COEVOLUTION OF BACTERIAL-PHAGE INTERACTIONS

By

Justin R. Meyer

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements
For the degree of

DOCTOR OF PHILOSOPHY

Zoology
and
Ecology, Evolutionary Biology and Behavior – Dual Major

2012

ABSTRACT

COEVOLUTION OF BACTERIAL-PHAGE INTERACTIONS

By

Justin R. Meyer

Bacteria and their viruses, phage, are the most abundant and genetically diverse group of organisms on earth. Given their prevalence, it is no wonder that recent studies have found their interactions important for ecosystem function, as well as the health of humans. Unfortunately, because of technical challenges with studying microbes, some of the most basic questions on their interactions, such as who infects whom, and how their relationships evolved in the first place, remain unanswered. Here I report six studies on bacterial-phage interactions, each focused on understanding their pattern and the underlying biophysical, ecological, and evolutionary processes that shape them. To do this, I tested a number of hypotheses using laboratory experiments and analyses of natural microbial diversity.

First, I tested whether *E. coli* cultured without phage would counter-intuitively evolve new interactions with phage. Typically bacterial traits responsible for phage resistance have pleiotropic consequences on growth, therefore as a side-effect of adapting to an abiotic environment, bacteria may also evolve to become more or less vulnerable to their parasites. After 45,000 generations of laboratory culturing without phage, *E. coli* gained resistance to λ phage, gained sensitivity to a mutant T6 phage, and remained resistant to wild type T6. Each response was explained by understanding the pleiotropic costs or benefits of resistance mutations. Because of pleiotropy, interactions may even evolve in the absence of one player.

For the rest of my studies I examined how interactions evolve when host and parasite co-occur. First, I found that when *E. coli* and phage λ are cocultured, *E. coli* evolves resistance by

reducing the number of phage genotypes that can infect it, whereas, λ evolves to increase the number of bacterial genotypes it can infect. This antagonism produces a interaction matrix with a *nested* form where less derived host-ranges fall one within another. To determine whether this nested pattern is an artifact of the labratory environment, or if the pattern is general to natural communities, I performed a metaanalysis on already published phage-bacterial interaction matrices. The majority of networks were significantly nested (28 of 38). Lastly, I examined the molecular basis of *E. coli* resistance to λ and found that resistance often evolves through mutations in *E. coli*'s *lamB*, the gene for the phage receptor. Also, the strength of resistance is correlated with how the mutation perturbs the orientation specific features of the protein structure, primarily loop four which extends out of the cell membrane.

For the final two chapters, I studied whether λ could evolve to target a novel receptor and the evolutionary consequences of such an innovation. Under particular laboratory conditions, *E. coli* evolves resistance by down-regulating LamB, which sets the stage for λ to evolve the necessary mutations to exploit a new protein receptor. When allowed to coevolve under this condition, λ evolved to exploit another outer-membrane protein, OmpF. This new function is the result of a particular combination of four mutations in J, the gene for the protein ligand λ uses to bind to its host. Once λ evolves this novel interaction, an evolutionary arms-race begins that drives rapid diversification of the bacteria and phage.

Overall, my studies show that coevolution between bacteria and phage, whether it be in the lab or in nature, produces nested interactions matrices. Secondly, antagonistic coevolution is a creative process able to generate new genotypes of host and parasite and promote the evolution of novel function. Lastly, costs for resistance have many important effects, from determining whether resistance will evolve or be lost, to the generation of diversity.

ACKNOWLEDGEMENTS

Hillary Clinton famously said that it takes a village to raise a child, which I think can be extended to students of all ages. My village has been extensive and I have many people to thank. To begin, my parents are the two most giving people I know. I often wonder why I have been so lucky to attend the great schools I have, worked with such extraordinary scientists, and have so many opportunities. My thoughts always lead to how giving and nurturing my parents are. As for the qualities they instilled in me that helped me through my PhD, I am grateful to my mom for giving me a love for the natural world and my first lessons in scholarship, and my dad for giving me the basics on networking and instilling the value of hard work. I hit the most important jack pot of all, a great family. Included in my family is my brother Adam, who himself is very giving and is one of the most intelligent and interesting people I know. He is also very good at giving his older brother a reality check now and again. Also included in my immediate family is our neighbor Dennis. Dennis is one of the most thoughtful, understanding, and bright people I know. It is his early lessons in critical thinking that have made my transition into becoming a scientist possible. Last, but not least, my family grew during my PhD to include my partner Cody. He is, without a doubt, the greatest person I have ever met. I admire his intelligence, balance, even-handedness, and enjoy his humor. And even though a future version of myself will regret writing this, I acknowledge that I have a lot to learn from him.

My village also includes incredible scientific mentors. Just a few are Nelson Hairston Jr., Steven Ellner, Rees Kassen, Francis Pick, and Richard Lenski. Nelson and Steve gave me my first experience with science and provided an outstanding interdisciplinary model for how to perform research. Rees and Frances helped an immature, but enthusiastic scientist move beyond

the early stages of curiosity and investigation, to communicating and conducting science effectively. Rich has opened my eyes to an incredible and diverse world of science and engineering. I have never met another individual with such broad and creative vision.

Last, but not least. I have had two *lab mothers*, Colleen Kearns and Neerja Hajela who have helped me maintain a happy and healthy life through intense and long days in the laboratory.

TABLE OF CONTENTS

LIST OF TABLES……..…...VIII

LIST OF FIGURES…….….IX

KEY TO SYMBOLS AND ABBREVIATIONS……....…XII

INTRODUCTION……...1

CHAPTER 1
PARALLEL CHANGES IN HOST RESISTANCE TO VIRAL INFECTION DURING 45,000 GENERATION OF RELAXED SELECTION
Main text……15
Tables and figures…….……35
Literature cited……...41

CHAPTER 2
INCREASED COSTS FOR VIRAL RESISTANCE INDUCE A COEVOLUTIONARY ARMS RACE BETWEEN *ESCHERICHIA COLI* AND PHAGE λ
Main text……....47
Tables and figures……..65
Literature cited……...73

CHAPTER 3
STATISTICAL STRUCTURE OF HOST–PHAGE INTERACTIONS
Main text……...78
Tables and figures……95
Literature cited……102

CHAPTER 4
BIOPHYSICAL CAUSES OF VIRAL RESISTANCE AND THEIR COSTS IN *ESCHERICHIA COLI*
Main text……111
Tables and figures ……..130
Literature cited……141

CHAPTER 5
REPEATABILITY AND CONTINGENCY IN THE EVOLUTION OF A KEY INNOVATION IN PHAGE LAMBDA
Main text……..145
Tables and figures ……...157
Literature cited……163

CHAPTER 6

KEY INNOVATION IN A VIRUS CATALYZES A COEVOLUTIONARY ARMS RACE
Main text……..168
Tables and figures ……...171
Literature cited……..…174

CONCLUSION……176

APPENDIX 1……..182

APPENDIX 2……...190

APPENDIX 3……..236

APPENDIX 4……...241

LIST OF TABLES

Table 1.1 Statistical analysis of resistance over 45,000 generations……..35

Table 2.3 Identity of λ resistance mutations……...70

Table 4.1: *lamB* single nucleotide replacements……131

Table 4.2: Regression parameters of spatial relationship *lamB* SNPs and phenotypes.....137

Table A1.1 Sensitivity of lab strains of bacteria to three phage……186

Table A2.1 Characteristics of PBINs…….200

Table A2.2 Characteristics of phage and bacteria sampled for metaanalysis……204

Table A2.3 Global parameters……...212

Table A2.4 PCA Analysis……..213

Table A2.5 Correlation analysis…….214

Table A2.6 Isolation bias……215

Table A4.1 Reference genome sequence of cI26……...243

Table A4.2 *E. coli* gene knock-outs studied……..248

Table A4.3 Protein homology to LamB…….……258

Table A4.4 Number of J mutations in each evolved λ genotype……...260

Table A4.5 Bacterial mutations across genome…….271

Table A4.6 Frequency of Man$^-$……..273

LIST OF FIGURES

Figure 1.1: Competitions of T6 sensitive and resistant *E. coli*……..36

Figure 1.2: Dynamics of resistance to T6* over 45,000 generations……..…….37

Figure 1.3: Dynamics of resistance to λ over 45,000 generations……..…….39

Figure 2.1: Models of Coevolution.…….....65

Figure 2.2: Fitness of λ-resistant genotypes……..66

Figure 2.3: Population dynamics……..67

Figure 2.4: Phage-bacterial interaction matrix……..69

Figure 2.5: Dynamics of λ and sensitive and resistant *E. coli*……..71

Figure 2.6: Dynamics of a range of bacterial and phage genotypes.……72

Figure 3.1: Models of coevolution in the networks they create……...95

Figure 3.2: Matrices of bacterial-phage interactions collected from the literature……..96

Figure 3.3: Examples of matrices resorted to estimate modularity……..97

Figure 3.4: Visualization of matrices modularity…….98

Figure 3.5: Visualization of matrices nestedness…….99

Figure 3.6: Values of modularity and nestedness compared to random distributions……100

Figure 3.7: Union of two matrices to demonstrate possible phylogenic scale effects…...101

Figure 4.1: Large-scale λ-*E. coli* interaction matrix……....131

Figure 4.2: Relationship between phage infectivity and host-range……132

Figure 4.3: Trade-off between resistance and fitness……..133

Figure 4.4: Distribution of *lamB* mutations……134

Figure 4.5: Spatial relationship between mutation location and resistance and fitness…..135

Figure 4.6: Examples of mutant LamB deformations…….....136

Figure 4.7: Relationship between LamB deformation and phenotypes……..…137

Figure 4.8: Procrustes distance and bacterial phenotypic values……139

Figure 4.9: Areas where protein shape change correlates with alterations in phenotype...140

Figure 5.1: Spot assays of OmpF+ phage……...157

Figure 5.2: Mutations found in the J gene of phage isolated on day 8……158

Figure 5.3: Mutations found in the J gene from 48 phage genotypes…….159

Figure 5.4: Evolutionary replay experiment results……160

Figure 5.5: Coevolutionary dynamics…….161

Figure 6.1: Interaction matrix of bacteria and phage isolated from two populations…….171

Figure 6.2: Dyanmics of resistance and host-range……173

Figure A1.1 T6* sensitivity after 40,000 generations…….185

Figure A2.1 PCA analysis of PBIN characteristics…….228

Figure A2.2 Correlation between connectedness and species number……229

Figure A2.3 K-means versus PCA……..230

Figure A2.4 Distribution of clustering validity of source types…….231

Figure A2.5 Options for displaying networks……232

Figure A2.6 Resorted PBIN nestedness value compared to original……..233

Figure A2.7 Random distribution of nestedness compared to original……..234

Figure A2.8 Random distribution of modularity compared to original……..235

Figure A3.1 Image analysis……228

Figure A4.1 malT$^-$……..252

Figure A4.2 Population dynamics……...254

Figure A4.3 Phage maintained with malT$^-$ but not lamB$^-$……..256

Figure A4.4 Distribution of timing for OmpF^{+}……..259

Figure A4.5 Adsorption assays……..262

Figure A4.6 λ genotype growth on two hosts…….264

Figure A4.7 Analysis of parallel molecular evolution……266

KEY TO SYMBOLS AND ABBREVIATIONS

Ara^{-}……..…….....Unable to use arabinose

Ara^{+}……..……..Capable of using arabinose

Ara^{-}##.....Designation for one of six LTEE lines founded with an Ara– ancestor

Ara^{+}##.....Designation for one of six LTEE lines founded with an Ara+ ancestor

Mal^{-}……..…….....Unable to use maltose

$MalT^{-}$…….……..……*E. coli* with mutated *malT* gene

$LamB^{-}$……..…… *E. coli* with a nonsense *lamB* mutation

$OmpF^{+}$……..…….....λ able to exploit OmpF

DM…….……Davis and minimal salts base

M9…….……M9 minimal salts base

DNA……...……Deoxyribonucleic acid

IS…….……..…….Insertion sequence element

LB……...Luria-Bertani broth

LTEE……...……...Long-term evolution experiment

L…….……Liter

mL…….....Mililiter

µl…….……Microliliter

g……..…….....Gram

mg…….....Miligram

µg…….……Microgram

PCR……..Polymerase chain reaction

5’……...…..……..5’ end of DNA fragment

3’…….……..3’ end of DNA fragment

SNP……....Single nucleotide polymorphism

TA……..…Tetrazolium arabinose indicator medium

SA……LB agar with half the concentration of agar

PBIN……Phage-bacterial interaction network

INTRODUCTION

Throughout history, naturalists have puzzled over the amazing diversity and complexity of the living world. Extraordinary efforts have been made to document species diversity on earth and today over 1.3 million species have been identified, although studies suggest that this is just a fraction of global diversity (1). Organisms have been discovered in every patch of earth, from as deep as the Mariano trench, high in the atmosphere, at extreme temperatures, and even near nuclear reactors. Besides documenting diversity, naturalists have taken on the audacious task of trying to understand how such complex living systems could have formed from natural processes of biotic evolution. Darwin, of course, uncovered the most well-studied hidden process, evolution by natural selection (2). Since Darwin, evolutionary biology has emerged as a scientific field unto itself, and has made many important advances. Some of the more outstanding work includes the discovery of yet another hidden process, evolution by genetic drift (3,4), uncovering the genetic and genomic underpinnings of adaptation (5,6) and speciation (7,8), deriving a detailed understanding of the evolution of animal development (9), and revealing a more realistic webbed view of the tree of life (10).

Provided this great progress, what mysteries remain to inspire my generation of naturalists? E. O. Wilson, an accomplished naturalist, believes the microbial world holds most of the unanswered questions (personal communication). This is particularly meaningful since Wilson dedicated his life to studying insects and the development of Sociobiology, and not microbes (11). Wilson gained this view after new metagenomic data was found that suggested microbes are the most abundant and genetically diverse group of organisms on earth (12-14). Microbes are ubiquitous, they facilitate an extraordinary number of chemical reactions, fill almost every niche from autotroph to parasite, and occur in even the most extreme environments. Because of their prevalence, microbes are critical for both human and ecosystem health. Unfortunately, due to microbes small size, they have been notoriously hard to study and little is known about how microbial diversity is organized in space, time, and into ecological communities (15).

Besides offering a relatively unexplored frontier for discovering biodiversity, microbes also provide a powerful tool to study the process of evolution. Traditionally, evolution was thought to occur over geological timescales and therefore could not be witnessed or directly studied with experiments. Instead, evolutionary biologists relied on extant patterns of species and genetic diversity to draw inference about the processes underlying the formation of biodiversity, which provides only limited potential to test hypotheses. Microbes however, have short generation times, which means their evolution can be witnessed over the course of days, months, and years. Added to this, controlled and replicated experiments can be performed on microbial populations because they are easily cultured in the laboratory (16). With these characteristics, many evolutionary hypotheses have been tested on microbes that once were thought impossible to directly test, including the role of sexual recombination in adaptation (17),

evolution of assortative mating (18), the evolution of cheating in social groups (19), and the evolution of multicellularity (20).

Other than providing study systems amenable to the 'Strong Inference' approach to science, microbes also offer systems for more open-ended exploration of the evolutionary process. Many important results have been gained from *ad hoc* analyses of evolution experiments, rather than testing specific *a priori* hypotheses (21-23). For instance, the longest running and most prolific evolution experiment, Richard Lenski's Long Term Evolution Experiment (LTEE) of *Escherichia coli* is largely motivated by exploration rather than specific hypothesis. The experiment does not have an end point or a specific expected result. However, many evolutionary hypotheses have been tested using the LTEE through *ad hoc* analyses. Microbial evolution experiments are amenable to these after-the-fact analyses because the evolutionary dynamics that occurred during the course of the experiment can be reconstructed by resurrecting frozen samples stored throughout the experiment. Additionally, because microbial genomes are small and the cost of genome sequencing reduced in recent years, it is feasible to find mutations for adaptation no matter where they occur in the genome. Lastly, even if adaptive mutations occur in a gene with no *a priori* expectation for why it may be adaptive, whether the mutation is beneficial can be inferred by allelic replacement through efficient microbial genetic engineering techniques.

My interest in the natural world has developed in a similar sequence as the field of evolutionary biology. As a child, I was obsessed with collecting and documenting insect and vertebrate diversity. Throughout high school and university I developed an interest in not just characterizing diversity, but in understanding the natural process that gave rise to it. This

interest in process led me to study microbes as an undergraduate and Master's student because they could be used to study the mechanisms of evolution through direct experimentation. I continued similar research for my PhD, where I studied the ecology and evolution of bacteria and viruses. My research goals were to characterize the extraordinary natural diversity of microbes and to take them into the laboratory in order to test hypotheses about evolution. The majority of the studies were laboratory based experiments, and of them, some were carefully designed experiments to test specific hypotheses and others were open-ended to explore unforeseen outcomes of evolution. Despite their various designs, all of my studies were run to investigate the ecology and evolution of interactions between bacteria and phage. Within this framework my research incorporated a range of subjects from genomics and protein structure, to ecological community assembly, coevolution, diversification, and the evolution of novelty.

For the rest of this introduction I will summarize the projects discussed in the following chapters, while also providing a narrative for their logical order. My goal for the Introduction is to provide context for each study and to enrich their conclusions by building connections between them.

One subject that has always fascinated me is the evolution of predator-defense traits. Prey are often trapped with the intriguing dilemma of avoiding being eaten at a price, or being able to out-compete their competitors (24, 25). Through my undergraduate and Master's research, I discovered that this tension could lead to interesting dynamics, such as evolutionary predator-prey cycles (26), and unintuitive interactions between competition and predation during the process of diversification (27). As a PhD student, I continued similar work that examined the ecology and evolution that emerges at the confluence of top-down and bottom-up pressures (28,

29). For my first project as a PhD student I studied evolutionary dynamics after predation pressure and therefore the dilemma to choose between growth and defense was removed. Given that defense traits are typically expensive, in the absence of predators, they should be lost over long evolutionary time. There are a number of natural patterns that support this prediction; plants and animals living in predator- and herbivore-free islands tend to be less resistant than their continental sister species (30, 31), and domesticated species living in artificial enemy-free space are more sensitive to disease (32), as well as bacteria living in axenic laboratory cultures (personal communication with Bruce Levin). However, short-term experiments with bacteria evolving without viruses revealed that viral resistance was not lost, but instead costs for resistance were ameliorated by compensatory mutations (33). This contradiction led Anurag Agrawal and I to question whether bacteria would lose resistance during experiments conducted over longer time-scales. We set out to test this by taking advantage of a long-term evolution experiment (LTEE) where *E. coli* had been evolving in the absence of predators or viruses for over 20 years and 45,000 generations (34). Our findings are reported in **Chapter 1: Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection.**

We studied the evolution of *E. coli* interactions over 45,000 generations to three viruses, T6, T6*, and λ. These strains were chosen because *E. coli* varied in its initial sensitivity to them; resistant to (T6), partially resistant to (T6*), and a completely sensitive to a third (λ). We found that *E. coli* evolved in similar ways across six experimental replicates, however the responses did not match the predictions of trade-off theory and the interactions for each phage showed very different responses. Resistance to T6 did not change, T6* decreased, and surprisingly, *E. coli* became completely resistant to λ. We found that even though classic ecological theory failed to make accurate predictions, the population responses made sense in light of the biochemical

details of resistance and whether resistance was costly under the peculiarities of the experimental conditions.

For the rest of my PhD I decided to study not just the evolution of prey, but their coevolution with parasites. Coevolution is thought to have many important effects on the ecology and evolution of populations, including the ability to drive rapid molecular change (35), the evolution of higher mutation rates (36), speciation (37), and the maintenance of costly sexual reproduction (38). Considering its important effects, I set out to create an experimental system to study coevolution and its broader ecological and evolutionary consequences. To do this, I developed a model coevolutionary system between *E. coli* and phage λ. I chose these species because I could turn their coevolution on and off by adjusting the culture conditions and then monitor the effects of coevolution. My advisor, Richard Lenski, predicted that their coevolution could be controlled by adjusting their resource environment. Lenski had previously shown that phage and bacteria would engage in an arms race when resistance was costly, however not when it was free (39). *E. coli* typically evolves resistance to λ through changes in the expression or amino acid sequence of the viral receptor, LamB (40). LamB is a porin *E. coli* uses to transport a specific class of sugars, maltodextrins, across its outer-membrane (41). Lenski predicted that if I cultured the bacteria and phage with maltodextrins they would coevolve, however if I cultured them in glucose, they would not. My results are reported in **Chapter 2: Increased costs for viral resistance induce a coevolutionary arms race between *Escherichia coli* and phage λ.** Resistance was more costly with maltodextrins and the microbes engaged in an arms race when presented with maltodextrins, but not glucose. As predicted, coevolution in maltodextrins had broad effects that included changes in population dynamics, community composition, diversity, and molecular evolution.

Another prediction Dr. Lenski and I made was that λ and *E. coli* would engage in matching-alleles coevolution. We expected *E. coli* to evolve resistance by altering the shape of LamB, which λ would counter by mutations in the gene for the J protein it uses to adsorb to LamB. Bacteria would continually evolve new LamBs that phage would counter with mutant J proteins (42,43). Matching alleles coevolution is a particularly interesting form of coevolution because it is predicted to drive rapid and never-ending evolutionary cycles (44). Instead, I found that λ and *E. coli* engaged in a relatively short coevolutionary sequence where bacteria evolved increased resistance and phage evolved broader host-ranges for only three rounds until bacteria evolved complete resistance (chapter 2). This form of coevolution is called extended host-range coevolution and is commonly observed between phage and bacteria cultured in the laboratory (43-47).

Conventional wisdom on bacterial-phage interactions is that phage tend to evolve specialized interactions 'One phage, one receptor rule' (48), not the broad host-ranges that emerge during extended host-range coevolution I observed in the lab. Colleagues at Georgia Technical Institute and I wondered if bacterial phage interactions were in fact specialized, or if they showed other patterns indicative of the extended host range coevolution. To test this, we preformed a meta-analysis on interaction matrices collected from the literature that included a wide range of habitats and sampled across large geographical distances. For each matrix we tested whether it possessed nonrandom patterns that correspond to predictions made by extended host-range and/or matching alleles coevolution. Our results are reported in **Chapter 3: Statistical structure of host–phage interactions**. We found that interaction matrices were on average, significantly nonrandom and in line with predictions made by the extended host-range

model. Patterns observed in the lab matched natural patterns, despite disagreeing with conventional wisdom.

After moving up in scale to study how general the extended host-range evolution was, I focused my attention on a much smaller scale to study the molecular basis of resistance. I was interested in determining how mutations in *lamB* drove increased resistance, thereby triggering the extended host-range form of coevolution, and not the specialized interactions first predicted to evolve. Additionally, I wanted to study the biophysical constraints that cause the pleiotropic costs for resistance observed in chapter 2. To do this, I isolated a number of *lamB* mutant *E. coli* that had a range of λ-resistance and pleiotropic costs for defense. Next, I developed a procedure to predict the mutant LamB structures and to determine what aspects of the structure correlate with changes in resistance and competitiveness. The results are reported in **Chapter 4: Biophysical causes of viral resistance and their costs in *Escherichia coli*.** I found that increased resistance was associated with perturbations of loop four that interacts with λ's J protein, and costs were correlated with deformations of the channel used to transport sugars into the cell. Mutations in *lamB* typically have the side effect of altering the pore and thus are costly, however a subset of *lamB* mutations have focused effects on loop four and are free. Therefore, resistance is typically costly, although there is not a strict trade-off between resistance and competitiveness.

Besides driving the evolution of increased values of defense and counter-defense along a single phenotypic axis as described in chapters 1-4, antagonistic coevolution is also thought to shift directions and produce evolutionary novelty (49). For example, coevolution between *E. coli* and λ could drive λ to exploit a novel receptor, rather than only evolving to counter new forms of LamB as observed in Chapters 2 and 4. To study the role of coevolution in driving

innovation I revisited my experiments conducted with glucose medium reported in Chapter 2. When *E. coli* and λ are co-cultured with glucose as the only carbon source, *E. coli* evolves nearly complete resistance by mutations that nearly stop expression of LamB. λ's populations suffer and *E. coli* dominates. Under these conditions, mutations that allow λ to access a novel protein receptor would be selected for. I adjusted the experimental conditions slightly so that the phage population would reach higher densities and reran the experiment. This time, the phage evolved counter-defenses to target a novel receptor, OmpF. Details on the genomics and evolutionary processes underlying the evolution of this key innovation are described in **Chapter 5: Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda.**

Evolutionary innovations are thought to be important moments in a lineage's history that by triggering rapid evolution and cladogenesis (50). I hypothesized that the innovation would trigger an arms race because the bacteria and phage had gained a new gene interaction that may open new evolutionary potentials. To study this, I constructed interaction matrices for two communities previously evolved and reported in Chapter 5, one that evolved the innovation and another that did not. The results are reported in **Chapter 6: Key innovation in a virus catalyzes a coevolutionary arms race.** *E. coli* and λ engaged in an arms race immediately following the innovation, meanwhile they remained static in the other community. Interestingly, the coevolutionary pattern took the form of extended host-range coevolution, just as in the other experiments.

LITERATURE CITED

LITERATURE CITED

1. Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. 2011. How Many Species Are There on Earth and in the Ocean? PLoS Biol 9:e1001127.

2. Darwin, C. 1859. The Origin of Species. Murray. London, UK.

3. Wright, S. W. 1948 On the roles of directed and random changes in gene frequency in the genetics of populations. Evolution 2:279-294.

4. Motoo, K. 1983. The neutral theory of molecular evolution. Cambridge, MA.

5. Hoekstra, H. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222-234.

6. Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. Plos Genetics 6:e1000862.

7. Bradshaw Jr, H., S. M. Wilbert, K. Otto, and D. Schemske. 1995. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers(Mimulus). Nature 376:762-765.

8. Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. PNAS 107:9724.

9. Carroll, S. B. 2005. Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom. WW Norton & Company. New York, NY.

10. Lake, J. A. and M. C. Rivera. 2004. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. Mol. Biol. and Evol. 21:681-690.

11. Wilson, E. O. 2011. Sociobiology: The new synthesis. Cambridge, MA.

12. Torsvik, V., L. Ovreas, and T. F. Thingstad. 2002. Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science 296:1064-1066.

13. Edwards, R. A. and F. Rohwer. 2005. Viral metagenomics. Nat. Rev. Microbio. 3:504-510.

14. Fraser, C., E. J. Alm, M. F. Polz, B. G. Spratt, and W. P. Hanage. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323:741-746.

15. Nemergut, D. R., E. K. Costello, M. Hamady, C. Lozupone, L. Jiang, S. K. Schmidt, N. Fierer, A. R. Townsend, C. C. Cleveland, L. Stanish, and R. Knight. 2011. Global patterns in the biogeography of bacterial taxa. Enviro. Microbio. 13:135-144.

16. Elena, S. F. and R. E. Lenski. 2003. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nat. Rev. Gen. 4:457-469.

17. Cooper, T. F. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. PLoS biology 5:e225.

18. Leu, J. Y. and A. W. Murray. 2006. Experimental evolution of mating discrimination in budding yeast. Current biology 16:280-286.

19. Velicer, G. J., L. Kroos, and R. E. Lenski. 2000. Developmental cheating in the social bacterium Myxococcus xanthus. Nature 404:598-601.

20. Ratcliff, W. C., R. F. Denison, M. Borrello, and M. Travisano. 2012. Experimental evolution of multicellularity. PNAS 109:1595-1600.

21. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

22. Rainey, P. B. and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394:69-72.

23. Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461:1243-1247.

24. Chase, J. M., P. A. Abrams, J. P. Grover, S. Diehl, P. Chesson, R. D. Holt, S. A. Richards, R. M. Nisbet, and T. J. Case. 2002. The interaction between predation and competition: a review and synthesis. Ecol. Let. 5:302-315.

25. Vamosi, S. M. 2005. On the role of enemies in divergence and diversification of prey: a review and synthesis. Can. J. of Zool. 83:894-910.

26. Meyer, J. R., S. P. Ellner, N. G. Hairston, L. E. Jones, and T. Yoshida. 2006. Prey evolution on the time scale of predator-prey dynamics revealed by allele-specific quantitative PCR. PNAS 103:10690-10695.

27. Meyer, J. R. and R. Kassen. 2007. The effects of competition and predation on diversification in a model adaptive radiation. Nature 446:432-435.

28. Hairston, N. G., F. E. Smith, and L. B. Slobodkin. 1960. Community structure, population control, and competition. Am. Nat. 94:421-425.

29. Osenberg, C. W. and G. G. Mittelbach. 1996. The relative importance of resource limitation and predator limitation in food chains. Food webs: integration of patterns and dynamics. Chapman and Hall, New York, New York, USA.

30. Brodie, E. D. 1991. Evolutionary response of predators to dangerous prey: reduction of toxicity of newts and resistance of garter snakes in island populations. Evolution 45:221-224.

31. Bowen, L. and D. V. Vuren. 1997. Insular endemic plants lack defenses against herbivores. Conservation Biology 11:1249-1254.

32. Gepts, P. 2004. Crop Domestication as a Long-Term Selection Experiment. Plant breeding reviews 24:1-44.

33. Lenski, R. E. 1988. Experimental Studies of Pleiotropy and Epistasis in Escherichia coli. II. Compensation for Maldaptive Effects Associated with Resistance to Virus T4. Evolution 42:433-440.

34. Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli* 1. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315-1341.

35. Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, and B. Libberton. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464:275-278.

36. Pal, C., M. D. Maciá, A. Oliver, I. Schachar, and A. Buckling. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. Nature 450:1079-1081.

37. Brown, J. S. and T. L. Vincent. 1992. Organization of predator-prey communities as an evolutionary game. Evolution 46:1269-1283.

38. Hamilton, W. D. 1980. Sex versus non-sex versus parasite. Oikos 35:282-290.

39. Lenski, R. E. and B. R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. 125:585-602.

40. Thirion, J. P. and M. Hofnung. 1972. Some genetic aspects of phage lambda-resistance in *Escherichia coli* K12. Genetics 71:207-&.

41. Lin, E. C. C. 1987. Dissimilarity pathways for sugars, polyols, and carboxylates. Pages 244-284 *in* N. F. C., editor. *Escherichia coli* and *Salmonella typhimurium*: Cellular Molecular Biology. American Society for Microbiology, Washington, DC.

42. Weitz, J. S., H. Hartman, and S. A. Levin. 2005. Coevolutionary arms races between bacteria and bacteriophage. PNAS 102:9535.

43. Forde, S. E., R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220-223.

44. Agrawal, A. and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Eco. Res. 4:79-90.

45. Braun-Breton, C. and M. Hofnung. 1981. In vivo and in vitro functional alterations of the bacteriophage lambda receptor in lamB missense mutants of Escherichia coli K-12. J. of Bac. 148:845-852.

46. Buckling, A. and P. B. Rainey. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc. Roy. Soc. B 269:931-936.

47. Mizoguchi, K., M. Morita, C. R. Fischer, M. Yoichi, Y. Tanji, and H. Unno. 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157: H7 in continuous culture. App. And Enviro. Microbio. 69:170.

48. Schwartz, M. 1980. Interaction of Phages with their Receptor Proteins.*in* K. Lonberg-Holm and L. Philipson, editors. Virus Receptors. Chapman and Hall London, UK.

49. Thompson, J. N. 2005. The geographic mosaic of coevolution. University of Chicago Press. Chicago, IL.

50. Mayr, E. 1963. Animal Species and Evolution. Belknap Press of Harvard University Press, Cambridge, MA.

CHAPTER 1

PARALLEL CHANGES IN HOST RESISTANCE TO VIRAL INFECTION DURING 45,000 GENERATIONS OF RELAXED SELECTION.

Authors: Justin R. Meyer, Anurag A. Agrawal, Ryan T. Quick, Devin T. Dobias, Dominique Schneider, and Richard E. Lenski

Originally published in the journal Evolution 64, 3024–3034 (2010).

Copyrights extended to publish in this dissertation by License Agreement #2977130112749 between author Justin R. Meyer and publishing company John Wiley and Sons on August 27, 2012. Full agreement supplied on request: justin.raymond.meyer@gmail.com

Abstract

The dynamics of host susceptibility to parasites are often influenced by tradeoffs between the costs and benefits of resistance. We assayed changes in the resistance to three viruses in six lines of *Escherichia coli* that had been evolving for almost 45,000 generations in their absence. The common ancestor of these lines was completely resistant to T6, partially resistant to T6* (a mutant of T6 with altered host range), and sensitive to λ. None of the populations changed with respect to resistance to T6, whereas all six evolved increased susceptibility to T6*, probably ameliorating a cost of resistance. More surprisingly, however, the majority of lines evolved complete resistance to λ, despite not encountering that virus during this period. By coupling our results with previous work, we infer that resistance to λ evolved as a pleiotropic effect of a beneficial mutation that down-regulated an unused metabolic pathway. The strong parallelism between the lines implies that selection had almost deterministic effects on the evolution of these

patterns of host resistance. The opposite outcomes for resistance to T6* and λ demonstrate that the evolution of host resistance under relaxed selection cannot be fully predicted by simple tradeoff models.

Introduction

The evolutionary dynamics of populations' resistance to their enemies have long been of interest (1- 3). The costs and benefits of resistance, especially coupled with spatiotemporal variation in selection exerted by natural enemies, can maintain genetic variation in host populations (4). Both historical and experimental approaches have been pursued to study these dynamics, although there have been few long-term studies.

Two comparative approaches have often been employed to study the evolution of resistance. First, populations living on islands or otherwise protected from their natural enemies were predicted to become more susceptible to those enemies, given relaxed selection for resistance and its presumed cost (5-8). These studies have frequently reported reduced resistance in the absence of enemies, although these studies typically lack evolutionary replication because they focus on a single host lineage. Second, the patterns of sequence variation at loci conferring resistance have been used to infer evolutionary dynamics (9). Positive selection for resistance in the face of a coevolving pathogen, for example, has been inferred from observing an excess of non-synonymous to synonymous substitutions in genes controlling resistance to infection.

Experimental approaches, including quantitative genetics, have also provided a way to assess the costs and benefits of resistance using single-generation performance assays and selection experiments (10-12). However, multi-generational studies are needed to understand the

long-term dynamics of resistance evolution because responses to parasite-mediated selection may be ameliorated or intensified by other ecological interactions (e.g., competition), by pleiotropic effects of resistance genes, and by selection at other life stages than those measured in short-term assays.

Experimental microcosms have proven to be powerful for the study of resistance evolution (13). For example, the interplay between the benefits and costs of resistance by algae to rotifers promoted rapid evolution in that system and, moreover, substantially influenced their predator-prey cycles (14-16). In various experiments with *Pseudomonas* bacteria, costs of resistance, predator-prey evolutionary dynamics, and community consequences have been well studied (17-19). Although costs of resistance are often detected, they are sometimes quite small, they may be ameliorated by selection (20), and they may vary depending on other genetic factors (21).

The evolution of resistance of the bacterium *Escherichia coli* to several viruses that infect it has also been well studied using chemostats (22). In these studies, resistance is usually discrete and qualitative (i.e., complete resistance or susceptibility), the mechanisms of resistance are well characterized, and the multi-generational fitness consequences of resistance have been investigated. For example, Lenski (1988) studied the evolution of resistance to bacteriophage T4. Resistance was initially very costly, although the magnitude of this cost varied among independent host mutants. After 400 generations of evolution in the absence of T4, however, the fitness of resistant lines had achieved nearly the same level as susceptible populations that were evolving under identical conditions. Interestingly, the resistant populations did not revert to susceptibility, but instead they had evolved compensatory changes that ameliorated the cost of resistance.

The goal of the present study is to examine the long-term fate of resistance in host populations where the resistance is unnecessary owing to the absence of parasites. We take advantage of an experiment, started over 20 years ago with 12 replicate populations of *E. coli* (24-26). Based on simple tradeoff models, we initially predicted the loss of resistance by these bacteria to viruses, or phages, that infect them. However, our increased mechanistic understanding of the costs of resistance and pleiotropic effects of certain mutations that confer resistance altered our expectations for three different phages, as described below. The ancestral strain was resistant to the lytic phage T6, and the replicate populations have evolved independently for nearly 45,000 generations in the absence of T6 or any other parasites. If resistance to T6 imposed even a small cost, then one would expect this resistance to decay in a convergent manner (i.e., across the replicate lines) over time. We test that prediction by reviving frozen stocks of six of these lines from several time points and assaying their resistance to phage T6. We also directly assessed the cost of resistance to T6 using reverse genetics to create isogenic resistant and susceptible *E. coli* strains. In the course of our preliminary work with T6, we discovered a mutant phage, which we call T6*, that can infect the ancestral strain, although its infectivity on the ancestor was much lower than it was on another T6-sensitive host. We therefore also examined the evolution of susceptibility to T6*, where that trait was evidently quantitative rather than discrete.

A third phage, called λ, is unrelated to the other two, but offers an additional contrast in that the ancestral strain was fully sensitive to this phage. Moreover, from previous research on *E. coli* (27,28), and these evolving lines in particular (29) we expected that the expression of the receptor protein exploited by λ to infect cells may have evolved. In particular, the LamB surface protein is part of a regulon that *E. coli* uses to transport and metabolize maltose and other

maltodextrin sugars that were also not present in the environment during the evolution experiment. Indeed, proteomic and genetic analyses revealed that reduced expression of this regulon, including the LamB protein, had evolved in some of these long-term lines (29), although the consequences of those changes for resistance to λ were not analyzed. We therefore examined the evolution of susceptibility to phage λ because it offered a potentially instructive counter-example to the intuition that parasite resistance should decline under relaxed selection.

Materials and Methods

The long-term evolution experiment. Detailed information on the long-term evolution experiment (LTEE) can be found elsewhere (24, 25, 30) Briefly, two ancestral clones, called REL606 and REL607, were used to found six populations each in 1988. These clones differ by a selectively neutral mutation in *araA* that serves as an easily scored marker in competition assays based on the arabinose-utilization phenotype. The 12 populations have been serially propagated for over 20 years in a glucose-limited medium. Each day, 0.1 ml is removed from the population and transferred into 9.9 ml of fresh medium, where it grows until the glucose is depleted. Given the 1:100 daily dilution and re-growth, the populations undergo ~6.6 (= $\log_2$ 100) generations of binary fission each day, and they had achieved some 44,500 generations at the time that we commenced this project.

Every 500 generations (75 days), a sample of each population was stored frozen at -80°C with glycerol, which serves as a cryoprotectant. These samples include essentially all the diversity present in each population at the corresponding generation, because they were obtained by adding glycerol to, and preserving, the majority of the population from which the transfer was

made. For our analyses, we revived bacteria from six populations using samples taken at 0, 2000, 10,000, 20,000, 30,000, 40,000 and 44,500 generations. Three clones (asexual genotypes) were randomly isolated from each sample to assess their resistance phenotypes. In choosing the populations for this study, we excluded some that had evolved to grow poorly and inconsistently on our experimental plates.

As noted in the Introduction, the ancestral strain used for the LTEE was resistant to phage T6 (24). That resistance reflects a spontaneous mutation that had been selected by a geneticist using an even earlier progenitor (31,32). In particular, a point mutation generated a premature stop codon in the *tsx* gene (33, 34), which encodes an outer membrane protein that is the target for adsorption of phage T6 (35). The mutation confers complete resistance to 'wild type' T6, but it provides only partial (quantitative) resistance to some extended host-range mutants of T6, as has been reported for certain other T6-resistant mutants (36). One of us (R.E.L.) obtained a stock of phage T6 many years ago from Bruce Levin (now at Emory University). For this study, we also isolated a host-range mutant from a rare plaque (zone of lysis) that originally formed on a lawn of a T6-resistant clone sampled at generation 44,500 from the Ara+5 population of the LTEE. We refer to this mutant phage as T6*.

The ancestral strain in the LTEE was sensitive to λ infection. Phage λ targets the outer-membrane porin LamB, which the cell uses to transport maltose and other maltodextrin sugars across the outer-membrane into the periplasmic space, after which another system actively transports these sugars into the cell (28). Wild type λ is a lysogenic phage that sometimes becomes incorporated into the host chromosome, which would unnecessarily complicate our analysis of changes in host resistance. Therefore, we used a strictly lytic mutant cI26, provided by Donald Court (National Cancer Institute). This mutant phage harbors at least one chemically-

induced mutation in the *cI* gene, which encodes a repressor protein required for lysogeny. This phage has the same mode of entry into the cell as wild type λ, but a successful infection is lethal to the host cell. For simplicity, we refer to this mutant phage as λ.

Assays of bacterial susceptibility. We assessed susceptibility to phage infection using 'spot' plates (37), on which a dense lawn of bacteria is immobilized in a matrix of soft agar. An aliquot of a liquid stock of phage is then spotted on the bacterial lawn and incubated overnight. If a virus infects a cell, and if that infection propagates to other cells, then a spot of clearing, known as a plaque, will occur where the cells have lysed in the otherwise turbid lawn of cells. This approach has been widely used for decades to quantify differences among bacterial strains in their susceptibility to phage infection, although a recent study described some potential problems with this inference (39). These problems were not an issue for our study, however, as we confirmed our results on differences in susceptibility using a second method, in which liquid cultures of bacteria were challenged with phage (see Appendix 1).

The spot assays were set up by pouring a mix of 2 ml of molten soft agar and 1 ml of *E. coli* culture over a Petri plate containing a base of hardened Luria Bertani (LB) agar. Cultures were grown overnight at 37ºC and shaken at 120 rpm in liquid LB medium. See Sambrook et al. (1989) for media recipes. After the soft agar had solidified, six evenly spaced 10-µl spots of a single phage lysate were spotted onto the top agar. Each spot after the first represented a further 100-fold dilution of the phage lysate; thus, the first spot has the original phage concentration, followed by relative concentrations of 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, and 10^{-10}. Phage infectivity can vary over many orders of magnitude, and the idea of the spot test is to include a dilution that allows some countable number of discrete plaques, such that host susceptibility can be

quantified. Phage lysates were standardized such that a 10-µl aliquot of the 10^{-6} dilution yielded ~100 plaques when tested on a lawn of an *E. coli* K-12 strain (JA221 from ref. 41) that is sensitive to phages T6, T6*, and λ. The plates were incubated overnight at 37ºC, and the plaques within each appropriate spot were counted.

Assays were performed on 108 evolved clones, including three clones from each of six populations (those designated Ara-3, Ara-5, Ara+1, Ara+2, Ara+4, and Ara+5) at six time points (generations 2,000, 10,000, 20,000, 30,000, 40,000 and 44,500). In addition, three clones were isolated from the marked Ara^{+} and Ara^{-} ancestral variants to measure the ancestor's susceptibility to the phage. Assays were conducted in a completely randomized design, with two replicate assays for each clone that were averaged to produce a single value for the statistical analyses. The susceptible K-12 strain was also included in the assays to verify the infectivity of the phage lysates, but it was not included in the statistical analyses. We preformed the T6 and T6* assays on the same day, while the λ assays were performed on a later date. All aspects of the assays, including the *E. coli* clones tested, were identical for the three phages.

We analyzed the susceptibility assays separately for each phage by performing a mixed-model ANOVA using restricted maximum-likelihood (JMP, version 7). Time (generation) was treated as a continuous effect, while line (population) and the line-by-time interaction were random effects. Random effects were tested using likelihood ratios (42).

Sequencing the tsx *and* malT *genes.* We sequenced the *tsx* gene, which encodes the receptor for T6 adsorption, in the ancestor of the LTEE, in three clones from each of the six lines at 40,000 generations, and in the T6-sensitive *E. coli* K-12 strain. Primers (5' ATCCCGGCATTTTCATAA 3' and 5' AAGGGGATTTCTGTGGAT 3') were designed to

cover all of *tsx* as well as 150 base pairs upstream and 129 base pairs downstream of the gene. The PCR products were then purified on a GFX column, and *tsx* and the adjacent sequences were obtained in their entirety for each clone using an ABI automated sequencer. Sequences were aligned using DNA Star SeqMan (version 1) and inspected visually for any differences.

We also sequenced the *malT* gene in clones sampled at 40,000 generations from two of the six LTEE lines, Ara-5 and Ara+5, as well as the ancestral strain. The *malT* gene encodes the transcriptional activator of the maltose regulon, which includes the *lamB* gene that encodes the LamB surface protein that phage λ targets in sensitive cells. Previous work had identified mutations in *malT* in several of the LTEE lines (29), and we sought to determine whether some of the changes we saw in resistance to λ might also involve mutations in this gene. Primers 5' CACCGGTTTGGCGAATGG 3' and 5' GCGGCGGTGGGGGAATA 3' were designed to cover all of *malT* as well as 392 and 211 base pairs upstream and downstream, respectively. The subsequent steps proceeded as above for the *tsx* gene.

Direct test of the cost of resistance. The long-term evolutionary fate of resistance is likely to depend strongly on how costly resistance is in the absence of selection for that phenotype. Therefore, we used reverse genetics to construct a strain, in the ancestral background, that is isogenic except for the mutation that confers resistance to phage T6. We then used this isogenic strain in competition assays to measure directly the cost of resistance to T6.

To produce a T6-sensitive variant of the Ara^- ancestor, we replaced its *tsx* gene containing the premature stop-codon with the wild type version of that gene by using the 'gene-gorging' procedure described by Herring *et al*. (2003) and modified by Sleight et al. (2008).

Primers 5' TAGGGATAACAGGGTAATGCTGTTCCCGCGAGTTTGT 3' and 5' GGATGCGCCGGTATTCTTC 3' were designed to flank 1,000 bp around the premature stop at codon 258 (out of 295) and used to PCR-amplify a fragment from *E. coli* K-12 (strain JA221) that has a wild type (susceptible) version of *tsx*. This fragment was incorporated into a TOPO10 plasmid and used to transform the LTEE ancestor (strain REL606) together with a second plasmid over-expressing the I-*Sce*I restriction enzyme (43). We then screened 800 transformed clones for sensitivity to T6 using spot plates; only two of the clones were sensitive. The transformed strain that we would use in experiments to assess the cost of resistance was designated JRM100, and the success of the allelic exchange was confirmed by sequencing the entire *tsx* gene. The only change was at the site that had encoded the premature stop codon. JA221 has one other difference in *tsx* from REL606, a synonymous change in codon 88. However, this synonymous difference was not integrated into the T6-sensitive JRM100.

We then measured whether there was any difference in fitness by competing T6-sensitive and T6-resistant strains under the same phage-free conditions as used during the LTEE (25). The arabinose-utilization marker was used to distinguish the competitors. In particular, the Ara^- T6-resistant ancestor REL606 and its T6-sensitive counterpart JRM100 competed separately against the Ara^+ T6-resistant ancestor REL607, each with 10-fold replication. These pairwise competitions were propagated for 6 days, with plating onto tetrazolium-arabinose indicator agar on days 0, 1, 3, and 6 to determine the relative abundance of the competitors. We then calculated the relative fitness of the competitors from those data, where relative fitness is calculated as the ratio of net growth rates achieved during their direct competition (25), and we compared the fitness values estimated for the otherwise isogenic T6-sensitive and T6-resistant strains.

Pelosi *et al.* (2006) performed similar assays to determine the fitness effect of mutations in the *malT* gene, also in the ancestral genetic background and culture conditions of the LTEE. They found that the two mutant strains they studied each had small, but significant, advantages over their sensitive counterpart, even though no phage were present in the cultures.

We were unable to measure directly the cost of resistance to T6* because we do not know which loci harbor the mutations responsible for the quantitative variation in resistance to this phage. Hence, we cannot make and compete otherwise isogenic strains that differ only by the relevant mutations.

Results

All of the clones that we tested from all populations and generations appeared to be completely resistant to the wild-type T6 phage. However, by using a high-density phage T6 lysate, we observed a few plaques on the evolved bacteria. We isolated phage from six of these plaques, and all six of these new lysates consistently produced plaques on the T6-resistant ancestor as well as on the evolved bacteria, indicating that they were host-range mutants. We chose one of these host-range mutants, designated T6*, for further analysis. We infer that most or all of the plaques we observed when we challenged the evolved bacteria with the wild-type T6 lysate were produced by spontaneous host-range mutants within that lysate. There is no compelling evidence, therefore, that the evolved bacteria have increased susceptibility to the wild-type phage.

The sequence of the *tsx* gene that encodes the receptor for the wild-type T6 phage was identical in the T6-resistant ancestor of the LTEE and in a T6-sensitive *E. coli* K-12 strain, with

the exception of two mutations, one of which is a synonymous substitution. The mutation that confers resistance is a premature stop at codon 258 of 295; this effect was confirmed by replacing that codon with the corresponding one from the K-12 strain. The resulting construct was just as sensitive to T6 as K-12 (see Table A1.1). None of the 18 clones we sampled from the six evolved lines at generation 40,000 had any other mutations in *tsx* or the adjacent regions that we sequenced.

This result ran counter to our expectation that the bacteria would evolve increased sensitivity to T6, perhaps even by reverting the mutation in *tsx* that made them resistant, in the prolonged absence of any phage. Considering the duration of the evolution experiment and the corresponding opportunity for slightly beneficial alleles to fix, there was substantial potential to ameliorate even a small cost to resistance. However, resistance to T6 was not costly in the environment of the LTEE. In fact, the resistance caused by the premature stop codon in the *tsx* gene turns out to confer a slight benefit during competition in the phage-free environment of the evolution experiment (Figure 1.1). The T6-resistant ancestor was ~1.3% more fit than its otherwise isogenic T6-sensitive counterpart ($t_s = 3.81$, d.f. = 18, two-tailed $p = 0.0013$). Thus, the maintenance of T6-resistance for almost 45,000 generations of the evolution experiment is consistent with the absence of any associated cost of resistance.

By contrast, all six evolving lines showed substantial increases in susceptibility to the host-range mutant phage T6* over this same period (Figure 1.2). The temporal trend was significant, as was the variation among the lines in susceptibility (Table 1.1). However, the line-by-time interaction was not significant in our analysis (Table 1.1), which indicates that the lines showed similar temporal trends (Figure 1.1). We note that all lines started from the same ancestral state with respect to phage susceptibility; hence, the significant between-line variation

implies some line-by-time interaction, although the statistical analysis was not powerful enough to show it.

In contrast to both the absence of any evolutionary change in resistance to T6 and the evolution of increased susceptibility to T6*, five of the six lines evolved increased resistance to phage λ (Figure 1.3). Of those five, four evolved complete resistance, such that no plaques were seen even at the highest phage concentration, and three of them had evolved this resistance by 20,000 generations (Figure 1.3). The temporal trend, the heterogeneity among lines, and the line-by-time interaction were all significant in the case of susceptibility to phage λ (1.1). We sequenced *malT*, a gene that positively regulates the expression of LamB, the surface protein that phage λ targets in susceptible cells, in clones sampled at generation 40,000 from two populations, Ara-5 and Ara+5, that evolved complete and partial resistance, respectively. Mutations in *malT* that greatly reduce levels of the LamB protein were previously reported in 20,000-generation clones from populations Ara+1, Ara+2, and Ara-3 (29), all of which had become completely resistant to λ by that time (Figure 1.3). Consistent with the other populations that had evolved complete resistance to phage λ in earlier generations, Ara-5 had a mutation in *malT*. In this case, the change was a C-to-T point-mutation at nucleotide position 547, which results in a premature translational stop early in the gene (codon 183 out of 902). This mutation likely renders the residual protein non-functional and, because the MalT protein is a transcriptional activator of the *lamB* gene, expression of the receptor protein LamB would be eliminated. Population Ara+5 also acquired a mutation in *malT*, though one with more subtle effects. In particular, it has a C-to-T point-mutation at nucleotide position 2570, which changes an alanine to a valine at amino-acid 856. These residues are chemically similar, each being non-polar, although valine is slightly larger with two more methyl groups. The difference in size

evidently affects the structure of the protein, but only slightly, allowing the altered MalT to promote the expression of LamB, though presumably at a reduced level, thereby conferring partial resistance.

Discussion

Resistance to parasites and pathogens often imposes a fitness cost to the host, and such costs have been widely reported for many organisms from bacteria to plants and animals. Resistance functions may be costly either because they are energetically expensive to produce or because they interfere with other cellular or organismal functions. However, some instances of resistance do not incur any measurable fitness costs. For example, resistance by *E. coli* to phage T5 appears to be cost-free under conditions of carbon limitation (45,46); the same mutations are probably deleterious under conditions of iron-limitation, because T5 adsorbs to a receptor protein involved in iron transport (47). In a related way, some mutations that confer resistance to phage λ might be neutral, even when the relevant resource is limiting, if the mutations affect the LamB protein's ability to adsorb phage but not its maltodextrin binding and transport functions (48). As a eukaryotic example of cost-free resistance, a study of the annual morning glory, *Ipomoea purpurea*, detected no costs of resistance to several insect herbivores (49), although later work suggested that some costs might have been missed because the resistance traits were not expressed constitutively but instead were inducible (50).

Because so much ecological theory assumes a cost for resistance, it is imperative that more studies are conducted to understand the prevalence of these costs and their long-term effects on the evolution of resistance. Studies extending over many generations have the ability

to detect subtle costs that, when accumulated over long periods, may have profound evolutionary effects on resistance traits. In particular, in the prolonged absence of parasites, host populations are expected to evolve increased sensitivity to the parasites even if there is only a small cost of resistance.

*Evolution of resistance to phages T6 and T6**

We tested this prediction by measuring changes in six *E. coli* populations that initially were resistant to phage T6, but moderately susceptible to a mutant called T6*, across almost 45,000 generations. Convergent increases in sensitivity would lend strong support to the hypothesis of a cost of resistance (i.e., some advantage to becoming susceptible in the absence of viruses), because evolutionary parallelism is a hallmark of adaptation (51-59).

However, we saw no discernible changes in host susceptibility to T6, as all six evolved lines appeared to be just as resistant as their ancestor. The genetic basis of the ancestor's resistance to T6 was a single point mutation in the *tsx* gene, so there should be little difficulty in evolving lower resistance if that would provide a selective benefit. We sequenced this gene in the six evolved lines, and there were no reversions or any other mutations, fully consistent with the absence of phenotypic changes in susceptibility to T6. We inferred, therefore, that resistance to virus T6 imposes little or no cost to the bacteria under the environmental conditions of the LTEE. This inference was then verified by a direct test that compared the competitive fitness of otherwise isogenic T6-resistant and T6-sensitive strains. This comparison not only found no cost of resistance but, in fact, it revealed an unexpected, albeit slight, competitive advantage to the resistant genotype.

By contrast, all six evolved populations we tested show substantially increased susceptibility to virus T6* (Figure 1.1), strongly supporting the hypothesis that resistance to that mutant parasite was costly to the bacteria. However, the physiological basis of that cost is unclear because both the T6* receptor and the genes responsible for the altered resistance are unknown. We detected significant variation among the six lines in their susceptibility to T6* (1.1), even though the direction of the change was the same in all cases. While such quantitative variation is less obvious in its phenotypic manifestation than all-or-none resistance, it can have important effects on the ecological and evolutionary dynamics of bacteria-phage interactions (60, 61).

A number of other strikingly convergent changes have previously been reported for populations in the LTEE, including competitive fitness, cell morphology, catabolic niche breadth, and gene expression (15, 24, 25, 29). These parallel changes have often been imperfect, in the sense of significant among-line variability in the magnitude of the changes (25), similar to the heterogeneity in susceptibility to T6* that we observed in this study.

Although such parallel evolutionary changes usually imply adaptation, the precise nature of that adaptation is not always clear. Some parallel changes may reflect selection acting directly on the traits that have been measured, while others might instead indicate correlated responses to selection on other traits. It can be difficult to distinguish between these alternatives without precise understanding of the genetic architecture of the relevant traits as well as of the selective forces acting on all the traits. In the case of the LTEE, parallel changes in DNA supercoiling, for example, appear to confer a direct advantage based on genetic manipulations (62). By contrast, losses of physiological functions that are unnecessary in the environment of the LTEE, including

some catabolic activities as well as growth capacity at other temperatures, may reflect selection to reduce the costs of unused functions (26).

Without knowing the mechanism of infection by the host-range mutant T6*, we cannot distinguish between two opposing hypotheses for the evolution of increased host susceptibility to that mutant phage. On the one hand, selection in the LTEE environment might favor increased production of a cell-surface receptor that provides a direct benefit, such as resource transport. On the other hand, selection might favor reduced production of some metabolically costly molecule that inhibits access by T6* to the receptor it uses for adsorption. However, the lack of subsequent mutations in the *tsx* gene excludes the possibility that increased susceptibility to T6* resulted from a compensatory change in the same gene. Compensatory mutations for costly resistance traits have been reported in studies of bacterial resistance to other viruses and antibiotics (63-65), although compensation often involves mutations in other genes (66).

Evolution of resistance to phage λ

In striking contrast to the evolution of increased susceptibility to T6*, most of the same lines evolved resistance to λ, despite the absence of any phage in the environment of the LTEE. The majority of the lines evolved sudden, qualitative resistance to λ, with that discrete phenotype strongly suggesting that a single mutation might be responsible. This convergent resistance to a parasite that was absent implies that this trait evolved as a pleiotropic side-effect of selection on some other trait, and indeed that is almost certainly the case based on the relationship between our findings and results previously reported in these same lines. In particular, the medium used in the LTEE contains glucose but not maltose, and selection favored mutations that reduce the

expression of genes involved in maltose acquisition and catabolism (29). One of the maltose-related genes that evolved lower expression encodes the outer-membrane protein LamB, which is the receptor that λ exploits to gain entry to the host. Thus, a mutation that was beneficial because of metabolic cost-savings conferred resistance to λ as a pleiotropic effect. In fact, there is a perfect correlation between the lines that we observed to be resistant to λ at generation 20,000 and those with reduced expression of LamB in the same generation (29).

Furthermore, Pelosi et al. (2006) found the mutations responsible for the lower LamB expression in lines Ara-3, Ara+1, and Ara+2, all of which were in *malT*, a regulatory gene that controls the transcription of *lamB* and other maltose-related genes. These mutations caused deletions and amino-acid substitutions, and the resulting mutant MalT proteins may be unable to bind to the *lamB* promoter sequence (29). Line Ara-5 also evolved a similar, discrete resistance phenotype, but after the generation examined by Pelosi et al. (2006). We therefore sequenced clones sampled at 40,000 generations from population Ara-5 as well as from population Ara+5, which had evolved low-level quantitative resistance to phage λ infection. The results are strikingly consistent with those of Pelosi et al. (2006). Population Ara-5 evolved a premature stop early in the *malT* gene that likely destroys its function and, as a consequence, eliminates expression of LamB, consistent with complete resistance to λ in that lineage. Population Ara+5 had a non-synonymous mutation in the distal portion of *malT*, one that involves two similar amino-acids, consistent with the subtle quantitative resistance in that case.

If such regulatory mutations are beneficial, then why has line Ara+4 not evolved resistance to phage λ? Given the population sizes and mutation rates in the LTEE, any particular point mutation has likely occurred multiple times in each population (30). One might imagine, therefore, that all beneficial mutations would have had ample opportunity to sweep each

population. However, the bacteria in the LTEE are evolving asexually, which leads to competition between co-occurring clones that carry different beneficial mutations, giving rise to the phenomenon of clonal interference (66, 67). In particular, a beneficial mutation that confers only a small advantage may repeatedly appear, increase in frequency, but be driven extinct before it reaches fixation by other mutations of larger effect. The fitness benefit associated with *malT* mutations has been shown to be significant but small, on the order of 1% (29), making it among the most subtle beneficial mutations discovered to date in the LTEE. Moreover, one of the three clones we tested from population Ara+4 at generation 10,000 was completely resistant to phage λ, yet resistant clones were not observed in later generations, indicating that at least one such mutant rose to some intermediate frequency but was eliminated. Given the demonstrated fitness advantage of reducing the expression of the maltose regulon, we anticipate that resistance to phage λ will eventually become fixed in this population as well.

Conclusions

A substantial body of ecological theory assumes a universal cost to resistance to parasites, such that resistance should decay when hosts evolve over long periods in the absence of parasites. We saw this dynamic in the case of phage T6*, where all six *E. coli* lines we tested became increasingly susceptible over nearly 45,000 generations in the absence of this parasite. However, we did not see this trend for two other phages. In the case of T6, we saw no changes in host susceptibility in any of those six lines, despite the fact that a single mutation would lead to sensitivity. This outcome suggested that T6-resistance imposes little or no fitness cost on the bacteria in the environment of the LTEE, although it might be costly in some other

environments. In fact, we measured a slight advantage for the resistant ancestor over its isogenic sensitive counterpart in direct competition. Even more problematic for the generality of the theory of costly resistance, we observed that most lines we tested evolved increased or complete resistance to phage λ, even though the ancestor was sensitive and none of the lines were exposed to this or any other phage during the LTEE. By combining our observations with other data, we showed that this resistance had evolved as a pleiotropic by-product of selection on another trait, namely reduced expression of an unused pathway for exploiting a resource that was not present in the experiment. These opposite and unexpected patterns in the evolution of susceptibility to T6* and of resistance to λ illustrate that the genetic and physiological mechanisms that underlie host-parasite interactions can be complex. These complexities may be especially apparent when organisms evolve in changing or novel environments, as in our study. Thus, ecological and evolutionary theories that assume the ubiquity of costs of resistance could lead to incorrect predictions. Whether a cost exists in any particular case may depend on other, seemingly unrelated, aspects of the environment.

When such costs do exist, they will certainly be important for the maintenance of variation in resistance traits in nature. What is sobering, however, is that most studies that attempt to detect tradeoffs associated with resistance to natural enemies have been conducted over only one or a few generations, which may be insufficient if the cost is small, standing genetic variation in resistance is absent, or both.

Table 1.1 Statistical analysis of the effects of evolved line (six lineages started from the same ancestor), time (seven points from 0 to 44,500 generations), and their interaction on the susceptibility of *E. coli* populations to parasitism by phages T6* and λ. Three clones were tested from each line at each time point. Time is a fixed effect and was evaluated using an *F* test; line and the interaction term are random effects, and they were evaluated by likelihood ratio tests.

	Phage T6*			Phage λ		
	Df	F or χ^2	P	Df	F or χ^2	P
Time	1, 5.89	21.93	0.004	1, 6.19	12.95	0.011
Line	1	5.33	0.010	1	34.71	<0.001
Line x Time	1	0.34	0.280	1	18.30	<0.001

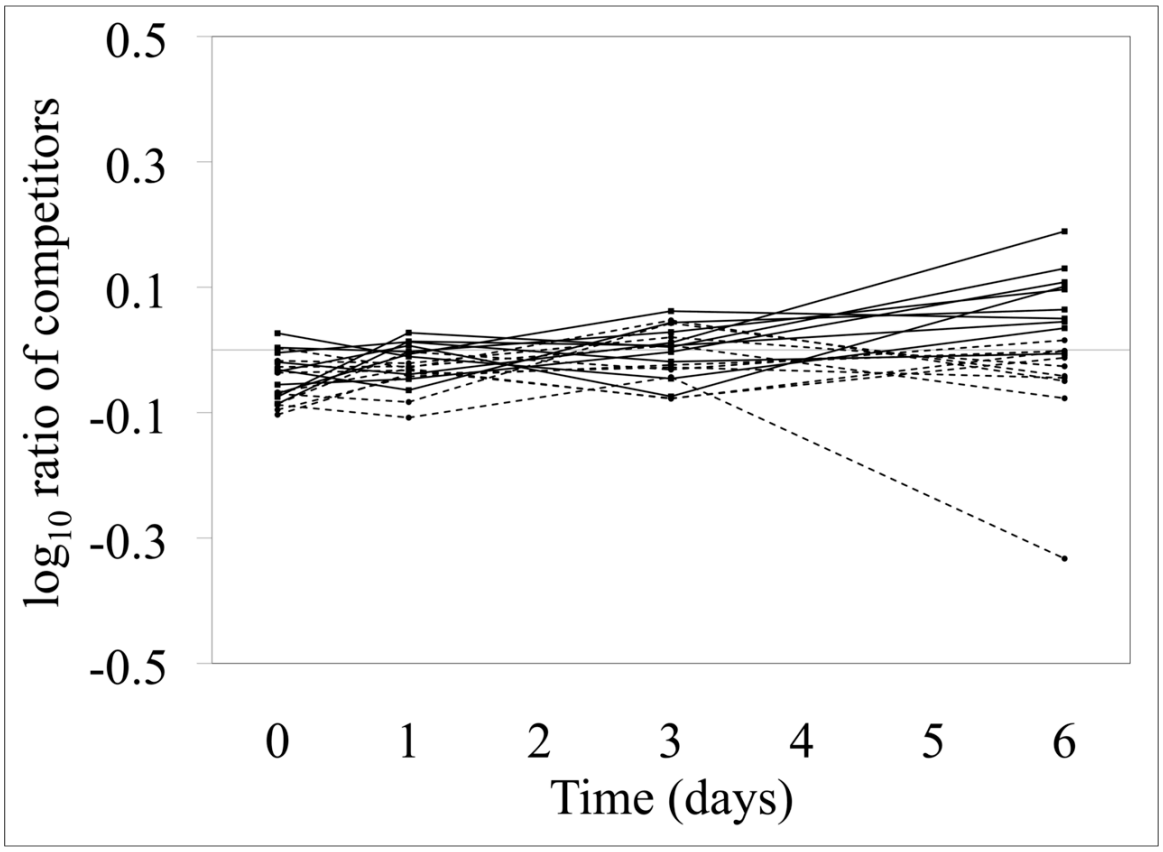


Figure 1.1. Log_{10} ratios of competing pairs of *E. coli* strains across six daily transfer cycles. For solid and dashed lines, the competitors of interest (reflected in numerator counts) are, respectively, the Ara^{-} T6-resistant ancestor and its T6-sensitive derivative. For both the solid and dashed lines, the common competitor (reflected in dominator counts) is an Ara^{+} mutant of the T6-resistant ancestral strain. Note that the resistant competitor of interest was relatively more abundant than its sensitive counterpart at the end of most assays, but not at the beginning, indicating a slight competitive disadvantage to being sensitive even in the absence of the phage. 0See text for statistics.

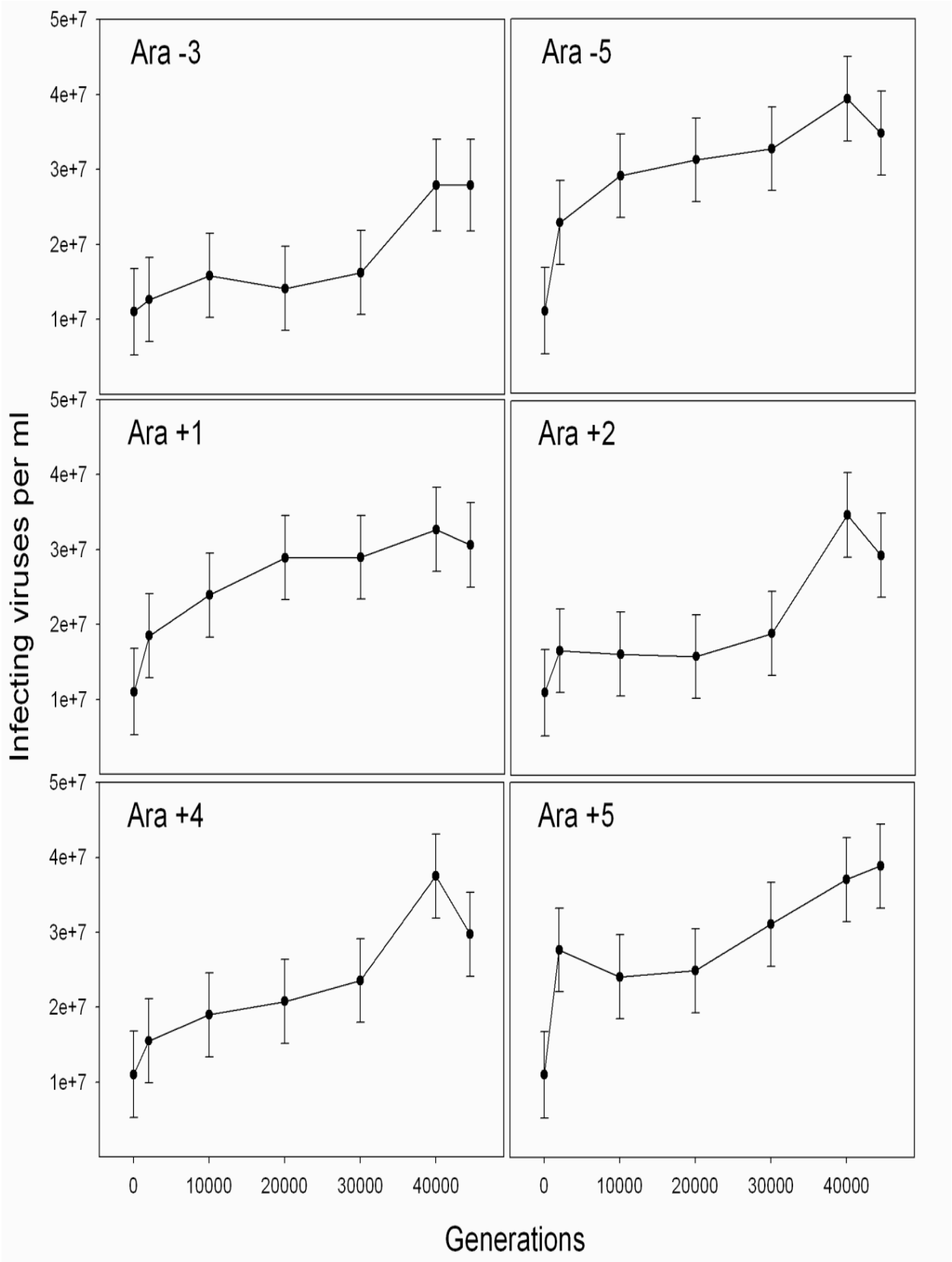


Figure 1.2 (caption on following page)

Figure 1.2. Assays of six evolving *E. coli* populations for susceptibility to infection by phage T6*. Three clones were tested at each time point for each population, except at generation 0, where assays for the same six ancestral clones are shown in all panels. The Y-axis shows counts of infecting viruses per ml based on plaque-forming units. These counts were all obtained using the same phage stock, and hence the differences across bacterial samples reflect differences in phage infectivity on those bacteria. Each point shows the least-squares mean and standard error based on the statistical analysis summarized in Table 1.1.

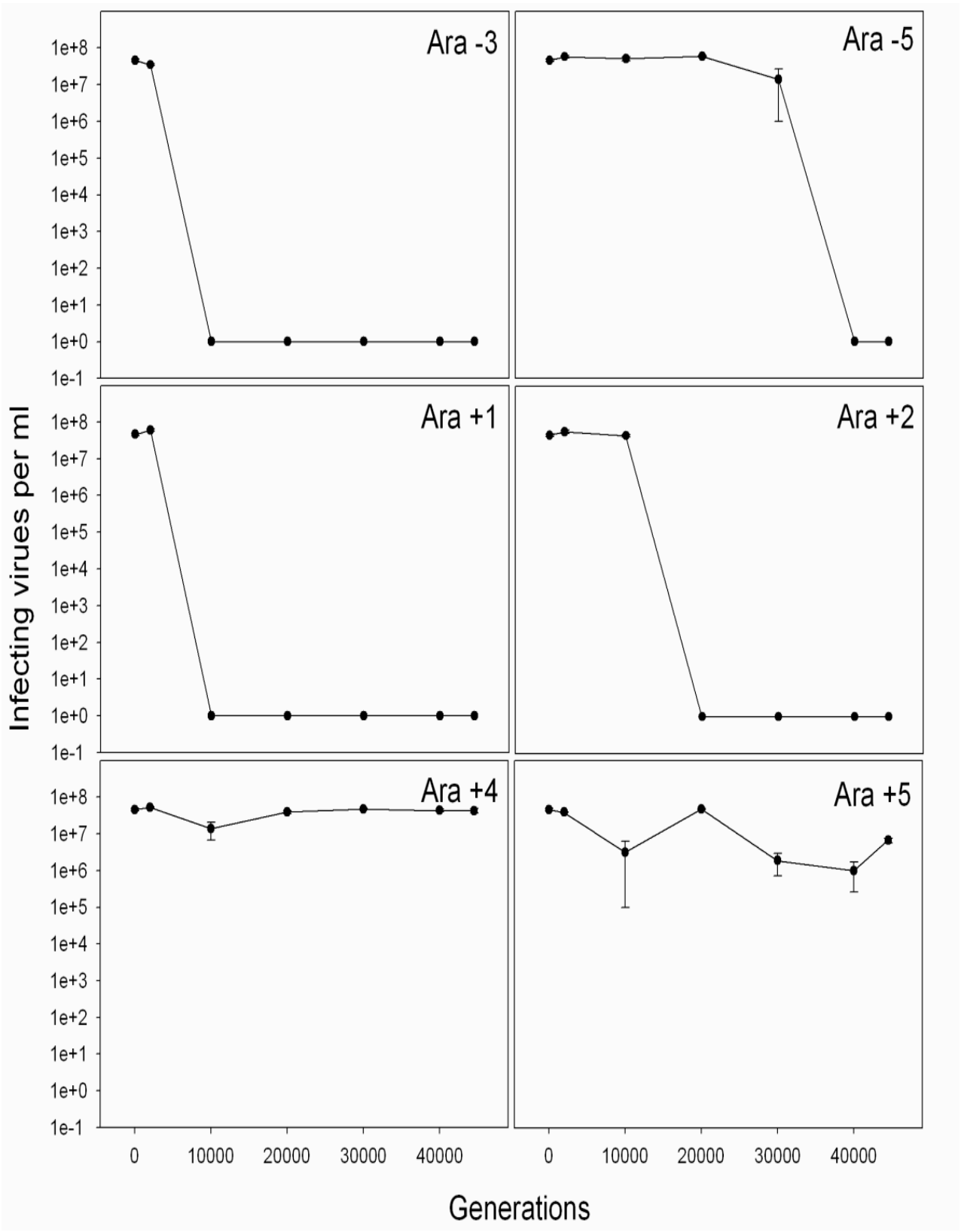


Figure 1.3 (caption on following page)

Figure 1.3. Assays of six evolving *E. coli* populations for susceptibility to infection by phage λ. Three clones were tested at each time point for each population, except at generation 0, where assays for the same six ancestral clones are shown in all panels. Each point shows the least-squares mean and standard error based on the statistical analysis summarized in Table 1.1. All means were transformed by adding one in order to plot the zero values on a log scale. Y-axis measures the same quantity as defined in Figure 1.2, but note the difference in scaling. When error bars are not visible, they are smaller than the symbols.

LITERATURE CITED

LITERATURE CITED

1. Ehrlich, P. R., and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586-608.

2. Van Valen, L. 1973. A new evolutionary law. Evol. Theory 1:1-30.

3. May, R. M., and R. M. Anderson. 1983. Epidemiology and genetics in the coevolution of parasites and hosts. Proc. R. Soc. Lond. B 219:281-313.

4. Gillespie, J. H. 1975. Natural selection for resistance to epidemics. Ecology 65:493-495.

5. Brodie, E. D. 1991. Evolutionary response of predators to dangerous prey: reduction of toxicity of newts and resistance of garter snakes in island populations. Evolution 45:221-224.

6. Bowen, L., and D. Van Vuren. 1997. Insular endemic plants lack defenses against herbivores. Cons. Biol. 11:1249-1254.

7. Stastny, M., U. Schaffner, and E. Elle. 2005. Do vigour of introduced populations and escape from specialist herbivores contribute to invasiveness? J. Ecol. 93:27-37.

8. Kato, T., K. Ishida, and H. Sato. 2008. The evolution of nettle resistance to heavy deer browsing. Ecol. Res. 23:339-345.

9. Stahl, E. A., and J. G. Bishop. 2000. Plant-pathogen arms races at the molecular level. Curr. Opin. Plant Biol. 3:299-304.

10. Simms, E. L. 1992. Costs of plant resistance to herbivores. Pp. 392-425 in R. S. Fritz, and E. L. Simms, eds. Plant Resistance to Herbivores and Pathogens. Univ. Chicago Press, Chicago.

11. Stowe, K. A. 1998. Experimental evolution of resistance in Brassica rapa: correlated response of tolerance in lines selected for glucosinolate content. Evolution 52:703-712.

12. Shonle, I., and J. Bergelson. 2000. Evolutionary ecology of the tropane alkaloids of Datura stramonium L. (Solanaceae). Evolution 54:778-788.

13. Abedon, S. T. 2009. Phage evolution and ecology. Adv. Appl. Microbiol. 67:1-45.

14. Yoshida, T., N. G. Hairston, and S. P. Ellner. 2004. Evolutionary trade-off between defense against grazing and competitive ability in a simple unicellular alga, Chlorelia vulgaris. Proc. R. Soc. Lond. B 271:1947-1953.

15. Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. Hairston. 2003. Rapid evolution drives ecological dynamics in a predator-prey system. Nature 424:303-306.

16. Meyer, J. R., S. P. Ellner, N. G. Hairston, L. E. Jones, and T. Yoshida. 2006. Prey evolution on the time scale of predator-prey dynamics revealed by allele-specific quantitative PCR. Proc. Natl. Acad. Sci. USA 103:10690-10695.

17. Poullain, V., S. Gandon, M. A. Brockhurst, A. Buckling, and M. E. Hochberg. 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62:1-11.

18. Escobar-Paramo, P., N. Faivre, A. Buckling, C. Gougat-Barbera, and M. E. Hochberg. 2009. Persistence of costly novel genes in the absence of positive selection. J. Evol. Biol. 22:536-543.

19. Scanlan, P. D., A.R. Hall, L. D. C. Lopez-Pascua *et al*. 2011 Genetic basis of infectivity evolution in a bacteriophage. Molec Ecol. 20:981-989

20. Lythgoe, K.A. and L. Chao 2003. Mechanisms of coexistence of a bacteria and a bacteriophage in a spatially homogeneous environment Ecol. Lett. 6:326-334.

21. Buckling, A., Y. Wei, R. C. Massey, M. A. Brockhurst, and M. E. Hochberg. 2006. Antagonistic coevolution with parasites increases the cost of host deleterious mutations. Proc. R. Soc. B 273:45-49.

22. Bohannan, B. J. M., and R. E. Lenski. 2000a. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Lett. 3:362-377.

23. Lenski, R. E. 1988a. Dynamics of interactions between bacteria and virulent bacteriophage. Adv. Microb. Ecol. 10:1-44.

24. Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315-1341.

25. Lenski, R. E., and M. Travisano. 1994. Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. USA 91:6808-6814.

26. Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving Escherichia coli populations. Nature 407:736-739.

27. Schwartz, M. 1987. The maltose operon. Pp. 1482-1502 in F. C. Neidhardt, ed. Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. American Society for Microbiology, Washington, DC.

28. Wang, Y. F., R. Dutzler, P. J. Rizkallah, J. P. Rosenbusch, and T. Schirmer. 1997. Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. J. Mol. Biol. 272:56-63.

29. Pelosi, L., L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173:1851-1869.

30. Lenski, R. E. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium Escherichia coli. Plant Breeding Rev. 24:225-265.

31. Lederberg, S. 1966. Genetics of host-controlled restriction and modification of deoxyribonucleic acid in Escherichia coli. J. Bacteriol. 91:1029-1036.

32. Daegelen, P., F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim. 2009. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J. Mol. Biol. 394:634-643.

33. Manning, P. A., and P. Reeves. 1976. Outer membrane of Escherichia coli K-12: tsx mutants (resistant to bacteriophage T6 and colicin k) lack an outer membrane-protein. Biochem. Biophys. Res. Comm. 71:466-471.

34. Bremer, E., A. Middendorf, J. Martinussen, and P. Valentinhansen. 1990. Analysis of the tsx gene, which encodes a nucleoside-specific channel-forming protein (Tsx) in the outer-membrane of Escherichia coli. Gene 96:59-65.

35. Schneider, H., H. Fsihi, B. Kottwitz, B. Mygind, and E. Bremer. 1993. Identification of a segment of the Escherichia coli Tsx protein that functions as a bacteriophage receptor area. J. Bacteriol. 175:2809-2817.

36. Luria, S. E. 1945. Mutations of bacterial viruses affecting their host range. Genetics 30:84–99.

37. Adams, M. H. 1959. Bacteriophages. Interscience, New York.

39. Loś, J. M., P. Golec, G. Wegrzyn, A. Wegrzyn, and M. Loś. 2008. Simple plating Escherichia coli bacteriophages forming very small plaques or no plaques method for under standard conditions. Appl. Environ. Microbiol. 16:5113-20.

40. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, NY.

41. Moyed, H. S., T. T. Nguyen, and K. P. Bertrand. 1983. Multicopy Tn10 tet plasmids confer sensitivity to induction of tet gene expression. J. Bacteriol. 155:549-556.

42. Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS System for Mixed Models. SAS Institute, Cary, NC.

43. Herring, C. D., J. D. Glasner and F. R. Blattner. 2003. Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. Gene 311:153–163.

44. Sleight S. C., C. Orlic, D. Schneider, and R. E. Lenski. 2008. Genetic basis of evolutionary adaptation by Escherichia coli to stressful cycles of freezing, thawing and growth. Genetics180:431-443.

45. Dykhuizen, D. E., and D. L. Hartl. 1983. Selection in chemostats. Microbiol. Rev. 47:150-168.

46. Lenski, R. E., and B. R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. 125:585-602.

47. Braun, V. 2009. FhuA (TonA), the career of a protein. J. Bacteriol. 191: 3431-3436.

48. Francis, G., L. Brennan, S. Stretton, and T. Ferenci. 1991. Genetic mapping of starch- and Lambda-receptor sites in maltoporin: identification of substitutions causing direct and indirect effects on binding sites by cycsteine mutagenesis. Mol. Microbiol. 5:2293-2301.

49. Simms, E. L. and M. D. Rausher. 1989. The evolution of resistance to herbivory in Ipomoea purpurea. II. Natural selection by insects and costs of resistance. Evolution 43:573-585.

50. Rausher, M. D., K. Iwao, E. L. Simms, N. Ohsaki and D. Hall. 1993. Induction and the cost of resistance in Ipomoea purpurea. Ecology 74:20-29.

51. Simpson, G. G. 1953. The Major Features of Evolution. Columbia Univ. Press, New York.

52. Harvey, P. H., and M. Pagel. 1991. The Comparative Method in Evolutionary Biology. Oxford Univ. Press, Oxford.

53. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

54. Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodriguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115-2118.

55. Crandall, K. A., C. R. Kelsey, H. Imamichi, H. C. Lane and N. P. Salzman. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol. Biol. Evol. 16:372-382.

56. Wichman, H. A., M. R. Badgett, L. A. Scott, C.M. Boulianne, and J. J. Bull. 1999. Different trajectories of parallel evolution during viral adaptation. Science 285:422-424.

57. Huey, R. B., G. W. Gilchrist, M. L. Carlson, D. Berrigan, and L. Serra. 2000. Rapid evolution of a geographic cline in size in an introduced fly. Science 287:308-309.

58. Schluter, D. 2004. Parallel evolution and inheritance of quantitative traits. Am. Nat. 163:809-822.

59. Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski. 2006. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proc. Natl. Acad. Sci. USA 103:9107-9112.

60. Lenski, R. E. 1984. Two-step resistance by Escherichia coli B to bacteriophage T2. Genetics 107:1-7.

61. Bohannan, B. J. M., and R. E. Lenski. 2000b. The relative importance of competition and predation varies with productivity in a model community. Am. Nat. 156:329-340.

62. Crozat, E., N. Philippe, R. E. Lenski, J. Geiselmann, and D. Schneider. 2005. Long-term experimental evolution in Escherichia coli. XII. DNA topology as a key target of selection. Genetics 169:523-532.

63. Lenski, R. E. 1988b. Experimental studies of pleiotropy and epistasis in Escherichia coli: I. Variation in competitive fitness among mutants resistant to virus T4. Evolution 42:425-432.

64. Schrag, S. J., V. Perrot, and B. R. Levin. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proc. Soc. Lond. B 264:1287-1291.

65. Moore, F. B. G., D. E. Rozen, and R. E. Lenski. 2000. Pervasive compensatory adaptation in Escherichia coli. Proc. R. Soc. Lond. B 267:515-522.

66. Muller, H. J. 1932. Some genetic aspects of sex. Am. Nat. 8:118-138.

67. Gerrish, P. J., and R. E. Lenski. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102/103:127-144.

CHAPTER 2

INCREASED COSTS FOR VIRAL RESISTANCE INDUCE A COEVOLUTIONARY ARMS RACE BETWEEN *ESCHERICHIA COLI* AND PHAGE λ.

Authors: Justin R. Meyer and Richard E. Lenski

Abstract

Trade-offs between disease resistance and competitive fitness are thought to influence many ecological and evolutionary processes, including coevolution between hosts and parasites. We experimentally manipulated the presence of a trade-off between viral resistance and host competitiveness in the bacterium (*Escherichia coli*) and examined how its coevolution with the virus, bacteriophage λ, was altered. *E. coli* and λ were co-cultured in three resource environments chosen because each differed in the pressure exerted on the bacterium to maintain the protein target of the virus, LamB. This constraint increased the likelihood that mutations for defense would be costly because λ-resistance typically evolves by mutations that modify expression or structure of LamB. By altering the presence of a trade-off, the resource environment affected a number of ecological and evolutionary processes, including phage population dynamics, the genetics of resistance, whether the host and parasite engaged in an arms race, and bacterial and viral diversification. This study provides explicit experimental evidence that the presence of trade-offs will affect coevolution and many other ecological and evolutionary dynamics between hosts and parasites.

Introduction

Antagonistic coevolution between parasites and their hosts is pervasive and impacts almost every species. By promoting evolutionary divergence (1), antagonistic coevolution can have a number of important effects, some of which include shaping the assembly of ecological communities (2), promoting diversification and speciation (3), and influencing macroevolutionary patterns observed in phylogenies of distantly related species (4). Some researchers have even hypothesized that the rapid evolution caused by coevolution could help explain classic biological paradoxes, including the prevalence of sexual reproduction (5), why phytoplankton communities are highly diverse (6), and the latitudinal gradient of diversity (7).

Despite its prevalence, it should not be assumed that coevolution always occurs, recent studies have revealed whether coevolution occurs between single species pair, can vary by environment, region, or in time (8-10). Considering the importance of coevolution, understanding what drives these differences is crucial. One of many possible explanations is that environmentally-determined differences in the costs for resistance (as observed by Jessup and Bohannan 2008) could alter coevolution. For example, if resistance to parasites is does not confer a cost, then hosts will quickly evolve complete resistance that the parasite cannot counter, ending any potential arms race. On the other hand, if resistance is expensive, then hosts will finely tune their level of resistance to avoid paying a cost without a benefit. As parasites evolve counter-defenses, the hosts will follow, causing an arms race (12).

There is circumstantial evidence to support a role for trade-offs in producing coevolutionary cycles. Costs are often observed in systems of coevolving species, from microbes to plants and animals (11, 13-15). Evolutionary theory has shown that costs expand the

range of conditions in which coevolution occurs (6). And experiments on the coevolution of bacteria and their viruses (phage) revealed that *Escherichia coli* engaged in arms races with phage species in which resistance is costly, but not with a phage for which resistance is free (12).

Here we present direct experimental evidence for the role of trade-offs in promoting coevolution. To do this, we manipulated the culture environment of a parasite, bacteriophage λ, and its bacterial host, *E. coli*, in order to augment the penalty the host experienced for evolving resistance, and then monitored their coevolution. Three environments were chosen that imposed increasing degrees of reliance on the bacterial protein LamB, the extracellular receptor of λ. LamB creates a pore in the outer-membrane that *E. coli* uses to transport maltodextrins, starches made of glucose molecules joined together in a linear chain (13). Previous studies have shown that the majority of λ-resistance mutations alter expression or the structure of LamB (14). Therefore, by culturing *E. coli* in environments where the carbon source varied in chain length, we were able to manipulate the presence and strength of trade-offs between resistance and resource uptake, while keeping all else equal.

We co-cultured *E. coli* and λ in minimal medium with only one carbon source; either glucose, maltose (two glucose molecules linked), or maltotriose (three molecules). Each sugar is commonly available to *E. coli* living in intestines, however the conditions are artificial because normally *E. coli* would encounter all three sugars simultaneously, as well as many more carbon sources. We chose this design to magnify differences in the resource environments that represent extremes of more subtle environmental variations that exist in nature. Trade-offs between resistance and competitiveness should not exist in the glucose-only condition because glucose is transported by OmpF (15), not LamB, making the receptor unnecessary for reproduction in the first environment. Maltose crosses the outer-membrane mostly through

LamB, although it can be transported by other porins, but less efficiently (16). Maltotriose's more cumbersome chain makes it even more reliant on LamB. We predicted that an arms race will be favored as the constraint and thus trade-off was magnified with each additional glucose segment.

Along with promoting an arms race, we also predicted that maltodextrins would alter the form of the race. Past studies on λ and *E. coli* showed their coevolution was primarily driven by directional selection for increased resistance and broader host-range (17-19). This escalation is known as gene-for-gene or extended host-range coevolution and is distinct from other forms, namely the matching alleles model (Figure 2.1) (20). Others have predicted that *E. coli* and λ should experience diversifying selection in environments with maltodextrins and undergo matching alleles coevolution (6,8, 12). In these environments, *E. coli* should evolve resistance through mutations that have only subtle effects on the structure of LamB in order to balance resistance to phage, yet maintain channel function. Structural changes in LamB would likely be countered by mutations that cause slight, yet coordinated alterations in the phage's J protein, the molecule λ uses to attach to LamB (21, 22). Coordinated, or *matching* changes would promote the evolution of highly specialized resistance in the *E. coli* and narrow host-ranges for the phage, similar to evolving *locks and keys,* and unlike the broad host-ranges and full-spectrum resistance evolved under the former model.

Gomez and Buckling (2011) presented anecdotal evidence that more restrictive environments will promote more specialized bacterial-phage interactions to evolve. They studied the coevolution of *Pseudomonas fluorescens* and one of its phages, SBW25φ2, which were previously shown to coevolve according to the extended host-range model in rich medium

(24), however Gomez and Buckling found specialized interactions evolved in a depauperate soil environment. In line with our hypothesis, they suggested that differences in resource-mediated effects on trade-offs likely drove the divergent forms of coevolution, although this could not be confirmed because many other factors varied between the experiments, such as spatial structure, population size, and chemistry that could also alter coevolution. A well-controlled experiment, such as ours, where only resources are manipulated is required to test this hypothesis.

The goals of our study were two-fold. First, we examined whether trade-offs between resistance and competitiveness would alter coevolutionary dynamics, and secondly, in light of predictions that coevolution can impact other ecological and evolutionary phenomenon, we studied how the resource-mediated differences in coevolution altered other ecological characteristics of the host-parasite community. To do this, we manipulated the presence and strength of costs for resistance by careful alteration of the resource environment and determined whether coevolution was affected. Differences observed between coevolution were also studied at the molecular level by sequencing candidate genes. Lastly, we monitored key ecological variables such as population dynamics, extinction, community network architecture, and genotypic diversity in order to test hypotheses for how coevolution alters other ecological processes.

Methods

Strains

We studied *E. coli* B strains REL606 and REL607 because their evolution in the lab has been well documented (25) and they readily coevolve with phage because they do not possess

generalized immunity through restriction modification systems, a cRISPR adaptive immune system, or *mucoid* cell defense (26, 27). The difference between REL606 and REL607 is a single mutation that inhibits REL606 from metabolizing arabinose (*ara*$^{-}$), which can be used as a neutral genetic marker (28, 29). The phage strain studied was cI26, an obligatorily lytic mutant provided by Donald Court (National Cancer Institute). cI26 has a single nucleotide deletion in the cI repressor gene responsible for signaling a phage lifecycle switch to lysogeny (30, 31).

Experimental conditions and procedure

λ and *E. coli* were co-cultured in modified Davis Medium (DM) (26) with 10x the magnesium sulfate (1 μg ml^{-1}) and 125 μg ml^{-1} of a single sugar; glucose, maltose, or maltotriose. Sugar mass was held constant between treatments to maintain the same bacterial carrying capacities between them. Six replicate flasks were created for each sugar and all flasks were inoculated with 100 λ particles and 1,000 *E. coli* cells grown from isogenic stocks that were preconditioned for a day in their experimental environments. Half of the flasks were inoculated with REL606 and the others with REL607 so that cross-contamination of bacteria between flasks would be more easily recognized than if all flasks had identical bacteria. Small initial population sizes were used to guarantee that mutations for resistance and host-range expansions evolved *de novo*. This procedure ensures that each replicate evolved independently, which is critical for assessing parallel evolution and local adaptation of hosts and parasites. After 24 hours of growth at 37 °C and shaking at 120 rpm, a random sample of 10^{-2} of the mixed community was transferred to a fresh flask and allowed to grow. At the end of each 24-hour growth period, two 1 ml samples of the mixed community were preserved with 15% glycerol and freezing at -80 °C. Additionally, bacterial and phage densities were determined by plating on LB (Luira-Bertani)

agar Petri dishes for bacteria (32) or by counting phage plaques on lawns of the bacterial ancestor (33). Serial transfers and sampling was repeated for 25 days.

E. coli genotypes were periodically isolated by picking a single colony from the plates described above, streaking them on a new LB Petri dish, picking another single colony after a day's growth, and repeating once more to eliminate possible phage contamination. A single colony was selected from the second streaked plate and grown in liquid LB (32) overnight, then a sample was created by freezing 1 ml in glycerol as described above. Phage clones were isolated by picking single plaques and culturing them on REL606 as described in Adams 1959. Phage isolates were stored as liquids with 4% chloroform added in 4 °C.

Competition experiments to determine costs for resistance

To determine whether maltodextrins encouraged the evolution of more costly resistance we competed a single resistant genotype isolated from each community against the ancestor with the opposite *ara* marker. The resistant clones were removed from day three of the experiment. We chose this early time-point because preliminary experiments showed that resistance had already evolved by then, and it was too early for the bacteria to acquire additional mutations that ameliorated the cost of resistance (compensatory mutations) to evolve. These mutations would hide the true costs for resistance. Competition trials were performed by inoculating each flask with 1.25×10^6 preconditioned cells of an experimental isolate and the ancestor, allowing them to grow and compete, and judging their relative fitness by the ratio of their growth rates calculated as their Malthusian parameters. For full details see (28, 29). Fitness was measured for each clone in the environment in which they evolved, over three days of culturing, and with three replicate competitions. To determine whether the fitness of genotypes varied by

environment we preformed an ANOVA using JMP version 7. Significant treatment differences were tested using a Tukey's test. Differences in the variance in fitness were analyzed with F-tests to examine whether the genetic variance in fitness increased when trade-offs were present by using the open access program PAST (http://folk.uio.no/ohammer/past/).

Bacterial resistance and phage host-range

Bacterial sensitivity to phage was determined by 'spot' assays (33), where bacteria are immobilized in a matrix of soft agar (LB with half the agar) and phage are dripped on top. If the phage strain is able to infect the given bacterial strain, then a clearing (spot) is observed under the drip. For each drip, we recorded whether there was no clearing, some clearing, or complete clearing. Each assay was replicated four times.

Sequencing

Genetic changes responsible for coevolution were identified by sequencing *lamB* (phage receptor) and *malT*, a positive regulator of *lamB* (34), from bacterial clones isolated throughout the experiment, as well as the gene for the phage ligand J in λ isolates. We used DNA fragments amplified by polymerase chain reaction as templates for sequencing. The fragments were first purified using GFX columns and sequenced using an automated ABI sequencer run by Michigan State University's Research Technology Support Facility. The primers used to amplify fragments containing *lamB* were amplified with oligo sequences 5' TTCCCGGTAATGTGGAGATGC 3' and 5' AATGTTTGCCGGGACGCTGTA 3', positioned 1,398 bases up and 504 bases downstream, *malT* was sequenced using primers 5' CACCGGTTTGGCGAATGG 3' and 5' GCGGCGGTGGGGGAATA 3' designed 424 bases up stream and 212 bases downstream, and primers for J were 5' CTGCGGGVGGTTTTGTCATT 3'

and 5' ACGTATCCTCCCCGGTCATCACT 3' placed 15 bases up and 318 bases downstream of the gene.

Dynamics of sensitive and resistant genotypes

The only environment in which we found genetic variation for resistance was in the maltotriose environment. In order to follow the genotypic dynamics of resistance in this environment we constructed new communities of genetically marked sensitive (*ara*$^{+}$, REL607) and resistant (*ara* −) bacteria, and ancestral phage cI26. *ara* marked bacteria allowed for efficient quantification of resistant and sensitive genotypes and more robust estimates of their population sizes. Three flasks were inoculated with a completely resistant genotype isolated from the 25th day of the maltotriose community labeled '-2' and another three flasks were initiated with a completely resistant genotype from '-3'. We allowed them to coevolve under the maltotriose conditions and took daily samples of the phage densities, total bacterial populations (number of colonies growing on LB agar plates), sensitive cell densities (colonies on minimal arabinose agar as in Lenski 1991), and the density of resistant bacteria was estimated as the difference in colonies counted on the LB and minimal arabinose agar. Each community was propagated for 15 days or until λ became extinct.

Results

Costs for resistance

Costs for resistance were on average higher in the maltodextrin environments (Figure 2.2), although these difference were not found to be significantly different (ANOVA $F = 3.532$, $p = 0.0553$, $DF = 15$), none of the environments were significantly different than one another by the

Tukey's test with an $\alpha = 0.05$ cutoff. Instead, genotypic variance in the costs were significantly larger in maltotriose (1.96×10^{-2}) than maltose (1.84×10^{-3}) or glucose (8.64×10^{-4}) (F-test maltotriose versus maltose: $F = 10.673$, $p = 0.0212$, $DF = 10$; or maltose versus glucose: $F = 2.125$, $p = 0.427$, $DF = 10$). Variance in fitness should increase because environments that encourage trade-offs should promote the emergence of genotypes with a range of values for resistance and fitness that fall at different points along the relationship between cost and defense.

Population dynamics

Phage population dynamics varied between environments, although bacterial densities did not (Figure 2.3). Phage cultured in glucose reproduced slowly for the first two days, then boomed and crashed. Their decline slowed and population densities stabilized at around 10^4 particles per ml. Once the populations entered this stage three remained at low densities and the others went extinct. We confirmed their extinctions with qPCR (35). Unlike the retarded growth of phage on hosts consuming glucose, phage in the maltodextrin treatments grew very quickly, experienced erratic population fluctuations reminiscent of coevolutionary cycles, and their densities were on average much higher than in glucose. In line with the prediction that resistance should be harder to evolve in maltotriose than maltose, the phage population was on average more dense in maltotriose than in maltose (of the populations that survived 25 days, maltose: average density = $8.10 \times 10^6 \pm 3.42 \times 10^6$ SD particles per ml; maltotriose: average density = $1.17 \times 10^7 \pm 1.08 \times 10^7$ SD particles per ml), although the differences were not significant (t-test; $t=0.528$, $df=5$, $p=0.62$). Phage populations cultured in maltodextrins also suffered extinctions (one half in maltose, and one third in maltotriose).

Bacterial-phage infection networks

To examine the pattern of infection produced through coevolution, we isolated phage and bacteria from each flask of the experiment, tested whether they could infect one another, and constructed a bacterial-phage interaction network. Two phage and two bacterial genotypes were isolated from the third day of coevolution in all 18 flasks. We focused on the third day because preliminary experiments revealed that diversity peaked then, therefore samples from this time point would provide a more complete picture of the diversity of interactions. Next, we constructed an interaction matrix between the bacteria and phage (37 by 37 interactions including the ancestors; Figure 2.4). Lastly, we used the patterns in the matrix to determine treatment differences in coevolution and community network structure.

Coevolution did not occur in the glucose environment, instead, bacteria acquired complete resistance to all phage genotypes and the phage did not evolve new host-ranges over this short time-scale (Figure 2.4). Similar to the glucose populations, all bacteria isolated from the maltose populations were resistant to all phage, however phage in maltose evolved host-range expansions to include resistant genotypes sampled from maltotriose. These phage likely adapted to include bacterial genotypes in their range that once existed in maltose, but were not sampled. Coevolution clearly occurred in the maltotriose treatment; three bacterial genotypes evolved higher levels of resistance, and two new phage genotypes evolved counter-defenses by broadening their host-ranges to included defended bacteria. Unlike our prediction that λ and *E. coli* would experience matching alleles coevolution, the broad host-ranges and resistance observed, combined with the nested pattern of host-ranges support the extended host-range model. Some of the most pronounced differences between the treatments are captured in the ecological network statistics, number unique host genotypes sampled (glucose = 1, maltose = 1,

and maltotriose = 4), parasite diversity (glucose = 1, maltose = 2, and maltotriose = 3), and the richness of the interactions (i.e. connectance = number of successful infections / total possible interactions; glucose = 0, maltose = 0, and maltotriose = 0.513).

Mutation identification

To determine the bacterial mutations responsible for the differences in coevolution we sequenced the *lamB* and *malT* genes from bacterial clones isolated on days 3, 10, and 25 from three separate communities of each environment. We were unable to sample all six replicates because only three λ populations survived for 25 days in the glucose and maltose treatments (Figure 2.4). Our results were striking, there were clear molecular differences between treatments: Resistance in glucose most often occurred by changes in the *malT* gene, a positive regulator of *lamB* (Table 2.1). In contrast, resistance typically evolved through changes directly in *lamB* in the maltodextrin environments. The *lamB* mutations evolved in maltose were nonsense mutations that eliminate function of LamB, an analogous phenotypic effect as the loss of LamB expression in glucose, whereas the majority of mutations isolated from maltotriose were missense mutations (6 out of 8) that have more subtle effects on the amino acid sequence and resistance. This difference in maltose and maltotriose supports the prediction that the sugars represent a gradient of constraints rather than a simple bifurcation between glucose and maltodextrins. Lastly, supporting the finding that an arms race occurred in the maltotriose treatment is the observation that two maltotriose isolates contained multiple mutations in *lamB* (Table 2.1).

Even though all bacterial strains were resistant to ancestral λ, some did not posses mutations in *lamB* or *malT*. Mutations in these strains likely occurred in regulatory regions of *lamB*, or other genes that affect λ reproduction (14, 31). Unlike the findings of other studies on

the genetics of λ host-range evolution (17, 31, 37), mutations in the gene for J were not responsible, at least in the 10 independantly evolved λ we sequenced. Evolved λ must have acquired mutations in other loci, likely other tail proteins that may also affect host-specificity.

Genotypic dynamics

Mathematical models of extended host-range coevolution predict low diversity communities will evolve because bacteria with the greatest resistance and phage with the broadest host-range should exclude other genotypes (20). However, when trade-offs occur between resistance and reproduction, then genotypic diversity can be maintained. We tested whether diversity was maintained in the maltotriose environment in line with trade-off theory by tracking genetically marked sensitive and resistant genotypes of bacteria. We found that the resistant genotypes quickly gained high density, although the sensitive populations remained relatively low and fluctuated with the rise and fall of phage populations (Figure 2.5). The dynamics reveal an example of cryptic predator-prey cycles where total host densities do not appear to change, yet parasite densities do, as well as a minority population of sensitive hosts. The term cryptic is used to describe the masking of the sensitive host dynamics by the dominant resistant genotype's superior abundance. Cryptic cycling was observed in other bacterial-phage coevolution experiments (38).

Fearing that more interesting coevolutionary dynamics were occurring that captured in Figure 2.5, we examined whether the sensitive bacteria had evolved resistance and whether the phage had gained broader host ranges. To do this, we examined the resistance and host-range profiles over time for one of the communities (Figure 2.5b). First, we isolated six genotypes of phage and three genotypes of *ara*$^{+}$ bacteria from each day of the experiment and determined their

host-ranges and sensitivity, respectively (Figure 2.6a and b). With this course sampling, we constructed a rough time-series of host-range and resistance evolution (Figure 2.6c). We found that the bacteria evolved increased resistance and according to trade-off theory, less defended host genotypes were maintained. Phage also evolved broader host-ranges, however the phage with the broadest range did not exclude the more specialized λ genotypes (Figure 2.6). Notably, in line with the prediction of an arms-race, the time-series of phage host-ranges increased stepwise.

Discussion

By controlling the length of the carbon molecule supplied to *E. coli* we altered its interactions and coevolution with one of its viruses, λ. This manipulation affected the likelihood that mutations for bacterial resistance would confer a pleiotropic cost, which caused the bacteria to evolve different sets of resistance mutations, some of which triggered a coevolutionary arms race and others that prohibited it from happening. The evolutionary path that each community took also had profound effects on ecological variables such as phage population dynamics, community structure, and diversity.

Glucose: *Ecological and coevolutionary dynamics*

After the initial inoculation, bacterial densities quickly rose to reach their peak, however the phage took a number of days to reach their maximum (Figure 2.3). This delay likely occurred despite the high abundance of sensitive hosts because bacteria limit their expression of LamB when glucose is present, making it difficult for λ particles to find a receptive cell. Once the phage became abundant, they applied a strong selective pressure on the bacterial population,

which responded by evolving resistance through mutations in *malT* (Table 2.1). This mutation rapidly evolved because it occurs at a high rate (14), confers broad resistance (31), and can provide a slight cost saving benefit when glucose is the only source of carbon (39). When the $malT^-$ allele fixes, the phage population suffers, but does not decline to extinction (Figure 2.3). A previous study showed that λ persists despite high levels of bacterial resistance because rare $malT^-$ cells stochastically express LamB, providing a small subpopulation of receptive hosts for λ to infect (31). At this stage, the phage does not evolve a counter-defense and an arms race does not occur (Figure 2.4). In the aforementioned study λ does evolve a counter-defense, however the λ in the current study does not because of its smaller population size (~10^4 per ml versus 10^6-10^7 per ml) prohibits adaptive evolution over the short time-scales of the experiment.

Maltotriose: *Ecological and coevolutionary dynamics*

Unlike the glucose environment, phage grew quickly on hosts supplied with maltotriose, likely because their receptor is highly expressed in *E. coli* when maltodextrins are present (Figure 2.3) (34). This large population of λ imposes strong selection for *E. coli* to evolve resistance, which it does through missense mutations in *lamB* (Table 2.1). *lamB*, instead of *malT* mutations evolve because *malT* mutations are fatal when maltodextrins are the only source of carbon (39). Next, phage evolve counter-defenses by expanding their host range to include the new forms of LamB (Figure 2.4). The bacteria counter again with additional mutations in LamB (Table 2.1). These cycles occurred multiple times generating a range of bacterial and phage genotypes. Even though completely resistant bacteria evolve, they do not fix (Figure 2.6) because of trade-offs between resistance and reproduction (Figure 2.2). Likewise, the phage genotype with the

broadest host-range also does not fix (Figure 2.6). Once this diversity is generated, selection on the phage seems to shift from promoting host-range expansions capable of generating new phage genotypes, to fluctuating selection on the already evolved λ genotypes. This dynamic was observed in another study on the coevolutionary dynamics between *Pseudomonas fluorescens* and its phage SBW25φ2 and seems to occur because extreme generalists pay too high of cost to out-compete phage with more specialized host ranges (40).

Maltose: *Ecological and coevolutionary dynamics*

Host-parasite dynamics in maltose, the molecule of intermediate length between glucose and maltotriose, shared characteristics of both treatments. Population dynamics, the gene target of resistance mutations, and phage evolution were more similar to maltotriose than glucose. However, the evolution of complete resistance, low genotypic diversity, and sparse interaction network resemble glucose. Hybrid dynamics in the maltose treatment reinforce the main thesis of this study that increases in sugar chain length constrain host evolution and alter patterns of coevolution.

Form of coevolution

Theory on λ-*E. coli* coevolution predicted that they would experience matching alleles coevolution when cultured in a minimal maltodextrin environment (6, 8, 12). This prediction was loosely based on the biophysics of their interactions. LamB shape was predicted to mutate and λ would counter with synchronized changes in its ligand J, creating a series of coevolved locks and keys. We found no support for this prediction, instead we uncovered a three-step extended host-range form. This type of coevolution was previously observed between a different strain of λ (wt or *PaPa*) and *E. coli* (K12) evolving in rich medium (17-19,41). Likewise, this

nested pattern is the most common motif observed in bacterial phage interaction matrices (42). While we have shown that the environment controlled whether coevolution occurred, we did not find evidence that environmentally-induced constraints could change the form of coevolution (ie. matching alleles versus extended host-range).

Interaction network varies by environment

Recent developments in network analysis have inspired new methods for analyzing ecological networks such as food webs (43). As more and more networks are measured and analyzed, it is clear that ecological communities demonstrate a wide spectrum of network structures (44). This spectrum is even observed when comparing communities of bacteria and phage (42). A growing question within the microbial and ecological literature is: What drives these differences? A similar coevolutionary study to ours showed that the diversity of genotypes and richness of interactions (connectedness) vary by productivity (42) which was confirmed for natural phage communities as well (45). The effect of productivity is thought to be mediated by its influence on costs for resistance. Our results reinforce this interpretation; increased costs for resistance yield much more diverse and connected matrices (Figure 2.4).

Biodiversity

Antagonistic coevolution is thought to encourage genotypic and species diversification by driving rapid divergence of hosts and parasites (3) and is even predicted to drive diversification (46). Likewise, the most prominent pattern of biodiversity, the latitudinal gradient of global biodiversity has been hypothesized to result from increasing importance of coevolution in tropical versus temperate zones (7). Our results demonstrate that coevolution can drive rapid gains in genotypic diversity (Figure 2.6). This study, along with other microbial experiments

have shown the diversifying effect of enemy-victim relationships (47) and antagonistic coevolution (24, 36, 49).

Conclusion

By making inferences about coevolution based on the knowledge of which gene products interact in the binding of a parasite to its host, along with information on the function and regulation of the host receptor, we were able to construct lab conditions that varied only slightly in their chemistries, but had profound effects on the ecology and evolution of a host and parasite. This study, along with a recent review demonstrate the predictive power of including molecular mechanisms for trade-offs in studies on ecology and evolution (50). With this approach, we showed how the potential for trade-offs between host resistance and reproduction can induce a coevolutionary arms race, which in turn had broad effects on the ecology of the host-parasite communities.

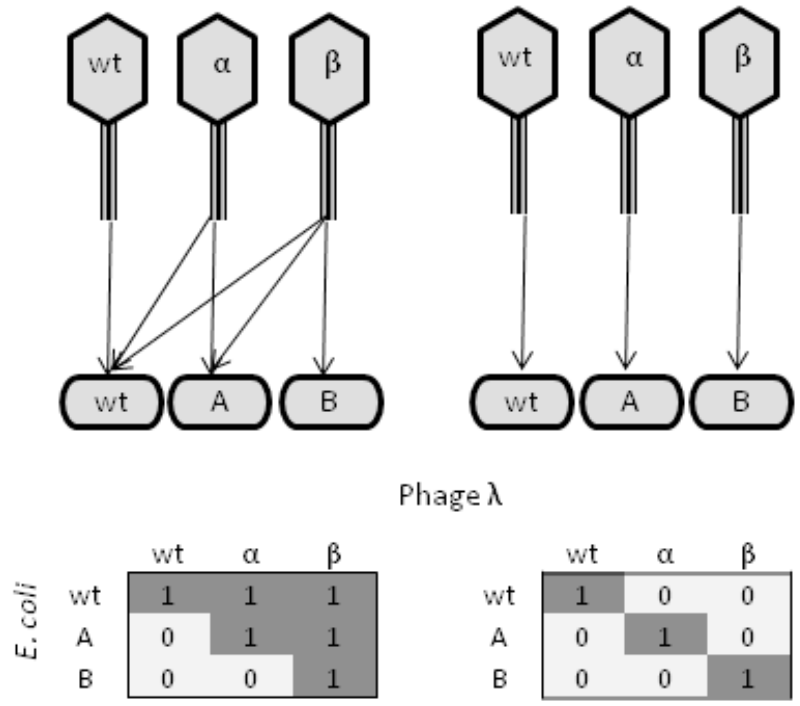


Figure 2.1: Extended host-range coevolution (left) versus the matching alleles model (right). Diagrams depict the ancestral 'wt' and evolved interactions of the phage and bacteria. Arrows or '1's describe a combination of phage and bacteria that result in a successful infection, '0's indicate an unsuccessful infection. Two-step extended host-range coevolution was shown previously to occur between λ and *E. coli* cultured in a permissive environment, however in more restrictive environments λ and *E. coli* are expected to undergo the matching alleles form of coevolution.

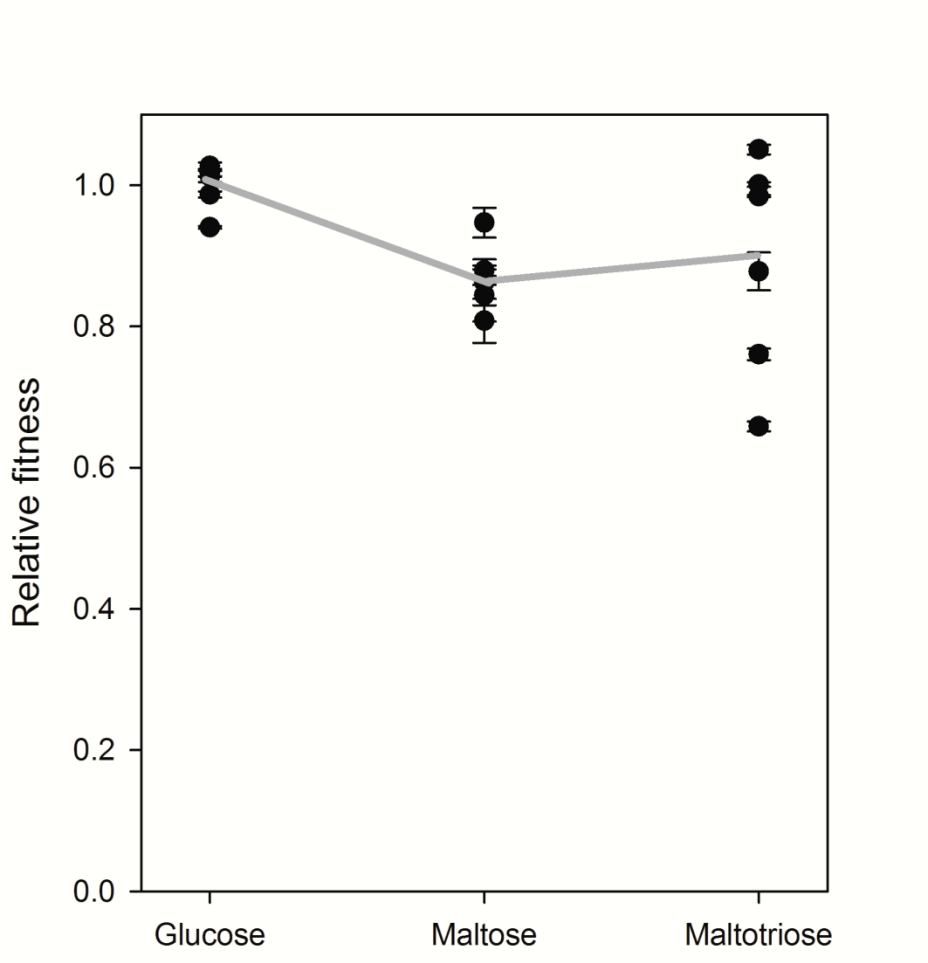


Figure 2.2: Fitness of λ-resistant genotypes evolved in three separate resource conditions. Each resistant genotype was competed in its native environment against the ancestor of the opposite *ara* marker. Points represents the mean of three replicate competition experiments and error bars describe s.e.m. One representative genotype was competed from each community (18 split among three treatments). The gray line links the mean value for fitness of all six isolates from each environment.

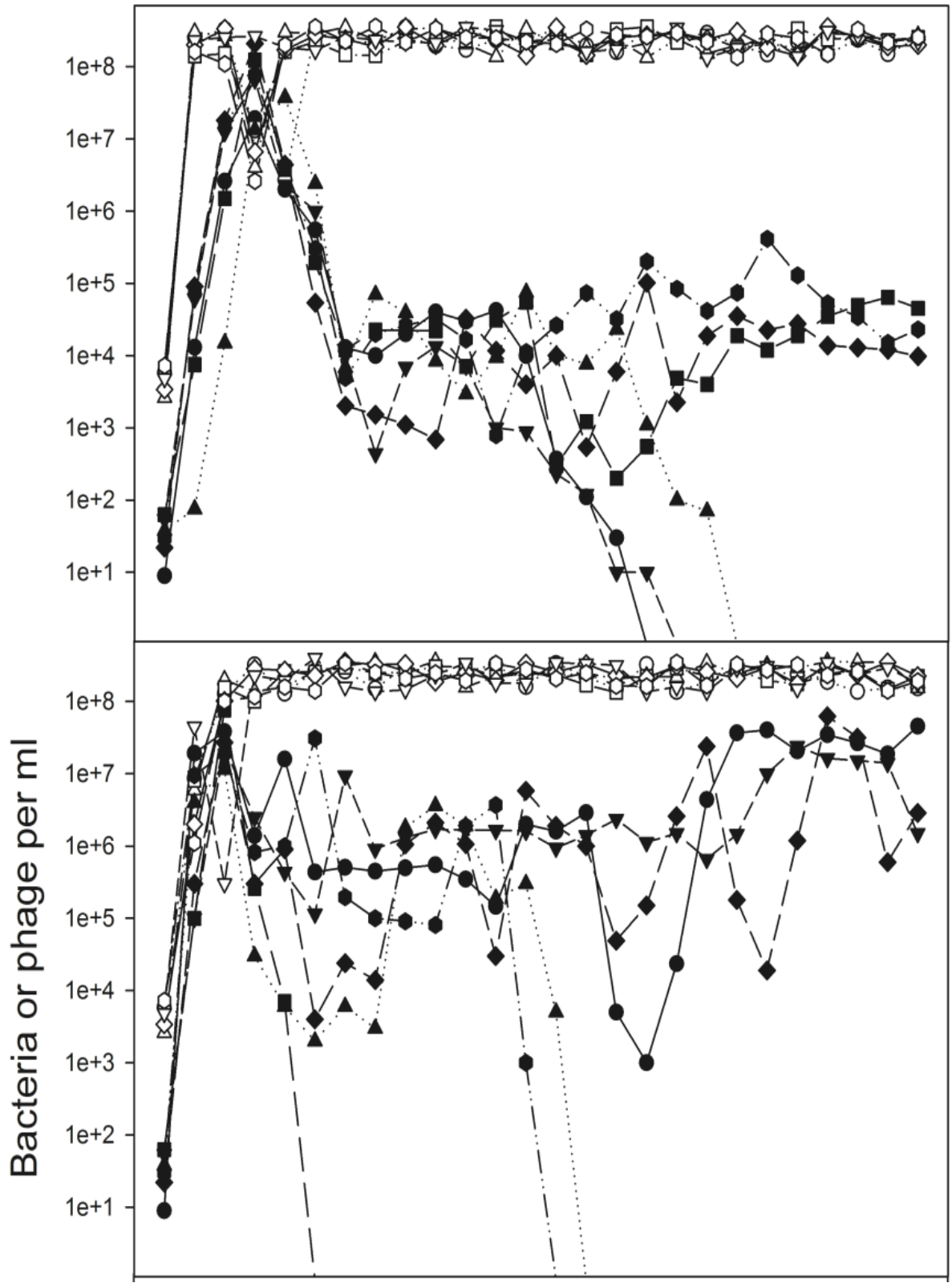


Figure 2.3 (continued on the following page)

Figure 2.3 continued from previous page

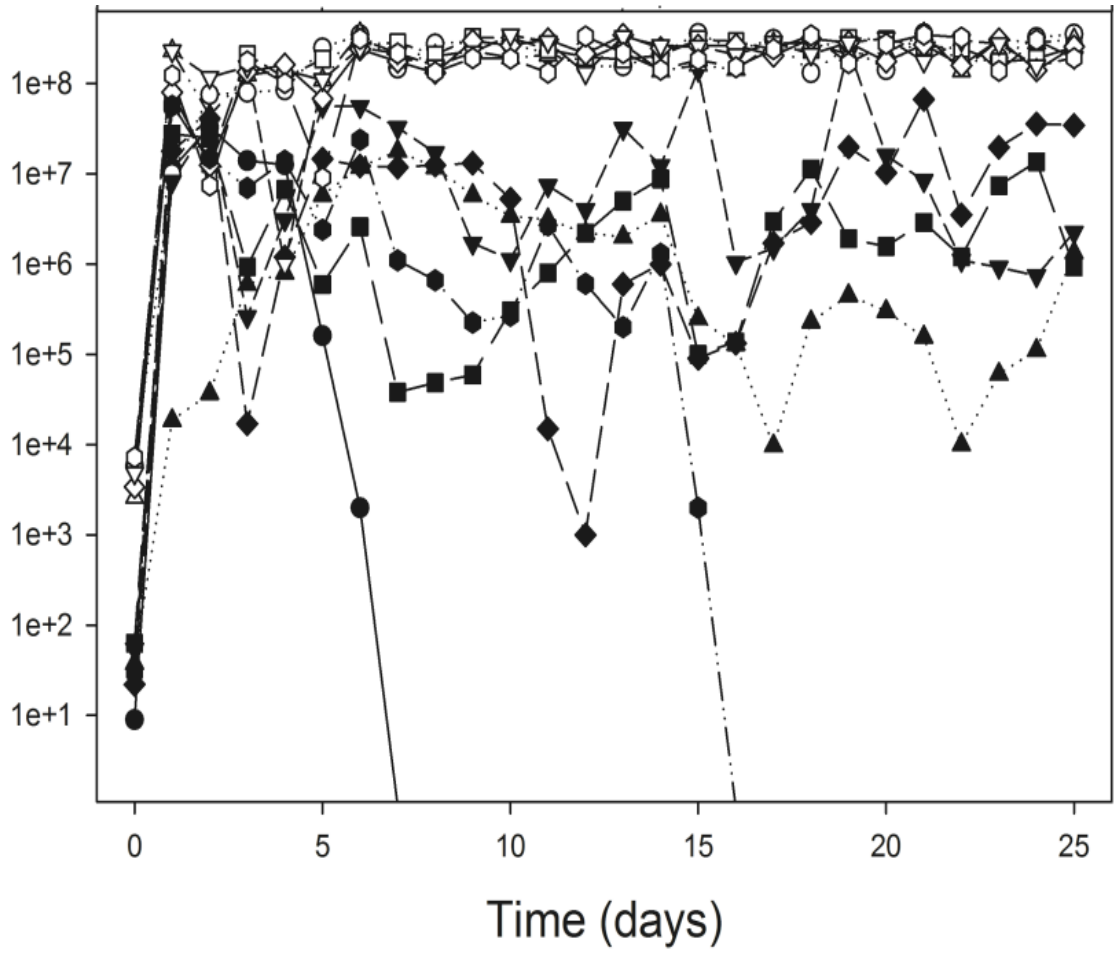


Figure 2.3: Phage population dynamics (filled symbols) vary by treatment (glucose: top, maltose: middle, maltotriose: bottom), however bacterial dynamics (open symbols) do not. The origin of the y-axis indicates the limit of phage detection (~3 particles per ml), where lines cross the x-axis phage went extinct. Note that the densities were measured at the end of the 24-hour cycle, just before they were transferred to a new flask to start a new growth cycle.

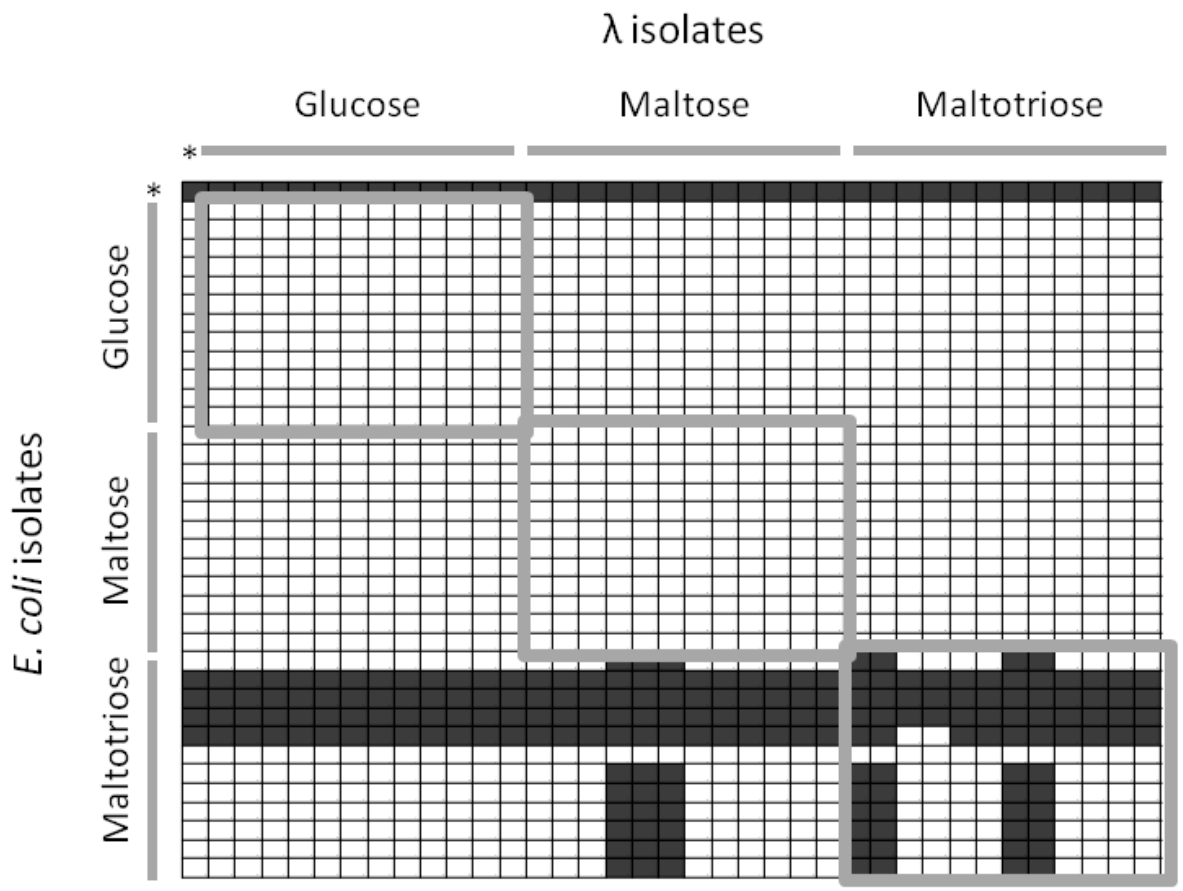


Figure 2.4: Interaction matrix of phage and bacteria sampled across all populations on day three of the first experiment. Two genotypes of phage and bacteria were assayed from each population, as well as their ancestors (asterisk). Rows describe the resistance profile of each bacteria (white = resistant, gray = sensitive) and columns are of phage host ranges (gray = included, white = excluded). Rectangles indicate regions of the matrix where bacteria are challenged by phage evolved in the same resource environment. The top-down and left-right orderings of isolates are mirrored such that the cells along the diagonal originating in the top left corner of the matrix describe interactions between co-occurring genotypes.

Table 2. 1: Mutations for λ-resistance evolved after 3, 10, and 25 days in three replicates of each treatment.

Environment	Population	Time (day)	gene mutated	nucleotide location*	mutation	Effect on protein
Glucose	-3	3	*malT*	992	25-bp duplication	frameshift
Glucose	1	3	*malT*	912	Δ1	frameshift
Glucose	3	3	unknown			
Glucose	-3	10	unknown			
Glucose	1	10	unknown			
Glucose	3	10	*malT*	992	25-bp duplication	frameshift
Glucose	-3	25	*malT*	T2681C	T → C	Leu → Pro
Glucose	1	25	unknown			
Glucose	3	25	*malT*	2177	A → T	Asp → Val
Maltose	-1	3	*lamB*	1162	Δ29	frameshift
Maltose	-2	3	*lamB*	1162	Δ11	frameshift
Maltose	1	3	*lamB*	308	Δ11	frameshift
Maltose	-1	10	unknown			
Maltose	-2	10	*lamB*	966	Δ11	frameshift
Maltose	1	10	*lamB*	429	C → G	stop
Maltose	-1	25	unknown			
Maltose	-2	25	*lamB*	966	Δ11	frameshift
Maltose	1	25	*lamB*	429	C → G	
Maltotriose	-2	3	*lamB*	697	Δ11	frameshift
Maltotriose	-3	3	unknown			
Maltotriose	1	3	*lamB*	1128	G → T	Trp → Cys
Maltotriose	-2	10	*lamB*	709 856	G → T T → A	Gly → Cys Leu → Gln
Maltotriose	-3	10	*lamB*	509	G → C	Arg → Pro
Maltotriose	1	10	unknown			
Maltotriose	-2	25	*lamB*	716	T → A	stop
Maltotriose	-3	25	*lamB*	1128 796	G → T +18	Trp → Cys +6 AA
Maltotriose	1	25	unknown			

* For deletions, insetions, or duplications the first base pair (bp) affected is indicated.

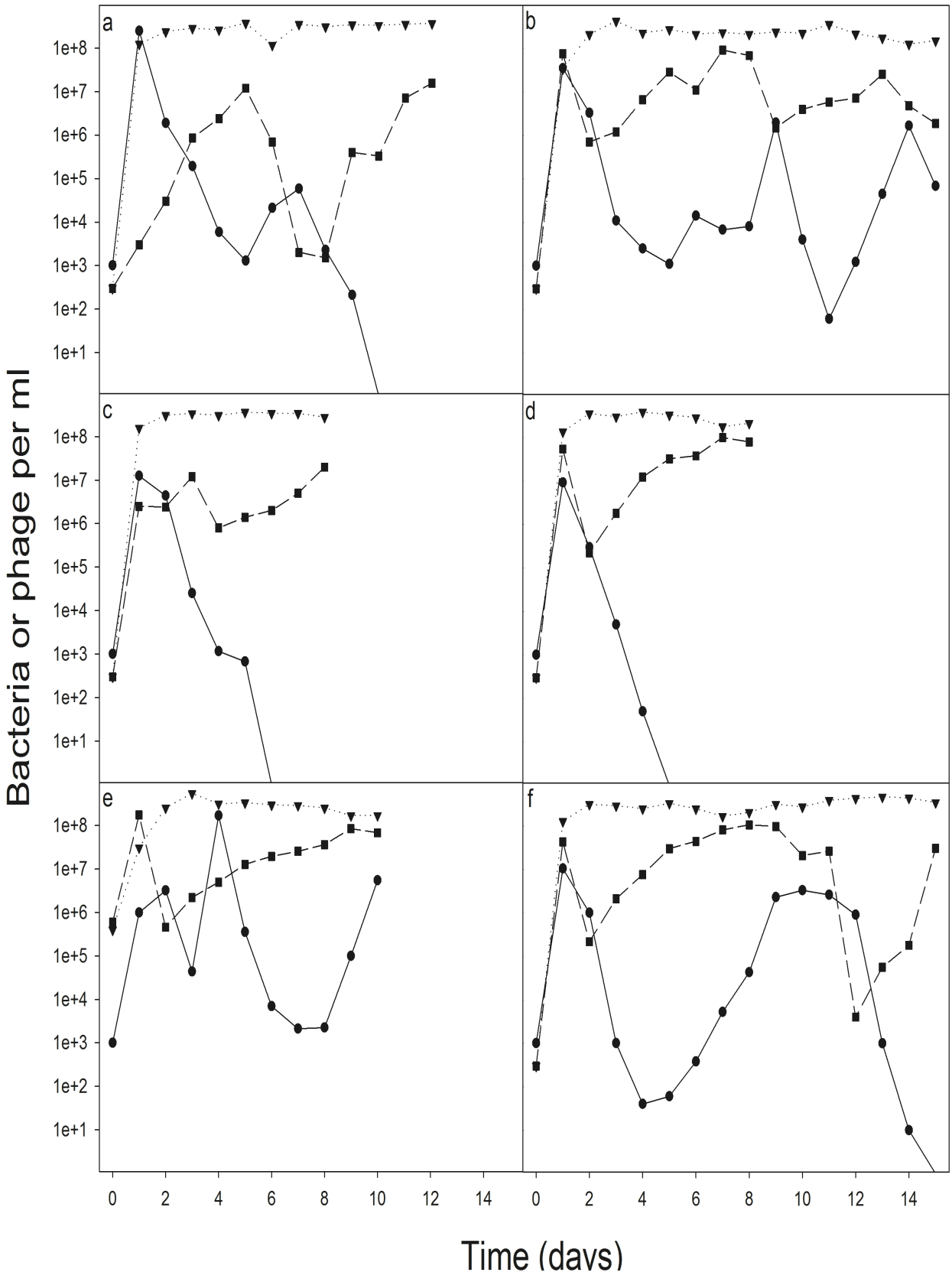
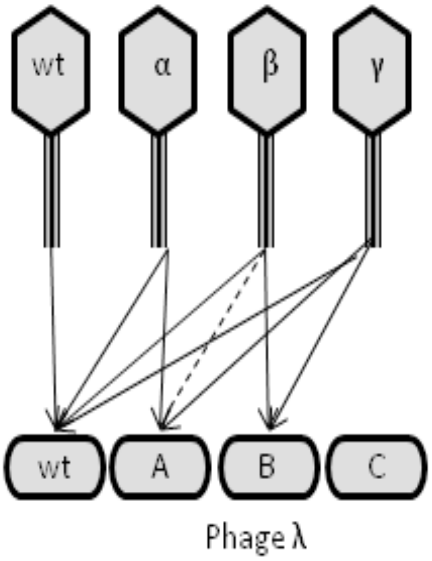


Figure 2.5 (caption on following page)

Figure 2.5: Population dynamics of phage (circles), sensitive bacteria (squares), and resistant bacteria (triangles) in maltotriose medium. Panels on the left depict communities initiated with a completely resistant genotype isolated from day 25 of the '-2' maltotriose replicate, while the panels on the right are from '-3'. Otherwise, all six replicates were run identically. The origin of the y-axis indicates the limit of phage detection (~3 particles per ml), lines that fall below show instances of phage extinction. The experimental replicate in panel **e** ended early because the flask became contaminated.

a



b

E. coli	wt	α	β	γ
wt	1	1	1	1
A	0	1	0.5	1
B	0	0	1	1
C	0	0	0	0

c

	E. coli				Phage λ			
Days	wt	A	B	C	wt	α	β	γ
0	0.5			0.5	1			
1	0.69			0.31	0.167	0.833		
2		0.002	0.001	0.997		0.167	0.833	
3	0.003			0.997			0.167	0.833
4	0.028			0.972			0.5	0.5
5	0.095			0.905				1
6			0.05	0.95				1
7	0.283			0.717				1
8			0.243	0.757				1
9		0.006		0.994				1
10		0.006	0.012	0.982		0.333		0.667
11	0.011	0.005		0.984		0.333		0.667
12	0.032			0.968		1		
13	0.086		0.043	0.872		0.333		0.667
14	0.024	0.012		0.963		1		

Figure 2.6: Resistance and host-range profiles (**a** and **b**) and their genotypic dynamics (**c**) underlying the population dynamics reported in Figure 2.5**b**. Arrows indicate whether a phage can infect a bacterium (solid), or only partially (dashed). Values in matrix **b** signify whether a phage-bacterial combination yields no infection '0', partial '0.5', or complete '1'. Values given in table **c** are for the frequencies of each bacterial and phage genotytype.

LITERATURE CITED

LITERATURE CITED

1. Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, and B. Libberton. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464:275-278.

2. Bohannan, B. J. M. and R. E. Lenski. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Let. 3:362-377.

3. Brown, J. S. and T. L. Vincent. 1992. Organization of predator-prey communities as an evolutionary game. Evolution 46:1269-1283.

4. Brooks, D. R. 1988. Macroevolutionary comparisons of host and parasite phylogenies. Ann. Rev. of Ecol. and Syst. 19:235-259.

5. Hamilton, W. D. 1980. Sex versus non-sex versus parasite. Oikos:282-290.

6. Weitz, J. S., H. Hartman, and S. A. Levin. 2005. Coevolutionary arms races between bacteria and bacteriophage. PNAS 102:9535-9540.

7. Schemske, D. W., G. G. Mittelbach, H. V. Cornell, J. M. Sobel, and K. Roy. 2009. Is there a latitudinal gradient in the importance of biotic interactions? Ann. Rev. of Ecol. and Syst. 40:245-269.

8. Forde, S. E., R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220-223.

9. Nuismer, S. and B. Ridenhour. 2008. The contribution of parasitism to selection on floral traits in Heuchera grossulariifolia. J. of Evol. Biol. 21:958-965.

10. Gómez, P. and A. Buckling. 2011. Bacteria-phage antagonistic coevolution in soil. Science 332:106-109.

11. Jessup, C. M. and B. J. M. Bohannan. 2008. The shape of an ecological trade-off varies with environment. Ecol. Let. 11:947-959.

12. Lenski, R. E. and B. R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat.:125:585-602.

13. Charbit, A. 2003. Maltodextrin transport through LamB. Frontiers in Bioscience 8:s265-274.

14. Thirion, J. P. and M. Hofnung. 1972. On some genetic aspects of phage Lambda-resistance in Escherichia coli K12. Genetics 71:207-216.

15. Lin, E. C. C., 1987 Dissimilatory pathways for sugars, polyols, and carboxylates, pp. 244-284 in Escherichia coli and Salmonella typhimurium: Cellulurand Molecular Biology, edited by F. C. Neidhardt. American Society for Microbiology, Washington, DC.

16. Dargent, B., J. Rosenbusch, and F. Pattus. 1987. Selectivity for maltose and maltodextrins of maltoporin, a pore-forming protein of Eschierchia coli outer-membrane. FEBS Letters 220:136-142.

17. Appleyard, R. K., J. F. McGregor, and K. M. Baird. 1956. Mutation to extended host range and the occurance of phenotypic mixing in the temperate coliophage Lambda. Virology 2:565-574.

18. Braun-Breton, C. and M. Hofnung. 1981. In vivo and in vitro functional alterations of the bacteriophage lambda receptor in lamB missense mutants of Escherichia coli K-12. J. of Bact. 148:845-852.

19. Spanakis, E. and M. T. Horne. 1987. Coadaptation of Escherichia coli and coliophage lambda-vir in continuous culture. J. of Gen. Micro. 133:353-360.

20. Agrawal, A. and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4:79-90.

21. Katsura, I. 1981. Tail assembly and injection. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

22. Berkane, E., F. Orlik, J. F. Stegmeier, A. Charbit, M. Winterhalter, and R. Benz. 2006. Interaction of bacteriophage lambda with its cell surface receptor: an in vitro study of binding of the viral tail protein gpJ to LamB (maltoporin). Biochemistry 45:2708-2720.

23. Gomez, P. and A. Buckling. 2011. Bacteria-phage antagonistic evolution in soil. Science 332: 106-109.

24. Buckling, A. and P. B. Rainey. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc. Roy. Soc. B 269:931-936.

25. Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461:1243-1274.

26. Schrag, S. J., V. Perrot, and B. R. Levin. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proc. Soc. Lond. B 264:1287-1291.

27. Studier, F. W., P. Daegelen, R. E. Lenski, S. Maslov, and J. F. Kim. 2009. Understanding the differences between genome sequences of Escherichia coli B Strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 Genomes. J. of Molec. Biol. 394:653-680.

28. Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in Echerchia coli. 1. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315-1341.

29. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

30. Hendrix, R. W. 1983. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

31. Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage Lambda. Science 335:428-432.

32. Sambrook, J., and D. W. Russell. 2001 Molecular Cloning, Third Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

33. Adams, M. H. 1959. Bacteriophages. Interscience Publishers, New York, NY.

34. Dippel, R., W. Boos. 2005. The Maltodextrin System of Escherichia coli: Metabolism and Transport. J. of Bac. 194:8322-8331.

35. Refardt, D. and P. B. Rainey. 2010. Tuning a genetic switch: experimental evolution and natural variation of prophage induction. Evolution 64:1086-1097.

36. Maynard, N. D., E. W. Birch, J. C. Sanghvi, L. Chen, M. V. Gutschow, and M. W. Covert. 2010. 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:e1001017.

37. Werts, C., V. Michel, M. Hofnung, and A. Charbit. 1994. Adsorption of bacteriophage lambda on the LamB protein of Escherichia coli K-12: point mutations in gene J of lambda responsible for extended host range. J. of Bac. 176:941-947.

38. Yoshida, T., S. P. Ellner, L. E. Jones, B. J. M. Bohannan, R. E. Lenski, and N. G. Hairston, Jr. 2007. Cryptic population dynamics: rapid evolution masks trophic interactions. PLoS Biology 5:1868-1879.

39. Pelosi, L., L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173:1851-1869.

39. Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. and Ocean. 45:1320-1328.

40. Hall, A., Scanlan, P.D. Morgan, A.D., and Buckling, A. 2011. Host–parasite coevolutionary arms races give way to fluctuating selection. Ecol. Let. 14: 635–642.

41. Hofnung, M., A. Jezierska, and C. Braunbreton. 1976. Lamb mutations in Escherichia coli K12 growth of Lambda host-range mutants and effect of nonsense supressors. Molec. & Gen. Genet. 145:207-213.

42. Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host–phage interactions. PNAS 108:E288-E297.

43. Hobson, K. A. and H. E. Welch. 1992. Determination of trophic relationships within a high Arctic marine food web using ?13C and ?15N analysis. Mar. Ecol. Prog. Series 84:9-18.

44. Pascual, M. and J. A. Dunne. 2006. Ecological networks: linking structure to dynamics in food webs. Oxford University Press, USA.

45. Poisot, T., G. Lepennetier, E. Martinez, J. Ramsayer, and M. E. Hochberg. 2011. Resource availability affects the structure of a natural bacteria–bacteriophage community. Biol. Lett. 7:201-204.

46. Whittaker, R. H. 1975. Communities and Ecosystems, Macmillan, New York, NY.

47. Meyer, J. R. and R. Kassen. 2007. The effects of competition and predation on diversification in a model adaptive radiation. Nature 446:432-435.

48. Marston, M. F., F. J. Pierciey Jr, A. Shepard, G. Gearin, J. Qi, C. Yandava, S. C. Schuster, M. R. Henn, and J. B. H. Martiny. 2012. Rapid diversification of coevolving marine Synechococcus and a virus. PNAS 109:4544-4549.

49. Bohannan, B. J. M. and R. E. Lenski. 1999. Effect of prey heterogeneity on the response of a model food chain to resource enrichment. Am. Nat. 153:73-82.

50. Gudelj, I., J. S. Weitz, T. Ferenci, M. C. Horner-Devine, C. J. Marx, J. R. Meyer, and S. E. Forde. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol. Lett. 13:1073-1084.

CHAPTER 3

THE STATISTICAL STRUCTURE OF HOST-PHAGE INTERACTIONS

Authors: Cesar O. Flores[x], Justin R. Meyer[x], Sergi Valverde, Lauren Farr, and Joshua S. Weitz
x These authors contributed equally to this work

Originally published an in open access manuscript in the Proceedings of the National Academy of Science 108, E288-297 (2011).

Abstract

Interactions between bacteria and the viruses that infect them (i.e., phages) have profound effects on biological processes, yet, despite their importance, little is known on the general structure of infection and resistance between most phage and bacteria. For example, are bacteria-phage communities characterized by complex patterns of overlapping exploitation networks, do they conform to a more ordered general pattern across all communities, or are they idiosyncratic and hard to predict from one ecosystem to the next? To answer these questions, we collect and present a detailed meta-analysis of 38 laboratory verified studies of host-phage interactions representing almost 12,000 distinct experimental infection assays across a broad spectrum of taxa, habitat, and mode of selection. In so doing, we present evidence that currently available host-phage infection networks are statistically different from random networks and that they possess a characteristic nested structure. This nested structure is typified by the finding that hard-to-infect bacteria are infected by generalist phages (and not specialist phages) and easy-to-infect bacteria are infected by generalist and specialist phages. Moreover, we find that currently available host-phage infection networks do not typically possess a modular structure. We explore possible underlying mechanisms and significance of the observed nested host-phage interaction structure. In addition, given that most of the available host-phage infection networks

examined here are composed of taxa separated by short phylogenetic distances, we propose that the lack of modularity is a scale-dependent effect and then describe experimental studies to test whether modular patterns exist at macroevolutionary scales.

Introduction

Bacteria and their viruses (phages) make up two of the most abundant and genetically diverse groups of organisms (35-37). The extent of this diversity has become increasingly apparent with the advent of community genomics. Microbial DNA isolated from oceans, lakes, soils, and human guts have revealed tremendous taxonomic diversity in a broad range of environmental habitats and conditions (38-45). The ongoing discovery of new taxonomic diversity has, thus far, out-paced gains in understanding the function of specific microbes and their most basic ecology of who interacts with whom. One of the starkest examples of this disparity is the lack of an efficient (bioinformatic or otherwise) approach for determining which viruses can infect which bacteria. Although it is well known that individual phages do not infect all bacteria, we have little understanding of what the precise host-range for any given phage is, or whether there are universal patterns or principles governing the set of viruses able to infect a given bacteria and the set of bacteria that a given virus can infect. This deficit is unfortunate given that phage-bacterial interactions are important for both human health and ecosystem function (46-50).

Phages have multifaceted effects on their hosts, they can lyse their hosts releasing new virions, transfer genes between hosts and form lysogens that can modify host function (51-53)**.** In some cases, phages can transfer genes for pathogenicity between pathogenic and labile strains, e.g., for both Cholera and Shigella, facilitating the spread of bacterial infections (54-56). Phages also alter ecosystem functions by the high levels of bacterial mortality they cause. Bacteria lysed by phage will release their biomass, which consequently is scavenged by other bacteria rather than being incorporated into bactivorous Eukaryotes (57, 58). This weakened connection early in the food chain can have effects that ripple throughout the ecosystem. Information on a general pattern of infection by phages on hosts could improve predictions of microbial population dynamics, ecosystem functioning, and microbial community assembly (59, 60).

What is our expectation for the general pattern of host-phage infection networks? Host-

phage infection networks have, in the past, been measured by performing pairwise infections of hosts by phages isolated from natural ecological communities, evolution experiments, or strain collections. The results of such pairwise infections can be represented as a network or as a matrix where the rows indicate host isolates and the columns indicate phage isolates, the cells within the matrix describe whether each combination results in a successful infection. We consider different classes of host-phage interaction networks as alternative hypotheses for an expected pattern (see Figure 3.1). First, phages may infect a unique host or a limited number of closely related hosts, leading to nearly diagonal matrices (Figure 3.1A) or block-like matrices that exhibit high degrees of modularity (Figure 3.1B). These patterns should occur if host-viral interactions are the result of coevolutionary processes that lead to specialization. Second, diversification of hosts and phages may result in nested matrices in which the most specialist phages infect those hosts that are most susceptible to infection, rather than infecting those hosts that are most resistant to infection (Figure 3.1C). The nested pattern is the predicted outcome of a prominent theory of "gene-for-gene" coevolution where phages evolve so as to broaden host ranges and bacteria evolve so as to increase the number of phage they are resistant to (18, 61). We should note that these two patterns and hypotheses for the forms of coevolution are not mutually exclusive and could in fact be scale-dependent. Nested patterns could form within modules if for instance microevolutionary changes result in nestedness, however genetic differences between species or genera that accumulate over macroevolutionary time may limit the exchange of viruses between these phylogenetic groups and create an overall modular structure. Finally, we consider a null model to be that matrices of host-phage infection are statistically indistinguishable from random matrices (Figure 3.1D).

Contrary to this null expectation, we show that currently available host-phage interaction matrices are, as a whole, statistically distinguishable from random matrices and possess a characteristic nested structure. We reach this conclusion by performing a meta-analysis on the patterns of host-phage infection matrices collected via a comprehensive search of the literature and supplementing these matrices by a novel experimental analysis of host-phage infection. The data we assemble consists of 38 matrices of host-phage infection assays representing the cumulative study of 1009 bacterial isolates, 502 phage isolates, and almost 12,000 separate attempts to infect a bacteria host with a phage strain (18, 62-97) (see Tables A2.1-2 for more information on the examined studies). This is the first attempt to subject host-phage infection

assays to a unified analysis and the first finding of a general pattern of host-phage interactions. We discuss biophysical, ecological, and evolutionary mechanisms that could lead to this nested (and not modular) pattern as well as future studies to explore how such a pattern may change as a function of phylogenetic scale.

Results

Compiling a Large-scale Host-Phage Interaction Dataset

We compiled a set of 37 studies with direct laboratory evidence of host-phage interactions using an extensive literature search, supplemented by an experimental study of an evolved *E. coli* and phage λ system (Tables A2.1-2 for complete details of all studies) (18, 62-97). The method of evaluating infection ability in assembling a host-phage infection matrix varies, however the most commonly used approach is that of spot assays, in which a single virus type is combined with a population of bacteria cells from a single strain. Infection is considered to have occurred given evidence that the phage has infected and lysed (part of) the bacterial population. Hence, the result of each study is a matrix of the infection ability for each phage on each host. The studies included in the host-phage infection assays analyzed here were isolated from one of three sources: co-occurring isolates within natural communities taken directly from the environment and then cultured (NAT), coevolutionary lab experiments where a single bacterial and phage clone were allowed to co-evolve for a fixed amount of time and then their evolved progenitors examined (EXP), and lab stocks of phages and hosts that were artificially combined (ART). Some of the matrices used were composed of bacteria and phage acquired from two separate isolation strategies. For these studies we classified the matrix by which isolation strategy represented the majority of matrix-cells and made a note of the other sources (see Supplementary Table A2.2). The criterion by which we searched for and catalogued these studies is explained in more detail in Appendix 2. Overall we identified and analyzed a wide-range of infection networks for organisms that varied in their phylogenic position, traits, and habitats. For example, the bacterial hosts included Gram-positives and Gram-negatives, heterotrophs and phototrophs, as well as pathogens and non-pathogens.

Some of the assays include graded information about infection**,** for example whether a

phage simply inhibits bacterial growth or forms regions of complete bacterial mortality, i.e., 'plaques'. In other studies, replicate phage populations were used to deduce whether phages always or only sometimes cause plaques**.** Details of the criteria for the interactions can be found in the original papers (18, 62-97) and the experimental methods for the novel experimental study of host-phage infection can be found in the Methods. Because graded information about infection was not uniformly available in all studies, assays were standardized using hand-curated extraction of original data into a single matrix of 1-s and 0-s with H rows (one for every bacterial host) and P columns (one for every phage), where a 1-valued cell represents evidence for infection (either full or partial) and a 0-valued cell represents no evidence for infection (see Figure 3.2 for a visual depiction of all host-phage interaction matrices).

Host-Phage Infection Statistics Do Not Vary with Study Type Nor Show Significant Cross-Correlations

We calculated a variety of global properties of these matrices: number of hosts (H), number of phages (P), number of interactions (I), number of species ($S=H+P$), size ($M=HP$), connectance ($C=I/M$), mean number of interactions across host species ($L_H=I/H$), mean number of interactions across phage species ($L_P=I/P$) (see Tables A2.1-3 for values of each property within each of the 38 studies). Importantly, on a per-study basis, we find that the average number of phages infecting a given host is 4.88 (median 3.04), whereas the average number of hosts that a phage can infect is 10.91 (median 6.13). Both results are inconsistent with the hypothesis that phages only infect one host and that hosts are only infected by one phage, see Figure 3.1A.

We first sought to establish whether the source type (NAT, EXP or ART) had any influence on basic characteristics of the matrices. We performed a Principal Component Analysis (PCA) (see Supplementary Materials and Methods in Appendix 2, Table A2.4 & Figure A2.1) using these 8 global properties. Despite the significant variation in global properties we find no statistically significant distinction between the three different types of studies. For example, the distributions of type-specific matrices do not cluster into three groups. We apply a Jaccard clustering validity index (98) and find that the degree of clustering validity is 0.26 (indicating poor separation of labeled classes into distinct clusters), which is not significantly different than random (p=0.34, see Supplementary Materials and Methods & Supplementary in

Appendix 2, Figures A2.3-4).

Not only do we not find evidence for clustering, we also find no evidence for significant and biologically meaningful correlations among the global properties of all matrices when grouped together. For example, prior work on the analysis of bipartite networks within plant and pollinator systems found inverse relationships between the total number of species in the network vs. the fraction of interactions that actually occurred (99, 100)**.** We do not find this relationship here. Supplementary Figure A2.2 plots connectance (*C*) vs number of species (*S*). The observed slope is small and non-significant (see Table A2.5). Moreover, the other correlations between connectance and the size of host-phage infection matrices are not significant (see Materials and Methods for details and Supplementary Table A2.5 for the correlation values).

Host-Phage Infection Assays Are Typically Nested and Not Modular

We measured higher-order properties of the host-phage interaction matrices, specifically modularity and nestedness. In this context, modularity is determined by the occurrence of groups of phages that infect groups of hosts significantly more often than they infect other hosts in the system. Modularity is typically found in biological systems in which groups of organisms preferentially interact with organisms within the group (e.g., plant-pollinator network (99, 100)) and is thought to be an important feature underlying the maintenance of biodiversity (101). Likewise, nestedness is determined by the extent to which phages that infect the most hosts tend to infect bacteria that are infected by the fewest phages (102, 103). Nestedness has been used to characterize species interactions because it is predicted to affect important properties of communities such as stability and extinction potential (100, 104). Both modularity and nestedness may emerge due to co-evolutionary adaptation of hosts and phages (61, 105). The individual host-phage infection studies collected here were not subjected to a network analysis with one exception (18). Hence, we examined each study to see if previously unrealized patterns existed within each host-phage interaction network (see Figures 3.3 and Supplementary Figure A2.5 for an example of how network properties are extracted from two matrices, Supplementary Files 1 and 2 for data corresponding to each matrix, and refer to the Methods for further details on how to calculate modularity and nestedness).

For the 38 matrices shown in Figure 3.2, the maximally modular re-labeling of each matrix is displayed in Figure 3.4 and the maximally nested re-sorting of each matrix is displayed in Figure 3.5. In order to evaluate the statistical significance of the modularity and nestedness values of observed host-phage matrices, we have to compare the observed values to those of random matrices. We generate random matrices that have the same size and number of interactions as the original data (see Appendix 2). In that way, we constrain our null model to have exactly the same global properties as detailed in Table A2.1 for each study, while the distribution of the nestedness and modularity will vary between realizations.

The titles of the study in Figure 3.4 (the maximally modular configuration) are in red if they are significantly modular, in blue if they are significantly anti-modular and black if they are non-significantly modular. The majority of studies are significantly anti-modular (where we used a p-value = 0.05, using 10^5 random matrices as our null). Our findings stand in contrast to expectations that groups of phages adsorb to non-overlapping groups of hosts, as would be expected if groups of phages had specialized on groups of hosts within the study systems. The titles of each study in Figure 3.5 (the maximally nested configuration) are displayed in red if they are significantly nested, in blue if they are significantly anti-nested, and black if they are non-significantly nested. The majority of studies are significantly nested ($p < 0.05$), where we used 10^5 random matrices as our null). Overall, we find 27/38 studies to be significantly nested, and when broken down by type we find significant nestedness in 7/9 artificial, 13/19 ecological and 7/10 experimental studies. Our findings corroborate, in one case, an earlier effort to characterize nestedness by Poullain et al (18) using a different nestedness metric. It is also apparent that some matrices are almost perfectly nested: e.g., that of Ceyssens (68), McLaughlin (82) and Seed (90). In some cases, like Middelboe (83), the data came from a mix of ecological and experimental studies, in that the bacteria were derived from wild-type and chemostat isolates, whereas the phages were wild from the same environment as the host. Does the finding of a strongly nested matrix mean, in this case, that in vitro evolution mimics selection in nature, suggesting that there exist robust principles underlying the emergence of nestedness?

Hence, given the number of studies we ask: what evidence is there that host-phage matrices are, as a whole, nested and not modular? We rank all 38 matrices from lowest to largest modularity and lowest to largest nestedness (see Figure 3.6A-B). It is evident that matrices tend to be more nested than their random counterparts but not more modular (and apparently anti-

modular) than their random counterparts. How often do we expect to find 27 significantly nested matrices in a sample of 38 random matrices, if each of the significantly nested matrices has a $p < 0.05$? Combinatorically, such a result is highly improbable and given by a binomial distribution with resulting $p << 10^{-10.}$. Likewise, the finding of an excess of anti-modular matrices (20 of 38) compared to a small number of modular matrices (6 of 38) is a highly improbable result. Moreover most of the significantly modular matrices have low values of modularity, suggesting that although modularity may be deemed significant in a few cases, it is not a driving mechanism underlying the structure of most of these matrices and may be incidental to other patterns. Together these results imply that currently available host-phage infection networks are typically nested and not modular.

Previously overlooked nested patterns uncovered

An additional power of subjecting host-phage infection networks to a unified analysis is that doing so can extract meaningful biological information about the organization of a system that may not have been possible given the original placement of hosts and phages in matrix format. For example, Zinno et al. (97) mention variability in phage infection, however the authors of that study make no mention of the fact that there are evidently groups of phages that preferentially infect groups of hosts (Figure 3.3A). Such block-like variability suggests that resistance mechanisms are less haphazard than they appear when network characteristics are not analyzed. Similarly, Holmfeldt et al. (77) highlighted the variability and possibly unique signature of infection for each host and phage. However, re-ordering hosts according to the number of infecting phages while also re-ordering phages based on the number of hosts they can infect leads to a nested pattern suggesting specific forms of infection rules may underlie infection variability (Figure 3.3B). To what extent is our finding of nestedness novel? As a reminder, nestedness is a property of a host-phage infection matrix as calculated for a given row and column ordering. Hence, we calculated nestedness for all of the matrices in the format as they were first reported in the literature and then compared these results to the nestedness calculated from our reshuffled matrices. We found that in 35/37 cases of the previously published studies the reshuffled matrix had a nestedness value higher than that of the original publication, whereas in 2 of the studies the nestedness was equal (80, 83) (see Supplementary Figure A2.6). Hence,

our results suggest that by and large prior efforts did not identify the extent to which their matrices were nested, or whether such nestedness was significant.

Addressing sample composition biases as potential drivers of network structure

We report a set of analyses to quantify the extent to which potential biases might impact our results. One potential bias in our study derives from the methods some researchers used for phage isolation. Phage require a bacterial host to reproduce, therefore which bacterial host(s) a researcher chooses can affect the form of the interaction matrix. For instance, if researchers used a single host to isolate phage and included this host in the matrix, then their matrix will necessarily posses a full row of positive infections, thereby introducing the first element of a perfectly nested matrix. We found only six studies that utilized such an approach (79, 80, 82, 83, 89, 91). To determine if phage isolation strategy biased our results towards nestedness we re-analyzed all six of these matrices having removed the isolation hosts. We found no significant difference between the nestedness and modularity for each of these six matrices with or without the excluded host (see Table A2.6).

Another potential bias is that studies included zero rows and columns which imply that there are hosts for which no phages infect and phages that infect no hosts, respectively. Note Inclusion of zero rows and columns has the potential to bias the structure toward a nested pattern. However, such zero rows and columns may be biologically meaningful if hosts or phages have evolved resistance that leads to non-interaction between particular sets of strains. Nonetheless, we redid the entire analysis by generating alternative matrices such that hosts and phages were only included if they had had at least one non-zero element in their row or column, respectively. Then, we recalculated nestedness for the modified matrices and compared it to the nestedness of appropriately re-sized null matrices. We found 26 of 38 studies were nested compared to 27 of 38 using the original analysis (see Supplementary Figure A2.7). Moreover, although the quantitative value of nestedness did decrease in one cases, that particular study (72) was in fact still highly nested and marginally significant at a $p = 0.067$ level. We also recalculated modularity for the modified matrices and find that 9/38 are modular compared to 6/38 in the original analysis (see Supplementary Figure A2.8). Hence, although there are minor changes in the number of significantly nested and modular networks, our finding that matrices

have a characteristic nested structure is robust to either of these sources of bias.

Finally, we ask whether there are certain characteristics of matrices that defy the general pattern of nestedness and if it is possible to learn from these outliers? Interestingly, the three matrices with the most significant modular structures (73, 88, 97) were determined for a single bacterial species, *Streptococcus thermophilus,* and its phages. This finding seems robust since different labs performed the studies and the microbes were isolated from three separate continents. Additionally, we did not find an example where a matrix that included *Streptococcus thermophilus* did not have the modular structure. We examined bacteria from the same taxonomic order (*Lactobacillales*) and isolated from the same environment (dairy products) but they lacked a modular structure. The consistent modularity observed for this species suggests that species-specific traits may have strong deterministic effects on the form their interactions with parasites take. We are unsure of what traits produce the modular interactions, however further research may help reveal if and what resistance mechanisms determine the shape of microbial interaction networks.

Possible scale-dependence of host-phage interactions: from nestedness to modularity?

The data we analyzed included almost 12,000 separate attempts to infect a host isolate with a phage isolate. Although the scale of the current data is beyond the scope of any individual project, it still pales in comparison to the number of possible interactions in a community, at local or regional levels. Scaling up to larger assays presents technical challenges aside from increasing the depth of sampling. Studying many host strains beyond the species (or genus) level often requires distinct culture conditions, a pre-requisite for studies that many labs cannot or do not want to reach. Here, we present an analysis of what such a hypothesized study may reveal. Consider for the moment an experiment in which the hosts from two groups of experiments were combined in a large cross-infection assay with the phages from the same two groups of experiments. If the original matrix sizes were $H_1 \bullet P_1$ and $H_2 \bullet P_2$, then the final matrix size is $(H_1 + H_2)\bullet(P_1 + P_2)$. A total of $H_1P_2+H_2P_1$ new experiments would need to be performed. If the hosts were of sufficiently distant types (e.g., *E. coli* and *Synecoccocus*) we should expect that nearly all of the new cross-infection experiments would lead to no additional infections. Hence, if the original matrices were nested, then the new matrix would have two modules, each of which

was nested (see Figure 3.7 for the results of such a numerical experiment). In other words, we predict that at larger, possibly macroevolutionary scales, host-phage interaction matrices should be typified by a modular structure, even if there is nested structure at smaller scales.

Discussion

Summary of major results

We have established, for the first time, a unified approach to analyzing host-phage infection matrices. In so doing, we find that a compilation of 38 empirical studies of host-phage interaction networks are nested, on average, and not modular (see Figures 3.4 and 3.5). In most cases, our finding of higher-order structure such as nestedness within an individual study was, itself, a novel finding, in that prior analyses of host-phage interaction matrices usually did not attempt to estimate the network characteristics examined here. We found that host-phage interaction networks are not perfectly nested, and that interactions that defy perfect nestedness are typical throughout nearly all of the data. Additionally we found no significant difference in nestedness or modularity based on taxa, sources, or isolation method. This dataset, though far larger than any individual study, is limited to (largely) microevolutionary scales, an issue which we addressed in the final Results section and which we return to later in this Discussion. Considering the large range of taxa, habitats, and sampling techniques used to construct the matrices, the repeated sampling of a nested pattern of host-phage infections is salient, although the process driving the nestedness is not obvious. It could result from multiple mechanisms or a single principle. Here, we examine three hypotheses to explain the nestedness pattern based on biochemical, ecological and evolutionary principles. Note that these hypotheses are not mutually exclusive and we have only limited ability to test them given our comparative approach. However, each of these hypotheses can be tested with additional lab-based or field experiments.

Mechanisms responsible for nestedness; biophysical, ecological, and evolutionary

Phage and bacterial infection matrices at microevolutionary scales may be constrained to a nested shape by the nature of their molecular interactions. Phage infect bacteria by using specialized proteins that target and bind to molecules on the outer-membranes of bacteria (receptor molecules). Nested infection matrices have been shown for T-phage that infect strains of *Escherichia coli* to be the result of the interactions of the phage proteins and receptor molecules (106). T-phage bind to the lipid-polysacharide (LPS) chains on the cell surface. Mutant *E. coli* have been observed with shortened LPS chains that confer resistance to some, but not all T-phage. There are T-phage that are able to infect these mutants because they require fewer segments of the LPS molecule to bind. If phage-bacterial molecular interactions are dominated by single traits and variation in these traits is constrained along a single hierarchical dimension such as LPS, then one should expect the nested pattern to arise. There are other examples of traits with physical characteristics that behave similarly; bacteria that evolve a thicker and thicker protective coating (107), phage that evolve increased host range by continually reducing its tail length (106), and bacteria that reduce their number of receptors and phage that target fewer receptors (108). While there are many examples of this type of one-dimensional interaction, the problem with this being a universal explanation for the form of bacterial-phage interactions is that host-phage interactions are governed by hundreds of other genes (109), bacteria can employ multiple strategies for resistance (107), and phage have complex mechanism to evade bacteria defenses (107, 110). Moreover, a recent discovery of an adaptive immune system in which bacteria acquire targeted sequences to prevent phage infection, and phages evolve to evade such immunity suggests a complex interaction space (111). Given the diversity of host-phage interactions it seems unlikely that the molecular details alone would constrain the form of their relationship (112). Instead we turn to the potential guiding forces of community assembly and coevolution to explain this reoccurring pattern.

The nested pattern may be common because the processes of microbial community assembly select for species with nested relationships. One could imagine that communities may settle into this pattern if this interaction structure is more stable than others (100, 104), noting that the stability of host-phage interaction structures may depend on ecological factors such as resource availability (113). Cohesive interaction structures such as nested patterns have been

shown to be more stable than other structures for mutualistic networks (114, 115). The regularity of the interactions and redundancies make these communities less susceptible to the random removal of nodes. However, these networks are thought to be susceptible to invasion by new species that violate the nested pattern, suggesting that migration of a species would perturb the nestedness. Further, the spatiotemporal complexity of microbial and viral communities suggests that prior theoretical efforts which consider community addition as a process in which invasions occur infrequently may not be widely applicable. Moreover, community assembly models rarely invoke the influence of evolutionary change at similar time scales as ecological change – an issue highly relevant to the study of microbial and viral communities.

Indeed, there may be an evolutionary explanation, for nestedness. Most attempts to characterize the form of coevolution with host-phage experiments to date have demonstrated a form of antagonistic evolution called expanded host-range (or gene-for-gene) coevolution (85, 116, 117). Under this model bacteria evolve ever-increasing resistance to more and more phage genotypes and phage evolve broader host ranges. If one were to sample a community of bacteria and phage coevolving under this model they would uncover a diversity of phage and bacteria that exhibit a nested interaction pattern: At any time-point there should exist the most derived bacteria which is either completely resistant or depending on the timing, is sensitive to the most derived phage. Given that selection by phage may be slow to alleviate the more sensitive ancestral variants or that there may be a trade-off between resistance and competitiveness, there will exist a diversity of bacteria with ever decreasing sets of phage they are resistant to. Similarly, the most derived phage will have the broadest host range and by the same logic as for the bacteria, its ancestors are likely to persist in the community and display ever-decreasing host-ranges. The nested pattern could be a product of taking a snapshot of a dynamically evolving community. While the majority of experimental results observed in artificial lab settings support this hypothesis, there is a single lab experiment (118) and models of bacterial host parasite coevolution suggest that other forms of coevolution are possible when there are bottom-up costs for modifications to resistance (119, 120). Further, if coevolution provided the only explanation, then the artificially assembled matrices would not have the nested pattern.

Dispelling and recognizing potential biases

Three sources of sampling bias challenge the generality of our findings. First, the taxa sampled may poorly represent microbial diversity given that they are subject to both human and methodological biases. If for instance only taxa associated with humans were selected or if all taxa were cultured similarly, then our results would only be relevant for a small group of microbes. Indeed the majority of microbial studies were preformed on the family *Enterobacteriaceae* which live within human digestive systems, however the spectrum of bacteria we examined is much broader and include both heterotrophic and photosynthetic species, Gram-negative and -positive bacteria, were isolated from six continents and in many disparate environments, from the extreme conditions of hot springs, the rich resource conditions of sewage, depauperate marine environments, the complex matrix of soil, and the simplified lab environment. While this study cannot feasibly test the full microbial diversity of the globe, it does include examples from much of it (See Tables A2.1-2).

Secondly, as previously discussed, the number of hosts used to isolate phage and the inclusion of non-interacting hosts and phages has the potential to alter the nestedness of a matrix. Ideally the same number of hosts studied in the matrix would be used to isolate phage, or if only a subset of hosts were used, then these hosts would not be included in the matrix. This is important to ensure that the pattern of infection is independent of how the parasites were isolated. We found that these biases were not a problem by: (i) testing matrices that were created by isolating phage on a single host; (ii) removing hosts and phages that were not interacting. We found that whether the matrices were significantly nested was not affected by including the isolation host in the matrix nor by removing non-interacting hosts and phages, which is strong support that isolation method did not enrich for nestedness.

The last category of bias, phylogenetic, is likely to mean that our results define a pattern at relatively narrow taxonomic scales. The majority of our studies were of closely related genotypes and species. As described in the results section, we anticipate that more complex patterns of infection may form at larger phylogenetic scales that likely include increasing compartmentalization. Hence, we hypothesize that a multi-scale view of host-phage infection networks will reveal nestedness at small scales and modularity at large scales. Our finding of nested interaction matrices are still relevant for characterizing patterns at short phylogenetic distances, which arguably are the most relevant for many ecological and evolutionary scenarios

because they likely share the richest connections.

A prospective view

Whatsoever the limitations of this dataset, it is important to point out that viewing host-phage interaction networks through a unifying lens will likely unveil other commonalities of microbial and viral communities. By way of analogy, over 25 years ago, the study of food webs was radically altered by the compilation of many small food webs that were subject to a unified analysis (121-124). The key finding of the earliest food-web studies was that the members of a community could be ranked, and that "larger" species would eat a random fraction of those species "smaller" than it. From this stage there were two ways forward. First, by studying larger food webs the original pattern was refined such that species ranking was found to be correlated with body size (but is not equivalent to body size), such that individuals eat prey that are smaller, although are a part of a well-defined size-class (125, 126). Secondly, the topology of food webs was then used as a target and basis for dynamic models of community behavior, i.e., what mechanisms can explain the patterns and how do the patterns influence community function (127). We hope and envision that a similar process unfolds here, in that the finding of a general principle in the current data set will stimulate the collection of more and larger host-phage infection networks in order to continue to provide a fuller picture of who infects whom across an entire community. In so doing, we caution that data completeness can alter the observed patterns of connectivity and refer readers to a number of recent papers that address this topic (128-132).

What do we expect to find when analyzing ever-larger host-phage interaction networks collected from within an ecological community, evolution experiments or from culture collections? We hypothesize that host-phage interaction matrices are likely characterized by modularity at larger taxonomic scales even if there is structure (e.g., nestedness) at small-taxonomic scales (see Figure 3.7). What would such a multi-scale phenomenon inform us about the structure and function of microbiological communities? First, it would suggest the existence of diversifying co-evolutionary induced selection that gave rise to (largely) independent host-phage communities. The molecular basis of such diversification could then be explored. Next, cross-infection assays or similar lab-based strategies (133) that test whether phages can infect or at least transmit their genes between phylogenetically divergent hosts have the potential to

provide significant advances in understanding patterns of global gene transfer. Such phages (and the bacteria they infect) may be critical to understanding the direct transfer of genes on a global scale. Instead of phages acting locally (in a taxonomic sense) to shuttle genes between closely related bacteria, a few rare links would permit greater cross-talk between bacterial taxa. Quantifying the frequency of such events may represent the small-world links that connect distant microbial populations (134), and is need of experimental testing.

Further, infections of distantly related groups by the same phages would imply that the bacteria are in indirect competition with one another, even if they do not seem to compete directly for the same set of carbon and nutrient sources. Although whole genome based approaches to infer host-range and phage-susceptibility may help provide candidates for such rare links, they are not the only solution. Rather, we suggest that the continued use of laboratory-based assays to catalog the life history traits of culturable host-phage pairs is essential if we are to improve our understanding of the population dynamics of host-phage communities in the wild. Of course, many (if not most) bacteria and phages are not currently culturable. Hence, in parallel, we recommend attention be given to the development of inverse methods to catalog the life history traits of phages based on community infection assays in those circumstances in which culturing is impossible or yet intractable.

Materials and Methods Summary

Network statistics

Modularity is estimated by reshuffling the rows and columns of the matrix to find groupings of highly interconnected phage and bacteria, labeling these groups and assessing matrix-wide the ratio between the number of within to outside group connections. This is done using a heuristic called the BRIM algorithm (135) to efficiently find the configuration that maximizes this ratio. We ported the BRIM algorithm to MATLAB from the original code in Octave and used the adaptive BRIM algorithm for all calculations here. By this definition a perfectly modular matrix is comprised of clusters of completely isolated groups and modularity declines as the number of cross-group connections increases. Nestedness is estimated by re-ordering the rows and columns (136, 137) to determine whether phage that infect fewer hosts are only able to infect a subset of bacteria which are susceptible to many phages. This reordering tries to maximize the position of

1-s in the matrix such that they cluster above a nullcline (see Figure 3.1C for a perfectly nested matrix). The value for nestedness depends on how frequently 1-s fall above rather than below this nullcline. Complete details are provided in Appendix 2.

Host-phage infection assay

Matrix #22 is the only dataset not previously published. We constructed the matrix by coevolving an obligately lytic phage λ strain with its host *E. coli*. The *E. coli* studied were of strain REL606, a derivative of *E. coli* B acquired from Richard Lenski, (Michigan State University) described in (138) and phage were of strain cI21 (λvir) provided by Donald Court (National Cancer Institute). The phage and bacteria were cocultured in 50ml Erlenmeyer flasks, with 10ml of liquid medium, shaken at 120 rpm, and incubated at 37 °C. This flask was incubated and the cycle of transfer and incubation was continued once more. Three 24 hour incubations were long enough for the bacteria to evolve resistance and the phage to counter it, however not long enough for a second round of coevolution. We randomly selected 150 bacteria and 150 phage isolates. We determined which of the 150 bacteria isolates were resistant to the 150 phage isolates. To do this we performed 'spot' plate assays. All bacterial phage combinations were replicated five separate times, a total of 28,125 spots were assayed. To make this processes more efficient, we placed up to 96 separate phage stocks onto a single dish (150mm radius). Phage stock replicates were never placed on the same plate in order to reduce the signal of any stochastic plating effects. The five replicates were combined and a phage was only determined to be able to infect a bacterium if three of five replicates were given '1's. Lastly, phage or bacteria that had identical infection or resistance profiles as their ancestors were removed from the matrix. Complete details are provided in Appendix 2.

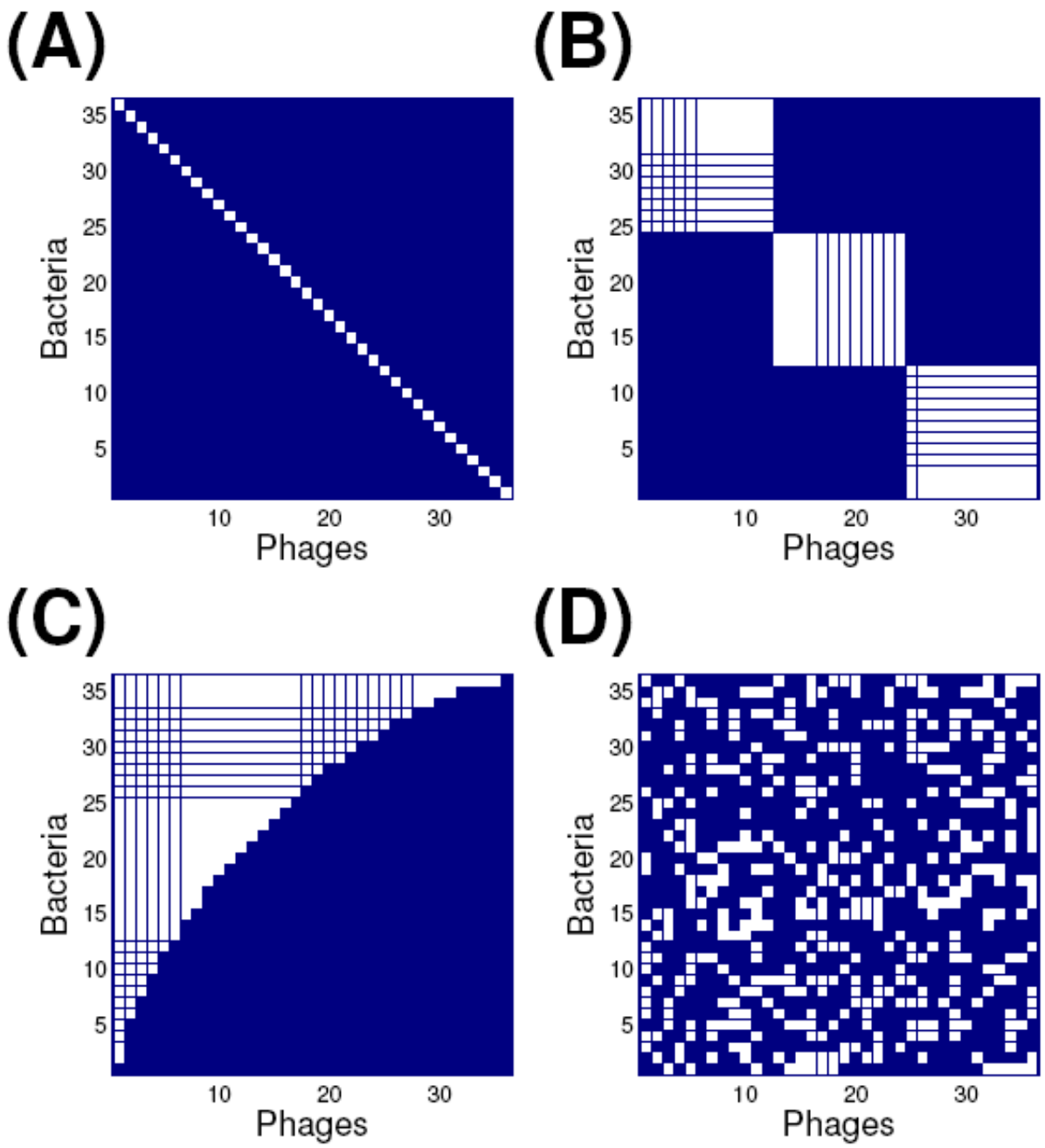


Figure. 3.1 Schematic of expected host-phage interaction matrices. (A) Host-phage interactions are unique, i.e., only one phage infects a given host and only one host is infected by a given phage. (B) Host-phage interactions are modular, i.e., blocks of phages can infect blocks of bacteria, but cross-block infections are not present. (C) Host-phage interactions are nested, i.e., the generalist phage infects the most sensitive and the most resistant bacteria, whereas the specialist phage infects the host that is infected by the most viruses. (D) Host-phage interactions are random and lack any particular structure. For B-D a connectance of 0.33 was used so that the total number of interactions are the same in each case. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

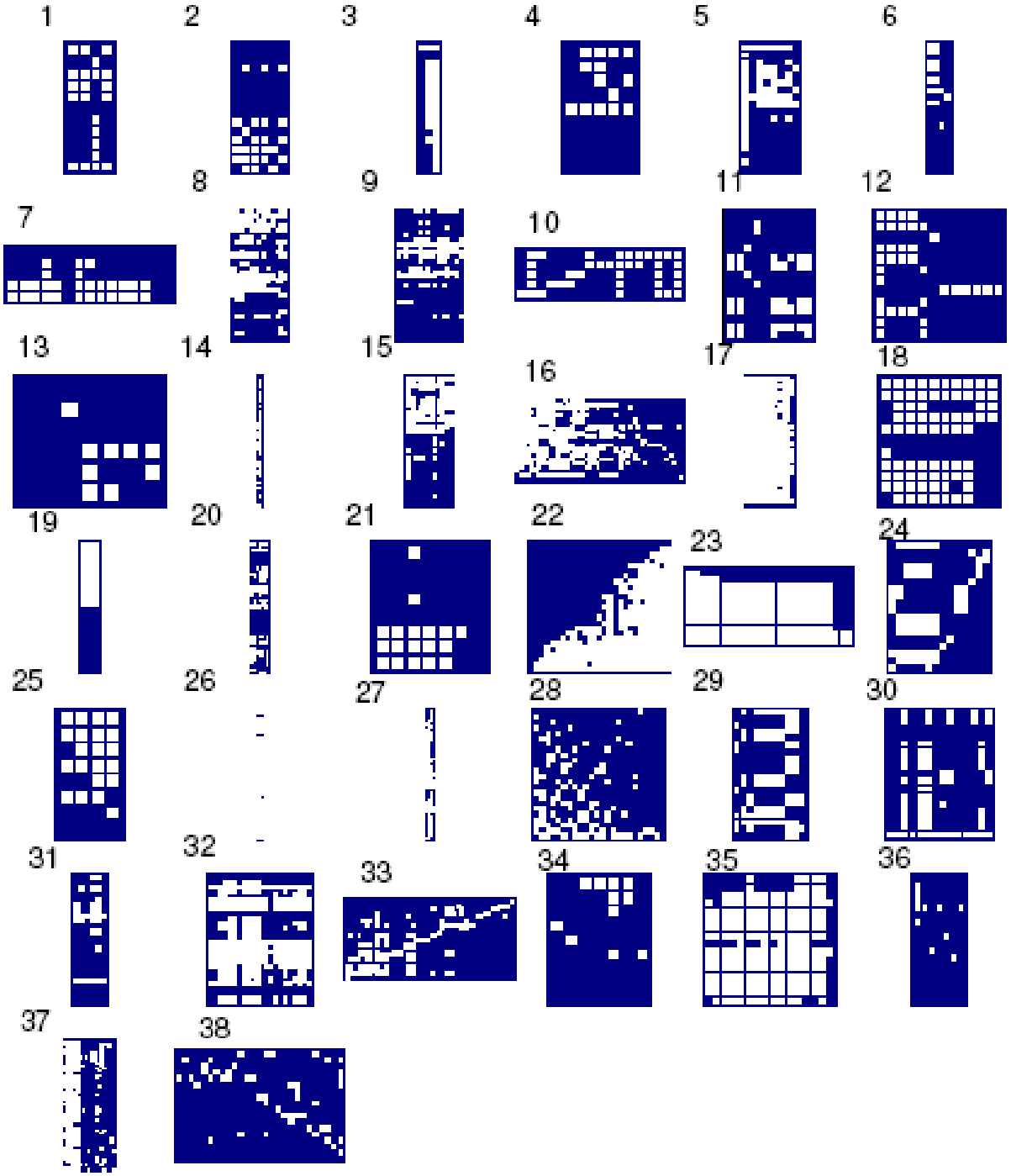


Figure. 3.2 Matrix representation of the compiled studies. The rows represent the hosts, and columns the phages. White cells indicate possibility of infection. Note the diversity in the size, shape, and connectedness (% white) of these matrices.

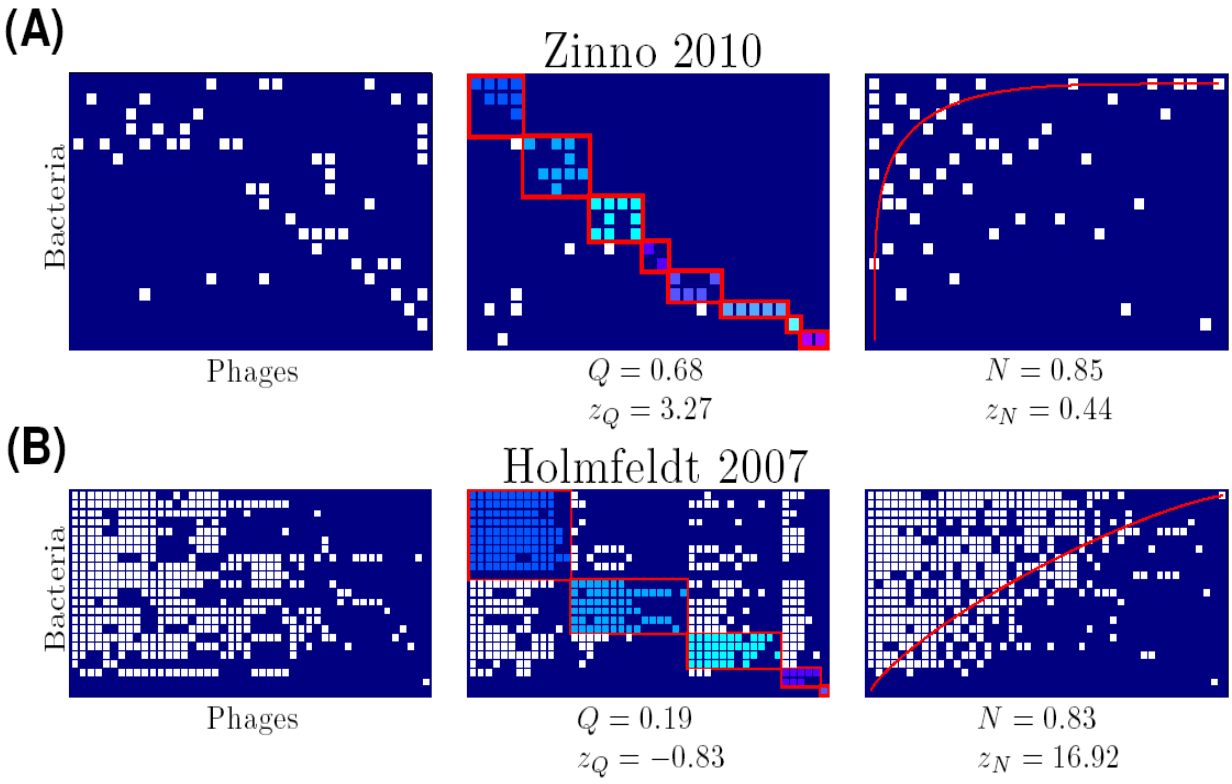


Figure. 3.3 Two example matrices re-sorted so as to maximize modularity and nestedness. The left matrix is the original data, the middle matrix is the output from the modularity algorithm (135), and the right matrix is the output from the modified nestedness algorithm (136, 137)**.** Colors represent different communities within the maximal modular configuration. (A) An example of a matrix with significantly-elevated modularity and insignificant nestedness. (B) An example of a matrix with insignificant modularity and significantly-elevated nestedness.

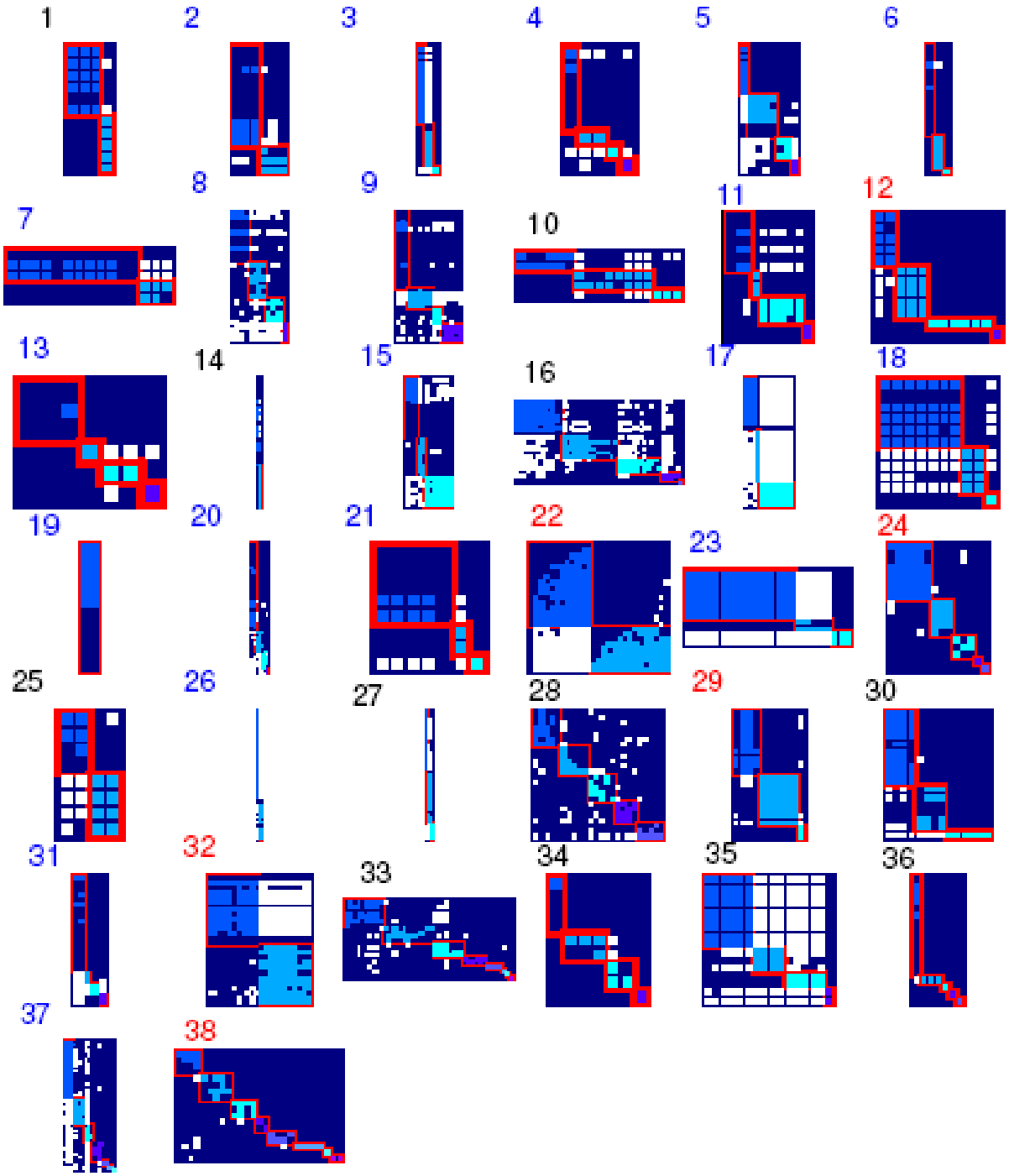


Figure. 3.4 Modularity sorts of the collected studies. Blue labels (20/38) represent studies statistically anti-modular and red (6/38) labels studies statistically modular.

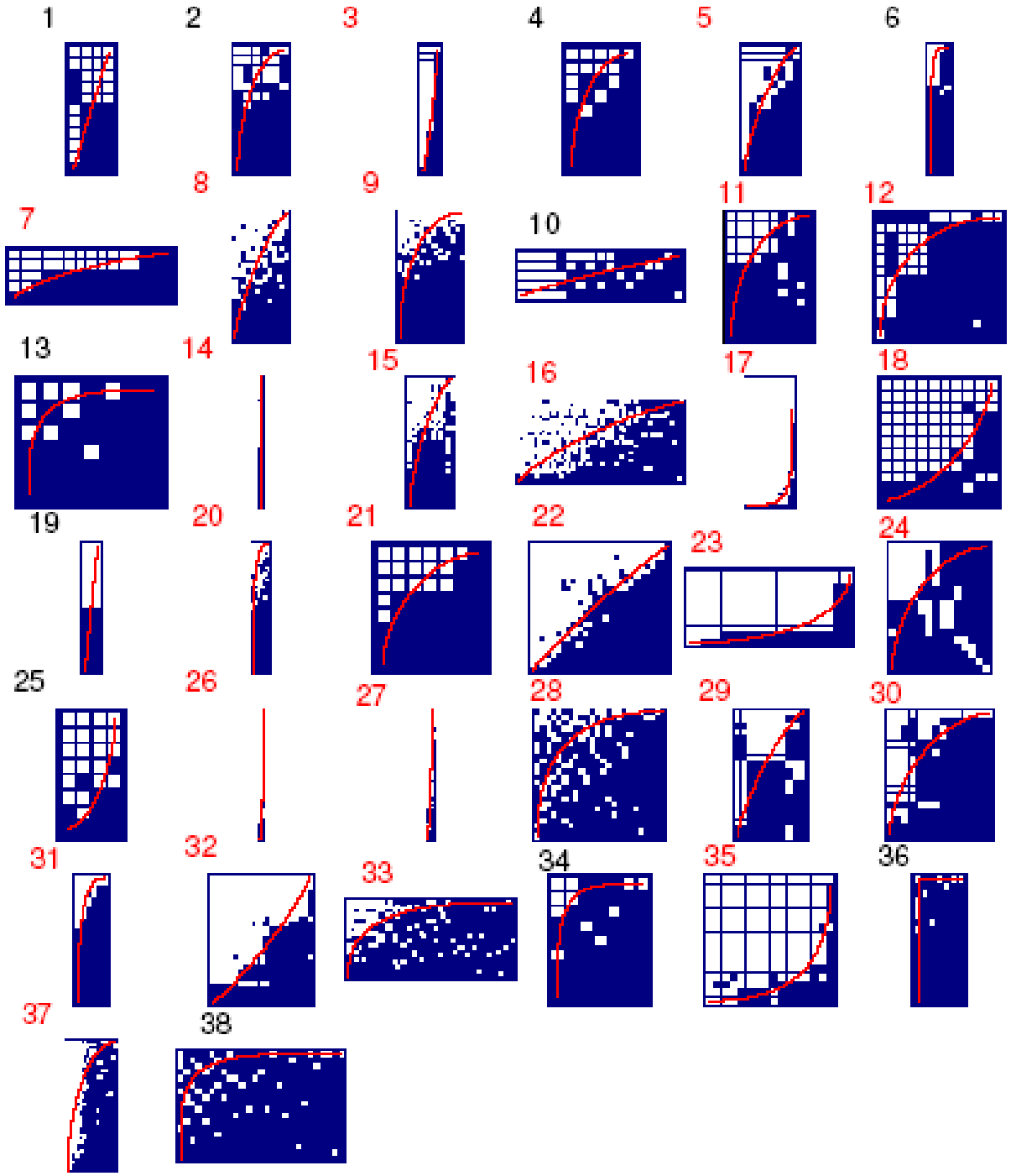


Figure. 3.5 Nestedness sorts of the collected studies. Red line represents the isocline. Blue labels (0/38) represent studies statistically anti-nested and red (27/38) labels studies statistically nested.

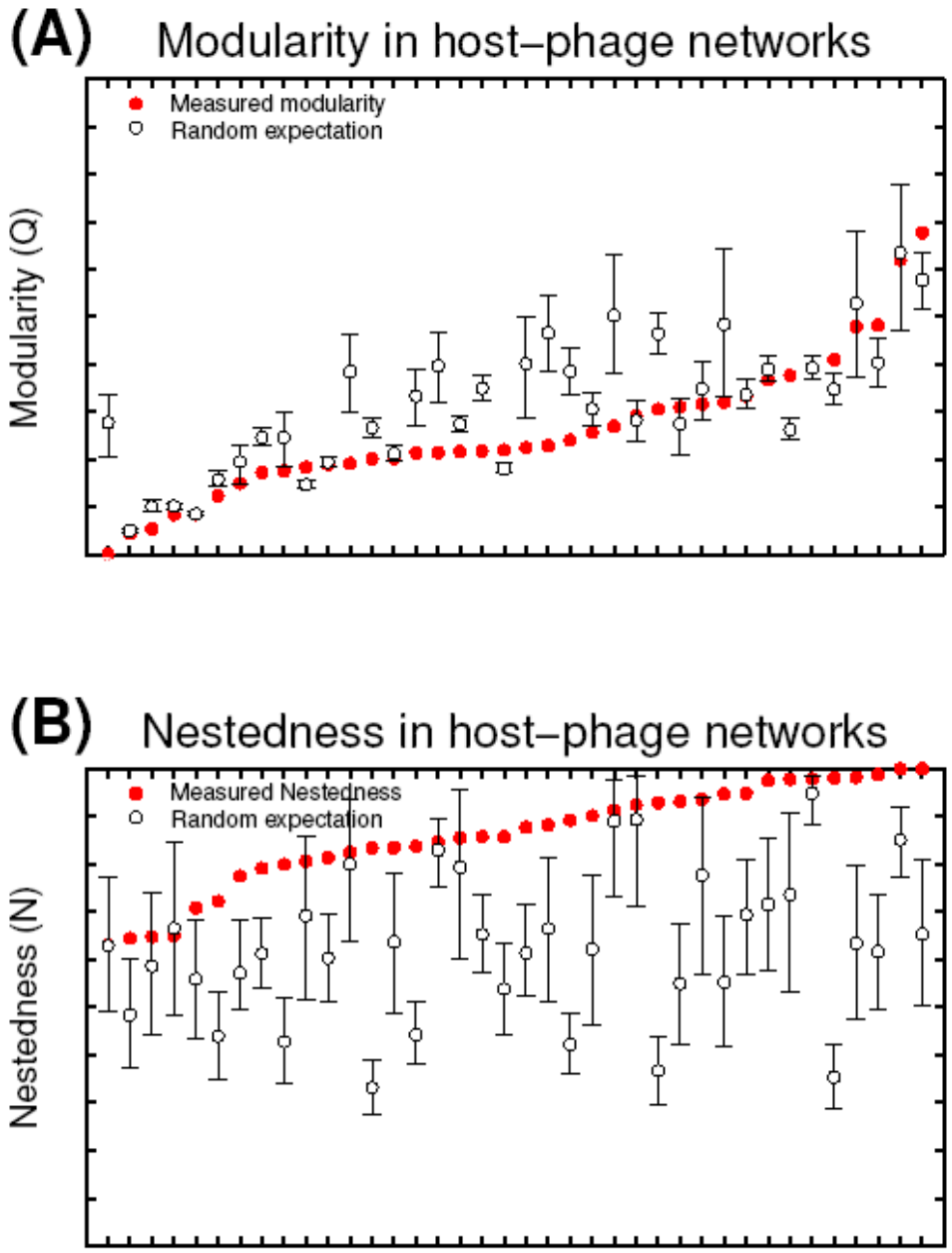


Figure. 3.6 Statistical distribution of modularity and nestedness for random matrices compared to that of the original data (increasing y-values indicate greater values of each). (A) Sorted comparison of modularity of the collected studies vs random networks. (B) Sorted comparison of nestedness of the collected studies vs random networks. In both cases, error bars denote *95 %* confidence intervals based on 10^5 randomizations.

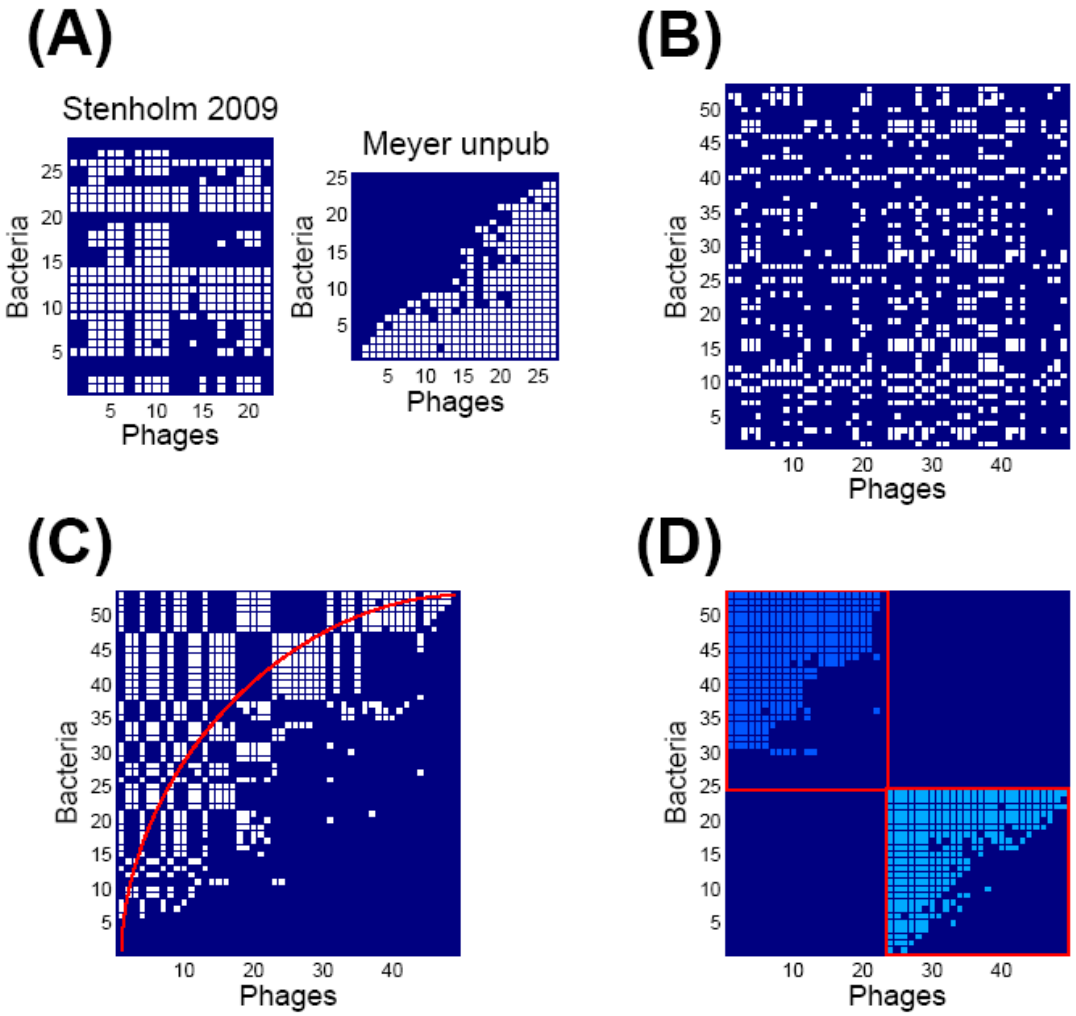


Figure. 3.7 Union of two nested matrices indicates possible host-phage interaction structure at larger, possibly macro-evolutionary scales. In this figure we selected two of the most nested studies and performed a union while presuming that there were no cross-infections of hosts by phages of the other study. In this case, *E. coli* and cyanobacteria were the host types. A) Depiction of the original matrices. B) Randomization of the union matrix. C) Nested sort of the union matrix. D) Modularity sort of the union matrix, with a nested sort of each module.

LITERATURE CITED

LITERATURE CITED

1. Edwards, R. A. & Rohwer, F. 2005. Viral metagenomics. Nature Reviews Microbiology 3: 504-510.

2. Torsvik, V., Øvreås, L. & Thingstad, T. F. 2002. Prokaryotic diversity -- magnitude, dynamics, and controlling factors. Science 296: 1064-1066.

3. Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323: 741-746.

4. Venter, J. C. et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74.

5. Huse, S. M. et al. 2008. Exploring microbial diversity and taxonomy using SSU rRNA Hypervariable Tag Sequencing. PLoS Genet. 4: e1000255+-e1000255+.

6. Angly, F. E. et al. The marine viromes of four oceanic regions. 2006. PLoS Biol. 4: 2121-2131 .

7. Yooseph, S. et al. 2007. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS Biol. 5: 432-466.

8. Cox-Foster, D. L. L. et al. 2007. A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. Science 318: 283-287.

9. Tyson, G. W. et al. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37-43.

10. Gill, S. R. et al. 2006. Metagenomic analysis of the human distal gut microbiome. Science 312: 1355-1359.

11. Tringe, S. G. et al. 2005. Comparative metagenomics of microbial communities. Science 308: 554-557.

12. Levin, B. R. & Bull, J. J. 2004. Population and evolutionary dynamics of phage therapy. Nature Reviews Microbiology 2: 166-173.

13. Rohwer, F. & Thurber, R. V. 2009. Viruses manipulate the marine environment. Nature 459: 207-212.

14. Faruque, S. M. et al. 2005. Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. Proc. Natl. Acad. Sci. U. S. A. 102: 6119-6124.

15. Faruque, S. M. et al. 2005. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc. Natl. Acad. Sci. U. S. A. 102: 1702-1707.

16. Goodridge, L. & Abedon, S. T. 2003. Bacteriophage biocontrol and bioprocessing: application of phage therapy to industry. SIM News 53: 254-262.

17. Weinbauer, M. G. 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28: 127-181.

18. Wommack, K. E. & Colwell, R. R. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64: 69-114.

19. Abedon, S. T. 2008. Bacteriophage ecology: population growth, evolution and impact of bacterial viruses. (Cambridge University Press).

20. Boyd, E. F., Davis, B. M. & Hochhut, B. 2001. Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol. 9: 137-144.

21. Silander, O. K. et al. 2005. Widespread genetic exchange among terrestrial bacteriophages. Proc. Natl. Acad. Sci. U. S. A. 102: 19009-19014.

22. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. 2004. Movement of viruses between biomes. Appl. Environ. Microbiol. 70: 5842-5846.

23. Fuhrman, J. A. & Noble, R. T. 1995. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol. Oceanogr. 40: 1236-1242.

24. Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M. & Sañudo-Wilhelmy, S. 1997. Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. Limnol. Oceanogr. 42: 1492-1504.

25. Bohannan, B. J. M. & Lenski, R. E. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Lett. 3: 362-377.

26. Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. 45: 1320-1328.

27. Poullain, V., Gandon, S., Brockhurst, M. A., Buckling, A. & Hochberg, M. E. 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62: 1-11.

28. Agrawal, A. & Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4: 79-90.

29. Abe, M., Izumoji, Y. & Tanji, Y. 2007. Phenotypic transformation including host-range transition through superinfection of T-even phages. FEMS Microbiol. Lett. 269: 145-152.

30. Barrangou, R., Yoon, S. S., F., Fleming, H. P. & Klaenhammer, T. R. 2002. Characterization of six Leuconostoc fallax bacteriophages isolated from an industrial sauerkraut fermentation. Appl. Environ. Microbiol. 68: 5452-5452.

31. Braun-Breton, C. & Hofnung, M. 1981. Invivo and invitro functional alterations of the bacteriophage-lambda receptor in LamB missense mutants of Escherichia coli K-12. J. Bacteriol. 148: 845-852.

32. Campbell, J. I. A., Albrechtsen, M. & Sørensen, J. Large 2006. Pseudomonas phages isolated from barley rhizosphere. FEMS Microbiol. Ecol. 18: 63-74.

33. Capparelli, R. et al. 2010. Bacteriophage therapy of Salmonella enterica: a fresh appraisal of bacteriophage therapy. The Journal of Infectious Diseases 201: 52-61.

34. Caso, J. L. et al. 1995. Isolation and characterization of temperate and virulent bacteriophages of Lactobacillus plantarum. J. Dairy Sci. 78: 741-750.

35. Ceyssens, P.-J. et al. 2009. Survey of Pseudomonas aeruginosa and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. Environ. Microbiol. 11: 1303-1313.

36. Comeau, A. M., Buenaventura, E. & Suttle, C. A. A 2005. persistent, productive, and seasonally dynamic vibriophage population within Pacific oysters (Crassostrea gigas). Appl. Environ. Microbiol. 71: 5324-5324.

37. Comeau, A. M., Chan, A. M. & Suttle, C. A. 2006. Genetic richness of vibriophages isolated in a coastal environment. Environ. Microbiol. 8: 1164-1176.

38. DePaola, A., Motes, M. L., Chan, A. M. & Suttle, C. A. 1998. Phages infecting Vibrio vulnificus are abundant and diverse in oysters (Crassostrea virginica) collected from the Gulf of Mexico. Appl. Environ. Microbiol. 64: 346-351.

39. Doi, K. et al. 2003. A comparative study and phage typing of silage-making Lactobacillus bacteriophages. J. Biosci. Bioeng. 95: 518-525.

40. Duplessis, M. & Moineau, S. 2001. Identification of a genetic determinant responsible for host specificity in Streptococcus thermophilus bacteriophages. Mol. Microbiol. 41: 325-336.

41. Gamage, S. D., Patton, A. K., Hanson, J. F. & Weiss, A. A. 2004. Diversity and host range of Shiga toxin-encoding phage. Infect. Immun. 72: 7131-7131.

42. Goodridge, L., Gallaccio, A. & Griffiths, M. W. 2003. Morphological, host range, and genetic characterization of two coliphages. Appl. Environ. Microbiol. 69: 5364-5364.

43. Hansen, V. M., Rosenquist, H., Baggesen, D. L., Brown, S. & Christensen, B. B. 2007. Characterization of Campylobacter phages including analysis of host range by selected Campylobacter Penner serotypes. BMC Microbiol. 7: 90-90.

44. Holmfeldt, K., Middelboe, M., Nybroe, O. & Riemann, L. 2007. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. Appl. Environ. Microbiol. 73: 6730-6739.

45. Kankila, J. & Lindstrom, K. 1994. Host range, morphology and DNA restriction patterns of bacteriophage isolates infecting Rhizobium leguminosarum bv. trifolii. Soil Biol. Biochem. 26: 429-437.

46. Krylov, V. N. et al. 2006. Ambivalent bacteriophages of different species active on Escherichia coli K12 and Salmonella sp. strains. Russ. J. Genet. 42: 106-114.

47. Kudva, I. T., Jelacic, S., Tarr, P. I., Youderian, P. & Hovde, C. J. 1999. Biocontrol of Escherichia coli O157 with O157-specific bacteriophages. Appl. Environ. Microbiol. 65: 3767-3773.

48. Langley, R. 2003. Lysogeny and bacteriophage host range within the Burkholderia cepacia complex. J. Med. Microbiol. 52: 483-490.

49. McLaughlin, M. R. & King, R. A. 2008. Characterization of Salmonella bacteriophages isolated from swine lagoon effluent. Curr. Microbiol. 56: 208-213.

50. Middelboe, M., Holmfeldt, K., Riemann, L., Nybroe, O. & Haaber, J. 2009. Bacteriophages drive strain diversification in a marine Flavobacterium: implications for phage resistance and physiological properties. Environ. Microbiol. 11: 1971-1982.

51. Miklič, A. & Rogelj, I. 2003. Characterization of lactococcal bacteriophages isolated from Slovenian dairies. Int. J. Food Sci. Technol. 38: 305-311.

52. Mizoguchi, K. et al. 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157: H7 in continuous culture. Appl. Environ. Microbiol. 69: 170-176.

53. Pantůček, R. et al. 1998. The polyvalent staphylococcal phage ϕ812: its host-range mutants and related phages. Virology 246: 241-252.

54. Paterson, S. et al. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464: 275-79.

55. Quiberoni, A. et al. 2003. Comparative analysis of Streptococcus thermophilus bacteriophages isolated from a yogurt industrial plant. Food Microbiol. 20: 461-469.

56. Rybniker, J., Kramme, S. & Small, P. L. 2006. Host range of 14 mycobacteriophages in Mycobacterium ulcerans and seven other mycobacteria including Mycobacterium tuberculosis--application for identification and susceptibility testing. J. Med. Microbiol. 55: 37-42.

57. Seed, K. D. & Dennis, J. J. 2005. Isolation and characterization of bacteriophages of the Burkholderia cepacia complex. FEMS Microbiol. Lett. 251: 273-280.

58. Stenholm, A. R., Dalsgaard, I. & Middelboe, M. 2008. Isolation and characterization of bacteriophages infecting the fish pathogen Flavobacterium psychrophilum. Appl. Environ. Microbiol. 74: 4070-4078.

59. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. 2003. Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature 424: 1047-1051.

60. Suttle, C. A. & Chan, A. M. 1993. Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. Mar. Ecol. Prog. Ser. 92: 99-109.

61. Synnott, A. J. et al. 2009. Isolation from sewage influent and characterization of novel Staphylococcus aureus bacteriophages with wide host ranges and potent lytic capabilities. Appl. Environ. Microbiol. 75: 4483-4490.

62. Wang, K. & Chen, F. 2008. Prevalence of highly host-specific cyanophages in the estuarine environment. Environ. Microbiol. 10: 300-312.

63. Wichels, A. et al. 1998. Bacteriophage diversity in the North Sea. Appl. Environ. Microbiol. 64: 4128-4133.

64. Zinno, P., Janzen, T., Bennedsen, M., Ercolini, D. & Mauriello, G. 2010. Characterization of Streptococcus thermophilus lytic bacteriophages from mozzarella cheese plants. Int. J. Food Microbiol. 138: 137-144.

65. Jaccard, P. 1912. The distribution of flora in the alpine zone. The New Phytologist 11: 37-50.

66. Memmott, J. 1999. The structure of a plant-pollinator food web. Ecol. Lett. 2: 276-280.

67. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. 2003. The nested assembly of plant-animal mutualistic networks. Proc. Natl. Acad. Sci. U. S. A. 100: 9383-9387.

68. Thompson, J. N. 2005. The geographic mosaic of coevolution. (University of Chicago Press).

69. Ulrich, W. & Gotelli, N. J. 2007. Null model analysis of species nestedness patterns. Ecology 88: 1824-1831.

70. Almeida-Neto, M., Guimarães, P. R. & Lewinsohn, T. M. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. Oikos 116: 716-722.

71. Bastolla, U. et al. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature 458: 1018-1091.

72. Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proceedings of the Royal Society of London Series B-Biologial Sciences 267: 2183-2188.

73. Forde, S. E. et al. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455: 220-223.

74. Labrie, S. J., Samson, J. E. & Moineau, S. 2010. Bacteriophage resistance mechanisms. Nature Reviews Microbiology 8: 317-327.

75. Hyman, P. & Abedon, S. T. 2010. in Adv. Appl. Microbiol. Vol. 70 (eds Allen I. Laskin, Sima Sariaslani, & Geoffrey M. Gadd) Ch. 7, 217-248 (Elsevier Academic Press Inc,).

76. Maynard, N. D. et al. 2010. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. PLoS Genet. 6: e1001017-e1001017.

77. Horvath, P. & Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167-170.

78. Karginov, F. V. & Hannon, G. J. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell 37: 7-19.

79. Gudelj, I. et al. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol Lett 13: 1073-1084.

80. Poisot, T., Lepennetier, G., Martinez, E., Ramsayer, J. & Hochberg, M. E. 2011. Resource availability affects the structure of a natural bacteria-bacteriophage community. Biol. Lett. 7: 201-204.

81. Bascompte, J. & Jordano, P. 2006. in Ecological networks: Linking structure to dynamics in food webs (eds M. Pascual & J. Dunne) 143-159 (Oxford University Press).

82. Bascompte, J. 2009. Disentangling the web of life. Science 325: 416-419.

83. Lenski, R. E. & Levin, B. R. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. 125: 585-602.

84. Buckling, A. & Rainey, P. B. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society of London Series B-Biologial Sciences 269: 931-936.

85. Gómez, P. & Buckling, A. 2011. Bacteria-phage antagonistic coevolution in soil. Science 332: 106-109.

86. Weitz, J. S., Hartman, H. & Levin, S. A. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc. Natl. Acad. Sci. U. S. A. 102: 9535-9535.

87. Spanakis, E. & Horne, M. T. 1987. Co-adaptation of Escherichia coli and Coliphage λvir in Continuous Culture J. Gen. Microbiol. 133: 353-360.

88. Cohen, J. E. & Newman, C. M. A. 1985. stochastic theory of community food webs I. Models and aggregated data. Proc. Roy. Soc. B 224:421-448.

89. Cohen, J. E., Newman, C. M. & Briand, F. A 1985. stochastic theory of community food webs II. Individual webs. Proceedings of the Royal Society of London Series B-Biologial Sciences 224: 449-461.

90. Pimm, S. L., Lawton, J. H. & Cohen, J. E. 1991. Food web patterns and their consequences. Nature 350: 669-674.

91. Cohen, J. E., Briand, F. & Newman, C. M. 1990. Community Food Webs: Data and Theory. (Springer-Verlag).

92. Allesina, S., Alonso, D. & Pascual, M. A. 2008. general model for food web structure. Science 320: 658-661.

93. Williams, R. J. & Martinez, N. D. 2000. Simple rules yield complex food webs. Nature 404: 180-183.

94. Pascual, M. & Dunne, J. A. 2005. in Santa Fe Institute Studies in the Sciences of Complexity Proceedings (Oxford University Press, New York).

95. Mestres, J., Gregori-Puigjane, E., Valverde, S. & Sole, R. V. 2008. Data completeness--the Achilles heel of drug-target networks. Nat. Biotechnol. 26: 983-984.

96. Olesen, J. M. et al. 2010. From Broadstone to Zackenberg: space, time and hierarchies in ecological networks. Adv. Ecol. Res. 42: 1-69.

97. Waser, N. M. & Ollerton, J. 2006. Plant-pollinator interactions: from specialization to generalization. (University of Chicago).

98. Fortunato, S. & Barthélemy, M. 2007. Resolution limit in community detection. Proc. Natl. Acad. Sci. U. S. A. 104: 36-41.

99. Genini, J., Morellato, L. P. C., Guimarães, P. R. & Olesen, J. M. 2010. Cheaters in mutualism networks. Biol. Lett. 6: 494-497.

100. Chiura, H. X. 1997. Generalized gene transfer by virus-like particles from marine bacteria. Aquat. Microb. Ecol. 13: 75-83.

101. Watts, D. J. & Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. Nature 393: 440-442.

102. Barber, M. 2007. Modularity and community detection in bipartite networks. Physical Review E 76, 066102.

103. Atmar, W. & Patterson, B. D. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96: 373-382.

104. Rodriguez-Girones, M. A. & Santamaria, L. 2006. A new algorithm to calculate the nestedness temperature of presence-absence matrices. J. Biogeogr. 33: 924-935.

105. Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. & Kim, J. F. 2008. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J. Mol. Biol. 394: 634-643.

CHAPTER 4

BIOPHYSICAL CAUSES OF VIRAL RESISTANCE AND PLEIOTROPIC COSTS IN *ESCHERICHIA COLI*

Authors: Justin R. Meyer, Devin T. Dobias, Robert Beardmore, Rafael Peña-Miller, Ryan T. Quick, Richard E. Lenski

Abstract

With improvements in biotechnology, the molecular details underlying large-scale ecological and evolutionary phenomena are coming into focus. One subject poised to gain from this new multi-scale approach is the study of ecological trade-offs. Trade-offs often originate from the thermodynamic constraints of biomolecules and have important effects on the ecology and evolution of species. Here we investigated the molecular basis of a trade-off between viral resistance and competitiveness among *lamB* mutants of the bacterium, *Escherichia coli*. LamB is an outer-membrane porin used by *E. coli* to transport sugars, however it is also the receptor for the virus, bacteriophage λ. All mutations studied provided some degree of resistance, although their strength varied, as did their pleiotropic costs. To study the molecular basis of this variation we developed a new procedure by combining protein shape modeling techniques with Geometric Morphometrics, a statistical approach able to relate protein structure to function. Resistance was strongly correlated with perturbations in LamB loop structure, whereas mutations that altered LamB's channel co-varied with competitiveness. Mutations evolved for defense typically had

the side effect of altering pore architecture, and thus produced a trade-off. However, many genotypes avoided the cost by acquiring mutations with focused effects on the loops. Here we describe the molecular basis of λ-resistance and its costs and discuss a new protocol to examine protein functional basis of phenotypes, including genetic diseases.

Introduction

A *Darwinian Demon* is a hypothetical organism that would evolve if evolution by natural selection proceeded unconstrained by the thermodynamic limits of living systems (1). If genetic variation existed that lacked plietropic costs, or was unbounded by physical limitations, then natural selection could optimize fitness in all phenotypic dimensions. This process would end in the evolution of a super organism that could occupy every niche and displace all existing species. Instead, the world observed is filled with millions of niche-differentiated species that because of physical constraints have evolved a particular set of functions, at the cost of performing others. Thus, these constraints are the causes of ecological trade-offs, and therefore are essential for the evolution and maintenance of biodiversity (2, 3).

Unfortunately, despite their crucial role in shaping biodiversity, the molecular mechanisms responsible for trade-offs are normally placed in a *black box* (2-4). This is despite the general appreciation that by knowing their molecular underpinnings, researchers would be better equipped to identify hidden trade-offs, predict how costs and benefits vary by environment, and improve the predictive power of ecological and evolutionary theory (3). Historically, the molecular underpinnings of trade-offs were ignored because the tools to study them were unavailable, however with the development of new genomic and biochemical

techniques, it is becoming possible. This is especially true for the study of microbes, who because of their short genomes and ease of genetic modification, are often at the leading edge of biotechnology.

A number of studies have demonstrated the strength of including a molecular description of trade-offs in their analyses (4-7). One outstanding example used modeled the coevolution of bacteria and their viruses, phage. They used coupled differential population dynamic models to describe the interactions between host and parasite, as well as the evolutionary dynamics by allowing the bacteria to evolve resistance and phage to evolve host-range expansions. They chose the effects of mutations based on a molecular understanding of the bacterial-phage interactions and how mutations conferred resistance and costs. With this model they were able to make accurate predictions for coevolutionary dynamics over a range of resource concentrations. They found that the trade-off between resistance and resource uptake resulted in a unimodal pattern of phage diversity. This was an important example of how one could use first costs derived from first principals of molecular biology to predict population biology level phenomenon.

One caveat to the success of this study is that depite making predictions for two bacterial-phage systems, their models only worked for one. The second, a model on *E. coli* and a another virus, bacteriophage λ, failed (Meyer and Lenski in prep). This model, and another like it (8), failed because their models were based on molecular studies of relatively few genotypes and not the diversity required to describe trade-offs and required to make accurate predictions about how the populations will coevolve. This failure represents a key problem with introducing molecular information into population models, which we argue arise from a fundamental conflict between the epistemological approach of molecular biology and the information required for population

models. Traditionally, molecular biologists have studied molecular interactions by examining a few mutations in great detail at the cost of studying genetic variation, however population genetic models require a comprehensive investigation of genetic variation to make predictions for evolution. On one hand, population biologists have failed to open the *black box*, however on the other hand, many molecular biologists have failed to characterize genetic variation. Here we describe research on λ and *E. coli* coevolution that accomplishes both. Our goal was to develop new methods to characterize the molecular basis of viral resistance and its costs for a large diversity of bacterial mutants.

The mechanics of infection and the genetics of resistance are well-established for λ and *E. coli*. λ targets and attaches to *E. coli* using the J protein at the end of the phage's tail (9,10). J docks to *E. coli*'s outer-membrane protein LamB and triggers infection. Typically mutations for λ-resistance stop infection at this early step by augmenting the regulation or amino acid sequence of LamB (11). LamB is used to transport a class of sugars (maltodextrins) through *E. coli*'s outer-membrane (12), and as a result, these mutations typically confer a pleiotropic cost when maltodextrins are present (Meyer in prep). Mutations that stop the expression of LamB endow complete resistance to most λ genotypes, but are fatal when maltodextrins are the only carbon source available (13,14). In contrast, mutations in the coding region of *lamB* have more subtle effects and provide varying degrees of resistance and costs (15). Genotypic variation in *lamB* is stably maintained, consistent with trade-off theory (Meyer in press). The source of the trade-off is likely an intrinsic biophysical constraint within LamB that prohibits simultaneous optimization of resistance and competitiveness, although its precise cause is unknown. LamB is a trimeric protein made of three beta barrel monomers, which produce channels that pierce the outer-membrane (16). Hydrophobic residues along the channels facilitate the translocation of sugars

(17) and large loops that extend outside the cell are thought to interact directly with λ's J protein (16).

For this study we characterized 52 *lamB* mutant *E. coli* that had evolved independently. The majority were unique, although a few mutations occurred in parallel in separate flasks. We also isolated 92 unique host-range mutant λ in order to measure *E. coli* resistance to a diversity of λ genotypes. Besides measuring resistance, we estimated how unfit each *lamB* mutant was compared to the sensitive ancestral genotype. Finally, we tested a series of three hypotheses for the molecular causes of resistance and costs.

Our first hypothesis was based on previous observations that mutations for resistance are most often located in the protein loops that extend outside of the cell and interact directly with the phage (15), whereas mutations that inhibit sugar transportation occur in the pore, the site that sugars are shunted into the cell through (18). We predicted that mutation location determines resistance and costs. Unfortunately, there appeared to be no relationships between location and the phenotypes they produce. Location alone may not be informative since changes of a single amino acid can cause deformation in protein shape that propagate like dominos throughout the protein (19), severing any detectible spatial associations between mutations and their effects. To study how changes in amino acids alter shape and their phenotypic effects, we modeled the mutant LamB shapes based on their mutated amino acid sequences, and then tested relationships between shape and resistance or competitiveness.

Our first shape-based analysis was designed to test whether mutations that were more disruptive to overall protein shape conferred greater resistance or costs. To do this, we calculated overall shape differences between each mutant and the ancestral LamB using methods

based on Euclidean distances. Next, we determined whether shape distances were correlated with resistance or competitiveness. We did not find a relationship, which lead us to hypothesize that rather than resistance being caused by global protein shape changes, there were particular deformations in shape that corresponded to increases in resistance. For example, shifts in the loop structures. To detect whether there were particular regions that correlated with increase of resistance of competitiveness, we performed a multivariate statistical test known as a Generalized Procrustes Analysis (GPA), which is a type of Geometric Morphometrics. Morphometrics is a technique typically used to characterize variation in animal morphology (20), however, when we applied it to study protein shape variation, the GPA uncovered linear deformations in LamB shape that were significantly correlated with resistance and competitive fitness. With this analysis, we were able to propose a molecular explanation for λ-resistance and the biophysical constraints underlying its costs.

Methods

Coevolution experiment protocol

We co-cultured *E. coli* strain B REL606 with the obligatorily lytic λ strain cI26. When co-cultured with cI26, REL606 experiences strong pressure to evolve resistance because *E. coli* B strains lack generalized phage defenses such as *mucoid* cell formation, restriction modification, or cRISPR adaptive immunity (21,22). This pressure is magnified by the lytic phage's increased virulence as compared to its lysogenic relatives (23). Once *E. coli* resistance evolved, the phage experiences selection to evolve counter-defenses, which triggers a rapid arms race within days.

100 flasks were initiated with few phage (~10^2 particles), and 10^3 bacterial cells that were preconditioned in the experimental environment for 24 hours. Small initial populations

increased the likelihood that mutations for defense and counter-defense arose *de novo*, which improved our chances of isolating unique *lamB* mutations and evolving divergent phage genotypes between replicates. The flasks were filled with 10ml of modified Davis Medium (DM) (24) (125 μg ml^{-1} maltotriose instead of glucose and 1 μg ml^{-1} of magnesium sulphate). Bacteria and phage were allowed to reproduce for 24 hours, at 37 °C, and shaken at 120 rpm. At 24 hours, a random 100 μl sample of each flask was added to a fresh flask and the bacteria and phage were allowed to grow again. This cycle was repeated once more and phage and bacteria were sampled after the third day of growth. We ended the experiment at this early time-point to ensure that the bacteria could only acquire a single mutation for defense, which made linking genotype to phenotype more tractable. Additionally, in previous experiments we observed the greatest genotypic diversity of bacteria and phage on the third day (Meyer and Lenski in prep). *E. coli* evolve resistance in this environment through many loci, however the most common is *lamB* (Meyer and Lenski in prep). For this study we will focus just on the *lamB* mutations.

Bacterial and phage isolation and storage procedures

To isolate bacteria we streaked a sample on Luria Bertani (LB) agar plates (25), randomly picked two colonies and then re-plated two more times to remove all phage. Finally we grew each colony in liquid LB overnight and preserved two 1ml samples in 15% glycerol and frozen at -80 °C. The entire phage population was preserved for each flask by chloroform preparation of the remaining volume of culture, ~8ml (26). Clonal isolates of the phage were created by picking plaques (miniature epidemics derived from a single phage particle) from bacterial *lawns* (films of bacteria immobilized in soft agar spread on top of Petri dishes) (26). For each phage population, we attempted to isolate phage from three separate lawns, one derived from the ancestral bacteria (REL606) and the two bacteria isolated from the very same flask. By

using the coevolved bacteria for phage isolation we increased the chance of sampling phage that had evolved specialized interactions with their coevolved bacteria, thereby improving our chances at uncovering more phage diversity than if we only used the ancestral bacteria. Two plaques from each flask were isolated and clonal cultures were created according to Adams 1959. When choosing phage, plaques were favored that formed on the lawns of the coevolved bacteria.

Estimating E. coli resistance to a library of phage

We measured resistance by challenging each bacterium with every phage isolate. To do this, we made *spot* plates by dripping ~2.5 µl of each phage stock on bacterial lawns (26). After 24 hours of growth at 37 °C the bacterial lawn would thicken, unless a phage was able to kill it, in this case a round clearing (spot) formed under the drip. Digital pictures were taken of each plate using an AlphaImager^{TM} 2200 by Alpha Innotech (pictures stored online in DRYAD) and analyzed with software designed to determine the level of clearing under each spot (See Appendix 3). This gave us a quantitative estimate for how well each bacterium resisted each phage. The full matrix of interactions was reported in a previous manuscript (27), for our current study we focused just on the bacteria that had evolved mutations in *lamB* and λ genotypes that had unique host-ranges in order to reduce the number of phage examined. We summed the clearing measurements across all phage genotypes to estimate the total sensitivity of each genotype. Sensitivity was natural log transformed for all of the analyses.

Measuring costs for resistance

Competitive fitnesses for *lamB* mutants were determined by competing each genotype head-to-head with a genetically marked version of the ancestor, REL607. REL607 can metabolize arabinose, whereas REL606 cannot because of a single nucleotide substitution that has little affect on the bacterium's fitness. Marked and unmarked genotypes can be distinguished on terazolium arabinose plates, which provides a tool to estimate the relative frequency of each in a mixed population. Full descriptions for competition experiments can be found (24, 28). In short, we initiated each flask with 50% of the resistant type and 50% of its *ara*$^{+}$ ancestor, cultured them for three days identically to the experiment except without phage, measured their initial and final densities, and calculated the ratio of the Malthusian parameters for the evolved versus the ancestor. We performed three replicates for each *lamB* genotype.

Sequencing

We sequenced *lamB* for at least one bacterial isolate from each flask, and the second isolate was sequenced if preliminary tests revealed the two sympatric isolates had different levels of resistance. Sequencing was performed with an automated ABI sequencer maintained at Michigan State University's Research Technology Support Facility. PCR amplified fragments purified with GFX columns were used as templates. Fragments were amplified with primer sequences 5' TTCCCGGTAATGTGGAGATGC 3' and 5' AATGTTTGCCGGGACGCTGTA 3', placed 1,398 bases up and 504 bases downstream of the gene, respectively.

Full genomes of 15 *E. coli lamB* mutants were sequenced to determine whether any other mutations occurred in the *E. coli*s' genomes that may influence resistance or competitive fitness. Isolates were chosen that we suspected had multiple λ-resistant mutations because they evolved high levels of resistance despite only possessing a single amino acid change in LamB or because

the genotype possessed a distinct resistance profile. Technicians at the Research Technology Support Facility at Michigan State University sequenced the genomes using an Illumina Genome Analyzer IIx. Genomic DNA samples were created by reviving frozen bacteria in LB medium, growing them over night, and then isolating DNA from several milliliters of the culture with Quiagen genome tips. Samples were fragmented by sonication, prepared with bar-coded attachments, and run as multiplexed samples over four lanes. Mutations were predicted from the resulting 75-base DNA single end reads using *breseq* v0.13 (www.sanger.ac.uk/resources/software/ssaha2/) and the ancestor's genome (Genbank accession: NC_012967.1) as the reference.

Protein shape prediction from amino acid sequence

Mutant protein shapes were predicted using the program Modeller (http://salilab.org/modeller/). Predictions were made in two steps: First, homology-based techniques were used to generate a protein structure by using the known ancestral LamB shape as a guide. LamB structure was determined by x-ray crystallography to 3.10-angstrom resolution (Genbank accession: YP_003047080) (16). Next, *de novo* loop refinement reorganized protein conformation to minimize entropy created by electrostatic conflicts introduced by the substituted amino acid.

Mutant-to-ancestor protein shape distances

Euclidean distance between the mutant LamBs and the ancestral form were calculated by comparing the X-, Y-, and Z-coordinates of each α-carbon for all 421 amino acids. Before comparing two proteins, we reoriented their structures to minimize the distance between the α-carbons. *E. coli* genotypes that possessed insertions, deletions, and premature stop codons in *lamB* were removed from the analysis, leaving only genotypes with a single amino acid

substitution (29 of the 52 *lamB* mutants), because this distance technique required conservation of all 421 amino acids.

Protein Geometric Morphometrics

In order to test whether there were specific LamB deformations that correlated with resistance or competitiveness, we used a Geometric Morphometrics version of linear regression, GPA. For this analysis we designated each α-carbon for all 421 amino acids as landmarks to compare mutant protein shapes. The same 29 mutant genotypes were used for this analysis as for the last. Next, we constructed a wireframe that matched the ordering of the polypeptide chain and performed a superimposition of all the 30 LamBs (29 mutants plus the ancestor) with Simple3D (freely available: http://www3.canisius.edu/~sheets/morphsoft.html). Superimposition is a technique that reorients and resizes shapes to minimize the distances between each landmark in order to isolate variation due to just changes in shape. Finally, we performed the regression based on the GPA of the LamB shapes against resistance and competitive fitness separately using IMP:ThreeDRegress (freely available: http://www3.canisius.edu/~sheets/morphsoft.html). This procedure determines which regions of protein deformation correlate with an independent variable by not just comparing the direct isometries of the landmarks as done with the Euclidean test, but also their reflections. Significance of the relationships were determined by generating Z-scores from the correlation coefficients. Lastly, for the analysis of resistance, but not competitiveness, we removed all mutant genotypes that were completely resistant. GPA requires independent variables that vary continuously and the '0' values represent levels of resistance that fell outside of our limit of detection. This reduced our sample size to 22.

Results

lamB and genome sequences

We isolated 52 independently evolved *lamB* single mutants, of which 29 had just single nucleotide replacements (SNPs) that resulted in a single amino acid replacement (Table 4.1). We focused the majority of our study on the 29 plus the ancestor because our analyses required conservation of peptide length. Of the 29 SNPs, 26 were unique. By sequencing the genomes of a subset of genotypes, we found that eight of fifteen possessed a second mutation, all of which were deletions of the ribose operon. Ribose mutations commonly occur in REL606 derivatives because the REL606 genome has a very active IS element (IS150) positioned near the ribose operon. These mutations do not affect phage resistance and are known to confer a 1% cost-saving advantage (29), and given the short time-scale of this experiment, are likely just genomic hitch hikers. Overall, the genome sequencing shows that the majority of change in resistance and fitness can be attributed to the *lamB* mutations.

Resistance by lamB mutations is low dimensional

Mutations in *lamB* provide a range of resistances, from genotypes that are only defended against the ancestral λ strain, to producing complete resistance (Figure 4.1). The pattern of resistance has two significant hierarchical structures: First, resistance is nested such that as *E. coli* gain resistance, they gradually loose sensitivity to the same sets of phage. This pattern continues along the continuum from completely resistant to sensitive creating a 'nested' appearance in their resistance profiles (Figure 4.1). Secondly, when two bacterial genotypes are both infected by the same virus, the bacterium that is resistant to more λ genotypes will also be more resistant to the shared virus (Figure 4.2). Together these observations show that resistance has evolved along a

single phenotypic dimension towards increased resistance. This is important to note because it suggests that the molecular mechanism for resistance should also vary in a single dimension.

Modest trade-off between resistance and competitiveness

Many of the *lamB* mutations were costly and we found a significant positive trend between sensitivity to λ and competitive fitness ($r^2 = 0.146$, p=0.0369, DF=28, slope = 14.157, intercept = -14.293.) indicating a modest trade-off (Figure 4.3). Costs likely occurred because the mutations interfered with sugar uptake, however we did not measure resource transport directly. Despite finding a significant trade-off, many genotypes did not pay a cost for resistance, including genotype '61B' which gained complete resistance without paying any price. This suggests that on average resistance in costly, although there is not a strict biophysical constraint prohibiting optimization of resistance and fitness.

No spatial pattern between mutation location and resistance or costs

In line with previous observations, the majority of mutations occur within the loops of the protein (58% of the mutations occurred in loops even though only 38% of LamB's amino acids occur in loops; $X^2=29.18$, $p<0.0001$, df=1), although visual inspection of the distribution of mutations does not indicate a clear pattern between location and resistance (Figure 4.4). We tested whether there were quantitative relationships between mutation location and resistance or competitiveness by drawling transects through the X-, Y-, and Z-axis of the protein and plotting the level of resistance or competitiveness with each mutation's location along the transect (4.2 and Figure 4.5). We found no relationship. This is best shown by examining genotypes 29A and 61A had mutations in the very same nucleotide site, however they caused different amino acid

substitutions (Table 4.1), which produce divergent LamB shapes (Figure 4.5 panels A versus B), levels of resistance (29A intermediate and 61A complete) and fitness costs (no cost for 29A, high cost for 61A). This pattern is repeated for other pairs of mutations as well (61B & 99A, 4A & 14A, and 22A & 62B).

Global perturbations to LamB shape are not correlated with resistance of competitiveness

As previously discussed, spatial relationships between mutation location and resistance may fail if amino acid changes have cascading effects in other protein regions. This was the case for the majority of mutant proteins, our LamB predictions show that multiple regions spread throughout the protein are perturbed by just single amino acid replacements (Figure 4.5). Considering that mutations may have broad effects on protein shape, we hypothesized that the more disruptive mutations will confer greater resistance, with substantial penalties on competitive fitness. We did not find support for either of these relationships (resistance $r^2 = 0.074$, p=0.1456 df= 27; fitness $r^2 = 0.070$, p=0.1590 df= 27), some of the most resistant and costly genotypes had LamB shapes predicted to be barely deformed from the ancestor (Figure 4.6).

Specific LamB deformations correlate with resistance and competitiveness

The global analysis of protein shape may have failed if only specific changes are required for changes in resistance or competitiveness, and if these changes were overshadowed by other incidental alterations. To test whether there are specific deformations in the protein that correspond to increased resistance or competitiveness we performed a standard three-dimensional shape-based regression technique, GPA. With this analysis we uncovered regions that share a relationship with resistance (Figure 4.7 top panel; N = 22, corr. coef. = 0.481, Z-score = 2.284, p=0.0111) and competitiveness (Figure 4,7 bottom panel; N = 30, corr. coef. =

0.350, Z-score = 1.899, p=0.0228). For resistance, the regions that are most correlated occur near amino acids 154-157 in loop four, but also include amino acids in adjacent loops five and six. Regions for competitiveness are not in the loops, but occur in the pore around amino acids 350-360 and augment the dimensions of the pore. There is very little overlap in the sites important for λ-resistance and competitive fitness (Figure 4.8).

Discussion

A significant barrier for incorporating molecular mechanisms into ecological and evolutionary theory is the scarcity of descriptions for the molecular basis of genetic diversity. However, new high throughput techniques are being developed through Systems Biology approaches that are suited to study genetic diversity. Here we propose one such strategy and used it to study *lamB* mutations in *E. coli* that confer varying degrees of defense and plietropic costs. With this strategy we were able to rejected two hypotheses and find support for a third hypothesis on the molecular basis of resistance and costs.

Molecular changes underlying λ-resistance

Specific changes in LamB structure were correlated with increases in resistance, particularly deformations in loop four (Figure 4.9). Interestingly, previous studies on *lamB* mutations had also implicated changes in loop four as causing defense, because more resistance mutations occurred there than anywhere else (16). This was not the case for our study, the majority of the mutations fell in the β barrels and loop 6 (Table 4.1), however we were able to uncover the importance of loop four because our analysis examined patterns of protein shape and not just mutation location. Traditionally, when new *lamB* mutations were uncovered for resistance, it was assumed that they represented a new binding site of λ. However, given that many λ-

resistance mutations occur in the β-barrels shielded from direct phage contact by the lipid bi-layer and our results from the shape analysis, we propose a new model for the molecular basis of resistance. Instead of many binding sites, there are relatively few locations required for λ adsorption, likely just the area near loop four, and resistance is evolved through mutations that augment loop four by landing directly within the loop or somewhere that also affects its conformation.

Beyond providing a more parsimonious model, this explanation is much more in line with our observation that resistance evolves along a single dimension. If there were in fact many possible binding sites, then coevolution would have the opportunity to promote resistance to evolve in many dimensions and should produce a very different pattern of defense. Interestingly, one-dimensional changes in another outer-membrane *E. coli* molecule (lipopolysaccharides) are responsible for conferring resistance to another phage, T7. As for coevolution with λ, *E. coli* also evolves nested resistance with T7 (30). In fact, the nested pattern in bacterial-phage infection matrices is wide spread, and one hypothesis for its ubiquity is that bacterial-phage molecular interactions may often be constrained to low dimensional spaces (28).

Biophysical constraints underlying costs for resistance

Much like our finding that loop four is important for resistance, our discovery that pore dimensions affect competitive fitness is in line with a previous study that revealed mutations seated within LamB's channel inhibit sugar transport (19). Additionally, because the authors studied mutations that did not confer λ-resistance, they too concluded that separate aspects of the protein were important for resistance and sugar transport. Unlike their study, we only examined *lamB* mutations that conferred λ-resistance, however we were able to draw the same conclusion

because our statistical analysis could disentangle the variation responsible for resistance from the variation most associated with costs. This result suggests that because separate regions of the protein are responsible for conferring resistance and sugar transport, there is not necessarily a conflict between optimizing each. However, because protein structure is highly interdependent, many mutations that confer resistance have a plietropic consequence on pore shape and therefore fitness. This provides a molecular explanation for why in general there is a trade-off between resistance and competitiveness, however not all resistant genotypes pay a cost.

Method for building causal and statistical connections between genotype and phenotype

Beyond our results for λ resistance, another significant contribution of this work is to introduce Geometric Morphometrics to the study of protein structural and functional variation. To the best of our knowledge, only a few protein studies have used morphometrics, but only to characterize variation and not to deduce function (31). However, like our study, another used multivariate statistics on shape to reveal how function covaried with aspects of protein shape. The researchers used machine learning techniques in place of morphometrics to find multivariate associations between mutant proteins and the occurrence of human disease (32). As with our study, their protein shapes were predicted from gene sequences that had SNPs.

One advantage of Geometric Morphometrics over machine learning is that changes in protein size and orientation would not affect the analysis. Although, an advantage to machine learning protocols are their ability to reveal more complex structural associations that involve interactions between landmarks, which our linear regression strategy cannot. As demonstrated by the machine learning study, we imagine that our protocol could also be automated to search

SNP databases and generate hypothesis for the molecular basis of phenotypes, including human genetic diseases (33).

One limitation with our analysis is that we did not measure protein structures directly, but instead predicted them from *lamB* sequences. Modeling structures is much more efficient than using x-ray crystallography, however modeling estimates are less reliable. Provided this caveat, our analysis likely avoided many of the pitfalls with protein modeling because the predicted structures only diverged from the known LamB by a single amino acid change. For a full discussion of the problems with protein modeling see Martí-Renom et al. 2000. Another limitation of protein modeling and structural biology in general, is that our analysis treats proteins as static shapes. Instead they are dynamic, especially when interacting with sugar substrates (35) and phage (36). Our analysis would miss changes in protein dynamics that are not reflected in its resting shape. Future work could be conducted to link bioinformatics techniques that predict protein qualities such as flexibility or stability to phenotypic qualities like resistance or sugar transport.

A second caveat to our procedure is that the multivariate statistics we used are susceptible to finding spurious correlations. The risk of type I error is inflated because the number of landmarks out-number the sample size. For this reason, the statistical test we used that was based on the correlation coefficients does not provide a reliable test for significance, and so our results should be interpreted as hypotheses for the molecular basis of resistance and costs and not the definitive test. Although, holding this caveat in mind, the strong agreement between the results from previous studies on *lamB* mutational effects and the regions GPA identified is compelling. For future analyses we are developing tools to reduce the number of free variables with PCA and then adding a permutation-based approach for determining more reliable

significance estimates. Ultimately, the strongest test of our hypotheses for the molecular basis of resistance and costs would be to generate new *lamB* mutants and make *a priori* predictions for their phenotypes by using the combination of protein structure models and the regression vectors we generated from the GPA.

Conclusion

Many ecological and evolutionary phenomena emerge from the dynamics of biomolecules. This is especially true for the study of ecological trade-offs since they originate from physiochemical constraints, yet contribute to large-scale processes such as ecological community assembly (37) and speciation (38). Here we developed a new method for generating hypotheses for the molecular source of functional variation among mutant proteins. This tool can be used to find the molecular underpinnings of ecologically relevant traits such as disease resistance and competitiveness, as well as the constraints that produce trade-offs. With this tool we uncovered the likely molecular sources for *E. coli* resistance to λ, protein constraints that yield pleiotropic costs, and a physiochemical explanation for why there is not a strict trade-off between resistance and competitiveness.

TABLE 4.1: *lamB* single nucleotide replacements.

Isolate ID	Nucleotide location	Nucleotide change	Amino acid location	Amino acid change	Protein structural location
1A	850	T→G	258	Phe→Val	loop 6
4A	854	C→A	259	Ala→Asp	loop 6
11A	556	T→G	160	Tyr→Asp	loop 4
14A	853	G→T	259	Ala→Ser	loop 6
18A	538	T→G	154	Phe→Val	loop 4
22A	1210	G→T	378	Asp→Tyr	β barrel
23B	793	A→C	239	Thr→Pro	β barrel
29A	518	A→G	147	Glu→Gly	β barrel
30A	571	G→A	165	Glu→Lys	loop 4
41A	284	A→C	69	Gln→Pro	β barrel
47A	850	T→A	258	Phe→Ile	loop 6
50B	874	G→T	266	Gly→Cys	β barrel
51A	824	C→A	249	Ser→Tyr	loop 6
51B	843	C→A	255	Asp→Lys	loop 6
52B	809	G→T	244	Gly→Val	loop 6
57B	788	C→T	237	Ser→Leu	β barrel
59A	1285	G→C	403	Gly→Arg	β barrel
60A	709	G→T	211	Gly→Cys	β barrel
61A	518	A→C	147	Glu→Val	β barrel
61B	559	G→T	161	Asp→Tyr	loop 4
62B	1211	A→G	378	Asp→Gly	β barrel
68A	815	C→T	246	Ser→Leu	loop 6
70A	793	A→C	239	Thr→Pro	β barrel
71A	793	A→C	239	Thr→Pro	β barrel
71B	809	G→T	244	Gly→Val	loop 6
95A	821	G→A	248	Gly→Glu	loop 6
96B	581	A→C	168	Asn→Thr	β barrel
97B	580	A→C	168	His→Thr	β barrel
99A	560	A→G	161	Asp→Gly	loop 4

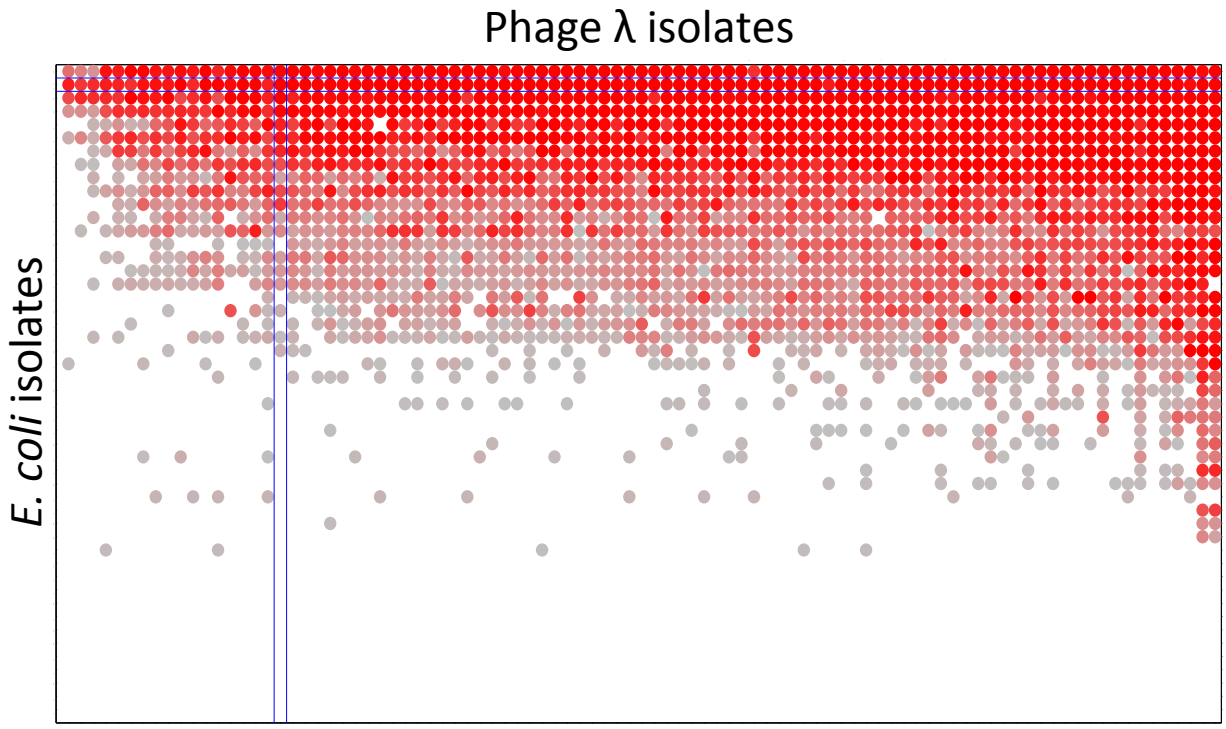


Figure 4.1: Interaction matrix between *lamB* mutant bacteria (aligned in rows) and coevolved phage λ genotypes (columns). A blank intersection indicates a combination that results in an unsuccessful infection. Dots indicate a successful infection, and their level of success, or relative clearing on the *spot* plate, is denoted by how bright red the dot is versus muted grey. Blue lines flank the row and column depicting the interactions with the ancestors.

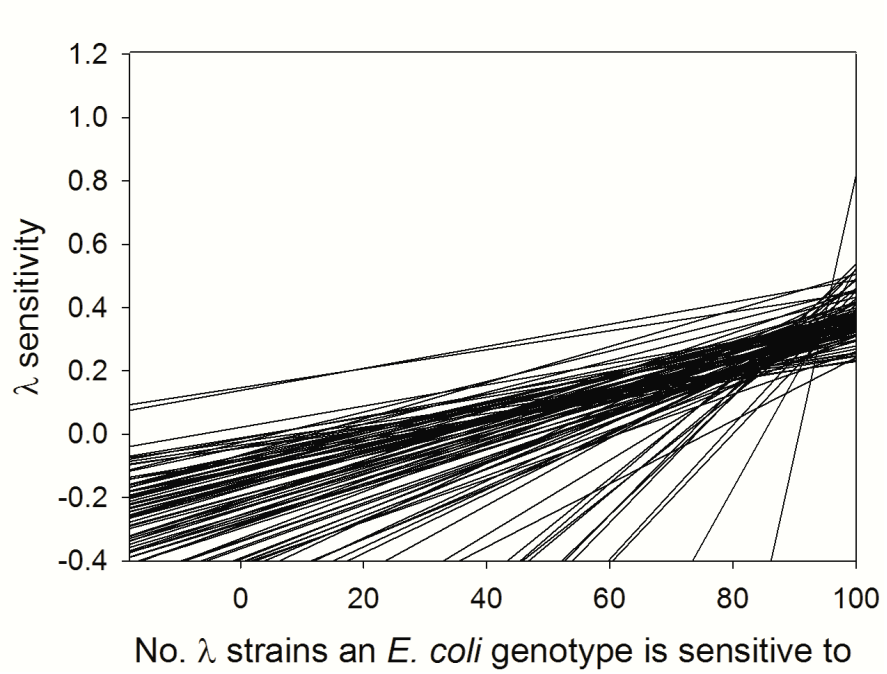


Figure 4.2: Relationship between how well a given phage can lyse a bacterial strain and how many phage the bacteria is sensitive to. Each line represents a separate phage. All lines have a positive slope, meaning as bacteria gain resistance to different genotypes of phage they increase their resistance to the set of phage they remain sensitive to. This pattern, along with the nested pattern in Figure 4.1 suggests defense evolves along a single axis of increased resistance.

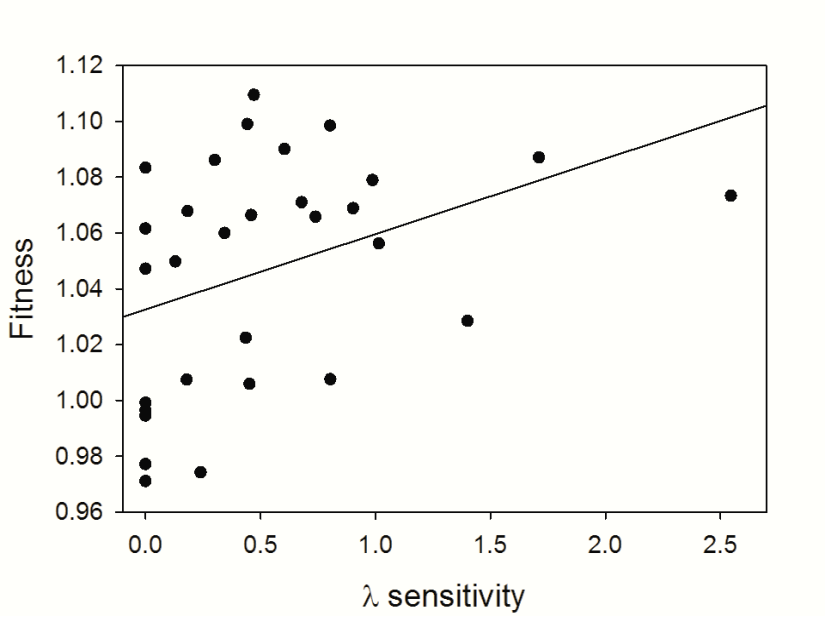


Figure 4.3: A positive relationship between sensitivity and fitness, indicating a trade-off between resistance and competitiveness. Each point represents an independently evolved *lamB* mutant, or the ancestor. λ-sensitivity is the total sum of clearing produced by all phage on the *spot* plates. Note, the REL607 strain we used had a fitness disadvantage of 7.3%, meaning any fitness value falling under 1.073 indicates a cost.

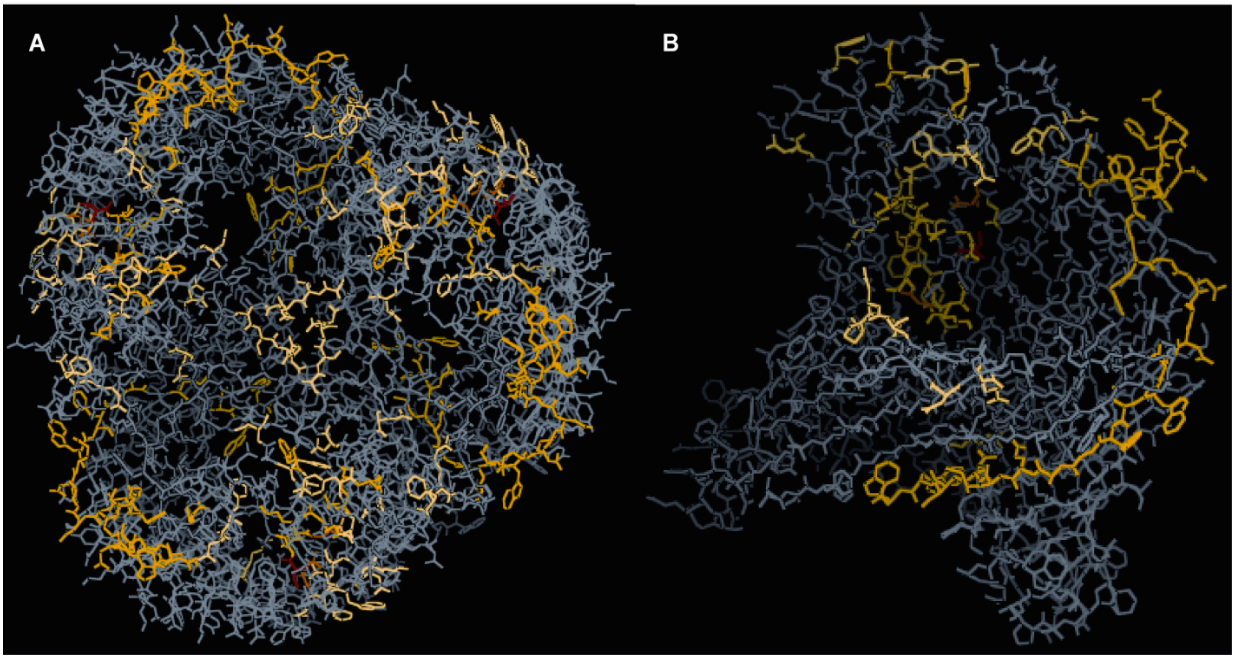


Figure 4.4: Distribution of mutations projected onto the three-dimensional confirmation of the viral receptor. A color spectrum from yellow to red was used to indicate the frequency of mutations observed at each amino acid; yellow = 1 through red = 5, gray = 0. Panel **A**, provides a 'phage eye' view of a trimer of LamB, this is the view from outside the cell. Panel **B** is a side view of a single monomer of LamB.

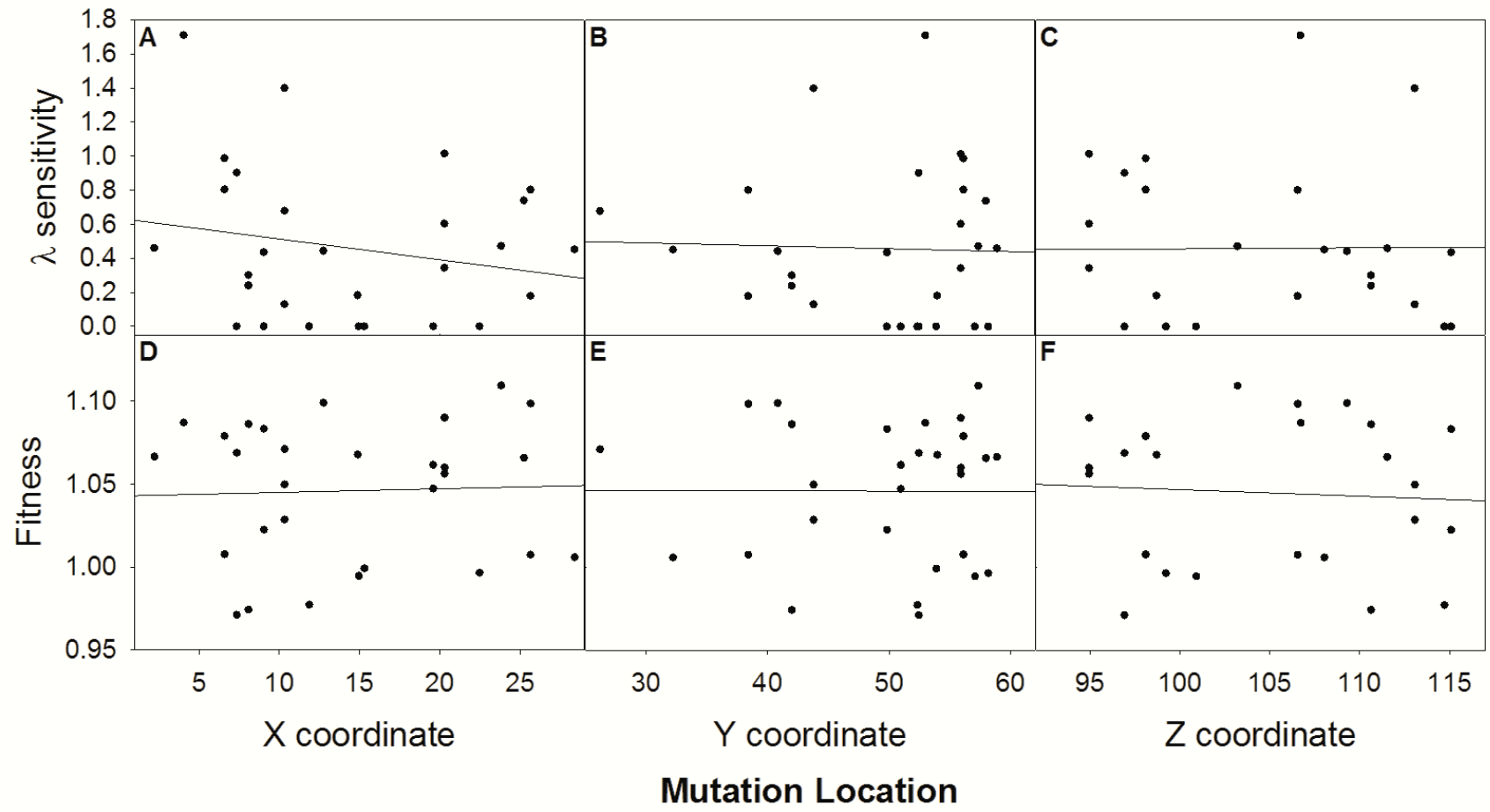


Figure 4.5: Spatial relationship between mutation location along the X (A and D), Y (B and E) and Z (C and F) axis with resistance (A, B, and C) or competitiveness (D, E, and F). Each dot represents an estimate for an independently evolved genotype.

Table 4.2: Regression results for relationships between mutation position and λ sensitivity or fitness relationships.

Figure 4.5 panel	Phenotype	Axis	r^2	p	slope	Intercept
A	λ sensitivity	X	0.04	0.2927	-0.0603	1.333
B	λ sensitivity	Y	0.001	0.868	-0.0534	3.102
C	λ sensitivity	Z	0.000	0.944	0.497	-4.629
D	Fitness	X	0.002	0.8398	0.00559	0.964
E	Fitness	Y	0.000	0.9857	-0.00494	1.291
F	Fitness	Z	0.007	0.6575	-0.00461	1.517

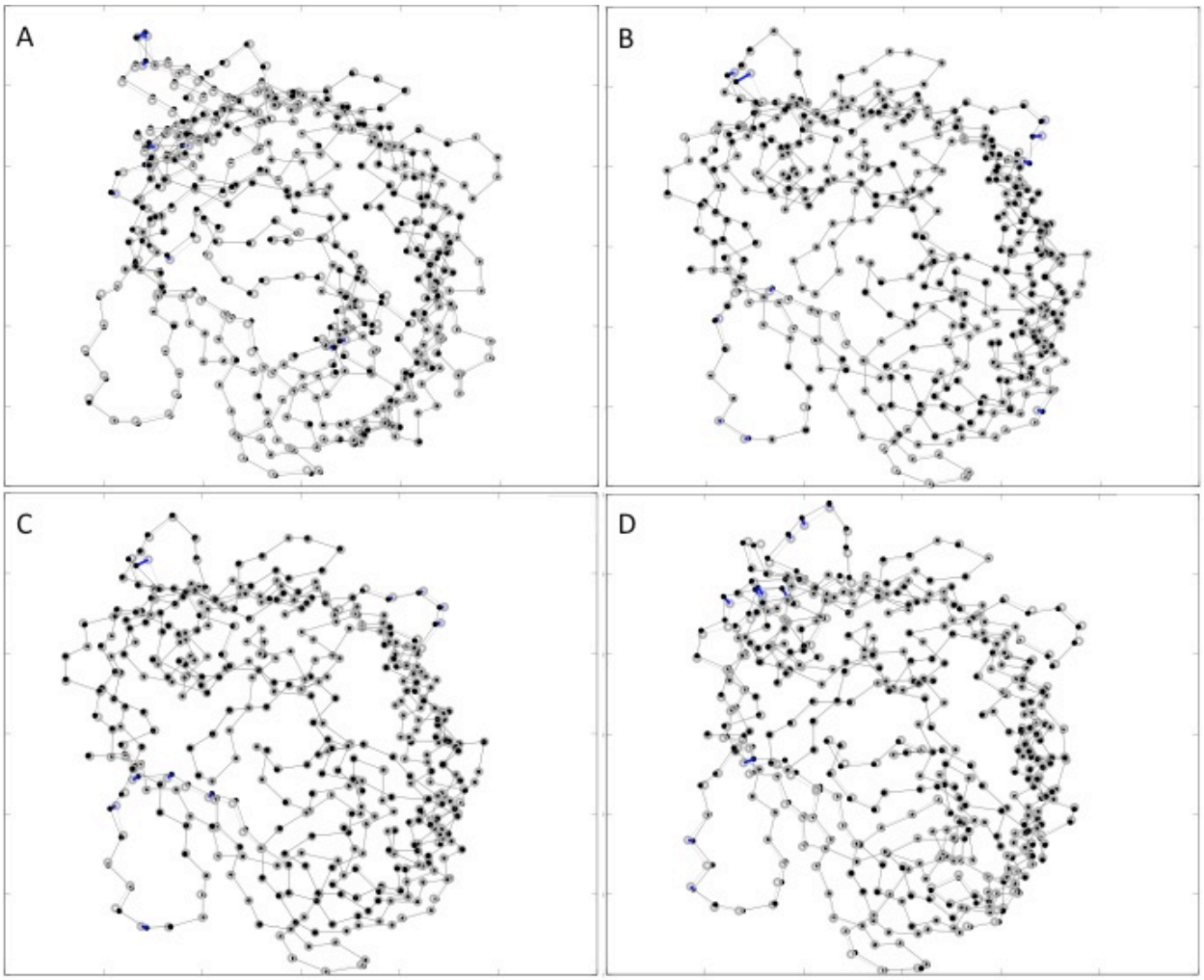


Figure 4.6: Four examples of LamB deformations predicted for different single nucleotide replacements (A: isolate 29a Glu→Gly at AA 147; B: isolate 61a Glu→Val at AA 147; C: isolate 18a Phe→Val at AA 154; D: isolate 59a Gly→Arg at AA 403;). Black dots indicate the ancestral location of each α-carbon for the ancestral LamB. Grey circles indicate the new predicted locations, blue lines highlight the top 8 largest changes in landmark location for each mutant.

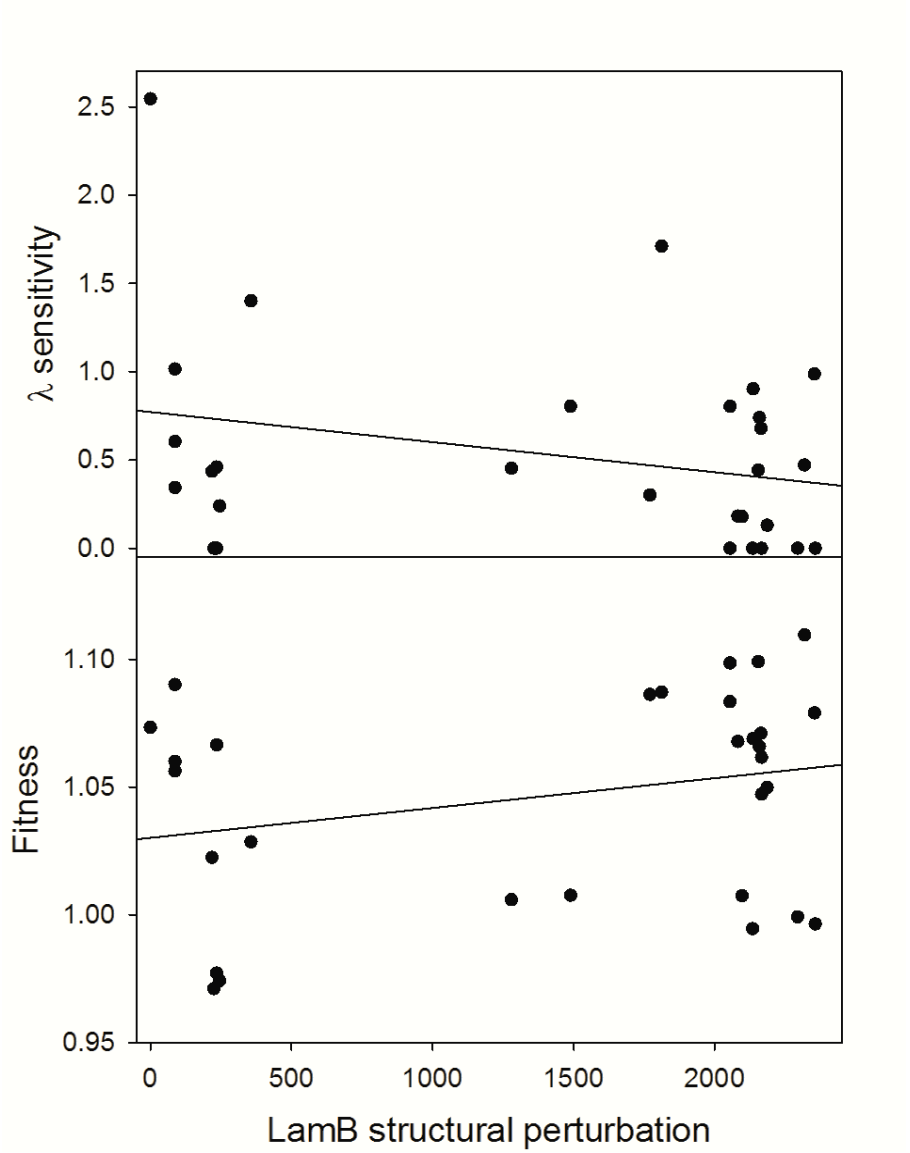


Figure 4.7: Relationship between overall protein deformations, calculated as the Euclidean distance between all α-carbons in a mutant LamB and the ancestor, and resistance (top panel) or competitiveness (bottom panel). Each point represents an estimate for an independently evolved *E. coli* genotype.

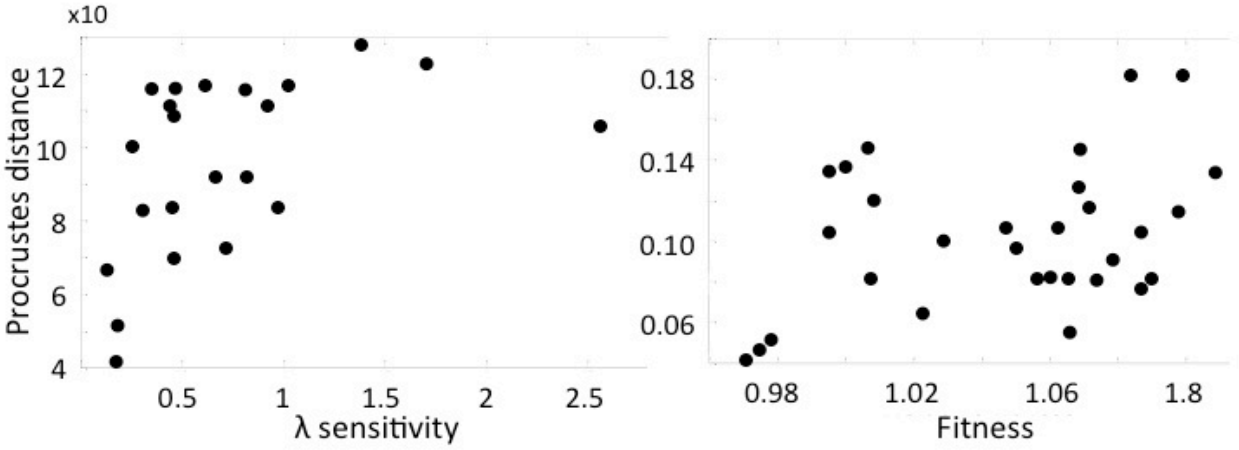


Figure 4.8: Correlations between Procrustes distances computed among mutant LamB shapes and sensitivity to λ or competitiveness. Each point represents an independently evolved *lamB* genotype.

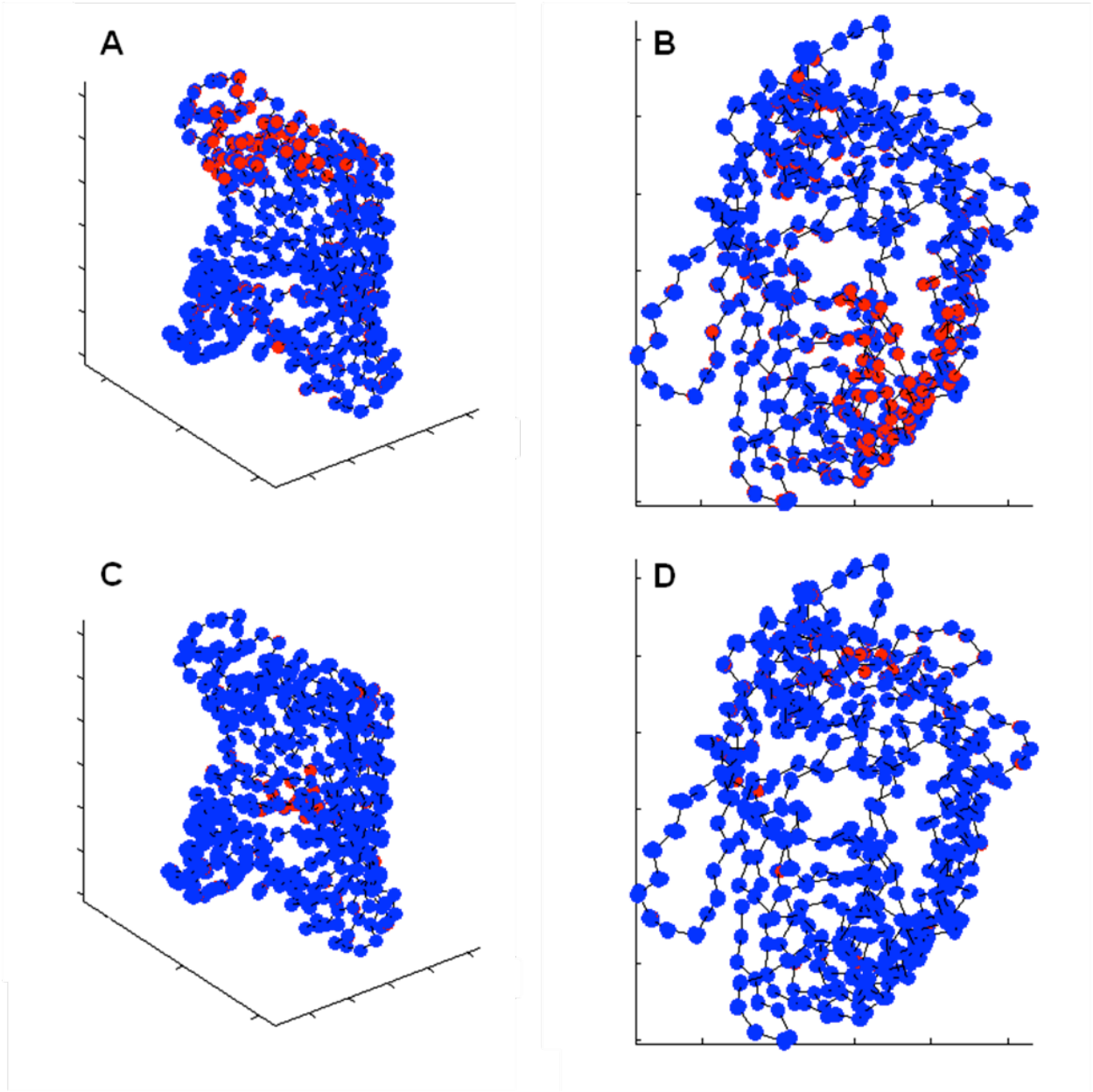


Figure 4.9: Two views of the particular protein deformations that most correlate with increased resistance (A and B) and fitness (C and D). Red areas indicate regions with stronger correlations between deformations and phenotype.

LITERATURE CITED

LITERATURE CITED

1. Silvertown, J. 2008. Demons in Eden: the paradox of plant diversity. University of Chicago Press, Chicago, IL.

2. Kneitel, J. M. and J. M. Chase. 2004. Trade-offs in community ecology: linking spatial scales and species coexistence. Ecology Letters 7:69-80.

3. Gudelj, I., J. S. Weitz, T. Ferenci, M. C. Horner-Devine, C. J. Marx, J. R. Meyer, and S. E. Forde. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecology Letters 13:1073-1084.

4. Litchman, E. and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annual Review of Ecology, Evolution, and Systematics 39:615-639.

5. Dekel, E. and U. Alon. 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588-592.

6. Ferenci, T. and B. Spira. 2007. Variation in stress responses within a bacterial species and the indirect costs of stress resistance. Annals of the New York Academy of Sciences 1113:105-113.

7. Weitz, J. S., Y. Mileyko, R. I. Joh, and E. O. Voit. 2008. Collective decision making in bacterial viruses. Biophysical journal 95:2673-2680.

8. Weitz, J. S., H. Hartman, and S. A. Levin. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proceedings of the National Academy of Sciences of the United States of America 102:9535.

9. Katsura, I. 1981. Tail assembly and injection. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

10. Berkane, E., F. Orlik, J. F. Stegmeier, A. Charbit, M. Winterhalter, and R. Benz. 2006. Interaction of bacteriophage lambda with its cell surface receptor: an in vitro study of binding of the viral tail protein gpJ to LamB (maltoporin). Biochemistry 45:2708-2720.

11. Thirion, J. P. and M. Hofnung. 1972. Some genetic aspects of phage Lambda-resistance in Eacherichia coli K12. Genetics 71:207-&.

12. Lin, E. C. C. 1987. Dissimilarity pathways for sugars, polyols, and carboxylates. Pages 244-284 in N. F. C., editor. Escherichia coli and Salmonella typhimurium: Cellular Molecular Biology. American Society for Microbiology, Washington, DC.

13. Pelosi, L., L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173:1851-1869.

14. Meyer, J. R., A. A. Agrawal, R. T. Quick, D. T. Dobias, D. Schneider, and R. E. Lenski. 2010. Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. Evolution 64:3024-3034.

15. Gehring, K., A. Charbit, E. Brissaud, and M. Hofnung. 1987. Bacteriophage lambda receptor site on the Escherichia coli K-12 LamB protein. Journal of Bacteriology 169:2103-2106.

16. Schirmer, T., T. A. Keller, Y. F. Wang, and J. P. Rosenbusch. 1995b. Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. Science 267:512-514.

17. Charbit, A. 2003. Maltodextrin transport through lamb. Front Biosci 8:s265-274.

18. Heine, H. G., J. Kyngdon, and T. Ferenci. 1987. Sequence determinants in the lamB gene of Escherichia coli influencing the binding and pore selectivity of maltoporin. Gene 53:287-292.

19. Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. Nature methods 7:248-249.

20. Zelditch, M. 2004. Geometric morphometrics for biologists: a primer. Academic Press. Oxford, UK.

21. Spanakis, E. and M. T. Horne. 1987. Coadaptation of Escherichia coli and coliphage lambda vir in continuous culture. Journal of General Microbiology 133:353-360.

22. Daegelen, P., F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim. 2009. Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). Journal of Molecular Biology 394:634-643.

23. Lenski, R. E. 1988. Dynamics of interactions between bacteria and virulent bacteriophage. Advances in microbial ecology 10:1-44.

24. Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term evolution experiment in Escherichia coli 1. Adaptation and divergence during 2,000 generations. American Naturalist 138:1315-1341.

25. Sambrook, J. and D. W. Russell. 2001. Molecular Cloning : A Laboratory Manual. 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

26. Adams, M. H. 1959. Bacteriophages. Interscience Publishers, New York, NY.

27. Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host-phage interactions. PNAS 108:E288-E297.

28. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

29. Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science 332:1193-1196.

30. Forde, S. E., R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220-223.

31. Adams, D. C. and G. J. P. Naylor. 2000. A new method for evaluating the structural similarity of proteins using geometric morphometrics. Currents in computational molecular biology). Universal Academy Press, Tokyo.

32. Emidio, C. and A. Russ. 2011. Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics 12:S3.

33. Wang, Z. and J. Moult. 2001. SNPs, protein structure, and disease. Human mutation 17:263-270.

34. Martí-Renom, M. A., A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Šali. 2000. Comparative protein structure modeling of genes and genomes. Annual review of biophysics and biomolecular structure 29:291-325.

35. Wang, Y. F., R. Dutzler, P. J. Rizkallah, J. P. Rosenbusch, and T. Schirmer. 1997. Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin1. Journal of Molecular Biology 272:56-63.

36. Moldovan, R., E. Chapman-McQuiston, and X. L. Wu. 2007. On kinetics of phage adsorption. Biophysical journal 93:303-315.

37. Bonsall, M. B., V. A. A. Jansen, and M. P. Hassell. 2004. Life history trade-offs assemble ecological guilds. Science 306:111.

38. Doebeli, M. 2011. Adaptive Diversification. Princeton Univ Press. Princeton, NJ.

CHAPTER 5

REPEATABILITY AND CONTINGENCY IN THE EVOLUTION OF A KEY INNOVATION IN PHAGE LAMBDA

Authors: Justin R. Meyer, Devin T. Dobias, Joshua S. Weitz, Jeffrey E. Barrick, Ryan T. Quick, Richard E. Lenski

Originally published in Science, permission was granted to publish in this dissertation per the license to publish agrrement:

http://www.sciencemag.org/site/feature/contribinfo/prep/SCIENCELicensetoPublish25Sep09.pdf

Abstract

Throughout the history of life, evolving lineages have acquired qualitatively new functions that enable organisms to expand their ecological opportunities and, in many cases, undergo further diversification (*1*). Explaining how these transitions have occurred is usually difficult, both because the responsible events typically occurred in the distant past and because their rarity suggests that they might involve atypical evolutionary processes. For example, natural selection is critical for the process of adaptation, yet its role in producing key innovations is less clear because, by fixing variants that improve existing functions, selection might strand populations on local adaptive peaks and thereby prevent them from discovering new functions (*2*). Darwin was well aware of the difference between improving an existing trait and evolving a new one (*3*), and he reasoned that new traits originate by co-opting previously existing structures and functions. Without an understanding of genetic mechanisms, however, he could not provide a detailed

account of how this process happens. Since then, others have proposed more explicit models of the origin of new functions that vary in two main respects: the structure of the adaptive landscape – including its dimensionality (*2*), genetic connectivity (*4*), and fluctuations caused by changing environments (*5*) including interactions with coevolving species (*6*) – and the relative importance of natural selection and random drift (*2, 4, 7*). However, no consensus has been reached owing, at least in part, to the paucity of cases with sufficient genetic and ecological data (*8*).

To that end, we examine the evolutionary forces responsible for the emergence of a novel trait in a microbial system, including data bearing on the genetic architecture of the adaptive landscape in which the novel capacity arose. Microbes are well suited for such research because their evolution can be observed in real time, experiments are easily replicated, and transitional states can be studied by reviving samples stored at different times during an experiment (*9, 10*). We investigated how a virus evolved the ability to infect its host through a new receptor that the ancestral virus cannot use. We tested competing hypotheses about the evolution of this new trait by determining the conditions that promoted its evolution, the mutations that conferred the new function, and the evolutionary forces that drove its emergence.

Study system. Viruses are genetically and morphologically diverse, and they infect all groups of organisms (*11, 12*). Viruses initiate infections by binding to receptors on the surface of host cells. The physicochemical properties of viral ligands determine which receptors they target and thereby influence the host range and ecological niche of the virus (*13*). Mutations in viral genes that encode the production of ligands can cause shifts in host range and thus are often associated with emerging diseases (*14-16*). The evolution of the ability to infect through a new receptor represents a key innovation for a virus.

The virus we studied, a strictly lytic derivative of phage λ called cI26 (Appendix 4, Table A4.1), is only known to infect one bacterial species, *Escherichia coli*, and has a specialized ligand, the J protein, at the end of its tail (*17*). J targets a single protein, LamB, on the *E. coli* outer membrane (*17, 18*). Phage λ requires only LamB to attach (*19*), and LamB is the only outer-membrane protein that affects λ reproduction (*20*).

Given interest in the fundamental question of how organisms evolve novel traits and in the practical problem of how emerging pathogens evolve to target new host receptors, we sought to determine whether λ could evolve to infect through an alternative receptor. Phage λ is well studied and amenable to experimentation (*21*). A related phage (Ur-λ) possesses side-tail fibers and can infect *E. coli* through a second receptor OmpC (*22, 23*), suggesting other receptors might be accessible to evolving λ populations. Moreover, we identified conditions that seemed suitable for promoting the use of a novel receptor. In particular, when *E. coli* and λ were cultured together in a glucose-limited environment, the bacteria evolved resistance by mutations in *malT*, which interfered with its role as a transcription factor that promotes *lamB* expression (*23*). The mutants arose and fixed within 8 days, generating highly resistant populations (Figure. A4.1), although the phage did not go extinct but instead persisted at densities of about 10^6 to 10^7 phage per ml, or about one phage per 10^2 to 10^3 host cells (Figure. A4.2). The phage evidently persisted on a subpopulation of cells that, despite their *malT* mutations, experienced spontaneous induction of the LamB protein (*24*). This explanation was confirmed by showing that phage were also sustained when they were grown on a *malT* mutant, whereas the phage went extinct when cultured with a *lamB* mutant that lacks the potential to produce the LamB protein (Figure. A4.3). We reasoned that mutant viruses able to infect through some protein other than LamB

would be favored after *malT* mutants arose because they could infect the entire host population rather than a small minority of cells.

Initial evolution experiment. We co-cultured a virulent (non-lysogenic) derivative of λ and *E. coli* B in 10 ml of a minimal glucose medium in six replicate flasks for 28 days with daily transfers of 1% of each community into a flask containing fresh medium, and we preserved samples weekly by freezing 10% of the mixture (*Appendix 4*). We tested whether the phage could infect cells through a new receptor by taking samples of the phage populations (typically ~5 x 10^4 virions) and inoculating them onto the surface of agar plates infused with a *lamB* mutant that does not produce the LamB receptor. A spot of clearing (lysis of host cells) provided evidence that some phage had evolved the ability to infect through a receptor other than LamB (Figure. 5.1, panels A-B). Such spots were observed in only one population, but this ability evolved quickly, such that ~0.01% of the phage could infect the *lamB*-negative mutants by day 8, including the isolate designated EvoC, and the majority did so by day 15.

Identification of the novel receptor. We used seven knockout strains (derivatives of K12 BW25113), each missing a gene encoding a different outer-membrane protein, to identify the new receptor (*Appendix 4, 25*). We then introduced *malT* mutations to these strains so that they also would not express the native LamB receptor. We inferred the new receptor by testing the ancestral and evolved λ against these double mutants to see which ones were resistant to the various phage isolates. The only host that was resistant to the EvoC isolate was the *ompF malT* double mutant that lacked expression of both OmpF and LamB (Figure. 5.1, Table A4.2), indicating that OmpF was the new receptor. This evolved phage could still infect the host strain expressing LamB but not OmpF, indicating that the phage retained the ability to infect through its native receptor (Figure. 5.1).

Both LamB and OmpF form trimeric porins composed of three identical β barrels (*26, 27*). This overall structure is probably essential for the J protein in the λ tail to bind because J, too, forms a trimer and is thought to attach with radial symmetry across the three pore domains (*28*). Although OmpF has the most similar crystal structure to LamB of any *E. coli* protein determined to date *(29)*, they are not the most similar pair by amino-acid sequence (table A4.3). This discordance suggests that the overall structure is at least as important for λ binding as the identity of specific amino-acid residues. Also, OmpF is the sole major porin in the *E. coli* B strain used in this study, and B expresses it constitutively during growth (*30, 31*). Hence, OmpF provided a substantial ecological opportunity to phage that evolved the ability to target it.

Genome evolution. We sequenced the genome of the evolved phage EvoC in order to identify the mutations that allowed it to use the OmpF receptor (*Appendix 4*). There were five mutations in total, and all of them were in the J gene (Figure. 5.2). Targeted sequencing of J showed that a single substitution (A to G at position 3034) differentiated EvoC from another evolved isolate from the same time point, EvoA, that could use only the ancestral LamB receptor, indicating that mutation contributed to the new receptor function. Another LamB-dependent phage from the same day, EvoB, differed from EvoC at five sites in the J protein.

Large-scale evolution experiment. We repeated our first experiment with 96 more communities to identify general principles of how λ evolves the capacity to target an alternative receptor (*Appendix 4*). We sampled daily for finer resolution of the evolutionary dynamics. As before, only some phage populations (24/96) evolved the ability to use a second receptor. This ability emerged about the same time (median 12 days; range 9-17 days; Figure. A4.4), and all isolates with altered receptor function infected hosts through the OmpF protein.

Parallel molecular evolution. We sequenced J alleles from 24 phage isolates that independently evolved the ability to target OmpF during the large-scale evolution experiment to determine whether the mutation at position 3034 or any others were required to use that receptor. The isolates were taken the same day the new function was detected (*Appendix 4*). For comparison, we sequenced phage from 24 populations that never evolved this trait; these isolates were sampled on the same days as those that evolved the new trait, so that the elapsed times were the same.

In total, there were 241 single nucleotide polymorphisms (SNPs) across the 48 J alleles, but no insertions or deletions (Figure. 5.3). However, there were only 40 unique mutations because many arose repeatedly in replicate populations. Moreover, all of them were non-synonymous. The alleles for phage able to target OmpF had on average 6.63 (± 0.51 95% confidence interval (CI)) SNPs, whereas the phage that required LamB had only 3.42 (± 0.50 95% CI) SNPs. This difference is highly significant based on a paired comparison between the two types of phage matched for the day of their isolation and, in the case of multiple equivalent pairs, matched arbitrarily by position in the experiment (t_s = 9.144, 23 d.f, two-tailed $p < 0.0001$; Table A4.4). Also, across both classes of phage, over 97% of the mutations were in the last 25% of the protein (C-terminal end), the region known to interact with LamB (*32*).

There are four striking cases of parallel evolution of the J protein in the phage that target OmpF. In two cases, the mutations were identical across all 24 populations, while in two others there were slight variations (Figure. 5.3). In particular, all J alleles from phage able to infect through OmpF had the A-to-G mutation at nucleotide position 3034 and G-to-A mutation at position 3319. Also, all of them had a mutation at either position 3320 or 3321, affecting the

same codon (amino-acid residue 1107) as the mutation at position 3319. Finally, all J alleles had at least one mutation between positions 2969 and 2999 (amino-acid residues 990 to 1000).

Each of these mutations or classes of mutation was also found in at least one of the phage that retained the ancestral host-range, although none of them had all four together (Figure. 5.3). Two LamB-dependent isolates, F2 and H4, had three of the mutations, as did EvoA from the initial experiment (Figure. 5.2), yet none produced clearing on lawns of *lamB* mutants.

The correspondence between the use of the OmpF receptor and the presence of these four mutations, coupled with the observation that phage having only three of the four cannot use OmpF, provides evidence that all four are required for λ to infect through OmpF. We performed two additional assays to confirm that only phage with all four mutations can infect *lamB* mutants (*Appendix 4*). The assays were performed using isolates EvoA, F2, and H4 that each had three of the four canonical mutations and D7 that had all four and no others. Only D7 exhibited a measureable adsorption rate on *lamB* mutant cells (Figure. A4.5), and it was also the only one that reproduced on *lamB* mutants in the medium used in the evolution experiments (Figure A4.6). These findings indicate an "all-or-none" form of epistasis among the four mutations responsible for the novel receptor phenotype.

Role of natural selection. In the λ population that evolved to use OmpF in the initial experiment, the A3034G mutation was the fourth and final step, and it was clearly advantageous because it conferred the ability to infect the entire cell population. However, the all-or-none epistasis among the mutations means that selection for that new capacity *per se* was not responsible for the rise of the three prior mutations. Nonetheless, there are several lines of evidence that selection drove their rise. First, all 248 independent mutations in the 51 sequenced J alleles were non-synonymous, whereas the expected ratio of non-synonymous to synonymous

changes is 3.19:1 under the null model for the ancestral J sequence (*Appendix 4*). This great excess is evident even if we include only the 82 non-synonymous mutations in the 24 isolates that did not evolve the new receptor function. Second, the mutations are highly concentrated in the region of the J protein that interacts directly with LamB (*18*). Third, there was parallel evolution at the genetic level across the populations. For those phage that evolved to exploit OmpF, an average of 61% (4.06/6.63) of mutations were shared across independently derived pairs (Figure A4.7), which greatly exceeds the fraction expected under a conservative randomization test (*Appendix 4*) using only the variable sites in J ($p < 10^{-5}$). Pairs of phage that remained dependent on LamB shared on average 17% (0.58/3.42) of their mutations (Figure 4A.7), and this fraction is again significant under the same test ($p < 10^{-5}$). Thus, it is clear that selection acted on the J protein even before the new capacity evolved. This selection presumably improved the interaction of the phage tail fibers with LamB.

Stochasicity and contingency. All of the λ populations had the same ecological incentive to exploit an alternative receptor, but only some evolved that ability. Why were some populations successful and others not? One possibility is that all of them would eventually have evolved that function, but there was insufficient time to do so. This explanation is consistent with the facts that the LamB-dependent isolates had fewer mutations than those able to target OmpF, and that the two groups shared many mutations. Alternatively, the evolution of the new receptor function might have been contingent on earlier events (*33-35*), such that particular changes in the phage or the host promoted or impeded the subsequent evolution of that function. To test these hypotheses, we replayed evolution (*35, 36*) using various combinations of phage and bacteria.

First, we tested whether certain mutations in the phage that might enhance performance on LamB would impede the evolution of the new OmpF function. We inoculated flasks with the

ancestral bacteria and one of six phage isolates. Three of the six phage isolates had different sets of three mutations that were present in multiple isolates that evolved to use the OmpF receptor, including one, zero, and two of the four canonical mutations. The other three isolates had three, two, and one mutations that were not observed in any isolate that previously evolved the ability to use OmpF; these isolates also each had one of the canonical mutations. The first set provided a positive control; the second set had candidate mutations for impeding the evolution of the new function. For each phage, we propagated 12 communities for 10 days and surveyed daily the phage's ability to lyse *lamB* mutant cells. There were as many or more successes in evolving the new function among the three phages that had the potentially interfering mutations as among the positive controls (Figure 5.4A). This experiment thus provides no evidence that some phage failed to evolve the new function because they had mutations that prevented them from doing so.

Next, we asked if the outcome was contingent on mutations in the evolving bacteria. To that end, we performed a similar replay experiment except that the initial phage type was held constant while the starting bacterial isolate was varied. For the phage, we used EvoA, an isolate that was one mutation away from using OmpF (Figure 5.2). For the bacteria, we used six clones: three from communities where λ evolved to use OmpF, and three where the phage retained their dependence on LamB. We observed a striking "all-or-none" pattern of outcomes, though not in accordance with our categories (Figure 5.4B). In particular, all 36 λ populations evolved the final mutation required to use OmpF in communities with three bacterial clones whereas none of the phage evolved that ability in 36 communities with the three other clones. For two of the latter class (EcA8 and EcF6), the phage were unable to reproduce and went extinct; in the other case (EcC3), phage persisted but none of the replays yielded phage able to use OmpF. It is clear that bacterial characteristics determined whether the phage would evolve the new receptor function.

However, the bacteria that promoted that outcome did not necessarily come from communities in which λ had previously evolved to exploit OmpF.

We sequenced the full genomes of these six bacteria to identify the mutations responsible for the differences in phage evolution. The six genomes harbor a total of 15 mutations (Table A4.5). Five have similar deletions that impact the *rbs* operon, which confers the ability to grow on ribose; previous work has shown that these deletions occur at an unusually high rate owing to a nearby insertion-sequence element in the ancestral strain (*37)*. All 10 other mutations are directly related to the interaction with λ. As expected, all six genomes have mutations in *malT* that confer resistance to the ancestral phage, as described earlier. One genome from a community where λ evolved to use the OmpF receptor has a non-synonymous mutation in the *ompF* gene, which might confer partial resistance to the evolved phage. The three remaining mutations disrupt *manY* or *manZ*, and they uniquely differentiate the three strains that prevented phage from evolving to use OmpF from the three strains that allowed that change (table S5). The *manY* and *manZ* genes encode the transmembrane channel of the ManXYZ mannose permease, which is required for λ DNA to cross the inner membrane (*38-40*). These mutations thus confer resistance by blocking a later step during infection, and they would render ineffective any phage mutations that altered the receptor function. Therefore, the evolution of phage that target OmpF is promoted by bacterial mutations in *malT* but impeded by mutations in *manYZ*, indicating contingency dependent on the host-parasite coevolution.

After discovering the *manYZ* mutations, we screened all 96 bacterial populations from the large-scale experiment to determine how many harbored mannose-deficient mutants that would block λ infections (*Appendix 4*). At least 80 populations (Table A4.6) had such mutations, including many from communities in which λ evolved the ability to exploit the OmpF receptor.

However, these mutations rarely fixed; instead, susceptible subpopulations persisted in 77 of the 80 communities that allowed the phage to continue to evolve. This finding suggests a complex interplay between coevolving phage and bacteria, one that depends on the entire community and its diversity. To test this hypothesis, we repeated the second replay experiment using the same phage and bacteria except with bacterial communities instead of clones (*Appendix 4*). Once again, some bacteria consistently impeded the evolution of phage that used OmpF while others consistently promoted that change (Figure 5.4C). Moreover, those outcomes differed for one clone (EcA8, Figure 5.4B) and its community (ComA8, Figure 5.4C), confirming the effect of bacterial diversity on the phage's evolution.

Repeatability, contingency, and the evolution of a key innovation. Phage λ often, but not always, evolved the ability to infect its *E. coli* hosts by targeting a new receptor, OmpF. Figure 5.5 summarizes the important steps in this process, including some that promoted this key innovation and others that impeded it. The fact that several mutations are required for λ to use OmpF may explain why no previous studies have reported this change, despite decades of intense study of this phage. However, by running evolution experiments rather than mutational screens, we observed 25 cases in which this new function evolved. Our experiments are not the first to demonstrate evolutionary transitions in viruses (*13, 41-43)*; one study found that a single mutation allowed phage φ6 to infect a new host species (*13*), while a study of phage SBW25Φ2 showed that coevolution experiments were more effective at producing host-range shifts than were screens using new hosts (*43*). The rich and complex dynamics of coevolving species may thus sometimes facilitate key innovations (*6, 44*).

More generally, our study shows some of the challenges that make it difficult to observe and explain the origin of many new functions including the requirement for multiple mutations,

the complex interactions of mutations within and between species, and the resulting historical contingency that can enable or impede the outcome of interest depending on the order in which mutations occur. The "all-or-none" epistasis among the four canonical phage mutations implies that it would have been unlikely for the new function to evolve on the scale of our experiments, except for the lucky fact that some of the mutations were beneficial to the phage in performing their current function, thereby pushing evolution toward the new function. The mutations in the bacteria, and how they influenced the phage's evolution, were also important. In particular, the initial resistance mutation generated a physiological subpopulation of hosts that allowed the phage to persist and adapt to the original receptor, thereby accumulating the required mutations. Yet, as fortuitous as these circumstances were, another mutation could – and often did – derail the emergence of the new function: namely, a mutation that conferred an alternative mode of host resistance eliminated the advantage to the phage of targeting the OmpF receptor. The interactions between bacteria and phage, which contributed so much to the development of microbial genetics and molecular biology (*45, 46*), continue to serve as powerful models to study ecology and evolution (*47, 48*).

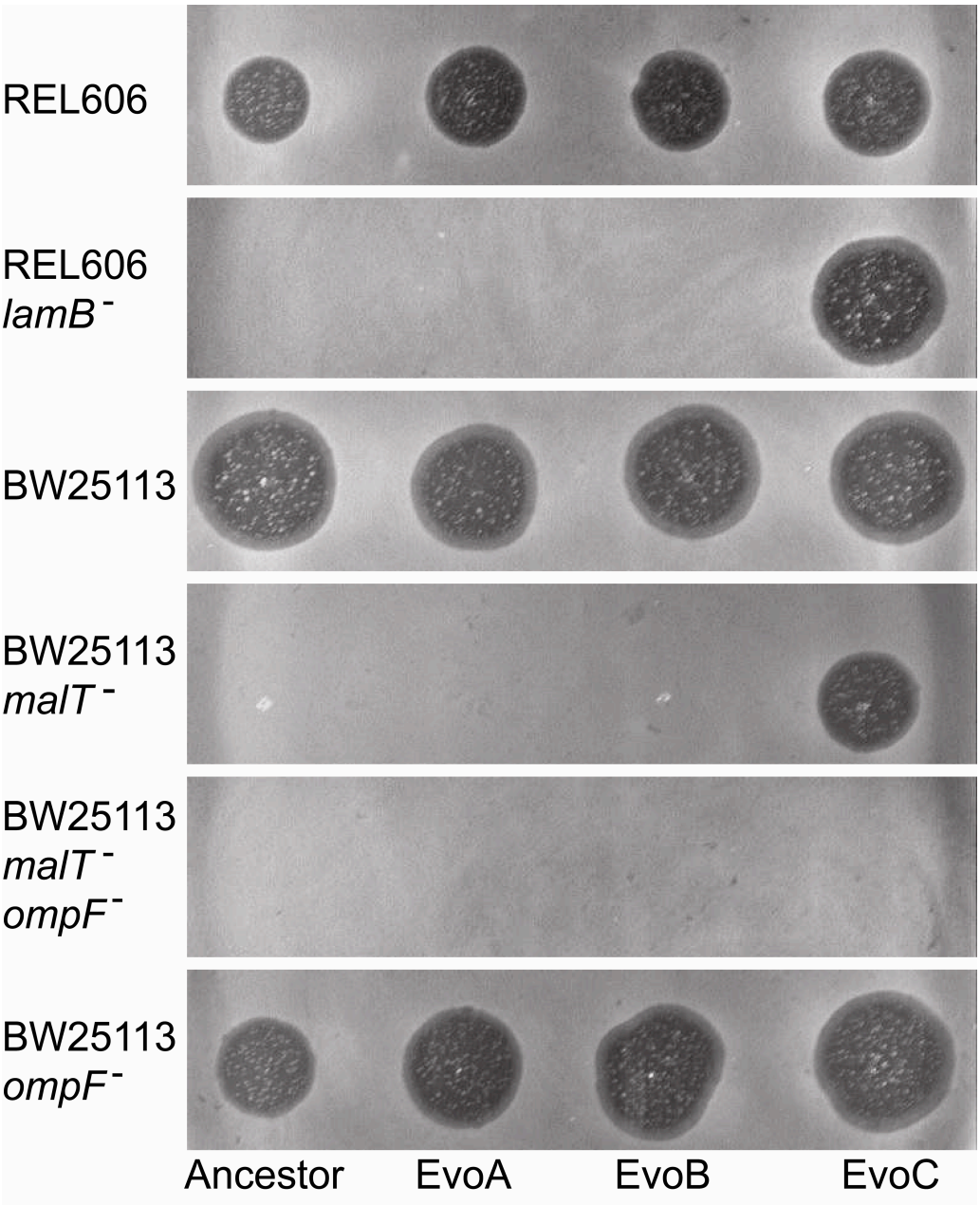


Figure 5.1. Infection assay for four λ isolates tested on six *E. coli* strains. Each panel shows a bacterial lawn with aliquots of the phage applied to it; darker regions indicate successful infections that clear the lawn. The phage isolates include the ancestor and three clones isolated from the same population on day 8 of the initial evolution experiment including one, EvoC, that can use the OmpF receptor. The bacterial strains include mutants that differ in the expression of LamB and OmpF porins on two genomic backgrounds, REL606 (the ancestral strain in the evolution experiments) and BW25113 (a derivative of K12). The *malT*$^-$ strains do not express LamB at appreciable levels.

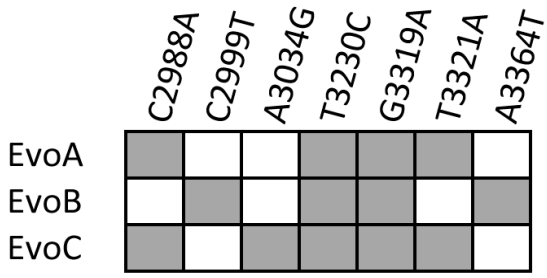


Figure 5.2. Mutations in the λ gene encoding the J protein in three isolates from the same population on day 8 of the initial evolution experiment. The isolates are shown in rows and the mutations in columns, with the first letter being the ancestral nucleotide, the number the nucleotide position, and the last letter the evolved nucleotide. The gray fill indicates that the phage isolate has the corresponding mutation.

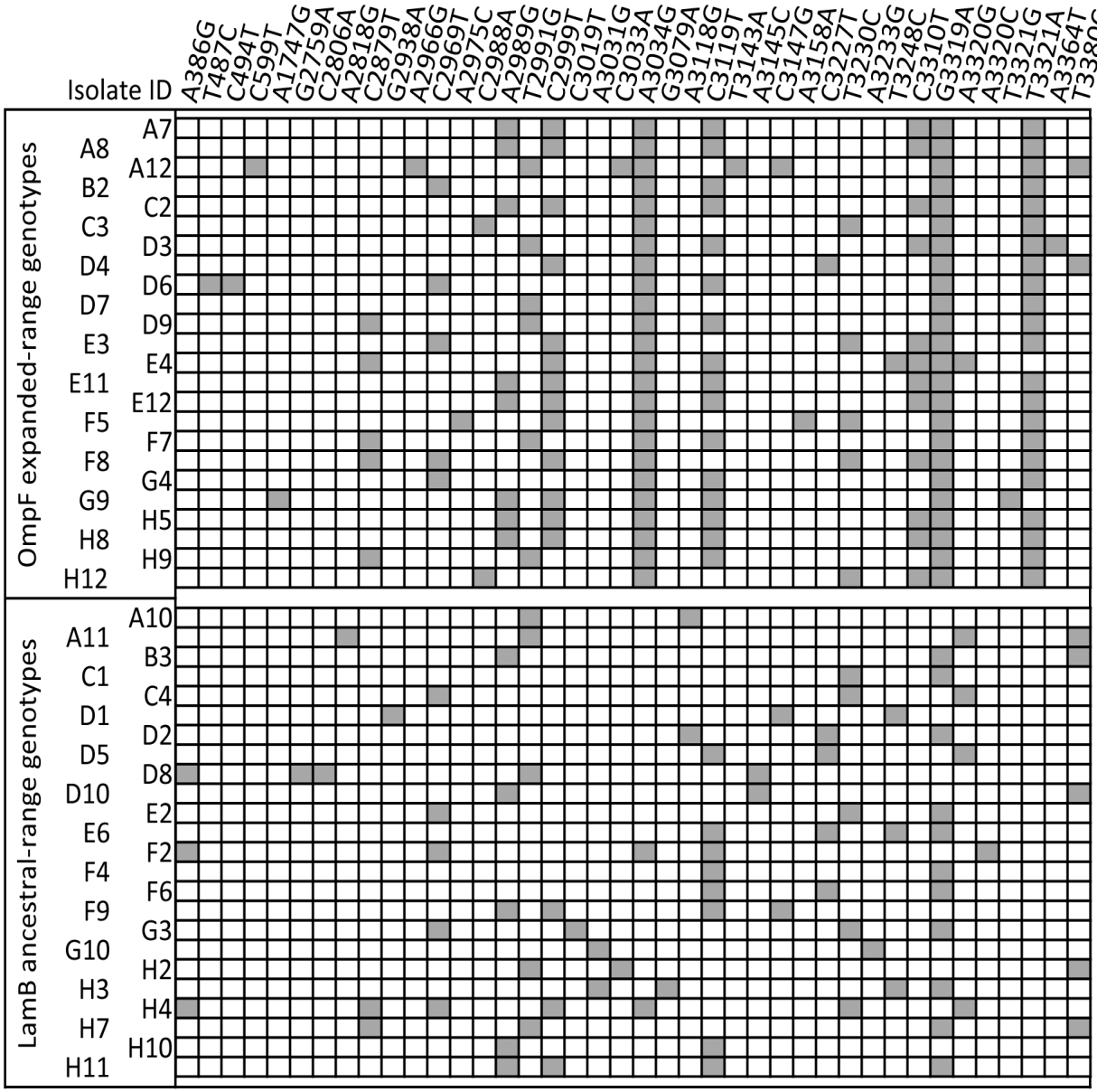


Figure 5.3. Mutations affecting the J protein in phage isolates from 48 independent populations of the large-scale experiment. Isolates are shown in rows (with alternate labels offset for readability) and mutations in columns; gray fill indicates an isolate has the mutation. The top 24 rows show phage isolates that can target the new OmpF receptor; the bottom 24 rows show phage that remain dependent on LamB.

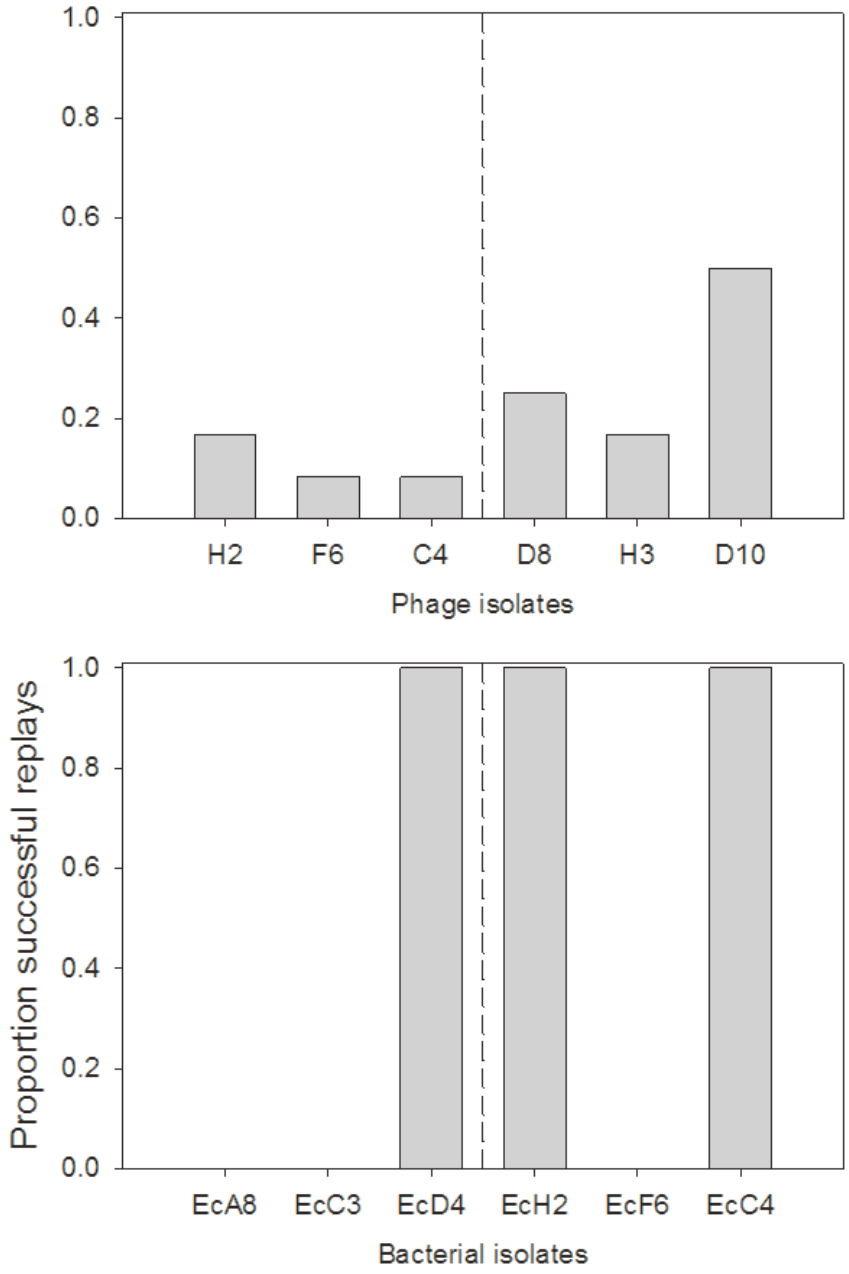


Figure 5.4 (continued on next page)

Figure 5.4 continued from previous page

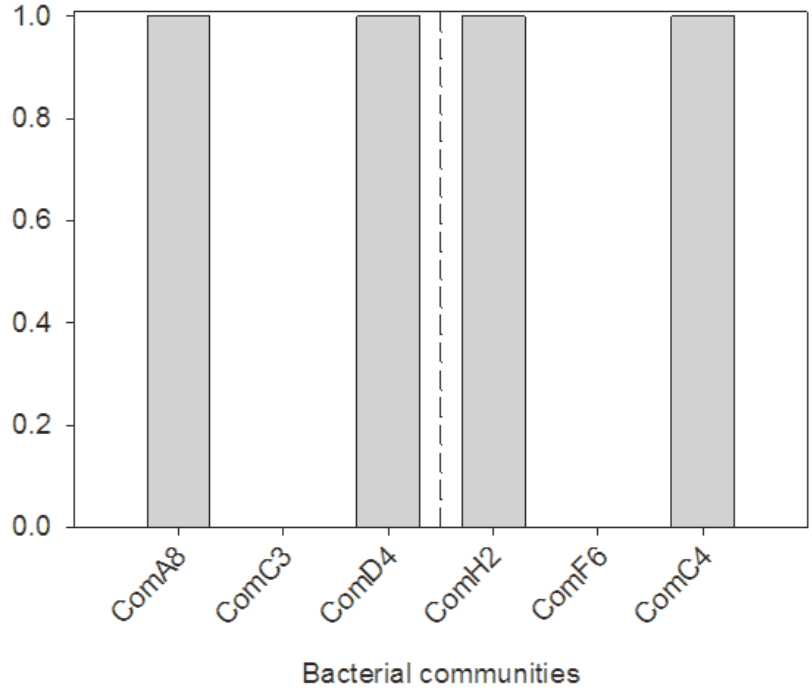


Figure 5.4. Replay experiments using different combinations of phage and bacteria. For each panel, the y-axis shows the proportion of replicate replays that produced phage able to target the new OmpF receptor. Top panel: Replays were initiated with the ancestral bacteria and six phage isolates. Each combination was replicated 12-fold. Three of the phage (H2, F6, and C4) had mutations shared by multiple lineages that evolved the capacity to target OmpF in the large-scale experiment. The other three (D8, H3, and D10) had mutations that were never observed in phage that targeted OmpF. The latter mutations were candidates for impeding the evolution of the new function, but that hypothesis was not supported. Middle panel: Replays were initiated with phage EvoA (which needs only one more mutation to use OmpF) and six bacterial clones. Each combination was replicated 4-fold. Three clones (EcA8, EcC3, and EcD4) were isolated from flasks in which phage evolved the capacity to target OmpF in the large-scale experiment. The other three (EcH2, EcF6, and EcC4) came from flasks in which phage did not evolve that function. The replay outcomes did not support these categories, but sequencing the bacterial genomes identified mutations that uniquely determined whether the phage would evolve the OmpF function. See text for details. Bottom panel: Replays were initiated using the same phage and bacteria used in the middle panel, except with full bacterial communities rather than individual clones. Each combination was replicated 12-fold. The different outcomes for one bacterial clone (EcA8, middle) and its source community (ComA8, bottom) show the effect of bacterial diversity on phage evolution.

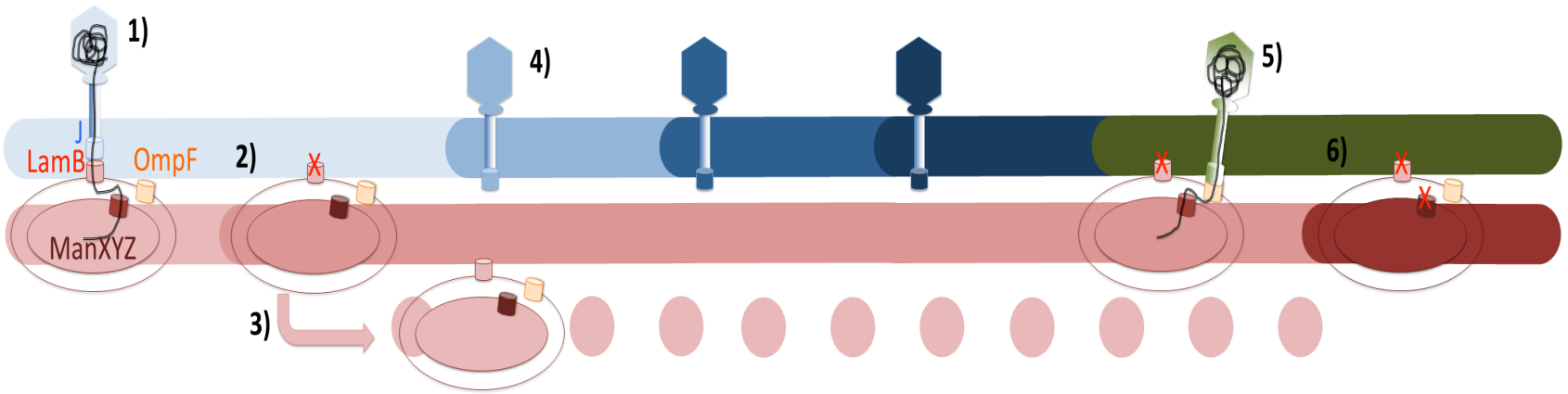


Figure 5.5. Steps in the coevolution of phage λ and its *E. coli* host leading to the phage's ability to target a new receptor, OmpF. (1) The ancestral phage targets the LamB porin using the J protein and injects its DNA into the periplasm, then the DNA is transported into the cytoplasm via the ManXYZ permease. (2) The bacteria evolve resistance by mutations in *malT*, a positive regulator of LamB expression. (3) However, spontaneous inductions of LamB generate a subpopulation of phenotypically sensitive cells that can sustain the phage population. (4) The phage evolves mutations in the J protein that improve performance on the LamB receptor. Some of these mutations are also required for the phage to target OmpF. (5) The phage eventually evolves the four mutations that enable it to use OmpF. (6) However, the bacteria may evolve additional resistance by mutations in *manY* or *manZ* that prevent transport of the phage DNA into the cytoplasm. When these mutants become sufficiently common, there is little or no benefit to mutant phage that can use OmpF.

LITERATURE CITED

LITERATURE CITED

1. Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University Press, Cambridge.

2. Gavrilets, S. 2010. High-dimensional fitness landscapes and the origins of biodiversity.in M. Pigliucci and G. Muller, editors. Toward an Extended Evolutionary Synthesis. MIT Press, Cambridge, MA.

3. Darwin, C. 1890. The origin of species by means of natural selection : or, the preservation of favored races in the struggle for life. Reprinted from the 6th London ed. with the additions and corrections. edition. Hurst, New York :.

4. Fisher, R. A. 1930. The genetical theory of natural selection. The Clarendon Press, Oxford.

5. Dieckmann, U. 2004. Adaptive speciation. Cambridge University Press, Cambridge, UK.

6. Thompson J. N. 2005. *The Geographic Mosaic of Coevolution* (Univ. Chicago Press, Chicago).

7. Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:0097-0159.

8. Pigliucci, M. 2008. What, if anything, is an evolutionary novelty? Philosophy of Science 75:887-898.

9. Dennehy, J. J. 2009. Bacteriophages as model organisms for virus emergence research. Trends in Microbiology 17:450-457.

10. Elena, S. F. and R. E. Lenski. 2003. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nature Reviews Genetics 4:457-469.

11. Fisher, R. A. 1930. The genetical theory of natural selection. The Clarendon Press, Oxford.

12. Torsvik, V., L. Ovreas, and T. F. Thingstad. 2002. Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science 296:1064-1066.

13. Duffy, S., C. L. Burch, and P. E. Turner. 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. Evolution 61:2614-2622.

14. Pepin, K. M., S. Lass, J. R. C. Pulliam, A. F. Read, and J. O. Lloyd-Smith. 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nature Reviews Microbiology 8:802-813.

15. Holmes, E. C. 2009. The Evolutionary Genetics of Emerging Viruses. Annual Review of Ecology Evolution and Systematics 40:353-372.

16. Parrish CR, Holmes EC, Morens DM, et al. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72:457-70.

17. I. Katsura, in *Lambda II*, R. W. Hendrix, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1981).

18. Berkane, E., F. Orlik, J. F. Stegmeier, A. Charbit, M. Winterhalter, and R. Benz. 2006. Interaction of bacteriophage Lambda with its cell surface receptor: An in vitro study of binding of the viral tail protein gpJ to LamB (Maltoporin). Biochemistry 45:2708-2720.

19. Grayson, P., L. Han, T. Winther, and R. Phillips. 2007. Real-time observations of single bacteriophage lambda DNA ejections in vitro. Proceedings of the National Academy of Sciences of the United States of America 104:14652-14657.

20. Maynard, N. D., E. W. Birch, J. C. Sanghvi, L. Chen, M. V. Gutschow, and M. W. Covert. 2010. A Forward-Genetic Screen and Dynamic Analysis of Lambda Phage Host-Dependencies Reveals an Extensive Interaction Network and a New Anti-Viral Strategy. PLoS Genet 6:e1001017.

21. R. W. Hendrix, Ed., *Lambda II* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1983).

22. Hendrix, R. W. and R. L. Duda. 1992. Bacteriophage-lambda – not the mother of all lambda-phages. Science 258:1145-1148.

23. Gallet, R., S. Kannoly, and I.-N. Wang. 2011. Effects of bacteriophage traits on plaque formation. BMC Microbiol. 11, 181 (2011).

24. Braunbreton, C. and M. Hofnung. 1978. Explanations accounting for transduction by bacteriophage-gamma in maltose negative bacteriophage gamma-resistant mutants of Escherichia-coli K12. Molec. & Gen. Genet. 159:143-149.

25. Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molec. Systems Biol. 2.

26. Schirmer, T., T. A. Keller, Y. F. Wang, and J. P. Rosenbusch. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1-anstrom resolution. Science 267:512-514.

27. Kefala, G., Ahn, C., Krupa, M., Esquivies, L., Maslennikov, I., et al., 2010. Structures of the OmpF porin crystallized in the presence of foscholine-12. Protein Sci. 19: 1117–1125.

28. Gurnev, P. A., A. B. Oppenheim, M. Winterhalter, and S. M. Bezrukov. 2006. Docking of a single phage lambda to its membrane receptor maltoporin as a time-resolved event. Journal of Molecular Biology 359:1447-1455.

29. Koebnik, R., K. P. Locher, and P. Van Gelder. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol. Microbiol. 37:239–253.

30. Schneider D , et al. 2002. Genomic comparisons among *Escherichia coli*strains B, K-12, and O157:H7 using IS elements as molecular markers. BMC Microbiol 2:18.

31. Crozat E, Hindre T, Kuhn L, Garin J, Lenski RE, Schneider D: Altered regulation of the OmpF porin by Fis in Escherichia coli during an evolution experiment and between B and K-12 strains. J. Bact. 2011:193:429-440.

32. Wang, J., M. Hofnung, and A. Charbit. 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. Journal of Bacteriology 182:508-512.

33. Gould, S. J. 1989. Wonderful life : the Burgess Shale and nature of history. 1st ed. edition. W.W. Norton, New York.

34. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

35. Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 105:7899-7906.

36. Woods, R.J. et al. (2011) Second-order selection for evolvability in a large Escherichia coli population. Science 331, 1433–1436

37. Cooper VS, Schneider S, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of E. coli B. J Bacteriol 183:2834 –2841.

38. B. Erni, B. Zanolari, H. P. Kocher, 1987. The mannose permease of Escherichia-coli consists of 3 different functions; transport, sugar phosphorylation, and penetration of phage –lambda DNA J. Biol. Chem. 262, 5238.

39. P. E. J. Saris, E. T. Palva. 1987. Regulation of manX (PTSL) and many (PEL) genes required for mannose transport and penetration of lambda-dna in Echerichia-coli K12 FEMS Microbiol. Lett. 44, 371.

40. M. Esquinas-Rychen, B. Erni, Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenolpyruvate : carbohydrate phosphotransferase system (PTS) *J. Mol. Microbiol. Biotechnol.* 3, 361 (2001).

41. Graham, R. L. and R. S. Baric. 2010. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. J. Virology 84:3134-3146.

42. Ferris, M. T., P. Joyce, and C. L. Burch. 2007. High frequency of mutations that expand the host range of an RNA virus. Genetics 176:1013-1022.

43. A. R. Hall, P. D. Scanlan, A. Buckling. 2011. Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens *Am. Nat.* 177:44.

44. J. N. Thompson. 1994 *The Coevolutionary Process* (Univ. Chicago Press, Chicago, IL).

45. D. I. Friedman, D. L. Court. 2001. Bacteriophage lambda: alive and well and still doing its thing *Curr. Opin. Microbiol.* 4, 201.

46. J. Cairns, G. S. Stent, J. D. Watson 2007. Eds., *Phage and the Origins of Molecular Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

47. Bohannan, B. J. M., and R. E. Lenski. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage.Ecology Letters 3:362-377.

48. Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host–phage interactions. Proceedings of the National Academy of Sciences 108:E288-E297.

CHAPTER 6

KEY INNOVATION IN A VIRUS CATALYZES A COEVOLUTIONARY ARMS RACE

Authors: Justin R. Meyer, Cesar Flores, Joshua S. Weitz, Richard E. Lenski

Originally published as an open access manuscript in A-Life Proceedings 13 (2012).

Extended Abstract

I believe that ... any replay of the tape would lead evolution down a pathway radically different from the road actually taken. – Stephen J. Gould (1)

Starting with identical organisms under identical conditions, replicate evolution experiments nonetheless sometimes produce very different outcomes (2,3). These differences arise because mutations occur randomly, such that variation in their timing and order of appearance can affect evolutionary trajectories. An example of contingency was reported to occur when a virus, phage λ, coevolved with its host, the bacterium *Escherichia coli* (3). Meyer *et al*. found that λ evolved the ability to infect *E. coli* cells through a new receptor, the host's OmpF protein, in about one-quarter of the replicate communities, but it did not do so in the others. This bifurcation in the virus's evolution occurred, at least in part, because of variation in the mutations that evolved in the host populations. Here we compare the patterns of coevolution in two communities including one where the virus evolved this key innovation and one where it did not. In the community where the virus evolved the innovation a coevolutionary arms race

immediately ensued, whereas the other community appeared static. This work illustrates how early variation in the sequence of evolutionary events can propagate to produce increasingly different outcomes.

When communities of *E. coli* and λ were cultured in a glucose-limited medium, the bacteria evolved resistance through mutations in *malT* within about a week. The *malT* mutations caused reduced expression of the protein LamB, which the phage uses to attach to and inject its DNA through the host's outer cell envelope (4). Over time, λ then evolved mutations in its J gene, which encodes the protein it uses to bind to LamB, and these mutations evidently improved its ability to infect the host cells (3). At this stage, the evolutionary dynamics diverged quite sharply. Some phage populations evolved a combination of four J mutations that conferred the ability to use OmpF as an alternative receptor. However, most of the bacterial populations evolved a mutation in *manXYZ*, a set of genes encoding a channel that the phage uses to transport its DNA across the host's inner membrane. If one of these mutations reached high frequency in the host population, then the phage population did not evolve the ability to target OmpF, thereby leaving it dependent on the ancestral receptor. This divergence typically occurred within about two weeks, after which λ and *E. coli* continued to coexist for at least two more weeks (3).

In the present study, we examined the coevolutionary dynamics associated with the phage's innovation or the lack thereof. To do so, we revived the ancestral *E. coli* and λ along with samples from two communities that had been frozen on days 8, 15, 22 and 28 of the initial experiment reported by Meyer *et al*. Phage λ evolved the ability to exploit OmpF in one of the communities (Figure 6.1A), but not in the other (Figure 6.1B). We isolated 10 phage and 10 bacterial clones from each time-point and community. We then challenged each bacterial clone with each phage isolated from the same replicate community. We measured the ability of each

phage to infect an evolved bacterium relative to its ability to infect the ancestral bacterium; this infectivity metric has also been called the efficiency of plaquing (5). Figure 6.1 shows these data as interaction matrices. Figure 6.2 summarizes how the average bacterial resistance and phage host-range changed over time in the community where the phage evolved the ability to use OmpF, including both contemporaneous and time-shifted interactions.

The differences between the two communities are striking. The virus evolved the ability to use the new OmpF receptor at about day 8, which led to the emergence of a diverse assemblage of bacteria and phage with different patterns of resistance and infectivity, respectively (Figure 6.1A). Over time, the bacteria evolved increasing resistance to the phage (Figure 6.2), including two clones from day 28 that were completely resistant to all 40 phage isolates. Also, the phage tended to evolve expanded host ranges (Figure 6.2) and increased infectivity on the bacteria, with these trends being more pronounced when phage from later generations were tested on bacteria from earlier generations.

Key innovations are not end-points in evolution, but instead they are hypothesized to spark further rapid change. For example, the processes of speciation and adaptive radiation are thought often to follow the evolution of new ecological functions (6). We have shown here that a key innovation in an evolving virus population catalyzed a coevolutionary arms race with its host, which led to the rapid diversification of both the host and parasite. Although the two communities that we studied began with identical ancestors and experienced identical abiotic environments, they followed radically different evolutionary trajectories, consistent with Gould's view of life (1).

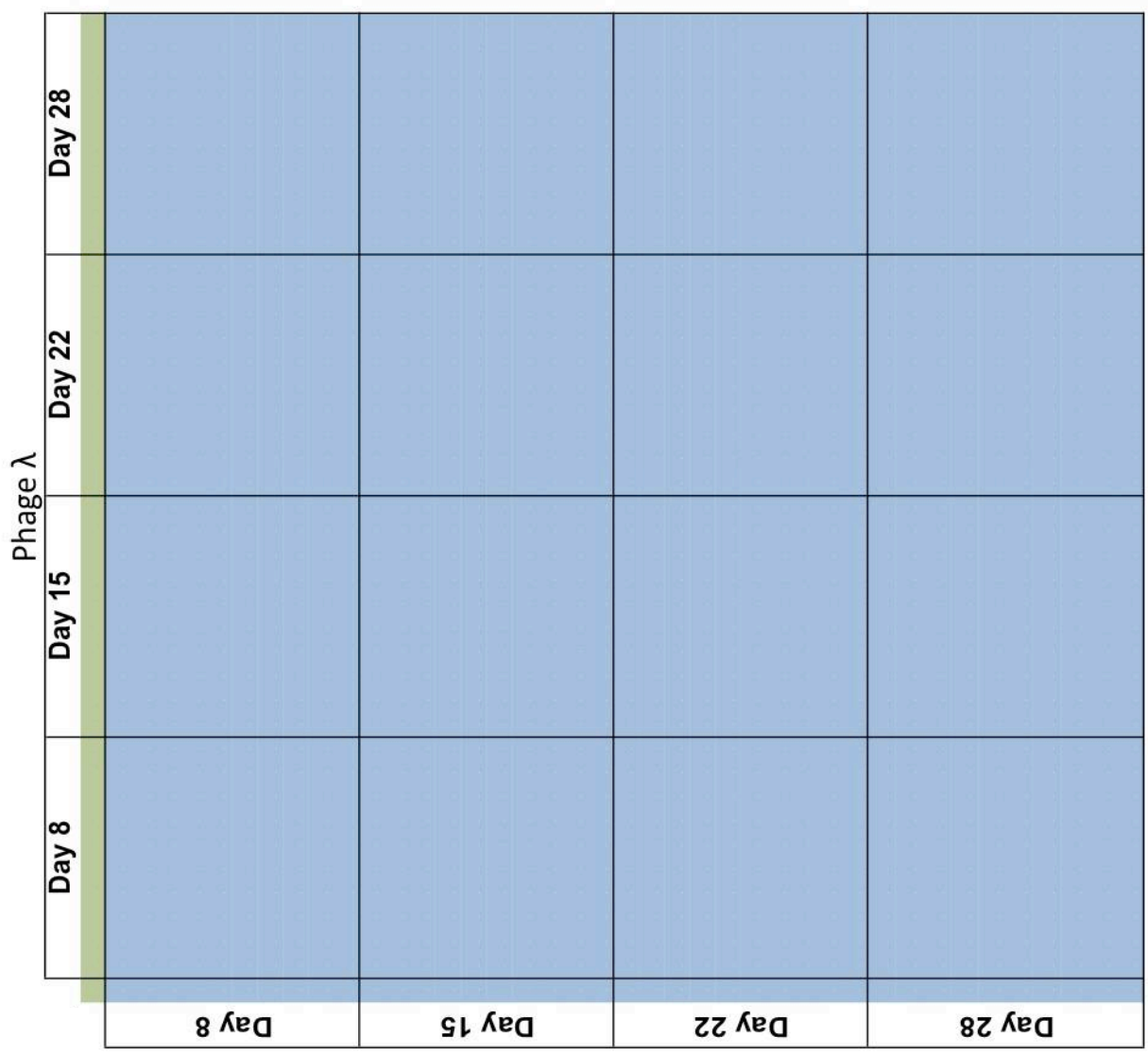


Figure 6.1 (continued on next page)

Figure 6.1 continued from previous page.

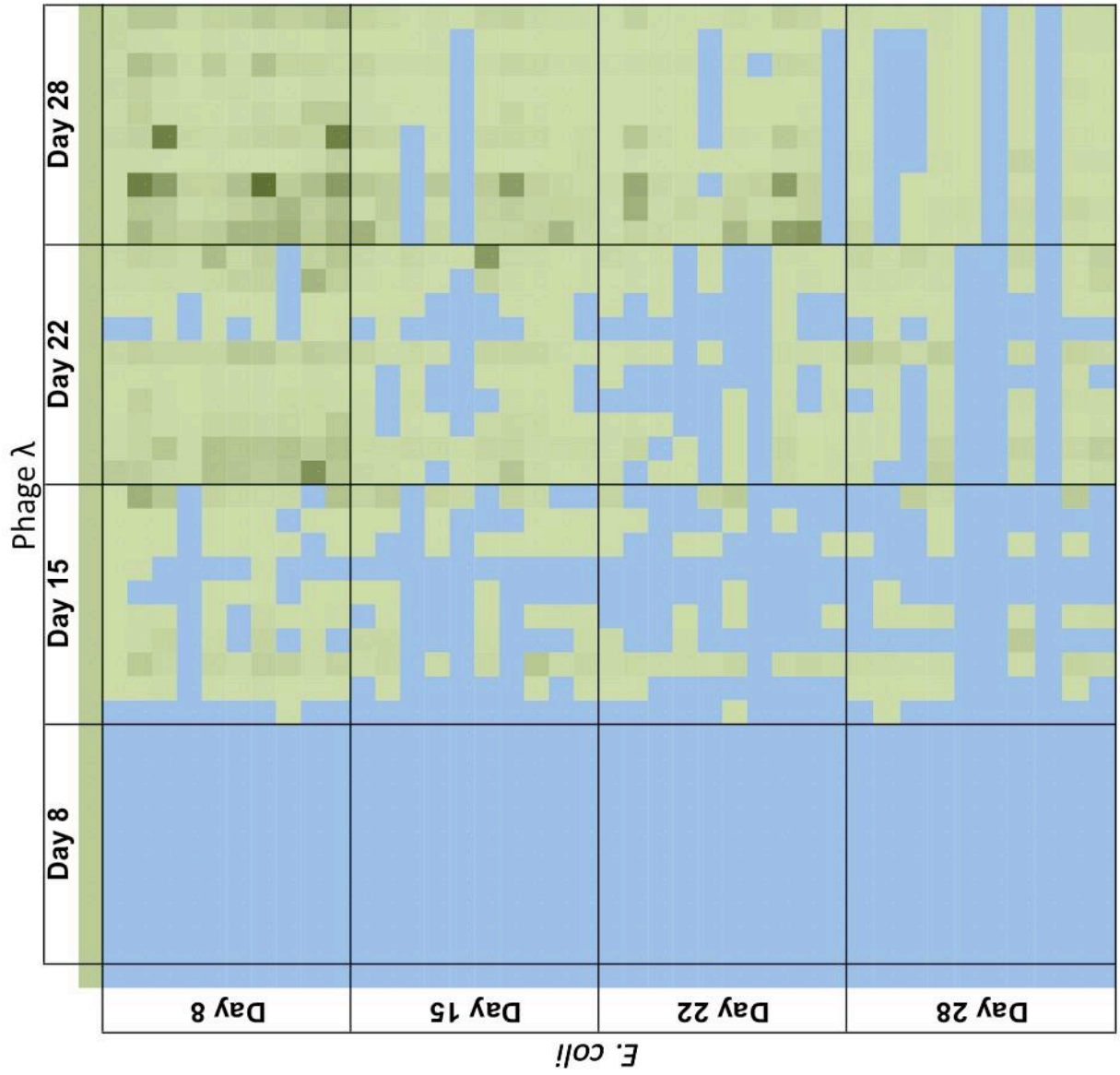


Figure 6.1. Interaction matrices for phage λ and *E. coli* from two replicate communities. **(A)** Community in which λ evolved the new ability to use OmpF as a receptor. **(B)** Community in which λ remained dependent on the ancestral receptor, LamB. In each panel, 41 bacterial clones (the ancestor and 10 from each of four later time points) are arranged vertically and 41 phage isolates (the ancestor and 10 each from the same time points) are arranged horizontally. Each cell shows the estimated ability of the phage isolate to infect the bacterial clone; blue indicates no observable plaques, light green indicates minimal plaque formation, and darker shades of green indicate greater infectivity based on plaque formation relative to the same phage isolate's performance on the ancestral host.

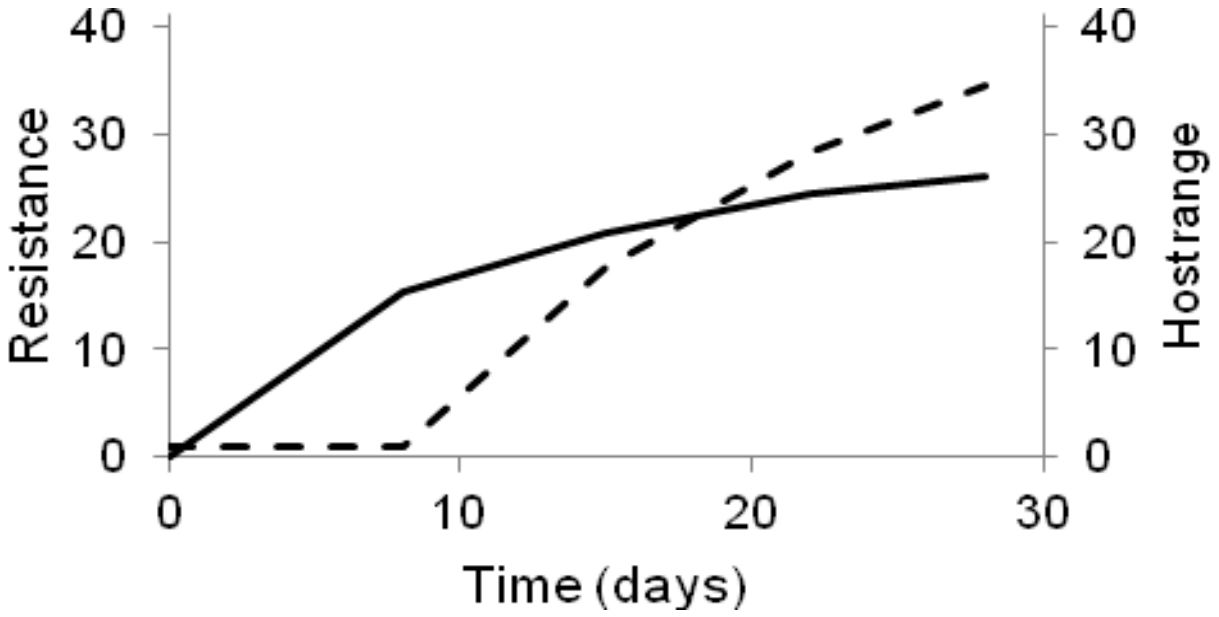


Figure 6.2. Summary statistics for bacterial resistance and phage host range for the community shown in Figure 6.1A**.** The solid line shows the average resistance of the bacterial clones over time based on the 10 clones at each time point (except day 0 for which there is one ancestral clone) and the number of the 41 phage isolates that could not infect that clone at all (blue cells in Figure 6.1A). The dashed line shows the average host-range of the phage isolates over time based on the 10 isolates at each time point (except day 0 for which there is one ancestral isolate) and the number of the 41 bacterial clones that the phage isolate could infect (all green cells in Figure 6.1A).

LITERATURE CITED

LITERATURE CITED

1 Gould, S. J. 1989 Wonderful Life: The Burgess Shale and the Nature of History (Norton, New York).

2 Blount, Z. D., Borland, C. Z. & Lenski, R. E. 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 105, 7899-7906.

3 Meyer, J. R. et al. 2012. Repeatability and contingency in the evolution of a key innovation in phage Lambda. Science 335, 428-432.

4 Katsura, I. Tail Assembly and Injection. 1981 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

5 Adams, M. H. 1959. Bacteriophages (Interscience Publishers, New York).

6 Schluter, D. 2000. The Ecology of Adaptive Radiation. (Oxford University Press, New York).

CONCLUSION

The struggle for bacteria to avoid its viruses and for bacteriophage to infect a host is unfolding in every gram of soil, milliliter of water, and inch of intestine. Each individual attempt by a phage to infect a bacterium is insignificant, however when summed over the incomputable number bacterial-phage interactions happening on earth at any given moment, its clear that their interactions matter. With the advent of new genomic tools we are just now learning of the incredible genetic and species diversity that exists in microbial communities. Unfortunately, the sequence diversity uncovered by genomic tools provides little information on microbial functional diversity and answers to the most basic questions on bacterial-phage interactions remain. For this PhD I studied a number of issues on the ecology and evolution of microbial interactions. Altogether I have characterized what are common features to microbial interaction networks, as well as many experiments on what processes drive the formation of these networks. My studies ranged in focus and scale, from biophysics of proteins to community assembly and macroevolution. Despite the diversity of topics, there are themes that reoccur in each chapter. Here I will discuss these coalescing themes. The first two are methodological problems that seem systemic to some fields of biology and the last three are more substantive about research findings.

First, molecules matter for the study of ecology and evolution. Traditional ecological theory proved too general to make accurate predictions for many of my experiments. Instead, details on the molecular interactions between the phage, bacteria, and their abiotic environments provided the tools to make much more accurate predictions for how coevolution would vary by resource environment (Chapter 2) and the conditions that would nurture the evolution of a novel function (Chapter 5). Likewise, when theory failed to predict the long-term dynamics of resistance (Chapter 1) and the appropriate form of coevolution (Chapters 2, 4, and 6), a more detailed understanding of the molecular basis of resistance provided an explanation for where the theory went wrong (Chapter 1 and 3). My dissertation is by no means the first publication to point out the importance of molecular details for understanding ecology and evolution (1-3), and hopefully with improvements in biotechnology, this practice will become more common.

Perpendicular to this lack of mechanistic detail provided by studies on ecology and evolutionary biology is the lack of statistical rigor and generalities produced by micro- and molecular biology**.** This problem was revealed in Chapters 3 and 4. As reported in Chapter 3, my colleagues from Georgia Technical Institute and I dispelled a long-standing microbiological myth of 'one phage, one receptor'. We did this not by conducting new experiments with the latest advances in biotechnology as are the tools typically relied on, but by performing a metaanalysis on data sets that had existed for decades. Similarly, I dispelled old ideas about how λ-resistance worked by creating a quantitative protocol to connect genotype to phenotype. These misconceptions were likely each the consequence of a scientific culture that emphasizes the importance of individual cases and not synthesis across a diversity genotypes and studies. Fortunately, this realization is not just my own, and the development of new Systems Biology approaches that emphasize the study diversity with high throughput techniques and to include

more quantitative rigor in molecular biology should provide the tools necessary to build more robust understandings of genetic diversity. When this happens, then the integration of ecology and evolution with molecular biology will be much smoother.

Even though much of my dissertation emphasized the importance of studying minute molecular details, I am pleased to report that there were three general themes on coevolution that reoccurred in multiple chapters irrespective of the molecules involved. First, antagonistic coevolution between bacteria and phage, at least in the short term, takes on the form of extended host-range coevolution and not matching alleles (4). Chapter 3's metaanalysis revealed this as well as the coevolution experiments between λ and *E. coli* described in Chapters 2-6. In Chapter 3 I discussed two hypotheses for why extended host-range coevolution is common, one molecular and one eco-evolutionary explanation. Despite spending Chapter 4 discussing the molecular basis of extended host-range coevolution, I expect that the molecular hypothesis does not provide an explanation for the generality of extended host-range coevolution. There are too many types of molecular interactions underlying bacterial-phage coevolution, not just the ligand-receptor dynamics I studied, for molecules to explain this result. Instead, it seems more likely that some aspect of the dynamics of antagonistic coevolution tend to drive directional selection for increased resistance and broader host-ranges which favors the emergence of trait variation in bacteria and phage that produce patterns of extended host-range coevolution rather than matching alleles.

Another overarching theme is that antagonistic coevolution is a creative process able to generate biodiversity, complexity, and innovation. Comparisons between experimental replicates in Chapter 2 and Chapter 6 where coevolution occurred versus those that had minimal evolution reveal that coevolution produced much more genotypic diversity. Besides selecting for

genetic diversity, the interactions between genotypes within the coevolved communities were much richer and complex than the evolutionary labile communities. Lastly, coevolution repeatedly promoted the evolution of phage novelty in Chapter 5, even though the novel function required the unlikely combination of four mutations. Coevolution was clearly the reason such an unlikely transition occurred, the host provided the stepping-stones necessary for the phage to evolve each mutation and the key innovation.

Lastly, trade-offs between resistance and competitiveness were themes visited in many chapters. Whether or not bacteria gained or lost resistance in chapter one depended on the cost of resistance. Likewise, for chapter two, trade-offs between nutrient uptake and resistance determined whether or not bacteria and phage would engage in an arms race. Even though trade-offs were not directly measured in chapter three, it was discussed that the existence of stable nested interaction matrices required trade-offs between viral host-range and some aspect of phage fitness. And lastly, I demonstrated the biophysics underlying trade-offs between resistance and competiveness in chapter four. Trade-offs are the primary cause of biological diversity and many other population biology phenomenon. It is my hope that this dissertation improves our understanding of the molecular causes and ecological and evolutionary effects of trade-offs.

Altogether I have explored many aspects of the interactions of bacteria and phage, but have only scratched the surface. My plan is to follow up on many of the studies reported here. Some of them include, 1) determining the fitness landscape for λ evolving to exploit OmpF and how this landscape changes as the host coevolves, 2) structural biology of J-LamB and J-OmpF interactions, 3) sympatric speciation of phage following the key innovation, and 4) answering why the nested pattern to interaction matrices is so prevalent.

LITERATURE CITED

LITERATURE CITED

1. Lenski, R. E. 1995. Molecules are more than markers: new directions in molecular microbial ecology. Molec. Ecol. 4:643-651.

2. Kneitel, J. M. and J. M. Chase. 2004. Trade-offs in community ecology: linking spatial scales and species coexistence. Ecol. Let. 7:69-80.

3. Gudelj, I., J. S. Weitz, T. Ferenci, M. C. Horner-Devine, C. J. Marx, J. R. Meyer, and S. E. Forde. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol. Let. 13:1073-1084.

4. Agrawal, A. & Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4: 79-90.

APPENDIX 1

Validation of 'spot' plate technique

A technique known as 'spot' plating has been used for many years to test the infectivity of phage on their bacterial hosts (1). To perform this technique, one 'spots' a small volume of phage lysate onto a lawn of bacteria suspended in a matrix of soft agar. The bacteria and phage are allowed to grow for 18 h on these plates. The bacteria will produce an uninterrupted lawn if the phage are unable to infect them. However, if the phage are able to infect, they will produce small regions of clearing in the region where the phage lysate contacted the bacterial lawn. If the phage are highly infective, then the entire region will be cleared. If the phage are rather poor at infecting the particular bacterial genotype, then only a few will have produced successful infections and only a few small regions will be cleared, thus forming a cluster of more or less distinct plaques. This technique has been widely used and accepted as a way to measure the infectivity of phage, although other parameters besides infectivity can sometimes affect the extent of clearing. For example, how fast the bacteria grow in the soft agar matrix can influence the apparent clearing (2). This effect could have been problematic for our study because the ancestral and evolved clones have very different growth rates in the liquid environment where they evolved (3) and, moreover, their adaptations to that environment also have diverse pleiotropic effects (4) that could influence how the evolved clones grow in soft agar.

Therefore, we performed supplementary experiments to measure bacterial sensitivity to phage infection in liquid culture, without the potentially confounding effects of growth-rate differences in the structured soft-agar medium. For these supplementary experiments, we grew the two ancestral variants, REL606 and REL607, and one clone isolated at 40,000 generations from each of the six populations in LB broth overnight. We diluted the bacteria in 0.14 M NaCl solution, plated samples on LB agar to estimate the total density of bacteria, added ~3 x 10^7 of phage T6*

per ml to each culture, and again plated samples on LB agar to estimate the density of uninfected bacteria shortly after the phage introduction. If the bacteria were highly sensitive, then very few or no colonies would form on the plate. By contrast, if the bacteria were relatively or fully resistant, then many colonies would grow because the phage had little time to adsorb to the bacteria before they were placed on solid medium.

We observed the same qualitative difference in these supplementary experiments as with the spot plates. In particular, the ancestral strains, REL606 and REL607, were much more resistant than the clones isolated at 40,000 generations (Figure A1.1). There were almost no differences in the number of bacterial colonies produced by the two ancestral strains before and after the phage were added. However, there were substantial declines in the number of colonies produced by the evolved clones after the phage were added. A Mann-Whitney U test indicates a significant difference in the proportion of infected cells between the ancestral and evolved strains ($n_1 = 2$, $n_2 = 6$, $U = 12$, one-tailed $p = 0.037$). These data thus support the results obtained from the spot plates, which show that the bacteria evolved increased sensitivity to phage T6* while they adapted to the phage-free environment of the long-term experiment.

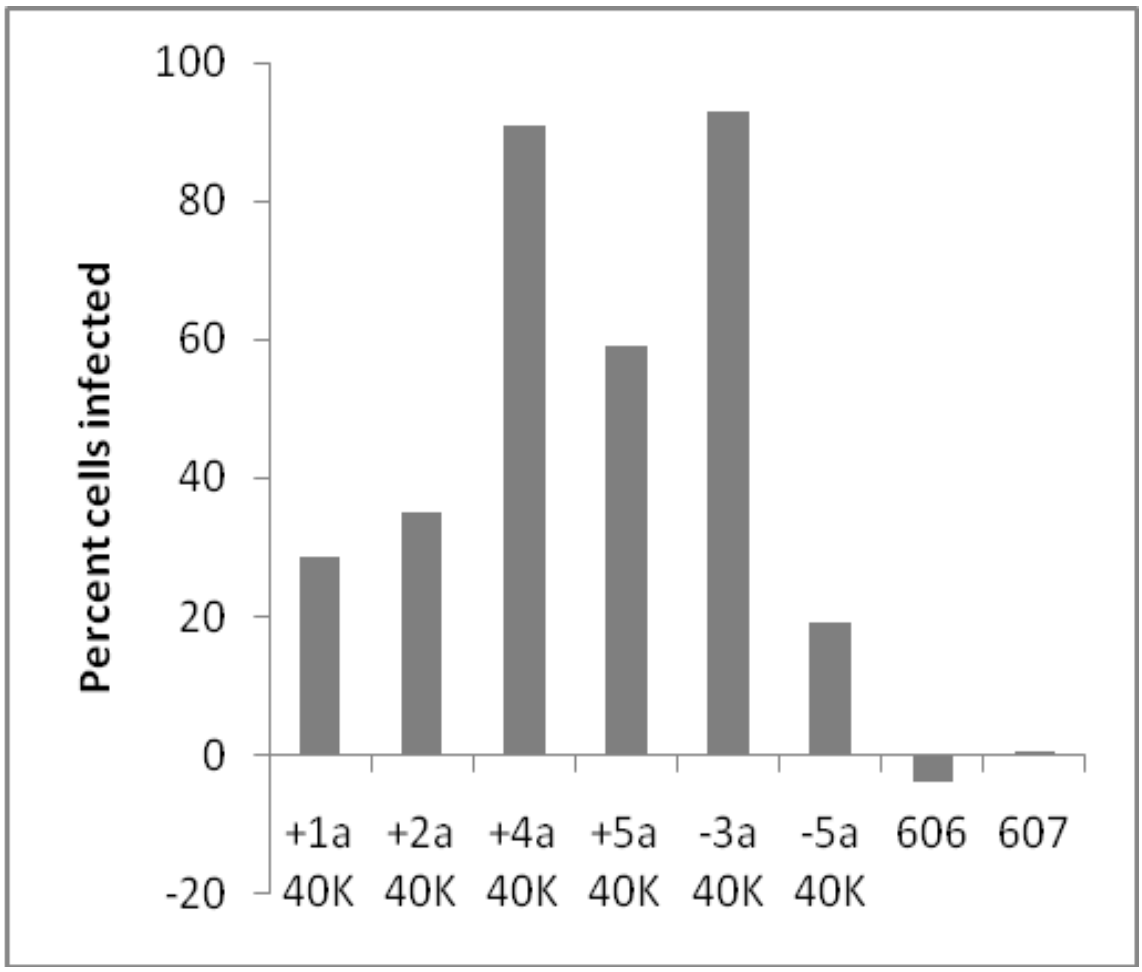


Figure A1.1: Percent cells infected, as estimated from the difference in colony-forming bacteria before and after phage T6* was added to liquid cultures.

Comparison of phage sensitivity of several *E. coli* strains

To better gauge the differences between the ancestral and evolved *E. coli* strains in their sensitivity to phages T6, T6*, and λ, we also quantified the variability in sensitivity among several reference strains. These data are summarized in Table A1.1. REL606 and REL607 are the *E. coli* B progenitors of the long-term lines, while REL606 T6-sensitive (JRM100) was

constructed by replacing the non-functional REL606 *tsx* allele with a functional *tsx* allele derived from *E. coli* K12. The strains C600 and JA221 are two derivatives of *E. coli* K12.

Three replicate spot plates were used to obtain plaque counts from the same phage lysates (stock cultures). All phage lysates were grown on REL606 cells. REL606 (5) and JA221 (6) lack the functional restriction enzymes characteristic of *E. coli* B and K12 strains, respectively. Hence, differences in plaque counts should largely reflect differences in phage adsorption to host cells, with the exception of strain C600, whose functional restriction system may reduce the overall infectivity of phage that were previously grown on another host that does not modify the phage DNA in a way that protects it from the relevant restriction enzymes.

The ancestral B strains used in the long-term experiment, REL606 and REL607, are both fully resistant to phage T6. The two K12 derivatives are both highly sensitive to T6. The B strain is made highly sensitive to T6 infection by replacing its defective *tsx* allele (7) with a functional version derived from K12. The three strains that are highly sensitive to T6 are also highly sensitive to T6*, the host-range mutant. Although T6* can infect the ancestors of the long-term lines, notice that its plaque production on the T6-resistant ancestors is several orders of magnitude lower than on the T6-sensitive B and K12 strains. Finally, all strains are highly sensitive to phage λ, with the exception of the K12-derivative C600. That strain retains a functional restriction system, so that most phage grown on another host that does not appropriately modify the phage DNA will be destroyed upon infecting C600 cells.

Table A1.1: Plaque formation of several phages on several *E. coli* strains. Values are shown as plaque forming units per ml of a stock lysate; error bars are standard errors based on three-fold replication of the assays. The relative values in any particular column can be compared with other phage-bacteria combinations previously reported (Studier 1979).

***E. coli* strains**	**Phage T6**	**Phage T6***	**Phage λ**
REL606	0	$7.47 \pm 2.49 \times 10^5$	$3.93 \pm 0.59 \times 10^8$
REL607	0	$5.80 \pm 1.45 \times 10^5$	$9.20 \pm 2.73 \times 10^8$
REL606 T6-sensitive	$7.93 \pm 3.60 \times 10^9$	$6.00 \pm 0.72 \times 10^9$	$7.27 \pm 1.81 \times 10^8$
C600	$3.03 \pm 0.80 \times 10^9$	$1.12 \pm 0.12 \times 10^9$	$3.26 \pm 0.93 \times 10^4$
JA221	$9.13 \pm 1.16 \times 10^8$	$2.17 \pm 0.15 \times 10^9$	$1.95 \pm 0.17 \times 10^9$

LITERATURE CITED

LITERATURE CITED

1. Adams, M. H. 1959. Bacteriophages. Interscience, New York.

2. Loś J. M., P. Golec, G. Wegrzyn, A. Wegrzyn, and M. Loś M. 2008. Simple method for plating *Escherichia coli* bacteriophages forming very small plaques or no plaques under standard conditions. Appl. Environ. Microbiol. 16:5113-20.

3. Lenski, R. E., and M. Travisano. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. USA 91:6808-6814.

4. Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving *E. coli* populations. Nature 407:736-739.

5. Studier, F. W. 1979. Relationships among different strains of T7 and among T7-related bacteriophages. Virology 95:70-84.

6. Clarke, L., and J. Carbon. 1977. Functional expression of cloned yeast DNA in *Escherichia coli*: specific complementation of argininosuccinate lyase (*arg*H) mutations. J. Mol. Biol. 120:517-532.

7. Studier, F. W., P. Daegelen, R. E. Lenski, S. Maslov, and J. F. Kim. 2009. Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. J. Mol. Biol. 394:653-680.

APPENDIX 2

Supplementary Materials and Methods

for

The Statistical Structure of Host-Phage Interactions

Quantitative estimation of nestedness and modularity

We represent the host-phage network with a *bipartite network* consisting of three sets $G = (U,V,E)$, where U and V are disjoint sets of nodes and $E = \{\{u_i, v_j\}\}$ is the set of edges connecting nodes of different type. For example, Supplementary Figure A2.5A shows the host-phage network described in Quiberoni (88). Define $P = |U|$ the number of phages and $H = |V|$ the number of hosts. The *adjacency matrix* of the bipartite network is $A_{ij} = 1$ if there is an edge $\{u_i, v_j\} \in E$ or $A_{ij} = 0$ otherwise (see Supplementary Figure A2.5b-c). The number of links attached to node u_i is the so-called *degree* $k_i = \sum_j A_{ij}$ (similarly, we can define the degree for v_j). Distinct colors indicate whether the node is a host (blue) or a phage (yellow) and bright (dark) shading depicts high (low) degree. Visual inspection of the network reveals significant structure, which can be rigorously detected by means of standard network measurements.

We have examined different properties of host-phage networks. Many real networks have a natural community structure, where disjoint subgroups of nodes exchange many internal connections among then than with the rest of nodes. Formally, we want to compute the optimal division of the network that minimizes the number of links between subgroups (also called communities). The raw number of links at the boundary does not give a good partition of the network. For example, the community structure can be a consequence of random variations in the density of links (145). A more reliable approach uses a null model to assess the quality of a given network partition. Newman and Girvan(146) defines the *modularity* as follows:

$$Q = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(g_i, g_j)$$

where $2m = \sum_{ij} A_{ij}$ is the number of links and g_i gives the label of the community the node i belongs to. Notice that maximizing the above function yields a partition that minimizes the

expected number of links falling between different communities, i.e., when $\delta(g_i, g_j) = 0$. Modularity Q takes values between *0* and *1*: low modularity indicates the number of links between distinct communities is not significantly different from the random distribution and high modularity indicates there is a strong community structure.

Our networks are different from the networks studied with the standard modularity measure Q (see above). Here, we study bipartite networks, i.e., networks having two distinct types of nodes and there are no links between nodes of the same type. Barber defines a new modularity quantity $Q_{bipartite}$ using a specific null model for bipartite networks:

$$Q_{bipartite} = \frac{1}{4m} \sum_{ij} \left(A_{ij} - b_{ij} \frac{k_i k_j}{2m} \right) \delta(g_i, g_j)$$

where $b_{ij} = 1$ if nodes i and j are of different type and *0* otherwise. Related studies of modularity in plant-pollinator networks have used the standard modularity Q *(147)***.** Empirical analyses of bipartite networks have shown that $Q_{bipartite} > Q$, that is, the bipartite modularity can often find better community divisions than the standard modularity when we do not consider the possibility to have links between nodes of the same type (135). We use the BRIM (135) (Bipartite Recursive Induced Modules) algorithm to maximize this bipartite modularity in our host-phage networks (see the paper by Barber for full details on the BRIM algorithm). For example Supplementary Figure A2.5A and 5D show the matrix and network representations of the optimal community structure found in a host-phage network. Figure 3.5B maps the four network communities found with BRIM into coherent matrix blocks of the (sorted) adjacency matrix. Alternatively, the network representation of community structure in Figure 3.7d suggests a geometrical interpretation of the maximization of bipartite modularity in terms of link crossing minimization, a hard problem that has been extensively studied in literature (148).

Fortunato and Barthélemy have pointed out that, in large networks, modularity optimization may fail to identify modules smaller than a characteristic size-dependent scale (149). A check of the modularity obtained through modularity optimization is thus necessary. When modularity optimization finds a module S with l_S internal links, it may be that the latter is a combination of two or more smaller modules. In this case:

$$l_s < \sqrt{2L}$$

where L is the number of links in the full network (see the paper by Fortunato and Barthélemy (149) for full details on the derivation). Modules close to this resolution limit can result from the random merging of two or more sub-modules. Then, modularity optimization might fail to detect the fine modularity structure in these situations.

An important measurement of ecological networks determines to what extent they form a nested network, i.e., when the specialist species only interact with proper subsets of the species interacting with the generalists (100). The computation of the degree of nestedness involves three steps: (i) computing the isocline of perfect order, which is the curve that separates all the non-zero entries in the adjacency matrix (above the isocline) from the absence of interactions (below the isocline) in a perfectly nested network, (ii) re-arrange all the rows and columns of the adjacency matrix in a way that maximizes the nestedness and (iii) compute the temperature T as the sum of distances d_{ij} between the expected and unexpected matrix entries and the isocline:

$$T = \frac{k}{HP} \sum_{ij \in \text{unexpected cells}} \left(\frac{d_{ij}}{D_{ij}} \right)^2$$

where D_{ij} is the diagonal that cross the unexpected cell and $k = 100/U_{max}$ with $U_{max} \approx 0.04145$ is a normalization factor that makes $0 \leq T \leq 100$ (136, 137). Finally, we have normalized the temperature T in such a way that the new range is $0 \leq N \leq 1$:

$$N = \frac{100 - T}{100}$$

Supplementary Figure A2.5C shows the sorted matrix corresponding to the optimal nestedness temperature. This matrix ordering indicates the network is highly nested. There are a few unexpected interactions below the isocline of perfect order, which correspond to the links of the right side of Supplementary Figure A2.5E.

Criterion for cataloging studies as Co-evolution (EXP), Natural Communities (NAT) or Host-phage typing (TYP):

Representative host-phage studies were found using a literature search using ISI Web of Science and tracking references (both to and from the original article). Productive search terms were as follows:

- (phage or bacteriophage) and host and range
- (phage or bacteriophage) and host and typing
- (phage or bacteriophage) and host and infectivity
- (phage or bacteriophage) and characterization

Searching cross-references were also a useful means of collecting infectivity matrices. Web of Science also generated the BibTex reference information for each article. The criteria of inclusion of a study was as follows:

1) Data is available in a matrix/table format in the paper
2) The matrix included interpretable quantitative information on infection
3) The matrix had no missing values
4) The matrix could be manually verified at each cell.
5) The matrix included at least 2 hosts and 2 phages.

Thirty-eight matrices were included in the analysis. Infectivity was indicated either with shading or a (+/-) system. Different amounts of shading would indicate the degree of infection. In the (+/-) system, a '+' generally indicated a positive infection, while a '-' indicated no infection. According to these criterion, we excluded three datasets because of missing data (150-152).

The criterion for cataloging studies was as follows:

Natural communities (NAT) – 19 studies:

This criterion was applied to studies in which both phages and hosts were isolated from the environment. These types of studies are indicative of community interactions within a natural

network. These studies were then divided into one of four sub-classes: aquatic, soil, microbiome, and food items. These sub-classes were based upon the environment from which the hosts and phages were isolated.

Co-evolution (EXP) – 10 studies:

This criterion was applie to studies in which phages and/or hosts were allowed to evolve in the lab. After phages were allowed to evolve, their host ranges were then tested. Sub-classes were based upon methodology of the study, and studies were classified as either serial dilution or chemostat experiments. Importantly, matrices of the EXP class need not be reflective of a given community at a fixed moment in time.

Artificial (ART) – 9 studies:

This criterion was applied to studies in which almost all hosts and phages were either generated within the lab or came from a collection. Sub-classes indicated the origination of the host strains. Host strains were either environmental or pathogenic.

Principal component analysis

The objective of PCA is to find a new coordinate system such that the maximal variance is explained in order of each coordinate (i.e., the principal components). Each variable was normalized to have zero mean and a standard deviation of 1 so that each contributed equally to the PCA. Supplementary Figure A1.1 shows the projection of each study onto the first two principal axes and Supplementary Table A2.4 shows the detailed coordinates underlying the principal components. Roughly, principal component 1 (PC1) corresponds to the size of the matrix, and so those studies to the right-side of Supplementary Figure A2.3 tend to be large matrices and those to the left tend to be small matrices. Roughly, PC2 corresponds to the asymmetry between number of phages and number of hosts, so that the top-most studies of Supplementary Figure A2.3 have more hosts than phages, whereas the bottom-most studies have more phages than hosts. Finally, the third principal component (not shown) corresponds, roughly, to the connectance of the study.

Statistical analysis of clustering validity using a re-shuffling approach

In order to find clusters the K-means algorithm (153) (with *k=3*) has been applied to the two main components of the PCA analysis output. This output is used as benchmark for study the subdivision of the studies and compare with those of random labels. The way in which this algorithm works is the next.

Given a set of observations ($\mathbf{x}_1, \mathbf{x}_1, \ldots, \mathbf{x}_n$), where each observation in our case represents a point in the PCA-analysis output, the k-means aims to partition the *n* observations into *k* sets ($k \leq n$) S = $\{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares:

$$\arg\min_{\mathbf{S}} = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \left\| \mathbf{x}_j - \mu_i \right\|^2$$

where μ_i is the mean of the points in S_i. In our case *n=38* and *k=3*. See Supplementary Figure A2.3 for the output of this algorithm.

In order to compare the three clusters found in this algorithm with the three real categories (NAT, EXP, ART) of our studies we used the *Jaccard Index* defined as:

$$J(C,K) = \frac{a}{a+b+c}$$

Where *C* represents the real labels and *K* the labels of the output in the k-means algorithm. *a* denotes the number of pairs of points with the same label in C and assigned to the same cluster in K, *b* denotes the number of pairs with the same label, but in different clusters and *c* denotes the number of pairs in the same cluster, but with different class labels. The index produces a result in the range (0,1), where a value of 1 indicates that *C* and *K* are identical.

We find that the three real categories when compared with the output of the k-means algorithm share a Jaccard Index of 0.26. This value indicates that there exist a poor clustering of labels of the studies with the labels of the k-means algorithm. And by consequence we can say (assuming that the k-means output is the perfect subdivision) that there is not significant subdivision in the three real categories (EXP, NAT and ART).

We subjected this index to a randomization test. We generated 10,000 trials where we relabeled the studies while retaining the number of each class (EXP, NAT and ART). The distribution of the Jaccard index of these random trials is showed in Supplementary Figure A2.4.

We found a p-value = 0.34 in the Jaccard index of the real labels. This indicates that there is not a statistically significant difference between the real subdivision of the studies and those that are labeled randomly.

Statistical analysis of correlations among global properties using a Bonferroni correction

We study the correlations coefficients among the global properties. These values are show in Supplementary Table A2.5. In that table is showed also the statistical significance of those values. For evaluate the statistical significance we used a Bonferroni correction, using both, the number of combinations and the number of global properties. This correction is used in statistics when one needs to address multiple comparisons. And comes by the fact that even when there is not statistical significance, we can find just by probability that some of the comparisons are statistically significant. Therefore this correction aims to avoid this problem. We can see in the indicated table that among the statistically significant values there is only a strong correlation between the number of hosts and the number of species. Another interesting result is that there is almost no correlation (no statistical significance) between the connectance and the number of species. This is contrary to the plant-pollinator networks where the relation follows a power law.

Experimental assays of host-phage infection

Conditions and microbial cultures

The phage and bacteria were cocultured in 50ml Erlenmeyer flasks, with 10ml of liquid medium, shaken at 120 rpm, and incubated at 37 °C. The medium was an altered version of Davis Medium (154), in which we added 10 times the magnesium sulfate (1g/L) to improve phage viability and 125 mg/L of maltotriose instead of glucose because *E. coli* and phage λ are predicted to undergo a coevolutionary arms-race when provided with maltodextrins as its only source of carbon (116, 119, 120). The medium was filtered and the magnesium was added just before use in order to stop crystallization of the magnesium during the experiment. 75 separate flasks were initiated with very small populations of bacteria(~1,000 *E. coli* cells) and phage (~100 phage λ particles) to assure that the initial populations were isogenic and that all mutant bacteria and phage arose *de novo,* this is important to make sure that each community has the potential to follow its own coevolutionary path. The *E. coli* studied were of strain REL606, a

derivative of *E. coli* B acquired from Richard Lenski (Michigan State University), described in (138) and phage were of strain cI21 (λvir) provided by Donald Court (National Cancer Institute). Most phage λ strains have two life cycles, lytic and lysogenic, the second includes a latent phase where the phage genome is incorporated into the bacterial chromosome at which time the bacteria acquires immunity to phage infection. Because the goal of this study was to characterize evolved phage resistance instead of acquired resistance, we used a phage that was unable to create the resistant lysogenic bacteria. cI21 is only able to reproduce through the lytic phase because it has a chemically induced mutation in the cI gene which is a repressor protein required for lysogeny. Each flask was cultured for 24 hours and then a random subsample of 100µl of the culture was removed and transferred to 9.9ml of fresh medium. This flask was incubated and the cycle of transfer and incubation was continued once more. Three 24 hour incubations were long enough for the bacteria to evolve resistance and the phage to counter it, however not long enough for a second round of coevolution.

Isolation strategies

After 72 hours of coculturing, two bacterial clones were isolated from each flask by streaking on LB (Luria Burtani medium, recipe found in (155)) agar plates and picking single colonies. These colonies were restreaked twice more to assure the bacteria was separated from the phage. A mixed phage stock of all coevolved genotypes was created from each flask by adding 500 µl of chloroform to the remaining culture in order to kill the bacterial cells, which were removed by centrifugation (143). Two phage clones were isolated from each of these mixed phage stocks by applying an aliquot of diluted stocks onto soft agar plates and picking isogenic 'plaques'. Soft agar plates are created by suspending an isogenic population of bacteria combined with the diluted phage stock in a thin agar matrix on top of a petri dish. When a single phage particle infects a bacterial cell trapped in the agar, the phage reproduces and spreads to nearby bacteria, this continues for a number of rounds and a clearing known as a plaque is produced in the 'lawn' of viable bacteria after 24 hours of incubations at 37 °C. This plaque contains an isogenic population of phage that can be removed to create a clonal stock of phage. We made three plates for each coevolved viral population; one from each bacterial clone isolated from the same population and then one of the ancestral bacteria REL606. Clonal phage cultures were created

by isolating single plaques from the soft-agar plates and following the procedure given by (143). Plaques on the coevolved bacteria were chosen over ones grown on REL606 to increase the chance of isolating phage that had evolve specialized counter-resistance strategies that have the plietropic consequence of losing the ability to infect the ancestral REL606. Despite this effort, none of the phage isolated lost the ability to infect REL606. Besides favoring plaques on the evolved bacterial plates, we tried to choose plaques from separate plates to improve our chances of picking different phage genotypes.

Evaluating patterns of infection and cross-resistance

We determined which of the 150 bacteria isolates were resistant to the 150 phage isolates. To do this we preformed 'spot' plate assays. Spot plates are created just as the soft agar plates above were, except instead of combining dilute samples of phage into the agar, one drops 2 µl of concentrated phage stock on top of the bacterial-agar matrix. If the phage is able to infect and reproduce on the bacterium, then a clearing or 'spot' larger than a single plaque will form in the bacterial lawn after 24 hours of incubations at 37 °C. If any clearing or inhibition of bacterial growth larger than a single plaque was observed a '1' was recorded. Plaque-sized clearings were excluded because they likely represent cross-contamination or a mutant phage that has a broader host-range than the originally isolated phage. All bacterial-phage combinations without '1's were given '0's. All bacterial phage combinations were replicated five separate times, a total of 28,125 spots were assayed. To make this processes more efficient, we placed up to 96 separate phage stocks onto a single dish (150mm radius). Phage stock replicates were never placed on the same plate in order to reduce the signal of any stochastic plating effects. The five replicates were combined and a phage was only determined to be able to infect a bacterium if 3 of 5 replicates were given '1's. Lastly, phage or bacteria that had identical infection or resistance profiles as their ancestors were removed from the matrix.

Supplementary tables

Table A2.1 Characteristics of complete host-phage networks included in the present study (18, 62-97). Please note that the table extends across four pages.

	Ref.	*Source Type*	*H*	*P*	*S*	*I*	*M*	*C*	*Lp*	L_h
1	(62) Abe (2007)	ecological	11	4	15	22	44	0.5	5.50	2.00
2	(63) Barrangou (2002)	ecological	14	6	20	25	84	0.3	4.17	1.79
3	(156) Braun-Brenton (1981)	experimental	18	3	21	30	54	0.56	10.00	1.67
4	(157) Campbell (1995)	experimental	9	5	14	14	45	0.31	2.80	1.56
5	(66) Capparelli (2010)	ecological	18	8	26	54	144	0.38	6.75	3.00
6	(67) Caso (1995)	experimental	23	4	27	17	92	0.18	4.25	0.74
7	(68) Ceyssens (2009)	artificial	5	15	20	29	75	0.39	1.93	5.80
8	(69) Comeau (2005)	experimental	30	13	43	152	390	0.39	11.69	5.07
9	(70) Comeau (2006)	experimental	32	16	48	118	512	0.23	7.38	3.69

Table A2.1 (cont'd)

10	(71) DePaola (1998)	ecological	5	17	22	39	85	0.46	2.29	7.80
11	(72) Doi (2003)	artificial	15	10	25	41	150	0.27	4.10	2.73
12	(73) Duplessis (2001)	artificial	12	12	24	37	144	0.26	3.08	3.08
13	(74) Gamage (2004)	ecological	6	7	13	9	42	0.21	1.29	1.50
14	(75) Goodridge (2003)	ecological	93	2	95	60	186	0.32	30.00	0.65
15	(76) Hansen (2007)	ecological	34	12	46	146	408	0.36	12.17	4.29
16	(77) Holmfeldt (2007)	artificial	23	46	69	418	1058	0.40	9.09	18.17
17	(78) Kankila (1994)	ecological	32	12	44	346	384	0.90	28.83	10.81
18	(79) Krylov (2006)	ecological	11	10	21	73	110	0.66	7.30	6.64
19	(80) Kudva (1999)	artificial	22	3	25	33	66	0.50	11.00	1.50

Table A2.1 (cont'd)

20	(81) Langley (2003)	ecological	66	9	75	99	594	0.17	11.00	1.50
21	(82) McLaughlin (2008)	ecological	8	7	15	18	56	0.32	2.57	2.25
22	Meyer (unpub)	experimental	25	27	52	314	675	0.47	11.63	12.56
23	(83) Middelboe (2009)	experimental	11	24	35	202	264	0.77	8.42	18.36
24	(84) Miklic (2003)	ecological	24	14	38	70	336	0.21	5.00	2.92
25	(158) Mizoguchi (2003)	experimental	8	4	12	21	32	0.66	5.25	2.63
26	(86) Pantucek (1998)	artificial	102	4	106	322	408	0.79	80.50	3.16
27	(159) Paterson (2010)	experimental	100	5	105	267	500	0.53	53.40	2.67
28	(18) Poullain (2008)	experimental	24	24	48	107	576	0.19	4.46	4.46
29	(88) Quiberoni (2002)	ecological	20	11	31	89	220	0.40	8.09	4.45
30	(89) Rybniker (2006)	artificial	17	14	31	70	238	0.29	5.00	4.12

Table A2.1 (cont'd)

31	(90) Seed (2005)	artificial	24	6	30	31	144	0.22	5.17	1.29
32	(91) Stenholm (2008)	ecological	28	22	50	348	616	0.56	15.82	12.43
33	(92) Sullivan (2003)	ecological	21	44	65	148	924	0.16	3.36	7.05
34	(93) Suttle (1993)	artificial	7	9	16	11	63	0.17	1.22	1.57
35	(94) Synnott (2009)	ecological	16	16	32	207	256	0.81	12.9	12.94
36	(95) Wang (2008)	ecological	18	7	25	11	126	0.09	1.57	0.61
37	(96) Wichels (1998)	ecological	59	23	82	318	1357	0.23	13.83	5.39
38	(97) Zinno (2010)	ecological	18	27	45	49	486	0.10	1.81	2.72
		Average	26.55	13.21	39.76	114.87	314.32	0.39	10.91	4.88
		Median	19.00	10.50	31.00	65.00	203.00	0.34	6.13	3.04
		Total	1009	502	1511	4365	11944			

*First column: These ID's corresponds to indexes in supplementary figures 3.1-3.

Table A2.2 Characteristics of complete host-phage networks included in the present study, including additional information on biological context of each study (1-37). Please note that the table extends for seven pages

ID	***Ref***	***Bacteria***	***Phage***	***Majority source***	***Additional source***	***Isolation Habitat***	***Bacterial association***	***Bacterial trophy***	***Geography***
1	(1) Abe (2007)	*Escherichia coli*	*T2 and PP01*	ecological	artificial		human pathogen	heterotrophic	
2	(2) Barrangou (2002)	*Leuconostoc*	*Caudovirales*	ecological	artificial	sauerkraut	Free	heterotrophic	North Carolina, USA
3	(3) Braun-Brenton (1981)	*Escherichia coli*	*λ*	experimental		lab-agar plates	human symbiont	heterotrophic	
4	(4) Campbell (1995)	*Pseudomonas*	*Myoviridae*	experimental	ecological	barley roots	plant symbiont	heterotrophic	Hojbakkegaard, Denmark
5	(5) Capparelli (2010)	*Salmonella*		ecological		gastroenteritis patients	human pathogen	heterotrophic	Europe

Table A2.2 (cont'd)

6	(6) Caso (1995)	*Lactobacillus*	*Siphoviridae*	experimental		food, fresh water, soil, sewage	free	heterotrophic	Spain
7	(7) Ceyssens (2009)	*Pseudomonas aeruginosa*		artificial		hospital sewage, fresh water	human pathogen	heterotrophic	global
8	(8) Comeau (2005)	*Vibrio*		experimental		marine	human pathogen / oysters	heterotrophic	British Columbia, Canada
9	(9) Comeau (2006)	*Vibrio*	*Siphoviridae and Podoviridae*	experimental		marine	human pathogen	heterotrophic	British Columbia, Canada
10	(10) DePaola (1998)	*Vibrio vulnificus*	*Podoviridae, Stylovidae, and Myoviridae*	ecological		marine	human pathogen / oysters	heterotrophic	Gulf of Mexico
11	(11) Doi (2003)	*Lactobacillus*	*Siphoviridae and Myoviridae*	artificial		silage (fermented bovine feed)	Free	heterotrophic	Japan

Table A2.2 (cont'd)

12	(12) Duplessis (2001)	*Streptococcus thermophilus*	*Myoviridae and Siphoviridae*	artificial		Industrial cheese plants	free	heterotrophic	Quebec, Canada
13	(13) Gamage (2004)	*Escherichia coli*		ecological		human and animal fecal isolates	human pathogen	heterotrophic	Ohio, USA
14	(14) Goodridge (2003)	*Enterobacteriaceae*	*Myoviridae*	ecological		human and animal	human pathogen	heterotrophic	global
15	(15) Hansen (2007)	*Campylobacter*	*Myoviridae*	ecological		poultry intestine	human pathogen	heterotrophic	Denmark
16	(16) Holmfeldt (2007)	*Flavobacteriaceae*	*Myoviridae, Siphoviridae, and Podoviridae*	artificial	ecological	marine	Free	heterotrophic	Scandinavia
17	(17) Kankila (1994)	*Rhizobium*		ecological		soil	Free	heterotrophic	Finland

Table A2.2 (cont'd)

18	(18) Krylov (2006)	*Escherichia and Salmonella*	*T-even superfamily*	ecological		sewage	human pathogen		heterotrophic	
19	(19) Kudva (1999)	*Enterobacteriaceae*		artificial		bovine and ovine feces	human pathogen		heterotrophic	North West USA
20	(20) Langley (2003)	*Burkholderia*	*T-even and λ - like*	ecological	artificial	soil, freshwater, plant	Mycorrhizal		heterotrophic	global
21	(21) McLaughlin (2008)	*Salmonella*		ecological	artificial	swine lagoon	human pathogen		heterotrophic	Mississippi, USA
22	Meyer (unpub)	*Escherichia*	*λ*	experimental		lab - batch culture	human symbiont		heterotrophic	
23	(22) Middelboe (2009)	*Cellulophaga baltica*	*Myoviridae, Siphoviridae, and Podoviridae*	experimental	ecological	marine	Free		photosynthetic	Scandinavia

Table A2.2 (cont'd)

24	(23) Miklic (2003)	*Lactococcus lactis*	*Siphoviridae*	ecological		dairy products	Free	heterotrophic	Solvania
25	(24) Mizoguchi (2003)	*Escherichia coli*	*PP01*	experimental		lab-chemostat	human pathogen	heterotrophic	
26	(25) Pantucek (1998)	*Staphylococcus*	*polyvalent staphylophage*	artificial		clinical isolates	human pathogen	heterotrophic	Brno, Czech Republic
27	(26) Paterson (2010)	*Pseudomonas fluorescens*	*φ2*	experimental		lab - batch culture	plant symbiont	heterotrophic	UK
28	(27) Poullain (2008)	*Pseudomonas fluorescens*	*φ2*	experimental		lab - batch culture	plant symbiont	heterotrophic	UK
29	(28) Quiberoni (2002)	*Streptococcus thermophilus*	*Siphoviridae*	ecological		yogurt industrial plant	Free	heterotrophic	Argentina

Table A2.2 (cont'd)

30	(29) Rybniker (2006)	*Mycobacterium*		artificial		soil	human pathogen		heterotrophic	global
31	(30) Seed (2005)	*Burkholderia*	*Myoviridae*	artificial		soil, freshwater, plant	human pathogen		heterotrophic	
32	(31) Stenholm (2008)	*Flavobacterium psychrophilum*	*Siphoviridae, Myoviridae, and Podoviridae*	ecological		fresh water	fish pathogen		heterotrophic	Denmark
33	(32) Sullivan (2003)	*Prochloroccus Synechococcus*	*Myoviridae and Podoviridae*	ecological		marine	Free		photosynthetic	Atlantic Ocean
34	(33) Suttle (1993)	*Synechococcus and Anacystis*	*Siphoviridae, Myoviridae, and Podoviridae*	artificial	ecological	marine	Free		photosynthetic	Texas, USA

Table A2.2 (cont'd)

35	(34) Synnott (2009)	*Staphylococcus aureus*	*Myoviridae*	ecological		sewage, dairy products	bovine pathogen		heterotrophic	Tokyo, Japan
36	(35) Wang (2008)	*Synechococcus and Prochlorococcus*	*Myoviridae and Podoviridae*	ecological		marine	Free		phtosynthetic	Chesapeake Bay, USA
37	(36) Wichels (1998)	*Pseudoalteromonas*	*Siphoviridae, Myoviridae, and Podoviridae*	ecological		marine	Free		heterotrophic	North Sea, Germany
38	(37) Zinno (2010)	*Streptococcus thermophilus*		ecological		dairy products	free		heterotrophic	Italy

First column: These ID's corresponds to indexes in supplementary figures 3.1-3.

Table A2.3 Global properties

Property	***Definition***
H	number of hosts
P	number of phages
I	number of interactions
$S = H + P$	number of species
$M = HP$	size
$C = I/M$	connectance
$L_H = I/H$	mean number of interactions across host species
$L_P = I/P$	mean number of interactions across phage species

Table A2.4 PCA Analysis

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}
H	0.352	0.446	-0.179	0.131	0.389	-0.131	-0.097	0.670
P	0.247	-0.534	-0.203	0.474	-0.461	-0.140	-0.279	0.279
I	0.470	-0.138	0.143	-0.474	0.008	0.517	-0.498	0.000
S = H + P	0.444	0.218	-0.257	0.320	0.192	-0.184	-0.208	-0.688
M = HP	0.397	-0.239	-0.359	-0.542	-0.078	-0.373	0.466	0.000
C = I/M	0.188	0.062	0.743	-0.093	-0.112	-0.601	-0.164	0.000
$L_H = I/H$	0.281	-0.449	0.359	0.313	0.504	0.224	0.435	0.000
$L_P = I/P$	0.353	0.431	0.177	0.177	-0.571	0.335	0.434	0.000
	48.95%	*27.98%*	*18.55%*	*2.03%*	*1.30%*	*1.07%*	*0.11%*	*0.00*

Table A2.5 Correlation analysis

	H	*P*	*S*	*I*	*M*	*C*	*Lp*	*Lh*
H	1.000	-0.146	0.916	0.458	0.394	0.125	0.847	-0.133
P	-0.146	1.000	0.264	0.535	0.744	-0.110	-0.191	0.697
S	0.916	0.264	1.000	0.664	0.686	0.077	0.748	0.154
I	0.458	0.535	0.664	1.000	0.752	0.466	0.553	0.716
M	0.394	0.744	0.686	0.752	1.000	-0.109	0.204	0.449
C	0.125	-0.110	0.077	0.466	-0.109	1.000	0.501	0.517
Lp	0.847	-0.191	0.748	0.553	0.204	0.501	1.000	0.035
Lh	-0.133	0.697	0.154	0.716	0.449	0.517	0.035	1.000

In Green p-value < 0.05/28

In Yellow 0.05/28 < p-value < 0.05/8

TableA2.6 Isolation bias

Study	*Modularity*		*Nestedness*	
	Original	*Recalculated*	*Original*	*Recalculated*
Krylov 2006	0.123	0.136	0.901	0.839
Kudva 1999	0	0	0.630	0.630
McLaughlin 2008 - Matrix minus TSB control	0.191	0.191	0.978	0.951
McLaughlin 2008 - Matrix minus TSB minus isolation host	0.191	0.313	0.978	1.000
Middleboe 2009	0.084	0.079	0.988	0.980
Rybniker 2006	0.333	0.274	0.931	0.908
Stenholm 2009	0.183	0.187	0.928	0.931

In Red significant modular/nested studies

In Blue significant anti- modular/nested studies

LITERATURE CITED

LITERATURE CITED

1. Edwards, R. A. & Rohwer, F. 2005. Viral metagenomics. Nature Reviews Microbiology 3: 504-510.

2. Torsvik, V., Øvreås, L. & Thingstad, T. F. 2002. Prokaryotic diversity -- magnitude, dynamics, and controlling factors. Science 296: 1064-1066.

3. Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323: 741-746.

4. Venter, J. C. et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74.

5. Huse, S. M. et al. 2008. Exploring microbial diversity and taxonomy using SSU rRNA Hypervariable Tag Sequencing. PLoS Genet. 4: e1000255+-e1000255+.

6. Angly, F. E. et al. The marine viromes of four oceanic regions. 2006. PLoS Biol. 4: 2121-2131 .

7. Yooseph, S. et al. 2007. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS Biol. 5: 432-466.

8. Cox-Foster, D. L. L. et al. 2007. A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. Science 318: 283-287.

9. Tyson, G. W. et al. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37-43.

10. Gill, S. R. et al. 2006. Metagenomic analysis of the human distal gut microbiome. Science 312: 1355-1359.

11. Tringe, S. G. et al. 2005. Comparative metagenomics of microbial communities. Science 308: 554-557.

12. Levin, B. R. & Bull, J. J. 2004. Population and evolutionary dynamics of phage therapy. Nature Reviews Microbiology 2: 166-173.

13. Rohwer, F. & Thurber, R. V. 2009. Viruses manipulate the marine environment. Nature 459: 207-212.

14. Faruque, S. M. et al. 2005. Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. Proc. Natl. Acad. Sci. U. S. A. 102: 6119-6124.

15. Faruque, S. M. et al. 2005. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc. Natl. Acad. Sci. U. S. A. 102: 1702-1707.

16. Goodridge, L. & Abedon, S. T. 2003. Bacteriophage biocontrol and bioprocessing: application of phage therapy to industry. SIM News 53: 254-262.

17. Weinbauer, M. G. 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28: 127-181.

18. Wommack, K. E. & Colwell, R. R. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64: 69-114.

19. Abedon, S. T. 2008. Bacteriophage ecology: population growth, evolution and impact of bacterial viruses. (Cambridge University Press).

20. Boyd, E. F., Davis, B. M. & Hochhut, B. 2001. Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol. 9: 137-144.

21. Silander, O. K. et al. 2005. Widespread genetic exchange among terrestrial bacteriophages. Proc. Natl. Acad. Sci. U. S. A. 102: 19009-19014.

22. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. 2004. Movement of viruses between biomes. Appl. Environ. Microbiol. 70: 5842-5846.

23. Fuhrman, J. A. & Noble, R. T. 1995. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol. Oceanogr. 40: 1236-1242.

24. Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M. & Sañudo-Wilhelmy, S. 1997. Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. Limnol. Oceanogr. 42: 1492-1504.

25. Bohannan, B. J. M. & Lenski, R. E. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Lett. 3: 362-377.

26. Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. 45: 1320-1328.

27. Poullain, V., Gandon, S., Brockhurst, M. A., Buckling, A. & Hochberg, M. E. 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62: 1-11.

28. Agrawal, A. & Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4: 79-90.

29. Abe, M., Izumoji, Y. & Tanji, Y. 2007. Phenotypic transformation including host-range transition through superinfection of T-even phages. FEMS Microbiol. Lett. 269: 145-152.

30. Barrangou, R., Yoon, S. S., F., Fleming, H. P. & Klaenhammer, T. R. 2002. Characterization of six Leuconostoc fallax bacteriophages isolated from an industrial sauerkraut fermentation. Appl. Environ. Microbiol. 68: 5452-5452.

31. Braun-Breton, C. & Hofnung, M. 1981. Invivo and invitro functional alterations of the bacteriophage-lambda receptor in LamB missense mutants of Escherichia coli K-12. J. Bacteriol. 148: 845-852.

32. Campbell, J. I. A., Albrechtsen, M. & Sørensen, J. Large 2006. Pseudomonas phages isolated from barley rhizosphere. FEMS Microbiol. Ecol. 18: 63-74.

33. Capparelli, R. et al. 2010. Bacteriophage therapy of Salmonella enterica: a fresh appraisal of bacteriophage therapy. The Journal of Infectious Diseases 201: 52-61.

34. Caso, J. L. et al. 1995. Isolation and characterization of temperate and virulent bacteriophages of Lactobacillus plantarum. J. Dairy Sci. 78: 741-750.

35. Ceyssens, P.-J. et al. 2009. Survey of Pseudomonas aeruginosa and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. Environ. Microbiol. 11: 1303-1313.

36. Comeau, A. M., Buenaventura, E. & Suttle, C. A. A 2005. persistent, productive, and seasonally dynamic vibriophage population within Pacific oysters (Crassostrea gigas). Appl. Environ. Microbiol. 71: 5324-5324.

37. Comeau, A. M., Chan, A. M. & Suttle, C. A. 2006. Genetic richness of vibriophages isolated in a coastal environment. Environ. Microbiol. 8: 1164-1176.

38. DePaola, A., Motes, M. L., Chan, A. M. & Suttle, C. A. 1998. Phages infecting Vibrio vulnificus are abundant and diverse in oysters (Crassostrea virginica) collected from the Gulf of Mexico. Appl. Environ. Microbiol. 64: 346-351.

39. Doi, K. et al. 2003. A comparative study and phage typing of silage-making Lactobacillus bacteriophages. J. Biosci. Bioeng. 95: 518-525.

40. Duplessis, M. & Moineau, S. 2001. Identification of a genetic determinant responsible for host specificity in Streptococcus thermophilus bacteriophages. Mol. Microbiol. 41: 325-336.

41. Gamage, S. D., Patton, A. K., Hanson, J. F. & Weiss, A. A. 2004. Diversity and host range of Shiga toxin-encoding phage. Infect. Immun. 72: 7131-7131.

42. Goodridge, L., Gallaccio, A. & Griffiths, M. W. 2003. Morphological, host range, and genetic characterization of two coliphages. Appl. Environ. Microbiol. 69: 5364-5364.

43. Hansen, V. M., Rosenquist, H., Baggesen, D. L., Brown, S. & Christensen, B. B. 2007. Characterization of Campylobacter phages including analysis of host range by selected Campylobacter Penner serotypes. BMC Microbiol. 7: 90-90.

44. Holmfeldt, K., Middelboe, M., Nybroe, O. & Riemann, L. 2007. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. Appl. Environ. Microbiol. 73: 6730-6739.

45. Kankila, J. & Lindstrom, K. 1994. Host range, morphology and DNA restriction patterns of bacteriophage isolates infecting Rhizobium leguminosarum bv. trifolii. Soil Biol. Biochem. 26: 429-437.

46. Krylov, V. N. et al. 2006. Ambivalent bacteriophages of different species active on Escherichia coli K12 and Salmonella sp. strains. Russ. J. Genet. 42: 106-114.

47. Kudva, I. T., Jelacic, S., Tarr, P. I., Youderian, P. & Hovde, C. J. 1999. Biocontrol of Escherichia coli O157 with O157-specific bacteriophages. Appl. Environ. Microbiol. 65: 3767-3773.

48. Langley, R. 2003. Lysogeny and bacteriophage host range within the Burkholderia cepacia complex. J. Med. Microbiol. 52: 483-490.

49. McLaughlin, M. R. & King, R. A. 2008. Characterization of Salmonella bacteriophages isolated from swine lagoon effluent. Curr. Microbiol. 56: 208-213.

50. Middelboe, M., Holmfeldt, K., Riemann, L., Nybroe, O. & Haaber, J. 2009. Bacteriophages drive strain diversification in a marine Flavobacterium: implications for phage resistance and physiological properties. Environ. Microbiol. 11: 1971-1982.

51. Miklič, A. & Rogelj, I. 2003. Characterization of lactococcal bacteriophages isolated from Slovenian dairies. Int. J. Food Sci. Technol. 38: 305-311.

52. Mizoguchi, K. et al. 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157: H7 in continuous culture. Appl. Environ. Microbiol. 69: 170-176.

53. Pantůček, R. et al. 1998. The polyvalent staphylococcal phage ϕ812: its host-range mutants and related phages. Virology 246: 241-252.

54. Paterson, S. et al. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464: 275-79.

55. Quiberoni, A. et al. 2003. Comparative analysis of Streptococcus thermophilus bacteriophages isolated from a yogurt industrial plant. Food Microbiol. 20: 461-469.

56. Rybniker, J., Kramme, S. & Small, P. L. 2006. Host range of 14 mycobacteriophages in Mycobacterium ulcerans and seven other mycobacteria including Mycobacterium tuberculosis--application for identification and susceptibility testing. J. Med. Microbiol. 55: 37-42.

57. Seed, K. D. & Dennis, J. J. 2005. Isolation and characterization of bacteriophages of the Burkholderia cepacia complex. FEMS Microbiol. Lett. 251: 273-280.

58. Stenholm, A. R., Dalsgaard, I. & Middelboe, M. 2008. Isolation and characterization of bacteriophages infecting the fish pathogen Flavobacterium psychrophilum. Appl. Environ. Microbiol. 74: 4070-4078.

59. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. 2003. Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature 424: 1047-1051.

60. Suttle, C. A. & Chan, A. M. 1993. Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. Mar. Ecol. Prog. Ser. 92: 99-109.

61. Synnott, A. J. et al. 2009. Isolation from sewage influent and characterization of novel Staphylococcus aureus bacteriophages with wide host ranges and potent lytic capabilities. Appl. Environ. Microbiol. 75: 4483-4490.

62. Wang, K. & Chen, F. 2008. Prevalence of highly host-specific cyanophages in the estuarine environment. Environ. Microbiol. 10: 300-312.

63. Wichels, A. et al. 1998. Bacteriophage diversity in the North Sea. Appl. Environ. Microbiol. 64: 4128-4133.

64. Zinno, P., Janzen, T., Bennedsen, M., Ercolini, D. & Mauriello, G. 2010. Characterization of Streptococcus thermophilus lytic bacteriophages from mozzarella cheese plants. Int. J. Food Microbiol. 138: 137-144.

65. Jaccard, P. 1912. The distribution of flora in the alpine zone. The New Phytologist 11: 37-50.

66. Memmott, J. 1999. The structure of a plant-pollinator food web. Ecol. Lett. 2: 276-280.

67. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. 2003. The nested assembly of plant-animal mutualistic networks. Proc. Natl. Acad. Sci. U. S. A. 100: 9383-9387.

68. Thompson, J. N. 2005. The geographic mosaic of coevolution. (University of Chicago Press).

69. Ulrich, W. & Gotelli, N. J. 2007. Null model analysis of species nestedness patterns. Ecology 88: 1824-1831.

70. Almeida-Neto, M., Guimarães, P. R. & Lewinsohn, T. M. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. Oikos 116: 716-722.

71. Bastolla, U. et al. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature 458: 1018-1091.

72. Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proceedings of the Royal Society of London Series B-Biologial Sciences 267: 2183-2188.

73. Forde, S. E. et al. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455: 220-223.

74. Labrie, S. J., Samson, J. E. & Moineau, S. 2010. Bacteriophage resistance mechanisms. Nature Reviews Microbiology 8: 317-327.

75. Hyman, P. & Abedon, S. T. 2010. in Adv. Appl. Microbiol. Vol. 70 (eds Allen I. Laskin, Sima Sariaslani, & Geoffrey M. Gadd) Ch. 7, 217-248 (Elsevier Academic Press Inc,).

76. Maynard, N. D. et al. 2010. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. PLoS Genet. 6: e1001017-e1001017.

77. Horvath, P. & Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167-170.

78. Karginov, F. V. & Hannon, G. J. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell 37: 7-19.

79. Gudelj, I. et al. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol Lett 13: 1073-1084.

80. Poisot, T., Lepennetier, G., Martinez, E., Ramsayer, J. & Hochberg, M. E. 2011. Resource availability affects the structure of a natural bacteria-bacteriophage community. Biol. Lett. 7: 201-204.

81. Bascompte, J. & Jordano, P. 2006. in Ecological networks: Linking structure to dynamics in food webs (eds M. Pascual & J. Dunne) 143-159 (Oxford University Press).

82. Bascompte, J. 2009. Disentangling the web of life. Science 325: 416-419.

83. Lenski, R. E. & Levin, B. R. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. 125: 585-602.

84. Buckling, A. & Rainey, P. B. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society of London Series B-Biologial Sciences 269: 931-936.

85. Gómez, P. & Buckling, A. 2011. Bacteria-phage antagonistic coevolution in soil. Science 332: 106-109.

86. Weitz, J. S., Hartman, H. & Levin, S. A. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc. Natl. Acad. Sci. U. S. A. 102: 9535-9535.

87. Spanakis, E. & Horne, M. T. 1987. Co-adaptation of Escherichia coli and Coliphage λvir in Continuous Culture J. Gen. Microbiol. 133: 353-360.

88. Cohen, J. E. & Newman, C. M. A. 1985. stochastic theory of community food webs I. Models and aggregated data. Proc. Roy. Soc. B 224:421-448.

89. Cohen, J. E., Newman, C. M. & Briand, F. A 1985. stochastic theory of community food webs II. Individual webs. Proceedings of the Royal Society of London Series B-Biologial Sciences 224: 449-461.

90. Pimm, S. L., Lawton, J. H. & Cohen, J. E. 1991. Food web patterns and their consequences. Nature 350: 669-674.

91. Cohen, J. E., Briand, F. & Newman, C. M. 1990. Community Food Webs: Data and Theory. (Springer-Verlag).

92. Allesina, S., Alonso, D. & Pascual, M. A. 2008. general model for food web structure. Science 320: 658-661.

93. Williams, R. J. & Martinez, N. D. 2000. Simple rules yield complex food webs. Nature 404: 180-183.

94. Pascual, M. & Dunne, J. A. 2005. in Santa Fe Institute Studies in the Sciences of Complexity Proceedings (Oxford University Press, New York).

95. Mestres, J., Gregori-Puigjane, E., Valverde, S. & Sole, R. V. 2008. Data completeness--the Achilles heel of drug-target networks. Nat. Biotechnol. 26: 983-984.

96. Olesen, J. M. et al. 2010. From Broadstone to Zackenberg: space, time and hierarchies in ecological networks. Adv. Ecol. Res. 42: 1-69.

97. Waser, N. M. & Ollerton, J. 2006. Plant-pollinator interactions: from specialization to generalization. (University of Chicago).

98. Fortunato, S. & Barthélemy, M. 2007. Resolution limit in community detection. Proc. Natl. Acad. Sci. U. S. A. 104: 36-41.

99. Genini, J., Morellato, L. P. C., Guimarães, P. R. & Olesen, J. M. 2010. Cheaters in mutualism networks. Biol. Lett. 6: 494-497.

100. Chiura, H. X. 1997. Generalized gene transfer by virus-like particles from marine bacteria. Aquat. Microb. Ecol. 13: 75-83.

101. Watts, D. J. & Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. Nature 393: 440-442.

102. Barber, M. 2007. Modularity and community detection in bipartite networks. Physical Review E 76, 066102.

103. Atmar, W. & Patterson, B. D. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96: 373-382.

104. Rodriguez-Girones, M. A. & Santamaria, L. 2006. A new algorithm to calculate the nestedness temperature of presence-absence matrices. J. Biogeogr. 33: 924-935.

105. Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. & Kim, J. F. 2008. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J. Mol. Biol. 394: 634-643.

106. Forde, S.E., et al., 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455: 220-3.

107. Labrie, S.J., J.E. Samson, and S. Moineau. 2010. Bacteriophage resistance mechanisms. Nature Reviews Microbiology 8317-327.

108. Hyman, P. and S.T. Abedon. 2010. Bacteriophage host range and bacterial resistance, in Advances in Applied Microbiology, A.I. Laskin, S. Sariaslani, and G.M. Gadd, Editors. Elsevier Academic Press Inc: San Diego. 217-248.

109. Maynard, N.D., et al. 2010. 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: e1001017-e1001017.

110. Horvath, P. and R. Barrangou. 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167-170.

111. Karginov, F.V. and G.J. Hannon 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Molecular Cell 37:7-19.

112. Gudelj, I., et al. 2010. An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol Lett 13:1073-84.

113. Poisot, T., et al., 2011. Resource availability affects the structure of a natural bacteria-bacteriophage community. Biology Letters 7:201-204.

114. Bascompte, J. and P. Jordano. 2006. The structure of plant-animal mutualistic networks, in Ecological networks: Linking structure to dynamics in food webs, M. Pascual and J. Dunne, Editors. Oxford University Press: Oxford. 143-159.

115. Bascompte, J. 2009. Disentangling the web of life. Science 325:416-419.

116. Lenski, R.E. and B.R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. 125: 585-602.

117. Buckling, A. and P.B. Rainey. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society of London Series B-Biologial Sciences. 269:931-6.

118. Gómez, P. and A. Buckling. 2011. Bacteria-phage antagonistic coevolution in soil. Science. 332:106-109.

119. Weitz, J.S., H. Hartman, and S.A. Levin. 2005. Coevolutionary arms races between bacteria and bacteriophage. PNAS 102: 9535-9535.

120. Spanakis, E. and M.T. Horne 1987. Co-adaptation of Escherichia coli and Coliphage λvir in Continuous Culture. J. Gen. Micro 133:353-360.

121. Cohen, J.E. and C.M. Newman. 1985. A stochastic theory of community food webs I. Models and aggregated data. Proceedings of the Roy. Soc. B 224:421-448.

122. Cohen, J.E., C.M. 1985. Newman, and F. Briand, A stochastic theory of community food webs II. Individual webs. Proceedings of the Roy. Soc. B 224:449-461.

123. Pimm, S.L., J.H. Lawton, and J.E. Cohen 1991. Food web patterns and their consequences. Nature 350:669-674.

124. Cohen, J.E., F. Briand, and C.M. Newman. 1990. Community Food Webs: Data and Theory. Berlin: Springer-Verlag.

125. Allesina, S., D. Alonso, and M. Pascual. 2008. A general model for food web structure. Science 320:658-661.

126. Williams, R.J. and N.D. Martinez. 2000. Simple rules yield complex food webs. Nature 404:180-183.

127. Pascual, M. and J.A. Dunne. 2005. eds. Ecological Networks: Linking Structure to Dynamics in Food Webs. Santa Fe Institute Studies in the Sciences of Complexity Proceedings. Oxford University Press: New York.

128. Mestres, J., et al., 2008. Data completeness--the Achilles heel of drug-target networks. Nature Biotechnology 26:983-984.

129. Olesen, J.M., et al. 2010. From Broadstone to Zackenberg: space, time and hierarchies in ecological networks. Advances in Ecological Research 42:1-69.

130. Waser, N.M. and J. Ollerton. 2006. Plant-pollinator interactions: from specialization to generalization. Chicago: University of Chicago.

131. Fortunato, S. and M. Barthélemy. 2007. Resolution limit in community detection. PNAS 104: 36-41.

132. Genini, J., et al. 2010. Cheaters in mutualism networks. Biol. Let. 6:494-497.

133. Chiura, H.X. 1997.Generalized gene transfer by virus-like particles from marine bacteria. Aquat. Micro. Ecol. 13: 75-83.

134. Watts, D.J. and S.H. Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393: 440-442.

135. Barber, M., 2007. Modularity and community detection in bipartite networks. Physical Review 76:1-9.

136. Atmar, W. and B.D. Patterson, 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96:373-382.

137. Rodriguez-Girones, M.A. and L. Santamaria. 2006. A new algorithm to calculate the nestedness temperature of presence-absence matrices. Journal of Biogeography. 33:924-935.

138. Daegelen, P., et al. 2009. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J. Mol. Biol. 394:634-43.

139. Gould, S.J. 1989. Wonderful Life : the Burgess Shale and Nature of History. 1st ed. ed. New York: W.W. Norton.

140. Blount, Z.D., C.Z. Borland, and R.E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. PNAS 105: 7899-7906.

141. Meyer, J.R., et al., 2012. Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. Science 335: 428-432.

142. Katsura, I., 1981. Tail assembly and injection. Lambda II, ed. R.W. Hendrix. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

143. Adams, M.H. 1959. Bacteriophages. New York: Interscience.

144. Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, USA.

145. Guimera, R., M. Sales-Pardo, and L. Amaral. 2004. Modularity from fluctuations in random graphs and complex networks. Physical Review E 70:25101.

146. Newman, M. and M. Girvan. 2004. Finding and evaluating community structure in networks. Physical Review E 69:26113.

147. Fortuna, M., et al. 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? Journal of Animal Ecology. 79:811-817.

148. Garey, M. and D. Johnson. 1983. Crossing number is NP-complete. SIAM Journal on Algebraic and Discrete Methods 4:312.

149. Fortunato, S. and M. Barthélemy. 2007. Resolution limit in community detection. PNAS 104:36-41.

150. Jensen, E.C., et al. 1998. Prevalence of broad-host-range lytic bacteriophages of Sphaerotilus natans, Escherichia coli, and Pseudomonas aeruginosa. Appl. Environ. Micro. 64:575-80.

151. Sillankorva, S., P. Neubauer, and J. Azeredo. 2008. Isolation and characterization of a T7-like lytic phage for Pseudomonas fluorescens. BMC Biotechnology 8:80-80.

152. Yu, M.X., M.R. Slater, and H.W. Ackermann. 2006. Isolation and characterization of Thermus bacteriophages. Archives of Virology. 151:663-79.

153. MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. California, USA.

154. Carlton, B.C. and B.J. Brown. 1981. Gene mutation, in Manual of methods for general bacteriology, P. Gerhard, Editor. American Society of Microbiology: Washington DC. 222-242.

155. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1998. Molecular Cloning. Plainview, NY: Cold Spring Harbor Laboratory Press.

156. Braun-Breton, C. and M. Hofnung. 1981. In vivo and in vitro functional alterations of the bacteriophage lambda receptor in lamB missense mutants of Escherichia coli K-12. Journal of Bacteriology. 148:845.

157. Campbell, J.I.a., M. Albrechtsen, and J. Sørensen. 1995. Large Pseudomonas phages isolated from barley rhizosphere. FEMS Microbiology Ecology 18:63-74.

158. Mizoguchi, K., et al., 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157: H7 in continuous culture. Applied and Environmental Microbiology 69:170-170.

159. Paterson, S., et al. 2010. Antagonistic coevolution accelerates molecular evolution. Nature, 464:275-278.

SUPPLEMENTARY FIGURES

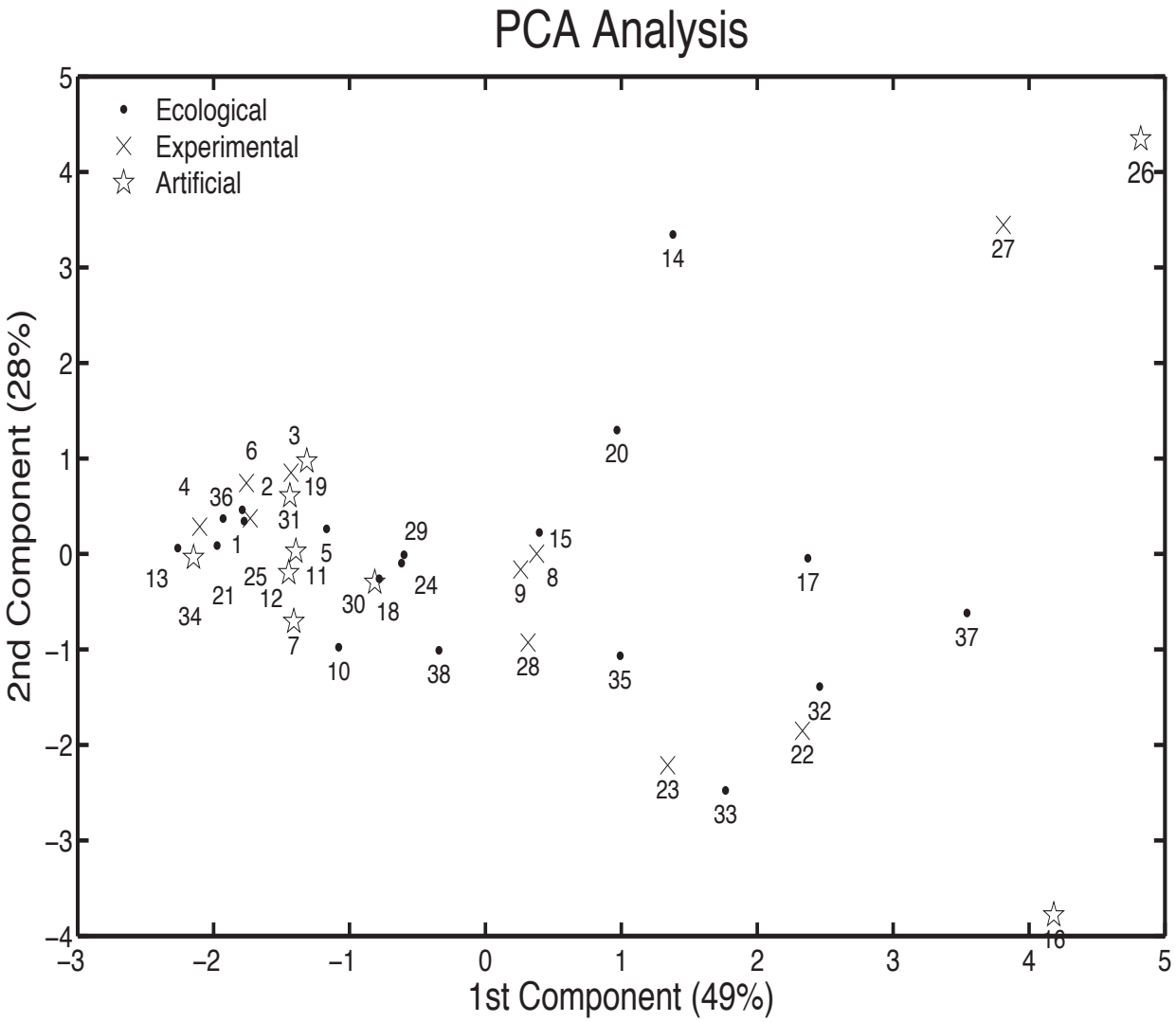


Figure A2.1 PCA Analysis in the global properties of the collected studies. Only the two main components are showed. There is no distinction between the three different type of studies.

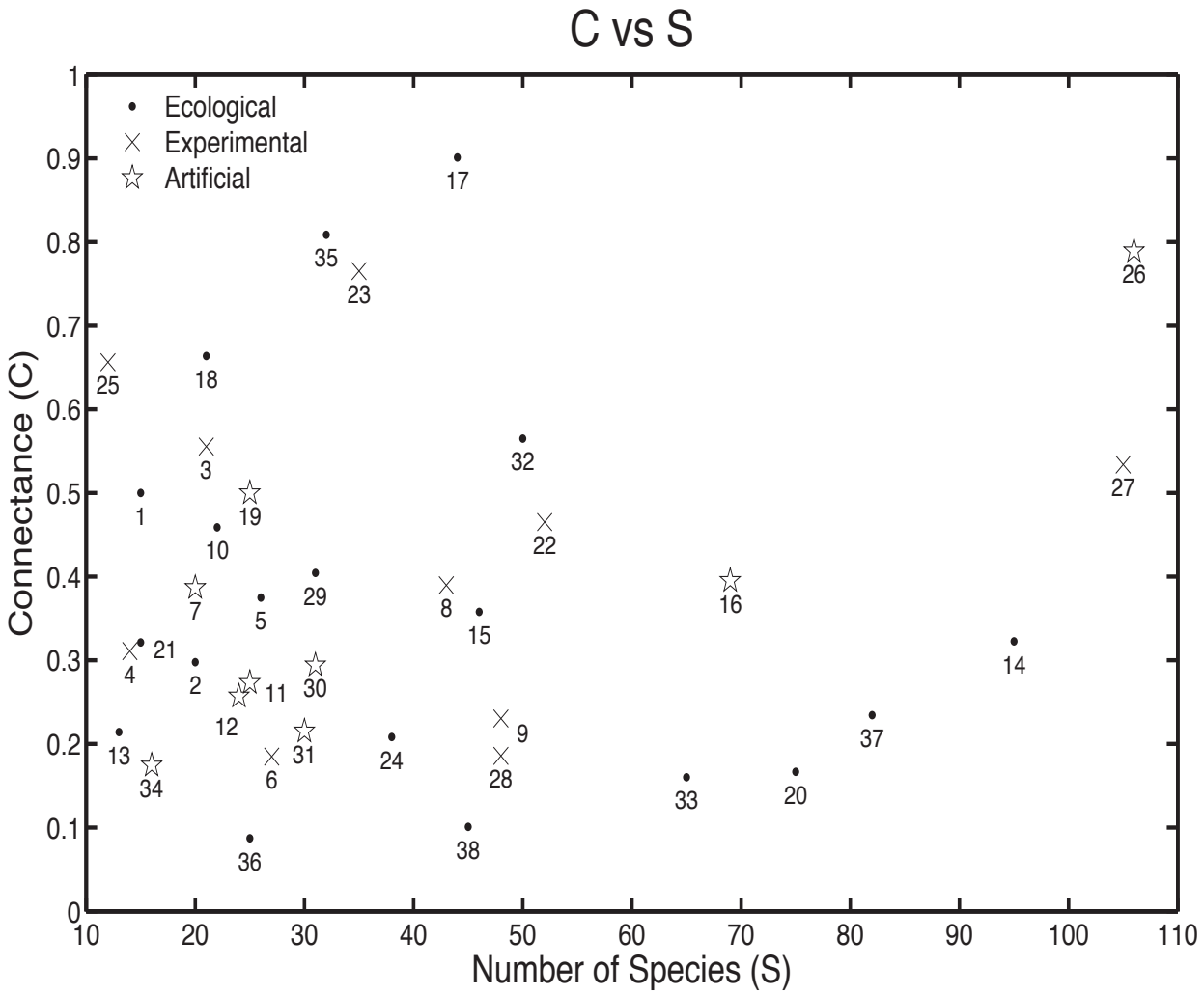


Figure A2.2 Correlation between connectance (*C*) and number of species (*S*). This plot shows that there is no relation between the connectance and the number of species. Numbers in both plots indicate the study id that can be consulted in the appendix

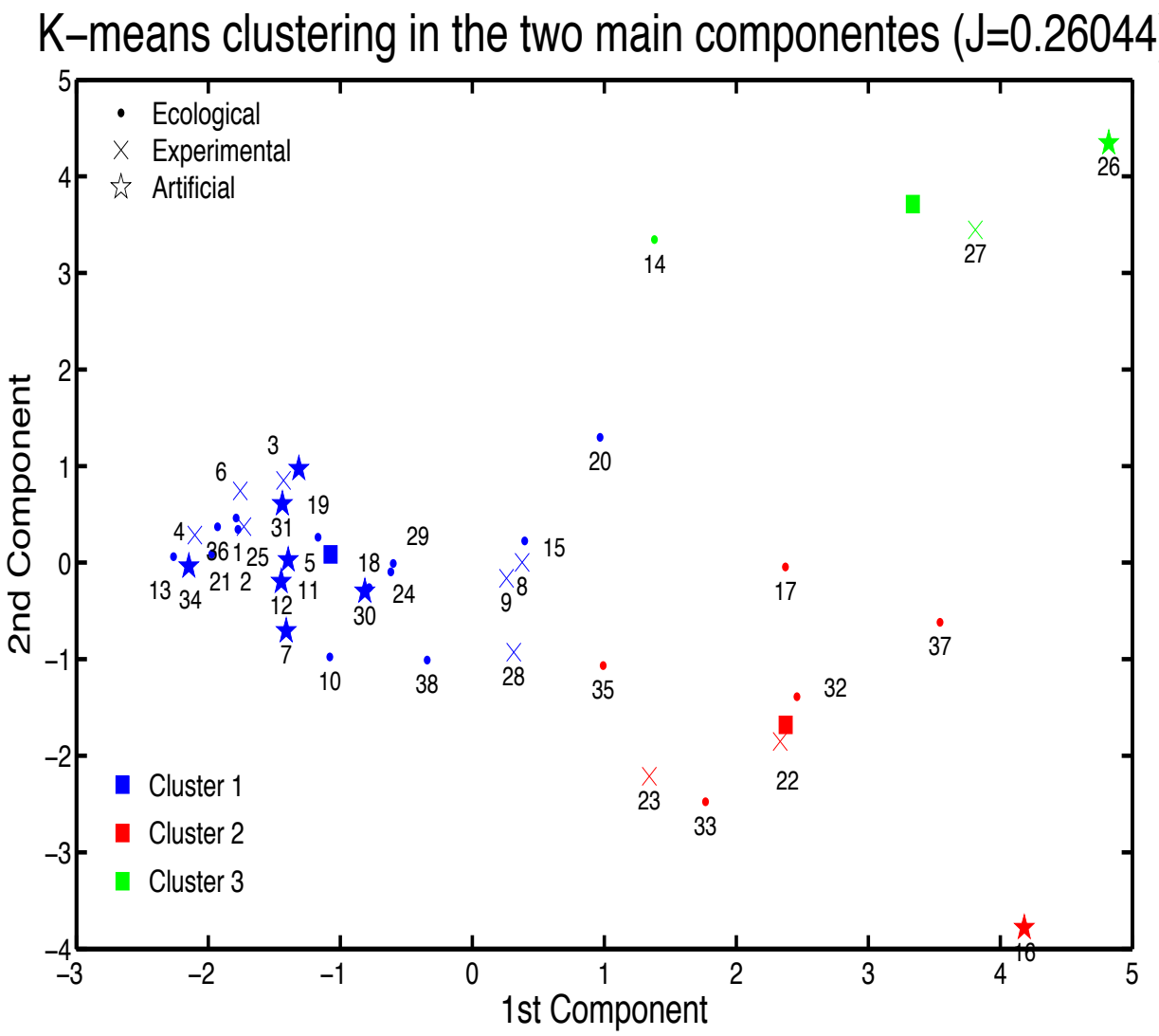


Figure A2.3 Output of the k-means (with k-3) algorithm when applied to the two main components of the PCA-analysis output.

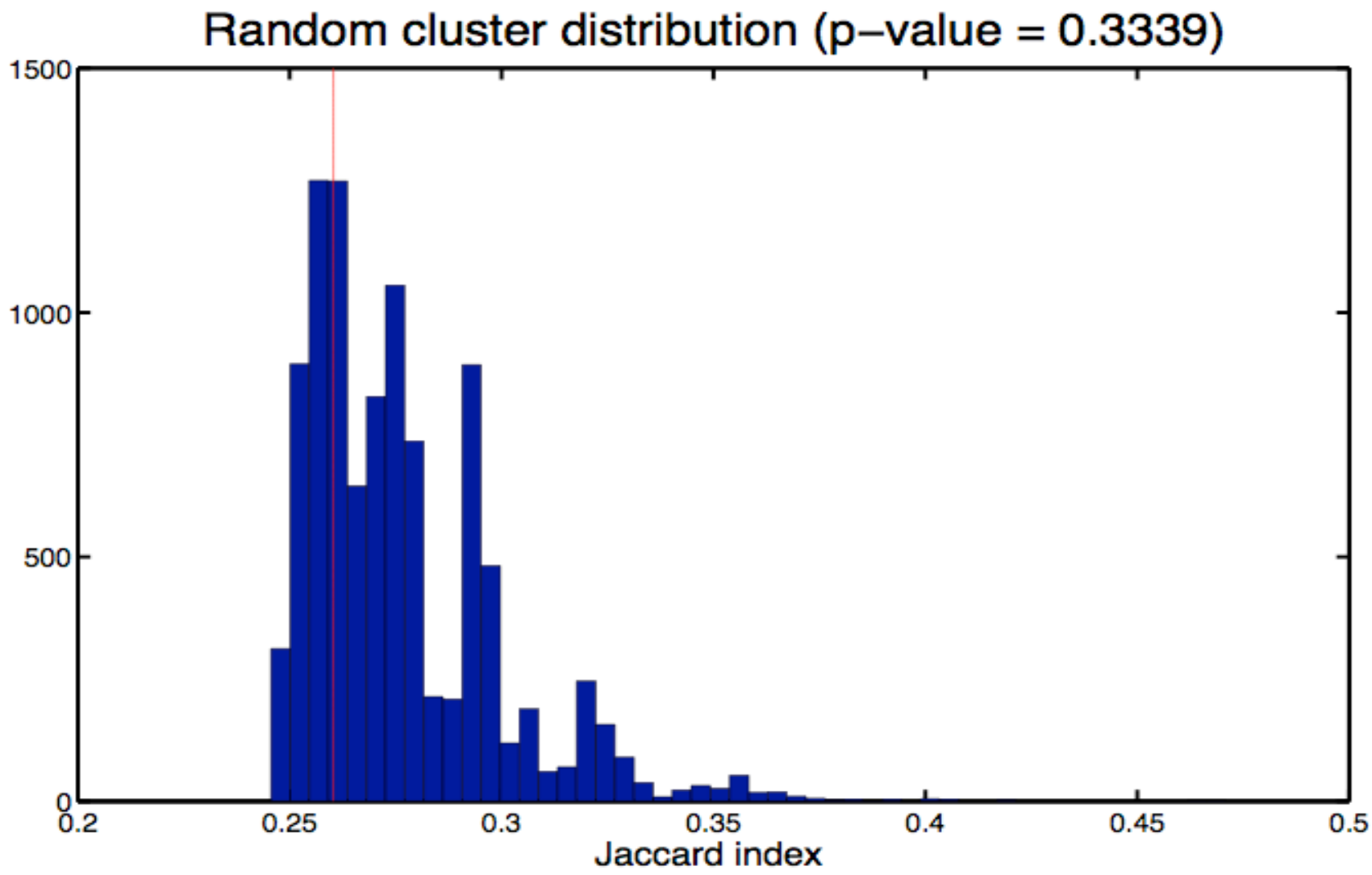


Figure A2.4 Distribution of clustering validity of source types (EXP, NAT and ART) based on global properties. The histogram denotes 10,000 randomization trials in which the labels of each study were relabeled while retaining the total number of each class (EXP, NAT and ART). The value on the x-axis is the Jaccard index of clustering validity (see Supplementary Materials and Methods). The red line denotes the observed clustering validity for the data set which is non-significant, $p = 0.34$.

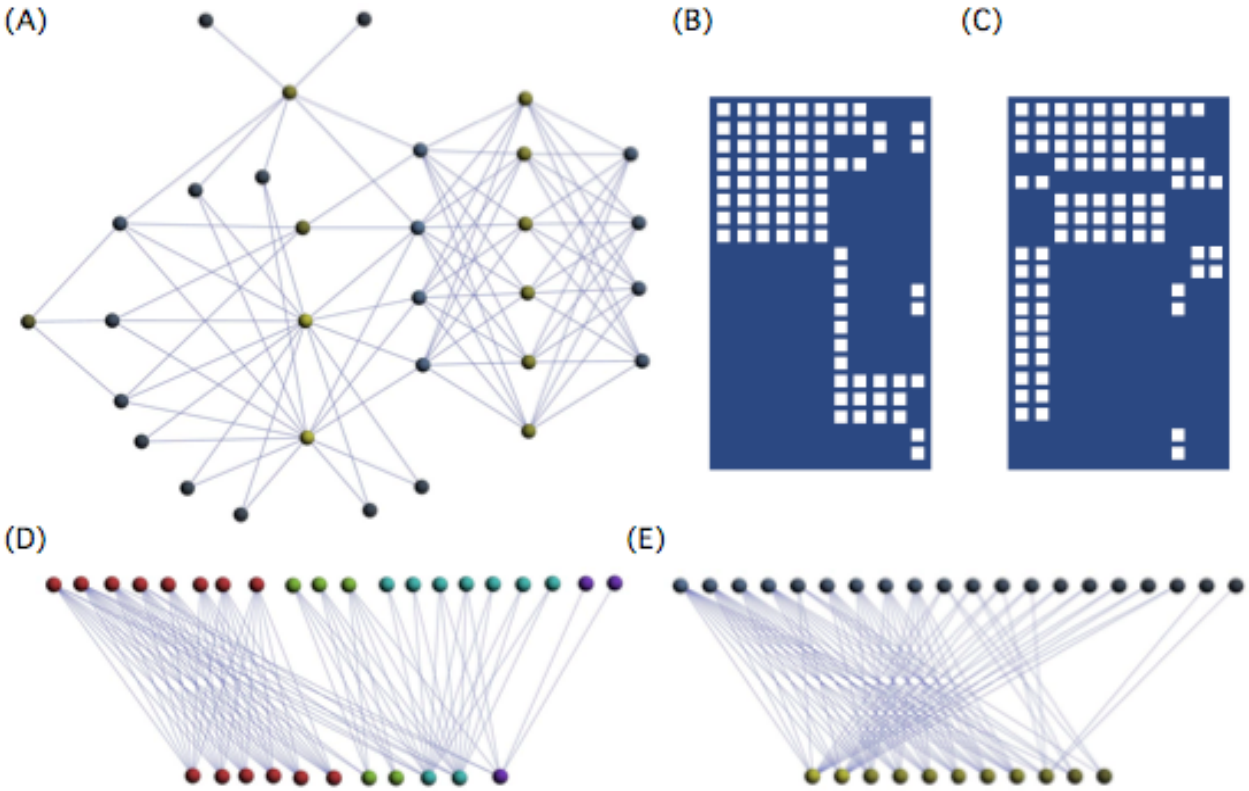


Figure A2.5 Matrix and network representations reveal non-random patterns in host-phage networks. (A) Force-directed layout of the host-phage network where yellow and blue nodes represent phages and hosts, respectively. Shading represents the number of node connections, or degree (see text). We can re-arrange the rows and columns of the adjacency matrix according to optimal network modularity (B) and degree of nestedness (C). (D) Strong modularity indicates the presence of subsets of nodes with the same color (communities) having many more internal links than external links (i.e., less crossings across different modules). (E) Network representation evidences a high degree of nestedness overall, with a few unexpected interactions between specialist species (on the right). Notice that generalist species have more connections and they are located on the left.

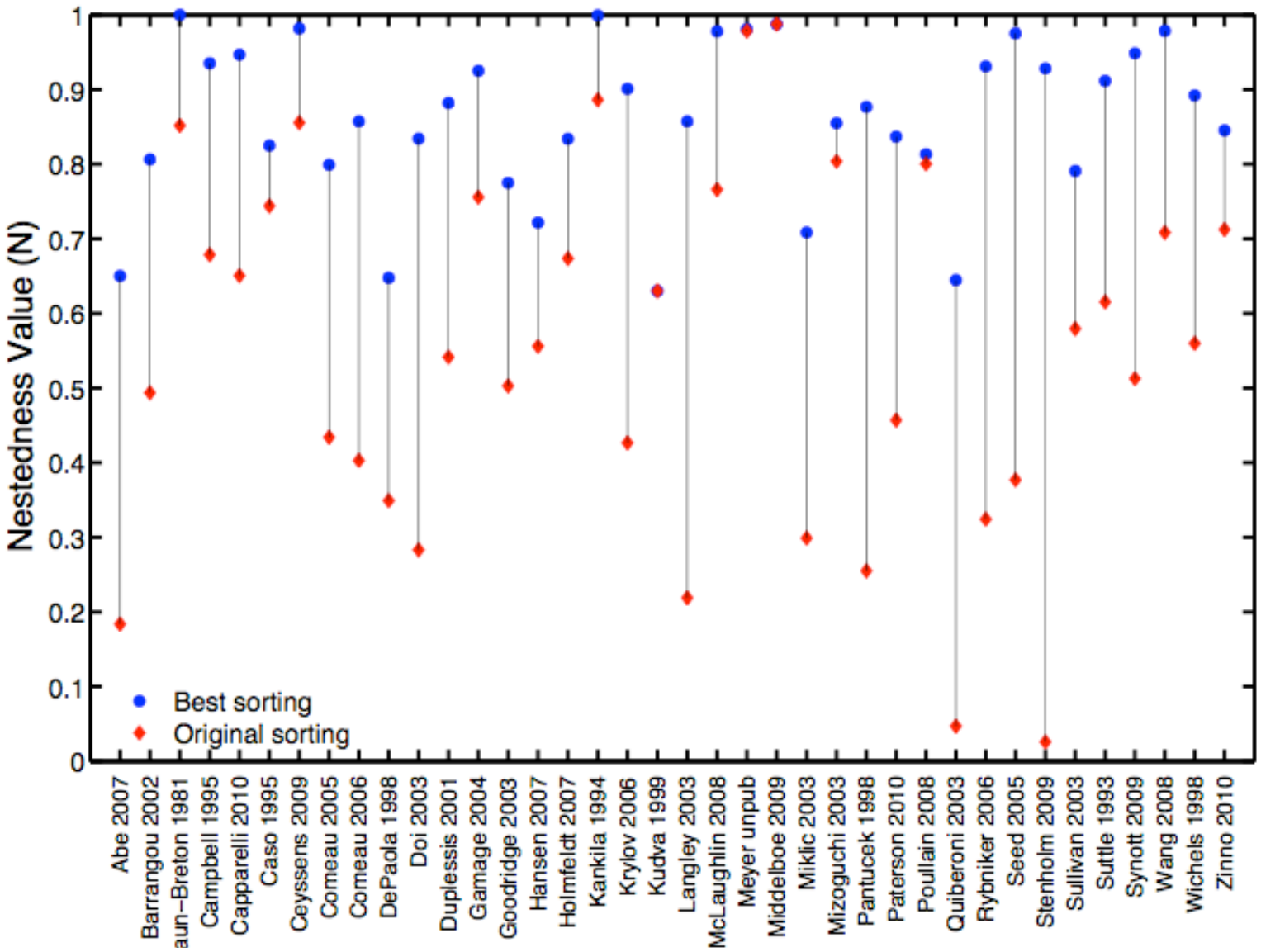


Figure A2. 6 Nestedness value compared for the original publication format of the matrix (red diamonds) vs. the value found in this study (blue circles). X-axis lists all studies in alphabetical order. Y-axis denotes the value of nestedness. Lines connect the points for ease of comparison. Note that in all cases the current value exceeded that of the original publication.

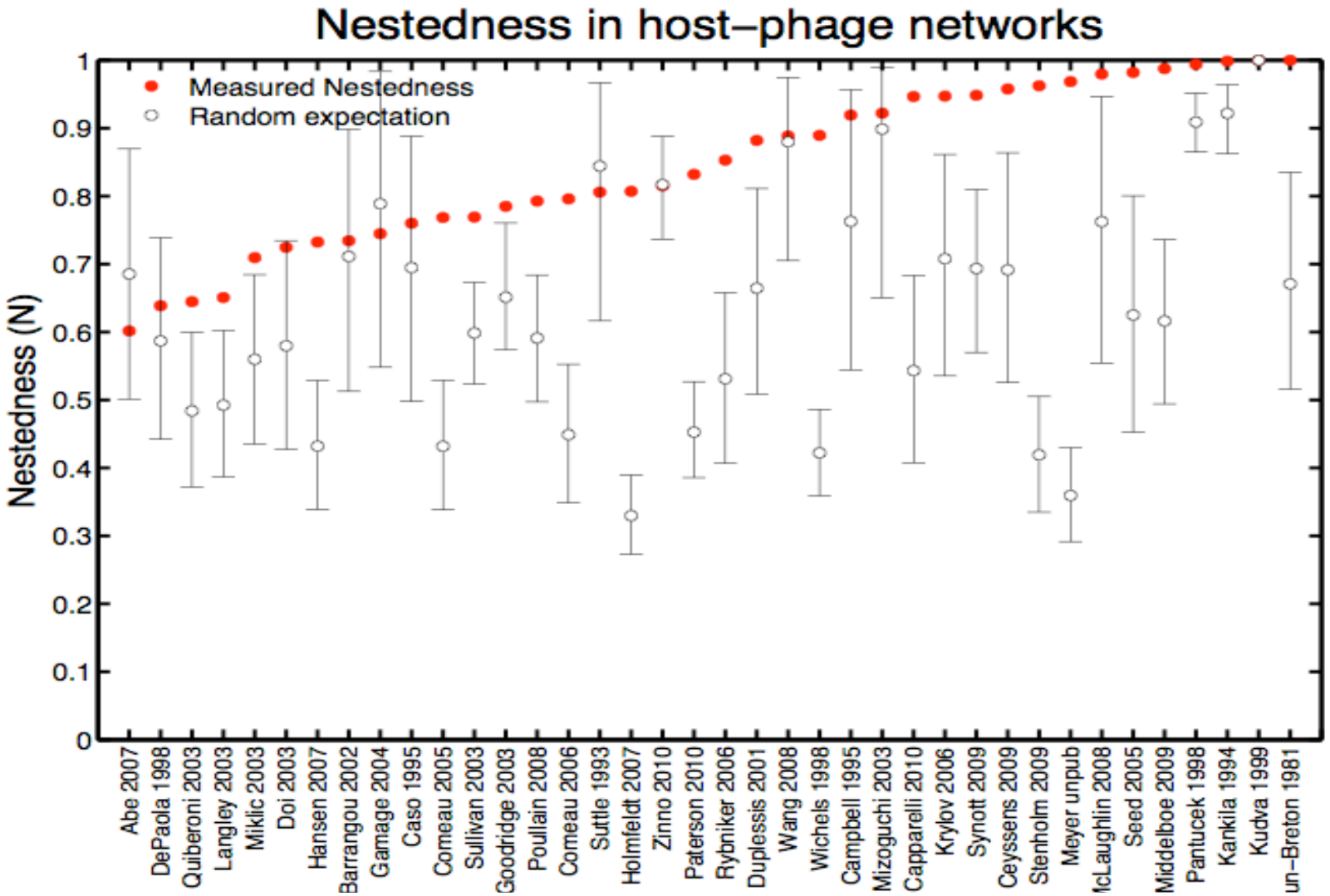


Figure A2.7 Statistical distribution of nestedness for random matrices compared to that of the original data. Here, empty rows/columns from all matrices were removed so that matrices only contain hosts that were infected by at least one phage and phages that infected at least one host. Error bars denote *95 %* confidence intervals based on 10^5 randomizations of appropriately randomized null networks. Here 26/38 are significantly nested, where Doi et al.(72) is the only study to no longer be significant at the 0.05 level compared to the original data, yet it remains highly nested (p = 0.067).

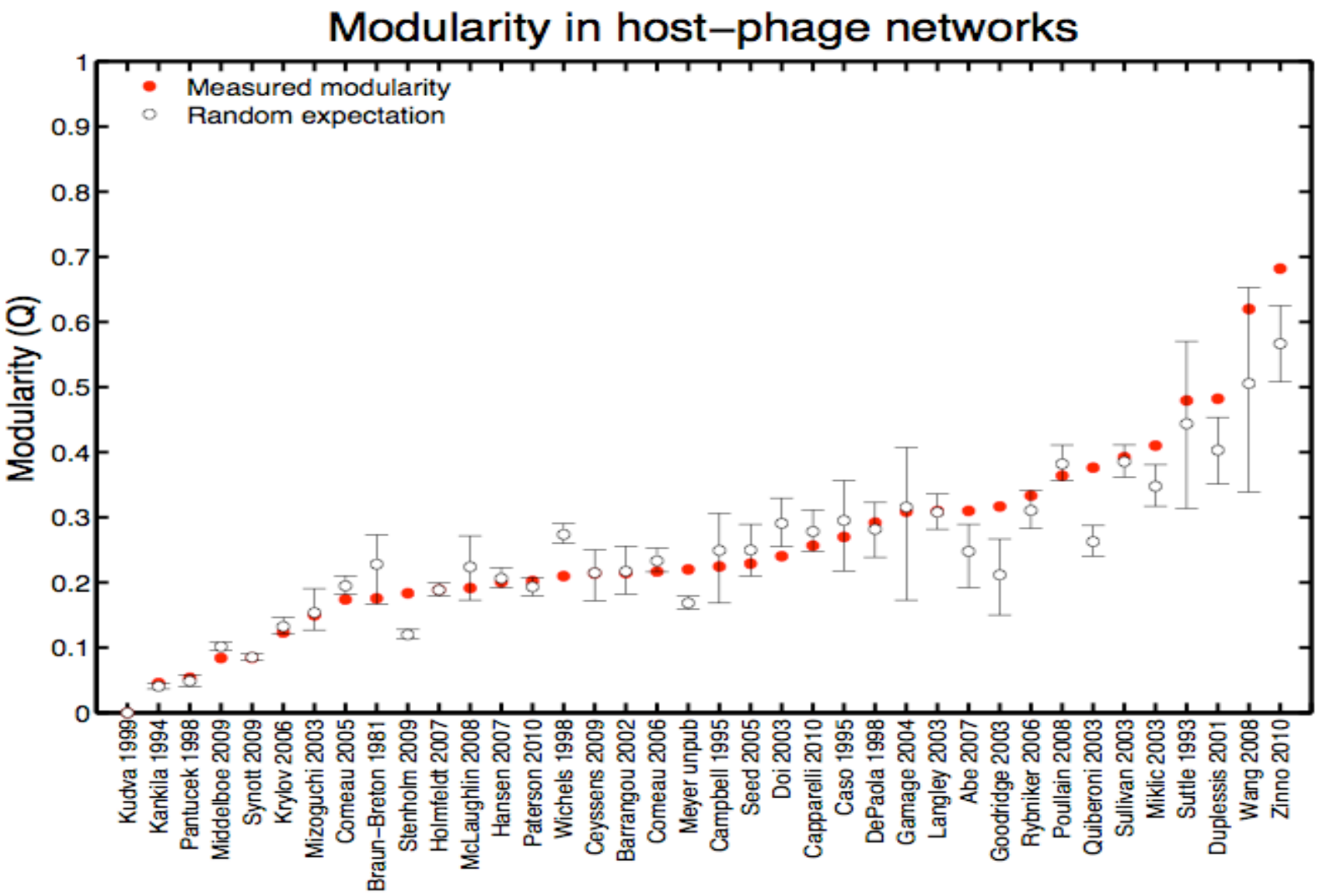


Figure A2.8 Statistical distribution of modularity for random matrices compared to that of the original data. Here, empty rows/columns from all matrices were removed so that matrices only contain hosts that were infected by at least one phage and phages that infected at least one host. Error bars denote *95 %* confidence intervals based on 10^5 randomizations of appropriately randomized null networks. Here 9/38 are significantly modular as opposed to 6/38 which were significantly modular in the original data.

APPENDIX 3

Quantifying phage infectivity patterns through image processing

The photograph of each plate was taken with a digital camera at a resolution of 72dpi (see Figure A3.1a) and processed using the public-domain software ImageJ (1). After acquisition, the picture was converted into an 8-bit image and _ltered to reduce the e_ect of non-uniform illumination. The local normalization algorithm used is based on local mean and variance values computed using local spatial smoothing (2). The obtained image was then smoothed using a gaussian blur (with _ = 4) and enhanced using a contrast stretching algorithm (normalized, with 3 saturated pixels). The resulting image was then thresholded using the iterative isodata algorithm (3), inverted and converted to a mask. By combining the resulting image with the original image we get the pre-processed image shown in Figure A3.1b). Then the region of interest (ROI) was established manually for each picture in order to determine the 8 by 12 grid where each phage was inoculated. From each cell in the grid a sub-image was automatically cropped and the area of each blob measured using a growing region algorithm with optimized thresholds.

The blob-detection algorithm can be summarized as follows: 1) determine a seed pixel, for instance bychoosing a local maximum near the center of the image; 2) then recursively add in neighbouring pixels that are above the optimized threshold, and thus increasing the size of the region; 3) convert the resulting blob into a polygonal ROI and compute its area. After repeating this process for each of the 96 cells in the grid, the obtained values are then normalized in order to measure the likelihood of infection of each phage type, a quantity that can be visualized using a heat map, as illustrated in Figure A3.1c).

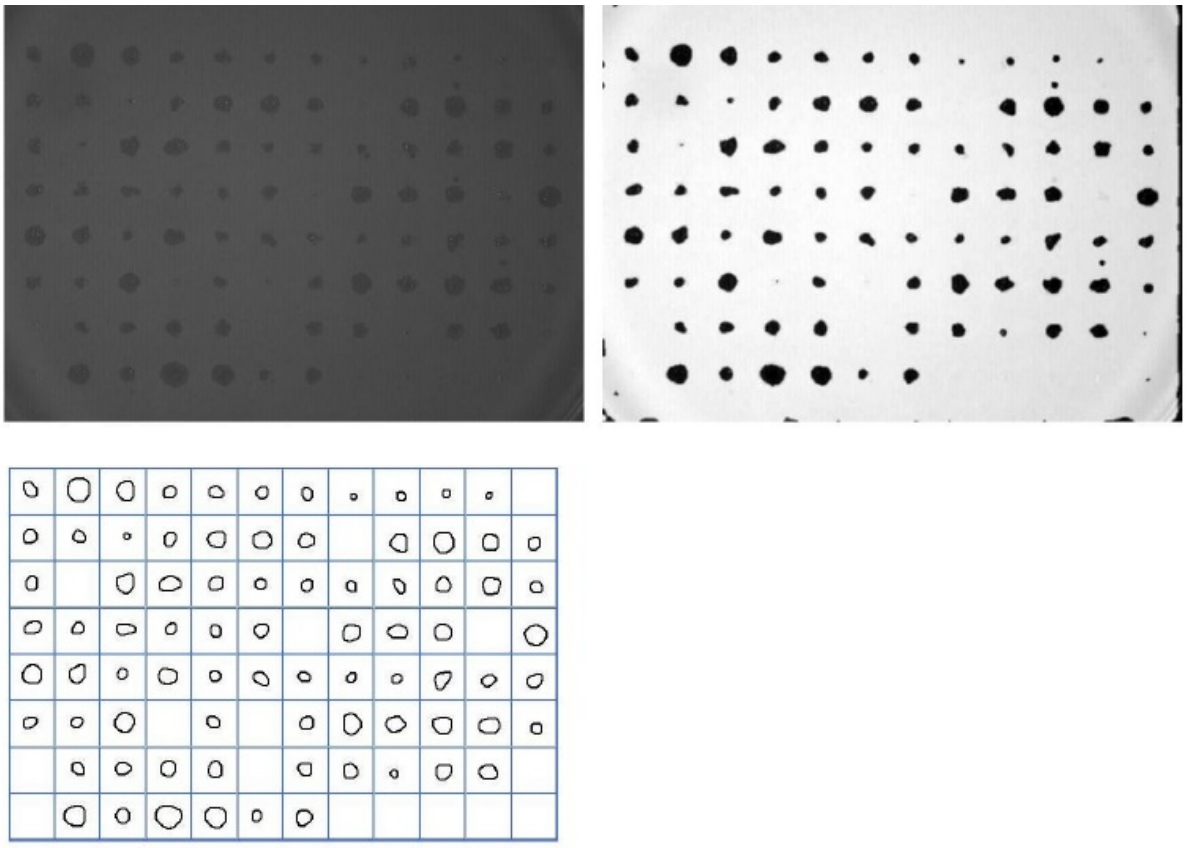


Figure A3.1: An example (bacterial type 23B) illustrating the image processing steps taken to quantify phage infectivity patterns from the picture of a plate. Top left) Photograph at 72dpi. Top right) Pre-processed image. Bottom left) Blob-detection algorithm which is converted into normalized infectivity data.

LITERATURE CITED

LITERATURE CITED

1. Abramoff, M.D., Magalhaes, P.J., Ram, S.J. Image Processing with ImageJ. Biophotonics International 11:36-42.

2. Sage, D. Local Normalization. http://bigwww.ep.ch/sage/soft/localnormalization/.

3. Ridler, T.W. and Calvard, S. 1978. Picture Thresholding Using an Iterative Selection Method. Systems, Man and Cybernetics 8:630-632.

APPENDIX 4

Supporting Online Material for

Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda

Materials and Methods

Strains. We studied *Escherichia coli* B strain REL606 because its evolution in the laboratory is well documented (*49*), its genome has been sequenced (*50*), and it is a permissive host for phage λ. Also, REL606 lacks generalized phage defenses such as a restriction modification system, CRISPR adaptive immunity, and mucoid cell formation (*50, 51*).

The phage λ strain that we used is cI26, which was provided to us by Donald Court (National Cancer Institute, MD). Most λ strains have two alternative life cycles, lytic and lysogenic. During the lytic cycle, a phage uses the bacterial cell to produce new phage particles and then lyses the host. When a phage enters the lysogenic cycle, its genome is incorporated into and replicated with the bacterial genome (*21*). The λ strain cI26 is strictly lytic as a consequence of a deletion that causes a frameshift in the *cI* gene, a repressor required for the phage to switch into the lysogenic mode. We sequenced the complete genome of cI26, which served as the ancestral reference for genetic comparisons with the evolved phage. Table A4.1 shows all of the differences between cI26 and the previously published λ reference genome (GenBank: NC_001416).

Table A4.1. Genomic differences between λ strain cI26, used as the ancestral phage in this study, and the λ reference genome (GenBank: NC_001416). Table extends for two pages.

Ref genome location	Mutation	Type	Gene position nucleotide (amino acid)	Amino acid change	Gene	Product
138	Δ1 bp	Noncoding	/–53	–	–/nu1	–/DNA packaging protein
14266	+G	Noncoding	+139/–10	–	L/K	tail component/tail component
20661	A→G	Substitution	1012 (338)	K→E	orf-401	Tail fiber protein
20835	+C	Frameshift	1186 (396)		orf-401	Tail fiber protein
21714	G→A	Substitution	686 (229)	S→N	orf-314	Tail fiber protein
21738	Δ5996 bp	Deletion			(orf-314) orf-194 ea47 ea31 ea59	Tail fiber, fiber assembly, and proteins of unknown function
31016	T→C	Substitution	9 (3)	E→E	orf61	hypothetical protein
34934	A→G	Substitution	453 (151)	G→G	lambdap48	Superinfection exclusion protein B
37818	Δ1 bp	Frameshift	123 (41)	–	cI	repressor
45618	T→C	Substitution	126 (42)	F→F	R	endolysin
46957	+A	Noncoding	-205/+85	–	bor/lambdap78	Bor protein precursor/putative envelope protein
46985	C→T	Noncoding	-233/+57	–	bor/lambdap78	Bor protein precursor/putative envelope protein

Table A4.1 (cont'd)

46992	C→T	Noncoding	-240/+50	–	bor/lambdap78	Bor protein precursor/putative envelope protein
47004	G→A	Noncoding	-252/+38	–	bor/lambdap78	Bor protein precursor/putative envelope protein
47129	A→G	Substitution	447 (149)	H→H	lambdap78	putative envelope protein
47143	C→T	Substitution	433 (145)	V→I	lambdap78	putative envelope protein
47243	G→A	Substitution	333 (111)	N→N	lambdap78	putative envelope protein
47315	G→A	Substitution	261 (87)	I→I	lambdap78	putative envelope protein
47360	G→A	Substitution	216 (72)	N→N	lambdap78	putative envelope protein
47398	C→T	Substitution	178 (60)	D→N	lambdap78	putative envelope protein
47509	T→C	Substitution	67 (23)	T→A	lambdap78	putative envelope protein
47529	C→T	Substitution	47 (16)	R→K	lambdap78	putative envelope protein
47575	C→A	Substitution	1 (1)	V→L	lambdap78	putative envelope protein
47669	T→C	Noncoding	-94/-69	–	lambdap78/ lambdap79	putative envelope protein/hypothetical protein
47878	A→G	Substitution	141 (47)	R→R	lambdap79	hypothetical protein
47973	T→C	Noncoding	+29/	–	lambdap79/–	hypothetical protein/
47977	G→A	Noncoding	+33/	–	lambdap79/–	hypothetical protein/

Table A4.1 (cont'd)

47978	T→C	Noncoding	+34/	–	lambdap79/–	hypothetical protein/
48160	T→C	Noncoding	+216/	–	lambdap79/–	hypothetical protein/

Evolution experiments. Bacteria and phage were cultured together in 50-ml Erlenmeyer flasks, each containing 10 ml of modified M9 medium (*52*) supplemented with 5 times the usual $MgSO_4$ concentration (1 g/L) to improve λ growth and 1 g/L of glucose to allow the bacteria to reach high density. We added ~10^2 phage particles and ~10^3 bacterial cells to each flask at the start of an experiment. These small numbers minimized the initial genetic variation; thus, beneficial mutations arose *de novo*, which allows one to evaluate the repeatability of evolutionary outcomes without the complicating effect of shared variation. Each flask was incubated for 24 h at 37°C and shaken at 120 rpm. After 24 h, a 100-µl sample of the community was transferred to a flask containing 9.9 ml of fresh medium. The initial experiment ran for 28 days and the large-scale experiment for 20 days. These experiments had 6 and 96 replicate communities, respectively.

Samples of the communities were periodically preserved by adding glycerol (~15% by volume) to the cultures, which were then frozen at –80°C. For the initial experiment, 1-ml samples were taken every week (days 1, 8, 15, 22, and 28), while 200-µl samples were stored daily for the large-scale experiment. Before freezing, each sample was tested for the presence of phage able to exploit a new receptor by plating a subsample (2-5 µl) onto a lawn of a mutant *E. coli* with defective LamB protein (*lamB*$^-$; derived from REL606 by a 1-bp insertion (T) after nucleotide position 610). The section on "Detection of λ that use new receptor" provides further details.

Isolation and culture techniques. To isolate bacterial clones from a community, we spread a portion of the appropriate sample on a Luria-Bertani (LB) agar plate (*52*) and, after incubation for 24 h at 37°C, picked individual colonies. The isolates were streaked twice more in the same manner to eliminate any phage particles that might be present. After the third cycle, a colony was picked and grown overnight at 37°C in liquid LB shaken at 120 rpm. Two ml of this culture was stored with 15% glycerol at –80°C. To revive cultures, ~3 µl of frozen stock was grown for one day in LB, then 10 µl was transferred to a flask containing 10 ml of modified M9 medium and grown for 24 h to acclimate the cells to the experimental conditions.

Phage were sampled by plating serial dilutions of the community culture into 4 ml of molten (~50°C) soft agar (LB agar except with only 0.8% w/v agar) infused with ~5 x 10^8 cells of the ancestral bacterial strain, REL606. The agar was poured over an LB agar plate, allowed to solidify, and incubated overnight at 37°C. We then picked individual plaques (~1 mm diameter), each the product of a single virus, from suitable dilutions. Phage stocks were grown on REL606 cells in modified M9 following procedures adapted from ref. *53*, then stored with 2% chloroform at 4°C. Aliquots were added directly to the experiments from these refrigerated stocks. Evolved phage stocks tended to decay, therefore they were stored for long term by freezing with glycerol as the bacteria were.

Estimating population densities. The density of *E. coli* cells was estimated by dilution in saline solution (8.5 g/L NaCl) followed by plating on LB agar, with a target count of 150-500 colonies per plate. The density of phage λ was estimated in a similar manner except dilutions were done in modified M9 without glucose and plaques were assayed on soft-agar plates.

Detection of λ that use new receptor. To determine when λ evolved the ability to use a new receptor, we performed "spot assays" (*52*) on lawns of a *lamB*$^-$ mutant derived from the ancestral strain, REL606. For this assay, ~5 x 10^8 *lamB*$^-$ cells were dispersed in soft agar and an undiluted sample of phage was dripped onto the agar. If some of the phage could exploit a receptor other than LamB, then a clear spot in the lawn would be observed after 24 h at 37°C.

Identifying the new receptor. To identify the new receptor used by some evolved λ isolates, we performed spot assays using lawns of *E. coli* mutants defective in the production of various outer-membrane proteins. Each test strain lacked LamB and one of seven proteins – OmpA, OmpC, OmpF, OmpG, OmpW, BglH, and PhoE – that share sequence or structural similarities to LamB. We used double mutants because the evolved phage retained the ability to use LamB. The double mutants were produced starting with seven knockout strains in the Keio Collection (*54*) (Table A4.2), and then introducing a *malT*$^-$ mutation to each strain so that it does not express LamB. (See the section on "Evolution of *malT*$^-$ mutants" for details of how these mutations affect the expression of LamB.) To generate the *malT*$^-$ mutants, we challenged populations (~10^6 cells) of each Keio strain with the ancestral λ (~10^8 particles) on LB plates and isolated colonies of resistant mutants. We confirmed the mutants were *malT*$^-$ by plating on tetrazolium maltose (TMal) agar plates (*55*).

Table A4.2. Set of *E. coli* knockout strains from the Keio Collection (*54*) used to identify the novel receptor used by some evolved phage λ. The CGSC number is the strain identifier used by the Coli Genetic Stock Center at Yale University.

Gene removed	CGSC no.	KEIO name
ompA	8942	JW0940-6
ompC	9781	JW2203-1
ompF	8925	JW0912-1
ompG	11793	JW1312-1
ompW	9125	JW1248-2
bglH	10702	JW3698-5
phoE	8466	JW0231-1

Phage genomics. To sequence λ strain cI26, which was the ancestral phage in our study, we pooled three 4-ml liquid stocks into a single 12-ml sample containing ~10^9 plaque-forming units (pfu) per ml. The same approach was used for evolved phage EvoC, except the final preparation had ~10^7 pfu per ml. Genomic DNA was purified from each sample by using a Qiagen Lambda Midi Kit, fragmented by sonication, prepared as bar-coded libraries, and sequenced on an ABI SOLiD 4 instrument at the University of Texas at Austin's Genome Sequencing and Analysis Facility. The paired 50-base and 35-base reads were mapped in color space to the reference genome (GenBank: NC_001416.1) using SHRiMP v2.1.1b (compbio.cs.toronto.edu/shrimp/). Only the top-scoring alignments of properly mapped read pairs were analyzed. The resulting SAM files were reformatted using a custom Perl script, then entered into the *breseq* pipeline v0.13 to predict consensus base substitutions, small indels, and larger deletions as well as to identify any genetic polymorphisms in the sample. The only site that showed heterogeneity was an A→G change at position 18,538 that was present in ~55% of the reads in the ancestral

sample. All other mutations were predicted to be consensus changes present in essentially all of the sequenced population (Table A4.1).

Bacterial genomics. Bacteria were revived from freezer stocks, grown overnight in LB medium, and genomic DNA was isolated from several ml using Qiagen genomic tips. DNA samples were fragmented by sonication, prepared as bar-coded libraries, and run as six of twelve multiplexed samples spread over four lanes on an Illumina GenomeAnalyzer IIx by the Research Technology Support Facility at Michigan State University. Mutations were predicted from the resulting 75-base paired-end DNA reads using *breseq* v0.13 and the genome sequence of the ancestral *E. coli* strain, REL606 (GenBank: NC_012967.1), as the reference. The *breseq* pipeline performs single-end read alignment to the reference genome with SSAHA2 (www.sanger.ac.uk/resources/software/ssaha2/). In addition to the types of mutations predicted from the phage sequence data by *breseq*, the detection of structural variation from reads with split alignments was enabled for the bacterial samples.

Data and software availability. The λ and *E. coli* genome-sequence data have been deposited in the NCBI Sequence Read Archive (SRA043942). The source code for *breseq* is freely available online (barricklab.org/breseq and code.google.com/p/breseq/).

Targeted sequencing of the J gene. To find mutations in the gene encoding the J protein of the λ tail (host specificity protein, GenBank: NP_040600), we sequenced DNA fragments using an automated ABI sequencer. The fragments were PCR-amplified and purified using a GFX column (GE Healthcare). Primer sequences were 5' CTGCGGGCGGTTTTGTCATT 3' and 5' ACGTATCCTCCCCGGTCATCACT 3', which complement sequences 15 bp upstream and 318 bp downstream of the J gene, respectively.

Null model for non-synonymous mutations. The sequence of the J gene was obtained from the reference λ genome (GenBank: NC_001416.1). All possible base substitutions and their effects on the encoded protein were enumerated using a custom Perl script to calculate the ratio of non-synonymous to synonymous mutations among all base substitutions, assuming equal rates.

Replay experiments. To examine whether specific steps along the evolutionary paths taken by the phage and bacteria influenced the likelihood that λ would evolve the ability to use OmpF, we replayed evolution by assembling communities with particular combinations of phage and bacteria that could reveal historical contingencies affecting that outcome. The *E. coli* and λ used in the replays were isolated from different populations at different time points, as described in the main text. The replays were run in the same manner as the other coevolution experiments, except using different strains. Each replay community was propagated for up to 10 days, and each was sampled daily to determine whether λ had evolved to target OmpF by plating 2-5 μl on lawns of the ancestral bacterial strain, the $lamB^-$ mutant of the ancestor, and the $ompF^-$ $malT^-$ derivative of BW25113. The replays were stopped early if the phage either acquired the ability to exploit OmpF or went extinct.

In the final set of replay experiments, we used diverse bacterial communities rather than clones. This approach required special procedures to include a representative sample of bacteria while excluding phage. For each community of interest we plated ~300 cells on LB agar, picked each colony with a sterile toothpick, and suspended them together in LB broth. We then grew the mixed culture overnight at 37°C with shaking at 120 rpm, and again plated ~300 cells. We repeated this process three times to eliminate phage from the mixed culture. To confirm the absence of phage, we took an aliquot of each mixture, added chloroform to kill the bacteria, let

the chloroform settle, and added 1 ml to a lawn of the λ-sensitive ancestor, REL606, in soft agar. No plaques were formed, confirming that this process had eliminated the phage. Finally, we stored 1 ml of each mixed bacterial culture by adding 15% glycerol and freezing it at –80°C. To start the replay experiments, we took 100 µl of the thawed mixture, let it grow overnight in LB, transferred 100 µl to modified M9, and let this culture grow overnight. We then used 100 µl of this culture to initiate each replay community. We expect that this technique was effective at isolating and propagating abundant bacterial genotypes from the source communities, although their frequencies may have shifted and most rare variants would be excluded. These effects might explain why the replay experiments, while highly reproducible, sometimes differed from the corresponding source communities in the initial experiment.

Supporting Information

Evolution of *malT*$^-$ mutants. Preliminary experiments showed that, in minimal glucose medium, *E. coli* strain REL606 generally evolved λ-resistance through *malT*$^-$ mutations. MalT is a positive regulator of *lamB*, which encodes the receptor LamB, so that mutations that disrupt MalT function prevent LamB expression (*56*). MalT also activates other genes required for growth on maltose and other maltodextrins (*56*), and these *malT*$^-$ mutations are therefore defective in growth on those substrates. However, these mutations are advantageous, even in the absence of phage, in glucose medium for the *E. coli* strain used in our study (*57, 58*), probably because they reduce the basal expression of unnecessary gene products. This additional benefit may explain why *malT*$^-$ mutations evolved in the coevolution experiments, rather than mutations in the *lamB*-encoded receptor that would not have yielded the metabolic cost-savings.

In any case, we tracked the evolution of *malT*$^-$ mutants in all 96 populations in the large-scale experiment to determine how often and how quickly these genotypes fixed. We plated a random sample of bacteria (50-100 cells) on TMal plates on days 5 and 8 of the experiment; *malT*$^-$ mutants produce red colonies on these plates; the *malT*$^+$ ancestor produces white colonies. Figure S1 shows the frequency of *malT*$^-$ cells in all 96 populations.

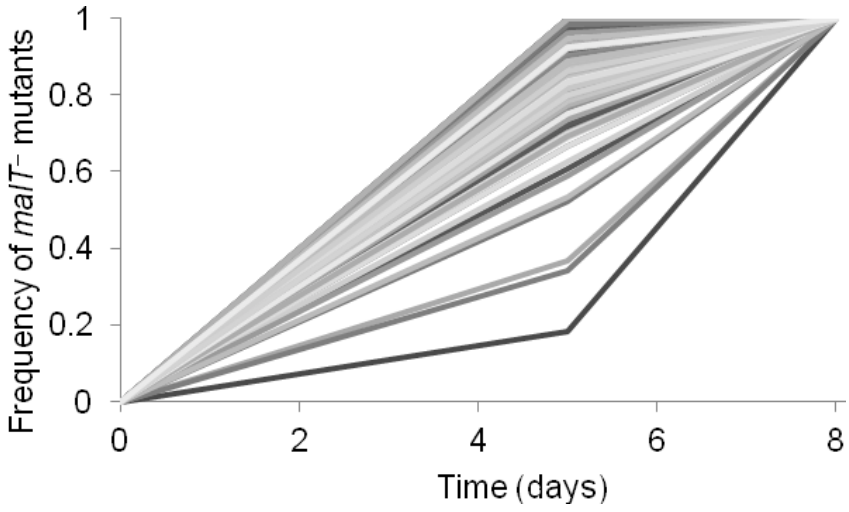


Figure A4.1. Rapid fixation of *malT*$^-$ mutants in the 96 populations of the large-scale experiment.

Population dynamics. We quantified the dynamics of the coevolving bacteria and phage in the initial evolution experiment to better understand the conditions under which the phage evolved the novel receptor function. Lytic phages can, in principle, exert top-down limitations on the density of bacteria (*59*). In our experiments, however, any such limitation was quickly overcome as the bacteria evolved high levels of resistance (Figure A4.1). As a consequence, the phage density was low compared with that of the bacteria (Figure A4.2). This difference meant that

any phage mutant that overcame the resistance would gain access to a large host population. Indeed, one phage population in this experiment evolved to use the OmpF receptor and transiently achieved a higher density (Figure A4.2: open triangles), but its density declined after the bacteria evolved further resistance in addition to the early *malT*$^{-}$ mutation.

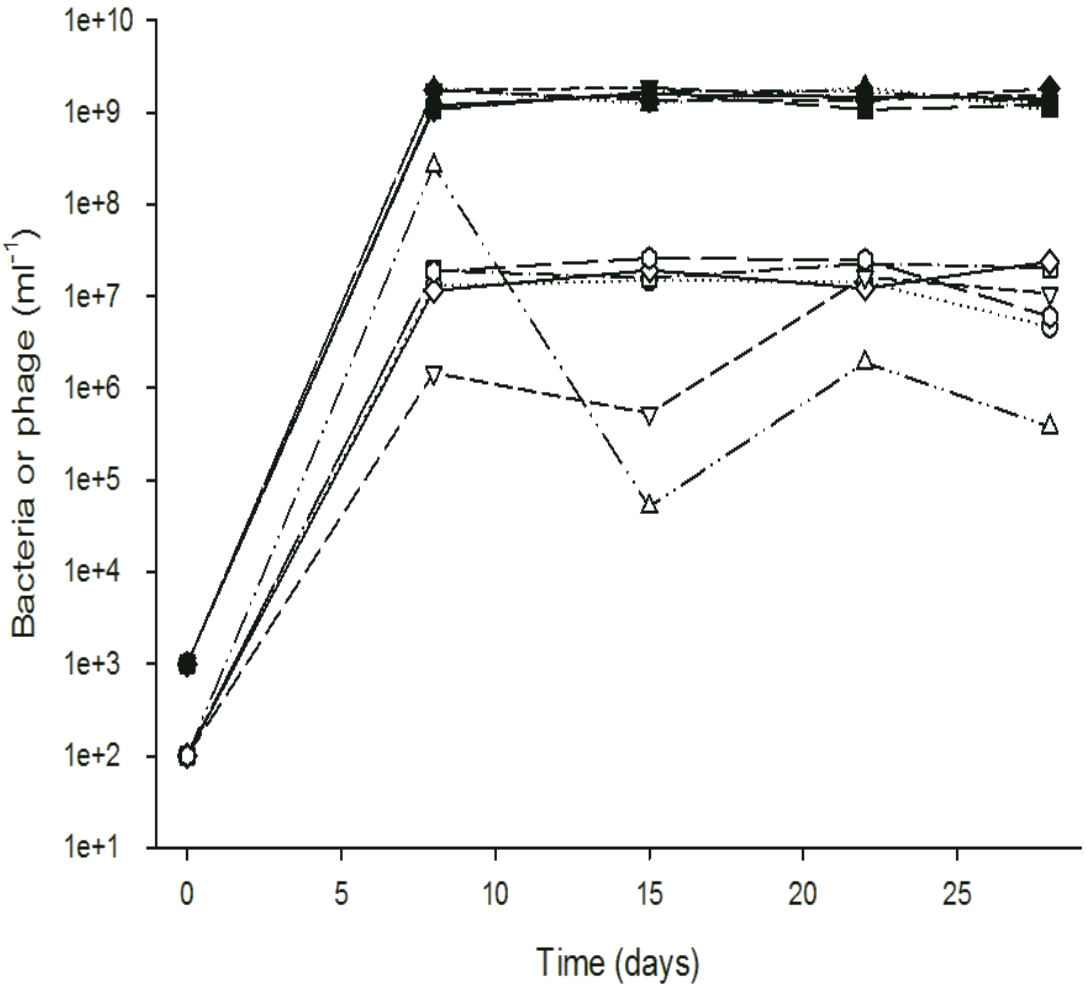


Figure A4.2. Population dynamics of *E. coli* (filled symbols) and phage λ (open symbols) from the six replicates of the initial evolution experiment. Except for day 0, bacteria and phage densities were based on colony and plaque counts, respectively, with the latter obtained by using lawns of the sensitive ancestral host. Day 0 values were calculated from corresponding densities in stock cultures multiplied by known dilution factors.

Mechanism of phage persistence following initial resistance. Phage λ persisted after the rise of *malT*$^{-}$ mutants (Figs. S1, S2), even though the mutants appeared to be completely resistant when the phage were spot-tested on bacterial lawns. We hypothesized that the ancestral λ could

infect rare *malT*$^{-}$ cells that spontaneously expressed LamB. This hypothesis is consistent with a study showing that *lamB* regulatory mutants were occasionally infected by wild-type λ (*60*). An alternative explanation is that *malT*$^{-}$ mutants may not have completely fixed in the population if the *malT*$^{+}$ ancestors had a growth-rate advantage that allowed them to maintain a small minority of sensitive cells that λ could exploit. This mechanism has been demonstrated in several studies of coevolving bacteria and phage (*59, 61, 62*). However, this explanation seemed unlikely in the present case because, as already noted, *malT*$^{-}$ mutants have a competitive advantage in glucose-limited media in the absence of phage (*57, 58*). A third possibility is that the ancestral λ could infect cells through some other receptor at a very low rate that would not allow plaque formation on lawns of the *malT*$^{-}$ mutants.

To discriminate among these hypotheses, we isolated a *malT*$^{-}$ mutant of REL606 that had a 25-bp duplication causing a frameshift in this gene. We propagated six communities of the ancestral λ with this bacterial mutant for seven days; we simultaneously ran six replicates with the same phage and the *lamB*$^{-}$ bacterial mutant. The three hypotheses make distinct predictions about whether λ can persist in these two treatments. Under the first hypothesis, in which spontaneous inductions generate a physiological minority of susceptible cells, λ should be maintained on the *malT*$^{-}$ mutant, but not on the *lamB*$^{-}$ mutant. Under the second hypothesis, which requires a subpopulation of genetically sensitive cells, λ should go extinct in both treatments. Under the third hypothesis, according to which even the ancestral phage can use an alternative receptor, λ should persist in both treatments. Figure A4.3 shows that the results

clearly support the first hypothesis, in which spontaneous induction and expression of LamB allow the ancestral phage to persist only on the *malT*$^{-}$ mutant.

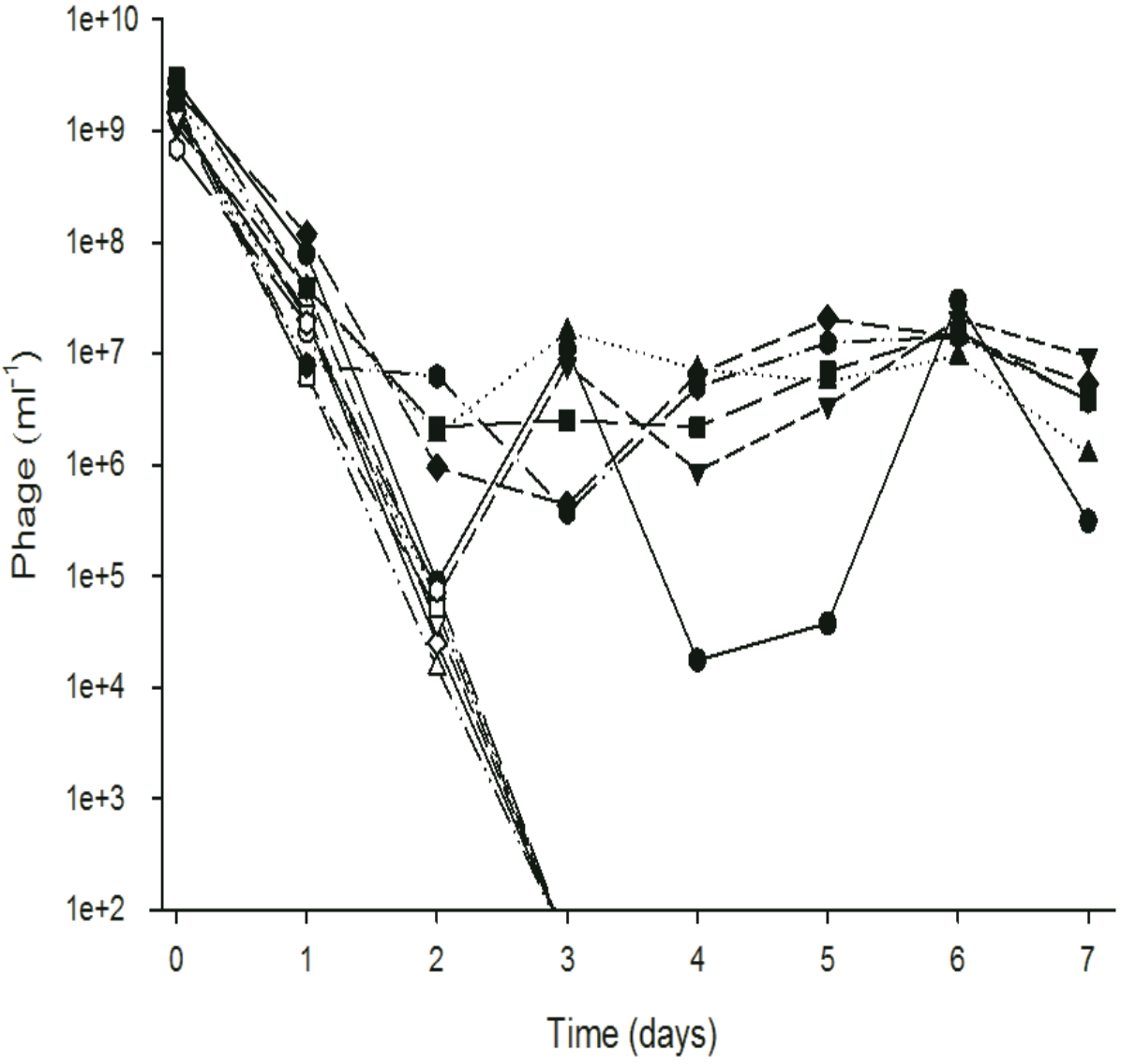


Figure A4.3. The ancestral λ strain persists on the *malT*$^{-}$ (closed symbols) but not the *lamB*$^{-}$ (open symbols) mutant host population. Phage densities were obtained from plaque counts on lawns of the sensitive ancestral bacteria. No phage were seen after day 2 on the *lamB*$^{-}$ hosts; the limit of detection was 10^2 phage per ml.

Properties of the new λ receptor. There are two hypotheses for what receptor properties are most important for phage binding: hydrogen-bond formation between specific amino acids on the ligand and receptor, or electrostatic interactions facilitated by complementary shape motifs between the two structures (*63*). Consistent with the latter hypothesis, λ evolved repeatedly to use OmpF, which has a similar structure to LamB (*64*), but a very different amino-acid sequence from LamB (Table A4.3). By contrast, the phage never targeted BglH, despite its more similar amino-acid sequence (Table A4.3), although the structural similarity of BglH to LamB is unknown. It may also be relevant that OmpF is more highly expressed than BglH under conditions similar to our experiments (*65*; see transcriptomic data at myxo.css.msu.edu/ecoli/arrays/arrays.txt).

To find proteins with similar amino-acid sequences to LamB, we performed a BLAST protein search (*66*) of LamB (Genbank accession: YP_003047080) against the genome of the ancestral strain, REL606 (Genbank accession: NC_012967). Many putative matches were found, although OmpF was not one of the top matches (Table A4.3). To compare the similarity of OmpF and LamB in light of these other proteins, we performed BLAST protein alignments. A few small sections of OmpF matched LamB, but only under the most relaxed settings, and OmpF was not the most similar outer-membrane protein to LamB (Table A4.3).

Table A4.3. Results from BLAST alignment of LamB to all proteins in the ancestral genome. The five top-ranked matches are listed and compared to OmpF. Default blastp parameters were used.

Protein	Cellular location	No. identical amino acids	Region of putative homology
yieC carbohydrate-specific porin (BglH)	outer membrane	118	462
Glycerate kinase II	cytoplasm	25	108
Cytidine deaminase	cytoplasm	12	23
Potassium proton antiporter	membrane	11	26
Glycerophosphodiester phosphodiesterase	periplasm	11	27
Porin protein OmpF	outer membrane	7	14

Time required for λ to target OmpF. We sampled the 96 communities in the large experiment daily to determine whether the phage had evolved to target a new receptor and, if so, whether that receptor was OmpF. We spotted 2-5 µl of each culture on lawns of the ancestral bacteria, the *lamB*$^-$ mutant of the ancestor, and the *ompF*$^-$ *malT*$^-$ derivative of BW25113. Regions of lysis on the first two lawns, but not on the third, indicated that evolved phage could infect cells using OmpF. This capacity evolved in 24 of the 96 replicates. If the third lawn also showed lysis, that would imply phage could infect cells using some other receptor than LamB or OmpF; however, that outcome was never seen. Figure A4.4 shows the timing of the evolution of the ability to use OmpF in the 24 populations that achieved this innovation. No population evolved the trait early in the experiment, in agreement with the finding that λ requires four mutations to use OmpF. The number of populations evolving the new function also appears to have declined toward the end of the experiment, even though 72 populations still had not evolved that function.

This latter observation is consistent with the finding that some bacteria evolved resistance mutations that rendered the phages unable to evolve the new function.

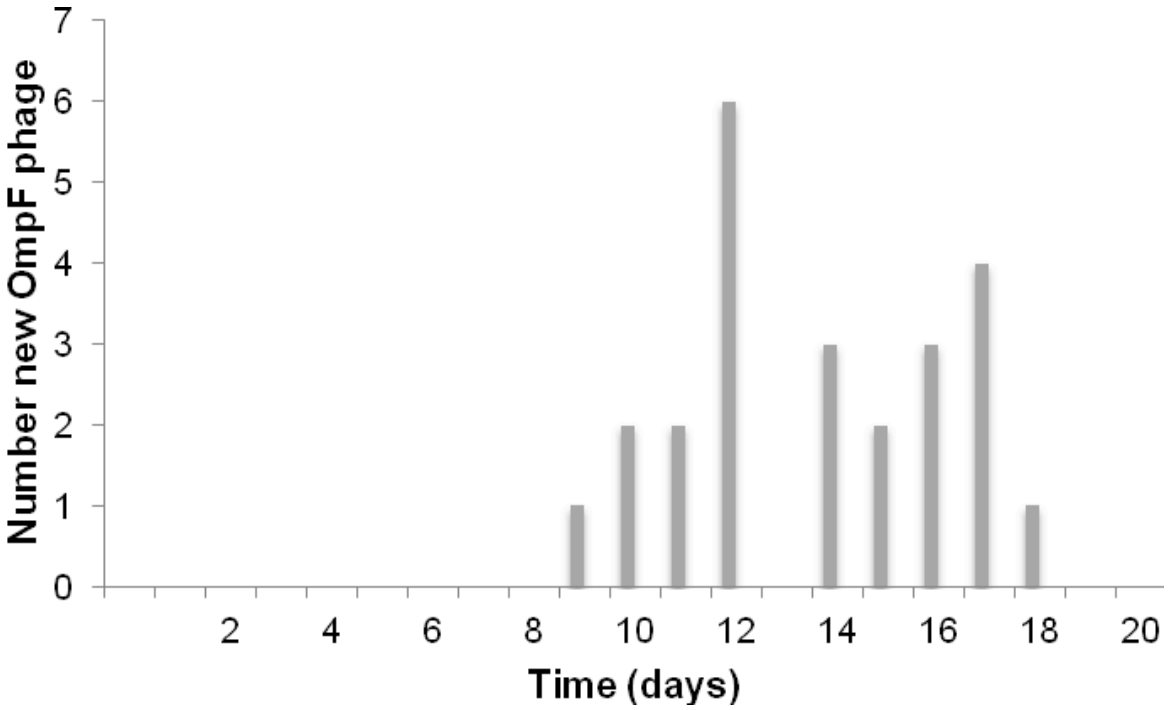


Figure A4.4. Distribution of times at which 24 λ populations in the large-scale evolution experiment first showed the capacity to infect *E. coli* through the OmpF receptor.

More mutations in phage that targeted OmpF. Table A4.4 shows the number of mutations in the J gene of phage isolates from the large-scale experiment. The left half of the table shows all 24 isolates that evolved the ability to use OmpF; these isolates were sampled on the first day this function was observed in the source population. The right half includes 24 isolates from other populations that did not evolve this function; each of these isolates was sampled on the same day as one of the isolates in the first group. Therefore, the rate of evolution for the two groups can be compared statistically by a paired test, with the time available for mutations to have accumulated being the same for the two members of each pair. The test results are presented in the main text.

Table A4.4. Mutations in the gene encoding the J protein in two groups of evolved λ phage, one that acquired the ability to exploit OmpF and the other that remained dependent on LamB. Table extends for two pages.

λ that can use OmpF			λ that use LamB only		
ID	Day isolated	No. of mutations	ID	Day isolated	No. of mutations
A7	14	7	H2	14	3
A8	11	7	A10	11	2
A12	14	10	A11	14	4
B2	17	5	B3	17	3
C2	12	7	C1	12	2
C3	9	5	C4	9	3
D3	16	7	D1	16	3
D4	10	6	D2	10	3
D6	16	7	D5	16	3
D7	12	4	D8	12	5
D9	12	6	D10	12	3
E3	12	7	E2	12	3
E4	17	8	E6	17	4
E11	15	7	G10	15	2
E12	18	7	F2	18	5
F5	16	7	F4	16	2
F7	12	6	F6	12	3
F8	14	8	F9	14	4
G4	11	5	G3	11	4
G9	15	7	H10	15	2
H5	17	7	H4	17	7
H8	17	7	H7	17	4

Table A4.4 (cont'd)

H9	10	6	H3	10	4
H12	12	6	H11	12	4

Two additional tests of receptor profile. Spot tests indicated that λ required four mutations in J to be able to exploit OmpF as a receptor. However, this method might not be sensitive enough to detect phage that can use that receptor but with very low efficiency. We therefore performed two additional tests – one based on phage adsorption and the other on phage growth – to verify that all four mutations are required for λ to exploit, even slightly, the OmpF receptor. For each test, we examined four informative phage including three genotypes (EvoA, F2, and H4) at the precipice of evolving the new function (each has 3 of the 4 canonical mutations) and one (D7) with all 4 canonical mutations and no others.

Adsorption assays. We measured the adsorption rates of the four evolved phage using the *lamB*$^-$ mutant of the ancestral host strain. This assay measures the rate at which phage adsorb to and infect cells by tracking how many phage remain free (unattached) in the medium over time (*67*). We added ~5 x 10^4 phage and ~2 x 10^9 exponentially growing bacteria to 10 ml of modified M9. We measured the density of free phage at six time points over ~25 minutes. We then fit a linear regression to the log ratio of free phage density at time t, $p(t)$, to their initial density, $p(0)$, i.e., $\ln(p(t)/p(0)) = b \times t$, where the slope, b, reflects the rate at which the phage adsorb. The intercept of the regression was constrained to 0 because all phage are unbound at $t = 0$ and densities are expressed relative to the initial value. With samples taken at 6 time points for each experiment, and with the intercept fixed, each regression has 4 degrees of freedom. A significant negative slope indicates that the phage can adsorb to some receptor other than LamB. Only the

D7 phage, which has all 4 canonical mutations, showed a significant decline indicative of its ability to use the OmpF receptor (D7: $p = 0.046$, $b = –0.128$; EvoA: $p = 0.970$, $b = 0.008$; F2: $p = 0.605$, $b = 0.004$; H4: $p = 0.999$, $b = –0.024$) (Figure A4.5). The adsorption-rate constant for D7 is estimated to be ~6 x 10^{-10} per ml per minute, where that rate is calculated as $–b/N$ and N is the bacterial density. This rate is similar to a previous estimate for wild-type λ using host cells that express LamB (*68*), which implies that the evolved phage D7 adsorbs quite well to OmpF.

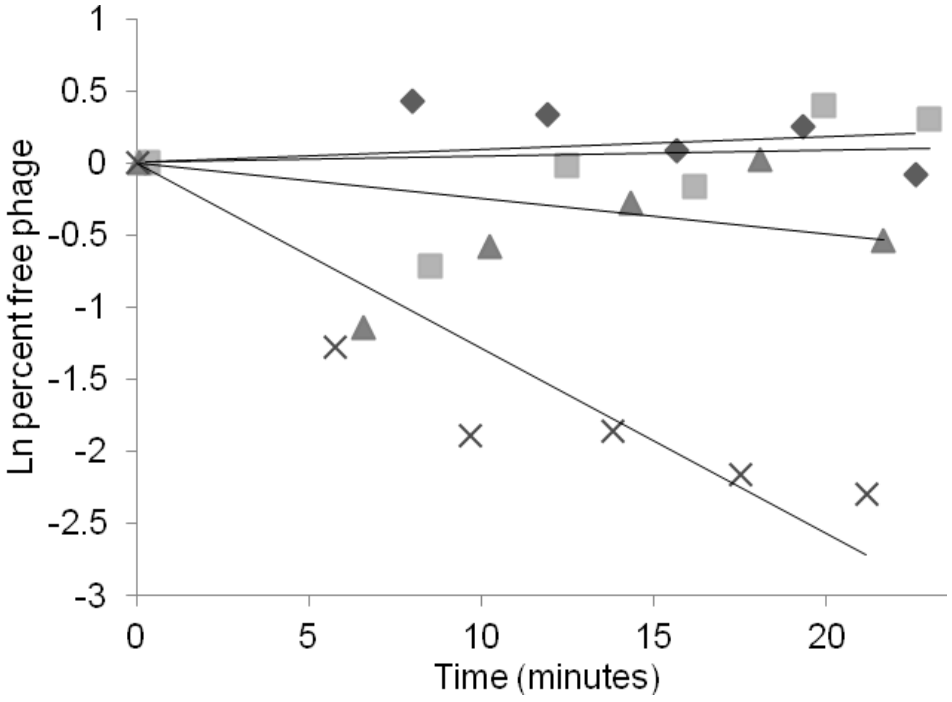


Figure A4.5. Adsorption assays using four evolved λ genotypes and a *lamB*⁻ bacterial mutant. The concentration of free phage should decline only if the phage can adsorb to cells using a receptor other than LamB. EvoA (diamonds), F2 (squares), and H4 (triangles) all have only 3 of the 4 canonical mutations needed to target OmpF, whereas D7 (crosses) has all 4 mutations.

Growth assays. We performed growth experiments to determine which of four evolved phage λ genotypes (EvoA, F2, H4, and D7) could reproduce on the *lamB*⁻ bacterial mutant. We included the ancestral phage as a negative control, and we also measured phage growth on isogenic *lamB*⁺

bacteria as a positive control. Each combination of phage and bacteria was replicated three-fold. We mixed the phage and bacteria in small volumes (1.2 ml) of modified M9 in glass tubes. We added ~8 x 10^5 exponentially growing cells to each tube; the initial phage numbers were kept low at ~250 particles (even fewer for F2 owing to its low-density stock) to limit the possibility that mutants derived from the genotypes with three canonical mutations might acquire the final mutation. The cultures were incubated at 37°C and shaken at 160 rpm for 24 h. Phage densities were assessed at the beginning and after 24 h by plaque assays on lawns of the ancestral bacteria. As expected, the ancestral phage and all four evolved types showed robust growth on the *lamB*$^+$ bacteria (Figure A4.6, top panel). Phage D7 also grew very well on the *lamB*$^-$ bacteria, but none of the other phage could reproduce at all on the mutant cells (Figure A4.6, bottom panel). These results confirm that all 4 of the canonical mutations are required for the evolved phage to use OmpF as an alternative receptor to LamB.

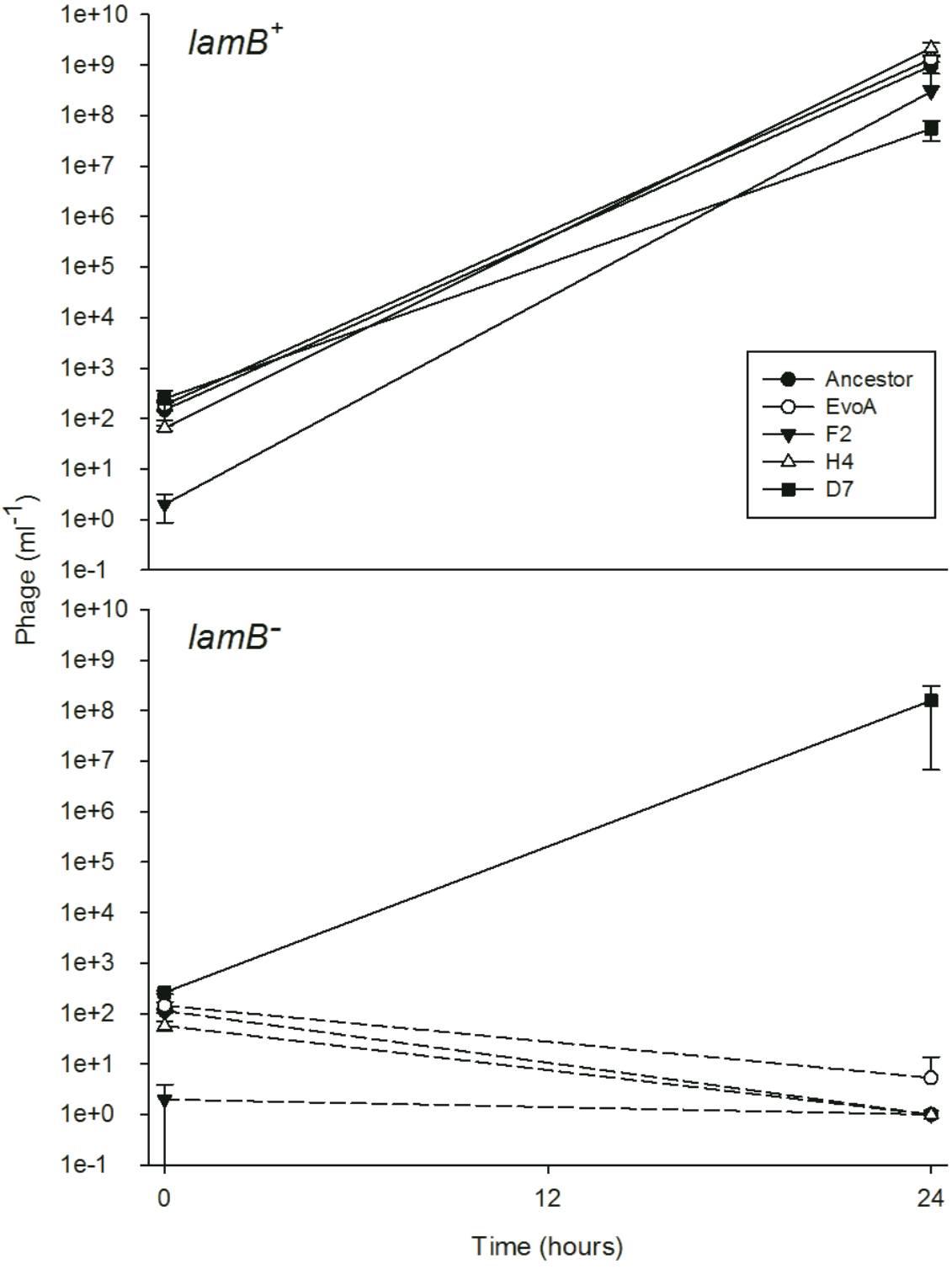


Figure A4.6 (caption on following page)

Figure A4.6. Population growth of five λ genotypes on two bacterial hosts. EvoA, F2, and H4 have three of the four canonical mutations required to use OmpF as a receptor, while D7 has all four mutations. Dashed lines in the bottom panel indicate that, when mixed with *lamB*$^{-}$ bacteria, the ancestral phage and evolved types EvoA, F2, and H4 dropped below the limit of detection (~3 pfu ml^{-1}) after 24 h, except for one replicate of EvoA that yielded some plaques. Error bars show 95% confidence intervals.

Parallel evolution in the J gene. Parallel evolution provides a strong signal of natural selection. Many studies have documented parallel changes in phenotypes (*57, 58, 69, 70*), and others have reported parallel evolution at the level of evolving genes (*48, 65, 71, 72*). Parallel changes at the level of nucleotide sequences are much less common, although a previous study with a different phage reported extensive parallelism at the nucleotide level (*74*). Figure 5.3 (main text) shows many parallel mutations in the gene encoding the J protein across independently evolved λ lineages. To determine if this parallelism was statistically significant, we compared the observed average number of mutations shared by pairs of evolved phage with the random expectation (Figure A4.7). We performed the analysis on two separate groups, the phage that evolved to exploit OmpF and those that did not. To generate the null-hypothetical distribution, we constructed 10^5 random matrices (24 by 40 cells, identical in size and shape to the top or bottom half of Figure 5.3 in the main text). We generated each random matrix by shuffling the cells while preserving the number of mutations in each row. This approach is highly conservative because it considers only those sites that differed from the ancestor in at least one sequenced allele, and thus it implicitly ignores all sites that did not vary. We then computed the average number of shared mutations for the actual matrix and for each of the randomized matrices (Figure A4.7). Among the 24 alleles from phage that evolved the capacity to use OmpF, all pairs shared at least two mutations and, on average, the pairs shared 4.07 mutations (Figure A4.7A). However, when the cells were randomized, the average pair shared only 1.07 (± 0.05 standard

deviation) mutations (Figure A4.7B). None of the randomized matrices showed parallelism close to the observed level; hence, the signal is highly significant ($p << 10^{-5}$). The evolved genotypes that continued to require LamB shared many fewer mutations; the average pair had 0.58 mutations in common (Figure A4.7C). Nonetheless, this value was higher than any random matrix (Figure A4.7D), again indicating highly significant parallelism ($p < 10^{-5}$).

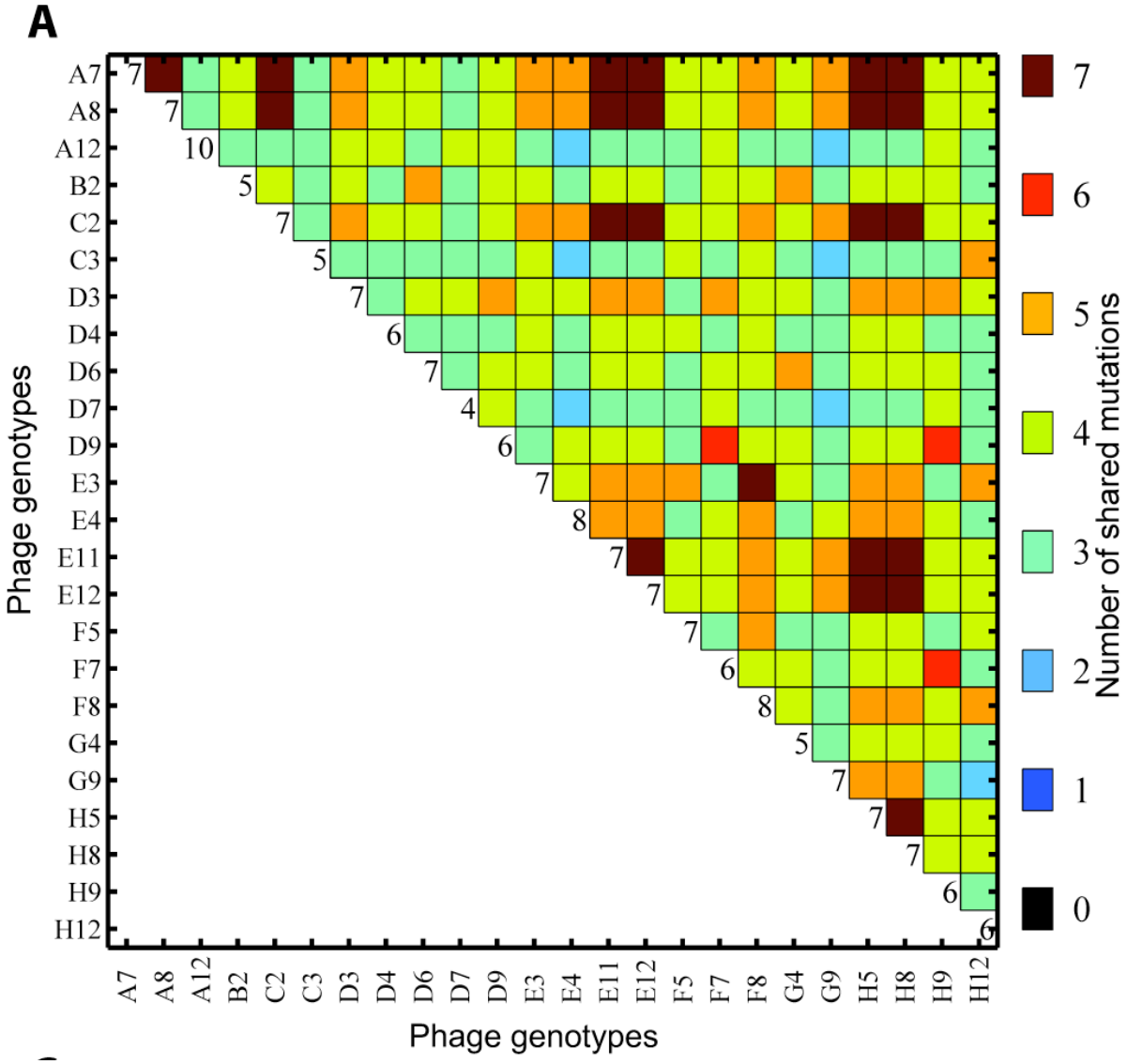


Figure A4.7 is continued from the previous page.

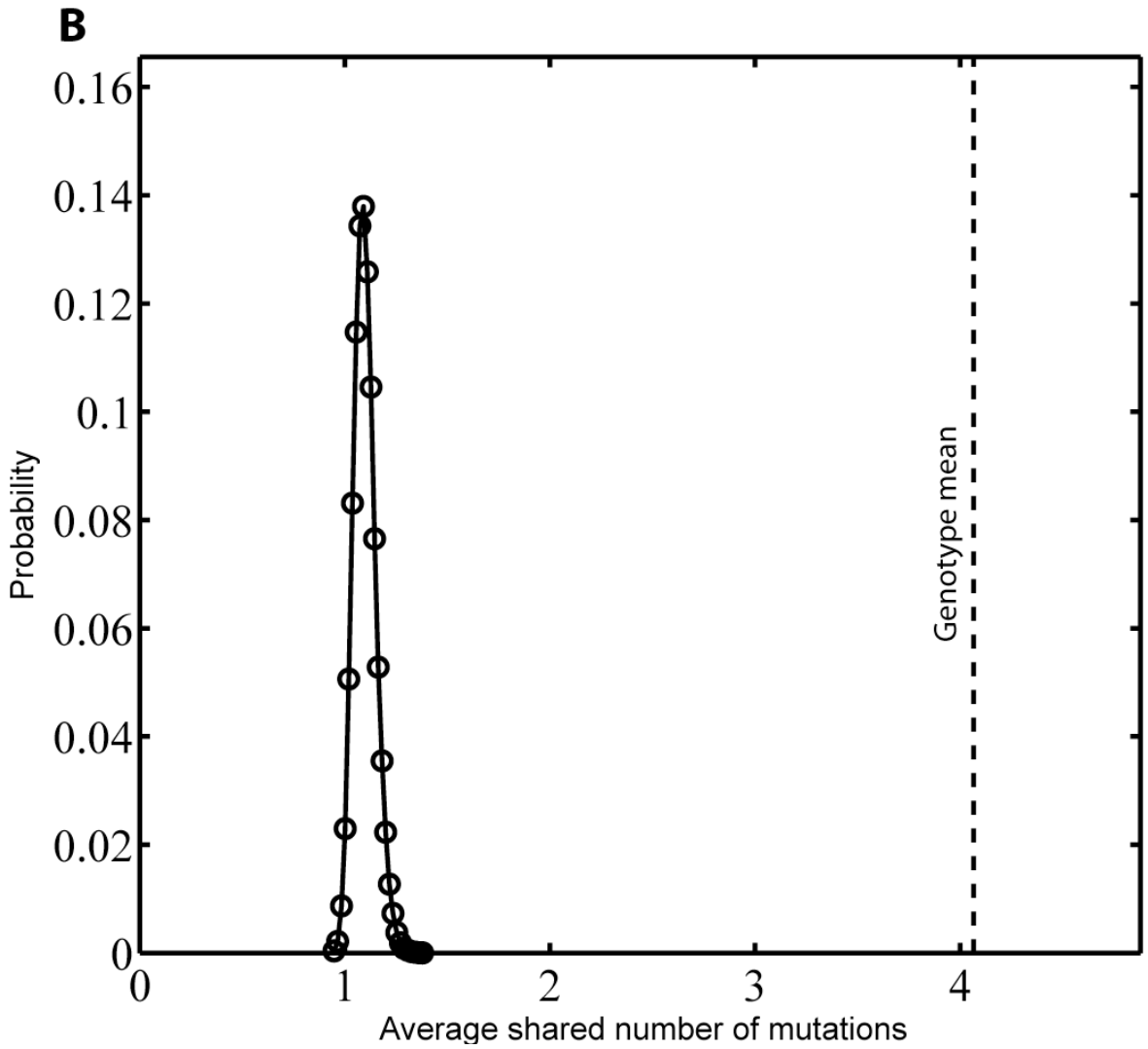


Figure A4.7 is continued from the previous page.

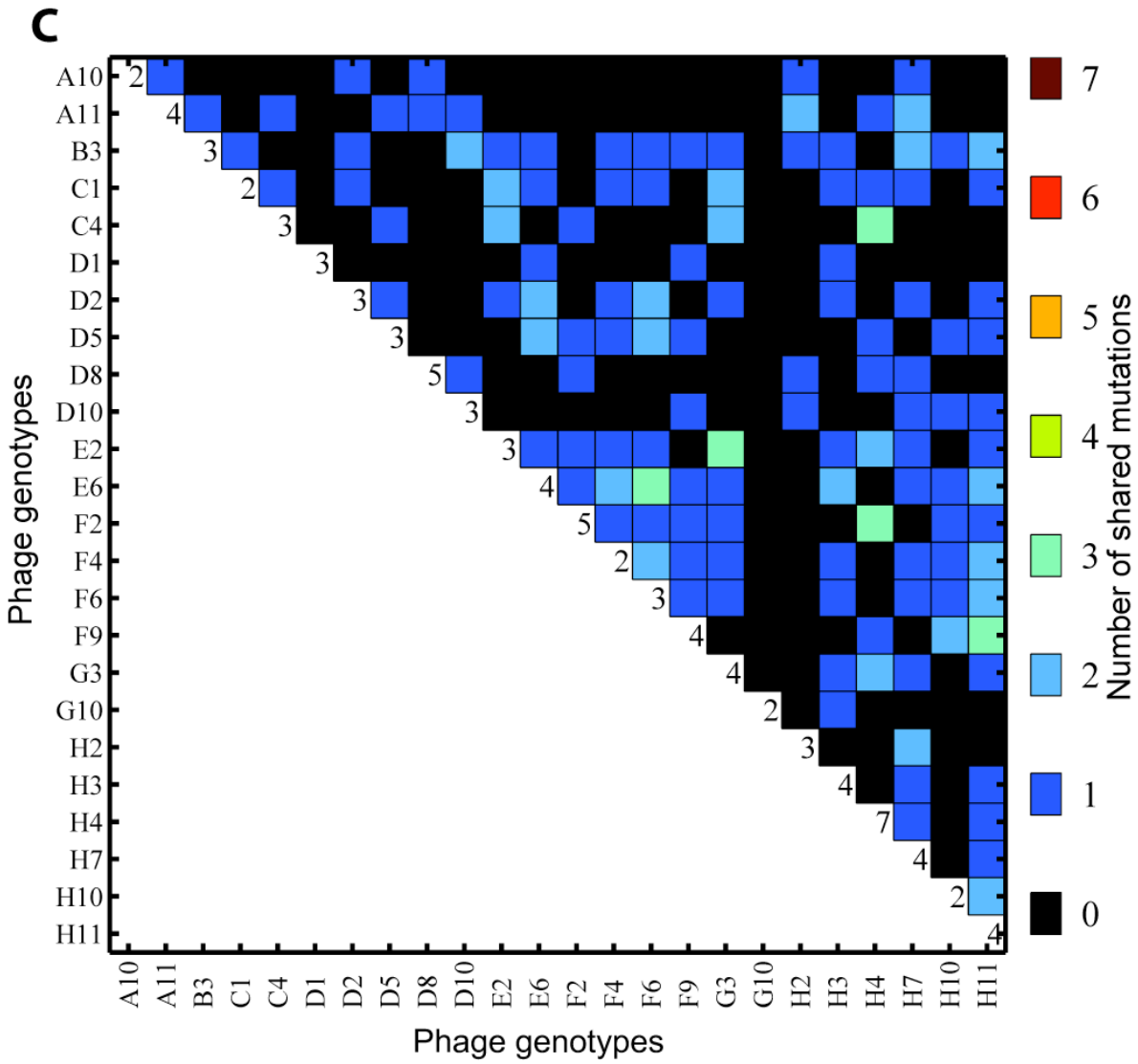


Figure A4.7 is continued from the previous page.

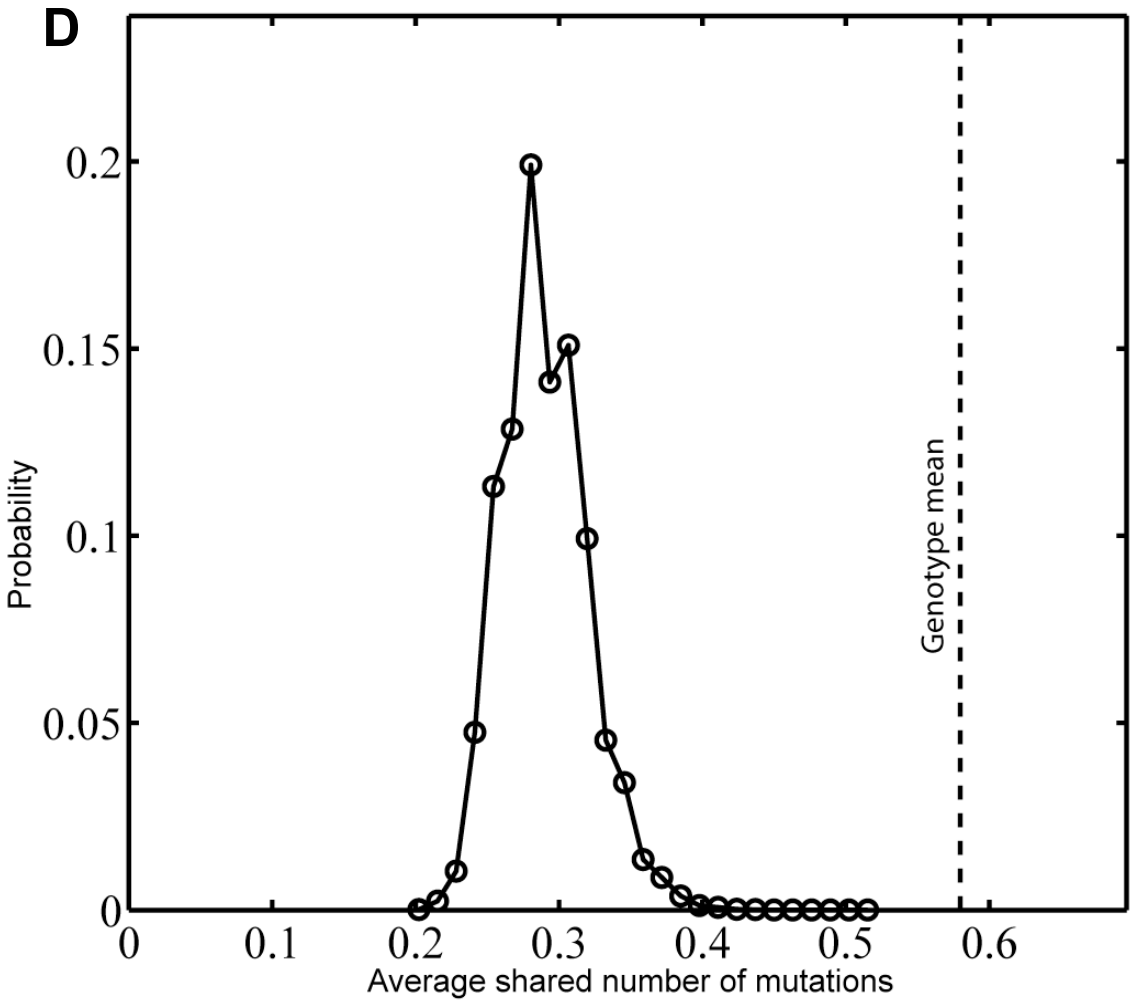


Figure A4.7. Parallel evolution in J protein. (A) Pairwise comparisons among 24 λ genotypes that independently evolved the ability to target OmpF, showing the number of shared mutations for each pair. The values along the diagonal show the number of mutations for each genotype. (B) Probability distribution for the average number of shared mutations based on 10^5 randomized similarity matrices; the vertical dashed line shows the observed average. (C & D) Same as (A & B), except showing the observed data and randomized distribution for 24 genotypes that retained their dependence on the LamB receptor.

Mutations in bacterial genomes and their effects on phage evolution. We sequenced the complete genomes of all six evolved bacterial clones used in the second replay experiment, and we compared them to the ancestral genome, as described in the Materials and Methods. We observed 15 mutations in total (Table A4.5). As explained in the main text, the mutations in *manY* and *manZ* uniquely distinguish the bacteria that blocked the evolution of phage able to use OmpF from those bacteria that allowed the phage to evolve that new function.

Table A5.5. Mutations and their phenotypic effects in six *E. coli* clones that evolved with phage λ. The first three clones prevented phage from evolving the ability to use OmpF as a receptor, while the last three clones allowed phage to evolve that novel trait.

Clone	Genome location*	Mutation	Genes affected	Effect on proteins**	Phenotypic effects***
EcA8	1,882,610	5-bp duplication	*manZ*	frameshift at AA 49	Man^-, λ^{all}-r
	3,482,737	G→A	*malT*	stop at AA 351	Mal^-, λ^{LamB}-r
	3,894,997	4,048-bp IS*150*-mediated deletion	*rbsD–rbsB*	partial deletion of ribose operon	Rbs^-
EcC3	1,881,820	16-bp duplication	*manY*	frameshift at AA 59	Man^-, λ^{all}-r
	3,482,677	25-bp duplication	*malT*	frameshift at AA 339	Mal^-, λ^{LamB}-r
	3,894,997	1,278-bp IS*150*-mediated deletion	*rbsD–(rbsA)*	partial deletion of ribose operon	Rbs^-
EcF6	1,881,721	G→T	*manY*	stop at AA 21	Man^-, λ^{all}-r
	3,482,677	25-bp duplication	*malT*	frameshift at AA 339	Mal^-, λ^{LamB}-r
	3,894,997	4,631-bp IS*150*-mediated deletion	*rbsD–(rbsK)*	partial deletion of ribose operon	Rbs^-
EcC4	3,482,567	C→T	*malT*	stop at AA 295	Mal^-, λ^{LamB}-r
EcD4	1,003,919	G→T	*ompF*	N→K at AA 52	Probably affects λ adsorption to OmpF
	3,482,677	25-bp duplication	*malT*	frameshift at AA 339	Mal^-, λ^{LamB}-r
	3,894,997	395-bp IS*150*-mediated deletion	*(rbsD)*	partial deletion of ribose operon	Rbs^-
EcH2	3,483,588	T→G	*malT*	L→R at AA 635	Mal^--, λ^{LamB}-r
	3,894,997	7,868-bp IS*150*-mediated deletion	*rbsD–(yieP)*	deletion of ribose operon	Rbs^-

* For deletions, location indicates the first base pair (bp) deleted. For duplications, location indicates the first bp of the duplicated region. For insertions, location indicates the last bp before the inserted bases.

** For insertions and duplications, the effect is reported as a frameshift at the first affected amino acid (AA), indicated by its codon number.

*** Phenotypes include Mal^- (unable to use maltose), Man^- (unable to use mannose), Rbs^- (unable to use ribose), λ^{LamB}-r (resistant to λ using LamB receptor), and λ^{all}-r (resistant to λ using LamB and OmpF receptors).

Genetic polymorphism for mannose utlization. We used tetrazolium mannose (TMan) agar plates to score Man^+ and Man^- cells in samples taken on day 20 from the 96 populations in the large-scale experiment. Man^+ and Man^- cells produce white and red colonies, respectively, on TMan plates. The ancestral strain is Man^+. Man^- cells that have been sequenced (Table A4.5) have mutations in the *manXYZ* operon that confer resistance to all λ phage, including those that evolved the ability to use the OmpF receptor. The vast majority of populations were genetically polymorphic for mannose use (Table A4.6) and, by extension, for *manXYZ*-mediated resistance to λ phage.

Table A4.6. Frequencies of Man$^-$ mutants in 96 bacterial populations on the last day of the large-scale experiment, with the community ID, number (n) of cells scored, and frequency of Man$^-$ cells shown for each population. Table extends for two pages.

ID	Number (n)	Frequency of *man*$^-$	ID	Number (n)	Frequency of *man*$^-$	ID	Number (n)	Frequency of *man*$^-$
A1	181	0.03	C9	96	0.73	F5	45	0.00
A2	40	0.25	C10	34	0.94	F6	84	0.10
A3	29	0.10	C11	77	0.21	F7	75	0.01
A4	37	0.78	C12	22	0.95	F8	136	0.16
A5	57	0.26	D1	220	0.71	F9	36	0.83
A6	38	0.47	D2	60	0.25	F10	37	0.30
A7	126	0.02	D3	123	0.07	F11	87	0.17
A8	54	0.94	D4	30	0.00	F12	107	0.90
A9	32	0.00	D5	61	0.26	G1	64	0.27
A10	71	0.15	D6	48	0.00	G2	95	0.48
A11	25	0.04	D7	210	0.95	G3	85	0.16
A12	8	0.00	D8	46	0.09	G4	85	0.33
B1	29	0.28	D9	152	0.98	G5	95	0.15
B2	44	0.34	D10	28	0.61	G6	65	0.26
B3	68	0.03	D11	148	0.17	G7	116	0.30
B4	22	0.00	D12	43	0.93	G8	64	0.16
B5	132	0.31	E1	59	0.88	G9	74	0.18
B6	31	0.00	E2	78	0.27	G10	77	0.48
B7	115	0.12	E3	35	1.00	G11	110	0.14
B8	110	0.95	E4	32	0.00	G12	79	0.49
B9	32	1.00	E5	29	0.00	H1	129	0.33
B10	70	0.16	E6	48	1.00	H2	68	0.07
B11	41	0.02	E7	34	0.74	H3	80	0.98

Table 4.6 (cont'd)

B12	76	0.00	E8	51	0.88	H4	142	0.10
C1	77	0.00	E9	69	0.35	H5	131	0.48
C2	157	0.07	E10	64	0.00	H6	98	0.29
C3	30	0.00	E11	120	0.02	H7	75	0.00
C4	56	0.09	E12	38	0.92	H8	69	0.88
C5	101	0.12	F1	95	0.01	H9	126	0.10
C6	100	0.14	F2	79	0.06	H10	142	0.02
C7	118	0.13	F3	45	0.00	H11	38	0.00
C8	138	0.07	F4	89	0.46	H12	30	0.80

LITERATURE CITED

LITERATURE CITED

1. Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University Press, Cambridge.

2. Gavrilets, S. 2010. High-dimensional fitness landscapes and the origins of biodiversity.in M. Pigliucci and G. Muller, editors. Toward an Extended Evolutionary Synthesis. MIT Press, Cambridge, MA.

3. Darwin, C. 1890. The origin of species by means of natural selection : or, the preservation of favored races in the struggle for life. Reprinted from the 6th London ed. with the additions and corrections. edition. Hurst, New York :.

4. Fisher, R. A. 1930. The genetical theory of natural selection. The Clarendon Press, Oxford.

5. Dieckmann, U. 2004. Adaptive speciation. Cambridge University Press, Cambridge, UK.

6. Thompson J. N. 2005. *The Geographic Mosaic of Coevolution* (Univ. Chicago Press, Chicago).

7. Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:0097-0159.

8. Pigliucci, M. 2008. What, if anything, is an evolutionary novelty? Philosophy of Science 75:887-898.

9. Dennehy, J. J. 2009. Bacteriophages as model organisms for virus emergence research. Trends in Microbiology 17:450-457.

10. Elena, S. F. and R. E. Lenski. 2003. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nature Reviews Genetics 4:457-469.

11. Fisher, R. A. 1930. The genetical theory of natural selection. The Clarendon Press, Oxford.

12. Torsvik, V., L. Ovreas, and T. F. Thingstad. 2002. Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science 296:1064-1066.

13. Duffy, S., C. L. Burch, and P. E. Turner. 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. Evolution 61:2614-2622.

14. Pepin, K. M., S. Lass, J. R. C. Pulliam, A. F. Read, and J. O. Lloyd-Smith. 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nature Reviews Microbiology 8:802-813.

15. Holmes, E. C. 2009. The Evolutionary Genetics of Emerging Viruses. Annual Review of Ecology Evolution and Systematics 40:353-372.

16. Parrish CR, Holmes EC, Morens DM, et al. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72:457-70.

17. I. Katsura, in *Lambda II*, R. W. Hendrix, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1981).

18. Berkane, E., F. Orlik, J. F. Stegmeier, A. Charbit, M. Winterhalter, and R. Benz. 2006. Interaction of bacteriophage Lambda with its cell surface receptor: An in vitro study of binding of the viral tail protein gpJ to LamB (Maltoporin). Biochemistry 45:2708-2720.

19. Grayson, P., L. Han, T. Winther, and R. Phillips. 2007. Real-time observations of single bacteriophage lambda DNA ejections in vitro. Proceedings of the National Academy of Sciences of the United States of America 104:14652-14657.

20. Maynard, N. D., E. W. Birch, J. C. Sanghvi, L. Chen, M. V. Gutschow, and M. W. Covert. 2010. A Forward-Genetic Screen and Dynamic Analysis of Lambda Phage Host-Dependencies Reveals an Extensive Interaction Network and a New Anti-Viral Strategy. PLoS Genet 6:e1001017.

21. R. W. Hendrix, Ed., *Lambda II* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1983).

22. Hendrix, R. W. and R. L. Duda. 1992. Bacteriophage-lambda – not the mother of all lambda-phages. Science 258:1145-1148.

23. Gallet, R., S. Kannoly, and I.-N. Wang. 2011. Effects of bacteriophage traits on plaque formation. BMC Microbiol. 11, 181 (2011).

24. Braunbreton, C. and M. Hofnung. 1978. Explanations accounting for transduction by bacteriophage-gamma in maltose negative bacteriophage gamma-resistant mutants of Escherichia-coli K12. Molec. & Gen. Genet. 159:143-149.

25. Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molec. Systems Biol. 2.

26. Schirmer, T., T. A. Keller, Y. F. Wang, and J. P. Rosenbusch. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1-anstrom resolution. Science 267:512-514.

27. Kefala, G., Ahn, C., Krupa, M., Esquivies, L., Maslennikov, I., et al., 2010. Structures of the OmpF porin crystallized in the presence of foscholine-12. Protein Sci. 19: 1117–1125.

28. Gurnev, P. A., A. B. Oppenheim, M. Winterhalter, and S. M. Bezrukov. 2006. Docking of a single phage lambda to its membrane receptor maltoporin as a time-resolved event. Journal of Molecular Biology 359:1447-1455.

29. Koebnik, R., K. P. Locher, and P. Van Gelder. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol. Microbiol. 37:239–253.

30. Schneider D , et al. 2002. Genomic comparisons among *Escherichia coli*strains B, K-12, and O157:H7 using IS elements as molecular markers. BMC Microbiol 2:18.

31. Crozat E, Hindre T, Kuhn L, Garin J, Lenski RE, Schneider D: Altered regulation of the OmpF porin by Fis in Escherichia coli during an evolution experiment and between B and K-12 strains. J. Bact. 2011:193:429-440.

32. Wang, J., M. Hofnung, and A. Charbit. 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. Journal of Bacteriology 182:508-512.

33. Gould, S. J. 1989. Wonderful life : the Burgess Shale and nature of history. 1st ed. edition. W.W. Norton, New York.

34. Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87-90.

35. Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 105:7899-7906.

36. Woods, R.J. et al. (2011) Second-order selection for evolvability in a large Escherichia coli population. Science 331, 1433–1436

37. Cooper VS, Schneider S, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of E. coli B. J Bacteriol 183:2834 –2841.

38. B. Erni, B. Zanolari, H. P. Kocher, 1987. The mannose permease of Escherichia-coli consists of 3 different functions; transport, sugar phosphorylation, and penetration of phage –lambda DNA J. Biol. Chem. 262, 5238.

39. P. E. J. Saris, E. T. Palva. 1987. Regulation of manX (PTSL) and many (PEL) genes required for mannose transport and penetration of lambda-dna in Echerichia-coli K12 FEMS Microbiol. Lett. 44, 371.

40. M. Esquinas-Rychen, B. Erni, Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenolpyruvate : carbohydrate phosphotransferase system (PTS) *J. Mol. Microbiol. Biotechnol.* 3, 361 (2001).

41. Graham, R. L. and R. S. Baric. 2010. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. J. Virology 84:3134-3146.

42. Ferris, M. T., P. Joyce, and C. L. Burch. 2007. High frequency of mutations that expand the host range of an RNA virus. Genetics 176:1013-1022.

43. A. R. Hall, P. D. Scanlan, A. Buckling. 2011. Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens *Am. Nat.* 177:44.

44. J. N. Thompson. 1994 *The Coevolutionary Process* (Univ. Chicago Press, Chicago, IL).

45. D. I. Friedman, D. L. Court. 2001. Bacteriophage lambda: alive and well and still doing its thing *Curr. Opin. Microbiol.* 4, 201.

46. J. Cairns, G. S. Stent, J. D. Watson 2007. Eds., *Phage and the Origins of Molecular Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

47. Bohannan, B. J. M., and R. E. Lenski. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage.Ecology Letters 3:362-377.

48. Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host–phage interactions. Proceedings of the National Academy of Sciences 108:E288-E297.

49. Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461:1243-1247.

50. Daegelen, P., F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim. 2009. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). J. Mol. Biol. 394:634-643.

51. Spanakis, E. and M. T. Horne. 1987. Coadaptation of *Escherichia coli* and coliophage lambda-*vir* in continuous culture. Journal of General Microbiology 133:353-360.

52. Sambrook, J., and D. W. Russell. 2001. Molecular Cloning, Third Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

53. Adams, M. H. 1959. Bacteriophages. Interscience Publishers, New York, NY.

54. T. Baba et al., *Mol. Syst. Biol.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection *Mol. Syst. Biol.* 2.

55. H. A. Shuman, T. J. Silhavy. 2003. The art and design of genetic screens: Escherichia coli. Nat. Rev. Genet. 4: 419-31.

56. W. Boos, A. Bohm. 2000. Learning new tricks from an old dog: MalT of the Escherichia coli maltose system is part of a complex regulatory network. Trends Genet. 16, 404-9.

57. Meyer, J. R., A. A. Agrawal, R. T. Quick, D. T. Dobias, D. Schneider, and R. E. Lenski. 2010. Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. Evolution 64:3024-3034.

58. Pelosi, L., L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173:1851-1869.

59. Bohannan, B. J. M., and R. E. Lenski. 1997. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology 78:2303-2315.

60. Braunbreton, C. and M. Hofnung. 1978. Explanations accounting for transduction by bacteriophage gamma in maltose negative bacteriophage gamma resistant mutants of Escherichia-coli K-12. Molec. & Gen. Genet. 159:143-149.

61. B. R. Levin, F. M. Stewart, L. Chao. 1977. Complex community in a simple habitat – experimental study with bacteria and phage Am. Nat. 111:3.

62. Lenski, R. E. and B. R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. American Naturalist 125:585-602.

63. M. Schwartz. 1980. in Virus Receptors, *Vol. 1*, K. Lonberg-Holm, L. Philipson, Eds. (Chapman and Hall, London, UK).

64. Kefala, G. Ahn, C. ; Krupa, M., Esquivies, L., Maslennikov, I., Kwiatkowski, W., Choe, S. 2010. Structures of the OmpF porin crystallized in the presence of foscholine-12 Protein Sci. 19, 1117.

65. Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *E. coli*. Proc. Natl. Acad. Sci. USA 100:1072-1077

66. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403.

67. Y. P. Shao, I. N. Wang. 2008. Bacteriophage adsorption rate and optimal lysis time *Genetics* 180:471.

68. M. De Paepe, F. Taddei. 2006. Viruses' life history: Towards a mechanistic basis of a trade-off between survival and reproduction among phages *PLoS Biology* 4:1248.

69. H. D. Rundle, L. Nagel, J. W. Boughman, D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks Science 287:306.

70. J. B. Losos. 2010 Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism. Am. Nat. 175, 623.

71. Ostrowski, E. A., R. J. Woods, and R. E. Lenski. 2008. The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. Proc. Roy. Soc. B 275:277-284.

72. M. Manceau, V. S. Domingues, C. R. Linnen, E. B. Rosenblum, H. E. Hoekstra. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function *Philos. Trans. R. Soc. B 365*:2439.

73. H. A. Wichman, J. Wichman, J. J. Bull. 2005. *Genetics* Adaptive molecular evolution for 13,000 phage generations: A possible arms race 170:19.