bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

# Leapfrog dynamics in phage-bacteria coevolution revealed by joint analysis of cross-infection phenotypes and whole genome sequencing

**Animesh Gupta[1+], Shengyun Peng[2+], Chung Yin Leung[3], Joshua M. Borin[4], Joshua S. Weitz[3,5]* and Justin R. Meyer[1]***

1 – Department of Physics, University of California San Diego, La Jolla, CA, USA

2 – Adobe Inc., 345 Park Avenue, San Jose, CA, USA

3 – School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

4 – Division of Biological Science, University of California San Diego, La Jolla, CA, USA

5 – School of Physics, Georgia Institute of Technology, Atlanta, GA, USA

[+]**Authors contributed equally**

[*]**To whom correspondence should be addressed.**
**E-mail:** jsweitz@gatech.edu and jrmeyer@ucsd.edu
phone: +1 (858) 822-3338
fax: +1 (858) 534-7108

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Abstract

Viruses and their hosts can undergo coevolutionary arms races where hosts evolve increased resistance and viruses evolve counter-resistance. Given these arms race dynamics (ARD), each species is predicted to evolve along a single trajectory as more recently evolved genotypes replace their predecessors. Here, by coupling phenotypic and genomic analyses of coevolving populations of bacteriophage λ and *Escherichia coli*, we find conflicting evidence for ARD. Virus-host infection phenotypes fit the ARD model, yet whole genome analyses did not. Rather than coevolution unfolding along a single trajectory, cryptic genetic variation emerges during initial virus-host coevolution. This variation is maintained across generations and eventually supplants dominant lineages. Our observations constitute a new type of 'leapfrog' coevolutionary dynamics (LFD), revealing weaknesses in the predictive power of standard coevolutionary models. The findings shed light on the mechanisms that structure coevolving ecological networks and reveal the limits of using phenotypic assays alone in characterizing coevolutionary dynamics.

## Introduction

Bacteria and their viruses (phages) are the two most abundant and genetically diverse groups of organisms on Earth[1-3]. Together they form ecological communities with complex networks of interactions whose structures have important implications that extend beyond the microbial world. For example, in oceans, phages may be responsible for 10-40% mortality of bacteria (with significant variation in lineage-specific mortality[4]). When bacterial cells are lysed by viruses, their biomass is diverted away from the rest of the food web, reducing productivity of macroscopic organisms and the full ecosystem[5-8]. Thus, if interactions within the network were to substantially change, such as the bacteria evolving resistance to the phage, then productivity may increase causing rippling effects throughput the marine ecosystem. Bacteria and phages are known to engage in arms races where bacteria evolve phage resistance[9], and phages evolve counter defenses[10]. This dynamic causes continual remodeling of the interaction network, which can impact ecosystem processes, the stability of ecological communities, and maintenance of microbial diversity[11-14]. There is a growing interest in characterizing the dynamics of phage-bacterial coevolution and to understand the molecular and ecological mechanisms that shape their coevolving networks[15].

A starting point to study phage-bacterial coevolution is to learn how their interactions change over time[16-19]. Models of coevolution tend to predict two types of dynamics[20,21]. The first is arms race dynamics (ARD) where bacteria evolve resistance to an increasing number of phages and phages counter by expanding their host-range. For the bacteria, this leads to an escalation where increasingly resistant bacteria replace their less-resistant predecessors. This causes rapid bacterial genomic divergence and leads to an imbalanced phylogenetic pattern with a single pronounced branch. Similarly, as the phage broadens its host range, the most evolved type will supplant its predecessors, resulting in the formation of a similarly imbalanced phylogeny. A second model is based on the evolution of specialized interactions, often described as *lock and key* interactions. In this model, as coevolution progresses, bacteria gain resistance to contemporary phages, but lose resistance to phages encountered in the past. Likewise, as phages evolve counter-defenses, they lose the ability to infect other host genotypes. Under this model, host genotypes rise and fall according to how abundant their corresponding parasite genotypes

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

are, while parasite genotypes track the abundance of their hosts creating a feedback loop and fluctuating selection dynamics (FSD)[22,23]. FSD produces negative frequency-dependent selection that promotes diversification and the formation of a balanced phylogeny with multiple branches. ARD and FSD represent two ends of a spectrum of possible coevolutionary dynamics and notably models have been created that span the space between the two end points[24].

One way to gain insight on whether phage and bacteria coevolve according to ARD or FSD is to quantify their interaction networks (phage-bacterial interaction networks; PBINs) and test for nonrandom nested and modular patterns[25]. Nestedness measures the extent to which interaction patterns form strict hierarchical subsets, analogous to nesting Russian dolls[26]. This pattern is produced by ARD since at each step the phage adds on to its existing host-range, expanding its range in a way that encapsulates its ancestors' ranges. Modularity arises in networks when interactions form into nonrandom clusters such that groups of phages and bacteria tend to interact more often within clusters. This pattern is consistent with FSD where interactions are highly specialized. The majority of PBINs are significantly nested supporting the prominence of ARD; however some PBINs are modular[25], and while rare, specialized interactions have been documented to evolve during phage-bacterial coevolution[27]. Surprisingly, nested patterns at short spatial scales can give way to modular patterns at large spatial scales[28], and ARD has been shown to give way to FSD during advanced stages of coevolution[28,29]. Together, this variation and scale-dependence provide a glimpse at how complicated phage-bacterial coevolution can be.

While the phenotypic predictions for ARD and FSD are often tested, assessments of the phylogenomic predictions are not as common (ARD: imbalanced, FSD: balanced), and we are unaware of an example where PBINs have been coupled with phylogenomic analyses. Given this, we decided to study a model phage-bacterial coevolutionary system; bacteriophage λ and its host, *Escherichia coli*. When these species are cultured under certain laboratory conditions, they rapidly coevolve with one another[30,31]. *E. coli* is known to evolve resistance through mutations in the regulatory gene *malT* that suppress expression of the host receptor, the outer-membrane protein LamB. λ counters this by evolving mutations in the binding domain of its host recognition protein J that allows it to use a new receptor, OmpF. *E. coli* then evolves additional mutations in OmpF or in an inner-membrane protein complex, ManXYZ, that transports λ DNA into the cytoplasm[32,33]. While much is already known about the molecular details of their coevolution, PBIN or phylogenomic analyses have yet to be performed on this pair.

For this study, we revived cryopreserved samples that were isolated from a coevolution experiment previously reported[30]. We focused our analyses on a single replicate; the first experimental community in which λ evolved to use OmpF. We isolated a total of 50 bacteria and 44 phages spread across multiple time points. Next, we constructed a PBIN of all combinations of pairwise phage and bacteria interactions and used multiple analyses to characterize their coevolution based on phenotypes. All three phenotypic-based analyses suggested that viruses and microbes engage in ARD. Lastly, we sequenced the full genomes of each isolate and reconstructed the isolates' phylogenetic relationships. The genome sequences revealed a phylogenetic pattern that was inconsistent with the ARD model. Our study demonstrates that phenotypic analyses are not sufficient to test hypothesis on coevolutionary dynamics and reveals a new type of coevolutionary dynamic we refer to as Leapfrog Dynamics (LFD).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Results

*Phage-bacterial infection network*

The pairwise interaction study revealed an incredible number of λ genotypes with phenotypically distinct host-ranges, and *E. coli* genotypes that vary in resistance (Fig. 1a). In line with the ARD model, we found that the interactions were highly nested (Fig. 1b) and had a low level of modularity (Fig. S1). Also in line with ARD, *E. coli* evolved increasing resistance (Fig. 1c, $R^2_{adj} = 0.5051$, $F_{1,48}$ = 51.01, $P$ = 4.453e-09 for linear model: response ~ time), and λ gained increasing host range and infectivity (Fig. 1d, $R^2_{adj} = 0.8131$, $F_{1,42}$ = 188.1, $P$ = 2.2e-16) through time.

To further test between indicators of ARD and FSD, we performed a time-shift analysis using EOP values to determine how phages' infectivity varies when presented with past, contemporary, or future bacteria (Fig. 2a)[46]. ARD predicts that phages will be able to infect past and contemporary, but not future hosts, while FSD predicts that phages will be best at infecting contemporary hosts. The time-shift analysis was conducted for each λ isolate by calculating its mean EOP value for all 10 bacterial isolates on each day (Fig. 2b). This analysis was repeated for the bacteria using the same EOP data but by calculating levels of resistance to λ isolates from different time points (Fig. 2c). In line with ARD, λ isolates from days 22 and 28 had higher infectivity on past hosts than contemporary or future hosts (Fig. 2b). The analysis for day 15 phages was inconclusive because the EOP values across time were not statistically significant. The pattern for bacteria was also in line with ARD: isolates from days 8, 15 and 22, had lower resistance (higher EOP) for phage samples from the future versus the phage isolated from the same time or in the past (Fig. 2c). A full time-shift analysis could not be conducted for isolates from days 28 and 37 since the phage went extinct between days 28 and 37, however isolates from these time points were the most resistant.

Note, all mean EOP values were zero for day 8 phages because all isolated hosts were resistant to all day 8 phages (Fig. 1a and 2a). Still, λ did not go extinct in the coevolution experiment because of a phenomenon known as 'leaky-resistance'[47]. Leaky resistance denotes a phenomenon by which a small number of resistant hosts revert to sensitive, thereby sustaining phage in the population until they are able to evolve to gain access to a new receptor, e.g., OmpF in the case of phage λ[30].

*Genome sequencing*

The genome sequencing revealed 22 unique *E. coli* genomes and 34 unique phage genomes. Among the *E. coli* strains, we found a total of 18 unique mutations; 6 missense mutations, 1 nonsense mutation, 1 intergenic point mutation, 7 deletions and 3 duplications (Fig. 3a). The most abundant mutation that occurred in 38 out of 50 host genomes was a frameshift mutation caused by a 25-base duplication in the *malT* gene, in line with the original study[30]. Disruptions in *malT* interferes with the expression of LamB protein which ancestral λ needs to bind to *E. coli* cells. We also observed one isolate with a *lamB* mutation (1-base deletion) *in lieu* of the typical *malT* mutation. The most resistant *E. coli* strains on day 37 have a number of mutations that are expected to confer resistance; a *malT* deletion, a nonsynonymous change in *ompF*, and a deletion in *manZ*[30].

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

One of the most common bacterial mutations observed in our dataset was a 777 bp deletion caused by the excision of an IS element. Despite occurring in 25 genomes, we had not observed the deletion in any previous studies and the deletion did not affect any genes known to protect against λ infection; *insB-22* which encodes for IS1 protein InsB, *insA-22* which encodes for IS1 protein InsA, and *ECB_02825* which encodes for a pyrophosphorylase[48]. Similar IS mutations are known to occur at high rates[49] and the deletion was only observed in genomes with a *malT* mutation, so it might have just been a neutral genomic hitchhiker. However, a follow-up experiment that we report in the online materials revealed that the deletion enhances resistance when it cooccurs with the *malT* mutation, suggesting the deletion was adaptive and epistatic with *malT* mutations (Fig. S2, Supplementary Information). We are unsure of how the deletion confers resistance; however, our results suggest that there are more genes in the λ-*E. coli* interactome than previously known and that experimental evolution can help reveal them.

In the λ genomes, we found a total of 176 unique mutations; 53 nonsynonymous point mutations, 87 synonymous point mutations, 2 insertions, 3 deletions and 31 intergenic mutations (Fig. 3b). While this level of molecular evolution may seem surprising for such a short-term experiment, similar levels have been observed for other phages evolving in the laboratory[50]. 116 of these mutations were in the host-recognition gene *J*. The J protein is positioned at the end of the phage's tail, and initiates infection by binding to *E. coli*'s LamB protein. Some of these J mutations have been shown to increase adsorption rates to LamB and allow λ to exploit a novel receptor, OmpF[51-53]. Interestingly, the extensive sequencing effort performed here revealed a mutation in another tail fiber protein called H (C→T substitution at nucleotide position 11,451). This mutation rises late in the experiment and likely plays a role in expanding λ's host range. *H* is called the tape-measure gene because it helps determine the length of λ's tail, and mutations in this gene have been shown to increase λ's host-range in other experiments[54].

*Phylogenomic reconstruction of coevolution*

Even though multiple analysis of the phenotypic data supported the ARD model for coevolution, the pattern produced by the phylogenies are in line with predictions of FSD (Fig. 4). The phylogenies of both *E. coli* and λ show that multiple lineages coexist for weeks, rather than a single dominant branch. A second unexpected observation was that the bacteria that had acquired the highest level of resistance at the end of the experiment on day 37 was not most closely related to isolates at previous time points (e.g., days 28, 22, or 15), instead it was most closely related to a common ancestor of isolates identified at the early stages of the experiment (on day 8). This finding suggests that a rare lineage leapt ahead of the dominant lineage, a process we term leapfrog dynamics (LFD). Similarly, for λ, we found that that the clade dominant at the final timepoint with the broadest host-range was more closely related to wildtype λ than the clade dominant at preceding timepoints. For both species, the clades that win out later in the arms race appeared to exist as cryptic subpopulations early in the coevolutionary experiment.

*Whole population sequencing at an early stage of coevolution*

To test the key prediction of LFD that cryptic lineages coexist with dominant lineages and can supply the genetic reservoir used for later stages of coevolution, we sequenced full populations of *E. coli* and λ from day 8 and searched for mutations that rose to prominence at the end of the study. Given the low mutation rates of *E. coli* ($8.9\times10^{-11}$ base$^{-1}$ replication$^{-1}$[55]) and λ ($7.7\times10^{-8}$ base$^{-1}$ replication$^{-1}$[56]), and their average population sizes ($10^9$ and $10^7$, respectively[30]) it is

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

unlikely that any single mutation would evolve twice, so the presence of these key mutations in an early sample would indicate that the lineages had established early in the coevolutionary LFD.

For *E. coli*, we were specifically searching for two mutations: a Δ16 bp deletion at position 1,882,915 in *manZ* and a non-synonymous mutation at 1,003,271 in *ompF*. These mutations are present in most of the day 37 isolates and are thought to confer resistance. The Δ16 bp deletion in *manZ* was detected, but not the *ompF* mutation (Table S1). We also found a 141 bp deletion in *malT* that cooccurs in the day 37 genomes with the *manZ* mutation. The *malT* deletion was at the same frequency, suggesting that these mutations were indeed linked and that they evolved sometime before day 8.

For λ, we focused on the mutation in *H* that rises to dominance between days 22 and 28. Indeed, this specific H mutation was present at day 8 (Table S2). Unlike *E. coli*, we did not find any other mutations present in the day 28 isolates, suggesting that the H mutation was the first adaptation to occur in this lineage.

Besides revealing the eventual winning lineages of the arms race, by sequencing populations we also discovered much more genetic diversity than through isolate sequencing. We found 52 unique mutations in *E. coli* and 38 mutations in λ from full population sequencing compared to 7 and 30 through isolate sampling, respectively. This shows that there is a significant amount of genetic diversity generated at the earliest phases of the arms race that can become the grist for subsequent adaptation as the host (for phage) or phage (for hosts) change as a result of coevolution.

**Discussion**

Through large scale phenotypic assays and whole genome sequencing, we were able to test existing paradigmatic models of coevolution and learn that each were inadequate to explain λ and *E. coli*'s coevolutionary dynamics. Three complimentary phenotypic analyses in Fig. 1 and Fig. 2 suggested that coevolution between λ and *E. coli* followed arms race dynamics (ARD). However, the phylogenetic pattern revealed by whole genome sequencing was in line with fluctuating selection dynamics (FSD) (Fig. 3). These observations lead us to develop a new model to characterize λ and *E. coli*'s coevolution: which we term leapfrog dynamics (LFD). In this model, selection operates similarly to the ARD model, where parasite genotypes with ever-expanding host-ranges are selected and hosts with ever-increasing resistance are favored. However, the difference is that in the LFD model there is a genetically diverse pool of hosts and parasites that evolve early, and on occasion, rare individuals are drawn from this pool with advantageous phenotypes and replace the dominant strains.

ARD models fall short in making accurate predictions for the phylogenies likely because of simplifying assumptions about the genetics of host-range expansion and resistance. Evolutionary models tend to assume that mutations have small additive effects on phenotypes (as highlighted in[20]). Applied to ARD, this would mean that the phage with the broadest host range is likely to expand its host-range faster than lagging genotypes, and for bacteria, the strain with the greatest resistance is most likely to acquire the next level of resistance and outcompete other strains. This genetic architecture favors the evolution of directed phylogenies with one dominant branch. Instead, λ J mutations are known to possess high-order epistasis and are nonadditive[37,53]. This

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

may allow rare lineages with an exceedingly beneficial combination of mutations to leap ahead. Non-additivity was also discovered in the bacteria in this study with respect to the interactions between *malT*⁻ and the Δ777 mutations. A second problem with evolutionary models is the assumption of small effect-size mutations since mutations like the *malT* mutation can cause nearly complete resistance to some λs (Fig. 1a). Acquiring such large effect mutations could help lineages leap ahead.

One question left unanswered by this study is how multiple lineages persisted in this population for long durations. We hypothesized that trade-offs between host-range and other viral traits for phage, and resistance and competitive fitness for bacteria, could explain the evolution of genomic diversity. Trade-offs between host-range and λ stability were previously observed[52]; however, we were unable to detect trade-offs in *E. coli* for phage resistance. A recent theoretical study showed that that there are many other mechanisms besides trade-offs that can select for genetic diversity during ARD, providing new hypothetical mechanism to pursue in the future[57].

Our findings also have important implications for understanding the mechanisms that structure nested PBINs. The nested pattern is ubiquitous in PBINs[25], as well as many other ecological networks[58-60], so it is important to understand the processes that produce this ordering. One hypothesis is that the structure is determined by the genetics of the interactions (gene-for-gene, see[20]). The second is ecological; nestedness emerges because of how the selection steadily shifts during an arms race to promote incremental increases in host-range and resistance. The conventional wisdom for which process controls the arms races is the underlying genetics. This stems from a pervasive idea that genetic mutation and evolution happen more slowly than changes in ecology, so coevolutionary systems must be constrained by their access to genetic variation. The coevolution experiment studied here was initiated with small population size of isogenic stocks of λ and *E. coli.* This should have favored genetic control because all of the variation had to evolve *de novo*. However, a tremendous amount of genetic variation was generated at the earliest phase of the arms race that was not deployed for many generations, suggesting that the ecology controlled the dynamics, not the availability of genetic variation.

Lastly, our results provide a cautionary tale for over-interpreting phenotypic data based on the phage-bacteria infection networks. Our initial prediction before performing genomic analysis was that λ-*E. coli* coevolution under these laboratory conditions would fit the ARD model. However, this proved to be erroneous due to the large amount of cryptic genetic variation and its role in driving late stages of coevolution. Interestingly, this is the second time that we learnt about the limitations of PBIN data for making prediction. Based on the matrix in Fig. 1a one would predict that the phage should go extinct on day 8, however leaky resistance fueled by host cells too rare to be sampled allowed λ to persist. Together, the LFD and leaky-resistance results show that host-phage dynamics revealed from PBINs alone miss the rich dynamics occurring at lower frequencies in the population and offer insufficient information to predict dynamics.

In studying this diversity with both phenotypic and genomic approaches we revealed that typical models of coevolution are insufficient to explain the complex dynamics in this community and that phenotypic assays alone fall short in characterizing coevolutionary dynamics. We also showed that highly organized ecological pattern like nestedness can emerge from the extremely

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

disordered population genomic pattern underlying them, demonstrating the power of selection in producing nonrandom ecological patterns.

## Methods

*Details on the initial coevolution experiment previously published*

Meyer *et al.*[30] performed the original coevolution experiment with *Escherichia coli* B strain REL606 and a lytic bacteriophage λ strain, cI26. This λ strain was chosen because it cannot enter lysogeny, a life cycle phase where λ confers immunity to additional λ infections. By choosing a lytic strain, we forced the bacteria to evolve genetic resistance. *E. coli* and λ were cocultured in a carbon-limited minimal glucose media at 37 °C for 37 days[30]. At the end of each day, 1% of the community was transferred to new flasks with fresh media, and, weekly, 2 ml of the community was preserved by adding ~15% of glycerol and freezing the mixture at -80 °C.

*Isolation of host and phage clones*

We randomly isolated ten host and eleven phage individuals from different timepoints from the cryopreserved samples. In total, 50 strains of *E.* coli and 44 strains of λ were isolated from days 8, 15, 22, 28 and 37 of the experiment (no phage were detected on day 37). Bacteria were isolated by streaking onto Luria-Bertani (LB) agar plates[34] and randomly picking 10 colonies. These colonies were re-streaked three times to remove phage particles and grown overnight in liquid LB to create stocks. Phages were isolated by plating an appropriate dilution of the population onto overlay plates[35] with the sensitive ancestral bacteria, REL606, and randomly picking 11 plaques. These plaques were grown overnight with REL606 in LBM9 medium and stocks were created using chloroform isolation technique[30]. All phage and bacteria stocks were stored at -80 °C with the addition of 15% *v/v* glycerol.

*Pairwise infection assays and efficiency of plating (EOP)*

We performed quantitative, pairwise infection assays for all combinations of host strains and phage strains that were isolated. Specifically, seven serial 1/10$^{th}$ dilutions were made of each phage isolate. 2 µl of each dilution plus undiluted phage stock was spotted on top of different host strain lawns including ancestor REL606. Thus, a total of 8*44 spots of phage were plated on 51 different types of bacterial lawns, leading to a total of 17,952 pairwise infections. This allowed us to measure how well each phage isolate infects every host isolate. We quantified phage infectivity by calculating efficiency of plating (EOP), defined as the ratio of density of phage isolate calculated on a coevolved isolate to the density of phage calculated on the REL606 ancestor.

*Analysis of PBIN nestedness and modularity*

*BiMat* was used to assess the nestedness and modularity of the PBIN[36]. The raw EOP value matrix was binarized into 0 for EOP = 0 and 1 for EOP > 0, and then *BiMat* was run with default settings. Here we report the statistics for a conservative version of the analysis where the rows and columns that contained all zeros were removed from the matrix to reduce any bias these entries cause in establishing significant nested patterns.

*Resistance and infectivity calculations*

For a total number of $n$ host samples and $m$ phage samples, we denote the EOP value for the $i$th host sample against $j$th phage sample as $e_{ij}$ where $i \in [1, n]$ and $j \in [1, m]$; $n = 50$ and $m =$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

44. We denote the five checkpoint days of day 8, 15, 22, 28 and 37 for host by $k$, where $k = 1,2,3,4,5$, and the four checkpoint days of day 8, 15, 22 and 28 for phage by $l$ where $l = 1,2,3,4$. Host resistance for a host sample $i$ is calculated as

$$r_i = \sum_{j=1}^{m} \mathbf{1}_{\{e_{ij}=0\}}$$

which measures the number of phage strains that the host is resistant to. The host range of a phage sample $j$ is calculated as

$$h_j = \sum_{i=1}^{n} \mathbf{1}_{\{e_{ij}>0\}}$$

which measures the number of host strains that the phage can successfully infect. The resistance percentage of host for each day is calculated as

$$RP_k = \frac{\sum_{i \in A_k} r_i}{m \times |A_k|}$$

where $A_k$ denotes the range of the host sample that belongs to the $k$th day and $|A_k|$ denotes the cardinality of the set $A_k$, i.e. the number of host samples at the $k$th day. Likewise, the host-range percentage of phage for each day is calculated as

$$HP_l = \frac{\sum_{j \in B_l} h_j}{n \times |B_l|}$$

where $B_l$ denotes the range of the phage sample that belongs to the $l$th checkpoint and $|B_l|$ denotes the cardinality of the set $B_l$, i.e. the number of phage samples at the $l$th checkpoint.

*Genomic DNA preparation for sequencing*

λ genome extraction for whole genome sequencing were previously reported in[37]. To summarize, λ particles were concentrated using PEG precipitation, the phage were treated with DNase I to remove free-floating DNA not protected by phage capsids, the DNase is denatured with heat, which also releases capsid-enclosed phage DNA. The DNA was extracted using Invitrogen's PureLink kit. *E. coli* genomic DNA was extracted and purified from a 1 ml sample of culture by using PureLink kit.

Genomic DNA was further processed by fragmenting the DNA and attaching adapters and barcodes using a method outlined in[38]. Sequencing was done at UC San Diego IGM Genomics using paired-end Illumina HiSeq 4000 platform.

*Construction of mutation profile tables*

After collecting the raw sequencing reads, we removed the adapters using cutadapt[39] and performed quality control (QC) for each isolated strain using FastQC[40]. The QC filtered sequencing reads were then analyzed using the *breseq* (v0.32.1)[41]. We ran *breseq* in the consensus mode with default parameters except for the consensus-frequency-cutoff, which was set to 0.5.

*Phylogenomics*

Due to the prevalence of large insertions and deletions in the host genomes, conventional nucleotide substitution models were not suitable for estimating the host phylogenetic tree. However, such models were suitable for estimating the maximum-likelihood phylogenetic tree for phage genomes. As a result, two different approaches were taken to reconstruct the evolutionary trajectories of the host and virus.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To construct the phage phylogeny, multiple sequence alignments were performed for all recovered genomes and the ancestral genome using *mafft* (v7.305b)[42] with default settings except that 'retree' was set to 2 and 'maxiterate' was set to 1,000. A maximum likelihood tree was constructed using *raxml-ng*[43]. Finally, the *TreeTime*[44] program was used to generate the phylogenetic tree.

To reconstruct the host's evolution, we constructed a Hamming distance matrix to calculate genetic distances between different host isolates. Neighbor-joining (NJ) trees were then built based on the hamming distance matrix using *T-REX*[45]. Finally, the *TreeTime* program was used to build the host phylodynamic tree.

*Whole genome whole population sequencing*

We sequenced the full population of phage and bacteria from Day 8 of the experiment to 3,726-fold and 142-fold coverage, respectively. This allowed us to uncover alleles in the bacterial and phage populations that existed at lower frequencies than we could detect by isolating individuals. To do this, λ and *E. coli* populations were revived by growing 120 μl of frozen stock of the whole community in the laboratory conditions from the original experiment[30]. Phage and bacteria were then separated, and their genomic DNA was extracted in the same manner as described before for clonal stocks. Genomic libraries were prepared using NexteraXT kit at UC San Diego IGM Genomics. IGM also sequenced the samples using 75 base single reads on the Illumina HiSeq 4000 platform. *breseq* v0.32.1 was used to analyze whole population sequencing data of Day 8. We ran *breseq* in polymorphism mode with default settings to construct the mutation profile tables.

**Acknowledgements**

We would like to thank Morgan Mouchka for her help with sequencing genomes of the whole community.

**Author Contributions:**

JRM isolated the viruses, isolated the bacteria, and measured the phage-bacteria infection matrix. AG and CYL sequenced whole genomes. AG and SP performed analyses, AG, CYL and JMB performed experiments, AG, SP, JSW and JRM wrote the manuscript, and JRM and JSW developed the project and oversaw the research. All coauthors helped edit the manuscript.

**Competing interests**

The authors declare no competing interests.

**Data accessibility statement:**

All sequencing data will be deposited in the NCBI Sequence Read Archive. Code will be made available at GitHub and data that support the findings of this study will be deposited at Dryad. Biological material is available from J.R.M. under a material transfer agreement with UC San Diego (http://blink.ucsd.edu/research/conducting-research/mta/index.html).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

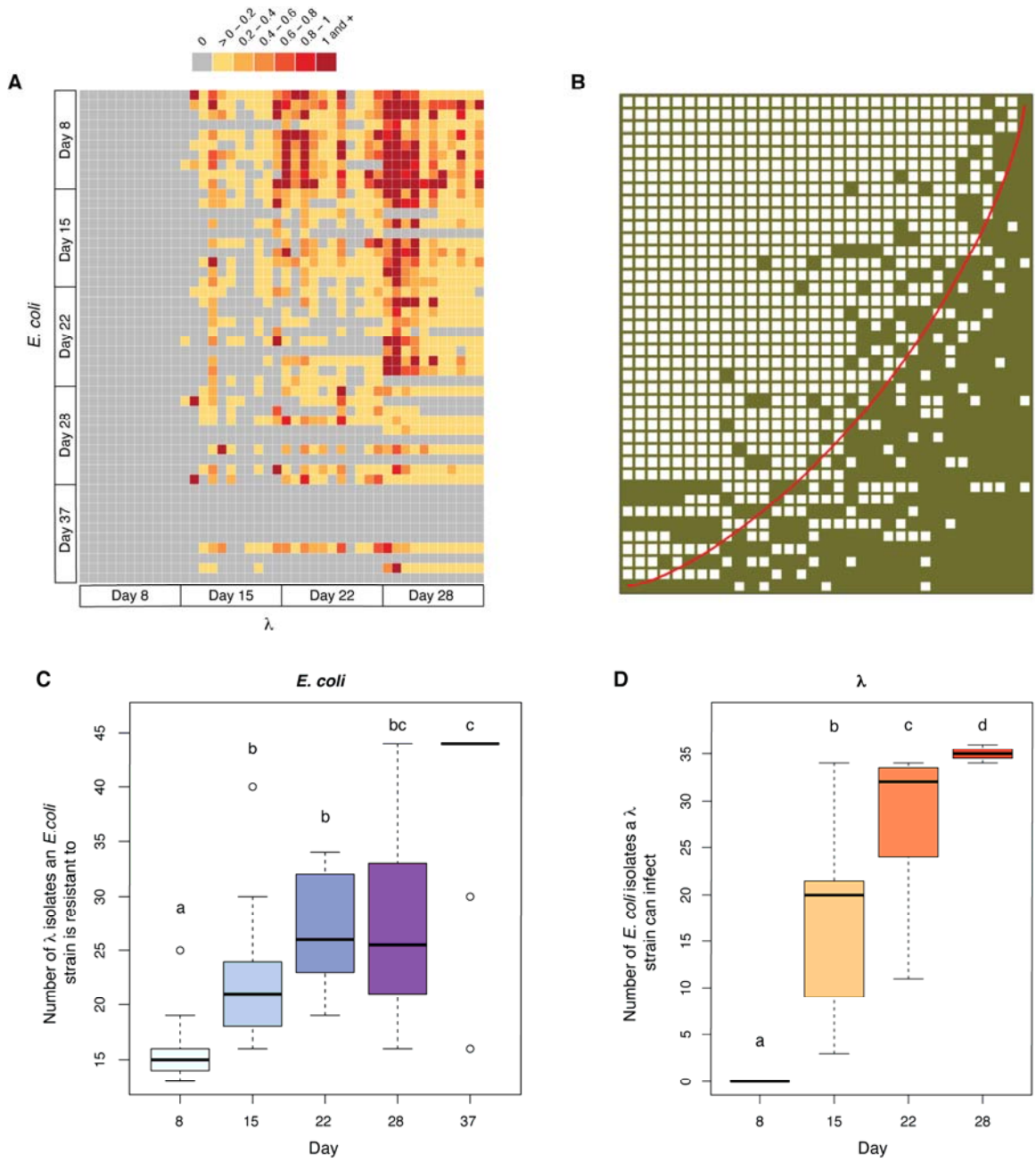


**Fig. 1 Host resistance and phage infectivity measured by pairwise plaque assays**. (**A)** Phage-bacteria infection network where the color of each cell is determined by the EOP values obtained for that host-phage interaction pair; grey cells represent no infection by λ on the given *E. coli* strain, yellow represents low infectivity and red represents high infectivity. (**B)** The original network in a) but reassembled to maximize nestedness using the software BiMat. Filled squares indicate a combination of host and phage that result in successful interactions (EOP > 0), and the red line highlights the isocline using the NTC algorithm. The nestedness value of the network based on the NODF algorithm was significantly greater than the null expectation when constraining the fill of the bipartite network (measured value of nestedness 0.839 vs. null value of nestedness 0.638 ± 0.011 based on 200 trials). **(C)** Boxplots showing the total number of λ

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

isolates from all days that *E. coli* genotypes are resistant to across different sampling days. **(D)** Boxplots showing the total number of *E. coli* isolates from all days that λ genotypes can infect across different sampling days. Lowercase letters in c) and d) denote significant difference between different days via Tukey's honest significance test: c) ANOVA: $F_{4,45}$ = 13.3, $P$ = 3.11e-07), d) ANOVA: $F_{3,40}$ = 67.05, $P$ = 1.17e-15. A simple linear regression model with time as the predictor variable was also used to test if *E. coli* evolved increasing resistance in c) and λ evolved increasing host range in d) (statistics in the main text).

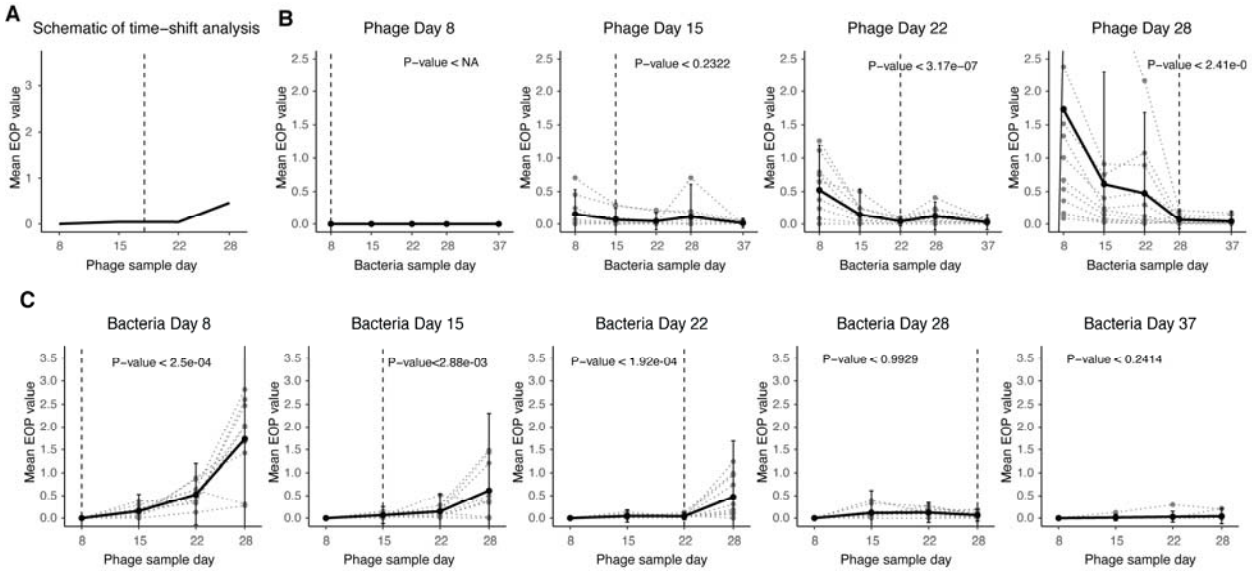


**Fig. 2 Time-shift analysis results from different checkpoints**. (**A)** Schematic for the time-shift analysis that compares the mean EOP from hosts or phages interacting with their counterparts from the past, contemporary and the future. **(B)** Time-shift results from phage checkpoints day 8, 15, 22 and 28 respectively. The gray dotted line shows the time-shift curve for each individual phage and the black line shows the average. The vertical dashed line represents the phage sample day. The *P*-values shown here are the maximum *P*-value from one-sided paired *t*-tests comparing the initial checkpoints with each of the later checkpoints. **(C)** Time-shift results from host checkpoints day 8, 15, 22, 28 and 37, respectively. The gray dotted line shows the time-shift curve for each individual host and the black line shows the average. The vertical dashed line represents the host sample day. The *P*-values shown here are the maximum *P*-value from one-sided paired *t*-tests comparing the final checkpoints with each of the previous checkpoints.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

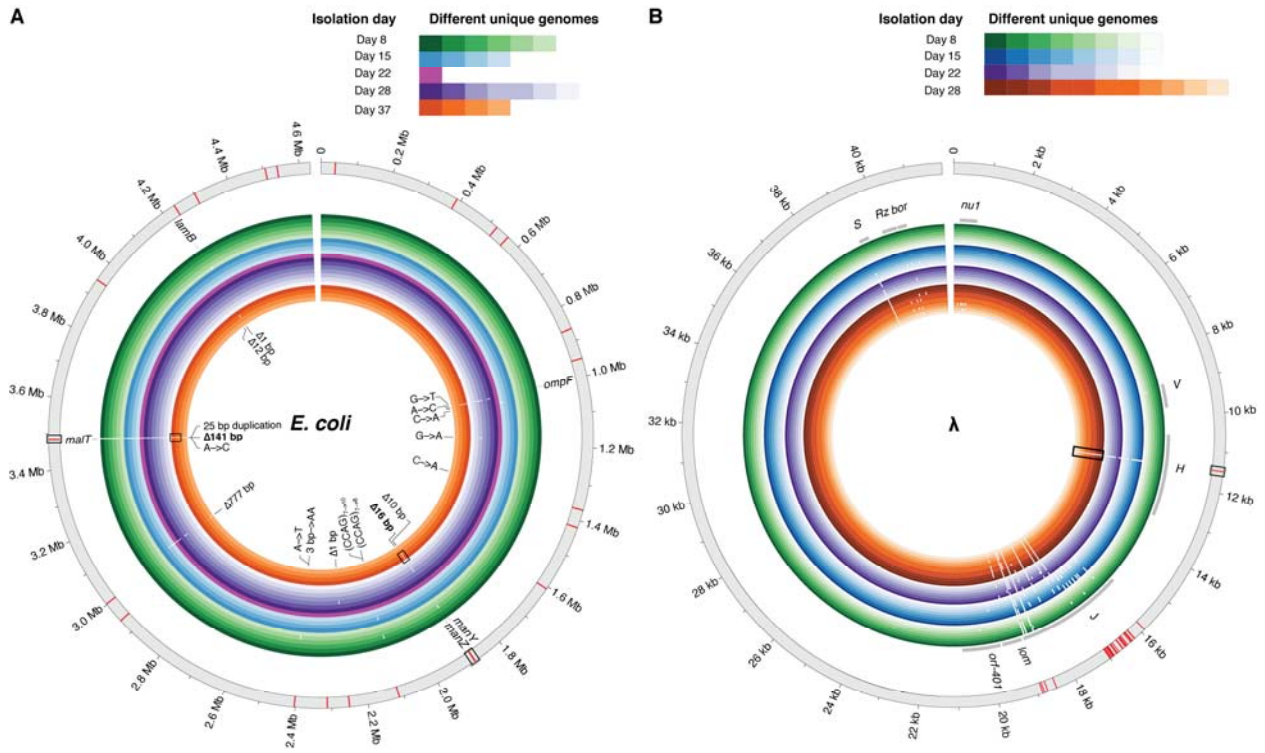


**Fig. 3 Genomic diversity in clones isolated from different days and full population sequencing for (A) *E. coli* and (B) λ.** The outermost gray ring represents the reference genome with red bars indicating the placement of mutations uncovered by the whole-population sequencing at day 8. The inner colored rings represent the isolates sequenced from different time points (outer rings are genomes isolated from earlier time points). Shades within each color depict unique genomes sequenced from each time point. White gaps in the genomic rings indicate the location of mutations. All mutations found in clonal isolates have been labeled for *E. coli* in A); however, due to the large number of mutations in λ, only the gene names that harbor mutations have been identified (gray bars). The mutations that become dominant at later stages of coevolution and were also found in day 8 population sequencing have been highlighted with rectangular boxes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

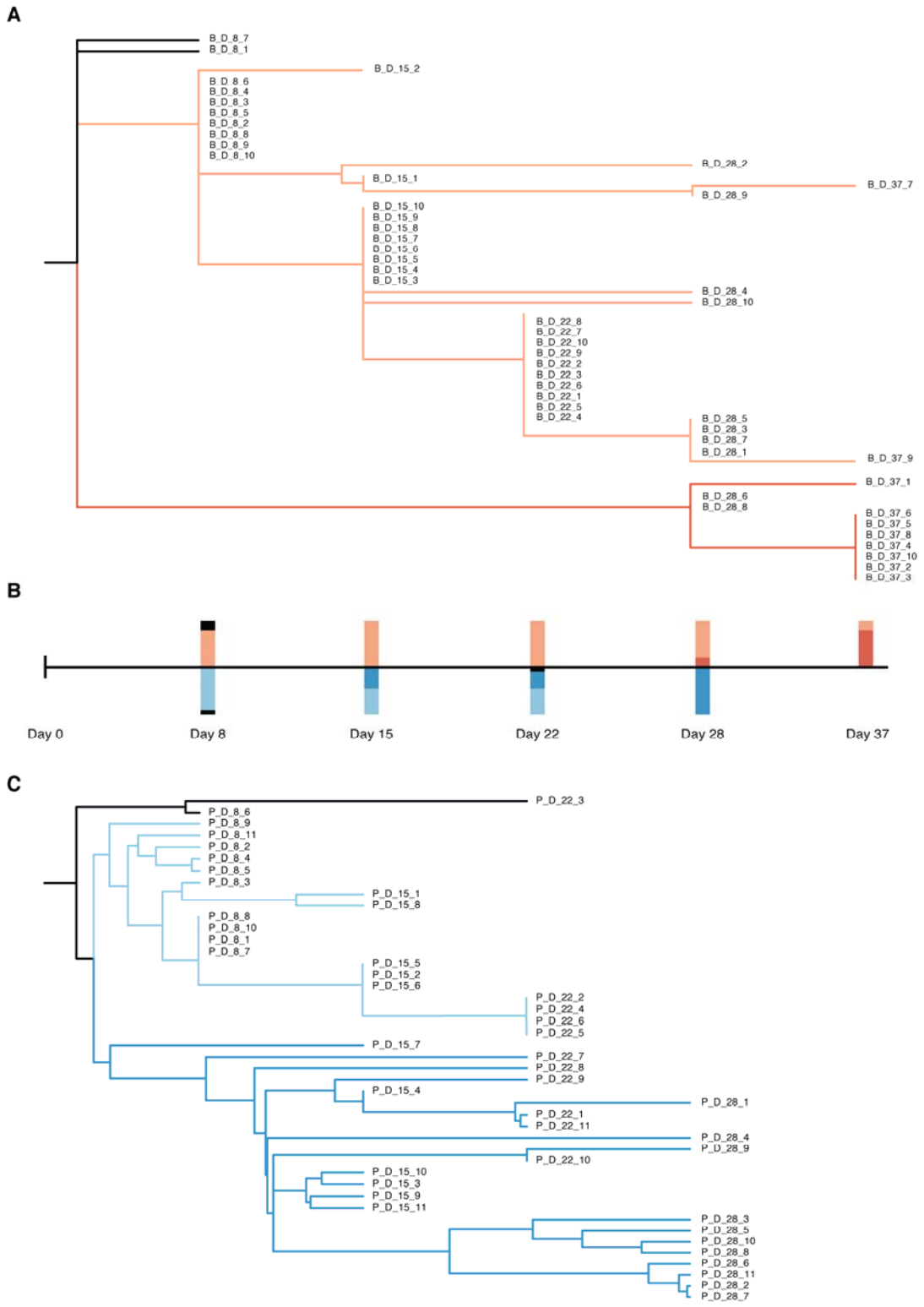


**Fig. 4 Reconstructed phylodynamic trees of the host and phage**. **(A)** The host phylodynamic tree based on host mutation profiles. All completely-resistant host strains are located on the red branch. Bars above the time scale in (**B**) represents the proportion of host strains from each colored branch across different checkpoints. **(C)** The phage phylodynamic tree based on the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

phage mutation profiles. All day 28 phage strains are located on the dark blue branch. Bars below the time scale in (**B**) represents the proportion of phage strains from each colored branch across different checkpoints.

## References


1 Torsvik, V., Øvreås, L. & Thingstad, T. F. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**, 1064-1066, doi:10.1126/science.1071698 (2002).

2 Thomas, R., Berdjeb, L., Sime-Ngando, T. & Jacquet, S. Viral abundance, production, decay rates and life strategies (lysogeny versus lysis) in Lake Bourget (France). *Environ Microbiol* **13**, 616-630, doi:10.1111/j.1462-2920.2010.02364.x (2011).

3 Clokie, M. R. J., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45, doi:10.4161/bact.1.1.14942 (2011).

4 Mruwat, N. *et al.* A single-cell polony method reveals low levels of infected Prochlorococcus in oligotrophic waters despite high cyanophage abundances. *The ISME Journal*, doi:10.1038/s41396-020-00752-6 (2020).

5 Suttle, C. A. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801-812, doi:10.1038/nrmicro1750 (2007).

6 Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal* **10**, 437-449, doi:10.1038/ismej.2015.125 (2016).

7 Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541-548, doi:10.1038/21119 (1999).

8 Weitz, J. S. & Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* **4**, 17, doi:10.3410/b4-17 (2012).

9 Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**, 317-327, doi:10.1038/nrmicro2315 (2010).

10 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336, doi:10.1038/s41586-019-1894-8 (2020).

11 Buckling, A. & Rainey, P. B. The role of parasites in sympatric and allopatric host diversification. *Nature* **420**, 496-499, doi:10.1038/nature01164 (2002).

12 Stern, A. & Sorek, R. The phage-host arms race: Shaping the evolution of microbes. *BioEssays* **33**, 43-51, doi:10.1002/bies.201000071 (2011).

13 Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**, 828-836, doi:10.1038/nrmicro2235 (2009).

14 Bohannan, B. J. M. & Lenski, R. E. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters* **3**, 362-377, doi:10.1046/j.1461-0248.2000.00161.x (2000).

15 Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS microbiology reviews* **38**, 916-931, doi:10.1111/1574-6976.12072 (2014).

16 Valverde, S., Elena, S. F. & Solé, R. Spatially induced nestedness in a neutral model of phage-bacteria networks. *Virus Evol* **3**, vex021, doi:10.1093/ve/vex021 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.


17 Childs, L. M., Held, N. L., Young, M. J., Whitaker, R. J. & Weitz, J. S. Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* **66**, 2015-2029, doi:10.1111/j.1558-5646.2012.01595.x (2012).

18 Weinberger, A. D. *et al.* Persisting Viral Sequences Shape Microbial CRISPR-based Immunity. *PLOS Computational Biology* **8**, e1002475, doi:10.1371/journal.pcbi.1002475 (2012).

19 Weitz, J. S., Hartman, H. & Levin, S. A. Coevolutionary arms races between bacteria and bacteriophage. *Proc Natl Acad Sci U S A* **102**, 9535-9540, doi:10.1073/pnas.0504062102 (2005).

20 Weitz, J. S. *et al.* Phage-bacteria infection networks. *Trends Microbiol* **21**, 82-91, doi:10.1016/j.tim.2012.11.003 (2013).

21 Agrawal, A. & Lively, C. M. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* **4**, 79-90 (2002).

22 Sasaki, A. Host-parasite coevolution in a multilocus gene-for-gene system. *Proc Biol Sci* **267**, 2183-2188, doi:10.1098/rspb.2000.1267 (2000).

23 Gandon, S., Buckling, A., Decaestecker, E. & Day, T. Host-parasite coevolution and patterns of adaptation across time and space. *J Evol Biol* **21**, 1861-1866, doi:10.1111/j.1420-9101.2008.01598.x (2008).

24 Leung, C. Y. & Weitz, J. S. Conflicting attachment and the growth of bipartite networks. *Physical Review E* **93**, 032303, doi:10.1103/PhysRevE.93.032303 (2016).

25 Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host-phage interactions. *Proc Natl Acad Sci U S A* **108**, E288-297, doi:10.1073/pnas.1101595108 (2011).

26 Almeida-Neto, M., Guimarães, P., Guimarães Jr, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227-1239, doi:10.1111/j.0030-1299.2008.16644.x (2008).

27 Gómez, P. & Buckling, A. Bacteria-phage antagonistic coevolution in soil. *Science* **332**, 106-109, doi:10.1126/science.1198767 (2011).

28 Flores, C. O., Valverde, S. & Weitz, J. S. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal* **7**, 520-532, doi:10.1038/ismej.2012.135 (2013).

29 Hall, A. R., Scanlan, P. D., Morgan, A. D. & Buckling, A. Host-parasite coevolutionary arms races give way to fluctuating selection. *Ecol Lett* **14**, 635-642, doi:10.1111/j.1461-0248.2011.01624.x (2011).

30 Meyer, J. R. *et al.* Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432, doi:10.1126/science.1214449 (2012).

31 Meyer, J. R. & Lenski, R. E. in *Evolution in Action: Past, Present and Future: A Festschrift in Honor of Erik D. Goodman* (eds Wolfgang Banzhaf *et al.*) 273-288 (Springer International Publishing, 2020).

32 Erni, B., Zanolari, B. & Kocher, H. P. The mannose permease of Escherichia coli consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage lambda DNA. *Journal of Biological Chemistry* **262**, 5238-5247 (1987).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.


33 Esquinas-Rychen, M. & Erni, B. Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenolpyruvate : carbohydrate phosphotransferase system (PTS). *Journal of Molecular Microbiology and Biotechnology* **3**, 361-370 (2001).

34 Sambrook, J. & Russell, D. W. *Molecular cloning : a laboratory manual*. 3rd edn, (Cold Spring Harbor Laboratory Press, 2001).

35 Adams, M. H. *Bacteriophages*. (Interscience Publishers, Inc., 1959).

36 Flores, C. O., Poisot, T., Valverde, S. & Weitz, J. S. BiMat: a MATLAB package to facilitate the analysis of bipartite networks. *Methods in Ecology and Evolution* **7**, 127-132 (2016).

37 Meyer, J. R. *et al.* Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* **354**, 1301-1304, doi:10.1126/science.aai8446 (2016).

38 Baym, M. *et al.* Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes (vol 10, e0128036, 2015). *Plos One* **10**, 1, doi:10.1371/journal.pone.0131262 (2015).

39 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, pp. 10-12 (2011).

40 Andrews, S. (2010).

41 Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* **1151**, 165-188, doi:10.1007/978-1-4939-0554-6_12 (2014).

42 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).

43 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

44 Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* **4**, vex042, doi:10.1093/ve/vex042 (2018).

45 Makarenkov, V. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics* **17**, 664-668 (2001).

46 Gaba, S. & Ebert, D. Time-shift experiments as a tool to study antagonistic coevolution. *Trends Ecol Evol* **24**, 226-232, doi:10.1016/j.tree.2008.11.005 (2009).

47 Chaudhry, W. N. *et al.* Leaky resistance and the conditions for the existence of lytic bacteriophage. *PLOS Biology* **16**, e2005971, doi:10.1371/journal.pbio.2005971 (2018).

48 Maynard, N. D. *et al.* A Forward-Genetic Screen and Dynamic Analysis of Lambda Phage Host-Dependencies Reveals an Extensive Interaction Network and a New Anti-Viral Strategy. *Plos Genetics* **6**, 15, doi:10.1371/journal.pgen.1001017 (2010).

49 Cooper, V. S., Schneider, D., Blot, M. & Lenski, R. E. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. *Journal of Bacteriology* **183**, 2834-2841, doi:10.1128/jb.183.9.2834-2841.2001 (2001).

50 Wichman, H. A., Millstein, J. & Bull, J. J. Adaptive Molecular Evolution for 13,000 Phage Generations. *Genetics* **170**, 19, doi:10.1534/genetics.104.034488 (2005).

51 Burmeister, A. R., Lenski, R. E. & Meyer, J. R. Host coevolution alters the adaptive landscape of a virus. *Proc Biol Sci* **283**, doi:10.1098/rspb.2016.1528 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.31.337758; this version posted November 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

52 Petrie, K. L. *et al.* Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* **359**, 1542-1545, doi:10.1126/science.aar1954 (2018).

53 Maddamsetti, R. *et al.* Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution* **72**, 2234-2243, doi:10.1111/evo.13586 (2018).

54 Scandella, D. & Arber, W. Phage λ DNA injection into Escherichia coli pel− mutants is restored by mutations in phage genes V or H. *Virology* **69**, 206-215, doi:https://doi.org/10.1016/0042-6822(76)90207-5 (1976).

55 Wielgoss, S. *et al.* Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia coli*. *G3: Genes|Genomes|Genetics* **1**, 183, doi:10.1534/g3.111.000406 (2011).

56 Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A* **88**, 7160-7164, doi:10.1073/pnas.88.16.7160 (1991).

57 Best, A. *et al.* Host–parasite fluctuating selection in the absence of specificity. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20171615, doi:doi:10.1098/rspb.2017.1615 (2017).

58 James, A., Pitchford, J. W. & Plank, M. J. Disentangling nestedness from models of ecological complexity. *Nature* **487**, 227-230, doi:10.1038/nature11214 (2012).

59 Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant-animal mutualistic networks. *Proc Natl Acad Sci U S A* **100**, 9383-9387, doi:10.1073/pnas.1633576100 (2003).

60 Guimarães, P. R., Jr., Sazima, C., dos Reis, S. F. & Sazima, I. The nested structure of marine cleaning symbiosis: is it like flowers and bees? *Biol Lett* **3**, 51-54, doi:10.1098/rsbl.2006.0562 (2007).