DISS. ETH NO. 18592

EXPLORING THE EVOLUTIONARY CAUSES AND CONSEQUENCES OF STOCHASTIC VARIATION IN GENE EXPRESSION IN BACTERIA

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

Presented by

NICOLE ELIZABETH FREED

MSc., The University of California San Diego

Date of birth: 18th of September 1978

American citizen

accepted on the recommendation of

Prof. Dr. Martin Ackermann, examiner

Prof. Dr. Urs Jenal, co-examiner

2009

To my family.

To Martin
The most talented and amazing boss

To Olin
For making every day better

Table of Contents

TITLE PAGE 1
DEDICATION 3
TABLE OF CONTENTS 5
SUMMARY 7
ZUSAMMENFASSUNG 9
CHAPTER 1: GENERAL INTRODUCTION 11
CHAPTER 2: A SIMPLE SCREEN TO IDENTIFY PROMOTERS CONFERRING HIGH LEVELS OF PHENOTYPIC NOISE 18
CHAPTER 3: SELF DESTRUCTIVE COOPERATION MEDIATED BY PHENOTYPIC NOISE 40
CHAPTER 4: EXPERIMENTAL EVOLUTION OF INCREASED NOISE IN BACTERIAL PROMOTERS 55
CHAPTER 5: EFFECT OF *FLGM* EXPRESSION ON VARIATION IN *FLIC* EXPRESSION IN SALMONELLA 75
CHAPTER 6: DISCUSSION 84
APPENDIX I: 91
APPENDIX II: 97
APPENDIX III: 107
ACKNOWLEDGEMENTS: 118
CURRICULUM VITAE 121

Summary

According to the conventional view, the characteristics of an organism are determined by nature and nurture - by its genes and by the environment it lives in. Consequently, one would expect that two organisms that share the same genes and live in the same environment have identical characteristics. Recently it has become clear that this expectation is often not borne out: clonal families of simple organisms living under constant conditions often show variation in biological traits, and sometimes even have markedly different properties and do different things.

In order to investigate such phenotypic variation and its possible biological relevance, I carried out laboratory experiments with the commensal bacteria *Escherichia coli* (*E. coli*) and bacterial pathogen *Salmonella enterica* ssp. I serovar Typhimurium (*S.* Typhimurium). The strain of *E. coli* used in these experiments is a harmless bacterium commonly found in the gut of most mammals. *S.* Typhimurium is an enteropathogen that can cause disease ranging from self-limiting enterocolitis to life-threatening sepsis. Both organisms are standard laboratory model organisms that contain characteristics important for this study: they are well characterized, fully sequenced, have a short generation time, and can be modified genetically to express fluorescent proteins.

In the first part of this thesis I describe a simple and fast method we have developed for identifying biological traits that have relatively high levels of phenotypic variation. Using this method, we screened for noisy traits in the bacterial pathogen *S.* Typhimurium that vary at a timescale of one day. We found that traits involved in interaction with the host are particularly noisy, suggesting that phenotypic noise might be important in pathogenesis. This method can be readily adopted for other organisms, and might contribute to elucidating the role of phenotypic variation in biology. We also begin to explore the genetic regulation that might amplify noise in particular genes.

In the second part of the thesis, we build on the result that phenotypic variation might play a role in salmonella pathogenesis. Using mathematical models coupled with experimental data, we investigate whether evolution of noisy genes is plausible, given that the expression of such genes might lead to the destruction of the bacteria that express them. We conclude that under some conditions, noisy genes that lead to “self-destructive” behavior could evolve. These results provide a new perspective on the significance of phenotypic noise in bacterial pathogenesis: it might promote the formation of cooperative sub-populations that die while

preparing the ground for a successful infection.

In the final part of the thesis we explore whether phenotypic variation it is a trait that selection can act on. We explore whether higher levels of variation in gene expression can be evolved experimentally in promoters from *E. coli* and *S.* Typhimurium. Our results indicate that there is limited evolution of increased phenotypic variation after seventeen selective events in the five promoters used in this study. This indicates that for the genes we investigated, the variability in protein concentration is robust to the alternating selection we imposed.

Zusammenfassung

In der herkömmlichen Betrachtungsweise sind die Eigenschaften von Organismen komplett von ihren Genen und ihrer Umgebung bestimmt. Daher erwartet man, dass zwei Organismen mit identischen Genen, die in der gleichen Umgebung leben, auch die gleichen Eigenschaften haben. Neuere Untersuchungen zeigen jedoch, dass dies nicht notwendigerweise stimmt: Klon-Familien einfacher Organismen, die unter konstanten Bedingungen leben, zeigen oft phänotypische Unterschiede. Es kommt manchmal sogar zu deutlich unterschiedlichen Merkmalen und Verhaltensweisen.

Um diese phänotypische Variation zu untersuchen und mögliche biologische Relevanz zu erkennen, habe ich Labor-Experimente mit dem symbiotischen Bakterium *Escherichia coli* (*E. coli*) und dem Bakterien-Pathogen *Salmonella enterica* ssp. I serovar Typhimurium (*S.* Typhimurium) durchgeführt. Der verwendete *E. coli*-Stamm ist ein harmloses Bakterium, das im Darm von den meisten Säugetieren vorkommt. *S.* Typhimurium ist ein Enteropathogen, das Krankheiten von self-limiting Enterocolitis bis zu lebensbedrohlichen Blutvergiftungen auslösen kann. Beide Organismen sind im Labor verbreitete Modell-Organismen und haben für meine Untersuchungen wichtige Eigenschaften: sie sind sehr gut verstanden, vollständig sequenziert, und haben eine kurze Generationszeit. Des Weiteren können sie durch genetische Modifikation fluoreszierende Proteine exprimieren.

Im ersten Teil der Arbeit beschreibe ich eine von uns entwickelte einfache und schnelle Methode um biologische Merkmale zu identifizieren, die ein relativ hohes Level an phänotypischer Variation zeigen. Wir benutzen diese Methode, um nach Merkmalen mit einem Rauschen im Bakterien-Pathogen *S.* Typhimurium zu filtern (auf einer ein-tägigen Zeitskala). Wir zeigen, dass Merkmale, die in Interaktion mit dem Wirt evolvieren, besonders stark rauschen. Dies legt die Vermutung nahe, dass phänotypisches Rauschen ein wichtiger Faktor in der Krankheitsentstehung ist. Unsere Methode kann direkt auf andere Organismen angewendet werden, und könnte dazu beitragen, die Rolle von phänotypischer Variation in der Biologie zu verstehen. Wir haben zudem angefangen, Gen-Regulationen zu untersuchen, welche Rauschen in bestimmten Genen verstärken könnte.

Im zweiten Teil der Arbeit untersuchen wir, ob und wie phänotypische Variation eine Rolle in der Salmonellen-Pathogenese spielt. Mit Hilfe von mathematischen Modellen und experimentellen Daten untersuchen wir, ob Gene mit hoher Expressionsvarianz evolviert werden können, gegeben dass die Expression solcher Gene zur Zerstörung eben dieser

Bakterien führen könnte. Wir zeigen, dass unter geeigneten Bedingungen Gene mit hoher Expressionsvarianz, die zu selbst-zerstörerischem Verhalten führen, tatsächlich entstehen können. Diese Resultate geben eine neue Sichtweise auf die Bedeutung von phänotypischem Rauschen in der bakteriellen Krankheitsentstehung: phänotypisches Rauschen könnte die Bildung von kooperierenden Sub-Populationen fördern, die zwar sterben, aber zugleich die Basis einer erfolgreichen Infektion legen.

Im letzten Teil der Arbeit untersuchen wir, ob phänotypische Variation ein Merkmal ist, auf dem Selektion agiert. Wir untersuchen für verschiedene Promoter in *E. coli* and *S.* Typhimurium ob ein höheres Level an Variation in Gen-Expression evolviert werden kann. Unsere Ergebnisse zeigen, dass nach siebzehn Selektionsschritten in den fünf Promotern, die wir benutzt haben, nur wenig Evolution hin zu erhöhter phänotypischer Variation stattgefunden hat. Das bedeutet, dass in den untersuchten Genen die Variabilität in Proteinkonzentration robust ist gegenüber der durchgeführten wechselnden Selektion.

Chapter 1: General Introduction

Isogenic populations of bacteria existing in identical environmental conditions can show marked phenotypic heterogeneity. Much of this variation in phenotype has been shown to arise from noise, or random fluctuations, in gene expression within each individual cell. Recent experimental and theoretical investigations have illuminated some of the mechanisms behind such noise in gene expression. However, less is known about the consequences of stochastic variation in gene expression for bacterial fitness and its evolutionary origins.

As early as 1976, it had been observed that bacterial cells show individual differences in chemotactic migration within a clonal population[1]. More recently, new methods and technologies have helped to further our understanding of the mechanisms that drive stochastic phenotypic variation[2,3,4,5,6,7,8]. Namely, small concentrations of molecules involved in molecular processes such as transcription and translation within a single cell can lead to fluctuations in the absolute number of proteins produced[9]. These fluctuations can be amplified even further in gene regulatory networks that feature positive feedback circuitry[10]. It has been shown experimentally that mutations to DNA affecting the rates of transcription and translation can result in changes in the level of phenotypic variation. Specifically, mutations that increase the rate of translation, causing random bursts in protein production, have been shown to increase phenotypic heterogeneity[5]. As well, reduced rates of transcription have been shown to increase noise in gene expression[2]. These studies help to establish that the level of phenotypic variation is subject to mutational change in DNA sequence and thus may respond to natural selection. One of the goals of this thesis is to explore the evolutionary pressures that might affect phenotypic noise and test if it can evolve in response to selection.

To begin to address the question of whether noise in gene expression is adaptive, it is necessary to obtain accurate measures of gene expression. Several techniques are available to detect differences in gene expression patterns in isogenic populations. Single cell analysis such as flow cytometry and fluorescent microscopy are the principal techniques used by others[11]. Flow cytometry provides a fast method to obtain several important pieces of data from tens of thousands of individual bacterial cells in just a few seconds. It records accurate measures of cell size, shape, complexity, and relative concentration of fluorescent proteins in single cells and can be used to distinguish heterogeneity within a clonal population[12]. In addition, bacterial cells can be sorted based on the expression of fluorescent reporters, and the subpopulations can be re-grown[13]. Similarly, time-lapse fluorescent microscopy can be

used for visualization of gene expression on the single cell level and provides an additional level of information concerning relative rates of switching in gene expression between on and off states[14]. A drawback to time-lapse microscopy is the relatively low number of cells that can be visualized in a given amount of time (on the order of hundreds of cells over several hours). In this thesis, we use both flow cytometry and time-lapse microscopy to identify noise in gene expression in clonal populations of bacterial cells.

As a first step in identifying whether phenotypic variation is an adaptive process or is just to costly for a cell to control, we set out to ask what biological processes in *S.* Typhimurium are noisy. One method to achieve this is to individually tag the promoter of each gene in a genome, record its expression, and calculate noise by dividing mean expression by standard deviation. A recent study by Newman *et al* did just this, systematically tagging each promoter in *Saccharomyces cerevisiae* with a green fluorescent protein (GFP)[3,15], and recording expression using flow cytometry. While this method is comprehensive, it is time consuming and beyond the scope of our laboratory capabilities. In an effort to simplify this process, we developed a method to enrich for noisy genes using a cell sorter. By pooling a plasmid library containing genomic fragments of *S.* Typhimurium driving the expression of GFP, and performing seven rounds of alternating selection (described further in Chapter 2), we were able to enrich for genome fragments containing noisy promoters. The majority of genes we identified as having increased levels of phenotypic variation were those that are involved in flagellar synthesis. Flagellar genes have been shown to be instrumental in infection of host tissue[16], and this suggests that phenotypic noise might play a role in pathogenesis. It should be noted here that we focus on phenotypic variation that arises from fluctuations in gene expression and protein concentrations and not on those that arise from genetic rearrangement[17]. It is important to clarify the difference between the two processes and note that *S.* Typhimurium is often used as a classic example of the latter. *S.* Typhimurium are known to undergo a site specific chromosomal inversion at low rates[18], whereby the bacteria can effectively switch between one of two antigenically different types of flagellin[19]. We will explicitly discuss how we control for effects of genetically-mediated flagellar switching in more detail in Chapter 2, as our results specifically look at noise in the expression of flagellar genes. It should be mentioned that is difficult to experimentally rule out genetic rearrangements at other locations and to easily distinguish between genetic rearrangements and intrinsic variation in gene expression in general.

It is important to then ask what the possible benefits of heterogeneity could be. From an evolutionary perspective, could stochasticity in gene expression have an adaptive role? Recent results, both theoretical and experimental, suggested that harnessing the stochastic variation that arises in the expression of a single gene or set of genes might provide a more optimal means of regulating certain processes than by purely deterministic means[20,21,22,23,24].

In general, two theories describe an adaptive explanation for the evolution of increased noise or bistabilty in some genes[20,25,26,27]. The first is concerned with the long-term benefits of having a small proportion of a population expressing an alternate phenotype[5]. For an organism living in a variable environment, where sudden environmental changes might not be preceded by a signal, it may be advantageous to produce diverse offspring in which one might be able to survive in the new environment. Such bet-hedging strategies have been shown to exist in bacteria and yeast [28,29,30]. One of the best known examples of bet hedging is bacterial persistence[26], where a subpopulation enters a metabolically dormant (non-dividing) state where it is less susceptible to environmental toxins and antibiotics. This confers an obvious growth disadvantage because cells in this state are not reproducing. However, because the transition between regular growth and persistence is a reversible random switch governed by a stochastic change, once the toxin is removed and the cell randomly switches back to regular growth, it is able to persist through the lethal environment[31]. Another example that is often cited is work by Bishop *et al* who show that heterogeneous populations of yeast cells can survive a stressful environment better than similar cells that exhibit more homogeneity[28].

Alternatively, there is a second adaptive explanation for the benefit of increased levels of stochastic phenotypic variation that has received less attention. Self-destructive cooperation, a special sub-type of division of labor, occurs when a subset of the population undergoes cell death for the benefit of the rest of the population, can evolutionarily be better understood when this process is mediated by stochastic phenotypic noise. For example, some individuals in a population of the bacterial pathogen *Streptococcus pneumoniae* produce pneumolysin, a toxin that is only released by cell lysis, thus killing the producer. This action however enhances colonization of the lung by other members in the population, providing an overall advantage[32]. We hypothesize in Chapter 3 that this process of stochastic mediated phenotypic variation may be acting in *S.* Typhimurium, where cells invading the host tissue are killed by the host immune system, yet at the same time induce an inflammation which

assists the other members who did not invade. Using a mathematical model, we show that the use of a stochastically mediated process to facilitate cooperative behavior where the costs to benefactor are large could evolve when there is spatial structure. In addition, we provide experimental evidence that hold with key assumptions of the model.

In addition to theoretical models explaining how stochastic variation may have originated, we set out to experimentally demonstrate that higher levels of phenotypic variation in single promoters can evolve using experimental evolution with a cell sorter and to identify the molecular mechanisms underlying any changes. Today, the use of laboratory evolution has expanded tremendously. The use of a fluorescent activated cell sorting as a selection technique for experimental evolution is relatively new, and has been used mostly for applied techniques[33,34,35]. The use of a FACS as method for selection has several advantages, the major one being that a single trait (GFP expression) can be selected on in a quantitative fashion. We developed a method of alternating selection which is carried out by isolating the five percent cells with the *brightest* GFP expression, subsequent growth, and then sorting of five percent cells with the *dimmest* expression. In this way, we set out to evolve specific promoters of *S.* Typhimurium and *E. coli* (see Chapter 4) to increase levels of phenotypic variation. For the genes we investigated, we did not observe any changes to noise in gene expression after the laboratory based evolution. This indicates that the noise in gene expression is robust to the alternating selection we imposed.

Finally, in Chapter 5, we briefly investigate the role of the anti-sigma factor FlgM and σ^{28} on noise in the expression of the flagellar filament gene *fliC* in *S.* Typhimurium. Collaborating with Kelly Hughes at the University of Utah, we hypothesized that the noise we observe in *fliC* expression might actually lie upstream and be propogated through to *fliC*. Previous results indicate a delicate balance between the upstream regulators of *fliC,* FlgM and σ^{28}. We set out to construct mutants with varied intracellular concentrations of FlgM and σ^{28} and observe noise in fliC expression. We found that increasing σ^{28} levels within the cell eliminated noise in *fliC* expression.

References

1. Spudich J, Koshland D (1976) Non-genetic individuality: chance in the single cell.

2. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297: 1183-1186.

3. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, et al. (2006) Noise in protein expression scales with natural protein abundance. Nature genetics 38: 636-643.

4. Austin D, Allen M, McCollum J, Dar R, Wilgus J, et al. (2006) Gene network shaping of inherent noise spectra. Nature 439: 608-611.

5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genetics 31: 69-73.

6. Cai L, Friedman N, Xie X (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440: 358-362.

7. Yu J, Xiao J, Ren X, Lao K, Xie X (2006) Probing gene expression in live cells, one protein molecule at a time. American Association for the Advancement of Science. pp. 1600-1603.

8. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences of the United States of America 99: 12795-12800.

9. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. Nature Reviews Genetics 6: 451-464.

10. Smits W, Kuipers O, Veening J (2006) Phenotypic variation in bacteria: the role of feedback regulation. Nature Reviews Microbiology 4: 259-271.

11. Brehm-Stecher B, Johnson E (2004) Single-cell microbiology: tools, technologies, and applications. Microbiology and molecular biology reviews 68: 538-559.

12. Davey H, Kell D (1996) Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiology and molecular biology reviews 60: 641.

13. Cormack B, Valdivia R, Falkow S (1998) FACS-optimized mutants of the green fluorescent protein (GFP). Google Patents.

14. Kaufmann B, Yang Q, Mettetal J, van Oudenaarden A (2007) Heritable stochastic switching revealed by single-cell genealogy. Plos Biology 5: e239.

15. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. Nature 441: 840-846.

16. Stecher B, Hapfelmeier S, Muller C, Kremer M, Stallmach T, et al. (2004) Flagella and Chemotaxis Are Required for Efficient Induction of Salmonella enterica Serovar Typhimurium Colitis in Streptomycin-Pretreated Mice Editor: JB Bliska. Infection and Immunity 72: 4138-4150.

17. Dybvig K (1993) DNA rearrangements and phenotypic switching in prokaryotes. Molecular Microbiology 10: 465-471.

18. Gillen K, Hughes K (1991) Negative regulatory loci coupling flagellin synthesis to flagellar assembly in Salmonella typhimurium. Journal of bacteriology 173: 2301-2310.

19. Moxon ER, Rainey PB, Nowak MA, Lenski RE (1994) Adaptive Evolution of Highly Mutable Loci in Pathogenic Bacteria. Current Biology 4: 24-33.

20. Ackermann M, Stecher B, Freed NE, Songhet P, Hardt WD, et al. (2008) Self-destructive cooperation mediated by phenotypic noise. Nature 454: 987-990.

21. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. Science 317: 526-529.

22. Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440: 545-550.

23. Dubnau D, Losick R (2006) Bistability in bacteria. Molecular Microbiology 61: 564-572.

24. Weinberger L, Burnett J, Toettcher J, Arkin A, Schaffer D (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122: 169-182.

25. Bull JJ (1987) Evolution of Phenotypic Variance. Evolution 41: 303-315.

26. Kussell E, Kishony R, Balaban NQ, Leibler S (2005) Bacterial persistence: A model of survival in changing environments. Genetics 169: 1807-1814.

27. Wolf DM, Vazirani VV, Arkin AP (2005) A microbial modified prisoner's dilemma game: how frequency-dependent selection can lead to random phase variation. Journal of Theoretical Biology 234: 255-262.

28. Bishop AL, Rab FA, Sumner ER, Avery SV (2007) Phenotypic Heterogeneity can Enhance Rare-Cell Survival in 'Stress Sensitive'Yeast Populations. Molecular Microbiology 63: 507-520.

29. Veening J, Smits W, Kuipers O (2008) Bistability, epigenetics, and bet-hedging in bacteria.

30. Booth IR (2002) Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. International Journal of Food Microbiology 78: 19-30.

31. Lewis K (2006) Persister cells, dormancy and infectious disease. Nature Reviews Microbiology 5: 48-56.

32. Paton JC (1996) The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae. Trends in Microbiology 4: 103-106.

33. Becker S, Schmoldt H, Adams T, Wilhelm S, Kolmar H (2004) Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. Current opinion in biotechnology 15: 323-329.

34. Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik D (2005) High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chemistry & Biology 12: 1281-1289.

35. Wang L, Jackson W, Steinbach P, Tsien R (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. Proceedings of the National Academy of Sciences 101: 16745-16749.

Chapter 2

A simple screen to identify promoters conferring high levels of phenotypic noise†

Nikki E. Freed[1*], Olin K. Silander[1*], Bärbel Stecher[2], Alex Böhm[3], Wolf-Dietrich Hardt[2], Martin Ackermann[1,4,*]

[1]Institute of Integrative Biology, ETH Zurich, Zurich Switzerland; [2]Institute of Microbiology, ETH Zurich, Zurich Switzerland; [3]Biozentrum, University of Basel, Basel, Switzerland. *Current address: Department of Environmental Microbiology, EAWAG, and Institute of Biogeochemistry and Pollutant Dynamics, Eidgenössische Technische Hochschule (ETH) Zurich, Switzerland, Corresponding author[4]

Abbreviations:

CV: coefficient of variation

GFP: green fluorescent protein

FACS: fluorescent-activated cell sorting

FSC: forward scatter

SSC: side scatter

S. Typhimurium: *Salmonella enterica* ssp. I serovar Typhimurium

†A modified version of this chapter appears as Freed NE, Silander OK, Stecher B, Böhm A, Hardt W-D, et al. 2008 A Simple Screen to Identify Promoters Conferring High Levels of Phenotypic Noise. PLoS Genet 4(12): e1000307. doi:10.1371/ journal.pgen.1000307

Abstract

Genetically identical populations of unicellular organisms often show marked variation in some phenotypic traits. To investigate the molecular causes and possible biological functions of this phenotypic noise, it would be useful to have a method to identify genes whose expression varies stochastically on a certain time-scale. Here, we developed such a method, and used it for identifying genes with high levels of phenotypic noise in *Salmonella enterica* ssp. I serovar Typhimurium (*S.* Typhimurium). We created a genomic plasmid library fused to a green fluorescent protein (GFP) reporter and subjected replicate populations harboring this library to fluctuating selection for GFP expression using fluorescent-activated cell sorting (FACS). After seven rounds of fluctuating selection, the populations were strongly enriched for promoters that showed a high amount of noise in gene expression. Our results indicate that the activity of some promoters of *S.* Typhimurium varies on such a short time scale that these promoters can absorb rapid fluctuations in the direction of selection, as imposed during our experiment. The genomic fragments that conferred the highest levels of phenotypic variation were promoters controlling the synthesis of flagella, which are associated with virulence and host-pathogen interactions. This confirms earlier reports that phenotypic noise may play a role in pathogenesis, and indicates that these promoters have among the highest levels of noise in the *S.* Typhimurium genome. This approach can be applied to many other bacterial and eukaryotic systems as a simple method for identifying genes with noisy expression.

Introduction

Clonal populations of unicellular organisms growing under constant conditions often show substantial variation in phenotypic traits. The rate at which some of these traits vary is so high that it cannot result from mutational change. Rather, this phenotypic noise has been shown to result from chance events in the cells, namely random fluctuation in the transcription and translation of genes [1,2,3]. Most of the research on phenotypic noise focuses on two questions. First, what are the molecular processes underlying this phenomenon? Second, are there cases in which phenotypic noise is beneficial? Can it provide a genotype with new biological functions and improve the chance that it will survive and reproduce?

To further our understanding of the biological significance of phenotypic noise, it would be helpful to have a simple method to identify genes whose expression varies stochastically at a

given timescale and under specific environmental conditions. So far, most of the research on phenotypic noise was based on the detailed analysis of individual biological traits [4,5,6]. It is interesting to complement these studies with a global analysis, so that one can ask whether the traits studied so far are indeed particularly noisy, or whether a substantial fraction of all genes show such high levels of noise. One possibility for a global analysis of phenotypic noise is the exhaustive characterization of ordered libraries of strains marked with reporter proteins [7]. Here, we have established a simple alternative that allows identifying promoters whose activity varies on a specific time-scale; we used this method to identify promoters in the bacterial pathogen *S.* Typhimurium that switch between active and inactive over the course of a few generations.

The method is based on subjecting a promoter library to selection for high levels of random variation on a short time scale. The screen was initiated with a genomic library consisting of short genomic fragments upstream of a gene encoding green fluorescent protein (GFP). Cells carrying a fragment with an active promoter expressed GFP. In order to select for promoters with a high level of phenotypic noise, we used fluorescence-activated cell sorting to select cells based on the cellular concentration of GFP, and alternated between selecting for high levels of GFP, and selecting for low levels of GFP. There was no signal indicating the direction of selection during a given round of the selection experiment; one would thus expect that promoters that randomly switch between expressing and not expressing GFP would increase in frequency.

This screen led to a strong enrichment of promoters with high levels of noise. The promoters that showed the highest levels of noise were found to be flagellar promoters, which are involved in the interaction with the host. These promoters have previously been reported to be heterogeneously expressed in clonal populations of *S*. Typhimurium. Our screen demonstrates that these promoters stand out in terms of the level of noise, and that they vary on a very short timescale. This method thus offers a simple and powerful approach to identify genes with high levels of noise, and allows for easily modulating timescale and environmental conditions under which such phenotypic noise manifests.

Results and Discussion

We established a population of approximately $7x10^6$ *S.* Typhimurium clones containing a library of genomic fragments ranging in size from 400 bp to 1200 bp linked to a GFP reporter

(see Methods). In order to enrich for clones exhibiting increased levels of phenotypic noise in GFP concentration, we used a regime of alternating selection. Cells were grown into exponential phase, and subjected to selection on GFP concentration in a fluorescence-activated cell sorter (FACS). First, we selected only those clones having a level of GFP expression in the highest 5% of the population; these clones were saved and used to inoculate fresh cultures that were grown overnight. In the next step, the opposite selection regime was imposed, such that only those clones having a level of GFP expression in the lowest 5% of the population were saved and grown. It is also possible to first select cells expressing low levels of GFP, and then high, which hypothetically would result in noisy promoters with lower average expression.

This process of fluctuating selection was repeated, until a total of seven alternating selection events had occurred. The fluctuating selection regime was performed on five separate populations; five control populations were also exposed to the same regime of growth and FACS sorting, but no selection occurred for the level of GFP concentration (a random subset of cells covering the entire range of GFP fluorescence was saved and grown). After the seven rounds of selection, clones from all populations were plated onto agar plates. Twenty-four clones from each of the ten populations were randomly selected for future analyses.

Phenotypic noise is a stable and consistent property of a clone

Selection for increased phenotypic noise can only be successful if the level of variation is a stable property of a clone. We thus first asked whether the level of phenotypic noise in GFP expression was a stable and consistent trait in these clonal isolates. We used the 240 frozen clonal stocks described above to seed fresh cultures of cells, and analyzed GFP concentration for about $5x10^5$ cells per clone (see Methods). We repeated the same procedure on a different day, and also gathered data on GFP expression for the same set of 240 clones. Phenotypic noise was quantified using the coefficient of variation in GFP expression from a subset of cells similar in size, shape, and cellular complexity (see Methods). We found that the level of phenotypic variation observed for a given clone on day 1 was highly correlated with the level observed on day 2 ($r^2 = 0.748$, $p<0.001$). This shows that the level of phenotypic noise is a consistent property of a clone (presumably reflecting the noise of the promoter on the genomic fragment it contains), and that this property is stably maintained in clonal populations that are repeatedly grown from an individual cell.

Fluctuating selection enriches for clones exhibiting increased phenotypic noise

Next, we asked whether fluctuating selection had led to an enrichment of clones exhibiting larger amounts of stochastic phenotypic variation. We compared the clones from the five selected populations to the clones from the five control populations. Among the clones from the selected populations, a sizable fraction showed high coefficients of variation in fluorescence, which is a measure of stochastic phenotypic variation. In contrast, the control population did not contain any clones with high coefficients of variation (Figure 1, Figure 2). An analysis of variance showed that the average coefficient of variation was higher in the selected than in the control populations (p-value = 0.016, by GLM univariate). This demonstrated that fluctuating selection on fluorescence enriched for strains with high levels of stochastic variation in this trait.

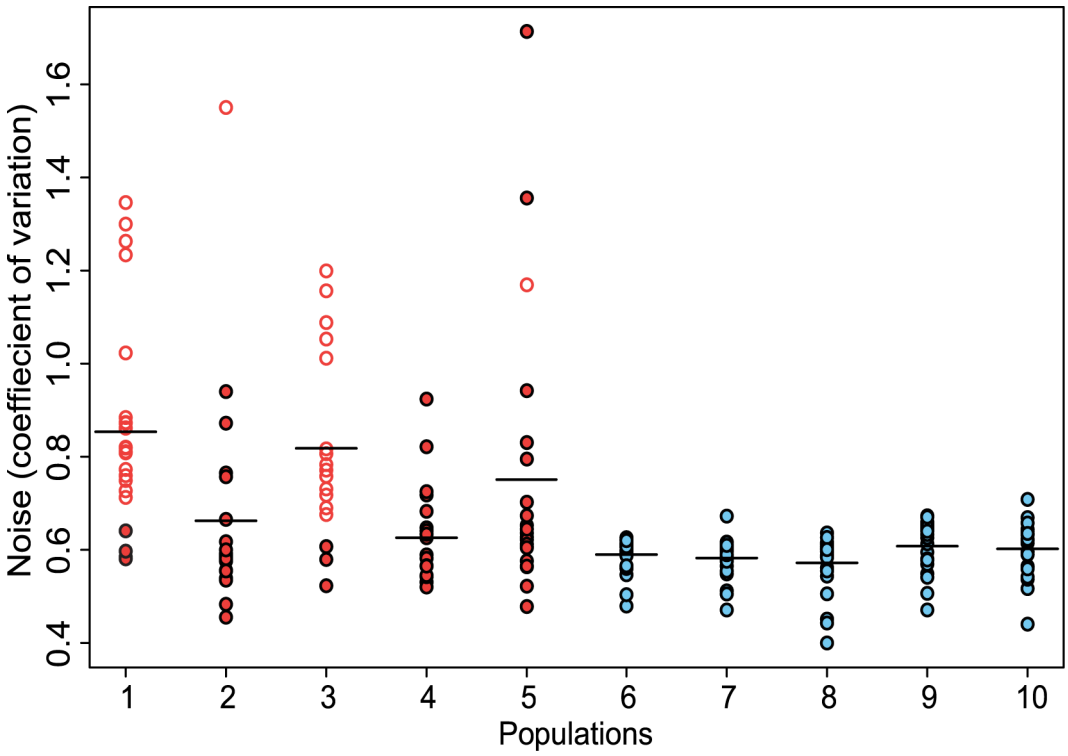


Figure 1: Noise in GFP expression in clones from selected and control populations.

Clones from selected populations (red) show a higher level of noise than do clones from control populations (blue) (univariate GLM, p = 0.016). Open circles indicate clones that contain the promoter sequence for the *fliC* gene driving GFP expression, which were significantly enriched in the selected

populations. Each data point represents the coefficient of variation of the GFP expression of several thousand individual cells.

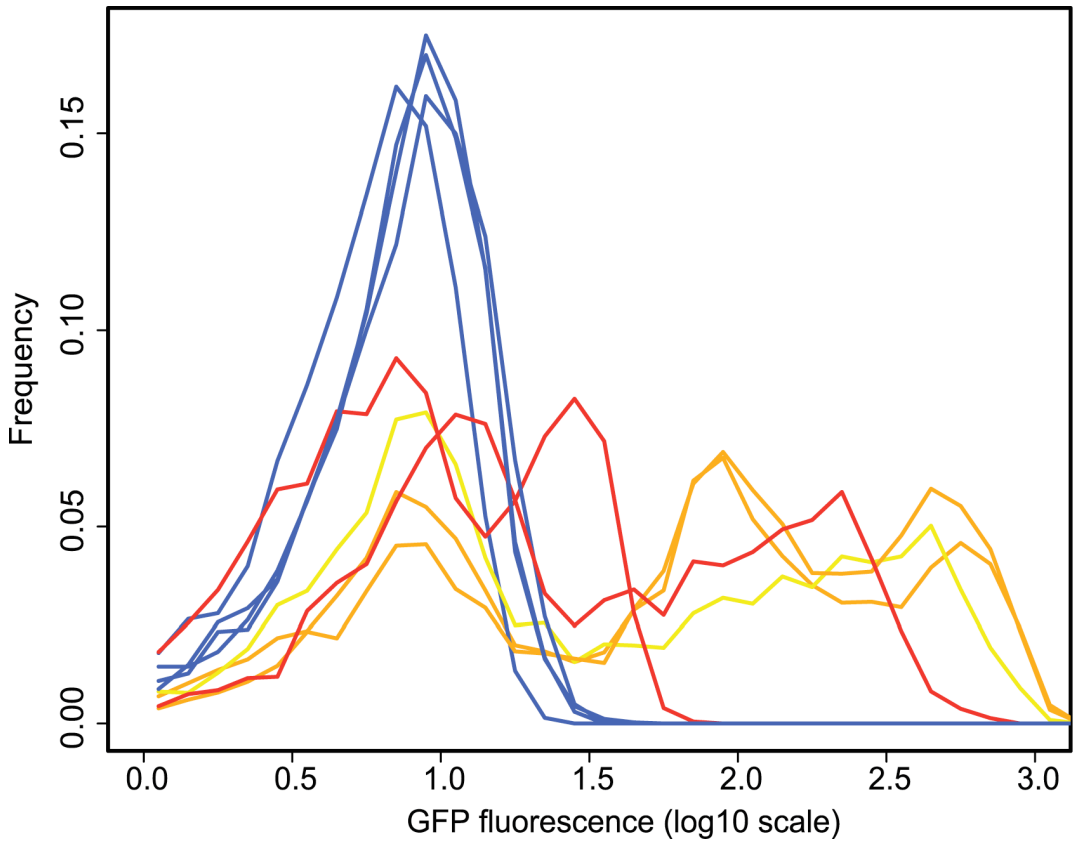


Figure 2: Histograms of GFP expression from clones exhibiting the highest level of noise in each population.

Clones from each of the ten populations were ranked according to the amount of noise in GFP expression produced. A histogram of GFP expression was plotted for a single clone from each population with the highest level of noise. Clones from selected populations (red, orange, and yellow lines) show a much higher level of noise than the control those from control populations (blue lines). Clones containing the *fliC* promoter are orange and a clone containing the *flgK* promoter is yellow.

Promoters of genes involved flagellar synthesis exhibit high levels of phenotypic noise

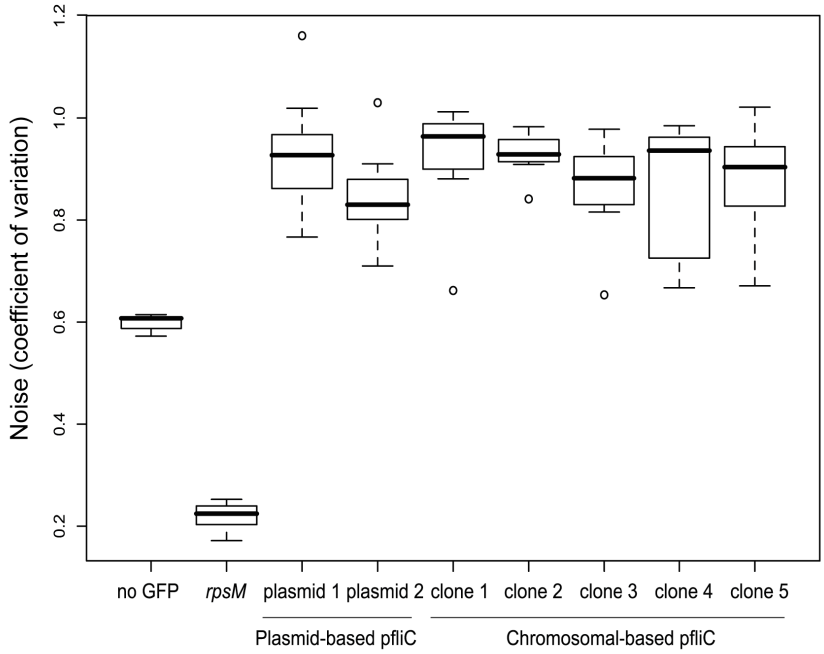
This simple selection scheme is thus a good tool for enriching for noisy promoters. Identifying the genes controlled by these promoters then gives a fairly unbiased look at genes whose expression is particularly variable, and might thereby provide new insights into the biological role of noise. In order to identify these genes, we sequenced the library inserts from the 240 frozen clonal stocks (24 from each experimental population). We found that the

clones exhibiting the highest levels of variation were dominated by two promoter sequences that regulate genes involved in flagellar synthesis, namely *fliC* and to a lesser extent *flgK* (Figure 1; Appendix II- Table 1). On the other hand, none of the inserts sequenced from the control populations contained promoters associated with the expression of flagellar or related genes, suggesting that this result was not due simply to overrepresentation of flagellar promoters in the genomic library.

We focused on *fliC* for two tests of the robustness of our results. First, we tested whether the *fliC* promoter is also noisy in the native chromosomal context. To do so, we constructed a transcriptional fusion of *gfp* to the *fliC* promoter at its native location in the chromosome. Clones from this chromosomal construction showed very similar levels of phenotypic noise to the plasmid-based *fliC* promoter (Figure 3).

Second, we asked whether GFP expression from the plasmid is correlated with actual protein production of *FliC*. Clones containing the *pfliC-GFP* insert in the plasmid pM968 with high levels of variation in GFP expression were sorted into three fractions (expression of GFP, no expression of GFP, and cells expressing all levels of GFP). Western blot analysis with *anti-FliC, anti-fljK* antibodies on these three cell fractions confirmed that GFP expression is positively correlated with FliC protein production (Figure 4). These two experiments indicate that the levels of noise we measured are, at least in the case of *fliC*, not an artifact of the plasmid-based reporter system, but do reflect actual differences in protein production between cells.

A



B

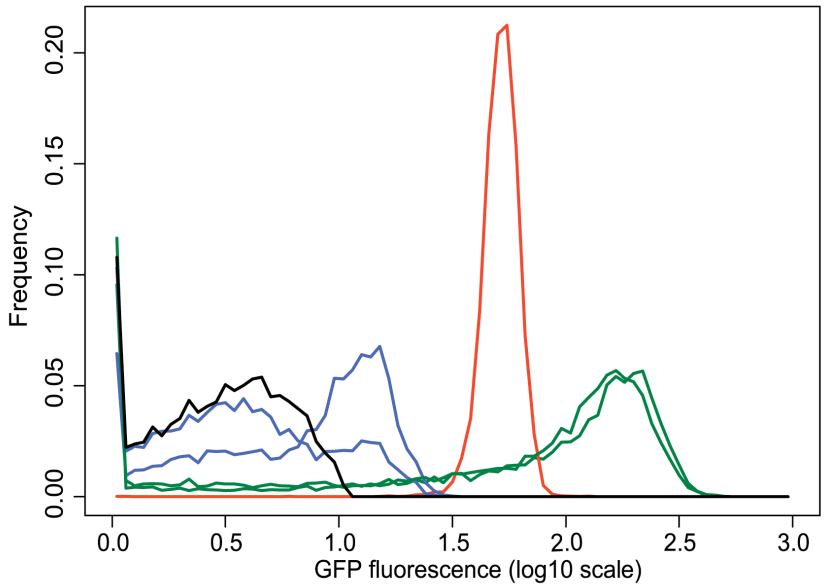


Figure 3: Comparison of noise in expression of chromosomal-based and plasmid-based *fliC* promoter.

A. Comparison of noise, as given by coefficient of variation in GFP expression, from the *fliC* promoter on the plasmid pM968 and in the native location on the chromosome of strain M557. Strain M557 (containing no *gfp* gene) and a *rpsM* promoter fused to *gfp*+[8] inserted in the chromosome of strain M557 serve as controls. There is no significant difference in noise between plasmid-based and chromosome-based expression of GFP under the control of the *fliC* promoter. B. Histograms of GFP expression from the *fliC* promoter on the plasmid pM968 (blue lines) and in the native location on the chromosome (green lines). These two strains differ in the average expression level and in the pattern of distribution of the expression levels in the

population. Strain M557 containing no *gfp* gene (black line) and a *rpsM* promoter fused to *gfp*+ (red line) inserted in the chromosome of strain M557 serve as controls.

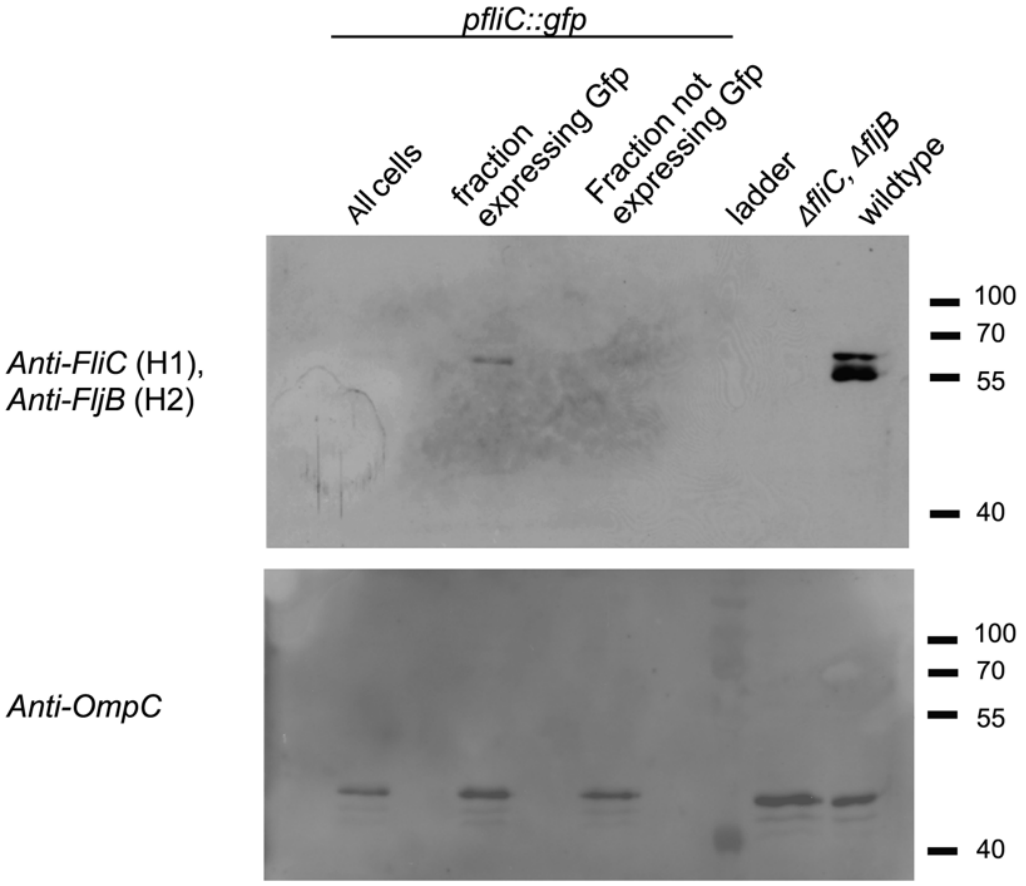


Figure 4: Western blot analysis shows that GFP expression correlates with the expression of *FliC*.

Cells containing the *pfliC::gfp* construct in plasmid pM968 were sorted based on expression of GFP using the FACS. Cells were sorted into three fractions, each containing the same number of cells: The first fraction contained cells with high levels of fluorescence; the second fraction contained cells whose fluorescence did not exceed background; the third fraction was a random sample of cells, chosen irrespective of their level of fluorescence. Cells were subjected to western blot analysis with staining using *anti-FliC, -FljB* antibodies and reprobed with *anti-OmpC* as a loading control. Only cells with high levels of GFP expression of GFP showed a band when stained with *anti-FliC,* indicating that GFP expression positively correlates with production of *FliC* protein. It is unclear why the fraction containing all cells does not also show a band; however, the lower intensity of the *anti-OmpC* band of this fraction and the fact that this fraction contains many cells that do not express gfp suggests that the *anti-FliC* band might be too faint to see.

High levels of phenotypic variation in the fliC promoter are not due to genetic phase variation

The variation in the expression of GFP under the control of flagellar promoters observed here is reminiscent of a genetic switch known as phase variation. *S.* Typhimurium express two distinct flagellin proteins, FliC and FljB [9], and switches between the two flagellar types using a site-specific recombination event in the chromosome. Can phase variation account for the phenotypic noise that we measured in the clones harboring the flagellar promoters? Site-specific recombination occurs at a rate of 10^{-3} to 10^{-5} per cell division [10,11,12]. In a clonal population grown from a cell in one phase, it thus takes many divisions until recombination-mediated phase variation has a reasonable likelihood of occurring. However, this is not what we observed in the clones with flagellar promoters: populations grown from single cells quickly attained substantial proportions of cells with both high and low expression of GFP (Figure 5, Appendix II- Movie 1). In contrast, clones isolated from the control populations maintained similar levels of GFP expression (Figure 6, Appendix II- Movie 2). This suggests that it is unlikely that the variation observed in these clones can be attributed to phase variation.

As a direct test of the effect of phase variation on stochastic phenotypic variation, we transformed the plasmid with the *fliC* promoter controlling GFP expression into a host strain that is incapable of phase variation [9] and into a wildtype strain. The resulting populations still showed strong variation in the amount of GFP between cells, and the coefficient of variation was not significantly different between the plasmid containing the *fliC* promoter in the wildtype background and the strain incapable of phase variation (t-test, $p = 0.199$, 95% Mean CV for wildtype background is 1.07, mean CV for *fljAB* promoter "locked" off background is 0.95, 95% Confidence interval for the difference is 0.068 and -0.299). This demonstrates that phase variation is not the main reason for the phenotypic noise observed here, and is most likely not involved.

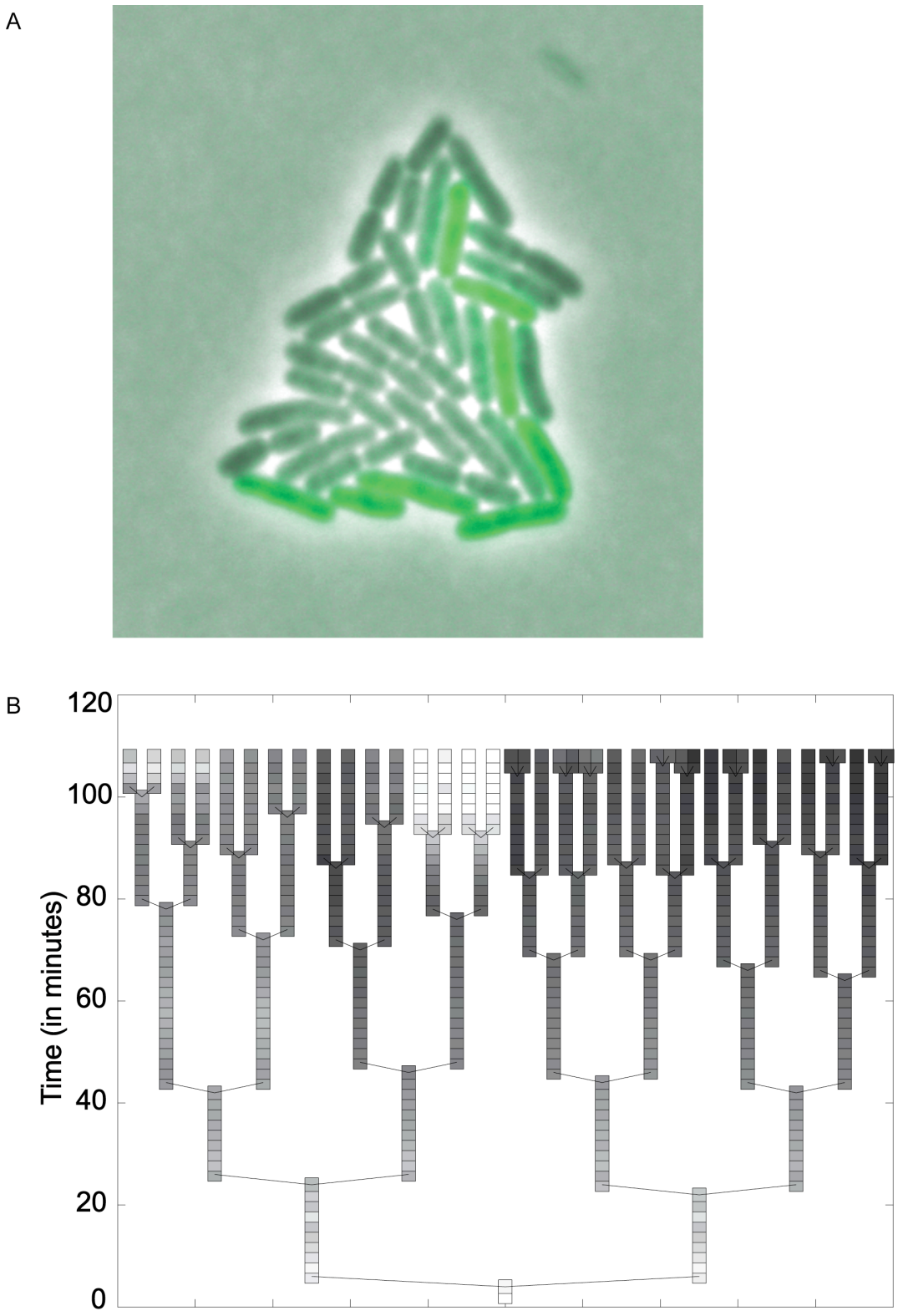


Figure 5: Phenotypic noise in a microcolony in *gfp* expression from the *fliC* promoter.

A. An image of a microcolony containing the plasmid-borne *fliC* promoter driving expression of GFP. The colony was started from a single cell and grown for about 6 generations. B. A lineage tree of this microcolony with

GFP expression plotted in green (light colored boxes represent high levels of GFP, and dark boxes represent low levels), illustrating the temporal pattern of switching of the *fliC* promoter. The image and the lineage tree are based on Appendix II- Movie 1.

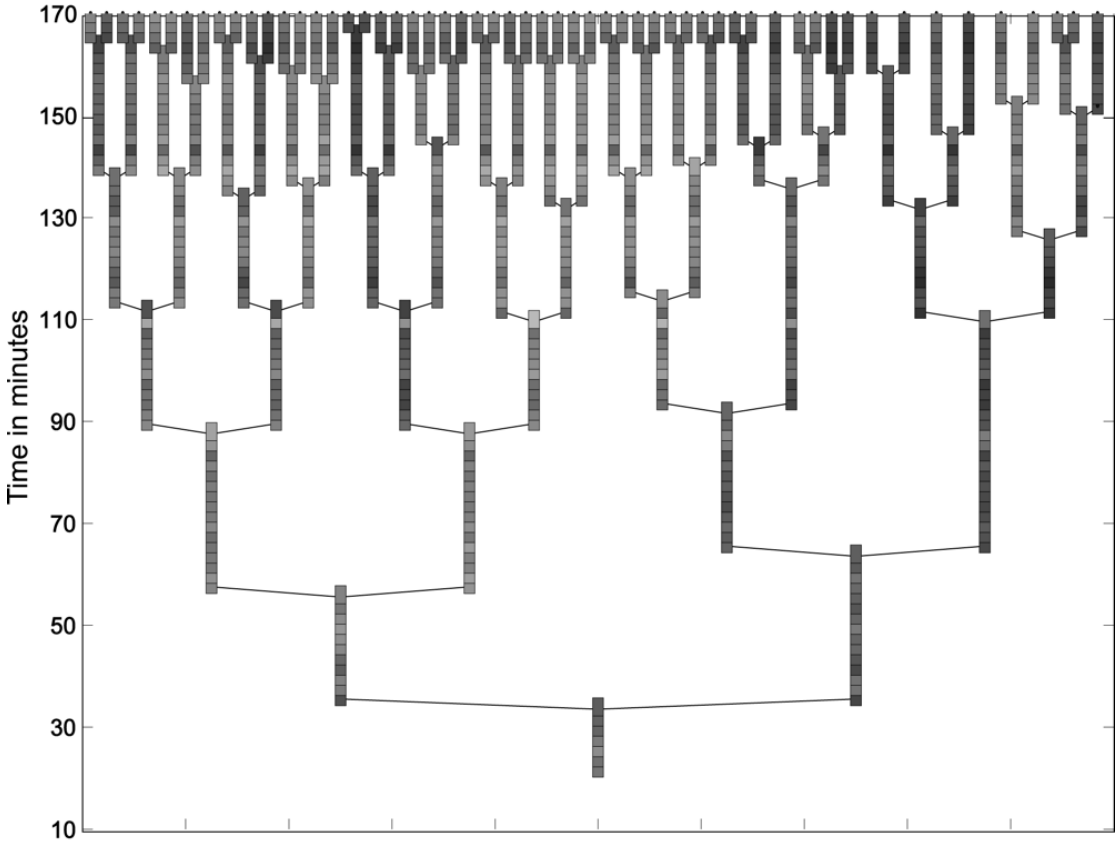


Figure 6: Lineage tree of microcolony growth and expression pattern of the *dcm* promoter.

GFP expression is plotted in grey (light colored boxes represent high levels of GFP, and dark boxes represent low levels), illustrating the temporal pattern of switching of the *dcm* promoter, isolated from a control population. The image and the lineage tree are based on Appendix II-Movie 2.

Possible biological roles of phenotypic noise in S. Typhimurium

Having identified promoters that are particularly variable, one can then ask whether variability in these promoters might serve a biological function. This question can be addressed by functional studies of the genes whose expression is particularly noisy. However, first insights can be gained from looking at the types of promoters that showed the highest levels of stochastic phenotypic variation.

By far the highest level of phenotypic noise observed in our experiment comes from flagellar promoters, most notably, *fliC*. This supports a previous report that the expression of *FliC* is

heterogeneous in clonal populations of *S.* Typhimurium [13]. Bacterial flagella are required for colonization and tissue invasion [14,15] and they interact with the host immune system in a myriad of ways, eliciting both innate and specific immune responses [16,17,18,19]. That variation in the expression of flagella might be advantageous is a well-established concept [20]; it usually refers to variation mediated by a site-specific recombination event, but has recently also been extended to variation that presumably does not involve changes in the DNA sequence [13,21].

The advantage that is usually postulated is mediation between conflicting selection pressures on flagellar expression in the host. During initial stages of gut infection by *S.* Typhimurium, flagella are instrumental for swimming towards the host's epithelial mucus layer [15]. During later stages of infection, a switch towards not expressing flagellin might be of advantage for bacteria that have invaded epithelial tissue, as it avoids recognition by the innate immune system [TLR5, Naip/Nalp][22]. There is a second possible biological function of phenotypic noise in flagella and other factors involved in the interaction with the host. A recent study suggested that heterogeneous expression of these traits in clonal populations of *S.* Typhimurium promotes the division of labor between two phenotypically different subpopulations. One subpopulation invades the gut tissue and elicits an inflammation of the gut; the other subpopulation remains in the gut and benefits from the fact that the inflammation reduces competition from commensal bacteria [23].

Two main insights emerge from this study. The first insight is that the activity of some *S.* Typhimurium promoters varies on such a short time scale that these promoters can absorb rapid fluctuations in the direction of selection, as imposed during our experiment. This is an important experimental test of one of the main ideas for why phenotypic noise can be adaptive: variation in the phenotypes encoded by a single genotype can increase the long-term growth rate of this genotype in fluctuating environments [24,25].

The second insight is methodological: fluctuating selection is a simple and fast tool to screen large pools of individuals in order to identify variable promoters in unicellular organisms, and thus complements exhaustive characterizations of individual genes [7]. Exhaustive characterizations require the construction of ordered libraries in which fluorescent markers are transcriptionally or translationally fused to every gene, as well as individual measurement of all resulting strains. In contrast, the method presented here only requires the relatively simple construction of a random genomic library, and sorting of the pooled library. It is thus also applicable to eukaryotic systems and organisms that are not genetic model systems, as

long as they can be stably transformed. It should thus be feasible to identify noisy promoters in a diverse range of environmental, commensal, and pathogenic organisms, and to ask whether differences in the lifestyle lead to consistent differences in the types of genes that are variable.

One particular advantage of this tool is that the time-scale at which the direction of selection changes can be varied. By changing the direction of selection every few cell divisions, on can impose selection for promoters that switch at a very high rate; changing the direction of selection less frequently selects for promoters that switch at lower rates. It should thus be possible to identify promoters that vary at different time scales, and to investigate whether they might be associated with responses to environmental conditions that vary at different frequencies.

Once noisy promoters are identified, functional studies are needed to investigate the biological consequences of their variation. This might lead toward new answers to one of the fundamental and most challenging questions about the biology of noise – whether phenotypic noise is beneficial, and what its possible benefits might be.

Materials and Methods

Growth of strains

Strains were grown at 37ºC on LB agar plates or in 1 ml of liquid LB broth in 5ml polystyrene round bottom tubes (BD Falcon), with shaking at 200 rpm until mid-exponential phase. Ampicillin (Sigma) was used at a concentration of 100 μg/ml in strains containing plasmid pM968 or its derivatives.

Construction of the plasmid library

A plasmid library (7 x 10^6 clones) was constructed by partially digesting *S.* Typhimurium SL1344 wildtype [26] chromosomal DNA with *Bsp*143I. Fragments within a size range of 400 bp to 1200 bp were ligated into *Bam*HI digested pM968. This plasmid is low copy number promoter-less derivative of pBAD24 containing promoterless *gfpmut2,* described in [27]. Plasmids were transformed into *E. coli* X6060, re-isolated by standard methods and electrotransformed into *S.* Typhimurium M324 (Δ *aroA invC*::*aphT ssaV*::*cat* [27]). Colonies were selected by growth on LB agar plates containing Ampicillin, harvested, and pooled.

Growth for flow cytometry and cell sorting

A 1:1000 dilution of an overnight culture of the plasmid library was split into ten equal populations; five populations were assigned to "selected" and five to "control" groups. Cells were grown for 2 hours to reach exponential growth. Cultures were spun down at 3000 x g for five minutes at 4ºC. Growth media was removed and cultures were re-suspended in ice cold PBS. Cells were kept on ice until sorted or analyzed as described below.

Fluctuating selection using cell sorting

We subjected the plasmid library to fluctuating selection on fluorescence intensity, where selection for bright cells alternated with selection for dim cells.

Cells were sorted using fluorescence-activated cell sorting (FACS) with FACS–Diva sorting software (Becton Dickinson, CA). Immediately prior to sorting, 5 x 10^5 cells from each of the ten populations were analyzed for GFP expression. Based on this analysis, on the first day, a gate was drawn for each population to include either the highest 5% of cells expressing GFP, or a gate that covered the entire range of GFP expression, for selected and control lines, respectively. From each gated area, 1x10^5 cells were collected into a sterile well of a 24-well plate. Cells were collected at a 2.0 flow rate and sorted on the basis of "single cell" and "purity". After sorting, cells were spun at 3000 *g* for ten minutes and any FACS buffer was removed. Cells were re-suspended in 1ml LB media containing Ampicillin and grown overnight. The following day the process was repeated; however the gates for the selected populations included only the lowest 5% of cells expressing GFP. This process was repeated for a total of seven rounds of selection, with gates being drawn for selected populations in a fluctuating manner: selection on the highest 5% of GFP expression, then lowest 5%, and back again to the highest 5% of the total. After the 5^{th} round of selection all populations were placed at 4ºC for 48 hours. After this time, selection was resumed as normal. After all rounds of selection were completed, the populations were plated on LB agar plates containing Ampicillin, and 24 single colonies from each experimental population were randomly selected (240 clones in total). These were grown overnight in 1ml of LB containing Ampicillin and frozen at -80ºC in 15% glycerol.

Analysis and Data processing

One day prior to analysis, the 240 frozen clonal stocks were used to inoculate 1ml of medium in 5ml polystyrene round bottom tubes (BD Falcon) and prepared in the same manner as described above (Growth for cytometry and cell sorting). For each clone, $5x10^4$ cells were analyzed for GFP expression on the FACS Calibur (BD, CA).

Raw data was exported from FlowJo 4.6.1 software (TreeStar, Ashland, OR) into custom software. The software was used to exclude data deemed to be extraneous and for performing calculations relating to noise in fluorescence intensity.

The following conventions were applied to calculate variation in GFP expression and to limit the influence from cellular aggregates, cell detritus, and undefined values. Modified from Newman et al [7]:

1. All SSC, FSC, and fluorescence zero values were excluded.
2. Data was excluded that fell within the forward scatter (FSC) and side scatter (SSC) region where significant counts appeared in "buffer only" controls.
3. Extreme values of FSC and SSC were excluded (the highest and lowest 2.5% of events) from total counts in order to limit influence from cell detritus and cell aggregates.
4. FSC and SSC medians were calculated and a series of circular gates expanding out from the FSC and SSC medians were applied. For each gate size the coefficient of variation (CV) was calculated for fluorescence. A single gate size was then chosen for all analyses; this gate resulted in the lowest average CV (in order to maintain a conservative estimate of noise) yet contained enough cells for robust analysis (a minimum of 950 cells).
5. Extreme values of the fluorescence channel (FL1) (the highest 1.0% of events) were excluded to limit only a very small number of cells having undue effects on the values of the mean and CV.

When calculating the correlation between the coefficients of variation in fluorescence on two consecutive days, two data points were excluded from the analyses because they were more than 3 standard deviations away from expected values.

Sequencing

The following primers (F: 5' GTCAGAGGTTTTCACCGTCATCAC 3'. R: 5'CAAGAATTGGGACAACTCCAGTG 3') were used to PCR amplify the genomic segments inserted into plasmid pM968. Both primers anneal to regions on pM968 that flank the insert region. Inserts were sequenced using the reverse primer. The insert sequences were blasted against the genomic sequence of *Salmonella typhimurium* LT2 genomic and plasmid sequence (accession numbers NC_003197 and NC_003277), and the single best hit was retained as a hypothetical promoter. For each of these hypothetical promoters, the two nearest downstream genes were checked to see if either were oriented in the same direction as the hypothetical promoter. If either of these genes were oriented in the correct direction, the name and distance to the closest gene was noted. If neither of these genes were oriented in the correct direction, we concluded that it was unlikely that the insert sequence was actively driving transcription.

Construction of single copy chromosomal insertion of GFP under control of the fliC promoter

The following primers

(F5':GGCAACAGCCCAATAACATCAAGTTGTAATTGATAAGGAAAAGATCATGAG TAAAGGAGAAGAACTTTTC 3' and

R5':CGCTGCCTTGATTGTGTACCACGTGTCGGTGAATCAATCGCCGGATTATTACG CCCCGCCCTGCCA 3'

were used to PCR amplify *gfp*+ and chloramphenicol resistance cassette from strain JH3016 [8]. Amplified DNA was electroporated into strain M557 [27] containing pKD46 for lambda red recombination as described in [28].Constructs were cured of the pKD46 plasmid by growth at 43°C on non-selective plates. Final clones were tested for sensitivity to ampicillin. Constructs were sequenced to confirm proper orientation and insertion at the *fliC* locus.

Comparison of noise in expression of chromosomal-based and plasmid based fliC promoter

Two plasmids isolated independently from the plasmid library containing the fliC promoter and five independent chromosomal constructs of *fliC::gfp*+ were used for analysis. Plasmids and chromosomal inserts were measured in the background of strain M557. *rpsM-gfp*+ [8] in the M557 background was used as a control known to have low levels of noise, and M557 (no

GFP) was used as a negative control. All clones were streaked onto agar plates and ten single colonies were randomly chosen. Each colony was grown up and measured using the FACS as described in the Materials and Methods section of the manuscript. Analysis was also done as described in the text. Noise levels, as given by the coefficient of variation in GFP expression, were compared (See Figure 3a). We found that overall GFP expression was lower for chromosomal-based expression as compared to plasmid-based. The noise profiles also showed differences between the two types (see Figure 3b). There was no significant difference in the noise of GFP expression between plasmid-based and chromosomal-based GFP expression under the control of the *fliC* promoter.

Western Blot analysis of cells sorted based on GFP expression

We performed a western blot analysis using *anti-FliC, -FljB* antibodies and *anti-OmpC* (as a loading control) to test whether GFP expression from the *fliC* promoter on the plasmid (*pfliC::gfp*) does accurately represent the concentration of *FliC* protein in the cell. A clonal population of cells containing the *pfliC::gfp* contruct in plasmid pM9568 exhibiting noisy expression were sorted with the FACS into three fractions, each fraction containing 1.4x 10^7 cells. The first fraction contained cells with high levels of fluorescence (above the 55^{th} percentile); the second fraction contained cells whose fluorescence did not exceed background (below the 35^{th} percentile); the third fraction was a random sample of cells, chosen irrespective of their level of fluorescence. Cells were concentrated into 15ul and 5ul of Laemmli buffer was added. Cells were heated to 95°C for 5 minutes and then frozen at -20°C. Cell fractions were then subjected to western blotting, with staining for *anti-FliC*, *-FljB*, then reprobed with staining with monoclonal mouse antibody (CM95) *OmpC*-antibody as a control for total number of cells. The western blotting shows that GFP expression does correlate with the expression of *FliC*. This is indicated in Figure 4.

Cell tracking and analysis of GFP expression in microcolony formation

Cell tracking software was used to track cell lineages and analyze GFP expression in individual cells during microcolony growth as described in [29].

Acknowledgments

We thank T. Bergmiller and A. Sturm for microbiology advice, the ETH Zurich FACS facility for help with cell sorting, Kelly Hughes for strains TH5971 and TH5862, Michael Elowitz and Nitzan Rosenfeld for cell tracking software, and Sebastian Bonhoeffer and the Theoretical Biology group at ETH for stimulating discussions.

References

1. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genetics 31: 69-73.

2. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297: 1183-1186.

3. Raser JM, O'shea EK (2005) Noise in gene expression: Origins, consequences, and control. Science 309: 2010-2013.

4. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. Science 317: 526-529.

5. Acar M, Becskei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. Nature 435: 228-232.

6. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422: 633-637.

7. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. Nature 441: 840-846.

8. Hautefort I, Proenca MJ, Hinton JCD (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. Applied and Environmental Microbiology 69: 7480-7491.

9. Bonifield HR, Hughes KT (2003) Flagellar Phase Variation in Salmonella enterica Is Mediated by a Posttranscriptional Control Mechanism. Journal of Bacteriology 185: 3567-3574.

10. Zieg J, Simon M (1980) Analysis of the nucleotide sequence of an invertible controlling element. Proc Natl Acad Sci US A 77: 4196-4200.

11. Stocker BAD (1949) Measurements of Rate of Mutation of Flagellar Antigenic Phase in Salmonella typhi-murium. The Journal of Hygiene 47: 398-413.

12. Gillen KL, Hughes KT (1991) Negative regulatory loci coupling flagellin synthesis to flagellar assembly in Salmonella typhimurium. J Bacteriol 173: 2301-2310.

13. Cummings LA, Wilkerson WD, Bergsbaken T, Cookson BT (2006) In vivo, fliC expression by Salmonella enterica serovar Typhimurium is heterogeneous, regulated by ClpX, and anatomically restricted. Mol Microbiol 61: 795-809.

14. Schmitt CK, Ikeda JS, Darnell SC, Watson PR, Bispham J, et al. (2001) Absence of All Components of the Flagellar Export and Synthesis Machinery Differentially Alters Virulence of Salmonella enterica Serovar Typhimurium in Models of Typhoid Fever, Survival in Macrophages, Tissue Culture Invasiveness, and Calf Enterocolitis. Infection and Immunity 69: 5619.

15. Stecher B, Hapfelmeier S, Muller C, Kremer M, Stallmach T, et al. (2004) Flagella and chemotaxis are required for efficient induction of Salmonella enterica serovar typhimurium colitis in streptomycin-pretreated mice. Infection and Immunity 72: 4138-4150.

16. Hackett J, Attridge S, Rowley D (1988) Oral Immunization with Live, Avirulent Fla+ Strains of Salmonella Protects Mice against Subsequent Oral Challenge with Salmonella-Typhimurium. Journal of Infectious Diseases 157: 78-84.

17. Cookson BT, Bevan MJ (1997) Identification of a natural T cell epitope presented by Salmonella-infected macrophages and recognized by T cells from orally immunized mice. Journal of Immunology 158: 4310-4319.

18. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, et al. (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410: 1099-1103.

19. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, et al. (2003) Bacterial Flagella. Nature Immunology 4: 1247-1253.

20. Henderson IR, Owen P, Nataro JP (1999) Molecular switches--the ON and OFF of bacterial phase variation. Mol Microbiol 33: 919-932.

21. Stecher B, Barthel M, Schlumberger MC, Haberli L, Rabsch W, et al. (2008) Motility allows S. Typhimurium to benefit from the mucosal defence. Cellular Microbiology 10: 1166-1180.

22. Salazar-Gonzalez RM, McSorley SJ (2005) Salmonella flagellin, a microbial target of the innate and adaptive immune system. Immunology Letters 101: 117-122.

23. Ackermann M, Stecher B, Freed NE, Songhet P, Hardt WD, et al. (2008) Self-destructive cooperation mediated by phenotypic noise. Nature 454: 987-990.

24. Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. Science 309: 2075-2078.

25. Booth IR (2002) Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. International Journal of Food Microbiology 78: 19-30.

26. Hoiseth SK, Stocker BA (1981) Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. Nature 291: 238-239.

27. Hapfelmeier S, Stecher B, Barthel M, Kremer M, Muller AJ, et al. (2005) The Salmonella Pathogenicity Island (SPI)-2 and SPI-1 Type III Secretion Systems Allow Salmonella Serovar typhimurium to Trigger Colitis via MyD88-Dependent and MyD88-Independent Mechanisms 1. The Journal of Immunology 174: 1675-1685.

28. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640-6645.

29. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. Science 307: 1962-1965.

Chapter 3

Self-destructive cooperation mediated by phenotypic noise[†]

Martin Ackermann[1], Bärbel Stecher[2], Nikki E. Freed[1], Pascal Songhet[2], Wolf-Dietrich Hardt[2] & Michael Doebeli[3]

Institute of Integrative Biology, ETH Zurich, 8092 Zürich, Switzerland[1]. Institute of Microbiology, ETH Zurich, 8093 Zürich, Switzerland[2]. Department of Zoology and Department of Mathematics, University of British Columbia, Vancouver BC V6T 1Z4, Canada[3].

[†] A modified version of this manuscript appears as Ackermann M, Stecher B, Freed NE, Songhet P, Hardt WD, Doebeli M. Self-destructive cooperation mediated by phenotypic noise. Nature. (2008) 454: 987-90

Abstract

In many biological examples of cooperation, individuals that cooperate cannot benefit from the resulting public good. This is especially clear in cases of self-destructive cooperation, where individuals die when helping others. If self-destructive cooperation is genetically encoded, these genes can only be maintained if they are expressed by just a fraction of their carriers, whereas the other fraction benefits from the public good. One mechanism that can mediate this differentiation into two phenotypically different sub-populations is phenotypic noise[1,2]. Here we show that noisy expression of self-destructive cooperation can evolve if individuals that have a higher probability for self-destruction have, on average, access to larger public goods. This situation, which we refer to as assortment, can arise if the environment is spatially structured. These results provide a new perspective on the significance of phenotypic noise in bacterial pathogenesis: it might promote the formation of cooperative sub-populations that die while preparing the ground for a successful infection. We show experimentally that this model captures essential features of *Salmonella typhimurium* pathogenesis. We conclude that noisily expressed self-destructive cooperative actions can evolve under conditions of assortment, that self-destructive cooperation is a plausible biological function of phenotypic noise, and that self-destructive cooperation mediated by phenotypic noise could be important in bacterial pathogenesis.

Introduction

Recent experimental work demonstrated that genetically identical organisms living in the same environment show surprisingly high levels of variation in phenotypic traits[1,2], and sometimes even switch between distinct phenotypic states[3]. Stochastic cellular processes are one source of such phenotypic noise. The level of phenotypic noise is subject to mutational change, and can thus evolve. This raises the question whether natural selection always acts towards minimizing phenotypic noise, or whether there are cases in which genotypes that encode variable phenotypes are favoured by selection. In the existing theory[4,5,6], the most prominent adaptive explanation for phenotypic noise is bet-hedging[5], according to which the stochastic expression of alternative phenotypes allows a genotype to survive changes in external conditions and thus to persist in fluctuating environments.

Here we focus on a fundamentally different adaptive explanation for phenotypic noise: self-destructive cooperation. In this scenario, the individuals that survive and form a successful lineage all express the same phenotype. Individuals that express an alternative phenotype do exist, but they do not contribute to future generations; instead, they die while contributing to a public good that benefits others. There are many examples of cooperative acts that prevent reproduction or survival of the actor, ranging from non-reproductive workers in mammals and insects to unicellular bacteria that lyse when releasing chemical substances that benefit others. Importantly, genotypes that have the propensity to express self-destructive cooperation can only persist if the expression is limited to a fraction of the individuals carrying the genotype, whereas another fraction does not express the cooperative behaviour and benefits from the public good produced. Sometimes, this differentiation into two fractions is mediated by signals. In other examples, notably in microorganisms, there seems to be no signal. In these cases, phenotypic noise could promote the differentiation required for self-destructive cooperation to persist.

We investigated the evolutionary dynamics of a self-destructive cooperative act that contributes to the generation of a public good and that is expressed in a stochastic manner. In general, cooperation can evolve if cooperative individuals benefit from cooperative acts of others more often than non-cooperative individuals[7]—a situation referred to as assortment. We used a simple model to quantify the level of assortment as a function of the external conditions, and to calculate how selection on the probability to express self-destructive cooperation depends on the level of assortment, as well as on the amount of public good generated by cooperative acts.

Results and Discussion

The model is based on the public goods game and assumes that there are two strategies: cooperate (C) and defect (D). C sacrifices itself with probability q, and only if it sacrifices itself, it contributes an amount b to the public good. The decision between sacrificing and not sacrificing is a chance outcome resulting from phenotypic bi-stability; every cooperator makes this decision independently of the environment or of the decisions of other individuals. D never contributes to the public good. The game is played in interaction groups of N players. In general, if there are k cooperators among the N members of an interaction group, the total amount of the public good produced in that group is kqb. The total amount of the public good is available to each surviving player in the interaction group (an alternative scenario where

the public good is divided among the surviving players gives very similar results). We assume that, in addition to the public good, all players also receive a non-zero baseline payoff w.

Consider first the payoff to a focal C player in a given interaction group. Because the focal C is one of the k cooperators, its social environment consists of k - 1 cooperators and N - k defectors. The focal C gets no payoff from the defectors, but if it survives it receives the benefit b with a probability q from each of the other k - 1 cooperators, as well as the baseline payoff w. The probability that the focal C does not sacrifice itself is $1-q$, and hence the expected payoff to C in the given interaction group is:

$$p_{\mathrm{C}}(N,k)=(1-q)\big((k-1)qb+w\big) \tag{1}$$

The social environment of a focal D in the given interaction group consists of k cooperators and N - k - 1 defectors, and its expected payoff is:

$$p_{\mathrm{D}}(N,k)=kqb+w \tag{2}$$

Because $p_{\mathrm{C}}(N, k) < p_{\mathrm{D}}(N, k)$, cooperators always do worse than defectors within a given group. Therefore, the only way in which cooperators can dominate on a population-wide scale is if cooperators have, on average, a different social environment than defectors.

The composition of the social environment of a focal C or D depends on the current frequency of cooperators in the population, x, and on how individuals come together to form interaction groups. Let $e_{\mathrm{C}}(x)$ be the average number of cooperators among the other N - 1 members in an interaction group containing a focal C. Similarly, let $e_{\mathrm{D}}(x)$ be the number of cooperators in an average interaction group containing a focal D. The expected population-wide payoffs to C and D are then:

$$P_{\mathrm{C}}(x)=(1-q)\big(e_{\mathrm{C}}(x)qb+w\big) \tag{3}$$

and

$$P_{\mathrm{D}}(x)=e_{\mathrm{D}}(x)qb+w \tag{4}$$

With random composition of groups, we have $e_{\mathrm{C}}(x) = e_{\mathrm{D}}(x) = x(N - 1)$. In this case, the population-wide payoffs always satisfy $P_{\mathrm{C}}(x) < P_{\mathrm{D}}(x)$ for all x, and hence defectors always win. Thus, for cooperation to thrive, cooperators must find themselves, on average, in interaction groups containing more cooperators than the interaction groups in which defectors

find themselves on average. In other words, there must be positive assortment between cooperators.

Assortment can result from different mechanisms, for example, spatial structure, reciprocity or kin recognition[7]. Here we consider a simple case of spatially structured populations inhabiting an infinite number of demes. In the case of pathogenic bacteria, a deme would represent a host. Each deme is seeded by M individuals from a common pool of individuals. The number of individuals then increases to the carrying capacity of the deme, which is assumed to be N, the interaction group size. After reaching carrying capacity, cooperators sacrifice themselves with probability q; if they do, they contribute b to the public good. Cooperators that did not sacrifice themselves and defectors then harvest the public good. The payoffs they receive determine how much they contribute to the pool of individuals from which the next generation of demes is seeded.

If the number of individuals seeding a deme, M, is small, then the degree of assortment is high, and a focal cooperator sees on average more cooperators than a focal defector. It is easy to see (see Appendix II) that the average social environment of a focal cooperator and a focal defector is:

$$e_{\mathrm{C}}(x)=\frac{(M-1)N}{M}x+\frac{N}{M}-1 \qquad (5)$$

and

$$e_{\mathrm{D}}(x)=\frac{(M-1)N}{M}x \qquad (6)$$

Thus, $e_{\mathrm{C}}(x) > e_{\mathrm{D}}(x)$ for all x, which enables the origin and maintenance of cooperation based on equations (3) and (4). The general picture is as follows (Fig. 1; derivation in Appendix II): First, there is a value such that for any $q < q_1^*$, C can invade D, that is, $P_{\mathrm{C}}(0) > P_{\mathrm{D}}(0)$ ($q_1^* = \left[b(N-M)-Mw\right]/\left[b(N-M)\right]$; see Appendix II). Second, there is a value $q_2^* < q_1^*$ such that for any $q < q_2^*$, C not only invades D but also goes to fixation, that is, $P_{\mathrm{C}}(1) > P_{\mathrm{D}}(1)$) ($q_2^* = \left[b(N-M)-Mw\right]/\left[b(NM-M)\right]$; see Appendix II). Third, for any q with $q_2^* < q < q_1^*$, there is coexistence between cooperators and defectors, that is, C can invade D and D can invade C.

These results show that the assortment generated by the deme structure changes the nature of the evolutionary game between cooperators and defectors. In well-mixed populations without demes, D always dominates C. With demes, this remains true if C players have large probabilities q of committing cooperative suicide. However, for intermediate q (that is, $q_2^* < q < q_1^*$), the structure of the game changes to a 'Snowdrift' scenario[8], in which both C and D can invade when rare. For even smaller $q\left(q < q_2^*\right)$,) assortment makes C the dominant strategy and eliminates D.

Defectors D can be viewed as cooperators with $q = 0$, which naturally leads to the consideration of q as a continuous trait. Evolution of such continuous traits can be studied using adaptive dynamics theory[9,10,11]. Starting with a population of defectors, the trait q evolves to a convergence stable and evolutionarily stable strategy q_3^* satisfying $0 < q_3^* < q_2^*\left(q_3^* = \left[b\left(N-M\right)-Mw\right]/\left[b\left(NM+N-2M\right)\right]\right.$; see Appendix II). Therefore, continuous evolution by small steps results in a population consisting of a single cooperative strategy that cannot be invaded by any nearby strategies q or by pure defectors. Our model thus provides a framework for understanding the evolution of self-destructive cooperation mediated by phenotypic noise. We note that our deme-structured model of cooperative suicide can also be interpreted in the context of kin or group selection[7,12]. However, these perspectives are not required for understanding the evolution of cooperation; simple considerations based on assortment and interaction environments are sufficient.

This model sheds new light on the role of phenotypic noise in the biology of microbial pathogens. Phenotypic noise or bi-stability is common in unicellular pathogens, as are acts of self-destructive cooperation[9]. Some bacterial toxins that are instrumental in pathogenesis can only be released if the cell producing the toxin lyses[10,11,12]. Some of these toxins induce inflammation in the host, and there is now growing evidence that pathogens can decrease competition by co-inhabitants of the same niche through manipulation of the host's immune system[13,14,15,16]. One example is pneumolysin from *Streptococcus pneumoniae*. This toxin is released through bacterial lysis and enhances lung colonization[10,17]. A second example is TcdA, a key virulence factor of *Clostridium difficile[12]*. TcdA lacks a standard secretion signal and is released by bacterial lysis. Purified TcdA toxin alone can trigger gut inflammation[18], and gut inflammation enhances intestinal *C. difficile* colonization[19]. In this case, TcdA released by self-destructive acts seems to provide the pathogen with a competitive advantage, presumably by decreasing competition from commensal bacteria.

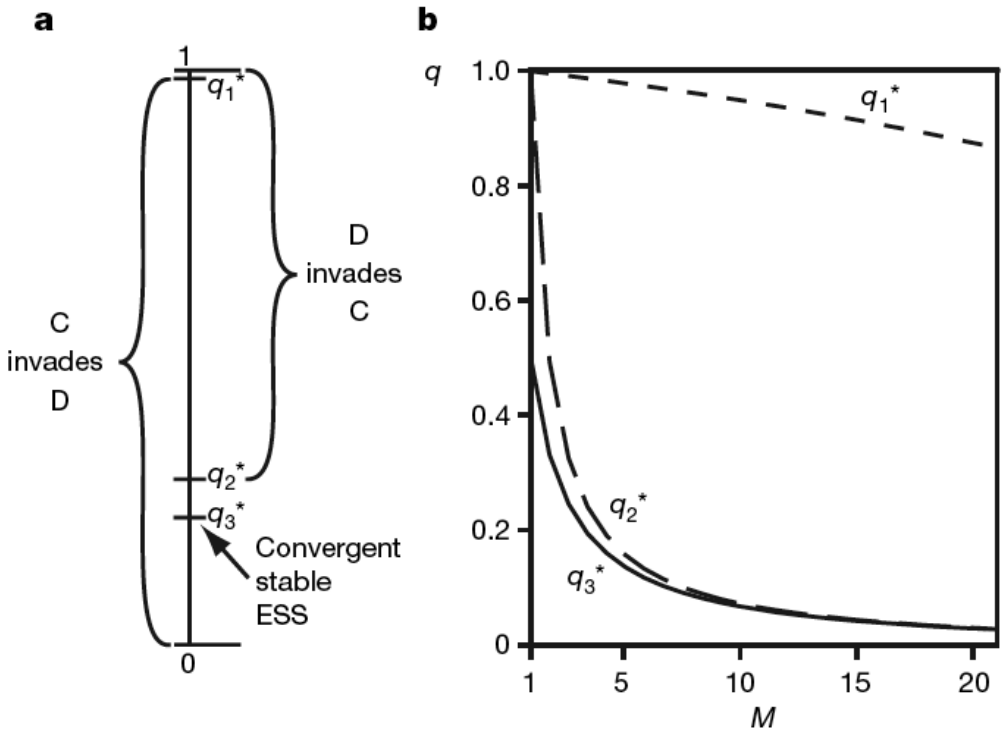


Figure 1: The evolutionary dynamics of stochastic self-destructive cooperation.

a) The dynamics are determined by three specific values for the probability q to self-sacrifice. The first value is q^*_1 . A population of pure defectors can be invaded by cooperators' that self-sacrifice with a probability $q<q^*_1$. The second value is q^*_2. A homogeneous population of cooperators can be invaded by defectors if the cooperators' probability to self-sacrifice is larger than q^*_2. Cooperators with a probability to self-sacrifice between q^*_1 and q^*_2 will thus stably co-exist with pure defectors. If *q* evolves through mutations of small effects, and the population initially consists of defectors, then *q* will evolve to the value q^*_3. This value is an evolutionarily stable strategy (ESS) and represents an endpoint for the evolutionary dynamics. The three values for q are determined by the parameters *M*, *N*, *b* and *w*, as described in the Appendix II. In the example shown here, the parameters are *M* = 3, *N* = 100, *b* = 5 and *w* = 2. b) When the number of individuals colonizing a deme, M, increases, and assortment therefore decreases, the three values q^*_1, q^*_2, and q^*_3 decrease. The parameter values are *N* = 100, *b* = 5 and *w* = 2.

We focused on *S. typhimurium* enterocolitis as a third example of bacterial pathogenesis, and tested experimentally whether central aspects of the infection process are captured by the model of self-destructive cooperation mediated by phenotypic noise. The establishment of *S. typhimurium* in the gut is hindered by the presence of the intestinal microflora. These competitors are removed by an inflammatory response in the gut triggered by *S. typhimurium* invading the gut tissue[20]. Gut tissue invasion and the triggering of inflammation depend on

S. typhimurium virulence factors, namely Type III secretion systems (TTSS) and flagella. Invasion factors—that is, the invasion-mediating TTSS-1 and the flagella—are heterogeneously expressed in *S. typhimurium* populations[21,22,23]. In our experiments, gut inflammation, which alleviates competition by commensals, was regarded as the public good. We focused on TTSS-1 expression as the phenotypic trait that is expressed stochastically, and tested three main assumptions of the mathematical model (Fig. 2; see Appendix II for experimental methods).

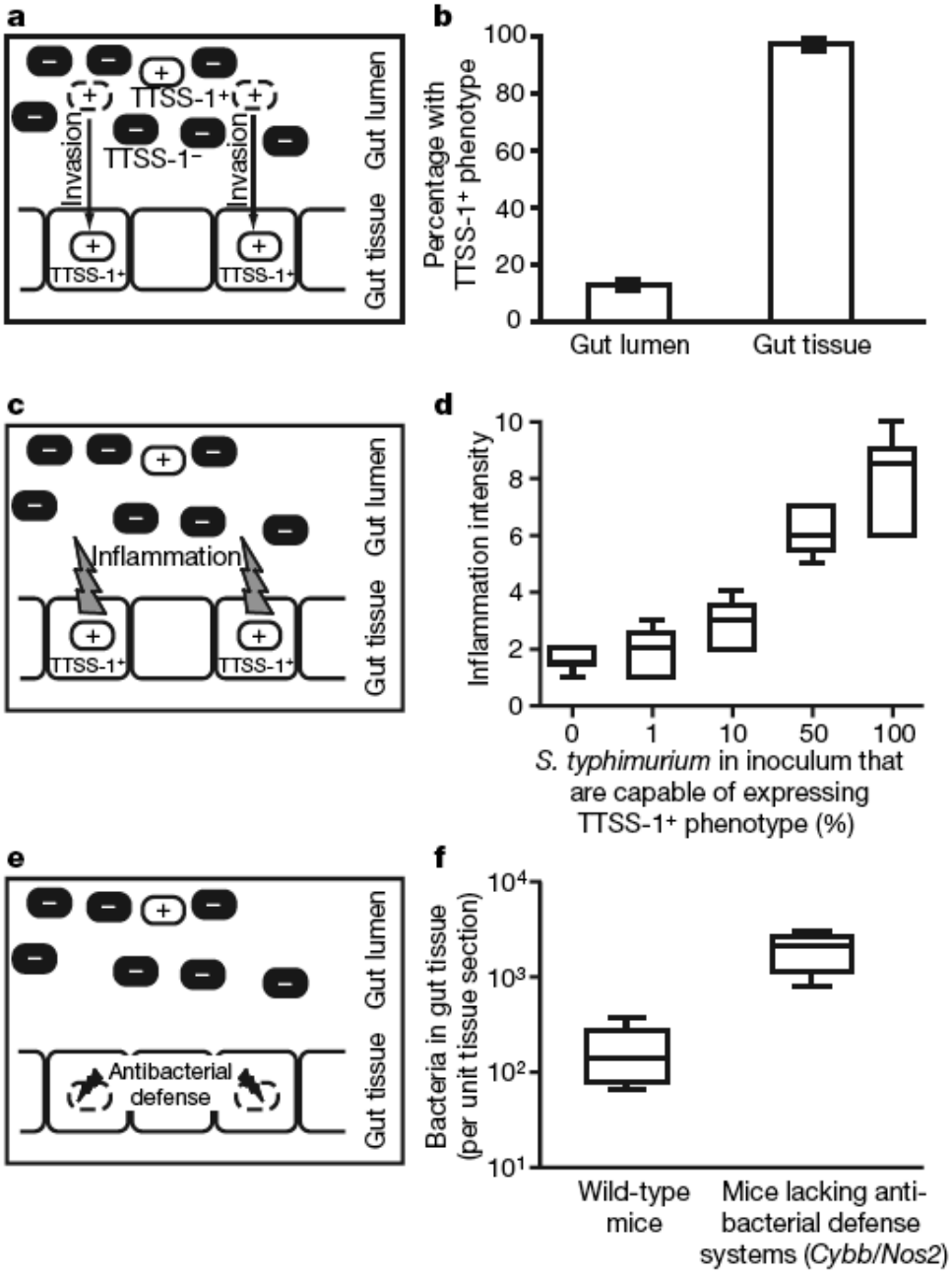


Figure 2: Testing biological assumptions of self-destructive cooperation mediated by phenotypic noise with a mouse model for *S. typhimurium* enterocolitis. For experimental details, see Appendix II.

a) The clonal bacterial population consists of two phenotypes; one of them (TTSS-1$^+$; white) expresses a cooperative act consisting of gut tissue invasion. b) We analysed the TTSS-1 phenotype in the gut lumen and in the gut tissue. In the gut lumen, about 15% of the bacteria were TTSS-1$^+$; in the gut tissue, almost all bacteria were TTSS-1$^+$ (Mann–Whitney U test for a difference between lumen and tissue, P<0.001; error bars, s.e.m.). The luminal *S. typhimurium* population differentiates into TTSS-1$^+$ and TTSS-1$^-$ phenotypes, and the TTSS-11 phenotype invades the gut tissue. c) The amount of public

good generated in an interaction group increases with an increasing number of cooperators. In the context of *S. typhimurium* infection, the public good is gut inflammation elicited by tissue invasion. d) To vary the number of bacteria that commit the cooperative act, we mixed wild-type *S. typhimurium* that express both TTSS-1$^+$ and TTSS-1$^-$ phenotypes (see b) with an isogenic strain that is incapable of expressing the TTSS-1$^+$ phenotype (ΔTTSS-1). The intensity of gut inflammation increased with an increasing fraction of wild-type *S. typhimurium* in the inoculum (box plots with median, quartiles and range; Spearman's rank correlation, P<0.001). Gut inflammation increases with an increasing number of individuals that express the cooperative act of tissue invasion. e) Contributing to the public good is a self-destructive act. In the context of *S. typhimurium* infection, most bacteria invading the gut tissue are expected to be killed by antimicrobial defence mechanisms, specifically by the *Cybb/Nos2*- encoded systems generating antimicrobial oxygen and nitrogen radicals. f) To test whether bacteria that invade the gut tissue are killed, we compared *S. typhimurium* loads in the gut tissue of wild-type mice and of mutant mice lacking the *Cybb/Nos2* systems. The bacterial loads in the gut tissue in wild-type mice were about ten times lower than in the *Cybb/Nos2*- knockout mice (box plots with median, quartiles and range; Mann–Whitney U test, P = 0.008; in the gut lumen, bacterial loads were not significantly different; see Appendix II). This suggests that in wild-type mice, most bacteria that invade the gut tissue are killed. Thus, triggering of gut inflammation by tissue invasion can be regarded as a self-destructive act.

Our experimental results show that: first, in a clonal population of *S. typhimurium*, in the gut lumen, only about 15% were phenotypically TTSS-1$^+$, which is in line with *in vitro* studies[22,23]. In contrast, almost all bacteria in the gut tissue expressed the TTSS-1$^+$ phenotype (Fig. 2a, b). This supports the assumption that TTSS-1 expression is variable in clonal populations of *S. typhimurium*, and that the TTSS-1 phenotype of a bacterium determines whether or not it will invade the gut tissue. (Other invasion factors probably also have an important role; TTSS-1 expression is therefore required but not sufficient for invasion.) Second, the intensity of inflammation increases as the proportion of bacteria that are capable of expressing the cooperative TTSS-1$^+$ phenotype increases (Fig. 2c,d). Taking inflammation as a proxy for the public good, this supports the assumption that the public good produced increases with increasing numbers of cooperators. Third, in the *S. typhimurium* enterocolitis model, most of the bacteria that invade the gut tissue and thereby contribute to the public good seem to be killed by the intestinal innate immune defenses (Fig. 2e, f). Thus, cooperation through invasion of the gut tissue is a largely self-destructive act.

Together, these experimental results indicate that the mathematical model of self-destructive cooperation mediated by phenotypic noise captures central features of *S. typhimurium*

enterocolitis. This gives a new perspective on how these and similar pathogens evolve self-destructive cooperative acts to infect their hosts. Our model predicts that this behaviour can evolve if the number of pathogens that infect a host is small. This number corresponds to the parameter *M*, and if *M* is small then bacteria that carry the genotype for stochastic cooperation without expressing this behaviour are surrounded by many cells with the same genotype that do cooperate, and hence this genotype will thrive. A first estimate for *M* is the number of bacteria required for an infection to establish. For *Salmonella* and *Escherichia coli* infecting humans, estimates can reach levels as low as one hundred[24]. Mortality during the passage through the stomach, as well as spatial structure of the environment from which the inoculum originates, can lead to a small effective *M* even if the infective dose is substantial. Over evolutionary time, a small *M* leads to the evolution of larger probabilities to express the cooperative act (Fig. 1) and, as one can easily show mathematically, to larger payoffs. This is in line with previous theoretical findings that clonal infections are beneficial for pathogens, whereas coinfection tends to be detrimental[25].

Self-destructive cooperation can be seen as an extreme form of the division of labour between two phenotypes, in which one of the phenotypes does not survive. The two phenotypes are encoded by the same genotype, which can persist because the expression of the self-destructive phenotype is stochastic. We thus conclude that self-destructive cooperation is a plausible biological explanation for certain instances of phenotypic noise. Establishing a link between phenotypic noise and cooperation gives new insights into how cooperation can persist despite its cost for the benefactor. At the same time, this link provides a new perspective into the significance of phenotypic noise in biological systems, and especially in microbial pathogens. Understanding why so many pathogens exhibit stochastic phenotypic variation is essential for developing efficient strategies for their control.

Acknowledgements

We are grateful to Markus Schlumberger, K. Aktories, I. Just, S. Hammerschmidt and J. Fletcher for discussions, and M. Barthel and the RCHCI team for professional help with the animal experiments. M.A. and N.E.F. were supported by the Swiss National Science Foundation, and M.A. was also supported by the Roche Research Foundation and the Novartis Foundation. M.D. is supported by NSERC (Canada). B.S. was supported by the UBS foundation. *Salmonella* work in the Hardt laboratory is supported by grants from the ETH

research foundation (TH-10 06-1), the Swiss National Science Foundation (310000-113623/1) and the European Union (SavinMucoPath number 032296).

Author Contributions

M.A., W.-D.H. and M.D. formulated the question; M.A. and M.D. wrote the mathematical model; M.D. analysed the mathematical model; M.A., B.S., N.E.F. and W.-D.H. planned the experiments and interpreted the results; B.S. performed the experiments for Figure 2b and 2d; P.S. performed the experiment for Figure 2f; M.A., W.-D.H. and M.D. wrote the manuscript.

References

1. Raser JM, O'shea EK (2005) Noise in gene expression: Origins, consequences, and control. Science 309: 2010-2013.

2. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. Nature Reviews Genetics 6: 451-464.

3. Dubnau D, Losick R (2006) Bistability in bacteria. Molecular Microbiology 61: 564-572.

4. Bull JJ (1987) Evolution of Phenotypic Variance. Evolution 41: 303-315.

5. Kussell E, Kishony R, Balaban NQ, Leibler S (2005) Bacterial persistence: A model of survival in changing environments. Genetics 169: 1807-1814.

6. Wolf DM, Vazirani VV, Arkin AP (2005) A microbial modified prisoner's dilemma game: how frequency-dependent selection can lead to random phase variation. Journal of Theoretical Biology 234: 255-262.

7. Nowak MA (2006) Five rules for the evolution of cooperation. Science 314: 1560-1563.

8. Hauert C, Doebeli M (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature 428: 643-646.

9. Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol: 577–587.

10. Paton JC (1996) The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae. Trends in Microbiology 4: 103-106.

11. Wagner PL, Livny J, Neely MN, Acheson DWK, Friedman DI, et al. (2002) Bacteriophage control of Shiga toxin 1 production and release by Escherichia coli. Molecular Microbiology 44: 957-970.

12. Voth DE, Ballard JD (2005) Clostridium difficile toxins: Mechanism of action and role in disease. Clinical Microbiology Reviews 18: 247-+.

13. Lysenko ES, Ratner AJ, Nelson AL, Weiser JN (2005) The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. Plos Pathogens 1: 3-11.

14. Raberg L, de Roode JC, Bell AS, Stamou P, Gray D, et al. (2006) The role of immune-mediated apparent competition in genetically diverse malaria infections. American Naturalist 168: 41-53.

15. Brown SP, Le Chat L, Taddei F (2008) Evolution of virulence: triggering host inflammation allows invading pathogens to exclude competitors. Ecology Letters 11: 44-51.

16. Stecher B, Hardt W (2008) The role of microbiota in infectious disease. Trends in Microbiology.

17. Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC (2007) Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of Streptococcus pneumoniae. Infection and Immunity 75: 350-357.

18. Lima A, Lyerly D, Wilkins T, Innes D, Guerrant R (1988) Effects of Clostridium difficile toxins A and B in rabbit small and large intestine in vivo and on cultured cells in vitro. Infection and Immunity 56: 582-588.

19. Rodemann J, Dubberke E, Reske K, Seo D, Stone C (2007) Incidence of Clostridium difficile Infection in Inflammatory Bowel Disease. Clinical Gastroenterology and Hepatology 5: 339-344.

20. Stecher B, Robbiani R, Walker A, Westendorf A, Barthel M, et al. (2007) Salmonella enterica Serovar Typhimurium exploits inflammation to compete with the intestinal microbiota. Plos Biology 5: e244.

21. Cummings L, Wilkerson W, Bergsbaken T, Cookson B (2006) In vivo, fliC expression by Salmonella enterica serovar Typhimurium is heterogeneous, regulated by ClpX, and anatomically restricted. Molecular Microbiology 61: 795-809.

22. Schlumberger M, Muller A, Ehrbar K, Winnen B, Duss I, et al. (2005) Real-time imaging of type III secretion: Salmonella SipA injection into host cells. Proceedings of the National Academy of Sciences 102: 12548-12553.

23. Hautefort I, Proenca MJ, Hinton JCD (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. Applied and Environmental Microbiology 69: 7480-7491.

24. Kothary MH, Babu US (2001) Infective dose of foodborne pathogens in volunteers: A review. Journal of Food Safety 21: 49-73.

25. May RM, Nowak MA (1995) Coinfection and the Evolution of Parasite Virulence. Proceedings of the Royal Society of London Series B-Biological Sciences 261: 209-215.

Chapter 4

Robustness of transcription networks to selection for increased noise

Nikki E. Freed, Olin K. Silander, and Martin Ackermann

Institute of Biogeochemistry and Pollutant dynamics, ETH Zurich, Zurich Switzerland

Abbreviations:

FACS: Fluorescent activated cell sorter

GFP: Green fluorescent protein

S. Typhimurium: *Salmonella enterica* ssp. I serovar Typhimurium

E. coli: *Escherichia coli*

Abstract

We set out to evolve increased levels of phenotypic noise in gene expression in specific bacterial promoters. We treated five different bacterial promoters to fluctuating selection based on GFP expression using a cell sorter, alternating between selection for bright and then dim expression. We did not see a response to selection and conclude that noise in gene expression is robust to our fluctuating selection.

Introduction

Clonal populations of bacteria or yeast living in constant environmental conditions can exhibit variation in their phenotype. Some of this variation has been shown to arise from noise, or stochastic fluctuations, in gene expression and protein production within each individual cell[1]. Much attention has been given to the mechanisms that drive this noise, however, less is known about how selection can act on this trait and if it is possible for organisms to evolve differences in variation of gene expression. We set out to evolve increased noise in expression of single genes, and if possible, determine the genetic basis underlying increased phenotypic noise.

Control of cellular processes, such as transcription and translation, involve a series of biochemical steps that are dependent on stochastic processes. The diffusion of molecules, the strength of binding between proteins, and variation in the concentration of each molecule can affect each of the steps involved in protein production. It is then possible to imagine that for two genetically identical cells, this stochasticity in protein production can translate into each cell having a distinct phenotype, even in an identical growth environment[2,3]. Recent advances in single cell imaging, single molecule assays, and high throughput techniques have begun to indicate which genes are noisy, and how noise might influence biological processes [4]. There are underlying genetic components to noise in gene expression. As a consequence, this trait could thus hypothetically respond to natural selection and evolve to different levels. Mutations that affect the rate of transcription and translation affect the level of noise [5], while the architecture of gene regulatory networks, such as positive feedback loops, can act as a noise amplifier[6].

Theoretically, there are foreseeable selective advantages of noise in gene expression for some genes in some environmental contexts. Current theory argues that phenotypic variation in some traits may be advantageous if an organism is living in a fluctuating environment

[6,7,8,9] and may provide an evolutionary strategy to "hedge bets" against a changing environment. We have argued elsewhere[10] that exploiting phenotypic variation may be evolutionarily advantageous in specific cases of cooperation within clonal bacterial populations. In contrast, there are indications that limiting phenotypic noise may be important for certain classes of genes, such as for housekeeping or essential genes. Fraser et al[11] observed that highly conserved genes have been shown to have significantly lower variation in their expression.

Measurement of Noise

Several techniques have facilitated the investigation of phenotypic noise. Often, fluorescent reporters such as green fluorescent protein (GFP) are transcriptionally or translationally linked to a gene of interest. When coupled with flow cytometry, which allows single cell measurements of GFP expression of thousands of cells in a few seconds, a quantitative analysis of variation in gene expression can be obtained. A simple and direct measure of noise in single cell expression experiments is then frequently given by the coefficient of variation (CV) where the standard deviation of GFP expression is divided by the mean ($CV=\sigma/N$)[3]. In addition, others have used noise strength ($\varphi=\sigma^2/N$) (discussed in [2]). In large scale studies where the expression of numerous genes are monitored the distance of each CV to a running median of other CVs has been used[4].

We set out to select for increased noise in five different promoters using fluctuating selection on expression, alternating between selection for high expression, then low, and high again allowing for approximately 20 generations between each selection event (See Materials and Methods). After selection, GFP expression is measured in individual clones and changes in expression are investigated. For 4 of 5 promoters we observed no change in either mean or variance in expression. In addition, the use of fitness as a metric to gauge a response to selection, also indicated that no significant change had occurred in any evolved or control populations.

Results

Pilot study

We performed a pilot study using a chromosomal construction of the *S.* Typhimurium *rpsM* promoter (P_{rpsM}) transcriptionally linked to a gene encoding a green fluorescent protein

(*gfp*$^{+}$[12]) in a neutral location in the chromosome [13]. P_{rpsM} is involved in controlling the expression of 30S ribosomal subunit protein S13. A single clone was grown up and split into eight replicate populations, four were designated as "selected" and four as "control" replicates. Selected replicates were treated to alternating selection using a fluorescent activated cell sorter (FACS) (See Methods and Materials and [14]). Briefly, on the first day of evolution, cells were grown to mid-log growth and 1 x 10^5 cells were sorted from the selected replicates expressing the highest 5% of GFP from the population. These cells were collected, allowed to grow for 4 hours, treated with the mutagen ethylmethane sulfonate (EMS), then allowed to grow to stationary phase overnight. The next day cells were diluted and upon reaching mid-log growth, were sorted again. However, this time 1 x 10^5 cells from the *dimmest* 5% of the population were sorted. This process was repeated for a total of 13 times; whereby each day the direction of selection alternated between high and low (See Figure 1). Control replicates underwent the same protocol, except a random selection of 1 x 10^5 cells (expressing any level of GFP) were sorted.

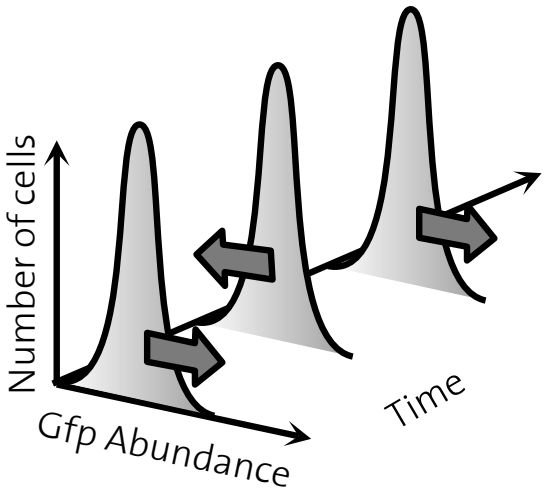


Figure 1. Cartoon scheme of selection.

Results of Pilot study

After 13 selective events, 24 individual clones were randomly chosen from the selected and control populations and from the ancestral population. These individual clones were grown to mid-log growth and the expression of GFP in 30,000 cells was measured using a flow cytometer (See Materials and Methods).

We observed a significant decrease in mean expression and increase in standard deviation in selected populations relative to the ancestor and control populations (stats here) (Figure 2). When using a simple metric for noise- the coefficient of variation (CV), our selected populations had a significant increase in CV ($p=0.026$, Wilcox test of median values of control and selected lines) and therefore seemed to respond as expected to our selection scheme. However, previous studies have shown that noise scales inversely with mean[15], and therefore it is difficult to conclude much about the observed changes in these evolved clones given that the mean expression changed so dramatically.

We chose to use competitive fitness (*w*) as a more accurate measure to test how selected and control populations responded to selection. During our selection experiment, competitive fitness of a mutant is determined by two factors. The first factor is the growth period between two consecutive FACS sortings. The second factors is the fraction of cells the mutant contributes to the fractions selected from the FACS sort – the 5% that show the highest GFP-concentration, and the 5% that show the lowest GFP concentration. If we assume that there are no growth differences between different genotypes, then competitive fitness is given by the second factor alone –by the number of cells a mutant contributes to the two selected fractions. As long as mutants are rare, then these fractions are determined by the ancestor. We thus calculated the competitive fitness of each selected and control clone as compared to the median ancestral GFP expression distribution (see Methods and Materials and Figure 3; a fitness of 1 indicates no change relative to the ancestor). Contrary to our expectations, clones from the selected populations had *lower* average fitness values than clones from the control populations. The average fitness values of the selected populations were also lower than 1, indicating that they had decreased fitness relative to the ancestor (Figure 3).

It is not clear how these selected clones were able to survive and compete against ancestral clones, especially during events of selection for high fluorescence. One possible explanation is that the growth period *between* selection events (which was not considered for our fitness measures) was important. For example, mutants that inactivate the second copy PrpsM upstream of gfp might have a growth advantage because the possibly minimize the titration of transcription factors for their native PrpsM. However, this effect would be expected to be equally important in the control lines; the difference between selected and control lines can thus not be explained by this effect. Also, such changes might have been recognized by the genetic analysis described below. A second hypothesis is that selection for low levels of

fluorescence was very efficient, so that the populations quickly became dominated by mutants with low levels of gfp expression.

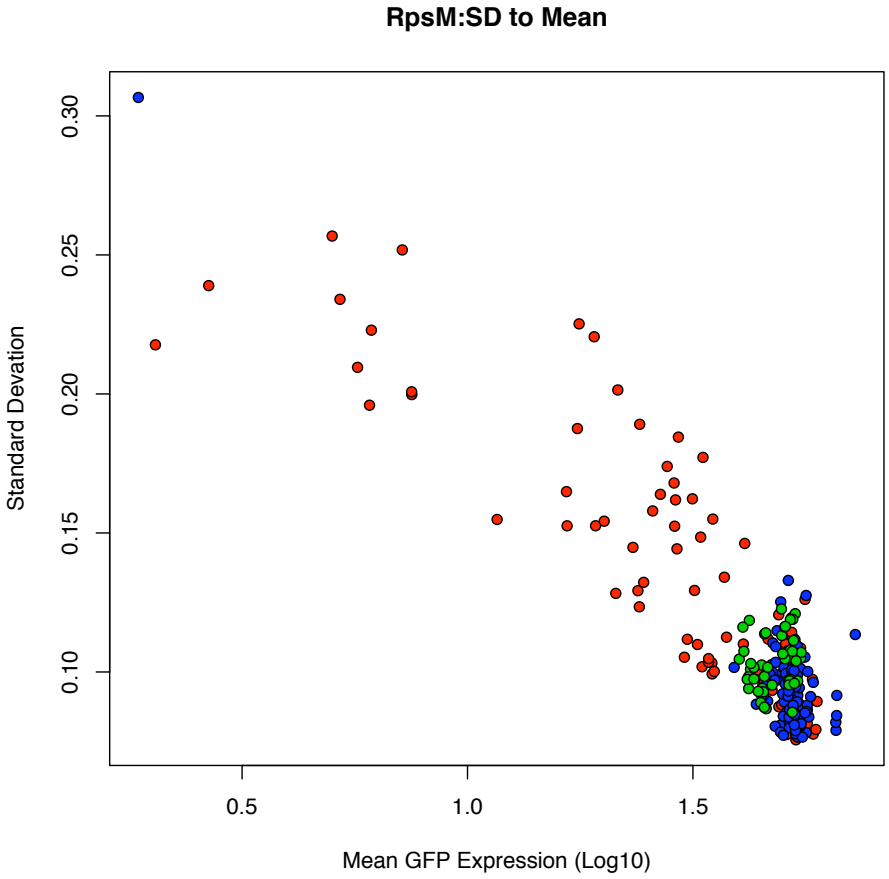


Figure 2:

Standard deviation versus mean expression for evolved (red), control (blue), and ancestral (green) clones.

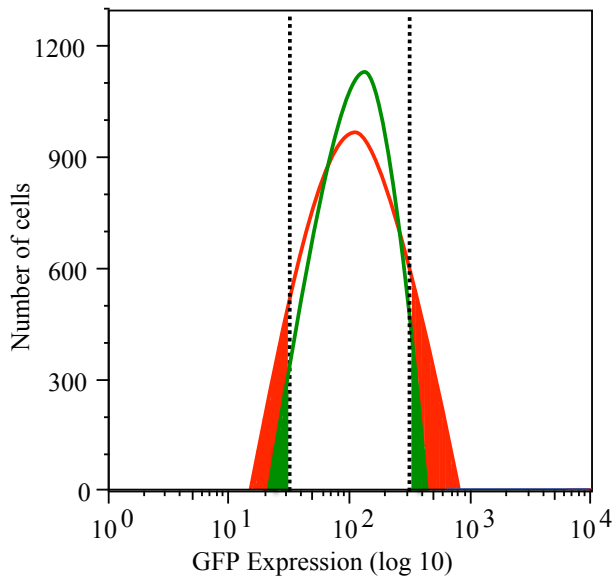


Figure 3: Illustration of fitness advantage of mutant with higher levels of noise in GFP expression.

A mutant clone with higher levels of noise (red) will have a fitness advantage over the resident ancestor population (green) when exposed to fluctuating selection. For example, on Day 1, the mutant clone has a 10 fold fitness advantage over the resident population. On Day 2, the mutant clone has a 5 fold fitness advantage. In total, the mutant clone will have a 50-fold (10*5) fitness advantage per two days over the resident population when exposed to the fluctuating selection as described here, assuming that mutant and resident grow equally fast between FACS selection.

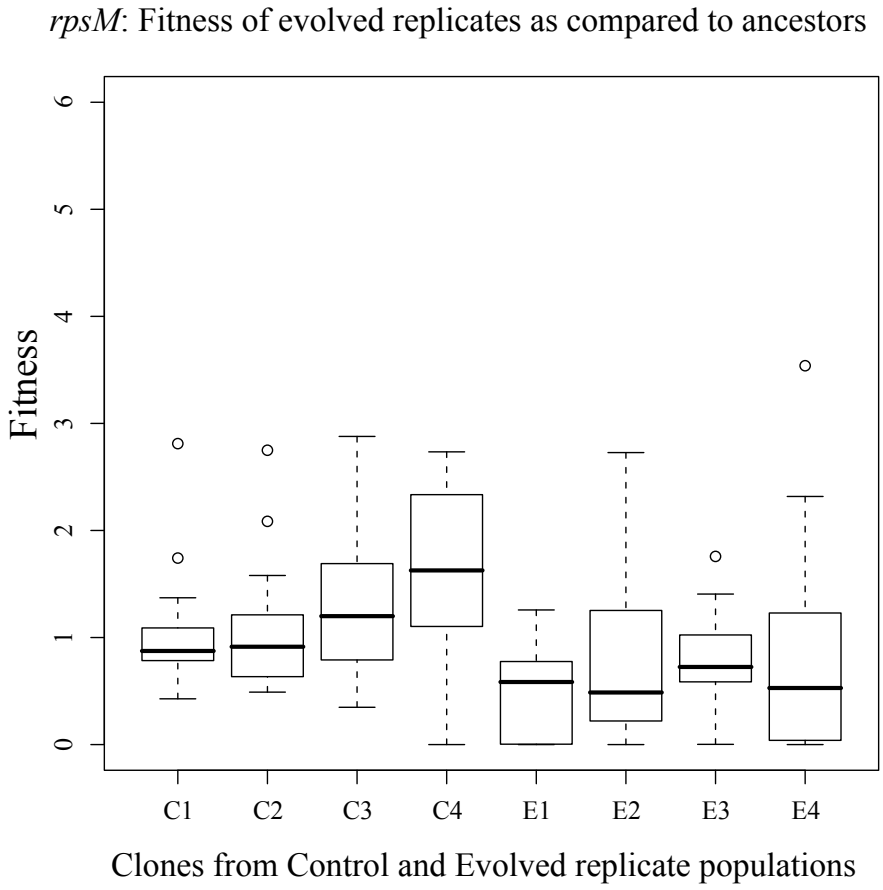


Figure 4: Fitness of evolved replicates as compared to ancestor for the P_{rpsM} promoter.

C1 through C4 are control replicates, while E1 through E4 are evolved replicates. Fitness of evolved replicates is below that of control replicates (p=0.02), indicating that the selection scheme was not effective in promoting mutants that contribute more cells to the 'high' and 'low' fractions than the ancestor.

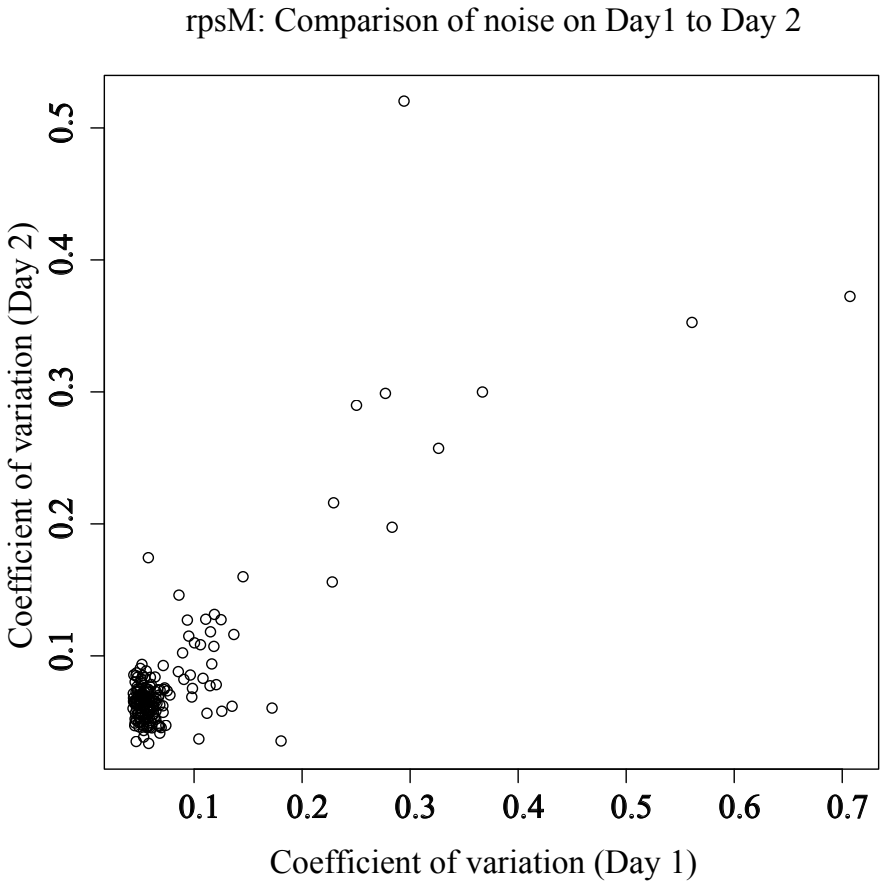


Figure 5: Repeatability of P_{rpsM} measurements on day 1 and day 2. Correlation between measurements of noise on two separate days is R^2 = 0.829, with 181 clones measured, indicating high level of heritability. There are approximately 15 generations between measurements on Day 1 and Day 2.

In order to better understand what caused the decreased expression and increased standard deviation in evolved clones (Figure 2) several steps were taken. First, all clones from the evolved and control populations were re-measured and re-analyzed to gauge reproducibility of the FACs measurement and heritability of the phenotype. The measurements were highly reproducible with an r^2 value of 0.829 (Figure 5). Secondly, the P_{rpsM} region and *gfp* of all selected clones (n = 94) were sequenced and analyzed for mutations. No mutations were found that were correlated with changes in expression (data not shown). To determine if the decrease in mean expression was due to changes within 43 kilobases of the P_{rpsM} region, P22 transductions were done on all selected (n=96), control (n=96), and ancestral (n = 48) clones. Clones were then re-measured by flow cytometry and GFP expression data collected (data not

shown). There was no significant difference in mean expression or CV between ancestral, selected, or control clones, indicating that mutations that caused the decreased expression and increase in CV were outside of the 43 kilobases transduced.

Main study

Building on the results of the pilot study, we decided to make several changes to the experimental design to improve the chance of evolving higher levels of noise in gene expression. Firstly, we increased the size of the experiment to include four promoters. The first three promoters we included control the expression of genes *gadB*, *hdeA*, and *rpoS*, from *Escherichia coli* and are linked to *gfp-mut2* on a plasmid[16]. We used these three promoters for several reasons. Firstly, we recognized that it might be prudent to use non-essential genes, because changes affecting regulation of or essential genes themselves can cause large detrimental fitness effects, and even non-viability[17]. While *rpsM* is not an essential gene, null mutants do show marked fitness decreases[18]. Secondly, each of these three promoters contained multiple transcriptional regulatory inputs, thus increasing the number of genomic targets on which selection can act. Finally, all three promoters have a medium level of expression[16] and thus can be more easily subjected to our selection process. The forth promoter used was from *S.* Typhimurium, and is involved in the control of *prgH*, a gene involved in pathogenecity and necessary for invasion of host tissues[19]. This construct is a chromosomal-based transcriptional fusion to GFP^{+} and has been shown to be expressed in a bimodal fashion in certain environmental conditions[13]. The number of selective events was increased to 15 for *prgH* and to 17 for *gadB*, *hdeA*, and *rpoS*. In addition, we changed the mutagen from EMS (primarily causes G:C to A:T transitions) to UV irradiation, known to cause a broader spectrum of mutations (frameshifts, base substitutions, and genetic rearrangements)[20]. The selection threshold for dim cells was relaxed, from 5% to 20%. This was done to limit selection for non-functional promoters and decreased expression. . For each promoter, we used four populations under fluctuating selection, and four control populations.

Results from main study

After 15 selection events on the *prgH* promoter and 17 on the *gadB*, *hdeA*, and *rpoS* promoters, 24 clones from each replicate were analyzed as well as 24 clones from each ancestral population. We analyzed the noise in expression of GFP, using the median value of CV for each population in order to account for psuedoreplication. We found no significant differences for any of the four genes in the coefficient of variation in GFP expression (gadB, p=0.67; rpoS, p=0.76; prgH p=0.06; hdeA, p=0.45), using a parametric test for significance.

Because of limited sample size, a non-parametric test, as used in the pilot study analysis, was not possible. Additional statistical tests can be found in the Appendix III. Additionally, we found no significant difference in fitness between selected and control clones for any of the four genes as compared to the median GFP expression distribution of the ancestral clones (Figure 6) (gadB, p=0.49; rpoS, p=0.95; prgH, p=0.57; hdeA, p=0.34). We found a low level of repeatability and heritability of in GFP expression between measurements on day 1 and day 2 (gadB r^2= -0.05, hdeA r^2= 0.23, rpoS r^2= 0.15, prgH r^2= 0.11, n=12 for each) (Figure 7), indicating that our experimental populations contained a limited number of different genotypes with consistent differences in the levels of phenotypic noise.

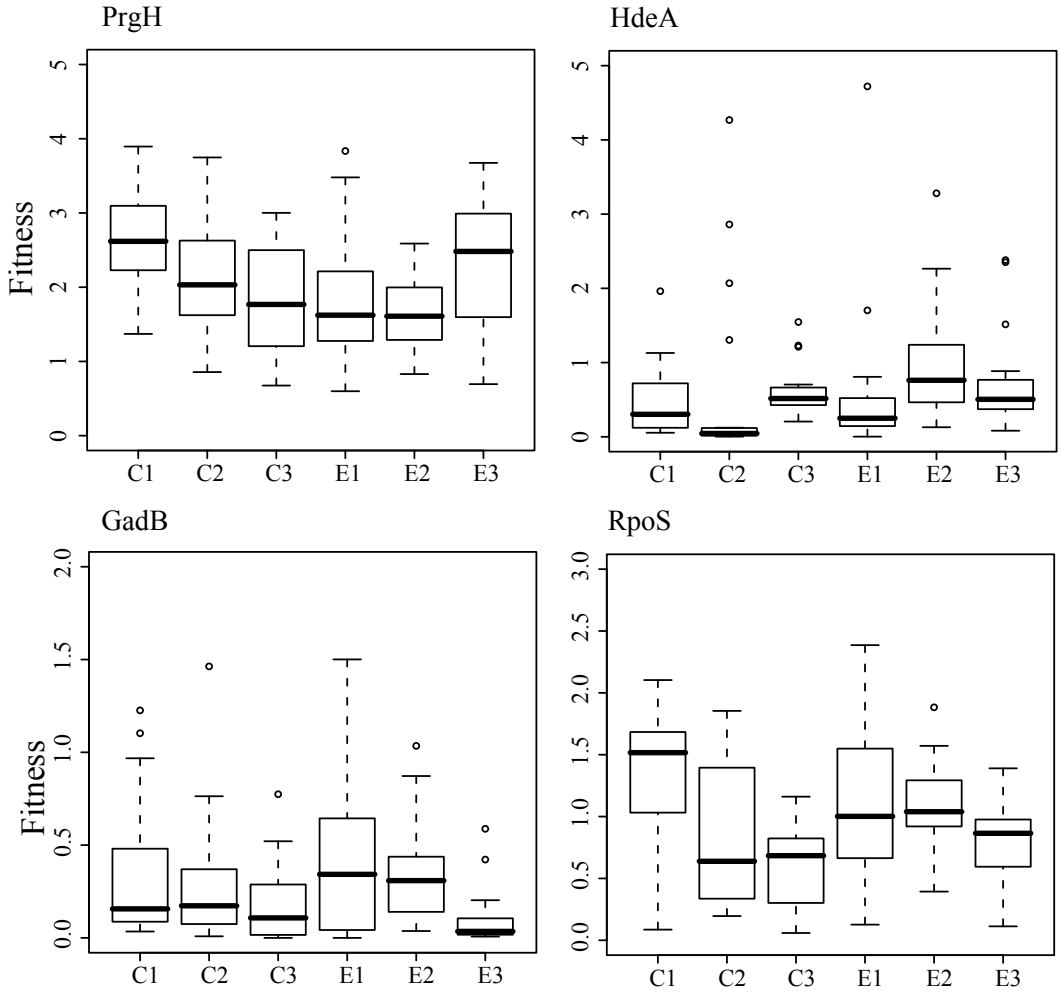


Figure 6: Fitness of selected and control replicates as compared to the ancestor.

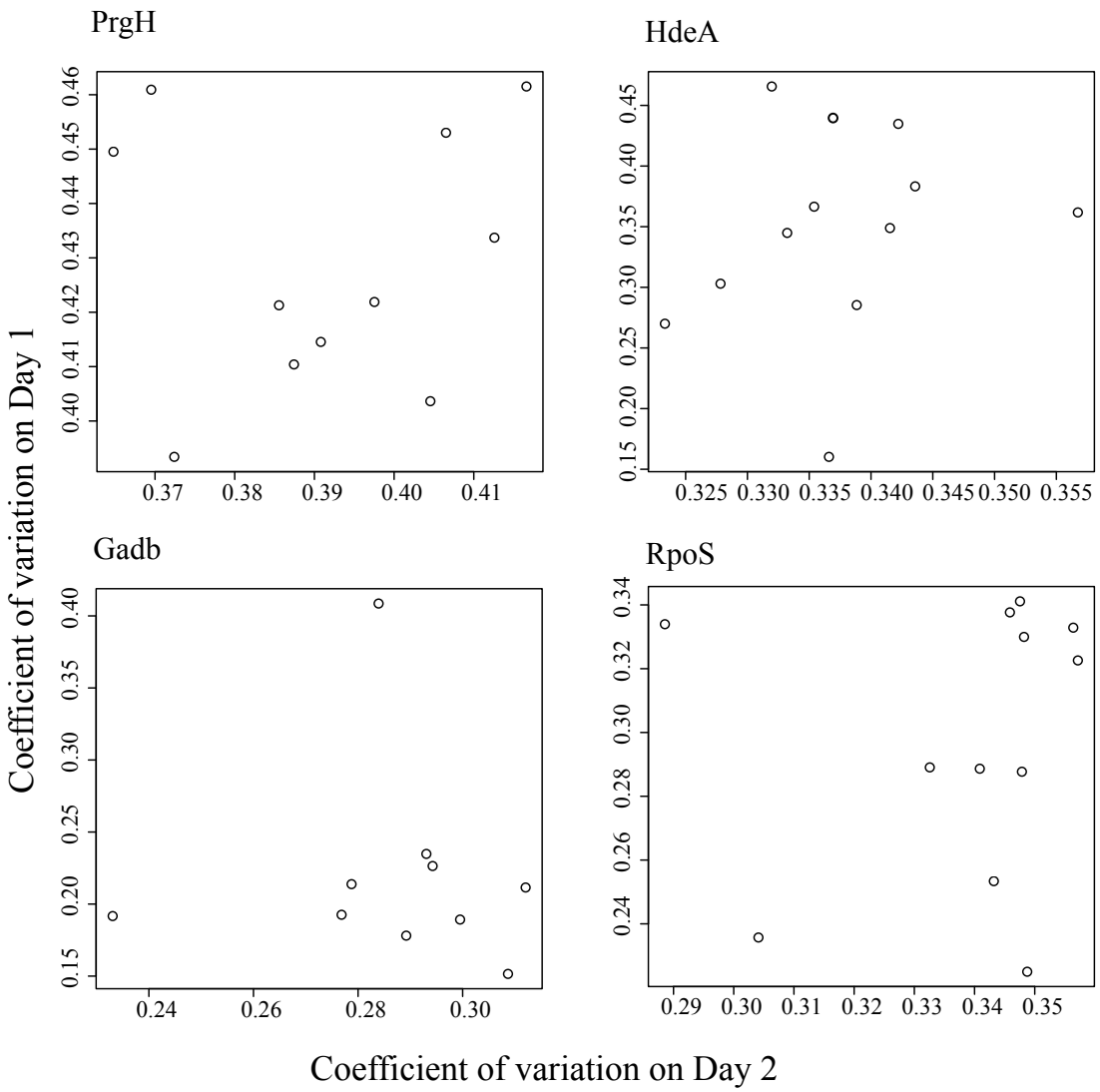


Figure 7: Repeatability of measurements on day 1 and day 2 in subset of clones (n=12) from each promoter in the main study. Correlation between measurements of noise on two separate days is (gadB r^2= -0.05, hdeA r^2= 0.23, rpoS r^2= 0.15, prgH r^2= 0.11) , indicating limited level of heritability. There are approximately 15 generations between measurements on Day 1 and Day 2.

Discussion

We selected five different promoters for increased levels of variation in gene expression, and found no consistent response. We used two metrics to quantify the response of selection – the coefficient of variation as a standard measure of noise in gene expression, and competitive fitness as a direct measure of the success of a mutant during fluctuating selection. Neither metric showed consistent evidence for an evolutionary change towards elevated levels of phenotypic noise.

There are two reasons that may explain why no significant change was observed: first, there was not enough genetic variation available for the action of selection to work upon, and second, that selection was neither long nor strong enough. Both explanations alone do not seem to be sufficient to fully explain our results for the following reasons. There are some heritable differences in noise levels (Figures 5 and 7) and it has been shown that the selection regime used in this study is effective for enriching for noisy promoters[14]. It is possible that given the relatively limited variation (as compared to the significant diversity selected present at the start of [14]) more selective events would be needed. In addition, we initially theorized that the four promoters used in the main study (P_{rpoS}, P_{gadB}, P_{hdeA}, and P_{prgH}) would be more amenable to the action of selection, given that each has several transcriptional regulators. We postulated that more transcriptional regulators translated into more genomic targets upon which selection could act. However, it could be that multiple inputs dampen noise generated from a single mutation and actually make it harder for selection to act. We conclude that lab evolution of increased variation in a promoter is not trivial. Other groups have had success using a FACS to evolve increased mean expression [21,22,23], however evolving increased variance in gene expression has proved to be more difficult. One of the main conclusions of our study is that gene expression is robust to changes in variance.

Materials and Methods

Strains used

Three of the five strains used in this study (P_{gadB}, P_{hdeA}, and P_{rpoS}) are from the *E. coli* plasmid promoter library from Zaslaver et al[16]. This plasmid contains a kanamycin resistance marker and strains were grown with 50 μg/ml Kanamycin at all times to ensure the plasmid was retained. The plasmids reside in E. coli K12 strain MG1655. The chromosomal-based transcriptional fusions of P_{rpsM} (used in the pilot study) and P_{prgH} to GFP^{+} (from strains

JH3016 and JH3010 respecitively[13]). These cassettes were moved via P22 transduction with selection for chloramphenicol to a biosafety level 1 strain M557 [24] for work in our lab, and were grown with 30 μg/ml chloramphenicol at all times. At the end of the evolution experiment, all clones used in this analysis were tested by specific PCR to ensure that no contamination between different strain replicates had taken place (i.e.: a P_{gadB} plasmid was not present in any clones from P_{hdeA} or P_{rpoS}, and vice versa).

Media and culture conditions

Strains were grown at 37ºC in 1 ml of liquid LB broth (Sigma, L3022) in 24 well flat bottom plates (Costar 354, Corning) or on plates containing LB agar (Sigma, L2897). To minimize slight variations in media content, all media for experimental evolution and analysis was made in a single large batch prior to the start of the experiment, autoclaved, appropriate antibiotic added, aliquots of 50ml were made, and subsequently frozen. Single aliquots were thawed and used for that day's experiments.

Experimental evolution: Strain growth

A single colony of each of the five strains (Table 1) was picked, grown up overnight, and split into equal populations; half of the population was assigned to "selected" and the other half to "control" replicates. Cells were diluted 1:1000 and grown for 3 hours to reach mid-log growth. Cultures were spun down at 6200 X g for five minutes at 4ºC. Growth media was removed and cultures were re-suspended in ice cold PBS. Cells were kept on ice until sorted or analyzed as described below.

Cell sorting, re-growth, and mutagenesis

We subjected each strain to fluctuating selection based on expression of GFP, where selection for bright cells alternated with selection for dim cells. Cells were sorted using fluorescence-activated cell sorting (FACS) as described[14]. For the first round of selection, a gate was drawn for each population to include either the highest 5% of cells expressing GFP, or a gate that covered the entire range of GFP expression, for selected and control lines, respectively. From each gated area, $1x10^5$ cells were collected into a sterile well of a 24-well plate. Cells were collected at a 1.0 flow rate and sorted with yield, purity, and phase masks set to 0, 32,

and 0 respectively to ensure that only cells matching the set GFP expression criteria were sorted. After sorting, cells were centrifuged and FACS buffer was replaced with 1ml LB media containing appropriate antibiotic and grown overnight. The following day, cultures were diluted 1:50 and grown 2 hours. Cultures were then exposed to mutagen (ethylmethane sulfonate (EMS) for pilot study, and 500 Joules/cm^3 UV for mutagenesis of strains in main study). 10ul of mutagenized culture was then diluted in 1ml of fresh LB media and allowed to grow to mid-log growth (3 hours). At this point, the cells were prepared for FACS sorting and the process was repeated; however the gates for the selected populations included only the *dimmest* cells expressing GFP (5% for pilot study, 20% for main study). This entire procedure was repeated for the specified number of selection events (13 for P_{rpsM} in the pilot study, 15 for P_{prgH}, 17 for P_{gadB}, P_{hdeA}, and P_{rpoS}). During weekends, when the FACS facility was closed, all cells were placed at 4ºC.

After the selection procedure was completed, all populations were plated on LB agar plates containing appropriate antibiotic, and 24 single colonies from each of the selected and control replicates were randomly chosen and grown up. As well, 24 clones from each ancestor population were plated and grown up, for a total of 249 clones used for analysis in the pilot study and an additional 672 clones for the main study. All clones were grown overnight at 37ºC in 96 well plates in 200ul of LB containing appropriate antibiotic and frozen at -80ºC in 15% glycerol.

Analysis and Data processing

One day prior to analysis, a subset of frozen clonal stocks was used to inoculate 200ul of LB medium and prepared in the same manner as described above (Growth for cytometry and cell sorting). For each clone, $3x10^4$ cells were analyzed for GFP expression on the FACS Calibur (BD, CA).

The following conventions were applied to calculate variation in GFP expression and to limit the influence from cellular aggregates, cell detritus, and undefined values. Modified from Newman et al [4] and Freed et al[14]. R scripts used for all analyses are given in the Appendix III.

1. All SSC, FSC, and fluorescence zero values were excluded.
2. Data was excluded that fell within the forward scatter (FSC) and side scatter (SSC) region where significant counts appeared in “buffer only” controls.

3. Extreme values of FSC and SSC were excluded (the highest and lowest 1.5% of events) from total counts to limit influence from cell detritus and cell aggregates. Extreme values of the fluorescence channel (FL1) (the highest 1.0% of events) were excluded to limit only a very small number of cells having undue effects on the values of the mean and CV.
4. The densest area of the FSC and SSC channels were calculated and an elliptical gate applied to include 5000 cells. GFP fluorescence values are $\log_{10}$ transformed. The mean, standard deviation, and coefficient of variation (CV) were calculated for fluorescence. This was done to ensure that cells being included in the calculations are of similar size, shape, and in a similar growth stage.

All clones from the pilot study were regrown from frozen stocks and measured on a separate day to examine repeatability of measurements and heritability of noise, as measured by coefficient of variation (Figure 5). Similarly, a subset of clones from the four selected populations in the main study (62 clones in all) were also grown on a second, separate day to repeatability of measurements and heritability of noise, as measured by coefficient of variation (Figure 7).

Fitness estimates

To obtain a measure of fitness of selected clones, we compare the expression of GFP from a selected clone to the median expression distribution its ancestor and ask, given our alternating selection regime, how many cells would survive? More specifically 20,000 cells from each clone are randomly chosen and a distribution of GFP expression is constructed. The median GFP distribution of the ancestral clones is constructed and fitness is calculated by:

$$\omega = \frac{N_{AN\ high}-N_{SEL\ high}}{N_{AN\ high}} * \frac{N_{AN\ low}-N_{SEL\ low}}{N_{AN\ low}}$$

$N_{SEL\ high}$ as the number of selected cells in the high gate, $N_{SEL\ low}$ as the number of selected cells in the low gate. $N_{AN\ high}$ and $N_{AN\ low}$ are the number of ancestral cells in high and low gates, respectively.

Acknowledgments

Thanks to Alex Boehm for providing technical advice, to the ETH Zürich FACS sorting facility for help with cell sorting and analysis, to Alon Zaslaver for information regarding the Uri Alon plasmid promoter library, to Wolf-Dietrich Hardt and Bärbel Stecher for salmonella strains and P22 lysates.

References

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic Gene Expression in a Single Cell. pp. 1183-1186.

2. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. Nature Reviews Genetics 6: 451-464.

3. Kaufmann B, van Oudenaarden A (2007) Stochastic gene expression: from single molecules to the proteome. Current opinion in genetics & development 17: 107-112.

4. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. Nature 441: 840-846.

5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genetics 31: 69-73.

6. Smits W, Kuipers O, Veening J (2006) Phenotypic variation bin bacteria: the role of feedback regulation. Nature Reviews Microbiology 4: 259-271.

7. Acar M, Mettetal J, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. Nature genetics 40: 471-475.

8. Salathe M, Van Cleve J, Feldman M (2009) Evolution of Stochastic Switching Rates in Asymmetric Fitness Landscapes. Genetics.

9. Fraser D, Kaern M (2009) A chance at survival: gene expression noise and phenotypic diversification strategies. Molecular Microbiology 71: 1333-1340.

10. Ackermann M, Stecher B, Freed NE, Songhet P, Hardt WD, et al. (2008) Self-destructive cooperation mediated by phenotypic noise. Nature 454: 987-990.

11. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. Plos Biology 2: 834-838.

12. Scholz O, Thiel A, Hillen W, Niederweis M (2000) Quantitative analysis of gene expression with an improved green fluorescent protein. European Journal of Biochemistry 267: 1565-1570.

13. Hautefort I, Proenca MJ, Hinton JCD (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. Applied and Environmental Microbiology 69: 7480-7491.

14. Freed N, Silander O, Stecher B, Böhm A, Hardt W, et al. (2008) A Simple Screen to Identify Promoters Conferring High Levels of Phenotypic Noise. PLoS Genetics 4.

15. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, et al. (2006) Noise in protein expression scales with natural protein abundance. Nature genetics 38: 636-643.

16. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, et al. (2006) A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nature methods 3: 623-628.

17. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology 2.

18. Bubunenko M, Baker T (2007) Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in Escherichia coli. Journal of bacteriology 189: 2844.

19. Pegues D, Hantman M, Behlau I, Miller S (1995) PhoP/PhoQ transcriptional repression of Salmonella typhimurium invasion genes: evidence for a role in protein secretion. Molecular Microbiology 17: 169-181.

20. Miller J (1992) A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria: Cold Spring Harbor Laboratory Pr.

21. Cormack B, Valdivia R, Falkow S (1998) FACS-optimized mutants of the green fluorescent protein (GFP). Google Patents.

22. Ito Y, Kawama T, Urabe I, Yomo T (2004) Evolution of an arbitrary sequence in solubility. Journal of molecular evolution 58: 196-202.

23. Wang L, Jackson W, Steinbach P, Tsien R (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. Proceedings of the National Academy of Sciences 101: 16745-16749.

24. Hapfelmeier S, Stecher B, Barthel M, Kremer M, Muller AJ, et al. (2005) The Salmonella Pathogenicity Island (SPI)-2 and SPI-1 Type III Secretion Systems Allow Salmonella Serovar typhimurium to Trigger Colitis via MyD88-Dependent and MyD88-Independent Mechanisms 1. The Journal of Immunology 174: 1675-1685.

Chapter 5

Effect of *flgM* Expression on Variation in *fliC* Expression in Salmonella

Nikki E. Freed[1], Olin K. Silander[1], Kelly Hughes[2], Martin Ackermann[1]

[1]Institute of Biogeochemistry and Pollutant dynamics, ETH Zurich, Zurich Switzerland

[2]Department of Biology, University of Utah, Salt Lake City, Utah, USA

Abstract

We investigated noise in the expression of the flagellar filament gene *fliC* in the bacterial pathogen Salmonella enterica serovar Typhimurium. We found that the heterogeneous expression of the flagellar filament *fliC* is dependent on a functional anti-sigma factor FlgM. Based on these results we hypothesize that the heterogeneity in *fliC* may be, in part, the result of variations in the equilibribrium of FlgM and its target of inhibition, the alternative sigma factor σ^{28}, encoded by the *fliA* gene.

Introduction

The flagellar organelle in Salmonella enterica serovar Typhimurium (*S.* Typhimurium) consists of three structural parts (basal body, hook, and filament) and is formed by the complex expression of over 60 genes[1]. The genes that make up the structural and regulatory components are subdivided into three promoter classes (Early, Middle, and Late) based on transcriptional hierarchy[2]. Briefly, at the top of this hierarchy lies the master regulator of flagellar synthesis, the *flhDC* operon[3]. The two proteins expressed from this operon, FlhD and FlhC, form a complex that activates the production of the Middle promoters. The Middle promoters produce structural proteins that make up the intial complex of the flagella- the basal body and hook. In addition to structural genes, several regulatory genes are also transcribed from the Middle promoter class, including *fliA* and *flgM*[4]. FlgM acts to repress the protein product of *fliA*, the alternative sigma factor σ^{28}. However, once the basal body and hook structures are complete, FlgM is secreted out of the cell. Once the intracellular level of FlgM is decreased, σ^{28} is no longer inhibited and RNA polymerase is recruited to initiate the expression of the Late promoters. These Late promoters produce the final flagellar protein products, like FliC, that complete the tail-like filament of the flagella.

Recent reports indicate that some Late promoters, notably *fliC* and *flgK* have noisy expression patterns[5,6]. In particular, the *fliC* gene shows high levels of noise and has been shown to switch from on to off and back on again in under 90 minutes [6]. Noise in gene expression in *fliC* occurs in clonal populations growing in constant environments and can be measured using the fluorescent reporter Green Fluorescent Protein (GFP) and flow cytometry or

microscopy. The precise molecular causes underlying the dynamic expression of *fliC* are not known. For example, it is not clear if the origin of the noise in expression is intrinsic to the gene itself [7] or if the noise originates upstream in the flagellar sythesis pathway. However, it is known that there is a balance in the intracellular concentration of FlgM and σ^{28} [8], and we thought that stochastic shifts in the concentration either protein may lead to noise downstream. Lending some support to this hypothesis, the regulation of FlgM and σ^{28} expression is structured as a feed forward loop, a network motif that is known to amplify noise in gene expression; σ^{28} increases expression of basal body proteins, which export FlgM out of the cell and thus relieve inhibition of σ^{28} by FlgM[9]. In addition, there are two other feed forward loops in the flagellar synthesis pathway[10]. In order to further investigate the contribution of *flgM* and *fliA* to noise in *fliC* expression, we used mutants of *flgM* and *fliA* and examined subsequent noise in expression of *fliC*.

Results and Discussion

We investigated the effects of disrupting the protein concentrations of FlgM and σ^{28} on noise in expression of *fliC*. We used three mutants in this study: 1) a mutant σ^{28} that is resistant to proteolysis, yet has wildtype activity in its interaction with FlgM and RNA polymerase (fliA5225(H14D))[9]. 2) A mutant σ^{28} that is defective in FlgM binding and therefore not inhibited by FlgM (fliA5230(N114K))[11], and 3) a *flgM* deletion mutant [3]. In essence, all three mutants should have an excess intracellular concentration of σ^{28} to FlgM.

It is important to note that *S.* Typhimurium can express two distinct flagellin proteins, FliC or FljB, a process called flagellar phase variation [12]. The expression either of the two flagellar types is mediated by a site-specific DNA inversion. This recombination event occurs rarely, at a rate of 10^{-3} to 10^{-5} per cell division [13,14,15], however, it is important to ensure that our results are not influenced by this phase variation, so all strains were placed, via P22 transduction, into a strain background that is defective for *hin*-recombination[12] (See Methods and Materials).

We found that for all three mutants investigated, an excess of σ^{28} to FlgM, eliminated the noise in *fliC* expression (Figure 1). Previous studies have shown that the expression of the *FliC* protein is heterogeneous[6,16]. Here we show that this heterogeneity is dependent on functional expression of *flgM*.

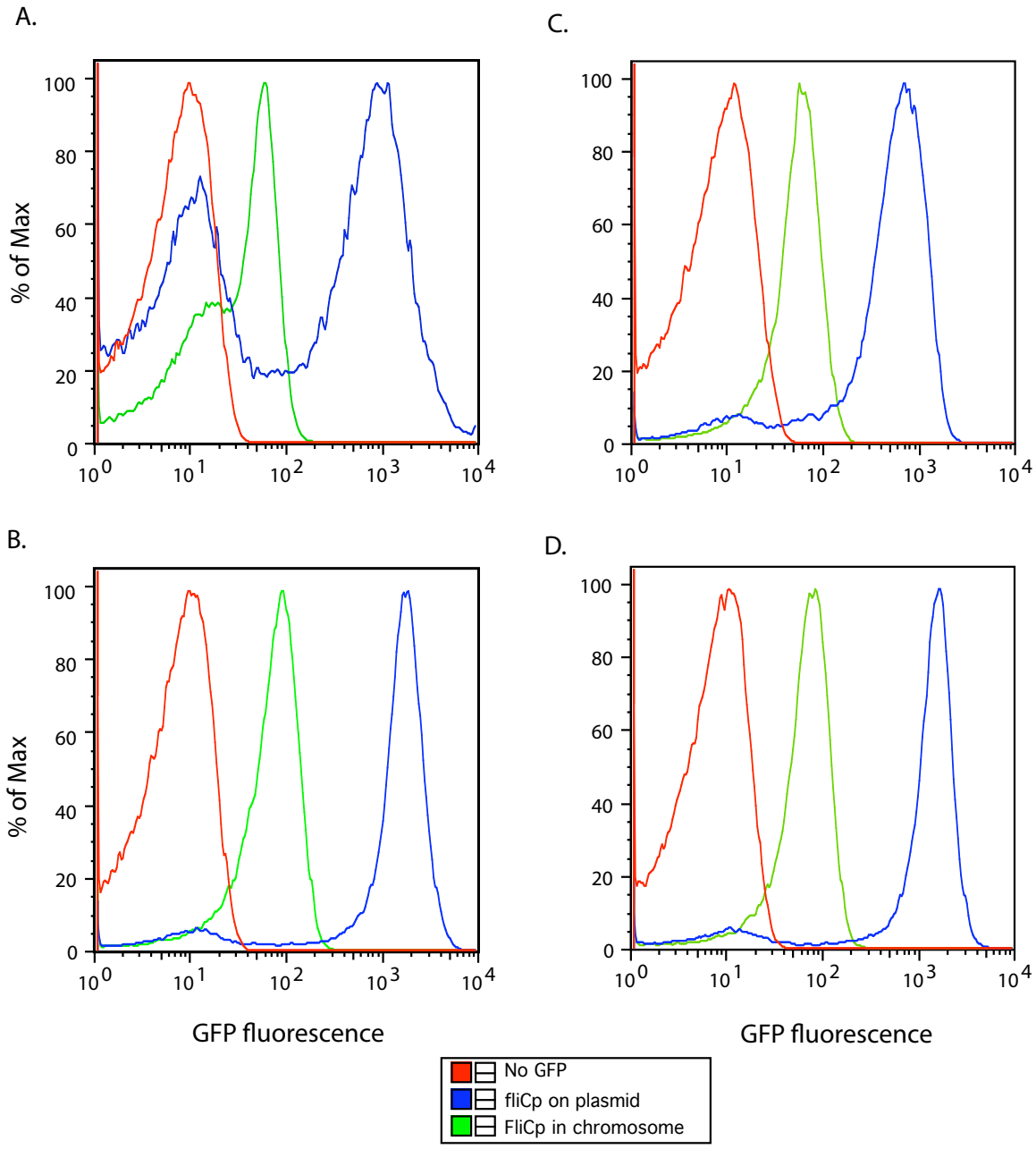


Figure 1. Decreased noise in expression of the *fliC* promoter in mutants with an excess σ^{28} to FlgM.

Histograms of GFP expression driven by the *fliC* promoter in plasmid pM968 (in blue) and from a chromosomal insertion (in green) at the native location, and the negative control lacking GFP (in red). A. *fliC* expression is noisy in NF003. B. Noise in fliC expression is decreased in strain NF8208, where σ^{28} is resistant to proteolysis, C. Noise in fliC expression is decreased in strain NF14827, a FlgM deletion mutant, and D. Noise in fliC expression is

decreased in strain NF14826, a σ^{28} mutant with decreased inhibition by FlgM.

We found that for all three mutants investigated, the excess of of σ^{28} to FlgM, eliminated the noise in *fliC* expression (Figure 1). Previous studies have shown that the expression of the *FliC* protein is heterogeneous[6,16]. Here we show that this heterogeneity is dependent on functional expression of *flgM*.

Materials and Methods

Strains and growth conditions

Media, growth conditions and transductional methods, were as described in Chapter 4. Strains were grown at 37ºC in 1 ml of liquid LB broth (Sigma, L3022) in 24 well flat bottom plates (Costar 354, Corning) or on plates containing LB agar (Sigma, L2897). The transducing phage of *S. enterica* serovar Typhimurium P22 was used in all transductions (ref). All genetic constructs were done in the strain background of the biosafety level 1 strain M556[17] for work in our lab. All strains are listed in Table 1.

Table of strains			
Name	**Description**	**Function**	**Reference**
M556	*sseD::aphT*	SPI-1 +, SPI2 -	[17]
NF003	Δhin-5717::FCF, sseD::aphT	Locked on for fljBA	this work
NF8208	fliA5225(H14D) Δhin-5717::FRT, sseD::aphT	mutant s^{28} resistant to proteolysis	this work
NF14827	fliA5230(N114K) Δhin-5717::FRT, sseD::aphT	s^{28} defective in FlgM binding	this work
NF14826	ΔflgM5628::FRT Δhin-5717::FRF, sseD::aphT	*flgM* deletion mutant	this work
P*fliC*-pM968	Amp^r; oripBR322; gfpmut2 wtih fliC promoter	plasmid based PfliC- GFP	[18]
PfliC-GFP^+	ΔfliC::GFP^+-cat cassette used for P22 transd.	chromosomal based P*fliC*- GFP	[6]

Table 1: List of strains.

Flow cytometry

Two days prior to analysis, a streak from frozen clonal stocks was made on LB agar plates containing appropriate antibiotic. Three individual colonies from each strain were chosen and each was inoculated 1ml of LB medium and grown overnight. On the day of analysis, overnight cultures were diluted 1 to 1000 and clones were grown for 2 hours to mid-log growth. All strains were measured in triplicate, with representative clones shown in Figure 1. For each clone, $5x10^4$ cells were analyzed for GFP expression on the FACS Calibur (BD, CA). All strains were gated in the same manner to exclude counts from buffer and to limit the influence from cellular aggregates, cell detritus, and undefined values.

References

1. Rosu V, Chevance F, Karlinsey J, Hirano T, Hughes K (2006) Translation inhibition of the Salmonella fliC gene by the fliC 5'untranslated region, fliC coding sequences, and FlgM. Journal of bacteriology 188: 4497.

2. Macnab R (1992) Genetics and biogenesis of bacterial flagella. Annual review of genetics 26: 131-158.

3. Frye J, Karlinsey J, Felise H, Marzolf B, Dowidar N, et al. (2006) Identification of new flagellar genes of Salmonella enterica serovar Typhimurium. Journal of bacteriology 188: 2233-2243.

4. Gillen K, Hughes K (1993) Transcription from two promoters and autoregulation contribute to the control of expression of the Salmonella typhimurium flagellar regulatory gene flgM. Journal of bacteriology 175: 7006-7015.

5. Cummings L, Wilkerson W, Bergsbaken T, Cookson B (2006) In vivo, fliC expression by Salmonella enterica serovar Typhimurium is heterogeneous, regulated by ClpX, and anatomically restricted. Molecular Microbiology 61: 795-809.

6. Freed N, Silander O, Stecher B, Böhm A, Hardt W, et al. (2008) A Simple Screen to Identify Promoters Conferring High Levels of Phenotypic Noise. PLoS Genetics 4.

7. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297: 1183-1186.

8. Kutsukake K, Iino T (1994) Role of the FliA-FlgM regulatory system on the transcriptional control of the flagellar regulon and flagellar formation in Salmonella typhimurium. Journal of bacteriology 176: 3598-3605.

9. Rosu V, Hughes K (2006) {sigma} 28-Dependent Transcription in Salmonella enterica Is Independent of Flagellar Shearing. Journal of bacteriology 188: 5196.

10. Kalir S, Mangan S, Alon U (2005) A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. Molecular Systems Biology 1.

11. Chadsey M, Hughes K (2001) A multipartite interaction between Salmonella transcription factor 28 and its anti-sigma factor FlgM: implications for 28 holoenzyme destabilization through stepwise binding. Journal of Molecular Biology 306: 915-929.

12. Bonifield HR, Hughes KT (2003) Flagellar Phase Variation in Salmonella enterica Is Mediated by a Posttranscriptional Control Mechanism. Journal of Bacteriology 185: 3567-3574.

13. Zieg J, Simon M (1980) Analysis of the nucleotide sequence of an invertible controlling element. Proc Natl Acad Sci US A 77: 4196-4200.

14. Stocker BAD (1949) Measurements of Rate of Mutation of Flagellar Antigenic Phase in Salmonella typhi-murium. The Journal of Hygiene 47: 398-413.

15. Gillen KL, Hughes KT (1991) Negative regulatory loci coupling flagellin synthesis to flagellar assembly in Salmonella typhimurium. J Bacteriol 173: 2301-2310.

16. Cummings LA, Wilkerson WD, Bergsbaken T, Cookson BT (2006) In vivo, fliC expression by Salmonella enterica serovar Typhimurium is heterogeneous, regulated by ClpX, and anatomically restricted. Molecular Microbiology 61: 795-809.

17. Hapfelmeier S, Stecher B, Barthel M, Kremer M, Muller AJ, et al. (2005) The Salmonella Pathogenicity Island (SPI)-2 and SPI-1 Type III Secretion Systems Allow Salmonella Serovar typhimurium to Trigger Colitis via MyD88-Dependent and MyD88-Independent Mechanisms 1. The Journal of Immunology 174: 1675-1685.

18. Stecher B, Hapfelmeier S, Muller C, Kremer M, Stallmach T, et al. (2004) Flagella and Chemotaxis Are Required for Efficient Induction of Salmonella enterica Serovar Typhimurium Colitis in Streptomycin-Pretreated Mice. Infection and Immunity 72: 4138.

Chapter 6: General Discussion

This thesis deals with the phenomenon of phenotypic variation in gene expression in bacteria and explores the mechanistic and evolutionary causes and consequences of that variation. Specifically, I have concentrated on noise in gene expression in the bacteria *S.* Typhimurium and *Escherichia coli*. A major aim of this thesis was to explore which genes are noisy in salmonella and to show the evolutionary advantage that variation in expression might provide for this pathogenic bacterium.

It has been known for some time that clonal populations of bacterial cells growing in the same environment do exhibit variation [1]. Recent advances in single-cell imaging, flow cytometry, and fluorescent reporters have allowed an abundance of new experimental and theoretical investigations into the phenomenon of noise in gene expression [2,3,4]. It has clearly been shown that noise does exist and that it is differential; for a given environment and a given gene there exists a certain, experimentally repeatable, level of noise [3,5]. The mechanism underlying some of this variation is known, and can be altered experimentally by changes to the DNA sequence in the promoter region [6]. As well, it is known that certain classes of genes are more noisy than others; essential genes have low noise levels[7], while genes involved in response to environmental stimuli and stress show more variation[3].

In Chapter 2 of this thesis I set out to investigate, using a novel method, which genes in *S.* Typhimurium were noisy. Using fluctuating selection with a cell sorter, I screened for genes that are variable on the time scale of one day from a library of *S.* Typhimurium promoters linked to GFP. I found that this alternating selection process is successful for enriching for promoters with significantly higher levels of noise than those from the control replicates. Upon sequencing promoters isolated in the screen, I found that genes involved in flagellar synthesis, namely *fliC* and *flgK*, have higher rates of switching from the on to the off state and back again than the other promoters isolated. In addition, I showed that the variation in expression for the 240 clones investigated is experimentally repeatable to a large degree (r^2 of 0.748) indicating consistent differences between promoters in the level of noise. Using time-lapse microscopy and cell tracking software, we were able to explore the rates of switching in *fliC* and observe how the 'On' and 'Off' states are propagated through the cell lineage as it grows from one cell into a small population. We hypothesize that this rapid switching in flagellar genes might be important for salmonella pathogenesis in a host environment and discuss a theoretical model that could explain the biological relevance of this phenomenon further in Chapter 3 (discussed below).

Investigations into the occurrence of and mechanistic causes of stochasticity in gene expression have been extensive. However fewer studies have dealt with the possible biological significance of noise in protein production[2,4,5,6,8,9,10,11]. This is due in part to the difficulty in experimentally demonstrating that noise in certain genes is beneficial (or detrimental) to an organism's fitness. The main hypothesis for an adaptive function of phenotypic noise is that it might promote a genotype's long-term survival and growth in fluctuating environments. The continuous production of phenotypic variants might increase the chance that some of these variants survive sudden environmental shifts, allowing the genotype to persist. This hypothesis is known as "bet-hedging". Specific examples that are suggestive of such "bet-hedging" include bacterial persistence, competence, and galactose utilization in yeast [4,12,13,14,15,16,17].

We proposed an alternative explanation to bet hedging as an adaptive biological function of noise in S. Typhimurium in Chapter 3. We suggest that the heterogeneity in the expression of virulence genes in this pathogen [18,19] may lead to the formation of different subpopulations that work together to establish an infection. The phenotype that expressed virulence factors produces a public good - a gut inflammation - that benefits the phenotype that does not express virulence factors. Individuals of the first phenotype are unable to benefit from that public good, because they are killed by the innate immune system of the host. According to this interpretation, noise thus promotes a special form of a division of labor, a form where one phenotype sacrifices itself in order to provide a benefit to the second phenotype.

We used a mathematical model coupled with experimental results to analyze this hypothesis in more detail. We provided experimental results that indicate that for S. Typhimurium infection in a mouse model, the central features of the model are captured. We experimentally show that there is heterogeneity in expression of the type three secretion system-1 (TTSS-1) virulence factors in the gut, and by increasing the number of TTSS-1+ cells, there is increased public good (gut inflammation). In addition, we showed the expression of TTSS-1+ increases the chance of being killed by the innate immune system of the mouse. The model indicates that the evolution of this type of cooperative behavior hinges on assortment, under which genotypes that express the cooperative phenotype with a certain probability have a greater chance of interacting each other, and thus have an increased chance to access the public good. We illustrated this mathematically, and showed that when the number of bacteria entering a given host is small (so that infections are almost clonal), cooperation mediated by phenotypic variation can arise and be maintained. This scenario holds true for salmonella infections,

where low inoculum sizes have been indicated[20]. Given that both phenotypic noise and self destructive cooperation also occur in other bacterial pathogens [21], we believe this model may explain some aspect of how bacterial pathogens might employ phenotypic noise in pathogenesis.

Building on the theoretical and experimental results described in Chapters 2 and 3, we set out to explore if we could experimentally demonstrate that selection can act to increase phenotypic variation in gene expression. Based on the method described in Chapter 2, we used laboratory-based evolution with five bacterial promoters linked to GFP to try and evolve higher levels of noise. After up to 17 rounds of fluctuating selection using a cell sorter, we measured the expression of GFP in individual clones from each replicate population. We found no significant differences in GFP expression or fitness in selected or control clones. We conclude that two main factors could have contributed to the failure to select for mutants with higher levels of noise: 1) the selection scheme did not work, and 2) that there was not enough genetic variation within the experimental populations. We have shown before that the selection scheme is efficient at enriching for noisy promoters (Chapter 2), and we found there are, albeit low, heritable differences between clones in the variation in GFP expression. This suggests that increasing the number of selection steps, or the genetic variation in the population, might increase the chance to successfully select for higher levels of noise in individual promoters.

Finally, we presented a short investigation into the molecular mechanisms behind the high level of noise observed in the *S.* Typhimurium gene *fliC* in Chapter 5. There are more than 60 genes involved in the formation of a functional flagella [22], and they can be further subdivided into three classes based on transcriptional hierarchy[23]. The expression of *fliC*, which lies in the third and final class of genes to be expressed in the formation of the flagella, is positively controlled by the class 2 alternative sigma factor *fliA* (σ^{28}). In turn, *fliA* expression is repressed by the anti-sigma factor FlgM. It is thought that the relative concentration of FlgM and FliA proteins in the cell may be responsible for the changes in expression of *fliC*[23]. We hypothesized that a shift in the balance of σ^{28} and FlgM may play a role in the downstream heterogeneity we observe in *fliC* expression. We found that disruption of FlgM/σ^{28} concentrations did eliminate variation in *fliC* expression as proposed. We would like to examine more precisely the role of *fliA* and *flgM* in pathogenesis, investigate when changes to the relative concentration of each protein in a cell might naturally occur, and more fully monitor other flagellar genes for heterogeneity.

These initial results suggest a number of follow-up studies for investigating noise in the expression of flagella and other virulence factors in *S*. Typhimurium. In a first step, it would be informative to measure relative concentration of FliA and FlgM in single cells, and more fully monitor other downstream flagellar genes for heterogeneity. A second step would be to examine how flagellar and pathogenecity genes in *S*. Typhimurium are co-expressed. Specifically, we have observed variation in *fliC* (involved in flagellar synthesis) and *prgH* (a pathogenecity factor). It would thus be interesting to know if the subset of cells that express *fliC* at the same time also express *prgH*. This might indicate that the two corresponding pathways are linked.

Future directions

The past ten or more years have seen an explosion of interest in the phenomena of stochastic phenotypic noise in gene expression and much has been learned about the mechanisms and possible biological relevance behind it. Recent technological advances, such as single cell transcriptomics[24], will provide a wealth of new information regarding the global level of noise in a single cell, given that previous studies have been limited to observations of one or two genes at a time. In addition, details concerning how noise is transmitted through genetic circuits, how noise changes in static versus dynamic environments, and how cells might exploit or control stochastic variation in expression may be elucidated with this new techniques.

Much remains to be discovered concerning how commonly employed phenotypic versus genotypically encoded variation occurs, and concerning the situations under which organisms recruit one mechanism versus the other. There is limited knowledge concerning noise in non-model systems, and it would be interesting to investigate whether these organisms use noise in order to hedge their bets in fluctuating environments, or as a means to produce different phenotypes that engage in the division of labor. Comparing organisms living in fluctuating environments versus constant environments will allow us to address the first hypothesis of bet-hedging, while investigating direct interactions between different phenotypes would address the role of noise in promoting the division of labor in clonal populations. Such studies will hopefully shed more light on whether noise in gene expression can give organisms new biological functions, and what these functions might be.

References

1. Spudich J, Koshland D (1976) Non-genetic individuality: chance in the single cell.

2. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297: 1183-1186.

3. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. Nature 441: 840-846.

4. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences 98: 8614.

5. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, et al. (2006) Noise in protein expression scales with natural protein abundance. Nature genetics 38: 636-643.

6. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genetics 31: 69-73.

7. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. Plos Biology 2: 834-838.

8. Austin D, Allen M, McCollum J, Dar R, Wilgus J, et al. (2006) Gene network shaping of inherent noise spectra. Nature 439: 608-611.

9. Pedraza J, van Oudenaarden A (2005) Noise propagation in gene networks. American Association for the Advancement of Science. pp. 1965-1969.

10. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. Science 307: 1962-1965.

11. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences of the United States of America 99: 12795-12800.

12. Balaban N, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. American Association for the Advancement of Science. pp. 1622-1625.

13. Acar M, Mettetal J, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. Nature genetics 40: 471-475.

14. Dubnau D, Losick R (2006) Bistability in bacteria. Molecular Microbiology 61: 564-572.

15. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. Science 317: 526-529.

16. Acar M, Becskei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. Nature 435: 228-232.

17. Fraser D, Kaern M (2009) A chance at survival: gene expression noise and phenotypic diversification strategies. Molecular Microbiology 71: 1333-1340.

18. Cummings L, Wilkerson W, Bergsbaken T, Cookson B (2006) In vivo, fliC expression by Salmonella enterica serovar Typhimurium is heterogeneous, regulated by ClpX, and anatomically restricted. Molecular Microbiology 61: 795-809.

19. Hautefort I, Proenca MJ, Hinton JCD (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during

infection of mammalian cells. Applied and Environmental Microbiology 69: 7480-7491.

20. Kothary MH, Babu US (2001) Infective dose of foodborne pathogens in volunteers: A review. Journal of Food Safety 21: 49-73.

21. Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol: 577–587.

22. Frye J, Karlinsey J, Felise H, Marzolf B, Dowidar N, et al. (2006) Identification of new flagellar genes of Salmonella enterica serovar Typhimurium. Journal of bacteriology 188: 2233-2243.

23. Kutsukake K, Iino T (1994) Role of the FliA-FlgM regulatory system on the transcriptional control of the flagellar regulon and flagellar formation in Salmonella typhimurium. Journal of bacteriology 176: 3598-3605.

24. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods.

Appendix 1

Movie 1: Time-lapse movie showing GFP expression under the control of the *fliC* promoter during the growth of a microcolony.

GFP is under the control of *fliC* promoter on plasmid M956. This movie lasts for 106 minutes in real time. The phase contrast and fluorescent images have been merged; a lineage reconstruction of this movie can be seen in Figure 5. It can be found online at www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000307

Movie 2: Time-lapse movie showing GFP expression under the control of the *dcm* promoter during the growth of a microcolony.

GFP is under the control of the *dcm* (a DNA cytosine methylase) promoter on plasmid M956. This clone was isolated from a control population. This movie lasts for 178 minutes in real time. The phase contrast and fluorescent images have been merged; a lineage reconstruction of this movie can be seen in Figure 6. It can be found online at www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000307

Table 1: Sequenced inserts from selected and control populations and corresponding levels of noise in GFP expression.

Sequence data from the 240 clones used for analysis.

Clone	Treatment	Genbank product annotation	Genebank gene annotation	cells1	cv1	mean1	cells2	cv2	mean2
01_01	Selected	flagellar biosynthesis protein	fliC	7640	0.79	317.13	16715	0.75	254.32
01_02	Selected	flagellar biosynthesis protein	fliC	8036	0.80	362.58	18991	1.02	204.57
01_03	Selected	ribulose-phosphate 3-epimerase		8048	0.58	5.30	5438	0.58	5.39
01_04	Selected	flagellar biosynthesis protein	fliC	8785	0.70	334.20	9433	0.73	243.37
01_05	Selected	flagellar biosynthesis protein	fliC	8620	0.74	316.17	9725	0.71	257.77
01_06	Selected	flagellar biosynthesis protein	fliC	9482	0.86	322.91	9981	0.77	278.59
01_07	Selected	flagellar biosynthesis protein	fliC	8998	0.86	314.16	10647	0.82	254.98
01_08	Selected	flagellar biosynthesis protein	fliC	9157	0.89	231.85	9379	0.88	197.73
01_09	Selected	no close genes on same strand		8763	0.61	4.78	6394	0.58	5.02
01_10	Selected	flagellar biosynthesis protein	fliC	8874	0.77	282.81	10377	0.76	231.63
01_11	Selected	flagellar biosynthesis protein	fliC	9802	0.81	270.81	10030	0.86	197.68
01_12	Selected	flagellar biosynthesis protein	fliC	10371	0.94	237.12	10148	1.30	102.53
01_13	Selected	no close genes on same strand		8812	0.62	4.52	3806	0.64	5.09
01_14	Selected	flagellar biosynthesis protein	fliC	10319	0.80	350.94	10867	0.81	269.50
01_15	Selected	flagellar biosynthesis protein	fliC	10435	0.78	334.30	11055	0.87	238.33
01_16	Selected	flagellar biosynthesis protein	fliC	9888	0.86	311.75	10383	0.82	261.95
01_17	Selected	flagellar biosynthesis protein	fliC	10122	0.79	285.53	10387	0.81	217.42
01_18	Selected	putative outer membrane receptor		8636	0.53	7.13	6738	0.58	7.36
01_19	Selected	flagellar biosynthesis protein	fliC	9578	0.87	294.79	10673	0.86	244.39
01_20	Selected	flagellar hook-associated protein		8187	1.17	131.84	6718	1.26	108.47
01_21	Selected	flagellar biosynthesis protein	fliC	9992	0.91	267.61	11662	1.23	126.98
01_22	Selected	flagellar biosynthesis protein	fliC	9513	0.83	251.66	10600	0.87	193.90
01_23	Selected	no close genes on same strand		9109	0.59	4.85	7070	0.60	5.08
01_24	Selected	flagellar biosynthesis protein	fliC	11379	1.40	135.34	10707	1.35	130.43
03_01	Control	putative integral membrane protein		5280	0.62	4.90	9870	0.62	5.14
03_02	Control	no close genes on same strand		5624	0.59	4.80	10130	0.60	4.81
03_03	Control	putative transcriptional regulator		5589	0.49	10.79	5559	0.48	11.12
03_04	Control	succinyl-CoA synthetase alpha subunit		5994	0.57	5.62	6107	0.59	5.50
03_05	Control	putative fimbrial usher	safC	5893	0.59	4.66	6620	0.61	4.66
03_06	Control	no close genes on same strand		6186	0.59	4.65	6059	0.61	4.95
03_07	Control	2,5-diketo-D-gluconate reductase B	yafB	6003	0.57	6.25	5736	0.60	6.32
03_08	Control	response regulator		5735	0.58	5.13	7005	0.61	5.12
03_09	Control	pyridoxine kinase		6061	0.48	11.16	6154	0.50	10.24
03_10	Control	no close genes on same strand		5130	0.58	5.00	6574	0.61	5.02
03_11	Control	ribulose-phosphate 3-epimerase		6419	NA	6.74	5639	NA	5.86
03_12	Control	arginine repressor		2890	0.66	8.33	5718	0.59	6.13
03_13	Control	tetrathionate reductase complex subunit C		5506	0.55	6.36	6351	0.56	6.01
03_14	Control	putative outer membrane receptor		6624	0.56	5.85	6216	0.59	6.28
03_15	Control	no close genes on same strand		6217	0.59	5.45	6190	0.61	6.03
03_16	Control	putative outer membrane receptor		6141	0.69	7.63	7020	0.62	7.53
03_17	Control	no close genes on same strand		5616	0.59	4.62	6432	0.61	4.90
03_18	Control	Pb/Cd/Zn/Hg-transporting ATPase		6154	0.60	4.48	7025	0.63	4.77
03_19	Control	no close genes on same strand		5836	0.59	4.80	6820	0.62	5.16
03_20	Control	no close genes on same strand		6170	0.60	4.83	7136	0.61	5.04
03_21	Control	no close genes on same strand		6247	0.54	8.01	7020	0.55	7.52
03_22	Control	no close genes on same strand		5806	0.60	4.54	6801	0.62	4.77
03_23	Control	homoserine O-succinyltransferase	metA	6460	0.54	6.58	7120	0.56	6.77
03_24	Control	putative outer membrane receptor		6605	0.55	6.25	6864	0.57	5.85
04_01	Selected	no close genes on same strand		4642	0.79	22.46	11356	0.94	14.50
04_02	Selected	no close genes on same strand		4591	0.76	16.22	11357	0.59	16.96
04_03	Selected	putative phosphotransferase system IIC component		5669	0.56	5.63	5699	0.59	5.49
04_04	Selected	no close genes on same strand		5187	0.56	6.07	5868	0.59	6.20
04_05	Selected	no close genes on same strand		5566	0.78	22.68	6726	0.77	20.01
04_06	Selected	phage shock protein G	pspG	5365	0.71	27.71	5370	0.67	21.98
04_07	Selected	Na+/H+ antiporter		5150	0.45	20.00	6099	0.46	18.84
04_08	Selected	putative cytoplasmic protein	2881972	5446	0.52	8.20	5630	0.59	7.98
04_09	Selected	putative inner membrane protein		5675	0.58	5.59	6899	0.59	5.40
04_10	Selected	no close genes on same strand		5777	0.54	6.95	7294	0.56	6.37
04_11	Selected	no close genes on same strand		5014	0.54	6.68	6287	0.58	6.33
04_12	Selected	putative transport protein		4962	0.59	5.16	NA	NA	NA
04_13	Selected	no close genes on same strand		5736	0.48	13.60	6262	0.48	12.51
04_14	Selected	no close genes on same strand		6388	0.54	7.00	6767	0.58	6.01
04_15	Selected	putative phosphotransferase system IIC component		6681	0.57	5.38	NA	NA	NA
04_16	Selected	hypothetical protein		5870	0.59	18.78	6982	0.54	15.31
04_17	Selected	no close genes on same strand		6262	0.80	19.95	7067	0.76	19.78
04_18	Selected	putative cytoplasmic protein	yrbL	6031	1.00	95.43	6611	0.87	64.93
04_19	Selected	hypothetical protein		6089	0.56	8.25	6100	0.58	7.41
04_20	Selected	putative cytoplasmic protein		6031	0.54	12.24	7071	0.53	9.82

04_21	Selected	homoserine O-succinyltransferase	metA	6346	0.52	7.28	6997	0.55	6.49
04_22	Selected	putative phosphotransferase system IIC component		6138	0.56	6.02	5685	0.62	5.96
04_23	Selected	phosphoenolpyruvate-dependent sugar phosphotransferase system component		5223	0.57	5.40	6973	0.60	5.31
04_24	Selected	flagellar hook-associated protein		5606	1.33	167.08	6388	1.55	101.71
06_01	Control	no close genes on same strand		5834	0.60	4.76	10734	0.60	4.71
06_02	Control	phosphoenolpyruvate-dependent sugar phosphotransferase system component		6350	0.57	6.52	12122	0.59	6.16
06_03	Control	hypothetical protein		7041	0.50	9.66	6239	0.51	9.87
06_04	Control	hypothetical protein		5825	0.51	8.99	6449	0.55	8.57
06_05	Control	putative regulatory protein		5662	0.56	7.59	6009	0.57	7.45
06_06	Control	outer membrane putative efflux transporter		7068	0.60	4.84	5623	0.60	5.03
06_07	Control	no close genes on same strand		7057	0.57	6.21	6417	0.58	5.80
06_08	Control	hypothetical protein		5991	0.53	11.22	6665	0.51	10.21
06_09	Control	terminase-like large protein		6349	0.62	5.00	6020	0.60	4.80
06_10	Control	no close genes on same strand		6109	0.61	4.89	5711	0.60	4.84
06_11	Control	histidine ammonia-lyase		6457	0.57	5.36	6359	0.58	5.23
06_12	Control	phosphoenolpyruvate-dependent sugar phosphotransferase system component		6433	0.62	5.57	5912	0.62	5.56
06_13	Control	no close genes on same strand		6830	0.56	6.70	5461	0.55	6.87
06_14	Control	phosphoenolpyruvate-dependent sugar phosphotransferase system component		6852	0.63	5.85	5709	0.59	5.14
06_15	Control	polyhedral body protein		7026	0.65	11.48	6137	0.67	16.25
06_16	Control	no close genes on same strand		6767	0.60	4.86	7215	0.60	4.72
06_17	Control	Na+/H+ antiporter		6425	0.55	34.84	6117	0.47	34.25
06_18	Control	no close genes on same strand		5477	0.63	4.85	5745	0.61	4.56
06_19	Control	serine endoprotease		5801	0.61	4.61	7331	0.61	4.56
06_20	Control	putative outer membrane receptor		6583	0.55	6.02	7648	0.58	5.80
06_21	Control	outer membrane putative efflux transporter		5725	0.59	4.83	6880	0.60	4.75
06_22	Control	terminase-like large protein		5528	0.60	4.63	6289	0.62	4.87
06_23	Control	Mg2+ ATPase transporter		5353	0.57	5.29	6584	0.59	5.17
06_24	Control	terminase-like large protein		5702	0.61	4.85	6547	0.61	4.74
07_01	Selected	flagellar biosynthesis protein	fliC	7163	0.66	479.21	19612	0.69	341.76
07_02	Selected	no close genes on same strand		5580	0.62	5.41	9815	0.61	5.15
07_03	Selected	flagellar biosynthesis protein	fliC	7973	0.66	452.71	11056	1.20	157.15
07_04	Selected	flagellar biosynthesis protein	fliC	8078	0.67	339.36	9889	0.77	251.12
07_05	Selected	flagellar biosynthesis protein	fliC	7847	0.75	324.88	NA	NA	NA
07_06	Selected	flagellar biosynthesis protein	fliC	8884	0.74	308.72	11226	1.05	177.80
07_07	Selected	flagellar biosynthesis protein	fliC	5073	0.87	317.20	NA	NA	NA
07_08	Selected	2,5-diketo-D-gluconate reductase B	yafB	2902	0.72	9.69	6646	0.58	6.13
07_09	Selected	flagellar biosynthesis protein	fliC	6010	0.72	296.33	10275	0.82	247.59
07_10	Selected	flagellar biosynthesis protein	fliC	6354	0.84	320.78	10475	0.78	308.84
07_11	Selected	putative metal-dependent hydrolase	3096557	3155	0.59	14.09	NA	NA	NA
07_12	Selected	flagellar biosynthesis protein	fliC	5284	0.71	365.44	NA	NA	NA
07_13	Selected	flagellar biosynthesis protein	fliC	6013	0.61	490.19	NA	NA	NA
07_14	Selected	sigma N		5104	0.75	9.64	NA	NA	NA
07_15	Selected	flagellar biosynthesis protein	fliC	5758	0.65	379.31	10482	0.76	274.41
07_16	Selected	flagellar biosynthesis protein	fliC	8677	0.70	333.82	11336	0.81	244.44
07_17	Selected	flagellar biosynthesis protein	fliC	9381	0.77	316.40	11746	1.09	161.10
07_18	Selected	flagellar biosynthesis protein	fliC	10181	1.06	180.31	10422	0.76	256.36
07_19	Selected	flagellar biosynthesis protein	fliC	9180	1.09	164.67	10670	0.68	287.46
07_20	Selected	flagellar biosynthesis protein	fliC	10289	0.80	246.24	10567	0.73	265.36
07_21	Selected	flagellar biosynthesis protein	fliC	10059	0.84	272.13	10541	0.72	330.65
07_22	Selected	flagellar biosynthesis protein	fliC	10161	0.80	225.13	10336	1.16	147.05
07_23	Selected	putative diguanylate cyclase/phosphodiesterase		4834	0.51	12.58	7454	0.52	10.33
07_24	Selected	flagellar biosynthesis protein	fliC	10680	0.82	283.77	9214	1.01	264.55
09_01	Control	hypothetical protein		4108	0.67	13.48	12440	0.54	7.61
09_02	Control	no close genes on same strand		4191	0.74	9.20	12314	0.61	4.65
09_03	Control	glucose-6-phosphate 1-dehydrogenase	zwf	4053	0.55	26.43	5484	0.51	23.34
09_04	Control	putative acetyltransferase		4476	0.50	24.36	7731	0.40	22.12
09_05	Control	outer membrane putative efflux transporter		4029	0.75	9.14	6013	0.60	4.69
09_06	Control	no close genes on same strand		3488	0.70	8.29	6544	0.59	6.07
09_07	Control	putative outer membrane receptor		2998	0.69	9.58	6953	0.56	6.89
09_08	Control	putative Na+-dependent transporter		3477	0.70	8.50	6179	0.55	6.39
09_09	Control	outer membrane putative efflux transporter		3622	0.71	7.73	6596	0.60	4.73
09_10	Control	glucose-6-phosphate 1-dehydrogenase	zwf	3909	0.49	25.82	7083	0.45	20.46
09_11	Control	no close genes on same strand		3621	0.71	7.44	7189	0.62	4.49
09_12	Control	sulfate permease A protein		5143	0.89	5.19	6900	0.61	4.43
09_13	Control	succinyl-CoA synthetase alpha subunit		7816	NA	5.43	8107	NA	4.79
09_14	Control	no close genes on same strand		8033	0.60	4.56	6761	0.60	4.57
09_15	Control	no close genes on same strand		7188	0.63	5.23	9165	0.64	5.02
09_16	Control	no close genes on same strand		7417	0.60	4.70	8133	0.61	4.71
09_17	Control	no close genes on same strand		7516	0.59	4.98	7345	0.61	5.04

09_18	Control	putative regulatory protein		6535	0.37	23.99	8378	0.44	18.78
09_19	Control	no close genes on same strand		7249	0.63	4.83	7381	0.61	4.52
09_20	Control	sulfate permease A protein		7834	0.61	4.61	8009	0.61	4.51
09_21	Control	no close genes on same strand		7175	0.60	5.42	7584	0.59	5.25
09_22	Control	no close genes on same strand		7779	0.61	4.61	7048	0.63	4.59
09_23	Control	tetrathionate reductase complex subunit C		6989	0.61	7.61	6714	0.58	7.33
09_24	Control	no close genes on same strand		5851	0.59	5.31	6716	0.60	4.92
10_01	Selected	no close genes on same strand		5596	0.58	5.32	NA	NA	NA
10_02	Selected	putative outer membrane receptor		5551	0.56	7.30	NA	NA	NA
10_03	Selected	hypothetical protein	smp	4658	0.46	29.32	3587	0.59	30.70
10_04	Selected	hypothetical protein		6167	0.51	9.60	6034	0.56	10.15
10_05	Selected	putative aminotransferase		5524	0.59	5.19	5613	0.65	6.64
10_06	Selected	molecular chaperone DnaK		6611	0.66	9.61	6321	0.68	11.20
10_07	Selected	hypothetical protein		6648	0.50	9.66	6522	0.53	10.54
10_08	Selected	deoxyguanosinetriphosphate triphosphohydrolase		5772	0.49	16.39	6404	0.53	18.36
10_09	Selected	hypothetical protein		6441	0.52	15.20	5800	0.52	16.62
10_10	Selected	putative aminotransferase		6172	0.60	4.85	6790	0.72	6.26
10_11	Selected	nucleoside/purine/pyrimidine transporter		6104	0.79	8.67	5594	0.64	11.64
10_12	Selected	sulfate permease A protein		5915	0.54	10.07	NA	NA	NA
10_13	Selected	SOS repair enzyme	dinG	6914	0.59	12.70	NA	NA	NA
10_14	Selected	SOS repair enzyme	dinG	6987	0.56	12.36	6244	0.58	14.22
10_15	Selected	putative transport protein	abc	6661	0.77	9.54	4742	0.82	12.14
10_16	Selected	acetylglutamate kinase		6851	0.90	11.14	7999	0.92	10.67
10_17	Selected	putative inner membrane protein	249289	6219	0.66	9.06	5762	0.73	9.30
10_18	Selected	hypothetical protein		6799	0.51	12.63	5651	0.54	13.20
10_19	Selected	SOS repair enzyme	dinG	6271	0.55	9.54	6011	0.56	11.40
10_20	Selected	no close genes on same strand		5520	0.57	12.86	5773	0.63	12.15
10_21	Selected	hypothetical protein		7618	0.53	8.75	5736	0.55	10.87
10_22	Selected	hypothetical protein		7395	0.61	12.33	6301	0.57	14.42
10_23	Selected	deoxyguanosinetriphosphate triphosphohydrolase		3723	0.60	17.10	4941	0.57	14.13
10_24	Selected	no close genes on same strand		6120	0.61	7.68	5668	0.63	7.42
12_01	Control	putative outer membrane receptor		6127	0.56	8.20	9324	0.58	8.25
12_02	Control	no close genes on same strand		6262	0.61	4.71	9647	0.65	5.45
12_03	Control	terminase-like large protein		5295	0.62	5.44	5151	0.65	5.40
12_04	Control	Na+/H+ antiporter		5350	0.46	31.96	NA	NA	NA
12_05	Control	putative outer membrane receptor		6632	0.55	6.81	5899	0.59	7.68
12_06	Control	hypothetical protein		6607	0.50	9.76	6110	0.55	10.22
12_07	Control	histidine ammonia-lyase		6716	0.59	5.44	5596	0.61	5.91
12_08	Control	DNA cytosine methylase	dcm	6113	0.41	42.26	4905	0.51	40.77
12_09	Control	no close genes on same strand		6597	0.61	5.08	4416	0.64	5.86
12_10	Control	no close genes on same strand		6616	0.47	22.62	5086	0.47	20.95
12_11	Control	tetrathionate reductase complex subunit C		3523	0.69	12.01	5965	0.58	8.11
12_12	Control	no close genes on same strand		6302	0.60	5.08	5268	0.66	5.68
12_13	Control	putative anaerobic dimethylsulfoxide reductase subunit A		6952	0.58	9.82	5181	0.57	14.96
12_14	Control	hypothetical protein		6747	0.53	8.71	5365	0.54	10.45
12_15	Control	no close genes on same strand		6739	0.62	4.89	4996	0.66	5.93
12_16	Control	response regulator		7503	0.58	5.45	5729	0.63	6.19
12_17	Control	putative Na+-dependent transporter		6237	0.57	6.43	4992	0.64	9.20
12_18	Control	secreted effector protein	sopD	6240	0.60	4.67	4968	0.65	5.80
12_19	Control	secreted effector protein	sopD	6198	0.60	4.68	6210	0.63	5.48
12_20	Control	putative secreted protein		6890	0.61	4.78	5488	0.67	6.16
12_21	Control	no close genes on same strand		8374	0.52	8.10	NA	NA	NA
12_22	Control	no close genes on same strand		7202	0.61	4.76	5143	0.67	5.92
12_23	Control	hypothetical protein		3984	0.65	15.31	6116	0.58	13.43
12_24	Control	no close genes on same strand		6813	0.58	5.30	5813	0.64	6.92
13_01	Control	terminase-like large protein		5681	0.59	4.85	6918	0.63	5.80
13_02	Control	ilvGEDA operon leader peptide		5590	0.57	5.78	9512	0.61	6.27
13_03	Control	threonine dehydratase		5880	1.46	60.30	4634	1.36	66.95
13_04	Control	Na+/H+ antiporter		6081	0.47	18.39	4539	0.48	18.97
13_05	Control	no close genes on same strand		5903	0.54	7.50	4742	0.58	8.74
13_06	Control	L-serine deaminase		6866	0.50	26.87	7378	0.56	25.71
13_07	Control	no close genes on same strand		6053	0.66	22.60	6881	0.70	18.44
13_08	Control	N-acetylmuramoyl-L-alanine amidase	amiC	6526	0.53	12.45	5907	0.56	12.14
13_09	Control	no close genes on same strand		5981	0.61	4.62	5991	0.65	5.68
13_10	Control	transaldolase		6241	0.60	4.77	5987	0.63	5.16
13_11	Control	transaldolase		6357	0.58	4.69	NA	NA	NA
13_12	Control	putative ATP-binding protein		6308	0.56	6.11	5474	0.64	7.01
13_13	Control	tetrathionate reductase complex subunit B		6608	0.59	5.58	NA	NA	NA
13_14	Control	putative anaerobic dimethylsulfoxide reductase subunit A		7090	0.51	10.83	6271	0.52	11.25

13_15	Control	ilvGEDA operon leader peptide		7951	0.58	5.45	6250	0.62	6.25
13_16	Control	terminase-like large protein		6401	0.59	4.59	6135	0.64	5.57
13_17	Control	terminase-like large protein		6718	0.63	5.02	4532	0.67	6.09
13_18	Control	putative cytoplasmic protein		6733	0.59	5.35	5518	0.61	6.09
13_19	Control	no close genes on same strand		7107	0.77	23.41	6247	0.79	20.78
13_20	Control	putative glycosyl transferase		7542	1.76	34.63	6634	1.71	30.09
13_21	Control	ilvGEDA operon leader peptide		6414	0.57	5.50	8006	0.60	6.03
13_22	Control	no close genes on same strand		6490	0.73	20.16	6057	0.83	12.79
13_23	Control	no close genes on same strand		7692	1.04	15.51	7324	0.94	17.04
13_24	Control	flagellar hook-associated protein		5971	1.19	180.76	5368	1.17	169.92
15_01	Control	phosphoenolpyruvate-dependent sugar phosphotransferase system component		4459	0.58	5.87	9875	0.63	6.39
15_02	Control	no close genes on same strand		4984	0.42	33.13	11780	0.44	32.52
15_03	Control	DNA helicase	recG	4925	0.58	5.52	5355	0.67	6.75
15_04	Control	putative 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase		5190	0.58	5.66	6692	0.61	5.94
15_05	Control	histidine ammonia-lyase		4884	0.60	5.89	5364	0.60	5.34
15_06	Control	putative transcriptional regulator		5494	0.51	10.58	6933	0.52	9.88
15_07	Control	hypothetical protein		5301	0.53	9.49	3857	0.59	8.55
15_08	Control	no close genes on same strand		4593	0.66	28.31	4902	0.71	23.24
15_09	Control	polyhedral body protein		4873	0.54	10.67	4120	0.64	11.39
15_10	Control	hypothetical protein		5065	0.56	8.36	6353	0.54	8.20
15_11	Control	putative transcriptional regulator		5741	0.51	8.65	6975	0.54	8.44
15_12	Control	transcriptional repressor		5291	0.57	6.67	6297	0.56	6.18
15_13	Control	2,5-diketo-D-gluconate reductase B	yafB	5871	0.56	6.16	5755	0.64	7.01
15_14	Control	Pb/Cd/Zn/Hg-transporting ATPase		5383	0.62	5.08	6427	0.64	5.05
15_15	Control	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase		5376	0.62	5.15	6620	0.62	4.76
15_16	Control	DNA helicase	recG	5288	0.61	5.66	7569	0.63	5.57
15_17	Control	no close genes on same strand		5574	0.60	5.29	7342	0.64	5.40
15_18	Control	putative inner membrane lipoprotein		5732	0.52	153.05	6237	0.56	158.59
15_19	Control	no close genes on same strand		4890	0.59	6.13	7074	0.60	5.87
15_20	Control	precorrin-8X methylmutase		6581	0.57	5.54	7016	0.66	6.91
15_21	Control	Pb/Cd/Zn/Hg-transporting ATPase		5024	0.60	4.90	5390	0.62	5.18
15_22	Control	no close genes on same strand		6151	0.62	5.10	5509	0.64	5.30
15_23	Control	no close genes on same strand		5644	0.54	7.76	5325	0.56	7.80
15_24	Control	no close genes on same strand		5214	0.58	7.22	6508	0.59	6.82

Appendix 2

Additional mathematical analysis of the model for the evolution of self-destructive cooperation mediated by phenotypic noise

In the following analysis, we use the within interaction group payoffs given by eqs. (1) and (2) for the case where the public good is fully available to everybody in a given interaction group. The case where the public good is distributed among the surviving members of the interaction group can be analyzed along the same lines, and gives qualitatively very similar results.

We assume that the population is spatially structured as described in the text, and hence that the functions $e_C(x)$ and $e_D(x)$ are given by eqs. (5) and (6). Recall that $e_C(x)$ and $e_D(x)$ are the average number of cooperators in an interaction group of a focal *C* or *D*, respectively, when the population frequency of *C* is x.

We want to investigate invasion of *C* into a population of *D* and vice versa, hence we consider the cases $x = 0$ and $x = 1$. When *C* is rare, we have $e_C(0) = N/M - 1$ and $e_D(0) = 0$, hence

$$P_C(0) = (1-q)\left(\frac{(N-M)qb}{M} + w\right) \tag{1}$$

and

$$P_D(0) = w. \tag{2}$$

C can invade *D* if and only if $P_C(0) > P_D(0)$, which is the case if and only if $q < q^*_1$, where q^*_1 is given by

$$q_1^* = \frac{b(N-M) - Mw}{b(N-M)}. \tag{3}$$

Thus, cooperators can invade defectors at least for some q if $w/b < (N-M)/M$. In particular, smaller baseline fitness w, larger benefit to cooperation b, larger interaction group size N and smaller neighborhood size M (i.e., more pronounced spatial structure) are all beneficial for cooperation, which makes sense intuitively.

Similarly, *D* can invade *C* if and only if $P_C(1) < P_D(1)$, which is the case if and only if $q > q^*_2$,

where q^*_2 is given by

$$q_2^* = \frac{b(N-M) - Mw}{b(NM-M)}. \tag{4}$$

Thus, for $q < q^*_2$, D cannot invade C. Clearly, $q^*_2 < q^*_1$, hence for $q < q^*_2$, C can invade D, but D cannot invade C. It is easy to see that this in turn implies that for $q < q^*_2$, C replaces D, whereas for $q^*_2 < q < q^*_1$, the two strategies can invade each other and hence they coexist at a mixed equilibrium. This proves claims 1)-3) in the main text.

To investigate the adaptive dynamics of the continuous trait q, we consider resident populations that are monomorphic for a particular trait value q into which a rare mutant with a mutant trait value q' attempts to invade. Because the mutant is rare, and given the spatial structure described in the main text, the environment of the mutant in an interaction group of size N consists of $N/M - 1$ mutant types q' and $N(M-1)/M$ resident types q. Thus, the payoff of the mutant q' in the resident q, which is the *invasion fitness* of the mutant q', is given by

$$f(q,q') = \left(\frac{(N-M)q'b}{M} + \frac{N(M-1)qb}{M} + w \right)(1-q'). \tag{5}$$

The adaptive dynamics is then given by the selection gradient

$$D(q) = \frac{\partial f(q,q')}{\partial q'}|_{q'=q} = \frac{b\left(N(1-q) - M(1-2q+Nq)\right)}{M} - w. \tag{6}$$

For any given resident value q, the direction of evolution is determined by the sign of $D(q)$. In particular, if $w < b(N-M)/M$, then $D(0) > 0$, and hence the trait q evolves away from 0. (Note that this is the same condition as the condition for $q^*_1 > 0$, and hence for the existence of a range of q-values for which cooperators can invade pure defectors.) In this case, the function $D(q)$ is linear with a negative slope. Therefore, the equation $D(q) = 0$ has a unique, positive solution

$$q_3^* = \frac{b(N-M) - Mw}{b(MN+N-2M)}, \tag{7}$$

and q^*_3 is a global attractor of the adaptive dynamics (i.e., starting with any q_0 in the interval

[0, 1], the
adaptive dynamics converges to q^*_3). Moreover, one easily checks that

$$\frac{\partial^2 f(q,q')}{\partial q'^2}|_{q'=q=q^*_3} = -\frac{2b(N-M)}{M} < 0, \tag{8}$$

and hence q^*_3 is evolutionarily stable. It is again clear from direct comparison of the denominators that if cooperation can persist, i.e., if $w < b(N-M)/M$, then $q^*_3 < q^*_2$. This means that the evolutionary dynamics comes to a halt before the probability to sacrifice is so high that pure defectors can invade and coexist with cooperators. Thus, continuous evolution in small steps (as is assumed in the adaptive dynamic framework) results in a population consisting of a single cooperative strategy.

To illustrate the change in the nature of the evolutionary game between *C* and *D* brought about by assortment due to spatial structure, consider the case $M = 2$, in which each local deme is seeded by 2 individuals, so that there are three types of demes: *CC*, *CD* and *DD*. In this case the game between cooperators and defectors can be viewed as a classical 2x2 game with two players and two strategies *C* and *D*. If a focal *C* colonizes a deme containing another *C*-seed, the payoff to the focal *C*-strategy will be $P_{CC} = (1-q)[(N-1)bq + w]$, and if the focal *C* colonizes a deme in which the other seed is a *D*, then the payoff to the focal *C*-strategy is $P_{CD} = (1-q)[(N/2-1)bq+w]$. Similarly, the payoff for focal *D* players are $P_{DC} = (n/2 - 1)bq + w$ and $P_{DD} = w$, depending on whether the other seed in the deme is a *C* or a *D*. In general, replicator dynamics of 2x2-games have three outcomes[1]: i) there are only the two boundary equilibria, and one strategy always dominates the other; ii) there is a stable interior equilibrium, at which the two strategies coexist; and iii) there is an unstable interior equilibrium, in which case one strategy wins, but the winner depends on initial conditions. In our model, case iii) does not occur. Instead, with $M = 2$ and the payoffs P_{CC}, P_{CD}, P_{DC} and P_{DD} given above, case i) holds with *D* the dominant strategy if $P_{CD} < P_{DD}$ and $P_{CC} < P_{DC}$, which occurs if and only if $q > q_1''$, where q^*_1 is given by eq. (9) for $M = 2$. If $q^*_2 < q < q^*_1$, with q^*_2 given by eq. (10), the payoffs satisfy $P_{CD} < P_{DD}$ and $P_{CC} > P_{DC}$, which generates scenario ii) and leads to coexistence of *C* and *D*. This situation corresponds to the well-known Snowdrift game[8]. Finally, case i) with *C* the dominant strategy occurs for $q < q^*_2$, for which $P_{CD} > P_{DD}$ and $P_{CC} > P_{DC}$. Overall, for $q < q^*_1$ assortment through spatial structure changes the nature of the game between *C* and *D* compared to the dominance of *D* observed in well-

mixed populations. For $M > 2$, the effect of assortment on the game between C and D can be similarly understood in a general framework for N-player games[2]. Part II: Additional information on the mouse infection experiments with S. typhimurium.

Experiment 1; Analyzing the TTSS-1 expression phenotype of individual bacteria in the gut lumen and the gut tissue in a mouse infection experiment (Figure 2b of the paper).

Under in vitro culture conditions, 10-30% of a clonal wild type S. typhimurium population expresses the TTSS-1+ phenotype, and the rest is TTSS-1- [3,4]. We employed an animal model for S. typhimurium enterocolitis [5,6] to analyze whether TTSS-1+ and TTSS-1- subpopulations also exist in vivo. In addition, we wanted to verify that only the TTSS-1+ subpopulation invades into the gut tissue.

Streptomycin-treated C57BL/6 mice were infected as described[5] with wild type S. typhimurium (SL1344; 5 * 10^7 cfu, intragastrically; n = 3 mice). The bacteria carried two reporter plasmids. pM974 allowed detecting TTSS-1+ bacteria. It encodes a GFP-reporter coupled to the TTSS-1 promoter of the sicAsipBCDA operon. The second plasmid, pDsRed, leads to bright red fluorescence in all bacteria located in intestinal tissues[7]. The mice were sacrificed at 12 hours post infection. Earlier work has indicated that 12h is long enough to allow significant tissue invasion by S. typhimurium. At the same time it is short enough to avoid pronounced bacterial growth within the host tissue which may have diluted out the TTSS-1-GFP signal. We determined total colonization levels of S. typhimurium in the gut lumen by plating appropriate dilutions of the cecal contents on McConkey agar. The density of bacteria was 2.1 x 10^9 +/- 2.5 x 10^8 (cfu/gram, +/-standard error). In addition, we analyzed gut inflammation using hematoxilin and eosin stained cecum tissue sections and a scoring scheme as described previously[8]. The average inflammation intensity was 5 +/- 1.7 (mean +/-standard error or the mean). This corresponds to a strong, acute gut inflammation. The results obtained by analyzing TTSS-1+ (i.e. GFP-) expression in individual S. typhimurium bacteria localized in the gut lumen or in the cecal tissue are depicted in Figure 2b. These data were obtained in the following way: Cecal tissue was fixed and embedded to preserve and visualize GFP and DsRed fluorescence as described previously37. For analyzing TTSS-1+ (i.e. GFP-) expression in the cecum tissue, tissue sections (20µm) were stained with phalloidin-Alexa-647 (infrared fluorescence; stains the actin brush border of the epithelium) and DAPI (4' 6'–diamidino-2-phenylindole, 0.5µg/ml, Sigma; blue fluorescence; stains DNA). Confocal images of individual bacteria inside of the cecum tissue (bright DsRed fluorescence) were

analyzed using a Perkin Elmer Ultraview confocal imaging system and a Zeiss Axiovert 200 microscope. We identified a total of 87 individual *S. typhimurium* bacteria in the cecum tissues from the 3 mice (three 20 μm thick tissue slices per mouse) and 86 of the 87 bacteria harboured significant levels of the GFP reporter, indicating that these cells express TTSS-1 (Figure 2b).

TTSS-1 expression by luminal bacteria was analyzed in the same way. However, bacteria were stained additionally with an anti-*S. typhimurium* antiserum (polyclonal rabbit α-LPS O antigen group B anti-serum; Difco; 1:500 in PBS, 10% goat serum) and a Cy3-conjugated goat α rabbit antibody (Milan; 1:300 in PBS, 10% goat serum; red fluorescence) to enhance sensitive and specific detection of *S. typhimurium* in the gut lumen. At least 100-300 individual bacteria (red$^+$) were scored per mouse. The fraction of luminal bacteria (open bars) expressing the GFP-reporter (= TTSS-1$^+$ phenotype) was calculated as: GFP$^+$red$^+$[%] = #GFP$^+$red$^+$/(#GFP$^+$red$^+$ + #GFP$^-$red$^+$) x 100. Approximately 15% of the luminal *S. typhimurium* population expressed the TTSS-1$^+$ phenotype (Figure 2b).

Experiment 2. Analyzing how decreasing factions of wild type S. typhimurium affect the elicitation of gut inflammation in mouse infection experiments (Figure 2d of the paper and Appendix. Fig. A1).

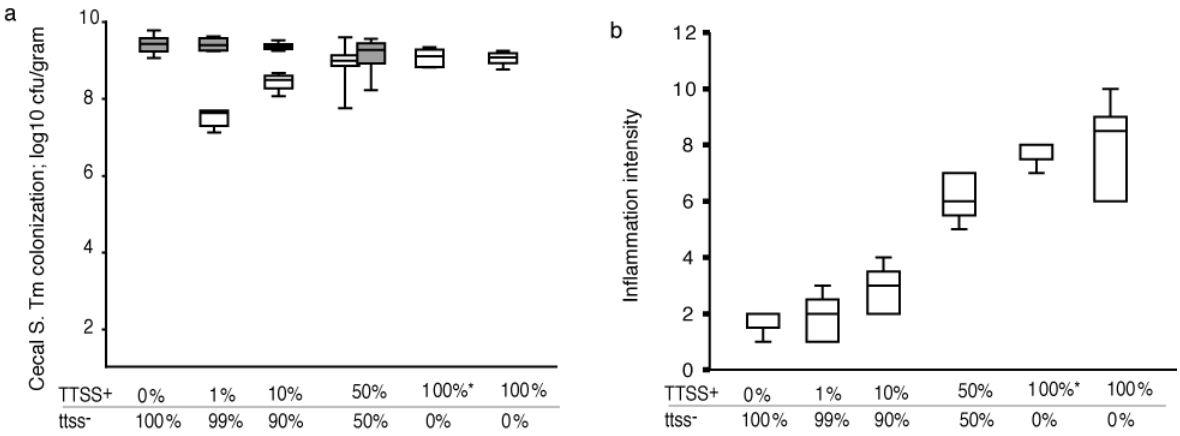
In the streptomycin mouse model, wild type *S. typhimurium* efficiently colonizes the large intestinal lumen and triggers pronounced gut inflammation within one day post infection[5]. An isogenic *S. typhimurium* mutant carrying a mutation disrupting TTSS-1 function (ΔTTSS-1, i.e. SB161; Δ*invG*) is still able to efficiently colonize the gut lumen, but it cannot efficiently invade the gut tissue and it is incapable of triggering inflammation at day one post infection[5,7].

In the gut lumen, 15% of the wild type *S. typhimurium* population expresses the TTSS-1$^+$ phenotype. In contrast, the gut luminal ΔTTSS-1 population never expresses the TTSS-1$^+$ phenotype. Therefore, infections with mixtures of wild type *S. typhimurium* and ΔTTSS-1 result in a proportional decrease in the total number of bacteria expressing the TTSS-1$^+$ phenotype in the gut lumen.

Streptomycin treated mice (n = 10 C57BL/6 mice per group) were infected intragastrically with a total of $5 * 10^7$ bacteria. This inoculum was composed of either 100% ΔTTSS-1, 1% wild type *S. typhimurium* and 99% ΔTTSS-1, 10% wild type *S. typhimurium* and 90%

ΔTTSS-1, 50% wild type *S. typhimurium* and 50% ΔTTSS-1, or 100% wild type *S. typhimurium*, as indicated. An additional control infection was performed with 5 * 10^5 bacteria (100% wild type *S. typhimurium*; 5 mice; marked as *). The wild type *S. typhimurium* strain (M939) was marked with a kanamycin resistance cassette integrated downstream of the *sopE* gene. This allowed quantifying wild type *S. typhimurium* ($kan^{resistant}$) and ΔTTSS-1 ($kan^{sensitive}$) in the inoculum (data not shown) and in the gut lumen (see, below).

The mice were sacrificed at day 1 post infection and we analyzed the following parameters: (a; Appendix. Fig. 1a) Total colonization levels by wild type *S. typhimurium* (M939; $kan^{resistant}$; white boxes) and ΔTTSS-1 (SB161; $kan^{sensitive}$; grey boxes) in the gut lumen. The densities (cfu/gram) were determined by plating appropriate dilutions of the cecal contents on McConkey agar plates harboring kanamycin (50μg/ml) or not. These data verified that the mixed inoculum approach was successful in adjusting the absolute density of wild type *S. typhimurium* in the gut lumen of the infected mice. (b; suppl. Fig. 1b). The gut inflammation was analyzed using hematoxilin and eosin stained cecum tissue sections and a scoring scheme as described previously[8]. These data verified that the intensity of gut inflammation increased with an increasing fraction of wild type *S. typhimurium* present in the gut lumen.



Appendix Figure 1

a, Infections with mixtures of wild type *S. typhimurium* and increasing fractions of ΔTTSS-1 result in a proportional decrease in the total number of bacteria expressing the TTSS-1+ phenotype in the gut lumen. Total colonization levels by wild type *S. typhimurium* (M939; kanresistant; white boxes) and ΔTTSS-1 (SB161; kansensitive; grey boxes) in the gut lumen are indicated. Gut inflammation levels decrease with decreasing proportions of wildtype *S. typhimurium* that are capable of expressing the TTSS-1+ phenotype. The data are the same as depicted in Figure 2d, but includes the additional control group infected with 5 * 10^5 wild type *S. typhimurium*

(marked as *). This control demonstrates that, in the absence of any ΔTTSS-1 bacteria, wild type *S. typhimurium* always grows up to densities of approx. 10^9 cfu/g cecum content, irrespective of the absolute size of the original inoculum.

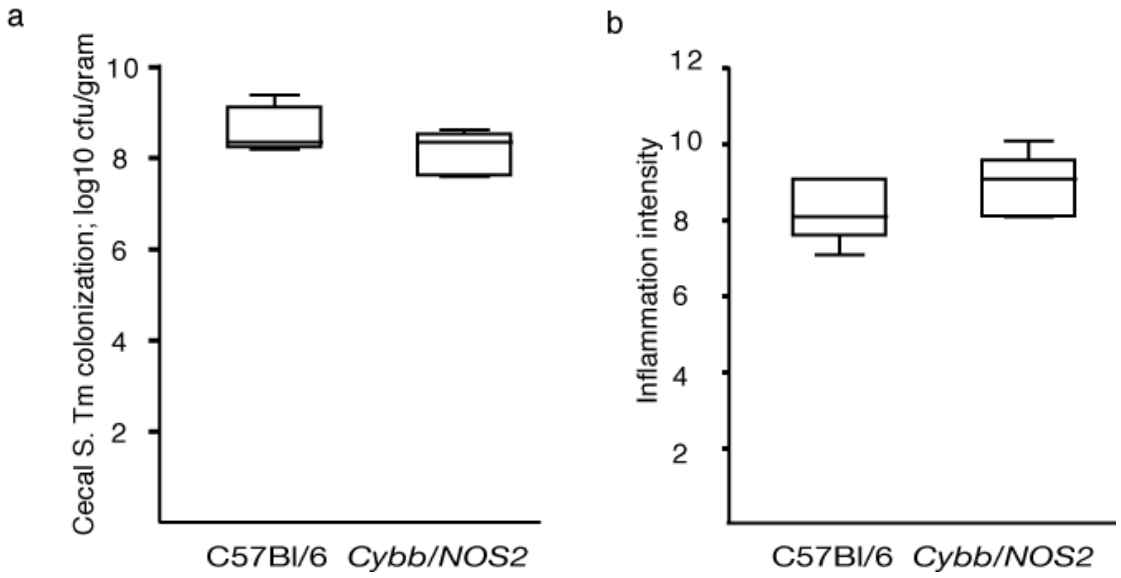
Experiment 3. Killing of S. typhimurium in gut tissues by key innate immune defenses in a mouse infection experiment (Figure 2f of the paper and Appendix Figure A2).

Bacteria that have invaded into intestinal tissues are killed by innate immune mechanisms which generate antibacterial nitrogen- and oxygen radicals (iNOS and NADPH-oxidase; gp91-phox protein of the phagocytic **NADPH** oxidase). It has been shown previously that mice lacking these key anti-bacterial defense systems cannot restrict pathogen growth in various infection models39. Thus, comparing *S. typhimurium* tissue loads in wild type and *Cybb/NOS2* knockout mice allows estimating the extent of bacterial killing in infected gut tissues.

Streptomycin treated wild type mice (n = 5 C57BL/6 mice per group) or *Cybb/NOS2* knockout mice (n = 5 mice per group, C57BL/6 genetic background; bred in the same colony as the wild type control mice) were infected with wild type *S. typhimurium* ($5 * 10^7$ bacteria, intragastrically). *Cybb/NOS2* double knockout mice were generated by crossing B6.129S6-*Cybb*^{tm1Din}/J [9] and B6;129P2 *Nos2*^{tm1Lau}/J [10] (both from Jackson Laboratory). The bacteria harbored a GFP reporter plasmid (pM973, [7]) allowing the quantification of *S. typhimurium* loads in the infected gut tissue. The mice were sacrificed at day 2 post infection and we analyzed the following parameters: (a; Appendix. Fig. 2a) Total colonization levels by wild type *S. typhimurium* in the cecum lumen34. The densities (cfu/gram) were determined by plating appropriate dilutions of the cecal contents on McConkey agar plates. Colonization densities in the gut lumen did not differ significantly between both groups of mice (p=0.730; Mann Whitney U-test) (b; Appendix. Fig. 2b) The gut inflammation was analyzed using hematoxilin and eosin stained cecum tissue sections and a scoring scheme as described previously37. Inflammation was pronounced and did not differ significantly between both groups of mice (p=0.484; Mann Whitney U-test).

Significant differences were detected in the numbers of *S. typhimurium* bacteria present in the cecum tissues of both groups of mice. These data are depicted in Fig. 2f and were generated in the following way: The cecal tissue was fixed and cryo-embedded as described38 and 20µm thick cryo-sections were stained with phalloidin-TRITC (stains actin; red fluorescence) and

DAPI (stains DNA, blue fluorescence). All bacteria (expressing GFPreporter of pM973) present in the gut tissue on one tissue section were enumerated. At least 3 tissue sections were analyzed per mouse. These data indicate that the innate immune system effectively kills approximately 90% of all bacteria entering the gut tissue. This suggests that host tissue invasion by *S. typhimurium* can be considered as a self-destructive act.



Appendix Figure 2

A, Density of *S. typhimurium* in the gut lumen, and B, inflammation intensity were very similar between wild type mice (C57Bl/6) and *Cybb/NOS2* knock-out mice.

References

1. Hofbauer JS, K (1998) Evolutionary Games and Population Dynamics: Cambridge University Press.

2. Hauert C, Michor F, Nowak M, Doebeli M (2006) Synergy and discounting of cooperation in social dilemmas. Journal of Theoretical Biology 239: 195-202.

3. Schlumberger MC, Muller AJ, Ehrbar K, Winnen B, Duss I, et al. (2005) Real-time imaging of type III secretion: Salmonella SipA injection into host cells. Proceedings of the National Academy of Sciences of the United States of America 102: 12548-12553.

4. Hautefort I, Proenca MJ, Hinton JCD (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. Applied and Environmental Microbiology 69: 7480-7491.

5. Barthel M, Hapfelmeier S, Quintanilla-Martinez L, Kremer M, Rohde M, et al. (2003) Pretreatment of Mice with Streptomycin Provides a Salmonella enterica Serovar Typhimurium Colitis Model That Allows Analysis of Both Pathogen and Host Editor: DL Burns. Infection and Immunity 71: 2839-2858.

6. Hapfelmeier S, Hardt W (2005) A mouse model for S. typhimurium-induced enterocolitis. Trends in Microbiology 13: 497-503.

7. Hapfelmeier S, Stecher B, Barthel M, Kremer M, Muller AJ, et al. (2005) The Salmonella Pathogenicity Island (SPI)-2 and SPI-1 Type III Secretion Systems Allow Salmonella Serovar typhimurium to Trigger Colitis via MyD88-Dependent and MyD88-Independent Mechanisms. J Immunol 174: 1675-1685.

8. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, et al. (2007) Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. Plos Biology 5: 2177-2189.

9. Pollock J, Williams D, Gifford M, Li L, Du X, et al. (1995) Mouse model of Xñlinked chronic granulomatous disease, an inherited defect in phagocyte superoxide production. Nature genetics 9: 202-209.

10. Laubach V, Shesely E, Smithies O, Sherman P (1995) Mice lacking inducible nitric oxide synthase are not resistant to lipopolysaccharide-induced death. Proceedings of the National Academy of Sciences 92: 10688-10692.

Appendix 3

Table 1. Comparison of CV between Ancester (AN), Control populations (C1, C2, C3, C4) and Experimental populations (E1, E2, E3, E4). Bold numbers are significant ($p<0.001$).

		rpoS						
		AN	C1	C2	C3	E1	E2	E3
Wilcox	AN	5.0E-01	1.2E-01	8.7E-01	7.6E-02	7.1E-01	**6.1E-04**	1.0E-02
	C1	8.8E-01	5.0E-01	9.9E-01	2.5E-01	9.5E-01	2.9E-02	1.3E-01
	C2	1.4E-01	7.6E-03	5.1E-01	2.5E-02	4.1E-01	**3.9E-05**	**7.7E-04**
	C3	9.3E-01	7.6E-01	9.8E-01	5.0E-01	9.3E-01	2.1E-01	4.0E-01
	E1	3.0E-01	5.4E-02	6.0E-01	7.6E-02	5.0E-01	2.2E-03	1.4E-02
	E2	1.0E+00	9.7E-01	1.0E+00	7.9E-01	1.0E+00	5.1E-01	6.7E-01
	E3	9.9E-01	8.8E-01	1.0E+00	6.1E-01	9.9E-01	3.4E-01	5.1E-01
t-test	AN	5.0E-01	1.1E-01	8.3E-01	2.4E-01	8.2E-01	**5.0E-04**	3.6E-02
	C1	8.9E-01	5.0E-01	9.8E-01	5.8E-01	9.6E-01	1.9E-02	2.1E-01
	C2	1.7E-01	2.1E-02	5.0E-01	8.0E-02	5.9E-01	**5.7E-05**	6.7E-03
	C3	7.6E-01	4.2E-01	9.2E-01	5.0E-01	9.1E-01	4.3E-02	2.0E-01
	E1	1.8E-01	4.2E-02	4.1E-01	9.1E-02	5.0E-01	1.2E-03	1.5E-02
	E2	1.0E+00	9.8E-01	1.0E+00	9.6E-01	1.0E+00	5.0E-01	8.0E-01
	E3	9.6E-01	7.9E-01	9.9E-01	8.0E-01	9.8E-01	2.0E-01	5.0E-01

		PrgH						
		AN	C1	C2	C3	E1	E2	E3
Wilcox	AN	5.0E-01	1.0E+00	1.0E+00	1.0E+00	9.7E-01	1.0E+00	1.0E+00
	C1	**1.7E-05**	5.0E-01	3.0E-01	2.2E-01	1.6E-01	1.5E-01	1.8E-01
	C2	**1.8E-04**	7.1E-01	5.0E-01	4.7E-01	3.0E-01	2.0E-01	4.2E-01
	C3	4.0E-03	7.9E-01	5.4E-01	5.0E-01	4.1E-01	3.6E-01	5.4E-01
	E1	3.6E-02	8.4E-01	7.1E-01	6.0E-01	5.0E-01	5.8E-01	6.6E-01
	E2	1.6E-03	8.6E-01	8.1E-01	6.4E-01	4.2E-01	5.0E-01	6.5E-01
	E3	1.8E-03	8.2E-01	5.9E-01	4.6E-01	3.5E-01	3.6E-01	5.0E-01
t-test	AN	5.0E-01	1.0E+00	1.0E+00	1.0E+00	9.9E-01	1.0E+00	1.0E+00
	C1	**1.2E-05**	5.0E-01	3.2E-01	1.9E-01	2.0E-01	1.0E-01	2.3E-01
	C2	**2.2E-04**	6.8E-01	5.0E-01	3.3E-01	3.3E-01	2.3E-01	3.9E-01
	C3	2.1E-03	8.1E-01	6.7E-01	5.0E-01	4.7E-01	4.1E-01	5.5E-01
	E1	1.0E-02	8.0E-01	6.7E-01	5.3E-01	5.0E-01	4.5E-01	5.8E-01
	E2	1.6E-03	9.0E-01	7.7E-01	5.9E-01	5.5E-01	5.0E-01	6.5E-01
	E3	1.4E-03	7.7E-01	6.1E-01	4.5E-01	4.2E-01	3.5E-01	5.0E-01

		gadB						
		AN	C1	C2	C3	E1	E2	E3
Wilcox	AN	5.1E-01	8.0E-01	**4.7E-06**	4.8E-01	**5.3E-05**	1.0E-02	1.5E-02
	C1	2.1E-01	5.1E-01	**2.7E-06**	2.8E-01	**4.3E-06**	**4.2E-04**	1.4E-03
	C2	1.0E+00	1.0E+00	5.0E-01	1.0E+00	4.7E-01	1.0E+00	1.0E+00
	C3	5.3E-01	7.3E-01	**1.7E-04**	5.0E-01	**2.8E-04**	3.4E-02	2.3E-02
	E1	1.0E+00	1.0E+00	5.4E-01	1.0E+00	5.1E-01	1.0E+00	1.0E+00
	E2	9.9E-01	1.0E+00	**2.4E-04**	9.7E-01	3.8E-03	5.0E-01	5.9E-01
	E3	9.9E-01	1.0E+00	**8.3E-04**	9.8E-01	4.7E-03	4.1E-01	5.0E-01
t-test	AN	5.0E-01	8.5E-01	2.0E-02	5.1E-01	4.2E-03	9.4E-03	4.0E-03
	C1	1.5E-01	5.0E-01	7.3E-03	2.1E-01	1.7E-03	7.7E-03	4.6E-03
	C2	9.8E-01	9.9E-01	5.0E-01	9.7E-01	4.2E-01	8.5E-01	8.1E-01
	C3	4.9E-01	7.9E-01	3.2E-02	5.0E-01	1.1E-02	6.6E-02	4.2E-02
	E1	1.0E+00	1.0E+00	5.8E-01	9.9E-01	5.0E-01	9.3E-01	9.1E-01
	E2	9.9E-01	9.9E-01	1.5E-01	9.3E-01	6.5E-02	5.0E-01	3.5E-01
	E3	1.0E+00	1.0E+00	1.9E-01	9.6E-01	9.4E-02	6.5E-01	5.0E-01

		hdeA						
		AN	C1	C2	C3	E1	E2	E3
Wilcox	AN	5.0E-01	2.2E-01	1.0E+00	**9.6E-04**	2.5E-02	1.3E-03	2.8E-01
	C1	7.8E-01	5.1E-01	1.0E+00	5.2E-03	1.2E-01	5.8E-03	5.2E-01
	C2	**5.7E-13**	**8.3E-07**	5.1E-01	**3.3E-10**	**6.6E-08**	**1.0E-08**	**2.2E-06**
	C3	1.0E+00	1.0E+00	1.0E+00	5.1E-01	1.0E+00	7.3E-01	9.9E-01
	E1	9.8E-01	8.8E-01	1.0E+00	1.6E-03	5.0E-01	4.1E-03	5.4E-01
	E2	1.0E+00	9.9E-01	1.0E+00	2.8E-01	1.0E+00	5.1E-01	9.7E-01
	E3	7.3E-01	4.9E-01	1.0E+00	8.8E-03	4.7E-01	2.8E-02	5.0E-01
t-test	AN	5.0E-01	4.8E-01	1.0E+00	1.0E-03	1.8E-01	1.0E-02	5.0E-01
	C1	5.2E-01	5.0E-01	1.0E+00	8.7E-03	2.4E-01	3.2E-02	5.2E-01
	C2	**1.4E-13**	**3.2E-08**	5.0E-01	**1.2E-14**	**3.0E-11**	**3.7E-12**	**3.0E-08**
	C3	1.0E+00	9.9E-01	1.0E+00	5.0E-01	9.7E-01	6.9E-01	9.9E-01
	E1	8.2E-01	7.6E-01	1.0E+00	3.1E-02	5.0E-01	1.0E-01	7.7E-01
	E2	9.9E-01	9.7E-01	1.0E+00	3.1E-01	9.0E-01	5.0E-01	9.7E-01
	E3	5.0E-01	4.8E-01	1.0E+00	8.8E-03	2.3E-01	3.2E-02	5.0E-01

		rpsM								
		AN	C1	C2	C3	C4	E1	E2	E3	E4
Wilcox	AN	5.0E-01	3.2E-02	**5.8E-12**	**7.6E-05**	**2.4E-04**	1.0E+00	1.0E+00	5.2E-01	5.7E-01
	C1	9.7E-01	5.0E-01	**2.0E-04**	1.8E-01	1.5E-01	1.0E+00	1.0E+00	7.5E-01	9.5E-01
	C2	1.0E+00	1.0E+00	5.0E-01	1.0E+00	9.9E-01	1.0E+00	1.0E+00	9.9E-01	1.0E+00
	C3	1.0E+00	8.2E-01	1.6E-03	5.0E-01	4.6E-01	1.0E+00	1.0E+00	9.3E-01	9.5E-01
	C4	1.0E+00	8.5E-01	1.3E-02	5.5E-01	5.0E-01	1.0E+00	1.0E+00	8.9E-01	9.6E-01
	E1	**1.9E-04**	**1.0E-04**	**1.6E-11**	**1.1E-07**	**3.7E-07**	5.0E-01	9.1E-01	1.5E-02	6.2E-02
	E2	**3.1E-08**	**1.6E-07**	**1.6E-11**	**3.0E-09**	**7.2E-08**	9.3E-02	5.0E-01	**5.0E-05**	1.0E-02
	E3	4.8E-01	2.6E-01	7.5E-03	6.9E-02	1.1E-01	9.9E-01	1.0E+00	5.0E-01	9.3E-01
	E4	4.4E-01	5.2E-02	**2.9E-05**	5.7E-02	4.0E-02	9.4E-01	9.9E-01	7.2E-02	5.0E-01
t-test	AN	5.0E-01	6.5E-02	**2.2E-13**	**3.9E-05**	8.0E-01	1.0E+00	1.0E+00	8.2E-01	9.6E-01
	C1	9.3E-01	5.0E-01	**2.6E-04**	8.5E-02	8.1E-01	1.0E+00	1.0E+00	9.2E-01	9.7E-01
	C2	1.0E+00	1.0E+00	5.0E-01	1.0E+00	8.6E-01	1.0E+00	1.0E+00	1.0E+00	9.8E-01
	C3	1.0E+00	9.2E-01	**9.4E-04**	5.0E-01	8.3E-01	1.0E+00	1.0E+00	9.7E-01	9.7E-01
	C4	2.0E-01	1.9E-01	1.4E-01	1.7E-01	5.0E-01	6.5E-01	6.9E-01	2.4E-01	6.7E-01
	E1	2.3E-03	1.5E-03	**5.3E-04**	1.0E-03	3.5E-01	5.0E-01	5.6E-01	5.3E-03	5.6E-01
	E2	**2.4E-04**	**1.5E-04**	**3.8E-05**	**8.6E-05**	3.1E-01	4.4E-01	5.0E-01	**7.5E-04**	5.3E-01
	E3	1.8E-01	7.9E-02	3.5E-03	2.5E-02	7.6E-01	9.9E-01	1.0E+00	5.0E-01	9.5E-01
	E4	3.8E-02	3.2E-02	2.0E-02	2.7E-02	3.3E-01	4.4E-01	4.7E-01	5.2E-02	5.0E-01

Table 2. Comparison of fitness between control populations (C1, C2, C3) and Experimental populations (E1, E2, E3). Significant ($p<0.001$) values are in bold.

		rpoS					
		C1	C2	C3	E1	E2	E3
Wilcox	C1	5.0E-01	1.5E-02	**2.0E-05**	1.3E-01	1.5E-01	1.5E-03
	C2	9.9E-01	5.0E-01	5.5E-02	8.7E-01	9.5E-01	4.0E-01
	C3	1.0E+00	9.4E-01	5.0E-01	1.0E+00	1.0E+00	9.8E-01
	E1	8.7E-01	1.3E-01	1.5E-03	5.0E-01	6.3E-01	4.7E-02
	E2	8.5E-01	5.4E-02	**9.3E-06**	3.7E-01	5.0E-01	3.8E-03
	E3	1.0E+00	6.0E-01	2.0E-02	9.5E-01	1.0E+00	5.0E-01
t-test	C1	5.0E-01	4.1E-02	**3.0E-05**	9.8E-02	5.0E-02	1.2E-03
	C2	9.6E-01	5.1E-01	2.5E-01	8.5E-01	9.7E-01	7.3E-01
	C3	1.0E+00	7.6E-01	5.0E-01	1.0E+00	1.0E+00	9.8E-01
	E1	9.1E-01	1.5E-01	2.5E-03	5.0E-01	6.6E-01	7.8E-02
	E2	9.5E-01	2.9E-02	**9.6E-06**	3.4E-01	5.1E-01	3.4E-03
	E3	1.0E+00	2.7E-01	2.0E-02	9.3E-01	1.0E+00	5.1E-01

		prgH					
		C1	C2	C3	E1	E2	E3
Wilcox	C1	5.0E-01	1.2E-02	**1.8E-04**	**7.1E-05**	**7.3E-07**	1.4E-01
	C2	9.9E-01	5.0E-01	1.1E-01	6.7E-02	2.5E-02	7.7E-01
	C3	1.0E+00	9.0E-01	5.0E-01	3.3E-01	2.5E-01	9.8E-01
	E1	1.0E+00	9.4E-01	6.8E-01	5.0E-01	3.9E-01	9.6E-01
	E2	1.0E+00	9.8E-01	7.5E-01	6.2E-01	5.0E-01	9.9E-01
	E3	8.7E-01	2.4E-01	2.2E-02	4.1E-02	6.9E-03	5.0E-01
t-test	C1	5.0E-01	1.1E-02	**8.3E-05**	**1.7E-04**	**7.3E-07**	7.4E-02
	C2	9.9E-01	5.0E-01	8.8E-02	9.7E-02	1.4E-02	7.5E-01
	C3	1.0E+00	9.1E-01	5.0E-01	5.0E-01	2.1E-01	9.7E-01
	E1	1.0E+00	9.0E-01	5.0E-01	5.0E-01	2.3E-01	9.7E-01
	E2	1.0E+00	9.9E-01	7.9E-01	7.7E-01	5.0E-01	1.0E+00
	E3	9.3E-01	2.5E-01	2.7E-02	3.1E-02	3.4E-03	5.0E-01

Table 2 (**continued**). Comparison of fitness between control populations (C1, C2, C3) and Experimental populations (E1, E2, E3). Significant ($p<0.001$) values are in bold.

		gadB					
		C1	C2	C3	E1	E2	E3
Wilcox	C1	5.1E-01	3.1E-01	2.9E-02	5.0E-01	7.7E-01	**1.1E-04**
	C2	7.0E-01	5.0E-01	8.6E-02	6.8E-01	9.0E-01	1.0E-03
	C3	9.7E-01	9.2E-01	5.0E-01	9.8E-01	9.9E-01	2.0E-01
	E1	5.1E-01	3.3E-01	2.6E-02	5.1E-01	5.9E-01	1.8E-03
	E2	2.3E-01	1.0E-01	6.8E-03	4.2E-01	5.0E-01	**5.3E-06**
	E3	1.0E+00	1.0E+00	8.1E-01	1.0E+00	1.0E+00	5.0E-01
t-test	C1	5.0E-01	2.9E-01	4.3E-02	6.5E-01	4.6E-01	4.3E-03
	C2	7.1E-01	5.0E-01	1.0E-01	8.2E-01	7.0E-01	8.5E-03
	C3	9.6E-01	9.0E-01	5.0E-01	9.8E-01	9.8E-01	5.1E-02
	E1	3.5E-01	1.8E-01	2.1E-02	5.0E-01	2.9E-01	2.2E-03
	E2	5.4E-01	3.0E-01	1.6E-02	7.1E-01	5.0E-01	**2.3E-04**
	E3	1.0E+00	9.9E-01	9.5E-01	1.0E+00	1.0E+00	5.0E-01

		hdeA					
		C1	C2	C3	E1	E2	E3
Wilcox	C1	5.1E-01	1.9E-03	9.6E-01	4.5E-01	1.0E+00	9.2E-01
	C2	1.0E+00	5.1E-01	1.0E+00	1.0E+00	1.0E+00	1.0E+00
	C3	4.6E-02	**3.7E-04**	5.1E-01	3.8E-03	9.2E-01	4.1E-01
	E1	5.6E-01	**7.3E-04**	1.0E+00	5.0E-01	1.0E+00	9.8E-01
	E2	3.9E-03	**1.8E-04**	8.6E-02	**2.9E-04**	5.1E-01	5.6E-02
	E3	8.6E-02	**3.1E-04**	6.0E-01	1.7E-02	9.5E-01	5.0E-01
t-test	C1	5.0E-01	6.1E-01	8.3E-01	6.1E-01	9.9E-01	8.8E-01
	C2	3.9E-01	5.0E-01	5.7E-01	4.8E-01	8.9E-01	6.7E-01
	C3	1.7E-01	4.3E-01	5.0E-01	3.8E-01	9.7E-01	7.0E-01
	E1	3.9E-01	5.2E-01	6.2E-01	5.0E-01	9.4E-01	7.3E-01
	E2	1.1E-02	1.1E-01	3.5E-02	5.9E-02	5.0E-01	1.1E-01
	E3	1.2E-01	3.3E-01	3.0E-01	2.7E-01	8.9E-01	5.0E-01

Table 2 (**continued**). Comparison of fitness between control populations (C1, C2, C3) and Experimental populations (E1, E2, E3). Significant ($p<0.001$) values are in bold.

		rpsM							
		C1	C2	C3	C4	E1	E2	E3	E4
Wilcox	C1	5.0E-01	5.0E-01	9.7E-01	1.0E+00	**2.0E-04**	9.1E-03	2.2E-02	3.1E-02
	C2	5.1E-01	5.0E-01	9.7E-01	1.0E+00	**8.3E-04**	5.1E-03	9.3E-02	2.2E-02
	C3	3.4E-02	2.9E-02	5.0E-01	9.8E-01	**5.9E-06**	**7.2E-04**	**9.1E-04**	1.5E-03
	C4	**3.9E-05**	**1.3E-04**	2.3E-02	5.0E-01	**2.3E-07**	**5.5E-05**	**1.9E-06**	**1.9E-04**
	E1	1.0E+00	1.0E+00	1.0E+00	1.0E+00	5.0E-01	6.9E-01	9.7E-01	8.3E-01
	E2	9.9E-01	1.0E+00	1.0E+00	1.0E+00	3.1E-01	5.0E-01	9.5E-01	5.0E-01
	E3	9.8E-01	9.1E-01	1.0E+00	1.0E+00	3.2E-02	5.1E-02	5.0E-01	1.4E-01
	E4	9.7E-01	9.8E-01	1.0E+00	1.0E+00	1.7E-01	5.1E-01	8.6E-01	5.0E-01
t-test	C1	5.0E-01	5.9E-01	9.7E-01	1.0E+00	**2.3E-04**	7.7E-02	5.5E-02	1.7E-01
	C2	4.1E-01	5.0E-01	9.5E-01	1.0E+00	**2.3E-04**	6.0E-02	4.1E-02	1.4E-01
	C3	2.8E-02	5.0E-02	5.0E-01	9.7E-01	**3.0E-06**	2.7E-03	**7.2E-04**	1.3E-02
	C4	**1.3E-04**	**3.3E-04**	2.7E-02	5.0E-01	**1.1E-08**	**1.6E-05**	**1.8E-06**	**1.9E-04**
	E1	1.0E+00	1.0E+00	1.0E+00	1.0E+00	5.0E-01	9.0E-01	9.9E-01	9.2E-01
	E2	9.2E-01	9.4E-01	1.0E+00	1.0E+00	9.9E-02	5.0E-01	6.1E-01	6.0E-01
	E3	9.4E-01	9.6E-01	1.0E+00	1.0E+00	1.3E-02	3.9E-01	5.0E-01	5.2E-01
	E4	8.3E-01	8.6E-01	9.9E-01	1.0E+00	8.3E-02	4.0E-01	4.8E-01	5.0E-01

R script for gating and extracting fluorescent expression data

This script is use

```
#! /usr/bin/env Rscript
#args <- commandArgs(TRUE)

# READ libraries. Note: You need to install several packages to get this script to run.
library(flowCore)
library(flowViz)
library(geneplotter)
library(MASS)

# DEFINE parameters!
MIN.SSC=1.5 # minimum SSC.H value (will not include anything below this value)
MIN.FSC=1.5 # minimum FSC.H value (will not include anything below this value)
MIN.FL1=1.001 # minimum FL1.H value (will not include anything below this value)
MIN.CELLS=5000 # minimum number of cells to gate
BINS=50

#size of pdf to print
PDF.W <- 2
PDF.H <- 2
PDF.C <- 12
PDF.R <- 8
```

```
# If you want plots of the histograms used for ananlysis <- 1, if not <-0
PLOT <- 1
# If you want FSC and SSC plots <- 1, if not <-0
FSPLOT <- 0

#### SPECIFY FILE NAMES BELOW.
### Get the DIRECTORY list and name the analysis location
### make a text file containing a list of DIRECTORIES here (even if it's only a single file)
### This program will pull **ALL** files from this directory that have an ".fcs" at the end
tab.dir.list <- paste("/Users/nikki/file_location/dir_list.txt", sep="")
### put the location of where you want the analysis files to be placed here
analysis.location <- "/Users/nikki/file_location/Analyzed.txt"


start.time = Sys.time()
dir.list <- scan(file=tab.dir.list, what="", sep=",")

# initialize vars
x=0
pdf.count <- 0
quartz()

### this is a loop around the whole thing for using multiple file directories ---
for(xx in 1:length(dir.list)) {

dir.name <- dir.list[xx]

# READ FILES
print ("reading filenames")

fcs.names=dir(dir.name,full.names=TRUE, pattern="")
#fcs.names=dir(dir.name,full.names=TRUE)

# How many files have been read
nof=length(fcs.names)
cat("number of files read: ")
cat(nof, "\n")

# DEFINE some useful transformations
# r and d are scale factors, usually set to 1
logT=logTransform(transformationId="log10-transformation", logbase=10, r=1, d=1)
biexpT=biexponentialTransform(transformationId="biexp")

#Define table
stat.all=data.frame(cbind(SAMPLE.ID=0,TIME=0,CELLS=0,
  FSC.H.Av=0,SSC.H.Av=0,GATE.D=0, GFP.H.Av=0,GFP.H.SD=0))

# loop through files and create statistics and figures
for (fcs.name in fcs.names){
        x=x+1
### open a new pdf file everytime PDF.R*PDF.C analyses are completed
        if(PLOT) {
        if (x %% (PDF.R*PDF.C) == 1) {
                pdf.count <- pdf.count+1
                dev.off()
```

```
            pdf.name <- paste(analysis.location, "/fcs", pdf.count, ".pdf", sep="" )
            while (file.exists(pdf.name)) {
                  pdf.count <- pdf.count+1
                  pdf.name <- paste(analysis.location, "/fcs", pdf.count, ".pdf", sep="" )
            }
            pdf(file = pdf.name, height=PDF.R*PDF.H,width=PDF.C*PDF.W)
            par(mfrow=c(PDF.R, PDF.C))
      }
      }

      if (x %% (5) == 1 && x>2) {
            cat(x, " of ", length(fcs.names), "files analyzed in", xx, "of", length(dir.list), "dirs.\n")
            cur.time = Sys.time()
            time.left <- difftime(cur.time, start.time, units="mins")
            mean.process <- time.left/((xx-1)*length(fcs.names) + x)
            files.left <- length(fcs.names)*length(dir.list)-(length(fcs.names)*(xx-1)+x)
            time.left <- mean.process*files.left
            cat("Appr.", round(time.left,1), "mins left", "-- per file:", round(mean.process, 4), "mins -
-", files.left, "files left.\n")
      }

  cat("\nReading file", x, "-- ")
  fs <- read.FCS(filename=fcs.name, alter.names=1)

# Gate away v. low values
not.debris.gate = rectangleGate(filterId="nonDebris",
"FSC.H"=c(MIN.FSC,262143),"SSC.H"=c(MIN.SSC,262143))
fs.not.debris = Subset( fs, not.debris.gate)
flh.gate = rectangleGate(filterId="flh", "FL1.H"=c(MIN.FL1,262143))
fs.not.debris = Subset( fs.not.debris, flh.gate)

# Gate the beads away
  cat("Gating beads -- ")
  #bead.gate=rectangleGate(filterId="beads",
"Crimson.1.H"=c(10000,262143),"Crimson.2.H"=c(2000,262143))
  bead.gate=rectangleGate(filterId="beads", "FSC.H"=c(500,262143),"SSC.H"=c(1000,262143))
  fs.cells = Subset (fs.not.debris, !bead.gate)
  fs.beads = Subset (fs.not.debris, bead.gate)
  #summary(fs)

fs.cellsT <- transform(fs.cells,`FL1.H`=logT(`FL1.H`), `FSC.H`=logT(`FSC.H`),
`SSC.H`=logT(`SSC.H`))
fs.beadsT <- transform(fs.beads,`FL1.H`=logT(`FL1.H`), `FSC.H`=logT(`FSC.H`),
`SSC.H`=logT(`SSC.H`))
#plot(fs.cellsT, c('FSC.H','SSC.H'), xlim=c(0,2.5), ylim=c(0,2.5))

# Do some gating on the cells via FSC and SSC
cat("Gating on FSC and SSC")
# Filter out the noise first
n2f <- norm2Filter(x=c("FSC.H", "SSC.H"), method="covMcd", scale.factor=2, n=50000,
filterId="defaultNorm2Filter")
temp.cells <- Subset(fs.cellsT, n2f)

# Get a robust estimate of the densest portion of the graph -- this will be the "mean"
fsc <- exprs(temp.cells$"FSC.H")
```

```
ssc <- exprs(temp.cells$"SSC.H")
cells.dens <- kde2d( fsc, ssc, n = BINS, lims = c(range(fsc), range(ssc)) )
fsc.d <- cells.dens$x[which(cells.dens$z == max(cells.dens$z)) %% BINS]
ssc.d <- cells.dens$y[floor(which(cells.dens$z == max(cells.dens$z))/BINS) + 1]
mean <- c("FSC"=fsc.d, "SSC"=ssc.d)

# Calculate the covariance (from which -- fs.cells or temp.cell) for the ellipse gate
#fsc <- exprs(fs.cellsT$"FSC.H")
#ssc <- exprs(fs.cellsT$"SSC.H")
cov <- matrix(c(var(fsc), cov(fsc,ssc), cov(fsc,ssc), var(ssc)), ncol=2, dimnames=list(c("FSC.H",
"SSC.H"), c("FSC.H", "SSC.H")))

dist = 0.08
total.cells=0
if(dim(fs.cellsT)[1] > MIN.CELLS) {
        while(total.cells<MIN.CELLS) {
                dist = dist + 0.01
                sim.gate <- ellipsoidGate(filterId="similar.cells", .gate=cov, mean=mean, distance=dist)
                fs.cellsA <- Subset(fs.cellsT, sim.gate)
                total.cells <- dim(fs.cellsA)[1]
        }
}
else { fs.cellsA <- fs.cellsT }

# DEFINE statistics
#cat("calculating statistics\n")

# Fill List with FCS VARS
  stat.all[x,1]=as.character(keyword(fs.cellsA,c("SAMPLE ID")))
  stat.all[x,2]=as.character(keyword(fs.cellsA,c("$BTIM")))
  stat.all[x,3]=as.double(dim(fs.cellsA)[1])

# Fill List with statistics (mean) fs.cells
  stat.all[x,4]=as.double(mean(exprs(fs.cellsA$FSC.H)))
  stat.all[x,5]=as.double(mean(exprs(fs.cellsA$SSC.H)))
  stat.all[x,6]=dist
  stat.all[x,7]=as.double(mean(exprs(fs.cellsA$FL1.H)))

# Fill List with statistics (SD) fs.cells
  stat.all[x,8]=as.double(sd(exprs(fs.cellsA$FL1.H)))

# Generate plots
plot.name <- paste(keyword(fs.cellsA,c("SAMPLE ID")), keyword(fs.cellsA,c("$BTIM")))
plot.label <- paste("mn:", round(stat.all[x,7], digits=2), "sd:", round(stat.all[x,8], digits=2), "N: ",
stat.all[x,3])
gfp.dens <- density(exprs(fs.cellsA$FL1.H))

plot(gfp.dens$x, gfp.dens$y, ty="l", main=plot.name, xlab ="log10 GFP (AU)", ylab="Kernel
Density", lwd=1, sub=plot.label, xlim=c(0.0,4.0))

if (FSPLOT){
        cat("\nplotting FSC-SSC\n")
        xyplot(`SSC.H` ~ `FSC.H`, fs.cellsT, nbin = 100, main=plot.name, smooth="T", filter=sim.gate,
xlim=c(0,2.5), ylim=c(0,2.5))
}
```

```
stat.all[x,]
}
}
dev.off()

# Sort data by time of measurement (this provides the DIVA layout)
stat.all.ordered = stat.all[order(stat.all$TIME,stat.all$SAMPLE.ID),]

# write out the data in csv-style
# filename=paste(args[2],".csv",sep="")

print ("writing results")
write.csv (stat.all.ordered, file=paste(analysis.location, "/stat_all.csv", sep=""))
```

Acknowledgements

I would like to thank the Swiss National Foundation for providing the funding for this research.

I am very grateful to many people who have provided enormous support the past four years. There have been many people who made an incredible impact on the work presented in this thesis. My family has provided much support over the last four years and have been very understanding of my decision to live so far away. A short conversation with my dad was critical for making the decision to do this PhD, and I am very grateful to him for always listening and offering good advice.

I would not have a working computer, pipette, tips, reagents, or a paycheck without the always cheerful Roland Loosli, Rita Jenny, and Christine Weber. I am continually amazed at their skill and efficiency, and I don't believe I have heard one of them complain about anything in four years. Nothing would run as smoothly around here without them. I am indebted to Regula Schmidt-Hempel, who always knows where everything is and is always very helpful in so many ways.

In terms of techniques, Tobias Bergmiller deserves very special recognition for bringing me up to date on many of the techniques I knew nothing about and providing an awesome starting strain collection from which to work with. Alex Böhm tirelessly fielded emails, phone calls, and even lab visits from me when I had no idea what I was doing. Alex is a powerhouse of reliable knowledge of all things bacterial. As well, Bärbel Stecher was incredibly helpful for providing knowledge, strains, and good advice concerning all things salmonella.

The numerous members (past and present) of Paul Schmidt-Hempel, Sebastian Bonhoeffer, and Wolf-Deitrich Hardt's groups have provided tremendous scientific input and social interaction along the way. From the first day I have felt welcomed here and many people went out of their way to make sure Olin and I were settled in Zürich. TB and Wednesday Group meetings were a critical part of my scientific experience at ETH and I appreciate all of the great minds who listened and commented during those talks. As well, the innumerable parties, lab hiking, sledding, and skiing trips were so much fun!

I am very lucky to share so much of my life and work with Olin Silander and Martin Ackermann, both of whom really make *everything* so so so much better. It is a true pleasure to work with two of the smartest people around and two of the happiest. Martin is the best boss a girl can ask for and Olin is an amazing collaborator (and husband!). Both have been

supportive of me when I have had emotional breakdowns, failed experiments, contaminations, and have gently guided me away from some bad scientific ideas.

Specifically, thanks to Martin for all of his tremendous help, cheerleading, and encouragement during my PhD. Also for showing me all the cool stuff there is to do in Zürich. And for finding us a great place to live. And for some great BBQs. And for Letten swims. And… the list goes on and on.

Words cannot capture or due justice to how thankful I am to Olin. I am so lucky for the amazing life we share. Every minute of my life (including most of the work in this thesis!) has been better because of you.

NICOLE (NIKKI) E. FREED

CHN J14 ETHZ, Universitätsstr. 16
8092 Zurich, Switzerland
http://nikkifreed.googlepages.com

work:+41 44 632 3983
freednikki@gmail.com
Citizenship: American

EDUCATION

2005-2009	ETH Zurich, Switzerland Ph.D., Institute of Integrative Biology, Microbial Evolution *Thesis title: Exploring The Evolutionary Causes And Consequences Of Stochastic Variation In Gene Expression In Bacteria*
2000-2002	University of California, San Diego (UCSD) M.S., Ecology, Behavior, and Evolution *Thesis title: Microsatellite-length polymorphisms : an experimental model to explore effects on cekl regulation of hyphal formation in Candida albicans*
1996-2000	University of California, San Diego (UCSD) B.S., Biochemistry and Cell Biology

WORK EXPERIENCE

2002-2005	Molecular lab researcher, respiratory disease detection. Full time. Naval Health Research Center, Department of Defense, San Diego, CA, USA.
1999	Lab assistant. Part time. Salk Institute, Infectious Disease Laboratory, HIV Infection, Dr. Ned Landau.
1996-1998	Lab assistant. Part time. Salk Institute, Plant Molecular and Cell Biology, Dr. Joanne Chory.

TEACHING EXPERIENCE

2009	Teaching assistant. Microbiology Lab Course. ETH Zurich.
2007	Guest Lecture. Cell and Molecular Biology for Engineers. ETH Zurich.
2000-2002	Teaching assistant. UCSD. - Biochemical Techniques Laboratory Course. Spring 2001, Fall 2001, Fall 2002. - Molecular Biology. Fall 2000. Including a guest lecture on DNA repair. - Structural Biochemistry. Winter 2001.

SKILLS

Expertise in a broad range of molecular methods: microbial genetics (cloning, production of gene knockouts, plasmid library production), microarrays, high-throughput methods using robotics, quantitative RT-PCR and PCR, sequencing and sequence analysis, flow cytometry, fluorescent microscopy. Proficient in mammalian cell culture, transfection, and transduction.

Excellent knowledge of bacterial gene expression, good familiarity with the requirements for clinical lab certification and accreditation (similar to Good Clinical Practices), Bio-Safety Level Three standards and certification, experience in writing peer-reviewed publications, and supervision/mentoring of Masters' students, undergraduates, and technicians.

AWARDS

2008	EMBO Best Poster Prize at *Bacterial Networks* conference
2000 and 2001	UCSD Biology Excellence in Teaching Award
2000	UCSD Julia Brown Undergraduate Research Scholarship

PUBLICATIONS

Freed NE, Silander OK, Stecher B, Boehm A, Hardt WD, Ackermann M
A simple screen to identify promoters conferring high levels of phenotypic noise. PLOS GENETICS (2008) 4(12).

Ackermann M, Stecher B, **Freed NE**, Songhet P, Hardt WD, Doebeli M
Self-destructive cooperation mediated by phenotypic noise. NATURE (2008) 454: 987-990

Blyn LB, Hall TA, Libby B, Ranken R, Sampath R, Rudnick K, Moradi E, Desai A, Metzgar D, Russell KL, **Freed NE**, Balansay M, Broderick MP, Osuna MA, Hofstadler SA, Ecker DJ
Rapid Detection and Molecular Serotyping of Adenovirus by Use of PCR Followed by Electrospray Ionization Mass Spectrometry. CLINICAL MICROBIOLOGY (2008) 46(2):644-51

Freed NE, Myers CA, Russell KL, Walter EA, Irvine M, Coon RG, Metzgar D.
Diagnostic discrimination of live attenuated influenza vaccine strains and community-acquired pathogenic strains in clinical samples. MOLECULAR AND CELLULAR PROBES (2007) 21(2):103-10

Russell, KL, Broderick, MP, Franklin, SE, Blyn, LB, **Freed, NE**, Moradi, E, Ecker, DJ, Kammerer, PE, Osuna, MA, Kajon, AE, Morn, CB, Ryan, MAK
Transmission dynamics and prospective environmental sampling of adenovirus in a military recruit setting. INFECTIOUS DISEASES (2006) 194:877-885

Vora, GJ, Lin, BC, Gratwick, K, Meador, C, Hansen, C, Tibbetts, C, Stenger, DA, Irvine, M, Seto, D, Purkayastha, A, **Freed, NE**, Gibson, MG, Russell, K, Metzgar, D.
Co-infections of adenovirus species in previously vaccinated patients. EMERGING INFECTIOUS DISEASES (2006) 12:921-930.

Russell KL, Ryan MAK, Hawksworth A, **Freed NE**, Irvine M, Daum LT
Effectiveness of the 2003-2004 influenza vaccine among US military basic trainees: a year of suboptimal match between vaccine and circulating strain. VACCINE 23 (2005) (16): 1981-1985

Ecker DJ, Sampath R, Blyn LB, Eshoo MW, Ivy C, Ecker JA, Libby B, Samant V, Sannes-Lowery KA, Melton RE, Russell K, **Freed N**, Barrozo C, Wu JG, Rudnick K, Desai A, Moradi E, Knize DJ, Robbins DW, Hannis JC, Harrell PM, Massire C, Hall TA, Jiang Y, Ranken R, Drader JJ, White N, McNeil JA, Crooke ST, Hofstadler SA
Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 102 (2005) (22): 8012-8017.

Blasiole DA, Metzgar D, Daum LT, Ryan MAK, Wu JG, Wills C, Le CT, **Freed NE**, Hansen CJ, Gray GC, Russell KL
Molecular analysis of adenovirus isolates from vaccinated and unvaccinated young adults. CLINICAL MICROBIOLOGY (2004) 42 (4): 1686-1693.

LANGUAGES

English: Fluent (mother tongue)
German: A2 level (currently enrolled)
Spanish: A1 level