EVOLUTION OF THE HOST RANGE OF THE BACTERIOPHAGE ϕ6

MARTIN THOMAS FERRIS

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology.

Chapel Hill
2007

Approved by:

Christina Burch, PhD

Ralph Baric, PhD

Robert Peet, PhD

David Pfennig, PhD

Maria Servedio, PhD

ABSTRACT

MARTIN THOMAS FERRIS: Evolution of the Host Range of the Bacteriophage ϕ6
(Under the direction of Christina L. Burch)

The emergence of novel viral disease into naïve host populations is a specific case of biological invasion. One critical component of viral disease emergence is the evolution of a virus' host range – the ability of a virus to utilize different hosts. In this dissertation I use the bacteriophage ϕ6 as a model system to investigate the evolution of virus host range.

The expansion of a virus' host range is often the first step in emergence. I found that ϕ6 expands its host range by acquiring mutations that allow viral attachment to a novel host, that many mutations can expand host range, and that the majority of these mutations cause a decrease in the fitness of ϕ6 on its original host. I also examined whether the genetic similarity between the original host of ϕ6 and a novel host was predictive of characteristics of host range mutations that are important to emergence. I found that as host genetic similarity decreases, fewer mutations are capable of expanding ϕ6's host range, the fitness costs these mutations cause ϕ6 on its original host decrease, and the fitness of these host range mutants on a novel host decreases. Taken together, these results suggest that as host genetic similarity decreases, viruses should be less successful in colonizing and persisting on a novel host.

Once a virus can grow on a novel host, viral adaptation is often necessary for persistence on this host. I adapted replicate populations of an expanded host range mutant of

ϕ6 to a novel host until each population acquired a single adaptive mutation. I found that all of these mutations greatly increased the fitness of ϕ6 on this novel host, that they occur in many genes in the ϕ6 genome, and that most of these mutations do not reduce the fitness of ϕ6 on its original host. In conjunction with the data I collected on host range expansion, these results suggest that the expansion of a virus' host range will often be the limiting step in successful emergence.

DEDICATION

To my friends and family who have supported me throughout graduate school, but most importantly for the support, love, and good times that my wife Amanda has provided me.

ACKNOWLEDGEMENTS

This work would not have been possible without the assistance of a large number of individuals. Christina Burch has been a wonderful advisor and provided excellent training, mentoring, and collaborations throughout my graduate career. Members of the Burch lab, especially Siobain Duffy, Sarah Joseph, and Jennifer Knies have provided valuable conversations, as well as experimental and writing assistance. A number of other individuals, notably Sarah Diamond, Amanda Ferris, Cris Ledon-Rettig, Ryan Martin, and Amber Rice all provided comments on various portions of this document. Finally, my committee has provided much-appreciated guidance and training throughout my graduate training.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Chapter 1: Background and Introduction

The emergence of disease causing organisms into naïve host populations is a major public health concern[1, 2]. Viral diseases accounted for 44% of recently emergent diseases in the human population [3, 4], including such noteworthy examples as HIV [5] and Influenza A Virus [6]. In addition, viruses are likely to continue to be an emergence threat, in large part because of the high mutation rates of RNA viruses [7]. Despite concerns over the impact of future emergent viral disease, difficulty in empirically studying the emergence process means that few studies have been able to examine factors associated with emergence (but see [3, 4, 8, 9]).

Epidemiological models of disease emergence [10, 11] and evolutionary ecology models of niche expansion [12-17] have both highlighted a small number of factors that are expected to determine the probability of viral emergence on a novel host. These factors include the rate of transmission of viruses to a novel host, the initial fitness of these viruses on this host, and the ability of these viruses to adapt to this host. In addition, there is a large body of literature on the molecular genetic basis and the evolution of viral host range. Results from these empirical studies should be well suited for investigating the factors that theoretical studies have predicted are important to emergence.

In this introductory chapter, I attempt to apply the findings of theoretical studies of emergence to empirical results from studies of virus host range. In order to do so, I first summarize the key predictions of emergence success from niche expansion theory. I then examine the literature on the genetic basis and evolution of viral host range to determine whether these studies can address predictions of niche expansion theory. I conclude by describing how my dissertation work attempts to further unify these two bodies.

Evolutionary Ecological theory of disease emergence

Emerging viruses can be defined as viruses which are increasing in frequency on a particular host [4]. Thus emergence includes both the scenario in which a virus first begins to infect a novel host, and the scenario in which the frequency of infection increases from an initial low level on a previously permissive host. Successful emergence can be further defined as occurring when a virus utilizes a host resource so that the virus' intrinsic rate of increase, R_0, on that host exceeds one (i.e. each successful infection produces more than one new infection) [11]. Under these broad definitions, emergence can involve interactions between different evolutionary and ecological factors, making predictions of emergence difficult [3]. Furthermore, data on which to base these predictions comes largely from those situations where a virus has emerged, and data on conditions which oppose emergence are not available. While the available examples of emergence have allowed for the identification of some emergence risk factors [3, 4, 9], theoretical studies have been able to more precisely define factors important in facilitating emergence.

The evolutionary-ecology theory of niche expansion [16] simplifies much of the complex nature of emergence by utilizing a classical ecology modeling approach: source-sink models [18]. Under this framework, the environments (hereafter hosts) to which a virus is well adapted (i.e. $R_0 \geq 1$) are source hosts, whereas hosts to which a virus is poorly adapted (i.e. $R_0 < 1$) are sink hosts. These models have been used to investigate simple scenarios under which transmission of viruses from source to sink hosts occurs, but not vice-versa. Virus populations on a sink host will be exposed to two types of extinction risks- stochastic risks (due to small population sizes) and deterministic risks (due to $R_0 < 1$). By using this framework, source-sink studies are able to examine the factors that favor the persistence of virus populations on a sink host, as well as the factors that determine whether a virus population can evolve so that a sink host becomes a source host.

Emergence requires the transmission of viral colonists from a source to a sink host. Both the number of individuals transmitted per transmission event and the frequency of transmission events are important to emergence, as these two factors buffer colonizing viruses from extinction on a sink host [12, 16]. Specifically, it has been shown that large numbers of colonists increase a virus' ability to successfully emerge on any sink host and are required for successful emergence on harsh sink hosts (i.e. $R_0 << 1$) [12].

Theory also suggests that adaptation to a sink host will play a large role in the probability of emergence, with several factors having been shown to influence the likelihood of this adaptation. First, the absolute fitness of viruses on a sink host influences the likelihood of adaptation because persisting longer on a sink host will increase the chance of

adaptive mutations appearing in this population [10]. Second, the rate of transmission of sink-adaptive mutations from source populations influences the likelihood of adaptation because continual migration to the sink host can provide more adaptive variation than mutations arising *de novo* on the sink host [13, 14, 16, 19]. Finally, the degree to which adaptive mutations improve fitness on the sink host will be important in determining emergence success because adaptive mutations of small effect are at an increased risk of elimination due to genetic drift and ecological effects (e.g. Allee effects) on the sink host [13, 14].

In sum, this collection of theoretical studies has highlighted several factors that can influence the emergence success of viruses in novel host populations. These factors include:

- The number of viruses transmitted to a sink host per transmission event
- The frequency of these transmission events
- The fitness of a virus on a sink host
- The ability of a virus to adapt to a sink host.

I now review empirical studies of virus host range to determine whether the abundant molecular genetic data sufficiently address the genetic basis of factors important to emergence.

Viral genetics literature review

To identify the available molecular genetic data on virus host range, I performed a literature search using PubMed and WebOfScience to find papers on animal viruses published between 1997 and 2004 that contained the key words ‘virus’ and ‘host range’. I

added relevant studies cited by these papers to this collection. An examination of these studies revealed three types of data relevant to virus emergence. The majority of these studies identified the genetic basis of virus host range. Other studies measured the change in a virus' fitness on a number of hosts following many generations of adaptation. Finally, other studies identified mutations which increased viral fitness on one host, and measured the effects of this mutation on the virus' fitness on a number of hosts. In total, I identified 72 studies of the genetic basis of host range, 25 studies of evolutionary patterns of host usage, and 15 studies which identified the effects of single mutations on host usage.

These studies include a range of viral species, experimental designs, and methodologies. Therefore it is important to determine how the results of these studies differ from theoretical expectations. While theoretical studies base their predictions on an explicit 2 host, source-sink framework, molecular genetic studies diverge from this framework. Due to a number of constraints, empirical studies of virus host range substitute the cellular tropism of viruses (the ability of a virus to bind, enter, replicate in, and escape from a host cell in tissue culture) for the more inclusive definition of host range (transmission between host organisms, avoidance of the immune system, and successful replication within a cellularly complex host). As the selective environments that viruses naturally experience are the complex cellular environments of their natural hosts, most viruses are not well adapted to any cell types in a laboratory setting. Additionally, many viruses have naturally broad cellular tropisms, and studies often include fitness measures on a number of different cell types.

To reconcile these differences between theoretical assumptions and empirical designs, I adjusted my analyses in two ways. First, although the studies I identified investigated the cellular tropisms of viruses, I consider (and refer to) the cell types used in these studies as hosts. Second, as source hosts were not explicitly defined in these empirical studies, when I examine the results of studies where virus fitness was measured on a number of non-selective hosts, I analyze the fitness effects on all non-selective hosts (pleiotropic fitness effects) used in the study. Although such an approach is not optimal for applying these empirical data to theoretical predictions, it does allow me to draw some conclusions regarding the theoretical framework of emergence because the fitness effects that sink-adaptive mutations have on a source host are one specific case of pleiotropic fitness effects. I now review the studies I collected to investigate: 1) the genetic basis of viral host range, 2) the pleiotropic fitness effects of host range evolution, and 3) predictors of these pleiotropic fitness effects.

Genetic basis of host range

Studies of the genetic basis of viral host range have identified the viral genes and points in the infectious cycle at which a virus' growth is prohibited on a non-permissive host. The point at which a block occurs is the target of selection in that virus' emergence on that non-permissive host. I found 72 studies that identified blocks to a virus' infection of a non-permissive host (Table 1.1). The majority of these studies (54/72, or 75%) identified blocks to host range at the attachment and entry of viruses into non-permissive cells. The remaining studies (18/72, or 25%) identified blocks at various points throughout the virus' cellular life cycle (e.g. replication, RNA and protein synthesis). Variation in the identity of blocks to infection occurred both between virus species (e.g. blocks to Vaccinia virus infection never

occurred at attachment and entry [20, 21], while blocks to Measles infection only occurred at attachment and entry [22, 23]), and also within a virus species, but between novel hosts (e.g. Foot and Mouth Disease Virus shows blocks at both attachment [24] and RNA synthesis [25] in different cell types).

Pleiotropic fitness effects of host range evolution

Studies of the evolution of virus host range have identified the pleiotropic fitness effects of mutations that adapt a virus to a host. In particular, we want to know whether mutations that are adaptive on a novel (sink) host are costly on a standard (source) host. The fitness effect that a sink-adaptive mutation has on a virus growing on a source host will determine this mutation's frequency in a source population. In turn, this frequency will determine how often such a mutation is likely to be transmitted from a source host to a sink host. In particular, if a sink-adaptive mutation is costly on a source host, it will be at a very low frequency in a source population and therefore unlikely to be transmitted.

I identified 15 studies containing 25 individual mutations that were beneficial on one host, and had their fitness effects measured on alternate hosts (Table 1.2). These mutations showed a variety of pleiotropic fitness effects with 14 increasing fitness, 12 causing no change in fitness, and 17 being deleterious on alternate hosts. Another way to study the pleiotropic fitness effects of sink-adaptive mutations is to examine the short-term pleiotropic fitness changes that occur when a virus has adapted to a host. Although these pleiotropic fitness changes are the cumulative result of a number of mutations, they can still identify the general nature of pleiotropic fitness effects associated with host adaptation. I identified 25

studies containing 46 individual lineages that were adapted to a host and had their patterns of pleiotropic fitness change measured across a number of hosts (Table 1.3). These lineages showed a variety of pleiotropic fitness changes on alternate hosts, with 19 lines showing adaptation to an alternate host, 8 lines showing no change in fitness on an alternate host, and 40 lines losing fitness on an alternate host.

Although these data sets do not allow determination of whether viral species exhibit different pleiotropic effects, there is abundant variation within a virus species, depending on the alternate hosts fitness was measured on. In addition, studies that identified multiple mutations showed that there was variation in pleiotropic fitness effects depending on the identity of the mutation (e.g., Murine Leukemia Virus mutations beneficial on one host were deleterious on different numbers of alternate hosts [26]). From an emergence standpoint, the variation in these data suggests that the frequency of a sink-adaptive mutation in a source population will depend on the identities of the source and sink hosts, as well as the mutation in question.

Predictors of pleiotropic fitness effects

The ultimate goal of studying disease emergence is to predict risk of emergence. I now investigate whether there are characteristics of the mutations I identified in the previous section that are predictive of deleterious pleiotropic fitness effects, effects that would reduce the frequency of sink-adaptive mutations in a source population. The studies of individual mutations I collected also identified: 1) the viral gene(s) these mutations occurred in, and 2) whether a mutation expanded a virus' host range or adapted a virus to a host it could already

use. I determined whether either of these categories is associated with deleterious pleiotropic fitness effects.

The viral genes in which the sink-adaptive mutations occur might cause differences in the pleiotropic fitness effects of these mutations. Specifically, extracellular host components (those involved in virus attachment and entry) are expected to be more divergent between species than intracellular host components. As a result, mutations that adapt a virus to a sink host's extracellular components might be expected to cause more deleterious pleiotropic fitness effects than mutations that adapt a virus to a sink host's intracellular components. I determined whether the mutations identified in Table 1.2 occurred in viral structural genes (those likely to interact with extracellular components), or non-structural genes (those likely to interact with intracellular components). 13 out of 19 structural mutations and 4 out of 6 non-structural mutations showed a deleterious pleiotropic fitness effect. A χ^2 test failed to show that structural and non-structural mutations did not have different chances of exhibiting a deleterious pleiotropic fitness effect (χ^2_1=0.006, P=0.94).

The transmission of sink-adaptive mutations from a source to a sink host is important both for mutations that expand a virus' host range (expansion mutations), and also for mutations that adapt a virus to a sink host (adaptive mutations). A study of emergence into human populations observed that zoonotic viruses (those already able to infect humans) are more likely to be emerging than viruses newly infecting humans [4]. One possible contributor to this pattern is differences in the transmission of sink-adaptive mutations from a source population. To determine if there are differences in deleterious pleiotropic fitness

effects between expansion and adaptive mutations, I classified the mutations identified in Table 1.2 as either expanding a virus' host range, or further adapting a virus to a host. 4 out of 7 adaptive mutations and 13 of 18 adaptive mutations showed a deleterious pleiotropic effect. A χ^2 test failed to show significant differences between expansion and adaptive mutations in the likelihood of showing a deleterious pleiotropic fitness effect (χ^2_1=0.527, P=0.47).

Conclusions

In this chapter, I have summarized a theoretical framework of niche expansion which can be used to address questions of disease emergence. I also reviewed results from studies of the genetic basis and evolution of virus host range. Taken as a whole, these empirical studies show that there is considerable variation between virus species, as well as within a virus species across hosts, in both the genetic basis of host range and the pleiotropic fitness effects of viral adaptation to a host. However, it is not obvious which viral or host characters are predictive of this variability. Furthermore, despite this variation, it is difficult to directly apply these empirical data to the theoretical framework of niche evolution.

The difficulty I found in reconciling empirical and theoretical literature is not surprising, as the reviewed empirical studies were not designed with testing niche expansion theory in mind. As a result, existing empirical data can not be used to directly address predictions derived from theoretical studies for a number of reasons. First, due to the fact that source hosts were not explicitly defined in empirical studies, conclusions regarding pleiotropic fitness effects were generalized across all non-selective hosts in these studies, not

just the source hosts relevant to emergence theory. Second, although I could classify the general nature of the pleiotropic fitness effects of sink-adaptive mutations, I could not compare the magnitude of these mutations effects. Finally, given the extensive variation observed in the pleiotropic fitness effects of sink-adaptive mutations, the fact that only one or a few mutations were identified in most of these studies makes it difficult to generalize from these results.

Despite drawbacks in the design of previous studies, designing empirical experiments with niche expansion theory in mind should allow empirical studies to precisely examine the genetic basis of factors important to emergence. Specifically, an ideal study would identify a large collection of sink-expansion or sink-adaptive mutations and characterize the magnitude of these mutations' fitness effects, both in the sink host and in a more reasonable approximation of the source host. Such a study would be able to address both the frequency of these mutations in the source population (a major component of transmission rate) and the fitness of these viruses on the sink host, both critical predictors of emergence success. In the chapters that follow, I investigate these genetic determinants of virus disease emergence using the bacteriophage ϕ6. Due to its long history as a molecular [27-31] and evolutionary [32-39] model system, ϕ6 is an ideal system with which to investigate genetic determinants of virus disease emergence.

In chapter 2, I investigate the genetic basis, phenotypic basis, and fitness consequences of mutations that have expanded the host range of ϕ6 to include the novel host *Pseudomonas syringae* pathovar *glycinea*. In this study, I identify a large collection of sink-

expansion mutations and characterize the fitness effects of these mutations on both the sink host and the source host of ϕ6. The resulting data enable me to estimate the frequency that these sink-expansion mutations will have in a ϕ6 population on the source host.

Having shown the suitability of ϕ6 for investigating the genetic basis of host expansion, in chapter 3 I investigate characteristics of mutations that expanded ϕ6 host range to include one of three novel host types. I identify collections of sink-expansion mutations that allow growth on one of three novel (sink) hosts, and characterize the fitness effects of these mutations on the sink hosts and the source host of ϕ6. Based on these data, I show that the genetic distance between ϕ6's source and sink hosts is predictive of both the frequency of sink-expansion mutations in a ϕ6 population on the source host and the fitness of these mutations on the sink host.

Finally, in chapter 4, I compare the characteristics of mutations that expand ϕ6 host range to include the novel host *P. syringae* pv. *glycinea* to those mutations that adapt ϕ6 to *P. syringae* pv. *glycinea*. I identify a collection of sink-adaptive mutations and characterize the fitness effects of these mutations on both the sink host and source host of ϕ6. I then determine that the frequency of sink-adaptive mutations in a ϕ6 population on the source host is typically greater than the frequency of sink-expansion mutations in a ϕ6 population on the source host.

Table 1.1: Blocks to viral infection of host cells

Virus	Experimental Assays	Block to infection of non-permissive cell type
Adeno-associated Virus Type 2 [40]	Growth, chimeric virus	Attachment/Entry
Amphotrophic Murine Retrovirus [41]	Growth, binding, chimeric host	Attachment/Entry
Avian Leukosis Virus [42]	Sequence analysis	Attachment/Entry
Avian Leukosis Virus [43]	Growth, integration, mutant virus	Attachment/Entry
Avian Sarcoma Leukosis Virus [44]	Binding, chimeric host	Attachment
Avian Sarcoma Leukosis Virus [45]	Growth, binding, mutant virus	Attachment
Avian Retrovirus [46]	Growth, mutant virus	Attachment
B Lymphotrophic Papovavirus [47]	Binding	Attachment
B19 Parvovirus [48]	Protein synthesis, mRNA production, transport, stability, and ribosomal association	Protein synthesis
Bovine Respiratory Syncytial Virus [49]	Growth, chimeric virus	Attachment/Entry
Bovine Viral Diarrhoea Virus [50]	Growth, chimeric virus	Attachment/Entry
Canine Parvovirus [51]	Binding	Attachment
Canine Parvovirus [52]	Growth	Attachment/Entry
Canine Parvovirus [53]	Binding	Attachment
Canine Parvovirus [54]	Binding	Attachment
Duck Hepatitis B Virus [55]	Binding, chimeric virus	Attachment
Duck Hepatitis B Virus [56]	Binding, growth, chimeric virus	Entry
Ebola Virus [57]	Growth, chimeric virus	Attachment/Entry
Feline Leukemia Virus [58]	Binding, growth, chimeric virus	Attachment
Feline Leukemia Virus [59]	Growth	Attachment/Entry

Feline Leukemia Virus [60]	Binding, growth, chimeric virus	Attachment
Feline Leukemia Virus A [61]	Binding	Attachment
Foot and Mouth Disease Virus [24]	Binding, growth, chimeric virus, mutant virus	Attachment
Foot and Mouth Disease Virus [25]	Growth, RNA synthesis, mutant virus	RNA synthesis
Hepatitis B Virus [62]	Growth, chimeric virus	Attachment/Entry
Hepatitis B Virus [63]	Growth, transfection, chimeric vurs	Attachment/Entry
HIV [64]	Growth, mutant virus	Attachment/Entry
HIV [65]	Growth, mutant virus	Attachment/Entry
HIV[66]	Entry, transcription, genome integration	Genome preparation
Human Adenovirus 2 [67]	mRNA production and processing	mRNA processing
Human Cytomegalovirus [68]	DNA replication, viral promoter function, protein production	DNA replication
Human Rotavirus RV-3 [69]	Binding, growth, chimeric virus	Attachment
Influenza A Virus [70]	Growth	Attachment/Entry
Influenza A Virus [71]	Binding	Attachment
Influenza A Virus [72]	Binding	Attachment
Influenza A Virus [73]	Binding, growth, chimeric virus, mutant virus	Attachment/Entry
JC Virus [74]	Growth, mutant virus	Attachment/Entry
Measles [75]	Binding, growth	Attachment/Entry
Measles [22]	Growth, chimeric virus	Attachment/Entry
Measles [23]	Binding	Attachment
Minute Virus of Mice [76]	Growth, DNA virus, chimeric virus	Attachment/Entry
Minute Virus of Mice [77]	Binding, entry, DNA replication	Uncoating
Moloney Murine Leukemia Virus [78]	Binding, gene expression, chimeric virus	Attachment/Entry
Moloney Murine Leukemia Virus [79]	Growth, DNA production, chimeric virus	Genome replication
Mouse Hepatitis Virus[80]	Binding, growth, mutant virus	Attachment/Entry

Murine Hepatitis Virus [81]	Growth	Attachment/Entry
Murine Hepatitis Virus [82]	Growth, chimeric virus	Attachment/Entry
Murine Hepatitis Virus [83]	Binding, growth, chimeric virus	Attachment
Murine Leukemia Virus [84]	Growth, chimeric virus	Attachment/Entry
Murine Leukemia Virus [85]	Growth, chimeric virus	Attachment/Entry
Murine Leukemia Virus [86]	Growth, chimeric virus	Attachment/Entry
Murine Parvovirus [87]	Growth, transfection, protein production, chimeric virus	DNA replication
Norwalk Virus [88]	Binding	Attachment/Entry
Parainfluenza Virus [89]	Protein synthesis, genome replication, virion assembly	Genome replication/Virion assembly
Polio Virus [90]	Binding, uncoating, RNA release	Uncoating
Polio Virus [91]	mRNA synthesis, protein synthesis, mutant virus	Protein synthesis
Polyomavirus [92]	DNA replication	DNA replication
Procine Endogenous Retrovirus [93]	DNA replication, RNA synthesis	DNA replication
RD114 virus [94]	Growth, RNA synthesis, protein synthesis	Attachment/Entry
Rhinovirus type 16 [95]	Growth, chimeric host	Attachment/Entry
Ross River virus [96]	Binding, growth	Attachment/Entry
Rous Sarcoma Virus [97]	Growth, chimeric virus	Attachment/Entry
Ruminant Lentivirus [98]	Binding, cell fusion, chimeric host	Attachment/Entry
SARS Coronavirus [99]	Growth, RNA synthesis, chimeric virus	Attachment/Entry
Simian Sarcoma-Associated Virus [100]	Growth, chimeric virus	Attachment/Entry
Sindbis Virus [101]	Growth, protein synthesis, virion assembly and exit, mutant virus	Virion assembly/exit
Sindbis Virus [102]	Growth, mutant virus, mutant host	Virion assembly
SIV [103]	Binding, cell fusion, chimeric virus, chimeric host	Attachment/Entry
	Growth, chimeric virus	Attachment/Entry

SIV [104]		
Vaccinia virus [105]	Protein synthesis	mRNA synthesis
Vaccinia virus [20]	Cell entry, genome replication, RNA synthesis, protein synthesis, chimeric virus	Protein synthesis
Vaccinia virus [21]	Growth, genome replication, protein synthesis, assembly	Virion assembly

Summary table of those studies identifying limits to viral infection of host types. For each study, the virus type, experimental assays used, and the point in the life cycle where the block occurs are identified.

Table 1.2: Pleiotropic fitness effects of mutations

Virus	Mutation Type	Gene occurring in	Fitness effect of mutation on alternate hosts			Selective Environment
			Beneficial	Deleterious	Neutral	
Avian Sarcoma Leukosis Virus[45]	Expansion	Structural	6	0	1	Passaged simultaneously on two host types
Dengue Virus [106]	Adaptive	Non-structural	2	1	0	Mutant Screen
Hepatitis A Virus [107]	Adaptive	Non-structural	0	0	1	Passaged on one host type
Hepatitis A Virus [108]	Adaptive	Non-structural	0	1	1	Passaged on one host type
	Adaptive	Non-structural	0	1	1	
HIV [65]	Expansion	Structural	0	1	0	Natural Isolate
HIV [109]	Expansion	Structural	0	0	1	Passaged on one host type
Influenza A Virus [72]	Adaptive	Structural	0	0	1	Natural Isolate
Influenza A Virus [110]	Expansion	Non-structural	0	0	1	Passaged on one host type
Moloney Murine Leukemia Virus [79]	Expansion	Non-structural	0	1	0	Passaged on one host type
Moloney Murine Leukemia Virus [111]	Adaptive	Structural	1	0	1	Passaged on one host type[a]
	Adaptive	Structural	1	1	0	

Murine Leukemia Virus [112]	Expansion	Structural	0	1	0	Mutant Screen
Murine Leukemia Virus [26]	Expansion	Structural	1	7	0	Mutant Screen
	Expansion	Structural	1	7	0	
	Expansion	Structural	2	5	1	
	Expansion	Structural	3	4	1	
Poliovirus Type 1 [90]	Expansion	Structural	0	1	0	Natural Isolate
	Expansion	Structural	0	1	0	
Poliovirus Type 1 [113]	Expansion	Structural	2	0	1	Passaged on one host type
	Expansion	Structural	2	0	1	
	Expansion	Structural	2	1	0	
	Expansion	Structural	2	1	0	
Ross River Virus [96]	Expansion	Structural	1	1	0	Passaged on one host type
	Expansion	Structural	2	1	0	

[a] Passaged on animal host

Table 1.3: Pleiotropic fitness effects of adapted virus lineages

Virus	Number of alternate hosts on which			Selective Environment
	fitness increased	fitness decreased	fitness did not change	
African Swine Fever Virus [114]	0	1	0	Passaged on one host type
	0	1	0	
	0	1	0	
Avian Sarcoma Leukosis Virus [45]	6	0	1	Co-passaged on two host types
Hepatitis A Virus [115]	1	1	0	Passaged on one host type
Hepatitis A Virus [107]	0	0	1	Passaged on one host type
Hepatitis A Virus [108]	0	1[a]	1	Passaged on one host type
	0	1[a]	1	
HIV [65]	0	1	0	Natural Isolate
HIV [109]	0	1	0	Passaged on one host type
Influenza A Virus [72]	0	1	0	Natural Isolate
Measles Virus [116]	0	1	0	Passaged on one host type
Moloney Murine Leukemia Virus [111]	0	1	0	Passaged on one host type[b]
Moloney Murine Leukemia Virus [79]	0	0	1	Passaged on one host type
Mouse Hepatitis Virus [117]	4	1	0	Co-passaged on two host types
Mouse Hepatitis Virus [118]	6	1	0	Passaged on one host type
Murine Leukemia Virus [85]	0	1	0	Natural Isolate
Poliovirus [90]	0	1	0	Natural Isolate
Polyoma Virus [119]	1	1	0	Passaged on one host type
Ross River Virus [96]	1	1	0	Passaged on one host type
	2	1	0	

Sendai Virus [120]	0	1	0	Passaged on one host type
Sindbis Virus [121]	0	1	0	Passaged on one host type
SIV [122]	0	1	0	Passaged on one host type
Vaccinia Virus [123]	2	10	3	Passaged on one host type
Vesicular Stomatitis Virus [124]	0	1	0	Passaged on one host type
Vesicular Stomatitis Virus [125]	0	2	0	Passaged on one host type
	0	2	0	
	0	2	0	
	0	2	0	
	0	2	0	
	0	2	0	
	0	2	0	
	0	2	0	
	1	1	0	Passaged on two host types
	1	1	0	
	1	1	0	
	1	1	0	
	1	1	0	
	1	1	0	
	1	1	0	
	1	1	0	
Vesicular Stomatitis Virus [126]	2	0	0	Passaged on one host type
	1	0	1	
	1	0	1	
Yellow Fever Virus [127]	0	1	0	Passaged on one host type[b]

[a] Fitness measured on animal host
[b] Lineage adapted to animal host

Chapter 2: High frequency of mutations that expand the host range of an RNA virus

The work described in this chapter was accomplished in collaboration with Drs. Christina Burch and Steve Joyce. This chapter has been published [128] as a paper in Genetics. I would like to thank Corbin Jones, Siobain Duffy, and members of the Burch lab for discussions that improved both the experimental design and the final manuscript.

Abstract

The ability of a virus population to colonize a novel host is predicted to depend on the equilibrium frequency of potential colonists (i.e. genotypes capable of infecting the novel host) in the source population. In this study, we investigated the determinants of the equilibrium frequency of potential colonists in the RNA bacteriophage ϕ6. We isolated 40 spontaneous mutants capable of infecting a novel *Pseudomonas syringae* host and sequenced their host attachment genes to identify the responsible mutations. We observed 16 different mutations in the host attachment gene and used a new statistical approach to estimate that 39 additional mutations were missed by our screen. Phenotypic and fitness assays confirmed that the proximate mechanism underlying host range expansion was an increase in the ability to attach to the novel host, and that acquisition of this ability most often imposed a cost for growth rate on two standard hosts. Considered in a population genetic framework, our data suggest that host range mutations should exist in phage populations at an equilibrium frequency (3×10^{-4}) that exceeds the phage mutation rate by more than two orders of magnitude. Thus, colonization of novel hosts is unlikely to be limited by an inability to produce appropriate mutations.

Introduction

The increasing threat of disease emergence, especially among RNA viruses, provides considerable incentive for predicting whether and when virus populations will acquire the ability to colonize and adapt to a novel host. To make such predictions we must identify the factors that explain why viruses like HIV and influenza successfully adapted to human hosts, whereas viruses like SARS caused outbreaks but failed to persist. Progress toward this goal

will likely come from the application of ecological models that describe the colonization of sink habitats to the study of emerging pathogens [129]. In this study, we focus on one of the primary predictors of colonization success in these models [13-15, 19]: the rate of migration into the novel habitat.

In particular, we consider the scenario in which the ability to infect a novel host requires a mutation. In this case, the migration rate will depend jointly on the rate at which viruses are transmitted to the novel host and on the equilibrium frequency of potential colonists (i.e. genotypes capable of infecting the novel host) in the source population. Although transmission rate is determined by ecological factors that must be measured in the field, the equilibrium frequency of potential colonists is determined by two evolutionary factors that can be investigated in the laboratory: mutation and selection. Mutation will act to increase the frequency of potential colonists. In contrast, if the ability to infect a novel host imposes a pleiotropic fitness cost on the standard host, selection will act to reduce the frequency of potential colonists. An equilibrium will be achieved when the effects of mutation are exactly balanced by the effects of selection. Therefore, the equilibrium frequency of potential colonists in a population growing on its standard host will depend on the mutation rate, the number of different mutations that confer the ability to infect the novel host, and on the abundance and magnitude of pleiotropic fitness costs among these mutations.

There have been numerous investigations of the identity and effects of mutations that expand host range [26, 45, 52, 65, 79, 109, 110, 112, 130]. However, several characteristics

of these investigations make them unsuitable for predicting equilibrium frequencies of potential colonists in natural populations. First, investigations of mutations that expand host range have tended to examine only one or a few mutations, making it difficult to infer whether other mutations are possible. Second, the mutations examined in these studies were usually the result of long term adaptation in a laboratory or natural setting (i.e. fixed mutations). Fixed mutations have been sieved by natural selection acting on one or both of the standard and novel hosts, and therefore, it is likely that the distribution of pleiotropic fitness costs among fixed mutations will differ from the distribution among new spontaneous mutations. Finally, in most of these investigations fitness was assayed in tissue culture, which likely mimics the natural environment to only a limited extent.

To overcome these obstacles we investigated the possible genetic bases of host range expansion in the RNA bacteriophage ϕ6, a model system in which it was possible to isolate a large random sample of mutants with an expanded host range and to measure fitness in a manner that more closely mimicked the natural environment. We screened spontaneous ϕ6 mutants for the ability to infect a novel *Pseudomonas syringae* host. We sequenced the host attachment gene of 40 of the resulting phage to identify the mutations responsible for host range expansion, and developed a statistical method for estimating the total number of ways the attachment gene can mutate to acquire the ability to infect the novel host. In addition, we determined the abundance and magnitude of the pleiotropic fitness costs associated with these mutations on two standard (permissive) hosts, and identified the phenotypic basis of the host range expansion.

Materials and Methods

Strain and Culture Conditions

The double-stranded RNA bacteriophage ϕ6 (*Cystoviridae*) used in this study is a laboratory genotype descended from the original isolate [131]. The standard laboratory host of ϕ6, *Pseudomonas syringae* pathovar *phaseolicola* strain HB10Y, was obtained from the American Type Culture Collection (ATCC; no. 21781); an alternate permissive host, *P. syringae* pathovar *japonica* strain M301072, was obtained from D. Guttman (University of Toronto, Toronto, CA); and the novel host *P. syringae* pathovar *glycinea* strain R4a was obtained from J. Dangl (University of North Carolina, Chapel Hill, NC). Hereafter, hosts will be referred to by their pathovar designations. Details of diluting, filtering, culture, and storage of phage and bacteria are published [27, 33]. Phage and bacteria were cultured in LC media (5 g yeast extract, 5 g NaCl, and 10 g Bacto-tryptone per L of H_2O), and stored in 4:6 glycerol:LC media (v/v) at -80^{o}C and -20^{o}C, respectively. For growth on plates, phage were mixed with the appropriate host bacteria in top agar (0.7% agar) and plated on LC plates (1.5% agar).

Mutant Isolation

ϕ6 clones were plated onto a lawn of the standard (permissive) host *phaseolicola*, and incubated overnight to allow the phages to reproduce and form plaques. After 24 hours, phages were harvested from a randomly chosen isolated plaque and plated onto a fresh lawn of 200 μL of a stationary phase culture of the novel host *glycinea*. Only phages that acquired a host range mutation during growth of the plaque on *phaseolicola* form plaques on *glycinea*. After 24 hours, an isolated plaque was chosen randomly from the *glycinea* plate, and phages

from this plaque were plated on a fresh lawn of *glycinea* to purify the mutant phages of wild type ϕ6. A single plaque was harvested and stored for later use in 4:6 glycerol:LC media (v/v) at -20° C. This process was repeated 40 times to obtain 40 independent host range mutants.

Sequencing

Genome amplification and sequencing were performed as previously described [130]. Briefly, phage were grown to a high titer and viral RNA was extracted using a QIAamp Viral RNA Mini Kit (QIAGEN Inc., Valencia, CA). Viral RNA was reverse transcribed using random hexamer primers and Superscript II RNase H- RT (Invitrogen, Carlsbad, CA), and the resulting cDNA was used as template for PCR with Taq Polymerase (Invitrogen, Carlsbad, CA). We amplified three sections of the medium genome segment, corresponding to bases 1298-2142, 2042-3052, and 2877-3873, which encompassed the host attachment gene, P3. PCR products were purified using EXO-SapIT (US Biological, Swampscott, MA) and sequenced in both directions using PCR primers and primers internal to each amplicon. Sequencing was performed using BigDye v3.1 (Applied Biosystems, Foster City, CA) either locally with an Avant-3100 Genetic Analyzer Sequencer (Applied Biosystems, Foster City, CA) or at the UNC Automated Sequencing Facility (University of North Carolina, Chapel Hill, NC).

Examining the characteristics of host range mutations

We used χ^2 tests to compare the chemical properties of the observed P3 mutations to the random expectation. Random expectations were determined from the frequencies of P3

codons with particular chemical properties: acidic (D,E), basic (K, R, H), hydrophilic (N, Q, S, T, Y), or hydrophobic (A, C, F, G, I, L, M, P, V, W).

Paired growth fitness assay

Paired growth assays are a standard method for assaying fitness in ϕ6 [32], and were used to assay fitness on the permissive host *japonica*. Each host range mutant was mixed with the wild type ϕ6 at a 1:1 ratio. This mixture was plated on a bacterial lawn and incubated for 24 hours. The ratio of phage genotypes before and after the incubation was determined by plating on a mixed lawn of 200 μL of a 1:1 mixture of *phaseolicola* and *glycinea*, on which the wild type ϕ6 forms turbid plaques and mutant ϕ6 form clear plaques. The relative fitness of mutant genotypes were then determined as $W = R_1/R_0$, where R_0 and R_1 are the ratio of mutant to wild type phage before and after the 24 hour incubation, respectively. Replicate assays (N=4) were collected in blocks on different days.

Plaque size fitness assay

We recently developed a plaque size assay as an alternative means of measuring fitness on *phaseolicola* [38], and this assay proved useful for measuring fitness on the novel host *glycinea*, on which paired growth assays are not possible (because wild type ϕ6 does not grow on this host). On *phaseolicola*, the relationship between the paired growth measure of log(fitness) and plaque area is described by the equation: $\log W = 0.044*PS - 0.34$, where W is a one generation measure of relative growth rate, and PS is plaque area in mm^2 [38]. To calibrate the relationship on *glycinea*, we measured plaque size and the number of phages per plaque for 8 host range mutants grown on lawns of *glycinea* for 24 hours. As on

phaseolicola, there is a linear relationship between log(fitness) and plaque area (log(phage/plaque) = 0.71498 × (plaque size) + 4.34418; $R^2 = 0.7721$, $F_{1,6} = 24.71$, $p = 0.0025$). Plaque sizes were determined by plating phages onto a lawn of the appropriate host at a low density (<50 phage per plate) to ensure non-overlapping plaques, incubating at 25°C for 24 hours, and taking digital pictures for analysis using ImageJ (NIH, Bethesda, MD, http://rsb.info.nih.gov/ij/). Each plaque size measure is a mean area of plaques on an individual plate. For each genotype, six replicate assays were collected in blocks on different days.

Attachment assays

Attachment assays were performed following the method of [132]. An exponentially growing culture of *glycinea* was incubated shaking at 25°C until it achieved an OD of 0.8 (~5x10^8 CFU/ml), at which point the bacteria were pelleted and resuspended in ½ the total volume of LC media. 10^3 phage were added to 1 mL of the concentrated bacterial culture and this mixture was incubated at 25°C with shaking. Immediately and after 40 minutes a 500 μL aliquot of this culture was centrifuged at 5000 rpm, 5°C for 1 minute, and 200 μL of the supernatant was plated on a lawn of *phaseolicola* cells to obtain a count of the unbound phage remaining in the supernatant. The attachment rate constant was then calculated as $k = -1/(40N) \times \ln(P_{40} / P_0)$, where N is the concentration of bacteria (determined by plating), and P_0 and P_{40} are the number of unbound phage at 0 and 40 minutes, respectively. Replicate assays ($n = 4$ for mutant genotypes, $n = 8$ for wild type ϕ6) were collected in blocks on different days.

Statistical analyses

Fitness data were analyzed in Microsoft Excel 2003 (Redmond, WA) and SASv9.1 (SAS Institute, Cary, NC) using Proc GLM and Proc Corr. All models in SAS included block effects, but none of these block effects were significant. To determine which mutations caused a significant reduction in fitness on permissive host types, relative to the wild type ϕ6, we calculated least significant differences (LSD). The LSD is the smallest difference between any two means that is statistically significant, and is used for pre-planned comparisons in ANOVA [133]. The t-statistic used to calculate the LSD is determined in the same manner as the t-statistic used in a two sample t-test, except that the Mean Square Error (MSE) is used in place of the sample variance and the degrees of freedom is based on the MSE.

We used Proc GLM (SASv9.1) to conduct a one-way ANOVA to test for an effect of genotype on attachment rate among the phage examined in this study, including the host range mutants and the wild type ϕ6. In addition, we implemented a bootstrap procedure in Matlab v6.5 (Mathworks, Natick, Massachusetts) to more directly compare the mean attachment rate of the wild type ϕ6 to that of the host range mutants. We pooled the replicate attachment rate measures for all the mutants, and drew 1,000,000 bootstrapped samples of 8 measures with replacement from this pool. The mean attachment rates for each bootstrapped sample were used to generate a frequency distribution that describes the expectation for the wild type ϕ6 mean if the wild type ϕ6 attachment rate did not differ from the attachment rate of mutant phage. We obtained a *p*-value by determining the percentage of the bootstrapped means that were lower than the observed mean attachment rate of wild type ϕ6 (also a mean

of 8 measures). This *p*-value is equivalent to the probability that the wild type ϕ6 attachment rate measures were drawn from the same distribution as the mutant measures.

Estimating the total number of mutations that allow growth on *glycinea*

Our experiment ran $n = 39$ independent trials in which one of an unknown number of possible P3 mutations that enable infection of a novel host was sampled randomly. We know that the observed number of mutations is a lower bound on the total number of possible adaptive mutants, and use the pattern of variability in the data to estimate how many such mutants may have been missed.

This statistical problem is analogous to the well known coupon collecting problem [134] in which we have collected a sample of n coupons and observed K distinct coupons, with some coupons appearing multiple times in our sample. We now want to estimate the total number of distinct coupons N from which our sample has been drawn. This estimation procedure has been used by wildlife biologists since the 70's [135] to estimate population sizes of wild populations from samples of trapped animals.

The standard coupon collecting problem makes several simplifying assumptions that must be adjusted to adapt the methodology for the problem at hand. In particular, the standard problem assumes that every coupon was equally likely to be sampled. However, we know that transition mutations are more likely than transversion mutations and so we expect to sample adaptive transversions less often. Wildlife biologists have made similar adjustments to their models to account for sampling heterogeneity among `trap happy' and

`trap shy' animals [136]. Below we develop likelihood and method of moments frameworks for estimating the number of mutants that enable infection of a novel host.

Maximimum likelihood (ML). The probability of obtaining any particular collection of mutations is described by a multinomial distribution governed by the following two rules: 1) each trial can result in one of $t + r$ possible outcomes, where t and r are, respectively, the total number of transition and transversion mutations that enable infection of the novel host; and 2) defining α as the ratio of transitions to transversions, the probabilities of sampling particular transitions and transversions are $\alpha/(\alpha t + r)$ and $1/(\alpha t + r)$, respectively. If we now let the random variables $\mathbf{J} = (J_1, J_2, \ldots J_T,)$ represent the number of times the T observed transitions occurred in our data set, and the variables $\mathbf{K} = (K_1, K_2, \ldots K_R)$ represent the number of times the R observed transversions occurred, then the likelihood of obtaining these observations is:

$$L(T, R, \mathbf{J}, \mathbf{K}) = \frac{n!}{J_1!J_2!\cdots J_T!K_1!\cdots K_R!}\binom{t}{T}\binom{r}{R}\left(\frac{\alpha}{\alpha t + r}\right)^{\sum J_i}\left(\frac{1}{\alpha t + r}\right)^{\sum K_i} \qquad (1)$$

Since α can be estimated from an external data set, we estimated the total number of mutations that enable infection of the novel host using observed values for α, T, R, $J_1, J_2, \ldots J_T$, and $K_1, K_2, \ldots K_R$, and then determining the values of t and r that maximize this likelihood (using the R statistical package; http://www.r-project.org/).

Method of moments (MM). Note that maximum likelihood estimates for t and r depend only on $\alpha, T, R,$ and $N_1 = \sum J_i$. (Note that $N_2 = \sum K_i = n - N_1$.) Recall that α is obtained from external data. Statistical theory guarantees that any estimation procedure

based on the sufficient statistics T, R, and N_1 will give the same quality of inference. Therefore, we developed a method of moments estimate based on T, R, and N_1 which is simpler to calculate and should give the same quality of inference as maximum likelihood. Using the method of moments estimator, the expectations for T, R, and N_1 are as follows:

$$T = \hat{t} - \hat{t}\left(1 - \frac{\alpha}{\alpha\hat{t} + \hat{r}}\right)^{N_1} \tag{2}$$

$$R = \hat{r} - \hat{r}\left(1 - \frac{1}{\alpha\hat{t} + \hat{r}}\right)^{N_2} \tag{3}$$

$$N_1 = n\left(\frac{\alpha\hat{t}}{\alpha\hat{t} + \hat{r}}\right) \tag{4}$$

The formulas are not too difficult to interpret. The number of transitions that you observe (T) approximates the average number of observed transitions $E(T)$, where $E(T)$ is total number of transitions that exist (t) minus the expected number of transitions that were missed due to sampling error. The probability of missing a particular transition in each of the N_1 trials is $(1 - \alpha/(\alpha t + r))^{N_1}$, so the mean number missed is $t(1 - \alpha/(\alpha t + r))^{N_1}$.

Rearranging equation (3) gives $N_1 / n\hat{t} = \alpha/(\alpha\hat{t} + \hat{r})$, and substituting this into equation (1) gives

$$\hat{t} = T + \hat{t}\left(1 - N_1 / n\hat{t}\right)^{N_1} \tag{5}$$

This equation was solved iteratively by starting with $\hat{t}_0 = T$ and defining

$$\hat{t}_{k+1} = T + \hat{t}_k\left(1 - N_1 / n\hat{t}_k\right)^{N_1} \tag{6}$$

and $\hat{t} = \lim_{k\to\infty} \hat{t}_k$. Because α was obtained from external data, $\hat{r}$ was determined by rearranging the definition of $\alpha = (N_1\hat{r})/(N_2\hat{t})$ to yield:

$$\hat{r} = \frac{\alpha N_2 \hat{t}}{N_1}. \quad (7)$$

We used a parametric bootstrap to determine a 95% confidence interval for these estimates. The bootstrap assumes that the estimates of $\hat{t}$ and $\hat{r}$ are the true values and generates simulated datasets based on $\hat{t}$, $\hat{r}$, and the known α. We generated 1000 simulated datasets of 39 sampled mutations, and estimated $\hat{t}$ and $\hat{r}$ for each dataset using equations 4 – 6. Upper and lower 95% confidence limits were calculated, respectively, as the 26th lowest and 975th highest bootstrapped estimates.

Results

Mutant Identification

We isolated a total of 40 host range mutants on the novel host *glycinea*. To determine which of these mutants carried unique mutations, we sequenced the P3 gene from all forty mutants. P3 encodes the host attachment spike of ϕ6 [137], and previous studies [130, 137] implicated P3 in host range expansion. The 40 host range mutants were comprised of 19 unique P3 genotypes, designated A thru S (Table 2.1). One genotype (A) had no mutations in P3, 16 genotypes had a single mutation in P3, and two genotypes (Q and R) had two mutations in P3. However, one of the mutations present in genotypes Q and R was identical to the single mutation possessed by genotype P. These data are consistent with the presence of 17 unique nucleotide mutations in our collection that confer the ability to grow on the novel host *glycinea* – 16 in P3 and one elsewhere in the genome. Of the 16 mutations in P3, only two (G and H) produced an identical amino acid change.

Number of mutations capable of expanding host range

Because several mutations were represented more than once in our collection, we could use the sampling distribution of particular mutations to estimate the total number of ways that the ϕ6 P3 gene can be mutated to allow infection of the novel host *glycinea*. This estimation problem is analogous to the 'coupon collecting problem' that is well known in probability and statistics, except that we divided the mutations into two rate classes: transitions and transversions. From an external data set we know that the relative rate of transitions per transition site to transversions per transversion site (α) is 24.5 [39]. We used the method of moments (MM) to estimate the total number of transitions ($\hat{t}$) and transversions ($\hat{r}$) that allow infection of *glycinea* from the sampling distribution and α (we report maximum likelihood (ML) estimates for comparison). Recall that we observed eleven transitions and 5 transversions in our mutation sample. The MM estimate of $\hat{t}$ was 11.9, with a 95% confidence interval of $11.0 \leq t \leq 14.0$ (using ML, $\hat{t} = 11.0$). The similarity between our estimate of $\hat{t}$ and the observed number of transitions (11) is consistent with the observation that particular transition mutations were represented as many as eight times in our collection. The MM estimate of $\hat{r}$ was 42.9, with a 95% confidence interval of $7.8 \leq r \leq 103.5$ (using ML, $\hat{r} = 41.9$). Our estimate of $\hat{r}$ was much higher than the observed number of transversions (5), an observation that is not surprising since no transversions were represented more than once in our collection. In combination, the total number of mutations in P3 estimated to allow growth on *glycinea* was approximately 55. P3 consists of 643 amino acids, and there are a total of 4380 potential non-synonymous changes possible in the

gene. This means that 55/4380, or 1.3% of non-synonymous mutations in P3 are predicted to confer the ability to grow on the novel host *glycinea*.

Note that the value of α used here is itself an estimate, and there is some degree of uncertainty associated with this estimate. However, since α was estimated from external data that will be published later, to incorporate the uncertainty in our estimate of α would require an extra layer of mathematical modeling and a complete discussion of the external data set. To stay on point and because it makes no difference to the interpretation of our results, we chose not to incorporate this uncertainty in our calculations of the confidence intervals surrounding $\hat{t}$ and $\hat{r}$. The estimates $\hat{t} = 11.9$ and $\hat{r} = 42.9$ would remain the same, the confidence intervals surrounding $\hat{t}$ would be affected only slightly, and the confidence intervals surrounding $\hat{r}$ were already sufficiently wide to indicate a low confidence in the exact estimate of *r*. In sum, we take our analysis to provide qualitative support for the intuition that many transversions were missed by our screen; our analysis does not indicate conclusively that exactly $\hat{t} + \hat{r} = 54.8$ mutations confer the ability to infect the novel host *glycinea*.

Mutation Characteristics

We investigated whether the observed mutations in P3 occurred in amino acid residues with specific chemical characteristics. We used a χ^2 test to compare the observed numbers of mutated residues which were acidic (6), basic (0), hydrophilic (2), or hydrophobic (3) to the expectation based on the amino acid composition of P3 (9.16% acidic, 8.69% basic, 24.53% hydrophilic, and 57.45% hydrophobic). The chemical properties of

amino acids that mutated differed significantly from the random expectation ($\chi^2 = 34.76$, df = 3, $p < 0.0001$), and resulted from the disproportionately high number of mutations that occurred in acidic residues.

Mutational effects

We measured the fitness of the 18 host range genotypes that resulted in different amino acid sequences (all genotypes except H) on the standard hosts *phaseolicola* and *japonica*, and the novel host *glycinea*. To narrow our focus to only the mutations that affected host range, we first compared the fitness of the genotypes with two mutations in P3 (Q and R) to that of the genotype with one of the two mutations (P). The fitness of these phages did not differ on any of the hosts ($p > 0.2$ by a t-test for all 6 comparisons), so we excluded mutants Q and R from all subsequent fitness analyses.

The fitnesses of the remaining 16 mutant genotypes and the wild type ϕ6 on the standard and novel hosts are shown in figure 1. Fitness improvements on *glycinea* were generally accompanied by fitness losses on *phaseolicola* (Fig 2.1A) and *japonica* (Fig 2.1B). To assess whether these losses were statistically significant we used ANOVAs to calculate the smallest difference between means required to achieve statistical significance – the least significant difference or LSD. 15/16 host range mutations imposed a significant fitness cost on *phaseolicola*, and 10/16 imposed a significant fitness cost on *japonica* (Figure 2.1; $p < 0.05$; 1-tailed LSD). If we, instead, use a Bonferoni correction to account for multiple comparisons (16 comparisons on each host), all but one of these comparisons remains significant.

We also examined whether the pleiotropic effects of mutations on *phaseolicola* and *japonica* were correlated with the direct effect of mutations on *glycinea*. The direct effects of mutations were not significantly correlated with pleiotropic effects on either *phaseolicola* (Pearson's $r = 0.2245$, df = 14, $p = 0.3704$) or *japonica* (Pearson's $r = 0.0466$, df = 14, $p = 0.8543$). However , there was a significant positive correlation between the pleiotropic effects on *phaseolicola* and the pleiotropic effects on *japonica* (Pearson's $r = 0.7452$, df = 14, $p = 0.0004$).

Phenotypic basis of host range expansion

The fact that most of the mutations responsible for growth on the novel host *glycinea* were found in the candidate gene P3 suggested host attachment as a candidate mechanistic basis of host range expansion. Measures of the attachment rate constants to *glycinea* for the wild type ϕ6 and the 16 focal mutants (genotypes H, Q and R were again excluded) are shown in Figure 2.2A. The mean attachment rate constant of the wild type ϕ6 was 7.35×10^{-13} (s.e.m. = 1.24×10^{-12}), a value that fell within the mutant genotype range of -1.53×10^{-12} to 6.77×10^{-12}. (Note that the lower bound is mechanically 0, but that negative values can result from error variance).

If host attachment was the mechanistic basis of host range expansion, we expect the attachment rates of mutant phage to differ from that of the wild type ϕ6, and we might also expect the attachment rates of mutant phage to differ from each other. However, using a one-way ANOVA to test for differences in attachment rate among these 17 phage genotypes, we

failed to find a significant difference ($F_{16,54} = 1.75$, $p = 0.0638$). This result suggests that there are few or no differences in attachment rate among these 17 genotypes, however, the ANOVA analysis was not designed to test specifically for a difference between the mutant phage and the wild type ϕ6.

Therefore, we performed a second analysis to more directly test the hypothesis that the attachment rates of mutant phage were higher than that of the wild type ϕ6. In this test, we resampled the mutant data to determine how often sampling effects, alone, could produce a mean attachment rate as low or lower than the attachment rate measured for ϕ6. The distribution of 10^6 resampled means is compared to the actual wild type ϕ6 mean in figure 2.2B. The proportion of resampled means that were lower than the actual mean was $p = 0.034$, confirming that the higher attachment rates observed in mutant phage relative to ϕ6 did not result by chance, but from a real increase in attachment rates in the mutant phage.

Discussion

In this study we investigated the frequency and nature of mutations that expand the host range of the bacteriophage ϕ6. Our results corroborate the finding of a recent study in ϕ6 [130], that host range expansion is usually, but not always, accompanied by a cost on the standard laboratory host, and expand on that finding in a number of ways. First, we identified 16 mutations in the host attachment gene P3, and predicted the existence of 39 additional mutations that confer the ability to infect the novel host *glycinea*. Second, we determined that costs of host range expansion were apparent, not only on the host to which ϕ6 was well adapted, but also on an alternative permissive host to which ϕ6 was not well

adapted. And third, we identified the phenotypic basis of host range expansion, and therefore the proximate cause of the fitness costs, as an increase in attachment rate to the novel host.

These data are particularly relevant to one of the central questions in ecology – understanding the factors that limit the ability of populations to colonize new environments. The answer to this question depends on the abundance of potential colonists, and the extent to which potential colonists are maladapted to novel hosts [19]. Although our results may address the extent of maladaptation on novel hosts, (absolute growth rate on *glycinea* was 10^4-fold lower than on *phaseolicola* over 24 hours, data not shown), we focus on the implications of our results for the abundance of potential colonists in natural populations. In particular, with an understanding of the mutation rate and number of mutations that expand host range, and of the abundance and magnitude of pleiotropic fitness costs, we make a population genetics prediction for the equilibrium frequency of potential colonists in natural ϕ6 populations.

Abundance of mutations that expand host range

Although one out of 40 mutants did not have a mutation in the host attachment gene P3, the presence of P3 mutations in the other 39 mutants in our collection provides strong evidence that the P3 mutations were responsible for the host range expansion. Indeed, 7 of the 18 observed P3 mutations were present in multiple mutants, ruling out any other possibility. Although 2 of these 18 mutations appeared together with another P3 mutation and could, therefore, be ruled out as the cause of host range expansion, there is strong reason to believe that the remaining 16 P3 mutations did cause the host range expansions. We

sequenced a total of 101,610 nucleotides in the mutant genomes and found only 2 second site mutations (one each in mutants Q and R). Assuming that the ϕ6 mutation rate is consistent across genes, we can infer that there were only 10 second site mutations spread among all 40 mutant genomes (2 second site mutations/101,610 sequenced bases*13,385 bases/genome*40 mutants). Thus, second site mutations appeared in a minority of genomes and do not make a likely alternative to our conclusion that the P3 mutations caused the host range expansions.

We estimated that 55 different nucleotide substitutions in P3 confer the ability to grow on the novel host *glycinea*. This number represents 1.3% of the possible non-synonymous mutations in P3. To our minds this estimate is surprisingly high. Imagine that 55 different mutations enabled avian influenza to infect and transmit between humans. It seems likely that the ease with which ϕ6 mutates to infect *glycinea* is particular to this virus-host pair, and we can think of two possible reasons why so many mutations confer the ability to grow on the novel host *glycinea*. First, the ability to grow on *glycinea* may be acquired through 'loss of character' mutations rather than 'gain of character' mutations. Second, the close relatedness of *glycinea* pathovars to our standard *phaseolicola* host [138] might mean that only slight modifications to P3 are required for growth on *glycinea*.

Although acquisition of the ability to infect a novel host can be thought of as a gain of function, our data suggest that infection of the novel host may be achieved by a proximate mechanism that entails loss of a character that prevents infection rather than gain of a character that allows infection. For instance, the ability to attach to the novel host may have resulted from loss of a structure that prevented attachment rather than gain of a structure that

enabled attachment. Consistent with this idea, charge loss contributed disproportionately to the observed amino acid substitutions. In addition, if the ability to grow on the novel host was acquired through loss of character mutations, it might explain why the effects of mutations on the standard hosts *phaseolicola* and *japonica* were correlated with each other, but uncorrelated with their effects on the novel host *glycinea*.

An alternative explanation for the large number of mutations that enable growth on *glycinea* is a close relatedness of the novel host, *glycinea*, and the standard host, *phaseolicola*. A recent 16s rRNA phylogeny [138] of *P. syringae* pathovars suggests that *glycinea* pathovars are closely related to our standard host *phaseolicola* (note that the *glycinea* pathovar used here was not examined in [138], but the 2 *glycinea* pathovars that were examined were both closely related to the *phaseolicola* host used here). ϕ6 infects its host through the type IV pilus [139] which is chromosomally encoded. Assuming that divergence in the pilus genes reflects divergence in 16s rRNA, the type IV pilus structures of *glycinea* and *phaseolicola* should be similar. In this case, only slight modifications to P3 may be required to bind to the similar type IV pilus of the novel host *glycinea*. If we had used a more distantly related novel host, we suspect that we would have observed fewer mutations capable of allowing growth on that host.

Abundance and Nature of pleiotropic fitness costs

Our results indicate that mutations that enable growth on a novel host are generally characterized by negative (antagonistic) pleiotropic effects for growth on standard (permissive) hosts. Although the prevalence and magnitude of negative pleiotropy differed

slightly between the two permissive hosts we examined, the general form of pleiotropic effects did not differ. Negative pleiotropy predominated on both *phaseolicola* and *japonica* despite a 10^7-fold difference in absolute fitness of the wild type ϕ6 on these two standard hosts (data not shown). The high frequency of negative pleiotropy among ours and a previous collection of mutations that expand host range in ϕ6 [130], provide consistent support for the expectation that adaptation to one host should generally be accompanied by loss of fitness on alternative hosts.

The consistency of negative pleiotropy among the mutations in our collection contrasts with the mixed results of previous studies of host range expansion in which individual mutations were approximately equally likely to exhibit positive and negative pleiotropy [52, 65, 71, 72, 90, 106, 110]. We suspect that the high frequency of AP among mutations that expanded host range in ϕ6 resulted in part because the proximate mechanism underlying host range expansion was an increase in the rate of attachment to the novel host. Acquiring the ability to attach to a novel host (a new function) is a common mechanism of host range expansion in viruses [140], probably because the host surface is more divergent than components of the host cytoplasm. Further adaptation to a novel host would likely involve adapting to less divergent host cell components and be less characterized by negative pleiotropy. A similar investigation of beneficial mutations in *E. coli* also indicated that the abundance and form of pleiotropic effects are highly dependent on the proximate mechanism of adaptation [141].

Implications for disease emergence

In this study, we investigated the genetic determinants of a major predictor of disease emergence in models of population ecology [129], the rate of transmission into the novel host. We focused on the scenario in which the ability to infect a novel host requires a mutation. In this case, transmission rate depends on the equilibrium frequency in a source population of genotypes capable of infecting the novel host (potential colonists).

By considering our data in a population genetics context, we can predict the equilibrium frequency of potential colonists in a source population growing on the standard host *phaseolicola*. Two forces act to determine the equilibrium frequency of potential colonists. Mutation will act to increase the frequency of potential colonists and, if the ability to infect the novel host imposes a pleiotropic fitness cost, selection will act to decrease the frequency of potential colonists. For individual mutations, the equilibrium frequency, $\hat{q}$, at which the two forces are exactly balanced is known from population genetics [142] to be $\hat{q} \approx \mu / s$, where μ is the mutation rate and s the selection coefficient on the standard host. We consider only the mutations in our collection that exhibited pleiotropic fitness costs on the standard host *phaseolicola*, and show the predicted equilibrium frequency of each mutation in figure 3. We used the selection coefficients measured on the standard host *phaseolicola*, and mutation rate estimates of $\mu_{ti} = 1.9 \times 10^{-6}$ for transitions and $\mu_{tv} = 1.5 \times 10^{-7}$ for transversions, both of which were measured in another study [39]. It is clear from these data that the distribution of negative pleiotropic effects among mutations in our collection does not precisely predict the distribution of negative pleiotropic effects among host range mutations segregating in natural phage populations. In particular, mutations with large negative pleiotropic effects were reasonably common in our collection. However, the

strength of selection acting against such mutations is expected to keep them at a relatively low frequency in natural populations.

In addition to predicting the equilibrium frequencies of individual mutations, we used our data to predict the total equilibrium frequency of host range mutations in a source population growing on the standard host *phaseolicola*, $\hat{f}_{HR}$, by summing the equilibrium frequencies over all the mutations in figure 2.3. In this manner, we estimated an equilibrium frequency of mutations that enable infection of the novel host to be $\hat{f}_{HR} = 3\text{x}10^{-4}$. Note that this equilibrium frequency is slightly underestimated because it does not include the (~39) mutations missed by our screen, but it is only slightly underestimated because most of the missed mutations were transversions. Notice that the estimate of $\hat{f}_{HR} = 3\text{x}10^{-4}$ is well above the phage mutation rate ($\mu \approx 2\text{x}10^{-6}$ [35]) because several of the mutations in our collection exhibit very small costs. 38% of the host range mutations present in equilibrium populations are expected to exhibit fitness costs on the standard host *phaseolicola* of less than 5% (e.g. $s < 0.05$).

The substantial variation in pleiotropic fitness costs observed here among mutations that expand host range may explain the observation in viruses and other host specialists that performance tradeoffs among hosts are more common in laboratory populations [124, 143] than in natural populations [144]. Adaptation in laboratory populations of microbes generally occurs via selection acting on novel mutations, whereas adaptation in natural populations should more often occur via selection acting on standing genetic variation (i.e. mutations present in populations that are at an equilibrium between mutation and selection).

If there is variation in pleiotropic fitness costs, then novel mutations will be characterized by larger pleiotropic costs on average than the standing genetic variation, and laboratory populations would be expected to exhibit larger fitness tradeoffs among hosts than natural populations. In other words, if pleiotropic fitness costs are not universal among the mutations that expand host range, we should not expect to observe large fitness tradeoffs in nature. Adaptation to a novel host need not impose fitness costs on the standard host, at least in the short term.

Finally, our data suggest an alternative to the accepted explanation for why RNA viruses are the major contributor to emerging disease. The accepted explanation is that the high mutation rate characteristic of RNA viruses allows adaptation to a novel host *after* the initial transmission [10]. We posit that the high mutation rate of RNA viruses ensures the existence of a high frequency of mutations that allow colonization of a novel host *before* the initial transmission. We note that this explanation is consistent with the accepted explanation for the rapid evolution of drug resistance in RNA viruses such as HIV. Drug resistance evolves rapidly because mutations that confer resistance are circulating in the viral population before the drug is administered [145, 146].

Table 2.1. Sequence changes in the attachment gene P3 of host range mutants.

mutant ID	nt mutation[a]	aa mutation[a]	# in collection	aa property[b]
A	None	None	1	Unknown
B	g13a	G5S	2	Hydrophillic
C	g22a	E8K	1	Acidic
D	a23g	E8G	5	Acidic
E	a434g	D145G	3	Acidic
F	a437g	N146S	6	Hydrophillic
G	g534c	E178D	1	Acidic
H	g534t	E178D	1	Acidic
I	c1016t	P339H	1	Hydrophobic
J	a1546g	T516A	4	Hydrophillic
K	a1598c	D533A	1	Acidic
L	g1603a	D535N	1	Acidic
M	a1661t	D554V	1	Acidic
N	g1660a	D554N	2	Acidic
O	a1661c	D554A	1	Acidic
P	a1661g	D554G	6	Acidic
Q	a1661g (& t779a)	D554G (& F260Y)	1	Acidic (& Hydrophobic)
R	a1661g (& c318t)	D554G (& L106L)	1	Acidic (N/A)
S	c1663t	L555F	1	Hydrophobic

[a] Nucleotide and amino acid substitutions are labeled according to their position in P3.

Second site mutations are shown in parentheses.

[b] Amino acid chemical properties correspond to the wild type residue.

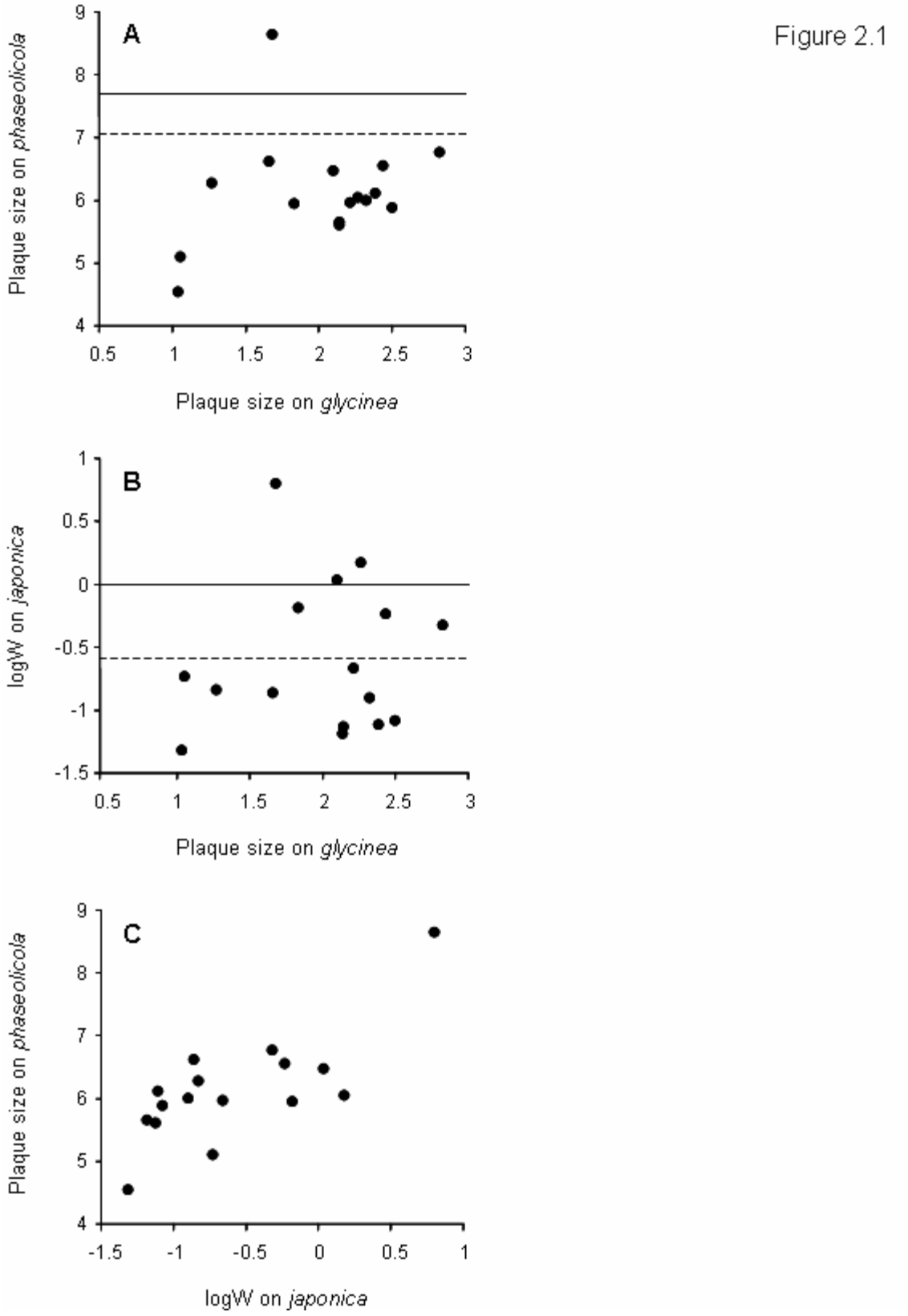

A
Plaque size on *phaseolicola*
Plaque size on *glycinea*
B
logW on *japonica*
Plaque size on *glycinea*
C
Plaque size on *phaseolicola*
logW on *japonica*

Figure 2.1

Figure 2.1: Correlations between the effects of mutations on different hosts. **(A)** and **(B)** illustrate the relationship between the fitness of mutant phage on the novel host *glycinea* and one of two standard hosts (*phaseolicola* or *japonica*). Data points are means of 4 replicate measures on host *japonica*, and means of 6 replicate measures on hosts *glycinea* and *phaseolicola*. The solid black lines are the mean value of the wild type ϕ6 on the standard host, and the dashed lines correspond to the value below which mutants are significantly lower than wild type (the LSD). **(C)** illustrates the relationship between fitness on the two standard hosts. To generate the data for these figures, fitness was measured using either a plaque size assay or a paired growth assay (relative growth rate = W). Plaque area increases linearly with log$W$.

Figure 2.2

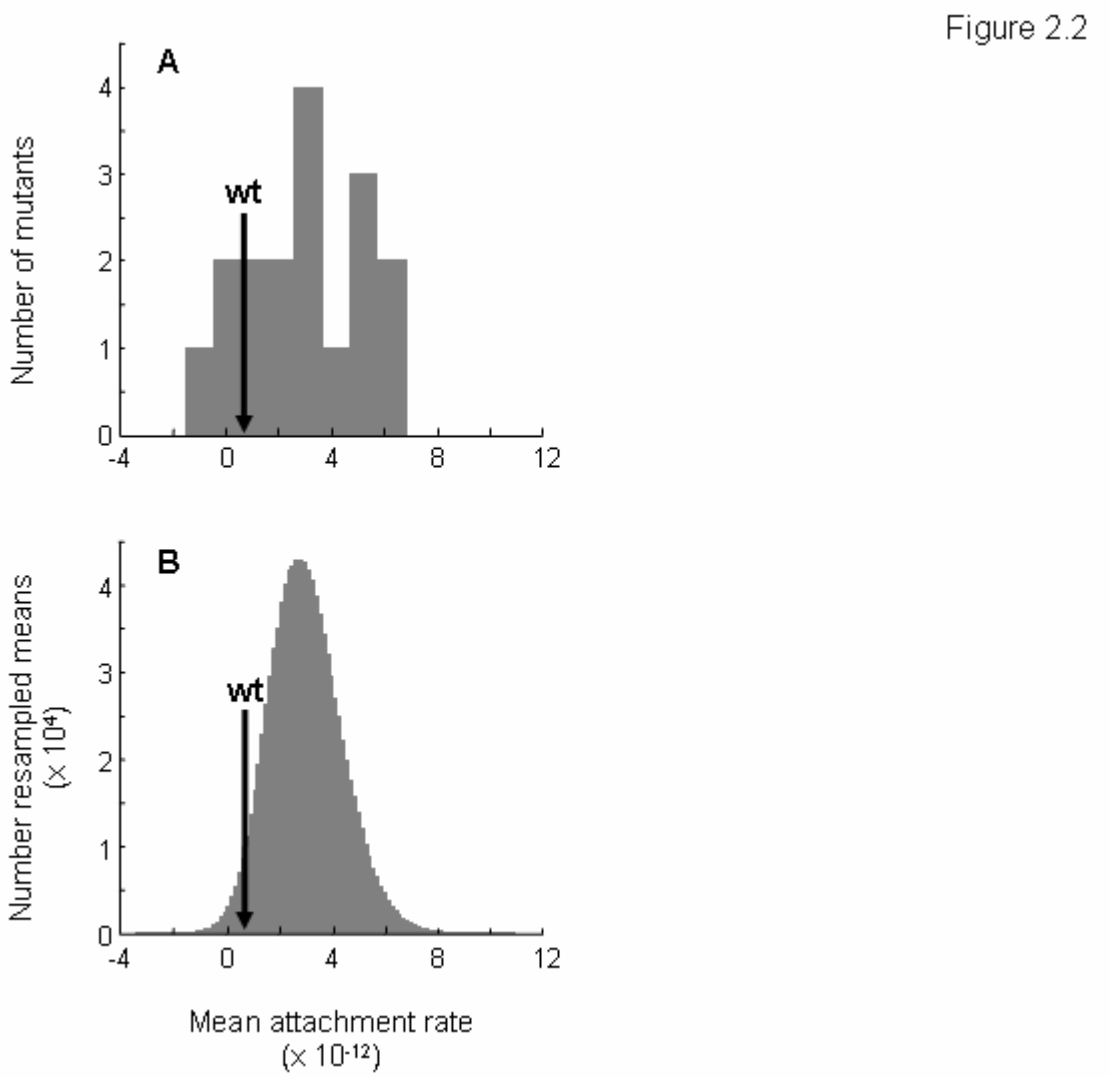


Figure 2.2: Attachment to *glycinea*. **(A)** Frequency distribution of the mean mutant attachment constant to *glycinea*. Mutant means are based on n=4 replicates. The mean wild type attachment constant's bin is indicated by an arrow. The wild type mean is based on 8 replicates. **B)** Histogram of the distribution of mean attachment rate constants resampled from the mutant attachment data. Each resampled mean is created by a draw of n=8 measured attachment rate constant values from the mutant attachment data. The red arrow indicates the measured mean wild type attachment rate constant. A proportion, p = 0.0342, of the resampled means lie to the left of the measured value.

Figure 2.3

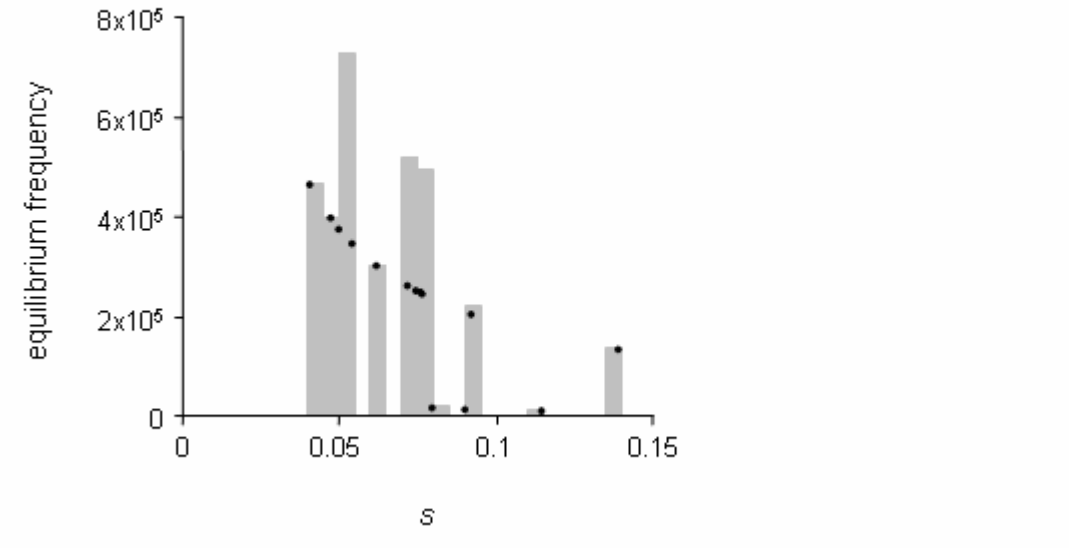


Figure 2.3: Predicted equilibrium frequencies of host range mutations before an encounter with the novel host. Equilibrium frequencies for each of the observed mutations were calculated as $\hat{q} = \mu / s$, where μ is either the transition or transversion mutation rate, and s is the observed fitness cost on the standard host *phaseolicola*. Dots indicate the equilibrium frequency of the individual observed mutations. Grey bars depict these data as a histogram by collecting the individual mutations into bins of width 0.005.

Chapter 3: Genetic divergence between a virus' original and novel host type predicts the characteristics of mutations which expand host range.

The work described in this chapter was accomplished in collaboration with Dr. Christina Burch. I would like to thank Siobain Duffy, Amanda Ferris, Sarah Joseph, and members of the Burch lab for discussions that improved both the experimental design and the final manuscript.

Abstract

I investigate the characteristics of mutations that expand ϕ6 host range to include a number of novel host types. I specifically investigate whether the genetic determinants of the ability of ϕ6 to colonize and persist on three different novel *P. syringae* hosts were predicted by the relationship between the original host of ϕ6, and the three novel hosts. I compared collections of host range mutants isolated on each of three novel host types which differed in their genetic distance from the original host of ϕ6. I found that multiple mutations allowed growth on each novel host type, that many of these mutations imposed a fitness cost on ϕ6 on its original host, and that these mutations had low fitness on the novel host from which they were isolated. I also found that these three characteristics differed depending on their novel host of isolation. Finally, I determined that the genetic distance between the original and novel host was predictive of the genetic determinants of both ϕ6's ability to colonize and persist on a novel host.

Introduction

The ongoing threat of emerging viral disease in novel host populations, such as the recent cases of HIV [5], West Nile Virus [147], and SARS Coronavirus [148] emergence into human populations, highlights the need to understand conditions that facilitate successful virus emergence [1-4, 8, 9, 149]. In the scenario where the wild type virus cannot infect a novel host, both the transmission of viruses with an expanded host range to this novel host [12], as well as the persistence of these viruses in this novel host population [10, 16, 150] are critical components of successful emergence. An accurate determination of emergence risk

depends on understanding how these two components predictably differ across novel host types.

Both transmission and persistence should be determined, in part, by the genetic similarity between a virus' original and novel host (hereafter, host similarity). Specifically, as host similarity decreases, the phenotypic similarity of these hosts will also decrease [151]. This decreasing phenotypic similarity should lead to both a decrease in the transmission of viruses with an expanded host range to a novel host population, and a decrease in the persistence of these viruses in this novel host population. These two expectations arise from considering the functional effects that host similarity should have on three characteristics of the mutations that expand host range: the number of mutations capable of allowing growth on a novel host, the fitness costs these mutations confer to a virus on its original host, and the absolute fitness viruses with these mutations have on the novel host.

Both the number of mutations that expand a virus' host range and the fitness costs these mutations confer to a virus on its original host will determine the frequency of viruses with an expanded host range in a population on the original host. This frequency will determine the rate of transmission of viruses with an expanded host range to a novel host population. As host similarity decreases, mutations that expand a virus' host range will have to cause larger changes in viral phenotype to allow infection of the novel host. There is a conceptual expectation that as the effect size of adaptive mutations increase, the number of mutations capable of causing such a change should decrease [152]. Furthermore, larger changes caused by a host range mutation should have a larger effect on that virus' interaction

with its' original host. This increase in effect size should increase the costs these mutations have on the original host. Therefore, with decreasing host similarity there should be fewer mutations which expand host range and these mutations should have greater costs on the original host. Together these results will lower virus transmission to a novel host type.

Once viral host range mutants are transmitted to a novel host, their persistence will depend on their absolute fitness, the ability to successfully produce progeny on that novel host [10, 150]. Progeny production is determined by the success of host range mutants in interacting with a number of novel host components. As host similarity decreases, host range mutants should be less successful in interacting with these novel host components. With a decreasing host similarity, host range mutants should produce fewer progeny, which will lower their persistence in a novel host population.

Evidence addressing the relationship between host similarity and the transmission and persistence of viruses with an expanded host range is both limited and indirect. Broad examinations of disease emergence and the host ranges of pathogens provide conflicting evidence on the importance of host similarity in emergence. A study of plant pathogenic fungi found that the ability of fungus to infect two host plants decreased as the genetic similarity between the two hosts decreased [153]. In contrast, a study of human pathogens [4] found that diseases emerging in the human population were not associated with any particular animal hosts more often than any other. Related viral species are also known to utilize entirely different host receptors [140], suggesting an evolutionary lability in receptor usage. However, all of these studies are unable to disentangle the evolutionary (e.g. the

genetic predisposition for host expansion) and ecological (e.g. the frequencies of different hosts in an area) factors which have interacted to determine current host ranges. As a result, these broad scale studies have only limited power to address the effects of host similarity on the characteristics of host range mutations.

Laboratory investigations of the genetic basis of viral host range have provided data on some of the characteristics of host range mutations that are important in viral emergence [26, 45, 65, 79, 90, 96, 109, 110, 112, 113]. These studies have identified mutations that expanded viral host range, and measured the fitness effects of these mutations on a number of previously permissive hosts, as well as on the novel host. However, by only examining the effects of one or a few mutations for a single virus-novel host pair and not a virus paired with many novel hosts, these studies cannot be used to investigate how host similarity affects the characteristics of host range mutants.

To investigate the effect of host similarity on the characteristics of host range mutations, we isolated collections of ϕ6 host range mutants on each of three novel host types with differing genetic similarities to ϕ6's original host, *Pseudomonas syringae* pathovar *phaseolicola*. We used these three collections of host range mutations to assess the number of mutations that allowed growth on each novel host type, the costs these mutations confer to a virus on the original host of ϕ6, and the absolute fitness of viruses with these mutations on their novel host. We found that these characteristics differed depending on the novel host, and that host similarity was often predictive of these characteristics.

Results

We screened for the ability of ϕ6 to produce host range mutants on 36 pathovars of the bacteria *Pseudomonas syringae*, and we identified three hosts on which ϕ6 produced host range mutants: *P. syringae* pvs. *glycinea*, *syringae*, and *tomato* (hereafter, all hosts will be referred to by their pathovar names). We determined the genetic similarity between the original host of ϕ6, *phaseolicola*, and these three novel hosts by sequencing a total of 2022 bases of each host from four loci (*acn*, *cts*, *gapA*, and *pilA*). The housekeeping genes *acn, cts,* and *gapA* have been previously used to determine the core genomic relationship between strains of *Pseudomonas syringae* [138]. The pilus protein encoded by *pilA* is required for ϕ6 infection [139], and mutations which expand the host range of ϕ6 do so by allowing attachment to the pilus of the novel host [128]. We found that the three novel hosts differed in their genetic similarity to *phaseolicola*. We also found that both *pilA* and the three housekeeping genes provided the same qualitative relationships between the novel hosts and *phaseolicola*, but these measures of genetic similarity were quantitatively different (Table 3.1).

We then investigated the characteristics of host range mutations by isolating 40 independently derived host range mutants on both of the novel hosts *syringae* and *tomato*, and combining these data with 40 independently derived host range mutants on the novel host *glycinea*, which had been collected for a prior study [128].

Identification of mutations expanding host range

Previous studies [128, 130] have identified the host attachment spike P3 as the main determinant of host range in ϕ6. We therefore sequenced the coding region of P3 to identify the unique host range mutations in our collection. We identified a total of 30 unique P3 genotypes: 18 in our *glycinea* collection, 5 in our *tomato* collection, 3 in our *syringae* collection, and 4 shared between host types (Table 3.2). Of the 4 that occurred in multiple collections, one was present in both the *glycinea* and *tomato* collections, and 3 were present in both the *tomato* and *syringae* collections. One genotype had no mutations in P3, 20 genotypes had 1 mutation in P3, 6 genotypes had 2 mutations in P3, and one genotype had 3 mutations in P3. Each of the 2 mutation genotypes shared a mutation with a single mutation genotype, suggesting that the mutations present in these single mutation genotypes are responsible for the expanded host range of the 2 mutation genotypes. The 3 mutation genotype did not share any of its mutations with another genotype. However one of its' mutations caused a substitution in an amino acid residue that was changed in other genotypes, marking this change as a likely candidate for the expanded host range of the 3 mutation genotype. Overall, these results are consistent with 17 mutations allowing growth on *glycinea* (16 occurring in P3), 6 allowing growth on *tomato*, and 5 allowing growth on *syringae* (Figure 3.1A).

Host range of isolated mutants

Four P3 mutations were isolated on more than one novel host, suggesting that some mutations allowed growth on more than one novel host. We assayed the P3 genotypes from each novel host collection for the ability to grow on each of the three novel hosts (Table 3.2). Most mutations only allowed growth on their novel host of isolation. Mutations a23g

(isolated in both the *glycinea* and *tomato* sets) and g1660t (isolated in the *tomato* set) allowed growth on both *glycinea* and *tomato*. Mutations g390c, g390t, and g1675a were isolated in both the *syringae* and *tomato* sets, however, these mutants were only able to grow on the specific host they were isolated on. That is, g390c isolated on *tomato* only grew on *tomato*, while g390c isolated on *syringae* only grew on *syringae*.

Fitness costs of host range mutations on original host

We wanted to determine which host range mutations conferred a significant fitness cost to ϕ6 on the original host *phaseolicola*. In order to do so, we compared the fitness of each unique host range mutation to the wild type ϕ6 on *phaseolicola* (Figure 3.1B). Fitness was assayed by either measuring average plaque area (*glycinea* collection) or by using paired growth assays (*syringae* and *tomato* collections). Plaque area provides a one generation measure of viral fitness, whereas paired growth assays provide a 5 generation measure of fitness.

We measured the plaque size of 16 of the unique host range mutations we isolated on *glycinea* as well as the wild type ϕ6, and used an ANOVA to calculate the Least Significant Difference (LSD) for these data (gly8 was measured alongside wild type ϕ6 separately). The LSD is the smallest difference between two means that is statistically significant. We determined which host range mutations isolated on *glycinea* had a mean cost on *phaseolicola* that was greater than the LSD when compared to wild type ϕ6. We found that 14 of these 16 (all but gly5 and gly7) mutations had a significant fitness cost on *phaseolicola*. Bonferroni correcting the LSD resulted in 13 of these 14 mutations (all but gly6) remaining significant.

The gly8 genotype did not show a significant fitness cost, as it had higher fitness than the wild type ϕ6 on *phaseolicola*.

Paired growth assays were used to measure the fitness of the unique host range mutations isolated on *syringae* and *tomato* relative to the wild type ϕ6. Each measure from a paired growth assay gives the log(relative fitness) of a mutation compared to the wild type ϕ6. Therefore, for these data we are not trying to compare two means to each other (as we did with the LSD), but rather the mean relative fitness of a mutation to zero, an equal fitness of the mutant and wild type ϕ6 on a log scale. We ran separate ANOVAs for mutations isolated on *syringae* and *tomato*, and calculated experiment-wise confidence limits (equivalent to the LSD) based on these ANOVAs. Mutations whose confidence limits did not encompass 0 had a significant fitness costs on *phaseolicola*. We found that 2 of the 6 mutations isolated on *tomato* (tom1, tom8) and 3 of the 5 mutations isolated on *syringae* (syr1, syr3, and syr6) had a significant fitness cost on *phaseolicola*. Bonferroni correcting these data did not change this result for *tomato* mutations, but resulted in none of the mutations isolated on *syringae* showing a significant fitness cost on *phaseolicola*.

We also determined the average selection coefficient, *s*, acting against the host range mutations on *phaseolicola*. Mutations isolated on *glycinea* had the greatest costs on *phaseolicola* ($\bar{s}$ =-0.062), followed by mutations isolated on *tomato* ($\bar{s}$ =-0.027), and then mutations isolated on *syringae* ($\bar{s}$ =-0.018). To determine whether costs on *phaseolicola* differed depending on the novel host of isolation, we conducted a series of 2-tailed t-tests between the *s* values of host range mutations isolated on each of the three novel hosts. Costs

differed between mutations isolated on *glycinea* and *syringae* (t_{21}=2.146, p=0.044), but did not differ between mutations isolated on *glycinea* and *tomato* (t_{22}=1.675, p=0.108), or between mutations isolated on *tomato* and *syringae* (t_{10}=0.446, p=0.665).

Fitness of host range mutants on novel host of isolation

We measured the absolute fitness, *W*, of each unique host range mutation on their novel host of isolation. The measure of absolute fitness we used was the log(phage/plaque) produced after 24 hours, a measure of the ability of a virus to infect host cells and produce progeny over ~ 5 generations (Figure 3.1C). The absolute fitness of these mutants on the novel hosts was at least one order of magnitude lower than that of wild type ϕ6 on *phaseolicola* (unpbl. data). Mutations isolated on *glycinea* had the highest absolute fitness ($\log \overline{W}$ =6.821313), followed by mutations isolated on *syringae* ($\log \overline{W}$ =6.277487), and then mutations isolated on *tomato* ($\log \overline{W}$ =5.978978). We then conducted a series of 2-tailed t-tests between the log(*W*) values of host range mutations isolated on each of the three novel hosts. Absolute fitness differed between mutations isolated on *glycinea* and *tomato* (t_{22}=3.965, p<0.001), as well as mutations isolated on *glycinea* and *syringae* (t_{21}=2.203, p=0.039), but did not differ between mutations isolated on *tomato* and *syringae* (t_{10}=1.466, p=0.173).

Discussion

In this study, we investigated several characteristics of host range mutations of the bacteriophage ϕ6 that allowed growth on one of three novel host types: *glycinea*, *tomato*, or *syringae*. Consistent with previous studies of host range expansion in ϕ6 [128, 130], we

found that multiple mutations are capable of allowing growth on a novel host type; that many, but not all host range mutations cause a reduction in viral fitness on the original host of ϕ6, *phaseolicola*; and that host range mutants are maladapted to their novel host types. We expanded upon these findings by showing that these characteristics differed depending on the novel host of isolation.

Expansion of RNA virus host range

The high mutation rates of RNA viruses have been identified as a likely explanation for the prevalence of emerging RNA viruses [7]. A high mutation rate allows an RNA virus population to rapidly explore its mutational neighborhood and produce host range mutants. Although high mutation rates should increase the *rate* at which RNA viruses produce host range mutants compared to DNA viruses, high mutation rates by themselves should not lead to any qualitative differences in the *ability* to generate host range mutants between DNA and RNA viruses.

However, RNA viruses would be expected to have a greater *ability* to evolve an expanded host range if the ability to infect a new host requires two mutations in combination. We believe we isolated several epistatic combinations of mutations which expand the host range of ϕ6. We identified three unique P3 genotypes (containing one of three mutations: g390c, g390t, or g1675a) which had different host ranges depending on their host of isolation. All three of these genotypes were isolated on both *tomato* and *syringae*, but could only grow on their host of isolation (e.g. g1675a isolated on *syringae* could only grow on *syringae* and not *tomato*, and vice versa). As each of these three genotypes was

phenotypically different depending on its host of isolation, it follows that they also must be genetically different, with an additional mutation occurring in a gene other than P3.

The simplest explanation for this observed pattern is that the 3 isolated P3 mutations have no effect on host range (i.e. they are neutral). However, this explanation is inconsistent with the high frequency of these mutations in our collection (g390c/t was isolated 37 times, g1675a was isolated 27 times). In addition, for each of these three mutations, the mutation was isolated many times on one novel host, rarely on a second, and never on a third (e.g. g1675a was isolated 25 times on *syringae*, 2 times on *tomato*, and never on *glycinea*). If these three mutations were neutral and their high frequency was explained by the presence of mutational hotspots, they should be distributed randomly across the three mutant collections. Finally, if these mutations were neutral it would mean that a second non-P3 mutation was responsible for the expanded host range phenotypes. If non-P3 mutations were capable of expanding host range to include *syringae* and *tomato*, we would expect to have isolated host range mutants on *syringae* and *tomato* with no mutations in P3, as we did in our *glycinea* collection. We did not find any such genotypes in our *syringae* and *tomato* collections.

A more plausible explanation for the observation that g1675a and g390c/t mutants isolated on *syringae* only grow on *syringae* and g1675a and g390c/t mutants isolated on *tomato* grow only on *tomato* is that these mutations affect host range, but the way they affect host range depends on the presence or absence of a 2nd mutation not in P3. Because g390c/t mutations occurred 35 times in the *tomato* set, we argue that a g390c/t mutation allows growth on *tomato* when it occurs alone. In this case, the fact that g390c/t mutants isolated on

syringae don't grow on *tomato* can only be explained by the presence of a second mutation not in P3 that interacts with a g390c/t mutation to enable infection of *syringae* and prevent infection of *tomato*. Similarly, the g1675a mutation occurred 25 times in the *syringae* set. Therefore, we suspect that the g1675a mutation allows growth on *syringae* when it occurs alone, and we suspect that a second mutation not in P3 interacts with g1675a to allow infection of *tomato* but prevent infection of *syringae* (Figure 3.2).

As host similarity decreases so much that single mutations are not sufficient to allow growth on a novel host, the epistatic effect of two mutations might allow growth on this novel host type. Due to their high mutation rates, RNA viruses will be able to sample from these double mutation combinations whereas DNA viruses will largely be prevented from sampling these combinations. This ability of RNA viruses might provide an explanation for the evolutionary lability of many animal viruses in their receptor usage [140]. Together with an increased likelihood of pleiotropy due to the prevalence of overlapping reading frames in RNA virus genomes [7], the ability of RNA viruses to create double mutant combinations presents a unique consideration of RNA virus emergence.

Implications for disease emergence

The process of disease emergence is complex and relies on the interaction of a number of evolutionary and ecological factors, as well as stochastic events. This makes predictions of emergence risk difficult [1, 3, 4, 8]. Based on functional considerations, we hypothesized that a decreasing genetic similarity between the original host of ϕ6 and a novel host would lead to a decreasing number of mutations that expanded host range, an increase in

the cost these mutations confer to a virus on the original host, and a decreasing absolute fitness of a virus with these mutations on the novel host, all critical components of emergence. To address these hypotheses, we took two measures of the genetic similarity between host types. These measures agreed in their ordering of novel host similarity to the original host (the original host, *phaseolicola* was most similar to *glycinea*, then *tomato*, then *syringae*), so we discuss our results as they relate to this ordering.

Consistent with our predictions, the number of mutations that expand ϕ6 host range decreased as host similarity decreased, while inconsistent with our predictions, the costs these host range mutations conferred to ϕ6 on its original host *phaseolicola*, did not increase as host similarity decreased. In fact, costs decreased as host similarity decreased. This deviation from our expectation indicates that genetic distance was not related to phenotypic distance in a straightforward and continuous way (i.e. as a continuous environmental variable, such as temperature). Instead, independently evolving host lineages should accumulate changes that affect different aspects of the host pilus. Under this more complex framework, we do not have an *a priori* assumption for predicting how the costs of host range mutations on *phaseolicola* would differ depending on the novel host of isolation. For example, it would be difficult to predict fitness costs if the difference between *phaseolicola* and *glycinea* is in protein charge, while the difference between *phaseolicola* and *tomato* is in protein conformation. Future attempts to gain predictive power of the costs that host range mutations confer to a virus on its original host will require a better understanding of how hosts differ.

As the change in the number of host range mutations was consistent with our hypothesis while the change in costs went against our expectations, it is worth considering how these two patterns will act on the frequency of host range mutations existing in a population on the original host *phaseolicola*. As in our previous study [128], we can use a population genetics context to predict this frequency. For an individual haploid mutation, the equilibrium frequency, $\hat{q}$, at which mutation and selection are exactly balanced is known from population genetics [142] to be $\hat{q} \approx \mu / s$, where μ is the mutation rate and *s* the selection coefficient on the original host *phaseolicola*. The overall equilibrium frequency of host range mutants able to grow on a given novel host will be the summation of the frequency of each individual mutation we isolated on that novel host.

For this analysis, we consider those mutations in our collection that a) exhibited fitness costs on the original host *phaseolicola*, and b) had their host range mutation identified (i.e. not the double mutations in the *syringae* and *tomato* collections or the unidentified mutation in the *glycinea* collection). We used the selection coefficients we measured on the original host *phaseolicola*, and mutation rate estimates of $\mu_{ti} = 1.9 \times 10^{-6}$ for transitions and $\mu_{tv} = 1.5 \times 10^{-7}$ for transversions, both of which were measured in another study [39] to calculate the expected equilibrium frequency of each individual mutation. We then summed the equilibrium frequencies of mutations isolated on each novel host type to give us $\hat{f}_{HR}$, the equilibrium frequency of host range mutants for each of the three novel hosts (Figure 3.3).

Despite the opposing patterns of change between the number of host range mutations and their costs, we can see that a decreasing host similarity leads to a decreasing equilibrium

frequency of host range mutants in a population on the original host *phaseolicola*. This result is not merely due to the exclusion of the double mutant genotypes from the *syringae* and *tomato* data sets. Even though these double mutants had small costs, their frequency should be several orders of magnitude lower than single mutants due to the low rate of double mutant production. Rather, the decrease in the equilibrium frequency of host range mutants is primarily driven by the decrease in the number of mutations capable of expanding host range. Therefore, as host similarity decreases, the frequency of host range mutants in a population on the original host will decrease, leading to reduced transmission of viral colonists to the novel host.

Our measures of the absolute fitness of host range mutants on their novel hosts were consistent with our expectation in that mutations on the two more distant novel hosts, *tomato* (core divergence = 0.0703) and *syringae* (core divergence = 0.0705) had lower absolute fitness than the mutations on the closest novel host, *glycinea* (core divergence = 0.0117). We note that, although not significant, the absolute fitness of mutants on *syringae* and *tomato* are reversed from our expectations (mutants on *syringae*, the least similar host, have a higher absolute fitness than mutants on *tomato*). *Syringae* and *tomato* have almost identical core genetic similarities to the original host, while being quite divergent from each other (core divergence = 0.08). This suggests that the difference in the absolute fitness of mutants on these two novel hosts is due to the independent evolutionary histories of these hosts. In general, viruses with an increased absolute fitness on a novel host will persist on that novel host for longer periods of time. Our finding suggests that viruses will have a higher absolute fitness on novel hosts which are more similar to their original host.

Previous examinations of disease emergence into human populations have found no relationship between host similarity and emergence [4, 9]. However, such studies have been unable to disentangle the interaction between evolutionary and ecological factors which contribute to emergence [1, 2]. By using a genetically tractable experimental system, we were able to show that the similarity between a virus' original and novel host appears to have an effect on two characteristics critical for emergence: the genetic potential for transmission to a novel host and the ability to persist in a novel host environment.

Materials and Methods

Strains and Culture Conditions

The double-stranded RNA bacteriophage ϕ6 (*Cystoviridae*) used in this study is a laboratory genotype descended from the original isolate [131]. The standard laboratory host of ϕ6, *Pseudomonas syringae* pathovar *phaseolicola* strain HB10Y, was obtained from the American Type Culture Collection (ATCC; no. 21781); the novel host *P. syringae* pathovar *glycinea* strain R4a was obtained from J. Dangl (University of North Carolina, Chapel Hill, NC); the novel host *P. syringae* pathovar *syringae* strain FF5 was obtained from L. Chao (University of California at San Diego, San Diego, CA); and the novel host *P. syringae* pathovar *tomato* strain Bakersfield was obtained from G. Martin (Cornell University, Ithaca, NY). Details of diluting, filtering, culture, and storage of phage and bacteria are published [27, 33]. Phage and bacteria were cultured in LC media (5 g yeast extract, 5 g NaCl, and 10 g Bacto-tryptone per L of H_2O), and stored in 2:3 glycerol:LC media (v/v) at -20^{o}C and -80^{o}C,

respectively. For growth on plates, phage were mixed with the appropriate host bacteria in LC top agar (0.7% agar) and plated on LC plates (1.5% agar).

Mutant Isolation

Isolation protocols have been previously described [128]. Briefly: ϕ6 clones were plated onto a lawn of 200 μl of the original host *phaseolicola*, and incubated overnight to allow the phages to reproduce and form plaques. After 24 hours, phages were harvested from a randomly chosen isolated plaque and plated onto a fresh lawn of 200 μL of a stationary phase culture of one of the novel hosts *syringae* or *tomato*. Only phages that acquired a host range mutation during growth of the plaque on *phaseolicola* form plaques on a plate of a novel host. After 24 hours, an isolated plaque was chosen randomly from a novel host plate and phages from this plaque were plated on a fresh lawn of the same novel host to purify the mutant phages. A single plaque was harvested and stored for later use in 2:3 glycerol:LC media (v/v) at -20° C. This process was repeated 80 times to obtain 40 independent host range mutants on the novel hosts *syringae* and *tomato*.

Phage sequencing

Genome amplification and sequencing were performed as previously described [130]. Briefly, phages were grown to a high titer and viral RNA was extracted using a QIAamp Viral RNA Mini Kit (QIAGEN Inc., Valencia, CA). Viral RNA was reverse transcribed using random hexamer primers and Superscript II RNase H- RT (Invitrogen, Carlsbad, CA), and the resulting cDNA was used as template for PCR with Taq Polymerase (Invitrogen, Carlsbad, CA). We amplified three sections of the medium genome segment, corresponding

to bases 1298-2142, 2042-3052, and 2877-3873, which encompassed the host attachment gene, P3. PCR products were purified using EXO-SapIT (US Biological, Swampscott, MA) and sequenced in both directions using PCR primers and primers internal to each amplicon. Sequencing was performed using BigDye v3.1 (Applied Biosystems, Foster City, CA) either locally with an Avant-3100 Genetic Analyzer Sequencer (Applied Biosystems, Foster City, CA) or at the UNC Automated Sequencing Facility (University of North Carolina, Chapel Hill, NC).

Bacterial sequencing

We sequenced portions of four *P. syringae* genes in each of the four host strains used in this study, to determine their genetic similarity. The housekeeping genes *acn*, *cts*, and *gapA* were used to determine the core genome similarities between strains, as in [138]. *pilA* was used to determine pilus similarity between host strains, as ϕ6 host range is determined at attachment to the host pilus [128, 139]. PCR and sequencing primers for the three housekeeping genes was taken from [138], and we designed degenerate primers using SCPrimer [154] for *pilA* based on the published sequences of *P. syringae pvs. tomato* DC3000 (Genbank Accession: NC_004578) *, syringae* B728a (NC_007005)*, and phaseolicola* 1448A (NC_005773). PCR and sequencing reactions were performed as described above, substituting a bacterial colony in place of cDNA for the initial PCR reaction and adding 0.2 μl of RNAse One (Promega, Madison, WI) to each 20 μl PCR reaction.

Mutant host range

We assayed the ability of the isolated host range mutants to grow on all three novel host types (e.g. the host range of each mutant). For each unique P3 genotype isolated on a given novel host type, we plated a sample of 10^3 and 10^2 plaque forming units (based on titers on *phaseolicola*) onto LC plates with 200 uL mixed lawns of either *phaseolicola*:*glycinea* (2:1), *phaseolicola*:*tomato*(20:1), or *phaseolicola*:*syringae* (100:1). These plates were incubated overnight at 25°C, and each genotype was then scored using an assay that determined whether clear (phages utilized both *phaseolicola* and the novel host) or turbid (phages utilized only the host *phaseolicola*) plaques formed.

Paired growth fitness assay

Paired growth assays are a standard method for assaying fitness in ϕ6 [32], and were used to assay fitness on the original host *phaseolicola* for host range mutants isolated on *syringae* and *tomato*. Each host range mutant was mixed with the wild type ϕ6 at a 1:1 ratio. This mixture was plated on a bacterial lawn and incubated for 24 hours. The ratio of phage genotypes before and after the incubation was determined by plating on a mixed lawn of 200 μL of a mixture of *phaseolicola* and either *tomato* (20:1 *phaseolicola*:*tomato*) or *syringae* (100:1 *phaseolicola*:*syringae*), on which the wild type ϕ6 forms turbid plaques and mutant ϕ6 form clear plaques. The relative fitness of mutant genotypes were then determined as $W = R_1/R_0$, where R_0 and R_1 are the ratio of mutant to wild type phages before and after the 24 hour incubation, respectively. These data were then log transformed to ensure homogeneous variances. Replicate assays (N=5) were collected in blocks on different days.

Plaque size fitness assay

A plaque size assay was recently developed as an alternative means of measuring fitness on *phaseolicola* [38]. On *phaseolicola*, the relationship between the paired growth measure of log(fitness) and plaque area is described by the equation: log W = 0.044*PS – 0.34, where W is a one generation measure of relative growth rate, and PS is plaque area in mm^2. Plaque sizes of the wild type ϕ6, and the mutations isolated on the novel host *glycinea* were determined by plating phages onto a lawn of the appropriate host at a low density (<50 phage per plate) to ensure non-overlapping plaques, incubating at 25°C for 24 hours, and taking digital pictures for analysis using ImageJ (NIH, Bethesda, MD, http://rsb.info.nih.gov/ij/). Each plaque size measure is a mean area of plaques on an individual plate. For each genotype, replicate assays (N=6) were collected in blocks on different days.

Determination of selection coefficients

We wished to determine the fitness effects that host range mutations had on *phaseolicola* relative to wild type ϕ6, and so we calculated s, the selection coefficients of these mutations. For paired growth assays, we measured the relative fitness of a host range mutant relative to the wild type virus over 24 hours. The equation $s = e^{\text{Log}(W)/5} - 1$ was used to determine a 1 generation selection coefficient for each host range mutation. As plaque size is already a one generation measure of fitness, we can take the difference in plaque size between a host range mutant, and the wild type virus, and multiply it by the scalar 0.044 (from the above section) to determine the selection coefficient for these host range mutants.

Phage per plaque fitness assay

We wished to determine the absolute fitness of the isolated mutants on their novel host of isolation, so we used phage per plaque assays. Phage from a single genotype are plated onto a 200 μl bacterial lawn and incubated for 24 hours. After 24 hours, the number of plaques on a plate was counted to determine the initial concentration of phage plated (C_0). The phages from the plate are then harvested, resuspended in LC media and filtered to remove bacteria. A sample of these phages are plated onto a 200 μl bacterial lawn of the identical host as the previous day, and incubated for 24 hours. After this 24 hour incubation, the number of plaques on a plate is counted (C_1). The value C_1/C_0 gives us the average number of progeny phages produced per plaque (single phage) over a 24 hour period. These data were then log transformed to ensure homogeneous variances. Replicate assays (N=6) were collected in blocks on different days.

Statistical analyses

Fitness data were analyzed in Microsoft Excel 2003 (Redmond, WA) and SASv9.1 (SAS Institute, Cary, NC) using Proc GLM. All models in SAS initially included a genotype by block effect. However these were never significant, and were therefore dropped from the models. To determine which mutations isolated on *glycinea* caused a significant reduction in fitness relative to the wild type ϕ6 on the original host *phaseolicola* we used an ANOVA, plaque size = genotype + block + ε, to calculate the least significant differences (LSD). The LSD is the smallest difference between any two means that is statistically significant, and is used for pre-planned comparisons in ANOVA [133]. The t-statistic used to calculate the LSD is determined in the same manner as the t-statistic used in a two sample t-test, except

that the Mean Square Error (MSE) is used in place of the sample variance and the degrees of freedom are based on the MSE.

To determine which mutations isolated on *syringae* or *tomato* caused a significant reduction in fitness on the original host *phaseolicola* we used ANOVAs, log(relative fitness) = genotype + block + ε, to calculate the experiment-wise confidence limits for each of these two sets of data. Experiment-wise confidence limits are calculated the same way as a standard confidence limit [133], except that the Mean Square Error (MSE) is used in place of the sample variance, and the degrees of freedom are based on the MSE.

Differences in the fitness effects of host range mutations across the novel hosts of isolation were assayed using t-tests. We first determined the mean fitness effect of each assayed host range mutation. Next, we grouped these mean effects based on their novel host of isolation. Finally, we conducted 2-tailed t-tests on these fitness effects between pairs of novel hosts of isolation.

Table 3.1: Per-nucleotide genetic divergence from original host *phaseolicola*

	glycinea	*tomato*	*syringae*
pilA[a]	0.0605	0.0816	0.1242
Core[b]	0.0117	0.0703	0.0705

[a] Determined from 314 nucleotides

[b] Determined from the weighted average of *acn* (457 ntds), *cts* (576 ntds), and *GapA* (675 ntds)

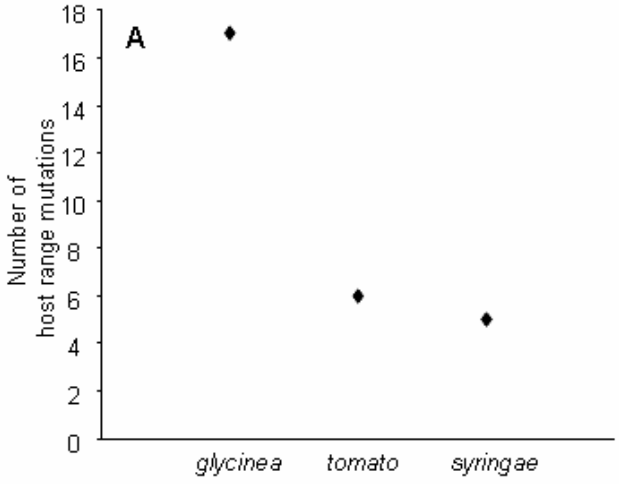
Table 3.2: Host range and sequence changes in the attachment gene P3 of host range mutants.

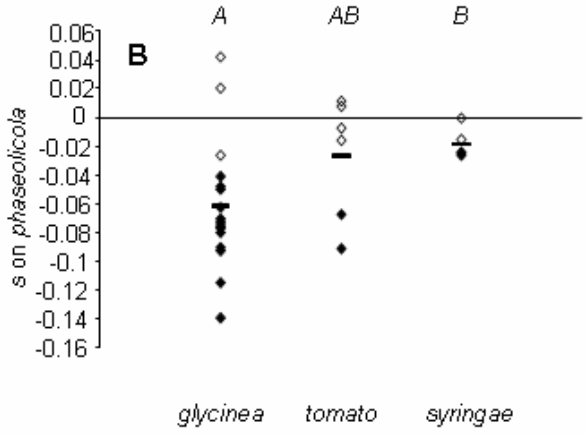
Mutant ID	nt mutation[a]	aa mutation[a]	Times isolated[b]	Permissive hosts[c]
Gly1	None	None	1	G
Gly2	g13a	Gly5Ser	2	G
Gly3	g22a	Glu8Lys	1	G
Gly4	a23g	Glu8Gly	5	G, T
Gly5	a434g	Asp145Gly	3	G
Gly6	a437g	Asn146Ser	6	G
Gly7	g534c	Glu178Asp	1	G
Gly8	g534t	Glu178Asp	1	G
Gly9	c1016t	Pro339His	1	G
Gly10	a1546g	Thr516Ala	4	G
Gly11	a1598c	Asp533Ala	1	G
Gly12	g1603a	Asp535Asn	1	G
Gly13	g1660a	Asp554Asn	2	G
Gly14	a1661t	Asp554Val	1	G
Gly15	a1661c	Asp554Ala	1	G
Gly16	a1661g	Asp554Gly	6	G
Gly17	a1661g (& c318t)	Asp554Gly (& Leu106Leu)	1	G
Gly18	a1661g (& t779a)	Asp554Gly (& Phe260Tyr)	1	G
Gly19	c1663t	Leu555Phe	1	G
Tom1	a23g	Glu8Gly	1	G, T
Tom2	a389g (& c911t, g1311a)	Gln130Arg (& Thr304Ile, Leu457Leu)	1	T
Tom3	g390c	Gln130His	15	T
Tom4	g390c (& t1719c)	Gln130His (& Pro573Pro)	1	T
Tom5	g390t	Gln130His	17	T
Tom6	g390t (& a1530g)	Gln130His (& Ala510Ala)	1	T

Tom7	g390t (& t357a)	Gln130His (& Ala119Ala)	1	T
Tom8	g1660t	Asp554Tyr	1	G, T
Tom9	g1675a	Asp559Asn	2	T
Syr1	g390t	Gln130His	1	S
Syr2	g390c	Gln130His	1	S
Syr3	g1675a	Asp559Asn	24	S
Syr4	g1675a (& c693t)	Asp559Asn (& Gly231Gly)	1	S
Syr5	g1675t	Asp559Tyr	8	S
Syr6	a1676c	Asp559Ala	5	S

[a] Nucleotide and amino acid substitutions are labeled according to their position in P3. Second site mutations are shown in parentheses.
[b] Number of times each genotype was isolated in that novel host collection.
[c] Novel hosts that this genotype can grow on. G=*glycinea*, T=*tomato*, S=*syringae*.

Figure 3.1


A
Number of host range mutations
18
16
14
12
10
8
6
4
2
0
glycinea
tomato
syringae


A
AB
B
B
s on phaseolicola
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.14
-0.16
glycinea
tomato
syringae

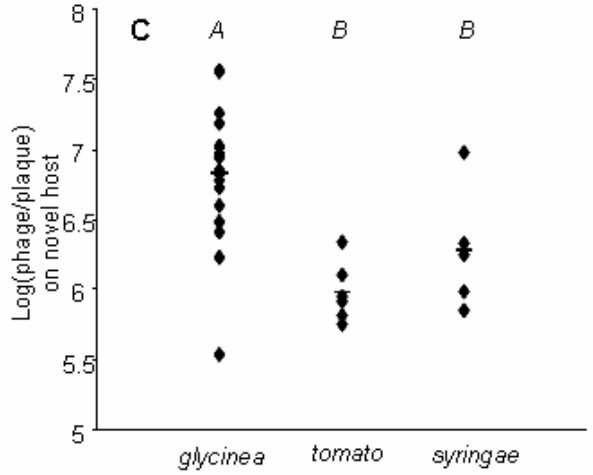

C
A
B
B
Log(phage/plaque) on novel host
8
7.5
7
6.5
6
5.5
5
glycinea
tomato
syringae

Figure 3.1: Characteristics of host range mutations depend on their novel host. Novel hosts are listed on the x-axis in order of their similarity from the original host *phaseolicola*. (**A**) The number of unique host range mutations isolated on a novel host. (**B**) The fitness effects, *s*, of host range mutations relative to the wild type ϕ6 on the original host *phaseolicola*. Each point is the mean of n=5 (*syringae* and *tomato*) or n=6 (*glycinea*) replicate measures of each unique host range mutation, and the bars are the grand mean of each group. The solid line (*s*=0) denotes the fitness of the wild type ϕ6. Filled points are those that have a significantly reduced fitness on *phaseolicola* when compared to the wild type ϕ6. Significance was determined by either using the LSD (for *glycinea* mutations), or experiment-wise confidence limits (for *tomato* and *syringae* mutations). Groups with the same overhead letter are not significantly different from each other based on 2-tailed t-tests. (**C**) The absolute fitness of host range mutants on their novel host. Each point is the mean of n=6 replicate measures of log(phage/plaque) for each unique host range mutation on its novel host, and the bars are the grand mean for each group. Groups with the same overhead letter are not significantly different from each other based on 2-tailed t-tests.

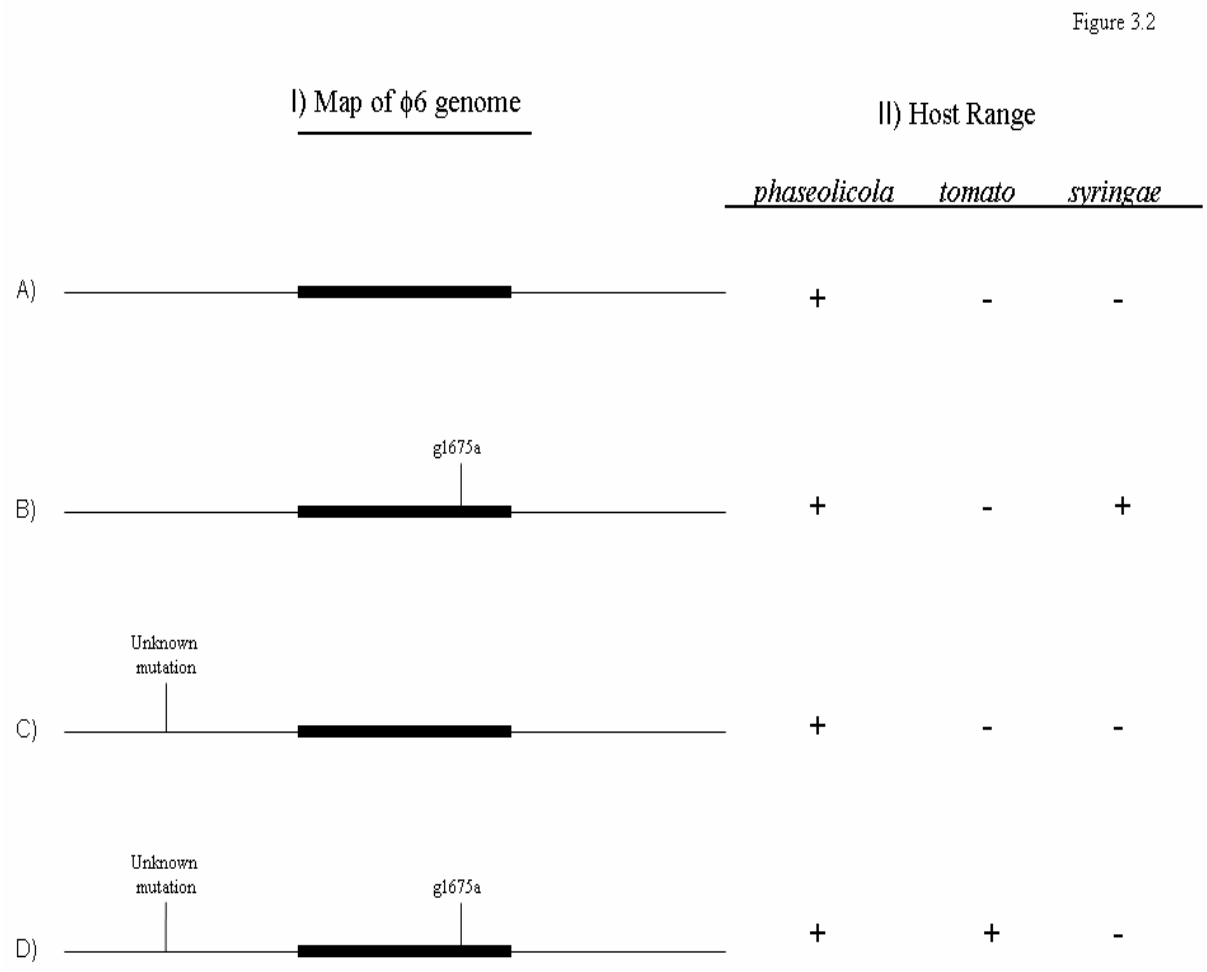


Figure 3.2: Epistatic basis of host range expansion. We identified three P3 genotypes with different host ranges depending on their novel host of isolation, and our data suggest that the most likely explanation for this pattern is an epistatic basis of host usage. We illustrate this example with one of the P3 mutations, g1675a. **I)** Linear representation of the ϕ6 genome, with the bolded segment representing P3, the gene we sequenced in this study. **II)** Host range of the genomes in I. The wild type virus (A) can only infect the standard host *phaseolicola*, and neither novel host. A host range mutant (B) containing only the g1675a

mutation will be able to grow on the novel host *syringae*, but not on *tomato*. The single mutant (C) containing the unsequenced mutation will not be able to infect either novel host. A host range mutant (D) containing both the g1675a mutation in P3, as well as the unsequenced mutation will be able to grow on *tomato*, but not *syringae*.

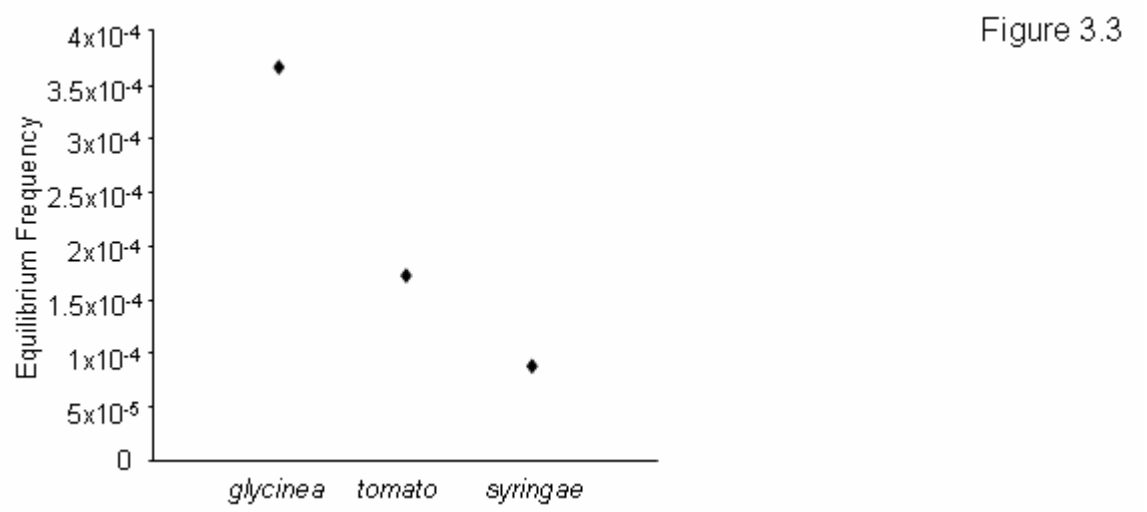


Figure 3.3: Predicted equilibrium frequency of host range mutants in a population on *phaseolicola*. The equilibrium frequency of each host range mutation were calculated as $\hat{q} \approx \mu / s$, where μ is either the transition or transversion mutation rate, and s is the observed fitness cost of a mutation on the original host *phaseolicola*. The individual frequencies of mutations isolated on each novel host type were summed, giving estimated frequencies of host range mutants allowing growth on each of these novel hosts, $\hat{f}_{HR}$, on the Y-axis. Novel hosts are presented on the X-axis in order of their similarity to the original host *phaseolicola*.

Chapter 4: Differing genetic basis and fitness effects of niche-expansion and niche-adaptive mutations.

The work described in this chapter was accomplished in collaboration with Dr. Christina Burch. I would like to thank Sarah Diamond, Siobain Duffy, Cris Ledon-Rettig, and members of the Burch lab for discussions that improved both the experimental design and the final manuscript.

Abstract

I compare characteristics of mutations that expand ϕ6 host range to those mutations that adapt ϕ6 to a novel host. I adapted 20 ϕ6 populations founded from the same ancestral virus on the novel host *P. syringae* pv. *glycinea* until these populations each acquired a single mutation that was beneficial on *glycinea*. I sequenced the host attachment genes of these adapted lines and measured their fitness on both the original host of ϕ6 and *glycinea*. I found that despite significantly adapting to *glycinea*, few lines showed a significant change in fitness on the original host of ϕ6, and that only 3 of the 20 adapted populations had a coding mutation in the host attachment spike. These findings contrast with those regarding the mutations that expanded the host range of ϕ6 to include *glycinea,* which nearly always imposed a fitness cost on the original host and nearly always occurred in the host attachment gene.

Introduction

The evolution of an organism's niche, the environmental conditions in which a species can persist without immigration [17], can influence a number of evolutionary and ecologically important processes, including the maintenance of variation [155], ecological speciation [156], and species invasion [157]. Niche evolution depends on both the direct response to selection a population experiences as well as the pleiotropic fitness effects (i.e., correlated fitness responses in other components of the niche) that a direct response to selection generates [158]. In particular, pleiotropic fitness costs are predicted to strongly constrain niche evolution [159, 160], though such costs are not always observed [125, 144].

Direct selection can act on an organism's niche in two ways: 1) by expanding the niche to include a novel environment (expansion) and 2) by improving fitness in a portion of the niche to which the organism is not optimally adapted (adaptation). Research on the thermal niche of *Eshcerichia coli* has suggested that niche expansion and niche adaptation are independent processes [161, 162]. If this independence is a general property of niche evolution, then a complete understanding of niche evolution requires determining both when and how the pleiotropic fitness effects of niche-expansion and niche-adaptive mutations differ.

In addition to being an important component of viral disease emergence into novel host populations [3, 4, 10, 149], the evolution of viral host range provides an excellent system for investigating whether the pleiotropic fitness effects of niche-expansion and niche-adaptive mutations differ. Viral hosts are discreet environments which comprise a major component of the viral niche. In addition, the simplicity of viral systems allows for both the dissection of the genetic basis of evolution as well as accurate measures of the fitness effects of mutations [125, 130, 163]. Finally, we have some intuition as to how the pleiotropic fitness effects of host-expansion and host-adaptive mutations might differ, based on a mechanistic understanding of virus host range expansion.

Host-expansion mutations should occur in only one or a few genes, as a virus will be blocked from infecting a novel host at only one point in the infectious cycle. Specifically, expansion of virus host range most often occurs via mutations that allow a virus to recognize and attach to an extracellular component of a novel host [140]. In contrast, host-adaptive

mutations should occur in a large number of virus genes, as a virus with low fitness on a host will likely be maladapted to this host at a number of points in the infectious cycle. Extracellular host components, including those involved in viral recognition and attachment, are typically more divergent between host species than intracellular components. We therefore expect that host-adaptive mutations, which can interact with both extra- and intracellular components, will tend to have a lower frequency and magnitude of pleiotropic fitness costs than host-expansion mutations.

Empirical evidence does not appear to support this expectation. A number of studies have investigated the pleiotropic fitness effects of host-expansion [26, 45, 65, 79, 90, 96, 109, 110, 112, 113] or host-adaptive [72, 106-108, 111] mutations. These studies have shown that both types of mutations commonly have pleiotropic fitness costs, although differences in the magnitude of costs between these two sets are not comparable. Additionally, a recent study of host adaptation of the bacteriophage ϕ6 found strong pleiotropic fitness costs, occasionally resulting in a complete loss of use of ϕ6's original host [164]. However, despite any intuition or evidence we may currently have, direct and controlled comparisons of the genetic basis and pleiotropic fitness effects of host-expansion and host-adaptive mutations have not been performed.

In this paper we build on a previous study in which we collected and characterized mutations that expanded the host range of the bacteriophage ϕ6 to include the novel host *Pseudomonas syringae* pathovar *glycinea* [128]. Here, we adapt replicate populations of a ϕ6 host range mutant to *P. syringae* pv. *glycinea*. We then compare both the genetic basis and

the pleiotropic fitness effects of *glycinea*-expansion and *glycinea*-adaptive mutations. By utilizing the same virus-host pair, we are able to directly compare the characteristics of host-adaptive and host-expansion mutations.

Materials and Methods

Strains and Culture Conditions

The wild type double-stranded RNA bacteriophage ϕ6 (*Cystoviridae*) used in this study is a laboratory genotype descended from the original isolate [131]. ϕG22 is a single nucleotide host range mutant of the wild type ϕ6 which has gained the ability to utilize the novel host *P. syringae* pv. *glycinea*. The standard laboratory host of ϕ6, *P. syringae* pv. *phaseolicola* strain HB10Y, was obtained from the American Type Culture Collection (ATCC; no. 21781), the novel host *P. syringae* pv. *glycinea* strain R4a was obtained from J. Dangl (University of North Carolina, Chapel Hill, NC), and marker strain *P. syringae* pv. *tomato* strain Bakersfield was obtained from G. Martin (Cornell University, Ithaca, NY). Hereafter, hosts are referred to by their pathovar designations. Details of diluting, filtering, culture, and storage of phage and bacteria are published [27, 33]. Phage and bacteria were cultured in LC media (5 g yeast extract, 5 g NaCl, and 10 g Bacto-tryptone per L of H_2O), and stored in 2:3 glycerol:LC media (v/v) at -20^{o}C and -80^{o}C, respectively. For growth on plates, phage were mixed with the appropriate host bacteria in LC top agar (0.7% agar) and plated on LC plates (1.5% agar).

Adaptation to novel host *glycinea*

Details of this protocol are published [33]. All populations were founded by phages from a different isolated plaque of the genotype ϕG22. A sample of the phages from this founder plaque (genotype) was plated on a 200 μL lawn of *glycinea*, and incubated at 25°C overnight to allow the phages to reproduce and form plaques. One hundred plaques from this population were harvested and these phages were then plated onto a fresh 200 μL lawn of *glycinea*. This cycle of population growth and bottlenecking was repeated for 10 days. ϕ6 undergoes approximately five generations every 24 hours so the adapting populations underwent ~50 generations of adaptation to *glycinea*. This entire process was replicated 20 times to produce independent populations (A-T) which were adapted to *glycinea*. Each day of passaging, a sample of the 100 harvested plaques (population sample), as well as a single plaque from that passage (individual isolate) were archived from each population.

Sequencing

Genome amplification and sequencing were performed as previously described [130]. Briefly, phages were grown to a high titer and viral RNA was extracted using a QIAamp Viral RNA Mini Kit (QIAGEN Inc., Valencia, CA). Viral RNA was reverse transcribed using random hexamer primers and Superscript II RNase H- RT (Invitrogen, Carlsbad, CA), and the resulting cDNA was used as template for PCR with Taq Polymerase (Invitrogen, Carlsbad, CA). We amplified three sections of the medium genome segment, corresponding to bases 1298-2142, 2042-3052, and 2877-3873, which encompassed the host attachment gene P3, as well as gene P13, a membrane protein. PCR products were purified using EXO-SapIT (US Biological, Swampscott, MA) and sequenced in both directions using PCR primers and primers internal to each amplicon. Sequencing was performed using BigDye

v3.1 (Applied Biosystems, Foster City, CA) either locally with an Avant-3100 Genetic Analyzer Sequencer (Applied Biosystems, Foster City, CA) or at the UNC Automated Sequencing Facility (University of North Carolina, Chapel Hill, NC).

Paired growth fitness assay

Paired growth assays are a standard method for assaying fitness in ϕ6 [32] and were used to assay fitness on the novel host *glycinea*. Each measured genotype was mixed with the marked competitor, ϕG22t (a mutant of ϕG22 which could also grow on the marker host *tomato*) at a 1:1 ratio. This mixture was plated on a lawn of *glycinea* and incubated at 25^{o}C for 24 hours. The ratio of phage genotypes before and after the incubation was determined by plating phages onto a 200 μL mixed lawn of a 50:1 mixture of *glycinea* and *tomato*, on which the measured genotype forms turbid plaques and ϕG22t forms clear plaques. The relative fitness of a measured genotypes were then determined as $W = R_1/R_0$, where R_0 and R_1 are the ratio of measured genotypes to the marked competitor before and after the 24 hour incubation on *glycinea*, respectively. This value was then log transformed to produce homogenous variances. Replicate assays (N=4) were collected in blocks on different days.

Plaque size fitness assay

A plaque size assay was recently developed as an alternative means of measuring fitness on *phaseolicola* [38]. On *phaseolicola*, the relationship between the paired growth measure of log(*W*) and plaque area is described by the equation: $\log W = 0.044*PS - 0.34$, where *W* is a one generation measure of relative growth rate, and PS is plaque area in mm^2 [38]. Plaque sizes of genotypes were determined by plating phages onto a lawn of

phaseolicola at a low density (<50 phage per plate) to ensure non-overlapping plaques, incubating at 25°C for 24 hours, and taking digital pictures for analysis using ImageJ (NIH, Bethesda, MD, http://rsb.info.nih.gov/ij/). Each plaque size measure is the mean area of plaques on an individual plate. For each genotype, replicate assays (N=4) were collected in blocks on different days.

Statistical analyses

Fitness data were analyzed in Microsoft Excel 2003 (Redmond, WA) and SASv9.1 (SAS Institute, Cary, NC) using Proc GLM. All models in SAS initially included a genotype by block interaction, but these were never significant and were dropped from these models. To determine which mutations had a significantly different fitness relative to the ancestral ϕG22, we used an ANOVA, fitness = genotype + block + ε, to calculate the least significant difference (LSD). The LSD is the smallest difference between any two means that is statistically significant, and is used for pre-planned comparisons in ANOVA [133]. The t-statistic used to calculate the LSD is determined in the same manner as the t-statistic used in a two sample t-test, except that the Mean Square Error (MSE) is used in place of the sample variance and the degrees of freedom is based on the MSE.

In order to compare the pleiotropic fitness effects of *glycinea*-adaptive and *glycinea*-expansion mutations, we calculated the mean plaque size on *phaseolicola* for each *glycinea*-adapted genotype, each unique *glycinea*-expansion mutation, and both ancestral genotypes (ϕG22 for the adapted genotypes and wild type ϕ6 for the expansion mutations). We calculated the difference in plaque size (ΔPS) between ϕG22 and each *glycinea*-adapted

genotype, as well as the ΔPS between wild type ϕ6 and each unique *glycinea*-expansion mutation. We conducted a 2-tailed t-test on the ΔPS values between the *glycinea*-expansion mutations and the *glycinea*-adapted genotypes.

To determine the number of beneficial mutations that fixed in a particular lineage, we used a forward stepwise least squares linear regression to indentify stepwise fitness increases in the daily fitness measures for that lineage (modified slightly from [39]). Our algorithm started with a model that assumed no mutations (i.e. no steps that increase fitness), and iteratively added mutations in the following manner. In each iteration one additional mutation is added to the set of existing mutations. The location of the mutation is chosen in such a way that addition of a step at that location produces the largest reduction in the Residual Sum of Squares (RSS). Mutations were added to the model until each population bottleneck was associated with a mutation. At the end of this process a nested sequence of fitted models was obtained. We then chose the "best" model as the one that gave the smallest value for the Bayesian Information Criterion (BIC). This criterion is a popular method for model selection proposed by Schwarz [165]. The BIC balances the RSS of a model and the number of parameters involved in fitting that model. Note that with the addition of each mutation to the model, there is a reduction in the RSS, but two parameters are added to the model, one for the step location and the other for the height of the added step. To reflect the underlying biological process, we implemented a constrained version of this algorithm that allowed only beneficial mutations (i.e. steps that increased fitness). In this case, when mutations were added to the model, we only considered locations that would result in increasing steps.

Results

Adaptation to the novel host *glycinea*

We wanted to investigate the genetic basis and the pleiotropic fitness effects of mutations that adapted ϕ6 to the novel host *glycinea*. To do so, we adapted 20 replicate populations of ϕG22 (a host range mutant of ϕ6 able to infect *glycinea*) to *glycinea* for 50 generations (10 days). Each population was founded from a different plaque (individual) to ensure that genetic variation that arose during growth of the founding plaque was not shared between populations. Each population underwent 5 generations of growth every day, with populations reaching ~10^9-10^{10} phages. Following this daily growth, these populations were bottlenecked down to 100 individuals before initializing a new daily growth cycle. Effective population size is largely determined by the size of the population bottlenecks, thus the adapting populations had an effective population size of ~ 100. By removing any initial genetic variation and by evolving these populations at a small effective population size, we maximized the likelihood that each population acquired a unique *glycinea*-adaptive mutation.

In order to determine that each population had adapted to *glycinea*, we used paired growth assays to measure the fitness of ϕG22 and the day 10 individual isolates (hereafter, adapted genotypes) from each evolved population. Individual isolates were used for this analysis and for all subsequent analyses in order to reduce error in fitness measures due to genetic variation present in population isolates. We used an ANOVA, $\log(W) = \text{genotype} + \text{block} + \varepsilon$ to calculate the smallest difference between means needed for statistical significance, the LSD. Based on this LSD, all 20 adapted genotypes had significantly

adapted to *glycinea* (Table 4.1), and a Bonferroni correction of the LSD did not change this result.

To ensure that the observed fitness gains on *glycinea* represented the effects of only a single adaptive mutation per population, we determined the number of adaptive steps that the four populations (F, M, Q, and R) which had shown the largest fitness increases on *glycinea* had undergone in their adaptation to *glycinea*. For each of these populations, the log(*W*) of ϕG22, the founder genotype, and the individual isolates from each day of passaging were measured using replicate (n=2) paired growth assays on *glycinea*. We fit stepwise models of fitness gain to these data, and the best fitting model in each lineage contained only a single adaptive step, suggesting the fixation of only a single beneficial mutation during the 50 generations of adaptation (Figure 4.1). As populations with the largest increase in fitness are those that are most likely to have acquired multiple adaptive mutants in a limited time period, we use the finding that these four populations only fixed a single adaptive mutant as evidence that all 20 populations adapted to *glycinea* by fixing only a single adaptive mutation.

Genetic basis of adaptation

The mutation that expanded the host range of ϕG22 and most mutations expanding the host range of ϕ6 occur in P3, the host attachment gene of ϕ6 [128, 130]. We therefore sequenced the coding region of P3 from each of the adapted genotypes to identify mutations that adapted these phages to *glycinea*. All 20 genotypes retained the host range mutation which ϕG22 possessed. Three of the 20 genotypes each acquired a single, unique coding mutation in P3, while another genotype acquired a single silent mutation (Table 4.1). The

primers we used to sequence P3 also covered the coding sequence of P13, a non-essential membrane protein [31]. One of the 20 genotypes contained a coding mutation in P13, which removed the start codon for that gene. The sequences from the remaining 15 genotypes revealed no other mutations in P3 or P13.

Pleiotropic fitness effects on *phaseolicola*

In order to determine the pleiotropic fitness effects that *glycinea*-adaptive mutations had on the original host *phaseolicola*, we measured the plaque size of ϕG22 and each adapted genotype on *phaseolicola*. An ANOVA, Plaque size = genotype + block + ε, was used to calculate the LSD and determine which of the adapted genotypes differed in fitness from the ancestral ϕG22. Only two adapted genotypes, J and R, had fitness significantly different from that of ϕG22, and both of these genotypes had an increased fitness on *phaseolicola* (Table 4.1). Bonferroni correcting the LSD resulted in no genotypes with a significantly different fitness from that of ϕG22. We also examined whether the pleiotropic fitness effects of *glycinea*-adaptive mutations differed between the adaptive mutations we identified in P3, and those not in P3. A comparison between these two classes of mutations showed that they did not differ from each other ($F_{1, 58}$=1.9, P=0.173).

Comparisons to *glycinea* expansion mutants

We compared the *glycinea*-adaptive mutations isolated in this study with the collection of *glycinea*-expansion mutations we had isolated previously [128]. To compare the genetic basis of these two classes of mutations, we used a χ^2 test to see if the likelihood of *glycinea*-expansion and *glycinea*-adaptive mutations occurring in P3 was equal. We found

that the likelihood of these two classes of mutations occurring in P3 was not equal (χ^2_1=43.44, P<0.0001). This result was due to the relative excess of *glycinea*-expansion mutants we isolated which had a P3 mutation (39/40) when compared to *glycinea*-adaptive genotypes which had a P3 mutation (3/20).

We also compared the pleiotropic fitness effects of the 20 *glycinea*-adaptive genotypes with those of the 17 unique *glycinea*-expansion mutations. The change in plaque size (ΔPS) between each mutant and their ancestor (wild type ϕ6 for the 17 unique expansion mutations, ϕG22 for the 20 adapted genotypes) measures the fitness effect that this mutation had on *phaseolicola* (Figure 4.2). We used a 2-tailed t-test to determine whether *glycinea*-expansion mutations and *glycinea*-adaptive mutations had different fitness effects on *phaseolicola*. We found a significant difference between these two groups (t_{36}=7.133, P<0.0001) with *glycinea*-expansion mutations having larger costs on *phaseolicola* than *glycinea*-adaptive mutations.

Discussion

Our goal in this study was to directly compare the characteristics of mutations that expand the host range of the bacteriophage ϕ6 to include the novel host *glycinea* (from [128]) with mutations that further adapt ϕ6 to *glycinea*. To do so, we adapted replicate populations of ϕ6 to *glycinea* until they had each acquired a single *glycinea*-adaptive mutation. We sequenced a region of the genome of each *glycinea*-adapted genotype that contained the host attachment gene. Consistent with our expectation, *glycinea*-adaptive mutations occur in a greater number of genes than *glycinea*-expansion mutations. We also determined the

pleiotropic fitness effects these *glycinea*-adaptive mutations had on the original host *phaseolicola*. Again consistent with our expectation, we found that these mutations showed a lower frequency and magnitude of pleiotropic fitness costs on *phaseolicola* than *glycinea*-expansion mutations did.

Genetic basis of niche-adaptive mutations

We identified 4 *glycinea*-adaptive mutations in the region of the ϕ6 genome we sequenced, with 3 occurring in the host attachment spike gene P3 and one in the membrane protein P13. Our *glycinea*-expansion data [128] shows that P3 is critical in allowing growth of ϕ6 on *glycinea*. P3 was therefore a logical candidate to examine for *glycinea*-adaptive mutations. In contrast, we had no *a priori* reason to expect that P13 would be involved in adaptation to *glycinea*. We therefore sampled both a candidate and non-candidate gene for *glycinea*-adaptive mutations and found that adaptive mutations were rarely acquired in either gene in our adapted populations. This result suggests that *glycinea*-adaptive mutations should be scattered across ϕ6 genes.

In addition to being scattered throughout the ϕ6 genome, our results also suggest that a large number of *glycinea*-adaptive mutations are available to ϕ6. We arrived at this conclusion by comparing our sequence data to two alternate scenarios. In the first scenario only one *glycinea*-adaptive mutation is available to ϕ6, while in the second scenario there are one hundred *glycinea*-adaptive mutations available to ϕ6. If we had adapted our 20 populations to *glycinea* under the first scenario, we would find that all 20 of our populations shared the same glycinea-adaptive mutation. In contrast, if we had adapted our 20

populations under the second scenario, we would not expect any of our populations to share *glycinea*-adaptive mutations. The number of *glycinea*-adaptive mutations we identified in our sequence data does not give us enough data for an estimate of the number of *glycinea*-adaptive mutations available to ϕ6 (as in [128]). However, the fact that none of the four adaptive mutations we identified were shared between adapted populations suggests that a large number of *glycinea*-adaptive mutations are available to ϕ6.

Pleiotropic fitness effects of adaptive mutations

Our finding that *glycinea*-adaptive mutations had a lower frequency and magnitude of pleiotropic fitness costs than *glycinea*-expansion mutations was consistent with our expectation. However, it appears that the reasoning behind our expectation was incorrect. We had predicted that mutations in ϕ6 genes which interact with extracellular host components would have a greater frequency and magnitude of pleiotropic fitness costs than mutations in ϕ6 genes which interact with intracellular components. The host attachment spike of ϕ6, P3, interacts with an extracellular host component. Therefore, we would expect that the adaptive mutations we identified in P3 would exhibit greater pleiotropic fitness costs than adaptive mutations occurring in other ϕ6 genes, which would interact with a mixture of extra- and intracellular host components. However, the adaptive mutations we identified in P3 did not differ in their pleiotropic fitness effects from the adaptive mutations occurring in genes other than P3. A lack of pleiotropic fitness costs therefore appears to be a general trait of *glycinea*-adaptive mutations.

The pleiotropic fitness effects associated with virus host range evolution have been shown to differ across hosts [45, 96, 106, 111, 112, 125] and similar results have been found in other systems, including bacterial sugar usage [141], bacterial thermal range [161, 162], and phage thermal range [166]. Our results showed that despite the strong and common pleiotropic fitness costs of *glycinea* expansion, only two of our adapted genotypes had significantly different fitness on *phaseolicola* from the ancestral genotype ϕG22, with both of these genotypes having increased fitness. These results add a new consideration to the study of pleiotropic fitness effects - both the magnitude and direction of the pleiotropic fitness effects of niche-expansion and niche-adaptive mutations can differ for the same environment.

Typically, studies of niche evolution examine the pleiotropic fitness effects associated with either niche-expansion [128, 130, 162] or niche-adaptation [124, 125, 141, 167]. When the results of these studies are generalized for the system of interest, they can misstate the effects that pleiotropic responses to selection will have on an organism's niche. In order to accurately determine how niche evolution occurs, future studies will have to consider the pleiotropic fitness effects of both niche-expansion and niche-adaptive mutations.

Implications for disease emergence

The data we collected are of particular relevance to viral disease emergence. Emerging viral diseases, viruses which have either recently appeared in a novel host population or are increasing in frequency in a previously permissive host population [4], present a major health concern to human populations. Studies of emergence have found that a number of evolutionary and ecological factors [1, 2], including the evolution of a virus'

host range [10], can influence the emergence process. Specifically, both the expansion of a virus' host range, as well as adaptation to a host that a virus uses inefficiently can contribute to emergence.

The frequency of host-expansion or host-adaptive mutations in a virus population on a reservoir host (i.e. a host to which the virus is well adapted) will be determined by two processes: mutation and selection. The virus mutation rate and the number of host-expansion or host-adaptive mutations available to a virus will determine the rate of creation of these two classes of mutations. Once created, selection will act on these mutations, reducing the frequency of mutations that cause the virus to suffer pleiotropic fitness costs on its reservoir host. The frequency of host-expansion or host-adaptive mutations in a reservoir population will determine how often these two types of mutations are transmitted to the host population a virus is emerging in. As the transmission of these two types of mutations are critical to emergence success [12-14, 16], it is useful to consider how the characteristics of these two types of mutations will effect their frequency in a reservoir population.

As we were not able to identify all of the *glycinea*-adaptive mutations in our populations, we cannot make a comparison between the number of *glycinea*-adaptive and *glycinea*-expansion mutations. However, our results here and in [128] suggest that both types of mutations should be common. We were able to measure the pleiotropic fitness effects both of these collections of mutations showed on the original host *phaseolicola* (the ϕ6 reservoir). Based on these observed pleiotropic fitness effects, we can see that selection will tend to reduce the frequency of *glycinea*-expansion mutations, but will not reduce, and

might even increase the frequency of glycinea-adaptive mutations. Therefore, the frequency of *glycinea*-adaptive mutations should be higher in a reservoir population than that of *glycinea*-expansion mutations.

This conclusion is in agreement with a striking pattern of disease emergence in human populations: zoonotic viruses, which naturally infect both human and animal hosts, are more likely to be increasing in frequency in human populations compared to viruses that have only recently begun infecting humans [4]. There are a number of demographic, ecological and evolutionary differences between these two scenarios of emergence. Due to experimental limitations, it is difficult to determine which of these differences contribute to the greater emergence probability of zoonotic viruses. By showing that there should be a difference in the transmission of host-adaptive and host-expansion mutations from a reservoir host, our results suggest a mechanism that might explain this pattern of emergence.

Table 4.1: Sequence changes and fitness measures of adapted lineages

Genotype	nt. mutation[a]	aa mutation[a]	Log(W) glycinea[b]	Plaque size phaseolicola
ϕG22	P3:g13a[c]	P3:G5S[c]	-0.25	6.36
A	none	none	0.57**	7.2
B	P13:t2g	P13:M1R	0.61**	6.26
C	none	none	0.57**	7.01
D	none	none	0.51**	7.25
E	none	none	0.42**	6.66
F	none	none	0.89**	6.87
G	P3:g433a	P3:D144N	0.47**	6.53
H	none	none	0.4**	5.77
I	none	none	0.4**	6.86
J	none	none	0.43**	7.77*
K	P3:t969c	P3:N323N	0.56**	7.24
L	P3:g382c	P3:G127R	0.36**	6.52
M	P3:c591g	P3:D197E	0.67**	6.34
N	none	none	0.61**	6.81
O	none	none	0.61**	6.33
P	none	none	0.47**	6.57
Q	none	none	0.65**	6.86
R	none	none	0.65**	7.69*
S	none	none	0.52**	6.25
T	none	none	0.45**	6.73

[a] Nucleotide and Amino acid substitutions are identified by their gene, and their position within that gene.

[b] relative to common competitor ϕG22t

[c] Ancestral residues, all adapted lineages retained this mutation

Significant differences from ancestor ϕG22 at the *P=0.05, or *P=0.0025 level

Figure 4.1

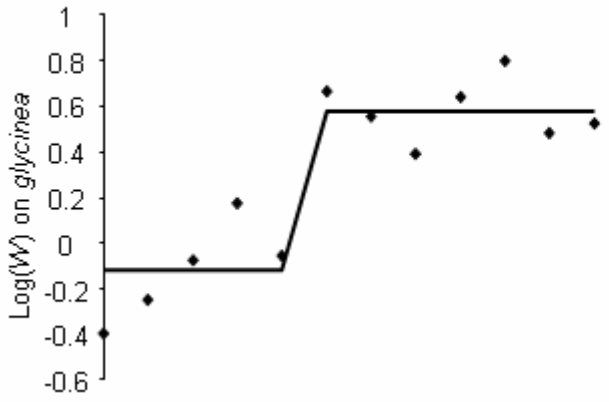


Figure 4.1: Adaptive trajectory of a population to *glycinea*. We measured the fitness trajectory on *glycinea* of four of our adapted populations and found that each of these four populations underwent a single adaptive step. We show one population's adaptive trajectory here for illustration. Individual points are the average of replicate (n=2) measures of log(*W*) on *glycinea* for ϕG22, the lineages' founder genotype, and each daily individual isolate from the lineage. The solid line shows the adaptive step that this population underwent and was identified by using a Bayesian Information Criterion (BIC).

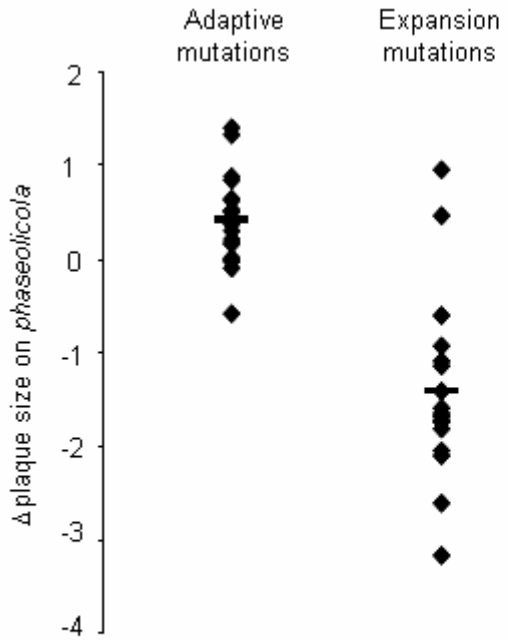


Figure 4.2: Pleiotropic fitness effects of *glycinea*-adaptive and *glycinea*-expansion mutations on *phaseolicola*. The difference in plaque size (ΔPS) was determined between each mutant genotype and its ancestor (wild type ϕ6 for the expansion mutations, ϕG22 for the adapted genotypes). Each point represents the average of n=4 (adaptive mutations) or n=6 (expansion mutations) replicate measures of ΔPS on *phaseolicola*, and the bars represent the mean ΔPS for all adaptive or expansion mutations. The difference between these groups was significant (t_{36}=7.133, P<0.0001).

REFERENCES

1. Schrag, S.J. and P. Wiener, *Emerging Infectious-Disease - What Are the Relative Roles of Ecology and Evolution.* Trends in Ecology & Evolution, 1995. **10**(8): p. 319-324.

2. Morse, S.S., *Factors in the Emergence of Infectious-Diseases.* Emerging Infectious Diseases, 1995. **1**(1): p. 7-15.

3. Taylor, L.H., S.M. Latham and M.E.J. Woolhouse, *Risk factors for human disease emergence.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2001. **356**(1411): p. 983-989.

4. Woolhouse, M.E.J. and S. Gowtage-Sequeria, *Host range and emerging and reemerging pathogens.* Emerging Infectious Diseases, 2005. **11**(12): p. 1842-1847.

5. Weiss, R.A., *Gulliver's travels in HIVland.* Nature, 2001. **410**(6831): p. 963-967.

6. deJong, J.C., E.C.J. Claas, A.D.M.E. Osterhaus, R.G. Webster and W.L. Lim, *A pandemic warning?* Nature, 1997. **389**(6651): p. 554-554.

7. Holmes, E.C. and A. Rambaut, *Viral evolution and the emergence of SARS coronavirus.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2004. **359**(1447): p. 1059-1065.

8. Woolhouse, M.E.J., *Population biology of emerging and re-emerging pathogens.* Trends in Microbiology, 2002. **10**(10): p. S3-S7.

9. Cleaveland, S., M.K. Laurenson and L.H. Taylor, *Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2001. **356**(1411): p. 991-999.

10. Antia, R., R.R. Regoes, J.C. Koella and C.T. Bergstrom, *The role of evolution in the emergence of infectious diseases.* Nature, 2003. **426**(6967): p. 658-661.

11. May, R.M., S. Gupta and A.R. McLean, *Infectious disease dynamics: what characterizes a successful invader?* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2001. **356**(1410): p. 901-910.

12. Gomulkiewicz, R. and R.D. Holt, *When Does Evolution by Natural-Selection Prevent Extinction.* Evolution, 1995. **49**(1): p. 201-207.

13. Gomulkiewicz, R., R.D. Holt and M. Barfield, *The effects of density dependence and immigration on local adaptation and niche evolution in a black-hole sink environment.* Theoretical Population Biology, 1999. **55**(3): p. 283-296.

14. Holt, R.D. and R. Gomulkiewicz, *How does immigration influence local adaptation? A reexamination of a familiar paradigm.* American Naturalist, 1997. **149**(3): p. 563-572.

15. Holt, R.D., M. Barfield and R. Gomulkiewicz, *Temporal variation can facilitate niche evolution in harsh sink environments.* American Naturalist, 2004. **164**(2): p. 187-200.

16. Holt, R.D., M. Barfield and R. Gomulkiewicz, *Theories of Niche Conservatism and Evolution: Could Exotic Species be Potential Tests?* in *Species Invasions: Insights into Ecology, Evolution, and Biogeography*, D. Sax, J. Stachowicz, and S.D. Gaines, Editors. 2005, Sinauer Associates: Sunderland, MA.

17. Holt, R.D. and M.S. Gaines, *Analysis of Adaptation in Heterogeneous Landscapes - Implications for the Evolution of Fundamental Niches.* Evolutionary Ecology, 1992. **6**(5): p. 433-447.

18. Holt, R.D., *Population-Dynamics in 2-Patch Environments - Some Anomalous Consequences of an Optimal Habitat Distribution.* Theoretical Population Biology, 1985. **28**(2): p. 181-208.

19. Holt, R.D., R. Gomulkiewicz and M. Barfield, *The phenomology of niche evolution via quantitive traits in a 'black-hole' sink.* Proceedings of the Royal Society of London Series B-Biological Sciences, 2003. **270**(1511): p. 215-224.

20. Hsiao, J.C., C.S. Chung, R. Drillien and W. Chang, *The cowpox virus host range gene, CP77, affects phosphorylation of eIF2 alpha and vaccinia viral translation in apoptotic HeLa cells.* Virology, 2004. **329**(1): p. 199-212.

21. Sancho, M.C., S. Schleich, G. Griffiths and J. Krijnse-Locker, *The block in assembly of modified vaccinia virus Ankara in HeLa cells reveals new insights into vaccinia virus morphogenesis.* Journal of Virology, 2002. **76**(16): p. 8318-8334.

22. Spielhofer, P., T. Bachi, T. Fehr, G. Christiansen, R. Cattaneo, K. Kaelin, M.A. Billeter and H.Y. Naim, *Chimeric measles viruses with a foreign envelope.* Journal of Virology, 1998. **72**(3): p. 2150-2159.

23. Bartz, R., R. Firsching, B. Rima, V. ter Meulen and J. Schneider-Schaulies, *Differential receptor usage by measles virus strains.* Journal of General Virology, 1998. **79**: p. 1015-1025.

24. Zhao, Q.Z., J.M. Pacheco and P.W. Mason, *Evaluation of genetically engineered derivatives of a Chinese strain of foot-and-mouth disease virus reveals a novel cell-binding site which functions in cell culture and in animals.* Journal of Virology, 2003. **77**(5): p. 3269-3280.

25. Pacheco, J.M., T.M. Henry, V.K. O'Donnell, J.B. Gregory and P.W. Mason, *Role of nonstructural proteins 3A and 3B in host range and pathogenicity of foot-and-mouth disease virus.* Journal of Virology, 2003. **77**(24): p. 13017-13027.

26. Kozak, C.A. and A. Chakraborti, *Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance.* Virology, 1996. **225**(2): p. 300-305.

27. Mindich, L., J. Cohen and M. Weisburd, *Isolation of Nonsense Suppressor Mutants in Pseudomonas.* Journal of Bacteriology, 1976. **126**(1): p. 177-182.

28. Mindich, L., I. Nemhauser, P. Gottlieb, M. Romantschuk, J. Carton, S. Frucht, J. Strassman, D.H. Bamford and N. Kalkkinen, *Nucleotide-Sequence of the Large Double-Stranded-Rna Segment of Bacteriophage-Phi-6 - Genes Specifying the Viral Replicase and Transcriptase.* Journal of Virology, 1988. **62**(4): p. 1180-1185.

29. Mindich, L., *Bacteriophage-Phi-6 - a Unique Virus Having a Lipid-Containing Membrane and a Genome Composed of 3 Dsrna Segments.* Advances in Virus Research, 1988. **35**: p. 137-176.

30. Mindich, L., *Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage phi 6.* Microbiology and Molecular Biology Reviews, 1999. **63**(1): p. 149-+.

31. Johnson, M.D. and L. Mindich, *Isolation and Characterization of Nonsense Mutations in Gene-10 of Bacteriophage-Phi-6.* Journal of Virology, 1994. **68**(4): p. 2331-2338.

32. Chao, L., *Fitness of Rna Virus Decreased by Muller Ratchet.* Nature, 1990. **348**(6300): p. 454-455.

33. Chao, L., T.T. Tran and T.T. Tran, *The advantage of sex in the RNA virus phi 6.* Genetics, 1997. **147**(3): p. 953-959.

34. Chao, L., K.A. Hanley, C.L. Burch, C. Dahlberg and P.E. Turner, *Kin selection and parasite evolution: Higher and lower virulence with hard and soft selection.* Quarterly Review of Biology, 2000. **75**(3): p. 261-275.

35. Chao, L., C.U. Rang and L.E. Wong, *Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage phi 6.* Journal of Virology, 2002. **76**(7): p. 3276-3281.

36. Burch, C.L. and L. Chao, *Evolution by small steps and rugged landscapes in the RNA virus phi 6.* Genetics, 1999. **151**(3): p. 921-927.

37. Burch, C.L. and L. Chao, *Evolvability of an RNA virus is determined by its mutational neighbourhood.* Nature, 2000. **406**(6796): p. 625-628.

38. Burch, C.L. and L. Chao, *Epistasis and its relationship to canalization in the RNA virus phi 6.* Genetics, 2004. **167**(2): p. 559-567.

39. Burch, C.L., S. Guyader, D. Samarov and H.P. Shen, *Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus phi 6.* Genetics, 2007. **176**(1): p. 467-476.

40. Girod, A., M. Ried, C. Wobus, H. Lahm, K. Leike, J. Kleinschmidt, G. Deleage and M. Hallek, *Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2.* Nature Medicine, 1999. **5**(9): p. 1052-1056.

41. Kozak, S.L., D.C. Siess, M.P. Kavanaugh, A.D. Miller and D. Kabat, *The Envelope Glycoprotein of an Amphotropic Murine Retrovirus Binds Specifically to the Cellular Receptor Phosphate Transporter of Susceptible Species.* Journal of Virology, 1995. **69**(6): p. 3433-3440.

42. Bai, J., K. Howes, L.N. Payne and M.A. Skinner, *Sequence of Host-Range Determinants in the Env Gene of a Full-Length, Infectious Proviral Clone of Exogenous Avian-Leukosis Virus Hprs-103 Confirms That It Represents a New Subgroup (Designated J).* Journal of General Virology, 1995. **76**: p. 181-187.

43. Valsesiawittmann, S., A. Drynda, G. Deleage, M. Aumailley, J.M. Heard, O. Danos, G. Verdier and F.L. Cosset, *Modifications in the Binding Domain of Avian Retrovirus Envelope Protein to Redirect the Host-Range of Retroviral Vectors.* Journal of Virology, 1994. **68**(7): p. 4609-4619.

44. Gilbert, J.M., P. Bates, H.E. Varmus and J.M. White, *The Receptor for the Subgroup-a Avian Leukosis-Sarcoma Viruses Binds to Subgroup-a but Not to Subgroup-C Envelope Glycoprotein.* Journal of Virology, 1994. **68**(9): p. 5623-5628.

45. Jonah, G., A. Rainey, A. Natonson, L.F. Maxfield and J.M. Coffin, *Mechanisms of avian retroviral host range extension.* Journal of Virology, 2003. **77**(12): p. 6709-6719.

46. Taplitz, R.A. and J.M. Coffin, *Selection of an avian retrovirus mutant with extended receptor usage.* Journal of Virology, 1997. **71**(10): p. 7814-7819.

47. Haun, G., O.T. Keppler, C.T. Bock, M. Herrmann, H. Zentgraf and M. Pawlita, *The Cell-Surface Receptor Is a Major Determinant Restricting the Host-Range of the B-Lymphotropic Papovavirus.* Journal of Virology, 1993. **67**(12): p. 7482-7492.

48. Pallier, C., A. Greco, J. LeJunter, A. Saib, I. Vassias and F. Morinet, *The 3' untranslated region of the B19 parvovirus capsid protein mRNAs inhibits its own mRNA translation in nonpermissive cells.* Journal of Virology, 1997. **71**(12): p. 9482-9489.

49. Buchholz, U.J., H. Granzow, K. Schuldt, S.S. Whitehead, B.R. Murphy and P.L. Collins, *Chimeric bovine respiratory syncytial virus with glycoprotein gene substitutions from human respiratory syncytial virus (HRSV): Effects on host range and evaluation as a live-attenuated HRSV vaccine.* Journal of Virology, 2000. **74**(3): p. 1187-1199.

50. Liang, D.L., I.F. Sainz, I.H. Ansari, L.H.V.G. Gil, V. Vassilev and R.O. Donis, *The envelope glycoprotein E2 is a determinant of cell culture tropism in ruminant pestiviruses.* Journal of General Virology, 2003. **84**: p. 1269-1274.

51. Hueffer, K., J.S.L. Parker, W.S. Weichert, R.E. Geisel, J.Y. Sgro and C.R. Parrish, *The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor.* Journal of Virology, 2003. **77**(3): p. 1718-1726.

52. LlmasSaiz, A.L., M. AgbandjeMcKenna, J.S.L. Parker, A.T.M. Wahid, C.R. Parrish and M.G. Rossmann, *Structural analysis of a mutation in canine parvovirus which controls antigenicity and host range.* Virology, 1996. **225**(1): p. 65-71.

53. Parker, J.S.L., W.J. Murphy, D. Wang, S.J. O'Brien and C.R. Parrish, *Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells.* Journal of Virology, 2001. **75**(8): p. 3896-3902.

54. Spitzer, A.L., C.R. Parrish and I.H. Maxwell, *Tropic determinant for canine parvovirus and feline panleukopenia virus functions through the capsid protein VP2.* Journal of General Virology, 1997. **78**: p. 925-928.

55. Spangenberg, H.C., H.B. Lee, J.S. Li, F.L. Tan, R. Skidgel, J.R. Wands and S.P. Tong, *A short sequence within domain C of duck carboxypeptidase D is critical for duck hepatitis B virus binding and determines host specificity.* Journal of Virology, 2001. **75**(22): p. 10630-10642.

56. Ishikawa, T. and D. Ganem, *The Pre-S Domain of the Large Viral Envelope Protein Determines Host-Range in Avian Hepatitis-B Viruses.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(14): p. 6259-6263.

57. Wool-Lewis, R.J. and P. Bates, *Characterization of Ebola virus entry by using pseudotyped viruses: Identification of receptor-deficient cell lines.* Journal of Virology, 1998. **72**(4): p. 3155-3160.

58. Brojatsch, J., B.S. Kristal, G.A. Viglianti, R. Khiroya, E.A. Hoover and J.I. Mullins, *Feline Leukemia-Virus Subgroup C-Phenotype Evolves through Distinct Alterations near the N-Terminus of the Envelope Surface Glycoprotein.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(18): p. 8457-8461.

59. Moser, M., C.C. Burns, S. Boomer and J. Overbaugh, *The host range and interference properties of two closely related feline leukemia variants suggest that they use distinct receptors.* Virology, 1998. **242**(2): p. 366-377.

60. Faix, P.H., S.A. Feldman, J. Overbaugh and M.V. Eiden, *Host range and receptor binding properties of vectors bearing feline leukemia virus subgroup B envelopes can be modulated by envelope sequences outside of the receptor binding domain.* Journal of Virology, 2002. **76**(23): p. 12369-12375.

61. Bupp, K. and M.J. Roth, *Altering retroviral tropism using a random-display envelope library.* Molecular Therapy, 2002. **5**(3): p. 329-335.

62. Chouteau, P., J. Le Seyec, I. Cannie, M. Nassal, C. Guguen-Cuillouzo and P. Gripon, *A short N-proximal region in the large envelope protein harbors a determinant that contributes to the species specificity of human hepatitis B virus.* Journal of Virology, 2001. **75**(23): p. 11565-11572.

63. Barrera, A., B. Guerra, H. Lee and R.E. Lanford, *Analysis of host range phenotypes of primate hepadnaviruses by in vitro infections of hepatitis D virus pseudotypes.* Journal of Virology, 2004. **78**(10): p. 5233-5243.

64. Takeuchi, Y., M. Akutsu, K. Murayama, N. Shimizu and H. Hoshino, *Host Range Mutant of Human-Immunodeficiency-Virus Type-1 - Modification of Cell Tropism by a Single Point Mutation at the Neutralization Epitope in the Env Gene.* Journal of Virology, 1991. **65**(4): p. 1710-1718.

65. Shioda, T., S. Oka, S. Ida, K. Nokihara, H. Toriyoshi, S. Mori, Y. Takebe, S. Kimura, K. Shimada and Y. Nagai, *A Naturally-Occurring Single Basic-Amino-Acid Substitution in the V3 Region of the Human-Immunodeficiency-Virus Type-1 Env Protein Alters the Cellular Host-Range and Antigenic Structure of the Virus.* Journal of Virology, 1994. **68**(12): p. 7689-7696.

66. Baumann, J.G., D. Unutmaz, M.D. Miller, S.K.J. Breun, S.M. Grill, J. Mirro, D.R. Littman, A. Rein and V.N. KewalRamani, *Murine T cells potently restrict human immunodeficiency virus infection.* Journal of Virology, 2004. **78**(22): p. 12537-12547.

67. Ross, D. and E. Ziff, *Defective Synthesis of Early Region-4 Messenger-Rnas During Abortive Adenovirus Infections in Monkey Cells.* Journal of Virology, 1992. **66**(5): p. 3110-3117.

68. Ellsmore, V., G.G. Reid and N.D. Stow, *Detection of human cytomegalovirus DNA replication in non-permissive Vero and 293 cells.* Journal of General Virology, 2003. **84**: p. 639-645.

69. Kirkwood, C.D., R.F. Bishop and B.S. Coulson, *Attachment and growth of human rotaviruses RV-3 and S12/85 in Caco-2 cells depend on VP4.* Journal of Virology, 1998. **72**(11): p. 9348-9352.

70. Suzuki, Y., T. Ito, T. Suzuki, R.E. Holland, T.M. Chambers, M. Kiso, H. Ishida and Y. Kawaoka, *Sialic acid species as a determinant of the host range of influenza A viruses.* Journal of Virology, 2000. **74**(24): p. 11825-11831.

71. Aytay, S. and I.T. Schulze, *Single Amino-Acid Substitutions in the Hemagglutinin Can Alter the Host Range and Receptor-Binding Properties of H1-Strains of Influenza-a Virus.* Journal of Virology, 1991. **65**(6): p. 3022-3028.

72. Kobasa, D., S. Kodihalli, M. Luo, M.R. Castrucci, I. Donatelli, Y. Suzuki, T. Suzuki and Y. Kawaoka, *Amino acid residues contributing to the substrate specificity of the influenza A virus neuraminidase.* Journal of Virology, 1999. **73**(8): p. 6743-6751.

73. Grassauer, A., A.Y. Egorov, B. Ferko, L. Romanova, H. Katinger and T. Muster, *A host restriction-based selection system for influenza haemagglutinin transfectant viruses.* Journal of General Virology, 1998. **79**: p. 1405-1409.

74. Chen, B.J. and W.J. Atwood, *Construction of a novel JCV/SV40 hybrid virus (JCSV) reveals a role for the JCV capsid in viral tropism.* Virology, 2002. **300**(2): p. 282-290.

75. Gerlier, D., G. Variorkrishnan and P. Devaux, *Cd46-Mediated Measles-Virus Entry - a First Key to Host-Range Specificity.* Trends in Microbiology, 1995. **3**(9): p. 338-345.

76. Ballgoodrich, L.J. and P. Tattersall, *2 Amino-Acid Substitutions within the Capsid Are Coordinately Required for Acquisition of Fibrotropism by the Lymphotropic Strain of Minute Virus of Mice.* Journal of Virology, 1992. **66**(6): p. 3415-3423.

77. Previsani, N., S. Fontana, B. Hirt and P. Beard, *Growth of the parvovirus minute virus of mice MVMp3 in EL4 lymphocytes is restricted after cell entry and before viral DNA amplification: Cell-specific differences in virus uncoating in vitro.* Journal of Virology, 1997. **71**(10): p. 7769-7780.

78. Hatziioannou, T., S. Valsesia-Wittmann, S.J. Russell and F.L. Cosset, *Incorporation of fowl plague virus hemagglutinin into murine leukemia virus particles and analysis of the infectivity of the pseudotyped retroviruses.* Journal of Virology, 1998. **72**(6): p. 5313-5317.

79. Grez, M., M. Zornig, J. Nowock and M. Ziegler, *A Single Point Mutation Activates the Moloney Murine Leukemia-Virus Long Terminal Repeat in Embryonal Stem-Cells.* Journal of Virology, 1991. **65**(9): p. 4691-4698.

80. Baric, R.S., E. Sullivan, L. Hensley, B. Yount and W. Chen, *Persistent infection promotes cross-species transmissibility of mouse hepatitis virus.* Journal of Virology, 1999. **73**(1): p. 638-649.

81. Kuo, L.L., G.J. Godeke, M.J.B. Raamsman, P.S. Masters and P.J.M. Rottier, *Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: Crossing the host cell species barrier.* Journal of Virology, 2000. **74**(3): p. 1393-1406.

82. Thackray, L.B. and K.V. Holmes, *Amino acid substitutions and an insertion in the spike glycoprotein extend the host range of the murine coronavirus MHV-A59.* Virology, 2004. **324**(2): p. 510-524.

83. Schickli, J.H., L.B. Thackray, S.G. Sawicki and K.V. Holmes, *The N-terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells.* Journal of Virology, 2004. **78**(17): p. 9073-9083.

84. Battini, J.L., J.M. Heard and O. Danos, *Receptor Choice Determinants in the Envelope Glycoproteins of Amphotropic, Xenotropic, and Polytropic Murine Leukemia Viruses.* Journal of Virology, 1992. **66**(3): p. 1468-1475.

85. Han, J.Y., P.M. Cannon, K.M. Lai, Y. Zhao, M.V. Eiden and W.F. Anderson, *Identification of envelope protein residues required for the expanded host range of 10A1 murine leukemia virus.* Journal of Virology, 1997. **71**(11): p. 8103-8108.

86. Stevens, A., M. Bock, S. Ellis, P. LeTissier, K.N. Bishop, M.W. Yap, W. Taylor and J.P. Stoye, *Retroviral capsid determinants of Fv1 NB and NR tropism.* Journal of Virology, 2004. **78**(18): p. 9592-9598.

87. Rubio, M.P., S. Guerra and J.M. Almendral, *Genome replication and postencapsidation functions mapping to the nonstructural gene restrict the host range of a murine parvovirus in human cells.* Journal of Virology, 2001. **75**(23): p. 11573-11582.

88. White, L.J., J.M. Ball, M.E. Hardy, T.N. Tanaka, N. Kitamoto and M.K. Estes, *Attachment and entry of recombinant Norwalk virus capsids to cultured human and animal cell lines.* Journal of Virology, 1996. **70**(10): p. 6589-6597.

89. Tao, T. and K.W. Ryan, *Host range restriction of parainfluenza virus growth occurs at the level of virus genome replication.* Virology, 1996. **220**(1): p. 69-77.

90. Couderc, T., F. Delpeyroux, H. LeBlay and B. Blondel, *Mouse adaptation determinants of poliovirus type 1 enhance viral uncoating.* Journal of Virology, 1996. **70**(1): p. 305-312.

91. Shiroki, K., T. Ishii, T. Aoki, Y. Ota, W.X. Yang, T. Komatsu, Y. Ami, M. Arita, S. Abe, S. Hashizume and A. Nomoto, *Host range phenotype induced by mutations in the internal ribosomal entry site of poliovirus RNA.* Journal of Virology, 1997. **71**(1): p. 1-8.

92. Eki, T., T. Enomoto, C. Masutani, A. Miyajima, R. Takada, Y. Murakami, T. Ohno, F. Hanaoka and M. Ui, *Mouse DNA Primase Plays the Principal Role in Determination of Permissiveness for Polyomavirus DNA-Replication.* Journal of Virology, 1991. **65**(9): p. 4874-4881.

93. Ritzhaupt, A., L.J.W. van der Laan, D.R. Salomon and C.A. Wilson, *Porcine endogenous retrovirus infects but does not replicate in nonhuman primate primary cells and cell lines.* Journal of Virology, 2002. **76**(22): p. 11312-11320.

94. Dunn, K.J., C.C. Yuan and D.G. Blair, *A Phenotypic Host-Range Alteration Determines Rd114 Virus Restriction in Feline Embryonic-Cells.* Journal of Virology, 1993. **67**(8): p. 4704-4711.

95. Harris, J.R. and V.R. Racaniello, *Changes in rhinovirus protein 2C allow efficient replication in mouse cells.* Journal of Virology, 2003. **77**(8): p. 4773-4780.

96. Heil, M.L., A. Albee, J.H. Strauss and R.J. Kuhn, *An amino acid substitution in the coding region of the E2 glycoprotein adapts Ross River virus to utilize heparan sulfate as an attachment moiety.* Journal of Virology, 2001. **75**(14): p. 6303-6309.

97. Dong, J.Y., M.G. Roth and E. Hunter, *A Chimeric Avian Retrovirus Containing the Influenza-Virus Hemagglutinin Gene Has an Expanded Host Range.* Journal of Virology, 1992. **66**(12): p. 7374-7382.

98. Hotzel, I. and W.R. Cheevers, *Differential receptor usage of small ruminant lentiviruses in ovine and caprine cells: Host range but not cytopathic phenotype is determined by receptor usage.* Virology, 2002. **301**(1): p. 21-31.

99. Giroglou, T., J. Cinatl, H. Rabenau, C. Drosten, H. Schwalbe, H.W. Doerr and D. von Laer, *Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein.* Journal of Virology, 2004. **78**(17): p. 9007-9015.

100. Ting, Y.T., C.A. Wilson, K.B. Farrell, G.J. Chaudry and M.V. Eiden, *Simian sarcoma-associated virus fails to infect Chinese hamster cells despite the presence of*

functional Gibbon ape leukemia virus receptors. Journal of Virology, 1998. **72**(12): p. 9453-9458.

101. Li, M.L., H.J. Liao, L.D. Simon and V. Stollar, *An amino acid change in the exodomain of the E2 protein of Sindbis virus, which impairs the release of virus from chicken cells but not from mosquito cells.* Virology, 1999. **264**(1): p. 187-194.

102. Boehme, K.W., J.C. Williams, R.E. Johnston and H.W. Heidner, *Linkage of an alphavirus host-range restriction to the carbohydrate-processing phenotypes of the host cell.* Journal of General Virology, 2000. **81**: p. 161-170.

103. Johnston, P.B., J.W. Dubay and E. Hunter, *Truncations of the Simian Immunodeficiency Virus Transmembrane Protein Confer Expanded Virus Host-Range by Removing a Block to Virus Entry into Cells.* Journal of Virology, 1993. **67**(6): p. 3077-3086.

104. Chen, J.L., E. Ido, M.H. Jin, T. Kuwata, T. Igarashi, A. Mizuno, Y. Koyanagi and M. Hayami, *Replication of human immunodeficiency virus type 1 (HIV-1), simian immunodeficiency virus strain mac (SIVmac) and chimeric HIV-1/SIVmac viruses having env genes derived from macrophage-tropic viruses: an indication of different mechanisms of macrophage-tropism in human and monkey cells.* Journal of General Virology, 1998. **79**: p. 741-745.

105. Rosales, R., G. Sutter and B. Moss, *A Cellular Factor Is Required for Transcription of Vaccinia Viral Intermediate-Stage Genes.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(9): p. 3794-3798.

106. Hanley, K.A., L.R. Manlucu, L.E. Gilmore, J.E. Blaney, C.T. Hanson, B.R. Murphy and S.S. Whitehead, *A trade-off in replication in mosquito versus mammalian systems conferred by a point mutation in the NS4B protein of dengue virus type 4.* Virology, 2003. **312**(1): p. 222-232.

107. Schultz, D.E., M. Honda, L.E. Whetter, K.L. McKnight and S.M. Lemon, *Mutations within the 5' nontranslated RNA of cell culture-adapted hepatitis a virus which enhance cap-independent translation in cultured African green monkey kidney cells.* Journal of Virology, 1996. **70**(2): p. 1041-1049.

108. Day, S.P., P. Murphy, E.A. Brown and S.M. Lemon, *Mutations within the 5' Nontranslated Region of Hepatitis-a Virus-Rna Which Enhance Replication in Bs-C-1 Cells.* Journal of Virology, 1992. **66**(11): p. 6533-6540.

109. Morris, J.F., E.J. Sternberg, L. Gutshall, S.R. Petteway and L.A. Ivanoff, *Effect of a Single Amino-Acid Substitution in the V3 Domain of the Human-Immunodeficiency-Virus Type-1 - Generation of Revertant Viruses to Overcome Defects in Infectivity in Specific Cell-Types.* Journal of Virology, 1994. **68**(12): p. 8380-8385.

110. Subbarao, E.K., W. London and B.R. Murphy, *A Single Amino-Acid in the Pb2-Gene of Influenza-a Virus Is a Determinant of Host Range.* Journal of Virology, 1993. **67**(4): p. 1761-1764.

111. Jung, Y.T. and C.A. Kozak, *Generation of novel syncytium-inducing and host range variants of ecotropic Moloney murine leukemia virus in Mus spicilegus.* Journal of Virology, 2003. **77**(9): p. 5065-5072.

112. Doi, K., A. Kawana, A. Iwamoto, H. Yoshikura and T. Odawara, *One base change is sufficient for host range conversion of murine leukemia virus from B to NB tropism.* Archives of Virology, 1997. **142**(9): p. 1889-1894.

113. Colston, E.M. and V.R. Racaniello, *Poliovirus Variants Selected on Mutant Receptor-Expressing Cells Identify Capsid Residues That Expand Receptor Recognition.* Journal of Virology, 1995. **69**(8): p. 4823-4829.

114. Zsak, L., Z. Lu, T.G. Burrage, J.G. Neilan, G.F. Kutish, D.M. Moore and D.L. Rock, *African swine fever virus multigene family 360 and 530 genes are novel macrophage host range determinants.* Journal of Virology, 2001. **75**(7): p. 3066-3076.

115. Graff, J., A. Normann, S.M. Feinstone and B. Flehmig, *Nucleotide-Sequence of Wild-Type Hepatitis a Virus Gbm in Comparison with 2 Cell Culture-Adapted Variants.* Journal of Virology, 1994. **68**(1): p. 548-554.

116. Kouomou, D.W. and T.F. Wild, *Adaptation of wild-type measles virus to tissue culture.* Journal of Virology, 2002. **76**(3): p. 1505-1509.

117. Baric, R.S., B. Yount, L. Hensley, S.A. Peel and W. Chen, *Episodic evolution mediates interspecies transfer of a murine coronavirus.* Journal of Virology, 1997. **71**(3): p. 1946-1955.

118. Schickli, J.H., B.D. Zelus, D.E. Wentworth, S.G. Sawicki and K.V. Holmes, *The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range.* Journal of Virology, 1997. **71**(12): p. 9499-9507.

119. Melin, F., R. Kemler, C. Kress, H. Pinon and D. Blangy, *Host Range Specificity of Polyomavirus-Ec Mutants in Mouse Embryonal Carcinoma and Embryonal Stem-Cells and Preimplantation Embryos.* Journal of Virology, 1991. **65**(6): p. 3029-3043.

120. Itoh, M., Y. Isegawa, H. Hotta and M. Homma, *Isolation of an avirulent mutant of Sendai virus with two amino acid mutations from a highly virulent field strain through adaptation to LLC-MK2 cells.* Journal of General Virology, 1997. **78**: p. 3207-3215.

121. McKnight, K.L., D.A. Simpson, S.C. Lin, T.A. Knott, J.M. Polo, D.F. Pence, D.B. Johannsen, H.W. Heidner, N.L. Davis and R.E. Johnston, *Deduced consensus sequence of Sindbis virus strain AR339: Mutations contained in laboratory strains which affect cell culture and in vivo phenotypes.* Journal of Virology, 1996. **70**(3): p. 1981-1989.

122. Labranche, C.C., M.M. Sauter, B.S. Haggarty, P.J. Vance, J. Romano, T.K. Hart, P.J. Bugelski and J.A. Hoxie, *Biological, Molecular, and Structural-Analysis of a Cytopathic Variant from a Molecularly Cloned Simian Immunodeficiency Virus.* Journal of Virology, 1994. **68**(9): p. 5509-5522.

123. Carroll, M.W. and B. Moss, *Host range and cytopathogenicity of the highly attenuated MVA strain of vaccinia virus: Propagation and generation of recombinant viruses in a nonhuman mammalian cell line.* Virology, 1997. **238**(2): p. 198-211.

124. Novella, I.S., D.K. Clarke, J. Quer, E.A. Duarte, C.H. Lee, S.C. Weaver, S.F. Elena, A. Moya, E. Domingo and J.J. Holland, *Extreme Fitness Differences in Mammalian and Insect Hosts after Continuous Replication of Vesicular Stomatitis-Virus in Sandfly Cells.* Journal of Virology, 1995. **69**(11): p. 6805-6809.

125. Turner, P.E. and S.F. Elena, *Cost of host radiation in an RNA virus.* Genetics, 2000. **156**(4): p. 1465-1470.

126. Cuevas, J.M., A. Moya and S.F. Elena, *Evolution of RNA virus in spatially structured heterogeneous environments.* Journal of Evolutionary Biology, 2003. **16**(3): p. 456-466.

127. Chambers, T.J. and M. Nickells, *Neuroadapted yellow fever virus 17D: Genetic and biological characterization of a highly mouse-neurovirulent virus and its infectious molecular clone.* Journal of Virology, 2001. **75**(22): p. 10912-10922.

128. Ferris, M.T., P. Joyce and C.L. Burch, *High frequency of mutations that expand the host range of an RNA virus.* Genetics, 2007. **176**(2): p. 1013-1022.

129. Sokurenko, E.V., R. Gomulkiewicz and D.E. Dykhuizen, *Opinion - Source-sink dynamics of virulence evolution.* Nature Reviews Microbiology, 2006. **4**(7): p. 548-555.

130. Duffy, S., P.E. Turner and C.L. Burch, *Pleiotropic costs of niche expansion in the RNA bacteriophage Phi 6.* Genetics, 2006. **172**(2): p. 751-757.

131. Vidaver, A.K., R.K. Koski and J.L. Vanetten, *Bacteriophage Phi-6 - Lipid-Containing Virus of Pseudomonas-Phaseolicola.* Journal of Virology, 1973. **11**(5): p. 799-805.

132. Stent, G.S., *Molecular Biology of Bacterial Viruses*. 1963, San Fransisco: W.H. Freeman.

133. Sokal, R.R. and F.J. Rohlf, *Biometry*. 1969, New York: W.H. Freeman and Company.

134. Ross, S., *A first course in probability*. 2002, Upper Saddle River, NJ: Prentice Hall.

135. Heck, K.L., G. Vanbelle and D. Simberloff, *Explicit Calculation of Rarefaction Diversity Measurement and Determination of Sufficient Sample Size.* Ecology, 1975. **56**(6): p. 1459-1461.

136. Miller, C.R., P. Joyce and L.P. Waits, *A new method for estimating the size of small populations from genetic mark-recapture data.* Molecular Ecology, 2005. **14**(7): p. 1991-2005.

137. Gottlieb, P., S. Metzger, M. Romantschuk, J. Carton, J. Strassman, D.H. Bamford, N. Kalkkinen and L. Mindich, *Nucleotide-Sequence of the Middle Dsrna Segment of Bacteriophage-Phi-6 - Placement of the Genes of Membrane-Associated Proteins.* Virology, 1988. **163**(1): p. 183-190.

138. Sarkar, S.F. and D.S. Guttman, *Evolution of the core genome of Pseudomonas syringae, a highly clonal, endemic plant pathogen.* Applied and Environmental Microbiology, 2004. **70**(4): p. 1999-2012.

139. Roine, E., D.M. Raineri, M. Romantschuk, M. Wilson and D.N. Nunn, *Characterization of type IV pilus genes in Pseudomonas syringae pv. tomato DC3000.* Molecular Plant-Microbe Interactions, 1998. **11**(11): p. 1048-1056.

140. Baranowski, E., C.M. Ruiz-Jarabo and E. Domingo, *Evolution of cell recognition by viruses.* Science, 2001. **292**(5519): p. 1102-1105.

141. Ostrowski, E.A., D.E. Rozen and R.E. Lenski, *Pleiotropic effects of beneficial mutations in Escherichia coli.* Evolution, 2005. **59**(11): p. 2343-2352.

142. Crow, J.F. and M. Kimura, *An Introduction to Population Genetics Theory*. 1970, New York: Harper & Row.

143. Turner, P.E., C.L. Burch, K.A. Hanley and L. Chao, *Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage phi 6.* Journal of Virology, 1999. **73**(3): p. 2420-2424.

144. Fry, J.D., *The evolution of host specialization: Are trade-offs overrated?* American Naturalist, 1996. **148**: p. S84-S107.

145. Mohri, H., M.K. Singh, W.T.W. Ching and D.D. Ho, *Quantitation of Zidovudine-Resistant Human-Immunodeficiency-Virus Type-1 in the Blood of Treated and Untreated Patients.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(1): p. 25-29.

146. Zhang, L.Q., P. Simmonds, C.A. Ludlam and A.J.L. Brown, *Detection, Quantification and Sequencing of Hiv-1 from the Plasma of Seropositive Individuals and from Factor-Viii Concentrates.* Aids, 1991. **5**(6): p. 675-681.

147. Johnson, R.T., *Emerging viral infections of the nervous system.* Journal of Neurovirology, 2003. **9**(2): p. 140-147.

148. Rota, P.A., M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M.H. Chen, S.X. Tong, A. Tamin, L. Lowe, M. Frace, J.L. DeRisi, Q. Chen, D. Wang, D.D. Erdman, T.C.T. Peret, C. Burns, T.G. Ksiazek, P.E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A.D.M.E. Osterhaus, C. Drosten, M.A. Pallansch, L.J. Anderson and W.J. Bellini, *Characterization of a novel coronavirus associated with severe acute respiratory syndrome.* Science, 2003. **300**(5624): p. 1394-1399.

149. Woolhouse, M.E.J., D.T. Haydon and R. Antia, *Emerging pathogens: the epidemiology and evolution of species jumps.* Trends in Ecology & Evolution, 2005. **20**(5): p. 238-244.

150. Dennehy, J.J., N.A. Friedenberg, R.D. Holt and P.E. Turner, *Viral ecology and the maintenance of novel host use.* American Naturalist, 2006. **167**(3): p. 429-439.

151. DeFilippis, V.R. and L.P. Villarreal, *An Introduction to the Evolutionary Ecology of Viruses*, in *Viral Ecology*, C.J. Hurst, Editor. 2000, Academic Press: San Diego, Ca.

152. Fisher, R.A., *The Genetical Theory of Natural Selection*. 1930, Oxford, U.K.: Oxford University Press.

153. Gilbert, G.S. and C.O. Webb, *Phylogenetic signal in plant pathogen-host range.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(12): p. 4979-4983.

154. Jabado, O.J., G. Palacios, V. Kapoor, J. Hui, N. Renwick, J.H. Zhai, T. Briese and W.I. Lipkin, *Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments.* Nucleic Acids Research, 2006. **34**(22): p. 6605-6611.

155. Bossart, J.L. and J.M. Scriber, *Maintenance of ecologically significant genetic variation in the tiger swallowtail butterfly through differential selection and gene flow*. Evolution, 1995. **49**(6): p. 1163-1171.

156. Schluter, D., *The ecology of adaptive radiation*. 2000, New York: Oxford University Press.

157. Lambrinos, J.G., *How interactions between ecology and evolution influence contemporary invasion dynamics.* Ecology, 2004. **85**(8): p. 2061-2070.

158. Via, S. and R. Lande, *Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity.* Evolution, 1985. **39**(3): p. 505-522.

159. Futuyma, D.J. and G. Moreno, *The Evolution of Ecological Specialization.* Annual Review of Ecology and Systematics, 1988. **19**: p. 207-233.

160. Hawthorne, D.J. and S. Via, *Genetic linkage of ecological specialization and reproductive isolation in pea aphids.* Nature, 2001. **412**(6850): p. 904-907.

161. Bennett, A.F. and R.E. Lenski, *Evolutionary Adaptation to Temperature.2. Thermal Niches of Experimental Lines of Escherichia-Coli.* Evolution, 1993. **47**(1): p. 1-12.

162. Mongold, J.A., A.F. Bennett and R.E. Lenski, *Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of Escherichia coli.* Evolution, 1999. **53**(2): p. 386-394.

163. Crill, W.D., H.A. Wichman and J.J. Bull, *Evolutionary reversals during viral adaptation to alternating hosts.* Genetics, 2000. **154**(1): p. 27-37.

164. Duffy, S., C.L. Burch and P.E. Turner, *Evolution of host specificity drives reproductive isolation among RNA viruses.* Evolution, 2007. **In Press**.

165. Schwarz, G., *Estimating Dimension of a Model.* Annals of Statistics, 1978. **6**(2): p. 461-464.

166. Knies, J.L., R. Izem, K.L. Supler, J.G. Kingsolver and C.L. Burch, *The genetic basis of thermal reaction norm evolution in lab and natural phage populations.* Plos Biology, 2006. **4**(7): p. 1257-1264.

167. Cooper, V.S. and R.E. Lenski, *The population genetics of ecological specialization in evolving Escherichia coli populations.* Nature, 2000. **407**(6805): p. 736-739.