Specialization and Trade-Offs in The Evolution of Resistance to Cheating in The Social Amoeba

by
Michael Ethan Miller

A dissertation submitted to the Department of Biology and Biochemistry,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biology

Chair of Committee: Ricardo Azevedo

Committee Member: Elizabeth Ostrowski

Committee Member: Blaine Cole

Committee Member: Gregory Velicer

Committee Member: Rebecca Zufall

University of Houston

December 2019

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee, especially to Dr. Elizabeth A. Ostrowski, for guiding me through this endeavor and for bestowing all of the vital knowledge and experience needed to develop my skills as an academic. I am extremely grateful to my undergraduate advisors, Drs. Michael J. Greene, Hannah C. Anchordoquy, and Charles A. Ferguson, without whom I would not have thought a doctorate an obtainable goal. Many thanks to my cohort and all my lab mates, especially Drs. Michael de la Flor, Jennifer L. Myers, and Jennifer J. Kuzdzal-Fick for their continued support and needed distractions along the way. I would also like to thank all of the members of the Dicty lab group at Baylor College of Medicine for the invitations to present and for the valuable feedback, and especially Drs. Mariko Kurasawa and Rafael D. Rosengarten, and Mr. Christopher Quang Dung Dinh for their technical support and counsels. Lastly, I am deeply grateful for the love and support of my partner and our family and friends, without whom this dissertation truly would have never made it to realization.

ABSTRACT

A central enigma of social evolution is the emergence and maintenance of cooperation and altruism in the face of selfish, free-loading individuals (i.e., cheaters), which take advantage of cooperators. One mechanism to counter cheating is the evolution of resistors. Both cheating and resistance appear to be persistent and prevalent in nature. However, whether resistance evolves to specifically outcompete local cheats, and whether it comes at the cost of reduced fitness in alternate environments is largely unknown. Here, I describe several experiments to: (i) test the cost of adaptation in replicate populations of social amoebae evolved in the presence of different cheaters, (ii) describe the function, transcriptomics, and population genomics of mutations identified in my evolved populations, and (iii) describe the population structure of the social amoeba at varying scales. Population structure analyses suggest significant differentiation in this species at spatial scales as small as one meter. Resistance readily evolved within each selection environment and strains performed better within their respective environments, revealing the capacity for cheating resistance with a variety of cheater in *D. discoideum*. I identified several genes mutated in the evolved strains that potentially influence resistance to cheating, and interestingly, that balancing selection might affect several of these genes. Trade-offs, thought to limit adaptation and drive specialization, were observed between selection groups. Thus, rapid adaptation in response to specific cheaters might prevent fixation of one cheater type in natural populations and might facilitate local adaptation of cheaters and resistors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Cooperation and Conflict

Cooperation has been fundamental to life's successful increase of complexity. Undeniably, the major transitions in evolution are contingent upon the organization of smaller replicating units into cooperative groups where individuals incur a direct fitness cost for the benefit of the collective (Smith and Szathmary 1997). Individuals that maintain cohesion ultimately benefit from the division of labor among unions afforded through specialization and improved efficiency (Crespi 2001). From genes cooperating within the genome, cells cooperating within an organism, social insects cooperating as superorganisms, to humans cooperating in societies, we can see examples of one of the great products of evolution. So long as cooperation provides a competitive advantage to participating individuals (be they genes or organisms), natural selection should promote increasing levels of cooperation (West et al. 2007a). However, taking advantage of cooperators by receiving benefits from societies without contributing might provide an even greater advantage to individuals (Axelrod and Hamilton 1981). Therefore, cooperative groups are at risk of exploitation by selfish non-cooperators. These selfish individuals, called cheaters, can potentially invade populations, spread to fixation and leave populations vulnerable to exploitation (Frankham 2005; van Elsas et al. 2012). Mechanisms to avoid, mitigate, punish, or otherwise reduce their spread might thus be favored (Özkaya et al. 2017). Nevertheless, associations between cheaters and non-cheaters can persist (Ferriere et al. 2002; Ferrière et al. 2007). Understanding the nature of cooperative systems and how they achieve cohesiveness is, therefore, paramount to evolutionary and social biologists.

1.2 Mutualism and Interspecies Cheating

Mutualistic interactions can occur when two species exchange goods or services, mainly nutrition, energy, protection, or transport. These interactions are defined as being beneficial to both parties, and they can play a central role in ecosystem function (Boucher et al. 1982). However, the opportunity for cheating is prevalent in interspecies mutualisms (Boucher 1988; Ferriere et al. 2002; Ferrière et al. 2007; Jones et al. 2015). Many animal systems provide fascinating examples of cheaters invading mutualisms. For example, nectar robbers (encompassing some insects, birds, and bats) remove nectar without transporting pollen (Maloof and Inouye 2000; Richardson 2004; Rojas-nossa et al. 2016). Sometimes seemingly mutualistic relationships arise through coercion, as seen in the lycaenid caterpillars and their attendant ants. In this relationship, ants are fed a solution by the caterpillar that induces physiological changes in the ants, causing them to protect the caterpillars from other insect attacks (Fiedler 1991; Hojo et al. 2015).

1.3 Cooperation and Cheating in Microbes

Microorganisms provide an efficient way to study fundamental questions in evolution, and microbial cooperation is proving to be as ubiquitous as microbes themselves (Strassmann et al. 2000; Crespi 2001; Rainey and Rainey 2003; West et al. 2007b). Microbial life can also involve interactions between individuals that lead to cooperation and conflict. There are several well studied systems of microbial cheaters (Strassmann et al. 2000; Fiegna and Velicer 2003; Zhang et al. 2009). Cheating can arise when some individuals fail to produce metabolically expensive communal goods (Ennis et al. 2000; Crespi 2001; Fiegna and Velicer 2003; Rainey and Rainey 2003; H. Koschwanez et al. 2011; Zhang and Rainey 2013). For example, the social bacterium

Myxococcus xanthus undergoes multicellular development. *Myxococcus* cheaters that fail to produce C-factor (the product of the *csgA* gene) can free-ride through development when enough producers are present, but lose their advantage when at high frequency (Fiegna and Velicer 2003). *Pseudomonas aeruginosa,* a common microbe found in cystic fibrosis patients cooperates through sharing of communal goods, such as pyoverdine, a virulence factor for iron sequestration, and elastase, a virulence-enhancing protease. *P. aeruginosa* defectors that fail to produce either product can cheat and thereby outcompete wild-type cells, which can lead to population collapse (Dandekar et al. 2012; Asfahl et al. 2015; Wang et al. 2015). Cooperative cells of the budding yeast *Saccharomyces cerevisiae* secrete the enzyme invertase into the environment to digest sucrose into simple sugars for absorption. Cheating yeast cells fail to produce invertase, but reap the benefit of neighboring cooperators (H. Koschwanez et al. 2011). These examples provide ample opportunity to study the maintenance of cooperation in the presence of potential instability caused by cheaters.

1.4 Mechanisms of Cheater Control

Various mechanisms are thought to control the spread of selfish individuals and maintain cooperation within groups, including quorum sensing (Miller and Bassler 2001; Waters and Bassler 2005), policing (Manhes and Velicer 2011), reciprocity (Axelrod and Hamilton 1981; Inglis et al. 2014; Hilbe et al. 2018), and kin discrimination (Roulston et al. 2003; Ostrowski et al. 2008; Strassmann and Queller 2011; Strassmann et al. 2011). As mentioned earlier, communities of *Pseudomonas aeruginosa* benefit from the production of multiple communal goods. *P. aeruginosa* cooperators can mitigate cheating, however, using a cell-to-cell communication mechanism (quorum sensing) that to measures the density of cooperators and

controls gene expression, ensuring expensive communal goods are produced only when the strategy is beneficial and defectors are low. Interestingly, co-infection by two cheating *P. aeruginosa* types can also rescue populations risking collapse, suggesting that reciprocal cheating can be one means to maintain cooperation (Özkaya et al. 2018). In human societies, defection can be deterred through policing, thereby promoting cooperation by punishing defectors. Policing has been shown to deter cheating in microorganisms as well (Manhes and Velicer 2011), but policing is likely to be a costly endeavor, and may lead to avenues for cheating.

Where policing uses active punishment to suppress defectors, reciprocity limits cheating by directing immediate or future exchange of altruistic acts between individuals (direct) or among a group of individuals (indirect) that are also altruistic (Axelrod and Hamilton 1981; Nowak and Sigmund 2005; Hilbe et al. 2018). Reciprocity requires invoking higher order behavioral responses and is limited to systems where there is high probability of repeated encounters between any two individuals (Riolo et al. 2001, although see Nowak and Sigmund 2005). Lastly, kin can help to promote the evolution of altruism (Hamilton 1964a; Axelrod and Hamilton 1981; Nowak and Sigmund 2005; Dyken et al. 2011; Strassmann et al. 2011; Ho et al. 2013), and can help to maintain cooperation within groups by denying altruistic acts to strangers. However, it is still possible for cheating to arise in highly related groups (Fiegna and Velicer 2003; Fiegna et al. 2006; Santorelli et al. 2008; Hibbing et al. 2010; Butaitė et al. 2017; Özkaya et al. 2017). Within many microorganisms, for which ample evidence of cooperation exists (Diggle et al. 2007; West et al. 2007b; Strassmann et al. 2011; Celiker and Gore 2013), there is potential for mixing between different genotypes and opportunity for encountering non-

kin (Fortunato et al. 2003b; Vos and Velicer 2009), which sets the conditions were cheating can occur and necessitates mechanisms of cheater control.

1.5 Cooperation and Altruism in *D. discoideum*

Dictyostelium discoideum is a free-living amoeba, found primarily among leaf litter in temperate forest floors (Swanson et al. 1999). The free-living cells spend part of their life as individuals feeding on bacteria, but upon starvation, tens to hundreds of thousands of individuals culminate into a fruiting body where division of labor and cell differentiation occurs. In fruiting bodies, approximately 20% of the cells vacuolize and altruistically die to form a rigid stalk that holds the reproductive spores aloft (Bonner 2009). Fruiting body formation is thought to aid in dispersal and protection of spores and is therefore a significant system for the study of altruism and cooperation (Raper 1984; Bonner 2009). Furthermore, *D. discoideum* can be indefinitely frozen at the spore stage and recovered from frozen stocks, allowing for direct comparison between ancestral and evolved populations and aiding experimental evolution (Santorelli et al. 2008; Khare et al. 2009; Kuzdzal-Fick et al. 2011; Hollis 2012; Levin et al. 2015).

1.6 Cheating in *D. discoideum*

In *D. discoideum,* cheating is defined as a disproportionate representation of spores in chimaeras and a reduced allocation to stalk (Strassmann et al. 2000). Co-occurrence of multiple *D. discoideum* genotypes has been observed in natural populations of *D. discoideum* (Fortunato et al. 2003b). In 12 different pairwise mixes between two natural isolates from North Carolina, Strassmann, et al. (2000) found that half of the mixes did not result in fair chimerae, meaning

that one strain per chimera contributed more, proportionately, to prespore cells than to prestalk cells when forming chimeric slugs (an intermediate social stage).

Work to identify genes involved in cheating behavior in *D. discoideum* began with the use of genetic screens, where experimenters selected for a given social behavior (i.e., increased spore proportioning in chimerae) from a pool of REMI mutants (Ennis et al. 2000; Khare et al. 2009; Khare and Shaulsky 2010; Santorelli et al. 2013). REMI (Restriction Enzyme-Mediated Integration) mutants are created by the random insertion of a selectable plasmid, which creates random null mutations (Kuspa 2006). The first described cheater was *chtA* (Ennis et al. 2000), which is unable to produce normal fruiting bodies on its own, as it lacks a developmental Fbox protein necessary for signaling the development of pre-stalk cells. In chimera with even a small percentage of wild-type cells, however, *chtA* mutants can utilize the wild-type cells' signal, resulting in normal looking fruiting bodies (although with somewhat smaller sori), where *fbxA*- contributes only to the spores. This type of cheater, known as an 'obligate parasite' (Buss 1982; Khare and Shaulsky 2010) because it cannot fruit on its own, serves as an early example of the possibility of mechanisms that could induce cheating behavior in *D. discoideum*.

A later genetic screen for cheating revealed the possibility of numerous mechanisms for cheating (Santorelli et al. 2008), with the additional requirement that they form fruiting bodies on their own. The *chtB* mutant identified in the screen was shown to inhibit wild-type from producing spores. Additionally, *chtB* had no measurable pleiotropic costs to cheating, and produced normal fruiting bodies when clonal (Santorelli et al. 2013), thus a facultative cheater. *ChtC* is one the more well studied cheater genes in *D. discoideum,* which is also a facultative

cheater identified by Santorelli et al. (2008). The wild type *chtC* gene is transcribed late in development and regulates the maintenance of pre-stalk cell fate. Cells lacking *chtC* fail to maintain their pre-stalk cell fate in chimeras and transdifferentiate into pre-spore cells (meaning they initially differentiate into pre-stalk and later change cell fate), leaving the deficiency to be filled by the partner cells, and thus form a disproportionate fraction of spores in fruiting bodies (Khare and Shaulsky 2010). This cheating by deficiency is different than that of *chtB,* in that it does not appear to coerce partners into producing stalk by secreting a signal to alter their cell proportioning, but rather takes advantage of an existing developmental pathway and passively cheats (Khare and Shaulsky 2010). In addition to this mutant, an additional 128 unique cheater mutants were estimated to have been produced from the screen, showing that there are abundant pathways for conflict within populations of *D. discoideum* (Santorelli, et al. 2008).

1.7 Resistance to Cheating

In a REMI screen similar to that of Santorelli et al. (2008), but using the *chtC* mutant as a selective competitor, Khare et al (2009) identified a mutant (*rccA*) that was resistant to *chtC*. A resistant strain is one that has increased spore production (compared to controls) when in chimaera with a cheating strain (Khare et al. 2009). *RccA* was not resistant to a different cheater (LAS1, also known as *chtB*-), suggesting that there are possibly different mechanisms to cheating, and that resistance may be specific to each cheater type. *RccA* is also a noble resistor, meaning that in mixes with the parental strain, it does not cheat. Furthermore, this screen resulted in isolation of six additional mutants that showed resistance to *chtC.* However, their resistant phenotypes could not be causally linked to the insertion site, suggesting that random mutations during the selection environment led to the observed resistant phenotypes. Further

support that selection through natural random variation can produce resistance to cheating is provided by the experimental evolution work of Hollis (2012). Hollis repeatedly mixed and co-developed an evolving strain of *D. discoideum* with a non-evolving cheater. Hollis' strains evolved resistance within 10 social generations in the presence of a strong cheater, and showed evidence of multiple competitive advantages derived through spontaneous mutation alone. Together these experiments demonstrate that cheating could be ubiquitous, can potentially involve multiple pathways, and that resistance might evolve as a counter measure to avoid cheating.

1.8 Discussion

The previously described experiments have demonstrated that resistance can evolve in response to cheating in the laboratory (Khare et al. 2009; Manhes and Velicer 2011; Hollis 2012). Hollis (2012) allowed the *Dictyostelium discoideum* lab strain to evolve in the presence of a strong cheater of and selected for spores of the evolving wild-type. After 10 rounds of selection, increased spore production of the evolved wild-type compared to the ancestor was observed, indicating that *D. discoideum* can rapidly evolve resistance to cheating through spontaneous mutations and that arms races might occur in nature. Evolutionary arms race dynamics are characterized by reciprocal selective pressures imposed by two interacting lineages (Dawkins and Krebs 1979) resulting in recurrent selective sweeps (Stahl et al. 1999; Bergelson 2001). One caveat to Hollis' study is that vegetative growth and social development were not isolated, meaning that the observed did not necessarily depend on social interactions. Additionally, the mutations that arose were not determined. Khare et al. (2009) showed that resistance to a specific cheater could be selected for from a pool of mutants. However, because of how the

mutations were created in these screens, the cheaters and resistor described thus far by Khare et al. (2009), Santorelli et al. (2008), and Ennis et al. (2000), are limited to null mutants, and their phenotypic anomalies are from lack of function, typically within developmental signaling pathways.

In Chapter 2, I describe a social evolution experiment where the only source of genetic variation random mutation, which might allow for different mutations than those in knock out screens. Life cycle stages prior to development were intentionally isolated from social competition to maximize selection on the social development stage, and I conducted fitness assays to examine the level of resistance and the extent to which adaptations to cheating confer costs among the evolved strains. I evolved replicate populations in the presence of four, genetically different REMI cheater mutants with three major goals in mind. (1) Can resistance evolve simultaneously to multiple different cheaters? (2) Is there a cost to cheating resistance? (3) Is the evolution of resistance repeatable and predictable?

There has been discussion in the literature about the definition of cheating, what constitutes cheating, and whether all observations of unequal investment during development equates to cheating (Ghoul et al. 2013; Wolf et al. 2015; Martínez-García and Tarnita 2016). Here, my measurement of cheating, the disproportionate contribution to the spores in chimeric fruiting bodies, fits within the confines of the definition laid out by Ghoul et al. (2013), who define cheating as " a trait that is beneficial to a cheat and costly to a cooperator in terms of inclusive fitness" and "when these benefits and costs arise from the actor directing a cooperative behavior toward the cheat, rather than the intended recipient". That is to say, in this system, stalk

production is a costly product for cooperators to produce and cheaters benefit from its use, thus gaining a fitness advantage. I use this definition here because even if a cheating strain does not coerce its victim into producing additional stalk, stalk production is unfavorable unless used by its producer. In other words, if stalk is an expensive communal good, and any one genotype can benefit by using more of it (by taking up a larger proportion of the spores), even if they also contribute to making the stalk, they gain a fitness advantage over, and thus take advantage of those strains with which they co-develop. So outside users of that stalk, even if they do contribute equally, still gain the advantage of the stalk produced by others because they make better use of it. However, there are examples of mechanisms in the literature of how cheaters exploit wild-type cells and cause them preferentially form stalk (Ennis et al. 2000; Khare and Shaulsky 2010; Santorelli et al. 2013). The mechanism used to achieve superior spore representation need not be from either coercion or failure to produce communal goods, however. For example, in *D. discoideum*, development is triggered by starvation, and the first cells to do so begin propagation of extracellular signaling. Kuzdzal-Fick, Queller, and Strassmann (2010), showed that proportioning within fruiting bodies can be determined in part by the timing of development among cells in mixes; those that starve first are more likely to become spores. In other words, a population that adapts to initiate early development might hold an advantage and cheat late developers. I will discuss some of the measures taken to investigate these different behaviors within Chapter 2, where I describe assays of several traits that could cause cheating under my definition.

In Chapter 3, I describe a bioinformatic and genomic analysis based on genome re-sequencing of the evolved strains from Chapter 2. Experimental evolution is a powerful tool to investigate

the genetic basis of traits of interest (Cooper et al. 2003; Rainey and Rainey 2003; Nidelet and Kaltz 2007; Hollis 2012; Kawecki et al. 2012; Jasmin and Zeyl 2013). I evolved replicate populations (described in Chapter 2) started from a single ancestor in the presence of each of four genetically different cheaters and observed that evolved populations showed improved performance against their focal cheater, and sometimes, to other unfamiliar cheaters as well. Evolved clones from each population were assessed using whole-genome resequencing to: (1) identify the genetic basis of evolved traits, (2) identify any genetic parallelism (i.e., whether similar mutations arise across multiple evolved populations), thereby narrowing the pool of potential mutations that caused the resistance to cheating, and (3) assess whether candidate genes might be subject to selection in nature. Finally, I describe what is known in the literature about the function of mutated genes and how they might confer their effects on cheating and resistance. Briefly, I identified several different mutations in each strain (Five mutations on average), some of which are likely to be neutral or deleterious, in addition to any beneficial mutations. Furthermore, some mutations may be beneficial in the evolution environment, but do not impact cheating resistance, instead conferring adaptations to some other component of the environment (e.g., temperature, humidity, media). I then further narrowed my candidate pool by assessing patterns parallelism at the level of genes and gene function in part by leveraging an existing dataset consisting of genome-wide expression during development and pre-spore versus pre-stalk expression bias (Parikh et al. 2010). Based on the pool of candidates that were mutated in my evolution experiment, I undertook a population genomic analysis designed to understand how selection might be acting on these genes in nature. Additionally, using an existing population genetic analysis based on genome sequencing of 20 natural isolates of *D. discoideum* (Ostrowski et al. 2015), I explored the sequence polymorphism in my mutated genes found in

nature and compared their patterns of polymorphism to those of other genes in the genome with the goal of determining whether they, either collectively or individually, show signatures consistent with a history of selection in nature. Together, the combination of experimental evolution, genome sequencing, and population genomics constitutes a powerful approach to understanding the genetic basis of social traits, especially those that potentially mediate social conflict, and their evolutionary history in nature.

Finally, in Chapter 4, I describe a hierarchical investigation of the population structure of *D. discoideum* wild isolates. Here my goals were to (1) identify what constitutes a population in nature, (2) evaluate gene flow among these populations, and (3) understand the degree of differentiation in populations. Using collections of natural isolates isolated at varying distances from 1 m to 1000 km, I investigated the genetic variation within and among populations using microsatellite fragment analyses. I describe the results of multiple different cluster analyses, differentiation analyses, and discuss the current knowledge of microbial population structure.

As I have shown here, there are many examples in nature of cooperation even though the potential for defection is high. I aim to provide insight into the underlying means by which evolution promotes altruism and cooperation despite these potential vulnerabilities. Specifically, I aim to elucidate how cheating and resistance might persist and how stable equilibria might exist between these two traits in nature.

CHAPTER 2: EXPERIMENTAL EVOLUTION OF CHEATER RESISTANCE

2.1 Introduction

Cooperative societies can be exploited by individuals that receive group benefits without contributing their fair share of the costs (Hamilton 1964a; Axelrod and Hamilton 1981). These selfish cheaters, having a proximate fitness advantage over cooperators, can potentially invade cooperative systems and can generate local extinctions in laboratory experiments (Fiegna and Velicer 2003), which begs the question as to how cheating is moderated in nature. If cheating persists in nature, the spread of cheating might be mitigated by traits that help to enforce fairness (Rainey and Rainey 2003; Travisano and Velicer 2004; Rankin et al. 2007).

One way to limit the spread of cheating is to cooperate only with kin (Hamilton 1964a; Kuzdzal-Fick et al. 2011; Strassmann and Queller 2011; Ho et al. 2013). Hamilton's theory of inclusive fitness (kin selection) is formalized by Hamilton's rule, which states that altruistic traits (those that benefit others at the expense of the actor) will be selected when the sum of positive impact of survival and reproduction on relatives (indirect fitness) and cost to self-survival and reproduction (direct fitness) exceeds zero (Hamilton 1964a). Kin recognition may indeed be an important factor allow for the evolution of cooperation and restrain cheating (Mehdiabadi et al. 2006; Ostrowski et al. 2008; Strassmann and Queller 2011; Ho et al. 2013). However, kin selection may not completely suppress cheating (Van Dyken and Wade 2012), and may be difficult in systems where interactions occur frequently between unrelated individuals (Rousset and Roze 2007). In the absence of relatedness, cheating might also be restricted through the evolution of cheater-resistance, which could manifest in different ways. On one hand, resistance could occur through the evolution of super cheaters, which cheat the

cheater, potentially leading to an escalating arms race (Hollis 2012; Ostrowski et al. 2015; Özkaya et al. 2018). Alternatively, the success or fitness of a cheater may be inversely related to the frequency of resistors in the population, and vice versa, akin to the negative frequency dependent selection seen in host-parasite co-evolution (Kassen 2002; Brandt et al. 2007; Pruitt and Riechert 2009; Tellier and Brown 2011; Van Dyken and Wade 2012). Moreover, if cheating or resistance are both costly (e.g., from trade-offs) it might result in a balanced polymorphism and maintenance of both traits in a population.

There are several possible outcomes of for the adaptation of resistance to cheating that may or may not involve costs to adaptation (see Figure 2.1.1). Costs of resistance could arise if there are trade-offs where mutations that confer resistance negatively affect other traits (i.e., antagonistic pleiotropy), which could lead to the evolution of specialists. One possible outcome is that increased fitness to a given genotype (the direct response, DR) trades-off with inferior fitness with alternate genotypes (the correlated or cross response, CR), thus creating specialists (Figure 2.1.1A). However, trade-offs from pleiotropic costs are not necessary for specialists to evolve (Figure 2.1.1B). Instead, specialists can result from opportunity costs, where adapting to one genotype leaves the specialist at a disadvantage in alternate environments. There are also intermediate possibilities, such as the emergence of superior generalists (Figure 2.1.1C), where there is no cost of adaptation to some genotypes.

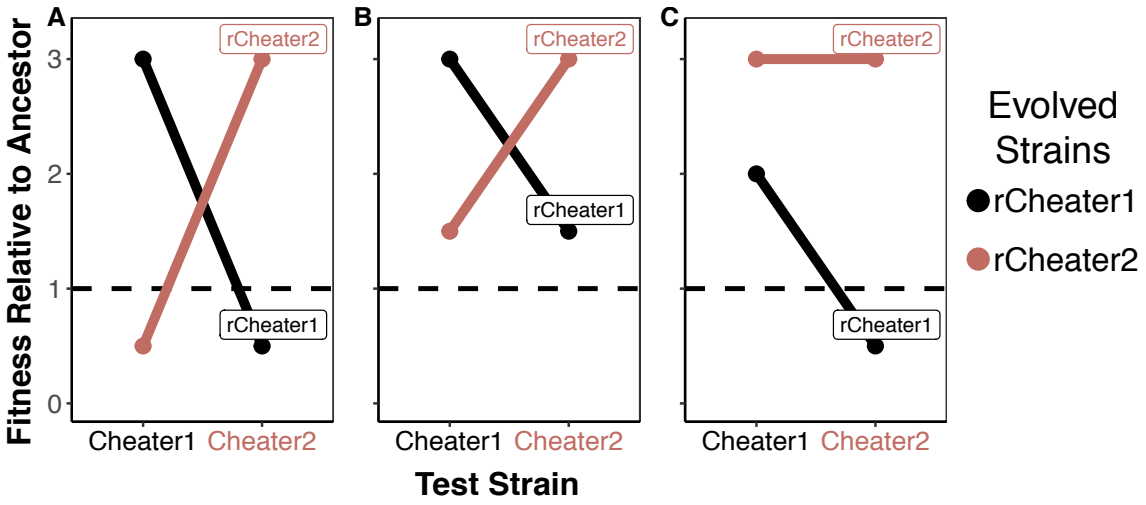

Figure 2.1.1 – Potential Reaction Norms. Genotype-by-genotype interactions between hypothetical evolved strains (Adapted from Nidelet & Kaltz, 2007). Lines represent two populations (rCheater1, rCheater2) evolved in the presence of two hypothetical cheaters (Cheater1, Cheater2). The dotted line indicates ancestral state (relative fitness of 1). (A) Adaptation to either cheater infers a cost in the form of reduced fitness to alternate cheaters, leading to specialists and local adaptation. (B) Opportunity cost of adaptation – evolved populations perform better than the ancestor with both novel and familiar cheaters, but Direct Response (DR) is greater than Correlated Response (CR). (C) Cost associated with adaptation to Cheater1, but not cost of adaptation to Cheater2, leading to superior generalists (rCheater2).

The social amoeba, *Dictyostelium discoideum,* is an excellent model system for studying the dynamics of the evolution of cooperation and conflict (Crespi 2001; Li and Purugganan 2011; Strassmann and Queller 2011; Celiker and Gore 2013; Jones et al. 2015). Upon starvation, the individual amoebae aggregate and form motile slugs (Raper 1984). These slugs develop into fruiting bodies consisting of a rigid stalk, which holds aloft and aids the dispersal of viable spores (Bonner 2009). The fate of stalk cells is to sacrifice themselves and die while providing a benefit to the reproductive spore cells by lifting them of the surface where they have a greater chance of dispersal and protection while in the dormant spore state (Bonner 2009).

Unlike many multicellular organisms, which pass through a single cell bottleneck at each generation, thereby maintaining high relatedness, *D. discoideum* amoebae become multicellular through aggregation. Aggregation potentially brings together diverse genotypes, which may or may not equally contribute to stalk formation, making *D. discoideum* vulnerable to cheaters (Ennis et al. 2000; Strassmann et al. 2000; Khare and Shaulsky 2010; Santorelli et al. 2013). When natural isolates of *D. discoideum* co-develop to form chimeric (mixed genotype) fruiting bodies, one genotype is sometimes preferentially represented among viable spores and underrepresented in the stalk (Strassmann et al. 2000; Buttery et al. 2010; Khare and Shaulsky 2010; Santorelli et al. 2013). This unequal sharing in the production of a communal good (the stalk) can be considered an example of cheating (Ennis et al. 2000; Strassmann et al. 2000; Foster et al. 2002; Fortunato et al. 2003a; Ostrowski et al. 2008; Santorelli et al. 2008; Khare et al. 2009; Buttery et al. 2010; Khare and Shaulsky 2010; Strassmann and Queller 2011; Kuzdzal-Fick et al. 2011; Hollis 2012; Santorelli et al. 2013; Ho et al. 2013; Levin et al. 2015).

In addition to studies of natural isolates, cheating in *D. discoideum* has been examined in the laboratory. For example, in an artificial selection experiment Santorelli et al. (2008) repeatedly passaged a genetically diverse pool of amoebae that had been subjected to random insertional mutagenesis. This process can enrich for those mutants that disproportionately form spores and avoid forming stalk. From the pool of mutants recovered, they identified over 100 genes, that when knocked out in the laboratory strain, caused cheating behavior, suggesting ample genetic opportunity for cheating in *D. discoideum*. In nature, multiple genotypes have been shown to co-occur in relatively small patches in the soil (Fortunato et al. 2003b). Fortunado et al. (2003b) collected 6 mm diameter soil samples averaging 0.2 g each and identified more than one

genotype in 63% of the samples. Furthermore, natural isolates can develop into chimerae with varying levels of cooperation and conflict (Strassmann et al. 2000). These findings suggest that different strains might encounter one another in nature and that these encounters might plausibly result in cheating.

Following Santorelli's work identifying genes for cheating behavior, Khare et al. (2009) performed a screen to identify mutants that suppress a known cheater (*chtC*; **ch**ea**t**er**C**), carried out by repeated rounds of co-development of the cheater with a pool of randomly mutated amoebae. The screen was designed to remove mutants that do not resist cheating and to enrich any cheater-resistant strains within the pool. The screen led to the identification of a single gene, *rccA* (**r**esistant to **c**heater **c**) that, when disrupted, conferred resistance to the *chtC*- mutant. In addition, although the *rccA*- mutant was resistant to the *chtC*- mutant, it was not resistant to a different cheater (LAS1). This result suggested that resistance could be specific to a particular target cheater. However, because Khare only selected for resistance to a single cheater (*chtC*-) and only looked into a single resistance mutation (*rccA*-), it is not known if this pattern will be generally found. Additionally, Hollis (2012) used laboratory experimental evolution to show that improvements in social fitness could arise rapidly. The experiment revealed that evolved populations can gain increased representation in the spores within 10 social generations, equivalent to approximately 160 cell generations in his experiment. Together, these experiments demonstrate that resistance can evolve rapidly as a counter measure to avoid cheating.

While it has been demonstrated that *D. discoideum* harbors numerous cheating pathways (Santorelli et al. 2008) and resistance can potentially evolve to counter cheating (Khare et al. 2009; Hollis 2012; Levin et al. 2015), it is not known how these adaptations evolve. On one hand, resistance could result in specialists that are resistant only to those cheaters they have encountered previously. Alternatively, resistance could result in generalists that are able to resist a wide variety of cheaters. Which of these patterns is found (generalism or specialism) may have important consequences for the evolutionary dynamics of cheating behaviors. For example, locally adapted specialists may emerge if the evolution of resistance is costly. In addition, novel cheaters may be able to invade populations when local genotypes are only resistant to the resident cheaters. Finally, specialists may only have a fitness advantage if their target cheater is at high frequency in the population. More generally, identifying adaptations to counter cheating, as well as potential fitness costs of resistance, may help to clarify why both cheating and cooperation appear to be persistent and pervasive in social systems (Axelrod and Hamilton 1981; Crespi 2001; Ferriere et al. 2002; Fiegna and Velicer 2003; Rainey and Rainey 2003; Griffin et al. 2004; West et al. 2007a,b; Diggle et al. 2007; Ferrière et al. 2007; Celiker and Gore 2013; Levin et al. 2015).

Here, I used experimental evolution to test the nature of resistance to cheating in *D. discoideum*. Starting from a labeled ancestor (AX4-GFP), I evolved six replicate populations in five different environments. Environments consisted of populations evolving in the presence of each of four different cheaters or a non-cheating control, for a total of 30 evolved lines (Figure 2.2.1). In each round of the experiment, the evolving populations (GFP-labeled, G418-resistant) were co-developed in equal proportions with their respective non-evolving cheaters (CD1-CD4) or the

ancestor (AX4, a non-cheating control). Following each round of development, the two competitors were separated by drug selection. The evolving populations were retained, regrown, and then co-developed again with the same, non-evolving competitor (either a cheater or the non-cheating control). Spontaneous mutation was the only source of genetic variation. Following 15 rounds of co-development, I isolated a single clone from each population. I first tested its level of direct resistance by determining how well it does in co-development with the cheater it experienced during its evolution. I then asked to what extent improvements against the focal cheater also confer improvements against each of the unfamiliar cheaters—those that the evolved strain did not experience during evolution. These assays were carried out as "cross resistance" assays, using a complete block design. In each block of the experiment, one or more evolved populations competed against all four cheaters (1 familiar, 3 novel) and the ancestor. Each block was replicated a minimum of three times. Second, I asked whether resistance is costly for germination rate or growth rate, and whether resistant strains cheat the ancestor, suggesting their resistance may really be a form of cheating.

2.2 Methods

2.2.1 Growth and Maintenance of Strains

The four cheater strains (CD1-4) were a gift of Adam Kuspa and Chris Dinh, Baylor College of Medicine. The strains were generated through random mutagenesis by using Restriction Enzyme Mediated Integration (REMI), whereby mutations are generated through the random integration of a selectable DNA fragment at common restriction sites (Kuspa 2006). These four strains were enriched and then isolated in an experiment where several mutant pools, each containing approximately 700 different mutants was propagated through multiple rounds of

development, a process that should select for strains that disproportionately form spores and avoid forming the stalk. The plasmid insertion site in each strain was identified by plasmid rescue and later confirmed using whole genome re-sequencing (Table 2.2.1). The ancestral strain for experimental evolution was a GFP-labeled axenic lab strain (AX4-GFP). This labeled strain was generated with the intent to induce multiple copies per cell of an extrachromosomal plasmid that expresses cytoplasmic GFP driven by the act15 promotor, which is maintained by the aminoglycoside G418 (geneticin). When maintained with 10 μg/ml G418, labeled cell cultures had an average of 97% fluorescence as measured by flow cytometry (Accuri C6, BD).

Table 2.2.1 – List of REMI disrupted genes in cheater mutants.

Strain	Gene Name	Name Description	Annotations
CD1	*DDB_G0272773*	NA	Contains a predicted signal peptide; there are two similar genes in *D. discoideum* (DDB_G0284765).
CD2	*srp54*	**S**ignal **R**ecognition **P**article **54** kDa subunit	Component of the signal recognition particle (SRP) - ensures correct targeting of nascent secretory proteins to the ER.
CD3	*gnt12*	**G**lc**N**Ac **T**ransferase	CAZy family GT49; similar to vertebrate acetylglucosaminyltransferase-like protein.
CD4	*DDB_G0289963*	NA	None

Frozen stocks were spotted on SM (Formedium, 2% Agar) supplemented with 400 μl Klebsiella pneumoniae (Kp) as a food source to obtain spores. Spores were inoculated into petri dishes containing 10 ml HL5 with glucose (Formedium) supplemented with PSV (10 μg/ml penicillin, 50 μg/ml streptomycin sulfate, 60 ng/ml cyanocobalamin, and 20 ng/ml folate). GFP-labeled strains were maintained with 10 μg/ml G418. Once cells reached confluence, the cultures were transferred to shaking flasks and maintained at 180 RPM and 22 °C. Cultures were maintained

in exponential phase through daily dilutions into fresh media for approximately one week prior to development assays.

2.2.2 G418 efficacy

Selection on 30 μg/ml G418 was found to be effective through two experiments. First, a mix of G418 resistant GFP-labeled WT with REMI mutant (with BSR) were subjected to drug selection. DNA was extracted and PCR confirmed that BSR was absent. In a second experiment, the culture was split after drug selection and half was allowed to grow without drug for 24 hours. I used FACS to collect cells below a detectable GFP expression level (to conservatively bias my sample towards potential remaining cheaters) and plated cells clonally on Kp/SM. I randomly chose 47 amoebae plaques, extracted DNA from the cells, and confirmed with PCR that all samples contained the G418 cassette sequence. This means that after G418 selection, even cells that are outside the gated threshold for GFP fluorescence are true G418 resistant cells.

2.2.3 Co-development Assays with Cheaters

Cells were harvested from shaking cultures during mid-exponential growth phase (3.5×10^6 cells/ml), washed twice in cold KK_2 buffer (per liter: KH_2PO_4 2.25 g, K_2HPO_4 0.67 g), and resuspended at 1×10^8 cells/ml in KK_2 buffer. GFP-labeled cells were then mixed with their non-labeled competitor at a 1:1 ratio and aliquots corresponding to 2×10^7 cells were deposited in a 6×6 square (corresponding to a density of 5.78×10^5 cells/cm) on 47 mm nitrocellulose filters (GVS Maine). Prepared filters were placed in 6 cm petri dishes atop Pall filter pads wetted with 1.5 ml PDF (20.1 mM KCl, 5.3 mM $MgCl_2\cdot6H_2O$, 9.2 mM K_2HPO_4, 13.2 mM KH_2PO_4, 0.5 g/L

streptomycin sulfate). At the same time, all strains were also developed clonally (without mixing with another strain). Plates were incubated at 22 °C in humid chambers with overhead light. After 48 hours filters were transferred to 50 ml Falcon tubes containing 5 ml detergent (KK_2 buffer, 0.1% IGEPAL, 20 mM EDTA), vortexed, and counted to determine total spore number using a hemocytometer or cell counter (Countess II, ThermoFisher). The percentage of cells expressing GFP was determined using a flow cytometer (Accuri C6, BD).

2.2.4 Experimental evolution

In each round of the experiment, GFP-labeled, G418-resistant evolving populations were co-developed at 1:1 with their respective unlabeled, G418-sensitive cheaters or the wild-type (WT) lab strain, as a control (Figure 2.2.1). Following co-development, spores were harvested from fruiting bodies and an aliquot was preserved in KK_2 buffer with 20% glycerol at -80 °C for later experimentation. The remaining spores were deposited in HL5 medium at a density of 5×10^4 spores/ml with 30 μg/ml of G418 to remove the cheater or non-cheating control strain. Selection using 30 μg/ml of G418 was determined to be sufficient at removing the cheater or control strain after co-development. During this time, each non-evolving cheater (CD1-CD4) and WT (cheaters and non-cheater, respectively) were inoculated onto SM plates from frozen stocks. Once all populations (the non-evolving cheaters, WT, and evolving populations) had reached exponential growth phase, the experiment was repeated. Each round of the experiment took approximately seven days, resulting in ~20 generations of vegetative growth, and was repeated for a total of 15 rounds of co-development. At the end of the experiment, there were 30 evolved populations (six populations/cheater × four cheaters, plus an additional six populations that evolved against the non-cheater control). To isolate a single clone from each evolved

population, the population was inoculated at low density on SM plates with 400 µl Kp to form well-spaced plaques. Spores were subsequently picked from the center of a single clearing, grown to high density, and the spores frozen for future experiments.

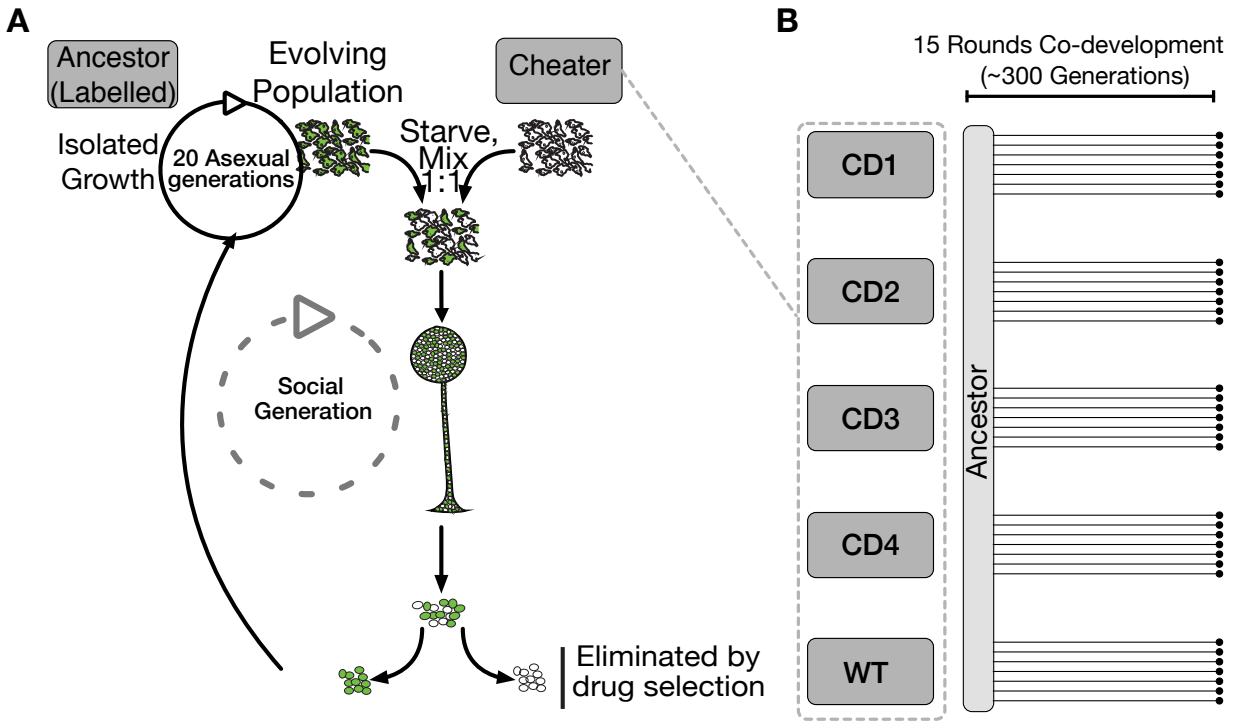

Figure 2.2.1 – Evolution Experiment Schematic. (A) GFP-labeled ancestor mixed and co-developed with naïve cheater or wild-type (unlabeled). In each round of the experiment the unlabeled strain was eliminated and only the evolving population was passaged to the next round. (B) Each evolution environment (consisting of a GFP-labeled evolving strain paired with an unlabeled cheater or control) was replicated in six parallel populations and allowed to evolve through 15 rounds of co-development.

2.2.5 Germination

To determine the germination efficiency, spores were plated from the freezer following the identical procedure used in the evolution experiment and then monitored by microscopy. Briefly, frozen stocks were spotted on SM agar plates (Formedium, 2% Agar) supplemented with 400 µl Klebsiella pneumoniae as a food source to obtain spores. Spores were diluted to

5×10^4/ml in HL5 and placed into wells of a 96 well plate. Wells were photographed immediately after plating and then again after incubation for 12 hours. Germination was calculated as (1-proportion of spores remaining after 12 hours incubation). The entire experiment was repeated three times.

2.2.6 Doubling time

To estimate the doubling time, the cell density of shaking cultures was measured throughout the co-development assays. The cell cultures were maintained at exponential growth throughout each assay by daily dilutions to 5×10^5 cells/ml. The cultures were grown in HL5 with glucose (Formedium) supplemented with PSV (10 μg/ml penicillin, 50 μg/ml streptomycin sulfate, 60 ng/ml cyanocobalamin, and 20 ng/ml folate) and GFP-labeled strains were maintained with 10 μg/ml G418. Cultures were kept at 22 ℃, shaking at 180 RPM with overhead light. At each dilution cell densities were measured by hemocytometer or automated cell counter (Countess II, ThermoFisher). The average of three cell density measurements per strain were then used to calculate the doubling time (dt) in hours with the formula: dt = duration×log(2)/(log(final density)-log(initial density). The entire process was repeated for each replication of the experiment.

2.2.7 Sporulation efficiency

To determine sporulation efficiency of each strain, each population was clonally developed alongside each co-development experiment, thus under identical conditions to the 1:1 mixes. After 48 hours, the filters were transferred to 50 ml Falcon tubes with 5 ml detergent (KK_2

buffer, 0.1% IGEPAL, 20 mM EDTA) to lyse any remaining cells and an aliquot was counted with a hemocytometer or using an automated cell counter (Countess II, ThermoFisher).

2.2.8 Statistical analyses

I first made linear mixed models with all factors (full models) for each test. I then dropped individual factors (reduced models) and estimated all models with the lmer package (Bates et al. 2015) in R (R Core Team 2018). Model selection was then carried out by comparing Akaike information criterion (AIC) for each model. The model having the lowest AIC was chosen as the best supported model. Finally, to test the significance of a given factor, I performed a likelihood ratio test (LRT) using the Chi-square statistic to compare the full and reduced models using the anova stats package (Chambers and Hastie 1992) in R.

2.3 Results

2.3.1 Experimental Mutants Cheat Wild-Type

To test whether the four REMI mutants (CD1-CD4) enriched in a prior screen for cheating behaviors do in fact cheat their parental strain (WT-GFP), I co-developed cells of each mutant with GFP-labeled wild-type cells (WT-GFP) in equal ratios. At the same time, I developed each strain clonally to test their sporulation efficiency. Note that although cells are starved before each experiment, cells that had undergone DNA replication before harvesting for co-development can divide after starvation, resulting in a value for sporulation efficiency that can be greater than 100%. When developed clonally, CD1-4 mutants do produce more spores than the wild-type (Figure 2.3.1), (*t*-tests, $P<0.05$). However, the observed shift in proportions (Figure 2.3.2, blue) is greater than estimated from sporulation efficiency (Figure 2.3.2, red). To

calculate the estimated proportions, I divided the clonal sporulation efficiency of WT-GFP by the total sporulation efficiency (clonal WT-GFP and clonal social partner) for each pairing on a given day. This calculation provides an estimate for the expected proportions in chimerae if no interaction occurs between strains. I prepared a linear mixed model with percent WT-GFP spores as the dependent variable, social partner (CD1-4) as a fixed variable, estimated or observed as a fixed variable, and block (date) as a random variable. I then performed a likelihood ratio test (LRT) to compare this model to a reduced model without the estimated vs observed factor. The LRT showed that observed proportions of WT-GFP spores were, on average, 4% lower than estimated proportions ($\chi^2(1) = 9.21$, $P < 0.01$). Finally, I recovered significantly fewer WT-GFP spores from chimerae with cheaters (CD1-CD4) compared to chimerae with WT (*t*-tests, all $P<0.05$), supporting the hypothesis that the four mutants cheat the wild-type (Figure 2.3.2, blue). Together, these results suggest that the overrepresentation of each mutant in the spores when co-developed with the wild-type strain is caused, at least in part, by a shift in its spore allocation that occurs in chimeras.

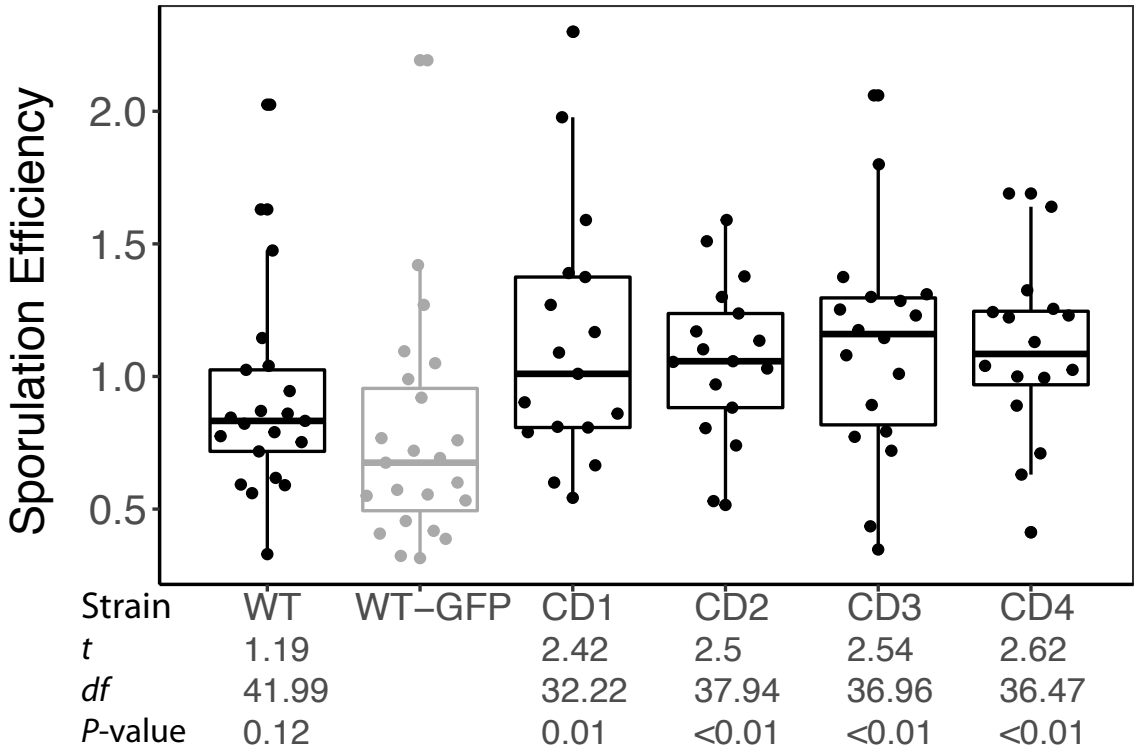

Figure 2.3.1 – CD mutants produce more spores than WT when developed clonally. Clonal sporulation efficiency of wild-type (WT), the labelled wild-type (WT-GFP), or the four mutants (CD1-CD4). Sporulation efficiency is calculated as the number of spores following development divided by the number of cells prior to development. In *t*-tests, the mutants CD1-CD4 produced more spores than WT-GFP when developed clonally.

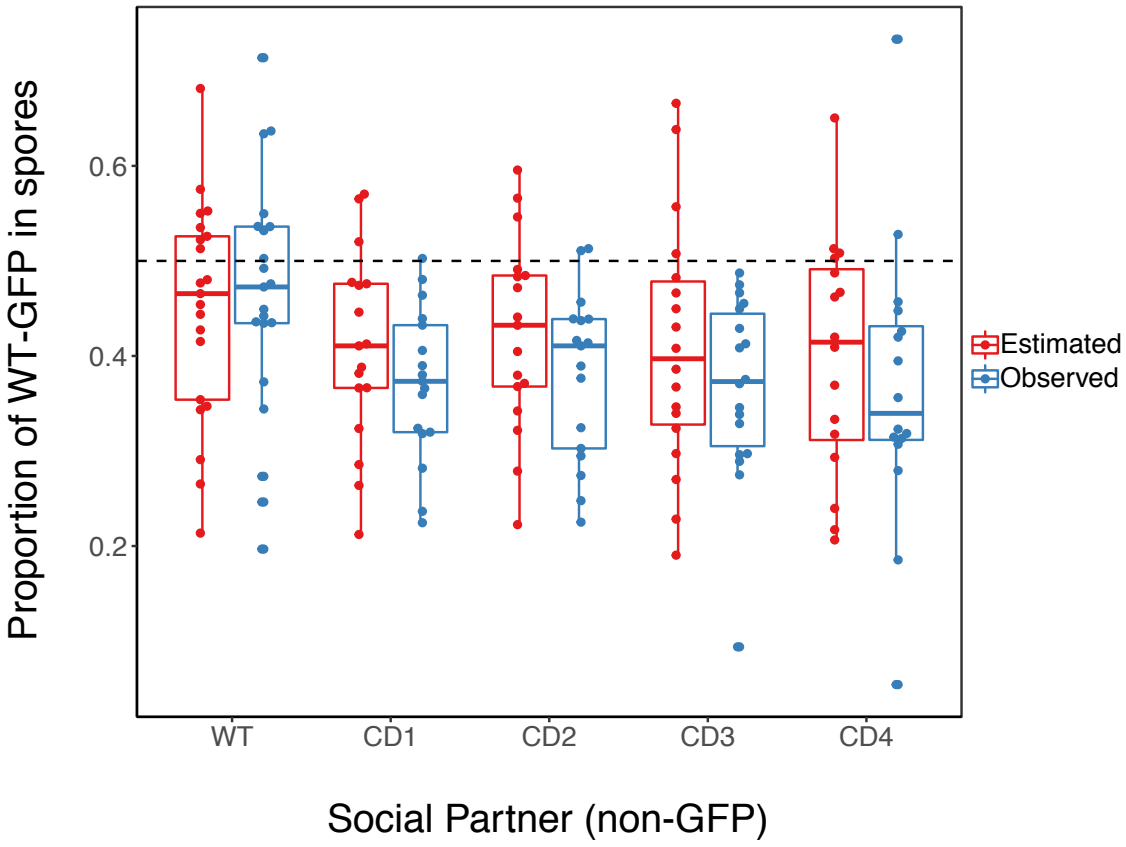

Figure 2.3.2 – Observed vs estimated cheating based on sporulation efficiency. Observed proportions (blue) indicate the measured proportion of WT-GFP spores recovered after co-development with cheaters or WT. Paired *t*-tests between each of the cheaters and WT for observed data (blue) were significant (all $P < 0.05$), suggesting that WT-GFP spores are suppressed when in chimera with cheaters. Estimated proportion of spores (red) indicates the estimated contribution of WT-GFP in spores when co-developed with each strain (WT or cheater) based on their development in pure culture (no interaction between strains). There was no significant difference between estimated means (red data points) for WT and any of the cheaters (paired *t*-tests, all $P > 0.05$). A mixed model ANOVA showed that, among cheaters (CD1-4), observed values (blue) were, on average, 4% lower than estimated values (red) and that the estimated vs observed factor significantly improves the fit of the model ($\chi^2 = 9.21$, $df = 1$, $P < 0.01$). These data suggest that although cheaters produce more spores than WT-GFP in pure culture, there is a significant shift in WT-GFP spores when in chimera with cheaters.

2.3.2 Evolved Strains Show Fitness Improvements Against Their Focal Cheaters

Each evolved strain was tested for its spore production in chimera with its focal cheater, and this performance was compared to that of the ancestor under the same conditions (Figure 2.3.3). Overall, evolved strains show a significantly higher representation in the spores compared to the ancestor ($\chi^2 = 6.2$, $df = 1$, $P = 0.013$), indicating response to selection. As a group, only CD1 and CD3 populations were significantly improved, whereas CD2 and CD4 populations were not (Table 2.3.1).

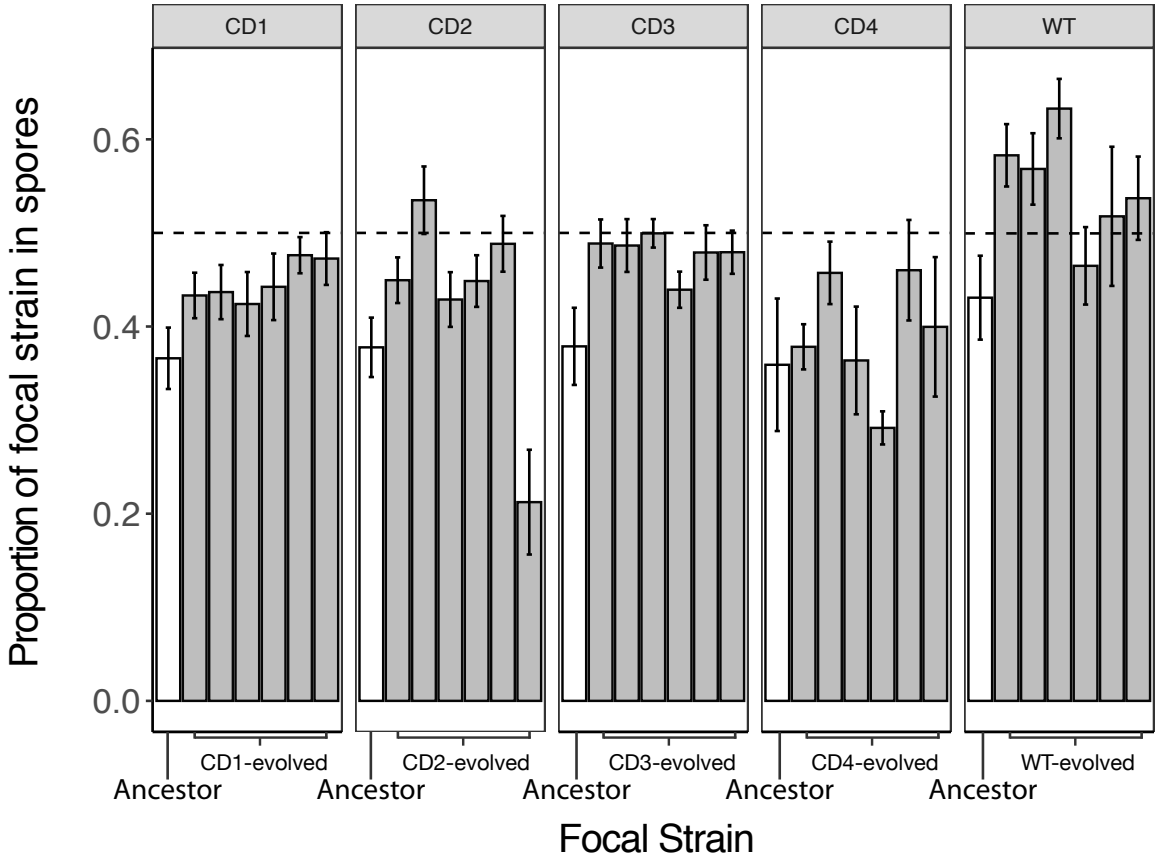

Figure 2.3.3 – Evolved populations have increased their share of the spores in chimeras with the target strain. Following 15 rounds of co-development, 28 out of 30 evolved populations show higher representation in the spores than the ancestor (sign-test, $Z = 3.1$, $P < 0.001$). Each panel shows the performance of the six lines that evolved with a given target strain (gray bars) as well as the performances of the ancestor against that strain (white bars) ± s.e.m. for comparison. The dotted line indicates the expected percentage of 50% when no cheating occurs.

Table 2.3.1 – Separate mixed model ANOVAs testing whether evolved strains show increased proportions in the spores compared to the ancestor when co-developed with their target strains. In the full models, the proportion of GFP spores is the dependent variable, whether the GFP strain was evolved or ancestor was considered to be a fixed effect, and population and block were treated as random effects. Significance is determined by comparing the fit of full and reduced models that lack a term of interest. In this case, the reduced model removed the term that indicated whether the strain was evolved or ancestral. These results suggest that evolved strains are significantly better only in the CD1 and CD3 environments.

Cheater Treatment	Model	*df*	AIC	dev	χ^2	*P*	
WT	Full	5	-73.87	-83.87	3.41	0.06	
	Reduced	4	-72.46	-80.46			
CD1	Full	5	-116.53	-126.53	6.77	0.01	*
	Reduced	4	-111.76	-119.76			
CD2	Full	5	-86.71	-96.71	0.36	0.55	
	Reduced	4	-88.35	-96.35			
CD3	Full	5	-111.83	-121.83	8.88	<0.01	*
	Reduced	4	-104.95	-112.95			
CD4	Full	5	-46.78	-56.78	0.62	0.43	
	Reduced	4	-48.16	-56.16			

I also asked whether there was significant variation among evolved lines within an environment (Table 2.3.2). These analyses indicate significant among-line variance in the non-cheating evolved lines (WT-evolved) and among those that evolved with CD2 (CD2-evolved).

Table 2.3.2 – Separate mixed model ANOVAs for each environment testing the variation among evolved lines within each.

Cheater Treatment	Model	df	AIC	*dev*	χ^2	*P*	
WT	Full	4	-73.47	-81.47	4.69	0.03	*
	Reduced	3	-70.77	-76.77			
CD1	Full	4	-107.64	-115.64	0.00	1	
	Reduced	3	-109.64	-115.64			
CD2	Full	4	-75.65	-83.65	11.99	<.01	*
	Reduced	3	-65.66	-71.66			
CD3	Full	4	-111.09	-119.09	<0.01	0.99	
	Reduced	3	-113.09	-119.09			
CD4	Full	4	-44.43	-52.43	0.52	0.47	
	Reduced	3	-45.91	-51.91			

To test whether each strain was fully resistant I performed *t*-tests for each strain against the hypothesized null of 50% spores for resistance. Of the 30 evolved strains, 10 formed significantly less than 50% of the spores in chimeras with their target cheaters, indicating that they had not evolved full resistance (Table 2.3.3).

Table 2.3.3 – One-sample t-tests based on a null hypothesized value of 0.5. Among the 30 evolved strains, 10 strains formed significantly <50% of the spores when co-developed with their focal cheaters, indicating that they have not evolved to fully counter the effects of the cheater. The remaining 20 strains are not significantly different from 50%.

					95% CI	
Strain	Estimate	Statistic	*P*	Parameter	Lower	Upper
rAX4.1_R15	*0.58*	*2.49*	*0.98*	*6*	*-inf*	*0.65*
rAX4.2_R15	0.57	1.79	0.94	8	-inf	0.64
rAX4.3_R15	0.63	4.18	1.00	6	-inf	0.69
rAX4.4_R15	0.46	-0.85	0.22	4	-inf	0.55
rAX4.5_R15	0.52	0.24	0.59	6	-inf	0.66
rAX4.6_R15	0.54	0.83	0.78	5	-inf	0.63
rCD1.1_R15	0.43	-2.75	0.01	7	-inf	0.48
rCD1.2_R15	0.44	-2.18	0.03	7	-inf	0.49
rCD1.3_R15	0.42	-2.22	0.03	7	-inf	0.49
rCD1.4_R15	0.44	-1.62	0.08	6	-inf	0.51

Table 2.3.3 Continued

					95% CI	
Strain	Estimate	Statistic	*P*	Parameter	Lower	Upper
rCD1.5_R15	0.48	-1.23	0.13	7	-inf	0.51
rCD1.6_R15	0.47	-0.97	0.18	7	-inf	0.53
rCD2.1_R15	0.45	-2.07	0.04	8	-inf	0.49
rCD2.2_R15	0.53	0.97	0.82	8	-inf	0.60
rCD2.3_R15	0.43	-2.44	0.02	8	-inf	0.48
rCD2.4_R15	0.45	-1.86	0.06	6	-inf	0.50
rCD2.5_R15	0.49	-0.39	0.35	6	-inf	0.55
rCD2.6_R15	0.21	-5.14	0.02	2	-inf	0.38
rCD3.1_R15	0.49	-0.44	0.34	6	-inf	0.54
rCD3.2_R15	0.49	-0.48	0.32	7	-inf	0.54
rCD3.3_R15	0.50	-0.03	0.49	6	-inf	0.53
rCD3.4_R15	0.44	-3.14	0.01	6	-inf	0.48
rCD3.5_R15	0.48	-0.72	0.25	5	-inf	0.54
rCD3.6_R15	0.48	-0.89	0.20	6	-inf	0.52
rCD4.1_R15	0.38	-5.04	<0.01	6	-inf	0.43
rCD4.2_R15	0.46	-1.28	0.13	4	-inf	0.53
rCD4.3_R15	0.36	-2.37	0.03	6	-inf	0.48
rCD4.4_R15	0.29	-11.78	<0.01	4	-inf	0.33
rCD4.5_R15	0.46	-0.74	0.24	7	-inf	0.56
rCD4.6_R15	0.40	-1.35	0.11	6	-inf	0.54

Finally, to ask whether the level of resistance varied depending on the target strain (i.e., whether some mutants are easier to adapt to than others), I examined the effect of the competitor's identity on the proportion of spores in each environment, relative to the ancestor in the same mix. In the full model, relative spore proportion is the dependent variable, target strain was considered to be a fixed effect, and population and block were treated as random effects. However, this analysis did not show a significant effect of the target strain on the proportion of evolved spores in mixes ($\chi^2 = 1.34$, $df = 4$, $P = 0.86$), suggesting that the magnitude of

evolutionary improvement did not differ among the groups, and thus that there were no differences in the ease or propensity to adapt to some targets over others.

2.3.3 Evolved Strains Have Greater Fitness in Familiar Environments

To test whether improvements were specific to the targeted strain or whether they were general adaptations to all strains or to abiotic aspects of the evolution environment, I co-developed all evolved strains with all four mutants. If adaptations are specific to the targeted strain, then I expect that mutants will do consistently worse when paired with the strains that evolved with it compared to strains that did not. Indeed, this pattern is what I observed—in all five environments, the cheater performed worse with the strains that evolved against it, compared to novice strains. For example, CD1 does worse, on average, against CD1-evolved strains than against non-CD1-evolved strains (a lowercase 'r' indicates strains that evolved in the presence of a particular cheater. For example, "rCD1" indicates those strains that evolved in the presence of CD1, Figure 2.3.4). This result is further supported by mixed model ANOVA, which shows that including direct vs correlated response as a factor significantly improves the fit of the model ($\chi^2 = 6.102$, $df = 1$, $P = 0.012$). On average, strains that had been exposed to a given cheater during their evolution had 3.59% more spores in chimera than strains that had not experienced that cheater (Table 2.3.4).

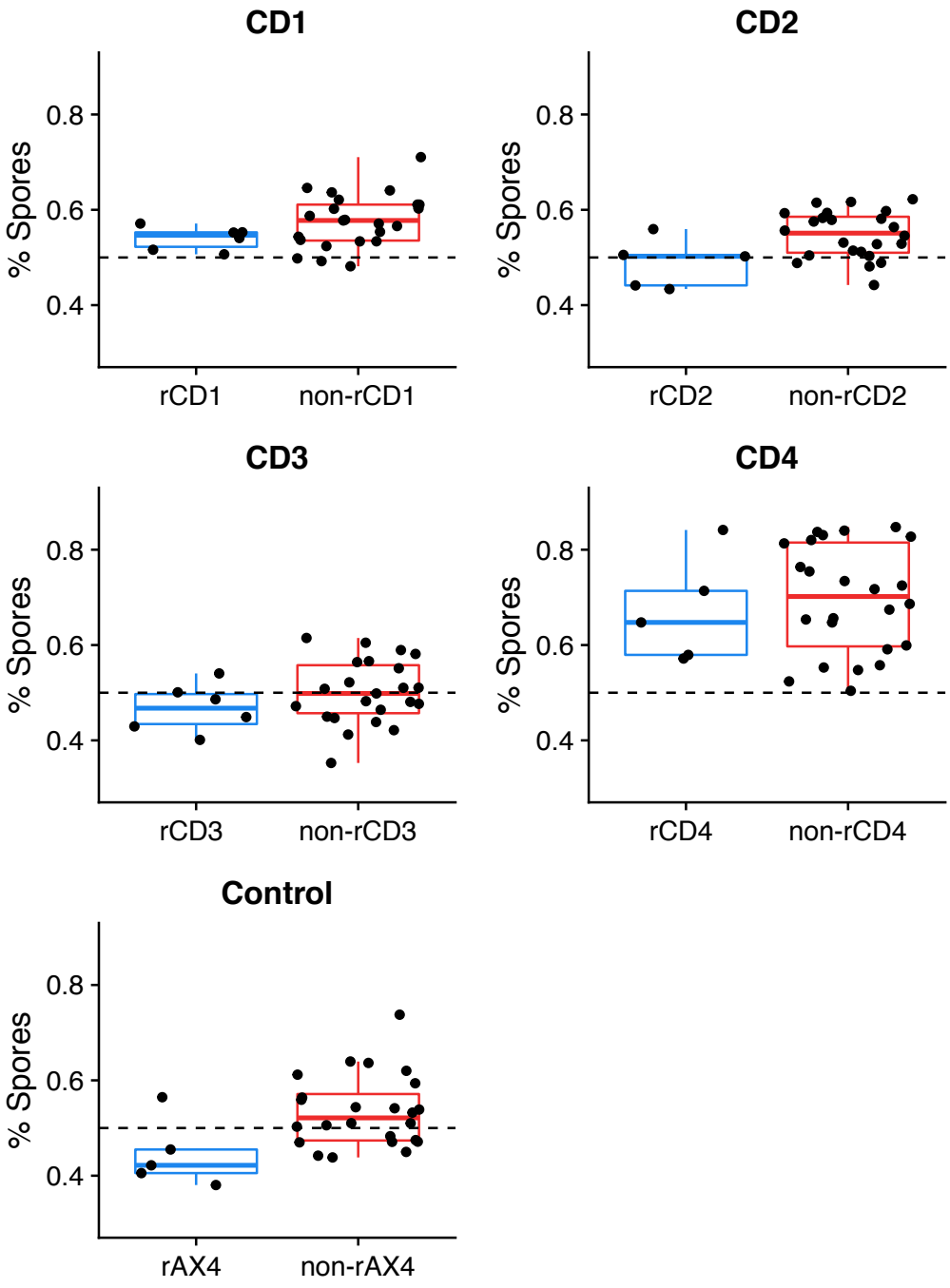

Figure 2.3.4 – Performance of each cheater (or the non-cheating control) against evolved populations. In all five cases, mean performance of a given cheater (CD1-CD4) or the non-cheating control (AX4) was worse when it was paired with strains that were directly selected to resist it (in blue) compared to strains that evolved to resist some other competitor (in red). This result suggests that the evolutionary improvements in performance were, at least in part, specific to the identity of the competitor, rather than being general adaptations to either the abiotic environment or aspects that were common to all competitors.

2.3.4 Likelihood Ratio Test of Correlated Response

I used linear mixed models with spore proportion as the dependent variable and CR/DR (i.e., whether the performance of a given pair is a direct or correlated response to selection) as the dependent variable. For example, the proportion of rCD1.1 spores when co-developed with CD1 is considered a direct response, whereas the proportion of rCD3.1 spores with CD1 is considered a correlated response, since rCD3.1 was not directly selected for its resistance to CD1. Additional factors included in the model were the evolved strain (nested within selection environment), the test strain (CD1-CD4 or the ancestor) and experimental block, all of which were random factors. Each factor was dropped to create four alternative models, which were estimated using the *lmer* package (Bates et al. 2015) in R (R Core Team 2018).

Table 2.3.4 – Mixed model ANOVAs, AIC, and likelihood ratio test (χ^2).

Models
Full: dependent variable = *spore proportion*, with *test strain*, *evolved* (nested within *environment*), *CR vs DR*, and *block* all considered random effects.
Drop CR/DR: dependent variable = *spore proportion*, with *test strain*, *evolved* (nested within *environment*), and *block* all considered random effects.
Drop environment: dependent variable = *spore proportion*, with *test strain*, *evolved* (nested within *environment*), *CR vs DR*, and *block* all considered random effects.
Drop test strain: dependent variable = *spore proportion*, with *test strain*, *evolved* (nested within *environment*), *CR vs DR*, and *block* all considered random effects.

Model	df	AIC	*dev*	χ^2	*P*	
Full	7	-541.97	-555.97			
Drop CR/DR	6	-537.87	-549.87	6.102	0.0135	*
Drop environment	5	-536.97	-546.97	8.997	0.0111	*
Drop test strain	6	-462.15	-474.15	81.82	<.0001	***

Table 2.3.4 shows the formulae for each mixed model. Model selection was carried out by comparing Akaike information criterion (AIC) for each model. Significance was assessed using likelihood ratio tests that compare the difference in fit of full and reduced models that drop a factor of interest (Chambers and Hastie 1992). The difference in AIC between the full model and drop DR/CR model is significant ($\Delta AIC = 4.1$), indicating that the performance of a given strain depends in part on whether its evolutionary improvement was a direct or correlated response to selection. This result confirms the graphical pattern shown in Figure 2.3.4 and suggests that there was some specificity in the response to given target strain.

Figure 2.3.5 shows the direct and correlated responses to selection for all pairs of selection environments with all cheaters as relative social fitness. Relative social fitness is the evolved strain's representation in the spores when co-developed with a given test strain divided by that of the ancestor when developed with the same strain. If there are trade-offs, then improvements relative to the ancestor against a target strain might cause reduced fitness against other strains (i.e., DR > 1 and CR < 1). I observed negative CR in only two cases: rCD3 and rWT strains, both when mixed with cheater CD4 (Figure 2.3.5H and 2.3.5J, respectively). A negative CR indicates a classic cost-of-adaptation scenario. In one-sample *t*-tests, the relative fitness of rCD3 strains ($\bar{x} = 0.24$, SD = 0.25) and rWT strains ($\bar{x} = 0.43$, SD = 0.27) were both significantly less than one; $t = -7.53$, $df = 5$, $P < .001$ and $t = -4.19$, $df = 3$, $P = 0.012$ against CD4, respectively. In some selection environments, the correlated response was greater than the direct response, suggesting the evolution of superior generalists in these environments. For example, the mean relative fitness was greater for rCD1 strains than rCD3 strains against the CD3 cheater ($t = 3.065$, $df = 10$, $P = 0.006$). Similarly, the rCD1 strains are more fit than the rCD4 strains against

the CD4 cheater ($t = 4.57$, $df = 8$, $P<0.001$). In these few cases, the correlated responses exceeded the direct responses, indicating that occasionally adaptation to one cheater can make a strain pre-adapted to be highly resistance to another.

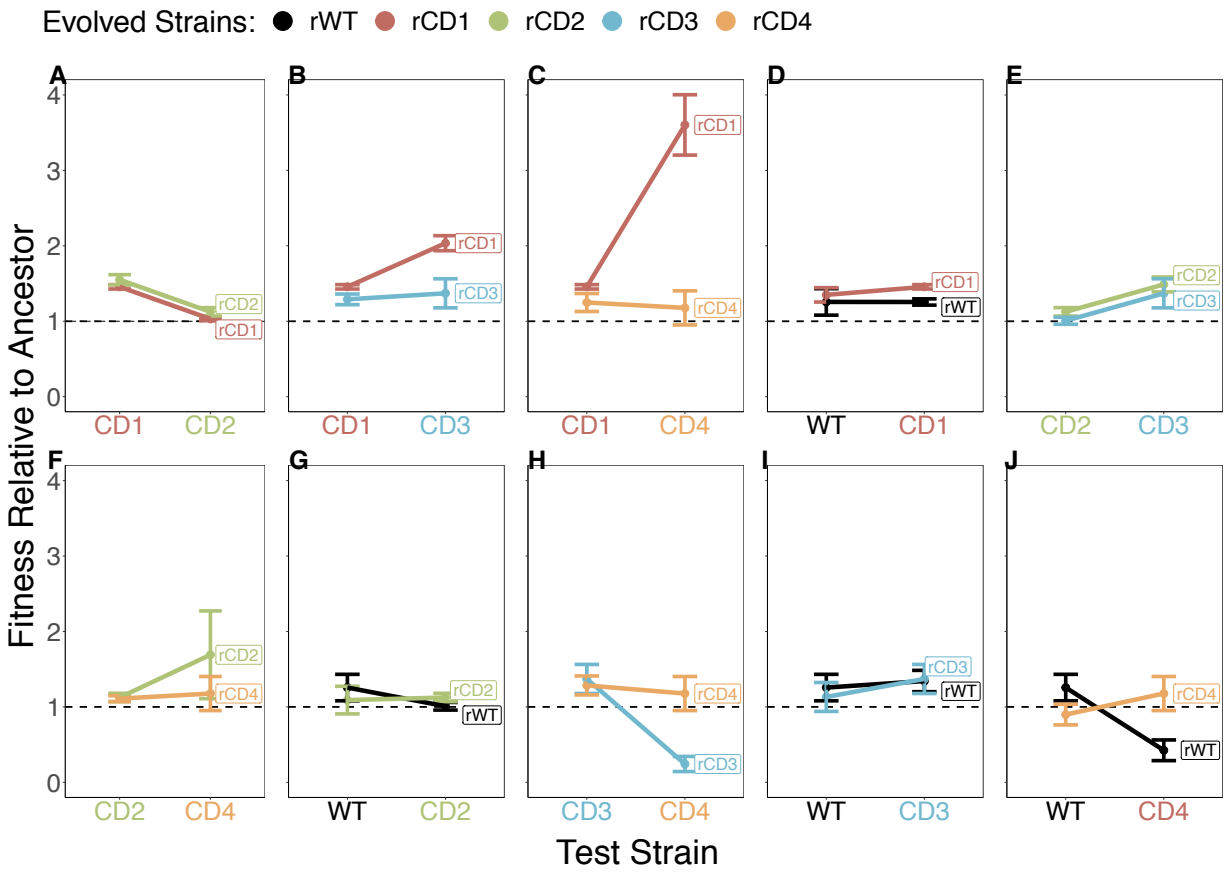

Figure 2.3.5 – Reaction norms of evolved strains with each competitor. Fitness relative to the ancestor is a function of the evolved strain's proportion in spores divided by ancestor's when mixed separately with the test strain (cheater or WT). Each point shows the relative social fitness (grand mean ± s.e.m.) of a group of evolved strains (labeled) when co-developed with a given cheater or the wild-type indicated on the x-axis. Each panel (A-J) shows one of each possible pairing between two environments.

2.3.5 Trade-Offs Among Selection Environments

To test whether evolutionary improvements during the social stage entail fitness trade-offs during other stages of the life cycle, I measured three traits in each evolved strain: doubling time during exponential growth, spore germination efficiency (proportion of spores that produce amoebae), and sporulation efficiency (the number of spores produced from a given number of starting cells). Analysis of variance for doubling time (Figure 2.3.6A), germination efficiency (Figure 2.3.6B), and sporulation efficiency (Figure 2.3.6C) was carried out to examine any differences in these traits as a function of the selection environment (CD1-4 or WT).

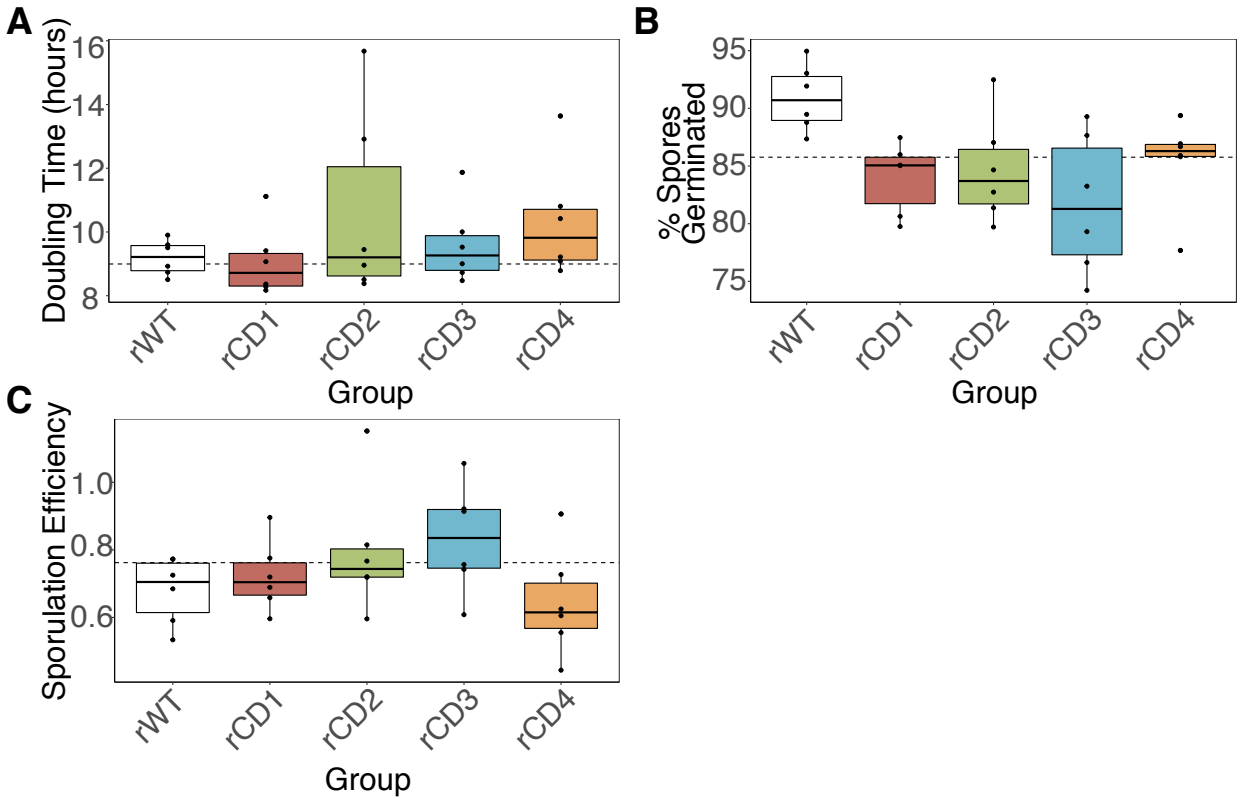

Figure 2.3.6 – Fitness traits among selection groups. Each point represents the mean value for each of the six strains that evolved in that environment. Boxplots show the mean for each selection group (rCD1, rCD2, rCD3, rCD4, or rWT). (A) doubling time, (B) germination efficiency, and (C) sporulation efficiency. The dotted line in each plot indicates the mean of the ancestor.

I found no significant differences in growth rate ($F = 0.95$, $df = 4,25$, $P = 0.45$) or for sporulation efficiency ($F = 1.73$, $df = 4,25$, $P = 0.18$) among the selection environments. However, I did find significant variation among selection environments for germination efficiency ($F = 3.82$, $df = 4,25$, $P = 0.02$), and *t*-tests show that this variation is caused by differences between WT and the four cheater environments (CD1-4, see Table 2.3.5).

Table 2.3.5 – Separate t-tests to compare the means between evolved strains within rWT (rAX4) and rCD1-4 selection environments for germination efficiency, testing the null hypothesis that the difference in means is equal to zero.

						95% CI			
Group1	$\bar{x}$	Group2	$\bar{x}$	*t*	*df*	Lower	Upper	*P*	
rAX4	90.92	rCD1	83.99	4.03	9.95	3.09	10.76	0.002	**
rAX4	90.92	rCD2	84.66	2.83	8.38	1.19	11.32	0.021	*
rAX4	90.92	rCD3	81.73	7.15	3.37	2.76	15.63	0.012	*
rAX4	90.92	rCD4	85.40	2.75	9.08	0.99	10.06	0.022	*

Specifically, populations that evolved in the control environment (WT) that lacked cheating, at least initially, evolved increased rates of spore germination, whereas the other populations (those evolving in the presence of a cheater) did not. This intriguing result is difficult to explain, but one possibility is that resistance mutations in the CD1-CD4 environments conferred bigger benefits than the germination mutations, and thus may have preferentially fixed. Regardless of the specific explanation, the result suggests more generally that populations evolving in the presence of a cheater were not able to simultaneously achieve the same degree of adaptation to their physical environment as those evolving in the absence of this additional hurdle.

2.4 Discussion

2.4.1 Summary

In this study, 30 populations within five cheater environments were evolved in the presence of four, genetically different cheaters. The results show that improvements in social fitness evolved readily in response to a variety of competitor strains that vary in the magnitude of their fitness advantages and possibly in the mechanisms, Trade-offs for social fitness among experimental environments, both in the form of opportunity cost and in one particular life cycle trait (germination) were observed. Overall, strains evolved to be better adapted within their selection environments (greater DR than CR). In other words, the data support the hypothesis of the evolution of specialists, even though general beneficial adaptive traits emerged during the experiment.

2.4.2 Evolved Resistance

Previous work has shown that improvements in social fitness rapidly evolve in lab experiments (Hollis 2012), but the extent to which resistance will be specific to a given cheater is less well understood. In these experiments, competition occurred primarily during the developmental stage. For this reason, the selection procedure was expected to enrich for mutations that confer resistance through processes that involve development in chimerae. I evolved six lines in parallel within each environment to test for repeated evolution and I simultaneously tested within five environments to test the specificity of the evolution of resistance to cheating. Most evolved showed improvements in their representation in the spores in chimeras with their target strain. Adaptation occurred through spontaneous mutations that conferred varying levels of improvement to each environment. Importantly, none of the evolved strains from cheater

environments showed significant evidence that they cheat their ancestor (Figure 2.4.1), suggesting that these strains may be noble resistors (Khare et al. 2009), meaning that they did not themselves become cheaters when adapting resistance. This finding is interesting in that, the only other strain known to resist cheating (*rccA*-) is also noble (Khare et al. 2009), further supporting the hypothesis that resistance does not necessarily involve the evolution of greater cheating.

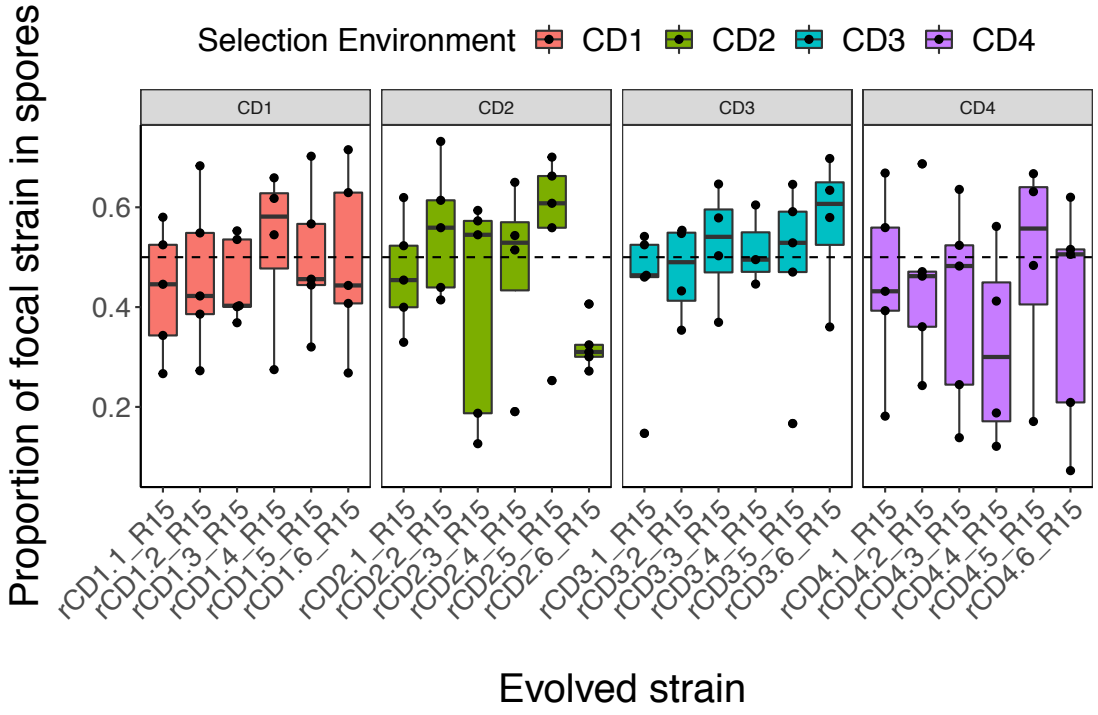

Figure 2.4.1 – Performance of non-rWT strains with WT. To test whether evolved strains cheat the WT, I co-developed the evolved strains from the four cheater treatments with WT and measured their representation in spores. In t-tests, none of the evolved strains were significantly better than WT (P>0.05).

2.4.3 Fitness Trade-Offs

One possibility is that cheaters have a numerical advantage during the social stage, but mutations that confer cheating ability will trade off with other traits. For example, Wolf et al. (2015) showed that strains that produce more spores in chimeras also produce spores of reduced

quality, reflected in their smaller size and lower rates of germination. Thus, higher spore production might trade off with spore viability, and trade-offs negate the benefits of increased spore representation in chimerae (Wolf et al. 2015). Consistent with some of these findings in studies of naturally occurring strains, I found that strains evolved in the absence of a cheater tended to evolve improvements in spore germination efficiency, meaning that a higher fraction of the spores germinated and produced amoebae (Figure 2.3.6), whereas populations that evolved with cheaters did not. This result suggests the evolution of adaptations to counter the effect of cheating might have hampered the ability of these populations to obtain other fitness improvements, such that a jack-of-all trades is effectively master of none (Whitlock 1996). In this way, resistance could be considered costly, in that evolving resistance results in a lost opportunity to adapt to other aspects of the environment.

Taken together, these findings demonstrate that cheating in social systems might be resolved through evolution of novel resistance traits. In addition, this work has thus demonstrated several ways in which resistance might be costly. The first is that adaptation to one cheater prevents maximal adaptation to other cheaters, such that there is some degree of specialization. Second, adaptations to reduce cheating or inequity in the spores prevented populations from adapting to other aspects of the environment. These subtle trade-offs and "opportunity" costs associated with the evolution of resistance may help to reconcile the seemingly conflicting observations that, on one hand, resistance to cheating evolves and yet variation in cheating and cooperation appears to be prevalent and persistent in nature.

CHAPTER 3: GENETIC BASIS OF RESISTANCE TO CHEATING

3.1 Introduction

Experimental evolution is a valuable tool to investigate the genetic basis of traits of interest (Kawecki et al. 2012). In these experiments, replicate populations are initiated from a common ancestor and evolved in a defined environment. In experimental evolution studies that use microorganisms, it is typically possible for the evolved populations, as well as the ancestor, to be frozen and subsequently revived, which enables direct comparison of evolved and ancestral populations (Elena and Lenski 2003; Bennett and Hughes 2009). In addition to the ability to observe directly the adaptations that result from a given selective environment, recent improvements in sequencing technology have made it possible to sequence the genomes of evolved populations to identify the genetic bases of the adaptations (Turner and Miller 2012; Kofler and Schlötterer 2014; Long et al. 2015; Schlötterer et al. 2015; Franssen et al. 2017).

Given that many different mutations can arise during experimental evolution, it can be difficult to know which mutations identified through whole genome sequencing are responsible for the adaptive phenotypes. One way is to introduce the mutations individually into the ancestral background, but constructing these strains can be difficult, and they might still not capture the phenotypic benefit that the mutation conferred in the particular genetic background in which it arose. One alternative way to further narrow the pool of candidate adaptive mutations is to look for the same changes in the genome across multiple populations that evolved in parallel. Parallelism (defined as similar mutations from the same starting point), convergence (defined as similar mutations from different starting points), and divergence across different environments are signatures of strong selection (Endler 1986; Harvey and Pagel 1991; Schluter 2000; Losos

2011; Conte et al. 2012) and can indicate a limited number of available pathways to achieve fitness improvements. Parallel or convergent evolution can be useful tools to distinguish functionally relevant changes from background mutations partly because deleterious mutations can fix only by genetic hitchhiking when populations are sufficiently large and thus are unlikely to arise in parallel across independently evolved lines (Smith and Haigh 1974; Wood et al. 2005; Woods et al. 2006; Ogura et al. 2009; Huse et al. 2010; Lieberman et al. 2011). For example, Ogura et al. (2009) identified important virulence genes in *E. coli* using a comparative genomics approach between two strains with different phylogenies, thus aiding potential strategies to control infections. Similarly, Huse et al. (2010) identified 24 genes linked to chronic colonization in cystic fibrosis lung infections by comparing gene expression profiles of parallel evolving *Pseudomonas aeruginosa* lines. Furthermore, when replicate populations are evolved in a variety of different environments, divergence between them can help to identify distinct mechanisms important to each through the unique mutations that arise in each environment (Deatherage et al. 2017). Thus, identifying the genetic basis of traits of interest can be aided through detection of similar changes among parallel evolving populations.

Once the putative genetic bases of traits of interest have been identified using evolve-and-re-sequence approaches (Kofler and Schlötterer 2014; Schlötterer et al. 2015), the methods of population genomics can then be applied to try to understand whether and how selection might be operating on these traits in nature (Hohenlohe et al. 2010; Jha et al. 2015; Long et al. 2015). Briefly, population genomics is an extension of population genetics that involves whole genome sequencing (WGS) of many individuals in a population to identify naturally occurring single nucleotide polymorphisms (SNPs), insertions and deletions, copy number variation, and

structural variation. When combined with analyses of molecular evolution, population genomics can be used to address the evolutionary history of variation at loci in nature. For example, regions in the genome with reduced genetic diversity compared to the rest of the genome can indicate a recent selective sweep, where positive selection has driven a particular allele to fixation (Grünwald et al. 2016). Selective sweeps can reduce diversity at closely linked loci because of genetic hitchhiking, the process by which allele frequencies change not from direct selection, but because of their proximity to loci under selection (Elena and Lenski 2003; Buskirk et al. 2017). On the other end of the spectrum, selection can also maintain multiple alleles in a population for long periods of time, through several different mechanisms, including negative-frequency dependence. Here the more common variants in a population have a selective disadvantage compared to rare variants. Signatures of balancing selection in the genome include shared similar polymorphisms between closely related species (excluding convergent evolution) (Klein et al. 1998), increased diversity around target loci (Roux et al. 2013), an excess of polymorphic sites at intermediate (balanced) frequencies, an even distribution of allele frequencies relative to expectations under neutrality (Weedall and Conway 2010), and increased linkage disequilibrium around more recent target loci (Charlesworth 2006). Thus, the combination of experimental evolution, genome re-sequencing, and population genomics can constitute a powerful approach to understanding what phenotypes might be favored, the genetic basis of these traits, and the evolutionary history of these traits in natural populations.

In the previous Chapter, I described a laboratory evolution experiment where I selected for resistance to cheating. Briefly, I evolved replicate populations of *D. discoideum* in the presence

of different cheaters or a non-cheating control. I assayed the evolved strains for their level of resistance to their respective cheaters, novel cheaters, and the ancestor. I found that strains in each experimental environment readily evolved improved social fitness against their evolutionary partner. Following evolution, I isolated a single evolved clone from each population and sequenced it with Illumina. Here, I describe the mutations that I identified. I describe what is known in the literature on a subset of them that might impact spore-stalk allocation and thus make them plausible candidates for conferring resistance to cheating. In addition, I leverage two existing datasets to better understand the function of these genes and whether selection has been operating on these genes in nature. The first dataset comes from a comparative transcriptomics study of social amoebae, in which Parikh et al. (2010) used RNA-seq to quantify the expression level of every gene in the genome every four hours during the 24 hour development cycle. Additionally, the dataset identifies which genes are differentially expressed between the two main cell types: prespore and prestalk cells. I used this dataset to assess whether the genes that were mutated in my experiment were preferentially prespore or prestalk biased in their expression as biases might indicate parallel mechanisms in the evolution of resistance. The second dataset comes from a population genomics study that used whole genome resequencing of 20 natural *D. discoideum* isolates to examine the evolutionary history of candidate loci involved in cheating behaviors (Ostrowski et al. 2015). The dataset provides several molecular evolution metrics for each gene in the genome, albeit based on only 20 strains. I leveraged this dataset to test the hypothesis that my candidate genes might be targets of selection in nature, and therefore might show distinctive signatures of molecular evolution compared to the rest of the genome. Finally, I found several genes among those mutated in my experiment that were of particular interest because of their involvement in development, which I

describe in detail. While I found only a few examples of parallelism, some molecular evolutionary metrics suggest the possibility of balancing selection at these loci. Together these analyses might aid in the identification of the genetic underpinnings of resistance to cheating and help to elucidate the importance of these components in nature.

3.2 Methods

3.2.1 Whole-Genome Resequencing and Variant Calling

Genomic DNA was prepared using phenol:chloroform extraction from axenic cell culture initiated by clones from each of the 30 evolved populations, the four cheaters, and the ancestral wild-type (AX4). DNA extraction was carried out on nuclei, which were prepared by spinning and washing cells in KK_2 and resuspending in 15 ml nuclei buffer (40 mM Tris-HCl, pH 7.8, 6 mM $MgCl_2$, 40 mM KCl, 0.1 mM EDTA, 5 mM DTT, 1.5% sucrose, 0.4% IGEPAL) while on ice for 10 minutes and centrifuging at 4000 RPM for 20 minutes. Nuclei were kept at -80 ℃ until DNA extraction. To purify DNA, nuclei were thawed on ice and resuspended in 100 µl EDTA [100 mM]. The following was added sequentially while incubating at 60 ℃: 450 µl STE solution (10 mM Tris-HCl, pH 8, 10 mM EDTA, 400 mM NaCl), 50 µl 10% SDS, 10 µl of 10 mg/µl Proteinase K. The solution was incubated for 1 hour at 60 ℃. Using cut tips throughout, the solution was transferred to a microcenterfuge tube, 500 µl of phenol:choloroform:isoamyl alcohol was added and tubes were spun at top speed for 10 minutes. The aqueous phase was transferred to a new tube and the process was repeated until no interface was observed. The supernatant was then treated with one round of chloroform followed by ethanol precipitation before adding Tris-EDTA (TE) with 100 µg/ml ribonuclease A (RNase) and incubating overnight at 4 ℃. Following

RNase treatment, phenol:chloroform extraction, and ethanol precipitation was repeated to remove remaining RNase.

Libraries were prepared using the Illumina NexteraXT kit according to the manufacturer's protocol, except that reaction volumes were reduced to 0.25×, before size selection using SPRIselect beads (Becker Coulter B23319) for a target size of 300-500 bp. Raw sequencing reads were mapped to the chromosomal sequences of AX4 (GFF files dated 30 November, 2016, available at: http://dictybase.org/download/gff3/) using BWA-MEM (Li and Durbin 2009). The AX4 genome contains a large (~750 Kb) duplication within chromosome 2, which was masked from the reference for the alignment. Duplicate reads were marked using Picard 2.18.13 MarkDuplicates (http://broadinstitute.github.io/picard). Variant calling, genotyping, and hard filtering was performed with GATK4.0.11.0 (McKenna et al. 2010) following GATK Best Practices (Depristo et al. 2011; Van der Auwera et al. 2013). Base quality score recalibration was performed on a join-call cohort and final variant calling was applied to individual samples. Variants were called using GATK4 Haplotype Caller with the PCR indel model set to hostile. This model attempts to correct for PCR errors by penalizing the variant likelihood based on the surrounding level of repetition. After initial variant calling, filtering thresholds were applied to remove variants likely to be false positives. These thresholds were set for eight statistics which describe each variant and provide evidence on the likelihood of each being a true variant. The threshold for depth (DP) and Read Position Rank Sum (ReadPosRandSum) were determined using Tukey's method to identify outliers scored above and below 1.5×IQR (inter quartile rage). QualitybyDepth (QD), Mapping Quality (MQ), StandOddsRatio (SOR),

MappingQualityRankSum (MQRankSum) were set according to the GATK4 hard filtering recommendations (available at: https://software.broadinstitute.org/gatk/documentation/article.php?id=6925). Variants were filtered by the following criteria: (QD > 2.0, FS < 5.0, MQ > 58.0, SOR < 2.0, DP > 10.0, DP < 260.0, MQRankSum > -0.5, MQRankSum < 0.5, ReadPosRankSum > -2.5, ReadPosRankSum < 2.5).

Following variant calling and hard filtering, the data set still contained numerous calls that were potentially false positives. The complex *D. discoideum* genome contains an abundance of duplicated genes and repetitive regions (Glöckner et al. 2001; Eichinger et al. 2005). These duplicates and repeat regions can lead to mis-mapping of reads and therefore erroneous variant calls. Additionally, the PCR amplification of libraries, as in the NexteraXT protocol, is known to introduce sequencing artifacts, especially within homopolymer and other repetitive regions (Kozarewa et al. 2009). To further reduce the likelihood of false positives, calls were manually reviewed and some were removed because they were within: (i) variant clusters (multiple variants in close proximity across more than one sample), (ii) repetitive regions (homopolymer runs, tandem repeats), (iii) gene duplicates (when reads map to multiple genes), or (iv) had a significantly high sample read depth (260×) – good indicators of improper read mapping or increased chance of PCR errors (Li 2014; Meynert et al. 2014). Annotations were added to variants using SnpEFF v4.0 (Cingolani et al. 2012) with AX4 GFF3 file generated November 30th, 2016 (available at: http://dictybase.org/download/gff3/).

3.3 Results

3.3.1 Summary

The 30 evolved strains were re-sequenced using Illumina HiSeq 150 bp paired end reads to an average coverage of 25×. After filtering, 196 variants remained across all strains, with a mean of 6.5 variants per strain. Of the total 196 variants, 118 were either synonymous, located within introns, intergenic, or single amino acid insertions/deletions within tandem repeats (such as microsatellite repeats), and thus considered low-impact and removed. The remaining 77 mutations consisted of 70 missense SNPs with a predicted moderate impact (according to snpEFF putative impact assessment), two nonsense mutations, one stop loss mutation, and three splice region variants.

Of the 77 remaining mutations (Table 3.3.1), 27 occurred within genes where the gene products have been described in the literature, as I discuss below. The annotations from the literature for these candidate genes are mostly based on studies of gene disruptions or knock outs, meaning that the annotations generally describe the phenotypic effect of null or inactivating mutations. However, most of the mutations in these same genes in my evolving populations are missense mutations, which potentially have different impacts on the phenotype, compared to null mutations. This is because point mutations (unlike insertional mutagenesis) can cause loss-of-function, gain-of-function (novel function), or increase-in-function.

Table 3.3.1 – Variants in candidate genes within the evolved strains. The predicted impact of the mutation is indicated with either a L (Low), M (Moderate), or H (High). Impact predictions were added using SnpEff 4.3 (Cingolani et al. 2012).

Sample	Chr	cDNA pos/length	Ref	Alt	Annotation	Impact	Gene Name
rAX4.6_R15	5	2914/3093	A	G	missense	M	adcF
rCD2.1_R15	2	103/3339	C	A	missense	M	argB
rCD4.1_R15	3		C	A	splice region	L	cog3
rCD1.5_R15	1	602/1313	A	AT	frameshift	H	DDB_G0269614
rAX4.6_R15	1	602/2439	A	C	missense	M	DDB_G0269732
rCD3.3_R15	1	602/2439	A	C	missense	M	DDB_G0269732
rCD3.1_R15	1	102/6135	G	T	missense	M	DDB_G0269934
rAX4.4_R15	1	285/909	G	T	missense	M	DDB_G0269944
rAX4.3_R15	1	394/1458	A	C	missense	M	DDB_G0270532
rCD1.4_R15	2	3006/5052	C	A	missense	M	DDB_G0272338
rCD2.1_R15	2	3006/5052	C	A	missense	M	DDB_G0272338
rCD3.2_R15	2	947/1434	G	A	missense	M	DDB_G0272432
rCD1.5_R15	2	2494/3241	G	A	missense	M	DDB_G0272558
rCD3.4_R15	2	203/2211	T	C	missense	M	DDB_G0272678
rAX4.6_R15	2	527/1368	C	A	missense	M	DDB_G0272686
rCD3.3_R15	2	590/1284	T	G	missense	M	DDB_G0272955
rAX4.6_R15	2	500/1989	C	G	missense	M	DDB_G0274349
rCD3.1_R15	2	1037/4074	G	C	missense	M	DDB_G0274795
rCD2.4_R15	2	2324/2514	T	G	missense	M	DDB_G0274981
rCD1.2_R15	2	7105/8607	T	A	missense	M	DDB_G0275305
rCD4.4_R15	2	2128/2607	A	C	missense	M	DDB_G0275509
rCD4.2_R15	2	2581/3813	A	G	missense	M	DDB_G0275937
rCD2.3_R15	2	385/1794	C	G	missense	M	DDB_G0276317
rAX4.2_R15	2	659/705	C	G	missense	M	DDB_G0276557
rAX4.5_R15	2	54/219	G	T	missense	M	DDB_G0276837
rAX4.6_R15	2	1587/2187	G	C	missense	M	DDB_G0277043
rCD3.2_R15	2	28/171	T	C	missense	M	DDB_G0277607
rAX4.1_R15	2	143/171	A	T	missense	M	DDB_G0277607
rCD2.4_R15	2	143/171	A	T	missense	M	DDB_G0277607
rCD1.5_R15	3	5247/5700	G	T	missense	M	DDB_G0278215
rCD4.6_R15	3	40/453	A	G	missense	M	DDB_G0278399
rCD1.5_R15	3	3358/5097	C	T	missense	M	DDB_G0279255
rCD3.1_R15	3	1709/3627	T	G	missense	M	DDB_G0279309
rCD4.5_R15	3	274/474	C	A	missense	M	DDB_G0279681

Table 3.3.1
(continued)

Sample	Chr	cDNA pos/length	Ref	Alt	Annotation	Impact	Gene Name
rCD2.4_R15	3	229/921	C	A	missense	M	DDB_G0280779
rAX4.3_R15	3	1552/2673	A	C	missense	M	DDB_G0280811
rAX4.4_R15	3	1268/1464	A	G	missense	M	DDB_G0281225
rCD4.2_R15	3	1105/1161	A	C	missense	M	DDB_G0281429
rCD4.4_R15	3	102/1128	A	C	missense	M	DDB_G0281653
rCD2.4_R15	3	275/423	T	C	missense	M	DDB_G0281811
rCD3.6_R15	3	1220/4923	A	C	missense	M	DDB_G0282115
rAX4.2_R15	3	644/4197	A	G	missense	M	DDB_G0282421
rCD3.2_R15	3	239/642	C	T	missense	M	DDB_G0282865
rCD1.5_R15	4	736/3960	C	A	missense	M	DDB_G0283325
rCD2.2_R15	4	1957/2067	G	T	missense	M	DDB_G0286807
rCD4.2_R15	5	3236/5637	G	A	missense	M	DDB_G0288007
rCD2.2_R15	5	544/4431	A	G	missense	M	DDB_G0288241
rCD4.4_R15	5	650/1540	G	T	missense	M	DDB_G0289053
rCD2.1_R15	5	3979/5952	C	A	missense	M	DDB_G0290289
rCD2.6_R15	5	3979/5952	C	A	missense	M	DDB_G0290289
rCD3.1_R15	6	2729/3492	A	C	missense	M	DDB_G0292124
rAX4.5_R15	6	1448/2460	A	T	missense	M	DDB_G0292936
rCD2.5_R15	6	1202/6132	T	C	missense	M	DG1098
rCD1.5_R15	4	6083/8379	T	G	missense	M	DG1104
rCD2.2_R15	1	2458/2637	C	G	missense	M	*glcS*
rCD3.2_R15	1	118/2859	T	A	missense	M	*gtaE*
rCD4.1_R15	2	10/1407	A	G	missense	M	*hgsB*
rAX4.3_R15	1	1750/2034	C	A	stop gained	H	*mkcF*
rAX4.5_R15	1	322/2034	G	A	stop gained	H	*mkcF*
rCD3.4_R15	3	10/2457	A	G	missense	M	*mybE*
rAX4.4_R15	2	1337/1464	C	A	missense	M	*nat6*
rCD3.2_R15	5	1925/3621	A	C	missense	M	*nup133*
rAX4.3_R15	2	1054/4353	T	G	missense	M	*pds5*
rCD4.5_R15	1	1810/1926	G	T	missense	M	*phg1a*
rCD4.5_R15	4	1687/1710	G	A	missense	M	*plbD*
rCD1.5_R15	5		GT	G	splice region	L	*rbx1*
rCD2.1_R15	3	2326/3357	A	G	stop lost	H	*rliB*
rCD4.3_R15	3	1454/3219	A	C	missense	M	*spt16*
rCD2.2_R15	6	1072/5259	G	T	missense	M	*tagA*

Table 3.3.1 (continued)

Sample	Chr	cDNA pos/length	Ref	Alt	Annotation	Impact	Gene Name
rCD3.3_R15	3	6897/13751	T	G	missense	M	*tra1*
rCD3.2_R15	3	52/1509	C	A	missense	M	*ugpB*
rAX4.6_R15	5	3350/5034	C	A	missense	M	*usp40*
rCD3.6_R15	2	1171/1338	T	G	missense	M	*vatH*
rCD3.6_R15	2	274/1752	T	C	missense	M	*wasB*
rCD3.6_R15	6	1962/3174	G	T	missense	M	*xpo1*

3.3.2 Gene descriptions

Several mutated genes are of particular interest because of their involvement in developmental phenotypes (Table 3.3.2). For example, the phg1A gene encodes a transmembrane signaling receptor involved in phagocytosis, cell-substrate adhesion, defense response to bacteria, and asexual reproduction (Benghezal et al. 2003; Gebbie et al. 2004). Disruption of phg1A has been shown to cause developmental arrest and loss of culmination, the final stage of fruiting body formation (Benghezal et al. 2003). The tagA gene was mutated in one strain from my experimental evolution cohort (rCD2.2_R15 - resistant to CD2, population 2 of 6, Round 15 of selection). This gene is necessary for cell type differentiation and is differentially expressed in pre-spore cells (Good 2003). The tagA gene is crucial in cell fate determination and required for the initial specification and maintenance of prespore cell lineage (Good 2003). In chimeras with the wild-type strain, the tagA- mutant forms very few prestalk cells (Khare and Shaulsky 2010).

Table 3.3.2 – Annotations for socially important genes.

Gene	Description	Annotations
mkcf	**M**AP **K**inase **C**ascade	*mkcf* null mutant preferentially forms prestalk cells (Parkinson *et al.*, 2011)
mybE	myb domain containing	*mybE* is necessary for correct differentiation of prestalk cells (Fukuzawa *et al*., 2006)
tagA	ABC transporter B family	*tagA* null mutant has abolished sorting to prestalk region (Cabral *et al*., 2006)
ugpB	**U**DP-**G**lucase **P**yrophosphorylase	both *ugpB* and *glcS* null mutants have aberrant stalk morphogenesis with similar phenotypes lacking glycogen for stalk structure (Tresse *et al*., 2008)
glcS	**GL**y**C**ogen **S**ynthase	
phg1A	**PH**a**G**ocytosis	*phg1A* null mutant has abolished culmination, decreased growth rate, decreased phagocytosis, and arrested developmentat tipped mound stage (Benghezal *et al*., 2003)

Moreover, its prestalk allocation was unaffected when co-developed with a known cheater (*chtC*-), meaning the *tagA*- is likely resistant to cheating by *chtC*- and a cheater of wild-type (Khare and Shaulsky 2010). Finally, the strain that harbored the *tagA* mutation (rCD2.2_R15) produced significantly more spores than CD2 in chimerae, making it one of very few strains that possibly countered cheating by becoming a cheater itself, and finding of a known cheater mutation in this strain furthermore suggests this interpretation is correct (see Figure 2.3.3).

In addition to *tagA*, other genes involved in stalk formation were also mutated, including *mkcF, mybE, ugpB, glcS,* and *phg1A.* The *mkcF* null mutant was identified in a genetic screen for insertional mutants that cause a loser phenotype, meaning that they disproportionately formed pre-stalk cells in chimerae (Parkinson et al. 2011). The *MybE* gene is necessary for the correct differentiation of prestalk cells (Fukuzawa 2006). Interestingly, two of the mutated genes, *ugpB*

and *glcS*, are both involved in the process of cell vacuolization, an important step in stalk formation. The *ugpB*$^{-}$ and *glcS*$^{-}$ null mutants both show abnormal stalk phenotype that causes short stalks and frequent delayed development (Tresse et al. 2008). Functional changes in these genes could alter the allocation to stalk or stalk morphology thereby countering cheating by either self-avoidance of stalk or by exploiting the stalk of others.

3.3.3 Parallelism

Four genes were mutated in more than one strain (Figure 3.3.1). In three cases (DDB_G0269732, DDB_G0272338, and DDB_G0277607) identical mutations were found in two different populations. One possibility is that these identical mutations are caused by cross-contamination of populations at some point during the evolution experiment. However, pairs of strains that shared an identical mutation also harbored, on average, 12 non-identical mutations, making cross-contamination an unlikely explanation, unless the mutation and the cross-contamination occurred early in the experiment.

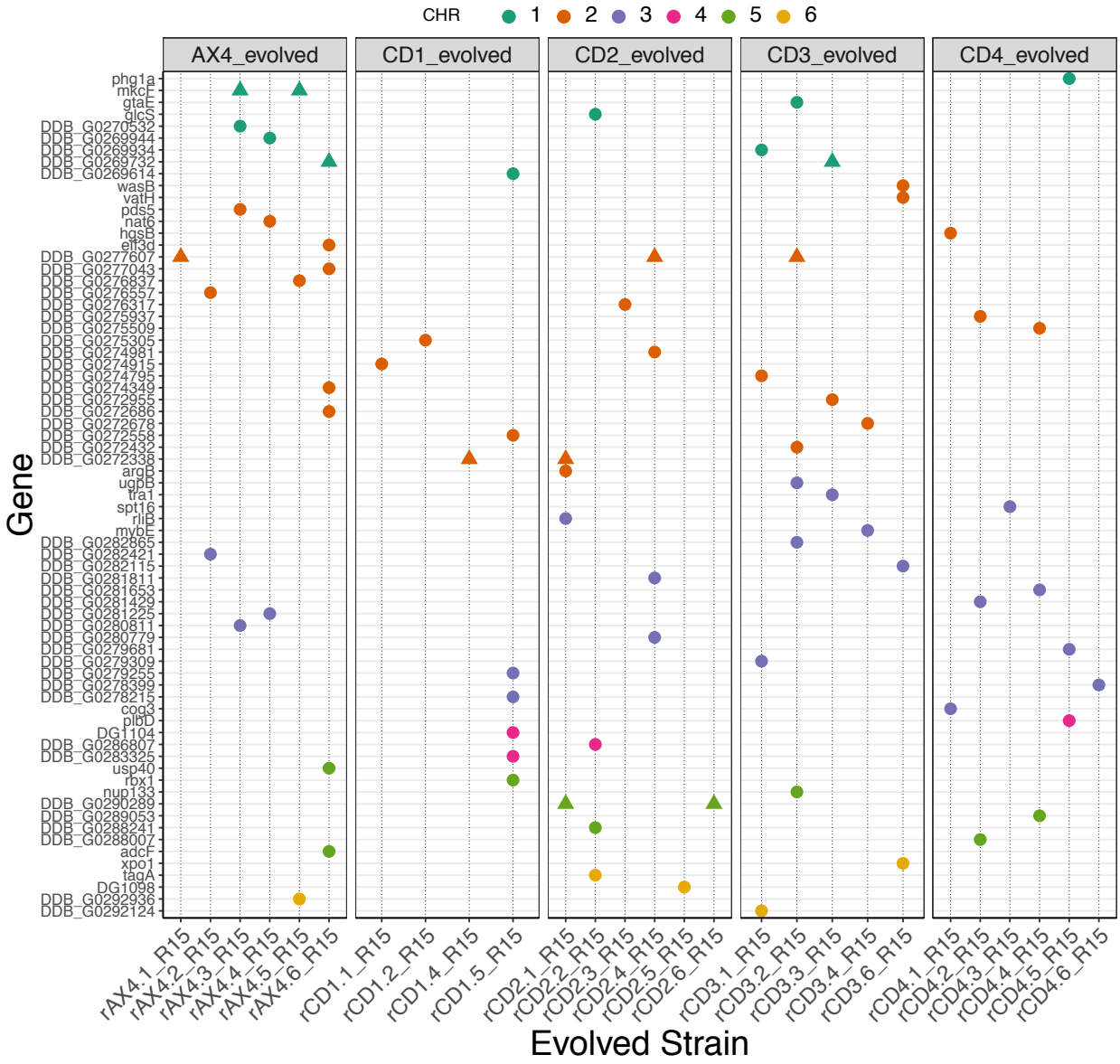

Figure 3.3.1 – Mutated genes within the 35 evolved strains. Rows indicate individual genes (ordered and colored by chromosome). Columns indicate individual strains (grouped by cheater treatment). Triangles (▲) indicate mutations that occurred in the same gene, in more than one strain.

In addition to the mutations that were identical, there were two genes with non-identical mutations. The *mkcF* gene was mutated in two populations that evolved in the presence of the wild-type (AX4) - in both cases, the mutations produced a premature stop codon. In rAX4.3_R15, the stop codon occurred towards the end of the gene (position 1750 of 2034). The

second *mkcF* mutant gained a stop codon early in the transcript (position 322 of 2034). In addition to the two identical mutations in the gene DDB_G0277607, there was a third strain with a unique mutation in this gene and all three mutations arose in different selection environments. Overall, there was no evidence of parallelism at the gene level, aside from distinct, and therefore, independent, mutations in *mkcF*. Given the occurrence of one mutation within the same gene out of 30 evolved strains (6 populations per environment), the probability of a second mutation in the same gene occurring in the same environment by chance is ~17% (5/29). Thus, even in this one instance, I do not have any evidence that parallel mutations were somehow more likely to occur within an environment than across.

3.3.4 Population Genomics

The 77 genes that harbored mutations following evolution were then assessed for several metrics of sequence diversity in natural populations that might help to identify whether any of these genes show unusual patters of polymorphism, suggestive of selection in nature. To do so, I use a dataset generated by Ostrowski et al. (2015) using WGS of 20 natural isolates of *D. discoideum*. With that dataset I created a null distribution of 10,000 random gene sets of the same size as my mutated gene set and compared them against my genes for several metrics of molecular evolution (Figure 3.3.2). The mean length of the mutated genes (3173 bp) was significantly greater than expected based on other genes in the genome (percentiles: 5^{th} = 1412, 95^{th} = 2036). This result also explains the increased number of singletons, segregating sites (*S*), the number of mutations, and the number of haplotypes that I observed in my mutated genes compared to random gene sets. When scaling these metrics to gene length, none differed significantly from the null expectation based on other genes in the genome. The mutated genes

showed elevated levels of intragenic linkage disequilibrium (low recombination, or Hudson's *C*). Consistent with what is expected in regions of low recombination, I also observed that the mean number of haplotypes (scaled for gene length) was somewhat lower than expected (6.5th percentile; Figure 3.3.2).

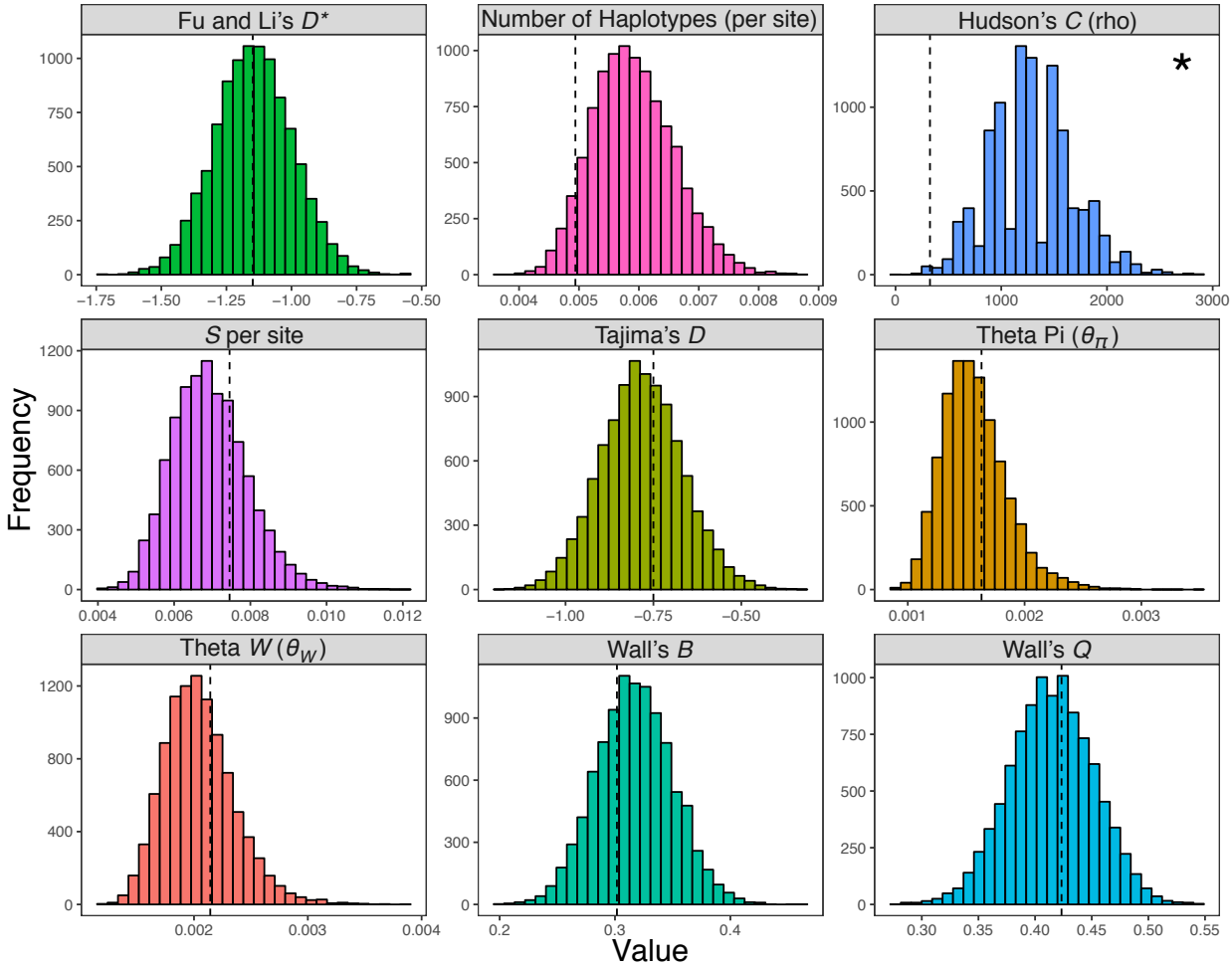

Figure 3.3.2 – Difference between candidate genes versus random gene sets for different evolutionary metrics. Dotted lines indicate the mean value for each molecular evolution statistic for mutated genes as a group, calculated by Ostrowski et al. (2015) using WGS of 20 natural isolates. The observed value for these genes is compared to a null distribution based on 10,000 random gene sets of the same size. Asterisks indicate statically significant results.

In addition, I did not observe overrepresentation of the mutated genes in either tail of the genome-wide distribution for any metric after correction for gene length (Table 3.3.3), nor

elevated variance for any of the molecular evolution metrics (Table 3.3.4), which might occur if extreme values at both ends of the genome distribution were averaged out.

Table 3.3.3 – Number of Candidate Genes that are "Extreme" (in the lower 5th or upper 95th percentile of the genome-wide distribution). Boldface indicates metrics where candidate genes are significantly overrepresented in either tail of the genome-wide distribution.

Statistic	Number (and percentage) of mutated genes in the lower 5th percentile of the genome-wide distribution[a]	Number (and percentage) of mutated genes in the upper 95th percentile of the genome-wide distribution[a]	*P*-value[b]
S	0(0)	**8(12.9)**	**0.024**
Singletons	0(0)	**8(12.9)**	**0.024**
NMut	0(0)	**8(12.9)**	**0.024**
nhap	0(0)	**10(16.7)**	**0.003**
hapdiv	0(0)	2(2.9)	0.871
WallsB	0(0)	1(1.5)	0.969
WallsQ	0(0)	2(3)	0.860
ThetaW	0(0)	2(2.9)	0.871
ThetaPi	0(0)	3(4.5)	0.687
Tajima's D	3(4.5)	4(6.1)	0.468
Fu and Li's D*	1(1.4)	4(6.1)	0.468
Fu and Li's F*	2(2.9)	4(6.1)	0.468
Hudson's $\hat{C}$ (rho)	0(0)	0(0)	1.000
Number of Segregating Sites[c]	0(0)	2(2.9)	0.871
Number of Haplotypes[c]	1(1.4)	1(1.4)	0.972
Number of Singletons[c]	0(0)	3(4.5)	0.687

[a]**Null expectation is that 5% of the candidate genes will reside in the top 95% and bottom 5% of the genome-wide distribution.**

[b]**P-value is the result of a Fisher's Exact test that compares the number of extreme genes versus not for candidate versus non-candidate genes.**

[c]**Per site (divided by gene length)**

Table 3.3.4 – Test of elevated variance in candidate genes compared to random genes. For each evolutionary metric, I tested whether the variance was lower or higher for mutated genes compared to 10,000 datasets consisting of genes chosen at random.

Metric	5th percentile	95th percentile	Observed value for mutated genes
Number of segregating sites	73.187	713.575	352.433
Number of Singletons	30.072	263.300	207.666
Number of Mutations	74.137	721.373	352.074
Number of Haplotypes	18.398	32.250	25.587
Haplotype Diversity	**0.100**	**0.152**	**0.032(low)**
Walls B	**0.056**	**0.096**	**0.04(low)**
Walls Q	**0.077**	**0.114**	**0.056(low)**
Theta W	2.10E-06	1.44E-05	3.45E-06
Theta Pi	1.31E-06	1.34E-05	1.39E-06
Tajima's D	0.681	1.200	0.780
Fu and Li's DStar	**1.253**	**1.917**	**1.24(low)**
Fu and Li's F*	1.380	2.138	1.405
Hudson's Ĉ (rho)	**5.50E+06**	**1.61E+07**	**2.83E+6(low)**
Number of Segregating Sites[a]	2.40E-05	1.54E-04	4.25E-05
Number of Haplotypes[a]	**1.31E-05**	**6.68E-05**	**8.70E-06(low)**
Number of Singletons[a]	9.86E-06	8.28E-05	3.12E-05

[a]Per Site (divided by gene length)

Despite seeing no consistent signals of selection in my mutated genes as a group, 17 of 77 genes had "extreme" values (defined as <5th or >95th percentile of the genome wide distribution based on n = 20) for at least one metric from the population genomics dataset (Table 3.3.5). Most of these genes (10 out of 14) show unusually high (i.e., positive) values for metrics of the site frequency spectrum, (Tajima's *D*, Fu and Li's *D** or *F**), which might reflect balancing selection.

Table 3.3.5 – Candidate genes with values in the tails of the genome-wide distribution for at least one metric.

Gene	Metrics in the lower 5th percentile of the genome wide distribution	Gene	Metrics in the upper 95th percentile of the genome wide distribution
DDB_G0269934	Tajima's *D*	DDB_G0281811	Wall's *B*,Wall's *Q*
DDB_G0272686	Tajima's *D*, Fu and Li's *F**	DDB_G0272432	Wall's *Q*,
DDB_G0292936	Tajima's *D*, Fu and Li's *D**, Fu and Li's *F**	DDB_G0267674	Tajima's *D*, Fu and Li's *D**, Fu and Li's *F**
DDB_G0279309	Number of Haplotypes/Site	DDB_G0277607	Tajima's *D*
		DDB_G0278399	Tajima's *D*, Fu and Li's *D**, Fu and Li's *F**
		DDB_G0280779	Tajima's *D*
		DDB_G0274627	Fu and Li's *D**, Fu and Li's *F**
		DDB_G0274871	Fu and Li's *D**, Fu and Li's *F**
		DDB_G0282865	*S*/site, Singletons/site
		DDB_G0292936	*S*/site, Singletons/site

3.3.5 Transcriptomics

If resistance impacts stalk or spore allocation, I might expect that the candidate genes (those mutated in the selection experiment for resistance) would be disproportionately expressed during the timepoint in development when cell fate is determined, or that they would be disproportionately expressed in one cell type or another. To test this hypothesis, I compared the gene expression timing during development between candidate genes and the whole genome

using the data from Parikh et al. (2010). These results indicated no difference in maximal expression timing (Table 3.3.6), meaning that candidate genes were not biased towards a particular life-cycle stage.

Table 3.3.6 – Timing of maximum expression. The number of mutated genes versus all genes whose maximal expression occurs at the indicated timepoint during the life cycle, according to Parikh et al. (2010). The mutated genes were not overrepresented in any of these categories. *P*-values are based on a two-tailed Fisher's Exact test.

Timepoint (hours)	Number (and percentage) of mutated genes	Number (and percentage) of all genes	*P*-value (two-sided)
0	10(14)	2160(17)	0.64
4	13(18)	1551(12)	0.14
8	12(17)	1903(15)	0.62
12	6(8)	1875(15)	0.18
16	15(21)	2105(16)	0.34
20	6(8)	1683(13)	0.29
24	10(14)	1592(12)	0.72

Table 3.3.7 – Candidate genes with differential expression. Genes with at least a two-fold change in RNA abundance and a P-value lower than 0.05 were considered to be significantly cell-type enriched (as determined by Parikh et al. 2010). There was no significant enrichment for either cell type among candidate genes (Fisher's exact test, $P = 0.73$).

Cell-type	Number of mutated genes (Proportion)	Number of all genes (Proportion)
Prespore	1(1.4)	850(7.2)
Prestalk	1(1.4)	915(7.7)

Two genes were differentially expressed between prespore and prestalk cells, based on data from Parikh et al. (2010). The evolved strain rAX4.3_R15 has a mutation in gene DDB_G280811 (unknown gene product) that is preferentially expressed in prespore cells. This same strain also has a nonsense mutation in *mkcF*, a gene that has been shown to be involved in prestalk proportioning (Parkinson *et al.*, 2011). Specifically, the *mkcF* knock-out causes a

"losing" phenotype, meaning it disproportionately allocates cells to the prestalk region in chimeras with the WT. Thus, both of these mutation present in rAX4.3 could conceivably impact spore or stalk allocation. However, it is not clear currently what impact these mutations have individually on this phenotype, nor whether their effects in combination differs from their effects singly. This is something that might be addressed in the future. Nevertheless, it is notable that the evolved strain that carries these two mutations is one of very few that I found that have evolved to cheat their ancestor (i.e., they produce >50% of the spores in chimera). Finally, strain rCD1.5_R15 has a mutation in a gene that shows a prestalk biased expression pattern in WT. This gene, DDB_G0278215, encodes an integral transmembrane protein with roles in chemotaxis (inferred through characterization of an ancestral gene) (Gaudet et al. 2011).

3.4 Discussion

The emergence and maintenance of cooperation has been a focus of evolutionary biologists (Axelrod and Hamilton 1981; Smith and Szathmary 1997; Crespi 2001). The social amoebae *Dictyostelium discoideum* is an excellent model system to investigate the maintenance of cooperation (Strassmann et al. 2000; Crespi 2001; Gilbert et al. 2007; Santorelli et al. 2008; Kuzdzal-Fick et al. 2011; Strassmann et al. 2011; Li and Purugganan 2011; Hollis 2012; Celiker and Gore 2013; Jones et al. 2015; Ostrowski et al. 2015). This work has helped to identify over 100 genes that, when knocked out in the lab strain, cause cheating in (Santorelli et al. 2008), and population genomic analysis showed that these genes have increased polymorphism and other signatures of balancing selection, suggesting the selective maintenance of multiple alleles in natural populations (Ostrowski et al. 2015). Finally, prior work has shown that cheating in *D. discoideum* can be countered by the evolution of cheater-resistance (Khare et al. 2009; Levin et

al. 2015), and that counter adaptations between strains can rapidly evolve through spontaneous mutation in lab experiments (Hollis 2012).

Although prior work demonstrated that resistance could evolve to cheating, there was still little knowledge of the genes involved in resistance to cheating – in fact, only a single gene, *rccA*, had been identified (see Khare et al. 2009) and nothing was known about whether selection might be operating on resistance in nature. Using a combination of experimental evolution and population genomic analyses, I have identified several genes that potentially influence resistance to cheating. I have also shown that balancing selection might affect several of these genes, which is interesting, given that balancing selection is the best explanation for polymorphism in loci that impact cheating behaviors in this species (Ostrowski et al. 2015). Signatures of balancing selection have also been identified in genes within co-evolving mutualists (Yoder 2016). Under balancing selection, long-standing genetic variation can persist because of heterozygote advantage (Carrington 1999), negative frequency-dependent selection (where rare alleles are favored) (Carius et al. 2001), or selection in a fluctuating environment (Gillespie and Turelli 1989). In *D. discoideum*, for example, rare resistance alleles may be more or less beneficial depending on the frequency and type of cheating alleles in the population. Although parallelism has commonly been found in experimental evolution studies (Rainey and Travisano 1998; West et al. 2006; Woods et al. 2006; Tenaillon et al. 2012; Deatherage et al. 2017; Zee and Velicer 2017), it is unclear why I found little evidence of parallel evolution at the molecular level in my experiments. One possibility is that there are many genetic avenues available to achieve resistance. However, some events were suggestive of parallel evolution – for example, in two cases I observed different mutations in the same gene (*mkcf* and

DDB_G0277607). Additionally, some of the candidate genes are known to involve similar phenotypic traits.

Looking at parallelism beyond the level of the gene, the analyses of differential expression by cell type or developmental timepoint revealed no evidence of overrepresentation for genes that are expressed in a particular developmental or cell type. This finding also suggests that there may be a variety of routes to available to achieve resistance to cheating. One important limitation is that I have not introduced each of these mutations into the WT background to determine their phenotypic effects apart from other mutations that arose during the experiment. Thus, many of these mutations might not be beneficial, or they could be beneficial but their benefits are unrelated to resistance. Finally, it may also be that these mutations work in tandem with other mutations. Future work is therefore necessary to determine whether and how these mutations might impact resistance by introducing them singly or in combinations in the ancestral genetic background. Additionally, population genomic analyses using a larger data set of sequences might also provide greater sensitivity to uncover the evolutionary history of these genes in nature. However, these analyses lend further support to the hypothesis that both cheating and resistance could be ubiquitous and possibly persistent in nature, and aid in our understanding of how social conflicts might be mitigated in natural populations of cooperators.

CHAPTER 4: POPULATION STRUCTURE OF THE SOCIAL AMOEBA

4.1 Introduction

Cooperative groups are vulnerable to conflict in the form of cheating, where some individuals benefit from the altruism or collective action of others, while failing to contribute their share of the cost (Ferriere et al. 2002; Ghoul et al. 2013; Zhang and Rainey 2013). However, the costs and benefits of cooperation will depend on the number and identity of social partners, and whether these social partners are cooperators or non-cooperators, kin or non-kin, as well as the frequency of these different interactions (Travisano and Velicer 2004; Ostrowski et al. 2008; Dyken et al. 2011; Frénoy et al. 2013; Gruenheit et al. 2017). For this reason, investigating the diversity and distribution of individuals among populations is a valuable tool to understand the impact of population structure on conflict and cooperation in nature.

The social amoebae, *Dictyostelium discoideum,* is an excellent model for investigating social interactions (Strassmann et al. 2000; Fortunato et al. 2003b; Brown and Buckling 2008; Santorelli et al. 2008; Khare et al. 2009; Buttery et al. 2010; Khare and Shaulsky 2010; Kuzdzal-Fick et al. 2010, 2011; Parkinson et al. 2011; Levin et al. 2015; Noh et al. 2018). In this organism, individual amoebae aggregate when starved and form a multicellular fruiting body. The fruiting body consists of a dead stalk, containing cells that altruistically sacrifice themselves to aid in survival and dispersal of the spores, which sit on top of the stalk (Raper 1984; Bonner 2009). Different strains of amoebae can potentially co-develop to form chimeric (multi-genotype) fruiting bodies with non-kin—and under these circumstances, there is potential for cheating, where one genotype does not contribute fairly to the stalk (Strassmann et al. 2000; Santorelli et al. 2008; Khare and Shaulsky 2010).

Although cheating has been well studied in this system (Ennis et al. 2000; Strassmann et al. 2000; Santorelli et al. 2008; Buttery et al. 2010; Flowers et al. 2010; Khare and Shaulsky 2010; Strassmann and Queller 2011; Santorelli et al. 2013; Ostrowski et al. 2015; Noh et al. 2018), knowledge of resistance to cheating in limited (Khare et al. 2009; Hollis 2012; Levin et al. 2015). Prior work has shown that cheating in *D. discoideum* can be countered by the evolution of cheater-resistance (Khare et al. 2009; Hollis 2012; Levin et al. 2015), defined as increased spore production (compared to ancestor) in chimera with a cheater (Khare et al. 2009). For example, Khare et al (2009) isolated a single mutant (*rccA*-) that was resistant to the cheater *chtC*-. To date, *rccA* is the only gene that, when disrupted in the lab strain, is known to confer resistance to a cheating.

While there are multiple examples of how these different genotypes interact in laboratory experiments, it is not known how different genotypes might interact in nature, or even how often they might encounter each other in natural populations. *Dictyostelium discoideum* is found in soil and leaf litter in temperate deciduous forests of the eastern United States, as well as in Japan, Mexico, and Costa Rica (Raper 1984; Cavender and Kawabe 1989; Douglas et al. 2011). Flowers et al. (2010) sequenced 137 gene fragments per strain spread throughout the genome in 24 wild strains (13 of which came from one locale in Virginia, 3 from TX, 1 each from IN, AR, MA, NC, IL, MO, KY, and TN) to investigate the genetic diversity and differentiation in *D discoideum*. Their analyses of SNP data among strains using bootstrap supported neighbor-joining clustering, multiple correspondence analysis, and Bayesian clustering found no clear geographic differentiation (extensive admixture between samples), and they found no evidence off isolation-by-distance (Mantel test $r = 0.11$, $0.05<P<0.11$). However, as the authors state,

they found very low variation among strains (average pairwise difference per site, π, = 0.08%), and *D. discoideum* is a species with very low genetic variation (Ostrowski et al. 2015; Noh et al. 2018), making studies of genetic variation difficult with small sample sizes. Douglas et al. (2011) investigated population differentiation among locations covering large geographic regions (comparing MA, NC, VA, TX, and central America) and found significant pairwise differentiation between regions (pairwise F_{ST} between 0.112 and 0.5). A more recent study looking at population differentiation among four locations separated between 40 and 1000 km, with sample size between 15 and 57 isolates, found that populations are moderately differentiated (mean: F_{ST} = 0.103 , G'_{ST} = 0.548) at this spatial scale (Douglas et al. 2016).

While population structure has been demonstrated over distances of hundreds of kilometers, little is known about population structure over smaller spatial scales, despite the potential existence and importance of smaller spatial scales for microbial social interactions and evolution. I use microsatellite genotyping to address the following questions: what constitutes a population in *D. discoideum* throughout its natural range? What amount of mixing occurs between these populations, and what is the range at which these populations might be interacting? What indication is there that social genes play a role in local adaptation and population structure? I approached these questions by genotyping between 7-21 strains from each of twelve 10 x 10 cm plots, separated by distances ranging from 2.4 m to 1,002 km. I genotyped the strains at 10 randomly chosen microsatellite loci, as well as two microsatellite loci that were in or linked to genes associated with cheating or resistance behaviors, *chtC* and *rccA*. I used these data to estimate and compare population structure across different spatial scales, ranging from the meter scale to ~1,000 km scale, thus covering 6 orders of magnitude. I

estimated migration rate between populations and the genetic diversity within different populations, thereby aiding the understanding of the spatial scale of interactions, which may have important consequences for the evolutionary dynamics of cooperation.

4.2 Methods

4.2.1 Collection

Soil samples were collected from twelve 10 x 10 cm plots in two regional locations between June and July of 2015 and 2016. Eight of the plots were located in the vicinity of Mountain Lake Biological Station ("MLBS") in Pembroke, Virginia. The remaining four plots were located at Proctor Academy in Andover, New Hampshire ("NHPA"). Within the MLBS site, there were eight plots, consisting of four sites (A-D; separated by distances of ~0.5 to 2 km), and within each site, there were two paired plots separated by a few meters (see Figure 4.2.1). MLBS is located in a 259 ha mixed deciduous forested reserve on a 1160 m ridge in the unglaciated Appalachians, with an average annual temperature of 13 °C and average annual rainfall of 1030 mm. Distance between MLBS plots ranges from 2 m to 5.7 m within sites at altitudes from 1167-1255 meters. NHPA is surrounded by a 1011 ha mixed forest on the southern slope of the Ragged Mountains, with an average annual temperature of 12.6 ℃ and average annual rainfall of 976 mm. NHPA plots are separated by distances of .25 to 1.4 km apart at altitudes between 249 and 309 meters. At each plot, the top 1 cm of soil in a 10 x 10 cm square was collected into sterile a Whirl Pak bag. The date, time, altitude, and coordinates were recorded using a handheld GPS unit (Magellan). Samples were stored at 4 ℃ until plating.

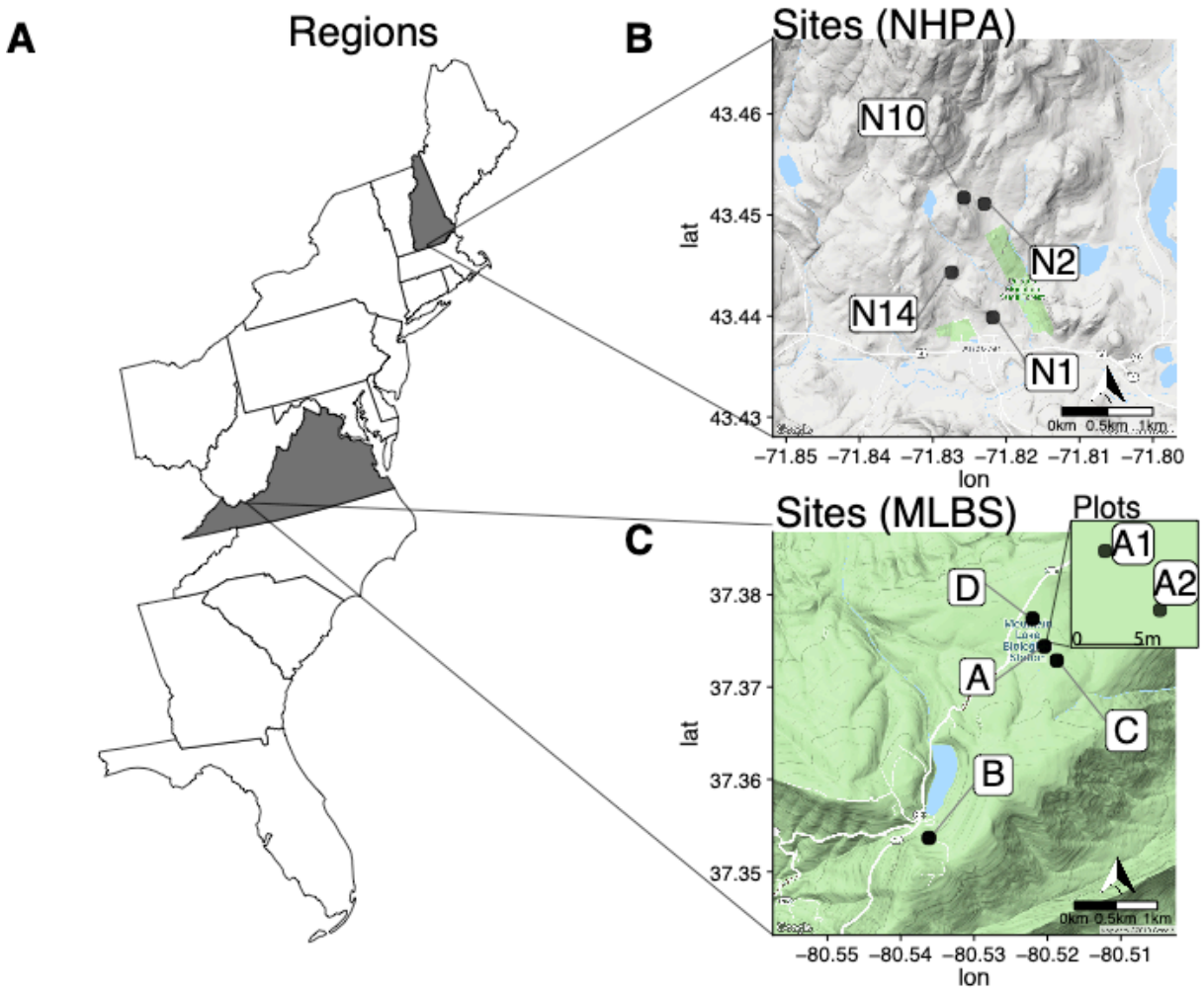

Figure 4.2.1 – Collection sites for *D. discoideum* strains. (A) Regions: New Hampshire and Virginia (B) Each NHPA (New Hampshire, Proctor Academy) sample was obtained from a 10 x 10 cm area, separated by .25 to 1.4 km (C) MLBS sites A-D, separated by 0.5 to 2 km. Within each MLBS site (A-D) a pair of 10 x 10 cm plots are separated by 2 to 6 meters.

4.2.2 Isolation

To isolate *D. discoideum* from soil samples, 6 g of soil were mixed with 30 ml of deionized and autoclaved water (ddH_2O) and agitated to loosen amoebae from soil particles. 0.3 ml of soil slurry was deposited onto hay infusion plates (1.5 g KH_2PO_4, 0.62 g Na_2HPO_4, and 20 g of agar to 1 L of filtered hay infusion, prepared by soaking 15 g hay in 1.5 L deionized water) with 0.4 ml of a *Klebsiella pneumoniae* (Kp) bacterial culture grown in SM broth (10 g of peptone, 1 g

of yeast extract, 10 g of glucose, 1.9 g of KH_2PO_4, 1.3 g of K_2HPO_4, 0.49 g of $MgSO_4$ anhydrous, and 17 g of agar in 1 L of ddH2O). The plates were dried in a laminar flow hood and then incubated at room temperature in the dark. Once slugs had formed, the plates were moved to overhead light to promote fruiting. Spores from fruiting bodies were collected with a sterile pin, transferred to 20% glycerol, and frozen at -80 ℃. Only a single fruiting body was collected per plate. Isolates were later cloned by plating spores at low density on SM plates with Kp and isolating cells or spores that formed from a single circular clearing (plaque) in the bacterial lawn.

4.2.3 DNA extraction and PCR

For each soil sample, twenty cloned isolates were inoculated from frozen stocks onto SM agar plates with 400 µl of *K. pneumoniae.* Approximately 10 fruiting bodies were collected with a sterile pin and deposited into wells of a standard PCR plate containing 50 µl 5% Bio-rad Chelex 100 resin, followed by 3.3 µl Proteinase K [20 µg/ml]. Samples were incubated for four hours at 58 ℃ followed by 1 hour at 98 ℃.

Microsatellite loci were amplified using either custom primers tagged with FAM, HEX, or NED fluorophores or M13-tagged primers (as described by Schuelke 2000) (Table 4.2.1). Microsatellite loci were chosen to cover all six chromosomes in addition to two loci, *chtC* and *rccA,* associated with cheating and resistance to cheating, respectively (Khare et al. 2009; Khare and Shaulsky 2010). PCR reactions were multiplexed such that each reaction resulted in amplification of alleles at three different loci, each tagged with a unique fluorophore to allow

them to be distinguished. Fragment analysis was carried out by capillary electrophoresis on an ABI 3730 sequencer at Arizona State University (School of Life Sciences, OKED DNAlab).

Table 4.2.1 – Microsatellite Primer Sequences for 13 loci. Primer pairs 307 – 366 were previously designed (described by McConnell et al. 2007). Primer pairs rccA and chtC were designed for this study using Tandem Repeat Finder (Benson 1999) and Primer3 (Koressaar and Remm 2007; Untergasser et al. 2012).

Primer	Dir	Primer sequence (5′-3′)	Size (bp)	Gene ID (DDB_G)	Motif	Chr
307	F	AGGATAGCTCTCAGCCATCAA	230	0277255	TCA	2
	R_1	AATGTTGGTTGGGATGATGA				
308	F	CCTGAACAAACACATTCCTCAA	239	0274889	TCA	2
	R_2	GGGGTTATTGTTGGTGCTGA				
317	F	CAATACCACCACCACCACAG	195	0293414	TCA	6
	R_1	GGTGGCGATGATGATGTAGTT				
319	F	GAGTCGATGTAATCAACCATCAG	384	0287449	AAT	5
	R_3	AAAACTGGTACTGCAACCACAA				
323	F	TTGGAAAAAGCCAACAACCT	440	0269582	AAT	1
	R_3	TCAAAGTCCATGGTACAAAACC				
327	F	TGGACAACAACCAATTCAACA	225	0280133	CAA	3
	R_1	TGTGGCTGAAAATTAGGGTCA				
328	F	TTGATCAAAAGATACATCATTATTTGG	238	0290527	CAA	5
	R_2	TGATCAACAGCAACAACAACAA				
329	F	CACAAACCTCAACTTCAACAACA	245	0281709	TCA	3
	R_2	TTGGTTTTGTTGATGACTCAA				
330	F	TTAATCAAAGTCAAATTGGTTTACAA	227	0284853	TCA	4
	R_1	ATTATTGTTATTTGATGATGATGATGT				
357	F	CAATTGGTGAATTTGCTCTAATTT	250	0284313	AAT	4
	R_3	AAAGAAGAAGAGATTGGTAATCAAGA				
366	F	TCAAATCAACTTTGGGAGCA	239	0268102	CAA	1
	R_2	TTTGTTGGTTGTTGTTGTTGC				
chtC	F	TCTCATACTTATGGAGTCCCTTCA	608	0290959	AAT	5
	R_3	ATAGGCAAAACCTTTACAGCA				
rccA	F	CTTCCATTGCTACTGCTGCT	315	0271758	TAT	2
	R_2	CACTTGAATTTGCACTTGCA				

4.2.4 Data analysis

Fragment peaks were analyzed and called using a combination of Peak Scanner 2.0 and the Geneious (v6) microsatellite package v1.4.4. As a control, a standard strain (the lab strain AX4), with known allele sizes, was included on each plate. The control (lab strain AX4) was analyzed on four separate fragment analyses consisting of three different loci. Of the three control loci, the greatest standard deviation across the four runs was 0.54 bp, indicating that the fragment analyses were consistent across runs. Electropherograms were scored for the tallest peak when assigning alleles (Figure 4.2.2).

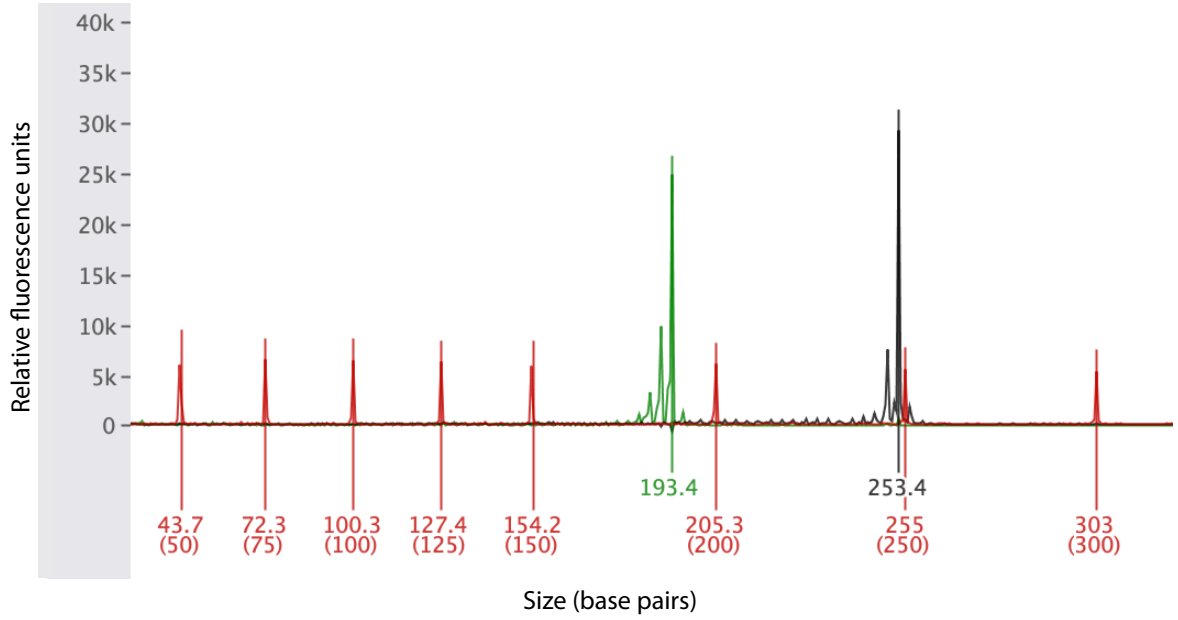

Figure 4.2.2 – Example electropherogram to illustrate peak calling. The reference ladder is shown in red, with the number on top indicating the observed peak location (in numbers of base pairs) and the expected fragment size in parentheses. Two PCR-amplified fragments are shown in green and black. Allele sizes were assigned according to the tallest peak within 3bp bin windows.

If multiple peaks could not be distinguished, the locus was discarded for that genotype. The completed data set consisted of 13% missing data, genotypes with missing data were removed for calculations that are not designed to handle missing data (PCA, AMOVA).

Pairwise F_{ST} was calculated in Arlequin (winarl 3.5) using molecular distance as the number of different alleles. Nei's G_{ST}, Hedrick's G'_{ST} and Jost's D were calculated using the mmod package in R (Nei 1973; Nei and Chesser 1983; Hedrick 2005; Jost 2008; Meirmans and Hedrick 2011; Winter 2012). G_{ST} is a generalized form of Wright's F_{ST} for use with multiple alleles and can thus handle the high mutation rates and diversity of microsatellites (Meirmans and Hedrick 2011). Estimates for these estimators range from 0, no differentiation or no difference in allele frequencies between populations, to 1, complete differentiation or populations are fixed for alternate alleles. However, with high within-population diversity, estimators F_{ST} and G_{ST} will never reach 1. Hedrick's G'_{ST} accounts for this limitation and creates a standardized estimate by dividing the differentiation estimate by the maximum theoretical G_{ST} based on the heterozygosity at each locus (Hedrick 2005). Jost's D is an alternative measure of differentiation that creates a standardized estimate by measuring the fraction of allelic variation among populations (Jost 2008).

To assess the correlation between genetic distance and geographic distance a matrix of genetic chord distance (Cavalli-Sforza and Edwards 1967) between plots was compared to the matrix of geographic distances between them using the mantel.rtest function within the ade4 (Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences) toolset (Bougeard & Dray, 2018; Chessel et al. 2004; Dray & Dufour, 2007) within the *poppr*R package (Kamvar et al. 2015; Kamvar et al. 2014). Analyses of molecular variation (AMOVA) were carried out using ade4. Principal component analyses were performed following centering

and scaling (as described previously by Odong et al. 2013; Putman and Carbone 2014), using the dudi.pca function within the ade4 R package.

4.3 Results

In total, 367 strains from two collection seasons were genotyped using short tandem repeats (STR), known as microsatellites. For the 2015 collection, 177 strains from eight plots were genotyped for nine loci. In the 2016 collection, 190 samples from 12 plots were genotyped for 11 loci (Table 4.2.1). An average of 18 and 24 alleles per locus were identified for the 2015 and 2016 samples, respectively. For the 2015 samples, I found 150 unique genotypes with a mean diversity within populations (HS) ranging from 0.55 for locus 308 to 0.73 for locus 327. For the 2016 samples I found 189 unique genotypes with a mean diversity within populations ranging from 0.61 for locus rccA to 0.83 for locus 329. Within the 2015 samples I found lower levels of genetic differentiation at the two loci with roles in cheating and resistance behaviors, rccA and chtC, (G'ST = 0.61 and 0.59, respectively). Both G'ST and Jost's D estimates for rccA and chtC fall below the 95% confidence intervals for random loci (see Table 4.3.1).

Table 4.3.1 – Genetic diversity analyses. Sets consisted of nine and 11 loci for 2015 (A) & 2016 (B) sample sets, respectively. Total number of alleles scored, genetic diversity of pooled populations (H_T), mean diversity within populations (H_S), Nei's coefficient of gene differentiation (G_{ST}) (Nei 1973), Hedrick's G'_{ST} (Hedrick 2005), Jost's D (Jost 2008).

A

Locus	# Alleles	*Hs*	*Ht*	G_{ST}	G'_{ST}	*D*
L317	19	0.640	0.815	0.215	0.674	0.569
L327	24	0.640	0.816	0.216	0.675	0.570
L357	24	0.628	0.866	0.275	0.825	0.747
L308	18	0.545	0.847	0.357	0.863	0.775
L329	25	0.693	0.868	0.202	0.743	0.667
L319	24	0.624	0.822	0.241	0.719	0.615
L323	20	0.635	0.869	0.269	0.824	0.748
rccA	13	0.546	0.725	0.247	0.610	0.461
chtC	14	0.627	0.779	0.196	0.593	0.477
Global		0.619	0.823	0.247	0.728	0.605

B

Locus	# Alleles	*Hs*	*Ht*	G_{ST}	G'_{ST}	*D*
L329	31	0.673	0.940	0.284	0.935	0.906
L357	28	0.688	0.941	0.269	0.929	0.900
L330	24	0.655	0.925	0.292	0.910	0.868
L366	22	0.590	0.908	0.351	0.915	0.863
L317	28	0.729	0.930	0.216	0.864	0.822
L328	30	0.647	0.934	0.307	0.935	0.903
L308	29	0.651	0.920	0.293	0.902	0.858
L327	25	0.691	0.918	0.248	0.866	0.817
rccA	15	0.477	0.790	0.396	0.806	0.665
L319	11	0.558	0.847	0.342	0.827	0.728
L307	27	0.714	0.879	0.188	0.716	0.642
Global		0.643	0.903	0.288	0.868	0.804

I calculated population differentiation at three hierarchical levels: plot (m-scale), site (km-scale), and region (~1000 km scale). Pairwise F_{ST} among plots ranged from 0.01-0.24, and all but three values were statistically significant based on a non-parametric permutation test (as

described by Excoffier et al. 1992). All three non-significant F_{ST} values consisted of comparisons to plot D2, which had the lowest sample number (n = 7 isolates). At larger spatial scales, all comparisons showed significant differentiation (mean pairwise F_{ST} = 0.137).

Table 4.3.2 – Pairwise F_{ST} among plots. The Virginia (MLBS) locations are indicated in bold, whereas the New Hampshire locations (NHPA) are indicated by italics.

Plot	**A1**	**A2**	**B1**	**B2**	**C1**	**C2**	**D1**	**D2**	*N1*	*N10*	*N14*
A1											
A2	0.09*										
B1	0.12*	0.13*									
B2	0.11*	0.05*	0.1*								
C1	0.1*	0.05*	0.1*	0.07*							
C2	0.13*	0.09*	0.19*	0.12*	0.1*						
D1	0.18*	0.11*	0.19*	0.13*	0.07*	0.18*					
D2	0.08*	0.04	0.1*	0.06*	0.01	0.08*	0.04				
N1	0.21*	0.18*	0.19*	0.18*	0.17*	0.24*	0.22*	0.15*			
N10	0.19*	0.15*	0.19*	0.17*	0.12*	0.22*	0.23*	0.14*	0.22*		
N14	0.17*	0.13*	0.16*	0.14*	0.13*	0.2*	0.2*	0.14*	0.18*	0.16*	
N2	0.12*	0.08*	0.11*	0.09*	0.07*	0.12*	0.15*	0.04*	0.14*	0.09*	0.08*

Table 4.3.3 – Pairwise F_{ST} values. (A) Among sites within regions. The Virginia (MLBS) locations are indicated in bold, whereas the New Hampshire locations (NHPA) are indicated by italics. (B) Among regions

A

Site	**A**	**B**	**C**	**D**	*N1*	*N10*	*N14*
A							
B	0.05*						
C	0.05*	0.07*					
D	0.08*	0.09*	0.07*				
N1	0.17*	0.15*	0.18*	0.18*			
N10	0.14*	0.15*	0.15*	0.18*	0.22*		
N14	0.12*	0.12*	0.15*	0.17*	0.18*	0.16*	
N2	0.08*	0.08*	0.08*	0.1*	0.14*	0.09*	0.08*

B

Region	MLBS
MLBS	
NHPA	0.058*

Significant isolation by distance (2015 samples: $r = 0.729$, simulated $p = 0.001$. 2016 samples: $r = 0.703$, simulated $p = 0.001$, 1000 permutations) was detected using Mantel tests between genetic (chord) distance and geographic distance (Figure 4.3.1).

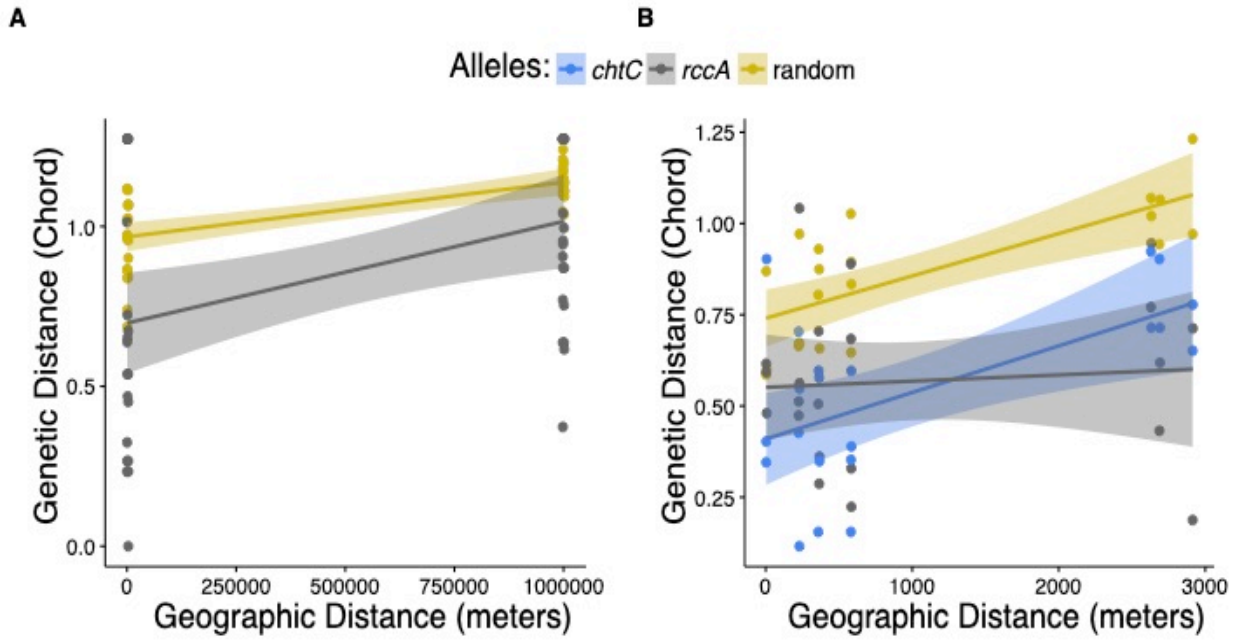

Figure 4.3.1 – Isolation by distance (IBD) plots. (A) 2016 samples and (B) 2015 samples grouped by random loci (yellow) and loci associated with cheating or resistance (blue or grey, respectively). Each point represents one comparison between two sample plots.

Further support for genetic differentiation at all spatial scales is found in analyses of molecular variation (AMOVA, Table 4.3.4). The AMOVA does not support differentiation among sites (*e.g.,* A vs B) within regions. According to the AMOVA, the majority of genetic variation within the 2016 samples (84.6%) was from variation within plots, although there was also a significant portion (8.4%) from variation among plots, within sites (*e.g.*, A1 vs A2). A small (3.4%) but significant portion of variation was from between region variation.

Table 4.3.4 – Analysis of molecular variance (AMOVA) for amplified microsatellite fragment length polymorphism data. The 12 *Dictyostelium discoideum* samples from MLBS, VA and Proctor Academy, NH, collected in 2016. Significance values were obtained by performing Monte-Carlo tests over 999 permutations on clone corrected data.

Hierarchical level	*df*	%	Obs	Std.Obs	*P*
Between Regions	1.00	3.34	0.35	3.84	0.025
Between Sites Within Region	6.00	3.48	0.36	1.51	0.073
Between Samples Within Sites	4.00	8.36	0.87	10.76	0.001
Within Samples	177.00	84.82	8.81	-43.44	0.001
Total	188.00	100.00			

To better understand the relationships among the sample populations of this study, I performed a principal component analyses (PCA) (Figure 4.3.2). In the PCA based on the 2016 samples, the first three axes explain 25.5% of the variation (Figure 4.3.2A). These components show that New Hampshire collectively harbors more overall variation than Virginia, indicated by the greater spread of the points (Figure 4.3.2A). Interestingly, there is greater distance among some pairs of sites within NHPA than there is among some pairs of sites between the two regions (NH and VA).

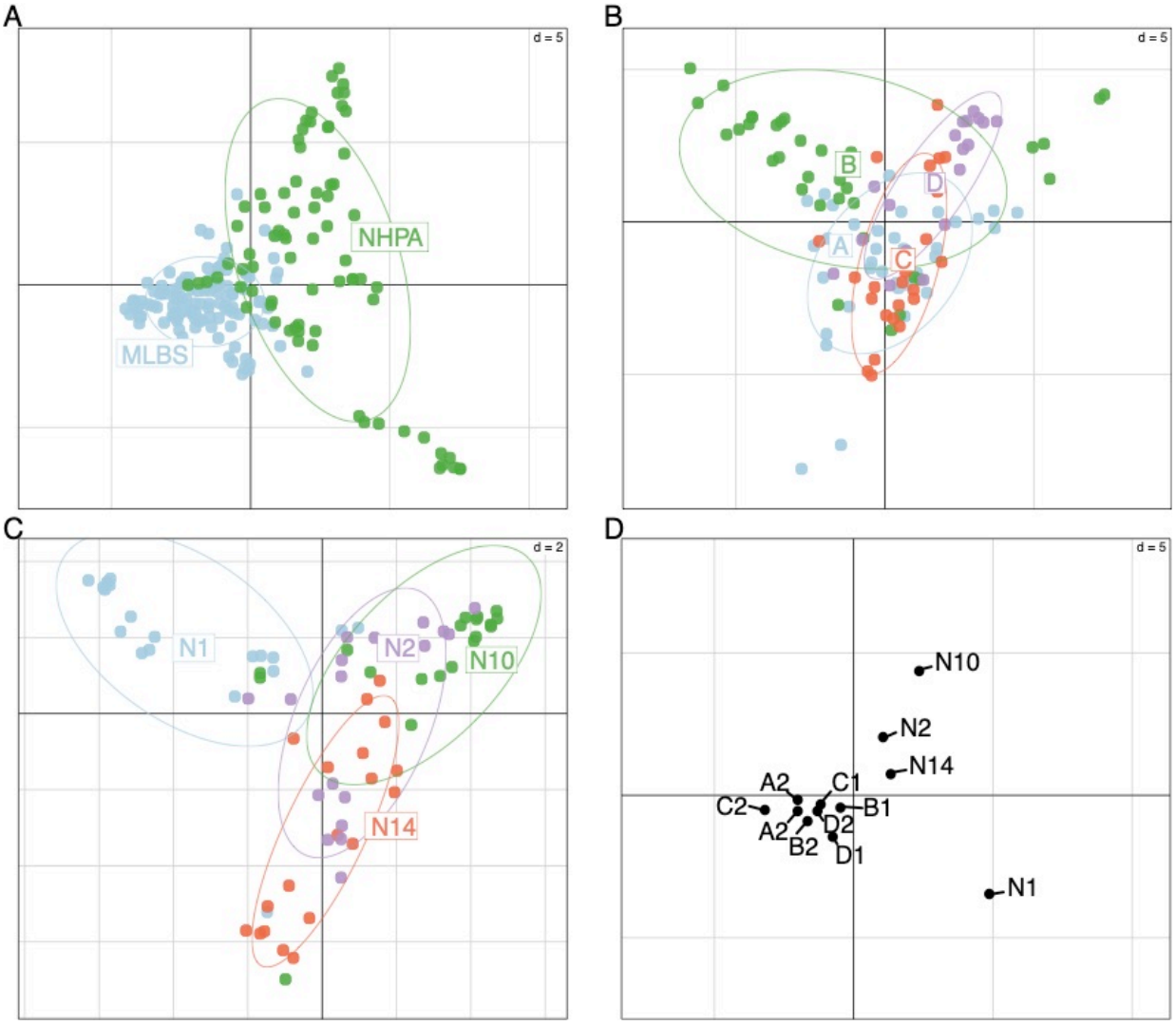

Figure 4.3.2 – Principal component analysis (PCA) showing positions of all individuals for 2016 samples. In A-C, each point represents one strain. In panel D, each point represents a single plot, calculated as the centroid (the arithmetic mean) of the strains within each plot. Panel A shows the distinct clustering of samples by region. Panels B and C show separate analyses for MLBS and NHPA, respectively. Analyses performed using the R package ade4 v1.7-13 (Dray and Dufour 2007).

The genetic population structure of the *D. discoideum* strains was also inferred using STRUCTURE 2.4.3 (Pritchard et al. 2000). The software uses a Bayesian model-based clustering method to infer population structure from multi-locus genotype data and assign individuals to *K* populations, where *K* is equal to the number of individual clusters or

populations. The log-likelihood of each individual belonging to a given cluster was averaged over 20 runs for each value of *K* ranging from *K* = 1 to *K* = #plots + 3, where #plots = the number of sampled plots (A1, A2, etc). To estimate *K,* I followed the Δ*K* method described by Evanno, Regnaut, & Goudet (2005). For the 2015 samples, the populations identified by STRUCTURE corresponded well to the collection sites, although STRUCTURE identified only three clusters instead of four, indicating some admixture (Figure 4.3.3C). A neighbor joining tree based on Hedrick's estimator of pairwise G'_{ST} (Figure 4.3.3B) resulted in a similar pattern of clustering compared to the STRUCTURE model.

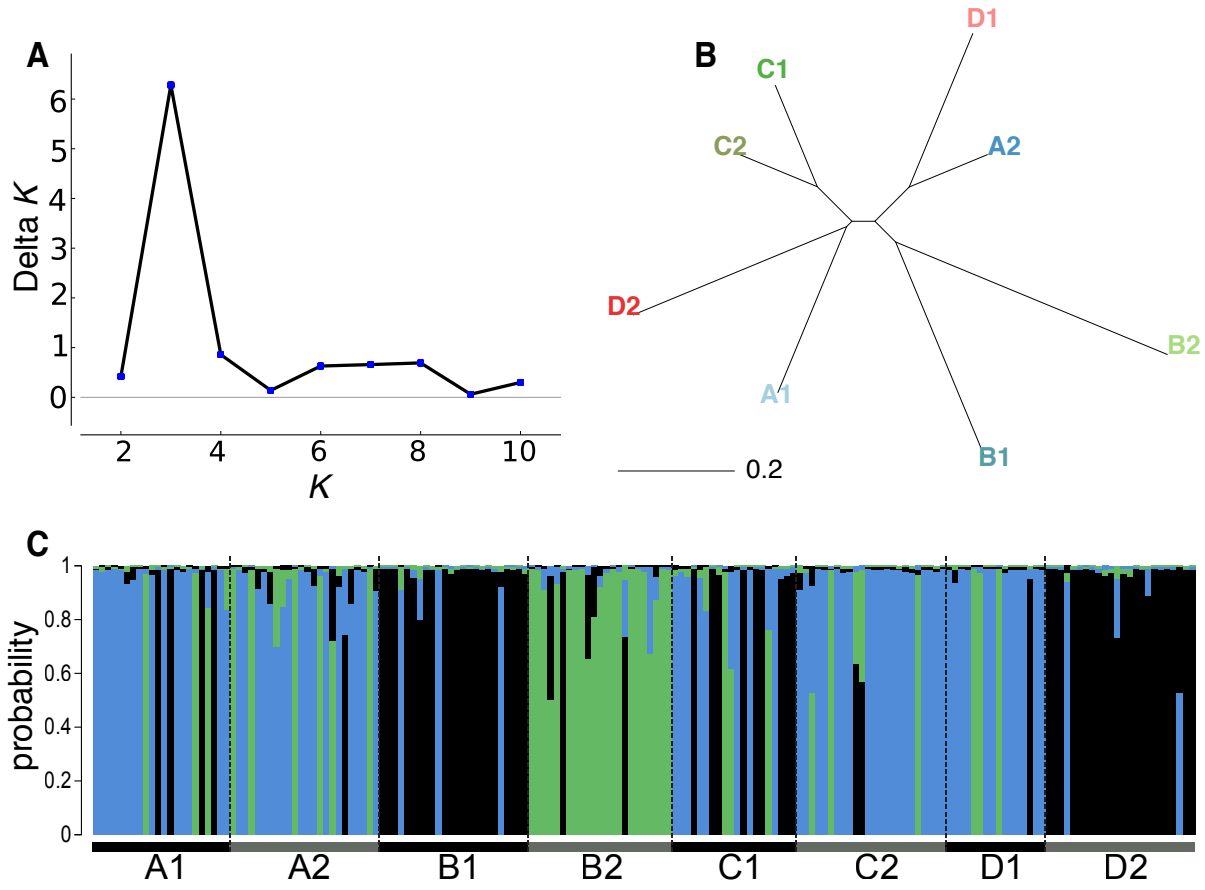

Figure 4.3.3 – Population structure analysis of 2015 samples. (A) Δ*K* for each value of *K* from 2 to 10. (B) Neighbor joining tree based on Hedrick's estimator of pairwise G_{ST} between each 10 x10 cm plot. Only MLBS. (C) Model-based clustering from multi-locus genotype data for *K* = 3 from Structure 2.3.4.

Similar to the PCA (Figure 4.3.2), these analyses reveal strong differentiation between New Hampshire and Virginia (e.g., see Figure 4.3.4B), with these sites not showing an intermingling within the tree. Similarly, there is little to no admixture between the New Hampshire and Virginia sites based on the STRUCTURE analyses. In contrast to strong differentiation at the regional level, there is greater evidence of gene flow within the regions. For example, the closest neighbors in the G'_{ST}-based NJ tree are sometimes (e.g., C1 and C2, D1 and D2 in 2016) closest geographically, but not always (e.g., A1 and A2 in 2016).

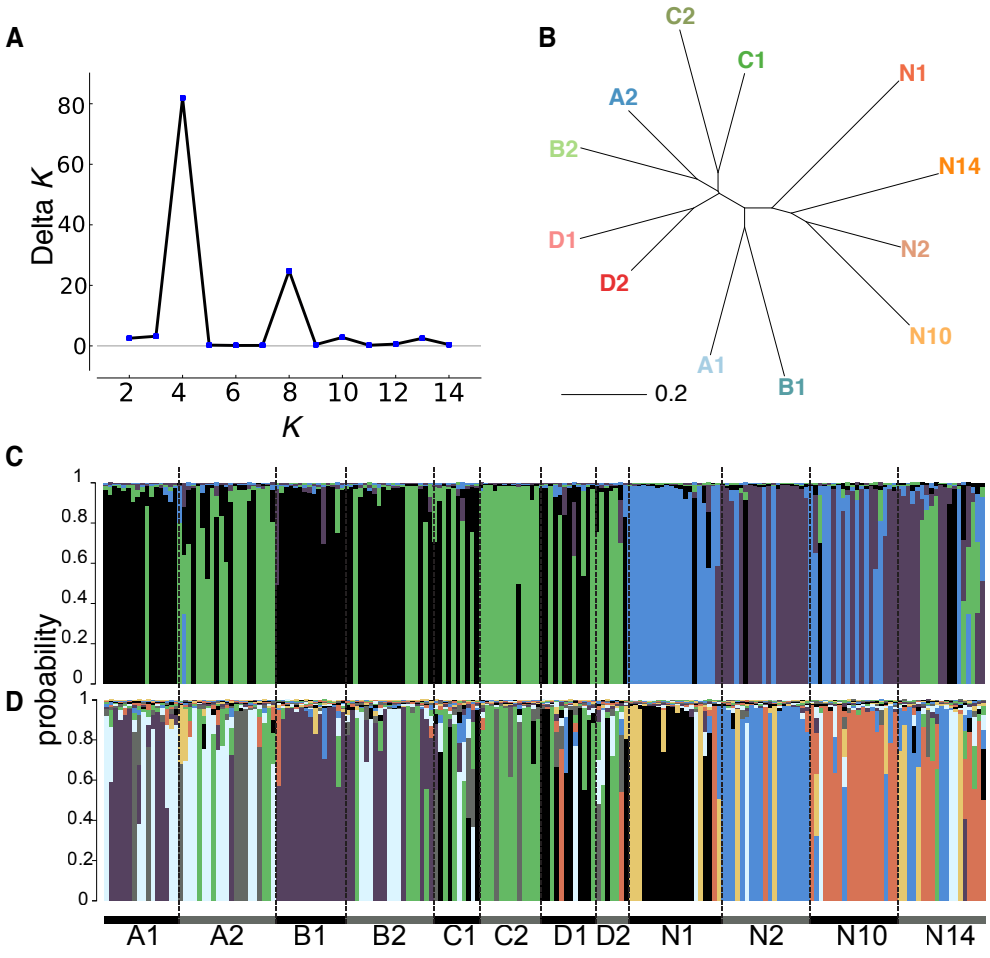

Figure 4.3.4 – Population structure analysis of 2016 samples. (A) ΔK for each value of K from 2 to 10. (B) Neighbor joining tree based on Nei's estimator of pairwise G_{ST} between each 10 x 10 cm sample plot. Model-based clustering from multi-locus genotype data for (C) $K = 4$ and (D) $K = 8$ from Structure 2.3.4.

4.4 Discussion

Classically, microorganisms were thought to have unrestricted migration because of their apparent lack of barriers to distribution, large population size, and small body size (Zwart et al. 1998; Finlay 2002; Finlay and Fenchel 2004). However, more recent evidence suggests microbes can have biogeographic patterns similar to macro-organisms (Bell 2010; Lindström and Östman 2011; Hanson et al. 2012; Albright and Martiny 2018). For example, a recent biogeography study of the model ciliate *Tetrahymena thermophila* investigated the limits to distribution and diversity of the microbial eukaryote to test whether these protists are ubiquitous or endemic. Zufall, Dimond, & Doerder, (2013) sampled ciliates from 43 ponds and streams downstream from ponds throughout the eastern United States and estimated significantly high population subdivision (F_{ST} between 0.14 and 0.5) and low rates of migration (<1 migrant per generation) between sample sites. My analyses support the existence of genetic structure in *Dictyostelium discoideum* among samples at all levels of organization (plot, site, and region). Similar to Douglas *et al*., (2011), I found differentiation at the largest spatial scale, which covered ~1000 km in this study. However, I found evidence of differentiation at spatial scales as small as a few meters. In fact, these results (supported by AMOVA) suggest that population structure in *D. discoideum* occurs at a scale of several meters or less. Fine-scale population structure was also observed in another social microbe, the bacterium *Mixococcus xanthus*, which also forms aggregative fruiting bodies in response to starvation. Vos and Velicer (2006) collected 78 *M. xanthus* isolates from a 16 by 16 cm grid of soil samples (each separated by 1cm) and found evidence of genetic population structure over spatial scales as small as 1cm. In *Caenorhabditis elegans,* high genetic diversity was observed at the scale of a few centimeters, which was surprising, given fairly low global genetic diversity. The authors suggest that this

pattern arises from a high migration rate in combination with local bottlenecks—in this case, a high migration rate prevents isolation by distance at large spatial scales, but bottlenecks cause allele frequency shifts (and thus population structure) at small spatial scales (Barrière and Félix 2005). Together this body of evidence supports that microbes are specially structured at small scale despite their seemingly abundant opportunities for dispersal.

In addition to fine-scale population structure, I found fewer alleles and lower levels of population structure (i.e., more gene flow) at the two loci with known roles in cheating and resistance behaviors, *rccA* and *chtC*, compared to randomly chosen loci. On one hand, these results might indicate stronger purifying selection on the social genes, which would reduce genetic diversity at these loci compared to others. On the other hand, Ostrowski et al. 2015 found lower levels of population structure, on average, for candidate genes that influence cheating behaviors compared to other genes in the genome, albeit in combination with other signatures of balancing selection.

Another factor that could influence population structure is seasonality. The sampling of multiple timepoints revealed a shift of genetic differentiation across seasons. For example, some plots within the MLBS samples (e.g., B1, B2 in figures 4.3.3B and 4.3.4B) cluster differently across season, but this finding would need to be confirmed with additional timepoints. Because the sampling locations experience freezing conditions in winter, these populations might experience population bottlenecks across seasons. If so, strains that form dormant spores before freezing temperatures arrive might survive better (Villa Martín et al. 2019). Alternatively, strains that recover faster from dormancy might lead a succession in following seasons (Lennon and Jones

2011). However, our preliminary population genomic sequencing results (Kuzdzal-Fick et al, in prep) indicate that multiple, genetically divergent strains are present within each 10 x 10 cm site, and these same strains are present across years. Although preliminary, this result suggests that there is some stability in population genetic composition across time and argues against a strong role for seasonality—at least in terms of losses of genetic diversity, although further analyses are needed.

Seasonality is also quite different between my two main sample locations, MLBS and NHPA (in Virginia and New Hampshire, respectively). The collection sites are at different elevations and longitudes and therefore experience different timing, intensity, and longevity of seasons. The STRUCTURE analyses also revealed greater admixture of strains among sites at MLBS location compared to among sites at NHPA, raising the question of how environmental differences between these locations might impact genetic variation and structure at these two locations.

Overall, my findings of small scale structure suggest that microbes like *D. discoideum* do have differentiation at spatial scales that should allow for interaction between diverse genotypes, yet other factors, such as social interactions and discrimination, could be preventing local establishment of rare migrants leading to structure at a small scale (Yanni et al. 2019). If migration occurs at this small scale, it is possible that the formation of chimera could occur between genetically distinct strains. If these interactions between genotypes occur in nature, and the strains have differing altruistic investment to stalk production (Votaw and Ostrowski 2017), this could allow for the invasion of less altruistic or cheating strains into new populations which

could then sweep through. Therefore, it is important to understand how often different individuals encounter each other in the soil and how often these encounters might be between individuals that are geographically and genetically divergent, potentially posing a risk to the cooperative individuals.

BIBLIOGRAPHY

Albright, M. B. N., and J. B. H. Martiny. 2018. Dispersal alters bacterial diversity and composition in a natural community. ISME J. 12:296–299.

Asfahl, K. L., J. Walsh, K. Gilbert, and M. Schuster. 2015. Non-social adaptation defers a tragedy of the commons in *Pseudomonas aeruginosa* quorum sensing. ISME J. 9:1734–1746.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Barrière, A., and M.-A. Félix. 2005. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr. Biol. 15:1176–1184.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48.

Bell, T. 2010. Experimental tests of the bacterial distance–decay relationship. ISME J. 4:1357–1365.

Benghezal, M., S. Cornillon, L. Gebbie, L. Alibaud, F. Brückert, F. Letourneur, and P. Cosson. 2003. Synergistic control of cellular adhesion by transmembrane 9 proteins. Mol. Biol. Cell 14:2890–2899.

Bennett, A. F., and B. S. Hughes. 2009. Microbial experimental evolution. Am. J. Physiol. Integr. Comp. Physiol. 297:R17–R25.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–80.

Bergelson, J. 2001. Evolutionary dynamics of plant R-genes. Science 292:2281–2285.

Bonner, J. T. 2009. The social amoebae. Princeton University Press, New Jersey.

Boucher, D. H. 1988. The biology of mutualism: ecology and evolution. Oxford University Press, New York.

Boucher, D. H., S. James, and K. H. Keeler. 1982. The ecology of mutualism. Annu. Rev. Ecol. Syst. 13:315–347.

Bougeard, S., and S. Dray. 2018. Supervised multiblock analysis in R with the ade4 package. J. Stat. Softw. 86:1–17.

Brandt, M., B. Fischer-Blass, J. Heinze, and S. Foitzik. 2007. Population structure and the co-evolution between social parasites and their hosts. Mol. Ecol. 16:2063–2078.

Brown, S. P., and A. Buckling. 2008. A social life for discerning microbes. Cell 135:600–603.

Buskirk, S. W., R. E. Peace, and G. I. Lang. 2017. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proc. Natl. Acad. Sci. 114:8330–8335.

Buss, L. W. 1982. Somatic cell parasitism and the evolution of somatic tissue compatibility. Proc. Natl. Acad. Sci. 79:5337–5341.

Butaitė, E., M. Baumgartner, S. Wyder, and R. Kümmerli. 2017. Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. Nat. Commun. 8:1–12.

Buttery, N. J., C. R. L. Thompson, and J. B. Wolf. 2010. Complex genotype interactions influence social fitness during the developmental phase of the social amoeba *Dictyostelium discoideum*. J. Evol. Biol. 23:1664–1671.

Carius, H. J., T. J. Little, and D. Ebert. 2001. Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. Evolution 55:1136–1145.

Carrington, M. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science 283:1748–1752.

Cavalli-Sforza, L. L., and A. W. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19:233–57.

Cavender, J. C., and K. Kawabe. 1989. Cellular slime molds of Japan. I. Distribution and biogeographical considerations. Mycologia 81:683–691.

Celiker, H., and J. Gore. 2013. Cellular cooperation: Insights from microbes. Trends Cell Biol. 23:9–15.

Chambers, J. M., and T. J. Hastie. 1992. Statistical models in S, Wadsworth & Brooks/Cole mathematics series. Springer.

Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2:379–384.

Chessel, D., A.-B. Dufour, and J. Thioulouse. 2004. The ade4 package -- I: One-table methods. R News 4:5–10.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 6:80–92.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proc. R. Soc. B Biol. Sci. 279:5039–5047.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. 100:1072–1077.

Crespi, B. J. 2001. The evolution of social behavior in microorganisms. Trends Ecol. Evol. 16:178–183.

Dandekar, A. A., S. Chugani, and E. P. Greenberg. 2012. Bacterial quorum sensing and metabolic incentives to cooperate. Science 338:264–266.

Dawkins, R., and J. R. Krebs. 1979. Arms races between and within species. Proc. R. Soc. London. Ser. B. Biol. Sci. 205:489–511.

Deatherage, D. E., J. L. Kepner, A. F. Bennett, R. E. Lenski, and J. E. Barrick. 2017. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. Proc. Natl. Acad. Sci. 114:E1904–E1912.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework

for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–501.

Diggle, S. P., A. S. Griffin, G. S. Campbell, and S. A. West. 2007. Cooperation and conflict in quorum-sensing bacterial populations. Nature 450:411–414.

Douglas, T. E., M. R. Kronforst, D. C. Queller, and J. E. Strassmann. 2011. Genetic diversity in the social amoeba *Dictyostelium discoideum*: Population differentiation and cryptic species. Mol. Phylogenet. Evol. 60:455–462. Elsevier Inc.

Douglas, T. E., J. E. Strassmann, and D. C. Queller. 2016. Sex ratio and gamete size across eastern North America in *Dictyostelium discoideum*, a social amoeba with three sexes. J. Evol. Biol. 29:1298–1306.

Dray, S., and A.-B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Dyken, J. D. Van, T. A. Linksvayer, M. J. Wade, J. D. Van Dyken, T. A. Linksvayer, and M. J. Wade. 2011. Kin selection–mutation balance: A model for the origin, maintenance, and consequences of social cheating. Am. Nat. 177:288–300.

Eichinger, L., J. A. Pachebat, G. Glockner, M.-A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabbinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, and A. Kuspa. 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43–57.

Elena, S. F., and R. E. Lenski. 2003. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4:457–469.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton.

Ennis, H. L., D. N. Dao, S. U. Pukatzki, and R. H. Kessin. 2000. *Dictyostelium* amoebae lacking an F-box protein form spores rather than stalk in chimeras with wild type. Proc. Natl. Acad. Sci. 97:3292–3297.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–91.

Ferriere, R., J. L. Bronstein, S. Rinaldi, R. Law, and M. Gauduchon. 2002. Cheating and the evolutionary stability of mutualisms. Proc. R. Soc. London. Ser. B Biol. Sci. 269:773–780.

Ferrière, R., M. Gauduchon, and J. L. Bronstein. 2007. Evolution and persistence of obligate mutualists and exploiters: competition for partners and evolutionary immunization. Ecol. Lett. 10:115–126.

Fiedler, K. 1991. Systematic, evolutionary, and ecological implications of myrmecophily within the Lycaenidae (Insecta: Lepidoptera: Papilionoidea). Issue 31 o. Zoologisches Forschungsinstitut und Museum Alexander Koenig, Bonn.

Fiegna, F., and G. J. Velicer. 2003. Competitive fates of bacterial social parasites: persistence and self–induced extinction of *Myxococcus xanthus* cheaters. Proc. R. Soc. London. Ser. B Biol. Sci. 270:1527–1534.

Fiegna, F., Y.-T. N. Yu, S. V Kadam, G. J. Velicer, N. Y. Yuen-Tsu, S. V Kadam, and G. J. Velicer. 2006. Evolution of an obligate social cheater to a superior cooperator. Nature 441:310–314.

Finlay, B. J. 2002. Global dispersal of free-living microbial eukaryote species. Science 296:1061–1063.

Finlay, B. J., and T. o. m. Fenchel. 2004. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 155:237–244.

Flowers, J. M., S. I. Li, A. Stathos, G. Saxer, E. A. Ostrowski, D. C. Queller, J. E. Strassmann, and M. D. Purugganan. 2010. Variation, sex, and social cooperation: Molecular population genetics of the social amoeba *Dictyostelium discoideum*. PLoS Genet. 6:1–14.

Fortunato, A., D. C. Queller, and J. E. Strassmann. 2003a. A linear dominance hierarchy among clones in chimeras of the social amoeba *Dictyostelium discoideum*. J. Evol. Biol. 16:438–445.

Fortunato, A., J. E. Strassmann, L. a Santorelli, and D. C. Queller. 2003b. Co-occurrence in nature of different clones of the social amoeba, *Dictyostelium discoideum*. Mol. Ecol. 12:1031–1038.

Foster, K. R., A. Fortunato, J. E. Strassmann, and D. C. Queller. 2002. The costs and benefits of being a chimera. Proc. R. Soc. London. Ser. B Biol. Sci. 269:2357–2362.

Frankham, R. 2005. Genetics and extinction. Biol. Conserv. 126:131–140.

Franssen, S. U., R. Kofler, and C. Schlötterer. 2017. Uncovering the genetic signature of quantitative trait evolution with replicated time series data. Heredity (Edinb). 118:42–51.

Frénoy, A., F. Taddei, and D. Misevic. 2013. Genetic architecture promotes the evolution and maintenance of cooperation. PLOS Comput. Biol. 9:1–12.

Fukuzawa, M. 2006. Regulation of *Dictyostelium* prestalk-specific gene expression by a SHAQKY family MYB transcription factor. Development 133:1715–1724.

Gaudet, P., M. S. Livstone, S. E. Lewis, and P. D. Thomas. 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief. Bioinform. 12:449–462.

Gebbie, L., M. Benghezal, S. Cornillon, R. Froquet, N. Cherix, M. Malbouyres, Y. Lefkir, C. Grangeasse, S. Fache, J. Dalous, F. Brückert, F. Letourneur, and P. Cosson. 2004. Phg2, a kinase involved in adhesion and focal site modeling in *Dictyostelium*. Mol. Biol. Cell 15:3915–3925.

Ghoul, M., A. S. Griffin, and S. a. West. 2013. Toward an evolutionary definition of cheating. Evolution 68:318–31.

Gilbert, O. M., K. R. Foster, N. J. Mehdiabadi, J. E. Strassmann, and D. C. Queller. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. Proc. Natl. Acad. Sci. 104:8913–8917.

Gillespie, J. H., and M. Turelli. 1989. Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121:129–38.

Glöckner, G., K. Szafranski, T. Winckler, T. Dingermann, M. A. Quail, E. Cox, L. Eichinger, A. A. Noegel, and A. Rosenthal. 2001. The complex repeats of *Dictyostelium discoideum*. Genome Res. 11:585–94.

Good, J. R. 2003. TagA, a putative serine protease/ABC transporter of *Dictyostelium* that is required for cell fate determination at the onset of development. Development 130:2953–2965.

Griffin, A. S., S. A. West, and A. Buckling. 2004. Cooperation and competition in pathogenic bacteria. Nature 430:1024–1027.

Gruenheit, N., K. Parkinson, B. Stewart, J. A. Howie, J. B. Wolf, and C. R. L. Thompson. 2017. A polychromatic 'greenbeard' locus determines patterns of cooperation in a social amoeba. Nat. Commun. 8:1–9.

Grünwald, N. J., B. A. McDonald, and M. G. Milgroom. 2016. Population genomics of fungal and oomycete pathogens. Annu. Rev. Phytopathol. 54:323–346.

H. Koschwanez, J., K. R. Foster, and A. W. Murray. 2011. Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. PLoS Biol. 9:1–10.

Hamilton, W. D. 1964a. The genetical evolution of social behaviour. I. J. Theor. Biol. 7:1–16.

Hamilton, W. D. 1964b. The genetical evolution of social behaviour. II. J. Theor. Biol. 7:17–52.

Hanson, C. A., J. A. Fuhrman, M. C. Horner-Devine, and J. B. H. Martiny. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat. Rev. Microbiol. 10:497–506.

Harvey, P., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hibbing, M. E., C. Fuqua, M. R. Parsek, and S. B. Peterson. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nat. Rev. Microbiol. 8:15–25.

Hilbe, C., Š. Šimsa, K. Chatterjee, and M. A. Nowak. 2018. Evolution of cooperation in stochastic games. Nature 559:246–249.

Ho, H.-I., S. Hirose, A. Kuspa, and G. Shaulsky. 2013. Kin recognition protects cooperators

against cheaters. Curr. Biol. 23:1590–1595.

Hohenlohe, P. A., P. C. Phillips, and W. A. Cresko (eds). 2010. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. Int. J. Plant Sci. 171:1059–1071.

Hojo, M. K., N. E. Pierce, and K. Tsuji. 2015. *Lycaenid* caterpillar secretions manipulate attendant ant behavior. Curr. Biol. 25:2260–2264.

Hollis, B. 2012. Rapid antagonistic coevolution between strains of the social amoeba *Dictyostelium discoideum*. Proc. R. Soc. B Biol. Sci. 279:3565–3571.

Huse, H. K., T. Kwon, J. E. A. Zlosnik, D. P. Speert, E. M. Marcotte, and M. Whiteley. 2010. Parallel Evolution in *Pseudomonas aeruginosa* over 39,000 Generations In Vivo. MBio 1:1–8.

Inglis, R. F., S. West, and A. Buckling. 2014. An experimental study of strong reciprocity in bacteria. Biol. Lett. 10:1–3.

Jasmin, J. N., and C. Zeyl. 2013. Evolution of pleiotropic costs in experimental populations. J. Evol. Biol. 26:1363–1369.

Jha, A. R., C. M. Miles, N. R. Lippert, C. D. Brown, K. P. White, and M. Kreitman. 2015. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in *Drosophila melanogaster*. Mol. Biol. Evol. 32:2616–2632.

Jones, E. I., M. E. Afkhami, E. Akcay, J. L. Bronstein, R. Bshary, M. E. Frederickson, K. D. Heath, J. D. Hoeksema, J. H. Ness, M. S. Pankey, S. S. Porter, J. L. Sachs, K. Scharnagl, and M. L. Friesen. 2015. Cheaters must prosper : reconciling theoretical and empirical perspectives on cheating in mutualism. Ecol. Lett. 18:1270–1284.

Jost, L. 2008. G ST and its relatives do not measure differentiation. Mol. Ecol. 17:4015–4026.

Kamvar, Z. N., J. C. Brooks, and N. J. Grünwald. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6:1–10.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:1–14.

Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. 15:173–190.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends Ecol. Evol. 27:547–560.

Khare, A., L. a Santorelli, J. E. Strassmann, D. C. Queller, A. Kuspa, and G. Shaulsky. 2009. Cheater-resistance is not futile. Nature 461:980–982.

Khare, A., and G. Shaulsky. 2010. Cheating by exploitation of developmental prestalk patterning in *Dictyostelium discoideum*. PLoS Genet. 6:1–12.

Klein, J., A. Sato, S. Nagl, and C. O'hUigín. 1998. Molecular trans-species polymorphism. Annu. Rev. Ecol. Syst. 29:1–21.

Kofler, R., and C. Schlötterer. 2014. A guide for the design of evolve and resequencing studies. Mol. Biol. Evol. 31:474–483.

Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291.

Kozarewa, I., Z. Ning, M. A. Quail, M. J. Sanders, M. Berriman, and D. J. Turner. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods 6:291–295.

Kuspa, A. 2006. Restriction Enzyme-Mediated Integration (REMI) Mutagenesis. Pp. 201–210 *in* *Dictyostelium discoideum* Protocols. Humana Press, New Jersey.

Kuzdzal-Fick, J. J., S. A. Fox, J. E. Strassmann, and D. C. Queller. 2011. High relatedness is necessary and sufficient to maintain multicellularity in *Dictyostelium*. Science 334:1548–1551.

Kuzdzal-Fick, J. J., D. C. Queller, and J. E. Strassmann. 2010. An invitation to die: initiators of sociality in a social amoeba become selfish spores. Biol. Lett. 6:800–802.

Lennon, J. T., and S. E. Jones. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat. Rev. Microbiol. 9:119–130.

Levin, S. R., D. A. Brock, D. C. Queller, and J. E. Strassmann. 2015. Concurrent coevolution of intra-organismal cheaters and resisters. J. Evol. Biol. 28:756–765.

Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843–2851.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, S. I., and M. D. Purugganan. 2011. The cooperative amoeba: *Dictyostelium* as a model for social evolution. Trends Genet. 27:48–54.

Lieberman, T. D., J.-B. Michel, M. Aingaran, G. Potter-Bynoe, D. Roux, M. R. Davis, D. Skurnik, N. Leiby, J. J. LiPuma, J. B. Goldberg, A. J. McAdam, G. P. Priebe, and R. Kishony. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43:1275–1280.

Lindström, E. S., and Ö. Östman. 2011. The importance of dispersal for bacterial community composition and functioning. PLoS One 6:1–7.

Long, A., G. Liti, A. Luptak, and O. Tenaillon. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. 16:567–582.

Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.

Maloof, J. E., and D. W. Inouye. 2000. Are nectar robbers cheaters of mutualists? Ecology 81:2651–2661.

Manhes, P., and G. J. Velicer. 2011. Experimental evolution of selfish policing in social bacteria. Proc. Natl. Acad. Sci. 108:8357–8362.

Martínez-García, R., and C. E. Tarnita. 2016. Lack of ecological and life history context can create the illusion of social interactions in *Dictyostelium discoideum*. PLoS Comput. Biol. 12:1–21.

McConnell, R., S. Middlemist, C. Scala, J. E. Strassmann, and D. C. Queller. 2007. An unusually

low microsatellite mutation rate in *Dictyostelium discoideum*, an organism with unusually abundant microsatellites. Genetics 177:1499–1507.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Mehdiabadi, N. J., C. N. Jack, T. T. Farnham, T. G. Platt, S. E. Kalla, G. Shaulsky, D. C. Queller, and J. E. Strassmann. 2006. Kin preference in a social microbe. Nature 442:881–882.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11:5–18.

Meynert, A. M., M. Ansari, D. R. FitzPatrick, and M. S. Taylor. 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 15:1–11.

Miller, M. B., and B. L. Bassler. 2001. Quorum sensing in bacteria. Annu. Rev. Microbiol. 55:165–199.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. 70:3321–3323.

Nei, M., and R. K. Chesser. 1983. Estimation of fixation indices and gene diversities. Ann. Hum. Genet. 47:253–259.

Nidelet, T., and O. Kaltz. 2007. Direct and correlated responses to selection in a host-parasite system: Testing for the emergence of genotype specificity. Evolution 61:1803–1811.

Noh, S., K. S. Geist, X. Tian, J. E. Strassmann, and D. C. Queller. 2018. Genetic signatures of microbial altruism and cheating in social amoebas in the wild. Proc. Natl. Acad. Sci. 115:3096–3101.

Nowak, M. A., and K. Sigmund. 2005. Evolution of indirect reciprocity. Nature 437:1291–1298.

Odong, T. L., J. van Heerwaarden, T. J. L. van Hintum, F. A. van Eeuwijk, and J. Jansen. 2013. Improving hierarchical clustering of genotypic data via principal component analysis. Crop Sci. 53:1546–1554.

Ogura, Y., T. Ooka, A. Iguchi, H. Toh, M. Asadulghani, K. Oshima, T. Kodama, H. Abe, K. Nakayama, K. Kurokawa, T. Tobe, M. Hattori, and T. Hayashi. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. 106:17939–17944.

Ostrowski, E. A., M. Katoh, G. Shaulsky, D. C. Queller, and J. E. Strassmann. 2008. Kin discrimination increases with genetic distance in a social amoeba. PLoS Biol. 6:1–7.

Ostrowski, E. A., Y. Shen, X. Tian, R. Sucgang, H. Jiang, J. Qu, M. Katoh-Kurasawa, D. A. Brock, C. Dinh, F. Lara-Garduno, Others, S. L. Lee, C. L. Kovar, H. H. Dinh, V. Korchina, L. R. Jackson, S. Patil, Y. Han, L. Chaboub, G. Shaulsky, D. M. Muzny, K. C. Worley, R. A. Gibbs, S. Richards, A. Kuspa, J. E. Strassmann, and D. C. Queller. 2015. Genomic signatures of cooperation and conflict in the social amoeba. Curr. Biol. 25:1661–1665.

Özkaya, Ö., R. Balbontín, I. Gordo, and K. B. Xavier. 2018. Cheating on cheaters stabilizes cooperation in *Pseudomonas aeruginosa*. Curr. Biol. 28:2070–2080.

Özkaya, Ö., K. B. Xavier, F. Dionisio, and R. Balbontín. 2017. Maintenance of microbial cooperation mediated by public goods in single- and multiple-trait scenarios. J. Bacteriol. 199:1–14.

Parikh, A., E. R. Miranda, M. Katoh-Kurasawa, D. Fuller, G. Rot, L. Zagar, T. Curk, R. Sucgang, R. Chen, B. Zupan, W. F. Loomis, A. Kuspa, and G. Shaulsky. 2010. Conserved developmental transcriptomes in evolutionarily divergent species. Genome Biol. 11:1–12.

Parkinson, K., N. J. Buttery, J. B. Wolf, and C. R. L. Thompson. 2011. A simple mechanism for complex social behavior. PLoS Biol. 9:1–15.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pruitt, J. N., and S. E. Riechert. 2009. Frequency-dependent success of cheaters during foraging bouts might limit their spread within colonies of a socially polymorphic spider. Evolution 63:2966–2973.

Putman, A. I., and I. Carbone. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol. Evol. 4:4399–4428.

R Core Team. 2018. R: A language and environment for statistical computing. R foundation for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rainey, P. B., and K. Rainey. 2003. Evolution of cooperation and conflict in experimental bacterial populations. Nature 425:72–74.

Rainey, P. B., and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394:69–72.

Rankin, D. J., K. Bargum, and H. Kokko. 2007. The tragedy of the commons in evolutionary biology. Trends Ecol. Evol. 22:643–651.

Raper, K. B. 1984. The Dictyostelids. First Edit. Princeton Univ Pr, Princeton, N.J.

Richardson, S. C. 2004. Are nectar-robbers mutualists or antagonists? Oecologia 139:246–254.

Riolo, R. L., M. D. Cohen, and R. Axelrod. 2001. Evolution of cooperation without reciprocity. Nature 414:441–443.

Rojas-nossa, S. V, J. M. Sánchez, and L. Navarro. 2016. Nectar robbing : a common phenomenon mainly determined by accessibility constraints, nectar volume, and density of energy rewards. Oikos 125:1044–1055.

Roulston, T. H., G. Buczkowski, and J. Silverman. 2003. Nestmate discrimination in ants: effect of bioassay on aggressive behavior. Insectes Soc. 50:151–159.

Rousset, F., and D. Roze. 2007. Constraints on the origin and maintenance of genetic kin recognition. Evolution 61:2320–2330.

Roux, C., M. Pauwels, M.-V. Ruggiero, D. Charlesworth, V. Castric, and X. Vekemans. 2013. Recent and ancient signature of balancing selection around the s-locus in *Arabidopsis halleri* and *A. lyrata*. Mol. Biol. Evol. 30:435–447.

Santorelli, L. A., C. R. L. Thompson, E. Villegas, J. Svetz, C. Dinh, A. Parikh, R. Sucgang, A. Kuspa, J. E. Strassmann, D. C. Queller, and G. Shaulsky. 2008. Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. Nature 451:1107–1110.

Santorelli, L. a, A. Kuspa, G. Shaulsky, D. C. Queller, and J. E. Strassmann. 2013. A new social gene in *Dictyostelium discoideum*, chtB. BMC Evol. Biol. 13:4.

Schlötterer, C., R. Kofler, E. Versace, R. Tobler, and S. U. Franssen. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity (Edinb). 114:431–440.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18:233–234.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.

Smith, J. M., and E. Szathmary. 1997. The major transitions in evolution. Oxford University Press, Oxford.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. Nature 400:667–671.

Strassmann, J. E., O. M. Gilbert, and D. C. Queller. 2011. Kin discrimination and cooperation in microbes. Annu. Rev. Microbiol. 65:349–367.

Strassmann, J. E., and D. C. Queller. 2011. Evolution of cooperation and control of cheating in a social microbe. Proc. Natl. Acad. Sci. 108:10855–10862.

Strassmann, J. E., Y. Zhu, and D. C. Queller. 2000. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. Nature 408:965–967.

Swanson, A. R., E. M. Vadell, J. C. Cavender, S. Journal, N. Jan, A. R. Swanson, E. M. Vadellt, and J. C. Cavendert. 1999. Global distribution of forest soil dictyostelids. J. Biogeogr. 26:133–148.

Tellier, A., and J. K. Brown. 2011. Spatial heterogeneity, frequency-dependent selection and polymorphism in host-parasite interactions. BMC Evol. Biol. 11:319.

Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. 2012. The molecular diversity of adaptive convergence. Science 335:457–461.

Travisano, M., and G. J. Velicer. 2004. Strategies of microbial cheater control. Trends Microbiol. 12:72–78.

Tresse, E., A. Kosta, C. Giusti, M.-F. Luciani, and P. Golstein. 2008. A UDP-glucose derivative is required for vacuolar autophagic cell death. Autophagy 4:680–691.

Turner, T. L., and P. M. Miller. 2012. Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. Genetics 191:633–642.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40:1–12.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine,

T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinforma. 11:11.10.1–11.10.33.

Van Dyken, J. D., and M. J. Wade. 2012. Detecting the molecular signature of social conflict: Theory and a test with bacterial quorum sensing genes. Am. Nat. 179:436–450.

van Elsas, J. D., M. Chiurazzi, C. A. Mallon, D. Elhottova, V. Kristufek, and J. F. Salles. 2012. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc. Natl. Acad. Sci. 109:1159–1164.

Villa Martín, P., M. A. Muñoz, and S. Pigolotti. 2019. Bet-hedging strategies in expanding populations. PLOS Comput. Biol. 15:1–17.

Vos, M., and G. J. Velicer. 2006. Genetic population structure of the soil bacterium *Myxococcus xanthus* at the centimeter scale. Appl. Environ. Microbiol. 72:3615–3625.

Vos, M., and G. J. Velicer. 2009. Social conflict in centimeter-and global-scale populations of the bacterium *Myxococcus xanthus*. Curr. Biol. 19:1763–1767.

Votaw, H. R., and E. A. Ostrowski. 2017. Stalk size and altruism investment within and among populations of the social amoeba. J. Evol. Biol. 30:2017–2030.

Wang, M., A. L. Schaefer, A. A. Dandekar, and E. P. Greenberg. 2015. Quorum sensing and policing of *Pseudomonas aeruginosa* social cheaters. Proc. Natl. Acad. Sci. 112:2187–2191.

Waters, C. M., and B. L. Bassler. 2005. Quorum sensing: cell-to-cell communication in bacteria. Annu. Rev. Cell Dev. Biol. 21:319–346.

Weedall, G. D., and D. J. Conway. 2010. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. Trends Parasitol. 26:363–369.

West, S. A., A. S. Griffin, and A. Gardner. 2007a. Evolutionary explanations for cooperation. Curr. Biol. 17:R661–R672.

West, S. A., A. S. Griffin, A. Gardner, and S. P. Diggle. 2006. Social evolution theory for microorganisms. 4:597–608.

West, S. a, S. P. Diggle, A. Buckling, A. Gardner, and A. S. Griffin. 2007b. The social lives of microbes. Annu. Rev. Ecol. Evol. Syst. 38:53–77.

Whitlock, M. C. 1996. The red queen beats the jack-of-all-trades: The limitations on the evolution of phenotypic plasticity and niche breadth. Am. Nat. 148:S65–S77.

Winter, D. J. 2012. *mmod* : an R library for the calculation of population differentiation statistics. Mol. Ecol. Resour. 12:1158–1160.

Wolf, J. B., J. A. Howie, K. Parkinson, N. Gruenheit, D. Melo, D. Rozen, and C. R. L. Thompson. 2015. Fitness trade-offs result in the illusion of social success. Curr. Biol. 25:1086–1090.

Wood, T. E., J. M. Burke, and L. H. Rieseberg. 2005. Parallel genotypic adaptation: when evolution repeats itself. Genetica 123:157–170.

Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski. 2006. Tests of

parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc. Natl. Acad. Sci. 103:9107–9112.

Yanni, D., P. Márquez-Zacarías, P. J. Yunker, and W. C. Ratcliff. 2019. Drivers of spatial structure in social microbial communities. Curr. Biol. 29:R545–R550.

Yoder, J. B. 2016. Understanding the coevolutionary dynamics of mutualism with population genomics. Am. J. Bot. 103:1742–1752.

Zee, P. C., and G. J. Velicer. 2017. Parallel emergence of negative epistasis across experimental lineages. Evolution 71:1088–1095.

Zhang, Q. G., A. Buckling, R. J. Ellis, and H. C. J. Godfray. 2009. Coevolution between cooperators and cheats in a microbial system. Evolution 63:2248–2256.

Zhang, X.-X., and P. B. Rainey. 2013. Exploring the sociobiology of pyoverdin-producing *Pseudomonas*. Evolution 67:3161–3174.

Zufall, R. A., K. L. Dimond, and F. P. Doerder. 2013. Restricted distribution and limited gene flow in the model ciliate *Tetrahymena thermophila*. Mol. Ecol. 22:1081–1091.

Zwart, G., W. D. Hiorns, B. A. Methé, M. P. van Agterveld, R. Huismans, S. C. Nold, J. P. Zehr, and H. J. Laanbroek. 1998. Nearly identical 16S rRNA sequences recovered from lakes in north america and europe indicate the existence of clades of globally distributed freshwater bacteria. Syst. Appl. Microbiol. 21:546–556.