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.

_____ _____

Nathan T. Jacobs Date

Competition and cooperation of respiratory pathogens in spatially structured environments

By

Nathan T. Jacobs
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution

Anice C. Lowen
Advisor

John Altman
Committee Member

Rustom Antia
Committee Member

Jacobus de Roode
Committee Member

Levi Morran
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Competition and cooperation of respiratory pathogens in spatially structured environments

By

Nathan T. Jacobs

B.S. Biology, The Pennsylvania State University, 2013

B.S. Immunology & Infectious Disease, The Pennsylvania State University, 2013

Advisor: Anice C. Lowen, Ph.D.

An abstract of a dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science

Population Biology, Ecology and Evolution

2019

Abstract

The focus of this dissertation is the interactions that occur between respiratory pathogens during infection, specifically between members of the same species. The human respiratory tract is home to many microorganisms, both commensal and pathogenic. Two common pathogens are *Streptococcus pneumoniae* and influenza A viruses. *S. pneuomoniae* forms biofilms in the upper respiratory tract during colonization, but causes severe disease when it disseminates to the middle ear, lungs, or bloodstream. Disease severity is correlated with capsular serotype, and many colonized individuals carry multiple serotypes, making the dynamics of competition during nasopharyngeal colonization important for predicting an individual's risk of disease. Co-colonization experiments were conducted, and it was determined that pneumococcal serotypes compete in a contact-dependent manner, but not through previously described mechanisms of fratricide in bacterial biofilms. The genome of influenza A virus comprises eight distinct RNA segments, with all segments being necessary for the production of progeny virions. At low concentrations, many cells infected by a single virus particle contain only a subset of the genome, but complementation through cellular co-infection can allow successful production of progeny. The frequency of incomplete genomes was measured in a model influenza A virus strain and the results were used to parameterize computational models to estimate the fitness costs of genome segment loss. Experimental investigations then revealed that the spatial structure inherent in replication and virus spread provide sufficient complementation, mitigating many of the costs of incomplete genomes. These results highlight the importance of spatial structure and intraspecific interactions in the dynamics of respiratory infections.

Competition and cooperation of respiratory pathogens in spatially structured environments

By

Nathan T. Jacobs

B.S. Biology, The Pennsylvania State University, 2013

B.S. Immunology & Infectious Disease, The Pennsylvania State University, 2013

Advisor: Anice C. Lowen, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science

Population Biology, Ecology and Evolution

2019

Acknowledgements

Completing this dissertation was no easy task, and while it required a great amount of effort on my part, I could not have done it without the guidance and support of several others. Primarily, I owe much gratitude to my advisor Anice Lowen. Joining a new lab as a 4th-year student was stressful, but choosing that lab could not have been easier. Having walked by the lab's posters every day and mulled over the interactions between genome segments, virions, and cells after seeing presentations by students, I had already been interested in influenza virus evolution, and when Anice and I talked more about the phenomenon of incomplete viral genomes, we quickly had ideas for early experiments and models. Having had no virology experience at the time I joined the lab, I'm amazed at how much I've learned in just two years, and can attribute much of it to Anice's guidance and continued support. I'm thankful to have been able to work in such a positive environment where I was also able to bring some different expertise in the form of programming and population biology methods.

In addition to the outstanding virology training and experience I received in Anice's lab, I have also been fortunate to work with Rustom Antia to incorporate mathematical modeling into my influenza virus research. Even before I joined Anice's lab, though, his advice was instrumental in applying for NIH and NSF fellowships, and making sure my ideas were as clear and organized as possible. More importantly, though, he knew that I would learn the technical skills and make the experiments work at some point, and helped me learn how to stay focused on the big picture and to write and think like a scientist.

The other members of my thesis committee were also essential in helping me reach this point. Jaap de Roode was incredibly helpful during the process of transitioning to Anice's lab, and made sure that I was poised to succeed there. Two NIH fellowship applications that I submitted as a student were well-received, and a substantial portion of their positive reviews can be attributed to John Altman's direction in the IMP grant-writing course, and his advice on larger issues during committee meetings was invaluable. Levi Morran joined later, but provided keen insight on gene flow and recombination as they related to my influenza work.

Several other faculty members were not on my committee, but were mentors in one way or another during my time at Emory. Jorge Vidal provided great advice during my extended rotation in his lab, and even after I had moved on, continued to provide helpful advice related to my other research project on the pneumococcus. John Steel, when his lab was at Emory, regularly attended meetings with Anice's lab and asked very insightful questions and provided outstanding feedback to help drive the direction of my influenza research. Finally, I cannot thank Nael McCarty enough for doing everything in his power as GDBBS director to mitigate the stress associated with changing labs, developing the Molecules to Mankind program in which I participated for two years, and more generally making the Division feel like an inclusive, supportive environment for students. As culture flows from the top, and it was apparent from every interaction that Nael had students' best interests at heart, the GDBBS is better for his leadership.

While Anice's guidance has been instrumental in the overall design and execution of this work, I've spent far more time with other students in the lab, and their contribution to the positivity of the environment cannot be overlooked. Specifically, I have to thank

Gabrielle Delima, Baptiste Elie, Meg Hockman, Kate Holmes, Kara Phipps, and Maria White for their company during late-night experiments, constant supply of memes, and carrying our trivia team ($E = mc^{Hammer}$) to victory week after week.

I also owe many thanks to my fellow students and friends from the IMP, MMG, and PBEE programs, namely Kelsie Brooks, Mary Bushman, Alice Cho, Ashley Cross, Paul Dunbar, Sean McMaster, Mariko Peterson, Sam Prezioso, Jess Trost, and Alex Wein, for their shared love of concerts, board games, cats, musicals, traveling, good food, good beer, and PBR. I'm lucky to have such great friends that understand how challenging science can be at times, but still make time to celebrate the good and help forget the struggles. I couldn't have done it without them.

Finally, I need to thank my family for their unwavering support, namely my parents Tom and Deb, my sister Amy, and my aunt Megan, for reminding me to make time for things other than science, keeping me sane from Pennsylvania and visiting Atlanta when possible, and sending me pictures of my little cousin Cassidy. Their support has been invaluable over the past 6 years and I couldn't have done it without them, either.

Table of Contents

Chapter 1. Introduction

1.1 Overview

The human respiratory tract is home to many microorganisms, both commensal and pathogenic. Among the most common of these pathogens are *Streptococcus pneumoniae* and influenza A viruses, which in the United States alone cause a combined 1.1 million hospitalizations and 48,000 deaths annually[1]. A large proportion of these severe cases are attributable to the fact that influenza A virus infection predisposes an individual to secondary bacterial pneumonia (reviewed in [2]), highlighting the importance of species interactions to the pathology of disease.

Streptococcus pneumoniae, or the pneumococcus, is a Gram-positive bacterium that typically colonizes the nasopharynx, but causes disease when it enters the middle ear (acute otitis media), lungs (pneumonia), and bloodstream (invasive pneumococcal disease). Over 90 serotypes have been described, with children frequently being colonized multiple times over the first 5 years of life. Simultaneous co-colonization with multiple strains is somewhat common, with up to 50% of colonized individuals carrying two serotypes in some studies [3, 4].

Influenza A viruses (IAV) are negative-sense RNA viruses of the family *Orthomyxoviridae*. Waterfowl serve as the primary reservoir of the virus, but certain lineages circulate in mammals such as swine and humans. The IAV genome comprises eight distinct RNA segments, which allows cells infected with multiple virus particles to produce chimeric progeny containing segments from different genetic backgrounds.

Competition

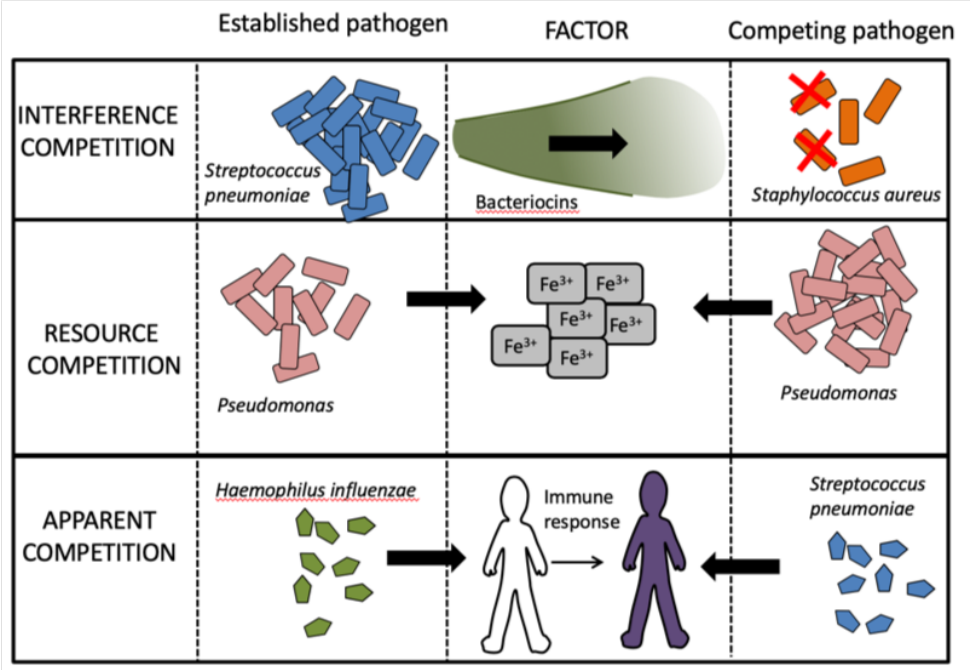


Figure 1 — Mechanisms of within-host competition. A) Interference competition occurs between locally co-occurring parasites. B) Resource competition occurs between parasites that exploit hosts in similar ways. C) Apparent competition is mediated by cross-reactive immune responses.

An individual infection is a population process because pathogens multiply within the host environment. Within-host competition occurs through three primary mechanisms: first, one organism may directly antagonize competitors through the production of antimicrobial compounds (interference competition). Secondly, organisms with similar requirements for growth may compete indirectly for shared resources (resource competition). And finally, immune responses to one pathogen may cross-react with other pathogens, acting in a manner similar to a shared predator (apparent competition).

Interference Competition

The most direct way for two pathogens to compete is through interference competition, in which one competitor secretes compounds that harm other pathogens. The classic example of this phenomenon is the production of bacteriocins, compounds secreted by bacteria that harm others nearby but to which they themselves are resistant. Such compounds allow *Escherischia coli* and *S. pneumoniae*, for example, to compete with other bacteria in the gut or upper respiratory tract [5-7]. Furthermore, they can be secreted in response to stress from inflammation caused by other bacteria [8] or the presence of specific competitors [9, 10], allowing an established population of bacteria to exclude or at least avoid displacement by an incoming competitor. As the effect of bacteriocins is concentration-dependent to the point that some require physical contact to operate [11, 12], such adaptations are expected to arise only between parasites that share a physical niche [13, 14]. Their benefit is primarily in helping a population maintain high within-host densities, but they also play a demonstrated role in facilitating transmission to new hosts. *Streptococcus mutans* strains with more potent and broader bacteriocin activity are favored in mother-to-child transmission [15], and the addition of bacteriocin-producing *E. coli* and other species successfully reduces shedding of pathogenic *E. coli* and *Campylobacter* by food animals [16-18]. Similarly, in avian malaria parasites, interference competition allows *Plasmodium juxtanucleare* to inhibit fertilization of *Plasmodium gallinaceum* in the mosquito, facilitating its own transmission and mitigating the costs of sharing a vector [19].

Resource Competition

As within-host population size increases, growth becomes limited by the resource which is present in the least amount. The size of a parasite's niche is determined by this limiting resource. For viruses and other intracellular parasites, the limiting resource may be target cells. For bacteria that colonize epithelial cells in the respiratory tract or gut, surface area may instead be limiting. The worst potential competition an organism can face is from members of its own species, which have the same basic requirements. When two populations compete for a shared resource, the one that makes more efficient use of the resource, suppressing the resource to a lower level, will generally be the superior competitor.

Apparent Competition

Even when pathogens share no common resources, both will elicit an immune response that limits their replication. Responses generated against one pathogen have the potential to cross-react with others, thereby killing co-infecting pathogens in a process known as apparent competition. As in the case of resource competition, the organism that better tolerates an immune response will fare better in a competitive interaction. The presence of *Haemophilus influenzae* in the nasopharynx, for example, induces the recruitment of neutrophils, which indiscriminately kill nearby bacteria. While *H. influenzae* is partially resistant to this neutrophil activity, *Streptococcus pneumoniae* is more sensitive, and thus *H. influenzae* adversely impacts the population size of *S. pneumoniae* in a neutrophil-dependent manner [20].

Evolutionary outcomes

Theory predicts that within-host competition for resources or as a result of cross-reactive immune responses selects for greater virulence [21]. As both competitors share

the cost of host death equally, but only the superior competitor benefits from increased exploitation, this asymmetry of costs and benefits favors continued scalation of virulence. *Plasmodium chabaudi* parasites that replicate faster and cause more severe anemia, for example, outcompete less virulent strains during mixed infections [22-25]. Experimental evolution of this species has substantiated this observation, showing that serial passage through mice results in parasites that are more virulent than the parental strain [26, 27]. Immunologically, mouse and epidemiological studies also reveal selection favoring more virulent *S. pneumoniae* serotypes through apparent competition with *Haemophilus influenzae* [20, 28-30]. Faster exploitation has the consequence of causing more pathology, making a parasite strain more virulent than is optimal in the absence of co-infection. However, shortening the infectious period also allows it to avoid spending too long in any one host. By accelerating exploitation in this manner, parasites trade optimal lifetime transmission success for the short-term gain of reaching new hosts faster.

Cooperation

Faster exploitation and tolerance to immunity allows parasites to avoid some of the deleterious effects of co-infection, but the increased virulence associated with such strategies is still costly, lowering lifetime transmission potential. Early theoretical work predicts that optimal virulence increases with the frequency of co-infection, as the need for competition becomes more common [31, 32]. But, if the frequency of co-infection increases because parasites are co-transmitted, creating repeated interactions between them [33], then parasites are predicted to cooperate instead, lowering the optimal virulence [34]. This would be especially useful to chronically infecting parasites, since the period of potential transmission is limited only by the natural lifespan of the host. Kin

selection predicts that such cooperative behaviors arise during chronic infection because parasites are closely related, and this prediction has been supported by experimental evolution [35]. Longitudinal studies of within-host adaptation show that chronically infecting bacteria such as *Pseudomonas aeruginosa* [36], *Burkholderia pseudomallei* [37], and *S. aureus* [38] become less virulent over time through loss of virulence factors among some fraction of the population. Because the genes responsible for synthesis are lost while the receptors remain intact [39, 40], social evolutionary theory characterizes these mutants as "cheaters", which have been shown to reduce overall virulence [41]. Beyond exploitation of public goods, extracellular resources that are available for use by all individuals in the host, chronically infecting parasites can reduce virulence through niche differentiation. Rather than every member of a large population competing for a single resource by escalating virulence, divergent subpopulations that utilize different, readily available resources, such that escalating virulence to compete for a single resource becomes a less viable strategy than specialzing on a different one. *P. aeruginosa* develops metabolic specialization during extended infections [42-44], reducing competition and allowing co-existence of less virulent subpopulations.

Co-infection often changes the resource or immunological landscapes for another parasite, but in the absence of competition for a shared resource, adaptation through redundancy and cooperation can be more beneficial than increased virulence. *H. influenzae*, for example, contains adherence genes that are dispensable for colonization of the murine lung in isolation, but essential for survival in a mouse recently infected with influenza A virus. Conversely, genes for oxidative stress resistance are required when colonizing alone, but not during secondary infection due to the impaired ability of

neutrophils to kill bacteria during influenza virus infection [45]. *S. pneumoniae*, by contrast, is capable of plasticity that allows it to express the different adherence factors needed to bind to pharyngeal cells depending on whether they have been infected with parainfluenza virus [46]. Together these studies suggest that parasites have not only evolved genes to contend with co-infection, but also the capacity to sense the presence of a second infection and modulate expression accordingly.

Spatial Structure

The competitive and cooperative interactions described previously all require that pathogens be co-localized within the same host, but the interactions themselves can be mediated by systemic processes. For example, hookworms reside in the gut, but feed on red blood cells, indirectly limiting replication of *Plasmodium* parasites by depriving them of a crucial resource. Other interactions, however, require much closer spatial organization. The effects of bacteriocins and "public good" compounds like siderophores, for example, are concentration-dependent {Galvez, 1998 #481}, and so bacteria must colonize the same organ or tissue space to interact through these mechanisms. Even when a host is infected with a single bacterium or virion, new individuals produced by replication will increase the local population size. This leads to the formation of biofilms by bacteria, or foci of infection in viruses. Given the tendency of microorganisms to behave differently in these spatially structured environments than in suspension culture {Gilley, 2014 #421;Chao, 2015 #404}, a thorough understanding of the role of spatial organization in collective pathogen interactions is important for predicting infection outcomes *in vivo*.

The role of spatial structure and intraspecific interactions in the population dynamics of these *Streptococcus pneumoniae* and influenza A virus are the subject of this dissertation. Chapter 2 examines competition between pneumococcal populations in mixed-strain biofilms both *in vitro* and in a simulated nasopharynx. Chapter 3 is devoted to the phenomenon of incomplete influenza A virus genomes, which make viral replication dependent on cellular co-infection and thus enforce cooperation between virus particles. Chapter 4 summarizes the work presented herein, highlights its broader implications, and suggests potential avenues of further inquiry. A brief introduction to Chapter 2 and 3 follow.

1.2 Competition in the pneumococcus

The primary niche of the pneumococcus is the nasopharynx, where bacteria form a biofilm comprised of cells, proteins, lipids, polysaccharides, and extracellular DNA. This manner of colonization is quiescent, with bacteria exhibiting reduced expression of virulence factors such as the polysaccharide capsule. While the behavior of biofilms containing one bacterial strain have been well characterized [47, 48], the behavior of multiple strains in the same biofilm is less understood. Up to 50% of colonized individuals carry multiple pneumococcal serotypes, however, making competition in mixed-strain biofilms an important element of the dynamics of infection [3, 4, 49].

When a bacterial community forms a biofilm, many cells are killed during colonization, either through autolysis or fratricide. These dead individuals release DNA into the extracellular matrix, which can be taken up by living bacteria through transformation. Exchange of DNA in this manner facilitates rapid transfer of genes in biofilm communities, such as those encoding antibiotic resistance.

The work described in Chapter 2 sought to characterize the competitive interactions between pneumococcal serotypes in mixed-strain biofilms. By competing three vaccine-targeted serotypes, S19F, S6B, and S23F, against prototypical strain TIGR4, we observed that S19F densities were markedly reduced in competition. This competitive effect was not mediated by differences in growth rates, quorum sensing, or production of stress-inducing competence-stimulating peptides.

1.3 Cooperation in influenza A virus

The influenza A virus genome comprises eight distinct RNA segments, with productive infection requiring the presence of at least one copy of each segment. At limiting dilutions, when cells are infected by single particles, cells often express a subset of viral genes [50].

Visualization of genome segments within virus particles by electron microscopy [51] and fluorescence in situ hybridization [52] show that particles often contain one copy of each segment. This suggests that segments often fail to be replicated after entering the cell, and that the phenomenon of incomplete viral genomes is mediated by inefficiency in the process of infection.

The segmented nature of the IAV genome allows co-infecting incomplete viral genomes (IVGs) to complement each other, so that even if no single particle delivers all eight genome segments, a cell will still produce progeny if it replicates all eight genome segments. When most particles in a population deliver IVGs, this creates a dependence on cellular co-infection for productive infection, causing the majority of progeny to be produced from co-infected cells. Free mixing of genome segments within co-infected

cells leads to the majority of progeny being reassortant, wherein virus particles contain genome segments from more than one parent.

The work described in Chapter 3 sought to explore the implications of IVGs for influenza A virus infectivity and replication in the context of a single host. We measured the frequency of IVGs in influenza A/Panama/2007/1999 H3N2 virus, then used the results to parameterize a probabilistic model of infectivity and an individual-based model of virus replication. These models predicted that the abundance of IVGs in virus populations create a need for complementation to establish infection, but that complementation occurs efficiently during subsequent rounds of replication due to the spatial structure inherent in viral spread within the respiratory tract. To test this hypothesis, we first manipulated spatial structure in cell culture infections and found that the need for complementation was reduced in spatially structured infections. To determine the contribution of incomplete genomes to infection *in vivo*, we next generated a virus that was entirely dependent on co-infection for productive infection, and observed that while its infectivity was markedly reduced, its ability to replicate in infected guinea pigs was only modestly affected.

Chapter 2. Competitive Dominance within Biofilm Consortia Regulates the Relative Distribution of Pneumococcal Nasopharyngeal Density

This chapter was previously published in *Applied Environmental Microbiology* [53].

2.1 Abstract

Streptococcus pneumoniae is a main cause of child mortality worldwide, but strains also asymptomatically colonize the upper airways of most children and form biofilms. Recent studies have demonstrated that ~50% of colonized children carry at least two different serotypes (i.e., strains) in the nasopharynx; however, studies of how strains coexist are limited. In this work, we investigated the physiological, genetic, and ecological requirements for the relative distribution of densities, and spatial localization, of pneumococcal strains within biofilm consortia. Biofilm consortia were prepared with vaccine type strains (i.e., serotype 6B (S6B), S19F, or S23F) and strain TIGR4 (S4). Experiments first revealed that the relative densities of S6B and S23F were similar in biofilm consortia. The density of S19F strains, however, was reduced to ~10% in biofilm consortia, including either S6B, S23F, or TIGR4, in comparison to S19F monostrain biofilms. Reduction of S19F density within biofilm consortia was also observed in a simulated nasopharyngeal environment. Reduction of relative density was not related to growth rates, since the Malthusian parameter demonstrated similar rates of change of density for most strains. To investigate whether quorum sensing (QS) regulates relative densities in biofilm consortia, two different mutants were prepared: a TIGR4Δ*luxS* mutant and a TIGR4Δ*comC* mutant. The density of S19F strains, however, was similarly

reduced when consortia included TIGR4, TIGR4Δ*luxS*, or TIGR4Δ*comC*. Moreover, production of a different competence- stimulating peptide (CSP), CSP1 or CSP2, was not a factor that affected dominance. Finally, a mathematical model, confocal experiments, and experiments using Trans-well devices demonstrated physical contact-mediated control of pneumococcal density within biofilm consortia.

Streptococcus pneumoniae kills nearly half a million children every year, but it also produces nasopharyngeal biofilm consortia in a proportion of asymptomatic children, and these biofilms often contain two strains (i.e., serotypes). In our study, we investigated how strains coexist within pneumococcal consortia produced by vaccine serotypes S4, S6B, S19F, and S23F. Whereas S6B and S23F shared the biofilm consortium, our studies demonstrated reduction of the relative density of S19F strains, to ~10% of what it would otherwise be if alone, in consortial biofilms formed with S4, S6B, or S23F. This dominance was not related to increased fitness when competing for nutrients, nor was it regulated by quorum- sensing LuxS/AI-2 or Com systems. It was demonstrated, however, to be enhanced by physical contact rather than by a product(s) secreted into the supernatant, as would naturally occur in the semidry nasopharyngeal environment. Competitive interactions within pneumococcal biofilm consortia regulate nasopharyngeal density, a risk factor for pneumococcal disease.

2.2 Introduction

Streptococcus pneumoniae, commonly known as pneumococcus, is a Gram-positive opportunistic pathogen that is the leading cause of bacterial pneumonia and acute otitis media [54, 55]. Despite its propensity for causing severe diseases, pneumococcus is a common commensal that quiescently colonizes the upper respiratory tract, forming

biofilms that adhere to the epithelium of the nasopharynx, ear epithelium, and lungs, rather than as planktonic cells that are associated with septicemia and meningitis[56-60]. In the nasopharyngeal environment, replication slows and bacteria show reduced expression of virulence factors, such as the polysaccharide capsule; instead, they form a biofilm structure made of extracellular DNA, proteins, lipids, and polysaccharides that facilitates asymptomatic carriage. This ability to colonize the upper respiratory tract and persist via biofilms, without causing disease, makes carriage of pneumococcal strains common [57, 59, 60].

To date, over 90 pneumococcal serotypes have been described, and while wide-spread immunization with conjugate pneumococcal vaccines (PCVs) targeting 7, 10, or 13 serotypes has been effective at reducing mortality, carriage of targeted serotypes still occurs [54, 61]. With the widespread use of new methodologies for pneumococcal serotyping, carriage of multiple S. pneumoniae strains (i.e., serotypes) was demonstrated to be as common as carriage of a single strain[3, 4, 49, 62]. For example, a high prevalence of multiple pneumococcal strain carriage was reported by Turner et al., who used both a sweep-latex agglutination method and microarray studies to demonstrate that 43% or 48.8%, respectively, of nasopharyngeal (NP) swabs from Thai children carried more than one pneumococcal serotype [4]. A similar prevalence of multiple serotype carriage (~40%) was observed recently in Spain; serotypes were detected based on a combination of Quellung reactions, latex serotyping, and multiplex PCRs, and in another study serotypes carried by Peruvian children were identified by using serotype-specific quantitative PCRs (qPCRs)[3, 49]. Therefore, simultaneous carriage of multiple serotypes is relatively common, meaning that these genetically distinct pneumococcal strains must

compete not only with resident microflora and other opportunistic pathogens, such as *Haemophilus influenzae* and *Staphylococcus aureus*, but also with members of their own species[3, 49, 62, 63].

Monostrain pneumococcal biofilms and biofilms made by pneumococcus and other species, such as *H. influenzae* or *Moraxella catarrhalis*, have been extensively studied during the last few years [56, 57, 63-66]. During early stages of formation of monostrain biofilms (i.e., within 8 h), the pneumococcus utilizes the quorum-sensing (QS) LuxS/AI-2 and Com systems as a proxy of population density to begin forming the biofilm structure [67-69]. Supporting the role of QS in the control of monostrain pneumococcal biofilms, Carrolo et al. [67] demonstrated that strains expressing competence- stimulating peptide 1 (CSP1), a quorum-sensing pheromone, produced denser biofilms than strains producing CSP2. When pneumococcal monostrain biofilms are formed in an enclosed system (i.e., in a polystyrene plate), quorum-sensing-regulated fratricidal factors, including some bacteriocins, accumulate in the microenvironment and cause irreversible death of biofilm cells that begins after 12 h of incubation [68, 69]. Biofilm lysis did not occur in a biofilm bioreactor with cultures of human pharyngeal cells [68] or in a plate biofilm model with immobilized pharyngeal cells where the culture medium was changed every 4 h (20, 21). With appropriate modifications, these biofilm models can be utilized to investigate chronic colonization or pneumococcal disease involving the upper organs (e.g., ear) or lower airways (e.g., lungs).

Despite compelling evidence of how monostrain pneumococcal biofilms are formed and how the biofilm structure is regulated, little is known about the coexistence and relative densities of pneumococcal strains within nasopharyngeal biofilm consortia.

In this work, we investigated the physiological, genetic, and ecological requirements for the relative distribution densities and spatial localizations of different pneumococcal serotypes during the formation of biofilm consortia. Mixtures of invasive strains recently isolated from pneumococcal disease cases and belonging to vaccine serotypes 6B (S6B), S19F, and S23F were utilized to produce biofilm consortia. The relative densities of individual strains forming biofilm consortia were investigated by serotype-specific qPCR and CFU counts. These studies demonstrated higher relative densities (i.e., dominance) of strains S6B, S23F, and TIGR4 in forming biofilm consortia with S19F strains. The relative densities were similar (i.e., tolerant) when S6B and S23F formed pneumococcal biofilms. The spatial arrangement of pneumococcal strains within biofilm consortia was investigated by using confocal microscopy imaging. Our data revealed that the QS systems LuxS/AI-2 and Com play no role in controlling dominance within a biofilm consortium. Moreover, a secreted molecule appeared not to be involved in the dominance of the relative densities, while spatial physical contact of pneumococcal strains within biofilm consortia was required for dominance.

2.3 Methods

Strains and bacterial culture media.

S. pneumoniae strains utilized in this study are listed in Table 1. Strains were routinely cultured on blood agar plates (BAP) or grown in Todd-Hewitt broth containing 0.5% (wt/vol) yeast extract (THY). When indicated, ampicillin (1 μg/ml), streptomycin (100 μg/ml), or erythromycin (0.5 μg/ml) was added to the culture medium. Antibiotics were purchased from Sigma.

Preparation of TIGR4 derivative strains.

SPJV16 and SPJV20, with a deletion within the *luxS* or *comC*gene, respectively, were prepared essentially as described for a previous study in our laboratory [68]. Mutation was confirmed by PCR [i.e., different PCR product sizes compared to wt products, due to deletion within the target gene and insertion of the *erm*(B) gene] and by sequencing (data not shown). SPJV21 was prepared by transforming SPJV16 with DNA from R6Ami9, which confers resistance to streptomycin, and plated onto BAP with the antibiotic. Transformation was done by standard methods [70] (Table 1).

Cell cultures.

Human pharyngeal Detroit 562 cells (ATCC CCL-138) were cultured in Eagle's minimum essential medium (EMEM; Lonza, Walkersville, MD) supplemented with non-heat-inactivated 10% fetal bovine serum (FBS; Atlanta Biologicals), 1% nonessential amino acids (Sigma), 1% glutamine (Sigma), penicillin (100 U/ml), and streptomycin (100 μg/ml), and the pH was buffered with HEPES (10 mM; Gibco). Cells were harvested with 0.25% trypsin (Gibco), resuspended in the cell culture medium, and incubated at 37°C in a 5% CO_2 humidified atmosphere.

Preparation of inocula.

S. pneumoniae strains were streaked on BAP and incubated overnight at 37°C in a 5% CO_2 atmosphere. Bacterial suspensions (OD_{600} of 0.05) were made in THY and further grown until they reached an OD_{600} of ~0.2; a 10% (vol/vol) final solution of glycerol was added to this culture, which was then stored at −80°C until used. Some aliquots were

removed from the freezer and then diluted and plated to obtain the CFU per milliliter data for inoculants.

Production of early biofilm consortia on abiotic surfaces.

Pneumococcal early biofilm consortia were generated by thawing inocula on ice, and then ~7×10^5 CFU/ml of each strain was inoculated into a CellBIND surface 24-well polystyrene plate (Corning) containing THY. Biofilm consortia and biofilm controls were incubated for 8 h in 5% CO_2, and after extensive phosphate-buffered saline (PBS) washes, biofilms were harvested by sonication for 15 s in a Bransonic ultrasonic water bath (Branson, Danbury, CT), followed by extensive pipetting to remove all attached bacteria. Biofilms were either counted by dilution and plating or frozen at −80°C for DNA extraction.

Production of early biofilm consortia on immobilized human pharyngeal cells.

The biofilm model using immobilized pharyngeal cells was developed by Marks et al. [71]and has been utilized in pneumococcal biofilm research by different laboratories[68, 71, 72]. Detroit 562 cells were grown until confluent (~5 days) on either 8-well glass slides (Lab-Tek), tissue culture treated 6-well polystyrene plates, or CellBIND surface 24-well polystyrene plates (Corning). Once confluent, cells were immobilized by fixation with 2% paraformaldehyde (PFA; Sigma) for 15 min at room temperature. After extensive washes with sterile PBS, immobilized pharyngeal cells were supplemented with cell culture medium without antibiotics and then inoculated with an aliquot containing ~7×10^5 CFU/ml of each strain. After 8 h of incubation, biofilms were harvested and counted.

Transwell experiments to physically separate biofilms.

A Transwell support (Corning) was installed in each well of a 6-well plate, creating a bottom and a top compartment within the same well. The Transwell system has a permeable membrane (pore size, ~0.4 μm) that allows the exchange of small molecules between compartments. The bottom compartment was inoculated with an S19F strain, whereas strain TIGR4 or the TIGR4Δ*luxS* mutant strain was inoculated directly into the Transwell device. The total volume was brought to 4 ml by addition of THY, and plates were incubated for 8 h in 5% CO_2. Then, the Transwell support containing a TIGR4 strain (top compartment) was removed, and S19F biofilms formed on the bottom of the well were harvested and quantified by colony counts.

DNA extraction.

DNA was extracted from 200 μl of the harvested biofilm samples with the QIAamp DNA minikit according to the manufacturer's instructions. Final elution was done with 100 μl of elution buffer. DNA preps were quantified using a nanodrop spectrophotometer and stored at −80°C until used.

Sequencing of the *comC* gene from clinical pneumococcal isolates.

Downstream and upstream sequences spanning the *comC* gene were amplified by PCR using primers JVS71L and JVS72R, and the amplified PCR product (~455 bp) was sequenced at Eurofins (Atlanta, GA). Sequences were analyzed with Lasergene software version 10.1.1 (DNASTAR)[58].

Calculation of the Malthusian parameter.

S4 (TIGR4), S6B (8655), S19F (5131), and S23F (8064) strains were cultured in THY at 37°C in a 5% CO_2 atmosphere until they reached the early log phase, and then bacteria were diluted to ~2.5×10^5 CFU/ml. An aliquot (300 µl) from each strain was added into seven different wells of a BioScreen C plate (Lab Systems, Helsinki, Finland), and the plate was incubated at 37°C in a BioScreen C reader. OD_{600} values were recorded by a BioScreen C plate reader every 5 min. A growth curve graphic was prepared using the R language and environment for statistical computing and graphics (http://www.gnu.org/). The Malthusian parameter, the rate of change in density, was finally calculated using the GrowthRates software (version 2.1)[73].

Antibiotic susceptibility testing.

Antibiotic resistance of invasive strains was investigated in order to count, using BAP with the appropriate antibiotic, individual strains when forming biofilm consortia. The Kirby-Bauer disc-diffusion method was used according to the Clinical and Laboratory Standards Institute (CLSI) guidelines. The following antibiotic discs (Becton-Dickinson, East Rutherford, NJ) were used: oxacillin (1 µg), erythromycin (15 µg), clindamycin (2 µg), chloramphenicol (30 µg), trimethoprim-sulfamethoxazole (1.25/23.75 µg), and tetracycline (30 µg). Isolates were regarded as susceptible, intermediate, or resistant by using the breakpoints set described by CLSI. Quality control was done with *S. pneumoniae* reference strain ATCC 49619.

Quantification of biofilm biomass by quantitative PCR.

Primers, probes, and the concentrations utilized are listed in Table 2. Total pneumococcal density was quantified using the pan-pneumococcus *lytA* assay[74], and densities of individual serotypes were quantified using primers and probes targeting serotype-specific sequences within the capsule (*cps*) locus [3, 74]. Reactions were run along serially diluted DNA standards corresponding to 4.29×10^5, 4.29×10^4, 4.29×10^3, 4.29×10^2, 4.29×10^1, and 2.14×10^1 genome equivalents per reaction mixture[75]. Reactions were carried out using a Bio-Rad CFX96 Touch real-time PCR detection system (Bio-Rad, Hercules, CA) and the following cycling parameters: 50°C for 2 min, 95°C for 2 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. The final number of genome equivalents per milliliter data were calculated using the CFX software (Bio-Rad, Hercules, CA).

Visualization of pneumococcal biofilms.

Biofilms inoculated into 8-well glass slides (Lab-Tek) were incubated for 8 h at 37°C in a 5% CO_2 atmosphere. Biofilms were then washed twice with PBS and fixed with 2% PFA for 15 min at room temperature. Once PFA was removed, biofilms were washed with PBS and blocked with 2% bovine serum albumin (BSA) for 1 h at 37°C. These biofilms were then incubated with serotype-specific polyclonal antibodies (~40 μg/ml; Statens Serum Institute, Denmark) for 1 h at room temperature. Antibodies had been previously labeled with Alexa 488 (anti-S4) or Alexa 555 (anti-S19) following the manufacturer's recommendations (Molecular Probes). Stained preparations were finally washed two times with PBS and mounted with ProLong Diamond antifade mountant with 4′,6-diamidino-2-phenylindole (DAPI; Molecular Probes). Biofilms produced on human pharyngeal cells were fixed with 2% PFA and washed three times with PBS. Nucleic

acids were stained with TO-PRO-3 (1 μM), a carbocyanine monomer nucleic acid stain (Molecular Probes), for 15 min; S4 and S19F strains were stained as mentioned above. Confocal images were obtained using an Olympus FV1000 confocal microscope and analyzed with ImageJ version 1.49k (National Institutes of Health, USA) or the Imaris software (Bitplane, South Windsor, CT).

Production of biofilm consortia in a bioreactor with cultures of human pharyngeal cells.

S. pneumoniae strains were inoculated as indicated above into a biofilm bioreactor, which simulated human airways. The bioreactor includes a continuous flow of nutrients to feed both the cell cultures and pneumococci but also to wash off toxic products, thus avoiding pneumococcus-induced cell cytotoxicity[68]. Briefly, human pharyngeal cells were grown on Snapwells (Corning) with a polyester membrane (0.4 μm) supported by a detachable ring. Once confluent (4 to 5 days), cells were inoculated with bacteria and immediately placed in sterile vertical diffusion chambers. The apical side (inner chamber) was perfused with sterile minimal essential medium with no antibiotics, using a Master Flex L/S precision pump system (Cole-Parmer, Vernon, IL). To avoid the accumulation of toxic substances but allow biofilm formation, a low flow rate (0.20 ml/min) was applied. Bioreactor chambers containing biofilm consortia on human pharyngeal cells were incubated for 24 h at 37°C under a sterile environment. At the end of the incubation period, chambers of the bioreactor were opened and biofilm bacteria were harvested as described above. Planktonic specimens were also collected from the outflow of the bioreactor. DNA from biofilms or planktonic cells was purified as described and was used as a template in qPCRs targeting serotype-specific sequences[75, 76]. Final genome equivalents (CFU) per milliliter data were obtained as described above.

Model for contact-mediated killing.

To run the contact-mediated killing model, we considered two populations with densities of N_1 and N_2 cells per milliliter, respectively. These populations grew at a rate proportional to the concentration of a limiting resource, R (in micrograms per milliliter), and parameter k(also in micrograms per milliliter), the Monod constant, which is the concentration of the resource when the population is growing at half its maximum rate, v_1 and v_2 (per cell per hour) for populations 1 and 2, respectively. As described by Levin and Udekwu[77], resources are consumed at a rate proportional to the growth rate and a conversion efficiency parameter, e (in micrograms), which is the amount of resource necessary to produce a new cell. There are lags L_1 and L_2 (in hours) for populations 1 and 2, respectively. We assumed population 2 kills population 1 at a rate proportional to the product of their densities and kill rate constant. The killing commences when the time t exceeds that of the lag for population 2 and is proportional to the concentration of the limiting resource. When the concentration of the limiting resource reaches a lower threshold, R_{MIN} (in micrograms per milliliter), autolysis sets in and populations 1 and 2 die off at rates of d_1 and d_2 (per cell per hour).

With these definitions and assumptions, the rates of change in the concentration of the limiting resource are given by the following equations.

$$\frac{dR}{dt} = -e\psi(R)\left(y_1 N_1 v_1 + y_2 N_2 v_2\right)$$

$$\frac{dN_2}{dt} = y_1 v_1 \psi(R) N_1 - y_2 \gamma N_1 N_2 \psi(R) - x d_1 N_1$$

$$\frac{dN_2}{dt} = y_2 v_2 \psi(R) N_2 - x d_1 N_2$$

$$\psi(R) = \frac{R}{(R+K)}$$

$$y_1 = 0 \text{ when } t < L_1 \text{ and } y_1 = 1 \text{ when } t \geq L_1$$

$$y_2 = 0 \text{ when } t < L_2 \text{ and } y_2 = 1 \text{ when } t \geq L_2$$

$$x = 1 \text{ when } R > R_{\text{MIN}} \text{ and } x = 0 \text{ when } R \leq R_{\text{MIN}}$$

where

To solve these equations numerically (to simulate the dynamics), we used the Berkeley Madonna program.

Statistical analysis.

The means of two pneumococcal densities were analyzed by a two-sample independent *t* test. For each *t* test, equality of variance was tested and, based on the result, equal or unequal variance was assumed for the corresponding *t* test. The means of more than two independent pneumococcal densities were analyzed by analysis of variance (ANOVA). Among samples that had two serotypes, proportions of a given serotype were compared using the Z test. The null hypothesis used was a proportion of 0.5, because equal amounts of each sample were added for the experiment. Two-tailed *P* values of <0.05 were considered statistically significant. Statistical analyses were performed using OpenEpi (Open Source Epidemiologic Statistics for Public Health [http://www.openepi.com/Menu/OE_Menu.htm]) and the software SigmaPlot version 12.0 (Systat Software, Inc.).

2.4 Results

Population dynamics of pneumococcal biofilm consortia produced by vaccine strains.

S. pneumoniae persists in the human nasopharynx in a biofilm state, and highly sensitive methods to investigate pneumococcal colonization include quantitative PCR assays[3, 76, 78, 79]. Therefore, we first evaluated by qPCR the density of monostrain pneumococcal biofilms produced by vaccine strains and compared the densities to those produced by a mixture of two strains of different serotypes. Pneumococcal strains chosen for these studies belonged to vaccine types (e.g., 6B, 19F, and 23F) (Table 1) and were recently isolated from invasive pneumococcal disease (IPD) cases in the United States. They had not been genetically modified since their isolation from patients. Experiments showed that vaccine-type strains inoculated alone formed robust biofilms (>8×10^8 genome equivalents/ml) at 8 h postinoculation on both abiotic substrates (Fig. 1A) and human pharyngeal cells (Fig. 1B). These biofilms produced by S6B, S19F, and S23F were similar, i.e., not statistically significantly different, on either substrate (Fig. 1). Almost identical results were obtained when the densities of planktonic bacteria or of biofilms were obtained by culturing (see Fig. S1 in the supplemental material). When S6B was coinoculated in the same well with either S19F or S23F, the relative density was similar to that produced in wells inoculated only with S6B. Similarly, biofilms produced by a mixture of S19F and S23F were similar to those produced in wells inoculated with only S19F or S23F (Fig. 1). These results indicated that *in vitro* formation of pneumococcal biofilm consortia reaches a plateau, and they led us to hypothesize that strains may partition a biofilm consortium to coexist.

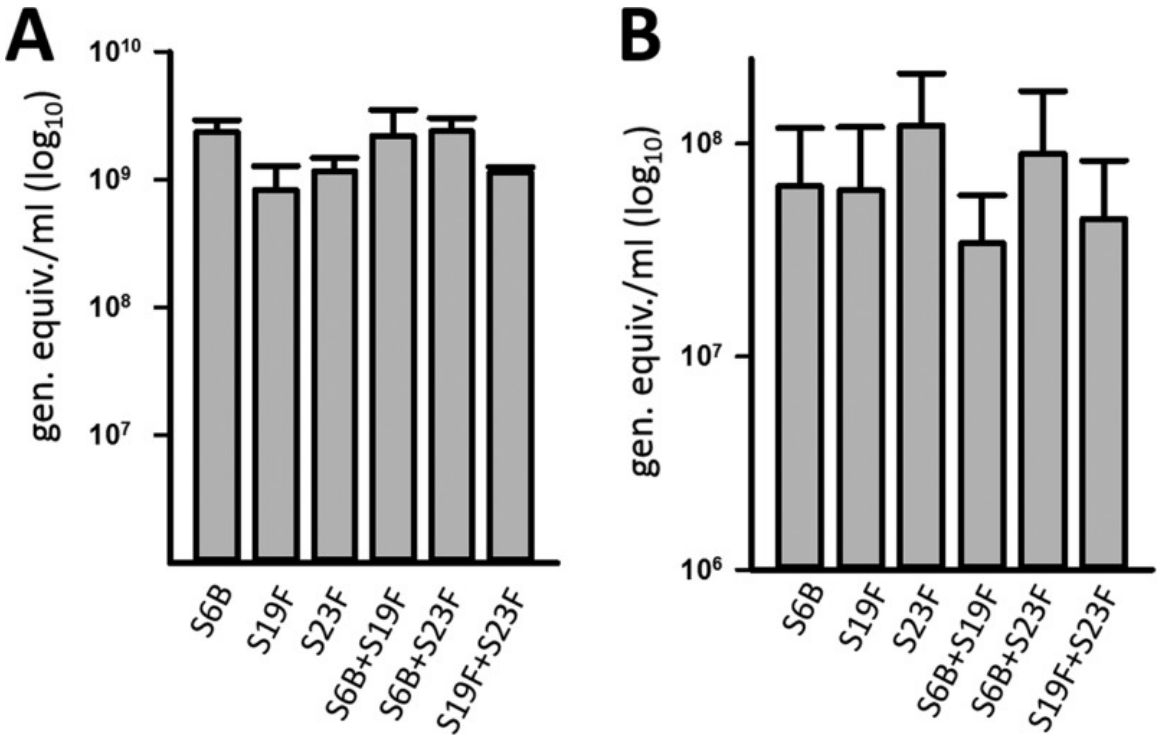


Figure 2.1 Formation of pneumococcal biofilm consortia by invasive strains. S. pneumoniae S6B (3875), S19F (5131), or S23F (8064) was inoculated alone or coinoculated with the indicated strain into 24-well plates (A) or immobilized human pharyngeal cells (B) and incubated for 8 h, after which biofilms (one strain) and biofilm consortia (mixtures) were harvested and DNA was extracted. DNA was used as the template in qPCRs targeting the pan-pneumococcus lytA gene, and the number of genome equivalents per milliliter was calculated. The error bars represent the standard errors of the means and were calculated using data from at least three independent experiments.

Relative densities of pneumococcal strains within biofilm consortia: dominance and tolerance.

To compare the relative density of each strain forming early biofilm consortia, we utilized serotype-specific quantitative reactions[3, 78]. The density of each strain within a

consortium was compared with biofilms formed by individual strains at 8 h postinoculation. As shown in Fig. 2A, the relative density of S6B was similar ($P = 0.57$) whether incubated alone or coincubated with S19F, whereas it significantly decreased when forming biofilm consortia with 23F. Similarly, biofilms made by S23F were significantly reduced when coincubated with S6B, in comparison to biofilms obtained in wells inoculated with S23F alone, but S23F density did not change when coinoculated with S19F (Fig. 2B). In contrast, when inoculated alone, the density of S19F was significantly higher than that of S19F coincubated with S6B or S23F ($P < 0.05$ for both cases) (Fig. 2C).

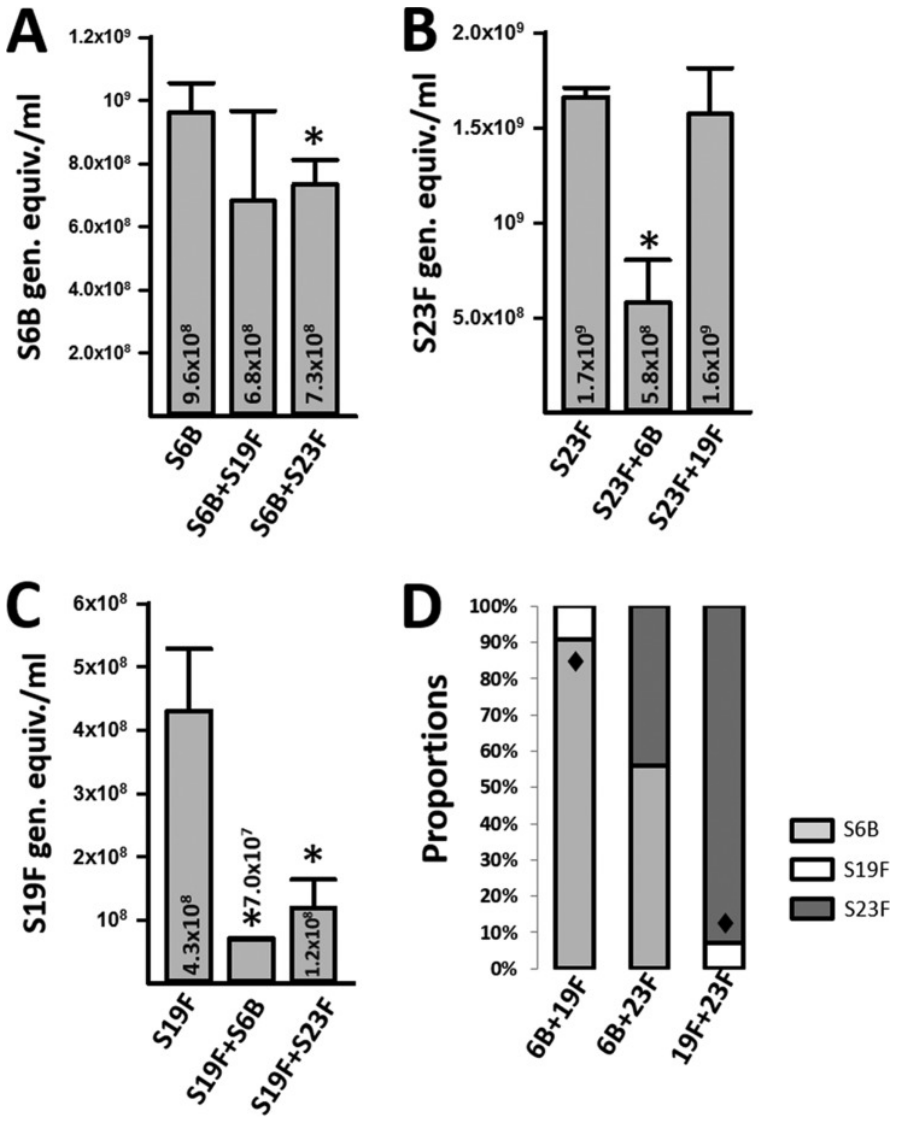


Figure 2.2 Dominance and tolerance within pneumococcal biofilm consortia. S. pneumoniae S6B (3875), S19F (5131), or S23F (8064) was inoculated alone or coinoculated with the indicated strain into 24-well plates and incubated for 8 h, after which biofilms were harvested and DNA was extracted. Serotype-specific qPCRs targeting serotype 6B (A), serotype 23F (B), or serotype 19F (C) were used to quantify the specific bacterial load of each strain. The error bars reflect the standard errors of the means and were calculated using data from three independent experiments. Numbers inside and above the bars are medians. *, statistically significant reduction ($P < 0.05$) in

biomass in comparison to biomass of the strain inoculated alone. (D) The observed proportion was compared to the null proportion of 0.5, as the same amounts of each strain were coinoculated together. Shown are the average proportions obtained from three independent experiments. ◆, statistically different ($P < 0.05$) proportion of strains were incubated together.

The observed proportion of each strain within consortial biofilms was analyzed next. The observed proportion was compared to the null proportion of 0.5, as the same amounts of each strain were coinoculated together. When S6B and S19F were coinoculated, the proportion of biofilms made by S6B (0.91) was significantly higher than that observed for S19F (0.09) (Fig. 2D). As expected, a significantly different proportion was observed when S19F and S23F were coinoculated: 0.07 and 0.93, respectively. The proportions of biofilms made by S6B and S23F, however, were similar: 0.56 and 0.44, respectively. Overall similar proportions were obtained when early biofilm consortia were produced on human pharyngeal cells (data not shown). Together, these results demonstrated dominance of S6B, or of S23F, within pneumococcal biofilm consortia produced along with S19F. Data also indicated that relative densities of S6B and S23F within biofilm consortia were similar (i.e., strains shared the niche equally).

Rate of change of density of pneumococcal strains.

To investigate if the increased relative densities of S6B, or of S23F, in consortial biofilms with S19F were related to an accelerated consumption of resources, the Malthusian parameter, which evaluates the rate of change in density per minute, was obtained. Growth curves (based on the optical density at 600 nm [OD_{600}]) first showed a delayed

growth of S6B and S23F in comparison to S19F at 5 h postinoculation (Fig. 3A). Accordingly, the calculated Malthusian parameter (i.e., growth rate per minute) was significantly lower for S23F and S6B than for S19F (Fig. 3B), thus indicating that faster consumption of resources by dominant S6B and S23F strains did not account for the observed increased relative densities.

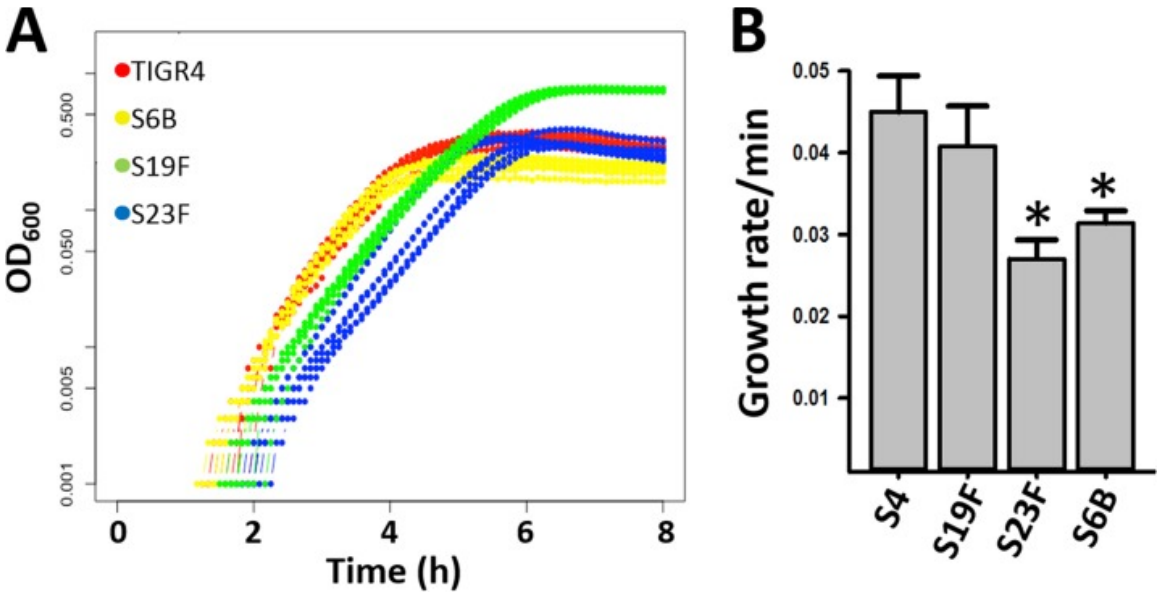


Figure 2.3 Rate of change of density of pneumococcal strains. (A) S4 (TIGR4), S6B (3875), S19F (5131), and S23F (8064) strains were cultured in THY until they reached early log phase, and then bacteria were diluted to ~2.5 × 105 CFU/ml. An aliquot (300 µl) from each strain was added to seven different wells of a BioScreen plate, and the plate was incubated at 37°C for 24 h. OD values were recorded by the plate reader every 5 min. (B) The Malthusian parameter (the growth rate per minute) was calculated as described in Materials and Methods. *, statistically significant difference ($P < 0.001$) in comparison to S19F.

Spatial localization of S19F strains within pneumococcal biofilm consortia.

Strains utilized in the above-described experiments were recently isolated from patients with pneumococcal disease, and therefore genetic information was not available. To gain insights into the potential mechanism controlling relative densities of biofilm consortia, we incubated S19F along with a reference genome sequenced strain, TIGR4 serotype 4

(vaccine type) [80]. These experiments showed that, in biofilm consortia including TIGR4 and S19F strain 5131, the relative density of S19F decreased to the same extent as that observed when S19F was coinoculated with S6B or S23F (Fig. 4A). The density of a different S19F strain, 4924, in consortial biofilms with TIGR4 was tested with similar results (Fig. 4B). This TIGR4-induced reduction of density of S19F strains was specific, since the relative density of reference strain D39 (serotype 2) was similar in both culture types when incubated alone or in biofilm consortia with TIGR4 (data not shown).

To visualize the spatial localization of each strain within biofilm consortia, we stained their capsule by fluorescence, using serotype-specific anticapsule antibodies and confocal microscopy. Micrographs first revealed that attachment to the glass substratum varied between S19F strains. At 4 h postinoculation, strain 5131 attached and formed a net-like pattern (Fig. 4C), whereas 4924 formed biofilm aggregates (Fig. 4E). After 8 h of incubation, however, the aggregates formed by 4924 had apparently disaggregated, covering just 70% of the substrate (Fig. 4F). A similar biofilm phenotype was produced by strain 5131 8 h postincubation (Fig. 4B). As expected, as we and others have reported that decreased transcription of capsule genes when biofilms mature [70], the pneumococcal capsule was clearly observed 4 h postinoculation but capsule staining decreased after 8 h of incubation. Selection of unencapsulated pneumococci binding to the substrate and then increasing in density after 8 h could also be a possibility. In the case of TIGR4 (Fig. 4O and andP),P), the capsule aggregated in bacterial poles (Fig. 4Q, arrows) or was apparently lost (Fig. 4Q, asterisks). Due to this effect, to further confirm the presence of bacterial cells, the DNA was also stained. When TIGR4 was incubated with S19F strains, there was a dramatic reduction in S19F bacteria at both 4 h and 8 h

postinoculation (Fig. 4, red bacteria [S19F] inoculated alone versus those inoculated with TIGR4). The density of TIGR4 bacteria in these consortia was not affected.

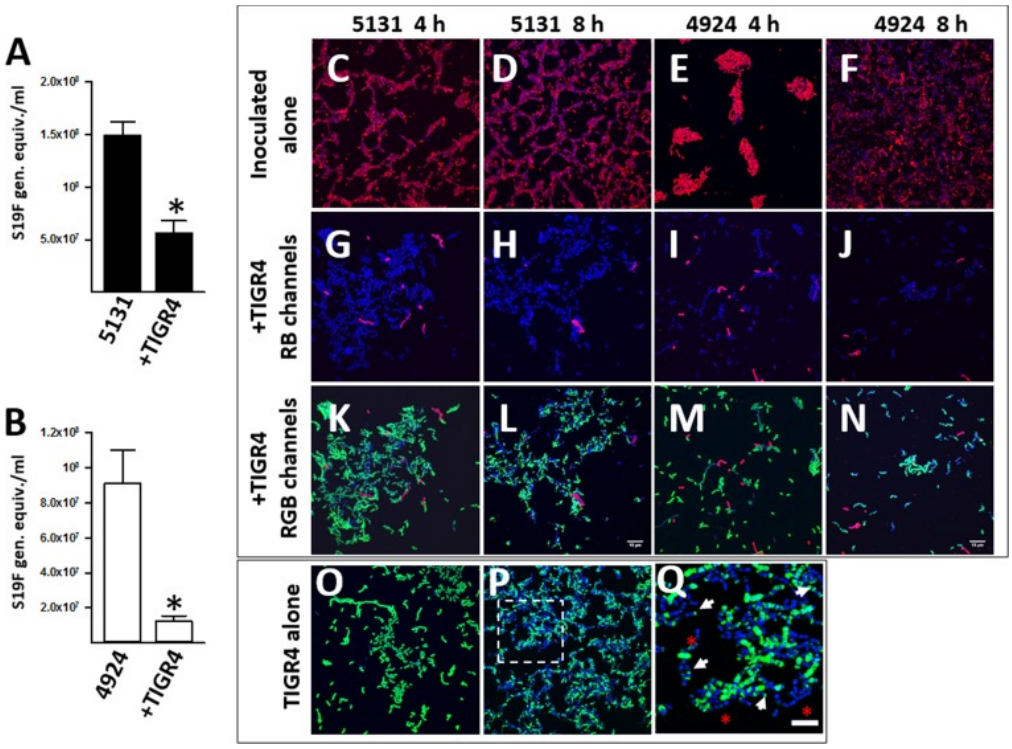


Figure 2.4 Biofilm consortia formed by invasive strains TIGR4 and S19F. (A and B) S19F (strain 5131) (A) or S19F (strain 4924) (B) was inoculated alone or with TIGR4 and incubated for 8 h, after which biofilms were harvested and DNA was extracted. Serotype-specific qPCRs targeting serotype 19F were used to quantify the specific bacterial load of S19F strains. (C to Q). Strains were incubated (as described for panels A and B) for 4 or 8 h, after which pneumococcal strains were stained with an anti-S19 antibody conjugated to Alexa 555 (red channel) or an anti-S4 antibody conjugated to Alexa 488 (green channel). Pneumococcal DNA was stained with DAPI (blue channel). Shown are confocal micrographs of S19F stains inoculated alone (C to F) or with TIGR4 and only showing red and blue channels (RB) (G to J) or showing all channels (RGB) (K to N). Results for strain TIGR4 incubated alone for 4 h (O) or 8 h (P) are also shown. (Q) An enlargement of the indicated area in panel P. Arrows point to zones where capsule staining concentrated in poles, and asterisks show pneumococci that lost their capsule.

Bars, 10 μm (shown in panels L and N and valid for all other panels except panel Q, in which the bar equals 5 μm).

Density of S19F decreases in nasopharyngeal biofilm consortia with TIGR4.

To investigate whether the relative density of S19F decreases in mature nasopharyngeal biofilm consortia along with TIGR4, we incubated both strains in a bioreactor containing living cultures of human pharyngeal cells for 24 h. Mature biofilms already form after 24 h of incubation in a biofilm bioreactor[68]. The density of S19F biofilms and numbers of S19F planktonic cells coming off the bioreactor chamber were significantly reduced when S19F was coincubated with TIGR4, in comparison with bioreactor chambers incubated only with S19F (Fig. 5A and andB).B). Confocal experiments additionally demonstrated that pneumococcal strains were in close proximity when incubated together (Fig. 5C); physical contact was confirmed through optical sections of confocal micrographs (Fig. 5D, yellow circles).

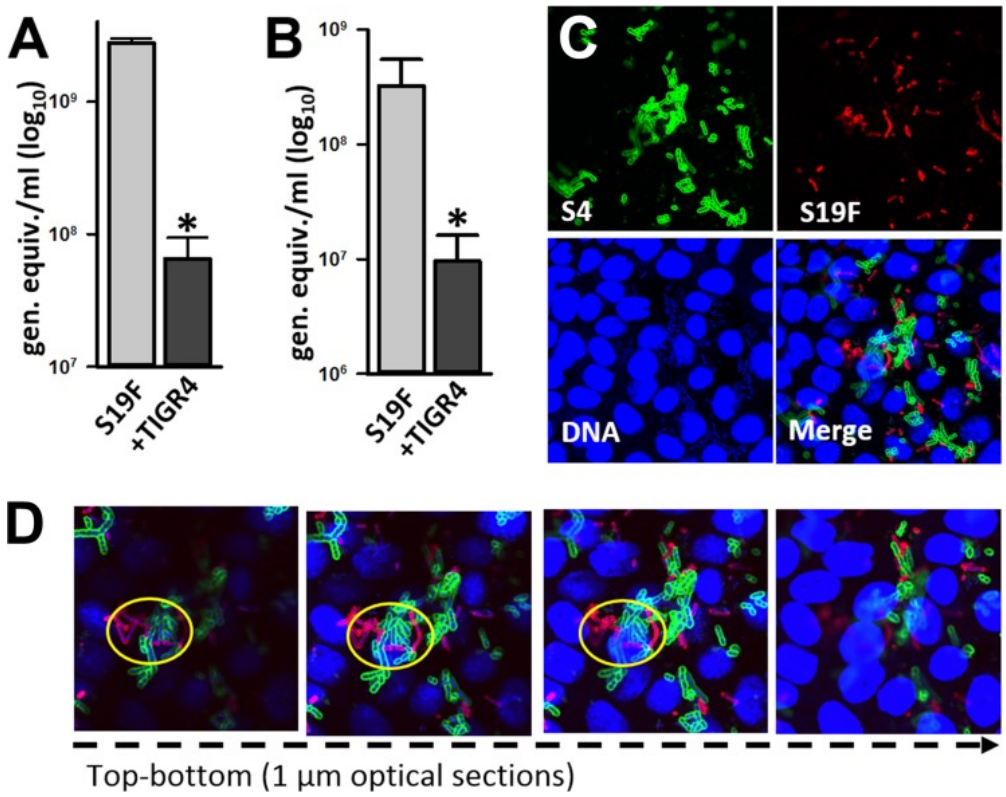


Figure 2.5 TIGR4 reduces biofilms of S19F on consortia formed on human nasopharyngeal cells. (A and B) S19F (strain 4924) was inoculated with TIGR4 and incubated for 24 h, after which biofilm consortia (A) or planktonic bacteria (B) were harvested and DNA was extracted. Serotype 19-specific qPCRs were used to quantify the number of genome equivalents per milliliter. The error bars reflect the standard errors of the means and were calculated using data from three independent experiments. *, statistically significant difference ($P < 0.05$) in comparison to S19F incubated alone. (C and D) Confocal micrographs of biofilm consortia formed by TIGR4 and S19F on human nasopharyngeal cells. Panel C shows the projection and panel D shows optical sections. Bacteria were stained with an anti-S19F antibody conjugated to Alexa 555 (red channel) or an anti-S4 antibody conjugated to Alexa 488 (green channel). Cells and pneumococcal DNA were stained with TO-PRO-3 (blue channel).

The factor(s) allowing reduction of S19F relative density is not regulated by quorum-sensing systems LuxS/AI-2 and Com.

Given that the quorum-sensing systems LuxS/AI-2 and Com have been implicated in formation of monostrain pneumococcal biofilms with strain D39 and are key regulators of fratricide and competence [68, 70, 81], TIGR4Δ*luxS* and TIGR4Δ*comC* mutant strains were prepared and evaluated for regulating the factor(s) that allows dominance of S19F biomass in biofilm consortia. Densities of monostrain biofilms were similar when 4924, 5131, mutant strain TIGR4Δ*luxS*, or mutant strain TIGR4Δ*comC* were compared ($P > 0.05$ for all comparisons) (Fig. 6A).

When biofilm consortia were produced with either the TIGR4Δ*luxS* or TIGR4Δ*comC* mutant strains, the relative density of S19F strains 4924 or 5131 decreased ~20% in comparison to the biomass produced by these strains when incubated alone (Fig. 6B and C). This reduction of S19F density was similar to that induced by the wild-type (wt) TIGR4 strain (data not shown). Therefore, the TIGR4 factor(s) that allows reduction of S19F density is not regulated by the QS systems LuxS/AI-2 and Com. Biofilms of the mutant strains TIGR4Δ*luxS* and TIGR4Δ*comC* were not affected by the presence of S19F strains ($P > 0.05$ for all comparisons) (data not shown).

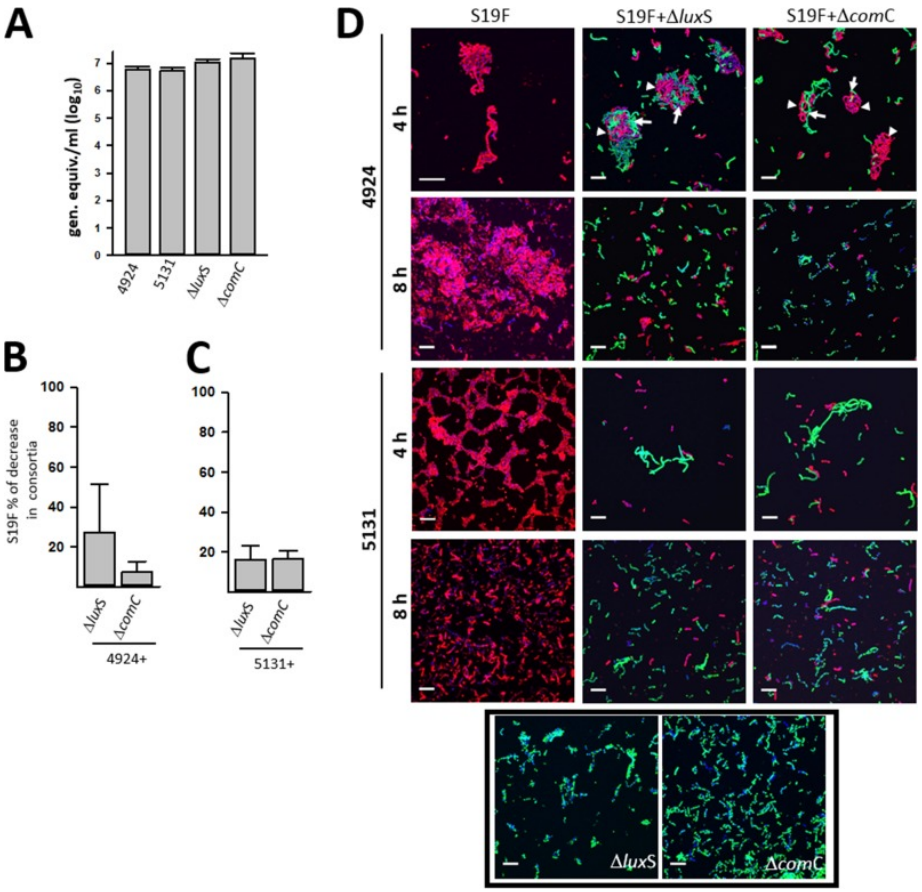


Figure 2.6 QS systems LuxS/AI-2 and Com do not regulate the dominant phenotype within biofilm consortia. (A to C) S19F strains (5131 and 4924) or mutant TIGR4ΔluxS or TIGR4ΔcomC strains were inoculated alone (A) or in mixtures containing strain 4924 and either TIGR4 derivative (B) or strain 5131 and either TIGR4 derivative (C). Biofilms were incubated for 8 h and then removed, diluted, and plated onto BAP containing ampicillin (for strain 5131), tetracycline (for strain 4929), erythromycin, or streptomycin (for the TIGR4ΔluxS and TIGR4ΔcomC mutant strains) when incubated with 5131 or 4929, respectively (i.e., 4929 is resistant to erythromycin). In panels B and C, the percent biomass decrease of S19F strains is presented for biofilm consortia, in comparison to strains inoculated alone. The error bars reflect the standard errors of the means and were calculated from three independent experiments. (D) S19F strains 5131 and 4924 were

inoculated alone or with mutant strain TIGR4ΔluxS or TIGR4ΔcomC and incubated for 4 or 8 h, after which S19F strains and TIGR4 derivatives were stained with an anti-S19 antibody conjugated to Alexa 555 or an anti-S4 antibody conjugated to Alexa 488, respectively. Pneumococcal DNA was stained with DAPI. Arrows indicate where TIGR4 derivatives localized within the biofilm consortium with S19F strain 4924, for which localization is indicated with an arrowhead. Bars, 10 μm.

Confocal micrographs were also obtained. As expected, at 4 h postinoculation S19F 4924 formed aggregates (Fig. 6D). In wells incubated for 4 h with 4924 and mutant strain TIGR4Δ*luxS* or TIGR4Δ*comC*, these aggregates were formed by a mixture of both strains (Fig. 6D, first row [red shows 4924 and green shows TIGR4 derivatives]). After 8 h of incubation, however, S19F aggregates had disappeared from wells incubated with either TIGR4 derivative QS mutant, reducing the population of 4924 bacteria. In contrast, S19F 4924 incubated alone formed robust biofilms. A similar reduction of 5131 biofilms attached to the substratum was observed when mutant strain TIGR4Δ*luxS*, or TIGR4Δ*comC* was incubated with 5131, while the strain growing alone was able to form robust biofilms at 8 h postinoculation (Fig. 6D). The biomass of mutant strain TIGR4Δ*luxS* or TIGR4Δ*comC*, whether incubated alone or with S19F, did not change.

Physical contact is required for dominance within pneumococcal biofilm consortia.

To investigate whether physical contact is required to dominate a biofilm consortium, we simulated physical interaction of an S4 strain and an S19F strain. (Parameters utilized to simulate the interactions are described in Materials and Methods.) The simulation showed

that after 4 h of coincubation, the S4 strain limited the growth of the S19F strain by about 1 order of magnitude, 2.1×10^7 versus 1.1×10^6 $\log_{10}$ CFU/ml (Fig. 7A). Biofilm dominance was not related to an increased consumption of resources by the dominant S4 strain, since the hourly growth rates of individual strains were similar (Fig. 3). Moreover, at 4 h, when S19F inhibition was observed, we detected a ~90% level of resources necessary to produce new cells (Fig. 7A). At 8 h postinoculation, the difference in densities of individual strains within a biofilm consortium made of S4 and S19F strains was about 2 orders of magnitude (Fig. 7A). Contact-dependent reduction of S19F density was verified by colony counts. To do this, we utilized S19F strain 4924, which is naturally resistant to erythromycin, and generated a TIGR4 strain encoding resistance to streptomycin, SPJV21. Figure 7B shows a statistically significant, ~2-log reduction of the S19F strain when incubated with TIGR4, whereas the density of TIGR4 was not affected by coincubation with the S19F strain (Fig. 7C).

To experimentally demonstrate whether reduction of the relative density of S19F required physical contact or whether it was mediated by a soluble factor(s), strain TIGR4 and S19F strain 4924, or S19F strain 5131, were inoculated into the same well but the strains were separated with a Transwell system device, i.e., TIGR4 was inoculated in the top chamber and the S19F strain (either 4924 or 5131) was inoculated in the bottom. Experiments with the TIGR4Δ*luxS* mutant were also included. The biomass of 5131 incubated with either TIGR4 strain (i.e., wt or the TIGR4Δ*luxS* mutant) was similar or actually increased after 8 h of incubation in comparison to wells where strain 5131 was inoculated alone (Fig. 7D). Similar results were obtained with S19F strain 4924. These data demonstrated that physical contact is required to reduce the density of S19F strains

in consortia with TIGR4. A similar density of TIGR4 was observed in incubation wells that also contained 5131, in comparison with wells where only TIGR4 was inoculated (data not shown). As expected, given that the TIGR4Δ*luxS* mutant does not produce AI-2, in Transwell experiments where the TIGR4Δ*luxS* mutant and 5131 were inoculated together, the density of the TIGR4Δ*luxS* mutant increased, perhaps due to LuxS/AI-2 produced by 5131 (Fig. 7E).

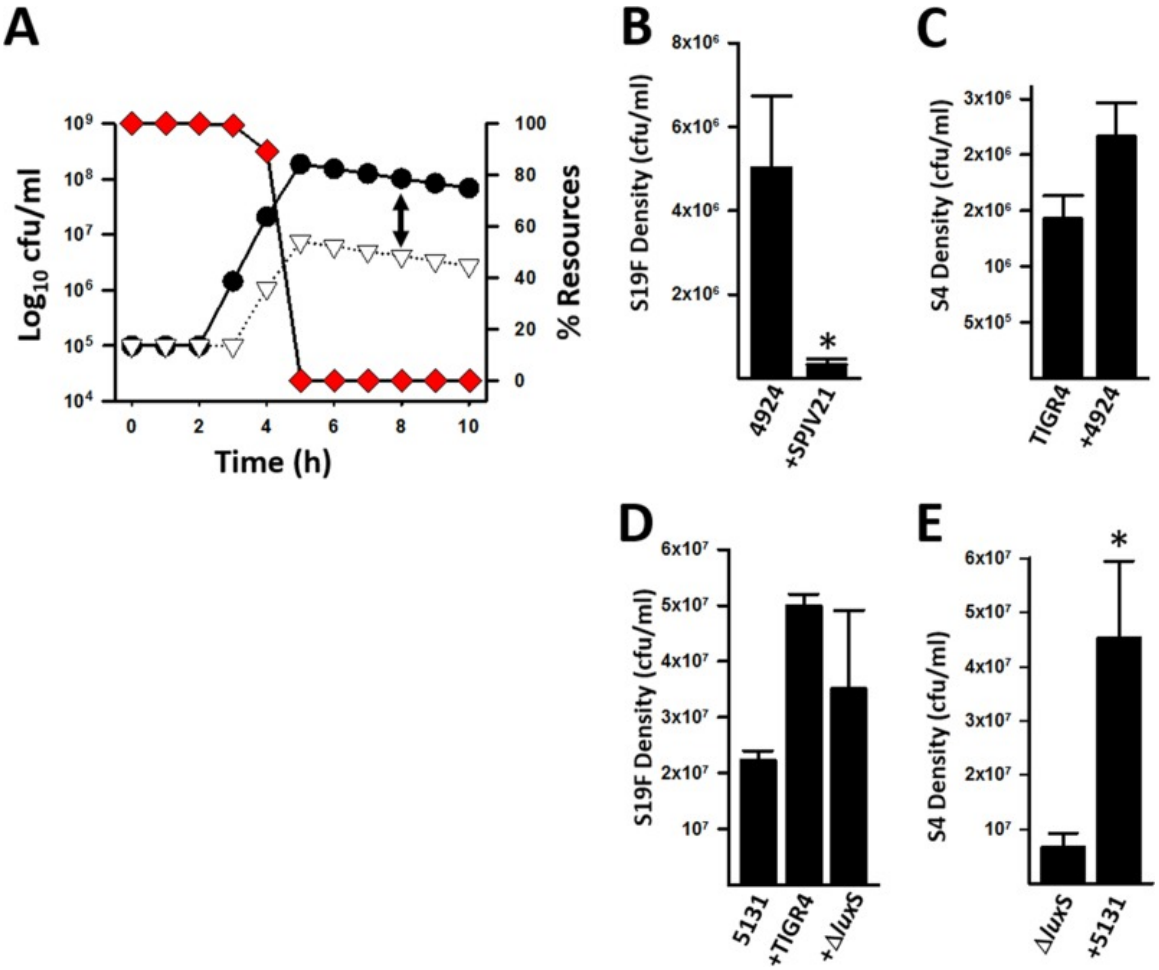


Figure 6.7 Physical contact-mediated dominance of pneumococcal biofilm consortia. (A) Simulation of biofilm physical contact-required dominance during a 10-h incubation period. Circles, S4; triangles, S19F; diamonds, resources. Parameters and formulas utilized to construct the model and prepare the graphic are described in Materials and

Methods. The arrow shows a 2-log difference in densities of S4 and S19F at 8 h. (B and C) S19F strain 4924 and a TIGR4 derivative streptomycin-resistant strain (SPJV21) were incubated alone or coincubated for 8 h, after which biofilms were harvested, serially diluted, and plated onto BAP with erythromycin to count S19F cells (B) or BAP with streptomycin to obtain counts of SPVJ21 cells (C). Error bars represent the standard errors of the means and were calculated using data from three independent experiments. *, statistically significant reduction ($P < 0.05$) of biomass in comparison to biomass of the strain inoculated alone. (D and E) TIGR4 wt or TIGR4ΔluxS was inoculated into the top chamber of a Transwell device, and S19F strain 5131 was inoculated into the bottom chamber. As a control, 5131 was incubated alone. Biofilms were incubated for 8 h and then harvested, and counts for S19F (D) and TIGR4ΔluxS (E) were obtained by dilution and plating. The error bars reflect the standard errors of the means and were calculated from three independent experiments. *, statistically significant increase in biomass ($P < 0.05$) in comparison to the strain inoculated alone.

2.5 Discussion

In this work, we recreated the dynamics of competitive interactions within pneumococcal nasopharyngeal biofilm consortia and demonstrated a dominance of the relative density of S19F strains compared to all other strains tested, i.e., S6B, S23F, and TIGR4. Dominance and tolerance have been observed in recent epidemiological studies where the specific bacterial density of pneumococcal serotypes was obtained (explained in detail below) (11, 28). In this study, dominance against S19F strains within consortia required physical contact, since in experiments conducted in Transwell devices, the TIGR4 strain was not able to decrease the biomass of S19F strains. Dominance of pneumococcal

cocolonization through a mechanism requiring physical contact recapitulates that it may naturally occur in the human nasopharynx, where pneumococcal strains cocolonize a nonaqueous microenvironment. Strain dominance was recently described in a mouse model of pneumococcal cocolonization (with use of serotype-specific qPCRs to evaluate serotype density) [82]. Pneumococcal colonization has also been inhibited *in vitro* by incubating pneumococcal strains with probiotic bacteria [83, 84].

Vaccine serotypes 6B, 19F, and 23F evaluated in this study are the most common serotypes isolated in IPD in countries where the pneumococcal vaccine is not available (i.e., India and China); together, they cause ~30% of IPD cases [85]. Unlike strains belonging to serotype 1 or 5, which are highly prevalent in IPD cases but display low prevalence in carriage studies, serotype 6B, 19F, and 23F strains are also the most prevalent serotypes carried in the nasopharynx of children [59, 85, 86]. While strains were able to form biofilms on both abiotic and biotic surfaces (i.e., human pharyngeal cells), only S6B and S23F dominated biofilm consortia when S19F colonized the same niche.

Recent studies have investigated nasopharyngeal colonization by multiple pneumococcal serotypes, showing that up to 50% of children are cocolonized by 2 or more strains (i.e., serotypes) [3, 87, 88]. In Malawian children, for example, more than 75% of cocolonization events included a vaccine type[87], whereas our study using single-plex qPCR demonstrated a vaccine type in 80% of cocolonized, nonvaccinated Peruvian children [3]. The study from our laboratory also quantified the specific bacterial load of pneumococcal serotypes in the cocolonized children and demonstrated dominance of a pneumococcal strain in ~85% of cocolonization events [3]. A semiquantitative

microarray approach resulted in similar evidence [3, 88]. We define strain dominance, in children cocolonized by two or more strains, as those cocolonization events where the bacterial load (i.e., biomass) of a specific strain accounts for at least 60% of the total pneumococcal load. Therefore, based on new recent epidemiological evidence and this study's findings, dominance appears to be the most common event during nasopharyngeal cocolonization by multiple pneumococcal strains.

Experiments comparing biofilm biomass formed by single strains versus the biomass of biofilm consortia suggest that dominance and tolerance within biofilm consortia are limited by the substrate. For example, whether S6B and S23F strains were inoculated separately or together, the total biomass was very similar (i.e., did not double in size). The *in vivo* situation in the human nasopharynx and oropharynx may be similar, although the host immune response should play an additional role in limiting the biomass of certain types. Experiments we have described here also demonstrated that when two strains cohabit a biofilm consortium produced on abiotic surfaces, the biomass of both strains will be proportionally similar (~50%). S19F strains, however, were dominated by all strains tested, whether grown on abiotic substrates or on human pharyngeal cells. Dominance of S6B, and S23F, against 19F correlates with the findings of an epidemiological study with Spanish children that demonstrated that S6B strains were more likely to cocolonize children with strains from other serotypes[49], and recent unpublished studies from our lab have shown that serogroup 6 strains and 23F are the most prevalent strains in cocolonized, nonvaccinated Peruvian children.

The specific mechanism(s) for dominance, or tolerance (if any), that is common to all pneumococci is under investigation by our laboratory and others. A candidate for

modulating densities of pneumococcal types had been the *blp* locus, which encodes a bacteriocin system that produces a potent bacteriocin, BlpC, but it has been recently demonstrated that BlpC plays a minor role in cocolonization [89, 90]. A secreted factor might not be involved in dominance of S19F strains by the S4 strain TIGR4, as demonstrated in experiments using Transwell systems, although our experimental design did not allow us to test for a factor(s) that could be only released when in close proximity. The factor(s) appears not to be regulated by the quorum-sensing LuxS/AI-2 and Com systems, since individual mutants prepared in the TIGR4 background were still able to reduce the population of S19F strains in biofilm consortia with S4 mutants. The possibility exists that a minimum amount of a QS molecule produced by the coincubated wt strain may complement the mutant strain. In the case of the TIGR4Δ*luxS* mutant, the absence of LuxS/AI-2 in TIGR4 could be supplied by that produced by S19F strains. This may be the same case for the absence of production of CSP (i.e., encoded by the *comC* gene) in the TIGR4Δ*comC* mutant when incubated along with S19F 5131, since both produced CSP2 but not in biofilm consortia with 4924, as this strain produced a different variant (i.e., CSP1).

It is likely that strains that dominate biofilm consortia induce cell lysis of the dominated strain, and therefore a source of DNA for recombination might be available. A study by Marks et al. [91] demonstrated higher frequencies of recombination in pneumococcal biofilms produced on human pharyngeal cells than on their planktonic counterparts. However, neither the recombination direction (i.e., when strains were incubated together) nor specific bacterial densities were obtained to allow us to determine whether, in mixtures of two different pneumococcal strains, the strain acquiring DNA from the other

one dominates the biofilm consortium [91]. We hypothesize that the strains which dominate biofilm consortia have an advantage, not just for colonization but also for acquiring DNA from other pneumococci. Studies in our laboratory are currently focused on investigating this hypothesis.

A very recent article from our laboratory demonstrated killing of *Staphylococcus aureus* biofilms and planktonic cells by TIGR4 and other pneumococci [92]. The unknown factor required physical contact and completely eradicated preformed biofilms made by the *S. aureus* reference strain and methicillin-resistant *S. aureus* strain USA300-0114; whether the factor(s) that allows TIGR4 to limit the biomass of S19F strains is similar to that eradicating *S. aureus* strains needs to be investigated. The need for physical contact might correlate with the *in vivo* situation of the upper airways, where pneumococcal strains have a limited liquid environment in which to secrete products and must instead release and acquire these products when in close physical proximity to each other.

Chapter 3. Incomplete Influenza A Virus Genomes are Abundant but Readily Complemented During Localized Viral Spread.

This work has been accepted for publication *Nature Communications* following formal copy editing. An earlier draft appears a preprint on *bioRxiv*.

3.1 Abstract

Viral genomes comprising multiple distinct RNA segments can undergo genetic exchange through reassortment. Segmentation also allows the generation of incomplete viral genomes (IVGs) within a cell, and evidence suggests that IVGs occur frequently for influenza A viruses. Here we quantified the frequency of IVGs using a novel single cell assay and then examined their implications for viral fitness. We find that each segment of influenza A/Panama/2007/99 (H3N2) virus has only a 58% probability of being present in a cell infected with a single virion. These observed frequencies account for abundant reassortment, and suggest that an average of 3.6 particles are required for replication of a full viral genome in a cell. To examine the extent to which this requirement for multiple-infection slows viral propagation, we first used theoretical models. In a well-mixed system, we find that IVGs carry high costs. These costs are substantially reduced by spatial structure, however, as complementation occurs more readily when spread occurs locally. This expectation is supported by experimental infections in which spatial structure was manipulated. Furthermore, a virus engineered to be entirely dependent on co-infection grows robustly in guinea pigs, suggesting that coinfection is sufficiently common *in vivo* to support propagation of IVGs. The infectivity of this mutant virus is, however, reduced and it does not transmit to contacts. Thus, while incomplete genomes

augment reassortment and contribute to within-host spread, the existence of rare complete genomes may be critical for transmission to new hosts.

3.2 Introduction

Pathogen evolution poses a continued threat to public health by reducing the effectiveness of antimicrobial drugs and adaptive immunity. In the case of the influenza A virus (IAV), this evolution results in seasonal outbreaks as new viruses emerge to which pre-existing immunity is weak. Each year requires a new vaccine as a consequence, and keeping pace with IAV evolution is challenging: unexpected emergence of new strains could render the vaccine obsolete before the flu season starts. IAV populations evolve rapidly in part because their mutation rates are high, on the order of 10^{-4} substitutions per nucleotide per genome copied[1]. The segmentation of the viral genome gives a second source of genetic diversity. The IAV genome is composed of eight single-stranded RNA segments, and so cells co-infected with two different IAV virions can produce chimeric progeny with a mix of segments from these two viruses. This process, termed reassortment, carries costs and benefits analogous to those of sexual reproduction in eukaryotes[2]. Reassortment can combine beneficial mutations from different backgrounds to alleviate clonal interference[3], and purge deleterious mutations to mitigate the effects of Muller's ratchet[4,5]. This combinatorial shuffling of mutations may accelerate adaptation to new environments such as a novel host[6]. But free mixing of genes through reassortment may also reduce viral fitness by separating beneficial segment pairings, as sexual reproduction carries this cost in eukaryotes[7]. Previous work has shown that reassortment occurs readily between closely related variants[8], but is limited between divergent lineages due to molecular barriers[9,10] or reduced fitness of

progeny[11,12]. Nevertheless, the contribution of reassortment to emergence of novel epidemic and pandemic IAVs has been documented repeatedly[13-17]. Factors that affect the frequency of co-infection and consequently reassortment are therefore likely to play an important role in viral evolution.

While the ability of a virus particle to enter a cell depends only on the proteins that line the virion surface, subsequent production of viral progeny requires successful expression and replication of the genome. A virion that does not contain, or fails to deliver, a complete genome could therefore infect a cell but fail to produce progeny. IAV particles outnumber plaque-forming units (PFUs) by approximately 10–100 fold[18], meaning that only a minority of particles establish productive infection at limiting dilution. Recent data suggest that IAV infection is not a binary state, however. Efforts to detect viral proteins and mRNAs at the single cell level have revealed significant heterogeneity in viral gene expression[19-22]. These data furthermore suggest that a subset of gene segments is often missing entirely from cells infected at low multiplicity of infection (MOI). Thus, many non-plaque-forming particles appear to be semi-infectious, giving rise to incomplete viral genomes (IVGs) within the infected cell[23].

Replication and expression of only a subset of the genome may be explained by two potential mechanisms: either the majority of particles lack one or more genome segments, or segments are readily lost in the process of infection before they can be replicated. Published data suggest that most particles contain full genomes: electron microscopy revealed eight distinct RNA segments in most virions[24], and FiSH-based detection of viral RNAs indicated that a virion typically contains one copy of each segment[25]. Loss of segments following delivery of a viral genome to the target cell

therefore seems likely to be an important mechanism. Inefficiencies inherent in the processes of cytoplasmic trafficking, nuclear import, and replication of incoming viral RNAs during the earliest stages of infection would all lead to loss of segments. Very likely, multiple mechanisms contribute to give rise to incomplete IAV genomes.

Regardless of the molecular mechanisms that lead to the phenomenon of incomplete IAV genomes, their frequent occurrence suggests that complementation by co-infection at the cellular level is an underappreciated aspect of the viral life cycle. The observation of appreciable levels of reassortment following co-infection at low MOIs suggested IVG reactivation through complementation occurs commonly during IAV infection[26]. Nevertheless, the extent to which IAVs rely on co-infection for replication, and how this need changes over the course of infection, remains unclear. Similarly, the existence of IVGs *in vivo* has been demonstrated[27], but their importance to the dynamics of infection within hosts is untested.

Here we investigate the biological implications of incomplete IAV genomes and the emergent need for cooperation at the cellular level. We first developed a novel single-cell sorting assay to measure the probability of each segment being delivered by an individual virion for influenza A/Panama/2007/99 (H3N2) [Pan/99] virus. Our data estimate that individual virus particles lead to successful replication of all eight gene segments only 1.22% of the time. When considering a well-mixed system in which virus particles are distributed randomly over cells, the potential fitness costs of incomplete genomes are high. In contrast, a model of viral spread that incorporates local dispersal of virions to nearby cells predicts that the spatial structure of virus growth mitigates costs of genome incompleteness. Testing of this model confirmed that infections initiated with

randomly distributed inocula contain more IVGs than those generated by secondary spread from low MOIs, in which spatial structure is inherent. To determine the potential for complementation to occur *in vivo*, we generated a mutant virus that required cellular co-infection for viral replication. We find that this virus is able to grow within guinea pigs, but unable to transmit to cagemates. Taken together, these results suggest that the abundance of incomplete genomes and the potential for complementation are important factors in the replication and transmission of IAV.

3.3 Results

Measurement of P_P

To better evaluate the implications of genome incompleteness for IAV fitness and reassortment, we sought to quantify the probability of successful replication for each of the eight genome segments within single cells infected with single viruses. To ensure accurate detection of IVGs, we devised a system that would allow their replication to high copy number. We applied our approach to the seasonal isolate, influenza A/Panama/2007/99 (H3N2) virus. In this assay, MDCK cells are inoculated with a virus of interest, referred to herein as "Pan/99-WT" or "WT", and a genetically tagged helper virus ("Pan/99-Helper" or "Helper"). This Helper virus differs from the WT strain only by silent mutations on each segment that provide distinct primer-binding sites. For example, qPCR primers targeting WT PB2 will not anneal to cDNA of Helper PB2, and vice versa. By co-inoculating cells with a low MOI of WT virus and a high MOI of Helper virus, we ensure that each cell is productively infected, but is unlikely to receive more than one WT virion. Following infection, one cell per well is sorted into a 96-well

plate containing MDCK cell monolayers. The initially infected cell produces progeny which then infect neighboring cells, effectively amplifying the vRNA segments present in the first cell. The presence or absence of WT segments in each well can then be measured by performing segment-specific RT-qPCR. As detailed in the Methods, the frequencies of Helper virus infection, WT virus infection, and each distinct WT segment were used to estimate the probability that a cell infected with a single WT virus would contain a given segment. The calculation used takes into account the known probability of multiple infection. We termed the resultant parameter "Probability Present", and refer to it hereafter as P_P. Segment-specific values are referred to as $P_{P,i}$ where $i = 1 – 8$, while P_P refers to the average P_P value across all segments, which is calculated as the geometric mean of eight segment-specific values to reflect the fact that productive infection requires independent delivery of all eight genome segments.

Using this assay, the $P_{P,i}$ values for each segment of Pan/99 virus were quantified (Fig. 1A). We observed that each segment was present at an intermediate frequency between 0.5 and 0.6, indicating that incomplete genomes may arise from loss of any segment(s). An average P_P value was also estimated for each experimental replicate by calculating the geometric mean of the eight segment-specific $P_{P,i}$ values. The arithmetic mean of each of these 13 summary P_P values was 0.58 (95% C.I. 0.54 – 0.61). When used to parameterize a model that estimates the frequency of reassortment, which we published previously[26], these $P_{P,i}$ values generated predicted levels of reassortment that align closely with experimental data (Fig. 1B). This match between observed and predicted reassortment is important because i) it offers a validation of the measured $P_{P,i}$

values and ii) it indicates that IVGs fully account for the levels of reassortment observed, which are much higher than predicted for viruses with only complete genomes[26].

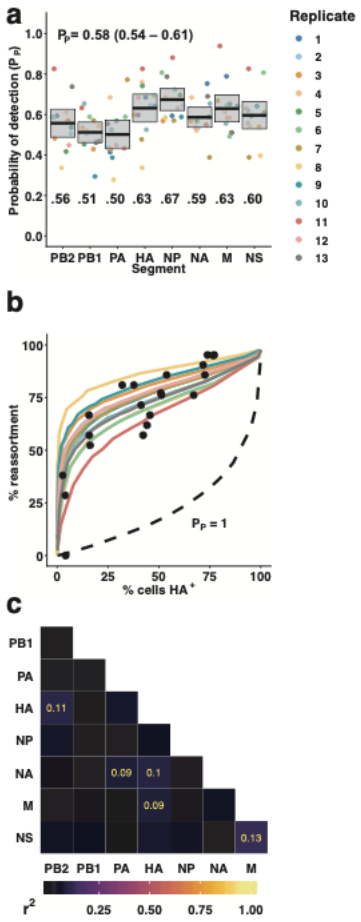


Figure 1 — Incomplete genomes are common in Pan/99 virus infection.

(A) Segment-specific $P_{P,i}$ values were measured by a single-cell sorting assay. Each set of colored points corresponds to eight $P_{P,i}$ values measured in a single experimental replicate, with thirteen independent replicates performed. Horizontal bars indicate the mean (written above each segment's name) and shading shows the 95% C.I. (mean ± 1.96 *S.E.). (B) Using each replicate's $P_{P,i}$ values as input parameters, the computational model from Fonville et al. was used to predict the frequency of reassortment across multiple levels of infection[26]. Black circles represent experimental data from Fonville et al. and show levels of reassortment observed following single-cycle coinfection of MDCK cells with Pan/99-WT and a Pan/99 variant viruses. Colored lines show the

theoretical predictions made by the model, with colors corresponding to the legend shown in panel A. (C) Pairwise correlations between segments (r^2) are shown as color intensities represented by a color gradient (below). r^2 values are shown in yellow for significant associations (Nr^2 (where N is the sum of p(1 virion) values) follows a χ^2 distribution with 3 degrees of freedom, $p < 0.05$ after Bonferroni correction for multiple comparisons).

Interactions between viral ribonucleoprotein (vRNP) segments are thought to play an important role in the assembly of new virions[10,28-31]. To determine whether similar interactions exist that could mediate the co-delivery of segments to the cell, the patterns of segment co-occurrence were analyzed. In performing this analysis, it was again important to take into account the known probability of multiple infection in our single cell assay. As shown in Supplementary Figure 1, cells containing more segments were likely to have been infected with multiple virions. Because such cells are less informative for this analysis, we applied a weighting factor to ensure that results relied more strongly on data from cells with fewer WT segments. Namely, we determined the probability that a given cell acquired its segment constellation by infection with a single virion, and weighted data according to this probability to calculate the pairwise correlation between segments. While some significant interactions were observed, they were relatively weak, with r^2 values below 0.15 (Fig.1C). Thus, our data suggest that associations among specific vRNPs do not play a major role during the establishment of infection within a cell. Given the independence of vRNP delivery and the similarity between $P_{P,i}$ values, the models described below use the average P_P value of 0.58 for simplicity.

Predicted costs of incomplete genomes for cellular infectivity

If singular infections often result in replication of fewer than eight viral gene segments, then multiple particles would be required to productively infect a cell. To evaluate the relationship between the frequency of IVGs and the number of particles required to infect a cell, we developed a probabilistic model in which the likelihood of segment delivery is governed by the parameter P_P. In Figure 2A we examine how P_P affects the frequency with which a single virion delivers a given number of segments. If P_P is low, singular infections typically yield few segments per cell. Even at the intermediate P_P that characterizes Pan/99 virus, the vast majority of singular infections give rise to IVGs within the cell. When P_P is high, however, most cells receive the full complement of eight segments. In Figure 2B we plot the relationship between P_P and the percentage of cells that are expected to be productively infected following singular infection. If only a single virus infects a cell, then the probability that all eight segments are present will be P_P^8. For Pan/99 virus, the frequency with which eight segments are present is approximately $0.58^8 = 1.22\%$ (95% C.I. 0 – 30.5%).

Importantly, however, if more than one virus particle infects the cell, then the probability that all eight segments are present will be considerably higher. This effect is demonstrated in Figure 2C, where the percentage of cells containing all eight IAV segments is plotted as a function of the number of virions that have entered the cell. Here we see that, even for low P_P, a high probability of productive infection is reached at high multiplicities of infection. Finally, in Figure 2D, the relationship between P_P and the average number of virions required to productively infect a cell is examined. We see that the number of virions comprising an infectious unit increases sharply at low values of P_P.

Based on our experimentally determined values of P_P for Pan/99 virus ($P_P = 0.58$), we estimate that an average of 3.6 (95% C.I. 1.0 – 6.5) virions must enter a cell to render it productively infected (Fig. 2D). Thus, as a result of stochastic loss of gene segments, the likelihood that a full viral genome will be replicated within a singularly infected cell is low. The fitness implications of this inefficiency may be offset, however, by complementation of IVGs in multiply infected cells.

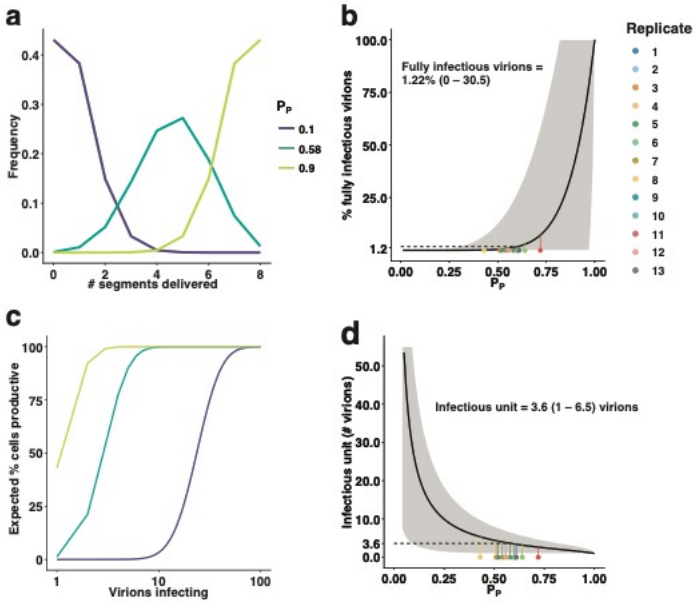


Figure 2 — Incomplete genomes require complementation for productive infection at the cellular level.

(A) The expected number of segments delivered upon infection with a single virion was calculated for two extreme values of P_P (0.10, 0.90) and the estimated P_P of Pan/99 virus (0.58, 95% C.I. 0.54 – 0.61). (B) The percentage of virions expected to initiate productive infection was plotted as a function of P_P. The 95% C.I. (mean ± 1.96 * S.D.) of this theoretical prediction is shown in gray. Colored points along the bottom of the plot correspond to the average P_P value of each experimental replicate in Fig. 1, with lines connecting them to their predicted value on the theoretical line, and therefore represent predicted frequencies for Pan/99 virus. (C) The percentage of cells expected to be productively infected following infection with a given number of virions was calculated for the same P_P values as in (A). (D) The expected number of virions required to make a cell productively infected is plotted as a function of P_P. As in (B), colored points

correspond to the average P_P value of each Pan/99 experimental replicate in Fig. 1, and the 95% C.I. (mean ± 1.96 * S.D.) is shown in gray.

Predicted costs of incomplete genomes for population infectivity

The potential for multiple infection to mitigate the costs of inefficient genome delivery will, of course, depend on the frequency of multiple infection. To evaluate the theoretical impact of IVGs on viral fitness, we therefore modeled the process of infection at a population level, where a given number of virions were delivered to a population of 10^6 cells. For viruses of different P_P values, the probability that at least one cell would receive all 8 genome segments and thus the population would become infected, was calculated over a range of MOIs (Fig. 3A). This was accomplished by calculating the probability that each cell received v virions based on the Poisson distribution $p(v)$, the conditional probability that a cell would contain 8 segments following infection with v virions, $p(8|v)$, and the sum of the joint probabilities across all values of v, $\lim_{N\to\infty} \sum_{v=1}^{N} p(v) * p(8|v)$, to give the probability that any given cell is infected at a given MOI and P_P. Multiplying this probability by the number of cells in the population gives the expected number of cells infected, and the probability of the population becoming infected is equal to this value or 1, whichever is lower. The resultant plot shows that viruses with lower P_P require markedly higher MOIs to ensure productive infection within a population of cells. Indeed, when we estimated the MOI required for a virus of a given P_P to have a 50% chance of infecting a population, we observed that the ID_{50} increases on a logarithmic scale as P_P decreases (Fig. 3B). We also estimated the ID_{50} when complementation cannot occur and infecting a population instead requires that at

least one virion deliver all 8 genome segments. This analysis revealed that ID_{50} is not affected by complementation when P_P is above 0.3, indicating that the frequency of fully infectious particles is the main determinant of infectivity under these conditions. Thus, a reliance on multiple infection in a well-mixed system is predicted to bear a substantial fitness cost, and the establishment of infection is likely driven by the rare minority of fully infectious particles. It is important to note that these calculations depend on the number of cells in the population being considered, and so the probability of establishing infection may be influenced by the size of the host respiratory tract.

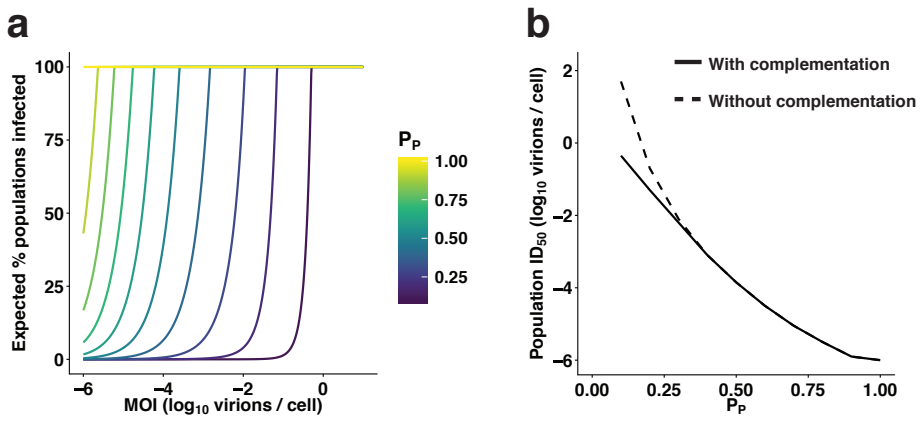


Figure 3 — Requirement for co-infection poses a barrier to establishing an infection in a population of cells.

To define the impact of IVGs on the ability of a virus to establish infection in a population of cells, the probability that a population of 10^6 cells became infected by a given number of virions was calculated. (A) The percentage chance of at least one cell containing 8 genome segments following delivery of virions was calculated for each P_P and across a range of MOIs. (B) The MOI that led to 50% of cell populations becoming infected (ID_{50}) was plotted as a function of P_P where complementation was possible (solid line) and where only complete viral genomes could initiate infection (dashed line).

Model of spatially structured viral spread

The estimates of viral infectivity made above assume that virus is distributed randomly over a population of cells. Following the initial infection event, however, viruses spread with spatial structure. We hypothesized that this structure may be very important for reducing the costs of genome incompleteness once infection is established. To test this idea, we developed a model of viral spread in which the extent of spatial structure could be varied.

The system comprises a spatially explicit grid of cells that can become infected with virus. The number and type of segments delivered upon infection is dictated by the parameter P_P and, if all eight segments are present, a cell produces virus particles. These particles can then diffuse in a random direction, with the distance traveled governed by the diffusion coefficient (D). D was varied in the model to modulate the spatial structure of viral spread: higher D corresponds to greater dispersal of virus and therefore lower spatial structure. We simulated replication of two virus strains under a range of diffusion coefficients, one with a frequency of IVGs characteristic of Pan/99 virus ($P_P = 0.58$) and one with complete genomes ($P_P = 1.0$).

Our results point to an important role for spatial structure in determining the efficiency of infection. When $P_p = 1.0$, replication proceeds faster at higher values of D, because virus particles reach permissive cells more efficiently (Fig. 4A and Supplementary Figure 2A–D). In contrast, when $P_P = 0.58$, replication proceeds fastest at intermediate values of D (Fig. 4A and Supplementary Figure 2E–H). An intermediate level of spatial structure is optimal for a virus with incomplete genomes for two reasons. At high values of D, virions diffuse farther and cellular co-infection becomes less likely, reducing the likelihood of complementation. At the other extreme, when D is very low, complementation occurs readily but spread to new cells becomes rare. Note, values of D greater than 10^3 um^2/s approximate even mixing.

The model allows the potential costs of incomplete genomes to be evaluated by comparing results obtained for a virus with $P_P = 1.0$ to those obtained for a virus with a lower P_P. In particular, we focused on $P_P = 0.58$ based on the measured values for Pan/99 virus. Thus, in Figure 4B, C and D, we evaluated three different measures of viral fitness

and plotted the relative values for a virus with $P_P = 0.58$ compared to a virus with $P_P = 1.0$. We show that reductions in the initial growth rate (Fig. 4B) and increases in the length of time taken to productively infect 100 cells or produce 10^5 virions (Fig. 4C and 4D, respectively) brought about by incomplete genomes vary with spatial structure. We see that costs of incomplete genomes are minimized at intermediate values of *D*. The greatest costs are seen at higher values of *D* as the process of virion dispersal approximates random mixing. The results of this model thus predict that the fitness effects of IVGs are dependent on the extent to which viral dispersal is spatially constrained.

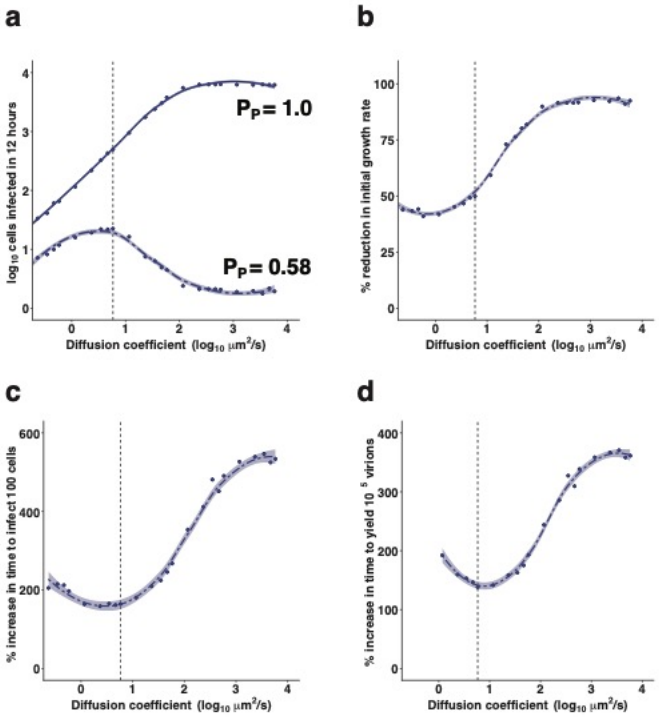


Figure 4 — The fitness costs of incomplete genomes may be mitigated by spatially structured spread.

The dynamics of multi-cycle replication in a 100x100 grid of cells were simulated, starting from a single cell in the center of the grid. (A) The initial growth rate (estimated by the log-transformed number of cells that are productively infected in the first 12 h) is shown across a range of diffusion coefficients for a virus with P_P = 1.0 (solid line) and P_P = 0.58 (dashed line). (B, C, D) The fitness cost of IVGs, as measured by the reduction in initial growth rate (B) or the increase in time taken to infect 100 cells (C) and produce 10^5 virions (D) are shown across a range of diffusion coefficients. The vertical dashed line represents the estimated value of D (5.825 um^2/s) for a spherical IAV particle in water. Each point shows the mean of 10 simulations. Curves were generated by local regression. Shading represents 95% C.I. of local regression (mean ± 1.96 * S.E.).

Impact of MOI on efficiency of virus production

Burst size, the average number of virions generated by an infected cell, is an important factor determining the potential for complementation of incomplete genomes. If an infected cell produces a larger number of viral progeny, the likelihood of coinfection in neighboring cells increases. We therefore measured this parameter experimentally for Pan/99 virus by performing single-cycle growth assays over a range of MOIs (1, 3, 6, 10, and 20 PFU/cell). Multiple MOIs were used to determine whether burst size is dependent on the number of viral genome copies per cell. We saw that higher MOIs resulted in earlier emergence of virus, suggesting that there is a kinetic benefit of additional vRNA input beyond what is required to productively infect a cell (Fig. 5A; Supplementary Figure 3). Despite these kinetic benefits, MOIs above 3 PFU/cell conferred no benefit in terms of percent infection (Fig. 5B) or total productivity (Fig. 5C). This growth analysis indicated that a maximum of 11.5 PFU per cell (95% C.I. 10.6 – 12.5 PFU) was produced during Pan/99 virus infection of MDCK cells. Based on measured P_P values, these data estimate that a single productively infected cell produces 962 virions, and this value was used as the burst size in our models.

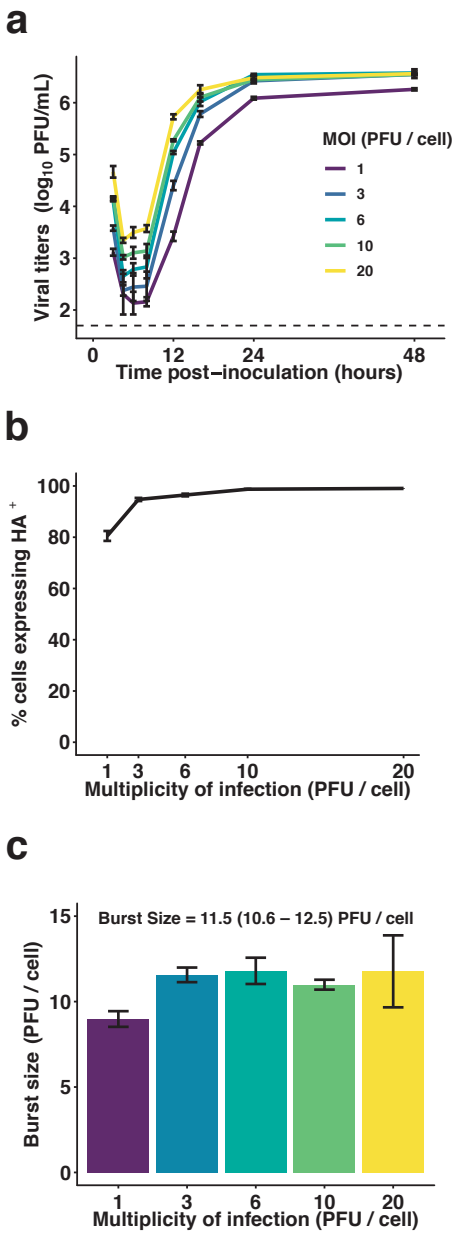


Figure 5 — Burst size of Pan/99 virus is constant over a range of high MOIs.

(A) MDCK cells were inoculated with Pan/99-WT virus at MOIs of 1, 3, 6, 10, and 20 PFU/cell under single-cycle conditions. Infectious titers at each time point are shown, with MOI indicated by the colors defined in the legend. Dashed line indicates the limit of detection (50 PFU/mL). (B) Fraction of cells expressing HA at each MOI, as measured by flow cytometry staining of cells 12 h post-inoculation. (C) Burst size in PFU produced

per HA^+ cell. In all panels, mean and standard error are plotted and colors correspond to the legend in panel A.

Impact of MOI and spatial structure on IVG complementation

Our models indicate that, for a virus of a given P_P, the frequency of infected cells containing IVGs is reduced i) at higher MOIs and ii) under conditions of high spatial structure. These predictions can be seen in Figure 6. In panel 6A we show how the proportion of infected cells that lack a complete viral genome is predicted to vary with MOI under single-cycle conditions. In panel 6B we show how this proportion varies with the structure imposed by diffusion during multi-cycle replication (Fig. 6B). We tested the predictions of these models experimentally by modulating MOI and spatial spread in IAV-infected cell cultures and gauging the impact of each manipulation on levels of IVGs.

First, to evaluate spatial structure under single and multi-cycle conditions, we used a reporter strain of Pan/99 virus with a tetracysteine tag on the NP protein (Pan/99-NP_TC virus) and visualized infected cell monolayers. The results confirmed that single-cycle inoculation results in random dispersal of virus across cells, while multi-cycle replication proceeds in a spatially structured manner resulting in foci of infection (Fig. 6C).

Next, we set up an experiment using Pan/99-WT and Pan/99-Helper viruses that lack a TC tag but carry different epitope tags fused to their HA proteins. To monitor levels of IVGs, we used flow cytometry targeting these epitope tags to measure the potential for complementation—that is, the benefit provided by the addition of Pan/99-

Helper virus. We hypothesized that, under single-cycle conditions, the potential for complementation would decrease with increasing WT virus MOI, since complementation between co-infecting WT viruses would occur frequently at high MOIs. In addition, under multicycle conditions initiated from low MOI, we predicted that the potential for complementation would be greatest at the beginning of infection, due to the random distribution of viral particles, and reduced by secondary spread. We hypothesized that the combination of local dispersal and high particle production during secondary spread would support co-infection in neighboring cells. To test our hypotheses, we inoculated cells with Pan/99-WT virus and either added Pan/99-Helper virus at the same time, or added the Helper virus after allowing time for secondary spread.

To evaluate the potential for complementation at the outset of infection and at a range of MOIs, cells were co-inoculated with Pan/99-Helper virus at a constant MOI and with Pan/99-WT virus at MOIs of 0.1, 0.3, 0.6, or 1 PFU/cell. Cells were then incubated under single-cycle conditions for 12 h to allow time for HA protein expression. Samples were processed by flow cytometry with staining for WT and Helper HA proteins (Supplementary Figure 4). In each co-infection, we quantified the benefit provided by Pan/99-Helper virus by calculating the enrichment of WT HA expression in Helper^{+} cells relative to Helper^{-} cells. Essentially, the enrichment measure works as follows. If the proportion of Helper^{+} cells that are WT^{+} is higher than the proportion of Helper^{-} cells that are WT^{+}, enrichment will be > 0%, indicating a cooperative interaction in which Helper virus allows the expression of WT HA genes present in incompletely infected cells. The results shown in Figure 6D revealed that the potential for complementation at the outset of infection was high at low MOIs, but decreased with increasing MOI. This result was as

expected, since complementation between WT virus particles was predicted to reduce the need for Helper virus (Fig. 6A).

To evaluate the impact of spatially structured secondary spread on IVG prevalence, cells were inoculated with Pan/99-WT virus at low MOI (0.002 or 0.01 PFU/cell) and then multicycle replication was allowed to proceed over a 12 h period. After this period, cells were inoculated with Pan/99-Helper virus to complement any incomplete genomes, and incubated for 12 h under single-cycle conditions to allow HA expression to occur. In contrast to the results seen when complementation was offered at the outset of infection, the enrichment of WT^+ cells in the $Helper^+$ fraction was significantly lower in these samples where multi-cycle replication occurred prior to the addition of Helper virus. This reduction in enrichment is clear when comparing infections performed under each condition in which ~50% of cells expressed WT HA (Fig. 6D). These data agree with our theoretical results (Fig. 6B) and indicate that the spatial structure of secondary spread facilitates complementation between WT particles as they infect neighboring cells at locally high MOIs.

To account for the alternative possibility that spread of the WT virus over 12 h would reduce the potential for complementation by super-infection exclusion, we analyzed the level of Helper virus infection in relation to the presence of WT virus (Supplementary Figure 4D). We observed that a similar fraction of cells expressed Helper HA in i) infections with Helper virus alone (data plotted at 0% WT HA^+), ii) simultaneous co-infections of WT and Helper viruses, and iii) infections where Helper virus was added 12 h after low MOI WT virus. If super-infection exclusion was indeed limiting the ability of Pan/99-Helper virus to complement IVGs among cells, we would

expect lower frequencies of Helper HA expression in populations where Pan/99-Helper virus was added after WT virus. Although super-infection exclusion develops in less than 12 h[8], this outcome was expected because, under the multicycle conditions used, the vast majority of infected cells present at 12 h were recently infected in a second or third round of multiplication.

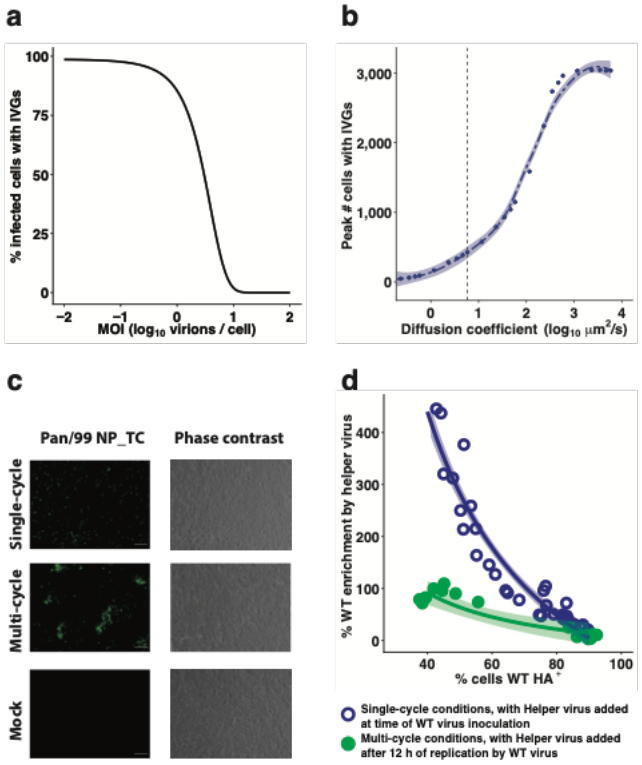


Figure 6 — Complementation of incomplete genomes occurs efficiently at high MOI and during secondary spread from low MOI.

(A) The model shown in Fig. 3 was used to calculate the probability that a given infected cell contained an IVG. The percentage of infected cells that contain fewer than 8 segments is shown at a range of MOIs for $P_P = 0.58$. (B) An infection in which multi-cycle replication occurs with spatial structure was simulated as in Fig. 4. The maximum number of cells that contain IVGs in each simulation is shown for a range of diffusion coefficients. Shading represents 95% C.I. of local regression (mean ± 1.96 * S.E.). (C) Visualization of spatially structured and unstructured infections. Cells were inoculated with Pan/99-NP_TC virus and incubated under single-cycle conditions for 12 h (top), multi-cycle conditions for 12 h followed by single-cycle conditions for 12 h (middle), or sham-inoculated and incubated under single-cycle conditions for 12 h (bottom), then

visualized by FlaSH staining (left) or phase contrast imaging (right). Scale bar represents 200 um. (D) The extent to which the presence of Pan/99-Helper virus increased WT HA positivity (% Enrichment) was evaluated at the outset of infection (open blue circles) and following secondary spread (filled green circles). To gauge potential for complementation at the outset of infection, cells were simultaneously inoculated with Pan/99-WT virus and Pan/99-Helper, then incubated under single-cycle conditions for 12 h. To test the impact of secondary spread on potential for complementation, cells were inoculated with Pan/99-WT virus at low MOI and incubated under multi-cycle conditions for 12 h, then inoculated with Pan/99-Helper virus and incubated under single-cycle conditions for 12 h. Curves represent estimates of a fixed effects model with the formula $\%Enrichment = \beta_1 \frac{Multi\text{-}cycle}{\%WT\ HA^+} + \beta_2 \frac{1}{\%WT\ HA^+} + \beta_3 * Multi\text{-}cycle$, with shading representing 95% C.I. (mean ± 1.96 * S.E.) of model estimate. IVGs = incomplete viral genomes.

A virus with absolute dependence on multiple infection

To evaluate the potential for complementation *in vivo*, we generated a virus that is fully dependent on complementation for replication. This was accomplished by modifying the M segment to generate one M segment which encoded only M1 (M1.Only) and a second one which encoded only M2 (M2.Only) (Fig. 7A). When combined with seven standard reverse genetics plasmids for the remaining viral gene segments, the plasmids encoding these two M segments allowed the generation of a virus population in which individual viruses encode functional M1 or M2, but not both. We called this virus Pan/99-M.STOP virus. Due to the rarity of recombination within segments in negative-

sense RNA viruses[32], it is unexpected that M1.Only and M2.Only segments will recombine to generate a WT M segment. Hence, this virus is reliant on both M segments being delivered to the same cell by co-infection. It is important to note that, in contrast to the more arbitrary multiplicity dependence of a wild type IAV, the complementation needed by Pan/99-M.STOP virus requires co-infection with two viruses of a particular genotype.

To characterize the Pan/99-M.STOP virus genetically, we used digital droplet PCR (ddPCR) to measure copy numbers of the two M segments and the NS segment in three virus samples (Fig. 7B). The total M segment copy number was found to comprise 30% M2.Only and 70% M1.Only. In addition, the total number of M segments was similar to the number of NS segments, as expected if each virion packages one NS and one M vRNA (Fig. 7B). To verify that M1.Only and M2.Only M segments were packaged into distinct virions, we performed infections of MDCK cells with serial dilutions of Pan/99-M.STOP virus under single-cycle conditions and analyzed expression of M1 and M2 by flow cytometry. We observed that, as dilution increased, cells expressing M1 were less likely to express M2, and vice versa (Fig. 7C). This result would be expected if expression of both proteins from the same cell required co-infection with M1.Only and M2.Only encoding virions. As a control, we monitored the effect of dilution on co-expression of HA and M1 or M2. Here, we found that co-expression of M1 or M2 and HA was much less sensitive to dilution, consistent with co-delivery of M and HA segments by single virions. At limiting dilutions, where only 2% of cells were infected, only one sample reached the point of absolutely no co-expression between M1 and M2, suggesting that stocks may contain some aggregates of virus particles

comprising fully infectious units. We cannot fully exclude this possibility, but note that co-expressing cells represented less than 10% of infected cells at this limiting dilution, suggesting that most co-expression of M1 and M2 at higher concentrations of virus is indeed mediated by co-infection.

M2 expression was shown previously to be non-essential for replication in cell culture[33,34]. To evaluate the extent to which Pan/99 virus relies on M2 expression for viral multiplication in the systems used here, we generated a virus unable to express M2 using the plasmid encoding the M1.Only segment. This virus was successfully recovered and formed small plaques in MDCK cells. Multiple attempts to culture the M1.Only virus in MDCK cells and eggs failed, however, indicating the importance of M2 to viral propagation in these substrates (data not shown).

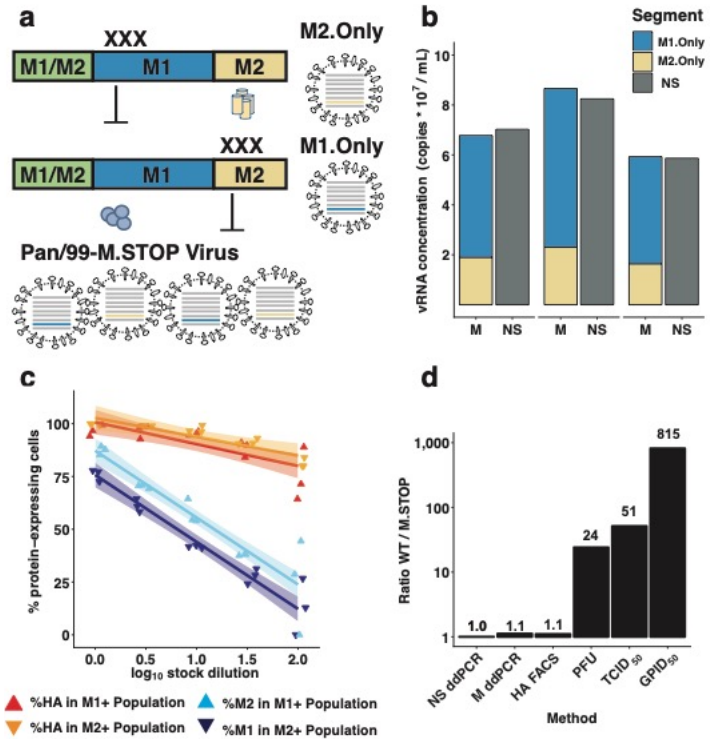


Figure 7 — Dependence on complementation hinders viral infectivity.

(A) Mutation scheme used to generate M1.Only and M2.Only segments, and Pan/99-M.STOP virus. (B) Copies of M1.Only, M2.Only, and NS segments in three separate aliquots of Pan/99-M.STOP virus stock were quantified by digital droplet PCR. (C) Cells were inoculated with Pan/99-M.STOP virus and incubated under single-cycle conditions before staining for HA, M1, and M2 expression. The percentage of cells expressing M1, M2, and HA within $M1^+$ or $M2^+$ subpopulations is shown at each dilution. Lines represent linear regression with shading representing 95% C.I. (mean ± 1.96 * S.E.). (D) Titers of WT and M.STOP virus stocks were quantified by ddPCR targeting the NS segment, ddPCR targeting (any) M segment, immunotitration by flow cytometry, plaque assay, tissue culture ID_{50}, and guinea pig ID_{50}. All results are normalized to the ratio of NS ddPCR copy numbers.

Consequences for infectivity of reliance on multiple infection

To test the hypothesis that a given number of Pan/99-M.STOP virus particles would be less infectious than a comparable number of Pan/99-WT virus particles, we characterized both viruses using a series of titration methods that vary in their dependence on infectivity and M protein expression. We first used ddPCR to quantify NS copy numbers of the WT and M.STOP viruses and then normalized all other comparisons to this ratio to account for the difference in virus concentration. As shown above, total M copy numbers were roughly equivalent when normalized to NS (Fig. 7B and 7D). Using immunotitration, in which cells are infected under single-cycle conditions with serial dilutions of virus and then stained for HA expression[35], we observed equivalent titers of both viruses (Fig. 7D). This was expected, as HA expression under single-cycle conditions is not dependent on M1 or M2 proteins. When titration relied upon multi-cycle replication, however, the WT virus was higher titer than the M.STOP virus. This difference was moderate in cell culture-based measurements, with PFU and $TCID_{50}$ titers 24- and 51-fold higher, respectively, likely because of the reduced importance of M2 in this environment. The full cost for infectivity of separating the M1 and M2 ORFs onto distinct segments was apparent *in vivo*, where 815 times as much M.STOP virus was required to infect 50% of guinea pigs compared to WT virus (Fig. 7D). Thus, although the M.STOP virus differs from a virus with very low P_P in that complementation can only occur when viruses carrying M1.Only and M2.Only segments co-infect, the prediction shown in Figure 3, that increased dependence on multiple infection decreases infectivity, held true in this system.

Potential for complementation in vivo

Having determined that the dependence of Pan/99-M.STOP virus on complementation impairs viral infectivity, we next sought to evaluate the potential for complementation to occur *in vivo* once infection had been established. Guinea pigs were inoculated intranasally with equivalent doses of Pan/99-WT or Pan/99-M.STOP virus in terms of NS vRNA copies. Specifically, a dose of 10^7 copies per guinea pig was used to ensure successful Pan/99-M.STOP virus infection in all animals. This dose represents 8 x $GPID_{50}$ of this mutant virus and $6.5x10^3$ x $GPID_{50}$ of the WT virus. Despite its reduced ability to establish infection, Pan/99-M.STOP virus successfully grew in guinea pigs, following similar kinetics to Pan/99-WT virus. Average peak virus production, measured as NS vRNA copies, was reduced by only 9-fold relative to WT (Fig. 8A). Sanger sequencing of viral cDNA from nasal washes confirmed that both M1.Only and M2.Only segments were present *in vivo*, and quantification of the the two alleles by ddPCR revealed a bias towards M1.Only segments (Supplementary Figure 5). Because the inoculum of WT and M.STOP viruses comprised the same dose in terms of vRNA copies, but different doses in terms of $GPID_{50}$, we also inoculated guinea pigs with WT virus at doses of 8 x $GPID_{50}$ and $6.5x10^3$ x $GPID_{50}$ (Fig. 8B). This experiment was designed to define the contribution of the effective dose to the differences observed between M.STOP and WT viruses in guinea pigs. Similar peak titers and kinetics of shedding were observed in both groups of WT virus infected guinea pigs, indicating minimal dose dependency (Fig. 8B).

Finally, we conducted an experiment to determine whether Pan/99-M.STOP virus was competent to undergo transmission to new hosts. In this case, guinea pigs were inoculated with equivalent doses in terms of $GPID_{50}$ with the goal of establishing

comparable infections in the donor hosts so that relative efficiency of transmission could be better evaluated. Thus, doses of 8 x $GPID_{50}$ of WT or M.STOP virus were used. At 24 h post-inoculation, each index guinea pig was co-housed with a naïve partner. As expected, WT virus transmitted to and initiated robust infection in each of the four contact animals. In contrast, only transient, low levels of the M.STOP virus were observed in nasal washings collected from contacts (Fig. 8C). These results suggest that the spatial structure inherent to multi-cycle replication mitigates the cost of incomplete genomes in an individual host, but that dependence on complementation is costly for transmission. This result is in line with the theoretical predictions displayed in Figures 2 - 4.

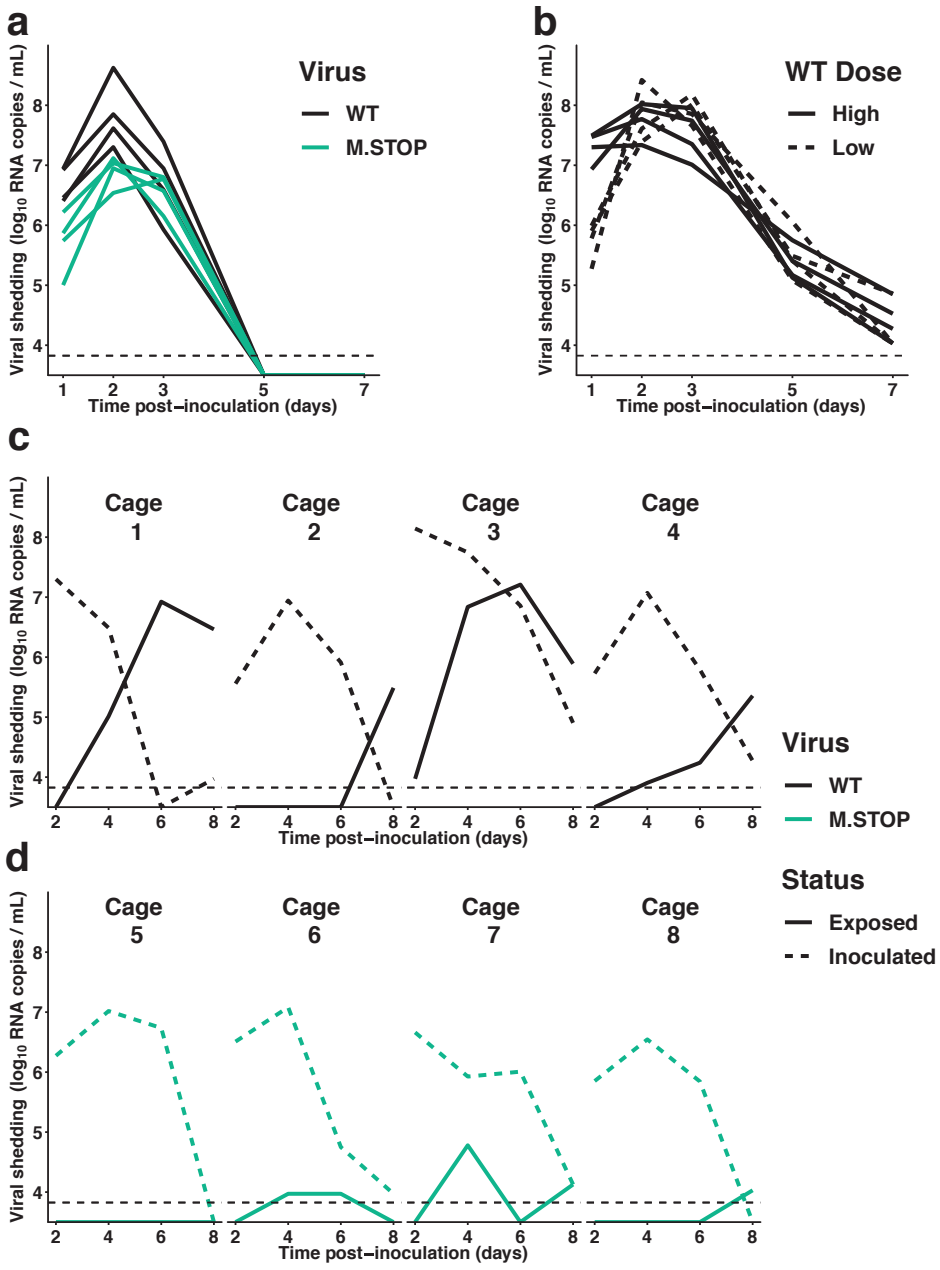


Figure 8 – Dependence on complementation hinders viral transmission but has a more modest effect on replication.

(A) Guinea pigs were inoculated with 10^7 RNA copies of Pan/99-WT virus or Pan/99-M.STOP virus, and nasal washes were collected over 7 days to monitor shedding. NS segment copy number per mL of nasal lavage fluid is plotted. (B) Guinea pigs were inoculated with 10^7 or $1.23x10^4$ RNA copies of Pan/99-WT virus, corresponding to

$6.5x10^3$ and 8 $GPID_{50}$, respectively. Nasal washes were collected over 7 days to monitor shedding. NS segment copy number per mL of nasal lavage fluid is plotted for high (solid lines) and low (dashed lines) doses. (C, D) Guinea pigs were inoculated with 8 x $GPID_{50}$ of Pan/99-WT virus (C) or Pan/99-M.STOP virus (D) and co-housed with uninfected partners after 24 h. Nasal washes were collected over the course of 8 days to monitor shedding kinetics and transmission between cagemates. NS segment copy number per mL of nasal lavage fluid is plotted. In all plots, horizontal dotted lines represents the limit of detection (6696 RNA copies/mL).

3.4 Discussion

Using a novel single-cell approach that enables robust detection of incomplete IAV genomes, we show that ~99% of Pan/99 virus infections led to replication of fewer than eight segments. The theoretical models we describe predict that the existence of IVGs presents a need for cellular co-infection, and that this need has a high probability of being met when spread occurs in a spatially structured manner. Use of silent genetic tags allowed us to experimentally interrogate cooperation at the cellular level to test these predictions. In agreement with our models, experiments in cell culture showed that co-infection and complementation occur readily when multiple rounds of infection are allowed to proceed with spatial structure. The high potential for complementation to occur *in vivo* was furthermore revealed by the robust within-host spread of a virus that is fully dependent on co-infection. Complementation was not observed during transmission, however, suggesting that fully infectious particles may be required to initiate infection in a new host.

The existence of incomplete genomes was previously predicted by Heldt et al., and these predictions are consistent with the experimental findings of our single-cell assay[21]. The parameter estimated by this assay, P_P, is defined as the probability that, following infection with a single virion, a given genome segment is successfully replicated. Previous work by Brooke et al. has shown that cells infected at low MOIs express only a subset of viral proteins[20]. While this failure of protein expression could be explained by a failure in transcription or translation, the results of our single-cell sorting assay indicate that the vRNA segments themselves are absent, as they should be amplified by the helper virus polymerase even if they do not encode functional proteins. As in Brooke et al., our method does not discriminate between the alternative possibilities that segments are absent from virions themselves or are lost within the cell, but published results suggest that a single virion usually contains a full genome[24,25]. Importantly, our single cell assay quantifies the frequencies of all eight segments, rather than only those that can be detected indirectly by staining for protein expression, and therefore allows for analysis of the associations between segments. Despite the importance of interactions among vRNP segments during virion assembly[10,28-31], we did not detect compelling evidence of segment co-occurrence at the level of vRNA replication within target cells. This observation suggests that interactions among segments formed during assembly are likely not maintained throughout the early stages the viral life cycle. While our measured P_P values are comparable for all segments, independent delivery of segments with distinct probabilities could serve as a mechanism to control gene dosage, similar to what has been observed in multipartite viruses[36]. Indeed, experiments with Pan/99-M.STOP virus showed a consistent bias towards M1-encoding segments *in vivo* when expression

of M1 and M2 was decoupled. This bias may have been important for maintaining low M2 protein levels, which we recently found to be an important predictor of viral fitness[37].

The results of our single cell assay indicate that 1.22% (95% C.I. 0% – 30.5%) of Pan/99 virions are fully infectious, which is consistent with our prior estimates based on observed levels of reassortment between Pan/99 wild type and variant viruses[26]. If these 1.22% of virions comprise the plaque-forming units present in a virus population, then the total number of virions present in a population is equal to to 82 (1 / 0.0122) times the number of plaque-forming units. This result is, however, lower than other reported estimates of the frequency of fully infectious particles. This difference may stem from our use of a helper virus, which likely allows more robust detection of IVGs than would be expected in a system dependent on the detection of non-replicating viral genomes or their mRNA transcripts[21,22]. In addition, quantitative differences among published reports may relate to use of different virus strains, as Brooke et al. observed that the frequency of IVGs is strain-specific[20]. Interestingly, the strain-specificity of P_P is likely to influence the relative representation of genome segments when two strains of varying P_P reassort during co-infection. As lower P_P values result in successful replication of fewer segments per cell, the genomes of reassortant progeny are likely to contain more segments from the virus with higher P_P.

Replication and secondary spread in an individual host involves inherent spatial structure, as virions emerge from an infected cell and travel some distance before infecting a new cell[38-40]. Localized spread is predicted to affect viral population dynamics in multiple ways, including by facilitating abundant co-infection, which we explored herein[39]. Our theoretical model predicts that local co-infection resulting from this spatial

structure mitigates the fitness costs of incomplete genomes, but that there is a trade-off between complementation and dispersal. Handel et al. explore a similar trade-off related to attachment rates in well-mixed (unstructured) populations, and find that an intermediate level of "stickiness" is optimal—virions that bind too tightly are slow to leave the cell that produced them, while those bind too weakly are unable to infect new cells[41]. We observe a similar effect with spatial structure: virions that diffuse faster, and hence disperse farther before infecting a new cell, are less likely to co-infect with enough virions to establish a productive infection. By contrast, when virions diffuse more locally, co-infection occurs more frequently than is required for productive infection and virions take longer to physically reach new cells, ultimately limiting spread. The optimal level of spatial structure for a virus with incomplete genomes is thus an intermediate one that allows a population of virions to efficiently reach new cells while ensuring enough complementation to minimize the frequency of incomplete infections.

In quantitative terms, our model predicts that a diffusion coefficient characteristic of a sphere of 100 nm diameter in water would give a near-optimal level of spatial structure. While this condition may approximate conditions for spherical virions in cell culture, the extracellular environment experienced by a virus *in vivo* would be different. Namely, virus replicating within the respiratory tract would be released into a layer of watery periciliary fluid, which underlies a more viscous mucous blanket[42]. The structure and composition of this epithelial lining fluid may act to limit dispersal the of virus particles relative to that expected in cell culture. Importantly, however, this fluid lining the airways is not static, but rather is moved in a directional manner by coordinated ciliary action[42]. This coordinated movement raises the interesting possibility that IAVs

may have evolved to depend upon ciliary action to mediate directional dispersal of virions to new target cells, while maintaining a high potential for complementation of IVGs. This concept will be explored in subsequent studies.

Our experiments designed to test the predicted role of spatial structure in enabling complementation confirmed that secondary spread allows Pan/99-WT virus to replicate efficiently even at low initial MOIs, diminishing the need for complementation after only 12 hours of multi-cycle replication. The potential for complementation of IVGs *in vivo* was furthermore evidenced by the replication in guinea pigs of Pan/99-M.STOP virus, which requires co-infection for productive infection. Importantly, however, Pan/99-M.STOP virus did not initiate productive infection in exposed cagemates. In interpreting this result, it is important to note that the complementation needed by Pan/99-M.STOP virus requires co-infection with two viruses of a particular genotype. This type of complementation has a lower probability of occurring than that typically needed for completion of a WT IAV genome. Despite this caveat, the failure of Pan/99 M.STOP virus to transmit suggests that the establishment of IAV infection requires at least some fully infectious virions. The delivery of multiple particles to a small area via droplet transmission may allow multiple virions to infect the same cell and establish infection, but our data suggest that this mechanism does not occur efficiently in a guinea pig model. The tight genetic bottleneck observed in human-human transmission events is furthermore consistent with a model in which infection is commonly initiated by single particles[43]. In prior work, mutations decreasing the frequency of fully infectious particles, but not eliminating them entirely, were observed to increase transmissibility[27]. This enhanced transmission was attributed to modulation of the HA:NA balance, which

enhanced growth in the respiratory tract. In contrast, the HA:NA balance of the Pan/99-M.STOP virus evaluated herein is not expected to differ from Pan/99 virus.

In summary, our findings suggest that incomplete genomes are a prominent feature of IAV infection. These semi-infectious particles are less able to initiate infections in cell culture and during transmission to new hosts, when virions are randomly distributed. In contrast, our data show that incomplete genomes actively participate in the within-host dynamics of infection. Within a host, IVGs are complemented by cellular co-infection, suggesting an important role for spatial structure in viral spread. This frequent co-infection leads to higher gene copy numbers at the cellular level, consequently promoting reassortment and free mixing of genes. Thus, a reliance of IAVs on co-infection may have important implications for viral adaptation to novel environments such as new hosts following cross-species transmission.

3.5 Computational Methods

Quantification of P_P values

A single-cell sorting assay was used to measure P_P, the probability that an individual genome segment from an infectious virion is successfully delivered and replicated within the infected cell. The technical details are described in the "Experimental Methods" section. Here we describe the mathematical analysis used to calculate P_P from the experimental data. The presence or absence of different viral genome segments was measured by qRT-PCR, with each well of a 96-well plate representing the viral RNA that was present in the cell that was initially sorted into the plate.

Given the MOI of Pan/99-WT virus used in the experiments, an appreciable number of wells are expected to receive two or more viral genomes, and so a mathematical adjustment is needed to estimate the probability of each genome segment being delivered by a signle virion. Using the relationship between MOI and the fraction of cells infected from Poisson statistics, i.e., $f = 1 - e^{-\mathrm{MOI}}$, the probability of the ith segment being present in a singly infected cell, or $P_{P,i}$ can be calculated from the 96-well plate using the following equation:

$$P_{P,i} = \frac{\mathrm{MOI}_i}{\mathrm{MOI}_{\mathrm{wt}}} = \frac{-\ln\,(1 - f_i)}{-\ln\,(1 - f_{\mathrm{wt}})} = \frac{\ln\,(1 - \frac{C_i}{A})}{\ln\,(1 - \frac{B}{A})}$$

where A is the number of Helper$^+$ wells, B is the number of WT$^+$ wells (containing any WT segment), and C_i is the number of wells positive for the WT segment in question. Wells that were negative for Helper virus segments were excluded from analysis. For each experimental replicate, the geometric mean P_P value was calculated to represent an average P_P.

Computational simulation to predict reassortment frequency

To predict reassortment frequency expected given the P_P values measured in the single-cell sorting assay, the model described by Fonville et al.[26] was used. Briefly, this model simulates infection of a population of cells with a 1:1 mixture of WT and var viruses across a range of MOI. When a virion infects a cell, it can deliver a single copy of each segment with probability P_P. In this model, P_P probabilities are segment-specific but identical for WT and var viruses. A cell is deemed "HA$^+$" if it contains at least one copy of each of the PB2, PB1, PA, HA, and NP segments, and "Productive" if it contains at

least one copy of each segment. The probability of a progeny virion being reassortant is calculated for each cell as

$\mathrm{p(Reassortant\ virion)} = 1$ -p(WT virion)-p(var virion), where

$p(\mathrm{WT\ virion}) = \frac{\#\ WT\ PB2\ copies}{\#\ Total\ PB2\ copies} * \frac{\#\ WT\ PB1\ copies}{\#\ Total\ PB1\ copies} * \ldots \frac{\#\ WT\ NS\ copies}{\#\ Total\ NS\ copies}$, and similar for *p*(var virion). The predicted percentage of reassortant virions resulting from a co-infection (% Reassortment) is the average of *p*(Reassortant virion) values among Productive cells. To generate the predictions shown herein, each simulation used a set of eight $P_{P,i}$ values measured in a single experimental replicate. Thus, a total of 13 separate predictions were made.

Analysis of pairwise associations between segments

To calculate the pairwise associations between segments, we first defined the probability that each cell acquired its combination of segments by infection with a single virion (*p*(1 virion | segment combination)), as follows. The results of this calculation are shown in Supplementary Figure 1.

$$p(1\ \mathrm{virion} \mid \mathrm{segment\ combination}) = \frac{p(\mathrm{segment\ combination} \mid 1\ \mathrm{virion}) * p(1\ \mathrm{virion})}{p(\mathrm{segment\ combination})}$$

The probability that one virion infected a given cell, p(1 virion), was calculated using the Poisson distribution, with $\lambda = -\ln(1 - \frac{\mathrm{WT^{+}Cells}}{\mathrm{Helper^{+}Cells}})$. The probability of a given segment combination arising following entry of *v* virions was calculated under the assumption of independent delivery of genome segments as

$$p(\mathrm{segment\ combination} \mid v\ \mathrm{virions}) = \prod_{i=0}^{8} \begin{cases} 1 - (1 - P_{P,i})^{v} & \text{if segment } i \text{ present} \\ (1 - P_{P,i})^{v} & \text{if segment } i \text{ absent} \end{cases}.$$

Using the estimated probability of segment combinations arising from infection with v virions and the frequency of infection by v virions, the overall probability of a cell containing each segment combination was calculated as $p(\text{segment combination}) = \sum_{v=0}^{\infty} p(\text{segment combination} \mid v \text{ virions}) * p(v \text{ virions})$. Because the number of cells infected and the measured P_P values varied between experimental replicates, each replicate was analyzed independently to calculate P_P, p(segment combination | v virions), and p(1 virion | segment combination) for each cell, and all experiments were pooled for the final analysis of segment associations. Using p(1 virion | segment combination) as a weighting factor for each cell, the pairwise correlations between WT segments were then calculated. Significant associations were defined as those with $p < 0.05$ after Bonferroni correction for multiple comparisons, where Nr^2 follows a χ^2 distribution with 3 degrees of freedom (N = sum of all p(1 virion | segment combination) values).

Probabilistic model to estimate costs of IVGs for cellular infectivity

To define the impact of incomplete viral genomes on viral infectivity, we first considered how specific infectivity, the ratio of plaque-forming units to virus particles, changes with P_P. The proportion of virions that can form plaques, or the probability of productive infection resulting from a single virion, was estimated as $p = P_P^8$. This is a Bernoulli process with a defined probability of success or failure, so the variance was estimated as $\sigma^2 = p(1 - p) = P_P^8(1 - P_P^8)$, and the 95% confidence interval given by $P_P^8 \pm 1.96 * \sigma$.

We next considered how the infectious unit (the average number of virions required to result in productive infection of a cell) varies with P_P. This model assumes that a single virion can deliver one copy of each segment to a cell with probability P_P and

that delivery of each segment is independent. The act of segment delivery by a single virion is therefore modeled as a binomial process where $p = P_P$, N = # of missing segments, and k = # of segments added by one virion. For an uninfected cell, $N = 8$ missing segments, and

$\mathrm{p}(k \text{ segments delivered}) = \binom{8}{k}({P_P}^k(1 - P_P)^{8-k})$. For a cell that has already received some segments, each successive virion can deliver one copy of each segment not already present. We model this process of infection by successive virions as a Markov chain in which a cell can exist in 9 states, containing between 0 and 8 genome segments. Transitions between states are governed by the 9x9 matrix **T**, in which each element is described by the binomial distribution:

$$\mathbf{T}_{i,j} = \binom{N}{k}(p^k(1-p)^{N-k}) = \binom{8-i}{j-i}({P_P}^{j-i}(1-P_P)^{8-j})$$

where i is the number of segments a cell contains before infection, and j is the number of segments it contains after infection. Since the binomial distribution is not defined for $k < 0$, all entries below the main diagonal are populated by 0s. The state of 8 segments, or productive infection, becomes the absorbing state, and it is assumed that each cell will obtain all 8 genome segments given the addition of enough virions.

To estimate how many virions are required to reach the state of productive infection, we first define a 1x9 vector representing the distribution of segments per cell. Each value in this vector gives the probability that a cell contains a given number of segments. To represent an uninfected cell, we set $\boldsymbol{\tau_0}$ = [1,0...0] to indicate that the probability that 0 segments are present is 1, and the probability that 1-8 segments are present is 0. The distribution of segments in a cell that has been infected with v virions is then given by:

$$\boldsymbol{\tau}_v = \boldsymbol{\tau}_0 * \mathbf{T}^v$$

With the 9th element of $\boldsymbol{\tau}_v$ representing the probability a cell contains 8 segments and is therefore productively infected.

Finally, we use survival analysis to calculate the expected number of virions that must infect a cell before it receives all 8 segments. We first define $\mathbf{T_{sub}}$ to represent the upper-left 8x8 matrix of **T** (in which a cell contains 0 – 7 segments), $\boldsymbol{\tau}_{\mathbf{sub}}$ as the first 8 columns of $\boldsymbol{\tau}_{\mathbf{0}}$, and $\mathbf{1_{sum}}$ as an 8x1 vector where each value is 1, which acts to sum each state into a single value. For a cell that starts with 0 segments, $\boldsymbol{\tau}_{\mathbf{sub}}$ = [1,0…0]. The probability distribution of a cell containing 0 – 7 segments is given by $\boldsymbol{\tau}_{\mathbf{sub}} * \mathbf{I}$ (where *I* is the identity matrix) for an uninfected cell, $\boldsymbol{\tau}_{\mathbf{sub}} * \mathbf{T}_{\mathbf{sub}}$ for 1 virion, $\boldsymbol{\tau}_{\mathbf{sub}} * \mathbf{T}_{\mathbf{sub}}^2$ for 2 virions, and so on. For an arbitrary number of virions (v), the distribution is given by $\boldsymbol{\tau}_{\mathbf{sub}} * \mathbf{T}_{\mathbf{sub}}^v$. The total probability that a cell contains 0 – 7 segments is then calculated as $\boldsymbol{\tau}_{\mathbf{sub}} * \mathbf{T}_{\mathbf{sub}}^v * \mathbf{1}_{\mathbf{sum}}$. As more virions infect a cell, this probability converges to 0. We therefore estimate the number of virions required to fully infect a cell using the equation:

$$\mathrm{E}(v) = \boldsymbol{\tau}_{\mathbf{sub}} * \lim_{n\to\infty}(\mathbf{I} + \mathbf{T}_{\mathbf{sub}}^1 + \mathbf{T}_{\mathbf{sub}}^2 + \cdots \mathbf{T}_{\mathbf{sub}}^n) * \mathbf{1}_{\mathbf{sum}} = \boldsymbol{\tau}_{\mathbf{sub}} * (\mathbf{I} - \mathbf{T}_{\mathbf{sub}})^{-1} * \mathbf{1}_{\mathbf{sum}}$$

This summary statistic represents the number of transitions required for a cell to reach the absorbing state, or more simply, the average number of virions required to infect a cell. The variance on this quantity can be calculated as $\sigma^2 = (2 * (\mathbf{I} - \mathbf{T}_{\mathbf{sub}})^{-1} - \mathbf{I}) * \boldsymbol{\tau}_{\mathbf{sub}} - \boldsymbol{\tau}_{\mathbf{sub,sq}}$, where each element of $\boldsymbol{\tau}_{\mathbf{sub}}$ is squared to generate $\boldsymbol{\tau}_{\mathbf{sub,sq}}$. The 95% confidence interval of this estimate is then given by $E(v) \pm 1.96 * \sigma$. A more detailed proof of these derivations can be found in Finite Markov Chains[44].

Probabilistic model to estimate costs of IVGs for population infectivity

To define the impact of incomplete viral genomes on the ability of a virus population to establish infection in a population of cells, the probabilistic model described above was adapted to account for the Poisson distribution of virions among a population of cells. In order for an infection to be established, at least one cell must receive all 8 genome segments. For a given MOI, the probability of a cell receiving v virions follows the Poisson distribution $p(v) = \frac{MOI^v e^{-MOI}}{v!}$. At each v, the probability that a cell received any given segment is equal to $1 - (1 - P_P)^v$, and so the probability that a cell is productively infected after infection with v virions is $p(8|v) = (1 - (1- P_P)^v)^8$. Iterating across values of v, the probability that any given cell is productively infected is $\lim_{N\to\infty} \sum_{v=1}^{N} p(v) * p(8|v)$. Multiplying this probability by the number of cells under consideration gives the expected probability that a population becomes infected following delivery of virions at a given MOI. The ID_{50} was estimated as the lowest MOI yielding a probability ≥ 50%. A similar analysis was used to estimate the ID_{50} when complementation was not allowed, with the $p(8|v)$ function being modified to $p(8|v) = 1 - (1 - P_P^8)^v)$ to reflect the fact that only complete viral genomes could initiate infection. Finally, the percentage of infected cells that contained incomplete viral genomes was calculated by estimating the probability that a cell infected by v virions contained between 1 and 7 segments, $p(1–7|v) = 1 - (1 - (1 - P_P)^v)^8$, and determining the total proportion of infected cells containing IVGs using the equation $\% \, Cells \; with \; IVGs = \lim_{N\to\infty} \sum_{v=1}^{N} p(v) * \frac{p(1-7|v)}{p(1-7|v) + p(8|v)} * 100$.

Individual-based model of replication

A cellular automaton model of viral spread was developed to investigate the relationships between spatial structure, prevalence of incomplete viral genomes, and viral

fitness. The system consists of a 100 x 100 grid of cells. Each cell contains 0 – 8 distinct IAV genome segments, and additional copies of the same segment are assumed to be redundant. Virions exist on the same grid, in a bound or unbound state. When a virion infects a cell, any missing segments may be delivered, with the probability of delivery defined by P_P, as derived in Figures 1–2.

The simulation begins with a single productively infected cell in the middle of the grid. The following events occur at each time-step (3 minutes), and the frequency of each of these events is governed by the parameters listed in Table 1.

1) All virions not currently bound to a cell will diffuse. First, the total distance traveled is randomly drawn from the normal distribution $Distance_{\text{Total}} = N(\mu = 0, \sigma = \sqrt{2Dt})$, where D is the diffusion coefficient, t is the length of the time-step, μ is the mean and σ is the standard deviation. Second, the direction traveled is randomly drawn from the uniform distribution $\theta = U(0, 2\pi)$. The total distance traveled is then converted to orthogonal distances $Distance_{\text{X}} = \cos(\theta) * Distance_{\text{Total}}$ and $Distance_{\text{Y}} = \sin(\theta) * Distance_{\text{Total}}$. Distance and direction of travel are calculated independently for each virion. These distances are then used to adjust the X and Y positions of the diffusing virions, which are tracked in absolute units (um). Cells are modeled as squares measuring 30 um x 30 um, and so a virion with an absolute position $0 \leq X < 30$ corresponds to cell position X = 1, one with absolute position $30 \leq X < 60$ corresponds to cell position X = 2, and so on. If a virion would diffuse beyond the border of the grid, it instead emerges from the other side (e.g. a virion that would be moved to the cell position [X = 105, Y = 69] is instead placed at the position [X = 5, Y = 69]).

2) Free virions may attach to the cell at their current position. While attached, virions are unable to diffuse.
3) Bound virions may be released, or infect the cell to which they are attached. Virions that are released become free-floating and are able to diffuse. When virions infect a cell, the number of segments added to that cell is determined by the probabilistic model described above. The number of segments added is calculated from the binomial distribution $B(N = 8 - S, p = 1 - (1 - P_P)^v)$ where S is the number of segments the cell already contains, P_P is the probability an individual segment is delivered successfully, and v is the number of virions infecting the cell at the current time-step. For example, if a cell already contains 5 segments, additional virions infecting it may add 0 – 3 segments ($N = 8 – 5 = 3$). If 4 virions are infecting this cell, then the probability of a given segment being delivered is high ($p = 1 – (1 – 0.58)^4 = 0.97$), and the probability of four virions all failing to deliver a given segment ($(1 – 0.58)^4 = 0.03$) is relatively low.
4) Infected cells (those containing 1 – 8 segments) may become refractory to super-infection. Diffusing virions cannot bind to non-susceptible cells, and any currently bound virions that attempt to infect these cells automatically fail to deliver all segments.
5) Productively infected cells (containing 8 segments) produce virions, which are initially bound to the producer cell's surface. The number of virions produced by each cell is independently drawn from the Poisson distribution with $\lambda =$ Burst_Rate (962 virions / day or 2.06 / 3-minute time-step). Each virion is placed

at a random location on the producer cell as determined by the uniform distribution.

6) Productively infected cells may die. These cells lose all segments, cannot produce virions, and cannot be bound by diffusing virions.

At each time-step, the number of virions that bind, release, or infect, and the number of cells that die or become refractory to super-infection, are calculated using the Poisson distribution with λ = Rate * N, where Rates are described in Table 1, and N represents the number of virions or infected cells present at that time.

To generate the data shown in Figure 4 and Figure 6B, these events were iterated over multiple rounds of infection up to 96 h post-infection. Ten simulations per (D, P_P) combination were conducted.

3.6 Experimental Methods

Cells

Madin-Darby canine kidney (MDCK) cells (contributed by Peter Palese, Icahn School of Medicine at Mount Sinai) were cultured in minimal essential medium (MEM) supplemented with 10% fetal calf serum (FCS), penicillin (100 IU), and streptomycin (100 ug/mL). 293T cells (ATCC, CRL-3216) were cultured in Dulbecco's modified essential medium (DMEM) supplemented with 10% FCS.

As used herein, "complete medium" refers to MEM supplemented with 10% FCS and penicillin/streptomycin at the above concentrations, which was used for maintaining cells in culture. Following infection with influenza viruses, cells were incubated with "virus medium", which herein refers to MEM supplemented with 0.3% bovine serum

albumin and penicillin/streptomycin at said concentrations. When their presence is indicated, TPCK-treated trypsin was used at 1 ug/mL, NH_4Cl at 20 mM, HEPES at 50 mM.

Viruses

All viruses were generated by reverse genetics following modification of the influenza A/Panama/2007/99 (H3N2) virus cDNA, which was cloned into pDP2002[45]. All viruses were cultured in 9–11 day old embryonated hens' eggs unless otherwise noted below. To limit propagation of defective interfering viral genomes, virus stocks were generated either from a plaque isolate or directly from 293T cells transfected with reverse genetics plasmids. Defective interfering particle contents were quantified using a digital droplet PCR (ddPCR) assay and confirmed to be minimal[46]. The only genetic modification made to the Pan/99-WT virus was the addition of sequence encoding a 6-His tag plus GGGGS linker following the signal peptide of the HA protein as previously described[8]. A genetically distinct but phenotypically similar virus, referred to herein as "Pan/99-Helper", was generated by the introduction of six or seven silent mutations on each segment, as well as the addition of the HA-tag (sequence: YPYDVPDYA) instead of the 6-His tag. The silent mutations are listed in Supplementary Table 1 and were designed to introduce strain-specific primer binding sites, allowing the presence or absence of each segment to be measured by qRT-PCR. Epitope tags in HA allowed identification of infected cells by flow cytometry.

To visualize infected cells by microscopy, a virus which expresses a tetracysteine (TC) tag on the NP protein, referred to herein as "Pan/99-NP_TC", was generated by the introduction of the amino acid sequence CCPGCC at the C-terminus of NP. To avoid

disruption of the packaging of the NP segment, 150 nt corresponding to the 3' end of the NP ORF was duplicated following the TC tag sequence and the stop codon.

A virus with two distinct forms of the M segment, referred to herein as "Pan/99-M.STOP virus", was constructed. Site-directed mutagenesis was used to introduce nonsense mutations into the pDP2002 plasmid containing the sequence of the M segment in order to abrogate expression of M2 but not M1 (M1.Only), or vice versa (M2.Only). For M2.Only, three in-frame stop codons were introduced downstream of the sequence encoding the shared M1/M2 N-terminus (the nucleotide changes introduced were T55A, C75G and A86T). An in-frame ATG at nucleotide 152 was also disrupted by mutation to TTG. For M1.Only, three in-frame stop codons were introduced in the M2 coding region downstream of the M1 ORF (T820A, T826A and CC786,787TA). In addition, four non-synonymous changes were made to M2 coding sequence in the region following the splice acceptor site and upstream of the introduced stop codons. These changes were synonymous in the M1 reading frame. Both plasmids were used in conjunction with pDP plasmids encoding the other seven segments to generate a mixed virus population in which each virion contained an M1.Only or M2.Only segment. At 24 h post transfection, 293T cells were washed with 1 mL PBS, then overlaid with $1x10^6$ MDCK cells in virus medium plus TPCK-treated trypsin, and incubated at 33°C for 48 h. Supernatant was used to inoculate a plaque assay, and after 48 h a plaque isolate was used to inoculate a 75 cm^2 flask of MDCK cells. Following 48 h of growth, this stock was aliquoted and used to inoculate a plaque assay. One plaque isolate was diluted and used to inoculate 10-day-old embryonated chickens' eggs for a third passage. Experiments were conducted with this egg passage stock.

Infections

6-well dishes (Corning) were seeded with 4 x 10^5 MDCK cells in 2 mL complete medium, then incubated for 24 h. Prior to inoculation, complete medium was removed and cells were washed twice with 1 mL PBS per wash. Inocula containing virus in 200 uL PBS were added to cells, which were incubated on ice (to permit attachment but not viral entry) for 45 minutes. After attachment, the monolayer was washed with PBS remove unbound virus before 2 mL virus medium was added and plates were incubated at 33°C. For multi-cycle replication, TPCK-treated trypsin was added to virus medium to a final concentration of 1 ug/mL. When single-cycle conditions were required, virus medium was removed after 3 h and replaced with 2 mL virus medium containing NH_4Cl and HEPES.

Single-cell sorting assay for measurement of P_P values

4*10^5 MDCK cells were seeded into a 6-well dish, then counted the next day just before inoculation. Cells were then washed 3x with PBS and co-inoculated with the virus of interest (Pan/99-WT, MOI = 0.5 PFU/cell) and helper virus (Pan/99-Helper, MOI = 3.0 PFU/cell) in a volume of 200 uL. Cells were incubated at 33°C for 60 minutes, after which they were washed 3x with PBS, and 2 mL of virus medium was added. After incubation at 33°C for 60 minutes, medium was removed and cells were washed 3x with PBS before addition of Cell Dissociation Buffer (Corning) containing 0.1% EDTA (w/v) to release cells from the plate surface. Cells were harvested by resuspension in complete medium, followed by a series of three washes in 2 mL FACS buffer (2% FCS in PBS). Cells were resuspended in PBS containing 1% FCS, 10 mM HEPES, and 0.1% EDTA and filtered immediately prior to sorting on a BD Aria II. After gating to exclude debris

and doublets, one event was sorted into each well of a 96-well plate containing MDCK cell monolayers at 30% confluency in 50 uL virus medium containing TPCK-treated trypsin. After sorting, an additional 50 uL of medium was added to a final volume of 100 uL per well, and plates were spun at 1,800 rpm for 2 minutes to help each sorted cell attach to the plate surface. Plates were incubated at 33°C for 48 h to allow outgrowth of virus from this single infected cell.

RNA was extracted from infected cells using a ZR-96 Viral RNA Kit (Zymo Research) as per manufacturer instructions. Extracted RNA was converted to cDNA using universal influenza primers (given in Supplementary Table 2), Maxima RT (Thermo Scientific, 100 U/sample) and RiboLock RNase inhibitor (Thermo Scientific, 28 U/sample) according to manufacturer instructions. After conversion, cDNA was diluted 1:4 with nuclease-free water and used as template (4 uL/reaction) for segment-specific qPCR using SsoFast EvaGreen Supermix (Bio-Rad) in 10 uL reactions. Primers for each segment of Pan/99-WT virus, as well as the PB2 and PB1 segments of Pan/99-Helper virus, are given in Supplementary Table 2, and were used at final concentrations of 200 nM each.

Flow cytometry

At 12 h post-inoculation, virus medium was aspirated from infected cells, and monolayers were washed with PBS. The monolayer was disrupted using 0.05% trypsin + 0.53 mM EDTA in Hank's Balanced Salt Solution (HBSS). After 15 minutes at 37°C, plates were washed with 1 mL FACS buffer (PBS + 1% FCS + 5 mM EDTA) to collect cells and transfer them to 1.7 mL tubes. Cells were spun at 2,500 rpm for 5 minutes, then resuspended in 200 uL FACS buffer and transferred to 96-well V-bottom plates

(Corning). The plate was spun at 2,500 rpm and supernatant discarded. Cells were resuspended in 50 uL FACS buffer containing antibodies at the following concentrations, then incubated at 4° C for 30 minutes:

1.) His Tag-Alexa 647 (5 ug/mL) (Qiagen, catalog no. 35370)

2.) HA Tag-FITC (7 ug/mL) (Sigma, clone HA-7, catalog no. H7411)

After staining, cells were washed by three times by centrifugation and resuspension in FACS buffer. After the final wash, cells were resuspended in 200 uL FACS buffer containing 7-AAD (12.5 ug/mL) and analyzed by flow cytometry using a BD Fortessa.

This approach was modified slightly when staining for M1 and M2. After staining for His and HA (where indicated), cells were washed once with 200 uL FACS buffer, then resuspended in 100 uL BD Cytofix/Cytoperm buffer and incubated at 4°C for 20 minutes. BD Cytoperm/Cytowash (perm/wash) buffer was added to each well, and cells were spun at 2,500 rpm for 5 minutes. After a second wash, cells were resuspended in 50 uL perm/wash buffer containing antibodies at the following concentrations:

1.) Anti-M1 GA2B conjugated to Pacific Blue (4 ug/mL) (ThermoFisher, catalog no. P30013)

2.) Anti-M2 14C2 conjugated to PE (4 ug/mL) (Santa Cruz, catalog no. sc-32238 PE)

Following another 30 minutes of staining at 4° C, cells were washed three times (as described above) with perm/wash buffer, then resuspended in FACS buffer without 7-AAD just prior to analysis on the BD Fortessa.

Single-cycle growth curves

Cells were inoculated with Pan/99-WT virus at MOIs of 1, 3, 6, 10, or 20 PFU/cell, and incubated with 2 mL virus medium at 33° C. After 3 h, virus medium was replaced with virus medium containing NH_4Cl and HEPES. 100 uL of medium was collected at 3, 4.5, 6, 8, 12, 18, 24, and 48 h post-inoculation (with replacement by fresh medium to keep volumes consistent) for virus quantification by plaque assay. At 12 h post-inoculation, cells were harvested and stained for analysis of HA expression by flow cytometry.

Impact of secondary spread on complementation of incomplete genomes

To optimize the approach of using Pan/99-Helper to activate and thereby detect cells containing incomplete genomes, we co-inoculated cells with a low MOI (0.01 PFU/cell) of Pan/99-WT virus and a range of Pan/99-Helper virus MOIs and measured expression of WT HA after 12 h. We observed a biphasic relationship between helper virus MOI and the benefit provided to WT virus (Supplementary Figure 4C). As more Helper virus was added, the percentage of cells expressing WT HA initially increased as more cells became co-infected and thus capable of expressing the WT HA protein. But, as the Helper MOI increased further, a competitive effect was observed and the probability of detecting WT HA expression was decreased. Observing that Pan/99-Helper virus provided the greatest benefit—a 2-fold increase in the frequency of WT HA expression—at an MOI of 0.3 PFU/cell, we used that amount in further complementation experiments. Based on measured P_P values, this dose is estimated to contain an average of 27 particles/cell.

In single-cycle replication conditions, cells were inoculated on ice with Pan/99-WT virus over a range of MOIs (0.1, 0.3, 0.6, 1 PFU/cell) and, at the same time, with

Pan/99-Helper virus (MOI = 0.3 PFU/cell) or PBS. After inoculation, cells were washed with 1 mL PBS, 2 mL virus medium (no trypsin) was added, and cells were incubated at 33°C for 3 h, after which initial virus medium was replaced with virus medium containing NH_4Cl and HEPES. At 12 h post-inoculation, cells were collected and stained for WT and Helper HA expression as described above.

In multi-cycle replication conditions, cells were inoculated on ice with Pan/99-WT virus at an MOI of 0.01 or 0.002 PFU/cell, and then incubated at 33°C with virus medium containing TPCK-treated trypsin to allow for multi-cycle growth. After 12 h, cells were washed with 1 mL PBS, then inoculated on ice with Pan/99-Helper virus (MOI = 0.3 PFU/cell), or PBS. After inoculation, cells were washed with 1 mL PBS, 2 mL virus medium (no trypsin) was added, and cells were incubated at 33°C for 3 h, after which initial virus medium was replaced with virus medium containing NH_4Cl and HEPES. At 12 h post-inoculation with Pan/99-Helper virus, cells were collected and stained for WT and Helper HA expression as described above. The amount of complementation provided by Pan/99-Helper virus was calculated using the equation:

$$\%Enrichment = \frac{\%WT^{+}|Helper^{+} - \%WT^{+}|Helper^{-}}{\%WT^{+}|Helper^{-}} * 100$$

where $\%WT^{+} \mid Helper^{+}$ denotes the percentage of cells expressing WT HA in the Helper HA^{+} sub-population.

Microscopy

To visualize foci of infection, cells were infected with Pan/99-NP_TCvirus at an MOI of 0.002 PFU/cell, and incubated under multi-cycle conditions. After 12 h, media was aspirated and changed to virus medium containing NH_4Cl + HEPES, and cells were incubated for another 12 h. To visualize randomly dispersed infections, cells were

inoculated with Pan/99-NP_TC virus at an MOI of 0.1 PFU/cell, and incubated under single-cycle conditions for 12 h (3 h with virus medium, followed by 9 h with virus medium containing NH_4 Cl + HEPES). To visualize FlaSH reagent background, cells were mock-infected with 200 uL PBS, then incubated under single-cycle conditions for 12 h (3 h with virus medium, followed by 9 h with virus medium containing NH_4 Cl + HEPES).

At the end of the infection, cells were washed twice with 1 mL PBS, then stained with 2 uM FlaSH reagent (ThermoFisher, cat. no. T34561) diluted in Opti-MEM. During staining, plates were incubated at 37°C for 30 min. To remove FlaSH, cells were washed three times by adding 300 uL of BAL wash buffer diluted to 250 uM in Opti-MEM and incubating at 37°C for 10 min per wash. After washing, cells were washed once with 1 mL PBS, then fixed by addition of 500 uL 4% paraformaldehyde and incubation for 10 minutes at room temperature. After fixation, cells were stained with 300 nM DAPI for 3 minutes at room temperature, then washed three times with 1 mL PBS, and visualized using a BioTek Lionheart FX. For each image, brightness was enhanced by 20%, and contrast by 40%.

Digital droplet PCR (ddPCR)

Primers and probes (listed in Supplementary Table 3) were diluted to concentrations of 900 nM and 250 nM per primer and probe, respectively. 22 uL reactions were prepared with 11 uL Bio-Rad SuperMix for Probes (1X final concentration), 6.6 uL of diluted primers (900 nM/primer, final concentration) and probes (250 nM/probe, final concentration), and 4.4 uL of diluted cDNA. 20 uL of each reaction mixture was partitioned into droplets using a Bio-Rad QX200 droplet generator per

manufacturer instructions. PCR conditions were: 1.) 95° C for 10 minutes, 2.) 40 cycles of A.) 94° C for 30 seconds and B.) 57° C for 1 minute, 3.) 98° C for 10 minutes, and hold at 4°C. Droplets were then read on Bio-Rad QX200 droplet reader, and the number of cDNA copies/uL was calculated.

Guinea pig infections

Female Hartley guinea pigs were obtained from Charles River Laboratories (Wilmington, MA) and housed by Emory University Department of Animal Resources. All experiments were conducted in accordance with an approved Institutional Animal Care and Use Committee protocol. For ID_{50} estimation and analysis of viral shedding, guinea pigs were anesthetized by intramuscular injection with 30 mg/kg ketamine/ 4 mg/kg xylazine, then inoculated intranasally with 300 uL virus diluted in PBS. Nasal washes were collected in 1 mL PBS on days 1, 2, 3, 5, and 7 as described previously[47], and titered by RT-ddPCR targeting the NS segment. For transmission experiments, inoculated guinea pigs were individually housed in Caron 6040 environmental chambers at 10°C and 20% relative humidity. At 24 h post-inoculation, one naïve guinea pig was introduced to each cage with one inoculated animal. Nasal washes were collected on days 2, 4, 6, and 8, and titered by RT-ddPCR.

Data availability

All raw data supporting these findings and used to generate figures and supplementary information are available on Github at the following URL: https://github.com/njacobs627/Pan99_IVGs_Spatial_Structure.

Code availability

All R code used for simualtions, data analysis, and visualizations is available on Github at the following URL: https://github.com/njacobs627/Pan99_IVGs_Spatial_Structure.

Chapter 4. Discussion

4.1 Overview

The research presented here had three aims:

(1) Characterize competition between pneumococcal strains during biofilm formation.

(2) Define the potential consequences of incomplete viral genomes for influenza A virus infectivity and replication.

(3) Determine the effectiveness of complementation in mitigating the fitness costs of incomplete viral genomes in cell culture and *in vivo*.

The work described in Chapter 2 represents forward progress in achieving Aim 1, and has laid the groundwork for more detailed studies of bacterial competition in biofilms [92, 137]. The work described in Chapter 3 represents an integrative approach to addressing Aims 2 and 3, and sheds light on some long-standing questions in influenza virus population biology.

4.2 Discussion of Chapter 2

The study described in Chapter 2 was motivated by observations that simultaneous co-colonization with multiple pneumococcal serotypes was common, raising the question of how two serotypes interact when placed in close proximity.

An important limitation of these studies is the absence of the host immune system. These experiments demonstrate that interference competition between TIGR4 and S19F strains occurs *in vitro*. While this likely also occurs *in vivo*, co-colonization of a human nasopharynx also involves the presence of neutrophils. Pneumococcal serotypes differ in their susceptibility to neutrophil-mediated killing due to variation in the structure of their polysaccharide capsule [138, 139]. Thicker capsules may provide more protection from neutrophils, but are energetically more costly to produce, presenting a trade-off between efficient growth and avoiding immune activity [138-140]. While our results are informative with respect to the interference competition that occurs during mixed-strain biofilm formation, they should not be taken as representative of what will occur during colonization of a human host.

While the work presented herein did not establish a definitive mechanism by which competition occurs between co-colonizing pneumococcal strains, experiments with trans-well devices demonstrated that physical contact is necessary. One potential mechanism of competition that may act at this scale is bacteriocins with limited stability. This property prevents them from diffusing far enough to kill spatially distant bacteria, but they been shown to affect competitive dynamics in mice, suggesting that competition does occur at local scales *in vivo* [141]. Continuations of this work have found that hydrogen peroxide, which is commonly secreted by *S. pneumoniae*, was not responsible for its ability to kill *Staphylococcus aureus* biofilms [92]. Competitive superiority of strains carrying genes for bacteriocins encoded by the *blp* locus during biofilm growth was recently observed by Wholey et al., which was not considered in Chapter 2 but provides a possible mechanistic explanation for the results of the present work [142].

4.3 Discussion of Chapter 3

The work described in Chapter 3 was motivated by the fact that, although the existence of incomplete IAV genomes had been characterized in work by Brooke et al., which suggested that an IAV populations existed as "a swarm of complementation-dependent particles," relatively little was known about the fitness costs of incomplete genomes and how much complementation was required to mitigate those costs.

An important limitation of the individual-based model of replication is that it tracks only the number of different genome segments present in each cell, not the copy number of each. Thus, it cannot account for the kinetic effect of gene dosage. While the effects of gene dosage are poorly characterized in the literature, our data reveal earlier emergence of viral progeny occurs when cells are infected at high multiplicities of infection. This limitation is a reasonable simplification because the present form of the model considers a genetically homogenous virus population. When considering genetic diversity in a replicating virus population, however, the fact that genome segments emerge more quickly in co-infected cells becomes important. Mechanisms that facilitate co-infection, such as intermediate dispersal rates, may improve viral fitness and therefore be selected for due to the kinetic advantage provided by multiple infection. Furthermore, a reliance on multiple infection could serve to buffer a virus population against the deleterious effects of Muller's ratchet, as free mixing of genome segments should facilitate the purging of deleterious mutations in multiply infected cells in a manner analogous to sexual reproduction in eukaryotes.

The work described in Chapter 3 demonstrated a kinetic benefit to high MOIs, showing that virus emerges earlier from multiply infected cells, but that maximum viral yield is reached once all cells become infected. The finite amount of virus that can be produced by an infected cell results in a trade-off to multiple infection. Thus, as the MOI increases past the point of saturating the number of infected cells, the amplification of the input virus declines to the point of less virus being produced than was used to inoculate the cells. This trade-off is likely to influence the optimal level of spatial structure for a virus population.

Our theoretical models of complementation assume that each particle contains a full complement of eight genome segments, and after a particle infects a cell, genome segments are either lost or replicated with independent probabilities. In measuring IVG frequencies, however, we detected slight associations among segment pairs.

It is interesting that that we observe such weak associations between segments in delivery to the cell, when the role of RNA-RNA interactions in IAV genome packaging has been well documented [143-145]. The RNA comprising the terminal regions of genome segments mediate efficient packaging of genomes into virus particles. Packaging is not perfectly efficient, but diverse methods have shown in multiple strain backgrounds that a majority of virus particles contain one copy of each genome segment. Molecular genetic methods such as fluorescence in situ hybriziation (FiSH) show tight pairwise associations between genome segments in particles [52].

The agent-based models and empirical results highlight the importance of spatial structure and co-infection in influenza A virus replication. It is important to note that this model did not consider genetic diversity of virions during an infection, but the

availability of barcoding technologies and methods of manipulating spatial structure in the systems tested make this area ripe for future investigation. Given the nature of diffusion, where the distance traveled by a given particle follows a normal distribution, some small fraction of particles will travel very far from the cell that produced them and be the only one to infect a new cell. While this is likely to leave that singly infected cell with an incomplete viral genome, in rare cases infection by a fully infectious particle in a long-range dispersal event would lead to the establishment of a satellite colony. Progeny virions from this new infection will then spread locally, facing much less competition from co-infecting particles that they would otherwise encounter at the wave front of the expanding mother colony, and thus genes from that original far dispersing virion would become overrepresented in the population. Depending on the frequency of these long-distance dispersal events, the corresponding founder effects they cause may result in the reduction of overall genetic diversity in the population.

Another important limitation, but potential avenue for future investigation, is the fact that our model does not fully consider the role of the host immune system, particularly the spatial structure inherent in type I interferon responses. Cells that mount an antiviral response secrete interferons, which signal to neighboring cells in a paracrine manner and induce them to upregulate antiviral factors. Depending on the rate at which this signal travels, diffusing virions may infect cells that are already refractory to infection, which would limit the growth of an expanding colony. The optimal rate of virion dispersal is likely, then, to be influenced by the rate of interferon signal diffusion, as a growing virus population must be able to diffuse quickly to infect new cells before interferon reaches them. Advances in longitudinal imaging of live cells over the duration

of infection, combined with reporter technologies to visualize interferon signaling and viral infection, may be useful in observing the interaction between these processes.

The generation of a virus strain that was entirely dependent on co-infection for replication *in vivo* served to demonstrate the ability of complementation to mitigate the fitness costs of incomplete genomes, with only a moderate reduction in peak viral shedding relative to wild-type Pan/99 virus in infected guinea pigs. The sharp reduction in infectivity and failure of this virus to transmit in a permissive transmission setting, however, highlight an important role for fully infectious particles in the establishment of infection. These results corroborate others by McCrone et al. suggesting that influenza A virus populations undergo a tight genetic bottleneck upon transmission [136]. Participation of incomplete genomes in infection may therefore be important for diversification within an individual host, but within-host diversity then contracts upon transmission to the next host, similar to the diversification-bottleneck cycles that occur during HIV transmission in humans, albeit with smaller magnitude [146].

A challenge to experimental studies with Pan/99 virus is that, because incomplete genomes appear to be an emergent property of the interaction between a virus particle and host cell, a population of virus particles will always be heterogenous, comprising some fully infectious particles and many that are semi-infectious. Furthermore, this makes it difficult if not impossible to differentiate between fully and semi-infectious particles. Abrogating the contribution of semi-infectious particles to infection would then require a host cell environment that is more permissive, such as expressing polymerase components *in trans* that are available to transcribe and replicate genome segments

shortly after they enter the cell, and altering the pH of fusion so that segments have only a short distance to travel between the endosome and the nucleus.

As discussed above and in Chapter 3, incomplete viral genomes require complementation to participate in infection, and the level of complementation required for productive cellular infection depends on the frequency of segments being absent. The work described in Chapter 3 demonstrates that the spatial structure inherent in secondary spread allows a virus population to meet this need through localized dispersal, but the optimal level of dispersal is influenced by the factors discussed previously, namely interferon responses, the kinetic effect of gene dosage, and competition between segments in co-infected cells. Factors governing virus dispersal can be broadly divided into two categories: 1) those that influence the diffusion coefficient of a virion, and 2) those that influence the time spent between a particle being produced and infecting a new cell.

The diffusion coefficient of a particle in liquid is governed by three specific factors: particle size, absolute temperature, and liquid viscosity. While the diameter of individual virus particles is generally consistent, substantial variation in shape has been reported, with particles ranging from spherical to filamentous [147], which may confer differences in dispersal as more complex processes affect diffusion of non-spherical particles. Fever may increase host body temperature and consequently virion diffusion coefficients, but the increase conferred by a 4°C increase in temperature is ~1%, making the contribution of body temperature to diffusion ultimately minimal. Finally, mucous is more viscous than water, which could make virion dispersal relatively slower than what is modeled, but as discussed in Chapter 3, airway mucous exists on top of a layer of

watery fluid, and diffusion to neighboring cells may be more likely to occur for this lower-viscosity fluid.

The expected impact of changing temperature, particle size, and liquid viscosity suggest that changes in a virion's dispersal kernel are more likely mediated by factors that govern a particle's ability to attach to and infect new cells, such as the HA:NA balance on the particle surface and the sialic acid content of the host cell surface. Handel et al. have explored the question "how sticky should a virus be?" by considering total fitness of a virus with different combinations of attachment and detachment rates, analogous to the strength of its hemagglutinin and neuraminidase proteins, respectively [134]. Modeling results suggest that a virus that is too "sticky" (i.e. has a high HA:NA balance) has difficulty releasing from the cell that produced it, and one that has too low an HA:NA balance has difficulty attaching to new cells long enough to infect them, and consequently viral fitness is maximized at intermediate levels of stickiness.

Appendix

5.1 Probabilistic model of cellular infection (Figure 3.2) code

```
Pp = seq(from = 0.01, to = 1, by = 0.01)
tau = matrix(nrow = 1,
             ncol = 9,
             data = c(1,0,0,0,0,0,0,0,0))
sub.tau = matrix(nrow = 1,
                 ncol = 8,
                 data = c(1,0,0,0,0,0,0,0))

trans = matrix(nrow = 9,
               ncol = 9,
               data = 0)
ident = diag(8)
one.mat = matrix(nrow = 8,
                 ncol = 1,
                 data = 1)

Segment.Probs = matrix(ncol = 5,
                       nrow = 100 * 100 * 9,
                       data = 0)

Inf.Unit = rep(0, 100)
Inf.Unit.sd = rep(0,100)

colnames(Segment.Probs) = c("Pp","Virion","Segments","Prob","Expected.Segments")
# Order = Pp -> Virion # -> Segments
Segment.Probs[,"Pp"] = rep(seq(from = 0.01, to = 1, by = 0.01), each = 900)
Segment.Probs[,"Virion"] = rep(1:100, each = 9)
Segment.Probs[,"Segments"] = rep(seq(0,8),100)
Segment.Mat = matrix(nrow = 9,
                     ncol = 1,
                     data = 0:8)
index = 1

for (pp in 1:100) {

 # Define Transition Matrix
 for (i in 0:8) {
  for (j in i:8) {
   trans[i+1,j+1] = dbinom(x = j - i, size = 8 - i, prob = Pp[pp])
  }
  trans[9,9] = 1
```

```
  }

  for (v in 1:100) {

    Prob.Vector = t(tau %*% (trans %^% v))

    Segment.Probs[index:(index + 8),"Prob"] = Prob.Vector
    Segment.Probs[index:(index + 8),"Expected.Segments"] = t(Prob.Vector) %*%
Segment.Mat
    index = index + 9
  }

  # Calculate Infectious Unit
  sub.trans = trans[1:8,1:8]
  Q = sub.trans
  N = solve(ident - Q)
  t = N %*% one.mat
  tsq = t^2
  Inf.Unit[pp] = sub.tau %*% solve(ident - sub.trans) %*% one.mat
  Inf.Unit.sd[pp] = sqrt(((2 * N - ident) %*% t - tsq)[1,1])

}

Segment.Probs = as.data.frame(Segment.Probs)

Geom = matrix(ncol = 3,
              nrow = 100,
              data = c(Pp,Inf.Unit,Inf.Unit.sd),
              byrow = FALSE) %>%
  as.data.frame
colnames(Geom) = c("Pp","Inf.Unit","Inf.Unit.sd")

Exp.Pp = read.csv(file = file.path(Data.Path,"Pp_Summary.csv"),na.strings = c("","-"))
%>%
  filter(Delay == 0) %>%
  dplyr::select(Rep,Avg) %>%
  mutate(Pp = as.numeric(round(Avg * 100)))

Exp.Pp$Rep = (Exp.Pp$Rep)

Inf.Unit.Sum = Geom[Exp.Pp[1:13,"Pp"],]
Inf.Unit.Sum$Rep = 1:nrow(Inf.Unit.Sum)
Inf.Dose = mean(Inf.Unit.Sum$Inf.Unit) %>% round(1)

Geom$Particle_PFU = (Geom$Pp ^ 8)
Inf.Unit.Sum$Full = (Inf.Unit.Sum$Pp ^ 8) * 100
```

```
Geom$Full.sd = sqrt(Geom$Particle_PFU * (1 - Geom$Particle_PFU))
Inf.Unit.Sum$Full %>% mean
Inf.Unit.Sum$Inf.Unit %>% mean
```

5.2 Individual-based model of replication (Figure 3.4) code

```
#### Load Packages ----

require(foreach)
require(doMC)
require(Rmisc)
require(ggplot2)
set.seed(666)

### Set File Paths ----

Proj.Home="/Users/Nate/Dropbox/Modeling" # Nate Macbook
Proj.Home="/home/rstudio/Dropbox/Modeling" # Rstudio Amazon
Data.Path = file.path(Proj.Home,"Amazon_Output")

#### Define Parameters ----

# Static parameters
Grid.X = 100
Grid.Y = Grid.X
Cell_Width = 30 # Assume each cell is a 30 um x 30 um square

# Rates from Handel et al. J R Soc Interface 2014
Death_Rate = 2 / 24
Birth_Rate = 2 / 24

Decay_Rate = -log(0.5) / 4 # Assume 1/2 life of 4 hours
Exclude_Rate = 1 / 8 # From Nicolle's thesis
Revert_Rate = 1 / 12

Infect_Rate = 1 / 6 # 6 hour eclipse phase
# Incorporates the 6-hour lag before segments are "useful"
# So "Segment #" is the number of useful segments (cells become productive instantly
once 8th segment is delivered)

# Varied parameters: first, set defaults
Virus.D = 5.825 * 60 * 60 # 5.825 um^2 / s based on Einsten-Stokes equation
Attach_Rate = 1 / (10 / 60) # 20 minutes
Detach_Rate = 10 / 24 # 10/day = 10/24 hr
Burst_Rate = 962 / 24
```

```
Start_Radius = 0

# Local Simulation Parameters
Pp = c(0.575,1)
Sim_Num = 2
Spread_Frac = c(0.5)
Virus.D = 5.825 * c(0.001, 0.01, 0.1, 1, 10, 100)
Virus.D = 5.825 * c(0.001, 0.1, 1, 100)
Virus.D = 5.825 * 60 * 60
Spread_Frac = c(0, 0.1, 0.25, 0.5, 0.75, 0.9, 1)
ts = 1/20

Duration = 1

#Amazon Simulation Parameters
Sim_Num = 10
Pp = c(0.575, 1.0)
Spread_Frac = c(0, 0.1, 0.25, 0.5, 0.75, 0.9, 1)
Virus.D = 5.825 * 60 * 60
Virus.D = Virus.D * c(rep(c(1, 0.1, 10, 100, 1e3),5) * rep(c(0.2, 0.4, 0.6, 0.8,1),each =
5),1e5, 1e8, 1e11)
Duration = 96
ts = 1/20
sort(Virus.D / 3600)

# Scaling to ts
Infect_Rate = Infect_Rate * ts
Attach_Rate = Attach_Rate * ts
Detach_Rate = Detach_Rate * ts
Burst_Rate = Burst_Rate * ts
Death_Rate = Death_Rate * ts
Birth_Rate = Birth_Rate * ts
Exclude_Rate = Exclude_Rate * ts
Decay_Rate = Decay_Rate * ts
Revert_Rate = Revert_Rate * ts
Virus.D = Virus.D * ts

Infect_Rate_Num = length(Infect_Rate)
Attach_Rate_Num = length(Attach_Rate)
Detach_Rate_Num = length(Detach_Rate)
Burst_Rate_Num = length(Burst_Rate)
Death_Rate_Num = length(Death_Rate)
Birth_Rate_Num = length(Birth_Rate)
Decay_Rate_Num = length(Decay_Rate)
Revert_Rate_Num = length(Revert_Rate)
Exclude_Rate_Num = length(Exclude_Rate)
```

```
Virus.D_Num = length(Virus.D)
Pp_Num = length(Pp)
Spread_Frac_Num = length(Spread_Frac)

Start_Radius_Num = length(Start_Radius)

#### Populations -----
# Initialize Cells ----

Cell.Headers =
c("Cell.ID","X","Y","Distance","Living","Segments","Virion.Hits","Infected","Complete
","Productive","Susceptible","Naive")
Cells = matrix(data = 0,
               nrow = Grid.X * Grid.Y,
               ncol = length(Cell.Headers))
colnames(Cells) = Cell.Headers
Cells[,"Cell.ID"] = 1:(Grid.X * Grid.Y)
Cells[,"X"] = rep(1:100, each = 100)  #1,1,1,1,1... 2,2,2,2,2
Cells[,"Y"] = rep(1:100, times = 100) #1,2,3,4,5...1,2,3,4,5
Cells[,"Distance"] = sqrt((Cells[,"X"] - 50) ^ 2 + (Cells[,"Y"] - 50)^ 2)
Cells[,"Living"] = 1
Cells[,"Susceptible"] = 1
Cells[,"Naive"] = 1
Init.Cells = Cells

# Initialize Virions ----
Virion.Headers =
c("Virion.ID","Cell.ID","X","Y","Cell.X","Cell.Y","Bound","Diffuse","Movable")

Virions = matrix(data = 0,
                 nrow = 0,
                 ncol = length(Virion.Headers))
colnames(Virions) = Virion.Headers
Init.Virions = Virions

#### Executive Function ----

Sim.Exec = function(Attach_Rate,
                    Detach_Rate,
                    Infect_Rate,
                    Birth_Rate,
                    Death_Rate,
                    Burst_Rate,
                    Exclude_Rate,
                    Revert_Rate,
                    Decay_Rate,
```

```
                Start_Radius,
                Virus.D,
                Pp,
                Spread_Frac,
                Cells,
                Virions,
                Duration,
                ts,
                Sim,
                Varied.Par) {
  set.seed(Sim)

  Cell.Summary.Headers = c("Sim","t",

"Attach_Rate","Detach_Rate","Birth_Rate","Death_Rate","Burst_Rate","Infect_Rate","E
xclude_Rate","Revert_Rate","Decay_Rate","Virus_D","Pp","N0","Spread_Frac", #
Parameters

"Virions","Free.Virions","Bound.Virions","New.Virions","Segments","Cell.MOI","Inf.M
OI","Semi.MOI","Prod.MOI","Productive","Semi.Cells","Dead.Cells","Refractory.Cells"
,"Wasted.Cells","Infected","Susceptible","Naive",

"Radius.25","Radius.50","Radius.75","Radius.90","Radius.95","Radius.975","Radius.Ma
x",
                      "Varied.Par")

  Cell.Summary = matrix(data = 0,
                        nrow = Duration/ts,
                        ncol = length(Cell.Summary.Headers))
  colnames(Cell.Summary) = Cell.Summary.Headers

  Cell.Summary[,"Sim"] = Sim
  Cell.Summary[,"Varied.Par"] = Varied.Par
  Cell.Summary[,"t"] = 1:(Duration / ts) * ts
  Cell.Summary[,"Virus_D"] = Virus.D / ts
  Cell.Summary[,"Pp"] = Pp
  Cell.Summary[,"Spread_Frac"] = Spread_Frac
  Cell.Summary[,"Attach_Rate"] = Attach_Rate / ts
  Cell.Summary[,"Detach_Rate"] = Detach_Rate / ts
  Cell.Summary[,"Birth_Rate"] = Birth_Rate / ts
  Cell.Summary[,"Death_Rate"] = Death_Rate / ts
  Cell.Summary[,"Burst_Rate"] = Burst_Rate / ts
  Cell.Summary[,"Infect_Rate"] = Infect_Rate / ts
  Cell.Summary[,"Exclude_Rate"] = Exclude_Rate / ts
  Cell.Summary[,"Revert_Rate"] = Revert_Rate / ts
  Cell.Summary[,"Decay_Rate"] = Decay_Rate / ts
```

```
  Cell.Summary[,"N0"] = (Start_Radius * 2 + 1) ^ 2

  Cell.Summary[,"New.Virions"] = 0

  # Initialize Infected Cells ----
  Start.X = 50 - Start_Radius
  End.X = 50 + Start_Radius
  Start.Y = 50 - Start_Radius
  End.Y = 50 + Start_Radius
  Start.Index = Cells[which((Cells[,"X"] %in% (Start.X:End.X)) & (Cells[,"Y"] %in%
(Start.Y:End.Y))),"Cell.ID"]

  Cells[Start.Index,"Segments"] = 8
  Cells[Start.Index,"Virion.Hits"] = 1
  Cells[Start.Index,"Infected"] = 1
  Cells[Start.Index,"Productive"] = 1

  Grid.X = max(Cells[,"X"])
  Grid.Y = max(Cells[,"Y"])
  Cell_Width = 30
  Grid_Width_X = Cell_Width * Grid.X
  Grid_Width_Y = Cell_Width * Grid.Y

  Inf.Cells = Cells[Cells[,"Infected"] == 1,,drop = FALSE]
  Prod.Cells = Inf.Cells[Inf.Cells[,"Productive"] == 1,,drop = FALSE]

  #Translocation probability for locally-dispersing virions
  Trans.Prob = matrix(data = 0,
                      ncol = 3,
                      nrow = 4)
  colnames(Trans.Prob) = c("Prob","X.Adjust","Y.Adjust")
  Trans.Prob[,"Prob"] = rep(1,4)
  Trans.Prob[,"X.Adjust"] = c(-1,0,1,0)
  Trans.Prob[,"Y.Adjust"] = c(0,-1,0,1)

  # Iterate Time Series ----
  for (t in 1:(Duration/ts)) { # Start Time

    # Exit if no more cells can be infected (Productive cells have died, and no more cells
can be infected)
    if ((sum(Cells[,"Susceptible"]) + sum(Cells[,"Productive"])) == 0) {
      Cell.Summary = Cell.Summary[1:(t - 1),]
      break
    }

    # Characterize Populations
```

```
  free = which(Virions[,"Bound"] == 0)
  diffuse.free = which(((Virions[,"Diffuse"]) * (1 - Virions[,"Bound"])) == 1)
  adherent = which(Virions[,"Bound"] == 1)
  diffuse.adherent = which(((Virions[,"Diffuse"]) * (Virions[,"Bound"])) == 1)

  # Characterize Cells ----
  Cells[,"Infected"] = (Cells[,"Segments"] > 0) * (Cells[,"Living"])
  Cells[,"Productive"] = (Cells[,"Segments"] == 8) * (Cells[,"Living"])

  # Obtain Infected, Productive, Semi-Infected, and Dead cell subpopulations
  Inf.Cells = Cells[Cells[,"Infected"] == 1,,drop = FALSE]
  Prod.Cells = Inf.Cells[Inf.Cells[,"Productive"] == 1,,drop = FALSE]

  Semi.Cells = Inf.Cells[Inf.Cells[,"Productive"] == 0,,drop = FALSE]
  Dead.Cells = Cells[Cells[,"Living"] == 0,,drop = FALSE]

  Refract.Cells = Cells[((1 - Cells[,"Living"]) + Cells[,"Susceptible"]) == 0,,drop =
FALSE]
  Wasted.Cells = Refract.Cells[Refract.Cells[,"Segments"] > 0,,drop = FALSE]

  # Free virions diffuse ----
  if (length(diffuse.free) > 0) {

    Virion.Num = length(diffuse.free)

    # Add diffusion distance
    Distance = rnorm(mean = 0, sd = sqrt(2 * Virus.D), n = Virion.Num)
    Theta = runif(min = 0 , max = 2*pi, n = Virion.Num)
    X.Distance = cos(Theta) * Distance
    Y.Distance = sin(Theta) * Distance

    Virions[diffuse.free,"X"] = Virions[diffuse.free,"X"] + X.Distance
    Virions[diffuse.free,"Y"] = Virions[diffuse.free,"Y"] + Y.Distance

    # Ensure virions stay in grid -- if anything is outside, it comes around from the other
side

    Virions[diffuse.free,"X"] = Virions[diffuse.free,"X"] %% Grid_Width_X
    Virions[diffuse.free,"Y"] = Virions[diffuse.free,"Y"] %% Grid_Width_Y

    # Virions[diffuse.free,"X"] = abs(Virions[diffuse.free,"X"])
    # Virions[diffuse.free,"X"] = Virions[diffuse.free,"X"] %% (2 * Grid_Width_X)
    # Virions[diffuse.free,"X"] = Virions[diffuse.free,"X"] - 2 * pmax(0,
    #                                                  Virions[diffuse.free,"X"] -
(Grid_Width_X))
    #
```

```
    # Virions[diffuse.free,"Y"] = abs(Virions[diffuse.free,"Y"])
    # Virions[diffuse.free,"Y"] = Virions[diffuse.free,"Y"] %% (2 * Grid_Width_Y)
    # Virions[diffuse.free,"Y"] = Virions[diffuse.free,"Y"] - 2 * pmax(0,
    #                                                  Virions[diffuse.free,"Y"] -
(Grid_Width_Y))

    # Update Cell locations of virions
    Virions[diffuse.free,"Cell.X"] = floor(Virions[diffuse.free,"X"] / Cell_Width) + 1
    Virions[diffuse.free,"Cell.Y"] = floor(Virions[diffuse.free,"Y"] / Cell_Width) + 1
    #Virions[diffuse.free,"Cell.X"] = pmin(Virions[diffuse.free,"Cell.X"],Grid.X)
    #Virions[diffuse.free,"Cell.Y"] = pmin(Virions[diffuse.free,"Cell.Y"],Grid.Y)
    Virions[diffuse.free,"Cell.ID"] = (Virions[diffuse.free,"Cell.X"] - 1) * Grid.X +
Virions[diffuse.free,"Cell.Y"]

  }

  # Free virions attach ----
  attach.num = round(Attach_Rate * length(diffuse.free), 0)
  attach.num = min(length(diffuse.free),
                   rpois(n = 1, lambda = Attach_Rate * length(diffuse.free)))
  if (attach.num > 0) {
    attach = sample(x = diffuse.free, size = attach.num, replace = FALSE)
    Virions[attach,"Bound"] = 1
  }

  # Bound virions detach ----
  detach.num = round(Detach_Rate * length(diffuse.adherent), 0)
  detach.num = min(length(diffuse.adherent),
                   rpois(n = 1, lambda = Detach_Rate * length(diffuse.adherent)))
  if (detach.num > 0) {
    detach = sample(x = diffuse.adherent, size = detach.num, replace = FALSE)
    Virions[detach,"Bound"] = 0
  }
  # Bound virions infect ----

  infect.num = round(Infect_Rate * length(adherent), 0)
  infect.num = min(length(adherent),
                   rpois(n = 1, lambda = Infect_Rate * length(adherent)))

  if (infect.num > 0) {
    infect = sample(x = adherent, size = infect.num, replace = FALSE)

    target = table(Virions[infect,"Cell.ID"])

    # Add segments to cell
    Cell.Target = as.numeric(names(target))
```

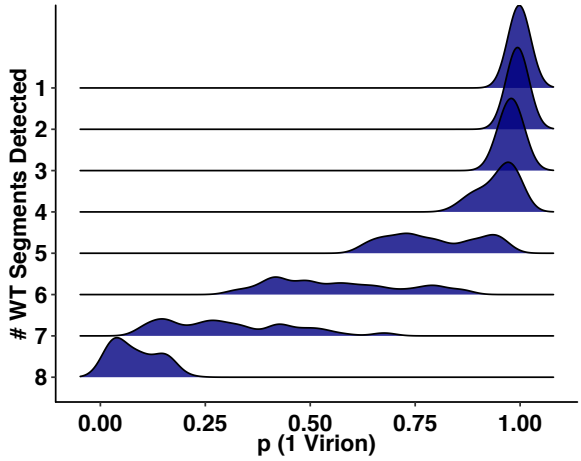
```
    Cell.Hit = as.numeric(target)

    Cells[Cell.Target,"Virion.Hits"] = Cells[Cell.Target,"Virion.Hits"] + Cell.Hit *
Cells[Cell.Target,"Susceptible"]
    Cells[Cell.Target,"Naive"] = 0
    Delivered = rbinom(n = length(Cell.Hit),
                       prob = 1 - ((1 - Pp) ^ Cell.Hit),
                       size = 8 - Cells[Cell.Target,"Segments"])
    Cells[Cell.Target,"Segments"] = (Cells[Cell.Target,"Segments"] + Delivered) *
Cells[Cell.Target,"Susceptible"]

    Virions[infect,"Bound"] = NA
    Virions = Virions[-infect,]
  }

  # Productive cells produce virions ----
  Burst.Size = rpois(lambda = Burst_Rate, n = nrow(Prod.Cells))
  Burst.Sum = sum(Burst.Size)

  if (Burst.Sum > 0) {
    New.Virions = matrix(data = 1,
                         nrow = Burst.Sum,
                         ncol = length(Virion.Headers))

    colnames(New.Virions) = Virion.Headers
    New.Virions[,"Bound"] = 1

    # Determine Location ----
    rows = rep(1:nrow(Prod.Cells), Burst.Size)
    New.Virions[,"Cell.ID"] = Prod.Cells[rows,"Cell.ID"]
    New.Virions[,"Cell.X"] = Prod.Cells[rows,"X"]
    New.Virions[,"Cell.Y"] = Prod.Cells[rows,"Y"]
    New.Virions[,"X"] = (New.Virions[,"Cell.X"] - 1) * Cell_Width + runif(min = 0,
max = Cell_Width, n = Burst.Sum)
    New.Virions[,"Y"] = (New.Virions[,"Cell.Y"] - 1) * Cell_Width + runif(min = 0,
max = Cell_Width, n = Burst.Sum)

    # Local Spread ----

    local.num = round(Spread_Frac * Burst.Sum,0)

    if (local.num > 0) {
      local = sample(x = 1:Burst.Sum, size = local.num, replace = FALSE)
      New.Virions[local,"Bound"] = 1
      New.Virions[local,"Diffuse"] = 0
```

```
      # Sample 1 - 4 for each to determine translocation direction of each locally
spreading virion
      adjust = sample(x = 1:nrow(Trans.Prob),
                      size = local.num,
                      prob = Trans.Prob[,"Prob"],
                      replace = TRUE)

      # Move to neighboring cell
      New.Virions[local,"Cell.X"] = New.Virions[local,"Cell.X"] +
Trans.Prob[adjust,"X.Adjust"]
      New.Virions[local,"Cell.Y"] = New.Virions[local,"Cell.Y"] +
Trans.Prob[adjust,"Y.Adjust"]

      # Ensure that virions are inside the grid
      New.Virions[local,"Cell.X"] = pmax(1, New.Virions[local,"Cell.X"])
      New.Virions[local,"Cell.Y"] = pmax(1, New.Virions[local,"Cell.Y"])

      New.Virions[local,"Cell.X"] = pmin(Grid.X, New.Virions[local,"Cell.X"])
      New.Virions[local,"Cell.Y"] = pmin(Grid.Y, New.Virions[local,"Cell.Y"])

      # Update Cell.ID
      New.Virions[local,"Cell.ID"] = (New.Virions[local,"Cell.X"] - 1) * 100 +
New.Virions[local,"Cell.Y"]

    }
    Virions = rbind(Virions,New.Virions)
  }

  # Productive cells become refractory to super-infection ----
  prod = which(Cells[,"Productive"] == 1)

  exclude.num = round(Exclude_Rate * length(prod), 0)
  exclude.num = min(length(prod),
                    rpois(n = 1, lambda = Exclude_Rate * length(prod)))

  if (exclude.num > 0) {
    exclude = sample(x = prod, size = exclude.num, replace = FALSE)
    Cells[exclude,"Susceptible"] = 0
  }
  # Productive cells die ----
  death.num = round(Death_Rate * length(prod), 0)
  death.num = min(length(prod),
                  rpois(n = 1, lambda = Death_Rate * length(prod)))
  if (death.num > 0) {
    death = sample(x = prod, size = death.num, replace = FALSE)
    Cells[death,"Living"] = 0
```

```
    Cells[death,"Productive"] = 0
    Cells[death,"Segments"] = 0
    Cells[death,"Susceptible"] = 0
    Cells[death,"Infected"] = 0
   }

   #Summarize system ----
   Cell.Summary[t,"Productive"] = sum(Cells[,"Productive"])
   Cell.Summary[t,"Dead.Cells"] = nrow(Cells) - sum(Cells[,"Living"])
   Cell.Summary[t,"Semi.Cells"] = sum(Cells[,"Infected"]) - sum(Cells[,"Productive"])
   Cell.Summary[t,"Refractory.Cells"] = nrow(Refract.Cells)
   Cell.Summary[t,"Wasted.Cells"] = nrow(Wasted.Cells)
   Cell.Summary[t,"Infected"] = sum(Cells[,"Infected"])
   Cell.Summary[t,"Susceptible"] = sum(Cells[,"Susceptible"])
   Cell.Summary[t,"Naive"] = sum(Cells[,"Naive"])
   #
   Cell.Summary[t,"Virions"] = nrow(Virions)
   Cell.Summary[t,"Bound.Virions"] = sum(Virions[,"Bound"])
   Cell.Summary[t,"Free.Virions"] = nrow(Virions) - sum(Virions[,"Bound"])
   Cell.Summary[t,"New.Virions"] = Burst.Sum
   Cell.Summary[t,"Segments"] = mean(Inf.Cells[,"Segments"])
   Cell.Summary[t,"Inf.MOI"] = mean(Inf.Cells[,"Virion.Hits"])
   Cell.Summary[t,"Radius.25"] = quantile(Prod.Cells[,"Distance"],0.25)
   Cell.Summary[t,"Radius.50"] = quantile(Prod.Cells[,"Distance"],0.50)
   Cell.Summary[t,"Radius.75"] = quantile(Prod.Cells[,"Distance"],0.75)
   Cell.Summary[t,"Radius.90"] = quantile(Prod.Cells[,"Distance"],0.90)
   Cell.Summary[t,"Radius.95"] = quantile(Prod.Cells[,"Distance"],0.95)
   Cell.Summary[t,"Radius.975"] = quantile(Prod.Cells[,"Distance"],0.975)
   Cell.Summary[t,"Radius.Max"] = max(Prod.Cells[,"Distance"])
  } # Next time step (t)

write.table(Cell.Summary[1,],file="/home/rstudio/Dropbox/Modeling/Amazon_Output/N
TJ_Sim_Progress.txt",row.names=F)
  return(Cell.Summary)
}

# Run Simulation ----

Sim.Start = Sys.time()

registerDoMC(cores = 2) # Nate Macbook
registerDoMC(cores = 64) # Amazon Server

Cell.Summary = foreach(pp = 1:Pp_Num, .combine = rbind) %:%
  foreach(d = 1:Virus.D_Num, .combine = rbind) %:%
  foreach(spread = 1:Spread_Frac_Num, .combine = rbind) %:%
```

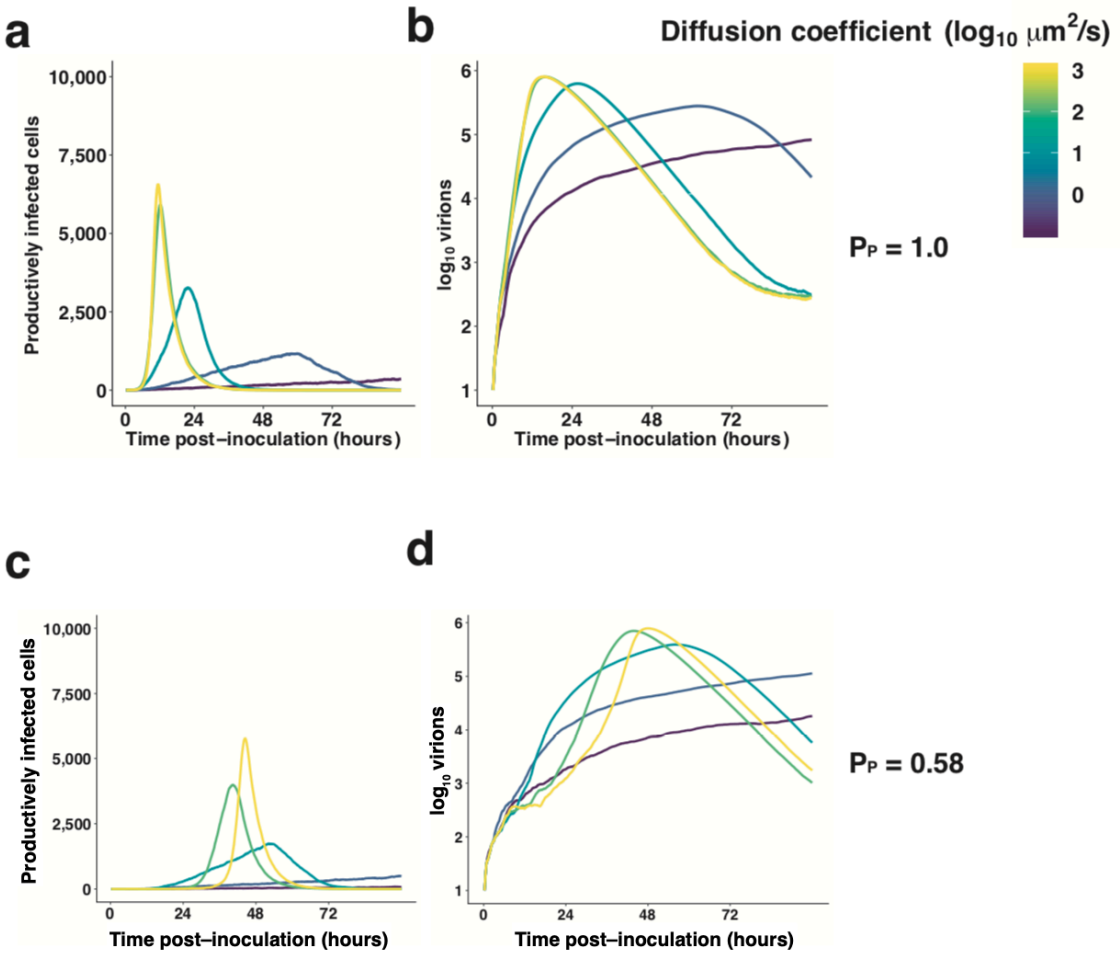
```
  foreach(sim = 1:Sim_Num, .combine = rbind) %dopar%
  try(Sim.Exec(Attach_Rate = Attach_Rate[1],
          Detach_Rate = Detach_Rate[1],
          Death_Rate = Death_Rate[1],
          Birth_Rate = Birth_Rate[1],
          Infect_Rate = Infect_Rate[1],
          Burst_Rate = Burst_Rate[1],
          Exclude_Rate = Exclude_Rate[1],
          Revert_Rate = Revert_Rate[1],
          Decay_Rate = Decay_Rate[1],
          Start_Radius = Start_Radius[1],
          Virus.D = Virus.D[d],
          Pp = Pp[pp],
          Spread_Frac = Spread_Frac[spread],
          Cells = Init.Cells,
          Virions = Init.Virions,
          Duration = Duration,
          ts = ts,
          Sim = sim,
          Varied.Par = "All"))

# QC ----
Sim.End = Sys.time()
```

5.3 Supplement to Chapter 3



Supplementary Figure 1

Bayes' rule was used to calculate the probability that each cell was infected with exactly 1 virion, based on the number of infected cells in each experiment, and each cell's combination of segment presences and absences. The distribution of probabilities is shown stratified by the number of segments present per cell.

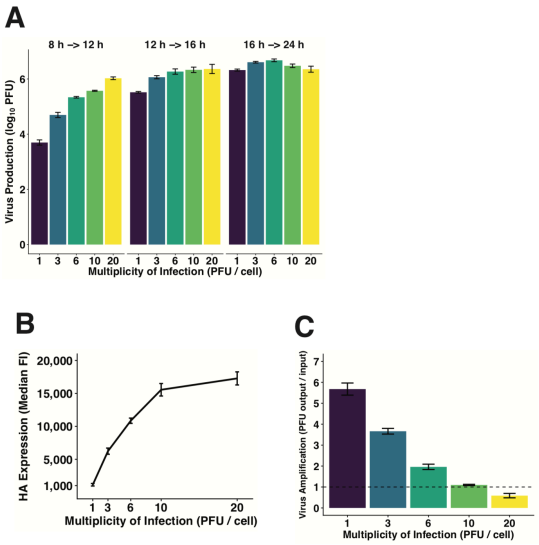


Supplementary Figure 2

(A, B) The dynamics of infection, in terms of productively infected cells (A) or virions present (B) are shown for a virus with complete genomes (Pp = 1.0).

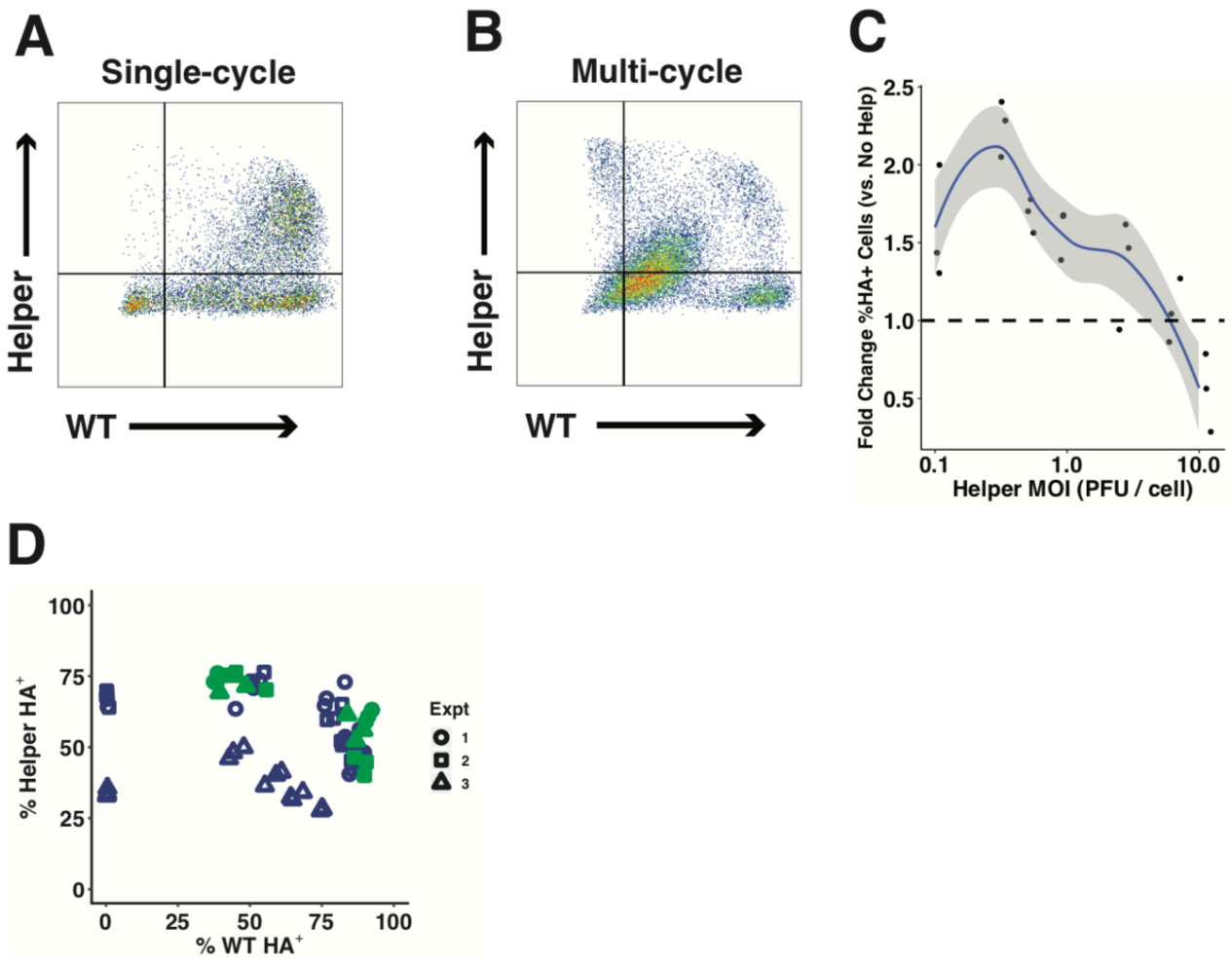
(C, D) The dynamics of infection, in terms of productively infected cells (C) or virions present (D) are shown for a virus with incomplete genomes (Pp = 0.58).

Lines are colored by diffusion coefficient.

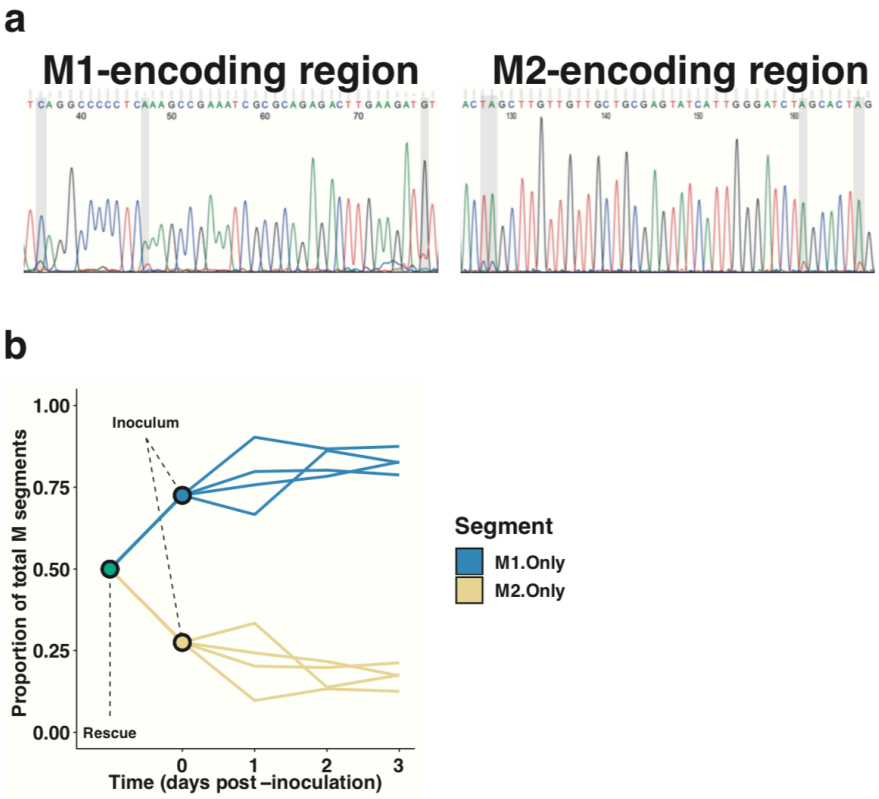


Supplementary Figure 3 (A) The amount of virus produced (in PFU) in three distinct time periods was calculated at each MOI. (B) The intensity of HA expression among HA^+ cells is shown across a range of MOIs.

(C) Amplification of virus was calculated by dividing the amount of virus produced (in PFU) after 48 hours by the number of PFU used to inoculate cells.



Supplementary Figure 4 (A,B) Representative flow cytometry measurement of Pan/99-WT HA when Pan/99-Helper virus is added simultaneously (A), or following 12 h of multi-cycle replication (B). Pan/99-WT MOI = 0.1 PFU/cell for simultaneous co-infection, 0.01 PFU/cell for multi-cycle replication. Pan/99-Helper virus MOI was 0.3 PFU/cell in both cases. (C) Cells were inoculated with Pan/99-WT (MOI = 0.01 PFU/cell) and Pan/99-Helper at a range of MOI, then incubated under single-cycle conditions before staining for expression of WT and Helper HA proteins. The extent to which Pan/99-Helper increased numbers of WT HA^+ cells (relative to controls infected with only Pan/99-WT) was calculated at each Pan/99-Helper MOI. Curve and ribbon represent mean and 95% confidence interval, respectively, of local regression.



Supplementary Figure 5 — Both M1.Only and M2.Only genome segments are maintained in vivo following infection with Pan/99-M.STOP virus.

(A) M segments of virus recovered from guinea pigs at 2 d post-inoculation with Pan/99-M.STOP virus were amplified by PCR, and the M1-encoding (left) and M2-encoding (right) regions were sequenced to verify maintenance of both alleles in vivo. Shaded gray rectangles highlight mutated regions.

(B) Guinea pigs were inoculated with Pan/99-M.STOP virus, and ddPCR was used to quantify copy numbers of M1.Only (blue) and M2.Only (yellow) segments, expressed as a proportion of all M segments. Blue and yellow dots represent the frequencies of M1.Only and M2.Only segments present in the virus stocks used to prepare inocula. The green dot represents the plasmid mixture used to generate the virus, in which equal molar quantities of M1.Only and M2.Only pDP plasmids were combined.

Supplementary Table 1 — Genotype of Pan/99-Helper virus

Segment	Mutations relative to Pan/99-WT
PB2	A550C, G552A, A555C, C556T, A617C, T621C, T622A, C623G
PB1	C346T, T348G, A351G, T441A, T444A, A447T
PA	G603A, T604A, C605G, C747T, T750C, G753A
HA	T308C, C311A, C313T, A464T, C467G, T470A
NP	C537T, T538A, C539G, G606C, A609T, G615C
NA	C418G, T421A, A424C, T511A, T514A, A517G
M	C413T, C415G, A418C, A517G, G523A, A526C
NS	A210C, G212A, G215A, T218C, C329T, C335T, A341G

Supplementary Table 2 — Primers for single-cell assay

Primer Name	Sequence (5' — 3')	Virus targeted
PB2 537F wt	TGAAGTGGGAGCCAGGATAC	Pan/99-WT
PB2 640R wt	ATGCAACCATCAAGGGAGAA	Pan/99-WT
PB1 332F wt	TTGAGAGCTCATGCCTTGAA	Pan/99-WT
PB1 459R wt	GTTGGCTAATGCAGTTGCTG	Pan/99-WT
PA 595F wt	TTTCGTCAGTCCGAAAGAGG	Pan/99-WT
PA 741R wt	AGCTTGCCCTCAATGCAGCCG	Pan/99-WT
HA 266F wt	ACCCTCATTGTGATGGCTTC	Pan/99-WT
HA 452R wt	GTTCCATTCTGAGCGACTCC	Pan/99-WT
NP 520F wt	ATGGATCCCAGAATGTGCTC	Pan/99-WT
NP 625R wt	TCAGCTCCATCACCATTGTC	Pan/99-WT
NA 408F wt	ATCAATTTGCCCTTGGACAG	Pan/99-WT
NA 528R wt	CCCAAATGAAATGGAACACC	Pan/99-WT
M 402F wt	GTTGCATGGGCCTCATATAC	Pan/99-WT
M 535R wt	ATTGGTTGTTGCCACCATTTG	Pan/99-WT
NS 173F wt	CCATGTTGGAAAGCAGATTG	Pan/99-WT
NS 321R wt	GGGCATTAGCATGAACCAGT	Pan/99-WT
PB2 537F var	TGAAGTGGGAGCCCGAATCT	Pan/99-Helper
PB2 640R var	ATGCAACCATCAACGGACTG	Pan/99-Helper
PB1 332F var	TTGAGAGCTCATGCTTGGAG	Pan/99-Helper
PB1 459R var	GTTGGCTAATGCTGTAGCAG	Pan/99-Helper

UnivF(A) + 6	GCGCGCAGCAAAAGCAGG	Pan/99-WT and Pan/99-Helper
UnivF(G) + 6	GCGCGCAGCGAAAGCAGG	Pan/99-WT and Pan/99-Helper

Supplementary Table 3 — Primers and probes for ddPCR

Primer/Probe Name	**Sequence (5' — 3')**
M2 F	ACTCATCCTAGCTCCAG
M2 R	CCGTGTTTGAAGAGTCG
M2.Only Probe (M2 WT)	HEX - CCATTCGTTTCTGATAGGTCTG - BHQ1
M1.Only Probe (M2 Mutant)	6-FAM - CCATACGCTTCTGGTACGTCTG - BHQ1
NS F	ACCTGCTTCGCGATACATAAC
NS R	AGGGGTCCTTCCACTTTTTG
NS Probe	6-FAM – AGAAACTGGTTCATGCTAATGCCCA - BHQ1

5.4 References

1. CDC, *CDC Wonder Database (Compressed Mortality File)*. 2013.
2. Mina, M.J. and K.P. Klugman, *The role of influenza in the severity and transmission of respiratory bacterial disease.* Lancet Respir Med, 2014. **2**(9): p. 750-63.
3. Sakal, F., et al., *Single-Plex Quantitative Assays for the Detection and Quantification of Most Pneumococcal Serotypes.* Plos One, 2015. **10**(3).
4. Turner, P., et al., *Improved Detection of Nasopharyngeal Cocolonization by Multiple Pneumococcal Serotypes by Use of Latex Agglutination or Molecular Serotyping by Microarray.* Journal of Clinical Microbiology, 2011. **49**(5): p. 1784-1789.
5. Devevey, G., et al., *First arrived takes all: inhibitory priority effects dominate competition between co-infecting Borrelia burgdorferi strains.* BMC Microbiol, 2015. **15**: p. 61.
6. Fgaier, H., et al., *An allelopathy based model for the Listeria overgrowth phenomenon.* Math Biosci, 2014. **247**: p. 13-26.
7. Lam, L.H. and D.M. Monack, *Intraspecies competition for niches in the distal gut dictate transmission during persistent Salmonella infection.* PLoS Pathog, 2014. **10**(12): p. e1004527.
8. Nedialkova, L.P., et al., *Inflammation fuels colicin Ib-dependent competition of Salmonella serovar Typhimurium and E. coli in enterobacterial blooms.* PLoS Pathog, 2014. **10**(1): p. e1003844.
9. Maldonado, A., R. Jimenez-Diaz, and J.L. Ruiz-Barba, *Induction of plantaricin production in Lactobacillus plantarum NC8 after coculture with specific gram-positive bacteria is mediated by an autoinduction mechanism.* J Bacteriol, 2004. **186**(5): p. 1556-64.

10. Maldonado-Barragan, A., et al., *Induction of bacteriocin production by coculture is widespread among plantaricin-producing Lactobacillus plantarum strains with different regulatory operons.* Food Microbiol, 2013. **33**(1): p. 40-7.
11. Khan, F., et al., *Streptococcus pneumoniae Eradicates Preformed Staphylococcus aureus Biofilms through a Mechanism Requiring Physical Contact.* Front Cell Infect Microbiol, 2016. **6**: p. 104.
12. Regev-Yochay, G., et al., *Interference between Streptococcus pneumoniae and Staphylococcus aureus: In vitro hydrogen peroxide-mediated killing by Streptococcus pneumoniae.* J Bacteriol, 2006. **188**(13): p. 4996-5001.
13. Gardner, A. and S.A. West, *Spite and the scale of competition.* J Evol Biol, 2004. **17**(6): p. 1195-203.
14. Gardner, A., S.A. West, and A. Buckling, *Bacteriocins, spite and virulence.* Proc Biol Sci, 2004. **271**(1547): p. 1529-35.
15. Gronroos, L., et al., *Mutacin production by Streptococcus mutans may promote transmission of bacteria from mother to child.* Infect Immun, 1998. **66**(6): p. 2595-600.
16. Stern, N.J., et al., *Isolation of a Lactobacillus salivarius strain and purification of its bacteriocin, which is inhibitory to Campylobacter jejuni in the chicken gastrointestinal system.* Antimicrob Agents Chemother, 2006. **50**(9): p. 3111-6.
17. Schamberger, G.P., et al., *Reduction of Escherichia coli O157:H7 populations in cattle by addition of colicin E7-producing E. coli to feed.* Appl Environ Microbiol, 2004. **70**(10): p. 6053-60.
18. Wisener, L.V., et al., *The use of direct-fed microbials to reduce shedding of Escherichia coli O157 in beef cattle: a systematic review and meta-analysis.* Zoonoses Public Health, 2015. **62**(2): p. 75-89.
19. Paul, R.E., et al., *Interspecific competition during transmission of two sympatric malaria parasite species to the mosquito vector.* Proc Biol Sci, 2002. **269**(1509): p. 2551-7.
20. Lysenko, E.S., et al., *The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces.* PLoS Pathog, 2005. **1**(1): p. e1.
21. May, R.M. and M.A. Nowak, *Coinfection and the evolution of parasite virulence.* Proc Biol Sci, 1995. **261**(1361): p. 209-15.
22. Bell, A.S., et al., *Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success.* Evolution, 2006. **60**(7): p. 1358-71.
23. de Roode, J.C., et al., *Dynamics of multiple infection and within-host competition in genetically diverse malaria infections.* Am Nat, 2005. **166**(5): p. 531-42.
24. de Roode, J.C., et al., *Virulence and competitive ability in genetically diverse malaria infections.* Proc Natl Acad Sci U S A, 2005. **102**(21): p. 7624-8.
25. Raberg, L., et al., *The role of immune-mediated apparent competition in genetically diverse malaria infections.* Am Nat, 2006. **168**(1): p. 41-53.
26. Barclay, V.C., et al., *The effect of immunodeficiency on the evolution of virulence: an experimental test with the rodent malaria Plasmodium chabaudi.* Am Nat, 2014. **184 Suppl 1**: p. S47-57.
27. Barclay, V.C., et al., *The evolutionary consequences of blood-stage vaccination on the rodent malaria Plasmodium chabaudi.* PLoS Biol, 2012. **10**(7): p. e1001368.

28. Lewnard, J.A., et al., *Epidemiological Markers for Interactions Among Streptococcus pneumoniae, Haemophilus influenzae, and Staphylococcus aureus in Upper Respiratory Tract Carriage.* J Infect Dis, 2016. **213**(10): p. 1596-605.
29. Lewnard, J.A., et al., *Density, Serotype Diversity, and Fitness of Streptococcus pneumoniae in Upper Respiratory Tract Cocolonization With Nontypeable Haemophilus influenzae.* J Infect Dis, 2016. **214**(9): p. 1411-1420.
30. Lysenko, E.S., et al., *Within-host competition drives selection for the capsule virulence determinant of Streptococcus pneumoniae.* Curr Biol, 2010. **20**(13): p. 1222-6.
31. van Baalen, M. and M.W. Sabelis, *The Dynamics of Multiple Infection and the Evolution of Virulence.* The American Naturalist, 1995. **146**(6): p. 881-910.
32. Bremermann, H.J. and J. Pickering, *A game-theoretical model of parasite virulence.* J Theor Biol, 1983. **100**(3): p. 411-26.
33. Sachs, J.L., et al., *The evolution of cooperation.* Q Rev Biol, 2004. **79**(2): p. 135-60.
34. Alizon, S., *Parasite co-transmission and the evolutionary epidemiology of virulence.* Evolution, 2013. **67**(4): p. 921-33.
35. Griffin, A.S., S.A. West, and A. Buckling, *Cooperation and competition in pathogenic bacteria.* Nature, 2004. **430**(7003): p. 1024-7.
36. Baldan, R., et al., *Adaptation of Pseudomonas aeruginosa in Cystic Fibrosis airways influences virulence of Staphylococcus aureus in vitro and murine models of co-infection.* PLoS One, 2014. **9**(3): p. e89614.
37. Viberg, L.T., et al., *Within-Host Evolution of Burkholderia pseudomallei during Chronic Infection of Seven Australasian Cystic Fibrosis Patients.* MBio, 2017. **8**(2).
38. McAdam, P.R., et al., *Adaptive evolution of Staphylococcus aureus during chronic endobronchial infection of a cystic fibrosis patient.* PLoS One, 2011. **6**(9): p. e24301.
39. Andersen, S.B., et al., *Long-term social dynamics drive loss of function in pathogenic bacteria.* Proc Natl Acad Sci U S A, 2015. **112**(34): p. 10756-61.
40. Winstanley, C., S. O'Brien, and M.A. Brockhurst, *Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections.* Trends Microbiol, 2016. **24**(5): p. 327-37.
41. Harrison, F., et al., *Cooperation and virulence in acute Pseudomonas aeruginosa infections.* BMC Biol, 2006. **4**: p. 21.
42. Ciofu, O., et al., *Genetic adaptation of Pseudomonas aeruginosa during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in mucA and/or lasR mutants.* Microbiology, 2010. **156**(Pt 4): p. 1108-19.
43. Hogardt, M. and J. Heesemann, *Adaptation of Pseudomonas aeruginosa during persistence in the cystic fibrosis lung.* Int J Med Microbiol, 2010. **300**(8): p. 557-62.
44. Hoboth, C., et al., *Dynamics of adaptive microevolution of hypermutable Pseudomonas aeruginosa during chronic pulmonary infection in patients with cystic fibrosis.* J Infect Dis, 2009. **200**(1): p. 118-30.

45. Wong, S.M., et al., *Genome-wide fitness profiling reveals adaptations required by Haemophilus in coinfection with influenza A virus in the murine lung.* Proc Natl Acad Sci U S A, 2013. **110**(38): p. 15413-8.
46. Kimaro Mlacha, S.Z., et al., *Transcriptional adaptation of pneumococci and human pharyngeal cells in the presence of a virus infection.* BMC Genomics, 2013. **14**: p. 378.
47. Chao, Y., et al., *Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease.* Front Cell Infect Microbiol, 2014. **4**: p. 194.
48. Galante, J., et al., *Quorum sensing and biofilms in the pathogen, Streptococcus pneumoniae.* Curr Pharm Des, 2015. **21**(1): p. 25-30.
49. Ercibengoa, M., et al., *Dynamics of pneumococcal nasopharyngeal carriage in healthy children attending a day care center in northern Spain. Influence of detection techniques on the results.* Bmc Infectious Diseases, 2012. **12**.
50. Brooke, C.B., et al., *Most influenza a virions fail to express at least one essential viral protein.* J Virol, 2013. **87**(6): p. 3155-62.
51. Nakatsu, S., et al., *Complete and Incomplete Genome Packaging of Influenza A and B Viruses.* MBio, 2016. **7**(5).
52. Chou, Y.Y., et al., *One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis.* Proc Natl Acad Sci U S A, 2012. **109**(23): p. 9101-6.
53. Wu, X., et al., *Competitive Dominance within Biofilm Consortia Regulates the Relative Distribution of Pneumococcal Nasopharyngeal Density.* Appl Environ Microbiol, 2017. **83**(16).
54. Rodgers, G.L. and K.P. Klugman, *The future of pneumococcal disease prevention.* Vaccine, 2011. **29**(Supplement 3): p. C43-C48.
55. Weiser, J.N., *The pneumococcus: why a commensal misbehaves.* Journal of Molecular Medicine-Jmm, 2010. **88**(2): p. 97-102.
56. Chao, Y.S., et al., *Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease.* Frontiers in Cellular and Infection Microbiology, 2015. **4**.
57. Gilley, R.P. and C.J. Orihuela, *Pneumococci in biofilms are non-invasive: implications on nasopharyngeal colonization.* Frontiers in Cellular and Infection Microbiology, 2014. **4**.
58. Hall-Stoodley, L., et al., *Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates.* Bmc Microbiology, 2008. **8**.
59. Shakl, J.R., J.E. Vidai, and K.P. Klugman, *Influence of bacterial interactions on pneumococcal colonization of the nasopharynx.* Trends in Microbiology, 2013. **21**(3): p. 129-135.
60. Simell, B., et al., *The fundamental link between pneumococcal carriage and disease.* Expert Review of Vaccines, 2012. **11**(7): p. 841-855.
61. Geno, K.A., et al., *Pneumococcal Capsules and Their Types: Past, Present, and Future.* Clinical Microbiology Reviews, 2015. **28**(3): p. 871-899.
62. Satzke, C., et al., *The PneuCarriage Project: A Multi-Centre Comparative Study to Identify the Best Serotyping Methods for Examining Pneumococcal Carriage in Vaccine Evaluation Studies.* Plos Medicine, 2015. **12**(11).

63. Domenech, M., E. Garcia, and M. Moscoso, *Biofilm formation in Streptococcus pneumoniae.* Microbial Biotechnology, 2012. **5**(4): p. 455-465.
64. Perez, A.C., et al., *Residence of Streptococcus pneumoniae and Moraxella catarrhalis within polymicrobial biofilm promotes antibiotic resistance and bacterial persistence.* Pathogens and Disease, 2014. **70**(3): p. 280-288.
65. Tikhomirova, A. and S.P. Kidd, *Haemophilus influenzae and Streptococcus pneumoniae: living together in a biofilm.* Pathogens and Disease, 2013. **69**(2): p. 114-126.
66. Weimer, K.E.D., et al., *Coinfection with Haemophilus influenzae Promotes Pneumococcal Biofilm Formation during Experimental Otitis Media and Impedes the Progression of Pneumococcal Disease.* Journal of Infectious Diseases, 2010. **202**(7): p. 1068-1075.
67. Carrolo, M., et al., *Pherotype Influences Biofilm Growth and Recombination in Streptococcus pneumoniae.* Plos One, 2014. **9**(3).
68. Vidal, J.E., et al., *Quorum-Sensing Systems LuxS/Autoinducer 2 and Com Regulate Streptococcus pneumoniae Biofilms in a Bioreactor with Living Cultures of Human Respiratory Cells.* Infection and Immunity, 2013. **81**(4): p. 1341-1353.
69. Wei, H. and L.S. Havarstein, *Fratricide Is Essential for Efficient Gene Transfer between Pneumococci in Biofilms.* Applied and Environmental Microbiology, 2012. **78**(16): p. 5897-5905.
70. Vidal, J.E., et al., *The LuxS-Dependent Quorum-Sensing System Regulates Early Biofilm Formation by Streptococcus pneumoniae Strain D39.* Infection and Immunity, 2011. **79**(10): p. 4050-4060.
71. Marks, L.R., G.I. Parameswaran, and A.P. Hakansson, *Pneumococcal Interactions with Epithelial Cells Are Crucial for Optimal Biofilm Formation and Colonization In Vitro and In Vivo.* Infection and Immunity, 2012. **80**(8): p. 2744-2760.
72. Wholey, W.-Y., et al., *Coordinated Bacteriocin Expression and Competence in Streptococcus pneumoniae Contributes to Genetic Adaptation through Neighbor Predation.* Plos Pathogens, 2016. **12**(2).
73. Hall, B.G., et al., *Growth Rates Made Easy.* Molecular Biology and Evolution, 2014. **31**(1): p. 232-238.
74. Carvalho, M.d.G.S., et al., *Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA.* Journal of Clinical Microbiology, 2007. **45**(8): p. 2460-2466.
75. Sakai, F., et al., *Expression of Streptococcus pneumoniae Virulence-Related Genes in the Nasopharynx of Healthy Children.* Plos One, 2013. **8**(6).
76. Hanke, C.R., et al., *Bacterial Density, Serotype Distribution and Antibiotic Resistance of Pneumococcal Strains from the Nasopharynx of Peruvian Children Before and After Pneumococcal Conjugate Vaccine 7.* Pediatric Infectious Disease Journal, 2016. **35**(4): p. 432-439.
77. Levin, B.R. and K.I. Udekwu, *Population Dynamics of Antibiotic Treatment: a Mathematical Model and Hypotheses for Time-Kill and Continuous-Culture Experiments.* Antimicrobial Agents and Chemotherapy, 2010. **54**(8): p. 3414-3426.
78. Pimenta, F.C., et al., *Sequential Triplex Real-Time PCR Assay for Detecting 21 Pneumococcal Capsular Serotypes That Account for a High Global Disease Burden.* Journal of Clinical Microbiology, 2013. **51**(2): p. 647-652.

79. Shak, J.R., et al., *Impact of Experimental Human Pneumococcal Carriage on Nasopharyngeal Bacterial Densities in Healthy Adults.* Plos One, 2014. **9**(6).
80. Tettelin, H., et al., *Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.* Science, 2001. **293**(5529): p. 498-506.
81. Trappetti, C., et al., *LuxS Mediates Iron-Dependent Biofilm Formation, Competence, and Fratricide in Streptococcus pneumoniae.* Infection and Immunity, 2011. **79**(11): p. 4550-4558.
82. Trzcinski, K., et al., *Effect of Serotype on Pneumococcal Competition in a Mouse Colonization Model.* Mbio, 2015. **6**(5).
83. Manning, J., et al., *Investigation of Streptococcus salivarius-mediated inhibition of pneumococcal adherence to pharyngeal epithelial cells.* Bmc Microbiology, 2016. **16**.
84. Wong, S.-S., et al., *Inhibition of Streptococcus pneumoniae adherence to human epithelial cells in vitro by the probiotic Lactobacillus rhamnosus GG.* BMC research notes, 2013. **6**: p. 135-135.
85. Johnson, H.L., et al., *Systematic Evaluation of Serotypes Causing Invasive Pneumococcal Disease among Children Under Five: The Pneumococcal Global Serotype Project.* Plos Medicine, 2010. **7**(10).
86. Adegbola, R.A., et al., *Carriage of Streptococcus pneumoniae and Other Respiratory Bacterial Pathogens in Low and Lower-Middle Income Countries: A Systematic Review and Meta-Analysis.* Plos One, 2014. **9**(8).
87. Kamng'ona, A.W., et al., *High multiple carriage and emergence of Streptococcus pneumoniae vaccine serotype variants in Malawian children.* Bmc Infectious Diseases, 2015. **15**.
88. Valente, C., et al., *Impact of the 13-valent pneumococcal conjugate vaccine on Streptococcus pneumoniae multiple serotype carriage.* Vaccine, 2016. **34**(34): p. 4072-4078.
89. Dawid, S., A.M. Roche, and J.N. Weiser, *The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo.* Infection and Immunity, 2007. **75**(1): p. 443-451.
90. Valente, C., et al., *The blp Locus of Streptococcus pneumoniae Plays a Limited Role in the Selection of Strains That Can Cocolonize the Human Nasopharynx.* Applied and Environmental Microbiology, 2016. **82**(17): p. 5206-5215.
91. Marks, L.R., R.M. Reddinger, and A.P. Hakansson, *High Levels of Genetic Recombination during Nasopharyngeal Carriage and Biofilm Formation in Streptococcus pneumoniae.* Mbio, 2012. **3**(5).
92. Khan, F., et al., *Streptococcus pneumoniae Eradicates Preformed Staphylococcus aureus Biofilms through a Mechanism Requiring Physical Contact.* Frontiers in Cellular and Infection Microbiology, 2016. **6**.
93. Jacobs, N.T., et al., *Incomplete influenza A virus genomes are abundant but readily complemented during spatially structured viral spread.* 2019: p. 529065.
94. Pauly, M.D., M.C. Procario, and A.S. Lauring, *A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses.* Elife, 2017. **6**.
95. Chao, L., T.T. Tran, and T.T. Tran, *The advantage of sex in the RNA virus phi 6.* Genetics, 1997. **147**(3): p. 953-959.

96. Ince, W.L., et al., *Reassortment complements spontaneous mutation in influenza A virus NP and M1 genes to accelerate adaptation to a new host.* J Virol, 2013. **87**(8): p. 4330-8.
97. Chao, L., T.R. Tran, and C. Matthews, *Muller Ratchet and the Advantage of Sex in the Rna Virus-Phi-6.* Evolution, 1992. **46**(2): p. 289-299.
98. Chao, L., *Fitness of RNA virus decreased by Muller's ratchet.* Nature, 1990. **348**(6300): p. 454-5.
99. Luijckx, P., et al., *Higher rates of sex evolve during adaptation to more complex environments.* Proc Natl Acad Sci U S A, 2017. **114**(3): p. 534-539.
100. Becks, L. and A.F. Agrawal, *The evolution of sex is favoured during adaptation to new environments.* PLoS Biol, 2012. **10**(5): p. e1001317.
101. Marshall, N., et al., *Influenza virus reassortment occurs with high frequency in the absence of segment mismatch.* PLoS Pathog, 2013. **9**(6): p. e1003421.
102. Essere, B., et al., *Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses.* Proc Natl Acad Sci U S A, 2013. **110**(40): p. E3840-8.
103. White, M.C., J. Steel, and A.C. Lowen, *Heterologous Packaging Signals on Segment 4, but Not Segment 6 or Segment 8, Limit Influenza A Virus Reassortment.* J Virol, 2017. **91**(11).
104. Phipps, K.L., et al., *Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny.* J Virol, 2017. **91**(17).
105. Villa, M. and M. Lassig, *Fitness cost of reassortment in human influenza.* PLoS Pathog, 2017. **13**(11): p. e1006685.
106. Yang, J.R., et al., *Reassortment and mutations associated with emergence and spread of oseltamivir-resistant seasonal influenza A/H1N1 viruses in 2005-2009.* PLoS One, 2011. **6**(3): p. e18177.
107. Simonsen, L., et al., *The genesis and spread of reassortment human influenza A/H3N2 viruses conferring adamantane resistance.* Mol Biol Evol, 2007. **24**(8): p. 1811-20.
108. Nelson, M.I., et al., *Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918.* PLoS Pathog, 2008. **4**(2): p. e1000012.
109. Holmes, E.C., et al., *Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses.* PLoS Biol, 2005. **3**(9): p. e300.
110. Hutchinson, E.C., et al., *Genome packaging in influenza A virus.* J Gen Virol, 2010. **91**(Pt 2): p. 313-28.
111. Dou, D., et al., *Analysis of IAV Replication and Co-infection Dynamics by a Versatile RNA Viral Genome Labeling Method.* Cell Rep, 2017. **20**(1): p. 251-263.
112. Heldt, F.S., et al., *Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection.* Nat Commun, 2015. **6**: p. 8938.
113. Russell, A.B., C. Trapnell, and J.D. Bloom, *Extreme heterogeneity of influenza virus infection in single cells.* Elife, 2018. **7**.
114. Diefenbacher, M., J. Sun, and C.B. Brooke, *The parts are greater than the whole: the role of semi-infectious particles in influenza A virus biology.* Curr Opin Virol, 2018. **33**: p. 42-46.

115. Fonville, J.M., et al., *Influenza Virus Reassortment Is Enhanced by Semi-infectious Particles but Can Be Suppressed by Defective Interfering Particles.* PLoS Pathog, 2015. **11**(10): p. e1005204.
116. Brooke, C.B., et al., *Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility.* Proc Natl Acad Sci U S A, 2014. **111**(47): p. 16854-9.
117. Kemeny, J.G.S., J.L., *Finite Markov Chains*, F.W.H. Gehring, P.R., Editor. 1976, Springer-Verlag: New York Berlin Heidelberg Tokyo.
118. Grantham, M.L., et al., *Tyrosines in the influenza A virus M2 protein cytoplasmic tail are critical for production of infectious virus particles.* J Virol, 2010. **84**(17): p. 8765-76.
119. Chen, H., et al., *Partial and full PCR-based reverse genetics strategy for influenza viruses.* PLoS One, 2012. **7**(9): p. e46378.
120. Schwartz, S.L. and A.C. Lowen, *Droplet digital PCR: A novel method for detection of influenza virus defective interfering particles.* J Virol Methods, 2016. **237**: p. 159-165.
121. Lowen, A.C., et al., *The guinea pig as a transmission model for human influenza viruses.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9988-92.
122. Gao, Q. and P. Palese, *Rewiring the RNAs of influenza virus to prevent reassortment.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15891-6.
123. Lakdawala, S.S., et al., *Influenza a virus assembly intermediates fuse in the cytoplasm.* PLoS Pathog, 2014. **10**(3): p. e1003971.
124. Marsh, G.A., R. Hatami, and P. Palese, *Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions.* J Virol, 2007. **81**(18): p. 9727-36.
125. Marsh, G.A., et al., *Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging.* J Virol, 2008. **82**(5): p. 2295-304.
126. Han, G.Z. and M. Worobey, *Homologous recombination in negative sense RNA viruses.* Viruses, 2011. **3**(8): p. 1358-73.
127. Cheung, T.K., et al., *Generation of recombinant influenza A virus without M2 ion-channel protein by introduction of a point mutation at the 5' end of the viral intron.* J Gen Virol, 2005. **86**(Pt 5): p. 1447-54.
128. Watanabe, T., et al., *Influenza A virus can undergo multiple cycles of replication without M2 ion channel activity.* J Virol, 2001. **75**(12): p. 5656-62.
129. Takeda, M., et al., *Influenza a virus M2 ion channel activity is essential for efficient replication in tissue culture.* J Virol, 2002. **76**(3): p. 1391-9.
130. Lonsdale, R., et al., *A rapid method for immunotitration of influenza viruses using flow cytometry.* J Virol Methods, 2003. **110**(1): p. 67-71.
131. Fukuyama, S., et al., *Multi-spectral fluorescent reporter influenza viruses (Color-flu) as powerful tools for in vivo studies.* Nat Commun, 2015. **6**: p. 6600.
132. Gallagher, M.E., et al., *Causes and Consequences of Spatial Within-Host Viral Spread.* Viruses, 2018. **10**(11).
133. Karlsson, E.A., et al., *Visualizing real-time influenza virus infection, transmission and protection in ferrets.* Nat Commun, 2015. **6**: p. 6378.

134. Handel, A., et al., *How sticky should a virus be? The impact of virus binding and release on transmission fitness using influenza as an example.* J R Soc Interface, 2014. **11**(92): p. 20131083.
135. Widdicombe, J.H., *Regulation of the depth and composition of airway surface liquid.* J Anat, 2002. **201**(4): p. 313-8.
136. McCrone, J.T., et al., *Stochastic processes constrain the within and between host evolution of influenza virus.* Elife, 2018. **7**.
137. Lattar, S.M., et al., *A Mechanism of Unidirectional Transformation, Leading to Antibiotic Resistance, Occurs within Nasopharyngeal Pneumococcal Biofilm Consortia.* Mbio, 2018. **9**(3).
138. Weinberger, D.M., et al., *Pneumococcal capsular polysaccharide structure predicts serotype prevalence.* PLoS Pathog, 2009. **5**(6): p. e1000476.
139. Li, Y., et al., *Surface charge of Streptococcus pneumoniae predicts serotype distribution.* Infect Immun, 2013. **81**(12): p. 4519-24.
140. Hathaway, L.J., et al., *Capsule type of Streptococcus pneumoniae determines growth phenotype.* PLoS Pathog, 2012. **8**(3): p. e1002574.
141. Dawid, S., A.M. Roche, and J.N. Weiser, *The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo.* Infect Immun, 2007. **75**(1): p. 443-51.
142. Wholey, W.Y., et al., *Characterization of the Competitive Pneumocin Peptides of Streptococcus pneumoniae.* Front Cell Infect Microbiol, 2019. **9**: p. 55.
143. White, M.C., et al., *H5N8 and H7N9 packaging signals constrain HA reassortment with a seasonal H3N2 influenza A virus.* Proc Natl Acad Sci U S A, 2019.
144. Bolte, H., et al., *Packaging of the Influenza Virus Genome Is Governed by a Plastic Network of RNA- and Nucleoprotein-Mediated Interactions.* J Virol, 2019. **93**(4).
145. Le Sage, V., et al., *Non-Uniform and Non-Random Binding of Nucleoprotein to Influenza A and B Viral RNA.* Viruses, 2018. **10**(10).
146. Carlson, J.M., et al., *HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck.* Science, 2014. **345**(6193): p. 1254031.
147. Seladi-Schulman, J., et al., *Filament-producing mutants of influenza A/Puerto Rico/8/1934 (H1N1) virus have higher neuraminidase activities than the spherical wild-type.* PLoS One, 2014. **9**(11): p. e112462.