Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Omar E. Cornejo Date

Population Dynamics and Population Genetics of Recombination in Bacteria

By

Omar E. Cornejo

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science

Population Biology, Ecology, Evolution

Bruce R. Levin, Ph.D.

Advisor

Rustom Antia, Ph.D.

Committee Member

Keith Klugman, Ph.D.

Committee Member

Sergei Pilyugin, Ph.D.

Committee Member

Mark Jensen, Ph.D.

Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.

Dean of the Graduate School

Date

Population Dynamics and Population Genetics of Recombination in Bacteria

By

Omar E. Cornejo

MSc., Instituto Venezolano de Investigaciones Cientifficas, 2004

B.S., Universidad Simon Bolivar, 2000

Advisor: Bruce R. Levin, Ph.D.

An abstract of a dissertation

submitted to the Faculty of the

Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological and Biomedical Sciences

(Population Biology, Ecology, Evolution)

2009

Population Dynamics and Population Genetics of Recombination in Bacteria

By

Omar E. Cornejo

Although a rare event and unnecessary for reproduction, for many species of bacteria, recombination broadly defined to include the acquisition of genes from other, at times very divergent, organisms plays a fundamental role as a source of variation for adaptive evolution. What about more mundane horizontal gene transfer (HGT) events: homologous gene recombination (HGR) among members of the same population? While there no question about central role of HGR has played as tool for understanding the genetics and molecular biology of bacteria, the role of this kind of gene shuffling for the evolution of bacteria is not at all clear. The motivation for the research in this dissertation was to increase our understanding of the contribution of HGR to adaptive evolution in bacteria, reviewed in the background Chapter 1. And, save for serendipitous observation, the profound toxin-mediated oscillations in the densities of *Streptococcus pneumoniae* in continuous culture described in Chapter 2, the population and evolutionary dynamics of HGR in bacteria is the focus of this dissertation. The third chapter explores the factors that constrain the extent to which different strains of S. pneumoniae engage in shuffling their genes by HGR. The fourth chapter, we use a semi-stochastic simulation of mutation, selection and recombination to ascertain the conditions under which, and extent to which, HGR accelerates the rate at which bacterial populations adapt to their environment. More importantly, we assess the conditions under which the capacity to increase rates of adaptive evolution will select for the evolution and maintenance of mechanisms for HGR while competing recombining vs non-recombining populations.

Population Dynamics and Population Genetics of Recombination in Bacteria

By

Omar E. Cornejo

MSc., Instituto Venezolano de Investigaciones Cientifficas, 2004

B.S., Universidad Simon Bolivar, 2000

Advisor: Bruce R. Levin, Ph.D.

A dissertation submitted to the Faculty of the
Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biological and Biomedical Sciences
(Population Biology, Ecology, Evolution)

2009

Acknowledgements

I believe, after three dissertations or theses (undergrad, master and this final one), that the acknowledgements is probably the hardest section to write. At this point in time I am tired, and I feel like I need some time to think about the results of these last 5 years of work. Nevertheless, the manuscript must be submitted and I have to move up to Cornell at the end of the week. I don't think that I will follow any particular order... if it feels like I have followed one, it has most likely been unconscious. I want to thank Bruce R. Levin, my advisor and friend. During the last four years and a half I have worked constantly arguing with Bruce. As I have told him repeatedly: "I do love you man, but you have been my advisor and my opponent during my whole PhD studies". Bruce has proven to be my hardest critic and the most difficult opponent. That is probably the most important training that I have had with him... beyond the scientific. Nevertheless, I know that I can count on Bruce, the advisor and the friend, in the long years to come. I expect he feels the same way about me. Don't give up Bruce, I might surprise you one day and become a full experimentalist.

I want to thank Rustom Antia, a member of my committee and friend who over the years have supported me and helped me overcome many frustrations. I do not know how to thank him and his family for the friendship that I will treasure the rest of my life. don't forget we have to keep climbing as much as we can.

I want to thank the members of my committee: Keith Klugman, Mark Jensen and Sergei

Pilyugin fro their support and readiness to keep up with my lack of organization. Sergei, I still owe you a paper man. I haven't forgotten.

I would like to thank all my friends in the PBEE program. Specially those that made my life bearable all these years: Sean, Karoun, Ellie, Yan, Alaine, Lynn. They know how much I felt miserable at times and I would like them to know how much their company ease my life. If I said that only them helped it I would be lying. I would like to thank Roland, Andy, Andreas, Pablito, David, Emilie, Katie, Helder, Julie, Maira, Josh, Klas, Pal, Pierre, Javier, Guaniri, Marcim, and I am afraid that if I keep naming I will eventually miss somebody and I will feel ashamed of myself for forgetting them while I write... Life, I believe, is made up of little moments that make it worth living, spaced by abysses of steadiness. You my friends are and will be part of those worthy moments. Ahora escribo en espanhol. Por que no? Claro que no esperaban esto pero es que los agradecimientos de una tesis deben ser hechos de la manera debida. Quiero agradecer a mis padres y mi hermano por el apoyo incondicional que siempre han tenido conmigo. Nada de esto habria sido posible sin que mis padres siempre dijeran: "si, por que no? si eso es lo que quieres hacer, pues dale". No es solo el decir: dale. Es el haberme apoyado activamente para que pudiera tener exito en la empresa que me propusiera. Mi hermano, para quien espero que este logro sea un estimulo y entienda que no existe otra manera de vivir mas que seguir adelante a pesar de las adversidades. Que entienda que lo unico que somos y que seremos es lo que hacemos con nuestras vidas y como nos relacionamos con el resto del mundo. Haz lo que quieres hacer, pero hazlo con seriedad y entrega mi gordo. Quiero agradecer a mi familia: tios, tias, primos, primas... porque siempre estan alli, pendientes del curso que sigue mi vida y que a pesar de la distancia me mantienen cerca de sus vidas como que si estuviera alli, con ellos.

Quiero agradecer a Tina... mi amor de tanto tiempo. Tan lejos y tan cerca a momentos. Me has apoyado tanto. Siempre he contado contigo para todo. Quizas es injusto lo que voy a escribir. Pero seria inclusive mas injusto no hacerlo. Que tanto mas dificil fue sin

ti. Lamento que no hallamos sido capaces de lograr un balance que nos permitiera estar juntos, seguir juntos. Todo esto me trae ecos de mis propios versos: ”Creo en la carencia de amores infinitos y de amores infimos/ se ama todo y de una vez sin la oportunidad de un para siempre”.

Entiendo que estos agradecimientos pareceran un poco grises cuando vuelva a leerlos en un tiempo. Pero espero recordar que ello es tan solo el reflejo de mi estado animico al final de este doctorado. Agotado, vacilante y con un desgaste intelectual y emocional que amerita tratamiento.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The main concern of Population and Evolutionary Genetics is to understand the nature and relative contributions of the different processes that determine the frequencies of genes in populations. A particular focus of this enterprise is elucidating those processes responsible for generating and maintaining the genetic variation upon which natural and not-so-natural selection operates. For all organisms, mutation in its various guises is the primary source of this variation. In eukaryotes, and particularly those that reproduce sexually, genes are continually being shuffled by recombination. As a result, the standing and generated genetic variation is well distributed among members of a population. In the bacteria, reproduction is by binary fission. Recombination is a rare event and populations are largely collections of genetically distinct lineages, clones, but not entirely so. For some (many?) species of bacteria recombination in the form of the acquisition of novel genes and genetic elements from other populations of the same and sometimes very different species plays a prominent role as source of genetic variation for adaptive evolution, i.e. Horizontal (lateral) Gene Transfer (HGT) (Shea et al. 1996; Lawrence and Ochman 1997; Bergstrom, Lipsitch, and Levin 2000; Campbell 2000; Levin and Bergstrom 2000; Koonin, Makarova, and Aravind 2001; Cazalet et al. 2004; Thomas and Nielsen 2005; Coleman et al. 2006; Gal-Mor and Finlay 2006). Less clear are

the evolutionary consequences of recombination shuffling homologous genes among members of the same species population, Homologous Gene Recombination (HGR).

Be it HGT moving a novel gene from an exotic source or more mundane HGR, recombination in bacteria is unidirectional process where recipient cells gain genetic material from a donor in a conservative or non-conservative fashion (see Figure 1.1). When recombination is mediated by conjugative plasmids, a copy of the gene is transferred from a donor cell to a recipient. (Smith 1988; Koonin, Makarova, and Aravind 2001). A small fraction of population bearing the plasmid carries this gene-transferring element integrated into its chromosome, making the chromosome "think' it's a plasmid. By conjugation with a cell free of this element a copy of the leading edge of the plasmid along with the adjacent chromosomal genes are moved to the recipient. The subset of cells carry the chromosome integrated plasmid transfers genes at a high rate and, when isolated as single population, this subset is known as a High Frequency Recombinant, Hfr (Hayes 1968). The genetic material obtained by the recipient cell might be degraded or may be incorporated into the chromosome of the recipient, presumably as an accident of rec-mediated DNA repair (Smith 1988; Fall et al. 2007). Once there is a rupture of the physical contact between the donor and recipient, gene transfer ceases.

There are two major forms of non-conservative recombination in bacteria. First, the transfer of the genes by phage, a phenomenon called transduction, in which the donor dies when the phage are released. Second the incorporation of exogenous DNA, transformation. Presumably, but possibly not surely, the donor of that DNA died when

it was released. Natural transformation is a highly evolved character and not one that is shared by all bacteria. Those with capability have to be "competent" for the uptake of exogenous DNA, a process that involves the coordinated action of a number of bacterial genes (Smith, Danner, and Deich 1981; Stewart and Carlson 1986; Smith 1988).

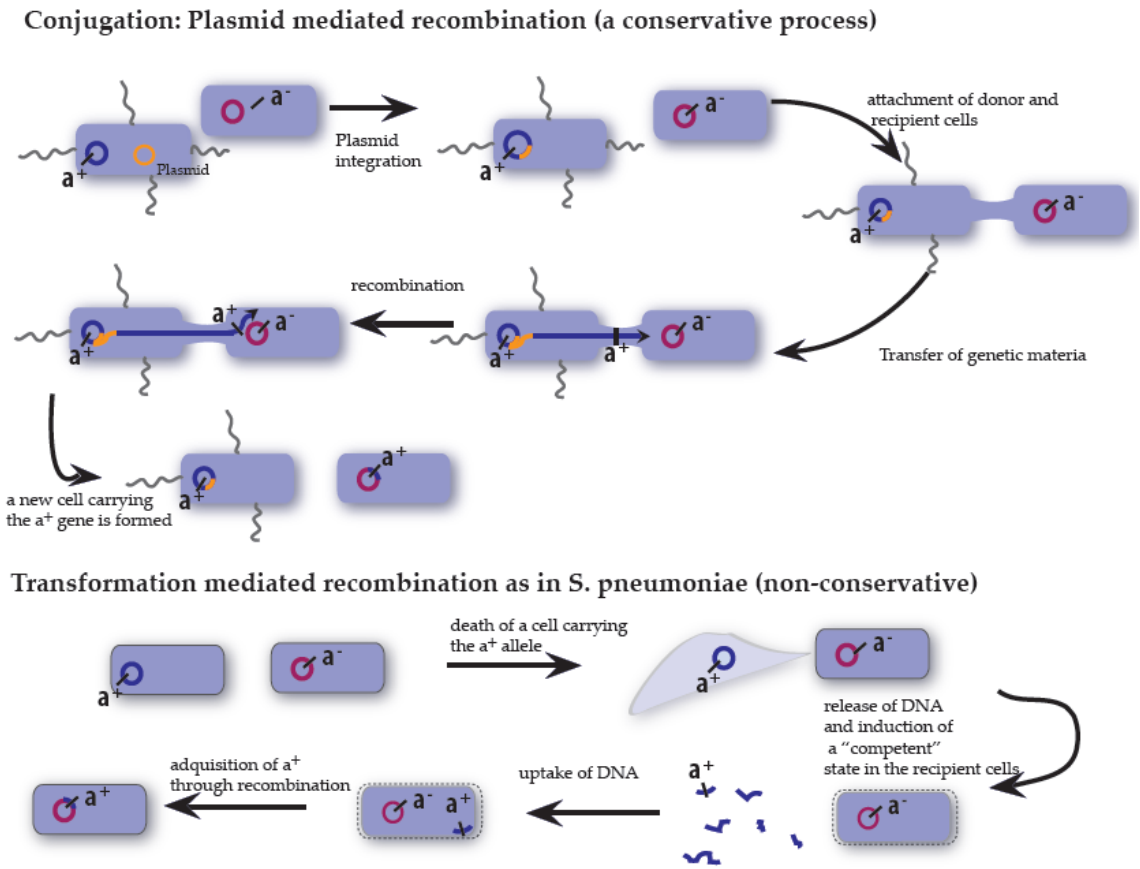

Figure 1.1. Simplified diagram representing Conservative (a), plasmid mediated, and non Conservative (b), natural transformation mediated, mechanisms of recombination in bacteria. In the conservative scenario (a), the donor bacteria keeps the gene (a+) that is transmitted to a recipient bacteria. In the case of non conservative transmission of the gene, the original carrier disappears from the population and only the recipient carries it.

Because recombination shuffles of genes among individual cells from a population genetic point of view, it can disrupt the linkage disequilibria (LD) associated with clonal reproduction. Moreover, in accord with classical population genetic theory, HGR enables bacteria to assemble in single cells adaptive combinations of genes present

in different cells (Felsenstein 1974; Felsenstein and Yokoyama 1976; Otto and Gerstein 2006). It also breaks down accumulating combinations of deleterious mutations, "Muller's ratchet " (Muller 1932; Muller 1964). The extent to which HGR breaks down LD and has positive contribution to adaptive evolution by avoiding the genetic decay by Muller's ratchet, depends on its rate of occurrence. If HGR occurs at too low a rate, the accumulating recombinants will be purged by the bottlenecks associated with the formation of new populations or replacement by higher fitness clones, periodic selection (Atwood, Schneider, and Ryan 1951b; Atwood, Schneider, and Ryan 1951a; Levin 1981). The rate at which genes from different cells are assembled in single cells and accumulating combinations of deleterious genes are broken down will also depend on the rate of HGR. But what is the rate of recombination in bacteria and how does one measure that rate?

One goal of the research for this dissertation was to develop procedures to estimate the rate of recombination mediated by transformation and conjugative plasmid and to validate those procedures as well as obtain estimates of those rates (Chapter 2). The second goal was to ascertain whether there are mechanisms that limit recombination in naturally transforming bacteria (Chapter 3). The third goal was to explore, in a quantitative way the contribution of HGR to the rate of adaptive evolution in bacteria and the under which accelerating the rate of adaptive evolution will provide a recombining population a competitive advantage when competing with populations that do not recombine or do so at lower rate (Chapter 4).

1.1 Some of the background behind these goals:

Homologous gene recombination has, of course, played a major role in classical as well as contemporary studies of the genetics and molecular biology of bacteria. By "mating" donor and recipient strains bacteria in vitro in either directly (conjugation) or indirectly (transformation and transduction) through HGR genes can be mapped (Smith 1988), and strains constructed to address a variety of molecular, biochemical, developmental and even evolutionary and ecological questions (REF). Although one can ascertain the relative rates of recombination of different genes from these investigations, they provide no information about absolute rate at which recombination occurs in laboratory strains of bacteria, much less those from natural populations. Moreover, save for studies of the mechanisms of recombination, for geneticists and molecular biologist HGR is a tool rather than the object of study. Although there are occasional references to evolution and ecology in some studies of the mechanisms of recombination, for the most part they restricted to the Discussions of articles, rather than the body of the report, see for example (Dubnau 1999; Chen and Dubnau 2003; Gonzalez-Pastor, Hobbs, and Losick 2003; Chen and Dubnau 2004; Chen, Christie, and Dubnau 2005; Ellermeier et al. 2006; Johnsen, Dubnau, and Levin 2009)

Most research on the population and evolutionary genetics of recombination in bacteria is through retrospective studies of either full genome sequences, or the analysis of the DNA sequences of selected "house keeping" genes multilocus sequence typing (MLST). One concern of these genomic studies is to elucidate the nature and extent to

which HGT contributes to the genomes of different strains and species of bacteria (Karlin and Mrazek 1997; Mrazek and Karlin 1999; Ochman, Lawrence, and Groisman 2000; Schubert et al. 2009; Touchon et al. 2009). While the MLST studies have been primarily concerned with the population genetic structure of bacteria, as part of these endeavors they explore the extent to which HGR can account for variation in house keeping genes (Maiden et al. 1998; Feil et al. 2000; Smith, Feil, and Smith 2000; Falush et al. 2001; Feil et al. 2001; Feil and Spratt 2001; Feil et al. 2003; Wirth et al. 2006; Didelot, Darling, and Falush 2009).

An early approach to ascertaining identify recombination between lineages were based on comparisons between gene trees. For instance, through retrospective analysis of DNA sequences, the lack of concordance between phylogenetic trees inferred with different genes has been used to suggest that recombination plays a significant role in the divergence and differentiation of bacteria (Maynard-Smith et al. 1993; Maynard Smith and Smith 1998). In perfectly clonal populations, all the genes would produce concordant topologies, with no differences in the relationships inferred between the taxonomic units employed for the comparisons. On the other hand, if shuffling genes from one lineage to another played an important role in maintaining the genetic variation observed then significantly different topologies would be inferred with different loci (Figure 1.2).

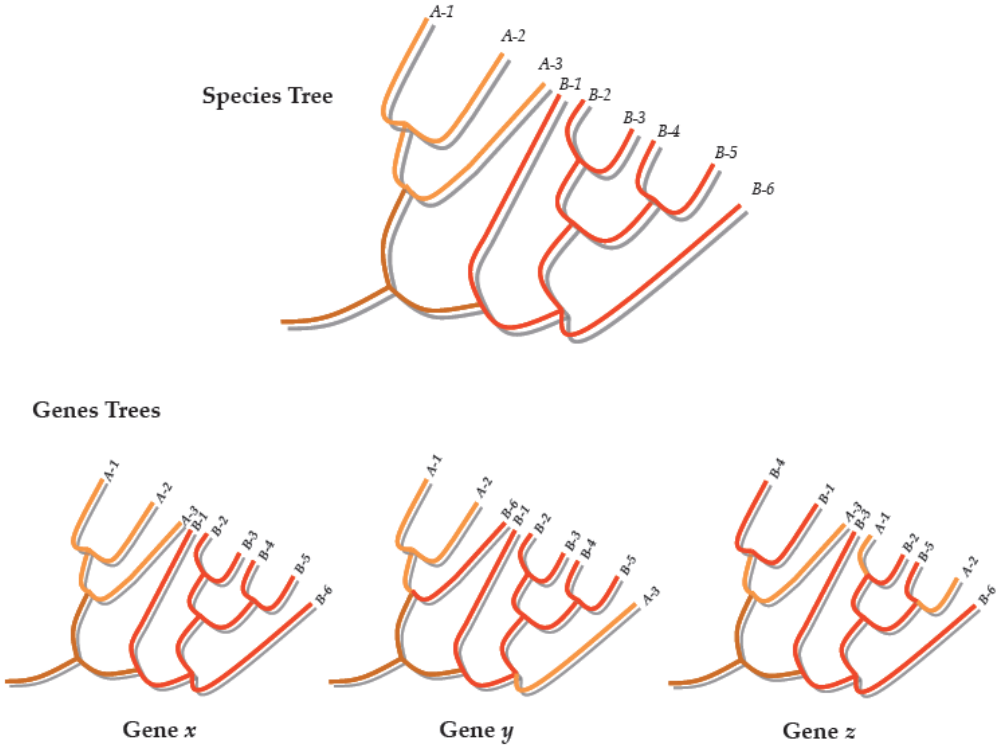

Figure 1.2. Cartoon showing the phylogenetic relationship between putative bacterial species (top, species A1-A3 and B1-B6). The phylogenetic relationships inferred with three different hypothetical genes (x, y z) show significant differences. It has been suggested that this pattern (different gene trees) are the result of frequent recombination.

Follow-ups to these initial works have led to the development of more refined analytic frameworks that have produced estimates of the amount of recombination that occurs in populations of *E. coli*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Helicobacter pylori* (Falush et al. 2001; Spratt, Hanage, and Feil 2001; Spratt 2004). According to this estimates *E. coli* is a very clonal organism, while in *Streptococcus pneumoniae* and *Haemophilus influenzae* recombination contributes 10 to 20 times more to the observed genetic variation than mutation does (Feil et al. 1999; Feil, Enright, and Spratt 2000; Feil et al. 2000); and *Helicobacter pylori* seems to recombine so much that its pattern of genetic variation resembles that one of a perfectly panmictic, fully recombining organism (Falush et al. 2001). These estimates, more than answering

questions have posed new ones: how can organisms that seem to recombine so much maintain their uniqueness and evolve as separate lineages? Does recombination play an important role in the spread of adaptive gene variants through these populations? Or does selection play a strong role in the spread of advantageous alleles in the population, how does it affect the estimates of recombination from retrospective analyses if it is not taken into account? How much beneficial recombination is in seemingly clonal populations like *E. coli*?

Many of these questions have stimulated the investigations presented in this work. The population dynamics of bacterial populations has a strong influence on demographic processes like bottlenecks and further expansions, as well as periodic selection of adaptive clones. This creates difficulties in the estimation of the rates at which recombination occurs in populations by retrospective studies, even when these mosaic patterns are recognized. In the simpler scenario, population bottlenecks are expected to reduce diversity and, if recombination is present, increase linkage disequilibrium, making it difficult to estimate the rate at which recombination occurs. Most importantly, periodic selective events in bacterial populations increase the difficulty of the estimation of recombination events, because it continuously generates strong linkage disequilibrium and the estimated rates of recombination are lower than the real recombination rates. Not knowing the strength and frequency of these selective events, it is difficult to separate both effects and obtain more precise estimates of the rate of recombination in bacteria. Given the potential relevance of recombination in giving

shape to the extant polymorphism, and the spread of adaptive variants, particularly drug resistance genes of importance for public health, it is of the utmost importance to estimate its rates in isolation of any other effects.

There have been few experimental studies exploring the contribution of HGR to the rate of adaptive evolution (Graham and Istock 1978; Graham and Istock 1979; Souza, Turner, and Lenski 1997; Bacher et al. 2007; Cooper 2007; Baltrus, Guillemin, and Phillips 2008). Their results can be seen as mixed. The studies of Graham and Istock and Baltrus and colleagues, with transformation-mediated recombination in B. subtilis and H pylor, respectively, and Cooper with F-plasmid mediated recombination in E. coli B are consistent with the hypothesis that recombination increases rates of adaptive evolution. Those of Souza and colleagues with an Hfr strain of E. coli K12 and E. coli B and Bacher and collaborators studies with Ascinetobacter, are interpreted to be inconsistent with this hypothesis. What can't be obtained from these experimental studies are the quantitative conditions under which recombination augments rates of adaptive evolution. These studies also don't ascertain the conditions under which HGR would be favored in competition with non-recombining populations. Stated another way, under what conditions would gene shuffling by recombination serve the selective force for the evolution of mechanisms for HGR and HGT.

1.2 Chapter 2. Population dynamics of *Streptococcus pneumonia*: Oscillations in continuous culture of *S. pneumoniae* and the evolution of clonal suicide

Streptococcus pneumoniae is a Gram-positive, α-hemolytic bacteria, naturally transformable in which widespread levels of recombination have been documented in retrospective analyses (Feil and Spratt 2001). Transformation in *S. pneumoniae* occurs through the process of competence gain, a physiological state that enables bacteria to further uptake DNA. The induction of this physiological state is mediated by complex genetic machinery that seems to work as a quorum sensing system (Chen and Dubnau 2004). This makes *Streptococcus* species a suitable candidate for estimating the rate of recombination in a controlled experimental setting and compare if the seemingly neutral rate of recombination estimated in the lab corresponds roughly to the estimated in retrospective analyses.

The motivation for this second chapter was to estimate recombination rates for *Streptococcus pneumonia* in continuous cultures. But alas, Mother Serendipity had other plans for this second chapter. We found an unanticipated dynamics when, probably for the first time ever, S. pneumonia was grown in continuous culture. Unlike well behaved bacteria, like E. coli in which yields a stable and almost invariant density at equilibrium, *Streptococcus pneumonia* cultures produced density oscillations with up to five orders of magnitude with an apparently constant period. This observation, which was certainly inconvenient for our purpose of estimating recombination rates was too delicious to not pursue. How can we explain these profound and seeming continous

oscillations in density? To answer this question we use both mathematical models and experiment.

A mathematical model was formulated to explain the dynamics observed according to which the oscillations could be explained by the waxing and waning of an agent (perhaps a toxin) autocatalytically produced that lyses cells. After several attempts to explain the oscillations with previously characterized lysins in *S. pneumonia*, we determined that the agent has not been previously described and proposed an ecological scenario, again in the form of a mathematical model for the maintenance of an agent that kills members of the same population that produces it. Experimental results consistent with such scenario are presented and succinctly discussed. A report of this study has been published in the Proceedings of the Royal Society of London B Series.

Not to be fully seduced by Ms. Serendipity, with the aid of a mathematical model and genetically marked strains of S. pneumonia we estimated the rate constant of recombination, X Levin 1981 for transformation-mediated recombination in this Gram positive bacteria. It was gratifying, albeit possibly coincidental, that the X estimated in these experiments are consistent with those obtained in retrospective studies using MLST data (Feil and Spratt 2001) This portion of chapter 1 will be separately published.

1.3 Chapter 3 Polymorphism in the competence peptide and population genetic structure in *Streptococcus pneumonia*

In order to understand the impact of recombination to the genetic makeup of bacteria populations, it is important not only to quantify how much recombination occurs in the populations but also discern which processes facilitate or limit the exchange of genetic material. For example, it has been shown how the frequency of homologous genetic recombination in *E. coli*, *Bacillus subtillis* and *S. pneumoniae* decays exponentially as the divergence of the to-be-incorporated sequence increases (Zawadzki, Roberts, and Cohan 1995; Fraser, Hanage, and Spratt 2007). The consequences of the limits imposed by sequence divergence have been nicely discussed in the case of *Streptococcus ssp*. (Fraser, Hanage, and Spratt 2007).

The only described limitation to recombination mediated by transformation in *S. pneumoniae* is the divergence of the free DNA segments available for up-take. This result has fostered the general belief that the limits to homologous recombination in this species are minimal. Nonetheless, recent molecular work has shown that the system controlling the induction of the competent state in *S. pneumoniae*, necessary for the uptake of DNA for recombination, is polymorphic and it has been suggested that such polymorphism limits the amount of recombination between groups of cells presenting different alleles. The development of natural competence is regulated by the action of the two-component signaling system encoded by *comC* and *comD*, coding a small peptide signal and its cognate receptor, respectively (Havarstein et al. 1996; Pestova,

Havarstein, and Morrison 1996; Cheng et al. 1997; Campbell, Choi, and Masure 1998). Furthermore, different alleles, encoded by *comC*, are only recognized by the receptor encoded by the matched allele of *comD*, and induction of competence is thought to be restricted to cells expressing the same alleles in peptide and receptor (Iannelli, Oggioni, and Pozzi 2005). It has been proposed that this polymorphism maintains genetically differentiated subpopulations of S. pneumonia carrying different alleles as they would be unable to cross induce competence and thus initiate transformation (REF).

In the second chapter of this thesis, we present the results of an investigation in which, employing a population genetic framework, we assess if clinical isolates carrying different competence peptide alleles present more differentiation than would be expected if there was no limit for the exchange of genetic material between subpopulations. We find that there is no significant differentiation between subpopulations carrying different alleles of the *comC/D* system. Furthermore we show that the best model explaining the polymorphism observed is one in which recombination explains significantly more variation between subpopulations than mutation. We also propose a hypothesis for the maintenance of the polymorphism in the competence peptide system based on the suggestions that loci under balancing selection tend to show higher levels of divergence than neutral loci along the genome.

1.4 Chapter 3 The Population and Evolutionary Dynamics of Homologous Gene Recombination in Bacterial Populations

Having done enough experimental work and wanting a rest, we elected to do purely theoretical study. There were two goals to this investigation, (i) to ascertain, in a quantitative way, the conditions under which HGR will increase rates of adaptive evolution and (ii) the conditions under which HGR increasing rates of evolution will provide a recombining population an advantage in competition with a population that is incapable of HGR or recombines and a lower rate.

Understanding the impact of recombination on the genetic makeup of the bacteria populations is a problem that has received increase attention over the years. As stated before, recombination could facilitate the spread of adaptive alleles through populations by shuffling genes that have been already selected for in other populations and reducing the problems created by clonal interference and the accumulation of background deleterious mutations (Muller 1932; Muller 1964; Evans 1986; Kim and Orr 2005). This is of particular importance when the bacteria species considered is a pathogen, and the adaptive genes considered confer resistance to antibiotics or provide with colonization and establishment advantages in the host. Nevertheless, the conditions under which recombination provides with a selective advantage to populations is not that clear.
In chapter three of the dissertation we assess, by simulation, the conditions under which recombination could provide an advantage when compared to non-recombining populations, under realistic parameter values for the recombination and mutation rate

as well as selective contribution of the alleles. Our results are in agreements with those found in similar studies (Kim and Orr 2005). Recombination increases the rate of adaptive evolution when compared to non-recombining populations. Once established, selection for the capacity to increase rates of adaptive evolution can maintain recombination mechanisms and prevent invasion of non-recombining populations even when recombination engenders some fitness cost. More interesting are the results of the conditions under which a recombining population could invade a non-recombining one. Most of the theoretical results and experimental work performed to explain the benefits of recombination in bacteria have been performed assuming that recombination evolved in isolation; and then compare the relative gain in fitness of recombining and non-recombining populations. The results obtained from such approach are useful in a first approximation but it seems unrealistic to assume that recombination could have originated in isolation. We address the question about the conditions for invasion of a recombinant subpopulation over a non-recombining population, and also the relative performance of recombining vs non-recombining populations at similar densities. Our results suggest that the conditions for invasion of recombination as a strategy that increases fitness are restrictive and it is very unlikely that recombination mechanisms could be favored initially based solely on their ability to spread selective variants in the population. We discuss briefly how the adaptive benefit of recombination could explain its maintenance, but it is insufficient to explain its origin. Also, we discuss how our simulation results provide an explanation for the lack of consistency of the experimental works that show support or lack of support for the adaptive hypothesis.

Not included as a chapter in this dissertation, but either done or initiated in the course of my utterly delightful years as a PBEE student are (i) an experimental study of The dynamics and rate of homologous recombination in E. coli and the frequency of naturally occurring *Escherichia coli* capable of recombination and the rate at which they recombine. This work is ongoing and will be among the first of my publications as Dr. Cornejo. (ii) The study of the age and geography of the polymorphism in *Plasmodium vivax*. This meta-analysis provides supporting evidence for *Plasmodium vivax* originating in Asia and published in Trends in Parasitology (Cornejo and Escalante 2006). (iii) A collaborative research in which the localization of multiple spliced variants of a zinc finger protein is assessed and it is shown that the multiple isoforms exist *in vivo* (Leung et al. 2009). In this work, I just provided support for the evolutionary claim that the gene targeted in the study is a conserved copy of a ZC3H14 family. (iv) As part of my ongoing collaborations in the molecular evolution and evolutionary history of *Plasmodium* ssp we have submitted a manuscript with new isolates of Plasmodium obtained from chimps and bonobos supporting an alternative hypothesis for the evolution of *Plasmodium falciparum*. (v) Also, two additional manuscripts will be hopefully submitted in the next couple of months studying the evolution of the apical membrane antigen (AMA-1) and the transmission blocking vaccine target Pvs-28 respectively.

Finally, we present general conclusions of the work presented in this dissertation. This work opens even more questions about the evolutionary biology of recombination in

bacteria, in particular in terms of its population genetics and population dynamics, than it answers. Although diverse in model organisms, experimental approaches, and analytic techniques, the common intent of this work is to understand how much recombination contributes to the genetic variation of bacteria populations.

CHAPTER 2

Oscillations in Continuous Culture Populations of *Streptococcus pneumoniae*: Population Dynamics and the Evolution of Clonal Suicide

Omar E. Cornejo, Daniel E. Rozen, Robert M. May and Bruce R. Levin
modified from Proc. R. Society London B Series, 276:999-1008 (2009)

2.1 Introduction

From an ecological and evolutionary perspective it is not at all surprising that bacteria produce toxic agents that kill other species (Wiener 1996) or genetically different clones of the same species (Riley and Gordon 1999; Czaran, Hoekstra, and Pagie 2002; Riley and Wertz 2002). More interesting evolutionarily are the toxins produced by bacteria that kill or induce suicide (apoptosis, autolysis) in genetically identical members of their own species. Included among these agents are the murein hydrolases and the choline-binding proteins responsible for autolysis and allolysis in *Streptococcus pneumoniae* (Steinmoen, Knutsen, and Havarstein 2002; Steinmoen, Teigen, and Havarstein 2003; Moscoso and Claverys 2004; Guiral et al. 2005), and the proteins produced by sporulation-committed *Bacillus subtilis* that kill non-committed members of the same genotype (Gonzalez-Pastor, Hobbs, and Losick 2003; Ellermeier et al. 2006; Claverys and Havarstein 2007).

Three general classes of hypotheses have been presented for the selection pressures responsible for the evolution and maintenance of mechanisms that are lethal to members of the same genotype. One has been likened to cannibalism, where the killed cells provide nutrients which in times of dearth enable the population at large to survive (Claverys and Havarstein 2007). A variant of this cannibalism hypothesis has been proposed for *Bacillus subtilis*, wherein the nutrients released by killed cells enable the remaining "killer" cells to postpone the terminal commitment to sporulation, an energetically costly process that delays vegetative growth until new resources become available (Gonzalez-Pastor, Hobbs, and Losick 2003). In the second class of hypothesis, gastronomy is not the reason for killing one's own, but rather the acquisition of genetic information via transformation (a more intellectual goal?). In accord with this hypothesis, killing members of one's own genotype is coincidental to the lysis of bacteria of the same or related species which may carry genes that provide an advantage to the killer population, perhaps by enabling them to invade new habitats or survive environmental stresses like antibiotics (Claverys and Havarstein 2007; Claverys, Martin, and Havarstein 2007). The third class of hypothesis of which we are aware considers lysis to be a stress response, which promotes the release of virulence factors (Pinas et al. 2008). Presumably, these factors would facilitate the ability of the surviving members of the killer population to colonize susceptible hosts.

In this report, we present evidence for a new mechanism by which bacteria kill members of their same genotype that we discovered serendipitously. While performing experiments to estimate the rate constant of recombination (Levin 1981) in chemostat

populations of *Streptococcus pneumoniae* (pneumococcus) strain R6, we observed oscillations in density of up to five orders of magnitude with an apparently constant period. These oscillations were not restricted to this domesticated laboratory strain but it was also obtained for the three clinical isolates of pneumococcus we examined. Although oscillations of substantial amplitude in the density of bacteria are expected when predators like lytic bacteriophage are present (Chao, Levin, and Stewart 1977; Yoshida et al. 2007), they are not anticipated in monoclonal cultures of bacteria free of a third trophic level of organisms. Under these conditions, bacteria are expected to maintain a constant density the magnitude of which depends primarily on the concentration of the limiting resource (Monod 1949; Kubitzchek 1970; Stewart and Levin 1973).

Our theoretical and experimental analysis indicates that these oscillations in *S. pneumoniae* R6 can be attributed to the waxing and waning in the concentration of an agent (a toxin) that is released by the bacteria via an autocatalytic process and either lyses cells directly or induces their autolysis Our experiments indicate that the toxin responsible for these oscillations is a protein that can also lyse or induce lysis in closely but not distantly related species of Streptococcus. Since we can rule out the six known suspects anticipated for this lysis, we conjecture that this oscillation-driving toxin is novel, i.e. has not been previously identified or characterized.

We present and, with the aid of a mathematical model, explore the properties of a fourth hypothesis for the evolution and maintenance of the observed, and other toxins, that kill

members of the same clone: allelopathy to prevent invasion of established populations by competing strains of the same and different species. We briefly discuss why resistance to the suicidal agent observed here has not evolved and the potential for using these toxins for the treatment of bacterial infections.

2.2 Materials and Methods

2.2.1 Bacteria

The stains and species of bacteria used in this investigation, their genetic markers and their source are listed in Table 2.1.

Table 2.1. List of Strains and Streptococci species used in this investigation. The source of each strain and the relevant genotypic features for this study are provided. Genetics marker employed (when relevant) is also provided.

Strain/Specie ID	Relevant genotypic features	Genetic Marker[b]	Source
R6	Non encapsulated derivative of D39		Laboratory isolate
LytA	Δ-*LytA* in R6 background		A. Tomasz
Δ-*ComC* originally FP5	Δ-*ComC* in RX1 background	Chloramphenicol resistant	F. Iannelli (Iannelli, Oggioni, and Pozzi 2005)
Δ-*ComD* originally FP48	Δ-*ComD* in RX1 background	Kanamycin resistant	F. Iannelli (Iannelli, Oggioni, and Pozzi 2005)
PMEN 6	PMEN 6 Multiply antibiotic resistant clinical isolate, serotype 23F		McGee
PMEN 18	PMEN 18 Multiply antibiotic resistant clinical isolate, serotype 14		McGee

Strain/Specie ID	Relevant genotypic features	Genetic Marker[b]	Source
PMEN 20	PMEN 20 Multiply antibiotic resistant clinical isolate, serotype 6B		McGee
TIGR4	encapsulated strain	Streptomycin resistant	Laboratory Strain (Regev-Yochay et al. 2007)
TIGR4 Δ-*spxB*	Δ-*spxB* in TIGR4 background	Kanamycin resistant	Lipstich (Regev-Yochay et al. 2007)
6A	encapsulated strain of serotype 6A		Dawid (Dawid, Roche, and Weiser 2007)
Δ-*blpR*	Δ-*blpR* in 6A background	erythromycin resistant	Dawid (Dawid, Roche, and Weiser 2007)
RH-1	R6 derivative, but *ebg*::*spc*	Ery^R, Spc^R,	Johnsborg (Johnsborg et al. 2008)
RH-3	RH1 but Δ-*ComE*	Ery^R, Spc^R, Cm^R, Kan^R	Johnsborg (Johnsborg et al. 2008)
RH-17	RH1 but Δ-*Cbdp*	Ery^R, Spc^R, Kan^R	Johnsborg (Johnsborg et al. 2008)
Streptococcus mitis			ATCC # 49456
Streptococcus oralis			ATCC # 35037
Streptococcus sanguinus			ATCC # 10556
Streptococcus salivarius			ATCC # 7073
Streptococcus gordonii			ATCC # 10558
Streptococcus pyogenes			ATCC # 21060
Streptococcus australis			ATCC # 700641

2.2.2 Culture Media

Liquid

THY - per litre: 30g Todd-Hewitt broth (Difco) plus 5g yeast extract (Difco)

Solid (agar) for sampling

TSA – Tryptic Soy Agar (Fisher) with 5% sheep blood (Difco) 16 g agar per liter.

MHA – Muller Hinton base (Difco) with 16 g agar per liter.

2.2.3 Chemostat culture and sampling

The chemostats used in these experiments are "home made"; see www.eclf.net for the design and photographs. Chemostat populations were established from -80°C freezer stocks of *S. pneumoniae* via a two step process: 200ul aliquots of frozen cells were thawed in 2 mls of THY and grown to an OD_{600} ~ 0.3 (corresponding to a density of ~ $2e^8$ cells/ml) at which time they were diluted 10-fold in THY broth and regrown to OD_{600} ~ 0.3. Aliquots of 100µl of these re-grown cultures were inoculated into 20 mls of THY broth in the chemostat vessels for an initial density of ~ 1e6 cells/ml. The chemostat vessels (tubes) were incubated at 35°C in a water bath. A peristaltic pump maintained a constant flow of fresh media at a rate of 2 ml per hour, which is the same rate at which the waste, excess resources and live or dead cells were removed from the vessel. Chemostat cultures were agitated and aerated by a vacuum-induced bubbling of filter-sterilized air. Samples were taken by removing ~500µl directly from the chemostat vessel. Viable cell densities (CFU) were estimated by serially diluting the samples in

0.85% saline and plating on Tryptic Soy Agar containing 5% sheep blood. The experiments with low densities of S. *pneumoniae* were performed in chemostats with a reservoir containing THY and 0.85% saline in a ratio of 1:28.

2.2.4 Bio-assay for toxin:

To test for the presence and relative concentration of a toxin, samples taken at different times were centrifuged for 2 minutes at 13,000 rpm and then the supernatant was stored at -20^{0}C. No CFUs were observed when 100µl samples of the centrifuged supernatants were plated on Tryptic Soy Agar with 5% sheep blood, indicating that supernatants were cell-free. Between 1.5.-2.0 µl of these supernatants were spotted onto soft agar THY lawns of target bacteria on a base of Mueller-Hinton (hard) agar. These lawns were prepared by growing *S. pneumoniae* R6 or other strains of this or other species of Streptococci in THY to an OD_{600} ~ 0.3 nm and adding 100µl to 4 ml of soft agar at ~ 55^o C (~5x10^6 cells per ml). The plates containing the supernatant-spotted lawns were scored after 24 hours of incubation at 35^o C in a CO_2 incubator. To test whether the zones of inhibition on the agar were due to replicating phage, material from zones of inhibition was added to fresh soft agar with *S. pneumoniae* and transferred to new lawns. No plaques were observed.

To ascertain whether the toxin is H_2O_2 we added approximately 5,000 units of catalase (Worthington Biochemical) to 3 mls of soft agar used for the lawns, or incubated the supernatants from the chemostats for 30 minutes with 5,000 units catalase before

spotting. To determine if the toxin had a protein nature we incubated the supernatant at 55°C, 65°C and 75°C for 30 minutes before assaying its activity. Also, prior to assaying for activity, supernatant was treated for 30 minutes at 37°C with Proteinase K at a final concentration of 1.5 mg/ml. Initial screening to determine the size range of the putative protein toxin was performed using membrane centrifugal filters (Microcon™) with cut-offs of 3 kDa, 10 kDa, 30 kDa and 50 kDa following manufacturers protocols.

2.2.5 Recombination Assay

In order to characterize the dynamics of recombination mediated by transformation in *S. pneumoniae* we inoculated chemostats with a mix culture of two otherwise isogenic **R6** strains carrying two different antibiotic resistance markers, rifampicin (**R6-Rif**) and streptomycin (**R6-Str**). We estimated the number of recombinants by counting assessing the number of doubly marked strains resistant to both rifampicin and streptomycin. The The culture conditions were as described previously for single strains chemostats. Before inoculation, the flow of the chemostats was stabilized and frozen aliquots of **R6-Str** and **R6-Rif** were grown in THY to OD_{600} ~ 0.3 *nm* and adding 100 μl. To follow the densities of the parental and recombinants strains, we plated samples at appropriate dilutions on THY (5 % blood) agar supplemented with streptomycin (40 μg/ml) to follow streptomycin resistant strains (**R6-Str**), on agar supplemented with rifampicin (25 μg/ml) to follow rifampicin resistant strains (**R6-Rif**); and agar supplemented with both antibiotics at the same concentrations mentioned before to follow recombinants carrying both markers. Sampling was performed at hourly intervals for 80 hours.

Control chemostats were followed to assure that the double resistant strains did not arise by mutation. For this individual chemostats carrying **R6-Str** and **R6-Rif** strains respectively were inoculated in a similar way, and in parallel to, the mixed culture chemostats. This chemostats were sampled hourly for 80 hours in THY (5% blood) agar to estimate total density of the culture and with THY (5% blood) agar supplemented with streptomycin (40 μg/ml) and rifampicin (25 μg/ml) to follow the rise of double resistant mutants.

In addition to the previously mentioned controls, we performed fluctuation tests on **R6-Str** and **R6-Rif** in order to estimate the mutation rate for the gain of rifampicin resistant in **R6-Str** backgrounds and streptomycin resistance in **R6-Rif** backgrounds. For this, 24 parallel cultures of **R6-Str** and 24 parallel cultures of **R6-Rif** were grown overnight in 2 ml of THY yeast (media as described earlier). 5 cultures were sample and plated at appropriate dilutions to estimate the average population size of the cultures and the totality of the cultures were centrifuged at 14,000 rpm for 10 minutes, re-suspended in 200 μl of saline solution and plated in the corresponding selective agar. Plates were incubated at 35°C for 48 hours, and the number of colony forming units in each culture was accounted for. The mutation rate was estimated as suggested by (Gerrish 2008).

2.2.6 Invasion Experiments

In order to evaluate if the proposed ecological model for the maintenance of a suicidal toxin could be supported, we performed two invasion experiments. In one of them, an established population of *S. pneumoniae* in a low resource concentration chemostat

established as described previously was challenged with an invading inoculum of *Streptococcus mitis*, a species sensible to the supernatant produced by pneumococcus in the chemostat (as it is shown in this work). A second experiment was performed by challenging an established population of *S. pneumoniae* with an inoculum of *Streptococcus pyogenes*, a species non-sensitive to the toxin produced by S. pneumoniae in the chemostat. Colonies formed by each one of this species show a very distinctive phenotype on THY 5% blood agar, which allows to distinguishing them in co-culture. Inoculation of the invading population was performed after *S. pneumoniae* reached an equilibrium (~ 65 hours after inoculation), and sampling was performed at approximately hourly intervals.

2.2.6 Numerical Analysis (Simulations)

Numerical solutions to the differential equations of our models (simulations) were programmed in **R (**a free software environment for statistical computing) and/or Berkley Madonna™ . The codes for these simulations are available on www.eclf.net.

2.3 Results

We open this report with the serendipitous observation that motivated this investigation, profound oscillations in the density of *Streptococcus pneumoniae* R6 in a chemostat (Fig 2.1a). Similar observations were obtained with all three clinical isolates of *S. pneumoniae* that we examined (Figure 2.2a). These dramatic oscillations in colony

forming units (CFU) are also apparent from changes in the optical density (OD 620nm) in these cultures. If these bacteria were killed but not did lyse, the OD of the chemostat would be expected to roughly halve every 6.9 hours at the flow rate employed ($0.1/hr^{-1}$) and the culture would remain turbid for more than 24 hours. Instead, we observed a complete loss of turbidity over roughly 10 hours, which we interpret as evidence that the decline in the CFU estimates of density in these chemostats can be attributed to the bacteria lysing.

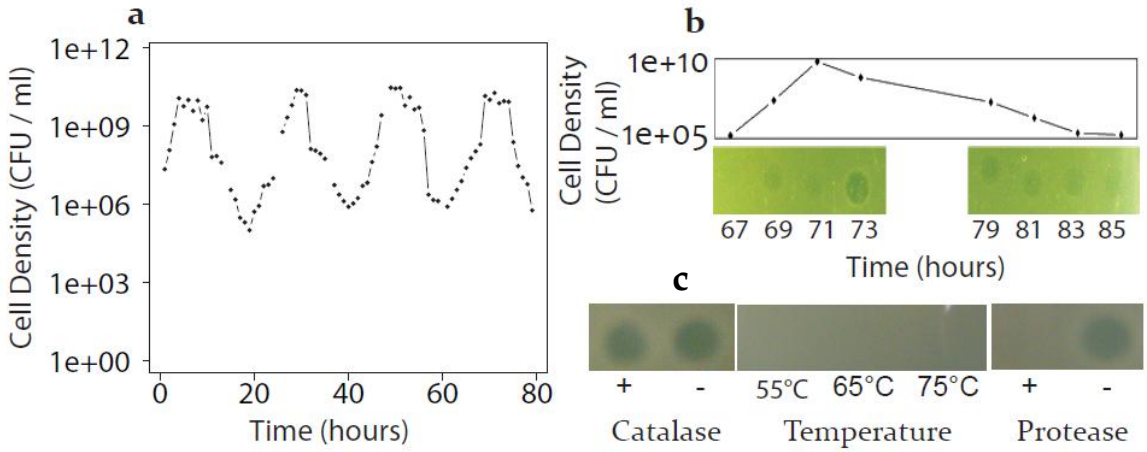

Figure 2.1. *Streptococcus pneumoniae* R6 in chemostat culture: (a) cell density estimated from colony forming units (CFU/ml) (b) Supernatant taken at different times during the course of a cycle spotted onto soft agar lawns of exponentially growing *S. pneumoniae* R6. (c) Catalase, heat-treated and proteinase K (protease) treated supernatant on lawns of exponentially growing *S. pneumoniae* R6. Unless otherwise noted, this and the other chemostats in this study were maintained at a dilution rate of approximately 0.1 per hour in Todd-Hewitt Broth plus 0.5% yeast extract (THY).

By spotting the supernatants of centrifuged samples on soft-agar lawns of *S. pneumoniae* R6 (see Materials and Methods) we assayed for the presence of an exogenous agent released by the bacteria, a "toxin", that could account for the declines in the densities. The results of these assays were positive; zones of inhibition were observed (Figure

2.1b). Based on the clarity of the zone, the concentration of this toxin appears to reach an apex shortly after the bacteria reach their maximum density (Figure 2.1b).

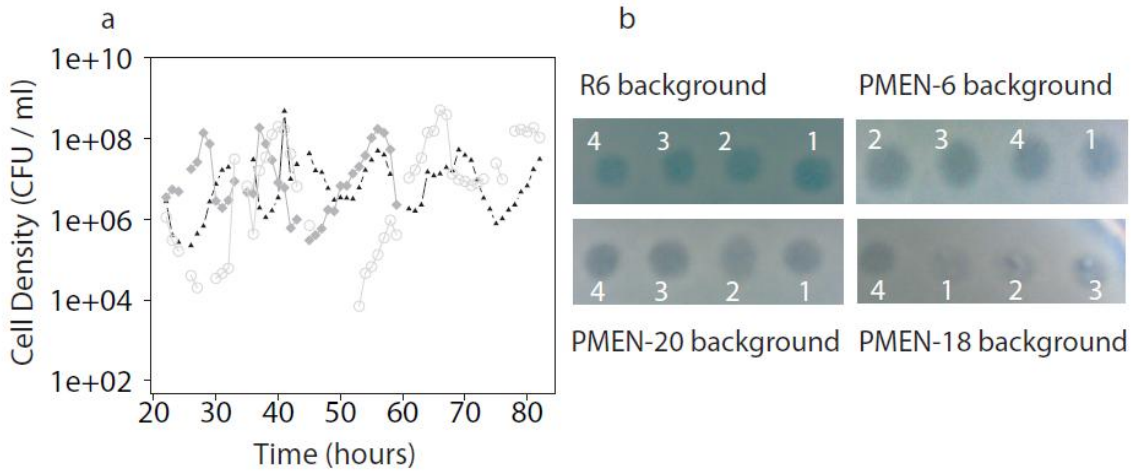

Figure 2.2. (a) Clinical isolates show oscillatory dynamics that qualitatively resemble those obtained with the lab strains (PMEN-18 in black solid triangles, PMEN-6 in grey solid squares, and PMEN-20 in grey open circles). The source of these and other strains used in this study are listed in Table 2.1. (b) Cross activity of supernatants among different strains. Spot assays performed on lawns (backgrounds) of the laboratory strain R6, and three clinical isolates (PMEN8, PMEN6 and PMEN20). The supernatant from single clone chemostats of *S. pneumoniae*: R6**(1)**, PMEN-20**(2)**, PMEN-18 **(3)**, and PMEN-6**(4)**, tested on lawns of each cell line.

In an effort to better understand the population dynamic mechanisms responsible for the observed oscillations in density and the conditions under which these oscillations would occur, we developed a simple model of these population dynamics. In this model, R is the concentration of a limiting resource ($\mu g/ml$), B is the density of the bacteria (*cells per ml*) and T the concentration of the toxin (*arbitrary units per ml*). We assume a Monod model for resource concentration dependent growth of the bacteria (Monod 1949), $\Psi(R) = vR/(R+K)$ where v is the maximum growth rate of the bacteria (hr^{-1}) and K is the

concentration of the resource at which the bacteria are growing at half their maximum rate ($\mu g/ml$). The limiting resource from a reservoir where it is maintained at a concentration of C ($\mu g/ml$) enters a vessel of unit volume (1 ml) at a constant rate w (hr^{-1}) which is the same rate at which excess resources, bacteria and toxin are removed. The bacteria take up the resource at rate proportional to their density, growth rate and a conversion efficiency parameter, e (μg) (Stewart and Levin 1973). In addition to being washed out with the flow, the toxin decays at a rate d (hr^{-1}).

The toxin kills the bacteria at a rate equal to the product of its concentration, the bacterial density and a constant x. We assume that the toxin is produced at a rate proportional to the product of its concentration, the density of bacteria and a constant y. Our biological justification for including concentration of the toxin in its rate of production is the observation that other known secreted products in *S. pneumoniae*, such as the quorum sensing peptide responsible for the induction of competence in *S. pneumoniae*, the competence-stimulating peptide (CSP), and the bacteriocin-like protein (BLP) involve a positive-feedback (autocatalysis) between the concentration of these agents and their production (Havarstein, Coomaraswamy, and Morrison 1995; Havarstein et al. 1996; Pestova, Havarstein, and Morrison 1996; de Saizieu et al. 2000; Berge et al. 2002).

With these definitions and assumptions, the rates of change in the density of bacteria and concentration of the resource and toxin are given by,

$$dR/dt = w(C-R) - \Psi(R)Be \qquad (1)$$

$$dB/dt = \Psi(R)B - xBT - wB \qquad (2)$$

$$dT/dt = yBT - dT - wT \qquad (3)$$

The density of bacteria that are killed by the toxin, D is calculated from a fourth differential equation, $dD/dt = xBT - wD$ (4).

To illustrate the properties of this model, we use numerical solutions to these differential equations. The growth rate parameter v used in these simulations is in the range estimated for *S. pneumoniae* R6 in THY broth. The rate of flow, w, corresponded to that used in our experiments. The value of the Monod constant, K, is similar to that estimated for *E. coli* in glucose limited minimal medium. For the maximum density (as determined by C and e), the toxin killing and production constants (x and y) and the toxin decay rate (d), we employed values that would produce the period and amplitude of the oscillations (Figure 2.3a) similar to that observed in our experiments (Figure 3.1a).

For an analytical understanding of the properties of this model and a dynamic summary of its properties see appendix 1. The central point of this analysis is straight-forward and we believe interesting. Most importantly, this analysis shows that the resource concentration, C, and conversion efficiency, e, are not only the primary determinants of maximum population density but also control the dynamical behavior of this system. To examine the transition between the different dynamic behaviors, we let $a \equiv ev(d+w)/wyC$, $b \equiv v/w$ (note that necessarily $b > 1$) and $f \equiv (d+w)/w$ (with $k \equiv K/C << 1$ -- typically $k \sim 10^{-2}$ -10^{-5} holding for the different dynamics considered below). In accord with our analysis if and only if $b > a$ (with the parameters used in Figure 2.3 this

corresponds to $C > 50$) there exists a "3 species" equilibrium with R, B and T all present. There is a very sharp transition in the dynamical behavior when $a \approx 1$; and for $a < 1$ (i.e., $C > 500$), the oscillations generated decay very slowly (on the order of 10^5 hours with parameters and initial values of the variables employed in our simulations), see Figure 2.3a. For $b > a > 1$, s would obtain when C is lower (300µg/ml), the oscillations are strongly damped (Figure 2.3b).

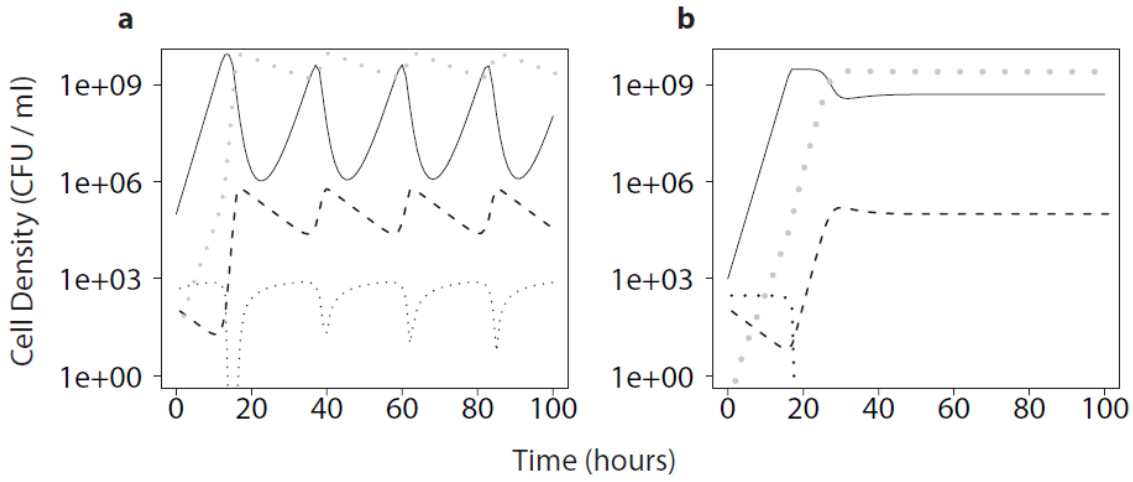

Figure 2.3. Simulation results: changes in the densities of bacteria B (solid black), killed bacteria D (dotted grey), and concentration of the resource R (dotted black), and toxin T (dashed black). Parameters, (a) C=1000, w=0.1, e=10^{-7} , x=5x10^{-6}, y=4x10^{-10}, d=0.10, v=1.0, k=0.25, (b) the same parameter values as in (a) save for the resource concentration in the reservoir, C=300.

To test the validity of the model's prediction that the oscillations would be damped if the cell density is lower, we inoculated *S. pneumoniae* R6 into a chemostat with the THY broth diluted 28-fold. The results of this experiment are presented in Figure 2.4. After the population recovered from an initial decline in density, which we believe is attributable to an extended lag period in this nutrient poor medium, the maximum density achieved is approximately two orders of magnitude lower than the maximum densities we observed in chemostats run with full-strength THY broth (Figure 2.1a). As

predicted by the model, the oscillations in diluted medium are considerably damped (Figure 2.4) relative to that in undiluted THY broth (Figure 2.1a). While we cannot rule out the possibility that the physiological effects of low nutrient concentrations, rather than the reduction in cell density, are responsible for the dampening of the oscillations, we admit to some satisfaction with the consistency between our theoretical predictions and our experimental results. After all, hypotheses can only be rejected.

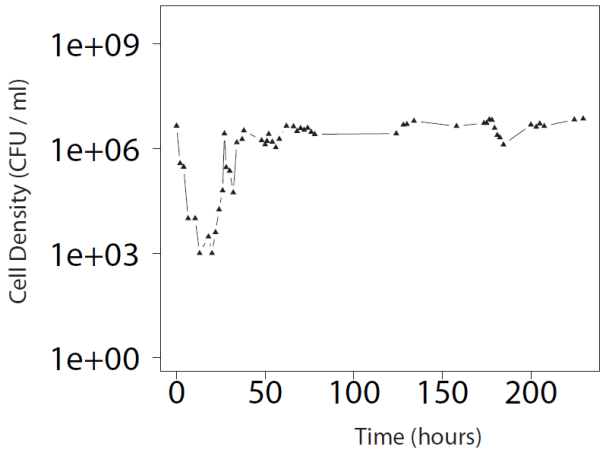

Figure 2.4. Changes in the density (CFU) of *S. pneumoniae* R6 grown in THY diluted by a factor of 28 in 0.85% saline. The dilution rate of this chemostat was approximately the same as that in Figure 2.1, ~0.1 per hour.

To date, our efforts to characterize this toxin have been through exclusion experiments using strains of *S. pneumoniae R6, TIGR4,* and *6A* that do not produce known candidates for this toxin and a limited biochemical characterization of the toxic supernatant. One candidate for this toxin is hydrogen peroxide (H_2O_2), which is released by *S. pneumoniae* at concentrations that are lethal to cells of this and other species of bacteria (Regev-Yochay et al. 2006; Regev-Yochay et al. 2007). To test whether H_2O_2 is responsible for the oscillations, we added approximately 5,000 units of catalase (Worthington

Biochemical) to the 3 mls of soft agar used to grow the lawns of target cells, or incubated the supernatants from the chemostats for 30 minutes with 5,000 units catalase before spotting. With both treatments the catalase would hydrolyze any H_2O_2. The results of this test indicated that H_2O_2 is not the toxin; lysis was observed on *S. pneumoniae* lawns containing catalase and with catalase-treated chemostat supernatants (Figure 2.1c). Additional evidence that the toxin is not H_2O_2 comes from the observation of density oscillations in chemostats inoculated with *S. pneumoniae* with a deletion in the gene encoding the synthesis of pyruvate oxidase (TIGR4-Δ*spxB*) rendering this strain unable to produce hydrogen peroxide (Regev-Yochay et al. 2006; Regev-Yochay et al. 2007), (see appendix 1).

Our initial characterization of supernatants taken at the peak of toxin production suggests this agent is likely to be a protein. These supernatants do not generate zones of inhibition on lawns of *S. pneumoniae* R6 following incubation under denaturing conditions at relatively high temperatures or after protease treatment (see Figure 2.1c). Size fractionation using Microcon™ filters and the spot assay indicates that the relative size of the toxin is between 30 and 50Kd (between 300 and 500 amino acids).

On first consideration it may seem the protein responsible for these oscillations is the same as that responsible for autolysis in batch cultures of *S. pneumoniae* (Sanchez-Puelles et al. 1986; Ronda et al. 1987; Havarstein et al. 2006). This does not appear to be the case. Mutants defective for the major pneumococcus lysin, N-acetylmuramoyl-L-alanine

amidase, *lytA* do not display autolysis in batch culture (Figure 2.5a) yet they oscillate in density when introduced into a chemostat (Figure 2.5b).

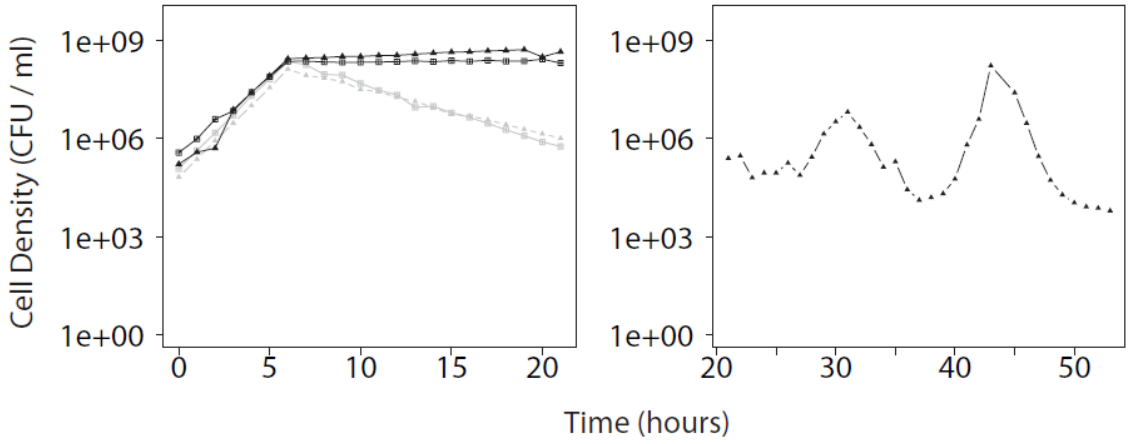

Figure 2.5. Changes in the density (CFU) of bacteria in batch and chemostat culture. (a) *S. pneumoniae* R6 LytA$^+$ (grey open squares and solid triangles) and LytA$^-$ in THY batch culture (black open squares and solid triangles). (b) A LytA$^-$ strain of *S. pneumoniae* R6 in chemostat culture with a flow rate of 0.1 per ml in THY.

It is somewhat gratifying that the products of *lytA* and hydrogen peroxide can be ruled out as the toxins, because there is no evidence for their being produced autocatalytically as our model predicts they would have to be. On the other side, our experiments have also ruled out "suspects" which directly or indirectly induce lysis in pneumococcus and for which there is evidence for autocatalytic production: the competence peptide (CSP), and the bacteriocin-like protein (BLP) (Claverys and Havarstein 2002; Peterson et al. 2004; Dawid, Roche, and Weiser 2007; Lux et al. 2007). In chemostat culture, mutants that are defective for the production of these secreted peptides (plus the CSP receptor), *comC, comD, blpR* oscillate in a manner similar to wild type strain R6 in chemostat culture (see appendix 1). In addition, a mutant defective in the competence response

regulator (Δ-*ComE*) which is unable to induce competence or activate downstream genes regulated by competence induction also shows oscillatory dynamics in a chemostat similar to the parental strains from which it was derived (see appendix 1).

A particularly appealing candidate for the toxin responsible for the density oscillations is the murine hydrolase coded for by the gene *CbpD*. As Johnsborg and colleagues recently demonstrated using a beta-galactosidase release assay, murein hydrolase lyses pneumococcus in liquid culture (Johnsborg et al. 2008). This enzyme is approximately the molecular weight we estimated for the toxin in the supernatants of our chemostats and, as we show below, the host range of bacteria sensitive to the *S. pneumoniae* R6 murine hydrolase resembles that of the supernatant from the oscillating S. pneumoniae R6 chemostats (Johnsborg et al. 2008). But alas, strains of pneumococcus R6 knocked-out for the production of murein hydrolase protein, Δ*CbpD,* oscillate to an extent no different than the *CbpD*+ strain from whence it was derived (see appendix 1), and the supernatant of these Δ*CbpD* chemostats produced zones of inhibition on R6 lawns similar to those we observed with *CbpD*+ strains (data not shown).

Using the spot assay we tested the cross sensitivity of R6 and the three clinical isolates to their respective secreted toxins. For each supernatant, zones of inhibition were observed on lawns of all four strains (Figure 2.2b). To further explore the host-range of the R6 toxin, we tested for lysis on lawns of the other Streptococcus species. The results of these assays suggest that closely, but not distantly related species of Streptococcus are susceptible to the *S. pneumoniae* R6 toxin (see appendix 1).

The observations that the activity of the toxin driving these oscillations does not appear to be strain specific and lyses other, closely related, species of Streptococcus raises the possibility (hypothesis) that the ecological role of these toxins is allelopathy. In this interpretation these toxins either prevent the invasion of established populations by competing strains and species sensitive to this toxin and/or facilitate invasion into established populations of toxin sensitive strains species. At this time we have not tested this allelopathy hypothesis experimentally. However, using an extension of the model equations (1)-(3) we have explored the a priori plausibility of this hypothesis by ascertaining the conditions under which self-killing toxins would prevent invasion of competitors into established populations of toxin-producing bacteria. In this extended version of our model there is second population of bacteria, B2 that is susceptible to the toxin produced by the first, B1, but does not produce that toxin. The system of equations describing this situation is:

$$dR/dt = w(C-R) - \Psi(R)e(V_1B_1 + V_2B_2) \quad (5)$$

$$dB_1/dt = \Psi(R)V_1B_1 - x_1B_1T - wB_1 \quad (6)$$

$$dB_2/dt = \Psi(R)V_2B_2 - x_2B_2T - wB_2 \quad (7)$$

$$dT/dt = yB_1T - d_TT - wT \quad (8)$$

where $\Psi(R) = vR/(R+K)$, V_1 and V_2 are the maximum growth rates of these bacteria, and x_1 and x_2 are the killing rates of the producing and invading bacteria respectively. For convenience we assume that the efficiency of conversion of resource into bacteria biomass, e, and the Monod constant, k, are identical for B_1 and B_2. As in the previous

model y is the rate constant of production of the toxin and d_T is the inverse of the half-life of the toxin.

As indicated in the analysis of the properties of this model in the appendix, there are conditions under which the production of a toxin that is lethal to the bacteria that produce it can prevent invasion by a toxin-sensitive population, even when that invader has a higher growth rate than the established population $V_2 > V_1$. This, however, requires that the ratio of the rates of toxin-mediated killing of the invading and established strain, x_2 / x_1, is larger than V_2 / V_1 and that the density of B_1 is sufficiently high to produce adequate concentrations of the toxin.

As can be seen in the simulations presented in Figure 2.6a, in the absence of the toxin the higher fitness clone, B_2, will invade and replace the established population B_1. If, however, the average density of B_1 is high and $x_2/x_1 > V_2/V_1$ the invasion of B_2 can be prevented. This obtains when the density of B_1 oscillates (Fig. 2.6b) and when the resource concentration is too low for this established population to oscillate (Fig 2.6c). However, if the density of B_1 and thus the concentration of toxin in the environment are too low, the clone with the greater growth rate will invade (Fig 2.6d), albeit at a rates lower than in the absence of toxin production (Fig 2.6a). It should be noted, however, that the conditions for B_2 to invade an established population of B_1 are to some extent dependent on the initial density of B_1 and thus the concentration of the toxin, T in the environment.

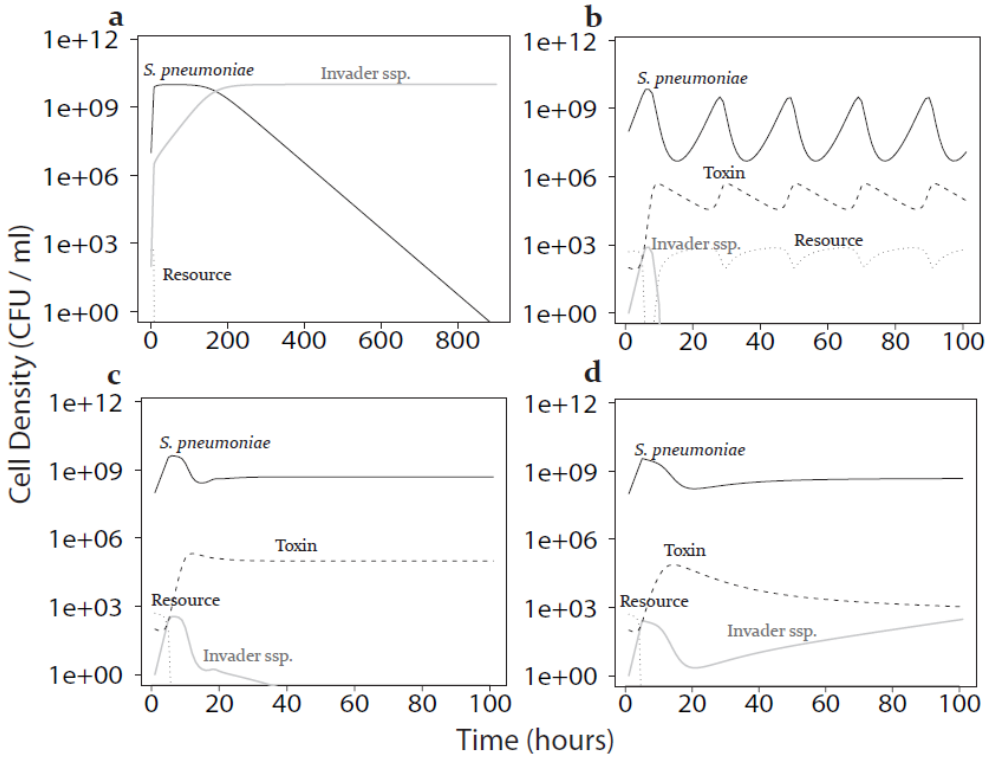

Figure 2.6. Simulation results: invasion of a clone with higher intrinsic fitness: Changes in the densities of the bacterial populations and concentrations of the toxin (dashed black) and resource (dotted black). The parameter values for this simulation are: $V_1 = 1.0$, $V_2 = 1.5$, $k = 0.25$, $e = 1\text{x}10^{-7}$, $y = 4\text{x}10^{-10}$, $\omega = 0.1$, $x_1 = 5\text{x}10^{-6}$, $x_2 = 9\text{x}10^{-6}$, and $d = 0.1$. In these figures a single cell of B_2 (solid grey) is introduced into populations of B1 (solid black) at initial density of 10^8. In (b), (c) and (d) the initial concentration of the toxin is 50 units per ml. (a) Invasion in the absence of toxin production (y=0), (b) Failure to invade a high density (C= 1000) oscillating toxin-producing population. (c) Failure to invade a lower density (C= 300) non-oscillating toxin-producing population; (d) Invasion in the absence of maintenance of toxin production because of very low resource concentration (C=50). In these simulations, when the density of the invading $B_2 < 0.5$, the simulation B_2 was set equal to zero.

If the density of B_1 is initially low, B_2 can invade under conditions where it would not invade a toxin producing B_1 population with an initially higher density (simulation results not shown). Our numerical simulation results indicate the same conclusion can be drawn when the invading strain, B_2, produces a toxin that kills members of the established strain, B_1; although it has higher fitness when rare it cannot colonize an

otherwise lower fitness high density toxin-producing established strain. (Simulation results not shown.)

Even, when we were unable to identify the suicidal agent responsible for the oscillations observed in continuous cultures of *S. pneumoniae*, there are questions that remained unexplored in our published manuscript. One of them is how plausible the ecological scenario for the maintenance of a suicidal toxin is. Our spot assays suggest that only closely related species are sensitive to the toxin produced by *S. pneumoniae*. Additionally, our theoretical model predicts that sensitive species would be unable to invade established populations of pneumococcus, while non-sensite species would succesfully establish. The results of our exploratory experiments in this direction are consistent with the models Figure 2.7. Although these experiments do not represent a test of the model proposed, these represent evidence consistent with it and encourages its further exploration.

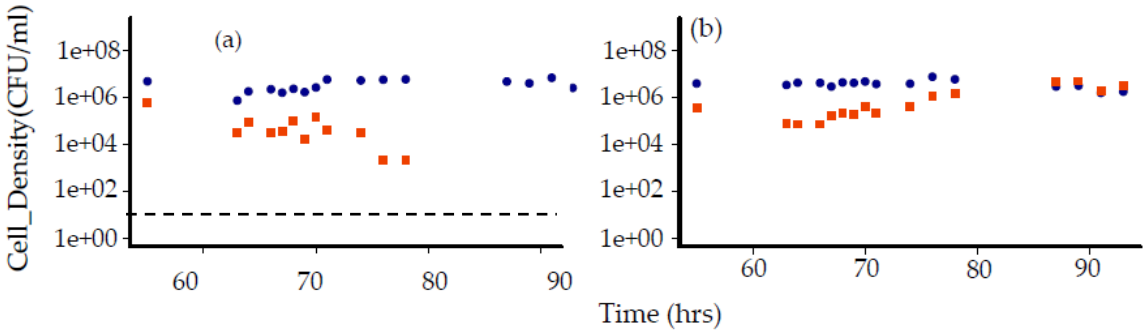

Figure 2.7: Invasion experiments of *S. mitis* (sensitive sp.) (a) and S. pyogenes (non-sensitive) (b) species (orange squares) on established populations of *S. pneumoniae* cli6 (dark blue circles) grown in THY diluted by a factor of 28 in 0.85 % saline. The dilution rate of this chemostat was approximately the same as that in (Figure 2.1), approx 0.1 per hour.

The second question that remained unexplored in our published article was the original motivation for this work. What is the seemingly "neutral" rate of recombination of *Streptococcus pneumoniae* in the laboratory? Our exploration of these question involved, again the expansion of the mathematical model proposed for the oscillations of *S. pneumonia* in the chemostat to include a second population B_{01} carrying a different marker to a population B_{10}, both equally sensitive to the activity of the toxin and, both contributing equally to the production of the lysing agent T. New recombinants B_{11} carrying both markers are produced at a rate χ that is proportional to the amount of DNA liberated by each population DB_{10} and DB_{01} available in the media and produced as a fraction f of DNA liberated per hour after lysis of the cells. The terms of other possible recombinants formed are not shown because they do not affect significantly the dynamics for the "observable" recombinants. The dynamics of recombination is described then by the following set of differential equations,

$$dR/dt = w(C-R) - \Psi(R)e(V_1B_1 + V_2B_2) \quad (9)$$

$$dB_{10}/dt = \Psi(R)V_1B_1 - xB_{10}T - wB_{10} - \chi B_{10}DB_{01} \quad (10)$$

$$dB_{01}/dt = \Psi(R)V_{01}B_{01} - xB_{01}T - wB_2 - \chi B_{01}DB_{10} \quad (11)$$

$$dDB_{10}/dt = f\,x\,B_{10}T - w\,B_2 \quad (12)$$

$$dDB_{01}/dt = f\,x\,B_{01}T - w\,B_{01} \quad (13)$$

$$dT/dt = y(B_{10} + B_{01} + B_{11})T - d_TT - wT \quad (14)$$

In Figure 2.8 it can be observed the observed (dots) and predicted (lines) dynamics of recombination between **R6-Str** and **R6-Rif** isogenic strains of *S. pneumoniae*. The rate of recombination was estimated as that one which provided the best explanation of the data ($\chi = 1.4 \times 10^{-14}$ *per cells*^{2} *per hour*). Because of the lysis, it is relatively difficult to estimate the generation time of pneumococcus in the chemostat. Restricting our observations to a few hours in the growing part of the curves it suggests that the generation time of *S. pneumoniae* in the chemostat was around 1 hour. This suggests that the recombination rate per generation is around 1.4×10^{-14} as well. The mutation rates for the gene markers employed in this experiment, according to our fluctuation tests, were around 10^{-9} mutants per generation. Because the harmonic mean of the population size of the donors throughout the experiment was ~ 10^6 cells then the rate of recombination per donor population throughout the experiment is 10^{-8} per recipients per generation. This implies that the rate of acquisition of new alleles is approximately 10 times greater by recombination than by mutation. This result is consistent with what has been estimated in retrospective genetic analyses (Feil, Enright, and Spratt 2000).

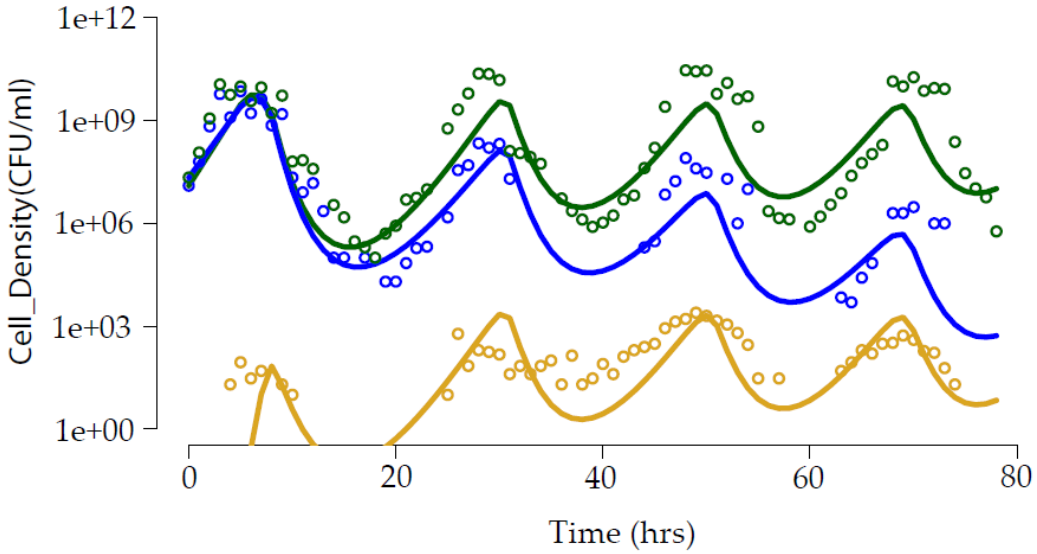

Figure 2.8: Density changes in populations of parental **R6-Str** (green) and **R6-Rif** (blue) strains of *S. pneumoniae*, and recombinants **R6-StrRif** (orange) in experiments (dots) and as predicted by our theoretical model of recombination (lines). In the numerical simulations, the maximum growth rate $v_1 = 1.17$, and $v_2 = 1.03$ were adjusted according to the best visual fit to single chemostats for each strain. The rest of the and the parameters were changed to improve the visual fit of this experiment. χ the recombination rate was χ 1.4×10^{-14} *per cells*^{2} *per hour*.

2.4 Discussion and Conclusions

When we set out to use chemostat cultures to estimate the rate of transformation-mediated chromosomal gene recombination in *Streptococcus pneumoniae* (now a tale for another report) we did not anticipate the profound oscillations in density reported here; this study is founded on a serendipitous observation. From a population dynamic perspective, these oscillations can be accounted for by the autocatalytic production of an agent, a toxin, which either directly lyses or induces autolysis in other cells of the same clone. The results of our experiments are consistent with the predictions of this model. The concentration of this toxin appears to reach its maximum shortly after the density of

the oscillating population reaches its apex and these oscillations are damped when the density of the culture is low.

We have yet to characterize the toxin responsible for these oscillations or determined the physiological and molecular mechanisms by which it acts. Our results, however, indicate that this toxin is likely to be protein of between 30 and 50 kDa. By exclusion, our experiments suggest that this toxin is novel in the sense of not being described or characterized earlier. Using strains of *S. pneumoniae* deleted for the genes coding for candidates for this toxin we have excluded the following potential toxins or genetic systems involved in its production:

(i) N-acytulmuramoyl-L-alanine amidase, *lytA* which is responsible for autolysis in batch culture.

(ii) The competence-inducing peptide, *comC*

(iii) The bacteriocin-like protein system *blpR* and

(iv) Murein hydrolase which lyses pneumococcus in liquid culture, *CdpB*.

(v) The cognate response regulator for the competence system, *comE*

The densities of strains that do not produce these peptides and proteins oscillate in a similar manner to that observed in the parental strains from which they were constructed. Moreover, using catalase and a strain that does not produce hydrogen peroxide we have also excluded this oxidizing agent as being responsible for these density oscillations.

Although density oscillations were observed for all the strains we examined, there appear to be strain differences in the densities reached in these chemostats with the same media and flow rates and in the amplitude and possibly the period of these oscillations. For two reasons, we have not explored the variation in these dynamics systematically. First, strain variation in these dynamics is secondary to the main focus of this investigation, which is the population dynamic processes responsible for these oscillations and the ecological/evolutionary reason bacteria produce the self-killing toxins generating these oscillations. Second, the frequent sampling of viable cell density (CFU data) required for a quantitatively accurate characterization of oscillations with periods of 30 or so hours is an onerous task that we cannot justify doing given the motivation for this study.

Probably for many readers and certainly for the authors the most intriguing questions raised by this study are the ecological role and thereby the selection pressure responsible for the evolution of this oscillation-driving toxin. Despite the fact that this phenomenon occurs with natural isolates as well as laboratory strains of *S. pneumoniae,* it is conceivable that density oscillations of the amplitude observed in Figure 2.1a are an artifact of laboratory culture. In their natural habitat within the human nasopharynx *S. pneumoniae* may never reach the densities where these oscillations occur. As predicted by our model (Figure 2.3b) and demonstrated experimentally (Figure 2.4), oscillations in density of the amplitude observed in Figure 2.1 are not anticipated in environments that can only support low density populations. However, whether these oscillations are an artifact of culture conditions or not it still remains necessary to account for why,

evolutionarily, *S. pneumoniae* produce this and the other toxins that kill genetically identical bacteria.

If the killing of genetically identical members of the same population were the sole function of the toxin, the capacity to produce it or respond to its action would not be favored by natural selection. If, however, the agent responsible for this killing provides a survival or growth advantage to the producing clone in its natural habitat, its production could be favored by selection even if some or even many cells of the producing clone are killed. In our introduction we briefly described three hypotheses that have been proposed to account for how the production of these suicidal agents could provide an advantage to the toxin-producing population:

(1) Cannibalism; to acquire resources (nutrients and carbon sources) from the killed cells that would increase the likelihood of survival of the whole clone in times of dearth (Sanchez-Puelles et al. 1986; Ronda et al. 1987; Lewis 2000; Claverys and Havarstein 2007) or postpone entry into a non-replicating state, sporulation (Gonzalez-Pastor, Hobbs, and Losick 2003).

(2) The acquisition of adaptive genes through transformation of DNA by lysing cells of other clones and species that bears these genes (Lewis 2000; Steinmoen, Knutsen, and Havarstein 2002; Steinmoen, Teigen, and Havarstein 2003; Claverys and Havarstein 2007).

(3) The release of other agents, like virulence factors that that would enable the population to colonize a new habitat or for parasitic bacteria an uncolonized host (Pinas et al. 2008).

Here we propose a fourth hypothesis for the production of these clone suicidal toxins, allelopathy to prevent invasion of established populations. We find this hypothesis appealing not only because we presented it, but also because the toxin responsible for the observed oscillations kills other clones of the same species as well as closely related and thereby potentially competing species. Our theoretical analyses suggest that there are relatively broad conditions under which the production of toxins that kill bacteria of the same genotype that produce them can prevent invasion of established populations by toxin-sensitive populations that would otherwise invade because of a growth rate advantage. The necessary conditions for this to occur are that the invader is more sensitive to killing by the toxin than the resident and the concentration of the toxin in the habitat is sufficiently great for toxin-mediated killing to override the growth rate advantage of the invading population. A testable prediction of this hypothesis is that strains would be less sensitive to killing by the toxins they produce than they are to the toxins produced by other strains and species. Our experimental exploration of this hypothesis show results consistent with it.

In this report, we have not formally explored the other face of this allelopathy hypothesis, i.e. to facilitate the invasion of established populations of bacteria. For two reason we would anticipate the conditions for the invasion of a clone producing a clonal

suicide toxin in mass (liquid) culture would be highly restrictive (Chao and Levin 1981; Levin 1988b). (i) Unlike bacteriocins, there is no evidence for immunity to this toxin by the producing clone. (ii) All of the strains and species of Streptococcus we have examined that are sensitive to this toxin also produce that toxin or a variant of it to which all the other strains and species are sensitive. To be sure, in physically structured habitats, where bacteria are maintained as colonies rather than planktonic cells, there are broad conditions under which bacteriocin production can facilitate invasion of established sensitive populations (Wiener 1996; Kerr et al. 2002). Whether this would also be the case for clonal suicide toxins, and how the production of these toxins would affect the population dynamics of bacteria in physically structured habitats remains to be seen and are intriguing problems for another time.

The four hypotheses for the evolution and maintenance of self-killing toxins considered here are not mutually exclusive and to our knowledge none of them has been adequately tested much less unambiguously supported experimentally. To be sure, there is good evidence that the rate of transformation-mediated recombination is greater for *S. pneumoniae* that produce a lysis-inducing murein hydrolase protein than for otherwise isogenic cells that do not (Johnsborg et al. 2008). On the other hand, we know of no evidence that this higher rate of recombination would provide a sufficient advantage to the murein hydrolase -producing population to overcome likely fitness costs associated with synthesizing this protein and its killing isogenic cells.

Interestingly, the rates of recombination estimated in a "neutral" setting in the laboratory does not seem to differ significantly from those estimated from DNA sequence data (Feil, Enright, and Spratt 2000). It seems that recombination contributes 10 times more to the maintenance of variation than mutation suggesting that other processes like selection and demography do not affect significantly the estimation if recombination in *S. pneumoniae*. This result is particularly interesting if we take into consideration that other processes like selection might increase the frequency of clones bearing a particular combination of alleles (see chapter 4), if that combination is advantageous. The concordance in the estimates could be, at least initially, considered as evidence for the lack of relevance of selection as an informative process while estimating recombination in this bacterial species.

Particularly intriguing from an evolutionary perspective is why pneumococcus remains susceptible to the toxins it produces. Even if the densities within natural habitats are too low for the toxin considered here to generate oscillations of the sort that motivated this study, if it is produced at all some cells would still be killed by its action. Saying this another way, there would be continuous and possibly intense selection for resistance to this toxin. Nevertheless resistance to this toxin has not been observed in *S. pneumoniae* chemostats maintained for three and a half months (DER and OEC, unpublished data). Perhaps more compellingly, since clinical (wild) isolates of *S. pneumoniae* also oscillated in chemostats and produce the toxin it is reasonable to assume that resistance to the toxin has not evolved even when there was plenty of time for this evolution to occur.

Could it be that viable and fit mutants resistant to the toxin cannot be generated by mutation or acquired by horizontal gene transfer?

Although a largely commensal bacteria that asymptomatically colonizes many humans, *S. pneumoniae* is also responsible for a number of invasive infections; e.g. otitis media, bacteremias, meningitis and pneumonia. While existing antibiotics have been successful in treating invasive pneumococcus disease and resistance infrequently prevents effective chemotherapy (Yu et al. 2003) this may not be true in the future. Anti-pneumococcus drugs with novel targets will be needed. Could the toxin responsible for the lysis reported here be developed into an effective and safe drug to treat pneumococcus infections? This would be particularly appealing if indeed viable strains of pneumococcus resistant to this toxin cannot be generated.

CHAPTER 3

Polymorphic competence peptide does not limit recombination in *Streptococcus pneumoniae*.

Omar E. Cornejo, Lesley McGee and Daniel Rozen
Re-submitted to Molecular Biology and Evolution

3.1 Introduction

The degree to which bacterial species are influenced by recombination will depend upon its rate of occurrence and the processes underlying it (Selander and Levin 1980; Maynard-Smith et al. 1993; Feil et al. 1999; Spratt and Maiden 1999; Feil, Enright, and Spratt 2000; Spratt 2004; Fraser, Hanage, and Spratt 2007). For example, while bacteria that are naturally transformable can promiscuously acquire free DNA present in the environment, often from phylogenetically disparate sources (Majewski et al. 2000; Cohan 2001; Cohan 2002), recombination mediated by conjugative plasmids or transducing phage is thought to be more restrictive. Consistent with this premise, naturally transformable species that undergo high rates of transformation mediated recombination, such as *Helicobacter pylori* and *Streptococcus pneumoniae*, have populations that are loosely genetically structured (Feil, Enright, and Spratt 2000; Falush et al. 2001), whereas species such as *Escherichia coli*, *Salmonella* ssp. or *Rhizobium meliloti* that recombine primarily via conjugation and transduction are more clonally structured (Maynard-Smith et al. 1993). Despite the promiscuity of naturally transformable species,

there are two well-known limits to inter-specific recombination mediated by transformation. First, some species, like *Haemophilus influenzae* and *Neisseria gonorrhoeae*, utilize short sequence tags to identify homotypic fragments of DNA, and only this DNA is recombined into the bacterial chromosome (Sisco and Smith 1979; Graves, Biswas, and Sparling 1982) . Second, high levels of sequence divergence across species, in otherwise homologous fragments of DNA, significantly reduce the frequency of inter-specific recombination (Zawadzki, Roberts, and Cohan 1995; Fraser, Hanage, and Spratt 2007). This latter limitation has been shown to exert a strong effect on population structure and the differentiation and maintenance of species within the genus *Streptococcus* (Fraser, Hanage, and Spratt 2007). However, their effects on limiting recombination within species, where sequence divergence is low and sequence tags are shared are expected to be small. This has fostered the general belief that, at least within species, the limits to homologous recombination mediated by transformation are minimal.

In *S. pneumoniae*, the development of natural competence is regulated by the action of the two-component signaling system encoded by *comC* and *comD*, specifying a small peptide signal and its cognate receptor, respectively (Havarstein et al. 1996; Pestova, Havarstein, and Morrison 1996; Cheng et al. 1997; Campbell, Choi, and Masure 1998) . The competence stimulating peptide (CSP) is secreted into the extra-cellular environment, where it binds to membrane bound *comD*, initiating a chain of events that ultimately culminates in the uptake and incorporation of free DNA (Tomasz 1965; Havarstein, Coomaraswamy, and Morrison 1995; Havarstein et al. 1996) . An intriguing and unexplained aspect of pneumococcal transformation is that both *comC* and *comD* are

genetically polymorphic (Pozzi et al. 1996; Whatmore, Barcus, and Dowson 1999). For both genes, there are two broad and highly concordant clades, corresponding to "pherotypes" designated CSP-1 and CSP-2 (Pozzi et al. 1996; Whatmore, Barcus, and Dowson 1999). Furthermore, each signal pherotype, encoded by *comC*, is only recognized by the receptor encoded by the matched allele of *comD*, and induction of competence is thought to be restricted to cells expressing the same pherotype (Iannelli, Oggioni, and Pozzi 2005).

The specific matching of ComC/D has led to the suggestion that pherotypes, acting as "mating types", facilitate a form of assortative mating or genetic exchange, which could maintain genetically diverse subpopulations within this species (Havarstein, Hakenbeck, and Gaustad 1997; Tortosa and Dubnau 1999; Steinmoen, Knutsen, and Havarstein 2002; Claverys and Havarstein 2007). This form of assortative mating would reduce or eliminate inter-pherotype recombination. If recombination in *S. pneumoniae* is limited to clones bearing the same pherotype, a clear prediction is that genetic differentiation among isolates belonging to different pherotype groups would be larger than expected assuming random recombination among pherotypes (Claverys, Prudhomme, and Martin 2006; Claverys and Havarstein 2007). An alternative and opposite prediction, suggested by recent results from Johnsborg et al (2008), is that there should be no such pherotype specific differentiation because transformation may be facilitated by the competence mediated lysis of cells bearing non-matching *comD* sequences. This "disassortative mating" would prevent any genetic sub-structuring of the population, and would result in a faster rate of gene exchange between members of

the population that carry different pherotypes than would be expected by random recombination with respect to pherotype.

In this work, we test whether there is significant genetic differentiation within *S. pneumoniae* according to their pherotype, characterized by the *comC* sequence, using a cosmopolitan sample of clinical isolates of *S. pneumoniae* and a standard population genetic analysis. We find that there is no significant genetic differentiation between groups defined by CSP pherotype and strong evidence for gene flow between CSP types. These results are inconsistent with the first hypothesis and provide indirect support for the second hypothesis: that there is facilitation of recombination with non-matching pherotype bacteria, mediated by lysis. We discuss why such a result is anticipated, given the current understanding of the mechanisms of transformation. Additionally, we present evidence in support for a selective hypothesis to explain the maintenance of the polymorphism in the competence system.

3.2 Materials and Methods

3.2.1 Clinical Isolates of *S. pneumoniae*: A geographically and serotypically diverse collection of clinical pneumococcal isolates from 2000 and 2001 were employed in this study (Yu et al. 2003).

3.2.2 Genetic characterization: 88 clones isolated in 2000 and 2001 were characterized by multilocus sequence typing (MLST). For MLST, internal fragments of

the *aroE*, *gdh*, *gki*,*recP*, *spi*, *ddl*, and *xpt* genes were amplified by PCR from chromosomal DNA, and the fragments were directly sequenced in both directions using the primers that were used for the initial amplification. The sequences (alleles) at each locus were compared to those on the publicly accessed MLST website (www.mlst.net) and were assigned allele numbers if they corresponded to sequences already submitted to the MLST database; novel sequences were submitted for new allele numbers and deposited in the database. The allele numbers at the six loci were compared to those at the MLST website, and sequence types (STs) were assigned. Allelic profiles that were not represented in the MLST database were submitted for assignment of new ST numbers and deposited in the database. The sequences of the loci for all isolates are available in www.eclf.net. Sequences of the *ddl* locus were excluded because it has been shown to be under selection (Enright and Spratt 1999).

The sequences of *comC* for the 88 isolates were obtained with the primers: FOR: 5'-CAATAACCGT CCCAAATCCA - 3', and REV: 5' – AAAAAGTACA CTTTGGGAGA AAAA - 3', producing a fragment of approximately 400 bp. The conditions for amplification were: 1× PCR buffer, 1.5 mm $MgCl_2$ 0.2 mm dNTP mix, 2U *Taq* Polymerase and 20 pmol of each primer, per 50ul reaction. The PCR cycling parameters were as follows: an initial denaturation step at 95°C for two minutes, 25 cycles of amplification performed as follows: denaturation at 94°C for 30 seconds, annealing temperature at 56°C for 30 seconds and extension temperature at 72°C for 1.0 minutes and finally completed with an extension at 72°C for five minutes. The isolates were assigned to a

given pherotype, by comparing their translated amino acid sequence to the types reported in (Kilian et al. 2008). The DNA sequences for the *comC* locus obtained in this work are available in www.eclf.net.

3.2.3 Population genetic analyses: Standard population genetic analyses were performed for all isolates using DNAsp v.4.20 (Rozas et al. 2003). Haplotype diversity, nucleotide diversity (π) and its standard deviation, were estimated for the sample as a whole, and for the sample stratified by CSP type (the overall estimates by geographic region of origin are provided in appendix 2).

Analysis of the Population Structure mediated by Pherotypes

In order to perform the population genetic structure analyses, CSP pherotype groups/subpopulations were assigned according to the criterion described above.

It is known that F_{st} statistics present some limitations, including the implicit assumption of uniform effective population sizes, and symmetric migration rates. These limitations can be problematic when structured populations have large effective population sizes, like in bacteria, and are weakly structured (Bossart and Prowell 1998), or for populations in which the subpopulations might differ significantly in size. For this reason, the assessment of genetic structure in the sample was performed under a coalescent framework that allows estimating Θ (the mutation parameter, proportional to the effective population size) and a migration parameter independently (Hudson 1991; Nath and Griffiths 1993; Beerli and Felsenstein 1999; Beerli and Felsenstein 2001).

Because of the definition of the sub-populations as carriers of different competence peptide/receptor alleles (CSP types), and because of the biology of recombination in bacteria, the number of migration events estimated in this way are informative of the levels of recombination (gene exchange) occurring between the two subpopulations. This approach, which uses sample frequencies rather than population frequencies, may render more reliable estimates of genetic differentiation than Fst and overcome some of its restrictions (Gonzalez, Beerli, and Zardoya 2008). Maximum likelihood estimates of Θ ($2N_e\mu$, with N_e being the effective population size and μ the mutation rate per site per generation) for each population, and immigration rates ($M_{1\rightarrow 2}$ and $M_{2\rightarrow 1}$, with $M = m / \mu$ and m is the rate of migration per generation) were obtained under a Markov Chain Monte Carlo model with importance sampling, employing 10 short chains (100,000 used trees out of 1,000,000 sampled) and 4 long chains (500,000 used trees out of 3,000,000 sampled), with adaptive heating of the chains as implemented in Migrate v3.0 (Beerli and Felsenstein 1999; Beerli and Felsenstein 2001). Initial values for Θ and M were obtained from F_{st} estimations. In order to assess the significance of the levels of migration inferred, we fitted a null model in which the migration is constrained to contribute as much as mutation to the differentiation between subpopulations ($M = m / \mu = 1$). The estimates under the null model were obtained employing 10 short chains (100,000 used trees out of 1,000,000 sampled) and 4 long chains (500,000 used trees out of 3,000,000 sampled). The comparison of the two models was done by a Likelihood Ratio test (LRT) with 2 degrees of freedom.

The relationships among the isolates were represented graphically by means of a haplotype network. The network was constructed with the concatenated MLST genetic sequences, employing a median joining algorithm of the pairwise distances among haplotypes, as implemented in the program Network v4.5.0.0 (Bandelt, Forster, and Rohl 1999).

Analysis of the Polymorphism and Assessment of selection in the comC *locus*

We estimated the synonymous and non-synonymous polymorphism of the comC locu and MLST loci within and between pherotypes. To facilitate the discussion of the results we will refer to the difference between CSP subpopulations as "divergence" between subpopulations. We also estimated the average number of synonymous and non-synonymous divergence between *S. pneumoniae* and a closely related species *S. mitis* for the homologous sequences. The *comC* sequences for *S. mitis* were obtained from the GenBank under the accession numbers: AJ240795, AJ000871, AJ000866, AJ000865, AY007503, AJ240764, AJ240763, AJ240762, and AJ000875. For the MLST data we used: *aroE* (accession numbers: EU075894, EU075893, EU075892, EU075891, EU075890, EU075889, EU075888, EU075887), *gki* (accession numbers: EU075788, EU075787, EU075786, EU075785, EU075784, EU075782, EU075783, EU075781), *recP* (accession numbers: EU075823, EU075822, EU075821, EU075820, EU075819, EU075818, EU075817, EU075816), *spi* (accession numbers: EU075860, EU075859, EU075858, EU075857, EU075856, EU075855, EU075854, EU075853), and *xpt* (accession numbers: EU075683, EU075682, EU075681, EU075680, EU075679, EU075678, EU075677, EU075676).

We assessed the neutrality of *comC* broadly and for each pherotype by performing the modified versión of the Hudson-Kreitman-Aguade (HKA) test proposed by Innan (Hudson, Kreitman, and Aguade 1987; Innan 2006) with the software ghka. This test asks if the levels of polymorphism and divergence observed in *comC* is significantly different than would be expected when compared to several reference neutral loci, here provided by the MLST loci.

We performed this test, in three ways: i) assessing *comC* broadly in *S. pneumoniae* and using *S. mitis* as an outgroup, ii) within *S. pneumoniae* assessing populations bearing pherotype 1, using pherotype 2 populations as the outgroup, and iii) assessing *S. pneumoniae* populations bearing pherotype 2, using pherotype 1 populations as the outgroup. In all cases, to obtain the probability distribution for the time of "speciation" we set 10,000 as the number of acceptances in the rejection-sampling algorithm, following the suggestion by Innan (Innan 2006). In order to obtain the probability value for the modified HKA test, 10,000 replicates were performed. The statistic obtained is the probability that ***r***, defined as the ratio of polymorphism to divergence of the locus presumably under selection (comC in this case) is consistent with the ratio estimated for the reference, neutral, regions.

Because of multiple testing, corrections were performed according to Benjamini-Yekutieli to maintain an overall significance of 0.05 (Benjamini and Yekutieli 2001).

3.3 Results

MLST profiles were obtained from 88 geographically diverse clinical isolates. Initial analyses of genetic diversity for each gene and the average over genes are shown in Table 3.1. In general, haplotype diversity is relatively high, as is nucleotide diversity which ranges from 0.9% – 1% per site. Because genetic structure among these clones

Table 3.1. Genetic diversity per gene and for the concatenated sequences in the ample as a whole andcategorized by CSP type.

	Gene	n (length)	S	h	Hd (s.d.)	π (s.d.)
Entire sample	*aroE*	88 (405 bp)	11	14	0.87(0.018)	0.0047(0.00035)
	gdh	88 (460 bp)	34	18	0.93(0.009)	0.0111(0.00147)
	gki	88 (483 bp)	32	17	0.86(0.022)	0.0156(0.00123)
	recP	88 (450 bp)	13	14	0.84(0.025)	0.0054(0.00035)
	spi	88 (474 bp)	40	18	0.79(0.041)	0.0090(0.00134)
	xpt	88 (486 bp)	31	24	0.92(0.016)	0.0091(0.00072)
	concatenated	88	161	56	0.97 (0.004)	0.0093 (0.0004)
By CSP type						
CSP1	*aroE*	65	10	12	0.86(0.023)	0.0047(0.0004)
	gdh	65	31	15	0.93(0.011)	0.0120(0.00189)
	gki	65	32	15	0.86(0.028)	0.0168(0.00151)
	recP	65	11	12	0.80(0.033)	0.0054(0.00042)
	spi	65	34	17	0.81(0.045)	0.0086(0.00155)
	xpt	65	26	19	0.91(0.019)	0.0093(0.00090)
concatenated		65	145	40	0.98 (0.007)	0.0095 (0.0005)

CSP2	*aroE*	23	7	10	0.85(0.052)	0.0042(0.00071)
	gdh	23	13	10	0.88(0.042)	0.0078(0.00081)
	gki	23	18	6	0.74(0.077)	0.0110(0.00150)
	recP	23	8	9	0.88(0.038)	0.0052(0.00065)
	spi	23	23	6	0.73(0.076)	0.0099(0.00252)
	xpt	23	15	11	0.89(0.043)	0.0078(0.00077)
concatenated		31	84	17	0.96 (0.027)	0.0078 (0.0007)

Note: **n** corresponds to the number of isolates, **length** is the nucleotide sequence length of the gene in base pairs **S** is the number of segregating or polymorphic sites, **h** is the number of haplotypes, **Hd** is the haplotypic diversity, π is the nucleotide diversity estimated as the average of the Jukes and Cantor pairwise distances. In all cases **s.d.** correspond to the standard deviations of the estimates.

arising from geographic subdivision could compromise our efforts to detect CSP mediated structure, we first determined if clones could be distinguished on the basis of sampling site. No strong signature of population genetic structure consistent with geographic location was detected. Although there is variability in the nucleotide diversity among loci, it shows consistency between groups of isolates bearing different types.

Of the 88 isolates, we determined that 65 (74%) presented comC amino acid sequences identical to pherotype type 1 while 23 (26%) presented comC amino acid sequences identical to pherotype type 2 (Table 3.2). It is important to notice that all amino acid changes are fixed between pherotypes with no polymorphism apparent at the amino acid level within *comC* pherotypes.

Table 3.2. Pherotypes (CSP types) identified by sequencing the *comC* locus.

Pherotype	n[a]	Amino acid sequence[b]
ComC 1	65	mkntvkleqfvalkekdlqkikgg **e m r L s K F F R d f I L Q r k k**
ComC 2	23	mkntvkleqfvalkekdlqkikgg **e m r I s R I I L d f L F L r k k**

a: The number of isolates presenting pherotype 1 and pherotype 2. **b** The amino acid sequence of the pherotypes is shown (conserved aa are shown in small case, aa in bold correspond to mature CSP peptide).

Our analysis of population migration under the coalescent fail to support a model with restricted migration, favoring a model in which frequent migration, or gene exchange, is maintained between subpopulations with different pherotypes (Table 3.3). The rates of effective migration (Nm) between subpopulations summarized over all loci considered in the analysis suggest that there are between 3.9 and 15.6 immigrants (lateral transfers) per site per generation between populations with different pherotypes. There are no apparent differences in the rates of immigration towards any particular subpopulation in our sample of clinical isolates on the basis of CSP pherotype.

If there were restrictions in the amount of gene exchange between bacteria carrying different pherotypes it would be expected that the haplotypes, based on MLST data, with the same pherotype grouped together. As it can be seen in Figure 3.1, this is not the case. The haplotype network estimated shows a structure that seems independent from the pherotype.

The analysis of the *comC* locus shows low genetic diversity within types (Table 3.4). The amino acid differences found in the *comC* locus are fixed between pherotypes within *S. pneumoniae*; but most of them are polymorphic when *S.pneumoniae* is compared to a

sister taxa, *S. mitis* (Table 3.5). Interestingly, the substitutions in the *comC* locus show an inverted pattern to that obtained with the neutral genes. In the neutral genes most of the changes are polymorphic between pherotype subpopulations with none or a few fixed differences. Similarly, most of the changes are polymorphic between species (*S. pneumoniae* and *S. mitis*), again with few fixed changes between species (Table 3.5). It can be observed that multiple cells contain no fixed differences (or polymorphic differences),

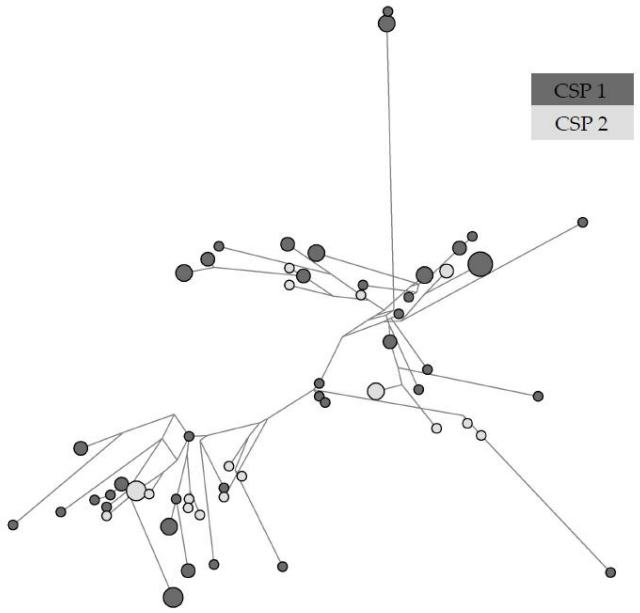

Figure 3.1. Haplotype network built with a median joining algorithm over the pairwise distance of the haplotypes. In grey are the haplotypes that belong to the CSP1 pherotype group, and in light grey the haplotypes belonging to CSP2 pherotype.. The size of the pies is relative to the frequency of the haplotypes. Some haplotypes show both pherotypes, and there is no clear clustering of the haplotypes with CSP type.

but the information provided by the MLST locus can still be used to assess if the pattern of substitution observed in the *comC* locus deviates significantly from neutrality. The comparison of the ratio of polymorphism to divergence (*r*) in the *comC* locus against the

reference loci, employing *S. mitis* as an outgroup, suggests that the observed polymorphism does not deviate significantly from what would be expected under neutrality ($r = 0.231$, p-value = 0.937).

It is known that the polymorphism observed in this locus is functional, and cells secreting one of the competence peptide pherotypes cannot induce competence in cells that secrete the alternative peptide, and express its corresponding receptor {Iannelli,

Table 3.3. Fitting to migration models: Full Model with migration estimated and Null model with migration restricted (Fixed to m / μ=1).

Model	Log L[a]	Θ_1 [b] (95 % CI)	Θ_2 [b] (95 % CI)	$N_em_{2\rightarrow1}$ [c] (95 % CI)	Nem 1→2 [c] (95 % CI)
Full model	-104.42***	0.013 (0.011 - 0.016)	0.0076 (0.005 – 0.010)	15.6 (7.60 – 21.3)	3.9 (2.56 – 6.14)
Null model	-136.22	0.014 (0.012- 0.016)	0.0083 (0.006 - 0.011)	n.a	n.a

a Maximum Likelihood value for the fitting. **b** Θ_i are the mutation parameter estimates for population i (CSP1 = 1, CSP2 = 2). **c** The number of immigrants per generation from population i to population j ($2N_em_{i\rightarrow j}$) and 95% confidence interval and the log likelihood are shown. **n.a** no estimates are provided because in the null model migration is fixed. *** significant p-value ≤ 0.001.

Table 3.4. Genetic diversity for the *comC* locus.

	n (length)	S	π (s.d.)
Total	88 (123 bp)	14	0.0405(0.0033)
CSP 1	65 (123 bp)	5	0.0077(0.0006)
CSP 2	23 (123 bp)	1	0.0007(0.0006)

n corresponds to the number of isolates, **length** is the nucleotide sequence length of the gene in base pairs **S** is the number of segregating or polymorphic sites **π** is the nucleotide diversity estimated as the average of the Jukes and Cantor pairwise distances. In all cases **s.d.** correspond to the standard deviations of the estimates

2005 #165}. A closer look to the polymorphism reveals that there is almost no variation within CSP types in the *comC* locus. Furthermore, there seems to be more fixed differences between types than polymorphic differences, when compared to the reference locus (Table 3.5). When reciprocal 2D-HKA tests were performed on the CSP1 and CSP2 populations, significant differences in the ratio of polymorphism to divergence were observed between the *comC* locus and the reference loci. There is significantly less polymorphism in the *comC* locus than would be expected, given the divergence within pherotype subpopulations (CSP-1: $r = 0.089$, p-value < 0.02; CSP-2: $r = 0.0078$, p-value < 0.001).

Table 3.5. Synonymous (Syn) and non Synonymous (NonSyn) fixed and polymorphic differences between pherotypes 1 and 2 (within *S. pneumoniae*) and between *S.pneumoniae* and *S. mitis*.

Locus		**Fixed**	**Polymorphic**
	Comparison	Syn / Non-Syn	Syn / Non-Syn
ComC			
	S. pneumoniae vs *S. mitis*	2 / 0	16 / 24
	CSP1 vs CSP2	**2 / 8**	**6 / 0**
aroE			
	S. pneumoniae vs *S. mitis*	1 / 2	60 / 27
	CSP1 vs CSP2	0 / 0	4 / 7
gdh	revisar		
	S. pneumoniae vs *S. mitis*	5 / 1	28 / 7
	CSP1 vs CSP2	0 / 0	28 / 7

gki		
S. pneumoniae vs *S. mitis*	2 / 1	79 / 6
CSP1 vs CSP2	0 / 0	27 / 5
recP		
S. pneumoniae vs *S. mitis*	1 / 0	58 / 4
CSP1 vs CSP2	0 / 0	10 / 3
spi		
S. pneumoniae vs *S. mitis*	0 / 0	93 / 9
CSP1 vs CSP2	0 / 0	39 / 3
xpt		
S. pneumoniae vs *S. mitis*	1 / 1	81 / 23
CSP1 vs CSP2	0 / 0	22 / 9

The pair-wise difference between CSP types is higher in the *comC* locus than in the neutral loci (see appendix 2). Such pattern could be simply the result of higher mutation rates in the *comC* locus. Nevertheless, the low polymorphism within the CSP types is inconsistent with a scenario of higher mutation rates in the *comC* locus. If this was the case, and the locus is evolving neutrally, it would be expected that the polymorphism in the *comC* locus were roughly equal to the difference ("divergence") between types. Clearly this is not the case, and the ratio of polymorphism over divergence in the reference loci is roughly 114 times that of the *comC* locus. Summary

tables showing the average pairwise polymorphic differences and "divergence" for the *comC* locus and the reference loci are available in appendix 2.

3.4 Discussion and Conclusions

Bacterial population genetic structure is influenced to varying degrees by recombination (Maynard-Smith et al. 1993; Fraser, Hanage, and Spratt 2007). In the case of *Streptococcus*, inter-specific recombination is known to be limited by high levels of genetic divergence across species (Fraser, Hanage, and Spratt 2007), a limit that is not believed to play an important role within species. Here we test the hypothesis that there may be genetic mechanisms, other than sequence divergence, that limit intraspecific recombination and as a consequence influence the genetic substructure of *S. pneumoniae* populations.

The observation that strains of *S. pneumoniae* can only induce competence among cells that present the same type of competence peptide (CSP pherotype) led to the suggestion that this could be a factor maintaining genetic differentiation among subpopulations of cells (Havarstein, Hakenbeck, and Gaustad 1997; Tortosa and Dubnau 1999; Steinmoen, Knutsen, and Havarstein 2002; Claverys and Havarstein 2007). Our results suggest that this is unlikely to be the case. If such a scenario were correct, higher levels of genetic differentiation should have been observed between populations expressing different pherotypes than within populations expressing the same pherotype. In contrast, we have shown that no significant genetic differentiation is observed

between pneumococcal subpopulations bearing distinct pherotypes. A model with restricted migration between pherotypes does not fit the data as well as a model in which there is considerable exchange of genes between pherotype subpopulations (Table 3.3). This result is consistent with a recently proposed alternative and opposite prediction (Johnsborg et al, 2008), that there should be no such pherotype specific differentiation owing to the fact that the induction of competence coincides with the targeted lysis of cells bearing non-matching *comD* sequences, and thus the opposite pherotype.

We believe that our results are consistent with the mechanism underlying transformation in this species, regardless of whether the issue is considered from the perspective that pherotypes cause either assortative or disassortative recombination. Consider a simple scenario in which two populations of cells with different pherotypes coexist within the human nasopharynx, a situation likely to be common for pneumococci given their high rates of co-colonization and clonal turnover (Bogaert et al. 2004). Cell type 1 produces CSP1 and cell type 2 produces CSP2, recognized by their respective receptor types. Now imagine that a certain proportion of each cell type initiates the competence cascade and secretes its strain specific CSP. These peptides are recognized by receptors in cells belonging to the same pherotype group and this causes: i) induction of competence in like-pherotype cells; and ii) production of bacteriocins or toxins that lead to the lysis of cells of both pherotypes that have not entered the competent state (Steinmoen, Knutsen, and Havarstein 2002; Kreth et al. 2005; Guiral et al. 2006). According to this scenario cells belonging to both pherotypes will

simultaneously become competent while the non-competent cells of both pherotypes will lyse and release DNA. Because competent cells are not discriminating in their uptake of DNA, it is easy to envision that the free DNA is taken up by competent cells without regard to its pherotype. This would cause non-specific recombination among pherotypes, thus preventing pherotype specific genetic differentiation. Recently Claverys et al. (Claverys, Prudhomme, and Martin 2006) have suggested that transformation driven by this process would prevent genetic homogenization within pherotypes and increase population wide genetic diversity. While this may be a consequence of the scenario we outline, it fails to explain the existence of the polymorphic competence system itself. It could be argued that inter-pherotype recombination could be prevented if there is sufficient divergence among clones with distinct pherotypes. However, our results indicate that this precondition is not met (Table 3.1), and moreover that pherotypes are insufficient to cause this differentiation to begin with (see figure 3.1 and Table 3.3).

It is important to mention caveats associated with the fact that our analyses were performed on a geographically diverse collection of clinical isolates. First, it is possible that local differentiation among pherotypes is present but that this is obscured at a regional scale because of increasing genetic variation in each group due to geographic subdivision. However, a preliminary analysis restricted to a geographic subset (South Africa) suggests that this is not the case. Nevertheless, a more intensive and localized analysis will be necessary in order to rule out such possibility. A second caveat derives from the fact that our clones are all clinical isolates rather than clones isolated from

carriage (non-disease causing), the predominant pneumococcal lifestyle. If clones causing disease are not a representative sample of pherotypes, this would limit our ability to detect localized genetic structuring. In future work we intend to address both caveats using a more geographically and temporally localized sample of exclusively carriage isolates.

Overall, our results indicate that polymorphism in the competence peptide does not maintain genetically differentiated subpopulations of pneumococci. It remains intriguing, however, that the polymorphism in the competence system exists and that the two dominant pherotypes are maintained at such high frequencies. It has been recently suggested that *comC* and *comD* sequences across Streptococci display substitution patterns indicative of positive selection (Ichihara, Kuma, and Toh 2006). It is possible that some form of balancing selection or frequency dependent selection maintains the polymorphism. It is expected for a region under balancing selection to show higher genetic diversity than loci evolving neutrally (Charlesworth 2006; Kawabe, Fujimoto, and Charlesworth 2007), and the estimated gene diversity of the *comC* locus seems higher than that estimated for the MLSTs (see Tables 3.1 and 3.4). Nevertheless, when we compared the polymorphism observed in the *comC* locus with that observed in neutral loci, we failed to detect an excess in polymorphic changes within *S. pneumoniae*, when compared with *S. mitis*, as would be expected under a scenario of balancing selection maintaining the polymorphism (HKA test not significant).

Further exploration of the polymorphism in the *comC* locus revealed that there is less variation within pherotype than it would be expected if the locus was evolving in a

neutral fashion. The contrasting pattern informed by the results of the two tests suggests that the polymorphism in this peptide could be maintained by selection and we propose an scenario under which this polymorphism could be maintained. If an inability to become competent is associated with deleterious fitness effects, then any loss of function or reduction in the efficiency of the function will be selected against. If this is so, any mutation in the signal peptide that induces competence (CSP), or its receptor, leading to reduction in the recognition of the two component system with a concomitant effect on competence will have two possible fates: i) it will be selected against and that mutant will be lost from the population; or ii) compensatory mutations in the receptor (or signal peptide) that restore the efficiency of the recognition will occur before the original mutant is lost, and will be maintained in the population because the process itself has been restored (negative selection maintaining the competence phenotype). The first possible fate, which may be the dominant fate due to stabilizing selection, is not of direct interest because this would not result in the development of a polymorphic system. On the other hand, the second fate is relevant because the restoration of wild-type competence by a form of compensation will have generated the initial variation necessary to cause polymorphism. This scenario for the evolution of lock and key components, or more generally for coevolving residues that structurally interact, in which mispairing of the alleles for signal and receptor significantly reduce the efficiency of the competence, could account for the emergence and further maintenance of polymorphism in this system. A similar hypothesis has been recently proposed for the

evolution of highly polymorphic self-incompatibility systems in crucifers (Chookajorn et al. 2004).

We argue that the reduced polymorphism found within pherotype subpopulations in the *comC* locus and the large number of fixed differences between pherotypes in the *comC* locus is consistent with the hypothesis we propose. The rapid fixation of the changes that maintain the function of the *comC* locus contribute to the low variability within CSP types. The reduced variability within CSP types represents a substantial difference with the mechanism proposed by Chookajorn et al., according to which variation within alleles is necessary for the diversification of the locus (Chookajorn et al. 2004). Of the 8 proposed CSP alleles reported in the literature (Kilian et al. 2008), only two have been assessed experimentally and have shown lack of cross recognition by the corresponding alleles. It is possible that some of the alternative alleles proposed are intermediate forms with reduced recognition but not completely differentiated from the well established ones, and the definition of the alleles should change after proper testing.

The testable hypothesis proposed in this work is consistent both with the polymorphism in *comC* loci, the low polymorphism within *comC* allele, and the lack of genetic differentiation between subpopulations of isolates presenting different pherotypes. We consider that further experimental and sequence analysis is necessary to adequately test the proposed hypothesis.

CHAPTER 4

The Population and Evolutionary Dynamics of Homologous Gene Recombination in Bacterial Populations

Bruce R. Levin and Omar E. Cornejo
accepted for publication PLoS Genetics (2009)

4.1 Introduction

Recombination in the form of the receipt and incorporation of genes and genetic elements from other strains and species of bacteria (Mazodier and Davies 1991) as well as archaea and eukaryotes (Nelson et al. 1999; Brown 2003; Hughes and Friedman 2003; Denker et al. 2008; Keeling and Palmer 2008; Pace et al. 2008) plays a prominent role as a source of variation for the adaptive evolution of many species of bacteria (Shea et al. 1996; Lawrence and Ochman 1998; Bergstrom, Lipsitch, and Levin 2000; Campbell 2000; Levin and Bergstrom 2000; Ochman, Lawrence, and Groisman 2000; Koonin, Makarova, and Aravind 2001; Cazalet et al. 2004; Thomas and Nielsen 2005; Coleman et al. 2006; Gal-Mor and Finlay 2006). Because of this ability to acquire genes and genetic elements from other organisms, horizontal gene transfer (HGT), the pace of adaptive evolution in bacteria need not be limited by the standing genetic variation within a population or the slow rate by which adaptive genes are generated by recurrent mutation. Through single HGT events bacteria can obtain chromosomal genes and gene clusters (islands) as well as plasmids, transposons and prophage bearing genes that have successfully traversed the gauntlet of natural selection in some other population of their own or other species. In this way, bacteria can expand their ecological niches; colonize new habitats and hosts, metabolize new energy sources, synthesize essential nutrients,

survive toxic agents like antibiotics, and alas, increase their virulence to human and other hosts.

Less clear are the ecological and evolutionary consequences of more mundane HGT events, such as homologous gene recombination (HGR) among members of the same population. In accord with classical population genetic theory, meiotic recombination of can increase the rate at which populations adapt to new environments by assembling in single organisms combinations of adaptive mutations occurring in different members of their population and by reducing the rate at which populations accumulate deleterious mutations, ("Muller's Ratchet"). For superb reviews of this classical theory and some of its more recent extensions see (Felsenstein 1974; Felsenstein and Yokoyama 1976; Otto and Gerstein 2006).

There are good theoretical reasons to anticipate that recombination among chromosomal genes already present and those generated by mutation within a bacterial population can augment its rate of evolution in a variety of ecological situations (Evans 1986). There is also experimental evidence in support of this prediction. Transformation occurs at measurable rates in *B. subtilis* maintained in a more or less natural setting and the resulting HGR appears to promote adaptive evolution in these populations (Graham and Istock 1978). Two recent experimental studies comparing rates of evolution among recombining and non-recombining populations provide direct evidence that capacity for F-plasmid mediated recombination in *E. coli* (Cooper 2007), and transformation-mediated recombination in *Helicobacter pylori* (Baltrus, Guillemin, and Phillips 2008), can increase the rate at which these bacteria adapt to culture conditions. However, there is also evidence from studies with experimental populations of *E. coli* (Souza, Turner, and Lenski 1997) and *Acinetobacter baylyi* (Bacher, Metzgar, and de Crecy-Lagard 2006) indicating that there are conditions where there are no differences in the rates at which recombining and non-recombining populations adapt to environmental conditions.

On first consideration, it would seem that if recombination increases the rate at which populations adapt to their environment, the capacity for shuffling homologous genes within a population would provide an advantage to the recombining strain when competing with populations without this capacity. Not so clear are the *conditions* under which selection will favor recombining populations in this way. When will a recombining population prevail over non- or more slowly- recombining populations, and do so in the face of fitness costs associated with the capacity for recombination?

Here, we present the results of a study using computer simulations of mutation, recombination, selection and inter-population competition to explore the conditions under which: i) recombination augments rates of evolution in bacterial populations and, ii) when the capacity for HGR will be favored in competition with non-recombining populations. We demonstrate that under broad conditions, HGR occurring at rates in a range estimated for *E. coli, H. influenza, S. pneumoniae,*, and *B. subtilis* can increase the rate of adaptive evolution in bacterial populations. We show that this capacity for increasing rates of evolution by shuffling chromosomal genes can provide a recombining population a selective advantage in competition with populations without this capacity even when the recombining population has a lower intrinsic fitness. On the other hand, we also demonstrate that because the rate of recombination in bacteria depends on the density of the recombining population, the conditions under which recombination can provide a population a selective advantage in competition with non-recombining populations are restricted to when the recombining population is relatively common and the total population density is high. Even in the absence of a fitness cost, when the recombining population is rare, it will not be favored despite its ability to acquire genes from the dominant non-recombining population. We discuss the implications of these simulation results to the role of recombination in the adaptive evolution of bacteria and the evolution and maintenance of different mechanisms for homologous gene recombination in bacteria.

4.2 Methods

SEMI-STOCHASTIC SIMULATIONS OF MUTATION, SELECTION AND RECOMBINATION IN A BACTERIAL POPULATION

Single population simulation: In this simulation we consider a population of bacteria in mass (liquid) culture and five loci each with three alleles designated, 1, 2 and 3. *P(I,J,K,L,M)* is the relative frequency of the *I,J,K,L,M* genotype where, *I, J, K, L* and *M* take values 1, 2, or 3 and $\sum_I\sum_J\sum_K\sum_L\sum_M P(I,J,K,L,M)= 1$; there are $3^5 = 243$ possible genotypes. We assume that the fitness of a genotype is proportional to the sum of the values of the alleles to some power, e $(e\geq 1)$, with all five genes contributing equally to fitness. We also assume that the contribution of the number 2 allele at any locus is intermediate between that of the number 1 and number 3 alleles. Thus the fitness of a genotype I,J,K,L,M,

$$W(I,J,K,L,M) = c + (1-c)\frac{(I^e + J^e + K^e + L^e + M^e)}{5(2^e)}$$

Where $(1- c)$ is a measure of the extent to which these five loci contribute to fitness. For example with c=0, and e=1, the fitness of a genotype, 1,2,2,3,1 would be 9/10 = 0.90. The total range of fitness values would be 0.5 to 1.5, with 1,1,1,1,1 being the genotype of lowest fitness and 3,3,3,3,3 being the genotype of highest fitness. If c=0.5 and e=2, the range of fitness values would be 0.625 for the 1,1,1,1,1 genotype to 1.625 for the 3,3,3,3,3 genotype. With this fitness function, the 2,2,2,2,2 genotype would have a relative fitness of 1.0 independently of the value of e. The course of this simulation is diagrammed in Figure 4.1. In the following we describe the different steps in the simulation.

Mutation: For individuals of all genotypes there is a probability μ per cell per generation of a mutation occurring in one of the five loci. For convenience we assume there is one generation per hour. To simulate this process, at each time interval, Δt, the

probability that a mutant will be generated is $PM = N^*\mu^*\Delta t$, where N *bacteria per ml* is total number of individuals in the population as well as the density. A pseudo random number, r, from a rectangular distribution $0 \le r \le 1$ is generated. If $r \le PM$, a mutation occurs **(YES)**, if $r > PM$ it does not **(NO)**. In these simulations, we use values of N and Δt such that at any given time $PM<1$. If a YES decision is made, we select the genotype

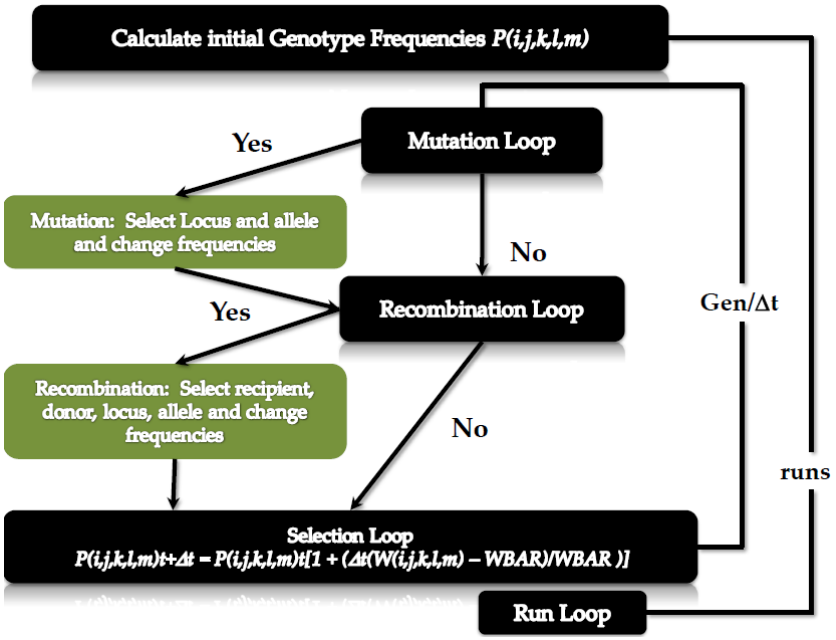

Figure 4.1. Five-locus, three allele simulation of mutation, recombination, selection and a bacterial population.

that may be changed by mutation. For this and similar decision processes, a pseudo random number is generated and sequentially compared to the sum of the probabilities of the different outcomes, which in this case are frequencies of the different genotypes. The cumulative sum of the frequencies of the different genotypes are continually calculated, $P(1,1,1,1,1) + P(1,1,1,1,2) + P(1,1,1,1,3)$ … and as soon as that sum exceeds the value of the random number, r, the process terminates, and the last genotype in the sequence is chosen to receive the mutation. Using a similar Monte Carlo decision process we then choose the locus for the mutation under the assumption that each locus has the same probability of being replaced, $y=1/5$. For example, if $r = 0.4567$, the third locus, J, is subject to change by mutation and if $r=0.9532$, the fifth locus, M, is subject to

change by mutation. Following the choice of the genotype and locus for mutation, another Monte Carlo decision process is used to determine the allelic state of the mutation. Here we use three probabilities, xm_1, xm_2 *and* xm_3, for the comparison, where $xm_1 + xm_2 + xm_3 = 1.0$. In this simulation a mutation need not result in the change an allele. For example, if the 3rd locus is chosen and bears a 2 allele, there is a probability xm_2 that that allele remains in a 2 state. Upon choosing the genotype to be changed, the frequency of the mutated genotype is reduced by $1/N$ and that of the mutant type is increased by $1/N$.

Recombination Loop– As in (Levin 1981) we assume recombination is a mass action process that occurs at a rate proportional to the product of the densities of the donor and recipient populations. We also assume that: (i) the five loci are sufficiently far apart that only a single gene is replaced in any HGR event, (ii) all loci have equal probabilities of being subject to allelic replacement by recombination, (iii) all individuals in the population are equally likely to serve as a donor and recipient. Thus, if the density of the population is N, the probability of a recombination event occurring during the finite time interval Δt is $PR = \chi * N^2 \Delta t$ where χ *ml/cell²/hour* is the rate parameter of recombination for those five loci (Chao and Levin 1981). When the value of χ is relatively low, Δt is set so that $0 \leq PR < 1$ and the decision regarding whether recombination occurs is random. If the random number $r \leq PR$, recombination occurs **(YES)** and if $r > PR$, it does not occur **(NO).** To reduce the number of random numbers generated, when χ is larger (usually 5×10^{-14} or greater) the recombination decision process is deterministic, and $\chi * N^{2*} \Delta t$ recombination events occur during the interval Δt.

Independently of the value of χ, the choice of the recipient, donor, locus and allele for recombination is stochastic. For this we use the random number decision process similar to that described above for choosing the genotype subject to mutation. The final step in the recombination loop is to choose the locus in the recipient that is subject to replacement by that locus from the donor. The recombination loop terminates by

increasing the frequency of the recombinant genotype by *1/N* and reducing that of the recipient genotype by *1/N*. The frequency of the donor genotype remains unchanged.

Selection Loop: The mean fitness of all the genotypes is calculated as the sum of the product of the post mutation post recombination frequencies of each genotype and their relative fitness,

$$\bar{W} = \Sigma_I \Sigma_J \Sigma_K \Sigma_L \Sigma_M P(I,J,K,L,M) * W(I,J,K,L,M)$$

Where I,J,K,L, and M take values 1, 2 or 3.

The frequency of that genotype in the next time interval $t+\Delta t$ is then calculated as

$$P(I,J,K,L,M)_{t+\Delta t} = P(I,J,K,L,M)_t \left[1 + \frac{\Delta t \left(W(I,J,K,L,M) - \bar{W} \right)}{\bar{W}} \right]$$

Once this is done for all genotypes, a new cycle of mutation, recombination and selection commences and the relative frequencies of the different genotypes in the population are adjusted accordingly. Unless otherwise indicated, iterations through this simulation continue for a defined number of generations and a defined number of runs.

Two competing populations: To explore the conditions under which recombination will be favored in the presence of a competing population with a different rate of recombination, we allow for a second population. The progress of both populations through the mutation, recombination, and selection process are identical to that described above for a single population. The two competing populations can, however, have different fitness functions, so that

$$W(I,J,K,L,M)_i = Z_i \left[c_i + (1 - c_i) \frac{(I^{e_i} + J^{e_i} + K^{e_i} + L^{e_i} + M^{e_i})}{5(2^{e_i})} \right]$$

The parameters Z_i (i=1, 2) are measures of the intrinsic fitness of two populations $(0 \leq Z_i \leq 1)$ and the e_i's are the exponents. In these two population simulations we also allow for different rate parameters of recombination, χ_1 and χ_2 and different densities, N_1 and N_2, but maintain a constant total density $N_1+N_2 = N_T$. In these simulations both the recipients or donors can come from the same population or the donors can come from either population. In the former case, the rate of recombination is proportional to the square of the densities of the population in which recombination is occurring, N_1^2 or N_2^2. In the latter case, the rate of recombination is proportional to the product of the density of the recipient population and the total density, N_1*N_T and N_2*N_T. In situations where all cells in the population can serve as donors, the likelihood of a population serving as a donor is proportional to its frequency in the community.

The densities of these two populations and thus the relative frequencies of these genotypes can change by clone level selection. For this we assume the mean fitness of the two populations $\overline{W}_B = q_1 * \overline{W}_1 + q_2 * \overline{W}_2$

Where $q_1 = N_1/N_T$ and $q_2 = N_2/N_T$ (since $N_1+N_2=N_T$, $q_1+q_2=1$).

For the next time interval, $t+\Delta t$,

$$q_{1_{ty+\Delta t}} = q_{1_t} + \left[\Delta t \frac{\bar{W}_{1_t} - \bar{W}_{B_t}}{\bar{W}_{B_t}} \right]$$

and

$$N_{1_{t+\Delta t}} = N_T * q_{1_{t+\Delta t}}, \text{ and } N_{2_{t+\Delta t}} = N_T - N_1$$

Starting conditions: The simulation runs to be considered were started in one of two ways: (i) with only a single genotype of intermediate fitness, $P(2,2,2,2,2)$=1 or (ii)with a random selection of nc genotypes, $(1 < nc < 3^5)$ for each of the two populations. For the second starting condition (ii) the random number decision process was run nc times

each time with picking an allele 1,2 or 3 at each of the five loci and assigning that genotype a random number *(0<r<1).* If the same genotype is selected multiple times, its relative frequency would be proportional to the sum of the random numbers chosen for that genotype. In simulations initiated with the second starting condition the relative frequencies of the genotypes in the population are normalized by dividing by the sum of the frequencies. Unless otherwise noted, at the start of each of the runs made with two competing populations, both initially had the same *nc* clones although the clones chosen varied between runs.

Parameter Values: The total population size was maintained at 10^8 for simulations with single populations and at $2x10^8$ for the simulations with two competing populations. Although we use different rates of mutation $\mu = 10^{-8}$, 10^{-7} and 10^{-6}, in all runs we assume that mutations to the lower fitness 1 allele occur at a rate greater than that for intermediate fitness allele 2, which in turn occur at a rate greater than that for the highest fitness 3 allele, i.e. $xm_1 = 0.80$, $xm_2 = 0.15$ and $xm_3 = 0.05$ independently of the existing state of the allele. The recombination rate parameters in these simulations are within range anticipated for *E. coli, H. influenzae*, *B. subtilis* and *S. pneumoniae* (see the Discussion).

In Figure 2, we illustrate the relationship between the number of 1, 2 and 3 alleles and the fitness *W(I,J,K,L,M)* of the different genotypes for the different simulations employed. For example, the genotype (1,3,2,2,1) has 2 #1 alleles, 2 #2 alleles and 1 number #3 alele and in this figure would be designated as $\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$. With the parameters, c=0.5 and e=2, the fitness of cells of this (1,3,2,2,1) genotype relative to the (2,2,2,2,2) genotype $\begin{pmatrix} 0 \\ 5 \\ 0 \end{pmatrix}$ would be, from equation 1.

$$W(1,3,2,2,1) = 0.5 + 0.5\,(1^2+3^2+2^2+2^2+1^2)/(5*2^2) = 0.975$$

In Figure 2A, the exponent, *e=1*, and the contribution of the five loci to fitness are either 0.1 or 0.5. In Figure 2B, the contribution of the five loci to fitness is 0.5, and the exponent e takes values1, 2, and 3. As this exponent increases the contribution of the higher index allele 3 to the fitness of a genotype increases, as does the magnitude of the fitness differential.

Simulations: The programs for this simulation study were written in quaint but fast FORTRAN 77. Copies of the FORTRAN code and/or executable files for Windows and Mac and instructions for their use are available from www.eclf.net.

4.3 RESULTS

We open our analysis with a consideration of the contribution of recombination to the rate of increase in the average fitness of single populations of bacteria. In these simulations, the five loci contribute equally to fitness and the three alleles at each locus, 1, 2 and 3 contribute additively. Mutation and recombination are random processes with all five loci equally likely to change in any given mutation or recombination event. In the case of mutation, the change in allelic state is independent of the genetic structure of the population. For recombination, the likelihood of a particular change in the allelic state of any of the five genes in recipient is proportional to the frequencies of those alleles in the population at large. Selection is a deterministic process with fitness being proportional to the frequency of high index alleles (see Figure 4.2). (For more details we encourage the reader to at least peruse the METHODS section, which we believe is written a way that would be amenable to those who prefer to hum equations than solve them.)

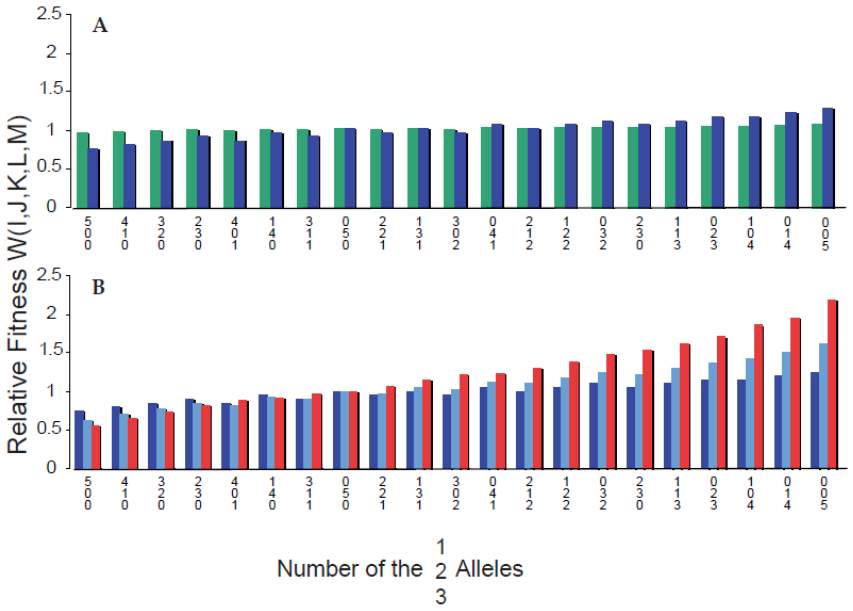

Figure 4.2. Relative fitness of the different genotypes *W(I,J,K,L,M)* as a function of the number of 1, 2 and 3 alleles as calculated from equations (1). (a) green: c=0.9, e=1.0, blue: c=0.5, e=1.0; (b) blue: *c*=0.5, e=1.0, light blue: c=0.5, e=2.0, red: c=0.5, e=3.0.

The rate at which a population adapts to its environment, as measured by the increase in its mean fitness is directly proportional to the rate of recombination and the relative magnitude to which the five loci contribute fitness, as measured by the parameter 1-*c* (Figure 4.3). If there is more variation in the population at the start of a simulation, as there is when we start with 10 or 50 randomly selected lineages (Figure 4.4), the rate of evolution is faster than when the population is initially monomorphic as in Figure 4.3. This is of course anticipated from Fisher's fundamental theorem of natural selection (Fisher 1930) as well as simple logic.

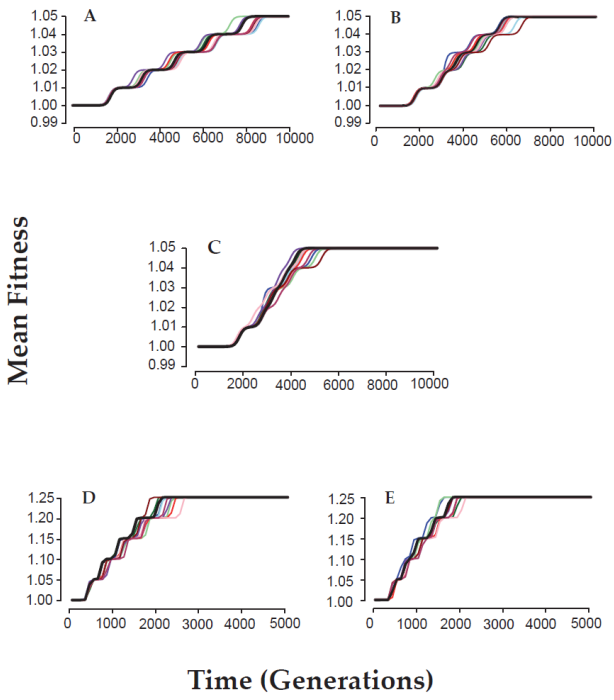

Figure 4.3. Change in mean fitness with different rates of recombination for ten independent runs initiated with a population monomorphic for the 2,2,2,2,2 genotype. Parameters, $N=10^8$, $\mu=10^{-8}$, *xm1=0.80, xm2=0.15, xm3=0.05*, $c=0.90$ in (a), (b), and (c) and $c=0.50$ in (d) and (e) the exponent e=1. (a) No recombination, (b) $\chi=10^{-15}$, (c) $\chi=10^{-14}$, (d) no recombination, (e) $\chi=10^{-15}$. The thick dark line in these figures is the mean of the 10 runs.

To provide an overview of the relative contributions of mutation and recombination to the rate of evolution in this model, we did fifty simulations with different mutation and recombination rates. We started these runs with either a single genotype of intermediate fitness, 2,2,2,2,2 genotype, or 10 randomly chosen clones (Figures 4.5A and 4.5B, respectively). In these figures we plot the mean and standard error of the time required for the mean fitness of the population to reach 0.001 less the maximum fitness, $\overline{W}_{\max}=1.25$. As anticipated from the results presented in Figures 4.3 and 4.4, the time to reach maximum fitness decreases with the rates of recombination and mutation. Saying this another way the rate of adaptive evolution increases with the rate of recombination and mutation. Although the mutation process is biased towards generating lower fitness alleles, with only five loci and the fitness and other parameters employed, the effects of generating less fit mutations on the average fitness are imperceptible. This is

the case even with a mutation rate of 10^{-5}, all of the variation in fitness determined by these five loci, *c=0*, and populations initiated with the highest fitness genotype 3,3,3,3,3 (data not shown). Although lower fitness mutants are produced, they are purged by "purifying" selection and do not accumulate.

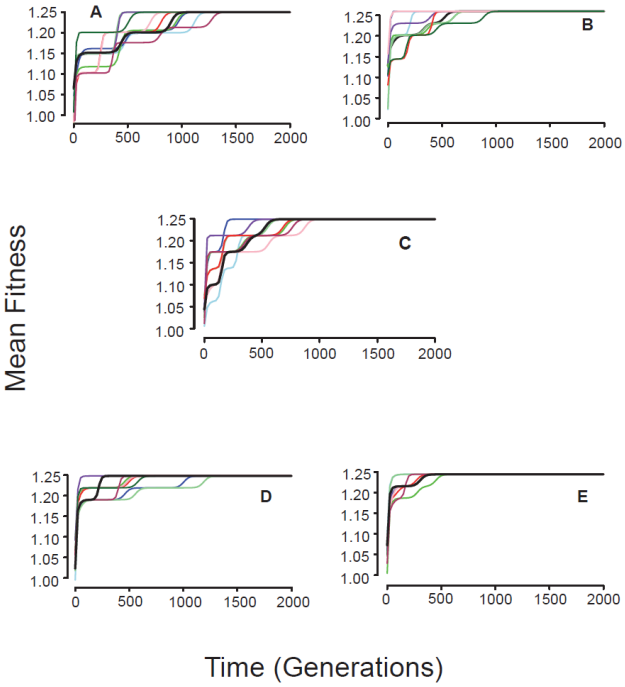

Figure 4.4. Increase in mean fitness for 10 independent runs with random start of nc =10 or 50 clones. In all cases, *c=0.5*, *e=1*, $\mu=10^{-8}$, $N=10^8$. (a) *c*=0, *nc=10* (b) $c=10^{-15}$ *nc=10*, (c) $\chi=10^{-14}$, nc=10, (d) $\chi=0$, *nc=50*, (e) $\chi=10^{-15}$, *nc=50*. . The thick dark line is the mean fitness of all 10 runs.

As noted in Figure 4.3, the extent to which recombination increases the rate of evolution is proportional to the intensity of selection at the loci subject to recombination, the selection differential. To explore this relationship a bit more and begin to consider the contribution of the form of the fitness function, we have performed simulations with *c=0.5* and exponents *e =1*, *e=2* and *e=3* (see Figure 4.2B). As *e* increases, the contribution of the higher number alleles becomes proportionally greater and the time to reach maximum fitness is reduced. The results of these "experiments" are presented in Figure 4.6.

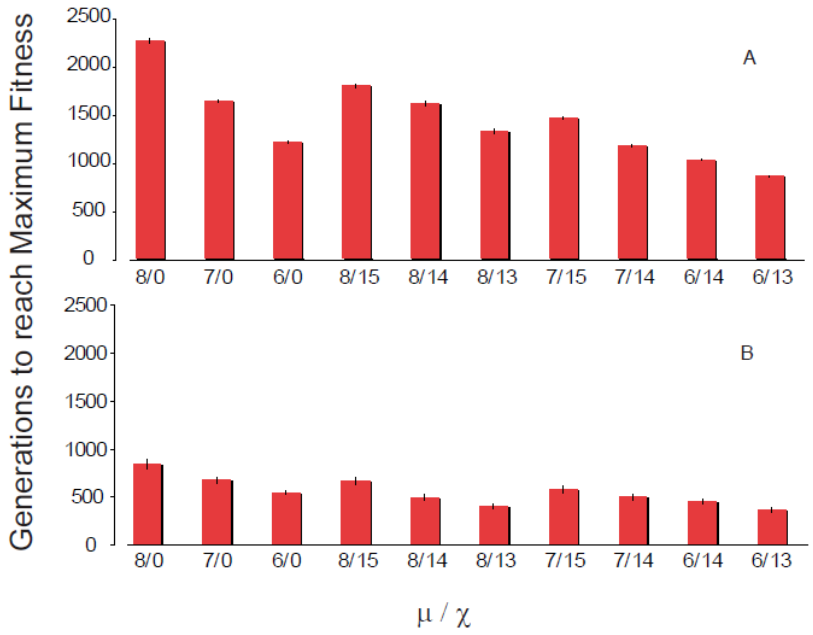

Figure 4.5. Time to reach the maximum mean fitness -0.001($\overline{W}_{\max} - 0.001$) for different rates of mutation and recombination. For the mutation rates (μ) -8 is 10^{-8}, and -7 is 10^{-7} per cell per hour. For the recombination rate constants (χ), 0 is no recombination, -15 is $\chi=10^{-15}$, -14 is $\chi=10^{-14}$. In all simulations $c=0.5$, $e=1$ and density and total population size is 10^8 cells per ml. Means and standard errors of the time to reach maximum fitness for 50 runs with each set of parameters. (a) Runs initiated with a single intermediate fitness clone, 2,2,2,2,2. (b) Runs initiated with 10 randomly chosen clones.

To provide a more detailed view of the contribution of the initial variability to the effects of recombination on the rate of evolution, we made 50 runs with each set of parameters. Each run would terminate when the mean fitness was nearly its maximum ($\bar{W}_{MAX} - 0.001$) or 5000 generations passed. The results of these simulation experiments are presented in Table 4.1. To better evaluate the relative contributions of the initial variability and the rate of recombination, we performed a two-way analysis of variance (ANOVA) on the first three rows (2 df) and four columns (3 df). Increasing the amount genetic variability in the population at the start of each run and the rate of recombination significantly increases the rate of evolution, ($p < 10^{-16}$). There is also a significant interaction $p<0.005$ for the combined effects of initial variability and rate of recombination.

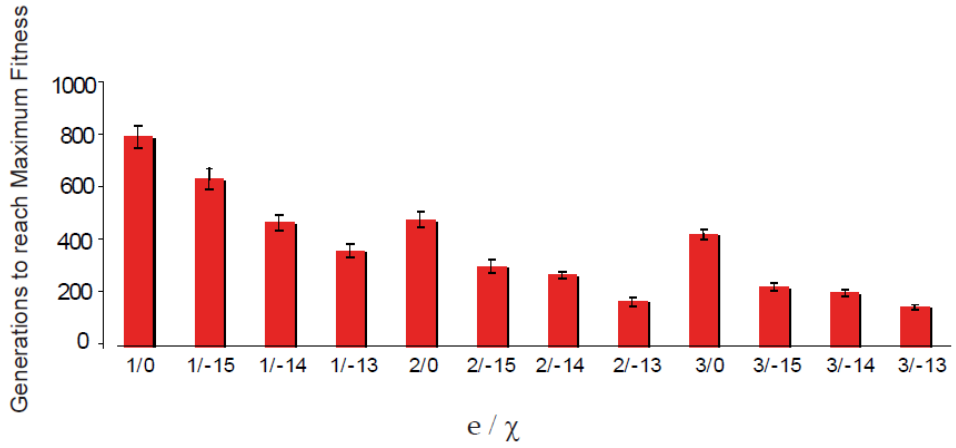

Figure 4.6. The contribution of intensity and form of the fitness function on the extent to which mutation and recombination augment rates of adaptive evolution; mean and standard error of the time before the population reaches its maximum value $\overline{W}_{\max} - 0.001$ for different values of the exponent e, and recombination rate parameter (χ). For the latter -15 is $\chi=10^{-15}$ and -14 is $\chi=10^{-14}$. In all simulations, $\mu=10^{-8}$, $N_T=10^8$, $c=0.5$. Each simulation was initiated with 10 randomly chosen genotypes and 50 independent simulations were run with each set of parameters.

Competition between populations with different rates of recombination:

To explore the conditions under which the capacity for homologous gene recombination will provide an advantage to a population, we consider mixtures of two genetically distinct populations, one of which does not recombine (in which variation is only generated by mutation), or recombines at a lower rate than the other. For each population, mutation, recombination and selection occur as described for the single population simulations. Although the total density of the two-population community remains constant, the densities and relative frequencies the two competitors change at a rate that depends on their respective mean fitness. Unless otherwise stated, recombination only occurs within a population. In these simulations, the rate parameter of recombination of the #1 population exceeds that of the #2 population, $\chi_1 > \chi_2$ and in most cases χ_2 is 0.

In Figure 4.7, we follow the changes in the ratio of the two populations for different situations with the #1 and #2 populations initially monomorphic for the intermediate fitness genotype, *2,2,2,2,2*. If there is no cost to recombination and initially the #1 and #2 populations are equally frequent, the recombining #1 population has an advantage over

the one that is not recombining, #2 (Figure 4.7A), i.e. in 9/10 runs the recombining populations prevailed. When there is 1% fitness cost associated with recombination and initially equal frequencies of the #1s and #2 populations, in the majority of runs the non-recombining population has an advantage (Figure 4.7B).

Table 4.1. The contribution of the initial number of clones and the rate of recombination to the time before the population reaches its maximum fitness W_{MAX} (mean ± standard error to the nearest generation). Standard parameters, c=0.5, μ=10^{-8}, $xm1$=0.80, $xm2$=0.15, $xm3$=0.05.

Initial Clones	χ =0	χ =10^{-15}	χ =10^{-14}	χ =10^{-13}
1 *	2248 ± 24	1785 ± 17	1603 ± 22	1383 ± 31
10	935 ± 49	668 ± 44	509 ± 29	450 ± 26
50	626 ± 41	395 ± 27	322 ± 22	278 ± 15

* All simulations were initiated with only the 2,2,2,2,2 genotype.

Although in the absence of an intrinsic fitness cost, HGR provides a clear advantage when the recombining population is common, this is not necessarily the case when the recombining population is initially rare (Figure 4.7C). On the other hand, the capacity for HGR can prevent the establishment of an initially rare, higher fitness, non-recombining population (Figure 4.7D).

In the simulations described above (Figure 4.7) the population are initially monomorphic and recombination does not come into play until sufficient variation builds up through recurrent mutation (see Figure 4.3). Qualitatively, the results obtained with runs stated with 10 randomly selected clones are similar to those initiated with no variability (compare Figures 4.7 and 4.8), but there are quantitative differences. The most conspicuous of the quantitative differences between the results presented in Figure 4.7 and Figure 4.8 is that when the populations are initially variable, the outcome of competition is more likely to end in a stalemate than the loss of the #1 or #2 population. This was particularly true for the runs initiated with a rare recombining population (Figure 4.8C). The reason for this stalemate is that the two populations both reach the maximum fitness before the run terminates and selection ceases.

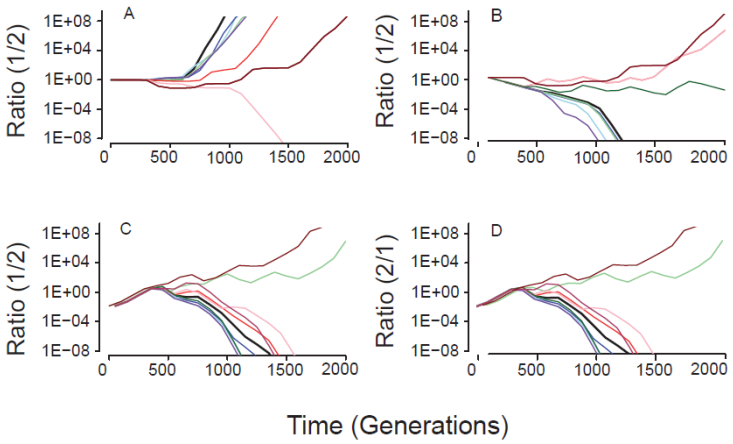

Figure 4.7 – Competition between populations with different rates of recombination. In (a), (b) and (c) we plot the changes in the ratio of the higher rate recombining population #1, relative to the lower rate recombining population, #2, 1:2 . In Figure (d) we plot change in the ratio of non-recombining #2 to recombining populations, #1, 2:1. The results of 10 independent runs for each initiated with a monomorphic, 2,2,2,2,2 population. In the runs depicted the total population size $N_T=2x10^8$, $\mu= 10^{-8}$, $\chi_1=5x10^{-15}$, $\chi_2 = 0$, $c= 0.50$, $e=1.0$. (a) Initially equal densities of #1 and #2 and no fitness cost associated with recombination. (b) Initial equal frequencies of #1 and #2 and a 1% fitness cost for the recombining population #1. (c) The recombining population is initially rare and there is no fitness cost associated with recombination. (d) The non-recombining population is initially rare and there is a 2% fitness cost associated with recombination.

To provide a broader and more quantitative perspective of the effects of recombination on the outcome of competition, including invasion and prevention of invasion, we made 50 independent runs with different rates of recombination, different initial frequencies of the #1 and #2 populations and different fitness costs. As controls for these "experiments" we used 100 simulations with the two populations having the same recombination rates. These simulations were run until the density of one population fell below 10^5 (the total density remained constant at $2x10^8$) or 2000 generations passed. The results of these experimental and control simulations are presented in Table 4.2. In these simulations a population "won" the competition when the density of the other population fell below 10^5 or when it had the highest relative frequency after 2000 generations passed.

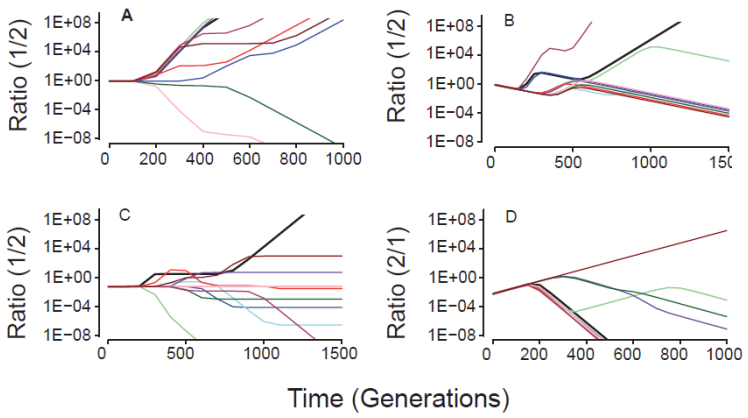

Figure 4.8– Competition between populations with different rates of recombination. In (a), (b) and (c) we plot the changes in the ratio of the higher rate recombining population #1, relative to the lower rate recombining population, #2, 1:2 . Runs initiated with 10 randomly chosen clones with identical starting populations for the 1 and 2 populations. In all runs the runs depicted the total population size $N_T=2x10^8$, $\mu= 10^{-8}$, $\chi_1=5x10^{15}$, $\chi_2 =0$, c= 0.50, e =1.0. (a) Initially equal densities of 1 and 2 and no fitness cost associated with recombination. (b) Initial equal frequencies of #1 and #2M and a 1% fitness cost for the recombining population 1. (c) The recombining population is initially rare and there is no fitness cost associated with recombination. (d) Ratio of non-recombining to recombining 2:1, the non-recombining population is initially rare and there is a 2% fitness cost for the recombining population.

For the simulations initiated with a monomorphic, all genotype *2,2,2,2,2* populations, in the absence of a cost and initially equal frequencies, the recombining population "won" in all 50 simulations. When the initial frequency of the recombining population was 0.05, the recombining population won about 20% of the time. When the initial frequency of the #1 was 0.005, the non-recombining #2 won in all 50 simulations. In the parallel no-fitness-cost runs with the initially variable population and initially equal densities in the majority (but not all) of the runs the #1 populations "won" or was "winning" when the runs terminated at 2000 generations. These runs illustrate how the advantages of recombination are reduced when the initial frequency of the recombining population is lower than that of the non-recombining population. The largest quantitative difference between the populations with and without initial variability is in the time required for this outcome (winning) to obtain, which, as suggested by the single population runs, is longer in the initially monomorphic populations. That recombination was responsible for the winning # 1 population in these no-fitness-cost runs with the recombining

population initially rare, can be seen from the controls where there was no recombination in the # 1 population.

Table 4.2 – Competition between a recombining (#1) and non-recombining population (#2): Effects initial variability, recombination rate, and fitness costs of recombination. Number of times each population Wins (the density of the competing population falls below 10^5) or is Winning (is the dominant population at the 2000th generation) in 50 or 100 independent runs. TF is the mean number of generations before the density of the losing population falls below 10^5 ±SE: standard error.

Runs initiated with a monomorphic 2,2,2,2,2 population **Freq of #1 / χ / s**	**Runs where # 1 Wins (TF±SE)**	**Runs where # 2 Wins (TF±SE)**	**Runs with #1 Winning at T=2000**	**Runs with #2 Winning at T=2000**
0.5 / $5x10^{-15}$/ 0.01	**50*** (1180±60)	**0**	**0**	**0**
0.05 / $5x10^{-15}$/ 0.01	**9*** (1151±9)$^{++}$	**41** (726 ± 47)	**0**	**0**
0.005 / $5x10^{-15}$/ 0.01	**0**	**50** (424±18)	**0**	**0**
0.5 / $5x10^{-15}$/ 0.01	**2** (1351)	**47** (747±33)	**0**	**1**
0.5 / $5x10^{-15}$/ 0.02	0	50 (391 ± 82)	0	0
0.5 / $5x10^{-14}$/ 0.02	0	50 (389±5)	0	0
0.5 / $5x10^{-13}$/ 0.02	0	50 (378±2)	0	0
0.95 / $5x10^{-15}$/ 0.02	**1**	**48** (720±35)	**0**	**1**
0.95/ 10^{-8}/ $5x10^{-14}$/ 0.02	**5** (960 ±65)$^{++}$	**43** (732 ±47)	**1**	**1**
0.95 /10^{-8}/ $5x10^{-13}$/ 0.02	**23*** (1023 ±47)$^{++}$	**25** (739 ±34)	**2**	**0**
0.5 / $5x10^{-15}$ / 0.01 **Cont.**	**0**	**100** (642±8)	**0**	**0**
0.5 / $5x10^{-15}$ / 0.02 **Cont.**	**0**	**100** (398 ± 4)	**0**	**0**
0.05 / $5x10^{-15}$/ 0.00 **Cont.**	**0**	**100** (479 ± 12)	**0**	**0**
0.005 / $5x10^{-15}$/0.00 **Cont.**	**0**	**100** (432 ± 7)	**0**	**0**
Runs initiated with 10 random clones **Freq #1 / χ / s**	**Runs where # 1 Wins (TF±SE)**	**Runs where # 2 Wins (TF±SE)**	**Runs with #1 Winning at T=2000**	**Runs with #1 Winning at T=2000**
0.5 / $5x10^{-15}$/ 0.01	**37*** (453±30)	**3**	**9**	**1**

0.05 / $5x10^{-15}$/ 0.01	**8*** (675±101)	**28** (521±44)	**9**	**5**
0.005 / $5x10^{-15}$/ 0.01	**0**	**50** (424±18)	**0**	**0**
0.5 / $5x10^{-15}$/ 0.01	**24*** (533±50) ++	**21** (1421±80)	**0**	**5**
0.5 / $5x10^{-15}$/ 0.02	**11*** (608±108)	**39** (739 ± 38)	**0**	**0**
0.5 / $5x10^{-14}$/ 0.02	**21*** (523±62) ++	**28** (1070±81)	**0**	**1**
0.5 / $5x10^{-13}$/ 0.02	**28*** (425±36) ++	**22** (942 ± 75)	**0**	**0**
0.95 / $5x10^{-15}$/ 0.02	**31*** (541±57) ++	**19** (1110±101)	**0**	**0**
0.95/ $5x10^{-14}$/ 0.02	**41*** (390 ±29) ++	**9** (1024 ±123)	**0**	**0**
0.95 / $5x10^{-13}$/ 0.02	**40*** (313±22) ++	**10** (1073 ±86)	**0**	**0**
0.5 / $5x10^{-15}$/ 0.01 **Cont.**	**0**	**100** (671±19)	**0**	**0**
0.5 / $5x10^{-15}$/ 0.02 **Cont.**	**0**	**100** (365 ± 3)	**0**	**0**
0.05 / $5x10^{-15}$/0.00 **Cont.**	**0**	**82** (415 ± 18)	**0**	**18**
0.005/ $5x10^{-15}$/0.00 **Cont.**	**0**	**100** (279 ± 9)	**0**	**0**

In all these simulations, $\mu=10^{-8}$, $c=0.5$, $e=1$, and $N_T=2x10^8$. The initial frequencies and selection coefficients, *s*, are for the **#1** population. The simulations labeled **Cont.** are controls in which both the #1 and #2 populations recombine with a rate constant $X=10^{-15}$. In the experimental runs, there is no recombination in the # 1 population. * significant with $p < 0.002$ (X^2 contingency table comparing number of wins of #1 and #2 in treatments vs. control). ++ significant with $p < 0.005$ (Student's t-test comparing TF between Population 1 and Population 2 within treatment).

With a 1% or 2% fitness cost associated with recombination, even when the recombining and non-recombining populations are initially equally frequent, the non-recombining population almost invariably prevails when the competitors are initially monomorphic. A very different situation obtains when at the start of the competition there is genetic variability (10 randomly selected runs). Under these conditions even with the lowest rate of recombination examined, $\chi = 5x10^{-15}$, a substantial fraction of the recombining populations wins even in the face of a 2% cost in intrinsic fitness. Moreover, the time before the recombining population wins is significantly shorter than that in the runs where the non-recombining, #2, population wins. This effect of initial variability also obtains in situations where the non-recombining population is initially common. With an initially variable population, recombination provides a substantial advantage in

competition with a rare but intrinsically fitter population. This is less so when the population is initially monomorphic. But even then with a sufficiently high rate of recombination the #1 population can prevail in competition with a high fitness non-recombining population.

To obtain more information about the contribution of the intensity and form of the fitness function to the conditions under which within-host selection would favor recombination, we performed simulated competition experiments with different values of the exponent ***e***. As noted in Figure 4.5A, the intensity of selection due to these five loci and the contribution of the highest fitness 3 allele to that increase is directly proportional to ***e***. The results of these simulations are presented in Table 4.3.

Table 4.3 – Competition between a recombining (#1) and non-recombining population (#2): Effects of initial variability, recombination rate, fitness costs and the relative contributions of the higher fitness alleles (e) on the outcome of competition. Number of times each population Wins (the density of the competing population falls below 105) or is Winning (is the dominant population at the 2000th generation) in 50 independent runs. TF is the mean number of generations before the density of the losing population falls below 105 ±SE: standard error.

Runs initiated with a monomorphic 2,2,2,2,2 population **Freq of #1 /χ /s /e**	**Runs where # 1 Wins (TF±SE)**	**Runs where # 2 Wins (TF±SE)**	**Runs with #1 Winning at T=2000**	**Runs with #2 Winning at T=2000**
0.5 /5x10^{-15}/ 0.02 / 3	18* (241±22)	32 (197±12)	0	0
0.5 / 5x10^{-14}/ 0.02 / 3	23* (253±18)	27 (255±53)	0	0
0.5 /0 / 0.02 / 3	5 (583±95) $^{++}$	45 (352±28)	0	0
0.5 / 5x10^{-15}/ 0.02 / 2	12* (283±41)	38 (230 ± 18)	0	0
0.5 / 5x10^{-15} / 0.02/ 1	0	50 (397±6)	0	0
0.05 /5x10^{-15}/ 0.00/ 3	3 (1151±9) $^{++}$	47 (198 ± 14)	0	0
0.005 /5x10^{-15}/0.00/ 3	0	47 (103±7)	0	3
0.95 / 5x10^{-15}/ 0.02/ 3	41* (183 ±14)	9 (391±119)	0	0
Runs initiated with 10 random clones **Freq #1 / χ / s /e**	**Runs where # 1 Wins (TF±SE)**	**Runs where # 2 Wins (TF±SE)**	**Runs with #1 Winning at T=2000**	**Runs with #1 Winning at T=2000**

0.5 / $5x10^{-15}$/ 0.02 /3	36* (222±22)++	14 (725±61)	0	0
0.5 / $5x10^{-14}$/ 0.02 /3	42* (162±7)++	8 (761±70)	0	0
0.5 / 0 / 0.02 / 3	5 (376±48)	45 (299 ± 30)	0	0
0.5 / $5x10^{-15}$/ 0.02 /2	28* (277±19)++	22 (691±68)	0	0
0.5 / $5x10^{-15}$/ 0.02/ 1	7* (472±95)++	45 (851±53)	0	0
0.05/ $5x10^{-15}$/ 0.00 /3	11* (293±24)	30 (236±20)	4	5
0.005 / $5x10^{-15}$/0.00/ 3	0	43 (140±5)	0	7
0.95 /$5x10^{-15}$/ 0.02/ 3	46* (149±5)++	4 (596±28)	0	0

In all these simulations, μ=10-8, c=0.5, and NT=2x108 and for the experimental populations the recombination rates, χ and selection coefficient, s, are for the #1 population. The parameter e determines the magnitude of the contribution of each allele (see Figure 2). * significant with $p < 0.002$ (χ 2 contingency table comparing number of wins of the # 1 and #2 population in recombining vs. control non-recombining). ++ significant with $p < 0.005$ (Student's t-test comparing TF between Population #1 and Population # 2 for a given set of runs).

The effects of having a greater range of fitness values and a greater contribution of the highest fitness allele, 3, can be seen by comparing the simulation runs in Table 4.3 with the parallel runs in Table 4.2. Most importantly, with a greater fitness range associated with these 5 loci and proportionally greater contribution of the #3 allele, when the competing populations are of roughly equal frequency, the increase in the rate of adaptation due to recombination is more likely to overcome the fitness burden associated with recombination it would with a more modest fitness range. This is true not only for the simulations initiated with genetically variable populations but also for those initiated with monomorphic, 2,2,2,2,2 populations (compare the outcomes of the e=1 runs in Table 4.2 with the corresponding e>1 simulations in Table 4.3).

On the other hand, even in the absence of a fitness cost associated with recombination and a greater contribution of these 5 loci and the number 3 allele to fitness, this capacity for gene shuffling does not enable the recombining population to invade when its frequency is low, 0.005. Although a greater contribution of the number 3 allele to fitness and a greater fitness differential associated with these five loci augments the likelihood

of the recombining population winning over an initially rare but higher fitness, non-recombining competitor.

Competition between genetically different populations: In the preceding, the recombining and non-recombining populations were at the start of each run genetically identical, either monomorphic for the same genotype, 2,2,2,2,2 or had the same set of genetically variable clones.

Table 4.4 – Competition between a recombining (#1) and non-recombining population (#2): Effects of the recombination rate, fitness costs and the initial frequency of the recombining population on the outcome of competition. Number of times each population Wins (the density of the competing population falls below 105) or is Winning (is the dominant population at the 2000th generation) in 100 runs. TF is the mean number of generations before the density of the losing population falls below 105 ±SE. In these simulations the at the start of each run the #1 and #2 populations have genetically different populations.

Initial Freq #1 / s / χ	**Runs where # 1 Wins (TF±SE)**	**Runs where # 2 Wins (TF±SE)**	**Runs with #1 Winning at T=2000**	**Runs with #2 Winning at T=2000**
0.5 / 0.00/ 0.00	47 (102±13)	51 (141±19)	1	1
0.5 / 0.00 / $5x10^{-15}$	73* (110±8)	24 (75 ± 9)	3	0
0.5 / 0.00 / $5x10^{-14}$	74* (102±8)	24 (78 ± 13)	2	0
0.5 / 0.00 / $5x10^{-13}$	66* (100±9)	29 (71±11)	4	1
0.05/ 0.00/ 0.00	38 (101±14)	60 (109±13)	1	1
0.05 / 0.00/$5x10^{-14}$	56* (134±14)++	40 (57 ± 9)	4	2
0.5 /0.02/ 0.00	44 (103±17)	56 (130±21)	0	0
0.5 / 0.02 / $5x10^{-14}$	54 (134±4)++	40 (57±9)	0	0
0.5 /0.10/ 0.00	18 (175±39)++	82 (94±11)	0	0
0.5 / 0.10 / $5x10^{-14}$	30 (141 ±24)	70 (100±11)	0	0

In all these simulations, μ=10-8, c=0.5, e=1.0, and NT=2x108 and for the experimental populations the recombination rate parameter χ and fitness costs, s, are for the recombining #1 population. * significant differences with $p<0.002$ for a Fisher exact test for the number of Wins of population #1 and #2 relative to the corresponding runs without recombination. ++ significant, $p < 0.005$ (Student's t-test comparing the TF between Population #1 and Population #2 in a given sent of runs).

To begin to explore the more realistic situation where the competing populations with and without sexual proclivity are initially different genetically, we performed simulations with 10 initially different random collections of genotypes for #1 and #2 populations. The results of these simulations are presented in Table 4.4.

In the absence of selection against recombination and initially equal frequencies of the #1 and #2 populations, as measured by the relative numbers of winners and losers the recombining population has an advantage over the non-recombining despite the initial genetic differences between these populations. This can be seen, by comparing the simulations for the recombining "experimental" populations (rows, 2, 3 and 4) and the non-recombining control (row 1). Notably, the rate of recombination seems to have no effect on the frequency of winning. Under these conditions, the initial fitness of the competing populations plays a more prominent role in determining the outcome of competition than the increase in fitness occurring during course of competition. By this same winning and losing criteria, in the absence of an intrinsic fitness cost, recombination increases the likelihood of the #1 population ascending to dominance when it is initially relatively rare (0.05). On the other hand, with initial genetic differences in the recombining and non-recombining populations and a cost associated with capacity for recombination, HGR does not provide a statistically significant advantage for the #1 population. Moreover, with initial differences in the genetic composition of the recombining and non-recombining populations, the time to winning by the non-recombining population is less than that of the recombining.

Both populations can serve as donors. In all of the preceding runs, we assumed that recombination only occurs within a population. It may well be that both populations can contribute as donors even when they both cannot serve as recipients, e.g. when recombination is through the uptake of exogenous DNA, transformation. To explore this situation, we used a version of the simulation where the donors for recombination can be chosen from both populations, with the choice dependent solely on their relative frequencies.

In Table 4.5 we compare the outcomes of simulations where only members of the recombining population #1 serve as donors as well as recipients with corresponding situation where members of both populations can serve as donors. In the simulation results presented in this table, the recombining population has a 2% intrinsic fitness cost. In the runs where the initial frequency of the recombining populations was 0.05, the acquisition of genes from the non-recombining population increased the likelihood of invasion. Although with a 1:1 ratio there was a significantly higher frequency of recombining populations winning when both populations served as donors in the initially monomorphic runs, this was not the case for the simulations initiated with 10 randomly selected clones. On the other hand, when the initial frequency of the recombining population was 5% for both the initially monomorphic and polymorphic populations, the recombining population was more likely to win when both populations served as the source of genes for recombination. When both populations served as donors and both were initially polymorphic, winning by the recombining population took less time then it did for the winning non-recombining populations. It should be noted, however, that for any value of χ when both populations serve as donors because of the greater density of the population,the frequency of recombination was greater than when only one competitor served as the donor.

Differences in recombination rates: The recombining and non-recombining populations considered in the preceding two population simulations are the extremes. It may well be that both competing populations are capable of recombination but do so at different rates. Based on the single population and the preceding mixed population results we would anticipate that if this were the case and all else were equal, the population with the higher rate of recombination would prevail. This is indeed confirmed by our simulation experiments. For example, for 100 simulations with initially monomorphic, intermediate fitness 2,2,2,2,2 populations in equal frequency, ($\mu=10^{-7}, c=0.5, e=3$) and recombination rate parameters $\chi_1=5x10^{-13}$ and $\chi_2=5x10^{-15}$, in 95 of the runs the #1 population won, or was winning at 2000 generation in the remaining 5 runs. On the

other hand, with these starting conditions, when the initial densities of the #1 and #2 populations were respectively 10^7 vs. $1.9x10^8$, the population with the lower recombination rate won in 99 out of 100 runs and the population with the higher rate of recombination won in only one run. The situation is different when both populations can serve as donors as well as recipients. In this 10^7 vs. $1.9x10^8$ contest between the populations with high (#1) and low (#2) rates of recombination, the score for 100 runs were #1 won 45 times, #2 won 41 times, and the numbers of # 1s and #2s winning at the 2000 generation termination were, respectively 11 and 3.

Table 4.5. Number of times each population Wins (the density of the competing population falls below 105) or is Winning (is the dominant population at the 2000th generation). TF is the mean number of generations before the density of the losing population falls below 105 ±SE: standard error.

Relative frequency of #1 Mono or Polymorphic	**Donor Populations**	**#1 wins (TF±SE)**	**# 2 wins (TF±SE)**
0.05: Mono	1 only	2 (472±24)	48 (123±7)
	1 and 2	15* (312 ±21) ++	35 (122±9)
0.05 – Poly 10	1 only	0	50 (354±39)
	1 and 2	11* (232 ±30) ++	39 (603±34)
0.50 – Mono	1 only	32 (333±20)	18 (457±86)
	1 and 2	49* (222±5)	1 (227 ±10)
0.50 – Poly 10	1 only	29 (205±7) ++	21 (670±38)
	1 and 2	34 (184±100) ++	16 (726 ±34)

In all runs there was a 0.02 fitness cost for the #1 population – $\chi 1=5x10\text{-}14$, $\mu=10\text{-}7$, NT=2x108, c=0.5, e=3. In all runs both populations were initially only the 2,2,2,2,2 genotype; Poly 10 - both populations were initiated with the same10 randomly selected clones. * $p < 0.002$ ($\chi 2$ contingency with Yates Correction) . ++ significant with $p < 0.005$ (t-student test comparing TF between Population 1 and Population 2 within treatment).

4.4 DISCUSSION

We interpret the results of this computer simulation study as support for the proposition that that there are realistic conditions where homologous gene recombination (HGR) will increase the rate at which bacterial populations adapt to their environment. These results are also consistent with the hypotheses that by increasing rates of adaptive evolution, HGR can provide a population a selective advantage when competing with otherwise identical or even somewhat more fit populations that are unable to shuffle homologous genes or do so at lower rates. Our mixed population simulations, however, also illustrate a major caveat to the hypothesis that homologous gene recombination in bacteria evolved in response to selection for increasing rates of adaptive evolution. Even in the absence of a fitness cost, the recombining population will only have an advantage over a non-recombining population when the recombining population is relatively common; HGR will not be favored when it is rare.

Homologous gene recombination and rates of adaptive evolution: The validity and generality of these predictions are, of course, empirical questions. They are however, questions that can be addressed experimentally. And, as noted in our Introduction, there have been at least four experimental studies testing the hypothesis that recombination increases the rate at which bacterial populations adapt to culture conditions. The results of two of these experiments are consistent with this hypothesis, Cooper's study with F-plasmid-mediated recombination in *E. coli B* (Cooper 2007) and Baltrus and colleagues study of transformation-mediated recombination in *Helicobacter pylori* (Baltrus, Guillemin, and Phillips 2008). The results of the other two reports, Souza and colleague's study of Hfr-mediated recombination in *E. coli* (Souza, Turner, and Lenski 1997) and Bacher and colleagues study of transformation-mediated recombination in *Acinetobacter baylyi* (Bacher, Metzgar, and de Crecy-Lagard 2006) are interpreted to be inconsistent.

How well do the results of this simulation study account for the outcomes of these recombination – rates of adaptive evolution experiments? We believe that at least at a qualitative level, the results of the three of these studies for which this model is a reasonable analog (Bacher, Metzgar, and de Crecy-Lagard 2006; Cooper 2007; Baltrus, Guillemin, and Phillips 2008) are consistent with the predictions of these simulations. The format of the experiments by Souza and colleagues (Souza, Turner, and Lenski 1997) were different from that of this model and therefore we do not believe these simulations are appropriate for interpreting their results. In their experiments, two genetically different *E. coli* strains were used; a Hfr strain of *E. coli* K-12 and a F- strain of *E. coli* B. Although the Hfr strain donated genes to the *E. coli* B, under the conditions of their experiments this donor did not replicate and it was not present throughout the course of the experiment as assumed in our model.

Although the details of the (Bacher, Metzgar, and de Crecy-Lagard 2006; Cooper 2007; Baltrus, Guillemin, and Phillips 2008) experiments were different from those specified by this simple model, their basic structure was similar to that of the single population simulations initiated with monoclonal (2,2,2,2,2) populations. In these experiments, which were initiated with single clones of either recombining (Rec+ or Com+) or non-recombining (Rec- or Com-) populations, the bacteria were growing in liquid media and reached densities of $5x10^7$ per ml or greater. Although the rate constants of recombination χ were not estimated in these experimental studies, it was clear that recombination was occurring at a substantial rate. The frequency of gene replacement by recombination in the Rec+ *E. coli B* and Com+ *H. pylori* experiments exceed that expected by mutation, and in the Cooper study the rate of gene replacement by recombination greater is greater than that of the elevated rate of mutation of a *mutS* strain. For recombination mediated by HFR, F', F+ plasmid in *E. coli,* χ, it seems reasonable to conclude that in the Cooper experiments $c > 10^{-13}$ (Cornejo and Levin, In Preparation- but available, see www.eclf.net). We would also expect $\chi > 10^{-13}$ for the *H. pylori* experiments and possibly in the *Acinetobacter baylyi* study as well. This is certainly

the case for the only two experimentally obtained estimates χ we know of for transforming bacteria, *H. influenzae* (Mongold 2001) and *B. subtilis* (Johnsen, Dubnau, and Levin 2009), both of which are on the order of $\chi \sim 10^{-12}$.

With population densities, mutation and recombination rate constants in the ranges of these experiments, our simulations show that recombining populations evolved more rapidly than those that did not have this capacity for shuffling homologous genes. For any given mutation and recombination rate parameters, the rate and magnitude of increase in mean fitness depended on the fitness function. Cooper's observation that recombination increased the rate of adaptation to culture conditions with a higher mutation rate ~ 3 times greater than it did with a lower rate (Cooper 2007) is also consistent with the predictions of this model; mutation and recombination act synergistically to increase rates of adaptive evolution. Although Bacher and colleagues (Bacher, Metzgar, and de Crecy-Lagard 2006) interpret the results of their experiments with *A. baylyi* to be inconsistent with the hypothesis that HGR increases rates of adaptive evolution, that is not the case for all the results they report. In their higher density experiments not only does the fitness of the population increase to a greater extent than in their low density experiments, but this increase in fitness was considerably as well as significantly greater ($p=0.00012$ for a two tailed t-test) for the transformation competent population than the non-competent controls.

Does homologous gene recombination increase rates of adaptive evolution in natural populations? While we are unaware of direct experimental evidence for an affirmative answer to this question from natural population studies, based on the predictions of the model we would anticipate a positive answer. Retrospective, multi-locus sequence studies suggest that the rates of gene of replacements by homologous recombination in species like *Streptococcus pneumoniae* exceed that by mutation by a factor of 10 or so (Feil et al. 2000; Feil et al. 2001; Fraser, Hanage, and Spratt 2007), and are even greater for some species, like *H. pylori* (Suerbaum et al. 1998; Falush et al. 2001).

To put these retrospective estimates of recombination rates into the context of our model and its parameters, consider the following intuitive argument. Assume a 1-hour generation time, a habitat of 1 ml, a population of 10^8 bacteria and a mutation rate of 10^{-8} per cell per generation. In the course of an hour in that population, for any given locus, an average of 1 mutant would be produced. If gene replacements by recombination occur at 10 times that rate, there would be 10 recombinants at that locus for a value of $\chi = 10/(10^8 x 10^8) = 10^{-15}$. As noted in our simulations, even at this low rate and an initially monoclonal population, recombination can increase the rate of adaptive evolution over that which would be anticipated by mutation alone. Moreover, natural populations of many bacteria are likely to be composed of multiple lineages and would be genetically variable at many loci. In accord with our simulations the pace at which recombination increases the rate of adaptive evolution would on average increase with the extent of genetic variability of the population, see Figure 4.5.

Accelerating adaptive evolution as a selective force for the maintenance and evolution of HGR: Processes, like homologous gene recombination, that increase rates of adaptive evolution would be to the advantage of a population and augment its prognosis for surviving the vicissitudes of an ever-changing environment. This is, of course, the most common explanation for ubiquity of HGR among extant species of eukaryotes. Indeed, the presumed lack of recombination, sex to be more provocative, in ancient groups of seemingly successful organisms like the bdelloid rotifers make them intriguing objects for study (Mark Welch and Meselson 2000). Whether accelerating rates of adaptive evolution is the selective force responsible for the evolution and maintenance of recombination in eukaryotes is a subject of some controversy (Felsenstein and Yokoyama 1976; Otto and Barton 1997), a subject that we are pleased to say is beyond the scope of this report. The population and evolutionary dynamics of recombination in bacteria are fundamentally different from that of sexually reproducing eukaryotes. In the bacteria, recombination depends on density and is not a part of the reproductive process. If they wish to procreate, sexually reproducing eukaryotes have no choice but

to find mates and generate recombinant progeny, independently of density of their populations.

Here, we postulate that once the mechanisms for HGR are established in a bacterial population, the advantage accrued by a more rapid rate of adaptation to environmental conditions can promote their maintenance, even if they engender a modest cost in fitness. The necessary condition for this to obtain is that the adaptive process is continuous. This may be the case when a population enters a new environment and/or is confronted by either physical or biological factors that reduce the rates of survival or reproduction (the fitness) of its members. As long as the population is continually confronted with situations where selection favors new genotypes, as was postulated for evolution of mutator genes (Tanaka, Bergstrom, and Levin 2003), recombination could continue to be favored and be maintained. This would not be the case if recombination engenders a fitness cost and the population is confronted with extensive periods of adaptive stasis. Under these conditions, the frequency of the recombining population will continue to decline. And, because of the frequency- and density- dependent nature of selection for recombination, the recombining population may not be able to recover. In this interpretation, the maintaining mechanisms of horizontal gene transfer by HGR by increasing rates of adaptive evolution is not an equilibrium outcome; on "equilibrium day" (Levin and Bergstrom 2000), recombination will be lost. Moreover, because HGR accelerating rates would not provide an advantage to a recombining population when it is initially rare, it is even less likely to have been a selective force for the original evolution of mechanisms for HGT than it is for maintaining those mechanisms once they evolved.

Models can be used to generate hypotheses and, in a quantitative way, evaluate their plausibility. They cannot be used to test those hypotheses! We are unaware of published empirical studies testing the hypotheses that selection for HGR is frequency- and density- dependent. These are, however, hypotheses that can be tested with experiments similar to the single clone studies testing the hypothesis that HGR increases

rates of adaptive evolution (Bacher, Metzgar, and de Crecy-Lagard 2006; Cooper 2007; Baltrus, Guillemin, and Phillips 2008). The idea would be to follow the changes in frequency of Com+ or Rec+ in competition with Com- or Rec- clones with different initial frequencies of these competitors and in populations of different densities. We postulate that under conditions where they accelerate rates of adaptive evolution in single clone culture and adjusting for intrinsic fitness differences: (1) when introduced at roughly equal frequencies, the recombining population will have an advantage over a non-recombining competitor and, (2) the recombining population will not have that advantage when it is initially rare (in our simulations much less than 1%.). We also postulate that because of a lower rate of production of mutants as well as the lower frequency of recombination (which would be proportional to the square of the density of the recombining population); (3) the rate of adaptive evolution would be less in recombining populations of low density than otherwise identical populations of higher density and, (4) the minimum frequency for a recombining population to have a selective advantage in competition with one that cannot recombine would be inversely proportional to the total density of the recombining population.

Using the long-term evolved strains of *E. coli* B developed by Richard Lenski and colleagues (Lenski et al. 1991; Lenski and Travisano 1994; Lenski et al. 1998) it should be possible to experimentally test the hypothesis that HGR will only be favored when there is relatively intense selection for adaptation to culture conditions. Although those experimental *E. coli* B populations continued to evolve in different ways as time proceeded the largest increase in mean fitness relative to the ancestral occurred within the first 5,000 or so generations. We postulate that if in an experiment similar to that in (Cooper 2007) the F'lac constructs were made with *E. coli* B taken from later generations, say >20,000, the recombining population will not evolve more rapidly than one that is not recombining.

<u>HGR and the maintenance and evolution of HGT in bacteria:</u> Two of the three major mechanisms responsible for HGT and HGR in bacteria, conjugation and transduction,

are not properties of the bacteria but rather that of their parasites, primarily conjugative plasmids and bacteriophage. One needn't postulate that these processes evolved and are maintained by selection favoring bacteria with the capacity for HGT. The most parsimonious hypothesis for recombination mediated by plasmids and phage is as a coincidental byproduct of the infectious transfer of these elements and the host's recombination repair system (Levin 1988a; Redfield 2001). This would also be the case for recombination resulting from cell fusion (Campbell 2000) or transformation mediated by natural electroporulation or cold shocks. In this interpretation, accelerating the rate of adaptive evolution by HGR mediated by these processes are a lagniappe rather than a product of adaptive evolution. To be sure we can make up and probably construct mathematical models illustrating ways by which bacteria evolve mechanisms to be more receptive to plasmids and phage carrying genes on their behalf, but we see no need to stretch our imaginations in that direction.

The third main mechanism for HGT and HGR in bacteria, the uptake and incorporation of exogenous DNA, i.e. competence and transformation, are intrinsic properties of bacteria rather that of their parasites. We postulate that under some conditions HGR accelerating rates of adaptive evolution will promote the maintenance of competence and transformation. HGR accelerating rates of adaptive evolution is, however, only one of at least three non-exclusive mechanisms that operate synergistically to maintain competence for the uptake of exogenous DNA. The other three are; (1) the acquisition of templates for the repair of double stranded breaks in DNA (Bernstein, Hopf, and Michod 1987; Hoelzer and Michod 1991); the uptake of nutrients and nucleotides (Stewart and Carlson 1986; Redfield 1993; MacFadyen et al. 2001; Redfield et al. 2005), and (3) episodic selection favoring transiently non-growing subpopulations of competent cells and rare transformants (Johnsen, Dubnau, and Levin 2009). In accord with these three hypotheses, transformation (recombination) is a coincidental byproduct of competence.

As is the case with meiotic recombination in eukaryotes, accounting for the selective pressures responsible for original evolution of competence and transformation is more problematic than explaining their maintenance once they have evolved. Competence is a complex character that requires the coordinated activity of a large number of genes (Berka et al. 2002; Barbe et al. 2004; Dagkessamanskaia et al. 2004; Chen, Christie, and Dubnau 2005; Thomas and Nielsen 2005). What are the selective pressures responsible for the evolution of these genes and coordinating their activity? Because recombination will only be favored when it is common, we postulate that HGR accelerating rates of adaptive evolution cannot account for the original evolution of natural competence and transformation. For the same reasons, we postulate that this is also the case for the episodic selection for competence. The DNA repair and food hypotheses have the virtues of selection operating at the level of an individual bacterium rather than populations and thereby allowing competence to be favored when it is rare, rather than only when it is common. On the other hand, these two hypotheses raise other issues about whether they can account for the original evolution of competence. For a recent critical consideration of these "other issues" we refer the reader to the Discussion in (Johnsen, Dubnau, and Levin 2009). At this juncture, we accept the selection pressures responsible for the origins of competence and transformation in bacteria as a delicious, but yet-to-be solved evolutionary problem.

A caveat: In our simulations we have restricted the theater of evolution to single populations. A long-standing argument for the evolution of recombination is that higher rates of adaptive evolution provide an advantage to the collective, the group, rather than individuals (Fisher 1930; Felsenstein 1974). Populations that evolve more rapidly are more likely to prevail and survive longer than those with lower rates of adaptive evolution. In theory there are conditions where group- or interpopulation-level selection can lead to the evolution of characters that are at a disadvantage within populations (Levins 1970; Eschel 1972; Boorman and Levitt 1973; Levin and Kilmer 1975). And, mechanisms of this type have been postulated to play a role in the evolution

of recombination in bacteria (Szollosi, Derenyi, and Vellai 2006). While we prefer individual-level selection operating within populations on the grounds of parsimony, we can't rule out the possibility that competence and transformation evolved and is maintained by some form of group- level selection.

CHAPTER 5

SUMMARY AND OUTLOOK

Despite the great diversity of projects and questions presented in this dissertation there is a common goal: estimate the rate of recombination in bacteria and assess how important is its contribution to: (i) the genetic variability of bacteria and, (ii) the adaptive evolution of bacteria.

The studies carried out in this dissertation have shown that the problem of estimating rates constant of recombination in laboratory conditions is more complex than originally conceived. At times, mother serendipity takes research out of the way and it is the job of the researcher to maintain a clear goal, despite the delicious distractions that might show up during the course of an investigation.

In particular, the population dynamics of *Streptococcus pneumoniae* in continuous culture has open the door for understanding the evolution and ecology of suicidal toxic agents. These toxic agents have not been yet characterized and that might serve as potential pro-biotic (antibiotic really) compounds of interest in the clinical treatment of streptococcal infections.

The rates of recombination of *Streptococcus pneumoniae* estimated in laboratory conditions correspond roughly to those that have been estimated in retrospective studies, suggesting that recombination rates in this species is high enough to overcome

the effects of continuous bottlenecks and periodic selection that might purged the variation generated.

In free recombining bacteria like *S. pneumoniae* is important to understand processes and mechanisms that might limit the uptake and incorporation of DNA by homologous recombination. In our studies we have found evidence that fail to support the proposition that the polymorphism of the competence system (peptide and recptor) might contribute to the maintenance of genetically differentiated subpopulations. More interestingly, such polymorphism seems to be maintained by positive selection. The results of this study have opened new questions about the evolution of two component systems; questions that can be addressed again with a combination of classy mathematical modeling as well as experimental work.

Finally, but not least, we have provided evidence to the conditions under which recombination might provided an advantage in bacterial populations. These finding provides with reasonable hypotheses that can be easily explored in laboratory settings. According to our results, the mechanisms of recombination in bacteria are unlikely to have evolved in response to selective pressures for maintaining variation to accelerate rates of adaptive evolution. The reasons for this is because, recombining populations have a minor advantage when rare, while competing against non recombining populations.

Beyond the results of the studies presented in this dissertation, the results of the studies carried during my PhD have opened new roads for the research of topics that range from the molecular evolution of parasites to the evolution of recombination in bacteria.

Some of which, we expect will serve to better understand the population genetics and evolution of bacteria and parasites.

Bibliography

Atwood, K. C., L. K. Schneider, and F. J. Ryan. 1951a. Selective mechanisms in bacteria. Cold Spring Harb Symp Quant Biol **16**:345-355.

Atwood, K. C., L. K. Schneider, and F. J. Ryan. 1951b. Periodic Selection in *Escherichia coli*. Proc. Nat. Acad. Sci. USA **37**:146-155.

Bacher, J. M., D. Metzgar, and V. de Crecy-Lagard. 2006. Rapid evolution of diminished transformability in Acinetobacter baylyi. J Bacteriol **188**:8534-8542.

Bacher, J. M., W. F. Waas, D. Metzgar, V. de Crecy-Lagard, and P. Schimmel. 2007. Genetic code ambiguity confers a selective advantage on Acinetobacter baylyi. J Bacteriol **189**:6494-6496.

Baltrus, D. A., K. Guillemin, and P. C. Phillips. 2008. Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori. Evolution **62**:39-49.

Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol **16**:37-48.

Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marliere, G. N. Cohen, and C. Medigue. 2004. Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res **32**:5766-5779.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci U S A **98**:4563-4568.

Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics **152**:763-773.

Benjamini, Y., and , and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics **29**:1165-1188.

Berge, M., H. Langen, J. P. Claverys, and B. Martin. 2002. Identification of a protein that inactivates the competence-stimulating peptide of Streptococcus pneumoniae. J Bacteriol **184**:610-613.

Bergstrom, C. T., M. Lipsitch, and B. R. Levin. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics **155**:1505-1519.

Berka, R. M., J. Hahn, M. Albano, I. Draskovic, M. Persuh, X. Cui, A. Sloma, W. Widner, and D. Dubnau. 2002. Microarray analysis of the Bacillus subtilis K-state: genome-wide expression changes dependent on ComK. Mol Microbiol **43**:1331-1345.

Bernstein, H., F. A. Hopf, and R. E. Michod. 1987. The molecular basis of the evolution of sex. Adv Genet **24**:323-370.

Bogaert, D., R. H. Veenhoven, M. Sluijter, E. A. Sanders, R. de Groot, and P. W. Hermans. 2004. Colony blot assay: a useful method to detect multiple pneumococcal serotypes within clinical specimens. FEMS Immunol Med Microbiol **41**:259-264.

Boorman, S. A., and P. R. Levitt. 1973. Group selection on the boundary of a stable population. Proc Nat Acad Sci US **69**:2711-2713.

Bossart, J., and P. Prowell. 1998. Genetic estimates of genetic structure and gene flow: limitations, lessons and new directions. Trends Ecology & Evolution **13**:202-206.

Brown, J. R. 2003. Ancient horizontal gene transfer. Nat Rev Genet **4**:121-132.

Campbell, A. M. 2000. Laterial gene transfer in prokaryotes. Thoretical Population Biology **57**:71-77.

Campbell, E. A., S. Y. Choi, and H. R. Masure. 1998. A competence regulon in Streptococcus pneumoniae revealed by genomic analysis. Mol Microbiol **27**:929-939.

Cazalet, C., C. Rusniok, H. Bruggemann, N. Zidane, A. Magnier, L. Ma, M. Tichit, S. Jarraud, C. Bouchier, F. Vandenesch, F. Kunst, J. Etienne, P. Glaser, and C. Buchrieser. 2004. Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. Nat Genet **36**:1165-1173.

Chao, L., and B. R. Levin. 1981. Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc Natl Acad Sci U S A **78**:6324-6328.

Chao, L., B. R. Levin, and F. M. Stewart. 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology **58**:369-378.

Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet **2**:e64.

Chen, I., P. J. Christie, and D. Dubnau. 2005. The ins and outs of DNA transfer in bacteria. Science **310**:1456-1460.

Chen, I., and D. Dubnau. 2004. DNA uptake during bacterial transformation. Nat Rev Microbiol **2**:241-249.

Chen, I., and D. Dubnau. 2003. DNA transport during transformation. Front Biosci **8**:s544-556.

Cheng, Q., E. A. Campbell, A. M. Naughton, S. Johnson, and H. R. Masure. 1997. The com locus controls genetic transformation in Streptococcus pneumoniae. Mol Microbiol **23**:683-692.

Chookajorn, T., A. Kachroo, D. R. Ripoll, A. G. Clark, and J. B. Nasrallah. 2004. Specificity determinants and diversification of the Brassica self-incompatibility pollen ligand. Proc Natl Acad Sci U S A **101**:911-917.

Claverys, J. P., and L. S. Havarstein. 2002. Extracellular-peptide control of competence for genetic transformation in Streptococcus pneumoniae. Front Biosci **7**:d1798-1814.

Claverys, J. P., and L. S. Havarstein. 2007. Cannibalism and fratricide: mechanisms and raisons d'etre. Nat Rev Microbiol **5**:219-229.

Claverys, J. P., B. Martin, and L. S. Havarstein. 2007. Competence-induced fratricide in streptococci. Mol Microbiol **64**:1423-1433.

Claverys, J. P., M. Prudhomme, and B. Martin. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. Annu Rev Microbiol **60**:451-475.

Cohan, F. M. 2001. Bacterial species and speciation. Syst Biol **50**:513-524.

Cohan, F. M. 2002. Sexual isolation and speciation in bacteria. Genetica **116**:359-370.

Coleman, M. L., M. B. Sullivan, A. C. Martiny, C. Steglich, K. Barry, E. F. Delong, and S. W. Chisholm. 2006. Genomic islands and the ecology and evolution of Prochlorococcus. Science **311**:1768-1770.

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

Cornejo, O. E., and A. A. Escalante. 2006. The origin and age of Plasmodium vivax. Trends Parasitol **22**:558-563.

Czaran, T. L., R. F. Hoekstra, and L. Pagie. 2002. Chemical warfare between microbes promotes biodiversity. Proc Natl Acad Sci U S A **99**:786-790.

Dagkessamanskaia, A., M. Moscoso, V. Henard, S. Guiral, K. Overweg, M. Reuter, B. Martin, J. Wells, and J. P. Claverys. 2004. Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. Mol Microbiol **51**:1071-1086.

Dawid, S., A. M. Roche, and J. N. Weiser. 2007. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infect Immun **75**:443-451.

de Saizieu, A., C. Gardes, N. Flint, C. Wagner, M. Kamber, T. J. Mitchell, W. Keck, K. E. Amrein, and R. Lange. 2000. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol **182**:4696-4703.

Denker, E., E. Bapteste, H. Le Guyader, M. Manuel, and N. Rabet. 2008. Horizontal gene transfer and the evolution of cnidarian stinging cells. Curr Biol **18**:R858-859.

Didelot, X., A. Darling, and D. Falush. 2009. Inferring genomic flux in bacteria. Genome Res **19**:306-317.

Dubnau, D. 1999. DNA uptake in bacteria. Annu Rev Microbiol **53**:217-244.

Ellermeier, C. D., E. C. Hobbs, J. E. Gonzalez-Pastor, and R. Losick. 2006. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. Cell **124**:549-559.

Enright, M. C., and B. G. Spratt. 1999. Extensive variation in the ddl gene of penicillin-resistant Streptococcus pneumoniae results from a hitchhiking effect driven by the penicillin-binding protein 2b gene. Mol Biol Evol **16**:1687-1695.

Eschel, I. 1972. On the neighbor effect and the evolution of altrisitc traits. Theor Popul Biol **3**:258-277.

Evans, R. 1986. Niche expansion in bacteria: can infectious gene exchange affect the rate of evolution? Genetics **113**:775-795.

Fall, S., A. Mercier, F. Bertolla, A. Calteau, L. Gueguen, G. Perriere, T. M. Vogel, and P. Simonet. 2007. Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. PLoS One **2**:e1055.

Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman, and S. Suerbaum. 2001. Recombination and mutation during long-term gastric colonization by

Helicobacter pylori: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A **98**:15056-15061.

Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day. 2003. How clonal is Staphylococcus aureus? J Bacteriol **185**:3307-3316.

Feil, E. J., M. C. Enright, and B. G. Spratt. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between Neisseria meningitidis and Streptococcus pneumoniae. Res Microbiol **151**:465-469.

Feil, E. J., E. C. Holmes, D. E. Bessen, M. S. Chan, N. P. Day, M. C. Enright, R. Goldstein, D. W. Hood, A. Kalia, C. E. Moore, J. Zhou, and B. G. Spratt. 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc Natl Acad Sci U S A **98**:182-187.

Feil, E. J., M. C. Maiden, M. Achtman, and B. G. Spratt. 1999. The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis. Mol Biol Evol **16**:1496-1502.

Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt. 2000. Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. Genetics **154**:1439-1450.

Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. Annu Rev Microbiol **55**:561-590.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics **78**:737-756.

Felsenstein, J., and S. Yokoyama. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics **83**:845-859.

Fisher, R. 1930. The Genetical Theory of Natural Selection. Clarendon, Oxford, UK.

Fraser, C., W. P. Hanage, and B. G. Spratt. 2007. Recombination and the nature of bacterial speciation. Science **315**:476-480.

Gal-Mor, O., and B. B. Finlay. 2006. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol **8**:1707-1719.

Gerrish, P. 2008. A simple formula for obtaining markedly improved mutation rate estimates. Genetics **180**:1773-1778.

Gonzalez-Pastor, J. E., E. C. Hobbs, and R. Losick. 2003. Cannibalism by sporulating bacteria. Science **301**:510-513.

Gonzalez, E. G., P. Beerli, and R. Zardoya. 2008. Genetic structuring and migration patterns of Atlantic bigeye tuna, Thunnus obesus (Lowe, 1839). BMC Evol Biol **8**:252.

Graham, J. B., and C. A. Istock. 1978. Genetic exchange in Bacillus subtilis in soil. Mol Gen Genet **166**:287-290.

Graham, J. P., and C. A. Istock. 1979. Gene exchange and natural selection cause Bacillus subtilis to evolve in soil culture. Science **204**:637-639.

Graves, J. F., G. D. Biswas, and P. F. Sparling. 1982. Sequence-specific DNA uptake in transformation of Neisseria gonorrhoeae. J Bacteriol **152**:1071-1077.

Guiral, S., V. Henard, C. Granadel, B. Martin, and J. P. Claverys. 2006. Inhibition of competence development in Streptococcus pneumoniae by increased basal-level

expression of the ComDE two-component regulatory system. Microbiology **152**:323-331.

Guiral, S., T. J. Mitchell, B. Martin, and J. P. Claverys. 2005. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. Proc Natl Acad Sci U S A **102**:8710-8715.

Havarstein, L. S., G. Coomaraswamy, and D. A. Morrison. 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A **92**:11140-11144.

Havarstein, L. S., P. Gaustad, I. F. Nes, and D. A. Morrison. 1996. Identification of the streptococcal competence-pheromone receptor. Mol Microbiol **21**:863-869.

Havarstein, L. S., R. Hakenbeck, and P. Gaustad. 1997. Natural competence in the genus Streptococcus: evidence that streptococci can change pherotype by interspecies recombinational exchanges. J Bacteriol **179**:6589-6594.

Havarstein, L. S., B. Martin, O. Johnsborg, C. Granadel, and J. P. Claverys. 2006. New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. Mol Microbiol **59**:1297-1307.

Hayes, W. 1968. The Genetics of Bacteria and their Viruses: Studies in Basic Genetics and Molecular Biology. Blackwell Scientific, London.

Hoelzer, M. A., and R. E. Michod. 1991. DNA repair and the evolution of transformation in Bacillus subtilis. III. Sex with damaged DNA. Genetics **128**:215-223.

Hudson, R. R. 1991. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology **7**:1-44.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153-159.

Hughes, A. L., and R. Friedman. 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. Mol Biol Evol **20**:979-987.

Iannelli, F., M. R. Oggioni, and G. Pozzi. 2005. Sensor domain of histidine kinase ComD confers competence pherotype specificity in Streptoccoccus pneumoniae. FEMS Microbiol Lett **252**:321-326.

Ichihara, H., K. Kuma, and H. Toh. 2006. Positive selection in the ComC-ComD system of Streptococcal Species. J Bacteriol **188**:6429-6434.

Innan, H. 2006. Modified Hudson-Kreitman-Aguade test and two-dimensional evaluation of neutrality tests. Genetics **173**:1725-1733.

Johnsborg, O., V. Eldholm, M. L. Bjornstad, and L. S. Havarstein. 2008. A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species. Mol Microbiol **69**:245-253.

Johnsen, P. J., D. Dubnau, and B. R. Levin. 2009. Episodic Selection and the Maintenance of Competence and Natural Transformation in Bacillus subtilis. Genetics **181**:1521-1533.

Karlin, S., and J. Mrazek. 1997. Compositional differences within and between eukaryotic genomes. Proc Natl Acad Sci U S A **94**:10227-10232.

Kawabe, A., R. Fujimoto, and D. Charlesworth. 2007. High diversity due to balancing selection in the promoter region of the Medea gene in Arabidopsis lyrata. Curr Biol **17**:1885-1889.

Keeling, P. J., and J. D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet **9**:605-618.

Kerr, B., M. A. Riley, M. W. Feldman, and B. J. Bohannan. 2002. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature **418**:171-174.

Kilian, M., K. Poulsen, T. Blomqvist, L. S. Havarstein, M. Bek-Thomsen, H. Tettelin, and U. B. Sorensen. 2008. Evolution of Streptococcus pneumoniae and its close commensal relatives. PLoS One **3**:e2683.

Kim, Y., and H. A. Orr. 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. Genetics **171**:1377-1386.

Koonin, E. V., K. S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol **55**:709-742.

Kreth, J., J. Merritt, W. Shi, and F. Qi. 2005. Co-ordinated bacteriocin production and competence development: a possible mechanism for taking up DNA from neighbouring species. Mol Microbiol **57**:392-404.

Kubitzchek, H. E. 1970. Introduction to Research with Continuous Culture. Prentice Hall, New Jersey.

Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the Escherichia coli genome. Proc Natl Acad Sci U S A **95**:9413-9417.

Lawrence, J. G., and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol **44**:383-397.

Lenski, R. E., J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt. 1998. Evolution of competitive fitness in experimental populations of E. coli: what makes one genotype a better competitor than another? Antonie Van Leeuwenhoek **73**:35-47.

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

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

Leung, S. W., L. H. Apponi, O. E. Cornejo, C. M. Kitchen, S. R. Valentini, G. K. Pavlath, C. M. Dunham, and A. H. Corbett. 2009. Splice variants of the human ZC3H14 gene generate multiple isoforms of a zinc finger polyadenosine RNA binding protein. Gene **439**:71-78.

Levin, B. 1988a. The evolution of sex in bacteria in bacteria. *in* B. Levin, and R. E. Michod, eds. The Evolution of Sex: A Critical Review of Current Ideas. Sinauer, Sunderland.

Levin, B. R. 1981. Periodic selection, infectious gene exchange and the genetic structure of E. coli populations. Genetics **99**:1-23.

Levin, B. R. 1988b. Frequency-dependent selection in bacterial populations. Philos Trans R Soc Lond B Biol Sci **319**:459-472.

Levin, B. R., and C. T. Bergstrom. 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proc Natl Acad Sci U S A **97**:6981-6985.

Levin, B. R., and W. L. Kilmer. 1975. Interdemic selection and the evolution of altruism: a computer simulation study. Evolution **28**:527-545.

Levins, R. 1970. Extinction. Some mathematical questions in biology: Lectures on Mathematics in the Life Sciences, Amer. Math. Soc, Providence RI **2**:75-108.

Lewis, K. 2000. Programmed death in bacteria. Microbiol Mol Biol Rev **64**:503-514.

Lux, T., M. Nuhn, R. Hakenbeck, and P. Reichmann. 2007. Diversity of bacteriocins and activity spectrum in Streptococcus pneumoniae. J Bacteriol **189**:7741-7751.

MacFadyen, L. P., D. Chen, H. C. Vo, D. Liao, R. Sinotte, and R. J. Redfield. 2001. Competence development by Haemophilus influenzae is regulated by the availability of nucleic acid precursors. Mol Microbiol **40**:700-707.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A **95**:3140-3145.

Majewski, J., P. Zawadzki, P. Pickerill, F. M. Cohan, and C. G. Dowson. 2000. Barriers to genetic exchange between bacterial species: Streptococcus pneumoniae transformation. J Bacteriol **182**:1016-1023.

Mark Welch, D., and M. Meselson. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science **288**:1211-1215.

Maynard-Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc Natl Acad Sci U S A **90**:4384-4388.

Maynard Smith, J., and N. H. Smith. 1998. Detecting recombination from gene trees. Mol Biol Evol **15**:590-599.

Mazodier, P., and J. Davies. 1991. Gene transfer between distantly related bacteria. Annu Rev Genet **25**:147-171.

Mongold, J. 2001. Experimental approaches to studying the population dynamics and evolution of microorganisms. Experimental Ecology: Issues and Perspectives: W.J.Resetarits, Jr. and J. Bernardo Eds:298-310.

Monod, J. 1949. The growth of bacterial cultures. Annual Review of Microbiology **3**:371-394.

Moscoso, M., and J. P. Claverys. 2004. Release of DNA into the medium by competent Streptococcus pneumoniae: kinetics, mechanism and stability of the liberated DNA. Mol Microbiol **54**:783-794.

Mrazek, J., and S. Karlin. 1999. Detecting alien genes in bacterial genomes. Ann N Y Acad Sci **870**:314-329.

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

Muller, H. J. 1964. The Relation of Recombination to Mutational Advance. Mutat Res **106**:2-9.

Nath, H. B., and R. C. Griffiths. 1993. The coalescent in two colonies with symmetric migration. J Math Biol **31**:841-851.

Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter,

and C. M. Fraser. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature **399**:323-329.

Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299-304.

Otto, S. P., and N. H. Barton. 1997. The evolution of recombination: removing the limits to natural selection. Genetics **147**:879-906.

Otto, S. P., and A. C. Gerstein. 2006. Why have sex? The population genetics of sex and recombination. Biochem Soc Trans **34**:519-522.

Pace, J. K., 2nd, C. Gilbert, M. S. Clark, and C. Feschotte. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A **105**:17023-17028.

Pestova, E. V., L. S. Havarstein, and D. A. Morrison. 1996. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol **21**:853-862.

Peterson, S. N., C. K. Sung, R. Cline, B. V. Desai, E. C. Snesrud, P. Luo, J. Walling, H. Li, M. Mintz, G. Tsegaye, P. C. Burr, Y. Do, S. Ahn, J. Gilbert, R. D. Fleischmann, and D. A. Morrison. 2004. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Mol Microbiol **51**:1051-1070.

Pinas, G. E., P. R. Cortes, A. G. Orio, and J. Echenique. 2008. Acidic stress induces autolysis by a CSP-independent ComE pathway in Streptococcus pneumoniae. Microbiology **154**:1300-1308.

Pozzi, G., L. Masala, F. Iannelli, R. Manganelli, L. S. Havarstein, L. Piccoli, D. Simon, and D. A. Morrison. 1996. Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone. J Bacteriol **178**:6087-6090.

Redfield, R. J. 2001. Do bacteria have sex? Nat Rev Genet **2**:634-639.

Redfield, R. J. 1993. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. J Hered **84**:400-404.

Redfield, R. J., A. D. Cameron, Q. Qian, J. Hinds, T. R. Ali, J. S. Kroll, and P. R. Langford. 2005. A novel CRP-dependent regulon controls expression of competence genes in Haemophilus influenzae. J Mol Biol **347**:735-747.

Regev-Yochay, G., K. Trzcinski, C. M. Thompson, M. Lipsitch, and R. Malley. 2007. SpxB Is a Suicide Gene of Streptococcus pneumoniae and Confers a Selective Advantage in an In Vivo Competitive Colonization Model. J Bacteriol **189**:6532-6539.

Regev-Yochay, G., K. Trzcinski, C. M. Thompson, R. Malley, and M. Lipsitch. 2006. Interference between Streptococcus pneumoniae and Staphylococcus aureus: In vitro hydrogen peroxide-mediated killing by Streptococcus pneumoniae. J Bacteriol **188**:4996-5001.

Riley, M. A., and D. M. Gordon. 1999. The ecological role of bacteriocins in bacterial competition. Trends Microbiol **7**:129-133.

Riley, M. A., and J. E. Wertz. 2002. Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol **56**:117-137.

Ronda, C., J. L. Garcia, E. Garcia, J. M. Sanchez-Puelles, and R. Lopez. 1987. Biological role of the pneumococcal amidase. Cloning of the lytA gene in Streptococcus pneumoniae. Eur J Biochem **164**:621-624.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Sanchez-Puelles, J. M., C. Ronda, J. L. Garcia, P. Garcia, R. Lopez, and E. Garcia. 1986. Searching for autolysin functions. Characterization of a pneumococcal mutant deleted in the lytA gene. Eur J Biochem **158**:289-293.

Schubert, S., P. Darlu, O. Clermont, A. Wieser, G. Magistro, C. Hoffmann, K. Weinert, O. Tenaillon, I. Matic, and E. Denamur. 2009. Role of intraspecies recombination in the spread of pathogenicity islands within the Escherichia coli species. PLoS Pathog **5**:e1000257.

Selander, R. K., and B. R. Levin. 1980. Genetic diversity and structure in Escherichia coli populations. Science **210**:545-547.

Shea, J. E., M. Hensel, C. Gleeson, and D. W. Holden. 1996. Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. Proc Natl Acad Sci U S A **93**:2593-2597.

Sisco, K. L., and H. O. Smith. 1979. Sequence-specific DNA uptake in Haemophilus transformation. Proc Natl Acad Sci U S A **76**:972-976.

Smith, G. R. 1988. Homologous recombination in procaryotes. Microbiol Rev **52**:1-28.

Smith, H. O., D. B. Danner, and R. A. Deich. 1981. Genetic transformation. Annu Rev Biochem **50**:41-68.

Smith, J. M., E. J. Feil, and N. H. Smith. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays **22**:1115-1122.

Souza, V., P. E. Turner, and R. E. Lenski. 1997. Long-term experimental evolution in Escherichia coli. V. Effects of recombination with immigrant genotypes on the rate of bacterial evolution. Journal of Evolutionary Biology **10**:743-769.

Spratt, B. G. 2004. Exploring the concept of clonality in bacteria. Methods Mol Biol **266**:323-352.

Spratt, B. G., W. P. Hanage, and E. J. Feil. 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr Opin Microbiol **4**:602-606.

Spratt, B. G., and M. C. Maiden. 1999. Bacterial population genetics, evolution and epidemiology. Philos Trans R Soc Lond B Biol Sci **354**:701-710.

Steinmoen, H., E. Knutsen, and L. S. Havarstein. 2002. Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. Proc Natl Acad Sci U S A **99**:7681-7686.

Steinmoen, H., A. Teigen, and L. S. Havarstein. 2003. Competence-induced cells of Streptococcus pneumoniae lyse competence-deficient cells of the same strain during cocultivation. J Bacteriol **185**:7176-7183.

Stewart, F. M., and B. R. Levin. 1973. Resource partitioning and the outcome of interspecific competition: a model and some general considerations. American Naturalist **107**:171-198.

Stewart, G. J., and C. A. Carlson. 1986. The biology of natural transformation. Annu Rev Microbiol **40**:211-235.

Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within Helicobacter pylori. Proc Natl Acad Sci U S A **95**:12619-12624.

Szollosi, G. J., I. Derenyi, and T. Vellai. 2006. The maintenance of sex in bacteria is ensured by its potential to reload genes. Genetics **174**:2173-2180.

Tanaka, M. M., C. T. Bergstrom, and B. R. Levin. 2003. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. Genetics **164**:843-854.

Thomas, C. M., and K. M. Nielsen. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol **3**:711-721.

Tomasz, A. 1965. Control of the competent state in Pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. Nature **208**:155-159.

Tortosa, P., and D. Dubnau. 1999. Competence for transformation: a matter of taste. Curr Opin Microbiol **2**:588-592.

Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M. E. Karoui, E. Frapy, L. Garry, J. M. Ghigo, A. M. Gilles, J. Johnson, C. Le Bouguenec, M. Lescat, S. Mangenot, V. Martinez-Jehanne, I. Matic, X. Nassif, S. Oztas, M. A. Petit, C. Pichon, Z. Rouy, C. S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Medigue, E. P. Rocha, and E. Denamur. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet **5**:e1000344.

Whatmore, A. M., V. A. Barcus, and C. G. Dowson. 1999. Genetic diversity of the streptococcal competence (com) gene locus. J Bacteriol **181**:3144-3154.

Wiener, P. 1996. Experimental studies on the ecological role of antibiotic production in bacteria. Evolutionary Ecology **10**:405-421.

Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. Maiden, H. Ochman, and M. Achtman. 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol **60**:1136-1151.

Yoshida, T., S. P. Ellner, L. E. Jones, B. J. Bohannan, R. E. Lenski, and N. G. Hairston, Jr. 2007. Cryptic population dynamics: rapid evolution masks trophic interactions. PLoS Biology **5**:e235.

Yu, V. L., C. C. Chiou, C. Feldman, A. Ortqvist, J. Rello, A. J. Morris, L. M. Baddour, C. M. Luna, D. R. Snydman, M. Ip, W. C. Ko, M. B. Chedid, A. Andremont, and K. P. Klugman. 2003. An international prospective study of pneumococcal bacteremia: correlation with in vitro resistance, antibiotics administered, and clinical outcome. Clin Infect Dis **37**:230-237.

Zawadzki, P., M. S. Roberts, and F. M. Cohan. 1995. The log-linear relationship between sexual isolation and sequence divergence in Bacillus transformation is robust. Genetics **140**:917-932.

APPENDIX

Appendix 1: Assisted Suicide in continuous culture populations of Pneumococcus

July 31, 2009

Contents

A1 Mathematical Model. Analysis

The simple version of the model is described by the following set of equations

$$\begin{aligned} \dot{R} &= \omega(C-R) - \Psi(R)eB \quad (1) \\ \dot{B} &= \Psi(R)B - \omega B - xBT \quad (2) \\ \dot{T} &= yBT - \omega T - dT \quad (3) \end{aligned}$$

where, $\Psi(R) = R/(K+R)$

These equations can be rewritten in dimensionless form by rescaling the resource (R), bacterium (B), and toxin (T) variables as:

$$R' = R/C, \quad B' = xB/f\omega, \quad T' = yT/\omega \quad (4)$$

This gives:

$$\begin{aligned} \dot{R'} &= 1 - R' - aB'\psi(R') \quad (5) \\ \dot{B'} &= B'(b\psi(R') - 1 - T) \quad (6) \\ \dot{T'} &= fT'(B' - 1) \quad (7) \end{aligned}$$

with $\psi(R')$ defined as $\psi(R') = R'/(R' + \kappa)$

Here we have also rescaled the time variable as $t' = \omega t$. The dimensionless combinations of parameters which determine the system's dynamical behavior are:

$$f = (\omega + d)/\omega, \quad \kappa = K/C, \quad a = ef\nu/yC, \quad b = \nu/\omega. \quad (8)$$

The possible stationary points (equilibrium points*) of this system, R^*, B^*, and T^*, are found by putting $\dot{R} = \dot{B} = \dot{T} = 0$, and solving the ensuing algebraic equations. There are three such stationary points.

From this point we dropped the superscript prime (′) on each variable. $B' \to B$, $R' \to R$, $T' \to T$.

The first –"resource only"– such point has:

$$B^* = T^* = 0, \quad \text{and} \quad R^* = 1 \tag{9}$$

A linearized stability analysis shows this point is stable if, and only if,

$$\kappa/(b-1) > 1 \tag{10}$$

Failing this, an arbitrary small introduction of bacterium will grow ($\dot{B} > 0$).

The second –"plant-hervibore"– point has

$$R^* = \kappa/(b-1), \quad B^* = (1-R^*)b/a, \quad T^* = 0 \tag{11}$$

This equilibrium is stable if, and only if,

$$(a/b) + \kappa/(b-1) > 1 > \kappa/(b-1) \tag{12}$$

Within this domain, any introduction of T decays to zero, and the system returns to the R^*, B^* equilibrium with damped oscillations if $\kappa > (b-1)^2/(2b-1)$, or via exponential damping otherwise.

Third, and the most interesting, is a "3 species" (a "plant-hervibore-predator") equilibrium solution with

$$B^* = 1, \quad R^* = -D + (D^2 + \kappa^2)^{1/2}, \quad T^* = (b/a)(1-R^*) - 1. \tag{13}$$

Here, D is defined for notational convenience as $D = (a + \kappa - 1)/2$. This equilibrium solution is stable with all the three components persisting together, if and only if

$$1 > (a/b) + \kappa/(b-1) \tag{14}$$

For the experimental system, representative values of the parameters are $\omega = 0.1\ ml\ hour^{-1}$, $x = 5 \times 10^{-6}\ ml\ cells^{-1}\ a.u.t.^{-1}$, $e = 10^{-7}\ \mu g$, $y = 4 \times 10^{-10}\ ml\ cells^{-1}\ a.u.t.^{-1}$, $K = 0.25\ \mu g\ ml^{-1}$, $\nu = 1\ cells\ hour^{-1}$, $d = 0.1\ hour^{-1}$ (the $a.u.t$ are defined as **"arbitrary units of toxin"**). The parameter characterizing the **resource/chemostat concentration**, C, take various values, but always $C >> K$. Given that C is the control variable and κ represents it, the definition of a can be changed to $a = \alpha\kappa$ with α defined below. For the dimensionless parameters of Eq.(8) we thus have $f = 2$, $b = 10$, $a = \alpha\kappa$ (with $\alpha = 2000$), and $\kappa = 0.25/C$, with C having values from a few tens to a few thousands. That is, $\kappa << 1$, whereas a can be less or greater than 1, and the defintion of α is as follows:

$$\alpha = \frac{ef\nu}{yK}, \tag{15}$$

and it is important to remember that according to our definitions $K = \kappa C$.

The dynamics of this 3-species system (represented in the diagram 1), as κ varies (corresponding to the chemostat resource concentration C varying), are interesting, with a rather sharp transition in dynamical behavior at the point where $a = 1 - \kappa \approx 1$. For the illustrative parameters above, this corresponds to $C = 500$. For a smaller than this ($a < 1$), and remembering $\kappa \ll 1$, for all C-values of interest, we have $D = (a + \kappa - 1)/2 < 0$, resulting in $R^* \approx 1 - a$, $B^* = 1$, $T \approx (b - 1)$. In this region the system approaches its stable equilibrium point via very weakly damped oscillations. The oscillations have a period, P, approximatelly given by

$$P \approx 2\pi[f(b-1)]^{-1/2}/\omega \tag{16}$$

whilst the damping time, τ, is very long of the order

$$\tau \sim 2[1 + f(b-1)][1-a]^2/[ab\omega\kappa] \tag{17}$$

Here, these expressions are given in absolute time (not rescaled), hence the factor ω.

The dynamical behavior changes abruptly when κ —- although still much less than 1 —- approaches the value such that $a = \alpha\kappa$ (with $\alpha \gg 1$) is essentially one: $a = 1+$ (a few multiples of)κ. In this tiny window, D $\sim K$ and $R^* \approx \kappa^{1/2}$, while still $B^* = 1$ and $T^* = b - 1$. The system still approaches equilibrium via damped oscillations, but now with the period (P) and the damping time (τ) are of comparable magnitude: P is much as above, but now $\tau \sim [4 + f(b-1)]/b\omega$.

Finally, for $a > 1$ (by a fe multiples of κ), we move into a regime where $R^* \approx \kappa/(a-1) << 1$, $B^* = 1$, $T^* = (b-a)/a$.

Perturbations to this equilibrium point decay back to it; the characteristic damping time is $\tau \sim 1/[(\beta - x)\omega]$, with the definitions $\beta = b/2a$ and $x = (\beta^2 - 2f\beta + f)^{1/2}$. As a increases, β and thence x decrease, until we have $x = 0$ when $\beta = f - (f^2 - f)^{1/2}$. Beyond this point, the system returns to its equilibrium point via damped oscillations, whose period ($T \sim 1/ \mid x \mid \omega$) and damping time ($\tau \sim 1/\beta$) have comparable magnitudes. Eventually, once a becomes sufficiently large that $a > b[1 - \kappa(b-1)]$, the 3-species equilibrium is no longer stable, and the toxin T can no longer be maintained in the system.

Figure S2 illustrates this range of behavior, and in particular the very abrupt transition around the point $a = 1$ (which for the illustrative parameter values outlined above, corresponds to $C = 500$): for C-values above this, there are very weakly damped oscillations; below it, there is exponential damping if $341 < C < 500$, and damped oscillations for $50 < C < 341$. Once $a > b[1 - \kappa/(b-1)]$, corresponding here to $C < 50$, the toxin can no longer be maintained, and we have a stable "resource - bacterium" system R^*-B^*. For implausibly small C-values ($C < 1/36$), the only stable state is the "resource-only" system, with neither bacterium nor toxin. Again, and at the risk of being repetitive, the composite parameteres a, b, and f allows us to examine the transition between the different dynamic behaviours that the system might experience as the result of changes in the resource concentration C.

It should be emphasized that the stability analysis presented above is, strictly speaking, valid only for small perturbations (i.e. a linearized analysis). But – although we have not found a Lyapunov function for the above system – previous experience suggests the global dynamical properties march with the local ones. Extensive numerical simulations support this suggestion.

Fig. S2

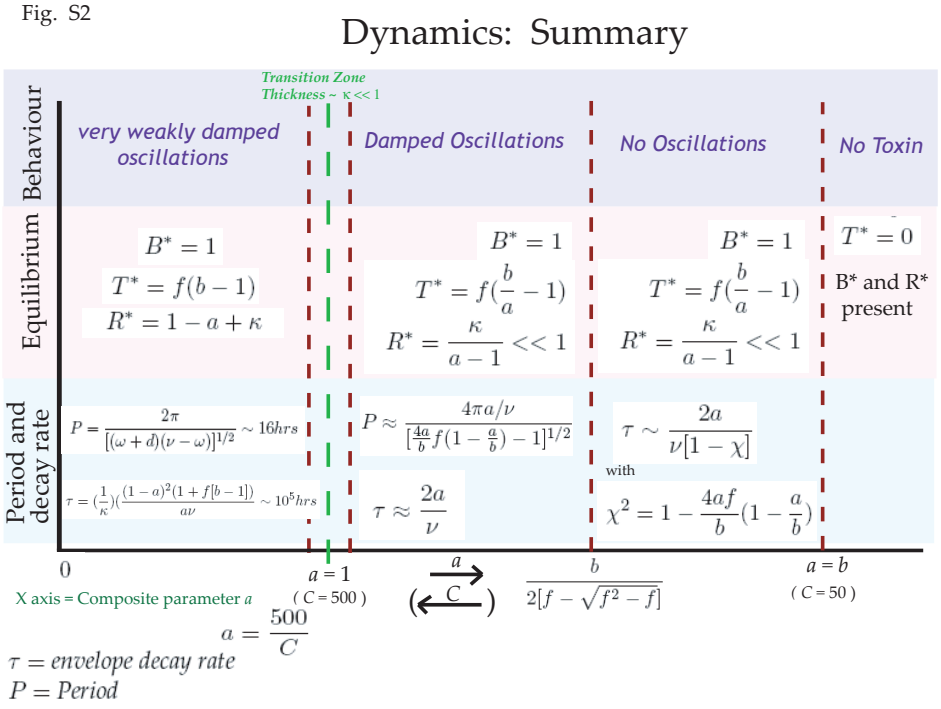

Figure A1: Diagram of the Dynamics predicted by the model. The space of possible dynamics is shown as a function of values the composite paramter a. It is shown in different boxes (from the top to the bottom) the Predicted Dynamics, the predicted equilibrium state for each range of a, The characteristic oscillation period (P) and envelope decay rate (τ) of the oscillations for the differen behaviours

A2 Strain deficient in the production of Hydrogen peroxide oscillates

Pneumococcus inhibits the replication or lyses *Staphylococcus aureus* by the release of H_2O_2 (Regev-Yochay et al. 2007). This raises the possibility that hydrogen peroxide could be the toxin responsible for the observed oscillations of *S. pneumoniae* R6 chemostat culture. The results of experiments with catalase reported in our article are inconsistent with this interpretation. Also inconsistent are the experiments we performed with a strain deficient in the production of pyruvate oxidase, an exzyme required for the production of hydrogen peroxide. In chemostat culture this strain oscillates, Figure S2.

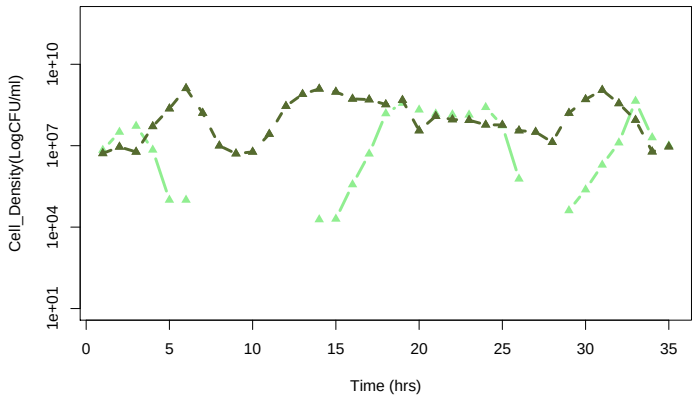

Figure A2: Changes in density (CFU) of *S. pneumoniae* strain (TIGR4 background) deficient for the production of pyruvate oxidase (Δ-*spxB*) an enzyme required for the production of H_2O_2 (dark green lines and triangles) and the *spxB*+ control (light green lines and spots).

A3 Strains deficient in the induction of competence and bacteriocin like protein (blp) systems oscillate

The competence peptide of *S. pneumoniae*, which appears to be produced in an autocatalytic manner similar to toxin postulated here, is a reasonable candidate for the agent (toxin or regulatory factor involved in the production of the toxin) driving the oscillations. To test this hypothesis, we inoculated chemostats with mutants defective in the production of the competence peptide (Δ-*ComC*) or in the receptor of the competence peptide (Δ-*ComD*). These strains oscillate (Fig S3).

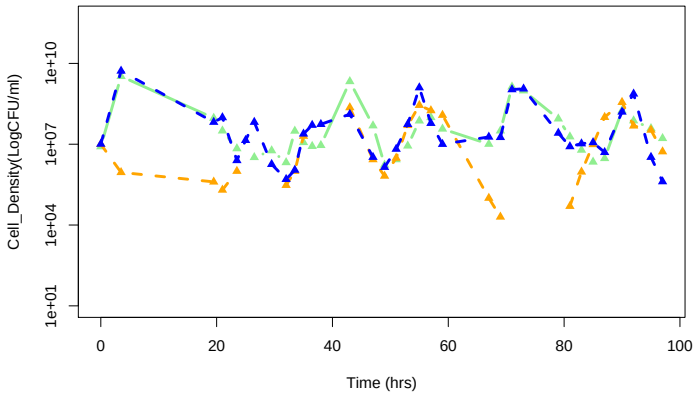

Figure A3: Changes in density in chemostat culture of *S. pneumoniae* wild type R6 (dark blue) and R6 deficient in the production of the competence peptide Δ-*ComC* (light green), and its receptor Δ-*ComD*) in light orange.

A similar behaviour is found in strains in which the regulatory gene of the *blp* system, also known to operate autocatalitically, has been knocked out. That is, it shows oscillatory dynamics similar to the parental strain (Fig S4)

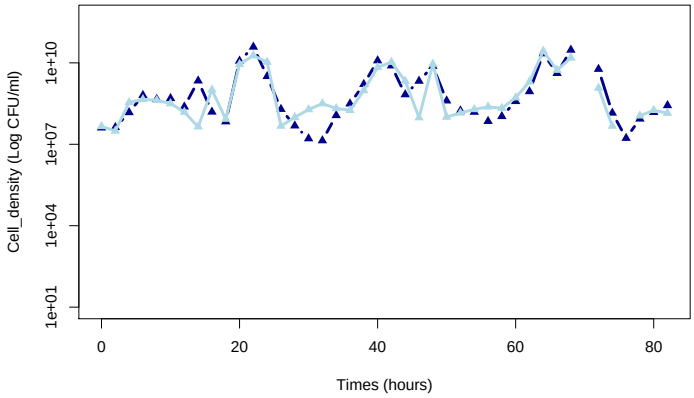

Figure A4: Chemostat dynamics of the *S. pneumoniae* 6A strain (serotype 23) deficient in the production of the regulatory peptide of the *blp* regulon (Δ-*blpR* Light Blue). The wild type is shown in Dark Blue.

A4 Strains deficient in the production of murein hydrolase *CbpD* also oscillates

As estated in the previous section the competence peptide of *S. pneumoniae*, which appears to be produced in an autocatalytic manner, does not oscillate. Additional evidence is found in a strain that lacks the production o the general response regulator *ComE* of the competence system (see Figure S5). Moreover, the lysing protein murein hydrolase, produced by *CbpD*, and seemingly regulated by the competence system, does not seem to prevent the strains from oscillating (see Figure S5), and we consider it can be ruled out as a potential oscillation driving agent in this system.

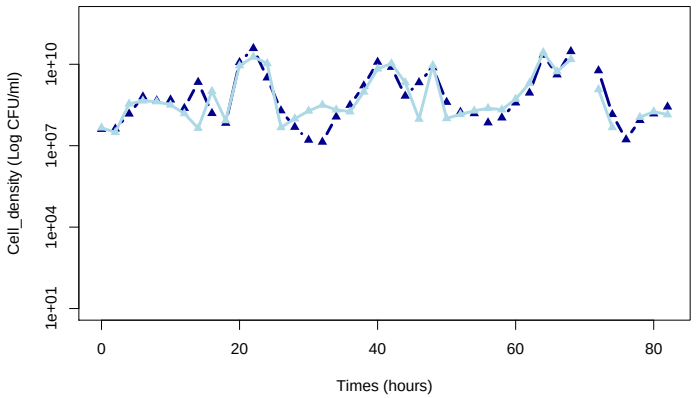

Figure A5: Changes in density in chemostat culture: *S. pneumoniae* RH1 wild type (blue). RH17 a derivative of RH1 deficient in the production of the murein hydrolase protein, Δ-*CbpD*, (dark green), RH3 a derivative of RH1 deficient in the production of the cognate response regulator of the competence system, Δ-*ComE*, (green) (Johnsborg et al. 2008)

A5 Assisted suicide protects established populations for invasion of higher fitness clones or species

In our report, we postulate that despite the killing of members of the same clone, assisted suicide could have evolved and be maintained as a mechanism to prevent established populations from invasion by bacteria of different clones or species with greater intrinsic fitness. The necessary condition for this is the invading clone is sufficiently sensitive to the toxin at its concentration in an established population of *Streptococcus pneumoniae* to have a net disadvantage relative to members of the established clone.

A5.1 Model and Analysis for conditions of invasibility of a second species with higher fitness

To begin to explore these invasion prevention conditions quantitative way, we expand our model (equations 1-3) to include a second population of bacteria, B_2 and add a subscript 1 to the variable designating the toxin-producing species, B_1. We assume both species are limited by the same resource which they consume in a Monod-like fashion with for simplicity the same values of k and e. These two populations can have different maximum growth rates, V_1 and V_2, and different sensitivities to the toxin. With these definitions and assumptions, the rates of change in the densities of the different populations are given by,

$$\dot{R} = (C - R)\omega - e\Psi(R)(B_1V_1 + B_2V_2) \quad (18)$$

$$\dot{B_1} = \Psi(R)V_1B_1 - \omega B_1 - x_1B_1T \quad (19)$$

$$\dot{B_2} = \Psi(R)V_2B_2 - \omega B_2 - x_2B_2T \quad (20)$$

$$\dot{T} = yB_1T - \omega T - dT \quad (21)$$

where, $\Psi(R) = R/(K + R)$ as in the previous model. V_1 and V_2 are the maximum growth rates of these bacteria, and x_1 and x_2 are the killing rates of the producing and invading bacteria respectively. For convenience we assume that the efficiency of conversion of resource into bacteria biomass, e, and the Monod constant, k, are identical for B_1 and B_2. As in the previous model y is the rate constant of production of the toxin and d_T is the inverse of the half-life of the toxin.

For analytical tractability in the above model we are assuming that B_2 does not produce the toxin. This assumption can be justified by our concern with the conditions for invasion of established populations of the toxin-producing strain B_1 by B_2 when it is rare. Even if the per-capita rate of production of the B_2 toxin was great and the established population highly sensitive to it, as long as the density of the invading population is low, the concentration of the toxin produced by it would be too low to cause a perceptible effect on the much more numerous established population.

Suppose B_2 tries to invade the $R - B_1 - T$ system. That is, for entry/growth of a tiny amount of B_2, we have:

$$\dot{B}_2 = B_2(\psi(R^*)\nu_2 - \omega - z_2 T^*) \tag{22}$$

where, R^*, T^* are given by the (equilibrium) solutions of the $R - B_1 - T$ system being invaded.
Clearly, invasion is possible if, and only if,

$$\nu_2 \psi^* > \omega + z_2 T^* \tag{23}$$

But ψ^* is given by $\nu_1 \psi^* = \omega + z_1 T$, so Eq.(23) becomes

$$\frac{\nu_2}{\nu_1}(\omega + z_1 T) > \omega + z_2 T \tag{24}$$

i.e

$$\left(\frac{\nu_2 - \nu_1}{\nu_1}\right) > \left(z_2 - \frac{\nu_2 z_1}{\nu_1}\right)\frac{T^*}{\omega} \tag{25}$$

with T^* from our earlier analysis of $R - B_1 - T$ system.

Clearly, if $\nu_2 \approx \nu_1$ and $z_2 > z_1$, this Eq.(25) can not be satisfied, and <u>cannot</u> invade. But, if ν_2 is sufficiently larger than ν_1, then the right hand side (RHS) of Eq.(25) gets bigger <u>AND</u> $\left(z_2 - \frac{\nu_2}{\nu_1} z_1\right)$ gets smaller — can even be negative if $\frac{\nu_2}{\nu 1} > \frac{z_2}{z_1}$ — and B_2 <u>CAN</u> invade.

To be more explicit, we need to put in the value of T^* (which, remember, is related to our earlier, RESCALED, T^* by $T^* = \frac{\omega}{g} T'^*$) for the $R - B_1 - T$ system with "$a > 1$" (smaller C values), we have the system stable at $T'^* \approx \frac{b}{a} - 1$, whence $T' \approx \frac{\left(\frac{\nu_1}{a} - \omega\right)}{g}$.

So, the condition for B_2 to be <u>UNABLE</u> to invade is (for $a > 1$)

$$z_2 > \frac{g\omega\left(\nu_2 - \nu_1\right)}{\nu_1\left(\frac{\nu_1}{a} - \omega\right)} + \frac{\nu_2}{\nu 1} z_1 \tag{26}$$

For $R - B_1 - T$ with "$a < 1$ let us first look at the corresponding criterion for B_2 unable to invade the stable equilibrium point $R^* - B_1^* - T^*$, and second consider the case when, in practice, we have cycles of large amplitude, weakly damped.

First, for "$a < 1$", $T'^{*} \approx b - 1 \equiv \frac{\nu_1 - \omega}{\omega}$
This gives the UNABLE to invade criterion as:

$$z_2 > \frac{g\omega\left(\nu_2 - \nu_1\right)}{\nu_1\left(\nu_1 - \omega\right)} + \frac{\nu_2}{\nu_1} z_1 \tag{27}$$

Notice that Eq.(27) has the larger denominator, $(\nu_1 - \omega)$, than Eq.(26) when $a > 1$, where the corresponding factor is $\left(\frac{\nu_1}{a} - \omega\right)$, so it is easier to keep B_2 out when $a < 1$.

It is also important to notice also that the long time in which the system is cycling will create conditions in which T will be often (roughly half the time) less than T^{*}. Even, so, it is likely that Eq.(26) remains a good approximation to the "**e**xclusion criterion" even when we are in the very-weakly-damped cycle phase of the system.

To make things slightly more clear, if we put explicit parameter values into equations Eq.(26) or Eq.(27):

$$\nu_2 \equiv \delta\nu_1 \quad with \quad \delta > 1, \quad and \quad z2 \equiv \theta z_1, \quad with \quad \theta > 1 \tag{28}$$

And $\nu = 1; \omega = 0.1$, $z_1 = 5 \times 10^{-6}$; $g = 4 \times 10^{-10}$, gives

for Eq.(25): $a > 1$ (small C) $\Rightarrow B_2$ cannot invade iff
$\theta > \psi + \left(\frac{4}{5} \times 10^{-5}\right)\left(\frac{\theta - 1}{\frac{1}{a} - 0.1}\right)$

or ie, $\theta > \psi$ to approximate neglection of 10^{-5}!

For Eq.(26): $a < 1$ (big $C \leftrightarrow$ most of the simulations)
(although complicated by oscillations)
cannot invade *iff* $\theta > \psi + \left(\frac{8}{9} \times 10^{-5}\right)(\psi - 1)$
idem, we can assume the term 10^{-5} as neglectable and the condition reduce to $\theta > \psi$

Even when more analysis can be done with this system our interest is to show that it is possible to find a condition under which a second species can not invade an established population that produces a self-killing toxin. The existance of such condition opens the gates to additional interesting questions on the way such system works and the toxin is

maintained.

A5.2 Simulation results

For the reader who prefers a less abstract presentation of these results, we present the following numerical solutions for Eq.(18- 21). In these simulations the invading clone has a growth rate advantage relative to the established toxin producing clone $V_2 > V_1$ but is more sensitive to the toxin, $z_2 > z_1$. In the absence of toxin production, the population with this higher intrinsic fitness (greater growth rate) invades the established population and replaces it (Fig. S7a). With toxin production by B_1 and a sufficient population density to generate toxin-mediated oscillations, invasion is prevented (Fig. S7b). As seen in Figure S7c, the invasion of the higher intrinsic fitness clone can also be when the toxin is produced but the density of the toxin producing population is too low to generate oscillations. However, if the density of the toxin producing population is too low, the concentration of toxin in the environment can be insufficient to overcome the intrinsic advantage of the invading clone (Fig S7d). The values of the parameters for these simulations are presented in the legend to figure S6.

As can be seen in the simulations presented in Figure 10a, in the absence of the toxin the higher fitness clone, B_2, will invade and replace the established population B_1. If, however, the average density of B_1 is high and $x_2/x_1 > V_2/V_1$ the invasion of B_2 can be prevented. This obtains when the density of B_1 oscillates (Fig. 10b) and when the resource concentration is too low for this established population to oscillate (Fig 10c). However, if the density of B_1 and thus the concentration of toxin in the environment are too low, the clone with the greater growth rate will invade (Fig 10d), albeit at a rates lower than in the absence of toxin production (Fig 10a). It should be noted, however, that the conditions for B_2 to invade an established population of B_1 are to some extent dependent on the initial density of B_1 and thus the concentration of the toxin, T in the environment. If the density of B_1 is initially low, B_2 can invade under conditions where it would not invade a toxin producing B_1 population with an initially higher density (simulation data not shown).

A6 Toxin sensitivity of other species of *Streptococcus*

As noted in our report, the naturally occurring clinical isolates of *S. pneumoniae* we tested oscillate when put into chemostat culture. Based on the spot assay on lawns, other species of *Streptococcus* are also susceptible to this toxin. To ascertain the relationship between

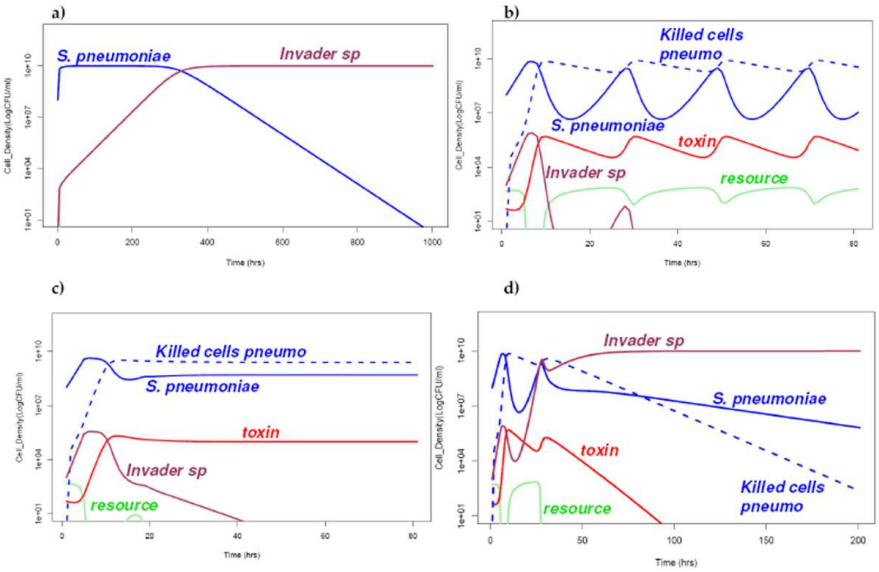

Figure A6: Simulation results: invasion of a clone with higher intrinsic fitness: Changes in the densities of the established (blue) and invading (maroon) bacterial populations, and concentrations of the toxin (red) and resource (green). The parameter values for this simulation are: $V_1 = 1.0$, $V_2 = 1.5$, $k = 0.25$, $e = 1x10^{-7}$, $y = 4x10^{-10}$, $\omega = 0.1$, $x^1 = 5x10^{-6}$, $x^2 = 9x10^{-6}$, and $d = 0.1$. In these figures a single cell of B^2 is introduced into populations of B^1 at initial density of 108. In (b), (c) and (d) the initial concentration of the toxin is $50 units per ml$. (a) Invasion in the absence of toxin production ($y = 0$), (b) Failure to invade a high density ($C = 1000$) oscillating toxin-producing population. (c) Failure to invade a lower density ($C = 300$) non-oscillating toxin-producing population; (d) Invasion in the absence of maintenance of toxin production because of very low resource concentration ($C = 50$). In these simulations, when the density of the invading B^2 ¡ 0.5, the simulation B^2 was set equal to zero.

this sensitivity and the genetic relationship of these assayed species we used a maximum likelihood reconstruction protocol for the phylogenetic reconstruction and the parsimony reconstruction of the sensitivity to the toxin overimposed to it.

The relationships inferred show that *S. mitis* and *S. oralis* are the closest related species to *S. pneumoniae* within the group of Streptococci species considered for the assays (see Fig. S8). The colors of the branches correspond to the activity of the supernatant on each one of the terminal species as well as the Maximum parsimony reconstruction of the activity in the deeper nodes (performed in Mesquite, Madison and Maddison 2007). In red it is shown the branches on which activity was detected or inferred, and in black the branches were no activivity was detected or inferred.

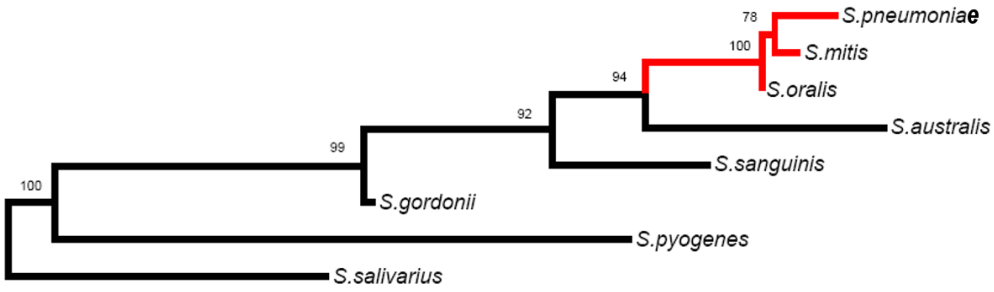

Figure A7: Maximum Likelihood (ML) phylogenetic tree of the Streptococcal species tested for sensitivity to the R6 toxin. Bootstrap support of the nodes for trees based upon 1000 pseudo-replicates, are presented on the top of the nodes. The species in red text are sensitive to the *S. pneumoniae* supernatant. Branches in light red are parsimonious reconstructions of expected activity of the toxin. Distances are measured in units of numbers of substitutions per site.

Appendix 2: Polymorphism in the competence peptide and the limits to recombination in *Streptococcus pneumoniae*

July 31, 2009

A1 Genetic diversity per geographic region

The following table shows the average genetic diversity across MLST categorized by geographical region.

Table **A1: Genetic diversity per gene and overall of the clinical isolates, categorized by geographic origin**

	n[a]	**S**[b]	**h**[c]	**Hd**[d] **(s.d)**[e]	π^f **(s.d)**[e]
By Region					
South America	11	64	9	0.95(0.07)	0.008(0.001)
North America	12	54	11	0.97(0.04)	0.008(0.001)
Europe	23	79	19	0.84(0.02)	0.009(0.001)
Africa	27	118	26	0.99(0.01)	0.011 (0.001)
New Zealand	9	56	7	0.92(0.09)	0.007(0.001)
New Zealand	6	47	5	0.93(0.12)	0.009(0.001)

[a] **n** is the sample size
[b] **S** is the number of segregating sites
[c] **h** is the number of haplotypes
[d] **Hd** is the haplotypic diversity of the sample
[e] **s.d** is the standard deviation of the estimates
[f] π is the pairwise genetic diversity estimated with a Jukes and Cantor correction

A2 Polymorphism and Divergence in reference loci and *comC*

The following Table shows the polymorphism and divergence within *Streptococcus pneumoniae* and between *S. pneumoniae* and *S. mitis* for the test locus *comC* on which we presume balancing selection might be maintaining polymorphism, and the reference loci used to perform Hudson-Kreitman-Aguade tests.

Table **A2: Genetic Polymorphism (average pairwise differences) within** *S. pneumoniae***, and Divergence (Average pairwise differences) between** *S. pneumoniae* **and** *S. mitis*

Locus	sequence lentgh	Polymorphism	Divergence
ComC	123	5.37	23.28
aroE	405	2.21	28.87
gdh	460	4.93	22.00
gki	483	7.23	26.37
recP	450	2.46	6.75
spi	474	4.91	24.40
xpt	486	4.38	27.45

The following Table shows the information of polymorphism and "divergence" within *S. pneumoniae* of CSP1 and CSP2 subpopulations and between subpopulations for the test locus *comC* on which we presume balancing selection might be maintaining polymorphism, and the reference loci used to perform Hudson-Kreitman-Aguade tests.

Table **A3: Genetic Polymorphism (average pairwise differences, Poly) within** *S. pneumoniae* **CSP1 and CSP2 subpopulations, and Divergence (Average pairwise differences) between CSP subpopulations**

Locus	sequence lentgh	Poly within CSP1	Poly within CSP2	Divergence
ComC	123	0.94	0.08	10.55
aroE	405	1.90	1.70	1.96
gdh	460	5.32	3.58	4.92
gki	483	7.96	5.23	7.08
recP	450	2.42	2.34	2.44
spi	474	4.01	4.64	4.41
xpt	486	4.48	3.79	4.39