REPORT DOCUMENTATION PAGE		Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
25-02-2016	Final Report	1-Oct-2012 - 30-Sep-2015

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Inferring Microbial Fitness Landscapes	W911NF-12-1-0552
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
	611102

6. AUTHORS	5d. PROJECT NUMBER
Joshua Plotkin, Charles Epstein	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Pennsylvania Office of Research Services 3451 Walnut Street, Suite P-221 Philadelphia, PA 19104 -6205	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	62345-MA.30

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Microbes and viruses evolve. Their evolution is often more rapid and of greater practical importance than our own evolution. How can we understand, or even predict, the evolutionary trajectory of microbes as they adapt? For example, what determines how quickly, and by what specific mutations, avian influenza viruses will adapt to novel human hosts; or how readily infectious bacteria will escape antibiotics or the human immune system?
In this research program we seek to combine mathematical models and statistical techniques to tackle this problem

15. SUBJECT TERMS
evolution, fitness landscapes, epistasis, microbial populations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU		JOSHUA PLOTKIN
UU	UU	UU			19b. TELEPHONE NUMBER
					215-898-7133

Report Title

Final Report: Inferring Microbial Fitness Landscapes

ABSTRACT

Microbes and viruses evolve. Their evolution is often more rapid and of greater practical importance than our own evolution. How can we understand, or even predict, the evolutionary trajectory of microbes as they adapt? For example, what determines how quickly, and by what specific mutations, avian influenza viruses will adapt to novel human hosts; or how readily infectious bacteria will escape antibiotics or the human immune system?

In this research program we seek to combine mathematical models and statistical techniques to tackle this problem head-on: to infer from data the determinants of microbial evolution with sufficient resolution that we can quantify their evolutionary trajectories, and sometimes even predict the details of their evolution.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received   Paper

02/23/2016 16.00   David E. Weinberg, Premal Shah, Stephen W. Eichhorn, Jeffrey A. Hussmann, Joshua B. Plotkin, David P. Bartel. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation,
Cell Reports, (02 2016): 1787. doi: 10.1016/j.celrep.2016.01.043

02/23/2016 14.00   Jakub Otwinowski, Joshua B. Plotkin, David M. McCandlish. Detecting epistasis from an ensemble of adapting populations,
Evolution, (09 2015): 2359. doi: 10.1111/evo.12735

02/23/2016 20.00   Michael B Schulte, Jeremy A Draghi, Joshua B Plotkin, Raul Andino. Experimentally guided models reveal replication principles that shape the mutation distribution of RNA viruses,
ELife, (01 2015): 3753. doi: 10.7554/eLife.03753

02/23/2016 18.00   Alexander Stewart, Joshua Plotkin. The Evolvability of Cooperation under Local and Non-Local Mutations,
Games, (07 2015): 231. doi: 10.3390/g6030231

02/23/2016 17.00   Alexey D. Neverov, Sergey Kryazhimskiy, Joshua B. Plotkin, Georgii A. Bazykin, Harmit S. Malik. Coordinated Evolution of Influenza A Surface Proteins,
PLoS Genetics, (08 2015): 1005404. doi: 10.1371/journal.pgen.1005404

02/24/2016 27.00   Alexander Stewart, Joshua Plotkin. From extortion to generosity, evolution in the IteratedPrisoner's Dilemma,
PNAS, (09 2013): 15348. doi:

02/24/2016 26.00   R. Der, J. B. Plotkin. The Equilibrium Allele Frequency Distribution for a Population with Reproductive Skew,
Genetics, (01 2014): 1199. doi: 10.1534/genetics.114.161422

02/24/2016 24.00   D. Gulisija, Y. Kim, J. B. Plotkin. Phenotypic Plasticity Promotes Balanced Polymorphism in Periodic Environments by a Genomic Storage Effect,
Genetics, (02 2016): 0. doi: 10.1534/genetics.115.185702

02/24/2016 23.00   David M. McCandlish, Charles L. Epstein, Joshua B. Plotkin. THE INEVITABILITY OF UNCONDITIONALLY DELETERIOUS SUBSTITUTIONS DURING ADAPTATION,
Evolution, (05 2014): 1351. doi: 10.1111/evo.12350

02/24/2016 22.00   Alexander J. Stewart, Joshua B. Plotkin. Collapse of cooperation in evolving games,
Proceedings of the National Academy of Sciences, (12 2014): 17558. doi: 10.1073/pnas.1408618111

02/24/2016 21.00   David M. McCandlish, Charles L. Epstein, Joshua B. Plotkin. Formal properties of the probability of fixation: Identities, inequalities and approximations,
Theoretical Population Biology, (02 2015): 98. doi: 10.1016/j.tpb.2014.11.004

02/24/2016 19.00   Premal Shah, David M. McCandlish, Joshua B. Plotkin. Contingency and entrenchment in protein evolution under purifying selection,
Proceedings of the National Academy of Sciences, (06 2015): 3226. doi: 10.1073/pnas.1412933112

07/03/2013 1.00   Premal Shah, Yang Ding, Malwina Niemczyk, Grzegorz Kudla, Joshua B. Plotkin. Rate-Limiting Steps in Yeast Protein Translation,
Cell, (06 2013): 0. doi: 10.1016/j.cell.2013.05.049

07/03/2013 5.00 Jeremy A. Draghi, Joshua B. Plotkin. SELECTION BIASES THE PREVALENCE AND TYPE OF
EPISTASIS ALONG ADAPTIVE TRAJECTORIES,
Evolution, (06 2013): 0. doi: 10.1111/evo.12192

07/03/2013 2.00 Etienne Rajon, David M. McCandlish, Premal Shah, Yang Ding, Joshua B. Plotkin. The role of epistasis in
protein evolution,
Nature, (05 2013): 0. doi: 10.1038/nature12219

08/12/2013 7.00 Etienne Rajon, Joshua Plotkin. The evolution of genetic architecturesunderlying quantitative traits,
Proceeding of the Royal Society B, (09 2013): 0. doi:

08/22/2014 10.00 A. F. Feder, S. Kryazhimskiy, J. B. Plotkin. Identifying Signatures of Selection in Genetic Time Series,
Genetics, (12 2013): 0. doi: 10.1534/genetics.113.158220

08/22/2014 9.00 J. B. Plotkin, J. Otwinowski. Inferring fitness landscapes by regression produces biased estimates of
epistasis,
Proceedings of the National Academy of Sciences, (05 2014): 0. doi: 10.1073/pnas.1400849111

TOTAL: **18**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u>   <u>Paper</u>

08/12/2013 6.00 Alexander Stewart, Joshua Plotkin. From extortion to generosity, evolution in the Iterated Prisoner's
Dilemma,
PNAS, (09 2013): 0. doi:

TOTAL: **1**

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received   Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received   Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper

02/23/2016 25.00 Alexander Stewart, Joshua Plotkin. Small games and long memories promote cooperation,
ArXiv (11 2015)

02/23/2016 15.00 Oana Carja, Joshua Plotkin. The evolutionary advantage of heritable phenotypic heterogeneity,
biorXiv (10 2016)

02/24/2016 28.00 Armita Nourmohammad, Jakub Otwinowski, Joshua Plotkin. Host-pathogen co-evolution and the
emergence of broadly neutralizing antibodies in chronic infections,
biorXiv (12 2015)

02/25/2016 29.00 David McCandlish, Joshua Plotkin, Mitchell Newberry. Assortative mating can impede or faciliatefixation
of underdominant alleles,
biorXiv ( )

07/03/2013 4.00 McCandlish David, Epstein Charles, Plotkin Joshua. The inevitability of unconditionally
deleteriousfixations during adaptation,
 (08 2013)

08/13/2013 8.00 David McCandlish, Charles Epstein, Joshua Plotkin. The inevitability of unconditionally deleterious
fixations during adaptation,
Evolution (09 2013)

08/22/2014 11.00 Mazzeo Rafe, Epstein Charles. Harnack Inequalities and Heat-kernel Estimates for Degenerate Diffusion
Operators Arising in Population Biology,
Analysis and PDE (09 2014)

08/22/2014 12.00 Epstein Charles, Pop Camelia. Regularity for the Supercritical Fractional Laplacian with Drift,
COMM. Math. Phys. (08 2014)

08/22/2014 13.00 Epstein Charles, Pop Camelia. Harnack Inequalities for Degenerate Diffusions,
Annals of Probability (08 2014)

TOTAL: **9**

Number of Manuscripts:

Books

Received		Book

TOTAL:

<u>Received</u>   <u>Book Chapter</u>

TOTAL:

Patents Submitted

Patents Awarded

Awards

Akira Okubo Prize of the Society for Mathematical Biology (to Joshua B. Plotkin)
Board of Reviewing Editors of Science Magazine / AAAS (to Joshua B. Plotkin)

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Mitchell Johnson	0.25	
Yang Ding	0.25	
FTE Equivalent:	**0.50**	
Total Number:	2	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
David McCandlish	0.80
FTE Equivalent:	**0.80**
Total Number:	1

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Joshua Plotkin	0.02	
Charles Epstein	0.08	No
FTE Equivalent:	**0.10**	
Total Number:	2	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME
Yang Ding

Total Number: 1

Names of other research staff

NAME   PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Attachment

Technology Transfer

ARO Project Summary Sheet
Proposal 62345-MA, *Inferring Microbial Fitness Landscapes*
17 February 2016

Investigators: Joshua B. Plotkin & Charles Epstein
                Departments of Biology and Mathematics
                University of Pennsylvania
                Philadelphia, PA

OBJECTIVE
　　　Microbes and viruses evolve. Their evolution is often more rapid and of greater practical importance than our own evolution. How can we understand, or even predict, the evolutionary trajectory of microbes as they adapt? For example, what determines how quickly, and by what specific mutations, avian influenza viruses will adapt to novel human hosts; or how readily infectious bacteria will escape antibiotics or the human immune system?
　　　In this research program we seek to combine mathematical models and statistical techniques to tackle this problem head-on: to infer from data the determinants of microbial evolution with sufficient resolution that we can quantify their evolutionary trajectories, and sometimes even predict the details of their evolution.

SCIENTIFIC BARRIERS
　　　The rules of evolution are simple: mutations introduce variants into a population, whose frequencies then change by genetic drift and natural selection. But the resulting evolutionary dynamics are extraordinarily complicated -- because they depend on the so-called ``fitness landscape'' that describes the fitness (reproductive rate) associated with each possible genetic type of the organism. Despite its central importance in evolution, very little is known about the actual fitness landscape of any biological organism.
　　　An expanding body of experimental data on microbial populations has begun to provide the empirical basis required to draw inferences about organismal fitness landscapes and how they shape evolution. Nevertheless, even for simple organisms, the number of possible genotypes is astronomically large, and therefore the fitness landscape is very high dimensional. High-throughput experiments on laboratory populations of microbes produce massive amounts of data, and yet still not nearly enough data to determine an entire fitness landscape directly.
　　　This presents the field with several pressing questions: how do we infer fitness landscapes from limited samples of genotypes? Do statistical approximations based on available data faithfully reproduce the true fitness landscape and accurately predict the dynamics of adaptation? Can we leverage time-series data to learn more about the fitness landscape and its effect on microbial evolution? These questions are intrinsically mathematical and statistical in nature. Answering these questions demands familiarity with the empirical literature on evolving microbes and how they are interrogated experimentally; as well as familiarity with the mathematical and statistical techniques required to draw meaningful inferences from these data.

SIGNIFICANCE
　　　Quantitative models of evolution have historically assumed simple models of the fitness landscape, with no serious attempt to determine its actual structure in nature. But the moment has arrived when empirical data, analytic sophistication, and computational tools make it feasible to determine the actual fitness landscapes of some organisms.

The payoffs of such a research program are potentially manifold -- both for the intellectual development of evolutionary theory and for practical applications to controlling viral and microbial disease. The practical payoffs hold particular interest for the Army, which regularly exposes its war-fighters to the insults and risks of novel pathogens.

APPROACH

Our approach to inferring microbial fitness landscapes combines mathematical models, statistical techniques, and detailed empirical data drawn from laboratory and wild populations of bacteria and viruses.

The quantitative methods we use are rooted in probability theory, stochastic processes, and PDEs. Such techniques are required because differences in fitness are understood as the deterministic, driving force in an evolving population, which is balanced against the stochastic forces of genetic drift and mutation. Inferring the fitness landscape thus requires that we discriminate between stochastic effects of drift, and the deterministic effects of selection.

More specifically, we are working to characterize the dynamics of adaptation in forward time on large families of mathematical fitness landscapes; and then, conversely, leverage empirical data to infer the fitness landscape on which an organisms is adapting. We are exploiting a variety of techniques, new and old, to describe fitness lanscapes -- including generalizations of the famous NK landscapes of Kauffman and Levin, as well as our recent technique describing a landscape as a family of distributions of mutational effects. In addition we are developing mathematical and computational tools, using infinite-population diffusion limits of standard Markov models, to simulate different fitness and mutational scenarios.

This approach is entirely novel. The combination of a precise, mathematical understanding of forward-time dynamics to provide a rigorous method for inferring the determinants of evolution from data has not yet been seriously attempted – and it has the potential to provide substantial and practical payoffs.

ACCOMPISHMENTS

We made tremendous progress towards the goals of our proposed ARO research program. Over the course of the grant we published 26 papers, ranging from topics such as the role of epistasis in protein evolution, the structure of epistasis along adaptive walks, the inference of fitness landscapes from time-series data or experimental evolution, the role of deleterious mutations during adaptation, as well as the importance of frequency-dependent effects, such as cooperation, in evolving populations. Below I will highlight just a few of these projects, and also provide a list of all ARO-funded publications.

Inferring epistasis from microbial evolution experiments

Recent years have seen a proliferation of controlled, laboratory experiments on evolving microbial populations. Although these experiments have produced examples of remarkable phenomena – *e.g.* the emergence of mutator strains, of long-term frequency-dependent selection, of novel metabolic capabilities, and even multi-cellularity – a synthetic understanding of how to draw inferences about the forces that shaped the course of evolution in these populations is still lacking.

Over the past year, I have begun initial foray into a long-term research program on how to draw principled inferences from laboratory evolution experiments. I have focused on how to infer the presence of epistasis – that is, interactions between genetic mutations that collectively influence phenotype and fitness. The role that epistasis plays during adaptation remains an outstanding problem, which has received considerable attention in recent years. Most of the

recent empirical studies are based on ensembles of replicate populations that adapt in a fixed, laboratory controlled condition. Researchers often seek to infer the presence and form of epistasis in the fitness landscape from the time evolution of various statistics averaged across the ensemble of populations. However, researchers lack a firm statistical framework for drawing such inferences.

Therefore, I have begun to develop a rigorous analysis of what quantities, drawn from time series of such ensembles of experimental populations, can be used to infer epistasis in the fitness landscape. Along with two post-docs in my group, we have analyzed the mean fitness trajectory—that is, the time course of the ensemble average fitness. We have shown that for any epistatic fitness landscape and starting genotype, there always exists a non-epistatic fitness landscape that produces the exact same mean fitness trajectory. Thus, the presence of epistasis is not identifiable from the mean fitness trajectory. By contrast, we have shown that two other ensemble statistics—the time evolution of the fitness variance across populations, and the time evolution of the mean number of substitutions—can detect certain forms of epistasis in the underlying fitness landscape. This work provides foundational guidance to experimentalists who wish to draw inferences about how genetic interactions shape evolution. A paper describing these results was published in *Evolution*. The topic remains a central focus on ongoing research in my group.

Epistasis along an adaptive walk

In another project, we have systematically studied how selection can bias the amount of epistasis observed among mutations that substitute while a population is adapting. Epistasis refers to non-additive interactions among loci that collectively determine the fitness of an organism. Such epistatic interactions are recognized as fundamental to shaping the process of adaptation in evolving populations. Although little is known about the structure of epistasis in most organisms, recent experiments with bacterial populations have concluded that antagonistic interactions abound and tend to de-accelerate the pace of adaptation over time.

We used the NK mathematical model of fitness landscapes to examine how natural selection biases the mutations that substitute during evolution, based on their epistatic interactions. We found that, even when beneficial mutations are rare, natural selection strongly biases the types of mutations that will fix; more importantly, the form of these biases change substantially throughout the course of adaptation. In particular, epistasis is less prevalent than the neutral expectation early in adaptation and much more prevalent later, with a concomitant shift from predominantly antagonistic interactions early in adaptation to synergistic and sign epistasis later in adaptation.

We confirmed our conclusions by analyzing data from a recent microbial evolution experiment. Our results show that when the order of substitutions is not known, standard methods of analysis may suggest that epistasis retards adaptation when in fact it accelerates it. These results have immediate implications for how researchers should interpret the observed fitness contributions of mutations that substitute in a population under selection. We published a paper describing these results in *Evolution*.

Inferring epistasis from genetic time-series

Over the past year, we have developed an entirely new approach to inferring selection on mutations, which will be especially useful for contemporary data on viruses and bacteria. Population geneticists typically seek to understand the selective forces responsible for patterns observed in contemporaneous samples of genetic data. Recently, however, there has been a rapid increase in the availability of dynamic data, where the frequencies of segregating alleles in an evolving population are monitored through time, both in laboratory experiments and and

natural populations. One important question is whether the changes in allele frequencies observed in such data are the result of natural selection or are simply consequences of genetic drift or sampling noise. In principle, it seems that dynamic data should provide researchers with more power to detect and quantify selective forces while avoiding the assumptions of stationarity that are required for many inference techniques based on static samples.

A standard chi-squared-based likelihood ratio test was previously proposed to address this problem. We have shown that the chi-squared test of selection substantially underestimates the probability of Type I error, leading to more false positives than indicated by its P-value, especially at stringent P-values. We developed two methods to correct this bias. The empirical likelihood ratio test rejects neutrality when the likelihood ratio statistic falls in the tail of the empirical distribution obtained under the most likely neutral population size. The frequency increment test rejects neutrality if the distribution of normalized allele frequency increments exhibits a mean that deviates significantly from zero. We characterized the statistical power of these two new tests for selection, and we applied them to three experimental data sets. We have shown that both of these new techniques have power to detect selection in practical parameter regimes, such as those encountered in fitness assays of microbial populations. A paper describing these results is in press at *Genetics*.

Deleterious mutations and adaptation

In a new theoretical direction this past year we have also studied the role of deleterious substitutions during adaptation – that is, the chance that deleterious mutations might fix, with no productive side effect whatsoever, while a population is adapting. The literature on the genetics of adaptation typically neglects the possibility that deleterious mutations will fix in a population. We have shown, by contrast, that even when a population is destined to adapt towards higher fitness over the long term, the first mutation to fix will often decrease fitness. In fact, in many regimes of populations undergoing long-term adaptation, the expected effect of the first substitution is actually to decrease fitness. We demonstrated these results under two of the most widely used models of fitness landscapes: the house of cards model of Kingman and Fisher's geometric model. Importantly, we also developed a simple intuition to help explain the surprising prevalence of deleterious substitutions during adaptation.

These results have implications for our understanding of adaptation. First, our results imply that the common practice of neglecting deleterious substitutions can lead to qualitatively incorrect predictions for the dynamics of adaptation. More generally, our analysis helps to dispel the widespread, but mistaken, impression that a population below its equilibrium mean fitness will increase in fitness as it approaches equilibrium. Finally, our results have practical implications for the expected pattern of substitutions in response to a change in population size. A paper describing these results and their implications is in press at *Evolution*.

Role of epistasis in protein evolution

An important question in molecular evolution is whether an amino acid that occurs at a given site makes an independent contribution to fitness, or whether its contribution depends on the state of other sites in the organism's genome known as epistasis. Work by Kondrashov and colleagues recently argued that epistasis must be pervasive throughout protein evolution, because the observed ratio between the per-site rates of non-synonymous and synonymous substitutions (dN/dS) is much lower than would be expected in the absence of epistasis. However, when calculating the expected dN/dS ratio in the absence of epistasis, Kondrashov assumed that all amino acids observed at a given position in a protein alignment have equal fitness. We relaxed this unrealistic assumption and found that any dN/dS value can in principle be achieved at a site, without epistasis; furthermore, for all nuclear and chloroplast genes in the

Kondrashov data set, we showed that the observed dN/dS values and the observed patterns of amino-acid diversity at each site are jointly consistent with a non-epistatic model of protein evolution. These results are important because they highlight the need for more nuanced techniques, such as the time-series methods discussed above, for inferring fitness landscapes. We published these results in *Nature*.

I have also worked to understand how epistasis between sites in a single protein may influence the course of its evolution. We used computational models of thermodynamic stability in a ligand-binding protein to explore the structure of epistasis in simulations of protein sequence evolution. Even though the predicted effects on stability of random mutations are almost completely additive, we found that the mutations that fix under purifying selection are enriched for epistasis. In particular, the mutations that fix are contingent on previous substitutions: Although nearly neutral at their time of fixation, these mutations would be deleterious in the absence of preceding substitutions. Conversely, substitutions under purifying selection are subsequently entrenched by epistasis with later substitutions: They become increasingly deleterious to revert over time. Our results imply that, even under purifying selection, protein sequence evolution is often contingent on history and so it cannot be predicted by the phenotypic effects of mutations assayed in the ancestral background. We published a study describing these results in *Proceedings of the National Academy of Sciences USA*.

Detecting epistasis between viral surface proteins.

In related work, I have also generalized earlier work to detect epistasis in evolving viral proteins. Previously, my group has detected epistasis between sites within individual viral surface proteins undergoing adaptation. However, the extent to which evolution of one viral protein affects the evolution of the other one is unknown. Therefore, working with colleagues from several countries, we developed a novel phylogenetic method for detecting the signatures of genetic interactions between mutations in different genes – that is, inter-gene epistasis. Using this method, we showed that influenza surface proteins evolve in a coordinated way, with mutations in Hemagglutinin affecting subsequent spread of mutations in Neuraminidase and vice versa, at many sites. Of particular interest was our finding that the oseltamivir-resistance mutations in NA in subtype H1N1 were likely facilitated by prior mutations in HA. Our results illustrate that the adaptive landscape of a viral protein is remarkably sensitive to its genomic context and, more generally, that the evolution of any single protein must be understood within the context of the entire evolving genome. We published a paper describing these results in *PLoS Genetics*.

Inferring epistasis from sampled genotypes

Measuring a microbe's fitness landscape is virtually impossible in practice, because of the coarse resolution of fitness measurements and because of episasis: the fitness contribution of one locus may depend on the states of other loci. To account for all possible forms of epistasis, a fitness landscape must assign a potentially different fitness to each genotype, the number of which increases exponentially with the number of loci.

To draw conclusions from a limited number of sampled genotypes whose fitnesses can be assayed, researchers fit statistical models to approximate the fitness landscape based on available data. This situation is perhaps best illustrated by recent studies of the HIV-1 virus. HIV genotypes were sampled from infected patients, and assayed for reproductive rate. Whereas the entire fitness landscape of HIV-1 consists of reproductive values for roughly 10^{600} genotypes, only ~70,000 genotypes were sampled. Researchers therefore approximated the fitness landscape, based on the measured data, by an expansion in terms of main effects of loci and epistatic interactions among loci. This presents the field with several pressing questions: Do

statistical approximations based on available data faithfully reproduce the relevant aspects of the true fitness landscape and accurately predict the dynamics of adaptation?

We have begun to address these fundamental questions about empirical fitness measurements and how they inform our understanding of the underlying fitness landscape. We have quantified the effects of approximating a fitness landscape from data in terms of main and epistatic effects of loci. We demonstrated that such approximations are subject to two distinct sources of biases that each tend to under-estimate high fitnesses and over-estimate low fitnesses. Biases in the inferred landscape distort commonly used measures of epistasis in the landscape. As a result, the inferred landscape will provide systematically biased predictions for the dynamics of adaptation. We have identified the same biases in a computational RNA-folding landscape, as well as in transcription factor binding data, treated to the same fitting procedure. Finally, we have developed a method to ameliorate these biases in certain circumstances. A manuscript describing these results is under consideration at *Proceedings of the National Academy of Sciences.*

Mathematical aspects of Kimura diffusions

The infinite-population limits of standard population-genetic models are diff usion processes that take place on a simplex, is a higher dimensional generalization of a triangle. A point of the simplex specifies the frequencies of the different genotypes. The coefficients of PDEs describing these limiting models contain information about the relative effects of genetic drift, fitness, mutation rates and migration. Because the paths of the stochastic process, which describe the time evolution of the frequencies of genotypes, are constrained to remain in a simplex, the partial differential operators that generate these diffusion processes degenerate at the boundary of the simplex. This makes the mathematical analysis and numerical simulation of these processes quite difficult. We have completed the foundational work on the mathematical analysis of these diffusion equations, and established the needed connections with stochastic differential equations and Markov processes. This produces explicit estimates for the transition kernel for the Markov process, and things like the stationary distribution and exit times. This work is described in a monograph published by Princeton University Press, and a series of publications and preprints.

COLLABORATIONS AND LEVERAGED FUNDING

Several publications from this ARO project directly impact the interpretation of evolution experiments performed by other scientists, including Chris Marx (Harvard), Rich Lenski (Michigan State), and Tim Cooper (U. Houston). As a result of these papers, we have begun to collaborate and draft a new research proposal that involves tight co-ordination of theory and experiment, with Marx and Lenksi, to understand and the outcomes of bacterial evolution.

TECHNOLOGY TRANSFER

None, yet. We are producing statistical techniques and algorithms for interpreting empirical data, which will likely be in the public domain.

PAPERS ACKNOWLEDGING SUPPORT FROM ARO GRANT W911NF-12-1-0552

1. Carja O, Plotkin JB. The evolutionary advantage of heritable phenotypic heterogeneity. *bioRxiv* doi:10.1101/028795 (pre-print)
2. Stewart A, Plotkin JB. Small games and long memories promote cooperation. *arXiv* 1407.1022 (pre-print)

3. Weinberg DE, **Shah P**, Eichhorn SW, Hussmann JA, **Plotkin JB**, Bartel DP. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Reports* 14:1-13 (2016)

4. **Gulisija D**, Kim Y, **Plotkin JB**. Phenotypic plasticity promotes balanced polymorphism in periodic environments by a genomic storage effect. *Genetics* (in press)

5. McCandlish M, Otwinowski J, Plotkin JB. Detecting epistasis from an ensemble of adapting populations. *Evolution* 69: 2359-2380 (2015)

6. Neverov AD, **Kryazhimskiy S**, **Plotkin JB**, Bazykin GA. Coordinated evolution of Influenza A surface proteins. *PLoS Genetics* 11: 1005404 (2015)

7. **Stewart A**, **Plotkin JB**. The evolvability of cooperation under local and non-local mutations. *Games* 6:231-250 (2015)

8. **McCandlish D**, Epstein C, **Plotkin JB.** Formal properties of the probability of fixation: identities, inequalities and approximations. *Theoretical Population Biology* 99:98-113 (2015)

9. **Shah P**, **McCandlish M**, **Plotkin JB**[*]. Historical contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences USA* 112:3226–3235 (2015)

10. Schulte MB, **Draghi JA**, **Plotkin JB**, Andino R. Experimentally guided models reveal replication principles that shape the mutation distribution of RNA viruses. *eLife* 4:3753 (2015)

11. **Stewart A, Plotkin JB**[*]. The collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences* 111: 17558-17563 (2014)

12. **Otwinowski J**, **Plotkin JB**[*]. Inferring fitness landscapes by regression produces biased estimates of epistasis. *Proceedings of the National Academy of Sciences USA* 111:2301-2309 (2014)

13. **Der R**, **Plotkin JB**. The equilibrium allele frequency distribution for a population with reproductive skew. *Genetics* 196: 1199-1216 (2014)

14. **McCandish D**, Epstein C, **Plotkin JB**. The inevitability of unconditionally deleterious substitutions during adaptation. *Evolution* 68:1351-1365 (2014)

15. **Stewart A**, **Plotkin JB**[*]. From extortion to generosity, evolution in the Iterated Prisoner's Dilemma. *Proceedings of the National Academy of Sciences USA* 110: 15348-15353 (2013)

16. **Shah P**, **Ding Y**, Niemczyk M, **Kudla G**, **Plotkin JB**[*]. Rate-limiting steps in yeast protein translation. *Cell* 153: 1589-1601 (2013)

17. **McCandlish D**, **Rajon E**, **Shah P, Ding Y**, **Plotkin JB**[*]. The role of epistasis in protein evolution. *Nature* 497: E1-E2 (2013)

18. **Feder A**, **Kryazhimskiy S**, **Plotkin JB**[*]. Identifying signatures of selection in genetic time series. *Genetics* 196: 509-522 (2013)

19. **Draghi J**, **Plotkin JB**. Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* 67: 3120–3131 (2013)

20. Li Y, **Bostick D**, Sullivan C, Myers J, Griesemer S, St. George K, **Plotkin JB**[*], Hensley S[*]. Single Hemagglutinin mutations that alter both antigenicity and receptor-binding avidity. *Journal of Virology* 87: 9904-9910 (2013)

21. **Stewart A**, **Plotkin JB**[*]. The evolution of complex gene regulation by low-specificity binding sites. *Proceedings of The Royal Society B* 280: 20131313 (2013)

22. **Rajon E, Plotkin JB**. The evolution of genetic architectures underlying quantitative traits. *Proceedings of The Royal Society B* 280: 20131552 (2013)

1.   **C.L. Epstein** and Camelia Pop, Regularity for the Supercritical Fractional Laplacian with Drift, *Jour of Geo. Anal.* (2015), DOI :10.1007/s12220-015- 9590-x.
2.   **C.L. Epstein** and R. Mazzeo, Harnack Inequalities and Heat-kernel Estimates for Degenerate Diffusion Operators Arising in Population Biology, arXiv:1406.1426, to appear in: *Applied Math. Research Express*, 2016, 57pp.
3.   **Charles L. Epstein**, and Camelia A. Pop, Harnack Inequalities for Degenerate Diffusions, submitted, arXiv:1406.4759, under revision for: The Annals of Probability 35pp.
4.   **C.L. Epstein** and Jon Wilkening Eigenfunctions and the Dirichlet problem for the Classical Kimura Diffusion Operator, 2015, http://arxiv.org/abs/1508. 01482v1