THERMAL ADAPTATION OF THE PHAGE G4 AND MOLECULAR EVOLUTION OF RNA SECONDARY STRUCTURE

JENNIFER LYNN KNIES

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 Curriculum in Genetics and Molecular Biology.

Chapel Hill
2007

Approved by:

Christina Burch, PhD

Joel Kingsolver, PhD

Maria Servedio, PhD

Ronald Swanstrom, PhD

Todd Vision, PhD

ABSTRACT

Jennifer L. Knies: Thermal Adaptation of the Phage G4 and Molecular Evolution of RNA Secondary Structures
(Under the direction of Christina Burch)

The focus of this dissertation has been to investigate the predictability of evolution. This has been done in two ways. The larger part of this thesis is devoted to using thermal adaptation as a model system for investigating adaptation to novel environments. The second and smaller part of this thesis makes and tests a prediction for evolution in RNA secondary structure.

Thermal adaptation is a useful model for studying common patterns of evolution, because the proximate mechanisms (e.g. biochemical rate processes) that determine the effects of temperature on growth or performance and the evolutionary causes of thermal adaptation are known. Two specific hypotheses about thermal adaptation are: (1) Thermal constraints on reaction rates will cause cold-adapted species to have lower maximal growth rates than hotter- adapted species at their thermal optima (i.e. "Hotter is better"), and (2) A trade-off between protein stability and activity underlies performance trade-offs observed at different temperatures.

These predictions were investigated by developing the bacteriophage G4 as a model experimental system. The growth rate of G4-like phages isolated from nature was examined over a wide temperature range and a positive correlation was detected between the phages maximal growth rates and optimal temperatures. By evolving multiple

independent phage populations at high and low temperatures, I was able to detect common patterns of adaptation to high temperatures (*e.g.* increased thermal stability), but I did not detect any loss of themal stability in the populations evolved at low temperatures. By combining analyses of lab evolved and natural isolates of these phage, I showed that the nature of thermal constraints are predictable and repeatable.

Lastly, I investigated interactions between within RNA secondary structure that place limits on the rate and trajectory of molecular evolution. A population genetics model of such interactions was used to predict that rate variation at interacting sites should be higher than rate variation at independent sites. This prediction was tested in eight RNA secondary structures by comparing the ratio of transition to transversion substitutions in paired sites to the ratio in unpaired sites. Six of the eight structures show an excellent match to the quantitative predictions of the population genetics model. These findings suggest use of the transition-transversion rate ratio as a simple diagnostic to validate proposed secondary structures.

DEDICATION

To my parents, Laurie and Leonard Knies, who inspired me by their strong work ethic and provided me with faith in family;

to my four sisters and one brother for being my dearest and lifelong friends,

and to my best friend, my husband - A. Ryan Hitch, for his steadfast love and for making me laugh.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my faculty advisor, Christina L. Burch, for her active involvement in all the studies described herein and for her patience, mentoring, and general good nature throughout these last 5 years. I would like to thank all those in the lab who not only made this work possible but also made my time spent in the lab very enjoyable, especially Martin Ferris, Kristen Dang, Sarah Joseph, and Sebastien Guyader. Furthermore, I would like to express my appreciation for the talent and dedication of the two undergraduates that I mentored, Katie Supler and Matthew Kasold. Finally, I would like to thank my committee members, Maria Serviedo, Todd Vision, Ron Swanstrom and Joel Kingsolver for their guidance and critical consideration of my research. I consider myself very fortunate for the collaborations I have enjoyed with many of my committee members.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

df	Degree(s) of freedom
MS	Mean sum of squares
SS	Sum of squares
geno	Genotype
temp	Temperature
TMV	Template mode of variation
H_o	Null hypothesis
se	Standard error

CHAPTER 1. Background and Introduction

In nature, organisms frequently encounter and adapt to novel environments. Examples in the public eye include pathogen adaptation to novel hosts (*e.g.* avian influenza adapting to humans [1]) and widespread adaptation to rising global temperatures (*e.g.* advances in the timing of bird migration [2]). Identifying general patterns of adaptation would improve our ability to predict the course of future adaptation events. For instance, we would be better able to evaluate the risks of disease emergence in humans and the risks of population extinctions due to global warming. A variety of common patterns of adaptation have been hypothesized, such as trade-offs between performance in alternative environments and genetic constraints that limit adaptation in specific environments, but there is little or inconsistent data for these hypotheses.

Adaptation to temperature is a useful model for studying adaptation to novel environments [3], both because organisms frequently experience variation in temperature [4,5] and because the proximate mechanisms (*e.g.* biochemical rate processes) that determine the effects of temperature on growth or performance [6] are known. Theoretical investigations of thermal adaptation make predictions about growth rates at different temperatures based on this knowledge. For example, organisms inhabiting cold environments are expected to have maximal growth rates that are lower than organisms inhabiting hot environments due to the temperature-dependent nature of enzyme kinetics.

These predictions have been difficult to test due to a lack of data collection and limitations in analysis. The first problem in testing these predictions is that growth rate (or other trait of interest) tends to be measured at only a few temperatures, which limits the available information. The second problem is that while growth rate varies continuously as a function of temperature, experimental studies tend to analyze growth rate at different temperatures as discrete traits. A more effective way to investigate these predictions is to examine thermal reaction norms, which are continuous curves describing how the growth rate of a genotype or species varies as a function of temperature. One reason why this approach has not been more extensively used is because it requires that growth rates be measured at many temperatures, which is too onerous for many model systems.

To overcome these difficulties, I developed the bacteriophage G4 as a model system for studying thermal adaptation. In chapter 2, I reexamined the outcome of an evolution experiment in which phage were adapted to a high temperature, and demonstrated that this system is useful for studying thermal adaptation because one can readily measure phages growth rate with a degree of high replication at numerous temperatures across a wide range of temperatures. By generating thermal reaction norms for evolved phage, I demonstrate that adaptation of a phage population to high temperatures results in a loss of fitness at lower temperatures. In addition, I characterized the genetic basis of adaptation in this high temperature adapted population and show that

mutations beneficial at high temperature in the laboratory are also beneficial at high temperature in a natural population of these phages.

Having developed G4 as a system for investigating patterns of thermal adaptation, I went on to test two specific hypotheses that arise from our knowledge of how temperature affects biochemical rate processes:

(1) The 'hotter is better' hypothesis posits that thermal constraints on reaction rates cause species adapted to low temperatures to have lower maximal growth rates than those adapted to high temperatures.

(2) Fitness trade-offs observed across temperatures are due to a trade-off between protein stability and catalytic activity.

In Chapter 3, I investigated the 'hotter is better' hypothesis, which stems from the knowledge of how temperature affects reaction rates. For example, low temperatures typically depress reaction rates simply because the collision rate of reactants is lower [4]. The 'hotter is better' hypothesis assumes that adaptation is unable to overcome the rate-depressing effects of low temperature, therefore causing populations adapted to lower temperatures to have lower maximal growth rates than those adapted to higher temperatures. Prior to my work, this hypothesis had been tested and the predicted pattern found in an inter-species insect study [7]. It was unknown whether the 'hotter is better' relationship characterized variation within a population as well as variation between populations, or whether the 'hotter is better' relationship is a general pattern across unrelated species.

In Chapter 3, I investigated this hypothesis by generating thermal reaction norms for the growth rates of G4-like phages isolated from nature. From these reaction norms, I estimated the optimum temperature and maximum growth rate for individual phage genotypes. As predicted by the 'hotter is better' hypothesis, I found a positive correlation between maximum growth rate and optimal temperature in this population of phages. I was able to show that the reaction norms of my phage do share a common shape, which allowed me to use a more powerful method of analysis. I also found that the reaction norms of the natural isolates of G4 and the reaction norms of the population characterized in Chapter 2 have a common general shape that may reflect thermodynamic constraints

In Chapter 4, I explore a second hypothesis that predicts the proximate mechanisms responsible for thermal adaptation and the fitness trade-offs seen across temperatures (*e.g.* as seen in Chapter 2). At high temperatures, stability of proteins is expected to be under strong selection to prevent denaturation. At low temperatures, protein flexibility/catalytic activity is expected to be under strong selection. Fitness trade-offs are expected to result from a biochemical dependence of catalytic function on enzyme flexibility [8]. At high temperatures selection for increased stability [9] is expected to cause proteins to become less flexible, and thus, less active at low temperatures. In contrast, at low temperatures selection for increased activity causes proteins to become more flexible, and thus, less stable at high temperatures. This trade-off between catalytic activity and stability has been observed in proteins from natural populations/species inhabiting contrasting thermal environments [8]. However, directed evolution experiments have been able to evolve enzymes that have both high

thermostability and high catalytic activity (reviewed in [10]. It is therefore not clear whether the observed trade-offs are due to biological constraints or due to natural populations having accumulated mutations that are deleterious in environments that they do not encounter.

In Chapter 4, I explored the proximate mechanisms of the initial adaptation to novel temperatures in laboratory populations of G4. I evolved 4 independent phage populations at temperatures higher and lower than G4's thermal optimum, and measured the thermal stability of the ancestral and evolved phages. Consistent with the predictions discussed above, lines evolved at high temperature show increased thermal stability as compared to their ancestor. However, lines evolved at low temperatures show no differences in thermal stability compared to their ancestor.

Together, the first three data chapters (Chapter 2-4) of this dissertation illustrate the power of the bacteriophage model system for investigating general patterns of thermal adaptation. Our demonstration of 'hotter is better' in the population of phage from nature is the first step towards uncovering the genetic bases of the correlation between optimal temperatures and maximal growth rates. Now that we have detected this pattern in a population of phage from nature, we can also look for this pattern in phage adapted to novel temperatures in the lab. Since the genetic differences between phage adapted in the lab are much fewer than the phages from nature, it should be easier to identify the genetic bases of 'hotter is better' in the lab adapted phage. In addition, by comparing adaptation

in the lab to polymorphism in nature, we hope to learn more about selection in nature and the role of polymorphisms in contributing to divergence between populations.

The last chapter takes a different approach to investigating the predictability of evolution. Here I, along with my collaborator Kristen Dang, examined evolution in RNA secondary structures to determine whether the constraint imposed by the requirement to maintain base pairing in stems makes sequence evolution predictable. Naively, one might assume that mutations that disrupt base pairing in stems will be removed by selection in large populations, but can become fixed in small populations via drift. However, Kimura [11] proposed a model, known as the compensatory neutral model, explaining how such deleterious mutations can become fixed in large and small populations. His model uses a special class of deleterious mutations – compensatory mutations – or mutations that are deleterious alone, but neutral or beneficial in combination together. In RNA secondary structures, compensatory mutations are those that restore base pairing that was disrupted by a deleterious mutation. A unique feature of compensatory evolution in RNA secondary structures is that if the mutation that disrupts base pairing is a transition, it can only be compensated for by another transition. Likewise, a transversion can only be compensated for by another transversion. Using this knowledge, I extended this compensatory neutral model of Kimura to predict that a difference in transition and transversion rates will cause rate variation to be elevated at paired (stem) sites as compared to unpaired (loop) sites in RNA secondary structure. Specifically, I expected the rate variation at paired sites to be the square of the rate variation at loop sites. I tested this prediction in 8 RNA secondary structures and found

that 7 structures are in excellent qualitative agreement with this prediction and 6 are in quantitative agreement with the model. Those structures that deviate from this model are best explained as either experiencing structural variation or having unpaired (loop) sites that actually base pair to sites elsewhere in the genome. The excellent fit of the model to the structures suggests that this quantitative prediction should be incorporated into programs that predict RNA secondary structure.

Chapter 2. The genetic basis of thermal reaction norm evolution in lab and natural phage populations.

The work described in this chapter was accomplished in collaboration with Katie Supler and Drs. Christina Burch, Joel Kingsolver, and Rima Izem. This chapter has been published [12] as a paper in PLoS Biology. I would like to acknowledge Steve Marron, Jim Bull, and the Burch and Kingsolver lab members for helpful discussions during the development of this project.

Introduction

The study of evolutionary responses to temperature has served as an important model for understanding the process of adaptation to novel environments [3], in part because the evolutionary and mechanistic causes of thermal adaptation are relatively transparent. Temperature is a component of the environment that varies predictably, and for which we have knowledge of the proximate mechanisms (*e.g.* biochemical rate processes) that determine the effects of temperature on growth or performance [6]. Theoretical explorations of thermal adaptation make predictions based on this knowledge [13-15]. For example, trade-offs in performance at different temperatures are expected to result from a biochemical trade-off between enzyme stability and function. Enzymes selected for stability at high temperatures are less functional at lower temperatures, while enzymes selected for function at lower temperatures are less stable at high temperatures [8]. However, although performance trade-offs have been observed for individual enzymatic reactions, they have rarely been observed for physiological traits [6]. Most of the studies that did observe trade-offs stand out because they measured performance at five or more temperatures across the entire thermal niche [16-22]. However, few of these studies were sufficiently powerful to describe the nature of the genetic constraint governing the trade-off [17,19].

Thermal adaptation and constraint are viewed most naturally as a type of continuous reaction norm, in which the phenotypic trait value (*e.g.* fitness or performance) of a genotype varies as a function of some continuous environmental variable (*e.g.* temperature). Reaction norms are routinely used to investigate

environmental effects on a variety of phenotypes. For example, reaction norms have been used to understand how bacterial survival varies as a function of antibiotic concentration [23], immune suppression varies as a function of mold toxin concentration [24], and photosynthesis rate in plants varies as a function of temperature [25].

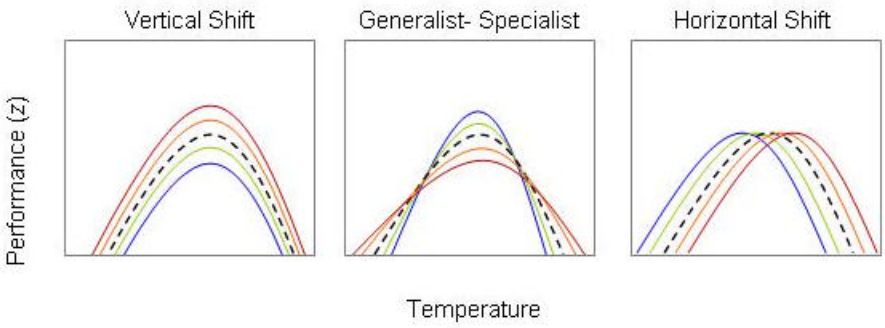


Fig. 2.1. TMV decomposes reaction norm shape variation into three modes of variation. Vertical shifts produce changes in average performance (a), horizontal shifts produce changes in the optimal temperature (m), and generalist-specialist variation produces changes in niche width (b). An identical template polynomial is shown in each panel as a black dashed reaction norm. The transitions from blue to red curves illustrate hypothetical contributions of the three modes of variation to adaptation to high temperature.

Thermal reaction norms typically have a common general shape in which performance increases with increasing temperature, reaches a maximum at some intermediate temperature, and then declines rapidly with further increases in temperature [26]. Based on knowledge of this common shape, evolutionary physiologists have proposed modes of variation in thermal reaction norms that are of particular biological interest (Fig. 2.1): vertical shift (in average fitness), horizontal shift (in optimal temperature), and generalist-specialist variation (niche width) [27]. A recent study of natural populations of *Peiris rapae* caterpillars successfully decomposed the quantitative genetic variation in thermal reaction norm shape into these three modes of biological interest, and found that most of the variation was due to generalist-specialist variation

[19]. Caterpillar growth rates were either intermediate across a wide range of temperatures, or high at intermediate temperatures and low at extreme temperatures. Studies like this one address the nature of standing variation in thermal reaction norm shape, however, the relative contributions of these different modes to the process of adaptation to novel thermal environments are unknown. For example, is adaptation to high temperatures due primarily to evolutionary changes in average performance, in optimal temperature, in niche width, or in a combination of these modes? How do particular genetic changes selected during evolution contribute to these modes?

Studies of species found naturally in contrasting thermal environments can provide some insight into these questions. For example, comparisons of the effect of temperature on growth rate of actinobacteria strains [28] and the photosynthetic ability of unrelated plant species [25] found in different thermal environments both suggest that adaptation to high temperature occurred primarily via a shift in the optimum temperature. However, these observed patterns are the result of a long history of selection on these species, and do not provide information as to which modes of variation were responsible for the initial adaptation to high temperature (*e.g.*, mutational decay could explain why species no longer perform well at temperatures they no longer experience).

Experiments in which laboratory populations have been adapted to a novel temperature should provide a more powerful means for assessing the contributions of changes in average performance, in optimal temperature, and in niche width to thermal adaptation. Indeed, bacteria and bacteriophage model systems have made considerable

progress toward this goal. Investigations of thermal adaptation in *E. coli* have provided thorough characterizations of phenotypic evolution, measuring thermal reaction norms with sufficient detail that a visual inspection suggests contributions from multiple modes of variation [17,22,29]. However, the relatively large genome of *E. coli* has made it difficult to associate reaction norm shape changes with particular genetic changes (but see [30]). In contrast, experiments that investigated thermal adaptation in bacteriophage provide detailed analyses of the genetic bases of adaptation, but only a limited characterization of reaction norm evolution [31,32]. Investigations in both model systems failed to analyze thermal performance using continuous reaction norms, thus, limiting their ability to characterize the nature of reaction norm shape changes [19].

To illustrate the power of analyzing evolutionary responses to temperature using continuous reaction norms, we revisit a recent study of high temperature adaptation in the bacteriophages G4 and phiX-174 [31] that stands out because it was able to associate adaptation with particular genetic changes. Holder and Bull demonstrated that phage populations readily evolved higher growth rates at high temperature, but characterized the correlated responses in growth rates only at the ancestral temperature optimum. Here we reexamine the adaptation that occurred in their G4 lineage, measuring performance over a wider temperature range in order to determine how the thermal reaction norm responded to selection for growth at high temperature. We employed a statistical method called Template Mode of Variation [19] to decompose the evolution of reaction norm shape into vertical shift, horizontal shift, and generalist-specialist variation. This approach enabled us to quantify the contribution of each mode to adaptation to high temperatures, to

identify how specific mutations affect variation in thermal reaction norms, and to identify genetic trade-offs that were not apparent in the earlier analysis. In addition, we investigated whether one of the mutations that had a large effect on reaction norm shape contributes similarly to genetic variation in thermal performance in natural bacteriophage populations.

Experimental Design

In the evolution experiment conducted by Holder and Bull [31], a single population of the bacteriophage G4 was adapted to the inhibitory temperatures of 41.5°C and 44°C and monitored for nucleotide changes throughout its entire 5 kb genome (experimental design is described in detail in Fig. 2.2). Phages were evolved by serial transfer into fresh cultures of rapidly dividing bacterial hosts to select for phage genotypes with the fastest rates of reproduction. Holder and Bull ensured rapid adaptation by selecting for growth beyond G4's thermal niche boundary, and by maintaining large population size ($N_e \sim 10^6$), to minimize the effect of genetic drift. Monitoring evolution during an initial 50 serial transfers (approximately 100 generations) at 41.5°C, and a further 50 serial transfers at 44°C, Holder and Bull confirmed that adaptation to high temperature occurred rapidly and that it was accompanied by a large number of genetic changes.

In the current study, we isolated five individual phages from this evolving population at transfers 20, 50, and 100. We sequenced the genomes of these phages, and used the isolate most representative of the population consensus at each transfer (as

described in Holder and Bull [31]) for all further analyses. Following the convention from Holder and Bull [31], the notation $G4_t$ will be used to represent the phage isolated from serial transfer *t*. Thus, $G4_0$ represents the phage genotype used to initiate selection at 41.5°C, and $G4_{100}$ represents the phage isolated following 100 transfers (50 transfers at 41.5°C plus 50 at 44°C). Our characterization of evolutionary responses to temperature in this population expanded on that of Holder and Bull [31] by measuring the growth rates of $G4_0$, and the evolved phage, $G4_{20}$, $G4_{50}$, and $G4_{100}$, at numerous temperatures that span the thermal niche of G4. Our measure of growth rate is a measure of absolute fitness in the experimental conditions (see Methods).

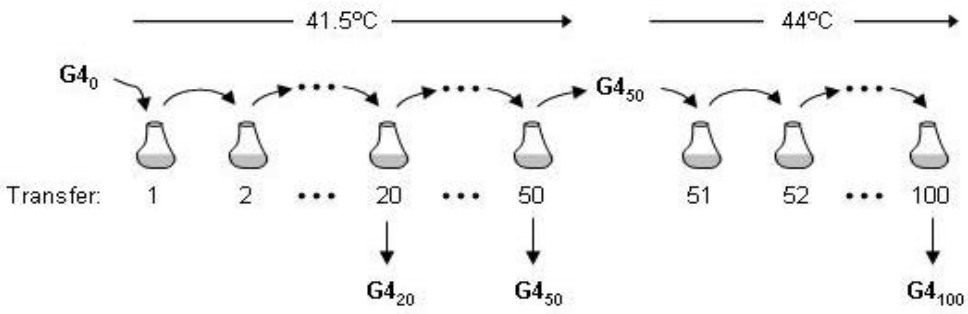


Fig. 2.2. Experimental Design. Holder and Bull used a clone of the bacteriophage G4 to found a population that was propagated through 50 serial transfers at 41.5°C, followed by 50 serial transfers at 44°C. In each serial transfer 10^6 phage were added to a 10 ml culture containing 10^9 exponentially growing *E. coli* hosts, incubated for approximately 45 minutes at the specified temperature, and then treated with chloroform to kill any remaining hosts. During the 45 minute incubation, the phage population size increased to approximately 10^{10} phage, and 10^6 of these phage were used to initiate the next serial transfer. The only exception occurred after the 50^{th} serial transfer, when the population was forced through a single phage bottleneck to homogenize the population. In the current study, we isolated single phage from this evolving population at transfers 20, 50, and 100 (designated $G4_{20}$, $G4_{50}$, and $G4_{100}$). We characterized the performance of these phage and the ancestral G4 (designated $G4_0$) across a range of temperatures.

Methods

Strains and Culture Conditions

G4 populations from the Holder and Bull [31] experiment were obtained from J. Bull (University of Texas, Austin, TX). Clonal isolates from these populations were designated $G4_{0,}$ $G4_{20,}$ $G4_{50}$, and $G4_{100.}$ The G4-like phages, NC6, NC19, ID52, and ID13, were isolated from natural phage populations in 2001 and 2002 [33], and were obtained from H. Wichman (University of Idaho, Moscow, ID).

Culture conditions were as in Holder and Bull [31]. Briefly, phage were grown on *E. coli* C, the standard laboratory host of G4, in LB (5 g Yeast extract, 10 g Bacto tryptone, 5 g salt/ 1 Liter) broth supplemented with 2 mM $CaCl_2$, and on LB plates (15% agar). LB top agar (0.7% agar) was also supplemented with 2 mM $CaCl_2$. Phage were stored for short term (< 1 week) at 4°C in this growth media, and for long term at -20°C in LB broth supplemented with 2 mM $CaCl_2$ and 40% glycerol.

Growth Rate Assays

Growth rate assays were designed to closely follow the selection conditions in Holder and Bull [31]. We used 300 µl of a stationary phase *E. coli* culture to inoculate 10 ml of LB, and incubated this fresh culture in a water bath at the experimental temperature until the culture reached an optical density (OD_{600}) of 0.6 ± 0.1, which corresponded to $2\text{-}3 \times 10^8$ cells/ml at most temperatures (at 44°C an OD_{600} of 0.6 corresponded to 9×10^7 cells/ml). Approximately 10^6 phages were added to this exponentially growing culture, and the mixture was grown for 45 minutes, at which point a 1ml sample was treated with

50 µl chloroform to kill bacteria. Phage numbers at the start (N_0) and end (N_{45}) of the assay were determined by plating. Growth rate was calculated as the increase in phage number over 45 minutes (N_{45}/N_0). Both the experimental serial transfers and the growth rate assays were initiated with a low ratio of phage to bacteria (~0.001) to minimize the possibilities for competition, and therefore, frequency dependent selection. In this way, we ensure that growth rate is a valid measure of absolute fitness in the experimental conditions. Although the temperatures reported here may have differed slightly from the temperatures reported by Holder and Bull [22], they were measured with a Factory-Calibrated Thermometer that is accurate to within 0.1ºC (Fisher Scientific, Hampton, NH).

Analysis of Reaction Norm Shape

Reaction norm shape was analyzed using a statistical method called Template Mode of Variation (TMV) [19]. TMV is a hypothesis driven decomposition of variation in which the principle modes of variation are predetermined by the experimenter. In our case these modes of variation are vertical shift, horizontal shift, and generalist-specialist variation. Because some of these modes are non-linear, the mean curve is not a good representation of the center of the variation in the data. Therefore, TMV estimates a template shape that represents the center of the distribution in the non-linear space of interest.

TMV constrains the template polynomial $P[t]$ in three ways to ensure a fit with biological reality. First, the first order polynomial coefficient of the template shape is

held at zero to cause the template shape to have a maximum at $t = 0$. This condition makes the m_i parameters identifiable and interpretable as the location of the maximum of the curve corresponding to genotype i. Second, TMV requires that $P[t] = 0$ for two values in the neighborhood of $t = 0$, so that there is a positive finite area under the curve. Finally, TMV requires that $\Sigma a_i = 0$, to ensure a unique fit. As a result, the template and all of the curves are constrained to have a positive maximum and a finite area under the curve in some neighborhood near the maximum. Even with these constraints, the observed data for a genotype could be monotonic within the measured temperature range (the m_i estimated for such a genotype would be outside the measured range). Thus, many possible curve shapes are allowed by the template.

We used TMV to fit the data to a model (eqn 1) in which both the common template polynomial $P[t]$ and the parameters a_i, m_i, and b_i for each genotype are unknowns. In practice, the model was fit by iteratively optimizing the template polynomial $P[t]$ and the variation parameters a_i, m_i, and b_i until the sum of squared errors are minimized. We evaluated the fit of the estimated template polynomial to the data by rescaling the data with respect to the parameter estimates (a_i, m_i, b_i) for each genotype, and measuring the residual variation around the template polynomial $P[t]$. We chose to model the common template shape using a polynomial of degree 3 because it was the lowest degree polynomial for which the residuals were random in direction and similar in magnitude across all temperatures, confirming that the polynomial gives a good approximation of the common template shape.

For further details, see a recent study that applied TMV to thermal performance data in natural populations of *Pieris rapae* caterpillars [19].

Sequencing

Genome sequences were obtained from polymerase chain reaction (PCR) products. Each genome was amplified in two overlapping segments using recombinant *Taq* DNA Polymerase (Invitrogen, Carlsbad, California), purified with ExoSap-It (United States Biological, Swampscott, MA) and its sequence was determined from 16 overlapping chain-termination sequencing reactions (UNC Genome Analysis Center and UNC Biology ABI 3100- Avant). The primers used for sequencing reactions were as follows, with the nucleotide position indicating the 5' end, plus or minus indicating which strand (positive or negative) is generated from that primer, and the length of the primer in parentheses. *PCR primers: 49+(21), 2702+(21)and 2495+(21), 225+(19); Sequencing primers: -49(21), 225(19), 548(19), 937+(21), 1447+(21), 1997+(21), 2495+(21), 2702+(21), 3162+(21), 3645+(19), 4050+(19), 4501+(19), 4945+(19), and 5435+(19).* Sequences were aligned and analyzed in Sequencher (version 4.5, Gene Codes Corporation, Ann Arbor, MI). In all cases, the genomic DNA used for PCR was from a single plaque isolate. Up to five isolates were sequenced from the evolved population at each time point. We chose the isolate most representative of the population consensus described in Holder and Bull [31] for use in the present study.

Characterization of genetic variation for thermal performance in natural phage populations.

We used an ANOVA [34] to assess the effects of phylogenetic relationship (as determined by [33]), genotype at the locus of interest, and temperature on the growth rate of natural phage isolates. Phylogenetic relationship and genotype were treated as class variables and temperature was treated as a quantitative variable. Replicates were collected in blocks (on different days), and block was treated as a random class variable. Residuals were normally distributed (Shapiro-Wilk test: $W = 0.98$, $p = 0.30$) and homogeneous across treatments. The analysis was conducted using the GLM procedure in SAS (version 8, SAS Institute, Cary, NC).

Results

Responses to Selection at High Temperature

To determine the consequences of adaptation to high temperature, we measured the growth rate of the ancestral phage and three evolved phages at six temperatures between 27°C and 44°C (Fig. 2.3A). The evolved genotypes were isolated from the Holder and Bull [20] evolving population after the 20^{th}, 50^{th}, and 100^{th} serial transfers at high temperature, and a visual inspection of the data indicates that evolution occurred between each of these time points (Fig. 2.3A). Reaction norms generated for all genotypes had the overall shape that is characteristic of thermal reaction norms for fitness and performance [27]. Growth rate increased with temperature until a maximum was achieved around 35°C and declined as temperature increased above this maximum (Fig. 2.3A). Note that either a difference in temperature calibration, or an inadvertent 2-fold

difference in salt concentrations produced a difference in phage growth rates between our study and that of Holder and Bull [31]. From our measures, phage did not exhibit positive growth rates at 41.5°C until transfer 20, or at 44°C until transfer 100. This calibration difference should not affect the measured shape of thermal reaction norms, it should only affect our ability to pinpoint the exact temperatures to which these phages were adapted.

An examination of the response to selection at 41.5°C (transfers 0–50) and at 44°C (transfers 50–100) revealed that evolutionary responses were greater at temperatures above 35°C than at temperatures below 35°C (Fig. 2.3B). Correlated responses to selection at both 41.5°C and 44°C were positive at temperatures above 35°C. By contrast, evolutionary responses at temperatures below 35°C were uncorrelated with the response to selection at 41.5°C, and negatively correlated with the response to selection at 44°C (Fig. 2.3B).

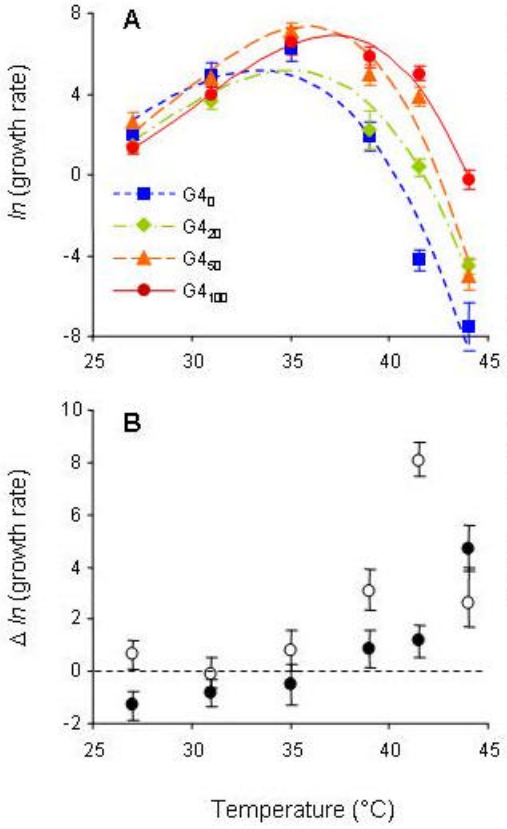


Fig. 2.3. Correlated responses to adaptation at high temperature. (A) Reaction norms estimated using Template Modes of Variation explain 71.79% of the sums of squares in the data. Data are mean growth rates ± s.e.m. for each phage genotype determined from 3 or 4 replicate measures at each temperature. (B) Correlated responses to selection at 41.5°C (transfers 1–50; open circles) and to selection at 44°C (transfers 51–100; closed circles). Data are the difference in mean ln(growth rate) ± s.e.d. between G450 and G40, and between G4100 and G450, respectively.

Evolution of Reaction Norm Shape

We analyzed reaction norm shape using a statistical method called Template Mode of Variation (TMV, [19]). Briefly, the TMV approach assumes that each of the measured reaction norms can be described by shifting and stretching a common template reaction norm, represented by a polynomial $P[t]$ of temperature t. The growth rate measures for all genotypes at all temperatures were simultaneously fit to the model:

$$P_i[t] = a_i + \tfrac{1}{b_i} P[\tfrac{1}{b_i}(t - m_i)], \qquad (1)$$

in which both the common template $P[t]$ and the parameters a_i, m_i, and b_i (representing, respectively, the average growth rate, optimal temperature, and niche width for each genotype i) are unknowns.

In this model, differences among genotypes in average growth rate (a_i) are modeled by vertical shifts of the template reaction norm (Fig 2.1A). Average growth rate is the only parameter that does not incorporate a trade-off in performance across temperatures. Differences in optimal temperature are modeled using horizontal shifts that move the entire reaction norm left and right (Fig 2.1B), positioning the optimum for each genotype at m_i without stretching or skewing the reaction norms. Differences in niche width are modeled by assuming a generalist-specialist tradeoff that constrains the positive area under the growth rate curve $P_i[t]$ to be constant. Thus, increases in niche width yield both a wider temperature range and a lower maximum growth rate (Fig 2.1C).

The model (equation 1) produced by the TMV analysis provides a good fit to the growth rate data at each time point (Fig. 2.3A). The reaction norms conformed to a common shape that was well approximated by a polynomial of degree 3 (see Methods). The contributions of the three modes to the variation in shape are illustrated in Figure 2.4. Optimal temperature showed the most consistent pattern of evolutionary change, increasing from 33.4 to 37.1°C during the experiment, and explained the largest proportion (47.38%) of the shape variation (Table 2.1). Niche width and average growth

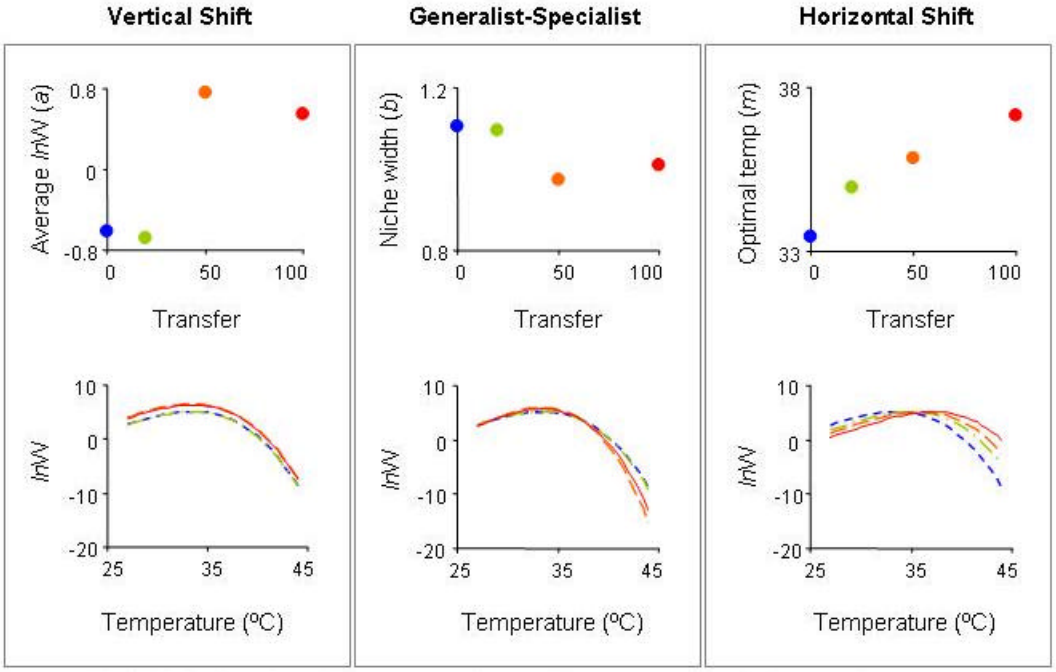


Fig. 2.4 Modes of variation among the reaction norms of evolving genotypes. TMV was used to decompose the variation among the reaction norms described in Figure 3 into three modes of variation. In each panel, the top graph depicts the values of the fitted parameters from the model and the bottom graph illustrates the impact of each parameter, if considered alone, on the shape of the reaction norm of $G4_0$. lnW is the ln(growth rate).

rate revealed more complex patterns of change (Fig. 2.4, top), and explained 12.65 and 11.75% of the shape variation, respectively. In total, the modes of variation explained 71.79% of the shape variation, with horizontal shift explaining most of the evolution of reaction norm shape (Table 2.1).

Table 2.1. Decomposition of Shape Variation

Source of Variation	Ratio of SS (%)[a]
Vertical Shift	11.75
Horizontal Shift	47.38
Generalist-Specialist	12.65
Modes of Variation Total	71.79

[a]Variance explained by the specified mode of variation relative to the total Shape Variation (SSVS + SSHS + SSGS + SSError).

Genetic Basis of Reaction Norm Evolution

In order to identify the mutations responsible for changes in reaction norm shape, we sequenced the genomes of the ancestral phage $G4_0$ and the evolved phages $G4_{20}$,

Table 2.2. Genetic basis of adaptation to high temperature.

Selective Temperature	Gene	Function	Mutation	$G4_{20}$	$G4_{50}$	$G4_{100}$	Frequency in Nature[d]
41°C	A	replication	D62D[b]	•	•	•	0.33
	A*	replication	T221I	•	•	•	0
	F	host attachment	P355S	•	•	•	0.07
	F	host attachment	H400Y	•[c]			
	H	DNA ejection	G142D	•	•	•	0
	H	DNA ejection	A47V		•	•	0.13
44°C	A	replication	T23A			•	0
	A*	replication	V239I			•	0
	B	morphogenesis	T104A			•	0
	D	morphogenesis	V9L			•	0
	intergenic		3916AG			•	0.13

[a]These mutations also affect gene A* which overlaps A, but is in the same reading frame.
[b]Synonymous mutation.
[c]Not detected by Holder and Bull [28].
[d]Frequency among 15 G4-like phages isolated from nature [30].

$G4_{50}$, and $G4_{100}$. Genome sequences indicated that 10 mutations fixed during the 100 serial transfers. Of these, four were present in $G4_{20}$, one additional mutation was present in $G4_{50}$, and the remaining 5 mutations appeared in $G4_{100}$ (Table 2.2). Although it was not present in phage from later time points, one additional mutation was present in $G4_{20}$. The location of the 10 mutations around the genome showed no significant deviation from a random distribution among the non-overlapping genes ($?^2 = 3.41$; df = 7; $p > 0.85$).

An examination of the genomes of 15 natural isolates of G4-like phage [33] indicated that four of the ten adaptive mutations are polymorphic in natural populations (Table 2.2). This observation allowed us to explore the possibility that variation at these sites might underlie variation in thermal performance in natural populations. To address this question, we focused on the single mutation present at transfer 50 that was not already present at transfer 20 – gene H: A47V. This is the only mutation whose acquisition was unambiguously associated with an increase in growth rate at the selective temperature (41.5°C). All other increases in growth rate were associated with the

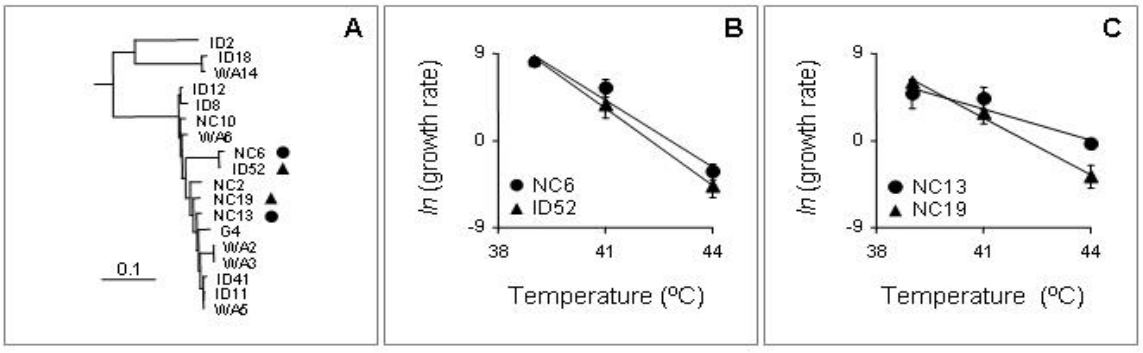


Fig. 2.5. Growth rates of phage isolated from nature at high temperatures. (A) Phylogenetic relationships of G4-like phage recently isolated from environmental samples, modified from [22]. The genotypes of two pairs of sister taxa that differ at a locus (gene H: A47V) determined to have a large effect on reaction norm shape are indicated with triangle (alanine) and circle (valine). **(B** and **C)** Thermal reaction norms of the two phage pairs. Data are means ± s.e.m. determined from four or five replicate measures at each temperature.

acquisition of multiple mutations. We identified two pairs of closely related phage from the natural isolates that differ at this gene H locus (Fig. 2.5A). In each pair, one genotype encodes the mutant amino acid (valine) and the other genotype encodes the ancestral amino acid (alanine). We measured the growth rates of these phages at 39, 41, and 44°C and found that growth rate depended on temperature, phylogenetic relationship (pair), and genotype at the locus of interest, with significant temperature × pair and temperature × genotype interactions (Table 2.3). Similar to the outcome of the evolution experiment, the temperature × genotype interaction among natural phage isolates resulted because the A47V mutation was associated with higher growth rates at high temperatures (41 and 44°C), but not at intermediate temperature (39°C) (Fig. 2.5).

Table 2.3. ANOVA describing the effect of gene H mutation A47V in natural phage isolates.

Source of Variation	SS	df	MS	Type III F	Pr(>F)
Block	17.65	4	4.41	1.01	.4118
Pair[a]	52.31	1	52.31	11.99	.0012
Temperature	934.32	1	934.32	214.10	<.0001
Genotype	18.58	1	18.58	4.26	.0449
Pair x Genotype	4.26	1	4.26	.98	.3283
Temp x Pair	50.99	1	50.99	11.68	.0013
Temp x Genotype	20.44	1	20.44	4.69	.0358
Temp x Pair x Genotype	4.36	1	4.36	1.00	.3227
Error	196.38	45	4.36		

[a]Each pair consists of two closely related wild phage (i.e. sister taxa), one with the ancestral allele (A) and the other with the evolved allele (V) at amino acid 47 in gene H.

Discussion

Holder and Bull [31] demonstrated that the phage G4 adapted rapidly to high temperatures (41.5-44°C) from an ancestral temperature of 37°C, and identified specific genetic changes associated with adaptation. They also showed that the increased fitness

at high temperatures was not correlated with decreased fitness at the ancestral temperature. In our study we considered adaptation of G4 to high temperature in the context of the evolutionary changes in thermal reaction norms over a wide temperature range (27-44°C). Our measurements indicate that evolutionary increases in fitness (growth rate) at the selected temperatures (41.5 and 44°C) were associated with correlated increases in fitness at temperatures of 39 ºC and above, but with no change or decreases in fitness at temperatures of 35°C and below (Fig. 2.3B). By measuring growth rate over a wider range of temperatures, we were able to detect correlated effects of selection on performance at high temperatures that were not detected by Holder and Bull [31]. However, our data are consistent with the conclusions of Holder and Bull because negative correlations arose over only a small subset of the temperature range, and from only a subset of the genetic changes.

To provide a more detailed picture of the evolutionary changes in thermal sensitivity during adaptation, we applied the TMV statistical method [19] to our data. TMV takes advantage of the continuous nature and common shape of thermal reaction norms to decompose reaction norm evolution into changes in three underlying parameters: average growth rate (*a)*, niche width (*b)*, and optimal temperature (*m*). Our analysis indicates that changes in average growth rate, niche width and optimal temperature all contributed to the evolutionary changes in thermal reaction norms in this study (Fig. 2.4, top), and that these parameters contributed differently to the evolutionary changes in growth rate at different temperatures (Fig. 2.4, bottom). For example, adaptation (increased growth rate) at higher temperatures was primarily due to

evolutionary increases in optimal temperature, with only small contributions from the other two parameters. By contrast, increased growth rates at temperatures near 35°C were due largely to evolutionary increases in average growth rate.

Both horizontal shifts (optimal temperature) and generalist-specialist tradeoffs (niche width) can generate patterns of positively and negatively correlated changes in performance at different temperatures (*e.g.* Fig. 2.3B); by using TMV were we able to quantify the relative contributions of these modes of variation to reaction norm evolution [19]. The TMV analysis confirmed that the largest and most consistent evolutionary changes were increases in optimal temperature (Fig. 2.4, top), which accounted for 47.38% of the shape variation in the data (Table 2.1). In contrast, generalist-specialist variation (niche width) and vertical shifts (average growth rate) contributed only 12.65% and 11.75%, respectively, of the shape variation (Table 2.1).

It is worth considering whether the contribution of vertical shifts to reaction norm shape evolution in this laboratory population was artificially low because it had been previously adapted to 37°C, the standard lab condition [31]. Adaptation to the standard lab condition is a common component of microbial evolution experiments, and in this case it was intended to exhaust the adaptive genetic variation that was not temperature specific – i.e. the adaptive genetic variation composed only of vertical shift. Our data are consistent with the expectation that vertical shifts that occur during adaptation to the experimental temperatures should occur only in combination with changes in another mode of variation.

Although the physiological basis of the thermal adaptation is unknown, horizontal shifts are consistent with the idea that increases in protein stability at high temperatures result in decreased protein activity (*e.g.* enzymatic reaction rates) at lower temperatures [8]. In addition, the random distribution of mutations among genes is suggestive of selection acting for increased protein stability because such selection may be expected to act equally on all proteins. In contrast, selection for particular activities relevant for growth high temperature would likely act with different strengths on different proteins. If this biochemical mechanism does explain most adaptation to novel temperatures, it is surprising that it has been so difficult to detect trade-offs in fitness in animals. It is tempting to think that the simplicity of the virus life cycle underlies the consistency of our finding of horizontal shifts and the expected biochemical underpinnings of thermal performance trade-offs. This is a hopeful perspective because horizontal shifts can be exploited in the development of live cold-adapted vaccines, as in [35]. This vaccine development strategy adapts viral strains for growth at temperatures substantially below human body temperature in an attempt to generate viruses that grow well in cell culture, but are avirulent in humans.

Investigating thermal adaptation and constraint

Temperature adaptation is generally assumed to be governed by performance trade-offs across temperatures [15,36], despite a shortage of data demonstrating their existence [6]. From the existing data, it is difficult to determine whether trade-offs are in fact rare, or whether the data and analysis methods (usually multivariate analyses of

variance in performance at a few temperatures) were not sufficiently powerful to identify trade-offs.

The primary evidence for the latter explanation is that studies that measure performance over wider temperature ranges more often identify trade-offs [16]. Whereas studies that measured performance across most of the thermal niche have identified thermal performance trade-offs in bacteriophage (this study), bacteria [17,20], and insects [16,21], counterparts that measured performance over only a portion of the thermal niche often failed to identify trade-offs (*e.g.* bacteriophage: [31]; bacteria: [37]; insects: [38]. Similarly, the power of our analysis derived from the measurement of performance over a wide temperature range. Had we measured performance in this study only at temperatures between 37°C and 44°C, the TMV method would have misleadingly attributed most of the variation in reaction norm shape to vertical shift, and we would have concluded that trade-offs did not influence reaction norm shape variation.

The genetic basis of adaptation in laboratory and natural populations

Whereas adaptation in the laboratory population described here resulted from selection acting on new mutations, adaptation in nature may often result from adaptive genetic variation that is not new, but already existing in the population. Whether the adaptive mutations identified here also contribute to genetic variation in natural populations will depend in part on whether their effects are neutral (or mildly deleterious) in the natural thermal environment. A comparison of the novel genetic variation generated in lab adaptation to the standing genetic variation present in nature suggests

this may be the case. We investigated the polymorphism present in natural populations at sites that contributed to adaptation in the lab by examining a collection of 15 G4-like phage recently isolated from nature [33]. Four of the 10 adaptive mutations were polymorphic among these phages. Most of the polymorphisms were rare, present at frequencies of 0.07 or 0.13, as would be expected if these mutations are neutral or slightly deleterious in nature. Furthermore, if we assume that the temperature optimum of 'lab-naïve' G4-like phage around 29^{o}C (unpublished data) is indicative of the natural thermal environment, then the minimal effects of the lab adaptive mutations at this temperature (Fig. 2.3, raw data) also suggest their near neutrality in nature.

To examine whether these mutations do contribute to variation in thermal performance in natural populations, we identified one mutation that both contributed conclusively to adaptation in the lab and existed in multiple natural phage isolates. Most of the mutations from the laboratory population of G4 occurred in temporal clusters so we could not identify their individual effects. However, a single adaptive mutation in gene H (A47V) occurred between transfers 20 and 50. This mutation is also present in two of the 15 G4-like natural isolates. Similar to its effects in lab adaptation, the presence of this gene H mutation in the natural isolates is associated with thermal reaction norm shape variation in phages with and without the mutation. In both the experimentally adapted population and among phages isolated from nature, this mutation was associated with higher growth rates at high temperatures, but not at low temperatures. Only one of the other mutations that contributed to lab adaptation was differentially represented among the hot adapted (NC6 and NC13) and the cold adapted

(ID52 and NC19) natural isolates. That mutation, gene F: P335S, was found only in phage NC19, which has an ancestral (cold adapted) allele at gene H: A47V. If anything, the fact that this gene F mutation appears in NC19 should have worked against the finding that A47V is associated with adaptation to high temperature in natural phage isolates.

The observation that experimental evolution can result in molecular changes that converge on natural isolates has been made before. Wichman et al. [39] made similar observations during adaptation of the related phages phiX-174 and S13 to alternative hosts. It is possible that these results are unique to phage because of their large population sizes and/or small genomes. However, it is as likely that these observations have been made only in phage systems because, until recently, it has only been possible to identify the genetic basis of lab adaptations in phage [40,41]. The ease with which ours and the previous study identified genetic variations that are associated with phenotypic differences among wild phage suggests that laboratory evolution experiments may often predict the precise genetic basis of adaptations in nature.

CHAPTER 3. Hotter is better: Investigating thermodynamic constraints on phage growth rates.

The work described in this chapter was accomplished in collaboration with Dr. Christina Burch. I anticipate submitting this work to American Naturalist sometime in the month after my defense. This work was greatly improved by conversations with Dr. Joel Kingsolver, the Function Valued Group, the UNC Evolunch group, and members of the Burch Lab – particularly Martin Ferris. Technical and statistical assistance was graciously provided by Drs. J. Steve Marron, Jack Weiss, Rima Izem, and Helen Olofsson. Katie Supler contributed to the collection of preliminary data.

Introduction

One central aim of evolutionary biology is to identify patterns of adaptation and use this knowledge to predict evolution. Adaptation to temperature is an important and useful model for investigating such patterns due to natural populations experiencing predictable and unpredictable variation in temperature [3,4] and its connection to biochemical reaction rates. The growth rates of individuals at different temperatures depends, in part, on their biochemical reaction rates at those temperatures [42,43]. For example, low temperatures typically depress reaction rates (and thus growth rate) in comparison to those at high temperatures simply because the collision rate of reactants is lower [42]. It is not known if adaptation at low temperatures can overcome this depression in rates of reaction.

In this paper, we focus on a hypothesis of thermal adaptation arising from this potential constraint on adaptation at cold temperatures. This hypothesis, termed 'hotter is better', stems from the assumption that adaptation is unable to overcome the rate-depressing effects of low temperature, therefore causing populations adapted to lower temperatures to have lower maximum growth rates than those adapted to higher temperatures. This hypothesis predicts a positive correlation between a species optimal temperature and its maximum growth rate [44] [45] [27] (Fig. 3.1). If this hypothesis is true, then exotherms adapted to warmer environments should have faster growth rates than those adapted to colder environments. It may also explain why endotherms regulate at high body temperatures [45].

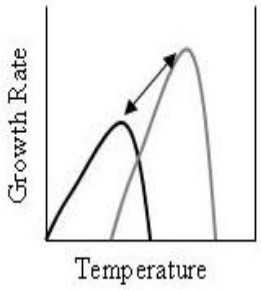


Fig. 3.1. Hotter is better. Shown is the predicted relationship between temperature and maximum growth rate. The cold-adapted genotype (black) has a lower thermal optima and a lower maximum growth rate than the hot-adapted genotype (gray).

While comparative studies on lizard sprinting speed [46], pine tree growth [47], and scallop metabolism [48] are consistent with 'hotter is better', to date only one study has explicitly tested this hypothesis. In this inter-specific meta-analysis of insect growth rate, Frazier and colleagues [7] were able to test 'hotter is better' because for each of 65 insect species they were able to estimate a continuous thermal reaction norm, which is a curve relating a continuous trait (*e.g.* growth rate) to temperature. From these reaction norms, the optimum temperature and maximum growth rate were estimated for each of the insect species. Consistent with the 'hotter is better' hypothesis, a significant and positive correlation existed between the species optimal temperatures and maximum growth rates. However, it is unknown whether the 'hotter is better' relationship also characterizes variation within a population or whether the 'hotter is better' relationship is a general pattern across unrelated species.

Here, we explore patterns of thermal adaptation, focusing on an intra-specific test of 'hotter is better' in a population of bacteriophages. The growth rates of 15 related G4-like phages that were isolated from nature were measured with a high level of precision over a wide temperature range. An analysis of these thermal reaction norms showed that most of the variation in the phages growth rates is at high temperatures. As predicted by

hotter is better, the correlation between the phages optimal temperatures and maximum growth rates is strong and positive – even after we correct for the relatedness of these phages. We are also able to rule out one alternative hypothesis for this pattern. Our results build on the findings of Frazier et al. [7] by confirming that hotter is better also characterizes intra-specific variation in thermal reaction norms. By demonstrating that this pattern is true of intra-specific variation, we have taken the first step towards determining the genetic basis of this pattern.

Methods

Strains and Culture Conditions

The natural isolates of the G4-like phage are described by Rokyta et al. [33], and were obtained from H. Wichman (University of Idaho, Moscow, ID). Culture conditions were as in Holder and Bull [31]. Briefly, phage were grown on *E. coli* C, the standard laboratory host of G4, in LB (5 g Yeast extract, 10 g Bacto tryptone, 5 g salt/ 1 Liter) broth supplemented to 2 mM $CaCl_2$, and on LB plates (15% agar). LB top agar (0.7% agar) was also supplemented to 2 mM $CaCl_2$. Phage were stored for short term (< 1 week) at 4°C in this growth media, and for long term at -20°C in LB broth supplemented to 2 mM $CaCl_2$ and 40% glycerol.

Growth Rate Assays

Growth rate assays were designed to mimic those of Knies et al. [12]. 60 µl of a stationary phase *E. coli C* culture was used to inoculate 2 ml of LB, and incubated at the experimental temperature until the culture reached an optical density (OD_{600}) of 0.6 ±

0.06, which corresponds to 2-3 $\times 10^8$ cells/ml [12]. Approximately 10^3 - 10^4 phage were added to this exponentially growing culture, and the mixture was grown for 45 minutes, at which point the culture was treated with 100 µl chloroform to kill the *E.coli.* Phage numbers at the start (N_0) and end (N_{45}) of the assay were determined by plating and incubating the plates until well formed plaques were present. Since several of the natural isolates were temperature sensitive at 37°C, plates were incubated at either 33°C (2.5 – 4hrs) or 25°C (16 – 24 hrs). Growth rate was calculated as the increase in phage number over 45 minutes: LN (N_{45}/N_0).

Estimation of Block/replicate Effect on Growth Rate

The growth rates of the 15 natural isolates of G4-like phages were measured three times at each of nine temperatures evenly spaced between 17 and 41 °C. A single block consisted of a single measurement of all 15 genotypes at one temperature.

We used an ANOVA (Sokal and Rolf 1995) to assess the effects of block, temperature, and genotype on the growth rate of natural phage isolates. Genotype and block were treated as class variables and temperature was treated as a quantitative variable. All the variables were treated as fixed effects. The analysis was conducted using the MIXED procedure in SAS (version 8, SAS Institute, Cary, NC). The effect of temperature on growth rate was modeled as a 3rd order polynomial that was estimated from the mean growth rate across genotypes in relation to temperature. The effect of block was modeled as an interaction between temperature and block due to block being

nested in temperature. Residuals were normally distributed (Shapiro-Wilk test: $W = 0.98$, $p = 0.30$) and homogeneous across temperatures.

The resulting estimate for each block effect was subtracted from the growth rate measures from that experiment to produce block corrected growth rate measures. The corrected growth rate measures were used for all subsequent analyses.

Analysis of Reaction Norm Shape

Reaction norm curves were estimated with two methods, Template Mode of Variation (TMV) and smoothing splines. The first method assumes that all the curves share a common shaped template. In contrast, the second method does not make the assumption of a common shaped template.

TMV

Reaction norm shape was first analyzed using the statistical method Template Mode of Variation (TMV) [19]. Briefly, the TMV approach fits the data to a model (equation 1) in which both the common template shape $P[t]$ and the parameters (a_i, m_i, b_i) for each genotype are unknowns. The model is fit by simultaneously optimizing the polynomial coefficients and the parameters a_i, m_i, and b_i for each curve to minimize the sum of squared errors. The first order polynomial coefficient of the template shape is held at zero to cause the template shape to have a maximum at zero. This condition makes the parameters m_i identifiable, and interpretable as the location of the maximum of curve i.

$$P_i[t] = a_i + \frac{1}{b_i} P[\frac{1}{b_i}(t - m_i)], \qquad (1)$$

This method has previously been applied to phage s thermal reaction norms [12] and is described in more detail in [19].

Smoothing Splines

Curves were also fit to the reaction norms by spline interpolation. Spline interpolation uses low-degree polynomials in each of the intervals (between each pair of temperature values), and chooses the polynomial pieces such that they fit smoothly together. The smoothness of the splines was estimated from the mean growth rate data (over all genotypes) by means of Leave One Out Cross Validation (LOOCV). LOOCV is a model evaluation method that estimates the error of a function by repeatedly creating the function using all but one of the data points. This single "left out" data point serves as a test which is used to compute the test error. The test error is the difference between the function output given the left out pattern and the desired output. The cross-validation error is the average of all individual errors calculated from the different functions tested on their left out pattern [49]. For our model fitting, the different functions tested were chosen to have different smoothing parameters: 0.05, 0.1, 0.2, 0.4, 0.8, and 1.6. The global smoothing parameter that minimized cross-validation error was 0.20. This smoothing error was then used to fit splines to each dataset using the *csaps* function in

Matlab (version 6.5, The Mathworks Inc.) Several of these curves were bimodal and thus had two optimal temperatures. For the purposes of this experiment, the optimum having the higher maximum growth rate was considered to be the thermal optimum.

Principal Component Analysis

Reaction norms fit to the growth rate data of the natural isolates are defined by a common template shape and the parameters a,b, and m (see equation 1). These parameters were subjected to principal components analysis (PCA) based on their correlation matrix. The analysis was conducted using the PRINCOMP procedure in SAS (version 8, SAS Institute, Cary, NC).

Correlation between thermal optima and maximum growth rates

The phages sequences were available on PubMed (accession numbers are reported in [33]). A phylogeny for these phages was obtained with MRBAYES v3.1.2[50] using the GTR + gamma + invariant model of nucleotide substitution and default parameters. Convergence was achieved in 100,000 generations with a 10% burn-in.

The correlation measure between the thermal optima and maximum growth rates was estimated using the method of least squares. This analysis was performed in Microsoft (Redmond, WA) Excel2004. This correlation was repeated after controlling for phylogenetic relatedness using standardized independent contrasts [51] computed with Phenotypic Diversity Analysis Programs [52] [53]. Independent contrasts were computed after exponentially transforming branch lengths to eliminate correlations

between the absolute values of the independent contrasts and their standard deviations [54].

Recombination Test

Recombination between the natural isolates was investigated using the Maximum Chi-Square method [55] [56], as implemented in RDP Version 2.0 [57]. Maximum Chi-Square was identified as one of the most powerful and conservative of numerous recombination detection methods [58] [56]. The window size used was 0.1% of variable sites, gaps were stripped, and the P values were estimated by randomizing the alignment 1,000 times. A maximum P value of 0.05 was considered significant.

Table 3.1. ANOVA describing the effect of block on growth rate of the natural phage isolates.

Source of Variation	Num df	Den df	Type III F	Pr(>F)
Genotype	14	358	7.44	<.0001
Temp	1	358	461.69	<.0001
Temp*Block	25	358	8.93	<.0001

Results

Variation in thermal adaptation

To explore patterns of thermal adaptation in a population of G4-like phages, thermal reaction norms were generated for the 15 G4-like phages by measuring their growth rate multiple times between 17 °C and 41°C (see Appendix A for a description of how this experimental design was determined). The relationship between temperature and mean growth rate (after subtracting block effects - see Table 3.1) is shown in Figure 3.2A. Reaction norms generated for all genotypes had the overall shape that is

characteristic of thermal reaction norms for fitness and performance [18]. Growth rate increased with temperature until a maximum was achieved around 32 °C and declined as temperature increased above this maximum. Substantial variation exists in growth rates between the natural phage isolates, especially at high temperatures.

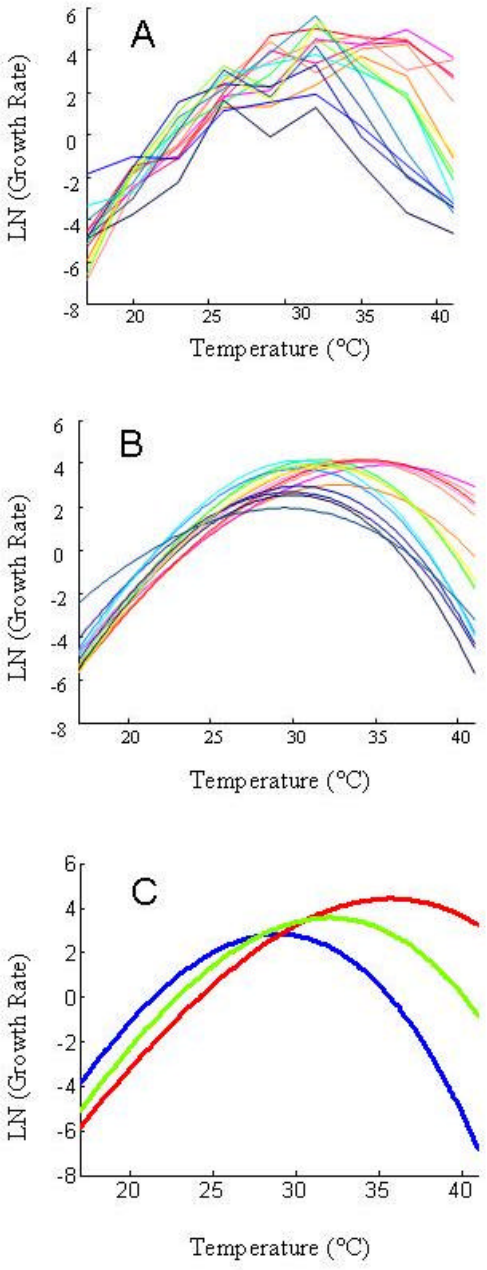


Fig. 3.2. Thermal reaction norms of natural isolates. The line colors are indicative of the thermal optima of the reaction norm, with blue being low and red being high. (A) Shown are the means of the growth rate measures for each genotype. Each line is a different genotype. (B) Continuous reaction norms fit to the raw data by TMV. The curves were constrained to be a polynomial of degree 3. (C) The general nature of these reaction norms as predicted by PC1 of a principal components analysis (PCA). A PCA was performed on the standardized parameters (a,b, and m) that define each curve.

To maximize the accuracy in estimating the shape of these reaction norms, curves were fit to these reaction norms using the statistical method, Template mode of variation (TMV). This method constrains all the curves to have a common template shape and allows the variation in curve shape to be decomposed into biologically relevant modes of variation. The variation in curve shape is achieved via shifting and stretching the curves relative to the common shape. Horizontal shifts in the curves produce changes in the optimum temperature, while vertical shifts and curve stretches produce changes in the maximum growth rate. The reaction norms fit to the data by TMV are shown in Figure 3.2B. Decomposition of the variation in curve shape by TMV explained 66% of the variation in curve shape. Of the total variation, 30.36% is due to horizontal shifts – or changes in the optimum temperature – while another 35.84% was explained by a combination of vertical shifts (17.27%) and stretches (18.57%) of the curve- which reflect changes in the maximum growth rate (Table 3.2). This decomposition indicates that there is variation between genotypes in their maximum growth rates and thermal optima.

Table 3.2. Decomposition of Shape Variation

Source of Variation	Lab: Ratio of SS (%)[a]	Nature: Ratio of SS (%)[a]
Vertical Shift	11.75	17.27
Horizontal Shift	47.38	30.36
Generalist-Specialist	12.65	18.57
Modes of Variation Total	71.79	66.20

[a]Percent of variance explained by the specified mode of variation relative to the total Shape Variation ($SS_{VS} + SS_{HS} + SS_{GS} + SS_{Error}$)

A Principal Components Analysis (PCA) was used to determine whether a general shape characterizes the thermal reaction norms of the phages. The PCA was

performed on the parameter values (a,b, and m) that define the phages reaction norms, according to TMV (see equation 1). PC1 explained 72.1% of the variation in these parameters. PC1 loaded similarly on all 3 parameters- though the load on b was negative (Table 3.3). Using the a,b, and m values that fall along the line define by PC1, it was possible to plot the characteristic shape of the thermal reaction norms (Fig. 3.2C). Genotypes with the highest optimal temperatures are expected to have the highest maximal growth rates. There also appears to be fitness trade-offs across temperatures, such that those genotypes with the highest growth rates at high temperatures have the lowest growth rates at low temperatures.

Table 3.3. Principal Component (PC) Analysis of parameters defining reaction norm shape

Parameter	Component Loadings		
	PC1	PC2	PC3
a	0.58	0.56	-0.59
b	-0.47	0.82	0.32
m	0.66	0.094	0.74
Cumulative Variance (%)	72.1%	97%	100%

To verify a crucial assumption of TMV, which is that the phages thermal reaction norms share a common shape, we also fit curves to the raw data using smoothing splines, which do not assume that the datasets share a common shape. The curves fit to the data using smoothing splines are shown to be very similar to those fit to the data with TMV (Fig. 3.4: compare solid lines to dashed lines).

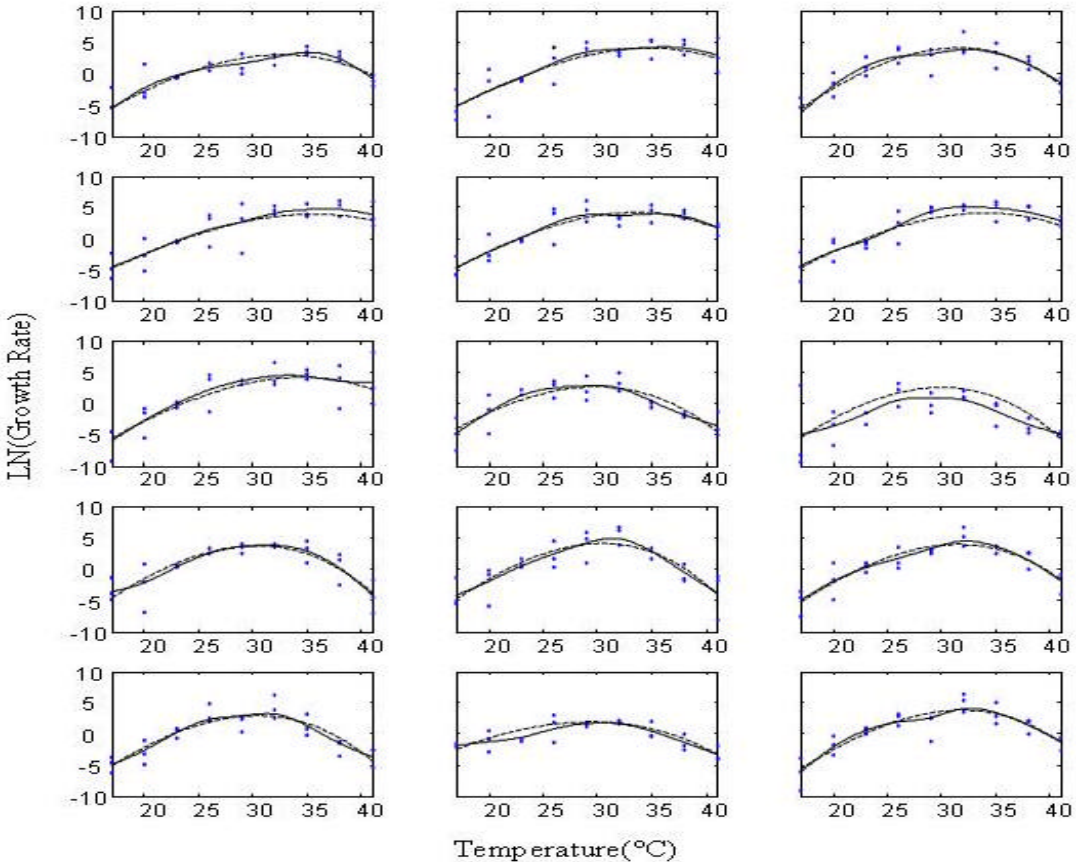


Fig. 3.3. A common shape is shared by thermal reaction norms. Each box shows the growth rate data for one of the G4-like phages and the curves fit to that dataset with either TMV (solid lines) and smoothing splines (dashed lines). TMV assumes all the curves share a common shape (here a polynomial of degree 3). The smoothing splines do not assume a common shape. The common shape of the reaction norms is validated by the close match of the curves fit by TMV and smoothing splines to each dataset.

Hotter is better: A positive correlation exists between a phages thermal optimum and maximum growth rate

To estimate the correlation between the thermal optima and maximal growth rates, these parameters were estimated from the continuous reaction norms fit to the raw data with the statistical method TMV. The estimated optimum temperature varies between 28.7°C and 35.5°C. The estimated *LN* maximum growth rate varies from ~2 – 5 (7.4 – 148.4 fold increase in phages/45 minutes). There exists a significant and positive

correlation between the optimum temperature and maximum growth rate (R^2 = 0.83, 13 df, p <0.0001) (Fig. 3.4). When this correlation is corrected for the relatedness of the phages using Felsenstein's independent contrasts, there is still a significant and positive correlation between optimum temperature and maximum growth rate (R^2 = 0.76, 13df, p < 0.0001). The correlation between the thermal optima and maximum growth rates is weaker, though qualitatively the same when these parameters are estimated from the smoothing spline functions fit to the raw data (R^2 = 0.39, 13 df, p = 0.0126).

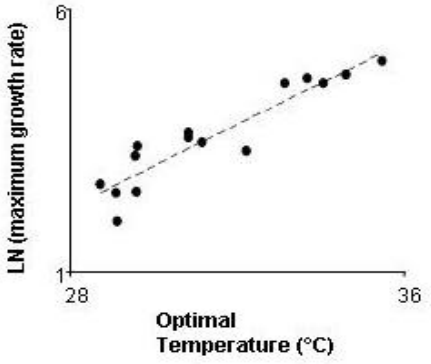


Fig. 3.4. A positive correlation exists between the phages optimum temperature and maximum growth rate. Both of these traits were estimated from the curves fit to each dataset by TMV.

To explore the possibility that this correlation is not due to "hotter is better", we considered whether our results could be due to colder environments favoring thermal generalists. If this was the case, then maximum growth rate might be lower in the natural isolates adapted to lower temperatures as a result of a trade-off between thermal niche and maximum growth rate [36] [59] [13] [60]. Here, thermal niche is defined as the temperature range over which a phages growth rate is greater than 0. We tested this alternative by calculating the thermal niche for each of the natural isolates and measuring the correlation between thermal niche and maximum growth rate. There exists a significant and negative correlation between thermal niche and maximum growth rate (R^2 =0.98, 13df, p< 0.00001), which allows to reject this alternative hypothesis.

It is possible that the mechanism underlying 'hotter is better' is that the phages maximum growth rate is simply a function of the growth rate of its host, *E.coli C*. To explore this alternative hypothesis, we have begun to measure the growth rate of *E.coli C* at high (35.6°C) and low (29.4°C) optimal temperatures of the phage. Pilot data indicates that *E.coli C* grows 57 % faster at the higher optimal temperature as compared to the lower.

Discussion

We explored patterns of thermal adaptation in a population of bacteriophages by measuring the growth the growth rates of these phages over a wide temperature range. A visual inspection (Fig. 3.2A and B) of the reaction norms generated for these phages showed that most of the variation in their growth rates is at high temperatures. Similarly, the phages showed considerable variation in the upper limit (35.7 - 45.87^{o}C) of their thermal niche, but little variation in their lower limit (20.8 - 23.7^{o}C). By characterizing the general nature of the phages reaction norms (Fig. 3.2C), we showed that genotypes with high optimal temperatures tend to have performance trade-offs at low temperatures. In addition these same genotypes are predicted to have higher maximal growth rates than genotypes with low thermal optima.

The general nature of the phages reaction norms (Fig. 3.2C) is consistent with the hypothesis that adaptation cannot overcome thermodynamic constraints at cold temperatures, or 'hotter is better'. These thermodynamic constraints are expected to cause the growth rate of cold adapted genotypes to be lower than the growth rate of hot

adapted genotypes at their thermal optimum. This hypothesis was qualitatively assessed by examining the general nature of the phages thermal reaction norms (as described above) and quantitatively tested by estimating the thermal optimum and maximum growth rates of the phages from their continuous reaction norms. A strong and positive correlation exists between the thermal optima and maximum growth rates of the phages (Fig. 3.4). The demonstration of the positive correlation between optimal temperature and maximum growth rate for very different organisms (phages: our study, insects: [7]; E.coli natural isolates (results not shown) from [20].) supports the mechanism proposed for this pattern – that there are underlying biochemical constraints that limit adaptation to cold temperatures across all organisms.

Alternative Hypotheses

The correlation between the thermal optima and maximum growth rates could be explained by other underlying causes besides thermodynamic constraint limiting phage adaptation at colder temperatures. Two alternative hypotheses for the observed correlation are that: (1) there is selection for thermal generalists at colder temperatures, and (2) the slower growth rates of the phages host (*E.coli C*) at colder temperatures limits the phages growth rates at these temperatures. The first alternative hypothesis makes two assumptions. The first is that phages in colder environments experience more thermal variation than those in warmer environments and the second assumption is that there is a trade-off between maximum growth rate and thermal niche width. The hypothesis would predict phages with lower maximum growth rates will have wider thermal niches and those with higher maximum growth rates would have narrow thermal niches. This trade-off between niche width and maximum growth rate is not evident in the curves describing

the general nature of the reaction norms (Fig. 3.2C). In addition, we tested this hypothesis by calculating the correlation between thermal niche and maximum growth rate as estimated from the continuous reaction norms of the phages. We found that these two parameters are strongly negatively correlated and so are able to reject this hypothesis.

The second alternative hypothesis for the observed correlation between optimal temperature and maximal growth rates is that the phages growth rate constrained by the growth rate of its host, *E.coli C*. Pilot data indicates that the growth rate of *E.coli C* differs 57 % between the lowest (29.4°C) and highest (35.6°C) optimal temperatures of the phages. At the highest optimal temperatures of the phages, *E.coli C* is probably dividing 1 additional and the phage may be getting 1 additional infection, which translates into 2-3 additional rounds of replication. Over this temperature range (29.4°C - 35.6°C), the maximum growth rate of the phages is predicted to differ by ~ 2000 %. To think about whether this difference of 2000% in phage growth rates can be explained by an extra 2-3 rounds of replication, assume that a bacterium infected by a single phage will yield 10 progeny phage. Two additional round of replication would result in a total of 1000 phage and three additional round of replication would result in 10,000 progeny phage, which is 1000 times more than a single round of replication. We can therefore not reject the possibility that differences in the host's growth rate explain most of the observed differences in phage growth rates.

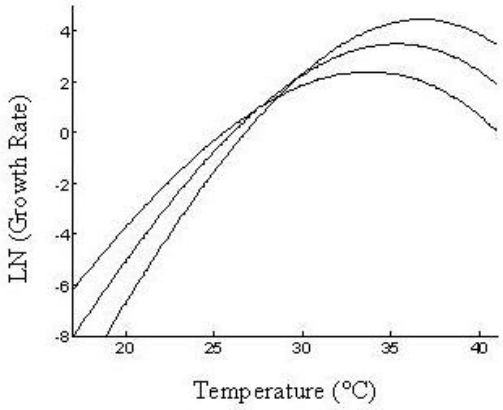


Fig. 3.5. The general nature of the reaction norms from the lab population of phage adapted to a constant temperature. A PCA was performed on the standardized parameters (a,b, and m) for the phage described in chapter 2. The a,b and m values that fall along PC1 were used to plot these curves.

Thermodynamic Constraints

We have the unique ability to compare patterns of thermal adaptation and constraint in natural isolates of G4 (this study) to a lab population adapted to a single temperature (Chapter 2). In Chapter 2, the growth rates of isolates from a lab adapted population were measured and these growth rates were fit with continuous curves using the same statistical method used here – TMV. In this chapter, the general shape of reaction norms of both datasets was characterized with a combination of Principal Component Analysis and TMV (compare Fig. 3.2C to Fig. 3.5). Obvious similarities exist between the reaction norms of these two datasets. The most noticeable pattern is the positive relationship between maximum growth and thermal optima (e.g. 'hotter is better'). In addition, those genotypes with the highest fitness at high temperatures have the lowest fitness at low temperatures.

These results provide strong support for two hypothesized constraints of thermal adaptation: (*i*) thermodynamic constraints at low temperatures prevent adaptation from overcoming the rate-liming effects of temperature on enzymatic reactions, and (*ii*) adaptation to higher (or lower) temperatures results in a loss of fitness at lower (or higher) temperatures.

CHAPTER 4. Exploring the Genetics and Proximate Mechanisms of Thermal Adaptation in the Bacteriophage G4

The work described in this chapter was accomplished in collaboration with Matthew Kasold, Katie Supler, and Dr. Christina Burch. At this point, I anticipate that this project would need more work (replication of the evolved phage lines) to be submitted for publication. This work was improved by conversations with members of the Burch lab.

Introduction

Identifying patterns of thermal adaptation provides useful insights for evaluating and predicting future thermal adaptation events. Observed patterns of thermal adaptation include constraints that limit adaptation at specific temperatures [7] and trade-offs in fitness at different temperatures [12,19]. Fitness trade-offs across temperatures most likely reflect underlying biochemical trade-offs, but the biochemical bases of these trade-offs is largely unknown.

One observed biochemical trade-off is between thermostability and catalytic activity in enzymes from organisms inhabiting different thermal environments. At high temperatures, selection is thought to favor increased thermostability at the cost of the selected enzymes being less flexible and thus less active at lower temperatures [8,61]. In contrast, at low temperatures, selection is thought to favor increased flexibility (for activity) at the cost of protein stability at high temperatures [62]. This trade-off between catalytic activity and stability has been observed in natural populations inhabiting different thermal environments [8]. However, directed evolution experiments have been able to evolve enzymes that have both high thermostability and high catalytic activity in the absence of a biological system (reviewed in [10]). It is not clear whether the observed trade-offs are due to biological constraints or due to natural populations accumulating mutations that are deleterious at temperatures that they do not experience [63].

Laboratory experiments under controlled selective temperatures are a powerful way to address the adaptive processes behind thermal adaptation [29,31,64]. Previous laboratory studies have been sufficiently powerful to demonstrate performance trade-offs across temperatures and have identified the genetic basis of thermal adaptation [12,65,66], but rarely address the biochemical mechanisms of thermal adaptation. A notable exception is the experiment conducted by Counago et al. [65], in which a bacterial population was adapted to an increasingly hot environment and the stability and activity of a single enzyme was monitored during adaptation. This single enzyme had been replaced by a homologue that performed well at low temperatures and did not perform well at high temperatures. As the temperature of the environment increased, this enzyme evolved to have a higher melting temperature and optimal temperature for activity. One difficulty inherent to this single enzyme approach is that it can only detect mutations occurring in the enzyme being examined. By focusing on a single trait, such as the ability of a particular enzyme to perform well at high temperatures, one loses the ability to study the genetics of adaptation for more complex traits underlain by multiple genes (*e.g.* fitness).

In order to investigate the genetics and proximate mechanisms of thermal adaptation of a complex trait - fitness, we adapted the bacteriophage G4 to four novel temperatures, two above and two below the optimal temperature of the ancestral genotype. This experimental design allowed us to compare and contrast adaptation at different temperatures through a combination of whole genome sequencing,

characterization of the nature of the accumulated mutations, and measurement of the thermal stability of the evolved and ancestral phages.

Two results suggest that adaptation to high temperatures occurs in a more consistent manner than adaptation to low temperatures. First, since increased hydrophobicity has been associated with increased thermal stability and adaptation to hot environments in natural populations [67], the change in hydrophobicity was measured at each residue at which a mutation occurred. Consistent with observations in natural populations, we found that the mutations that accumulated in the hot lines consistently increased hydrophobicity at the affected residues. In contrast, the mutations that accumulated in the cold lines had no consistent effect on hydrophobicity. Secondly, the sensitivity of the phages to heat was assessed to determine if the accumulated mutations affected the thermal tolerance of the evolved phage. An analysis of the thermal stability of the evolved and ancestral phages using Gompertz survival curves showed that the hot adapted lines are less sensitive to heat than their ancestor. The lines evolved at low temperatures showed inconsistent changes in their sensitivity to heat – one line was virtually unchanged compared the ancestor and the other was slightly less sensitive to heat.

Materials and Methods

Phage and Culture Conditions

The bacteriophage G4 is a single stranded lytic phage with a 5.5kB DNA genome that infects the bacterium *Escherichia coli.* Culture conditions were as in Knies et al.

[12]. Briefly, phage were grown on *E. coli* C, the standard laboratory host of G4, in LB (5 g Yeast extract, 10 g Bacto tryptone, 5 g salt/ 1 Liter) broth supplemented to 2 mM $CaCl_2$, and on LB plates (15% agar). LB top agar (0.7% agar) was also supplemented to 2 mM $CaCl_2$. Phage were stored for short term (< 1 week) at 4^oC in this growth media, and for long term at -20^oC in LB broth supplemented to 2 mM $CaCl_2$ and 40% glycerol.

Selective Conditions

Four independent populations of the bacteriophage G4 were adapted to the following temperatures: 25°C, 31°C, 39°C, and 41°C. The populations at 31°C and 39°C were evolved for 50 transfers of 45 minutes in length each, while the populations at 25°C and 41°C were evolved for 100 transfers of 60 minutes in length (Fig. 4.1). The longer transfer times were required so that the populations at the extreme temperatures (25°C and 41°C) could successfully be transferred at the beginning of the experiment without going extinct due to slow growth rates at these temperatures. A single transfer consisted of adding approximately $1x10^6$ phages to an exponentially growing culture of *E.coli C*, allowing the phages to grow for 45 or 60 minutes, and then treating the culture with 500 µl chloroform to kill the bacteria. A dilution of the phages was then transferred to another flask containing exponentially growing bacteria. Phage density was determined at the start (N_0) of every transfer by plating. Growth rate per transfer was calculated by estimating the density of phage at the end of every transfer from the density of phage at the beginning of the next transfer and calculating the natural log of the increase in phage number over the transfer time (*e.g.* $(\ln (N_{45}/N_0))/45$.

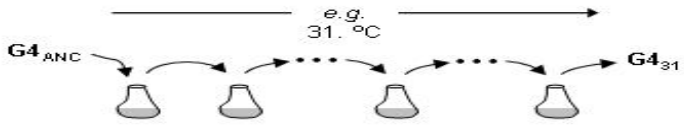


Fig. 4.1. Experimental Design
Each serial transfer had ~ $1x10^6$ phage from the previous transfer placed into 10mL LB with $1x10^9$ rapidly dividing *E. coli* hosts. The 25°C and 41°C populations were evolved for 50 transfers and at the 31°C and 39°C populations were evolved for 100 transfers.

Sequencing

Genome sequences were obtained from PCR products using sequence specific primers. Each genome was amplified in two overlapping segments, and its sequence was determined from 12 overlapping chain-termination sequencing reactions off the same strand. The following lists the G4 primer pairs, with the nucleotide position of each primer based on GenBank accession number J02454 ,followed by its length in parentheses: 492+(18), 1039+(18), 1593+(18), 2094+(18), 2610+(18), 2746+(19), 3295+(18), 3782+(18), 4336+(18), 4831+(18), 5317+(18), 5557+(20). Sequences were aligned and analyzed in Sequencher (version 4.5, Gene Codes Corporation, Ann Arbor, Michigan, United States). Any ambiguous bases were sequenced again. An isolate of each evolved population was selected as the most representative isolate based on the consensus sequence of the population and used for all subsequent analyses.

Thermostability Assays

The phage DNA is encased in a protein coat, called the capsid, and the maintenance of the capsid structure is integral to the phages survival. The capsid structure is maintained by intra and inter protein bonds among the proteins composing the

capsid (the F and G proteins). Mutations accumulating in these proteins may affect these bonds and strengthen or weaken the stability of the capsid at high temperatures.

To measure the sensitivity of the phages capsid to heat, approximately $5x10^8$ phages in ½ ml of media were subjected to a gradual increase in temperature (from 35^oC to 80^oC at the rate of 1^oC/minute). As the temperature increases, the intra and inter molecular bonds stabilizing the phage capsid are expected to become destabilized, leading to a disassociation of the phage components and a loss of viability. The titer of the surviving phages in this first aliquot was measured at 45°C. This entire process was repeated to obtain independent measures of the surviving phage once the temperature reached 50, 55, 60, 65, 70, 72, 74, 76, 78 and 80°C. This experiment was conducted simultaneously for all 5 genotypes (the 4 evolved genotypes and their ancestor) at the same time, referred to as a block. This block experimental design was performed three times.

Statistics

Adaptation of the evolved lines was assessed in two ways. For all the evolved lines, a linear regression was used to find the best fitting line describing growth rate in relation to transfer number. The null hypothesis was that adaptation did not occur in which case the slope of the best fitting line would not be significantly different from zero. This analysis was conducted using the REG procedure in SAS (version 8, SAS Institute, Cary, NC). In all cases, a visual assessment of the model residuals in relation to transfer number confirmed the appropriateness of a linear model. For the 31^oC evolved line,

which did not have a significantly positive slope, a t-test was used to compare the difference between replicate (n = 3) growth rate measures of the ancestral and evolved genotype at the selective temperature. The null expectation was that adaptation did not occur in which case the difference between growth rate of the ancestral and evolved line at the selective temperature would not be significantly different from zero. This analysis was conducted in Microsoft (Redmond, WA) Excel2004.

Chi-squared tests were performed to determine whether the location of mutations accumulated during adaptation differed significantly from a random distribution among genes. Since each evolved line had only accumulated 2-5 mutations, we did not perform chi-squared tests on each line, but rather combined the cold lines into one analysis and the hot lines into another for a total of two chi-squared tests. These analyses were conducted in Microsoft (Redmond, WA) Excel2004.

The change in hydrophobicity at each affected residue was calculated and then paired one-tailed t-tests were performed to determine whether the mean change in hydrophobicity differed significantly from zero. These analyses were conducted in Microsoft (Redmond, WA) Excel2004.

To analyze the results of the thermal stability experiment, the data were fitted to three functions traditionally used [68,69] to analyze survival data: the Weibull, Gompertz, and Logistic functions [70-72]. Each function was fit to the log of the

survival data and the equations shown below are those that were solved for using the NLIN procedure in SAS (version 8, SAS Institute, Cary, NC).

Weibull:

$\text{Log(Survival)} = \log(\exp(-(a/g)((\text{Temperature}-44.99)^{g})));$

Gompertz:

$\text{Log(Survival)} = \log(\exp((A/G)*(1-\exp(G*(\text{Temperature}-45)))));$

Logistic:

$\text{Log(Survival)} = \log(1/(1+(((\text{Temperature}-44.99)/v)^{w}))).$

In these equations, a and g are respectively the shape and scale parameters describing the Weibull curve shape, A and G can be used to calculate the mortality rate from the Gompertz curve, and v and w are respectively the median temperature of inactivation and a measure of the rectangularization of a survival curve. These functions are normally used for survival data that begins at time 0, but for this analysis, it was necessary for the function to be shifted over to start at 45°C (thus, 'Temperature-44.99').

To determine a global best fit function, each of these functions was fit to a combined dataset having survival data for all genotypes at all temperatures and from every experiment. All functions were confirmed to have converged on parameters that minimized the sum of squares. The Gompertz function was found to minimize the residuals and the data from each genotype was then fitted by the Gompertz function.

Results

Phage populations were adapted to 2 temperatures above, 39°C ($G4_{39}$), and 41°C ($G4_{41}$), and below, 25°C ($G4_{25}$) and 31°C ($G4_{31}$), the thermal optima of their ancestor $G4_{ANC}$ (~37°C). To confirm that each population adapted to its selective temperature, the relationship between ln(Growth rate) and transfer number was determined using least squares linear regression. In all of the adapted lines, except $G4_{31}$, the slope of this line is significantly greater than zero (Table 4.1). A separate experiment measured the growth rate of the 31°C population after evolution for 50 transfers and $G4_{ANC}$ at 31°C and found that the growth rate of $G4_{31}$ was significantly greater than that of $G4_{ANC}$ (p-value = 0.02).

Table 4.1. Linear Regression of ln(Growth Rate) on Transfer Number

Temperature	Slope (m) [a]	t-value	Pr > |t| [b]	Intercept
25	0.02	3.63	0.0007	5.17
31	0.00083	0.16	0.88	8.45
39	0.01	2.21	0.03	6.95
41	0.08	7.68	<.0001	3.11

[a] All of the data was used in this regression (*e.g.* growth rate measures from every transfer).
[b] H_o: $m = 0$.

Genetic Basis of Thermal Adaptation

Consensus population sequencing allowed us to identify the mutations accumulated during evolution and to choose a representative isolate to use for all analyses. The representative isolates from the thermal regimes will be referred to as $G4_{25\text{-}I}$, $G4_{31\text{-}I}$, $G4_{39\text{-}I}$, and $G4_{41\text{-}I}$. We identified a total of 13 mutations across all 4 lines (Table 4.2). The low temperature lines had a total of 5 mutations, 2 in $G4_{25\text{-}I}$ and 3 in $G4_{31\text{-}I}$. The high temperature lines had a total of 8 mutations, 3 in $G4_{39\text{-}I}$ and 5 in $G4_{41\text{-}I}$. There was no parallel genetic adaptation between the hot and cold lines and only one parallel mutation between the hot lines (gene F: P355S). This mutation was also seen in

a previous experiment that adapted a single G4 population to hot temperatures ($G4_{44}$) [31]. Two additional mutations (gene B: T103A and gene H: G424D) that accumulated in $G4_{41}$ and one mutation (gene H: A47V) that appeared in $G4_{31\text{-}I}$ also fixed in $G4_{44}$.

Table 4.2 Genetic basis of adaptation to high and low temperatures

Temperature	N_t Site[a]	AA	Gene	Δ HPI[b]
39	3618A>G	D339G	F	3.1
39	**3665C>T**	P355S	F	0.8
39	3845C>T	H415Y	F	1.9
41	**1585A>G**	T103A	B	2.5
41	**3665C>T**	P355S	F	0.8
41	4216A>G	N65S	G	2.7
41	4741A>G	T59A	H	2.5
41	**4988G>A**	G424D	H	-3.1

Temperature	N_t Site[a]	AA	Gene	Δ HPI[b]
31	2149C>T	C57C	D	0
31	3594A>G	K331R	F	-0.6
31	**4703C>T**	A47V	H	2.4
25	2937T>C	I112T	F	-5.2
25	5062A>G	T498A	H	2.5

[a] Bolded sites show mutations in $G4_{44}$.
[b] Hydrophobicity indexes (HPI) are from Kyte and Doolittle (1982) and positive

Neither the hot nor the cold lines show evidence of clustering of mutations among non-overlapping genes (hot lines: $?^2 = 5.18$; df = 7; $p > 0.50$; cold lines: $?^2 = 5.32$; df = 7; $p > 0.50$). Since it is possible that there were too few mutations to detect nonrandom clustering of the mutations, we also analyzed an expanded dataset of high temperature adaptive mutations. This expanded dataset included 19 mutations and combined all the mutations from $G4_{39}$, $G4_{41}$, and $G4_{44}$. However, an analysis of this expanded dataset also found no evidence of clustering of mutations among non-overlapping genes ($?^2 = 2.41$; df = 7; $p < 0.95$).

Because increases in amino acid hydrophobicity have been associated with adaptation to thermophilic environments in nature [67], we examined the change in

hydrophobicity at every amino acid that accumulated a change during adaptation. In the high temperature treatment, 7 out of 8 mutations increased the hydrophobicity at the amino acid at that position according to the Kyte and Doolittle hydropathy index [73]. In contrast, in the low temperature treatment, there is no apparent pattern in hydrophobicity change. One-tailed paired t-tests confirmed that the mean change in hydrophobicity is significantly positive ($p = 0.045$) and not significantly different from zero ($p = 0.45$) in the high and low temperature lines respectively.

Table 4.3. Survival functions fit to thermo tolerance data

Model	Parameters		**Residuals**
Weibull	a	0.000194	317.8
	g	3.2634	
Gompertz	A	0.0160	298.0
	G	0.1110	
Logistic	v	22.3849	371.7
	w	13.6295	

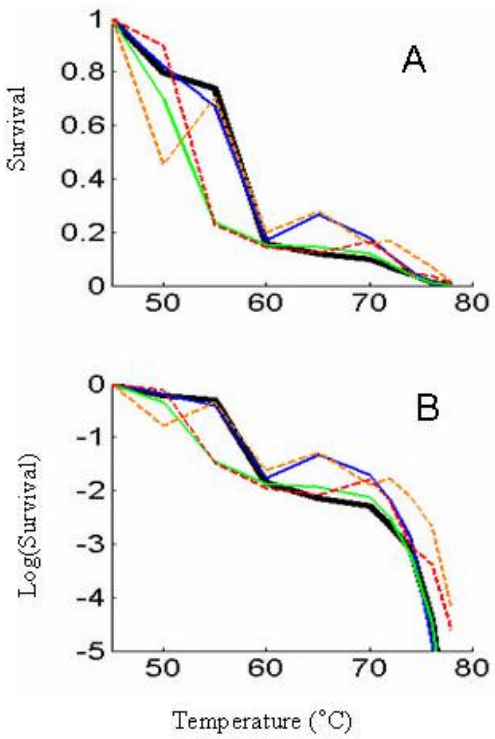


Fig. 4.2. Thermostability of Evolved and Ancestral Phage (A) Shown is the survival of the phages as a function of increasing temperature. (B) Shown is the log of the survival as a function of increasing temperature. Both linear and log survival axes are shown, to allow a full view of variation between genotypes at high and low survival levels. The black (thickest) solid line is the data for $G4_{ANC}$, the green and blue lines are the data for $G4_{25\text{-}I}$ and $G4_{31\text{-}I}$ respectively, and the orange and red dashed lines are the data for $G4_{39\text{-}I}$ and $G4_{41\text{-}I}$ respectively.

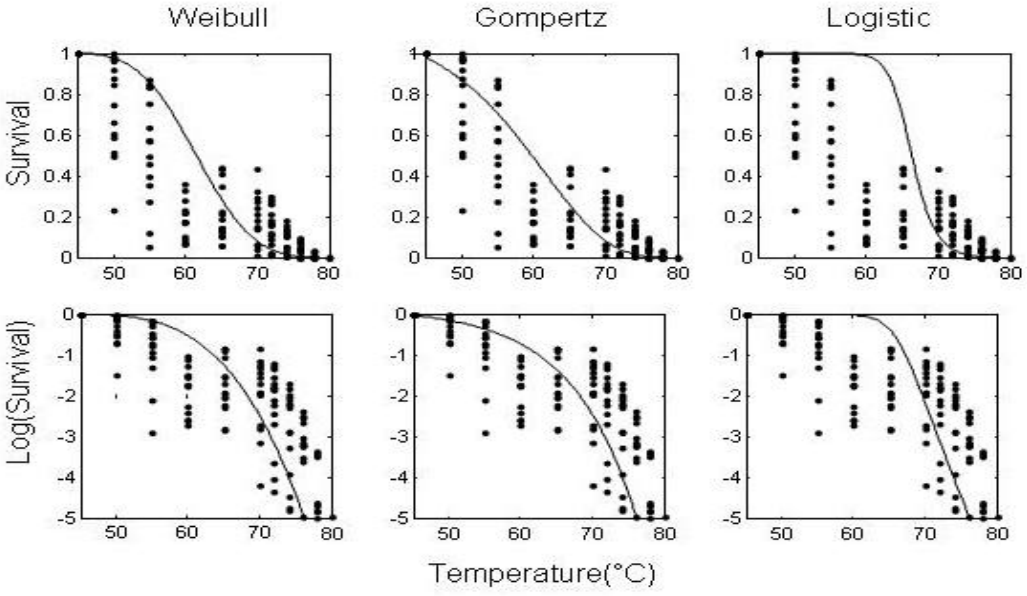


Fig. 4.3. Survival and mortality curves for global thermal stability dataset. Data is shown as closed circles. Both linear and log survival axes are shown, to allow a full view of the fit at high and low survival levels. From left to right, the panels show the fit of the Weibull, Gompertz, and Logistic curves to the global dataset.

Thermo tolerance/ stability

To determine the effect of accumulated mutations on the sensitivity of the phages capsid to heat, phages survival was measured as the temperature increased from 35°C to 80°C. It was expected that with increasing temperature, the intra and inter molecular bonds maintaining the capsid structure would become destabilized and eventually result in phage inactivation. As expected, as temperature increase, the fraction of the surviving phage decreases (Fig. 4.2). We explored methods for analyzing the survival data by fitting the Weibull, Gompertz, and Logistic survival functions to the complete survival dataset (5 genotypes x 10 temperatures x 3 replicate measures/temperature). All three of the best fit curves fit the survival data well at high temperatures (70-80°C), but none of these curves were able to satisfactorily fit the survival data between 55 and 60°C. The Gompertz function was determined to be the best fit of these 3 functions to the phages survival data because it explained the greatest amount of variation in the global survival dataset (Fig. 4.3 and Table 4.3).

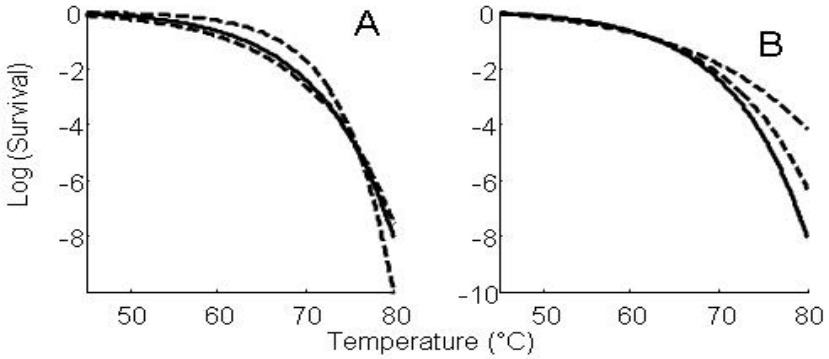


Fig. 4.4. Gompertz mortality curves for the evolved and ancestral phage. In both graphs, the solid line is Gompertz fit to the ancestor and the dashed lines are the Gompertz fits to the cold (A) and hot (B) adapted lines.

Table 4.4. Gompertz Parameter Estimates

Genotype	A	G	Residuals
$G4_{ANC}$	0.0138	0.1179	64.61
$G4_{25\text{-}I}$	0.00304	0.17772	70.49
$G4_{31\text{-}I}$	0.0173	0.1075	55.84
$G4_{39\text{-}I}$	0.0234	0.0735	12.73
$G4_{41\text{-}I}$	0.0167	0.1021	56.17

To analyze evolved differences in capsid sensitivity, we fit the Gompertz function to the survival data of each genotype. If the accumulated mutations affect the sensitivity of the phage capsids to heat, then the evolved phages survival would differ as compared to their ancestor and that survival differences would be magnified with increasing temperature. The best-fit curves as determined by the Gompertz function are shown in Figure 4.4 (parameter estimates are in Table 4.4). There are two ways that the best fit curves of the evolved phages differ from their ancestor. First, the capsid of $G4_{25\text{-}I}$ appears to have a reduced sensitivity to heat from 45-70°C, but this difference is not magnified with increasing temperature. Second, the capsids of the high temperature lines, $G4_{39\text{-}I}$, and $G4_{41\text{-}I}$, are less sensitive to heat above 65C° and this difference in survival is magnified with increasing temperature.

Discussion

In this study we used the bacteriophage G4 to investigate the genetics and proximate mechanisms of thermal adaptation. In contrast to the study by Counago [65] in which a single-enzyme approach was taken to investigate adaptation to high

temperatures, we chose to look at thermal adaptation of a more complex trait, fitness, at high and low temperatures. The results of our study indicate that adaptation to high temperatures is more predictable than adaptation to low temperatures and is associated with increased amino acid hydrophobicity and thermo tolerance. It may be that we have not found the traits that characterize adaptation at low temperatures.

Increased hydrophobicity is positively correlated with increased thermal stability in enzymes from natural populations inhabiting thermophilic environments and in enzymes evolved at extreme temperatures in directed evolution experiments [67,74,75]. Consistent with this data, in our high temperature lines, the accumulated mutations increased hydrophobicity at their amino acid residues, but in the low temperatures lines, the accumulated mutations had no significant effect on hydrophobicity.

Increased thermo tolerance/stability is also associated with adaptation to thermophilic environments [8]. At high temperatures, stability of proteins is expected to be under strong selection to prevent denaturation. Consistent with this hypothesis, mutations in the high temperature lines confer increased thermo tolerance at high temperatures as compared to the ancestral phage. In contrast, the genetic changes in the cold temperature lines had more complex effects on thermo tolerance. In one line, the thermo tolerance of the evolved phage was actually greater than that of the ancestor at low-intermediate temperatures. In the other line, the thermo tolerance of the evolved phage was consistently a little bit lower than that of the ancestor.

While few parallel genetic changes were observed between the evolved lines, one notable mutation (gene F: P355S) was shared between our two high temperature lines ($G4_{39,}$ $G4_{41}$) and the line ($G4_{44}$) from [31]. We can hypothesize the effect of this mutation at high temperatures given that is effect on fitness at high temperature has previously been shown to be beneficial [31] and from knowledge of the function of gene F. The advantage conferred by this mutation at high temperature may be that it increases the stability of G4 during phage morphogenesis. During phage morphogenesis, the phage capsid assembles via aggregation of its component proteins – including the F protein – and binding between F proteins is thought to make these aggregations energetically favorable. The site that was mutated in our lines, gene F: 355, is adjacent to two site (354 and 356) known to respectively be involved in binding to other F proteins and be important for protein- water interactions [76]. We hypothesize that the amino acid change at this site stabilizes the binding between the adjacent amino acids and their binding partners. This hypothesis could be tested by creating phage identical except for this single mutation and then measuring the rate at which the capsid intermediates assemble for phage with and without this mutation.

While we did not anticipate that the populations adapted to low temperatures would show no change or increases in thermo tolerance, this result is consistent with directed evolution experiments on enzymes (reviewed in [10]). In these experiments, enzymes are selected based on their performance at a selective temperature and their evolution is independent of a biological system. In the absence of biological constraints, it has been possible to evolve enzymes at low temperatures that maintain a high stability and a high catalytic activity. Similar results have been achieved for the evolution of

enzymes at high temperatures. Our results, combined with those of Counago et al. [65], and those from the directed evolution experiments on enzymes [10] indicate that the trade-offs observed between stability and catalytic activity may not due to biochemical constraints and might simply be a consequence of natural populations living in their thermal environment for many generations.

Future Directions

While our experiment is an improvement over previous studies of thermal adaptation, it has a few limitations that have impacted our findings. First, if more populations had been adapted to a particular thermal regime, a greater number of mutations would have accumulated, and we would have more power to detect nonrandom clustering of the mutations accumulated during evolution. This problem is mostly pertinent to the cold adapted lines: these two lines only accumulated 5 mutations. Another reason for evolving additional lines at cold temperatures is that the temperatures at which the cold populations evolved at may not have been extreme enough to result in the predicted decrease in thermostability at high temperatures. The second and greater limitation of our study is in our analysis of the thermo tolerance experiment. We used Gompertz curves to fit curves to the survival data, however, these curves can not capture the shape of the raw data between 50 and 60°C. One solution to this problem would be to fit the survival data with splines, which are more flexible, yet have less interpretable parameters.

In sum, our data suggests that there are predictable mechanisms of adaptation to high temperatures- namely increased thermal stability. We demonstrated this both in our genetic data and phenotypic characterization of the phages thermal stability. Increased replication of the number of populations and more thermal stability experiments would strengthen this conclusion.

CHAPTER 5. Compensatory evolution in RNA secondary structures increases substitution rate variation among sites.

The work described in this chapter was accomplished in collaboration with Kristen Dang and Drs. Todd J. Vision, Noah G. Hoffman, Ronald Swanstrom, and Christina L. Burch. I anticipate that this chapter will be submitted to Molecular Biology and Evolution before my defense date. Technical assistance and advice was graciously provided by Drs. Stefanie Hartmann, Sergei Kosakovsky Pond, and Derrick Zwickl.

Introduction

Compensatory mutations, or mutations that are individually deleterious but neutral or beneficial in combination, permit deleterious mutation to be fixed without causing a net fitness loss (Poon and Otto 2000). Experimental evidence from laboratory populations shows that most deleterious mutations can be compensated by numerous mutations at alternative sites (Burch and Chao 1999; Poon and Chao 2005) and that fitness recovery following the fixation of a deleterious mutation most often occurs via compensatory rather than back mutations (Schrag, Perrot, and Levin 1997; Maisnier-Patin et al. 2002; Hoffman, Schiffer, and Swanstrom 2005).

Kimura developed a population genetics model in which deleterious and compensatory mutations arise within a single genome during the time when both mutations are rare in the population and they occasionally drift to fixation as a pair (Kimura 1985). Because genomes with low fitness are not required to become fixed in this process, compensatory mutations are predicted to contribute to divergence even if their individual deleterious effects are large, and even in large populations. However, because the waiting time until both compensatory mutations arise in the same genome is longer than the waiting time for an independent neutral mutation to arise, the rate of molecular evolution is predicted to be lower at compensatory sites than at independently evolving neutral sites.

The contribution of compensatory mutations to molecular evolution in natural

populations has been most thoroughly investigated in regions of RNA secondary structure. RNA secondary structure offers a convenient model for investigating compensatory evolution because individual sites are readily identified as independently evolving (unpaired sites) or involved in a compensatory interaction (stems or paired sites). Compensatory evolution clearly plays a role in the evolution of these regions because mutations in stems are generally accompanied by compensating mutations that maintain base pairing in the stem (Kirby, Muse, and Stephan 1995; Wilke, Lenski, and Adami 2003). In addition, the rates of substitution at compensatory and independent sites in RNA secondary structures show the pattern predicted by Kimura's compensatory neutral model (Stephan 1996; Innan and Stephan 2001). This difference in substitution rates at compensatory and independent sites is used to predict secondary structure {Pedersen et al 2004; Muse 1995}.

In addition to the slowdown in molecular evolution predicted at compensatory sites, inherent differences in rates are also expected to be exaggerated at paired sites in RNA. Specifically, we expect the ratio of transitions to transversions to be exaggerated in the pairing regions. This expectation is derived from the biochemical constraints of base pairing, where a transition mutation can only be compensated for by other transition and likewise for transversions. Since transitions occur more frequently than transversions, transition will be compensated for more quickly than transversions, resulting in an elevated transition: transversion rate ratio (k) at compensatory sites.

Here, by extending Kimura's model specifically to molecular evolution in RNA secondary structures, we make the prediction that the transition to transversion substitution rate ratio (k at paired sites should be the square of that for unpaired sites, all

other factors being equal. We tested this prediction in eight functionally and taxonomically diverse RNA molecules and found common, but not universal, quantitative agreement with the model. The prediction that we test may be useful in increasing the accuracy of methods of RNA secondary structure prediction.

Kimura's model of compensatory neutral evolution

Following Kimura (1985), we consider the substitution process at a pair of loci involved in a compensatory interaction in a diploid population of size N (equivalent results are obtained for a haploid population of size $2N$). Let $??$ represent the rate of mutation from the wild type to the mutated allele at both loci, and ignore back mutation by assuming that selection is sufficiently strong to keep both mutations at a low enough frequency that back mutations are improbable. Selection is assumed to act equivalently on mutations at both sites, so that the fitness of genomes containing either one of the two mutations is $1 - s$. Because the two mutations are involved in a compensatory interaction, however, the fitness of genomes that contain both mutations is equivalent to the wildtype (i.e. fitness = 1). Finally, we assume that the loci are sufficiently close together that recombination between them can be ignored. This last assumption is reasonable for pairing sites within RNA secondary structures.

In Kimura's model, the deleterious mutations are assumed to be present initially at an equilibrium frequency determined by the balance between mutation and selection. At this equilibrium the expected number of alleles carrying either one of the two deleterious mutations is $4Nm/s$. Compensating mutations are assumed to arise in

genomes that already carry an initial deleterious mutation at rate m, and the probability that the newly arisen deleterious – compensatory pair of mutations drifts to fixation in the population is $1/2N$. Combining these effects, the substitution rate at sites involved in compensatory interactions is predicted to be:

$$d_C = \left(\frac{4Nm}{s}\right) \cdot m \cdot \left(\frac{1}{2N}\right) = \frac{2m^2}{s} \quad (1).$$

Equation (1) is equivalent to equation (8b) of the bidirectional, symmetric model of Stephan (Stephan 1996). We can compare this compensatory substitution rate to the expectation at independently evolving neutral sites (Kimura 1985):

$$d_I = 2Nm \cdot \left(\frac{1}{2N}\right) = m \quad (2)$$

Now we consider evolution in two classes of sites – one class in which both sites participating in the compensatory interaction mutate at a faster rate m_1, and one class in which both sites mutate at a slower rate m_2. We find that the ratio of the substitution rate between the two classes of sites:

$$\frac{d_{C,1}}{d_{C,2}} = \frac{2m_1^2/s}{2m_2^2/s} = \left(\frac{m_1}{m_2}\right)^2 \quad (3)$$

is the square of the ratio at independently evolving neutral sites:

$$\frac{d_{I,1}}{d_{I,2}} = \frac{m_1}{m_2} \quad (4)$$

To adapt this scenario to the evolution of RNA secondary structures, we make use

of the widespread observation that transition mutations (purine-to-purine or pyrimidine-to-pyrimidine) are generally more common than transversion mutations. Furthermore, we note that a transition mutation in one side of an RNA stem structure can only be compensated by another transition mutation, and likewise for transversions. Thus, we expect two rates of compensatory evolution, one for transitions ($d_{C,Ti} = 2m_{Ti}^2 / s$) and another for transversions ($d_{C,Tv} = 2m_{Tv}^2 / s$). By assuming that selection against both types of deleterious intermediates acts with the same strength, we predict that the rate ratio of transition to transversion substitutions (k) in paired regions of RNA secondary structure (stems) should be:

$$k_p = \frac{d_{C,Ti}}{d_{C,Tv}} = \left(\frac{m_{Ti}}{m_{Tv}} \right)^2 \qquad (5)$$

which is again the square of the rate ratio in unpaired regions (loops):

$$k_u = \frac{d_{I,Ti}}{d_{I,Tv}} = \frac{m_{Ti}}{m_{Tv}} . \qquad (6)$$

Methods

To test the predictions of the model, we selected eight RNA molecules with well-documented secondary structures for which a large number of diverse sequences are available. For each set of sequences, a multiple alignment and phylogeny was inferred. The value of k was then estimated for paired vs. unpaired sites and a test was performed to determine if these two values differed significantly.

Sources of sequence data and secondary structures

Sequences were obtained from a variety of sources, as listed in Table 5.1. For each molecule, sequence positions were classified as either paired or unpaired. In most cases we used structures reported in the literature: RRE (Phuphuakrat and Auewarakul 2003), CRE (Tuplin et al. 2002; Tuplin, Evans, and Simmonds 2004), 12SrRNA (Springer, Hollar, and Burk 1995). Structures for IRES, 5S rRNA (Fox and Woese 1975), (Specht, Wolters, and Erdmann 1991), and RNase P (Haas et al. 1991) were obtained, respectively, from the Viral RNA Structure Database (Thurner et al. 2004), the 5S Ribosomal RNA Database (Szymanski et al. 2002), and the RNase P database (http://www.mbio.ncsu.edu/RNaseP/home.html) (Brown 1999; Harris et al. 2001). Finally, tRNA structures were obtained using Mfold v. 3.0 (http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html) (Zuker 2003) with manual adjustments to fit the canonical model described in Sprinzl (Sprinzl et al. 1998).

Alignment and phylogenetic Inference

The sequences were either obtained having already been aligned or aligned *de novo* in ClustalW (Chenna et al. 2003) or MAFFT (Katoh et al. 2005). The alignments used for all analyses are available on request. We constructed phylogenies in MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) using the GTR + gamma + invariant model of nucleotide substitution and otherwise using default parameters. For the pestivirus, IRES, 5SrRNA, and RNaseP alignments 100,000 generations with a 10% (10,000 generations) burn-in was sufficient for convergence. For the 12SrRNA alignment and the mitochondrial alignments, convergence was achieved in 500,000 generations with a 10%

and 20% burn-in respectively. For the CRE and RRE alignments, convergence was achieved in 1 million generations with a 20% and 50% burn-in respectively., Three independent runs were conducted for each alignment and the log likelihood values of these runs were compared to confirm that the chains converged on the same posterior distribution. The consensus tree was obtained by majority rule. There was no evidence of saturation as the distance from leaf to tip was <1 in all phylogenies.

In two cases, we were unable to use of all of the available sequences because the phylogenetic method did not converge (RRE) or the resulting phylogenetic tree was poorly supported (5S rRNA). In these cases, we used a subset of the downloaded sequences, choosing the first 199 sequencing (RRE) or based on the number of sequences available from particular genera (Bacillus and Clostridium) that resulted in a well supported phylogenetic tree (5S rRNA). The tree constructed for RRE had bw interior branch support (1-2%), which indicates substantial convergent evolution or recombination between taxa. For HIV, recombination commonly occurs within patients and between closely related genotypes, which will not affect the phylogeny. The low branch support for the HIV phylogeny is most likely due to convergent evolution. There is no reason to expect convergent evolution to affect the underlying k parameters.

Estimation of substitution rate parameters

We used the program HyPhy (Kosakovsky Pond, Frost, and Muse 2005) to estimate k separately for paired and unpaired regions of each molecule. We

incorporated rate heterogeneity with a discretized gamma distribution of mutation rates (four rate classes), and used the HKY85 model of nucleotide substitution (Hasegawa, Kishino, and Yano 1985), which allows for unequal base frequencies, one substitution rate for all transitions and one for all transversions. We chose this model in order to obtain a single estimate for the transition-transversion rate ratio (k) with reasonably small confidence intervals, even though it is not necessarily the best fit for each alignment (see below). For each alignment, we report k and the approximate 95% confidence intervals derived from the Fisher information matrix. Parameter estimates of k are expected to be robust to minor errors in tree topology {Hillis 1999}.

We also investigated whether the HKY85 model, which allows one rate for all transitions and one for all transversions, gives a reasonable approximation of the observed substitution patterns. For each alignment, we found the best nucleotide substitution model using the model testing procedure implemented in HyPhy (Kosakovsky Pond, Frost, and Muse 2005), which is based on that of ModelTest (Posada and Crandall 2001). We fit models with discrete gamma distributed rate variation and a fraction of invariant sites.

Tests of predictions

Likelihood ratio tests were used to decide whether the data supports estimation of a separate k for paired and unpaired regions. Formally,

H_O: $k = k_p = k_u$

H_A: $k_p \neq k_u$

where κ is the transition-transversion ratio estimated from the entire molecule, k_p is the transition-transversion ratio estimated from an analysis of paired positions only, and k_u is the transition-transversion ratio estimated from an analysis of unpaired positions only. We calculated a test statistic, ?, for each alignment in the following way:

$$l = -2(\ln Lk - (\ln Lk_p + \ln Lk_u)) \quad (5)$$

The significance of ? ?was evaluated assuming a χ^2 distribution with one degree of freedom. For each molecule (or alignment), the likelihood values were calculated by holding constant the phylogenetic tree, the rate variation parameter a (estimated from the complete molecule), and the nucleotide substitution model (HKY85).

.

Table 5.1. Description of Structures

Structure [a]	C/N[b]	Organism	Source of Sequences
RRE	C	HIV	Los Alamos HIV Database (http://hiv-web.lanl.gov/content/index)[e]
IRES	N	Pestivirus	Viral RNA Structure Database (http://rna.tbi.univie.ac.at/cgi-bin/virusdb.cgi)
CRE [c]	C	Hepatitis C	Los Alamos HCV database (http://hcv.lanl.gov/content/hcv-db/index)
5srRNA	N	Firmicutes Bacteria	5S ribosomal RNA database (http://www.man.poznan.pl/5SData/)
tRNA [d]	N	Amphibian mitochondria	Organellar Genome Retrieval system (http://ogre.mcmaster.ca)
tRNA [d]	N	Mammalian mitochondria	Organellar Genome Retrieval system (http://ogre.mcmaster.ca)
12SrRNA	N	Mammalian mitochondria	GenBank (AB074968, AY172335, U33494 – UU3948)
RNaseP	N	Mammal	RNaseP database (http://www.mbio.ncsu.edu/RNaseP/home.html)

[a] Abbreviations: RRE: Rev response element; IRES: Internal ribosome entry site; CRE: cis-acting replication element.
[b] C and N are abbreviations for coding and Noncoding respectively.
[c] The sequences used were non-recombinant, non-related type 1 HCV NS5B sequences (positions 7394-9170 of reference sequence M62321).
[d] The tRNAs analyzed were a concatenated alignment of Alanine, Cysteine, Glutamic acid, Asparagine, Glutamine, and Tyrosine.
[e] Only one sequence per patient was downloaded.

Results

We tested the prediction that compensatory evolution in regions of RNA secondary structure should result in the transition to transversion rate ratio (k) being approximately squared at paired relative to unpaired sites by examining eight different RNA secondary structures. These molecules were selected from a diverse group of organisms: viruses, bacteria, amphibians, and mammals (Table 5.1). In order to accurately estimate ?, we used alignments that had at least 20 variable paired and 20 variable unpaired sites and for which the secondary structure was known to be conserved and functionally important. The structures used here were predicted from either comparative sequence analyses (12S rRNA, 5S rRNA, and tRNA's), experimental evidence (RRE), or both (RNaseP P, IRES, CRE). The alignments had varying amounts of sequence diversity (Table 5.2) and complexity of secondary structures ranging from relatively simple tRNAs with 3 stem-loops to the 12S rRNA with approximately 20 stem-loops.

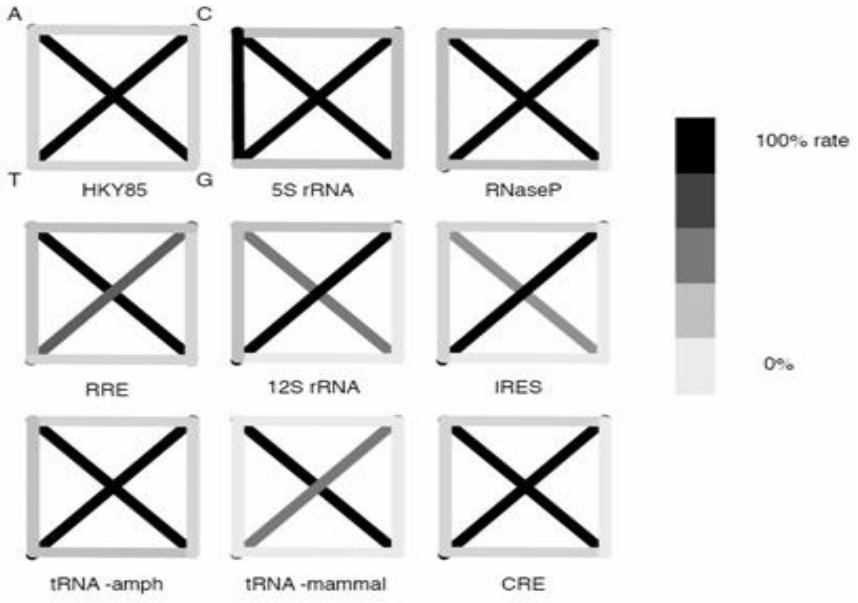


FIG. 5.1. Best-fit nucleotide substitution models for each alignment. Shown is a cartoon illustration of the rate categories of the best-fit nucleotide substitution models for each molecule. Within a molecule, rates were scaled to the maximum rate (black). Diagonal lines depict transitions; the edges of the square depict transversions. The HKY85 model, which was used for the rate ratios reported throughout this paper, is shown for comparison in the upper left. Models were chosen using the nucleotide model testing procedure implemented in HyPhy.

Testing Kimura's model of compensatory evolution.

Our approach assumes that transition substitution rates (G? A, C? T) are more similar to each other than they are to transversion substitution rates, and vice versa, and that the two rates differ substantially. To examine these assumptions, we used a model-testing procedure that allowed any combination of the six time-reversible substitution rates (see Methods) to find the best-fitting nucleotide substitution model for each alignment. The best-fit models are shown in Figure 5.1. Although the HKY85 model, which specifies one rate for transitions and one rate for transversions, was not statistically the best-fit model for any alignment (as determined by AIC scores), the best-fit models did show differences between transition and transversion rates, and that transition rates

were almost universally more similar to each other than to transversion rates. Only in the case of the 5S rRNA alignment was one of the transversion rates (A? T) similar to the transition rates. Thus, it is justifiable to use the HKY85 model as a reasonable approximation for the purposes of comparing estimates of k.

We estimated k in two ways. First, we considered the molecule as a whole, and estimated a single k to describe all sites (paired and unpaired). Second, we divided the molecule into paired and unpaired sites, and estimated k_p from all paired sites and k_u from all unpaired sites. In seven of the eight structures, k_p was higher at paired than unpaired sites (Fig. 5.3), qualitatively supporting the prediction of Kimura's model. Only in the RRE structure was $k_p << k_u$. Likelihood ratio tests (see Methods for details) led to rejection of H_O: $k = k_p = k_u$ in favor of H_A: k_p ? k_u for all eight molecules at $p<0.0001$ (Table 5.2).

We also examined the quantitative fit of the estimates to the model predictions. A visual inspection of Figure 5.3 confirms the close match of most structures to the expectation that $k_p = k_u^2$. For six structures (12S rRNA, 5S rRNA, mammalian tRNAs, RNase P, IRES, and CRE) estimates of k_p can not be statistically distinguished from k_u^2, *i.e.* 95% confidence intervals estimated from likelihoods overlap the $k_p = k_u^2$ line. Only the HIV RRE structure, and to a lesser extent the mammalian tRNAs, deviate significantly from the model prediction.

Table 5.2 Description of Alignments and Likelihood Ratio tests for k estimation

Structure	# Taxa	# Sites[a]	Stems[a]	Loops[a]	k	k_p	k_u	λ
RRE	199	231 (140)	161 (91)	70 (49)	5.19	4.21	9.01	546.05*
IRES	11	347 (162)	198 (92)	149 (70)	6.50	15.34	3.60	73.46*
CRE	98	255 (119)	189 (72)	66 (45)	12.52	22.36	2.93	177.32*
5srRNA	15	117 (86)	73 (61)	44 (25)	3.70	4.44	2.82	35.05*
tRNA	40	414 (344)	266 (224)	148 (120)	6.04	9.48	3.30	204.73*
tRNA	40	419 (187)	243 (98)	176 (89)	3.90	6.69	2.83	131.93*
12SrRNA[b]	7	930 (406)	457 (173)	473 (233)	11.98	18.78	9.65	122.24*
RNaseP	10	493 (352)	291 (218)	202 (134)	3.38	5.07	1.75	80.77*

[a] Numbers in parentheses indicate the number of variable positions.
[b] Excluding the region between stem 38 and its complement 38' , which is highly variable and difficult to align.
* LRT value significant at $p < .0001$

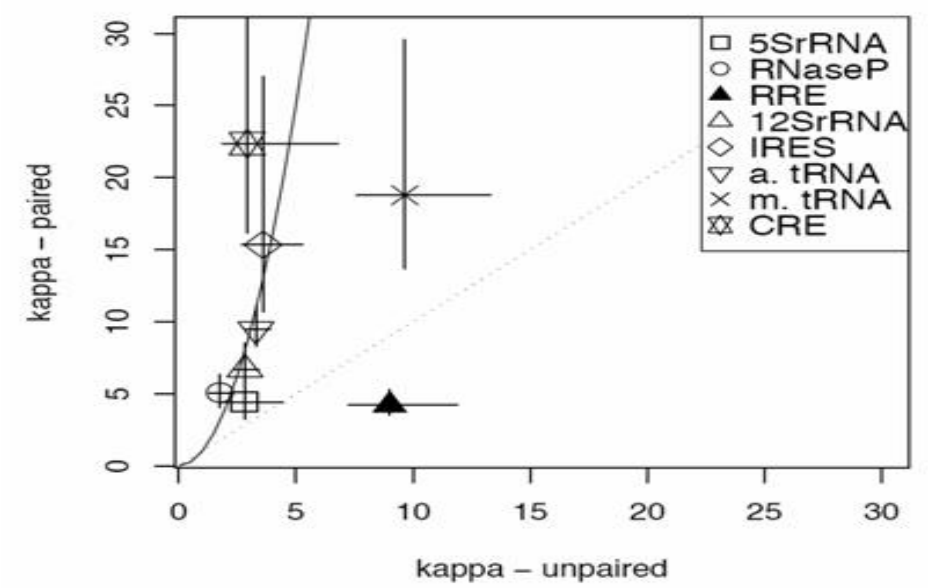


FIG. 5.2. Transition-transversion rate ratios (κ) for each alignment. The dotted line represents a 1:1 relationship between kappa – paired and kappa-unpaired. The solid line represents the predicted squared relationship between these two parameters. Note that the CRE data-point is from the analysis of four-fold degenerate sites in paired and unpaired regions.

Substitution patterns in RRE

We examined three possible explanations for the surprising result that $k_u > k_p$ in RRE. First, since both the RRE and CRE secondary structures occur within coding regions, we examined the possibility that the difference between k_u and k_p is diminished by selection on the protein sequence. We recalculated k_u and k_p for both molecules using only data from four-fold degenerate sites in paired and unpaired regions. In CRE, coding affects the estimates in the predicted direction (four-fold degenerate sites: $k_p = k_u^{2.89}$; all sites: $k_p = k_u^{1.45}$), though the four-fold sites overshoot the predicted pattern. We had less power to compare four-fold degenerate sites at the paired and unpaired sites of RRE because there were too few four-fold degenerate unpaired sites and there was insufficient sequence variability at these sites. However, the four-fold degenerate paired sites did show a higher k_p (k_p = 7.61 with 95% confidence interval [4.79, 18.48]) than the paired sites as a whole (k_p = 4.21 with 95% confidence interval [3.51 – 5.28]). This suggests that the presence of protein coding constraints does impede compensatory evolution at paired sites in RNA secondary structures, although it does not explain why k_u would be *greater* than k_p in RRE.

Second, we examined the possibility that we had used a non-representative sample of RRE sequences. To confirm that the observed substitution patterns in RRE were not specific to the particular set of HIV sequences we examined (which were all derived from subtype B), we estimated k_p and k_u from two additional RRE alignments of sequences drawn from higher taxonomic levels: sequences from different subtypes (A, B, C, F, G, H, J, and K) and sequences from different groups (M, N, and O). In both of

these alignments, the results were qualitatively similar to those for subtype B: k_u was significantly higher than k_p (Table 5.3).

Third, we considered whether the RRE estimates were disproportionately influenced by a portion of the molecule that experiences a type of selection that differs from the as a whole. We systematically removed each stem-loop of RRE and re-estimated k_p and k_u for the resulting partial structures. The k_p and k_u estimates were qualitatively similar for all of these partial structures (Table 5.4).

Table 5.3. Transition–transversion rate ratios (k) for RRE alignments from different taxonomic levels

RRE alignment	k_p [a]	k_u [a]
Subtype B	4.21 (3.51 – 5.28)	9.01 (7.25 – 11.89)
Subtypes	2.10 (1.45-3.09)	7.26 (3.88-15.09)
Groups	5.56 (3.21-10.46)	15.96 (7.00-46.29)

[a] Numbers in parentheses are the 95% confidence intervals for the k estimates.

Table 5.4 Transition-transversion rate ratios (k) for partial RRE structures

RRE structure [a]	k_p [b]	k_u [b]
Complete	4.21 (3.51 – 5.28)	9.01 (7.25 – 11.89)
I	4.70 (3.67 – 6.54)	8.03 (6.38 – 10.77)
II	5.43 (4.37 -7.19)	9.73 (7.40 – 14.17)
III	3.36 (2.78 – 4.24)	8.18 (6.55 – 10.92)
IV	4.47 (3.70 – 5.66)	11.21 (8.77 – 15.51)
V	4.02 (3.34 -5.05)	8.35 (6.70 – 11.07)

[a] This analyzed RRE structures had the listed (*e.g.* I,II, see Fig. 5.1) stem-loop removed. The complete structure is shown for comparison.
[b] Numbers in parentheses are the 95% confidence intervals for the k estimates.

Discussion

We extended Kimura's population genetic model of compensatory evolution to

make the prediction that substitution rate variation due to underlying mutation rate differences is exaggerated by compensatory interactions. We made use of the fact that transition rates generally exceed transversion rates to test this prediction in regions of RNA secondary structure. Specifically, we predict that k , the ratio of transition to transversion rates, should be higher in paired than in unpaired regions. Seven of the eight RNA secondary structures we examined confirmed this qualitative prediction. Moreover, six of the eight structures (RNase P, 5S rRNA, IRES, tRNA A, 12S rRNA, and CRE) closely matched the quantitative prediction that the ratio in paired regions should be the square of the ratio in unpaired regions (i.e. $k_p = k_u^2$).

We consider the implications of the close match between the model and the prediction by focusing on the structural and molecular evolution of these secondary structures. These results tell us that over long time periods, there has been minimal structural evolution of these molecules. Had structural evolution occurred, then some sites identified as paired would actually be unpaired in other sequences and vice versa, which would decrease the difference between k_u and k_p. The conservation of these structures allows them to be useful for establishing relatedness between different species. These results also effectively illustrate that the measured transition and transversion rates in secondary structures are governed by a history of compensatory evolution. The observed rates in secondary structures do not represent the actual mutation rates- but may sometimes be treated as such.

Structures not explained by the model

Why might the data for RRE and mammalian tRNAs not fit this model? The only feature that these two molecules share is a high k_u value compared to the other molecules we investigated, but a high ratio of m_{Ti} to m_{TTv} does not, in itself, violate the model assumptions. We focus on how the assumptions of the model could be violated in a way that preferentially elevates k_u with a particular focus on mechanisms that could explain the observation in the most deviant structure (RRE) that $k_p << k_u^2$. .In addition, we recognize that our estimates are sensitive to misidentification of paired and unpaired sites, as well as inaccuracies in the alignment and/or phylogeny

The model assumptions are (*i*) the only determinant of molecular evolution in these molecules is the functional constraints on the RNA secondary structures, (*ii*) recombination is infrequent, and (*iii*) the underlying m_{Ti} is the same in paired and unpaired regions and likewise for m_{TTv}. As discussed in the Results, coding may affect the estimates of k for CRE and RRE, but the size of this effect is not sufficient to explain a reversal of the relative magnitudes of k_u and k_p in RRE. Recombination is also unlikely to explain the results for the mammalian tRNA and RRE due to recombination rarely occurring between mitochondria and recombination between HIV particles primarily occurring between closely related genotypes, which will have a minimal effect on k_u (or k_p). In RRE, the reversal of the relative magnitudes of k_u and k_p may be due to the underlying m_{Ti} being unequal in paired and unpaired regions due to the actions of a cytodine deaminating enzyme – APOBEC3. APOBEC3 recognizes specific motifs

and preferentially deaminates C to T (a transition) in unpaired regions of DNA, which is then visible as a G to A transition on the + strand of HIV. HIV goes through a DNA stage in its lifecycle and at this stage, the unpaired regions of the folded RRE may be vulnerable to deamination by APOBEC3 and thus experience an elevated m_{Tii} as compared to paired regions. If the unpaired regions experienced very high rates of deamination by APOBEC enzymes, then this might explain the reversal of k_u and k_p in RRE. Consistent with this hypothesis, we find that that the nucleotide A is overrepresented in unpaired regions (~ 40%) of RRE as compared to paired regions (~19%).

We also recognize that our estimates of k are sensitive to misidentification of paired and unpaired sites, as well as inaccuracies in the alignment and/or phylogeny. It is possible that sites in loops, which we have classified as unpaired, could be pairing with sites elsewhere (e.g. through unknown pseudoknots) and thus artifactually elevate k_u. In all molecules, known pseudoknots were included in our analysis as paired. While we cannot rule out that a sufficiently large number of sites pair outside of the known RRE secondary structure to influence our estimates, we find this possibility unlikely, and the removal of individual stem-loops from the RRE alignment did not measurably affect k_u. Errors in the underlying phylogeny and alignment could also contribute to inaccurate estimates of k. Although the RRE phylogeny had very low branch support (1-2%) on some of the interior branches, it is not apparent how a poorly supported phylogeny could could elevate k_u in RRE. While alignments were manually checked for accuracy, it is not

possible to ensure that every site in every sequence is correctly identified as paired or unpaired. In particular, minor pervasive structural variation in the analyzed HIV sequences and mammalian tRNA would result in individual sites that are paired in some lineages and unpaired in others. While this phenomenon may contribute to a quantitative deviation from expectation, such as that seen in the mammalian tRNA's, it still fails to explain the reversal of magnitude seen in RRE, where $k_p < k_u$.

In sum, only one phenomena – the action of APOBEC3 enzymes on unpaired regions of the HIV genome – has the ability to explain conceptually the observation in RRE that $k_p << k_u^2$.

Implications for RNA secondary structure prediction

Recently developed methods for predicting RNA secondary structure from sequence data capitalize on the expectation that substitution rates should be lower at paired than unpaired sites (Muse 1995; Pedersen et al. 2004). These methods have been shown to be as strong or stronger in validating known structures and predicting new ones than previous methods that relied solely on the identification of co-variations (Muse 1995; Parsch, Braverman, and Stephan 2000; Pedersen et al. 2004). Our analysis parallels these recent prediction methods in that we also capitalize on the expectation that substitution rates should be lower at paired than unpaired sites. Where we differ from these methods is that we assess an exact expectation for the difference between substitution rates at unpaired and paired sites.

In our extension of Kimura's model, the pattern $k_p \approx k_u^2$ results because transversions are suppressed in stems to a greater extent than transitions (i.e. $d_{C,Tv} / d_{I,Tv} < d_{C,Ti} / d_{I,Ti}$). The secondary structure prediction method of Muse (Muse 1995) assumes that transition and transversion rates are equally reduced in stems. This method maintains $k_p = k_u$ and we show that this assumption is not realistic. The secondary structure prediction method of Pedersen et al (Pedersen et al. 2004) does assume that transition and transversion rates are reduced in stems and the reduction in these rates is based on empirically measured substitution patterns in known stems. This method allows $k_p \neq k_u$, but the measure of k_p is specified identically for all structures in all organisms. We show that this assumption is also not realistic. Our finding that the transition:transversion rate ratio works in most cases as a simple diagnostic for the identification of RNA secondary structure suggests that these prediction methods could be refined by incorporating exact expectations for the substitution rate differences between unpaired and paired sites.

Appendix A: Simulation used to determine experimental design for data collection in Chapter 3

A set of simulation experiments was conducted in MatLab (version 6.5, The Mathworks Inc.) to evaluate the experimental design that yielded the most accurate and least variable measures of the phages thermal optima and maximum growth rates. We considered seven reaction norms with different thermal optima (Fig. A1A). The experimental designs considered ranged from measuring growth rate singly at many (26) temperatures to many (7) times at only a few (4) temperatures. The reaction norm shape was assumed to be known, as was the standard deviation of growth rate at any single temperature. The assumed shape was based on pilot data collected for the natural isolates and the standard deviation based on growth rate measures from previous published work [12].

A single experiment consisted of generating approximately 27 growth rate measurements, in which each measure of growth rate (GR) was:

$GR = GR_i + e,$

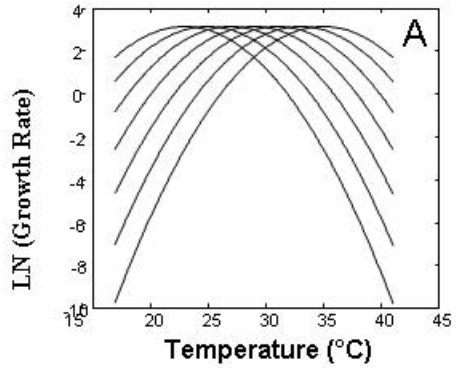
in which GR_i is the true growth rate at temperature i as calculated from the known reaction norm shape and the error (e) was drawn from a normal distribution with mean 0 and standard deviation =1.41.

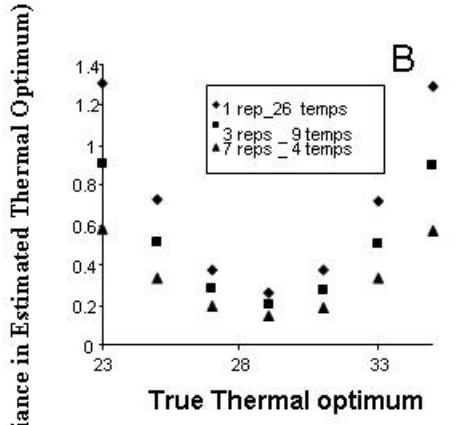
The least squares method was used to fit a second order polynomial to each experiment. The thermal optimum and maximum growth rates were calculated from these polynomials. This process was repeated 1 million times for each experimental

design. The precision of a particular experimental design in estimating a trait (*e.g.* the thermal optima) was calculated as the variance and bias in the estimates for that trait.

Bias = $\sum$(Estimated optima – True optima)/ Number of Simulations.

Reaction norm data collection over 17 to 41^{o}C was simulated in Mat Lab to determine the optimal experimental design for estimating the thermal optima. More extreme values of the optima resulted in increased variance and bias in the estimates of the optima. We found that increasing the number of temperatures did decrease the bias or variance in the parameter estimates (Figs. A1B and C). The experimental design that minimized the variance and bias in the optima consisted of 7 replicate growth rate measures at each of 4 temperatures evenly spaced between 17 and 41^{o}C (Figs. A1B –C). Since the experimental design would have limited our ability to detect reaction norms with a nonsymmetrical shape, the experiments were conducted using the next most optimal design: 3 growth rate measures at each of 9 temperatures evenly spaced between 17 and 41^{o}C.





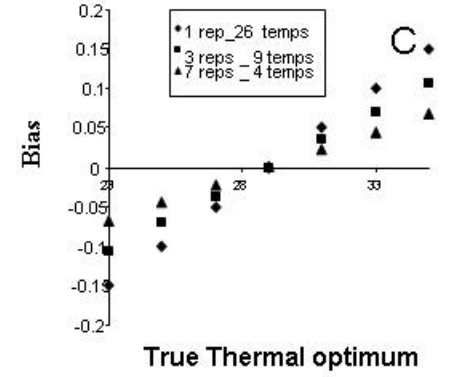


Fig. A1. Simulations used to find the experimental design yielding the best measures of thermal optima. (A) The reaction norms considered in the simulation. (B) The estimates of variance in estimated thermal optima in relation to the true thermal optima. (C) The estimates of bias in the estimated thermal optima in relation to the true thermal optima.

BIBLIOGRAPHY

1. Russell CJ, Webster RG (2005) The Genesis of a Pandemic Influenza Virus. Cell 123: 368.

2. Jonzen N, Linden A, Ergon T, Knudsen E, Vik JO, et al. (2006) Rapid advance of spring arrival dates in long-distance migratory birds. Science 312: 1959-1961.

3. Johnston IA, Bennett AF, editors (1996) Animals and Temperature. Phenotypic and evolutionary adaptation. Cambridge [England]; New York: Cambridge University Press. 435 p.

4. Bruck K (1973) In: Precht H, editor. Temperature and Life. New York: Springer-Verlag. pp. 779.

5. fix m (1996) fix me. In: Bennett AF, Johnston IA, editors. Animals and temperature: phenotypic and evolutionary adaptation. New York: Cambridge University Press. pp. 419.

6. Angilletta MJ, Niewiarowski PH, Navas CA (2002) The evolution of thermal physiology in ectotherms. Journal of Thermal Biology 27: 249-268.

7. Frazier MR, Huey RB, Berrigan D (2006) Thermodynamics constrains the evolution of insect population growth rates: "Warmer is better". American Naturalist 168: 512-520.

8. Fields PA (2001) Review: Protein function at thermal extremes: balancing stability and flexibility. Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology 129: 417-431.

9. Kumar S, Nussinov R (2001) How to thermophilic proteins deal with heat? Cellular and Molecular Life Sciences 58: 1216-1233.

10. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A (2001) How enzymes adapt: lessons from directed evolution. Trends in Biochemical Sciences 26: 100-106.

11. Kimura M (1985) The role of compensatory neutral mutations in molecular evolution. Journal of Genetics 64: 7-19.

12. Knies JL, Izem R, Supler KL, Kingsolver JG, Burch CL (2006) The genetic basis of thermal reaction norm evolution in lab and natural phage populations. Plos Biology 4: 1257-1264.

13. Gilchrist GW (1995) Specialists and generalists in changing environments. 1. Fitness landscapes of thermal sensitivity. American Naturalist 146: 252-270.

14. van der Have TM (2002) A proximate model for thermal tolerance in ectotherms. Oikos 98: 141-155.

15. Lynch M, Gabriel W (1987) Environmental Tolerance. American Naturalist 129: 283-303.

16. Kingsolver J, Izem R, Ragland G (2004) Plasticity of size and growth in fluctuating thermal environments: Comparing reaction norms and performance curves. Integrative and Comparative Biology 44: 450-460.

17. Cooper VS, Bennett AF, Lenski RE (2001) Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution 55: 889-896.

18. Kingsolver JG, Ragland GJ, Shlichta JG (2004) Quantitative genetics of continuous reaction norms: Thermal sensitivity of caterpillar growth rates. Evolution 58: 1521-1529.

19. Izem R, Kingsolver J (2005) Variation in continuous reaction norms: quantifying directions of biological interest. American Naturalist 166: 277-289.

20. Bronikowski AM, Bennett AF, Lenski RE (2001) Evolutionary adaptation to temperature. VII. Effects of temperature on growth rate in natural isolates of *Escherichia coli* and *Salmonella enterica* from different thermal environments. Evolution 55: 33-40.

21. Gilchrist GW (1996) A quantitative genetic analysis of thermal sensitivity in the locomotor performance curve of *Aphidius ervi.* Evolution 50: 1560-1572.

22. Mongold JA, Bennett AF, Lenski RE (1996) Evolutionary adaptation to temperature.4. Adaptation of Escherichia coli at a niche boundary. Evolution 50: 35-43.

23. Boulos A, Rolain JM, Maurin M, Raoult D (2004) Measurement of the antibiotic susceptibility of Coxiella burnetii using real time PCR. International Journal of Antimicrobial Agents 23: 169-174.

24. Keblys M, Bernhoft A, Hofer CC, Morrison E, Larsen HJ, et al. (2004) The effects of the Penicillium mycotoxins citrinin, cyclopiazonic acid, ochratoxin A, patulin, penicillic acid, and roquefortine C on in vitro proliferation of porcine lymphocytes. Mycopathologia 158: 317-324.

25. Salvucci ME, Crafts-Brandner SJ (2004) Relationship between the heat tolerance of photosynthesis and the thermal stability of rubisco activase in plants from contrasting thermal environments. Plant Physiology 134: 1460-1470.

26. Huey RB, Stevenson RD (1979) Integrating thermal physiology and ecology of ectotherms: a discussion of approaches. American Zoologist 19: 357-366.

27. Huey RB, Kingsolver JG (1989) Evolution of Thermal Sensitivity of Ectotherm Performance. Trends in Ecology & Evolution 4: 131-135.

28. Hahn MW, Pockl M (2005) Ecotypes of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. Appl Environ Microbiol 71: 766-773.

29. Bennett AF, Lenski RE (1993) Evolutionary Adaptation to Temperature.2. Thermal Niches of Experimental Lines of *Escherichia-Coli*. Evolution 47: 1-12.

30. Riehle MM, Bennett AF, Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli.* Proc Natl Acad Sci U S A 98: 525-530.

31. Holder KK, Bull JJ (2001) Profiles of adaptation in two similar viruses. Genetics 159: 1393-1404.

32. Bull JJ, Badgett MR, Wichman HA (2000) Big-benefit mutations in a bacteriophage inhibited with heat. Molecular Biology and Evolution 17: 942-950.

33. Rokyta DR, Burch CL, Caudle SB, Wichman HA (2006) Horizontal gene transfer and the evolution of microvir id coliphage genomes. Journal of Bacteriology 188: 1134-1142.

34. Sokal RR, Rohlf FJ (1995) Biometry; Brook SUoNYaS, editor. New York: W.H. Freeman and Company. 887 p.

35. Romanova J, Katinger D, Ferko B, Vcelar B, Sereinig S, et al. (2004) Live cold-adapted influenza A vaccine produced in Vero cell line. Virus Research 103: 187-193.

36. Levins R (1968) Evolution in Changing Environments. Princeton: Princeton University Press. 132 p.

37. Bennett AF, R.E. Lenski (1999) Experimental evolution and its role in evolutionary physiology. American Zoologist 39: 346-362.

38. Scheiner S, Caplan R, Lyman R (1989) A search for trade-offs among life history traits in *Drosophila melanogaster*. Evolutionary Ecology 3: 51-63.

39. Wichman HA, Scott LA, Yarber CD, Bull JJ (2000) Experimental evolution recapitulates natural evolution. Philos Trans R Soc Lond B Biol Sci 355: 1677-1684.

40. Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB (2002) Adaptive divergence in experimental Populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics 161: 33-46.

41. Lenski RE, Winkworth CL, Riley MA (2003) Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. Journal of Molecular Evolution 56: 498-508.

42. Bowler K, Cossins AR (1987) Temperature biology of animals. New York: Chapman and Hall. 339 p.

43. Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL (2001) Effects of size and temperature on metabolic rate. Science 293: 2248-2251.

44. Hamilton W (1973) Life's color code. New York: McGraw-Hill.

45. Heinrich B (1977) Why Have Some Animals Evolved To Regulate A High Body-Temperature. American Naturalist 111: 623-640.

46. Bauwens D, Garland T, Castilla AM, Vandamme R (1995) Evolution of Sprint Speed in Lacertid Lizards - Morphological, Physiological, and Behavioral Covariation. Evolution 49: 848-863.

47. Rehfeld GE, Tchebakova NM, Parfenova YI, Wykoff WR, Kuzmina NA, et al. (2002) Intraspecific responses to climate in *Pinus sylvestris*. Global Change Biology 8: 912-929.

48. Heilmayer O, Brey T, Portner HO (2004) Growth efficiency and temperature in scallops: a comparative analysis of species adapted to different temperatures. Functional Ecology 18: 641-647.

49. Fukunaga K, Hummels DM (1989) Leave-One-out Procedures for Nonparametric Error-Estimates. Ieee Transactions on Pattern Analysis and Machine Intelligence 11: 421-423.

50. Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

51. Felsenstein J (1985) Phylogenies And The Comparative Method. American Naturalist 125: 1-15.

52. Maddison WP, Maddison DR (2006) Mesquite: A modular system for evolutionary analysis. Version 1.1 ed.

53. Midford PE, Garland Jr. T, Maddison WP (2005) PDAP Package of Mesquite. Version 1.07 ed.

54. Garland T, Harvey PH, Ives AR (1992) Procedures for the Analysis of Comparative Data Using Phylogenetically Independent Contrasts. Systematic Biology 41: 18-32.

55. Smith MD (1992) Comparing the Exact Distribution of the T-Statistic to Students Distribution When Its Constituent Normal and Chi-Squared Variables Are Dependent. Communications in Statistics-Theory and Methods 21: 3589-3600.

56. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proceedings of the National Academy of Sciences of the United States of America 98: 13757-13762.

57. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21: 260-262.

58. Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: Empirical data. Molecular Biology and Evolution 19: 708-717.

59. Huey RB, Hertz P (1984) Is a jack-of-all-temperatures a master of none? Evolution 38: 441-444.

60. Portner HO (2004) Climate variability and the energetic pathways of evolution: The origin of endothermy in mammals and birds. Physiological And Biochemical Zoology 77: 959-981.

61. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. Microbiology and Molecular Biology Reviews 65: 1-+.

62. Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A Relationship between Protein Stability and Protein Function. Proceedings of the National Academy of Sciences of the United States of America 92: 452-456.

63. Kawecki TJ (1994) Accumulation of Deleterious Metations and the Evolutionary Cost of Being a Generalist. American Naturalist 144: 833-838.

64. Bennett AF, Lenski RE (1997) Evolutionary adaptation to temperature.6. Phenotypic acclimation and its evolution in Escherichia coli. Evolution 51: 36-44.

65. Counago R, Chen S, Shamoo Y (2006) In vivo molecular evolution reveals biophysical origins of organismal fitness. Molecular Cell 22: 441-449.

66. Bennett AF, Lenski RE (2007) Colloquium Papers: An experimental test of evolutionary trade-offs during temperature adaptation. PNAS 104: 8649-8654.

67. Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, et al. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. Proceedings of the National Academy of Sciences of the United States of America 96: 3578-3583.

68. Finch CE (1990) Longevity Senescence, and the genome. Chicago: University of Chicago Press. 922 p.

69. Comfort A (1979) The Biology of Senescence. New york: Elsevier.

70. Gompertz B (1825) On the nature of the function expressive of the law of human mortaility and on a new mode of determining the value of life contingencies. Phil Transact R Soc 115: 513-585.

71. Weibull W (1951) A statistical distribution function of wide applicabilit. Journal of Appl Mech 18: 293-297.

72. Wilson DL (1994) The Analysis of Survival (Mortality) Data - Fitting Gompertz, Weibull, and Logistic Functions. Mechanisms of Ageing and Development 74: 15-33.

73. Kyte J, Doolittle RF (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. Journal of Molecular Biology 157: 105-132.

74. Dello Russo A, Rullo R, Nitti G, Masullo M, Bocchini V (1997) Iron superoxide dismutase from the archaeon Sulfolobus solfataricus: average hydrophobicity and amino acid weight are involved in the adaptation of proteins to extreme environments. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1343: 23-30.

75. Miyazaki K, Takenouchi M, Kondo H, Noro N, Suzuki M, et al. (2006) Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. Journal of Biological Chemistry 281: 10236-10242.

76. McKenna R, Bowman BR, Ilag LL, Rossmann MG, Fane BA (1996) Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced

structural changes in the presence of calcium ions and functional implications. J Mol Biol 256: 736-750.

77. Muller H (1950) Our load of mutations. Am J Human Genet 2: 111-176.

78. Dobzhansky T (1955) A Review of Some Fundamental Concepts and Problems of Population Genetics. Cold Spring Harbor Symposia on Quantitative Biology 20: 1-15.

79. Kimura M, Ohta T (1971) Protein Polymorphism as a Phase of Molecular Evolution. Nature 229: 467-&.

80. Kimura M, Ohta T (1971) Theoretical Aspects of Population Genetics. Princeton: Princeton Univ Press.

81. Poon A, Otto S (2000) Compensating for our load of mutations: freezing the meltdown of small populations. Evolution Int J Org Evolution 54: 1467-1479.

82. Poon A, Chao L (2005) The Rate of Compensatory Mutation in the DNA Bacteriophage {phi}X174. Genetics Genetics 170: 989-999.

83. Burch C, Chao L (1999) Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics 151: 921-927.

84. Schrag SJ, Perrot V, Levin BR (1997) Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proceedings of the Royal Society of London Series B-Biological Sciences 264: 1287-1291.

85. Maisnier-Patin S, Berg OG, Liljas L, Andersson DI (2002) Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. Molecular Microbiology 46: 355-366.

86. Hoffman NG, Schiffer CA, Swanstrom R (2005) Covariation of amino acid positions in HIV-1 protease (vol 314, pg 536, 2003). Virology 331: 206-207.

87. Parsch J, Braverman J, Stephan W (2000) Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics 154: 909-921.

88. Kern AD, Kondrashov FA (2004) Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. Nature Genetics 36: 1207-1212.

89. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, et al. (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics 171: 1813-1822.

90. Belle EMS, Piganeau G, Gardner M, Eyre-Walker A (2005) An investigation of the variation in the transition bias among various animal mitochondrial DNA. Gene 355: 58-66.

91. Wilke C, Lenski R, Adami C (2003) Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. BMC Evol Biol 3: 3.

92. Kirby DA, Muse SV, Stephan W (1995) Maintenance of Pre-Messenger-Rna Secondary Structure by Epistatic Selection. Proceedings of the National Academy of Sciences of the United States of America 92: 9047-9051.

93. Innan H, Stephan W (2001) Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. Genetics 159: 389-399.

94. Stephan W (1996) The rate of compensatory evolution. Genetics 144: 419-426.

95. Chenna R, Sugawara H, Koike T, Lopez R, Gibson T, et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497-3500.

96. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511-518.

97. Phuphuakrat A, Auewarakul P (2003) Heterogeneity of HIV-1 Rev response element. Aids Research and Human Retroviruses 19: 569-574.

98. Tuplin A, Evans D, Simmonds P (2004) Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J Gen Virol 85: 3037-3047.

99. Tuplin A, Wood J, Evans D, Patel A, Simmonds P (2002) Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. RNA 8: 824-841.

100. Springer M, Hollar L, Burk A (1995) Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. Mol Biol Evol 12: 1138-1150.

101. Fox GE, Woese CR (1975) 5S-Rna Secondary Structure. Nature 256: 505-507.

102. Specht T, Wolters J, Erdmann VA (1991) Compilation of 5S Ribosomal-Rna and 5S Ribosomal-Rna Gene-Sequences. Nucleic Acids Research 19: 2189-2191.

103. Haas ES, Morse DP, Brown JW, Schmidt FJ, Pace NR (1991) Long-Range Structure in Ribonuclease-P Rna. Science 254: 853-856.

104. Thurner C, Witwer C, Hofacker IL, Stadler PF (2004) Conserved RNA secondary structures in Flaviviridae genomes. J Gen Virol 85: 1113-1124.

105. Szymanski M, Barciszewska M, Erdmann V, Barciszewski J (2002) 5S Ribosomal RNA Database. Nucleic Acids Res 30: 176-178.

106. Brown J (1999) The Ribonuclease P Database. Nucleic Acids Res 27: 314.

107. Harris J, Haas E, Williams D, Frank D, Brown J (2001) New insight into RNase P RNA structure from comparative analysis of the archaeal RNA. RNA 7: 220-232.

108. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

109. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S (1998) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 26: 148-153.

110. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.

111. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160-174.

112. Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. Systematic Biology 50: 580-601.

113. Hillis DM (1999) Phylogenetics and the study of HIV. In: Crandall KA, editor. The evolution of HIV. Baltimore: The Johns Hopkins University Press. pp. 504.

114. Hanly SM, Rimsky LT, Malim MH, Kim JH, Hauber J, et al. (1989) Comparative-Analysis of the Htlv-I Rex and Hiv-1 Rev Trans-Regulatory Proteins and Their Rna Response Elements. Genes & Development 3: 1534-1544.

115. Lesnik E, Sampath R, Ecker D (2002) Rev response elements (RRE) in lentiviruses: an RNAMotif algorithm-based strategy for RRE prediction. Med Res Rev 22: 617-636.

116. Zhang M, Dayton A (1996) Two secondary structures for the RRE of HIV-1? J Acquir Immune Defic Syndr Hum Retrovirol 13: 403-407.

117. Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. Journal of Molecular Evolution 54: 156-165.

118. Muse SV (1995) Evolutionary Analyses of DNA Sequences Subject to Constraints on Secondary Structure. Genetics 139: 1429-1439.

119. Pedersen J, Meyer I, Forsberg R, Simmonds P, Hein J (2004) A comparative method for finding and folding RNA secondary structures within protein-coding regions. Nucleic Acids Res 32: 4925-4936.