail, rather than global properties of the network, and the expected success of the cocktails can be gauged by comparing PBINs and PCNs.

For fully successful cocktails, all the bacteria in a PBIN should also be part of the final PCN. With this in mind, the expected cocktail efficacy was estimated as the fraction (%) of bacteria susceptible to at least one phage in the cocktail by calculating the ratio of bacteria in every PCN/PBIN pair (Fig. 3d). These calculations revealed that 26% of the PCNs comprised all the bacterial strains from the original PBINs, which corresponds to the fraction (13 out of 50) of fully successful phage cocktails. While the fraction of bacteria lysed is generally high for all cocktail sizes (the median of all datasets is 74%), the variability of the expected efficacy decreases as MCS increases. This, however, is likely due to the asymmetric distribution of cocktail sizes (Fig. 3b). Interestingly, large phage cocktails did not exhibit low efficacy values (Fig. 3d) evincing that the algorithms used to design the cocktails do not spuriously increase the MCS. In the case of the heuristic algorithm, MCS is determined by favoring generalist phages with broad host range over specialists with narrow host range. In this regard, several experiments have shown that host range is a highly evolvable trait that relies on diversity, density and quality of hosts²⁹. However, although host range expansion seems advantageous for phages, generalism may come with some tradeoffs such as a lower propagation rate³⁰. Taking this into consideration, it would be interesting to develop alternative strategies for designing phage cocktails that maximize phage productivity, for example by selecting phages with large burst sizes and/or short latent periods, rather than those that minimize cocktail size.

Phage-phage interactions: PBINs vs. PCNs. Although all the datasets evaluated in this work comprise experimental data, the actual performance “in vivo” of the phage cocktails designed here might be affected by phage-phage interactions that could hamper or facilitate coinfection (simultaneous infection of one host cell by several phages). To evaluate these putative interactions, PCNs were grouped by their MCS and then used to calculate the percentage of bacteria lysed by a given number of phages in all cocktails of the same size (Fig. 4a). This analysis revealed the intensity and frequency of phage-phage interactions for different cocktail sizes. Notably, the bulk of the bacterial hosts were lysed by a single phage, even in large cocktail networks, and only in the group comprising 3-phage cocktails there were more bacteria lysed by two phages than by a single one. Additionally, no bacterial host was lysed by 7, 9, 10 or 11 phages in the PCNs containing 7, 8 or 11 phages. These results suggest that, even though search algorithms preferably select generalist phages, the nestedness of the PCNs might be lower than that of the original PBINs²⁵. Consequently, despite the fact that phages in nature span the continuum from specialist to generalist³¹, cocktail design might alter this equilibrium by favoring the selection of broad-host-range phages.

To further compare the structure of PBINs and PCNs, we studied the redundancy variation (rv) of phage coinfections. The redundancy of phage-host interactions (r), defined as the fraction of phages lysing a given bacterial strain in a network (see Methods), can increase, decrease or remain constant when comparing PBINs and PCNs (see examples in Fig. 4b). In turn, rv allows comparison of the redundancy found in a PCN compared

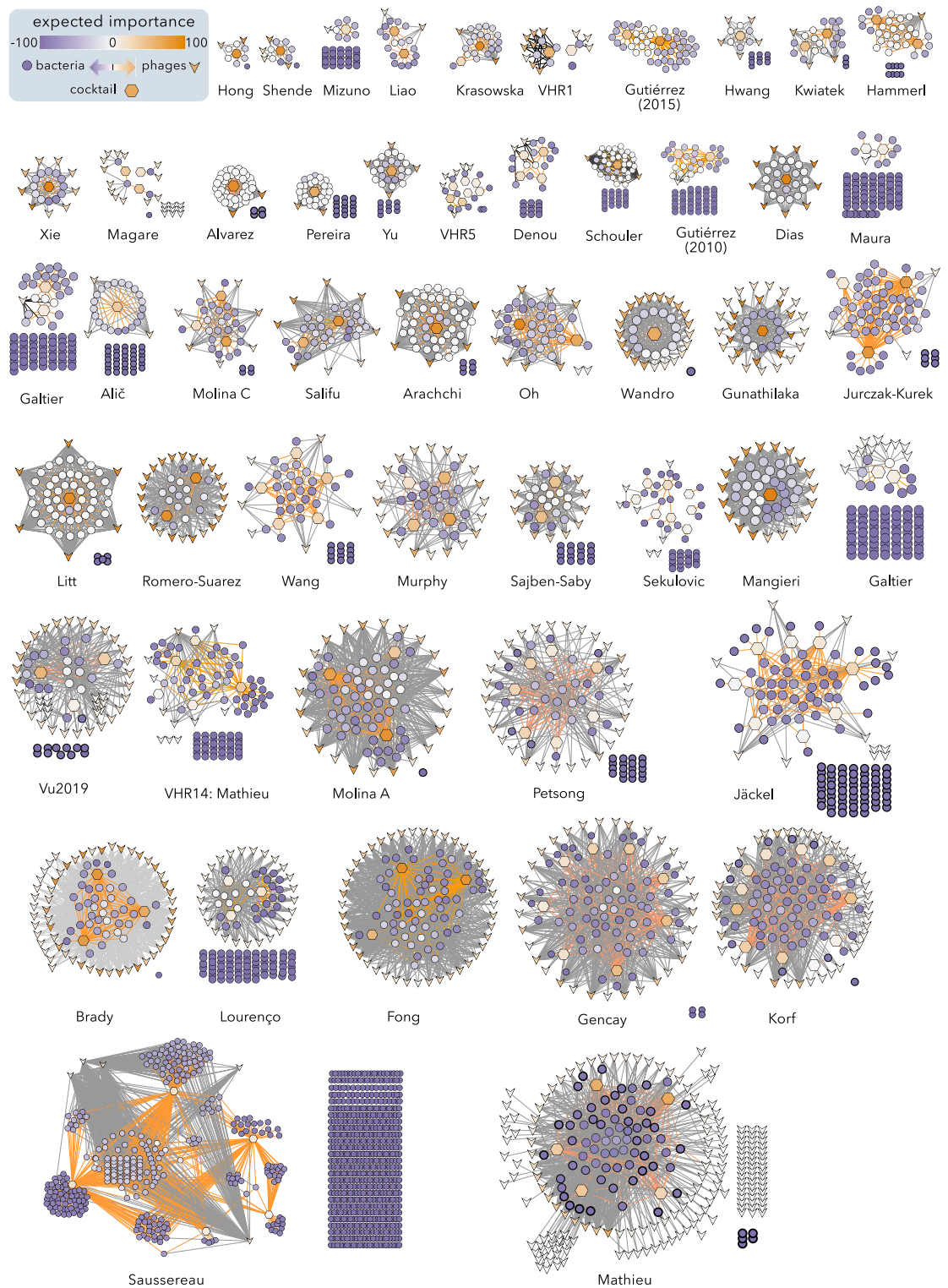


Figure 2. Graphical representation of Phage Bacteria Infection Networks (PBINs) and candidate Phage Cocktail Networks (PCNs) depicting the Minimum Cocktail Size (MCS). A total of 50 PBINs, each harboring a PCN subnetwork, were built and phage cocktails were designed using an exhaustive algorithm as detailed in Methods and sorted by increasing matrix size (Table 1). The shape of the nodes represents bacteria (●), unselected phages (▼) and cocktail phages (●). The expected importance (EI) of each node represented by color shading so that more relevant nodes show more intense colors. Bacteria not susceptible to any phage ($EI_b = -100$) are clustered in grids. Lysis is indicated by dark (unselected phages) or orange (phage cocktail) lines.

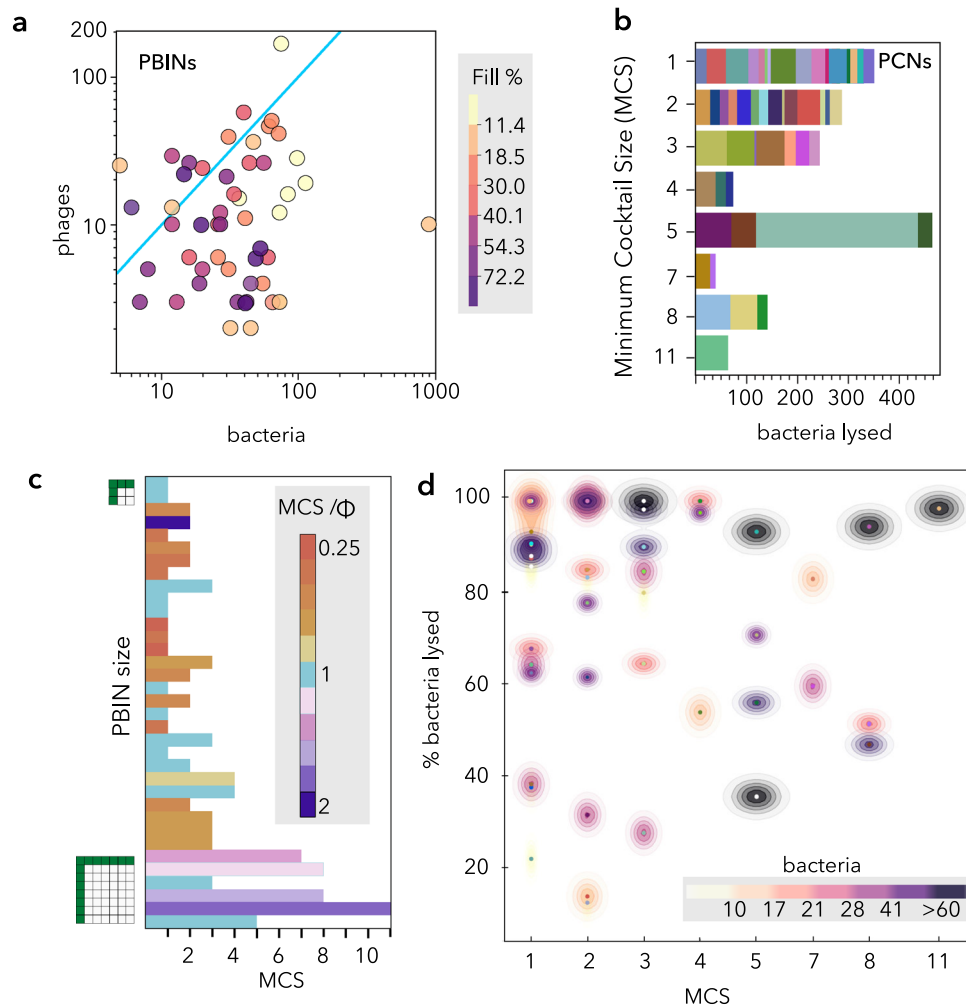


Figure 3. Characterization of networks (PBINs and PCNs) complexity and expected cocktail efficacy. **(A)** Analysis of PBINs complexity and symmetry. Each dot represents a single matrix (see Fig. 2) and the fill (%) is represented by color intensity. The cyan line indicates the position of symmetric matrices. **(B)** Distribution of PCNs grouped by the number of phages (MCS). Each bar sector represents a PCN and its length correlates with the number of bacteria lysed by the cocktail. **(C)** Comparison of MCS and Φ estimators. 34 PBINs, taken from Molina et al.²⁵, were sorted by decreasing size. Bar length indicates the MCS, whereas color corresponds to the MCS/ Φ ratio. **(D)** Expected cocktail efficacy vs. MCS. The fraction (%) of bacteria susceptible to at least one phage of the cocktail is shown for every PBIN. The number of bacteria is grouped by percentiles and represented by a density plot.

to that of the original PBIN. This parameter ranges between a minimum (rv_{\min}), when only one phage in the cocktail infects the host, and a maximum (rv_{\max}), when all the phages lysing the host are considered. However, for single phage formulations $rv = rv_{\min} = rv_{\max}$. Consequently, rv can never decrease (Fig. 4b) and its value depends on how many phages infect the host in the PBIN (see Methods). After grouping the PCNs by MCS the redundancy change (rv) of the different bacterial strains was determined as $\log_{10}(r_{\text{PCN}}/r_{\text{PBIN}})$ (Fig. 4c). It can be perceived that the median values of rv increased (up to 2.5-fold) for all but single phage cocktails. This result implies that most bacterial strains were lysed by all the phages in the PBINs corresponding to single phage cocktails. For the rest of the networks there was a fraction of bacteria showing negative rv values while others exhibited an increase of up to 35 times (1.5 log). However, rv did not reach 0.8 in large (7, 8 and 11-phage) cocktails, i.e., redundancy increased less than 6.3 times. Concurrently, a paired samples t-test revealed (data not shown) a statistically significant 0.17 log difference between the means of rv and rv_{\min} and a 0.27 log difference between rv and rv_{\max} , evincing the asymmetric effect of phage cocktails on rv . Remarkably, when rv values were grouped by the number of lysing phages instead of the MCS (Fig. 4d), the variability observed for rv decreased as the number of lysing phages increased. This result is partially explained by the low number of bacteria susceptible to a high number of phages in the cocktails, but it also reflects that these host strains are susceptible to be infected by a high number of phages both in the PBINs and PCNs. Conversely, bacteria lysed by a single phage in the cocktails were heterogeneous regarding their susceptibility to the other phages in their respective PBINs. Thus, some host strains susceptible to a single phage in the cocktails can be infected by multiple phages in the corresponding PBIN.

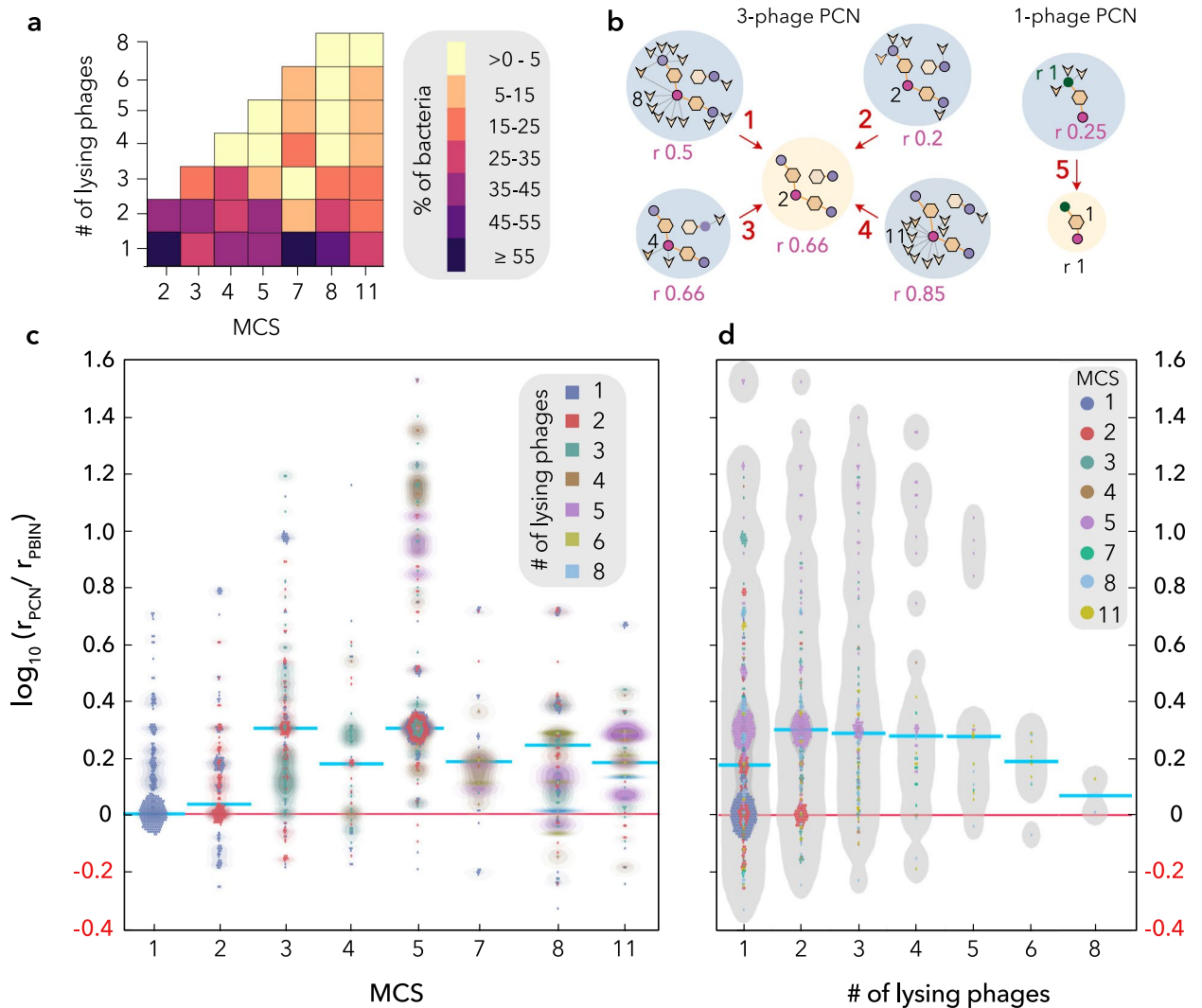


Figure 4. Analysis of phage-phage interactions: PBINs vs. PCNs. **(A)** Number of phages lysing bacterial strains for each cocktail size. The PCNs were grouped by MCS values and the fraction (%) of bacterial strains lysed by different number of phages is shown as a heatmap. **(B)** Examples of redundancy variations. Original PBINs and resulting PCNs are represented by blue and yellow shaded circles, respectively. Redundancy (r), the fraction of phages infecting a bacterial strain, is shown for one specific strain (red circle) in five PBINs and two PCNs. Additionally, the redundancy of a second strain is shown in PBIN #5. Examples of redundancy increase (1, 2, 5 ●), constancy (3 ●; 5 ●) and decrease (4 ●) are shown. **(C)** Redundancy variation (rv) of phage-host networks. PCNs were grouped by MCS and the redundancy change of the bacterial strains is represented as dots and density lines. The color indicates the number of lysing phages. The median of each distribution is shown by a cyan line. **(D)** Redundancy variation (rv) vs. number of lysing phages. Bacterial strains (dots) were grouped by the expected number of lysing phages and highest density region (HDR) box plots were generated. The color of the different dots indicates the MCS values and the cyan line represents the median of each distribution.

Recently, Niu et al.¹⁹ found different combinations of phages in cocktails displaying neutral, synergistic or antagonistic effects. Phages coinfecting the same host gain access to a common pool of proteins, which can be considered as intracellular public goods. Subsequently, a plethora of social behaviors may arise including cheaters, which do not contribute to the production of public goods, and cooperators, such as phages producing depolymerases that facilitate the adsorption of other phages to the host²⁹. On the other hand, the phage-phage arms race might entail various strategies to prevent superinfection, such as triggering the premature lysis of bacterial cells infected by competitors³² or phage-encoded CRISPR-Cas systems³³. Therefore, phage-phage interactions are worth examining carefully, as they might be detrimental or beneficial to the outcome of phage cocktail application. With this in mind, redundancy is a useful parameter to be considered in order to formulate cocktails with superior efficacy.

Nonetheless, the intensity of phage-phage interactions also depends, to some extent, on phage-host dynamics. For instance, the bacteriophage multiplicity of infection (MOI), ratio of phages to bacteria, is frequently employed to calculate the amount of phages that should be applied during dosing. Additionally, bacterial densities can change, and target bacteria may not be homogeneous regarding phage access³⁴. As a result, redundancy variation

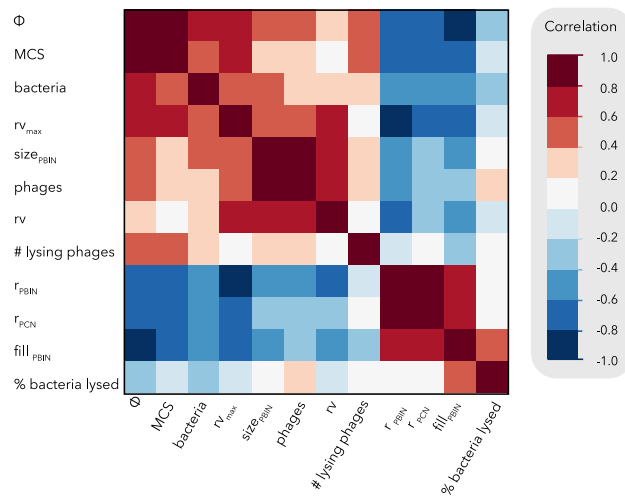


Figure 5. Correlation heatmap of different phage-host network parameters. Correlation values correspond to Spearman's r . All strong positive and negative correlations were statistically significant (P value < 0.001).

analysis could be customized by incorporating additional information such as the MOI, the ratio of phageome to microbiome in the original environment, the burst size of different phages or their virulence index score³⁵ to further improve its value during cocktail design.

Size is not a determining factor for cocktail efficacy. Next, we assessed the potential correlations among the different parameters that characterize the networks by building a heatmap (Fig. 5). Interestingly, the expected success of the cocktails (% of bacteria lysed) displays a moderately positive correlation with the fill of the network, whereas its correlation with the number of phages, bacteria or cocktail size (MCS) is weak (< 0.4). A previous comparison of 31 datasets (Molina et al.²⁵, data extracted from Chan et al.³⁶) had already shown that the observed bacterial load reduction does not increase significantly with the size of the phage cocktail.

As shown in Fig. 3C, there is a strong positive correlation (Spearman's $r = 0.84$) between MCS and Φ . Also, both parameters show a negative correlation with the fill of the PBIN, being particularly strong (> 0.8) in the case of Φ (Fig. 5). Additionally, MCS has a positive correlation with the number of target bacteria and rv_{max} . In contrast, cocktail size is not correlated at all with the experimental rv (Fig. 4c), probably due to its negative correlation with redundancy in both PBINs and PCNs. These results indicate that datasets comprising more bacteria correspond to low filled networks, which leads to larger cocktail sizes and low redundancy values. To fully distinguish whether this reflects the different structure of small and large networks or some bias in data gathering by researchers, untrimmed datasets would be required.

Concluding remarks. This study provides a simple and straightforward method to tackle cocktail design commencing with host range matrices. However, it must be noted that the use of binary networks, which only consider host range, leads to the loss of valuable information regarding infection efficacy³⁷. It has been well established that a broad host range is not necessarily a synonym of success for bacteriophages³⁸, especially because niche expansion may involve facing some antagonistic pleiotropic costs³⁹, such as the loss of virulence⁴⁰ and/or infection efficacy⁴¹. Moreover, the co-evolution of phages and their hosts often entails changes in bacterial community composition and structure⁴² or a decrease in the phage growth rate⁴³. These strategies would ensure preservation of the host bacterial population for long-term exploitation, as well as increase extracellular stability of the newly-formed virions⁴⁴. Likewise, the existence of synergy and/or antagonism among phages co-infecting the same host¹⁹ suggests that the development of new algorithms for phage cocktail design should involve the analysis of quantitative matrices and co-infection values. This method would allow the differentiation between simple short-term and more complex long-term biocontrol of host populations.

Immunity networks, built from CRISPR-Cas sequences, have been shown to influence the structure of complex PBINs due to their effect on virus diversification and host control⁴⁵. Moreover, network proximity analyses have been used for rapid identification of potential drug combinations targeting 2019-nCoV/SARS-CoV-2⁴⁶. Therefore, network-based methodologies can be applied not only to the control of microbes in foods, industrial settings and phage therapy applications, but also to study the dynamics of the microbiome, the evolution of host and viral populations³⁷, and even to fight the SARS-CoV-2 pandemic⁴⁷. Notwithstanding these examples of implementation of biological networks comprising viruses, the systematic analysis of PCNs is still far from being widespread. The present work attempts to tackle the potential offered by these strategies towards building a well-defined pipeline for phage cocktail design. However, further work will still be necessary to incorporate additional information regarding phage-phage interactions so that phage-bacteria networks accurately reflect the complexity of synergies and antagonisms among phages.

Methods

Data collection. A total of 50 datasets (Table 1) were gathered from three different sources: 34 tables were obtained from Molina et al.²⁵, 14 were downloaded (<https://viralhostrangedb.pasteur.cloud>) from the Viral Host Range database⁴⁸, and 2 were randomly selected after searching MEDLINE using the term “phage cocktail” and restricting the date from January 2019 onwards.

Generation of phage-bacteria infection networks (PBINs). The host range matrices (Table 1) were processed in the form of bipartite phage-host interaction matrices, reducing the lytic spectrum to either lytic or non-lytic interactions. These matrices were later transformed into directed networks (PBINs) by importing them into the Cytoscape platform⁴⁹ using the plugin aMatReader. The Expected Importance (EI) of each phage (EI_p) and bacterium (EI_b) within the generated networks was calculated according to Eqs. 1 and 2, respectively:

$$EI_p = 100 \cdot \frac{out_d}{|B|}, \quad (1)$$

$$EI_b = -1 \cdot \left[100 \cdot \left(1 - \frac{in_d}{|P|} \right) \right], \quad (2)$$

where $|\cdot|$ is the cardinality of the bacteria (B) and phages (P), and out_d and in_d refer to the outdegree and indegree of the nodes, that is, the number of arrows leaving source nodes (phages) or entering target nodes (bacteria), respectively. These equations consider that phages with a broad host range (higher outdegree values) are initially better candidates to constitute the cocktail, and that bacteria susceptible to be infected by fewer phages (lower indegree values) are more problematic for cocktail design. Additionally, it must be noted that the values generated by these two equations have opposite signs, which simplifies color labelling of source (phages) and target (bacteria) nodes in the networks (Fig. 1). An exception to this are phages that do not infect any bacterium or bacteria infected by every single phage in the dataset because both receive an EI of zero.

Heuristic search of minimum sized phage cocktails. First, bacteria lysed by a single phage ($EI_b = 1/|P|$) were identified for every network, and the corresponding phages were incorporated into the cocktail. Subsequently, the phage lysing the most bacteria (i.e., with the highest EI_p value) was included in the cocktail, unless no additional bacterial strains were lysed by the newly added phage. This process was iterated until all susceptible bacteria were predicted to be lysed by the cocktail, using the Network Metrics algorithm implemented in the Cytoscape app PhageCocktail 1.1.²⁶

Exhaustive search of minimum sized phage cocktails. The Exhaustive Search option of the Cytoscape app PhageCocktail 1.1 was used to perform a combinatorial calculation of all the possible phage combinations from size 1 up to a 12-phage cocktail size. Then, it returns the best phage combination (the one that lyses the most bacteria) for each size until reaching the maximum number of lysed bacteria. The last cocktail shown in the generated output file corresponds to the highest percentage of lysed bacteria, and, hence, was selected as the best candidate. The number of phages in this cocktail was then considered to be the Minimum Cocktail Size (MCS).

Analysis of redundancy variation. Phage Cocktail Networks (PCNs) were built from the original PBINs by selecting susceptible bacteria and the phages constituting a given cocktail. Redundancy (r) was defined as the number of phages lysing a bacterial strain (indegree) relative to the number of phages in the network, and redundancy variation (rv) was measured for each bacterial strain by comparing the redundancy in the PCN (r_{PCN}) to that in the whole PBIN (r_{PBIN}) as indicated in Eq. 3.

$$rv = \log_{10} \left(\frac{r_{PCN}}{r_{PBIN}} \right) = \log_{10} \left(\frac{\frac{in_d}{|MCS|}}{\frac{in_d}{|P|}} \right), \quad (3)$$

This metric varies between a minimum (rv_{min}), when only one phage in the cocktail infects a given bacterial host, and a maximum (rv_{max}), when every phage infecting each host strain in the PBINs is considered (Eq. 4).

$$rv_{min} = \log_{10} \left(\frac{|P|}{|MCS| \cdot in_d} \right); \quad rv_{max} = \log_{10} \left(\frac{|P|}{|MCS|} \right), \quad (4)$$

Data availability

The authors confirm that the data supporting the findings of this study are available within the article.

Received: 17 November 2021; Accepted: 31 January 2022

Published online: 14 February 2022

References

1. Moyer, Z. D., Woolston, J. & Sulakvelidze, A. Bacteriophage applications for food production and processing. *Viruses* **10**, 205 (2018).
2. Principi, N., Silvestri, E. & Esposito, S. Advantages and limitations of bacteriophages for the treatment of bacterial infections. *Front. Pharmacol.* **10**, 513 (2019).

3. Bruttin, A. & Brüssow, H. Human volunteers receiving *E. coli* phage T4 orally: a safety test of phage therapy. *Antimicrob. Agents Chemother.* **49**, 2874–2878 (2005).
4. Chadha, P., Katare, O. P. & Chhibber, S. *In vivo* efficacy of single phage versus phage cocktail in resolving burn wound infection in BALB/c mice. *Microb. Pathog.* **99**, 68–77 (2016).
5. Hamdi, S. *et al.* Characterization of two polyvalent phages infecting *Enterobacteriaceae*. *Sci. Rep.* **7**, 40349 (2017).
6. Chan, B. K. & Abedon, S. T. Phage therapy pharmacology phage cocktails. *Adv. Appl. Microbiol.* **78**, 1–23 (2012).
7. Pirnay, J.-P. *et al.* The phage therapy paradigm: Prêt-à-Porter or sur-mesure? *Pharm. Res.* **28**, 934–937 (2011).
8. Merabishvili, M., Pirnay, J.-P. & Vos, D. D. Guidelines to compose an ideal bacteriophage cocktail. *Methods Mol. Biol.* **1693**, 99–110 (2018).
9. Tanji, Y. *et al.* Toward rational control of *E. coli* O157:H7 by a phage cocktail. *Appl. Microbiol. Biotechnol.* **64**, 270–274 (2004).
10. Altamirano, F. L. G. & Barr, J. J. Unlocking the next generation of phage therapy: the key is in the receptors. *Curr. Opin. Biotechnol.* **68**, 115–123 (2021).
11. Friman, V. *et al.* Pre-adapting parasitic phages to a pathogen leads to increased pathogen clearance and lowered resistance evolution with *Pseudomonas aeruginosa* cystic fibrosis bacterial isolates. *J. Evol. Biol.* **29**, 188–198 (2016).
12. Ross, A., Ward, S. & Hyman, P. More is better: selecting for broad host range bacteriophages. *Front. Microbiol.* **7**, 1352 (2016).
13. Yehl, K. *et al.* Engineering phage host-range and suppressing bacterial resistance through phage tail fiber mutagenesis. *Cell* **179**, 459–469.e9 (2019).
14. Ando, H., Lemire, S., Pires, D. P. & Lu, T. K. Engineering modular viral scaffolds for targeted bacterial population editing. *Cell Syst.* **1**, 187–196 (2015).
15. Zschach, H. *et al.* What can we learn from a metagenomic analysis of a georgian bacteriophage cocktail? *Viruses* **7**, 6570–6589 (2015).
16. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* **10**, 2854–2866 (2016).
17. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci. USA* **109**, 16300–16305 (2012).
18. Abedon, S. & Thomas-Abedon, C. Phage therapy pharmacology. *Curr. Pharm. Biotechnol.* **11**, 28–47 (2010).
19. Niu, Y. D. *et al.* Efficacy of individual bacteriophages does not predict efficacy of bacteriophage cocktails for control of *E. coli* O157. *Front. Microbiol.* **12**, 616712 (2021).
20. Forti, F. *et al.* Design of a broad-range bacteriophage cocktail that reduces *Pseudomonas aeruginosa* biofilms and treats acute infections in two animal models. *Antimicrob. Agents Chemother.* **62**, e02573–e2617 (2018).
21. Turner, P. E., Burch, C. L., Hanley, K. A. & Chao, L. Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage $\phi 6$. *J. Virol.* **73**, 2420–2424 (1999).
22. Abedon, S. T. Bacteriophage secondary infection. *Virol. Sin.* **30**, 3–10 (2015).
23. Schmerer, M., Molineux, I. J. & Bull, J. J. Synergy as a rationale for phage therapy using phage cocktails. *PeerJ* **2**, e590 (2014).
24. Stokar-Avihail, A., Tal, N., Erez, Z., Lopatina, A. & Sorek, R. Widespread utilization of peptide communication in phages infecting soil and pathogenic bacteria. *Cell Host Microbe* **25**, 746–755.e5 (2019).
25. Molina, F. *et al.* A new pipeline for designing phage cocktails based on phage-bacteria infection networks. *Front. Microbiol.* **12**, 564532 (2021).
26. Menor-Flores, M., Vega-Rodríguez, M. A. & Molina, F. Computational design of phage cocktails based on phage-bacteria infection networks. *Comput. Biol. Med.* <https://doi.org/10.1016/j.combiomed.2021.105186> (2022).
27. Zenil, H., Kiani, N. A. & Tegnér, J. A review of graph and network complexity from an algorithmic information perspective. *Entropy* **20**, 551 (2018).
28. Hyman, P. Phages for phage therapy: isolation, characterization, and host range breadth. *Pharmaceuticals* **12**, 35–23 (2019).
29. Chevallereau, A., Pons, B. J., van Houte, S. & Westra, E. R. Interactions between bacterial and phage communities in natural environments. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-021-00602-y> (2021).
30. Ford, B. E. *et al.* Frequency and fitness consequences of bacteriophage $\Phi 6$ host range mutations. *PLoS One* **9**, e113078 (2014).
31. Weitz, J. S. *et al.* Phage-bacteria infection networks. *Trends Microbiol.* **21**, 82–91 (2013).
32. Hays, S. G. & Seed, K. D. Dominant *Vibrio cholerae* phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. *eLife* **9**, e53200 (2020).
33. Pausch, P. *et al.* CRISPR-Cas Φ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
34. Abedon, S. T. Phage therapy dosing: The problem(s) with multiplicity of infection (MOI). *Bacteriophage* **6**, e1220348 (2016).
35. Haines, M. E. K. *et al.* Analysis of selection methods to develop novel phage therapy cocktails against antimicrobial resistant clinical isolates of bacteria. *Front. Microbiol.* **12**, 613529 (2021).
36. Chan, B. K., Abedon, S. T. & Loc-Carrillo, C. Phage cocktails and the future of phage therapy. *Fut. Microbiol.* **8**, 769–783 (2013).
37. Beckett, S. J. & Williams, H. T. P. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. *Interface Focus* **3**, 20130033–20130033 (2013).
38. Fernández, L., Gutiérrez, D., García, P. & Rodríguez, A. The perfect bacteriophage for therapeutic applications—a quick guide. *Antibiotics* **8**, 126 (2019).
39. Duffy, S., Turner, P. E. & Burch, C. L. Pleiotropic costs of niche expansion in the RNA bacteriophage $\phi 6$. *Genetics* **172**, 751–757 (2006).
40. Molina, F. *et al.* Diversity and local coadaptation of *E. coli* and coliphages from small ruminants. *Front. Microbiol.* **11**, 1–12 (2020).
41. Poullain, V., Gandon, S., Brockhurst, M. A., Buckling, A. & Hochberg, M. E. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. *Evolution* **62**, 1–11 (2008).
42. Fernández, L. *et al.* Low-level predation by lytic phage ϕ IPLA-RODI promotes biofilm formation and triggers the stringent response in *Staphylococcus aureus*. *Sci. Rep.* **7**, 40965 (2017).
43. Keen, E. C. Tradeoffs in bacteriophage life histories. *Bacteriophage* **4**, e28365–e28367 (2014).
44. Paepe, M. D. & Taddei, F. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol.* **4**, e193–e199 (2006).
45. Pilosof, S. *et al.* The network structure and eco-evolutionary dynamics of CRISPR-induced immune diversification. *Nat. Ecol. Evol.* **4**, 1650–1660 (2020).
46. Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 14 (2020).
47. Hasan, M. Z., Islam, S., Matsumoto, K. & Kawai, T. SARS-CoV-2 infection initiates interleukin-17-enriched transcriptional response in different cells from multiple organs. *Sci. Rep.* **11**, 16814 (2021).
48. Lamy-Besnier, Q. *et al.* Viral host range database, an online tool for recording, analyzing and disseminating virus-host interactions. *Bioinformatics* **17**, btob070. <https://doi.org/10.1093/bioinformatics/btob070> (2021).
49. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
50. Mangieri, N., Picozzi, C., Cocuzzi, R. & Foschino, R. Evaluation of a potential bacteriophage cocktail for the control of shiga-toxin producing *E. coli* in Food. *Front. Microbiol.* **11**, 1801 (2020).
51. Fong, *et al.* Diversity and host specificity revealed by biological characterization and whole genome sequencing of bacteriophages infecting *Salmonella enterica*. *Viruses* **11**, 854–918 (2019).

Acknowledgements

This research study was supported by funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 813439. We thank Encarna Ferrera for her excellent technical help.

Author contributions

F.M., P.G. and L.F. wrote the manuscript. F.M. conceived and performed the experiments. M.M.F. and M.A.V.R. developed and optimized the software. F.M. and M.M.F. performed the statistical analysis. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022