Title
Long-Distance Non-Self Recognition and Control of Intercellular Communication in Neurospora crassa

Author
Rosenfield, Gabriel Rappaport

Publication Date
2019

Peer reviewed|Thesis/dissertation

Long-Distance Non-Self Recognition and Control of Intercellular Communication in *Neurospora crassa*

by

Gabriel R Rosenfield

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor N. Louise Glass, Chair
Professor Arash Komeili
Professor Peter Quail

Summer 2019

Abstract

Long-Distance Non-Self Recognition and Control of Intercellular Communication in *Neurospora crassa*

by

Gabriel R Rosenfield

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor N. Louise Glass, Chair

All multicellular organisms face a challenge: resources must be distributed throughout their bodies to ensure continued growth and coordinate cellular activities, but once resources are socialized, how can cells be coerced into working for the organism as a whole, rather than individually reproducing at the expense of the body they inhabit? Solutions to this freeloader problem usually rely on aligning the reproductive interests of all cells in a body by ensuring all are genetically identical. However, such solutions pose problems for organisms with indeterminate colonial growth habits, a lifestyle characterized by vegetative expansion, fragmentation, and reintegration of separated fragments. The filamentous fungus *Neurospora crassa* exemplifies this lifestyle. Its body is a cross-linked syncytial network produced by the fusion of many cells. Individual nuclei flow throughout the colony, making all the products of each nucleus potentially available to freeloaders. To keep the reproductive interests of all nuclei aligned, *N. crassa* must restrict fusion to genetically cells. The fungus prevents non-self fusion using an array of non-self recognition (NSR) systems.

NSR systems are encoded by polymorphic kind recognition loci. Each NSR locus enables cells to behave differently toward other individuals with the same haplotype than they do toward individuals with distinct haplotypes. At least two haplotypes at an NSR locus must exist in a population for the system to function. If many NSR loci are spread throughout an organisms genome, the net output of all NSR systems transitions to kin recognition; this allows an organism to identify genetically identical and distinct individuals with high fidelity. *N. crassa* encodes 15 confirmed NSR loci, with dozens more suspected based on sequencing and phenotypic investigations. Different NSR systems function at various stages in *N. crassa*'s life cycle, and during distinct phases of sociality. The first phase of sociality is called communication, and occurs when two cells detect each other and reorient their growth to intersect. The Determinant Of Communication (DOC) NSR system restricts communication between non-identical cells before they touch.

We published our discovery of the DOC system in 2016. At the time, we knew haplotypes at the *doc* locus were necessary and sufficient to specify which cells will be recognized as self. We also reported that DOC-2 localized to the cellular periphery and DOC-1 oscillated between the cell bodies and growing tips of cells along with the MAK-2 complex during intercellular communication. We found the DOC system was not required for self-communication, implying it suppresses communication until a compatible communication group (CG) signal is received. Finally, we identified five distinct communication group haplotypes (CGHs) at the *doc* locus in an *N. crassa* population from Louisiana; *doc* alleles from different CGHs were less than 50% identical at the nucleotide level, and exhibited trans-species polymorphism among *N. crassa* and other members of its genus.

Little else was certain regarding the DOC system. With neither characterized homologues nor identifiable functional domains, the DOC proteins remained particularly opaque. The primary goal of my doctoral work was to advance our understanding of the DOC system. I attempted to improve our model of DOC-mediated long-distance NSR by investigating how different *doc* genes and alleles interact *in vivo*, and evaluating the roles played by various regions of the DOC proteins in regulating communication. In pursuit of these aims, I spent considerable time and effort developing improved methods for quantifying intercellular communication.

After developing and validating a flow cytometry-based communication assay and analysis pipeline, I used this assay to evaluate the communication behavior of *N. crassa* strains expressing incompatible alleles of the *doc* genes, multiple incompatible DOC systems, truncated and otherwise manipulated alleles of the *doc* genes, and CGH1/CGH3 chimeric alleles of *doc-1*. The results of these experiments support the following model for DOC-mediated long-distance NSR:

As cells begin to communicate, general and CG-specific signals must be exchanged. CG signals are validated through interactions between the middle region of DOC-1 and DOC-2: matched CGH-variants of the DOC proteins are generally required for proper CG specificity. If the DOC system does not receive a compatible CG signal, DOC-1 and DOC-2 both prevent reinforcement of MAK-2 complex oscillations and suppress communication. This suppression involves the actions of the middle and C-terminal regions of DOC-1, and does not require the N-terminal region of DOC-2. If a compatible CG signal is detected, the DOC system allows MAK-2 complex oscillations and chemotropic growth to continue. DOC-1's N-terminal region is required to properly derepress communication, and is probably involved in validating CG signals.

Table of Contents

Acknowledgements

First, I want to acknowledge important academic contributions to my work. Thank you to Professor Marcus Roper from UCLA. His help with the first version of my flow cytometry data analysis software was essential to all my later work. Thank you to the National Science Foundation and UC Berkeley for funding my work and paying my salary. Next, I'd like to acknowledge several people's personal contributions to my work and life.

Like most graduate students, I did not find the PhD process easy. The electronic format of this dissertation belies the blood, sweat, and tears that would surely stain its pages if it could fully capture the graduate school experience. I was fortunate to have academic and scientific support from a number of wonderful colleagues without whom I could not have managed the degree of success I've achieved. Thank you to all the members of the Glass lab whose time overlapped with mine and from whose work I benefited, and a special thanks to everyone who took part in our lively lunchtime discussions. Not every lab is so social and filled with personable eccentrics, and I feel lucky to have spent the last six years with all of you.

I'd like to specifically thank a few lab mates whose help and support was truly indispensable. First, thank you to Jens Heller, PhD. Jens' work spawned my doctoral project, and, more importantly, Jens was my primary mentor throughout grad school. He shared his protocols and wisdom without reservation, and I still wonder how he accomplishes so much so quickly. Next, thank you to Darae Jun, PhD, for your continuous emotional support, for housing me in my time of need, and for traveling with me when I was a heart-broken, crippled man. I couldn't have asked for a better lab-sister. Thank you to Monika Fischer, PhD, for your incredible attitude and energy. You were a model graduate student and I tried to emulate you with limited success. Never stop taking yoga breaks! Finally, thank you to Professor N. Louise Glass for running the show. You've always treated me kindly and with respect, which is more than I can say for many professors. I especially appreciate you sending me to Bochum, Germany to collaborate with the Kück lab. None of us would be here without you!

Surprisingly, I've also had some semblance of a life outside of lab. Since starting graduate school, I've moved three times, taken up SCUBA diving, broken up with a long-term partner, temporarily lost the function of my right leg in a skiing accident, and gone on a number of wonderful backpacking trips with friends. I'd like to thank a few of my friends for joining, commiserating with, and tolerating me these last six years. Thank you to Tyler Helmann, almost PhD (like me!), for always letting me complain to you on our walks. We should have walked more often than we did! Thank you to Magdalena Stoeva, PhD, and Geoffrey Maines, M.A.Sc., for housing me, feeding me, and generally spending so much time with me. I can only assume I must be pretty cool to have such amazing friends.

Next, I must thank my parents, Laurence Rosenfield, MD, and Janet Rappaport, M.Ed. MSLP. They raised, loved, and supported me for more than the last six years. They taught me to enjoy learning and fostered my interest in all things science. Most importantly, they never discouraged me when I asked "why?" Without them, I wouldn't be here, literally!

To anyone I've forgotten, thank you and I'm sorry! It takes a village to create a PhD. Finally, thank you to the staff in the Department of Plant and Microbial Biology. None of this would be possible without your organization and guidance.

Chapter 1. Non-self recognition and intercellular communication in Neurospora crassa

1.1. Introduction to non-self recognition

Many organisms recognize other individuals as genetically identical (self) or genetically different (non-self) and modulate their interactions depending upon their identities. Although there is some ambiguity regarding the definition of non-self recognition, I will use a relatively limited definition in this dissertation:

- Non-self recognition (NSR), sometimes called allorecognition, is an intercellular communication process during which information about the genetic identity of an individual is received by and alters the behavior of another, conspecific individual. Multiple interaction modes between individuals using NSR must be possible (e.g. cooperation versus antagonism), and a mode must be chosen based on the genetic identities of the interacting individuals.

Note that this definition excludes the recognition of pathogens by organisms' immune systems because pathogens are not conspecific with their hosts. The secretion of toxins in response to interspecies competition is likewise excluded. Social organisms throughout the tree of life have independently developed NSR[1–7]. From fruiting body formation[8] and social motility[9] in bacteria to tissue-graft rejection[6] in humans, NSR evolves when many cells cooperate to produce social goods.

In the following subchapters, I present an overview of the evolution and genetics of NSR, a brief introduction to NSR in *Neurospora crassa*, and a summary of the publication that instigated my PhD project. At the end of this chapter, I outline the open questions and predictions that guided the work presented in the remainder of this dissertation.

1.2. Evolution of non-self recognition

Why would an organism evolve NSR? If a single-celled organism remains solitary throughout its life cycle, or only associates with a relatively small number of clonal sisters, NSR provides no obvious fitness advantage[10]. However, at least three potential drivers for NSR evolution in social and multicellular organisms have been identified: freeloaders, inbreeding, and disease transmission.

If multicellular or social phases of an organism's life cycle have the potential for genetic heterogeneity, freeloader genotypes gain a fitness advantage[11]. Freeloader (or cheater) genotypes are named by analogy to the "free-rider problem" and "tragedy of the commons" situations in economics[12]. Freeloaders participate in an organism's social phase, but receive social goods without contributing[13]. This behavior results in an increase in the freeloader's relative fitness at the expense of the social group. Although freeloading confers a fitness advantage relative to more altruistic genotypes, social parasitism by freeloaders may result in lower absolute fitness for all genotypes including the freeloader[14]. Selection for freeloaders is considered one of the most important impediments to the evolution of multicellularity, and a primary driver of NSR evolution[13]. NSR reduces the freeloader

problem by permitting organisms to limit social behaviors to genetically similar individuals. If social goods can only be shared among highly similar individuals, selection for cheating is reduced[15].

Sex generates genetic diversity by shuffling parental genotypes to generate chimeric offspring genomes. This enhances population fitness by providing a wider range of characteristics on which selection can act[16,17]. However, sex between closely related parents yields little diversity and largely abolishes the evolutionary advantages of sex. Worse yet, multiple generations of inbreeding can result in the accumulation and unmasking of deleterious mutations, decreasing fitness[16]. These mechanisms select for an outcrossing preference in sexual organisms, which requires NSR. NSR in a sexual context has been well studied in flowering plants. Many angiosperms place male and female gametes from the same parent in close proximity; this arrangement makes self-fertilization likely and puts great evolutionary pressure on such plants to avoid habitual inbreeding. A common solution among these plants has been the development of self-incompatibility: an NSR system to identify and reject pollen from closely related parents[4].

The Chestnut Blight fungus, *Cyphonectria parasitica*, provides examples of two potential drivers of NSR. First, *C. parasitica* colonies are syncitial, a lifestyle that selects for freeloaders[13]. Second, *C. parasitica* can be infected with viruses from the family Hypoviridae that reduce its fitness[18]. Mycoviruses, including the Hypoviridae, appear to lack external vectors and are transmitted by cytoplasmic mixing with infected cells during intercellular fusion[19]. Wild isolates of *C. parasitica* exhibit vegetative incompatibility (VIC), a form of NSR that prevents non-sexual mergers between genetically different strains. Multiple *vic* genes have been studied in *C. parasitica*, only some of which reduce viral transmission between incompatible strains[20]. Disease transmission pressures may explain why some organisms have developed multiple NSR checkpoints that operate at various levels of intercellular intimacy[20,21].

1.3. Characteristics of non-self recognition genes

NSR systems are highly diverse and often share no homology[1–7]. Nonetheless, NSR genes share some general characteristics. Typically, genes encoding NSR systems exhibit evidence of balancing selection[1]: long term maintenance of multiple alleles at similar frequencies in well-mixed populations. This occurs when rare alleles offer a fitness advantage. Evidence of balancing selection includes high allelic diversity, signatures of positive selection, and trans-species polymorphism[1,21]: a phenomenon in which some alleles from different species are more closely related to each other than they are to other intra-species alleles. Trans-species polymorphism results in an NSR genealogy and a species genealogy that do not correspond (Fig. 1.3), and indicates that multiple NSR alleles have been preserved over evolutionary time[1]. The regularity with which NSR genes are under balancing selection suggests that NSR enhances fitness under many conditions, and provides a population genomics approach for finding candidate NSR genes[1,22].

Phylogenetically diverse model organisms have been developed to investigate NSR across the tree of life: model prokaryotes include *Myxococcus xanthus*[8], *Burkholderia thailandensis*[3,23], and *Proteus mirabilis*[9,24,25], while model eukaryotes include the colonial animals *Hydractinia symbiolongicarpus*[26] and *Botryllus schlosseri*[2,27,28], the flowering plants *Petunia sp.*[29,30], *Brassica campestris*[31,32], and *Papaver rhoeas*[4,33], the slime mold

Dictyostelium discoideum[15,34], and the fungi *Cryphonectria parasitica*[18,20,21], *Podospora anserina*[35], and *Neurospora crassa*[7,36–40]. NSR systems composed of multiple genes are often tightly linked into functional modules. This reduces the probability of recombination between a module's components, encouraging coevolution of the components and divergence of different alleles[21,41]. While the fitness advantages, evolutionary origins, and mechanisms of NSR modules vary, each module acts as a kind selection or green-beard locus, allowing organisms to react differently to individuals expressing differing alleles of that module[10,42–45].

As more kind selection loci are distributed more evenly throughout an organism's genome, the combined output of all green-beard loci transitions to kin selection[46]. In other words, many NSR modules are required for reliable self-identification among genetically similar individuals. The *N. crassa* genome encodes the most confirmed, unlinked NSR modules of any ascomycete[20,40,47], suggesting that selection for NSR fidelity on this fungus is unusually high. Furthermore, most documented NSR modules require cell-to-cell contact to exchange identifying information[10]; *N. crassa* is the only eukaryote known to use an NSR module that operates before intercellular contact during somatic growth*[42]. These qualities, along with its high growth rate in culture[48], established scientific community[49], and plethora of research tools[50,51] make *N. crassa* an ideal model organism for NSR research.

* The familiar ABO blood type NSR system in humans (and other tissue type NSR systems in vertebrates) arguably operates prior to intercellular contact. The initial NSR occurs when antibodies in the blood bind to foreign carbohydrate antigens on the surface of non-self cells. Hostile NSR responses are triggered once these antibody-marked cells are detected by immune cells[6,52,53]. However, physical contact between marked cells and immune cells is required for an NSR response.

1.4. Neurospora crassa

While I assume any reader of this dissertation has a working knowledge of microbiological lingo, I recognize mycologists are especially fond of arcane jargon. If you're unfamiliar with mycology, you may refer to the *Illustrated Dictionary of Mycology* by Ulloa, Hanlin, and Acosta[54]. Unless other citations are given, information in the remainder of this subchapter is summarized from Rowland Davis's *Neurospora: Contributions of a Model Organism*[48], occasionally supplemented with my own observations.

Neurospora crassa is a filamentous ascomycete in the Sordariales. The *Neurospora* genus contains two groups: a better-studied fire-associated clade and a less-studied dung-associated clade[55,56]. *N. crassa* is adapted to fires, and its sexual spores (ascospores) require heat or pyrolytic compounds to germinate. The fire-associated *Neurospora* species are only found in the wild after forest fires, where they are among the first organisms fruiting on dead vegetation[48,55,56]. Little else is certain regarding the ecology of *Neurospora*, although sequencing evidence suggests members of the fire-associated clade may affiliate with lichens between forest fires[55,56].

N. crassa reproduces both sexually and asexually (Fig. 1.4). In response to nutritional cues, a colony will produce female structures known as protoperithecia, each with a specialized female hypha called a trichogyne. Although colonies are hermaphroditic, they are self-infertile and have two mating types referred to as "A" and "a". The trichogyne

follows mating pheromones released by a conidium of the opposite mating type until the two cells fuse and a male nucleus is transferred down the trichogyne into the protoperithecium. There, karyogamy occurs in a hook shaped structure known as a crozier. This is the only diploid stage in the life cycle, and the zygotic nucleus immediately undergoes meiosis without cell divisions. Each crozier generates one meiotic compartment and a new crozier as the protoperithecium matures into a perithecium. The meiotic compartments grow into sac-like cells called asci, from which the phylum Ascomycota takes its name. In *N. crassa*, meiosis is followed by one round of mitosis to produce eight ascospores per ascus. When mature, one ascus at a time is pressed to the apical opening of the perithecium where it forcibly ejects the ascospores it contains, firing them several centimeters into the air[48]. *N. crassa* ascospores are highly resistant to environmental stress, and can remain viable for years under laboratory conditions.

Ascospores must be heated to at least 60^{o}C or incubated with pyrolytic compounds to break their dormancy. Once an ascospore germinates and grows into a sufficiently large colony, it begins producing asexual spores called conidia. *N. crassa* produces three types of conidia[57], none of which have confirmed biological roles except as male gametes. Macroconidia are borne by conidiophores on aerial hyphae that grow up from the colony in response to light and nutritional cues. These abundant multinucleate spores give mature *N. crassa* colonies their distinctive orange hue. Microconidia are produced on tiny conidiophores emanating from otherwise typical hyphae. Unlike macroconidia, they contain only one to a few nuclei, making them useful for purifying mutants from heterokaryons (discussed below). As the macroconidia they produce mature and septate, aerial hyphae also thicken their septa: eventually the cells separate to form arthroconidia[57]. These are the least studied type of spore produced by *N. crassa*, and not many authors mention their existence. However, the distance between septa in aerial hyphae increases with distance from the conidiophore, suggesting arthroconidia may be produced as a byproduct of some unidentified diffusible conidiation factor[57]. Because macroconidia are the only spores that played a major role in the work described here, I will use the term conidia exclusively in reference to macroconidia for the remainder of this dissertation. Conidia can be passaged to generate a clone of the parent, or can serve as male gametes in the sexual cycle.

From the perspective of NSR, *N. crassa*'s morphology is most interesting. Like most filamentous fungi, its' thallus consists of a cross-linked network (mycelium) of tubular cells (hyphae) sharing a continuous cytoplasm. However, the perforated septa between cell compartments in *N. crassa* allow the entire cell contents, including nuclei, to flow around the network[7,48,58]. Although mature ascospores are also multinucleate, developing ascospores are the only cells produced by *N. crassa* that reliably pass through a uninucleate stage. All other stages of its' life cycle are syncytial[48]. This makes all the products of each nucleus potential social goods, strongly selecting for freeloaders[13]. *N. crassa* generates its interconnected mycelial network via fusion between hyphae, a process called anastomosis[58].

Hyphal anastomosis and the mycelium it generates are thought to enhance fitness by increasing colony growth rates as germlings and small mycelia fuse, and by distributing resources throughout the colony. Cell fusion appears especially beneficial in heterogeneous or resource-poor environments[59]. *N. crassa* regulates cell fusion using nutritional cues and NSR[58,60]. While mycelia containing multiple genotypes (heterokaryons) can be generated in

the lab, colonies of wild isolates will not merge and remain homokaryotic (containing a single genotype). The inability of genetically distinct isolates to somatically merge, termed heterokaryon or vegetative incompatibility (VIC), is a common NSR phenomenon in anastomosing fungi[40]. Experimental and simulated evolution studies indicate VIC evolution and maintenance is probably driven by the need to counteract selection for freeloaders in *N. crassa*[11,13].

In the lab, *N. crassa* is easy to grow on defined media, achieving a maximum linear growth rate of twelve cm per day. Its' sexual cycle is understood and crosses on defined media are routine[48]. The genome sequence of the laboratory strain (FGSC2489) is published[61], and most single-gene deletion strains as well as a number of "classical" mutants are available from the Fungal Genetics Stock Center[51]. Genetic transformation via electroporation of conidia is straightforward, although backcrossing is required to generate homokaryotic mutants because the conidia are multinucleate[48]. *N. crassa*'s characteristics and available resources make it well suited as a model for investigating NSR.

1.5. Non-self recognition checkpoints in the life cycle of *Neurospora crassa*

For ease of preparation, quantification, and to synchronize cell growth, many NSR experiments are performed on newly germinated conidia termed germlings. Genetically identical *N. crassa* germlings within ~15 μm of each other communicate, grow toward one another, and fuse, then repeat this process until a small mycelium has formed[42]. Germlings pass through three known NSR checkpoints during this initial growth phase, with two more checkpoints at later developmental stages[62,63]. I will refer to the first three NSR checkpoints as pre-contact, post-contact, and post-fusion, and the later checkpoints as hyphal-fusion and fertilization.

The first NSR checkpoint occurs at the initiation of sociality: pre-contact germling communication. Immediately after germination, cells use unidentified chemical signatures to signal their presence to others. If communicating germlings share compatible alleles at the determinant of communication (*doc*) locus, they will exhibit homing growth and eventually contact each other. If they have incompatible *doc* alleles, germlings will abort their communication and ignore each other (further details in 6.1. Long distance non-self recognition in *Neurospora crassa*). Unlike the post-contact, post-fusion, and hyphal-fusion NSR checkpoints, failure to pass the pre-contact checkpoint does not result in VIC. If cells expressing incompatible *doc* alleles inadvertently make contact, they may fuse and appear to form stable heterokaryons[42].

After cell contact, germlings test the second NSR checkpoint: post-contact cell wall breakdown. Like most fungal cells, *N. crassa* germlings are encased in a cell wall that must be degraded before membrane merger can occur. Germlings with incompatible alleles at the *cell wall remodeling* (*cwr*) locus cannot degrade their cell walls at the contact point, and fail to fuse after communicating. This NSR mechanism relies on interactions between incompatible variants of CWR-1, a putative polysaccharide monooxygenase, and CWR-2, a multipass transmembrane protein. Recognition appears to require catalytic activity of CWR-1. Manipulation of the *cwr* genes indicates another independent NSR locus must also operate at this checkpoint, but this second locus has yet to be confirmed. Although failure

to pass the post-contact NSR checkpoint does not result in cell death, *cwr* incompatibility does prevent heterokaryon formation like other VIC modules[63].

Once germlings break down their cell walls at the contact point, they merge their membranes and mix cytoplasm. Cell fusion triggers the third NSR checkpoint: post-fusion germling regulated death (GRD). If germlings have incompatible alleles at GRD loci, they rapidly lyse after cytoplasmic mixing. Two unlinked GRD modules have been identified as of this writing: *sec-9*/*plp-1*, and *rcd-1*. Incompatible variants of SEC-9, an essential SNARE, and patatin-like phospholipase-1 (PLP-1), a NOD-like receptor, interact in fused cells to activate and oligomerize PLP-1[64]. This triggers programed cell death either directly by altering phospholipids in the cell membrane or via an uncharacterized signaling process likely involving secondary messengers generated by PLP-1 activity. Alleles of *regulator of cell death-1* (*rcd-1*) encode inactive pore-forming toxins similar to mammalian gasdermins[65]. When incompatible variants of RCD-1 interact, at least one variant activates, oligomerizes, and inserts into the cell membrane causing lysis.

The fourth hyphal-fusion NSR checkpoint is activated when colonies mature. In *N. crassa*, canonical heterokaryon incompatibility (*het*) loci act as developmentally regulated VIC modules with their NSR functions suppressed in germlings[63,66,67]. As a result, germlings differing only in *het* alleles will successfully fuse, but will sector into separate colonies once the heterokaryotic mycelium develops. This contrasts with GRD loci, which are active during all vegetative phases of *N. crassa*'s life cycle[64,65]. After maturation, interacting colonies must encode compatible alleles at all *het* loci in order to form stable heterokaryons[40,66,68]. Anastomosis between hyphae from incompatible mycelia triggers septal plugging and programmed cell death of fused cells. Interestingly, most confirmed *het* genes contain a conserved HET domain, which has enabled a population genomics survey for candidate *het* genes[69]. Results suggest there may be over 70 *het* loci spread throughout *N. crassa*'s pan-genome, with at least 34 showing signs of balancing selection[47]. Assuming only two incompatible alleles at each locus, this gives 2^{34} (just over 17 billion) unique *het* allele combinations. This simplistic estimate relies on sequencing data without functional evidence, and ignores that some *het* genes have more than two alleles and that many are linked. However, Goncalves et al., 2019, considered only haplotypes of the 12 experimentally confirmed *het* loci, the *doc* locus, the *cwr* locus, and the *sec-9*/*plp-1* locus, and still calculated over a billion possible incompatible genotypes[63]. Calculations like these illustrate that NSR modules should stop any wild isolates from forming stable heterokaryons at maturity.

A final fertilization NSR checkpoint is triggered during sexual development. *N. crassa* uses a self-incompatibility system encoded by genes in the mating-type (*mat*) region to prevent selfing. Unlike most NSR modules, *N. crassa*'s two *mat* haplotypes are composed of evolutionarily unrelated genes, termed "idiomorphs"[70]. Alternative *mat* idiomorphs correspond to mating types designated "A" and "a", and strains must express differing *mat* idiomorphs to mate successfully. During vegetative growth, *mat* is also part of a VIC system: products from both *mat* idiomorphs must interact with each other and the product of an unlinked gene called *tol* to trigger a VIC reaction[37,71,72]. Expression of *tol* is repressed during sexual development, switching the *mat* locus from VIC to self-incompatibility mode. This is an important note, because VIC modules must be inactive in sexual tissues, by definition. So although strains with incompatible alleles at any VIC module fail to form heterokaryons, they will readily mate so long as they differ at the *mat* locus. This shuffles

the many unlinked NSR modules at each sexual generation, making non-self vegetative fusions nearly impossible, even among siblings.

1.6. Long-distance non-self recognition in *Neurospora crassa*

This work focuses on the DOC-system: the genetic locus controlling the pre-contact NSR checkpoint. This subchapter recounts findings from Heller et al., 2016, in which the discovery and initial characterization of the DOC-system was published[42]. Unless otherwise cited, all statements in this section refer to this publication.

Initial phenotypic investigations of pre-contact NSR among isolates from a population in Louisiana revealed at least three mutually exclusive communication groups (CGs). Strains within each group communicate and grow towards each other as if they were genetically identical, while strains from different groups largely ignore each other. We defined FGSC2489 (the lab strain, originally from the Louisiana population) as CG1, crossed it with a CG2 isolate, and tested the offspring's communication with both parents. Half of the progeny communicated with one parent, and half with the other parent: this suggested a single locus was responsible for the CG phenotype. A bulked segregant analysis revealed an ~14 kb region of *N. crassa*'s genome segregated completely with CG phenotype. In FGSC2489, this region contained four adjacent genes (NCU07191-NCU07194 following the FGSC2489 nomenclature) that exhibited various rearrangements, duplications, and deletions in the Louisiana isolates. These rearrangements fell into five categories we called communication group haplotypes (CGHs) suggesting this population contained more CGs than we initially identified (Fig. 1.6-1). We found that CGH1 and CGH3 correlated perfectly with CG1 and CG3, respectively, while CGH2, CGH4, and CGH5 were all initially classified as CG2 because all communicated with the isolate we selected as a representative CG2 strain in our screen. However, after retesting the CG phenotypes of wild isolates with CGH2, CGH4, and CGH5 architectures, we noticed that CGH4 strains communicated well with both CG2 and CG3, while CGH5 strains had high rates of self-communication and intermediate communication rates with CG2. This left us with five CGs corresponding to the five CGHs, but only CG1 remained completely exclusive.

Between CGHs, alleles of NCU07191 and NCU07192 were <50% identical, while flanking genes (NCU07190, NCU07193, and NCU07194) showed >90% nucleotide sequence identity. Furthermore, protein-variants encoded by alleles of NCU07191 and NCU07192 within CGHs shared >95% amino acid sequence identity, but this dropped to ~35% between CGHs. We named these polymorphic genes determinants of communication (*doc*) one and two, respectively. In CGH1 through CGH4, *doc-1* and *doc-2* were paralogs, while CGH2 and CGH4 contained another paralog of *doc-2* we named *doc-3*. CGH5 only contained *doc-1*. Bioinformatic tools failed to reliably predict any functional domains in the DOC proteins, and no homologs could be identified outside the Sordariales. However, comparing alleles of the *doc* genes from the *N. crassa* Louisiana population with sequences from isolates of the closely related species *N. discreta* revealed evidence of trans-species polymorphism: a signature of long-term balancing selection common to NSR genes.

Given five CGHs with differing numbers of *doc* genes, we chose to focus on the simplest comparison: CG1 and CG3. Each has single copies of *doc-1* and *doc-2*, with alleles sharing less than 50% DNA identity, and communication between CGH1 and CGH3 strains was consistently low. We tested the communication phenotypes of *Δdoc-1*, *Δdoc-2*, and

Δdoc-1 Δdoc-2 in the FGSC2489 background against representative strains from CG1, CG2, and CG3 (Fig. 1.6-2 panel A). All *doc* deletion strains appeared morphologically indistinguishable from FGSC2489. Either single deletion caused drops in self-communication and communication with CG1. The *Δdoc-1* strain also increased communication with CG2 and CG3 to intermediate rates. We assumed the double deletion would result in a total loss of communication, but to our surprise *Δdoc-1 Δdoc-2* displayed a high self-communication rate, intermediate communication with CG2, and a loss of communication with CG1. We tested *Δdoc-1 Δdoc-2* with other representatives of CGH2 through CGH5 and found that it exactly mimicked the CG5 phenotype, suggesting the *doc-1* allele in CG5 strains may be non-functional (Fig. 1.6-2 panel B). These data indicate the DOC-system acts as a negative regulator of communication, repressing communication between strains that do not share compatible CGHs.

To confirm the *doc* genes are sufficient for specifying CG, we phenotyped strains expressing *doc-1*^{CGH3} and *doc-2*^{CGH3} from the wild isolate P4471 at the *his-3* locus in an FGSC2489 *Δdoc-1 Δdoc-2* background. This CG3-swap strain did not communicate with CG1 or CG5, but displayed high communication rates with CG3, confirming the *doc* genes fully control CG phenotype (Fig. 1.6-2 panels A and C).

We tagged *doc-1* and *doc-2* with gfp to observe their locations *in vivo*. Both C-terminally tagged genes complemented their respective deletions. DOC-2-GFP showed a peripheral localization, but we didn't confirm whether it was directly affiliated with the membrane. DOC-1-GFP appeared in cytoplasmic punctae that accumulated at the growing tips of communicating cells in an oscillatory manner during intercellular communication. This oscillatory pattern was reminiscent of the *N. crassa* protein SOFT and components of the MAK-2 complex, which oscillate between the cell body and growing tip of communicating cells in an alternating fashion. Because these proteins are required for cell fusion in *N. crassa*, an interaction between DOC-1 and SOFT or the MAK-2 kinase cascade could explain how the DOC-system influences communication. We imaged heterokaryons expressing DOC-1-GFP with SOFT-mCherry or MAK-2-mCherry. We found DOC-1 colocalizes with MAK-2 (Fig. 1.6-3) and, like other components of the MAK-2 complex, DOC-1 oscillates perfectly out of phase with SOFT. We also expressed DOC-1-GFP in a *Δmak-2* background to assess whether DOC-1 oscillation is dependent upon MAK-2: without MAK-2, DOC-1-GFP failed to oscillate and the strain phenocopied the *Δmak-2* mutant. We concluded that DOC-1, but not DOC-2, is a component of the MAK-2 complex.

The primary hypothesis explaining how genetically identical cells can communicate using identical signals is known as the "ping-pong" model[73,74]. This model predicts that oscillatory recruitment of intracellular communication machinery to the apical tips of communicating cells is coupled to the release of discrete signal pulses. We observed whether MAK-2-GFP oscillations occurred between germlings from different CGs. Initial oscillations between germlings of all CGs were common, but these oscillations quickly dampened and stopped between strains from incompatible CGs. This suggested cells must be sending and receiving two units of information: a general signal recognized by all strains, and a CG-specific signal for the DOC system*. It occurred to me that CG information could be encoded in the frequency of general signal pulses. This one-signal-two-messages hypothesis could allow the DOC system to modulate communication without requiring a CG-signal molecule or receptor. I tested this hypothesis by timing the oscillation frequencies of SOFT-GFP in FGSC2489 and *Δdoc-1 Δdoc-2* genetic backgrounds (Fig. 1.6-4).

Although these two strains do not communicate, their SOFT oscillation frequencies were statistically identical, refuting the one-signal-two-messages hypothesis. Therefore, some unidentified CG signal must be exchanged between communicating cells.

Based on these results, we hypothesized the following model in which the DOC-system negatively regulates communication between cells via an interaction with the MAK-2 complex (Fig. 1.6-5):

1. Cells secrete some unidentified general signal independent of CG. This notifies cells of each other's presence and initiates communication machinery oscillations and pulsatile signal release.
2. After initiating communication, cells must also secrete an unidentified CG-specific signal, which is received and interpreted by the cells' DOC-systems. If a cell does not receive a compatible CG signal, its DOC-system alters the MAK-2 complex to inhibit oscillatory recruitment of these components to the growing tip. This interrupts communication between the cells, leading to an incompatible CG phenotype in which cells appear to ignore each other.
3. If both cells receive compatible CG signals, communication proceeds. DOC-1 continues to oscillate with the MAK-2 complex during communication in both cells, concluding with intercellular contact.

*Heller et al., 2016, do not explicitly claim there are two communication signals or name the one-signal-two-messages hypothesis. But in personal communications with my co-authors, we mostly agree that communicating cells must exchange two units of information to explain the DOC system's behavior, and the publication describes the one-signal-two-messages hypothesis and the results of my tests of this hypothesis.

1.7. Questions and predictions that guided this work

The main goal of the work presented in this dissertation is to improve our model of *doc*-mediated CG specificity. Although it's admittedly vague, the model outlined in section 1.6 makes several important predictions and inspires some interesting questions. Below, I've listed a few predictions and questions:

- If the DOC-system negatively regulates communication solely via interactions between DOC-1 and the MAK-2 complex, the *Δdoc-1* mutant should have parental levels of self-communication and should communicate with the *Δdoc-1 Δdoc-2* mutant.
    - However, the *Δdoc-1* mutant had reduced self-communication, implying DOC-2 must have the ability to influence communication independently of DOC-1.
- Direct or indirect interactions between matched variants of DOC-1 and DOC-2 must be important for CG specificity; otherwise we would expect recombination between haplotypes, which we did not detect. The lack of recombination between CGHs also suggests that coexpressing a *doc-1* allele from one CGH with a *doc-2* allele from another CGH should result in aberrant communication behavior and fitness costs. I will refer to this conjecture as the non-allelic interaction hypothesis.

- Because the DOC-system negatively regulates communication, coexpression of *doc* alleles from incompatible CGHs should inhibit communication with strains sharing either CGH. I will call this prediction the competition hypothesis.
- The CG5-mimicing behavior of the *Δdoc-1 Δdoc-2* mutant suggests *doc-1*^{CGH5} has lost its communication functions. This is the CGH5 non-functionality hypothesis. However, if the CGH5 variant of DOC-1 were totally non-functional, one would expect the *doc-1*^{CGH5} allele to have accumulated mutations and become a pseudogene. But RNAseq confirmed this allele's expression, suggesting it may have some function(s) we were unable to detect. Is the CGH5 allele truly non-functional? If not, does it retain some function unrelated to communication?
- If DOC-1 suppresses communication through interactions with the MAK-2 complex, all functional *doc-1* alleles must encode a domain that interacts with this complex. This domain should be relatively conserved between alleles, and should be required for DOC-1 to influence communication. I will refer to this assumption as the response domain hypothesis.
- Which members of the MAK-2 complex interact with DOC-1? Unpublished co-immunoprecipitation experiments by Jens Heller and immunoprecipitation-mass-spectrometry experiments I performed (data not shown) designate HAM-5, the MAK-2 complex scaffold protein[75], as the most likely candidate for direct interactions with DOC-1.
- Some region(s) of DOC-1 and DOC-2 must vary with CGH, controlling CG specificity. I'll call this assertion the specificity domain hypothesis.
- Can the inhibitory and recognition functions of DOC-1 be separated?
- The *doc* genes must encode a CG-specific signal, but are they catalyzing the production of a small molecule or is the CG signal a peptide cleaved from a DOC protein? Does the signal enter the cytoplasm, or does DOC-2 act as a cell surface receptor? What are the biochemical functions of the DOC proteins? Do they have any functions unrelated to communication?
- What are the signals? They must be short-lived to avoid self-stimulation, so are they degraded or volatile? Is there some way to convey two units of information with one signal molecule (other than the defunct one-signal-two-messages hypothesis)? If not, are there two separate signal molecules, or a modular molecule with general and CG-specific components?

These questions and predictions guided the work presented in this dissertation. I wasn't able to validate or answer them all, but my work has added detail to our model of CG specificity.

1.8. Summary and discussion

By working together, social cells can gain an advantage over their asocial rivals[10]. However, if social goods can be received without contributing, freeloaders will parasitize their altruistic neighbors and reduce the cellular society's overall fitness[11–14]. To counteract

freeloaders and other potential dangers of sociality, social cells try to limit their cooperation to genetically similar individuals[15]. This requires cells discriminate between self and non-self others, and has driven organisms across the tree of life to independently evolve NSR: the ability to adjust social behaviors according to the genetic identities of interacting conspecific individuals.

Although NSR genes are diverse and often share no homology, they usually share some characteristics[1–7]. If multiple genes must interact to endow cells with an NSR identity, they are often tightly linked into functional modules. Each NSR module acts as a green-beard or kind selection locus[10,42–45], and the action of many unlinked modules is required for cells to reliably distinguish self from genetically similar individuals[46]. NSR modules always produce multiple variants of at least one component throughout the population. Differences between variants confer identities and NSR alleles show signs of balancing selection as a result of rare-allele advantage[1]. In many cases, balancing selection on NSR genes survives speciation, resulting in trans-species polymorphism[1,21].

The filamentous ascomycete *N. crassa* lives as a syncytium and undergoes frequent vegetative cell fusions to construct its mycelium[48]. This highly social lifestyle has driven the evolution of many NSR modules that function during five developmental checkpoints[62,63]: pre-contact, post-contact, post-fusion, hyphal-fusion, and fertilization. The pre-contact checkpoint is unusual because most NSR modules require intercellular contact for recognition. My work focused on the *doc* genes; the single NSR module controlling pre-contact NSR in *N. crassa*. Cells expressing compatible alleles at the *doc* locus communicate and grow together, while those expressing incompatible alleles ignore each other. Cells that communicate are said to share a CG, while those that fail to communicate are classified into different CGs.

Previous work[42] confirmed the *doc* genes are necessary and sufficient to specify CG and identified five *doc* haplotypes conferring different communication patterns. These patterns were organized into five CGs, only one of which was exclusive. The different haplotypes contained between one and three paralogous *doc* genes, none of which revealed any identifiable functional domains. Localization studies of proteins encoded by the two *doc* genes in FGSC2489 placed DOC-2 at the cell's periphery while DOC-1 oscillated between the cell body and growing tip with the MAK-2 complex. Although single deletions of either *doc* gene reduced self-communication in FGSC2489, deleting both genes converted the strain from CG1 to CG5. This suggests the lone *doc* gene in CGH5 strains may be non-functional and confirms the DOC-system represses communication between strains until a compatible CG signal is received.

However, earlier work left mechanistic questions unanswered: how do the DOC proteins dampen MAK-2-complex oscillations, produce and perceive CG signals, and interact with each other and additional cellular components? This dissertation presents my attempts to improve the model of DOC-mediated pre-contact NSR in *N. crassa* developed by Heller et al[42]. Further investigations will be required for a full understanding of DOC-mediated NSR. Along the way, I developed automated methods for quantifying communication and fusion between germlings using flow cytometry assays[65]. These methods may prove useful for NSR researchers working in fungal and non-fungal systems.

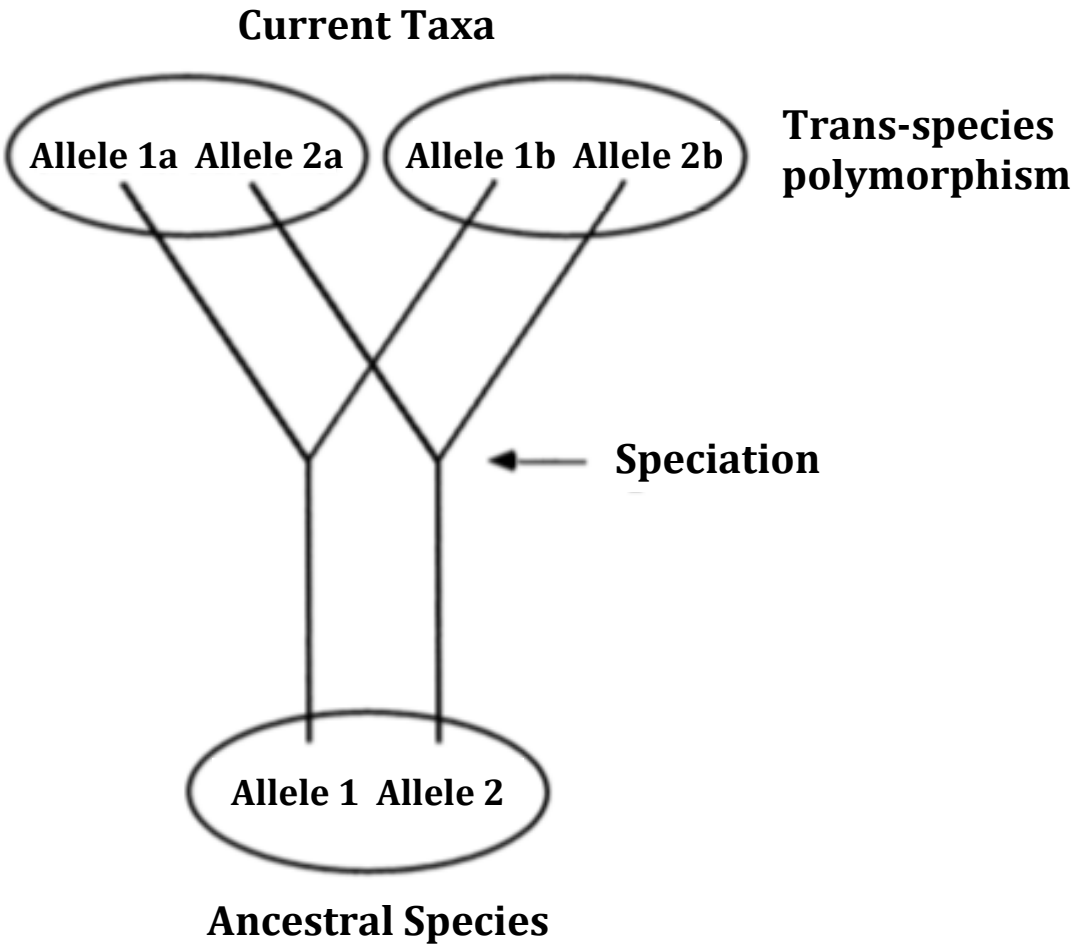


Figure 1.3-1. Trans-species polymorphism
Modified from "Fig. 1. Trans-specific evolution as a consequence of long persistence of allelic lineages" from Richman, 2000. Trans-species polymorphism results when two allelic lineages present in an ancestral species are transmitted to each of two daughter species. In such cases, the closest relative of an allele in one daughter species is found not in that species, but in the other daughter species.

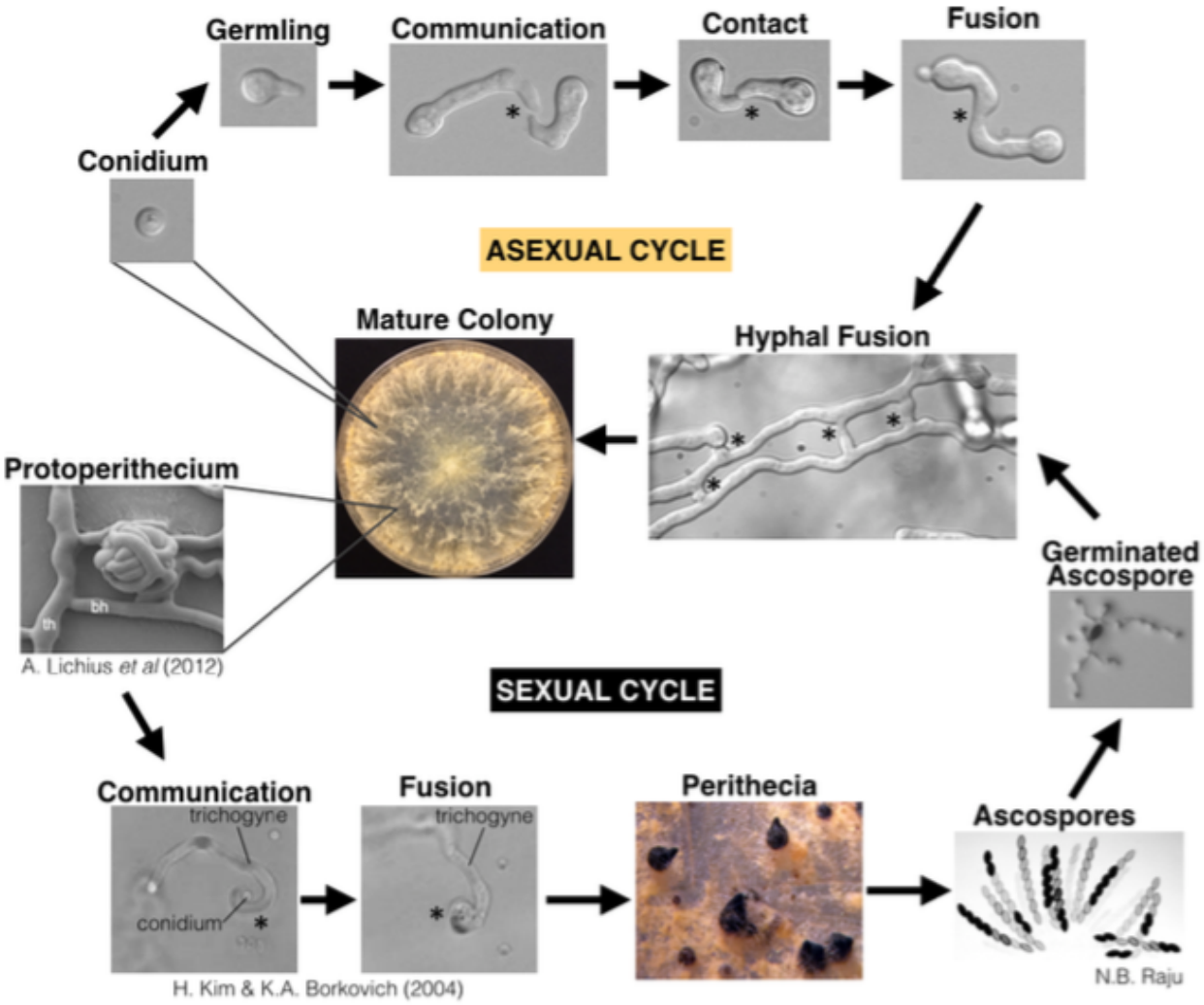


Figure 1.4. Life cycle of *Neurospora crassa*

Taken from "Fig. 1-1. The life-cycle of *Neurospora crassa*" from Fischer, 2018 (dissertation). *N. crassa* is a heterothallic Ascomycete with a distinct sexual cycle and asexual cycle. Conidia are clonal asexual propagules that can either generate a new colony on their own, or conidia serve as the "male" partner to the "female" trichogyne during mating. Ascospores are the result of meiosis, which occurs inside the perithecium, and then ascospores are forcibly shot out of the top of the perithecium. This lifecycle specifically highlights chemotropic interactions and cell fusion events. Stars indicate chemotropic interactions and fusion. Germlings, hyphae, and the trichogyne all undergo chemotropism and cell fusion. Protoperithecium image is from (Lichius *et al.*, 2012), trichogyne images are from (Kim and Borkovich, 2004), and the image showing many ascospores is available at https://web.stanford.edu/group/neurospora.

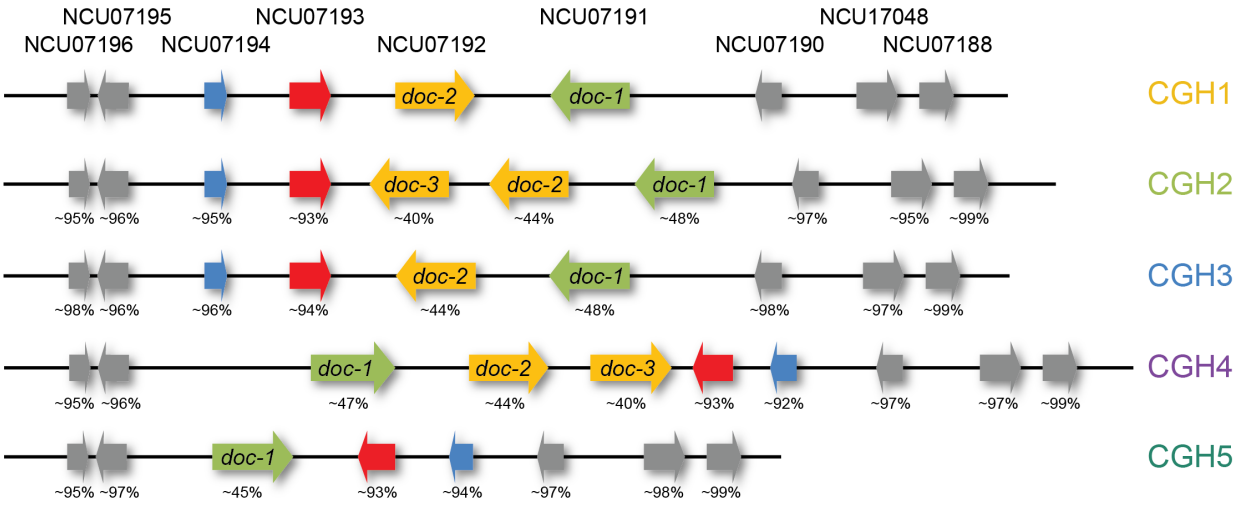


Figure 1.6-1. Communication group haplotypes at the *doc* locus
Taken from panel B of "Fig 2. Communication group haplotypes" from Heller et al., 2016. Genomic rearrangements within the CGHs spanned the genetic interval between NCU07191 and NCU07194 and included duplications of NCU07192 (*doc-2*, *doc-3*), a deletion of NCU07192, and inversions. Alleles at NCU07191 (*doc-1*) and NCU07192 (*doc-2*, *doc-3*) within a CGH show high DNA sequence identity, but are polymorphic between CGHs. The percent DNA identity between alleles in members of the different CGH groups across the genetic interval in comparison to FGSC2489 (a member of CGH1) are shown.

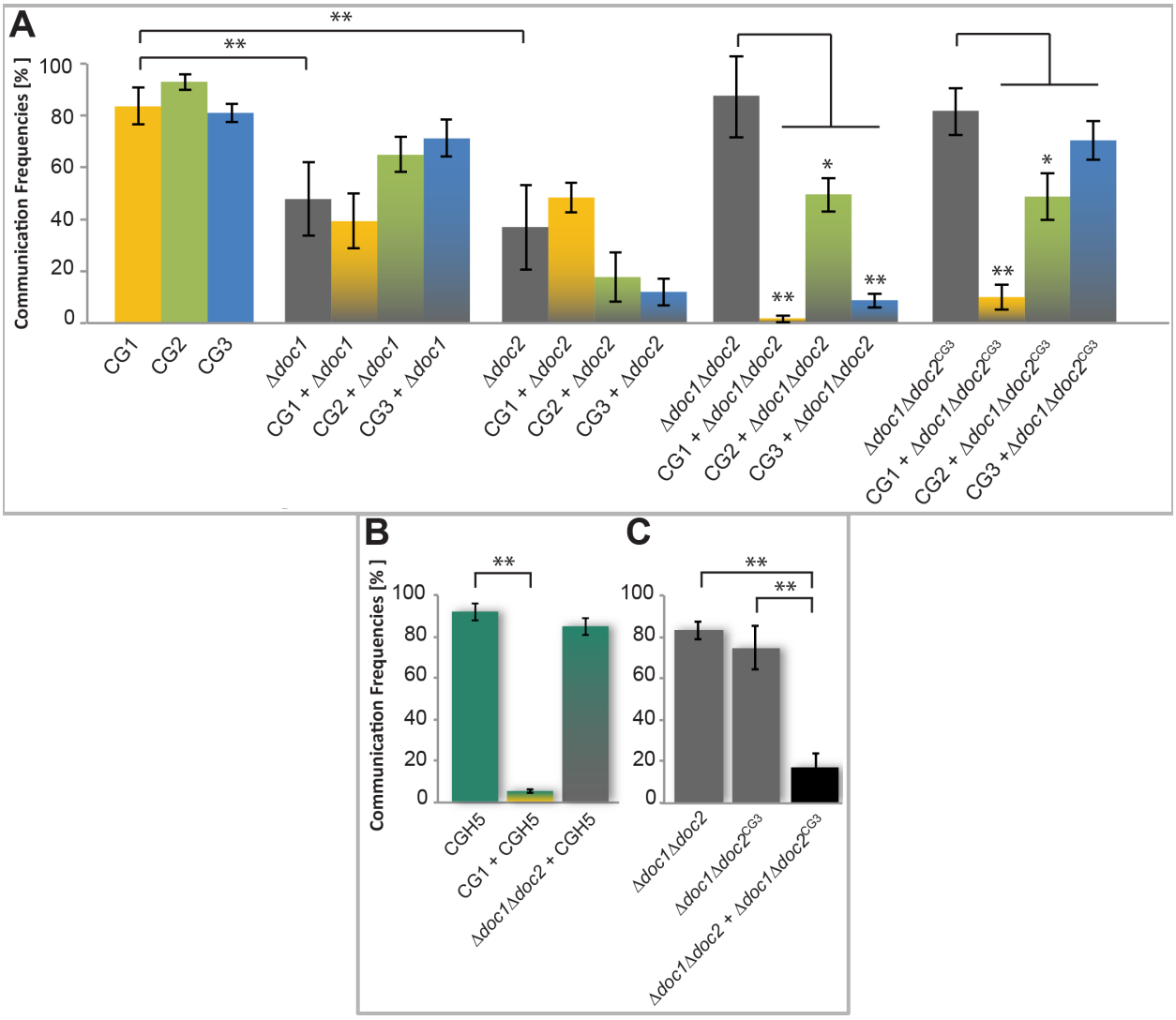


Figure 1.6-2. Microscopy communication assay results from *doc* deletion and CG3-swap strains

Bar graphs taken from panels A, B, and C of "Fig 3. Communication interactions of *Δdoc-1*, *Δdoc-2*, *Δdoc-1 Δdoc-2*, and *Δdoc-1 Δdoc-2* (*his-3::doc-1*^{CG3} *doc-2*^{CG3}) germlings" from Heller et al., 2016. Single-color bars depict self-communication and two-color bars depict non-self communication. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483, and CGH5 = JW220. All experiments were performed in triplicate with each replicate counting at least 100 germling pairs. Pairwise comparisons were made using Student's *t*-tests: * = $p < 0.05$, ** = $p < 0.01$. **A)** Self and non-self-communication rates of CG1, CG2, CG3, *Δdoc-1*, *Δdoc-2*, *Δdoc-1 Δdoc-2*, and the CG3-swap strain. **B)** Self-communication rate of CGH5 and communication rates between CG1 and CGH5, and between *Δdoc-1 Δdoc-2* and CGH5. **C)** Self-communication rates of *Δdoc-1 Δdoc-2* and the CG3-swap strain, and communication rates between those two strains.

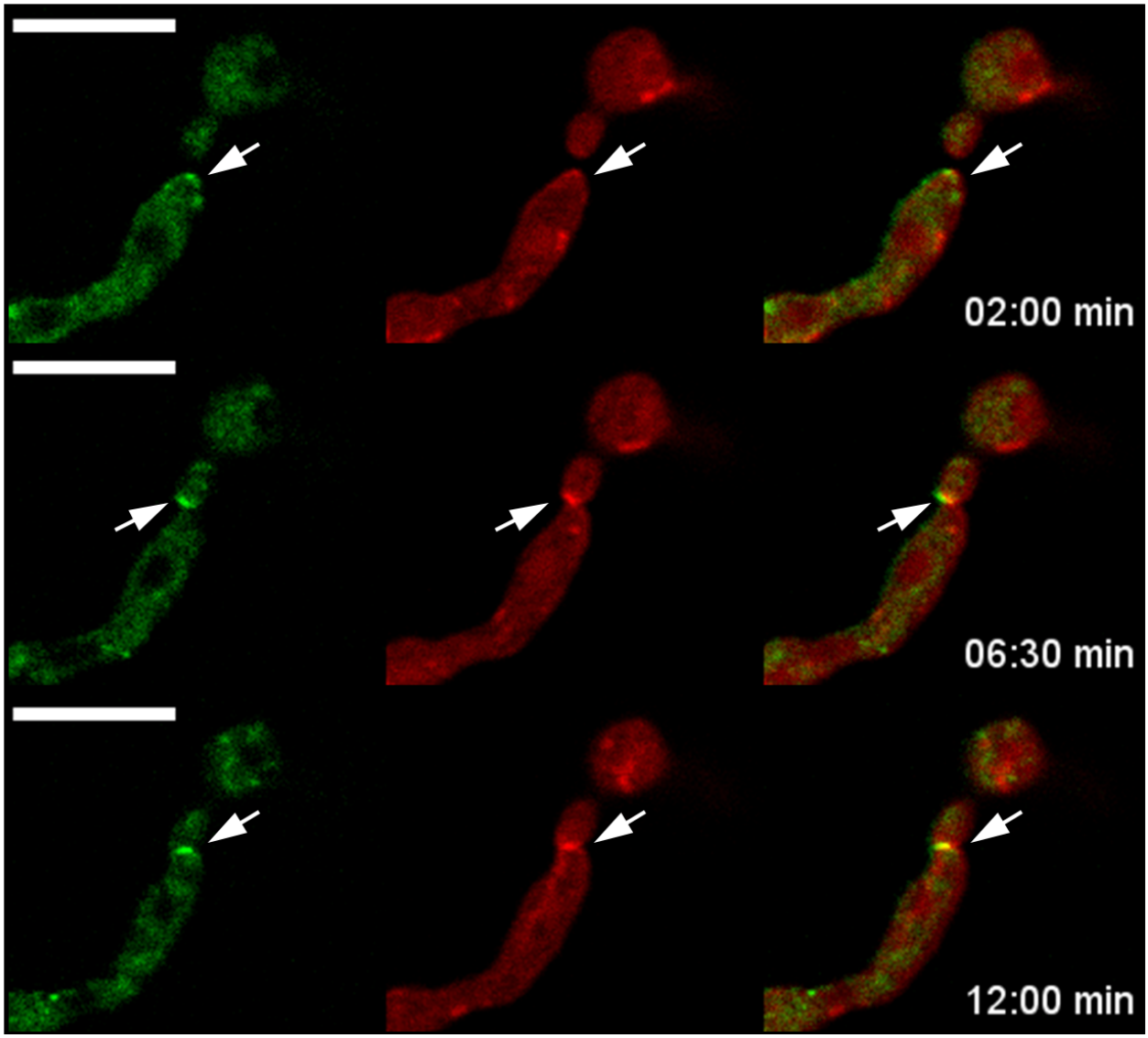


Figure 1.6-3. DOC-1-GFP co-localizes with MAK-2-mCherry during chemotropic interactions

Taken from panel C of "Fig 5. DOC-1-GFP co-oscillates with MAK-2 during chemotropic interactions" from Heller et al., 2016. Co-localization and co-oscillation of DOC-1-GFP (left panel) with MAK-2-mCherry (middle panel) in heterokaryotic germlings undergoing chemotropic interactions (overlay, right panel).

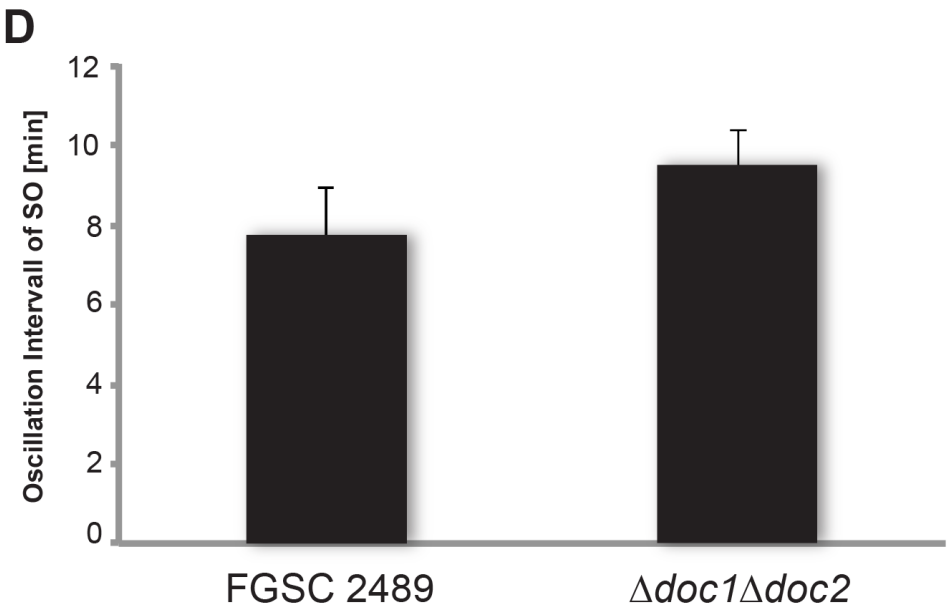


Figure 1.6-4. SOFT-GFP oscillation frequencies in FGSC2489 and *Δdoc-1 Δdoc-2* genetic backgrounds

Taken from panel D of "S5 Fig. Oscillation dynamics of DOC-1-GFP in hyphae and germlings" from Heller et al., 2016. SOFT-GFP oscillation intervals were measured in communicating FGSC2489 germlings (CG1, n = 3) and in *Δdoc-1 Δdoc-2* germlings (CG5, n = 4). There was no significant difference in oscillation timing detectable ($p > 0.5$).

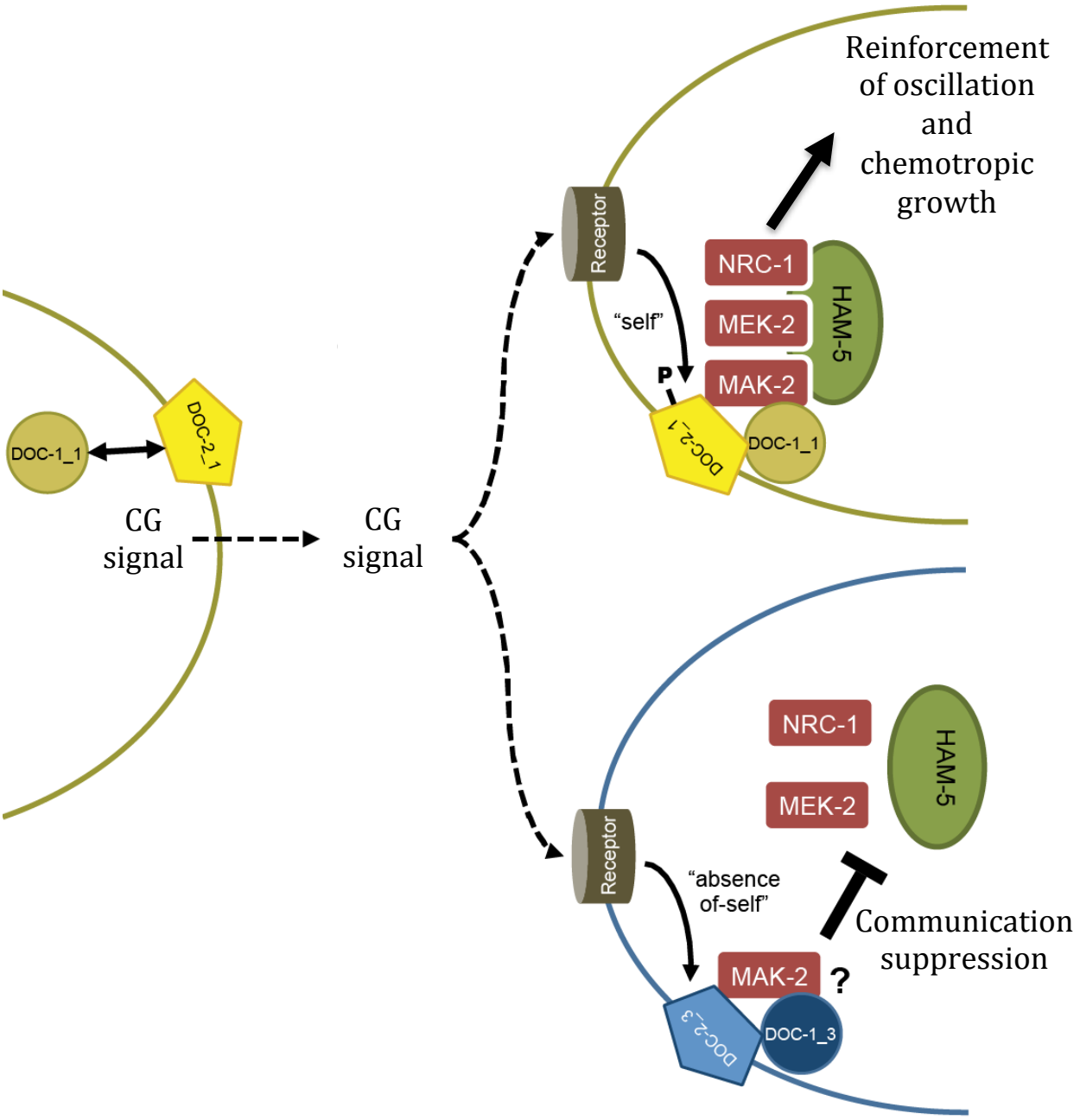


Figure 1.6-5. Model for the DOC system from Heller et al., 2016
Modified from "Fig 8. Model for DOC-1 and DOC-2 function in long-distance kind discrimination" from Heller et al., 2016. N. crassa germlings release a general signal for potential interaction partners (not shown) and a CG-specific signal encoded by the DOC system of the sending cell. If the CG signal is recognized by the receiving cell's DOC system, oscillation of the assembled MAK-2 complex is enforced and the signal-receiving cell shows chemotropic interactions (top). If the CG signal is not recognized, the DOC system of the receiving cell prevents enforcement of MAK-2 oscillation, and the receiving cell does not respond to the presence of a potential fusion partner (bottom).

Chapter 2. Improving and automating communication assays

2.1. Quantifying intercellular communication and fusion in *Neurospora crassa*

The first *N. crassa* mutant incapable of undergoing cell fusion was documented in 1999[38]. Since then, over 70 genes with roles in cell fusion have been identified[76]. Many genes required for cell fusion are termed *hyphal anastomosis mutant* (*ham*) genes, but I will refer to them as fusion genes when making broad generalizations. Loss-of-function mutations in many fusion genes cause pleiotropic effects, commonly including the loss of aerial hyphae ("flat" phenotype) and sexual defects[58]. The correlation between easily observed morphological phenotypes and mutations in known fusion genes have been used in forward genetic screens to identify candidate fusion genes[62]. Strains bearing mutations in candidate genes are observed microscopically for evidence of self-fusion; if no such evidence can be found, the candidate is validated as a fusion gene[77]. Germlings are typically used in microscopy fusion assays because they occupy fewer focal planes than colonies, and exhibit more synchronous behavior and development. Fusion between strains can be similarly assayed, with dyes or fluorescent markers identifying one or both strains[42,63,64.]

Microscopy assays for intercellular fusion are cumbersome, as fusion between cells in physical contact isn't always obvious. This is especially relevant for fusion mutants that retain their ability to communicate, and those with intermediate fusion rates. Quantification is improved when strains are divided into two fluorescently labeled groups: one group is stained with a cell wall or membrane dye, and the other expresses a cytoplasmic fluorescent marker. If germlings from the two groups fuse, the cytoplasmic marker will flow from one to both cells while the wall or membrane dye is restricted to one cell (Fig. 2.1-1). However, the use of a fluorescent marker requires further strain construction and limits the application of this improvement.

The problems with microscopy-based fusion assays are compounded for communication assays. Observing the flow of a fluorescent marker is not applicable, and communication is far more ambiguous than fusion. For example, are two cells a few μm apart and growing toward each other communicating, or growing that way by chance (Fig. 2.1-2)? Time-course observations can clarify whether germling pairs are communicating as mutual chemotropic growth and eventual cell contact can be confirmed; but this drastically increases the assay's complexity and time requirement. Optical tweezers can also be used to move germlings and the subsequent reorientation in their growth can unambiguously identify communicating pairs[77]. However, this method makes the assay even more complicated and time-consuming than a time-course. The ambiguity inherent in identifying communicating cells causes variable quantification of the same microscopy images by different researchers, although trends remain consistent (unpublished data from the Glass lab). Blind analysis is also required to prevent experimenter bias from influencing the quantification. These challenges prompted us to seek alternative communication and fusion assays that are faster, easier, and more reliable.

2.2. Automated microscopy and image analysis

Automation could provide one solution to the problems with microscopy assays. Automated fluorescence microscopes are capable of capturing multi-channel time-series images at multiple locations in a sample concurrently[78–80]. Such microscopes are able to capture time-courses of germling communication with minimal intervention. While this makes microscopy assays easier, manual quantification of the images remains problematic. An ideal solution would pair automated image analysis and communication quantification with automated microscopy. Advancements in image analysis have allowed the automated localization of fluorescently tagged proteins and automated phenotyping of various cell types[78,81,82]. Most of these analysis programs only work for non-overlapping, regularly shaped cells, although software capable of "untangling" and analyzing still images of overlapping nematodes has been developed[83]. Unfortunately, *N. crassa* germlings are far more irregularly shaped and sized than nematodes, and no existing software can reliably differentiate fused cells from larger single cells.

Marcus Roper's lab at UCLA has made great strides combining automated microscopy time-courses with automated image analysis of *N. crassa*. By counting conidia in the first image of a time-series, his system can simultaneously track the germination and growth of many cells. Although it remains to be realized, he should also be able to incorporate fluorescence images into his analysis software to automate the detection of cell fusion via the fluorescent marker flow technique described in the previous subchapter. However, Dr. Roper hasn't developed an automated method to identify homing growth between germlings, and therefore his assay cannot be used to quantify intercellular communication (personal communication).

Since the complex image analysis problems preventing automated quantification of germling communication via microscopy have yet to be solved, we began investigating quantification methods that do not rely on microscopy. While many of the following assays require additional strain construction and ultimately rely on cell fusion as a proxy for communication, I was eventually able to develop an assay that requires neither strain construction nor cell fusion.

2.3. Quantitative heterokaryon assay

Perhaps the simplest method for quantifying germling fusion relies of the ability of heterokaryotic nuclei to complement each other's auxotrophies. If conidia from two FGSC2489 mutants lacking the ability to synthesize unrelated metabolites (e.g. uridine and histidine) are spread on separate minimal media (MM) colony-inducing plates, few colonies will grow. However, if conidia from the mutants are plated together, the heterokaryons they form will be prototrophic and many colonies will grow. The number of colonies on both individual plates and the mixed plate can simply be counted to quantitate fusion rates. Because only the heterokaryons can grow in this assay, we call this technique a quantitative heterokaryon assay[42].

By combining complementary selection markers with mutations in genes potentially involved in cell fusion, one can quantify how the mutations affect self-fusion rates compared to an FGSC2489 control using quantitative heterokaryon assays. Fusion between strains can be similarly quantified by adding complementary selection markers into the different strains. If the mutations being tested are known to leave cell-wall breakdown and cell fusion unaffected, the prototrophic colony count becomes a proxy for intercellular

communication. Heller et al., 2016, confirmed this assay returns data comparable to those from microscopy assays: strains that do not communicate form very few prototrophic colonies[42].

This technique has two major limitations. First, mutants must share a genetic background. If strains to be tested have differing alleles at any VIC loci, the heterokaryons they form will sector or die and no prototrophic colonies will appear regardless of their fusion rate. Second, selection markers must be added to the strains, typically via crossing the mutant to be tested with an FGSC2489 auxotroph. These drawbacks limit the forced heterokaryon assay to FGSC2489 mutants because marker-bearing strains are only available in this background. However, most experiments are performed in this background anyway, so a different assay is only required for wild isolates and is helpful for sterile mutants.

A third minor drawback of this assay is the time required from start to quantification. While data from other assays takes a single day to obtain, the forced heterokaryon assay requires time for colonies to grow. This typically takes about a week from the time plates are inoculated, although no work is required from the researcher during this delay. Counting the colonies once they mature is much faster and simpler than manual analysis of microscopy images. Although the forced heterokaryon assay is simple and elegant, the ability to screen wild isolates is essential for identifying NSR loci. We continued to search for a more flexible assay that overcomes the problems with microscopy.

2.4. Flow cytometry assays

2.4.1. Introduction to flow cytometry

After identifying the *sec-9*/*plp-1* GRD locus, Heller et al. used flow cytometry to quantify programmed cell death induced by that NSR module (see 1.5. Non-self recognition checkpoints in the life cycle of *Neurospora crassa* for details about GRD)[64]. This technique proved to be very useful for quantifying germling communication and fusion. This subchapter will introduce you to flow cytometry and the measurements we take while assaying germling communication or fusion. Unless otherwise cited, information is sourced from Howard Shapiro's *Practical Flow Cytometry*[84].

Flow cytometry passes individual cells through a series of lasers and records how each cell interacts with the light (Fig. 2.4.1-1). First, a suspension of cells is fed into the instrument where a fluidic system mixes the input with sheath fluid (e.g. phosphate-buffered saline, PBS) and guides the cells into a single-file line through a capillary tube. As cells flow through the tube, they pass through a series of laser beams while detectors record scattered and fluorescent light. Sensors in-line and perpendicular to the first laser's path detect cells and record measurements called forward scatter (FSC) and side scatter (SSC), respectively. FSC measures the amount of light scattered through a small angle (~0.5° to 5°) by cells as they pass through a laser. SSC measures the amount of light deflected enough to reach the detector positioned perpendicular to the laser's path (~15° to 150°). After the first laser, cells pass through other beams that excite fluorescent molecules while detectors quantify the emitted fluorescence. When cells exit the capillary

tube, they are sent either to waste or a cell-sorter that separates cells based on their measured characteristics.

As data is collected from each cell, pulse height, width, and area metrics can be recorded from each measurement channel (Fig. 2.4.1-2). The maximum signal strength recorded in each channel for each cell is the pulse height. The pulse width is the time from the initial detection of a cell until the signal returns to baseline. Area metrics are obtained by integrating measurement signals over time (pulse width). Given any two of these metrics, the third can be estimated by assuming signal pulses' shapes will approximate triangles (or any other well-defined shapes, e.g. Gaussian distributions).

FSC-area (FSCA) roughly corresponds to a cell's volume, although many other factors affect FSCA. FSC-width (FSCW) is a more reliable metric for cell size, but electrical impedance is considered superior. SSC-area (SSCA) measures a cells internal complexity (also called granularity) and correlates with a cell's refractive index. Fluorescence-area measures a cell's total fluorescence in a given channel. Although all these data can be gathered using microscopy, the benefit of flow cytometry is speed: modern flow cytometers can analyze thousands of cells per second, returning population-level distributions of relevant parameters.

Using flow cytometry on filamentous fungi presents two challenges: size limitations and suspending cells. If even a small percentage of the cells being analyzed are too big, the flow cytometer will clog; this situation can require time-consuming remediation. Clogging can be avoided by filtering conidial suspensions to remove hyphal fragments and carefully timing experiments to give cells enough time to fuse without allowing excess growth. Assaying a mixture of slow and fast-growing strains requires the slower strain be inoculated first to permit sufficient growth without the faster strain over-growing.

As mentioned above, flow cytometry requires cells be suspended in liquid. Although *N. crassa* can be grown in liquid culture, we've found communication and fusion rates drop significantly under those conditions, even without shaking. This is probably because cells must be stationary for communication signal-gradients to form, and the larger volume and lower viscosity of liquid cultures may alter these gradients. However, germlings adhere to the surface of solid media and there is no way to resuspend them without damage. Heller et al. solved this problem using Pluronic F127 (Sigma-Aldrich), a gelling agent that is semisolid at 30^{o}C and liquefies at 4^{o}C. Growing germlings on Vogel's minimal media solidified with Pluronic F127 (MMP) allows them to communicate and fuse on a solid surface, then to be harvested and concentrated via centrifugation[63–65].

In the remainder of this chapter, I will describe the assays and automated analyses we developed for quantifying germling communication and fusion using flow cytometry. In general, we used FSC and fluorescence measurements in our analyses.

2.4.2. Flow cytometry fusion assay with induced death

The experimental design and concept for using flow cytometry to quantify death rates induced by fusion between germlings expressing incompatible GRD alleles are detailed in Heller et al. 2018[64]. After that proof of principle, we generated CG death-inducer (DI) strains by crossing the *Δplp-1 Δplp-2 sec-9^{GRD3}* strain (*sec-9*-swap strain CG1), which induces death upon fusing with FGSC2489, with the CG3-swap strain (CG3) or *Δdoc-1 Δdoc-2* (CG5)[42,64]. We could then use the induced death flow cytometry assay (IDFC) to quantify

cell death in mixtures of an arbitrary strain and the CG1, CG3, or CG5 DI strains as a proxy for communication between the tested strain and CG1, CG3, or CG5 strains (see Fig. 2.4.2-1 for microscopy images depicting how the flow cytometer detects fusion events). Because fusion and death induction with the *sec-9*-swap strain are required for IDFC to work, any strain tested with this assay must have an FGSC2489 genetic background.

My IDFC protocol is slightly modified from the one presented by Heller et al.[64], and standard protocols for *N. crassa* can be found on the Neurospora homepage at the FGSC (www.fgsc.net/Neurospora/NeurosporaProtocolGuide.htm). The workflow for flow cytometry communication and fusion assays is diagrammed in Fig. 2.4.2-2. Strains to be tested were grown in Vogel's minimal media slants for three days at 30°C in dark and seven days at 20-25°C in light. Conidia from each strain were then suspended in water and filtered through cheesecloth to remove hyphal fragments. Conidial titers were adjusted to $3x10^7$ conidia/mL and equal volumes of suspensions from strains whose post-fusion death rates would be measured were mixed. 88 µL of individual or mixed conidial suspensions were plated on 60 mm x 15 mm MMP plates, or 30 µL of suspensions were plated on 35 mm x 10 mm MMP plates (both ~935 conidia/mm^2), and equal volumes of single-strain conidial suspensions were saved at 4°C for germination controls. Plates were incubated at 30°C in dark for four hours. Four hours post-inoculation (4 HPI), plates were liquefied at -20°C for ten minutes before centrifugal harvesting and two washings in cold PBS. Germling suspensions and conidial suspensions saved at 4°C were resuspended in 0.5 or 1 mL (depending on plate size) PBS with 0.1 µM SYTOX Blue (SB, Life Technologies) and/or 0.15 µM propidium iodide (PI, Sigma–Aldrich). Stained conidial and germling suspensions were then analyzed on a BD LSR Fortessa X-20 flow cytometer using BD FACSDiva software (BD Biosciences), recording FSCA, FSCH, and SB and PI fluorescence areas (Pacific Blue and PerCP-Cy5-5 channels).

Data from at least 20,000 cells were recorded in each experiment, and each experiment was repeated at least three times. In each experiment, mixtures of the tested strain and the CG1 DI, the CG3 DI, and the CG5 DI strains were analyzed. In retrospect, I should have also generated a DI version of each strain I analyzed and used the IDFC assay to quantify self-communication rates. The time and effort required to construct DI versions of every strain so their self-communication can be assayed using IDFC was a major incentive for my development of dyeFC, described in sections 2.4.4 and 2.4.5.

Many controls are required to extract usable information from flow cytometry communication assays. Conidial control data from each strain were used to remove ungerminated cells from death rate calculations. Two vital dyes, SB and PI, were used as technical controls to ensure staining and fluorescence worked as expected. Each experiment includes FGSC2489 + CG1 DI (both CG1) as a positive control and either FGSC2489 + CG3 DI or FGSC2489 + CG5 DI (different CGs) as a negative control to confirm DI strains behaved as expected.

2.4.3. Automated analysis of induced death flow cytometry data

Heller et al., 2018, manually removed ungerminated conidia and defined fluorescence thresholds to analyze IDFC data[64], I wanted to use a faster, higher throughput, and less biased method. To that end, I sought to develop an automated IDFC analysis pipeline using MATLAB™ (version R2018b, MathWorks) with the help of Professor Marcus

Roper from UCLA. After many iterations, I settled on the following automated workflow for each experiment:

1. Remove ungerminated conidia from germling samples (Fig. 2.4.3-1).
    a. Match conidial and germling (4 HPI) samples. For strain-mixtures, get conidial samples from each individual strain.
    b. Generate 200 cell size bins using FSCW as a size metric. Estimate FSCW* by dividing FSCA by FSCH for each cell. If FSCH was not recorded in an experiment, FSCA was used instead of FSCW. Divide germling sizes into 200 equally spaced bins and normalize cell counts between germling and conidial samples. Divide conidial sizes into the bins defined for germlings.
    c. Subtract conidial size bins from germling size bins and set all negative values and values from bins smaller than the 50^{th} percentile of conidial size to zero. For strain-mixtures, use the mean conidial size distribution. Divide the value in each subtraction bin by the value in the corresponding germling bin to generate per-bin germination ratios.
    d. Calculate how many cells from each germling size bin have germinated by multiplying the number of cells in each bin by its per-bin germination ratio. Randomly select this number of cells from each germling size bin as germinated. For example, if a 4 HPI germling FSCW bin contains 200 normalized counts and the corresponding conidial bin contains 100 normalized counts, then 50% of the 4 HPI cells from this bin will randomly chosen as germinated.

       *Note that FSCA/FSCH = 2 x FSCW, assuming triangular pulses. However, since size estimates of all cells will contain the extra factor of two, there is no reason to reduce these estimates.
2. Define fluorescence gates using the positive control sample (Fig. 2.4.3-2).
    a. Generate a kernel distribution with normal smoothing function for the natural logarithm of fluorescence data from germinated positive control cells in each dimension (SB and PI).
    b. Find the most fluorescent and largest maxima in the kernel distribution. Confirm these maxima are sufficiently separated and in reasonable positions in the distribution.
    c. Set the fluorescence gate at the minimum value in the kernel distribution between these two maxima. Scrap the gate if it suggests an unreasonably high percentage of cells are dead (e.g. over 60%).
    d. If two reasonable maxima cannot be found in the kernel distribution, flag the data from that experiment in that channel (SB or PI) as unreliable and set a speculative fluorescence gate halfway between the most fluorescent kernel distribution maximum and the 99^{th}

percentile of fluorescence.

3. Calculate relative death rates (Fig. 2.4.3-3).
    a. For each mixture of strains, get the fluorescence data from the mixture and from each individual strain. Calculate the percentage of all three fluorescence distributions that exceed the fluorescence gate in each channel (SB and PI). These are the death rates for the three samples.
    b. Calculate the average percent dead in each channel (SB and PI) for the individual strains.
    c. Calculate the mixture's relative death rate in each channel (SB and PI) by dividing its percent dead in each channel by the average percent dead for the individual strains in the corresponding channel.

Using relative death rates helps account for variation in death rates caused by uncontrolled variables. Once relative death rates for three or more replicate experiments were obtained, values were statistically analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Analyzing IDFC experiments in this way, I can detect up to five distinct levels of communication with statistical support at $p = 0.05$ or better: low communication (same as negative control), intermediate-low communication (between negative and positive control, but closer to negative), intermediate-high communication (between negative and positive control, but closer to positive), high communication (same as positive control), and super communication (higher than positive control). Thus, a strain's complete CG phenotype may be given as a communication level with itself, CG1, CG3, and CG5. For example, FGSC2489's CG phenotype is self^{hi}CG1^{hi}CG3^{lo}CG5^{lo}.

2.4.4. Flow cytometry communication assay with two dyes

The IDFC assay works well, but it has three notable shortcomings. While only the first of these issues directly affected my work, all motivated me to develop a more general, easier to use flow cytometry based assay:

- IDFC requires strain construction to generate DI versions of each strain if self-communication rates are desired.
- All tested strains must be in the FGSC2489 genetic background, which precludes the use of IDFC on wild isolates.
- IDFC requires a rapid, robust death induction system, which limits its use on other organisms.

Shortly after succeeding with IDFC, I tried using a two-dye flow cytometry assay (dyeFC). The concept is alluringly simple: take two strains whose communication you want to assay, stain them with different fluorescent dyes, coinoculate a plate with them, and use flow cytometry to quantify how many cells are fluorescent in both channels. Such an assay would require no strain construction, could work on strains from any background, and could be used in many organisms. It would also measure communication more directly than IDFC because it only requires cell-cell adherence, rather than intercellular fusion.

Optionally, the assay can be modified to use a vital dye so both communication and fusion can be simultaneously assessed via dyeFC and IDFC, respectively.

The dyeFC protocol I used is similar to the IDFC protocol described in section 2.4.2, except conidia were dyed with calcofluor white (CFW, also called Fluorescence Brightener 28, Sigma-Aldrich) or concanavalin A, Alexa Fluor™ 488 conjugate (CAF, Invitrogen™) before plating, only PI was used as a vital dye (if any vital dye was used), and CFW and CAF fluorescence areas were recorded using the BUV 496 and FITC cytometer channels, respectively. I used CFW and CAF because they irreversibly stain fungal cell walls and are still detectable after germination, growth, and fusion. However, after my initial tests with dyeFC I deemed it unworkable. The biggest problem was defining fluorescence gates such that cells fluorescing in excess of both gates could be defined as having communicated. Gates that allowed statistical differentiation between positive and negative controls could not be drawn algorithmically.

I returned to the dyeFC assay in an attempt to quantify self-communication rates of a number of mutants without waiting for crosses and screening progeny for DI versions of the mutants. This time I found a metric that tracks communication rates and doesn't rely on fluorescence gates: the relative saddle-to-max probability density ratio (RSM). Like IDFC, dyeFC uses matched conidial controls, but the positive control can be any two strains known to communicate and stained with CFW and CAF, respectively. Conversely, the negative control can consist of any two oppositely stained strains known not to communicate. If one wishes to run simultaneous dyeFC and IDFC assays, standard IDFC positive and negative controls can be used. Two single-dye-alone samples are also needed for an *in silico* negative control, and I typically used the individual strain-dye combinations from the positive control (e.g. FGSC2489 CFW and FGSC2489 CAF if the positive control sample was FGSC2489 CFW mixed with FGSC2489 CAF). The *in silico* control is produced from oppositely stained samples, grown and run through the flow cytometer separately. Data from the two samples is then combined and analyzed to find the minimum theoretical overlap between oppositely stained cell populations. This minimum theoretical overlap may be affected by the choice and quality of fluorescent staining, the flow cytometer's settings, and other uncontrolled environmental variables. Data from at least 20,000 cells were recorded in each experiment, and each experiment was repeated at least three times.

2.4.5. Automated analysis of dye flow cytometry data

After removing ungerminated conidia from fluorescence data as described in step 1 of 2.4.3, I used the following automated workflow to analyze data from each dyeFC experiment:

1. Calculate a probability density (PD) surface for the fluorescence data.
    a. Fit two-dimensional fluorescence data (CFW and CAF) from germinated cells with a bimodal, bivariate Gaussian mixture model (GMM). For the *in silico* negative control, combine fluorescence data from each single dye alone sample before modeling.
    b. Calculate the GMM PD at a million evenly spaced points covering the two-dimensional range of the fluorescence data. This will approximate a PD surface for the data.

2. Find the maximum PD of the GMM.
    a. Search through the PD values calculated in the previous step and record the maximum value.
3. Find the saddle point on the PD surface and record its PD value.
    a. Divide the PD value space into 1000 equal slices and, starting from the bottom, check the two-dimensional (fluorescence space) contour of each slice.
    b. When the contour separates into two ellipses, record the PD value as the saddle PD value.
    c. If no saddle point can be found, record the saddle PD value as the maximum PD value.
4. Calculate the relative saddle-to-max PD ratio (RSM).
    a. Divide the saddle PD value by the max PD value to calculate the saddle-to-max PD ratio.
    b. Divide the saddle-to-max PD ratio by the saddle-to-max PD ratio found for the *in silico* negative control from the same experiment. The resulting number is the RSM.

See figures 2.4.5-1 through 2.4.5-3 for dyeFC control examples. The saddle-to-max PD ratio of the *in silico* negative control represents the minimum communication rate that is theoretically possible for each experiment. Normalizing ratios for other samples to the *in silico* negative control's ratio helps account for non-biological variation between replicates. Once RSMs for three or more replicate experiments are obtained, values are statistically analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Analyzing dyeFC experiments in this way, I can detect only two distinct levels of communication with statistical support at $p = 0.05$ or better: low communication (same as negative control) and high communication (same as positive control). However, I have not used dyeFC to test strain-pairs that communicate at intermediate rates according to IDFC, so I cannot confirm whether dyeFC can statistically differentiate intermediate communication phenotypes. The p-values between positive and negative controls for IDFC and dyeFC experiments suggest IDFC yields greater statistical power (IDFC p-values between controls were usually less than $1x10^{-3}$, dyeFC p-value between controls was $\sim 1x10^{-2}$). But this apparent difference in statistical power may be an artifact, as I performed many replicate IDFC experiments and only one dyeFC experiment in triplicate. In any case, thorough validation of the dyeFC assay is required before comparisons between dyeFC data and data from other communication assays can be confidently compared.

2.5. Induced death flow cytometry assay validation

2.5.1. Extending IDFC's validation from previous work

Although the initial confirmation that *sec-9/plp*-induced VIC death can be quantified using flow cytometry has been published[64], and IDFC protocols using simplified versions of my automated analysis program have also been published[63,65], I wanted to confirm that

IDFC could reproduce previously published communication phenotypes for *doc* mutants[42]. The remainder of section 2.5 reports my experimental validation that IDFC yields similar results to those obtained using microscopy assays.

2.5.2. Materials and methods

Strains, cloning, and growth conditions

Basic protocols for cultivation and manipulation of *N. crassa* can be found on the Neurospora homepage at the FGSC (www.fgsc.net/Neurospora/NeurosporaProtocolGuide.htm). Strains were grown on Vogel's minimal medium[85] (MM) or on Westergaard's synthetic cross medium[86] for crosses. FGSC2489 was the parent of all strains used in this study and served as a CG1 control for all experiments.

Table 2.5.2-1 lists the strains used in experiments discussed in this chapter. FGSC2489, classical mutants, and single gene deletion strains are available through the Fungal Genetics Stock Center (FGSC). Construction details for strains published in Heller et al., 2016, or Heller et al., 2018, can be found therein[42,64]. Strains constructed for this study were produced as follows.

doc alleles were cloned from FGSC2489 into a modified pMF272 plasmid, with *ccg-1* promoter and terminator and C-terminal GFP tag, using *Xba*I and *Pac*I restriction sites[42]. This vector is designed to recombine into *N. crassa*'s *his-3* locus, removing a premature stop codon from an auxotrophic point-mutant and restoring histidine prototrophy[87]. Sequences of primers used for cloning are in Table 2.5.2-2.

For *doc-1^{CGH1}-V5*, the *ccg-1* terminator in pMF272:*Pccg1*:*doc-1^{CGH1}-gfp*:*Tccg1* was swapped with *V5* from pMF272:*Ptef1*:*adv-1-V5*[88] using *Pac*I and *Apa*I to create pMF272:*Pccg1*:*doc-1-V5*. *adv-1-V5* in pMF272:*Ptef1*:*adv-1-V5* was then swapped for *doc-1-V5* from pMF272:*Pccg1*:*doc-1-V5* using *Xba*I to generate pMF272:*Ptef1*:*doc-1-V5*. Next, the *ccg-1* terminator was amplified from pMF272:*Pccg1*:*doc-1-gfp*:*Tccg1* using primers EcoRI-TGA-PsiI-Tccg1-f and ApaI-Tccg1-noEcoRI-r, and cloned into pCR-Blunt-II (Invitrogen™). Finally, the *ccg-1* terminator was subcloned into pMF272:*Ptef1*:*doc-1-V5* using *Eco*RI and *Apa*I to generate pMF272:*Ptef1*:*doc-1-V5*:*Tccg1*. See Fig. 2.5.2-1 for plasmid maps.

Before transformation, all constructs were linearized using *Nde*I and/or *Ssp*I. Prepared conidia of the *his-3* strains FGSC6103 (*mat A*) or FGSC9716 (*mat a*) were then transformed using electroporation following standard protocols. Prototrophic transformants were obtained and backcrossed to His$^-$ versions of FGSC2489, *Δdoc-1*, *Δdoc-2*, or *Δdoc-1 Δdoc-2*, as required. Histidine auxotrophic versions of *doc* deletion strains were generated via backcrosses to FGSC6103 or FGSC9716. The CG3-swap strain (*Δdoc-1 Δdoc-2* with *doc-1^{CGH3}* and *doc-2^{CGH3}* under native promoters expressed at the *his-3* locus) was created by Heller et al., 2016[42]. Other strains expressing CGH3 alleles of *doc-1* and *doc-2* were generated by crossing the CG3-swap strain with His$^-$ versions of FGSC2489, *Δdoc-1*, or *Δdoc-2*. The CG1 DI strain (*Δplp-1 Δplp-2 sec-9^{GRD3}*) was created by Heller et al., 2018[64]. Crossing the CG1 DI strain with the CG3-swap strain and *his-3*; *Δdoc-1 Δdoc-2* produced the CG3 and CG5 DI strains, respectively.

Throughout this chapter, unless otherwise indicated, CG1 and CGH1 will refer to the strain FGSC2489 and its *doc* alleles, CG3 and CGH3 will refer to the strain P4471[89] and its

doc alleles, and CG5 will refer to the *Δdoc-1 Δdoc-2* mutant. All epitope-tagged alleles were integrated into the genome at *his-3*.

Western blots

Protein extractions from germlings and western blots were performed as described in Jonkers et al., 2014[75]. 250 mL Erlenmeyer flasks with 100 mL liquid MM were inoculated to $1x10^6$ conidia per mL. Flasks were incubated at 30°C shaking at 220 rpm for 2.5 hours, then incubated another 2.5 hours at 30°C without shaking. Germlings were harvested by vacuum filtration over a nitrocellulose membrane and frozen in liquid nitrogen. Frozen germlings were then bead beaten with 0.5 mm glass beads at liquid nitrogen temperatures for one minute. 300 µL protein extraction buffer (described in Pandey et al., 2004[90]) was added to each sample and they were bead beaten at room temperature for 15 seconds. Samples were then centrifuged for 30 minutes at 4°C and the protein extract was separated from the cell debris.

Protein concentrations in the extracts were estimated using a NanoDrop™ spectrophotometer (Thermo Fischer Scientific™). Volumes of extracts were adjusted such that all samples run on a single gel contained similar total protein. Extracts were denatured at 70°C in 1x NuPAGE™ LDS sample buffer (Invitrogen™) with 5% β-mercaptoethanol (by volume) for 10 minutes, then run on 7% tris-acetate-SDS polyacrylamide gels (NuPAGE™, Invitrogen™) with PageRuler™ pre-stained protein ladder (10 to 180 kDa, Thermo Fischer Scientific™). Gels were blotted onto PVDF membranes with transfer buffer containing 20 mM tris base, 150 mM glycine, 20% methanol (by volume).

Blots were incubated in TBST (1x TBS (VWR), 0.5% Tween-20 (Sigma Aldrich)) with 5% milk for one hour at room temperature, with a buffer change after 30 minutes. Then blots were probed overnight at 4°C with primary monoclonal antibodies in TBST with 0.5% milk. Blots were washed three times with TBST, and then probed for one hour at room temperature with secondary polyclonal antibodies linked to HRP in TBST. After three more washes with TBST, blots were developed with SuperSignal™ West Pico developer (Thermo Fischer Scientific™) and imaged on a ChemiDoc™ XRS+ with ImageLab™ software (Bio-Rad). See Fig. 2.5.2-2 for western blots showing expression of epitope tagged proteins.

Flow cytometry analyses

Flow cytometry assays and analyses were conducted as described in sections 2.5 and 2.6 of this chapter. In most cases, relative death rates as measured by propidium iodide (PI) and SYTOX Blue (SB) fluorescence agreed, in which case only PI results are shown. In some cases, controls in one channel exhibited anomalously low or variable fluorescence levels, in which case results from the channel in which controls were more statistically separated are shown. Self-communication rates of strains expressing *doc-1-gfp*, *doc-1-V5*, or *doc-2-gfp* were not assessed via IDFC.

2.5.3. Reanalyzing microscopy communication data from Heller et al., 2016

Heller et al., 2016, analyzed the communication phenotypes of FGSC2489, *Δdoc-1*, *Δdoc-2*, *Δdoc-1 Δdoc-2*, and *Δplp-1 Δplp-2 sec-9^{GRD3}* (the CG3-swap strain) using microscopy (Fig. 1.6-2)[42]. However, they analyzed their results using statistical techniques I deemed inappropriate for IDFC data. In order to compare results from their microscopy assays to

those from my IDFC assays, I reanalyzed microscopy data collated from the S1 Data table in Heller et al., 2016, using one-way ANOVA and Tukey-Kramer multiple comparison tests. My analyses highlighted CG phenotype differences between *Δdoc-1* and *Δdoc-2* that weren't explicitly mentioned in our previous publication, and refuted our initial claims that GFP-tagged DOC proteins complement *doc* gene deletions. That particular claim aside, the other results of my analysis of microscopy data from Heller et al., 2016, should be interpreted as extending our published conclusions rather than contradicting them. For easy reference, Table 2.5.3-1 summarizes the results described below.

ANOVA and Tukey-Kramer analyses of microscopy communication data from FGSC2489 supports our published conclusion that the lab strain communicated well with itself and other CG1 isolates, but poorly with strains from all other CGs (CG4 not tested, Fig. 2.5.3-1)[42]. CG1 strains, including FGSC2489, have uniquely exclusive communication preferences. To summarize, microscopy data from Heller et al., 2016, indicates the CG phenotype of FGSC2489 is self^{hi}CG1^{hi}CG2^{lo}CG3^{lo}CG5^{lo}.

My analysis of microscopy communication data from the *Δdoc-1* mutant using one-way ANOVA and Tukey-Kramer multiple comparison tests showed this mutant's self-communication rate decreased to intermediate levels (Fig. 2.5.3-2). Communication between *Δdoc-1* and CG1 or CG2 decreased to levels similar to *Δdoc-1* self-communication, although the mutant communicates better with CG2 than CG1. However, communication between *Δdoc-1* and CG3 was comparable to CG3 self-communication and significantly greater than communication between *Δdoc-1* and CG1 ($p < 0.001$) or *Δdoc-1* self-communication ($p < 1x10^{-4}$). These analyses indicate the *Δdoc-1* mutant communicates preferentially, but not exclusively, with CG3, and deleting *doc-1* impairs self-communication. To summarize, microscopy data from Heller et al., 2016, shows the communication phenotype of the *Δdoc-1* mutant is self^{int}CG1^{int}CG2^{int}CG3^{hi}.

Although Heller et al., 2016, concluded *doc-1-gfp* was "fully functional in restoring communication frequencies in *Δdoc-1*" germlings, my analysis of their data indicates the gfp-tagged allele only partially complements *Δdoc-1* (Fig. 2.5.3-2)[42]. Relative to the *Δdoc-1* mutant, expressing the gfp-tagged allele reduced communication with CG2 and CG3. However the tagged allele failed to fully restore self-communication or communication with CG1. Although my analysis contradicts our original conclusion regarding DOC-1-GFP's ability to complement *Δdoc-1*, this discrepancy does not affect the key insights from our 2016 publication. The primary purpose of expressing the DOC-1-GFP was to observe its subcellular location, and I remain confident the colocalization of DOC-1-GFP with the MAK-2 complex was not an artifact. To summarize, data from Heller et al., 2016, gives *Δdoc-1 doc-1^{CGH1}-gfp* the CG phenotype self^{int}CG1^{int}CG2^{lo}CG3^{int}.

My analysis of microscopy communication data from the *Δdoc-2* mutant showed significant reductions in self-communication and communication with CG1 relative to FGSC2489 (Fig. 2.5.3-3). Although communication between the *Δdoc-2* mutant and CG1 was similar to the mutant's self-communication and significantly greater than communication between the *Δdoc-2* mutant and CG2 or CG3, all tested communication rates for the *Δdoc-2* mutant were statistically reduced compared to wild types. Based on these analyses, the *Δdoc-2* mutant retains a preference for CG1 with global communication deficits. To summarize, microscopy data from Heller et al., 2016, shows the communication phenotype of *Δdoc-2* is self^{int}CG1^{int}CG2^{lo}CG3^{lo}.

ANOVA and Tukey-Kramer tests also contradict the conclusion in Heller et al., 2016, that *doc-2-gfp* fully complements *Δdoc-2* (Fig. 2.5.3-3)[42]. Relative to the *Δdoc-2* mutant, the gfp-tagged allele significantly increased self-communication, but not to wild type levels. Non-self-communication rates were not statistically affected, although communication between the *Δdoc-2 doc-2-gfp* strain and CG1 did increase. As I explained for the *doc-1-gfp* complementation strain, although my analysis refutes the claim that expressing DOC-2-GFP fully complements the deletion of *doc-2*, I remain confident in the biological significance of the peripheral localization Heller et al. observed for DOC-2-GFP. To summarize, data from Heller et al., 2016, indicates the CG phenotype of the *Δdoc-2 doc-2-gfp* strain is self^{int}CG1^{int}CG2^{lo}CG3^{lo}.

Analyzing microscopy data from the *Δdoc-1 Δdoc-2* mutant as I did for each single mutant, I confirmed our original conclusion that strains without a DOC-system specify CG5 (Fig. 2.5.3-4)[42]. Like other CG5 strains, the double mutant didn't communicate with CG1 or CG3, but communicated well with itself and other CG5 strains and moderately with CG2. To summarize, microscopy data from Heller et al., 2016, shows the communication phenotype of the *Δdoc-1 Δdoc-2* mutant is self^{hi}CG1^{lo}CG2^{int}CG3^{lo}CG5^{hi}.

Finally, my analysis supports our original conclusion that the CG3-swap strain phenocopies wild CG3 isolates (Fig. 2.5.3-5)[42]. The CG3-swap strain retained high self-communication, had intermediate communication with CG2, and did not communicate with CG1, the *Δdoc-1 Δdoc-2* mutant, or another CG5 isolate (JW242, data not shown). However, the CG3-swap strain communicated well with CG3. This confirms expressing CGH3 alleles of *doc-1* and *doc-2* converted *Δdoc-1 Δdoc-2* from CG5 to CG3. To summarize, microscopy data from Heller et al., 2016, demonstrates the CG3-swap strain has the CG phenotype self^{hi}CG1^{lo}CG2^{int}CG3^{hi}CG5^{lo}.

The results of my analyses of microscopy communication data from Heller et al., 2016, are summarized in Table 2.5.3-1. Next, I used IDFC to assay the communication phenotypes of the same strains analyzed we previously analyzed via microscopy. My IDFC assay will be validated if it returns results comparable those summarized in Table 2.5.3-1.

2.5.4. IDFC reproduces results from microscopy communication assays

IDFC results described in this chapter are presented in Table 2.5.4 for easy reference. First, I tested FGSC2489's communication phenotype using IDFC (Fig. 2.5.4-1). FGSC2489's self-communication and communication with CG1 were high (in this case, self-communication and communication with CG1 were determined using the same strain-pairing: FGSC2489 + the *sec-9*-swap death-inducing strain). FGSC2489 did not communicate with CG3 or CG5. These results reproduce those obtained by Heller et al., 2016[42]. To summarize, IDFC shows FGSC2489 has the CG phenotype self^{hi}CG1^{hi}CG3^{lo}CG5^{lo}.

While the *Δdoc-1* mutant displayed somewhat reduced self-communication when assayed with IDFC (Fig. 2.5.4-2 panel A), this could not be statistically differentiated from positive or negative controls. Measurements of the mutant's self-communication were reasonably precise, and the loss of statistical resolution was due to positive control variability. Although the *Δdoc-1* mutant's self-communication rate was statistically much closer to the negative control than the positive control (P_a = ~0.8, P_b = ~0.1), I conservatively interpreted this mutant's self-communication rate as intermediate. The *Δdoc-1* mutant communicated well with CG3, but poorly with CG1 and CG5 (communication

between *Δdoc-1* and CG5 was not assayed using microscopy by Heller et al.). These results mostly agree with my analysis of published microscopy data[42]. To summarize, IDFC indicates the *Δdoc-1* mutant has the CG phenotype self^{int}CG1^{lo}CG3^{hi}CG5^{lo}.

IDFC tests of the *Δdoc-1 doc-1-gfp* complementation strain generated by Heller et al., 2016, showed intermediate communication rates with CG1 and CG3, and low communication rates with CG5 (Fig. 2.5.4-2 panel B). These results agree with my analysis of published microscopy data[42], and indicate the expression of DOC-1-GFP partially complements the deletion of *doc-1*. To summarize, the CG phenotype of the *Δdoc-1 doc-1-gfp* strain by IDFC is CG1^{int}CG3^{int}CG5^{lo}.

The *Δdoc-2* mutant exhibited low self-communication when assayed with IDFC (Fig. 2.5.4-3). The mutant did not communicate well with CG3 or CG5, but communication with CG1 was high. These results are similar to those obtained by microscopy, but not identical: the mutant's intermediate levels of self-communication and communication with CG1 by microscopy appeared statistically low and high, respectively, when assayed with IDFC. Communication between the *Δdoc-2* mutant and CG5 was not assayed using microscopy. To summarize, IDFC partially agrees with microscopy results, and measures the *Δdoc-2* mutant's CG phenotype as self^{lo}CG1^{hi}CG3^{lo}CG5^{lo}. I did not assay the *Δdoc-2 doc-2-gfp* strain's communication phenotype using flow cytometry because my later work focused on *doc-1*.

I assayed the communication phenotypes of the *Δdoc-1 Δdoc-2* double mutant in two separate series of IDFC experiments: one series using FGSC2489 + a *Δdoc-1 Δdoc-2* death-inducing strain as a negative control, and another series using FGSC2489 + a CG3-swap death-inducing strain as a negative control (Fig. 2.5.4-4 panels A and B, respectively). Together, these experiments reproduced the self- and non-self-communication phenotypes previously observed via microscopy for the double mutant[42]: deleting both *doc* genes restored self-communication (equivalent to communication with CG5 in this case) and abolished communication with CG1 and CG3. In summary, IDFC and microscopy assays agree that a strain lacking a DOC-system has the CG phenotype self^{hi}CG1^{lo}CG3^{lo}CG5^{hi}.

The final strain I tested to validate the IDFC assay was the CG3-swap strain generated by Heller et al., 2016[42]. The CG3-swap strain did not communicate with CG1 or CG5, but communicated with itself at intermediate to high levels (Fig. 2.5.4-5, slightly different results were obtained for communication between the CG3-swap strain and two different CG3-swap death-inducer strains). Because the CG3 death-inducer strain was produced by crossing the CG3-swap strain with the death-inducing *Δplp-1 Δplp-2 sec-9^{GRD3}* strain made by Heller et al., 2018[64], communication between the CG3-swap and CG3 DI strains can be interpreted as CG3-swap strain self-communication or communication with CG3. These results more-or-less agree with published microscopy data for the CG3-swap strain[42]. To summarize, IDFC assays give the CG3-swap strain a self$^{hi\text{-}int}$CG1^{lo}CG3$^{hi\text{-}int}$CG5^{lo} CG phenotype.

IDFC results from strains tested in this section are summarized in Table 2.5.4. Although IDFC and microscopy assays did not always return identical communication phenotypes for identical strains, communication patterns were very similar. I wondered why DOC-1-GFP and DOC-2-GFP failed to fully complement the deletion of *doc-1* and *doc-2*, respectively, so I tested several potential explanations for why expression of gfp-tagged *doc* alleles only partially restored the CG phenotype of their respective *doc* mutants.

2.5.5. GFP tags partially interfere with DOC functionality

The specificity domain hypothesis, outlined in chapter 1, predicts matched variants of DOC-1 and DOC-2 must interact to properly specify CG. The response domain hypothesis, also outlined in chapter 1, suggests at least DOC-1 must interact with other communication machinery (e.g. the MAK-2 complex) to influence communication behavior. Using the *Δdoc-1 doc-1-gfp* strain as an example, DOC-1-GFP could fail to fully complement the deletion because it cannot properly interact with DOC-2, or because it cannot properly interact with other cellular components, or both. One can distinguish between these possibilities by examining the communication phenotypes of a *Δdoc-1 Δdoc-2* double mutant strain expressing DOC-1-GFP: if the failure of the *Δdoc-1 doc-1-gfp* strain to mimic FGSC2489's CG phenotype is primarily due to aberrant interactions between DOC-1-GFP and DOC-2, a *Δdoc-1 Δdoc-2 doc-1-gfp* strain should phenocopy a *Δdoc-2* mutant. Conversely, if DOC-1-GFP can't properly interact with other communication machinery, a *Δdoc-1 Δdoc-2 doc-1-gfp* strain should behave differently than the *Δdoc-2* mutant. A similar logic applies to DOC-2-GFP: if a *Δdoc-1 Δdoc-2 doc-2-gfp* strain phenocopies the *Δdoc-1* mutant, DOC-2-GFP is probably only deficient in interacting with DOC-1. Otherwise, DOC-2-GFP must be impaired in other ways as well. Results from this section are summarized in Table 2.5.5 for easy reference.

I tested the CG phenotype of a *Δdoc-1 Δdoc-2 doc-1-gfp* strain using IDFC (Fig. 2.5.5-1 panel A). This strain did not communicate with a CG1 death-inducer, and communication rates were intermediate with CG3 and CG5 strains. However, communication rates between the *Δdoc-1 Δdoc-2 doc-1-gfp* strain and CG3 and between the *Δdoc-1 Δdoc-2 doc-1-gfp* strain and CG5 were statistically distinguishable, with CG3 communication nearly low, and CG5 communication nearly high. This CG phenotype does not resemble the behavior of a *Δdoc-2* mutant, implying DOC-1-GFP's ability to interact with other communication machinery must be impaired. Overall, the *Δdoc-1 Δdoc-2 doc-1-gfp* strain has the CG phenotype $CG1^{lo}CG3^{int}CG5^{int}$.

I also used IDFC to assay the CG phenotype of *Δdoc-1 Δdoc-2 doc-2-gfp* (Fig. 2.5.5-1 panel B). However, I could not detect DOC-2-GFP from this strain via western blot (data not shown). The *Δdoc-1 Δdoc-2 doc-2-gfp* strain phenocopied the *Δdoc-1 Δdoc-2* double mutant, communicating well with a CG5 strain, but poorly with strains from CG1 and CG3. Because I couldn't confirm this strain expressed DOC-2-GFP, I cannot tell if the *Δdoc-1 Δdoc-2 doc-2-gfp* strain doesn't phenocopy a *Δdoc-1* mutant because DOC-2-GFP's ability to interact with other cellular machinery is impaired, or if DOC-2-GFP simply isn't expressed in this strain. Given the CG phenotype observed for the *Δdoc-1 Δdoc-2 doc-1-gfp* strain, I suspect the former explanation but cannot be confident. Tentatively, the *Δdoc-1 Δdoc-2 doc-2-gfp* strain has the CG phenotype $CG1^{lo}CG3^{lo}CG5^{hi}$.

Presumably, the failure of DOC-1-GFP to properly interact with other communication machinery was caused by a combination of two factors: over-expression from a non-native promoter at a non-native locus, and/or reduced functionality of DOC-1 due to the GFP-tag. To help determine which of these factors was more important, I used IDFC to phenotype a *Δdoc-1* mutant that over-expressed *doc-1-V5* from the *his-3* locus. If the GFP-tag is primarily responsible for the failure of DOC-1-GFP to fully complement the *doc-1* deletion, the much smaller V5-tag should yield a more functional fusion protein.

However, if over-expression was the main problem, DOC-1-V5 should not complement the deletion better than its GFP-tagged counterpart.

Fig. 2.5.5-2 shows over-expression of *doc-1-V5* fully complemented the non-self-communication defects of the *Δdoc-1* mutant, restoring FGSC2489 behavior: high communication rates with CG1, and low rates with CG3 and CG5. These results indicate the bulky GFP-tag partially interferes with DOC-1's functions, and over-expression is not a major impediment to DOC-1's function. In summary, IDFC tests give the *Δdoc-1 doc-1-V5* strain a CG phenotype of $CG1^{hi}CG3^{lo}CG5^{lo}$.

IDFC results from this section are summarized in Table 2.5.5. I confirmed the GFP-tag reduces DOC-1's ability to influence CG phenotype by interfering with interactions between DOC-1 and other components involved in intercellular communication. Interactions between DOC-1-GFP and DOC-2 may also be impaired, but aberrant interactions between these two proteins are not sufficient to explain the reduced functionality of DOC-1-GFP in a *Δdoc-1 Δdoc-2* genetic background.

2.5.6. Summary and discussion of IDFC validation experiments

Microscopy assay results from Heller et al., 2016, are summarized in Table 2.5.3-1, and IDFC assay results for the same strains are summarized in Table 2.5.4. Comparing these tables, excluding strain-pairs not tested using both assays, IDFC and microscopy assays returned nearly identical communication phenotypes for FGSC2489, the *Δdoc-1* mutant, the *Δdoc-1 doc-1-gfp* complementation strain, the *Δdoc-1 Δdoc-2* double mutant, and the CG3-swap strain. Results differed most between the two assays for the *Δdoc-2* mutant, with IDFC returning a lower self-communication rate and higher communication rate with CG1. The discrepancy in measured self-communication rates between the two assays may be explained by the extreme variability Heller et al. observed for the *Δdoc-2* mutant's self-communication rate (Fig. 2.5.3-3, nine replicates spread evenly between ~12% and ~65% of cells communicating)[42]. High variability in the communication rates observed for mutants may or may not be biologically meaningful, but make statistical comparisons difficult.

However, I suspect discrepancies between microscopy and flow cytometry assays depend more on differences between the media used in the assays. As mentioned in section 2.4, FGSC2489's self-communication rate decreases in liquid cultures, and the MMP media used in flow cytometry communication assays is much less solid than MM agar media used in microscopy assays. This difference in media solidity may disproportionately affect mutants with pre-existing communication defects, although how media viscosity affects germling communication isn't understood well enough for me to hypothesize why the *doc-2* deletion strain was most affected. In any case, IDFC assays mostly reproduced results from published microscopy assays. This validates the IDFC assay's ability to assess the CG phenotypes of *N. crassa* strains with an FGSC2489 genetic background.

I also investigated why over-expressing DOC-1-GFP doesn't fully complement a *doc-1* deletion. DOC-1-GFP must fail to properly interact with cellular components other than DOC-2. In contrast, over-expressing DOC-1-V5 fully complements the non-self-communication defects observed for a *Δdoc-1* mutant, indicating the GFP tag, rather than over-expression, reduces DOC-1 functionality. Unfortunately, when I was planning and performing the experiments I will discuss later in this dissertation, I did not know small

epitope tags could relieve the problems caused by fusing DOC proteins to GFP. I felt epitope tags were necessary to confirm the expression of manipulated *doc* alleles. By the time I found the V5 solution, I did not have time to reconstruct all the strains and repeat all the experiments that were affected by GFP-tags. Thus, this finding is an important limitation to remember when interpreting results in my other chapters.

2.6. Summary and discussion

There are many ways to assay communication between *N. crassa* germlings. In this chapter I've described microscopy assays, forced heterokaryon assays, and flow cytometry assays. Table 2.6 lists the strengths and weaknesses of these methods. I confirmed IDFC and microscopy assays generate comparable statistically differentiable communication phenotypes. But I believe comparable qualitative communication phenotypes can be obtained using any of these assays, although their quantitative measures can't be directly compared. The communication phenotypes I discuss in the remainder of this dissertation were all obtained using flow cytometry (mostly IDFC).

For researchers attempting to analyze germling communication among a small number of strains, a microscopy assay will probably be best because it is versatile, requires no strain construction, and utilizes commonly available media and equipment. For those wanting to characterize communication among many strains, I recommend dyeFC because it is also versatile and requires no strain construction, but permits higher throughput analysis and reduced susceptibility to researcher bias. However, further testing with various permanent fluorescent dyes may improve dyeFC, and more experiments are needed to determine if dyeFC yields sufficient statistical power to differentiate intermediate communication phenotypes. As image analysis software improves, automated time-course microscopy and analysis may outperform flow cytometry-based assays because it can directly quantify chemotropic growth, intercellular contact, fusion, and programmed cell death in a single assay without introducing researcher bias or requiring complicated protocols.

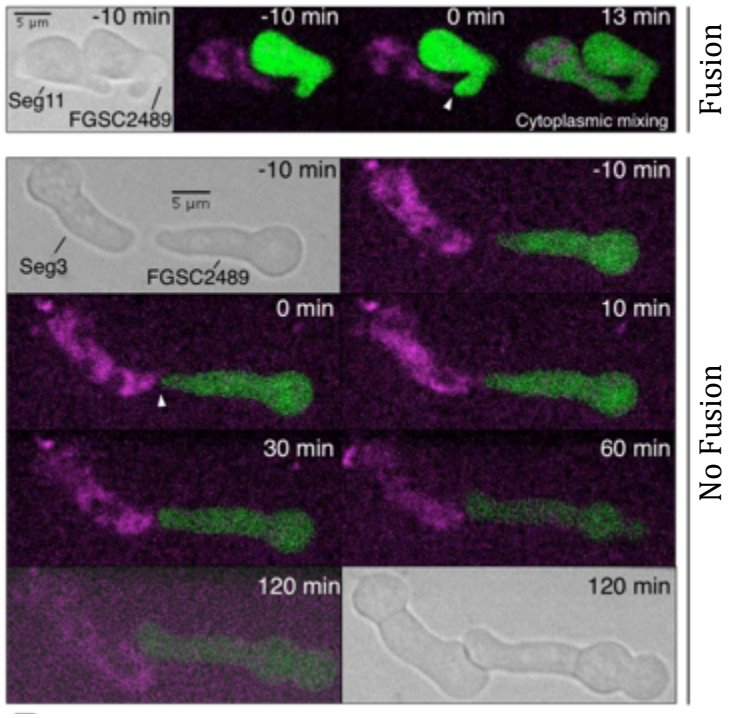


Figure 2.1-1. Fluorescent markers can help identify cell fusion
Microscopy images of *N. crassa* germlings adapted from panel A of "Figure 1. Cell wall-associated arrest is triggered to prevent cell fusion of nonself cells." from Goncalves et al., 2019. One cell from each pair is stained with FM4-64 (magenta) and the other expresses cytoplasmic GFP (green). Images marked "0 min" show the cell contact point with a white arrowhead. The upper series shows the GFP signal flowing into the FM4-64-stained cell 13 minutes after the cells make contact. This demonstrates the cells have fused. In the lower series, even 120 minutes after intercellular contact, the GFP signal is still limited to the cell that produced it, indicating the cells have not fused.

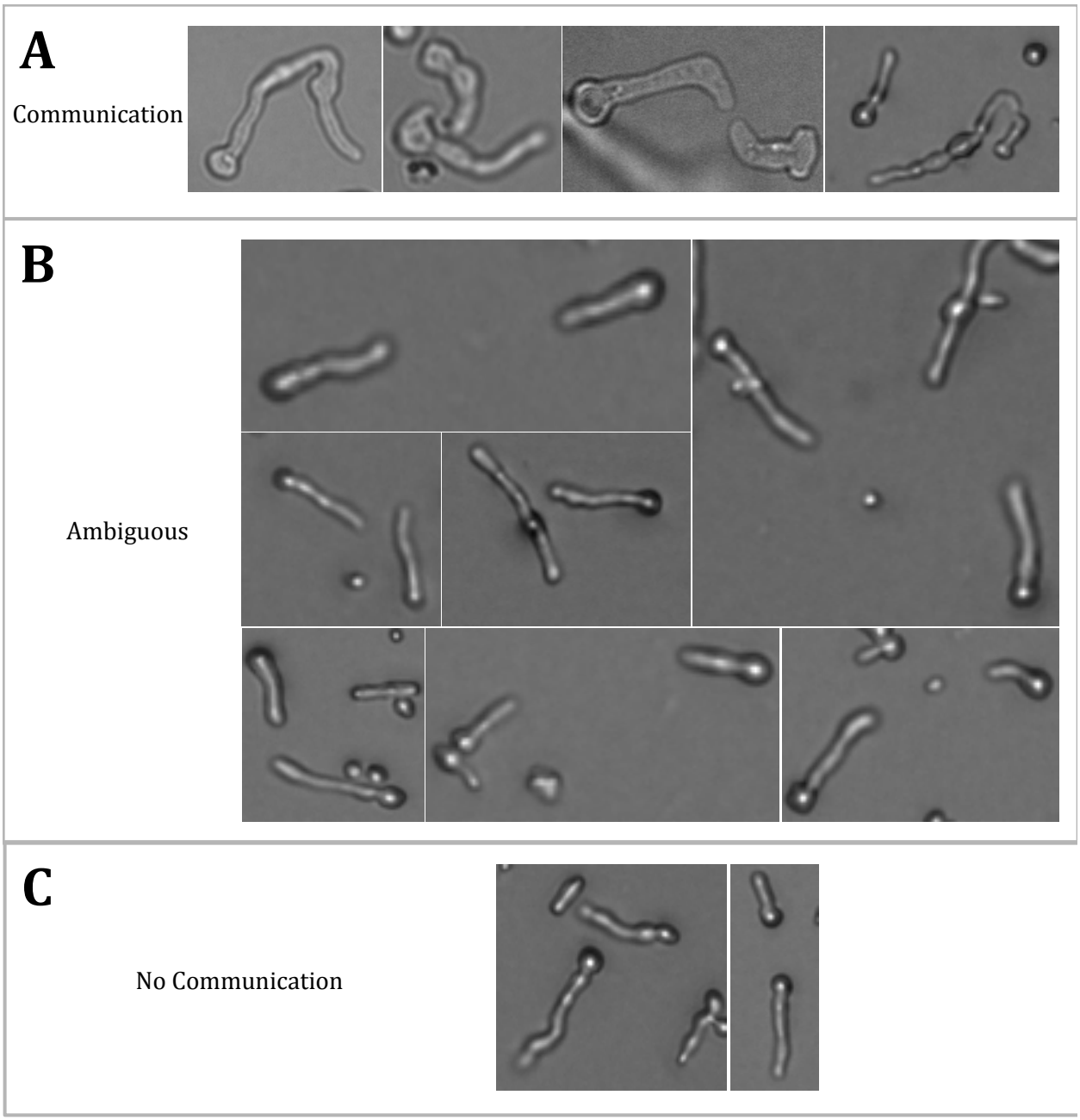


Figure 2.1-2. Communication can be ambiguous in microscopy images
Microscopy images of *N. crassa* germlings. **A)** Images showing obviously communicating germlings. Some pairs have already made contact, and both members of all pairs exhibit dramatic growth reorientation. **B)** Images showing more-or-less ambiguously communicating germlings. Some pairs are growing towards each other, but show no sign of growth reorientation. Other pairs show reorientation that is not as extreme as the pairs in panel A. In some images, one cell is on course to intersect the other, but growth reorientation is not obvious. The upper right image shows three germlings, two of which appear to be reorienting, but it isn't clear which two are communicating. **C)** Images showing germlings that are clearly not communicating. Germlings in these images appear to avoid each other.

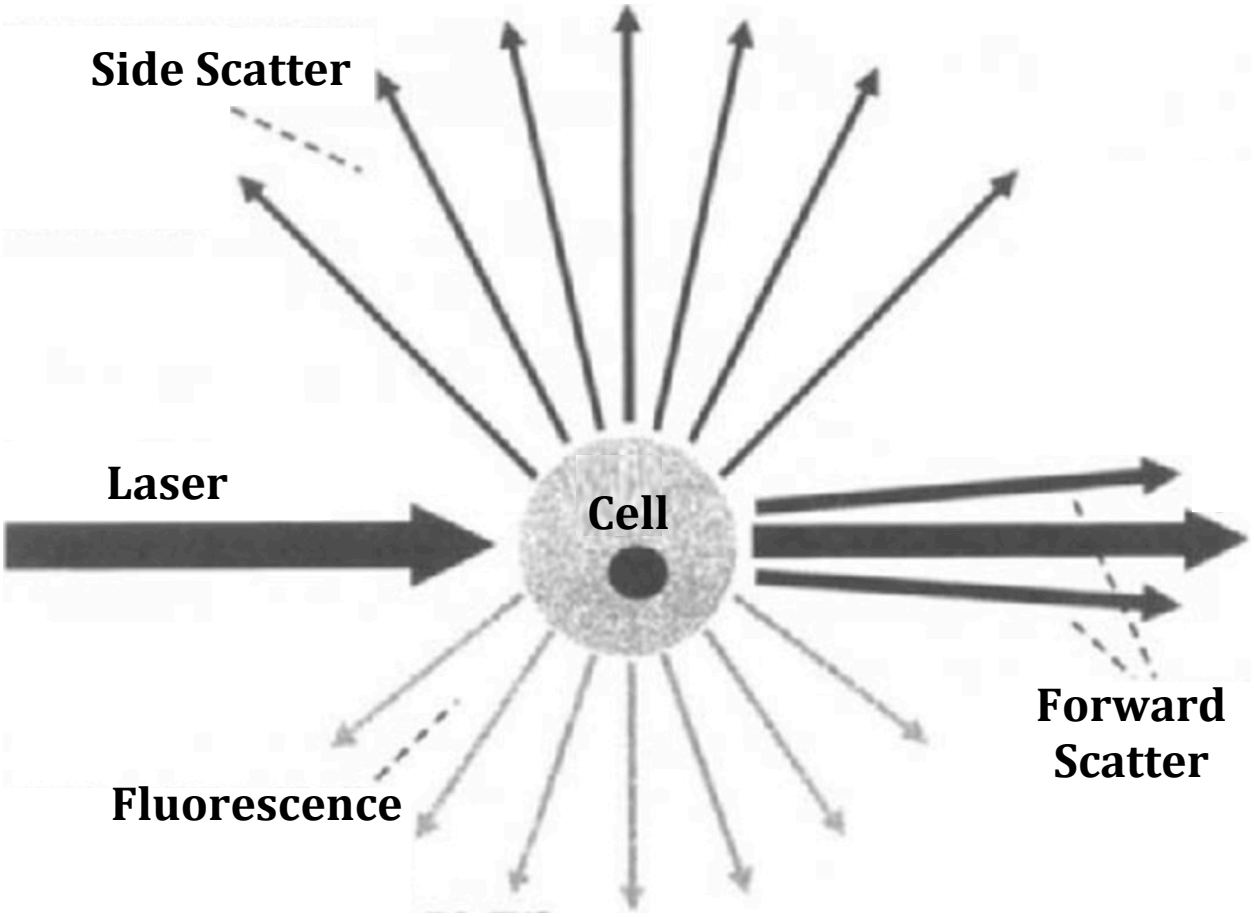


Figure 2.4.1-1. How light interacts with cells during flow cyto
Adapted from "Figure 1-1 Interaction of Light with a Cell" from *Practical Flow Cytometry* by Howard Shapiro. Light scattered through a small angle (~0.5° to 5°) by a cell is detected as forward scatter. Light scattered through a large angle (~15° to 150°) by a cell is detected as side scatter. Fluorescence emitted by a cell is detected by sensors perpendicular to the laser's path. A typical flow cytometer passes cells through a series of laser beams to excite a range of fluorophores. Forward and side scatter are usually only measured from the first laser. Some flow cytometers also record extinction, electrical impedance, and/or bright field images of the cells.

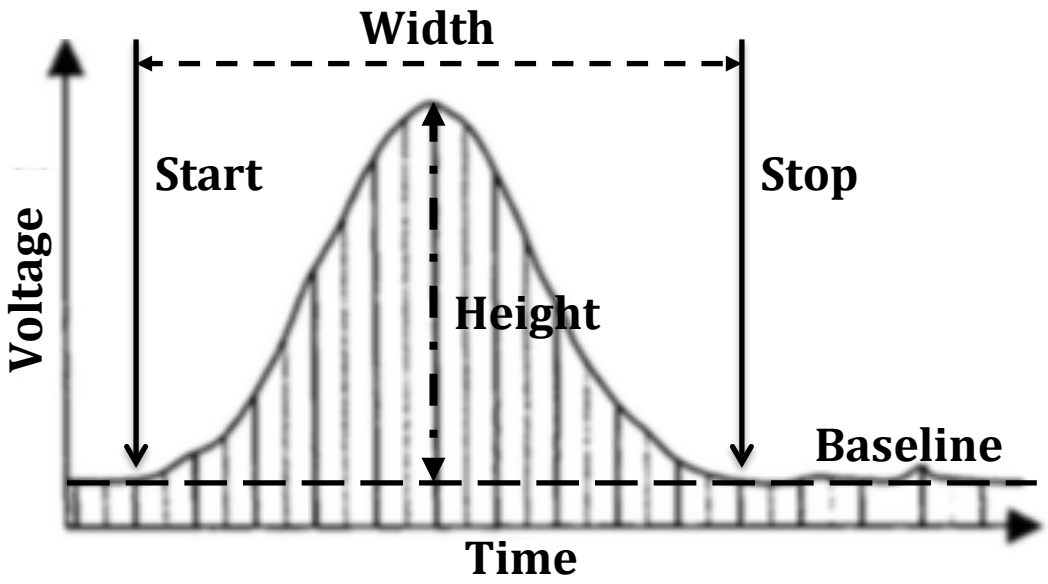


Figure 2.4.1-2. Flow cytometry pulse height, width, and area
Adapted from "Figure 4-59. Digital pulse processing: "slicing" a slightly noisy Gaussian pulse with a baseline." from *Practical Flow Cytometry* by Howard Shapiro. The graph depicts a hypothetical voltage readout from the forward scatter detector of a flow cytometer as a cell passes by. Voltage measurements are taken thousands of times per second (solid vertical lines under the curve). Integrating these voltage measurements from "Start", when the voltage rises above baseline (horizontal dashed line), to "Stop", when the voltage returns to baseline, gives the pulse area. The largest voltage measurement is recorded as the pulse height (vertical dot-dashed double arrow). The time between "Start" and "Stop" is recorded as the pulse width.

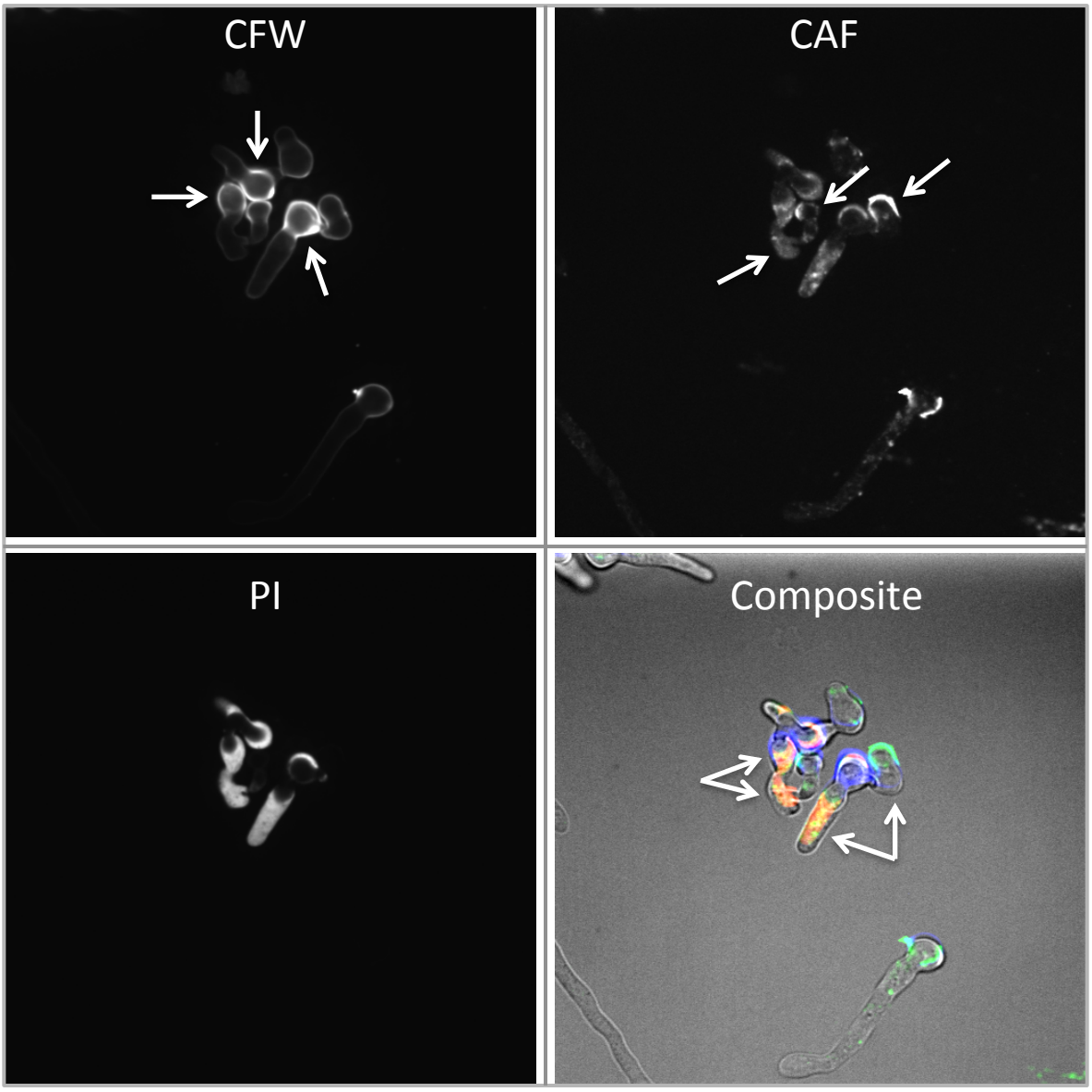


Figure 2.4.2-1. Microscopy images showing what flow cytometry assays detect

Conidia of FGSC2489 were stained with 5 mg/mL calcofluor white (CFW) and conidia of the *sec-9*-swap strain were stained with 150 ug/mL Concanavalin-A Alexafluor™-488 (CAF). These dyes permanently stain the conidial cell walls, but will not stain later growth. After washing, conidia from the two strains were mixed and allowed to grow for 4 hours on MM at 30°C. The plate was then flooded with 1 ug/mL propidium iodide (PI) and imaged. Arrows in the CFW and CAF panels indicate cells stained with those dyes. DyeFC detects these fluorescent cells, and the fluorescence levels of each cell in both channels. Pairs of arrows in the composite panel show fused, dead cells. IDFC detects these dead cells via their PI fluorescence. The unstained cells and germ tubes of stained cells show some fluorescence in the CFW panel, likely due to bleed-over from stained conidia. Unstained cells in the CAF panel also show some fluorescence, but this is auto-fluorescence dead *N. crassa* cells typically exhibit under 488 nm illumination.

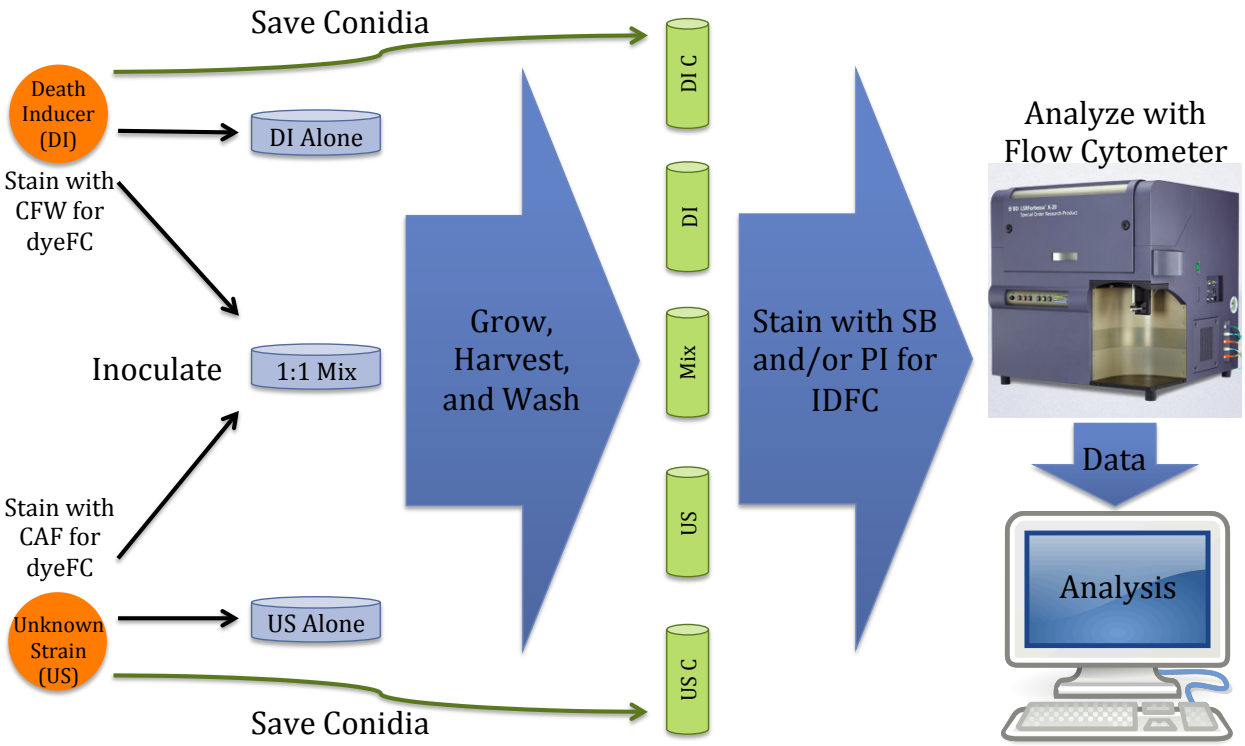


Figure 2.4.2-2. Communication and fusion flow cytometry assay workflow

Diagram depicting the workflow for flow cytometry communication assays. A strain with an unknown communication phenotype (US) is paired with a tester strain of known communication phenotype (DI). For IDFC assays, the DI must induce death upon fusion with the US. For dyeFC assays, conidia (C) of the two strains must be stained with compatible dyes (e.g. CFW and CAF) before inoculating plates. MMP plates are inoculated with each strain alone and with an equal mixture of the two strains. Conidia from each strain are also saved at 4°C for analysis. Plates are incubated for four hours before centrifugal harvesting and washing. For IDFC assays, harvested cells are stained with vital dyes (e.g. SB and/or PI). IDFC and dyeFC assays can be performed simultaneously if compatible fluorophores are used.

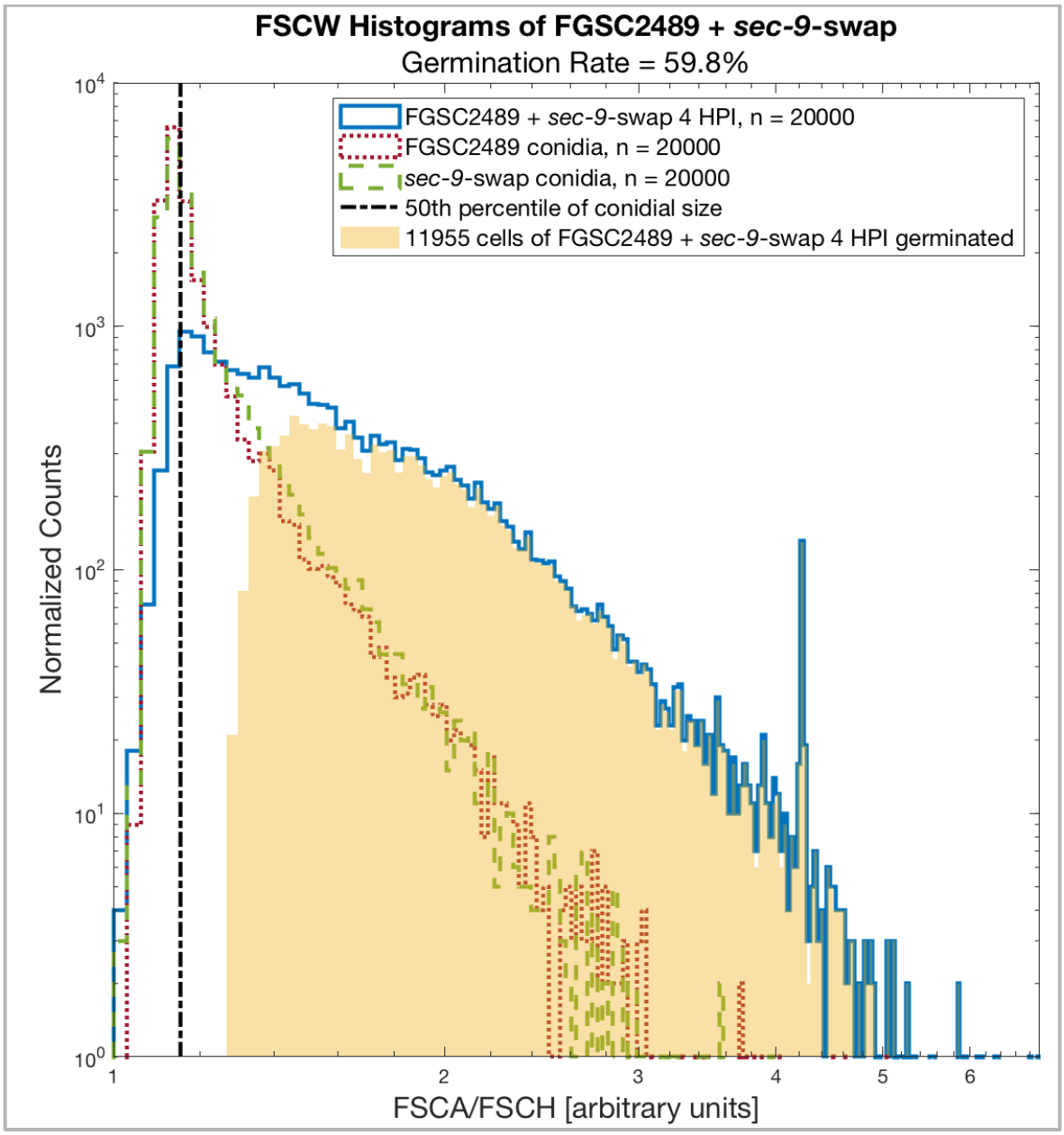


Figure 2.4.3-1. Flow cytometry germination gating

Germination gating figures using Forward Scatter Width (FSCW), calculated for each cell by dividing Forward Scatter Area (FSCA) by Forward Scatter Height (FSCH). FSCW histograms of FGSC2489 conidia (red dotted line), *sec-9*-swap conidia (green dashed line), and a mixture of both conidia grown for 4 hours (blue solid line, as described in the Methods section). For mixtures of strains, 4 HPI cells were selected as germinated based upon their FSCW bin overlap with bins in the mean conidial distribution. The mean conidial FSCW histogram was subtracted from the 4 HPI histogram and divided by the 4 HPI histogram to generate germination ratios for each bin in the 4 HPI histogram. Negative ratios and ratios from bins smaller than the 50th percentile of the mean conidial distribution (black dot-dashed line) were set to zero. Finally, 4 HPI cells from each bin were randomly selected as germinated according to these ratios (yellow bars).

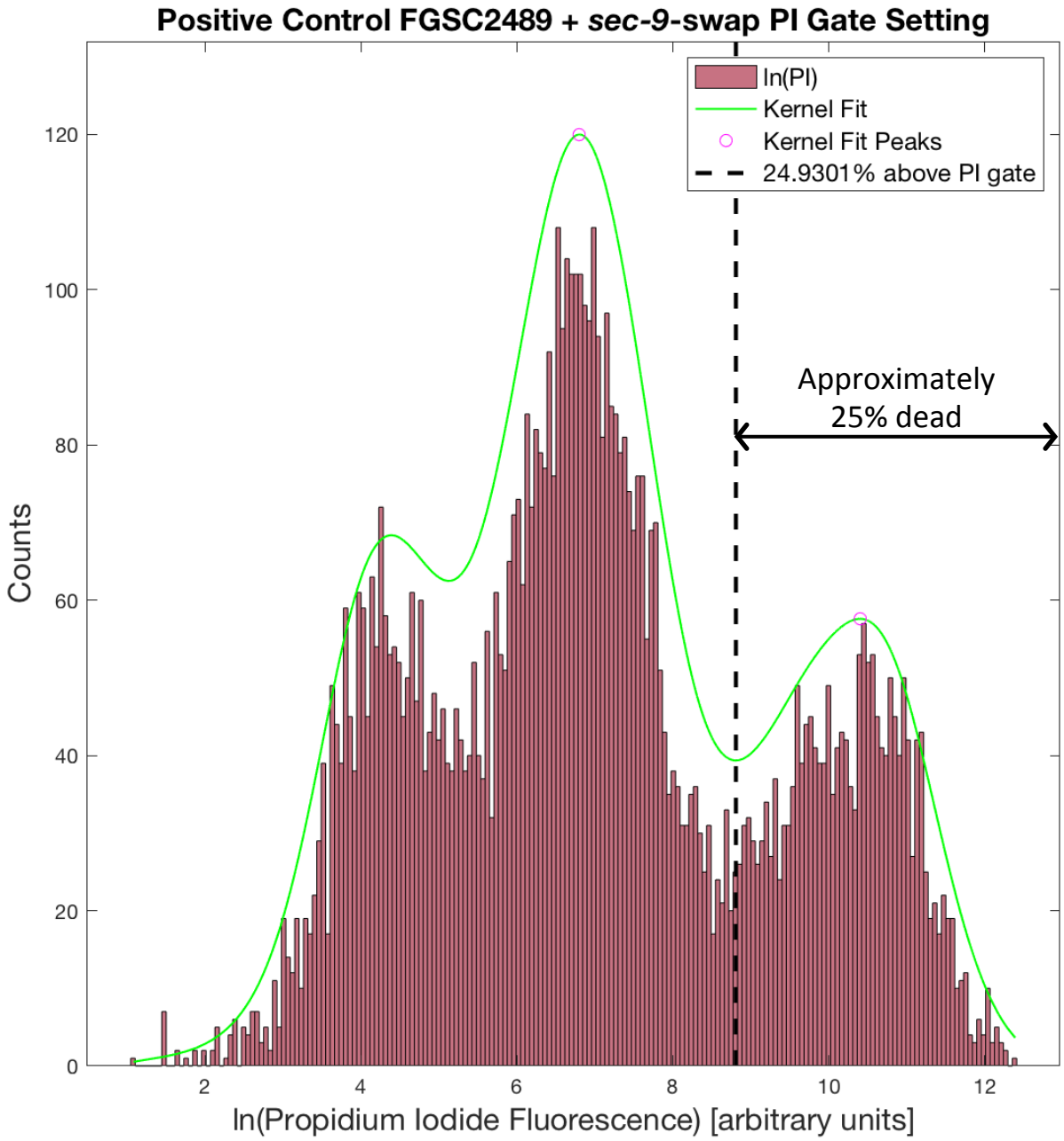


Figure 2.4.3-2. Induced death flow cytometry fluorescence gating
Natural log of Propidium Iodide (PI) fluorescence histogram of FGSC2489 + *sec-9*-swap (red bars). Fluorescence gates were defined based on the distribution of fluorescence values from germinated cells in the positive control sample for each experiment. PI fluorescence data from the positive control was fit with a kernel density estimator using a normal smoothing function and default bandwidth (green solid line). The two most fluorescent local maxima of the resulting kernel fit were identified (magenta circles) and the gate (black dashed line) was set at the PI fluorescence level corresponding to the kernel fit minimum between these local maxima. In this sample, just under 25% of germinated cells are above the PI fluorescence gate and are therefore identified as dead.

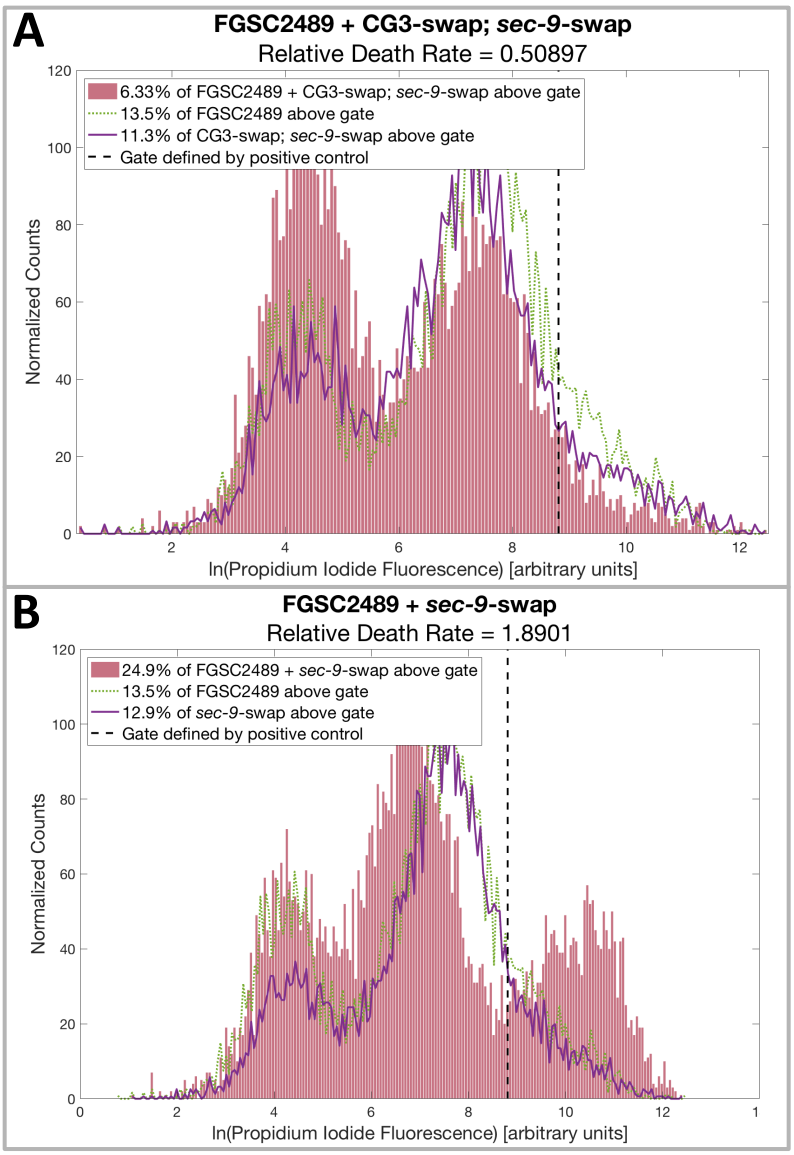


Figure 2.4.3-3. Induced death flow cytometry relative death rates
Induced Death Flow Cytometry negative and positive control examples. Fluorescence gates were defined using the positive control for each experiment (see Fig. 2.8). Relative death rates for strain-death-inducer pairs were calculated as the percentage of dead germinated 4 HPI cells from the mixed sample divided by the average percentage of dead cells in the constituent single strain samples. **A)** Representative negative control figure. 6.33% of the negative control mixture (FGSC2489 + CG3-swap:*sec-9* swap, red bars) is above the fluorescence gate (black dashed line), while only 13.5% and 11.3% of the FGSC2489 (green dotted line) and CG-swap:*sec-9* swap (purple solid line) strains alone are dead, respectively. This gives a relative death rate of 6.33/[(13.5+11.3)/2] = 0.510. **B)** Representative positive control figure. 24.9% of the positive control mixture (FGSC2389 + *sec-9* swap, red bars) is above the fluorescence gate (black dashed line), while only 13.5% and 12.9% of the FGSC2489 (green dotted line) and *sec-9* swap (purple solid line) strains alone are dead, respectively. This gives a relative death rate of 24.9/[(13.5+12.9)/2] = 1.89.

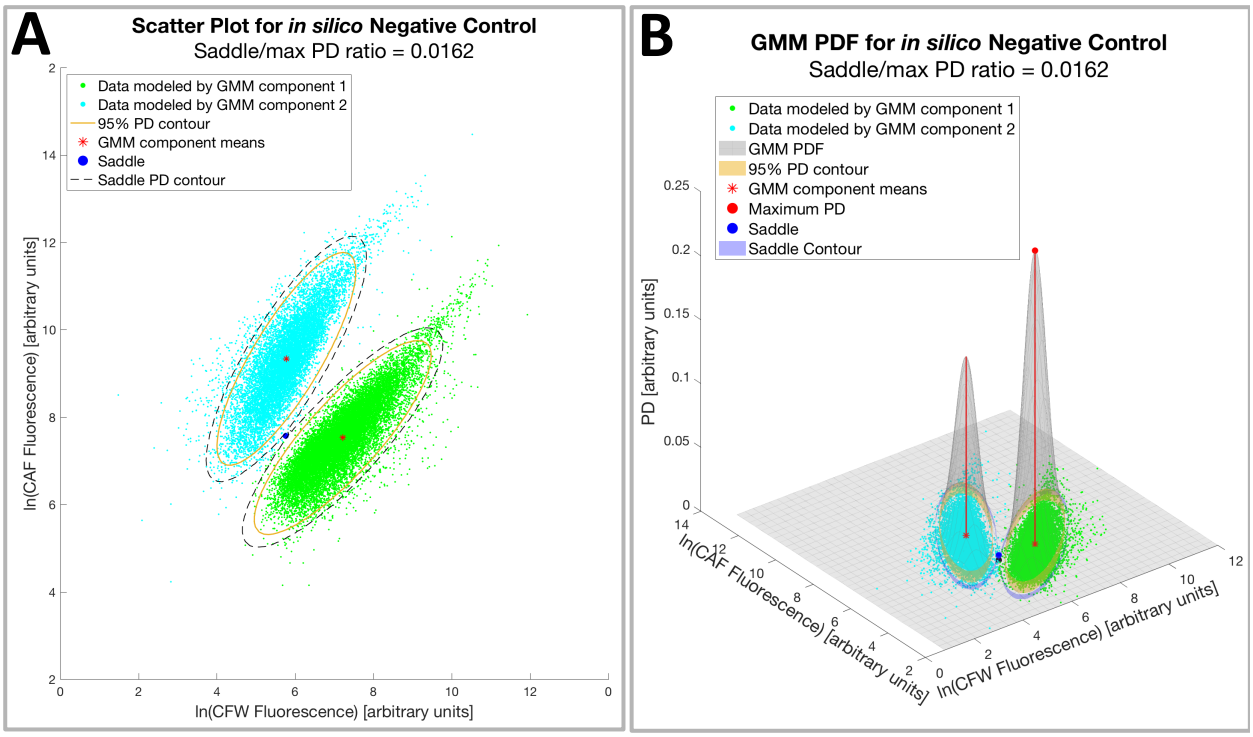


Figure 2.4.5-1. DyeFC *in silico* negative control example

Natural log of CFW versus natural log of CAF fluorescence scatterplots of a dyeFC *in silico* negative control in two and three dimensions. Saddle points and Gaussian mixed model (GMM) probability density function (PDF) contours are indicated. Oppositely stained conidia were grown for 4 hours, and ungerminated cells were excluded from analyses (described in section 2.6). **A)** CFW versus CAF fluorescence scatterplot from an *in silico* negative control. Two samples of FGSC2489 conidia were stained, grown, and analyzed separately and their fluorescence data was combined *in silico*. Because oppositely stained cells were unable to interact in this experiment, this data represents the minimal possible overlap of oppositely stained cell-populations. The combined data was fit with a bimodal Gaussian mixed model and colored green or cyan according to the component Gaussian mode to which the data corresponds. Single-dimension marginal histograms colored to match the modeled data are also displayed. The 95% probability contour (yellow solid line), saddle point (blue filled circle), GMM component means (red asterisks), and saddle point GMM PDF contour (black dashed line) are plotted. **B)** Scatterplot from panel A with GMM PDF (gray mesh), saddle point (blue filled circle), component Gaussian means (red asterisks emitting red lines), max PD (red filled circle), 95% probability contour (yellow vertical surface), and saddle contour (blue vertical surface) displayed in three dimensions. The PDF for this *in silico* negative control consists of two bivariate Gaussian distributions with minimal overlap. Although PD values shown here are arbitrary, the ratio between the PD at the saddle point and the maximum PD of the GMM PDF relates to the degree of overlap between cell populations in the underlying data. For this *in silico* negative control, the saddle/max PD ratio is 0.016, indicating no communication.

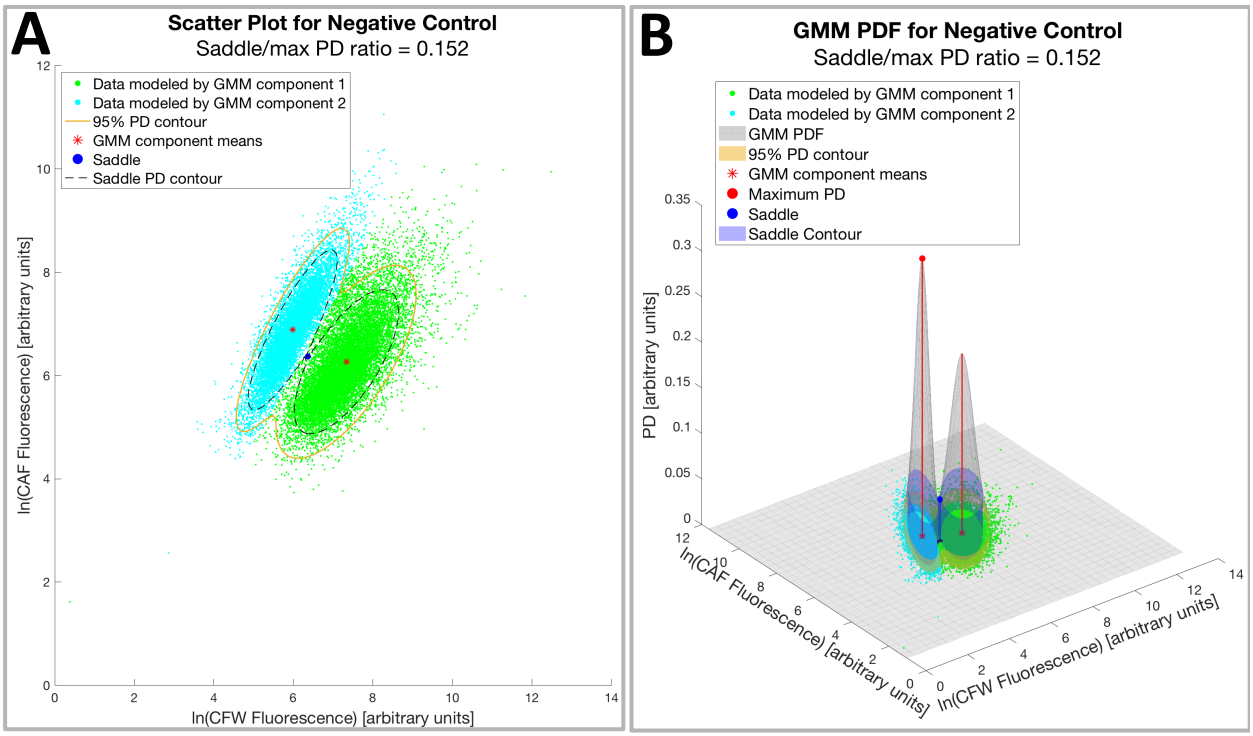


Figure 2.4.5-2. DyeFC negative control example

Natural log of CFW versus natural log of CAF fluorescence scatterplots of a dyeFC negative control in two and three dimensions. Saddle points and Gaussian mixed model (GMM) probability density function (PDF) contours are indicated. Oppositely stained conidia were mixed and grown for 4 hours, and ungerminated cells were excluded from analyses (described in section 2.6). **A)** CFW versus CAF fluorescence scatterplot from a negative control. Samples of FGSC2489 and *Δdoc-1 Δdoc-2 doc-1*$^{\Delta N220}$*-gfp* conidia were stained with CFW and CAF, respectively, and grown and analyzed together. Although these strains do not communicate with each other (see Chapter 4), their populations overlap much more than did those in the *in silico* negative control (Fig. 2.10). The data was fit with a bimodal Gaussian mixed model and colored green or cyan according to the component Gaussian mode to which the data corresponds. The 95% probability contour (yellow solid line), saddle point (blue filled circle), GMM component means (red asterisks), and saddle point GMM PDF contour (black dashed line) are plotted. **B)** Scatterplot from panel A with GMM PDF (gray mesh), saddle point (blue filled circle), component Gaussian means (red asterisks emitting red lines), max PD (red filled circle), 95% probability contour (yellow vertical surface), and saddle contour (blue vertical surface) displayed in three dimensions. Although PD values shown here are arbitrary, the ratio between the PD at the saddle point and the maximum PD of the GMM PDF relates to the degree of overlap between cell populations in the underlying data. For this negative control, the saddle/max PD ratio is 0.152, giving a relative saddle/max PD ratio of 0.152/0.016 = 9.5. This is the background communication rate detected by this assay.

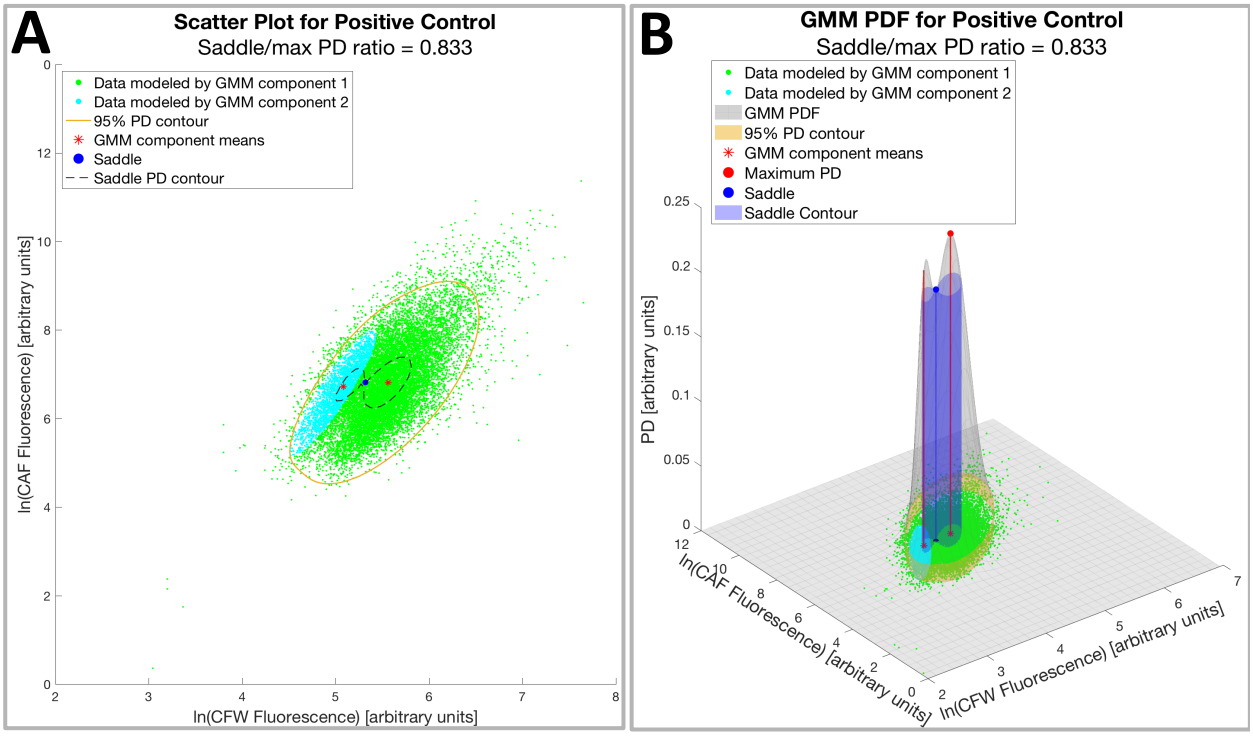


Figure 2.4.5-3. DyeFC positive control example

Natural log of CFW versus natural log of CAF fluorescence scatterplots of a dyeFC positive control in two and three dimensions. Saddle points and Gaussian mixed model (GMM) probability density function (PDF) contours are indicated. Oppositely stained conidia were mixed and grown for 4 hours, and ungerminated cells were excluded from analyses (described in section 2.6). **A)** CFW versus CAF fluorescence scatterplot from a positive control. Separate samples of FGSC2489 conidia were stained with CFW and CAF, then grown and analyzed together. Because these cells are genetically identical and wild-type, they communicate as frequently as possible for the FGSC2489 genetic background. The data was fit with a bimodal Gaussian mixed model and colored green or cyan according to the component Gaussian mode to which the data corresponds. The 95% probability contour (yellow solid line), saddle point (blue filled circle), GMM component means (red asterisks), and saddle point GMM PDF contour (black dashed line) are plotted. **B)** Scatterplot from panel A with GMM PDF (gray mesh), saddle point (blue filled circle), component Gaussian means (red asterisks emitting red lines), max PD (red filled circle), 95% probability contour (yellow vertical surface), and saddle contour (blue vertical surface) displayed in three dimensions. Although PD values shown here are arbitrary, the ratio between the PD at the saddle point and the maximum PD of the GMM PDF relates to the degree of overlap between cell populations in the underlying data. The saddle/max PD ratio for this positive control is 0.833, giving a relative saddle/max PD ratio of 0.833/0.016 = 52.1.

Table 2.5.2-1: Strains used in chapter 2		
Wild isolates and classical mutants		
Strain	**Genotype**	**Reference**
FGSC2489 (CG1)	Laboratory Wild Type, *mat A*	FGSC
P4471 (CG3)	Wild isolate from the Louisiana Population	Bhat & Vyas, 2003
FGSC4200	Laboratory Wild Type, *mat a*	FGSC
FGSC6103	Laboratory Wild Type, *his-3 mat A*	FGSC
FGSC9716	Laboratory Wild Type, *his-3 mat a*	FGSC
Manipulated strains (all in FGSC2489 genetic background, all deletions marked with hygR)		
Strain	**Genotype**	**Reference**
Δdoc-1	*Δdoc-1 mat a*	Dunlap et. al., 2007, & Heller et. al., 2016
Δdoc-1 his- A	*his-3; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 his- a	*his-3; Δdoc-1 mat a*	This study
Δdoc-1 Δdoc-2 (CG5)	*Δdoc-1 Δdoc-2 mat A*	Heller et. al., 2016
Δdoc-1 Δdoc-2 his- A	*his-3; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 his- a	*his-3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG3-swap strain	*his-3:: doc-1-CGH3 doc-2-CGH3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG1 DI strain	*his-3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat A*	Heller et. al., 2018
CG3 DI strain	*his-3:: doc-1-CGH3 doc-2-CGH3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat ?*	This study
CG5 DI strain	*Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+; Δdoc-1 Δdoc-2 mat ?*	This study
Δdoc-1 Δdoc-2 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 doc-1-CGH1-V5	*his-3:: doc-1-CGH1-V5; Δdoc-1 mat ?*	This study

Table 2.5.2-2: Primers used for cloning in chapter 2		
Name	**Sequence (5'-3')**	**Purpose**
EcoR1-TGA-Psi1-Tccg1-f	GAATTCTAATTATAAGCGACTTTACCAACAGTC	Forward primer to amplify the *ccg-1* terminator, adding EcoRI and PsiI sites and a stop codon
Apa1-Tccg1-noEcoR1-r	TATAGGGCCCAAGCTTGATATCGTATTCGC	Reverse primer to amplify the *ccg-1* terminator, adding an ApaI site and deleting an EcoRI site

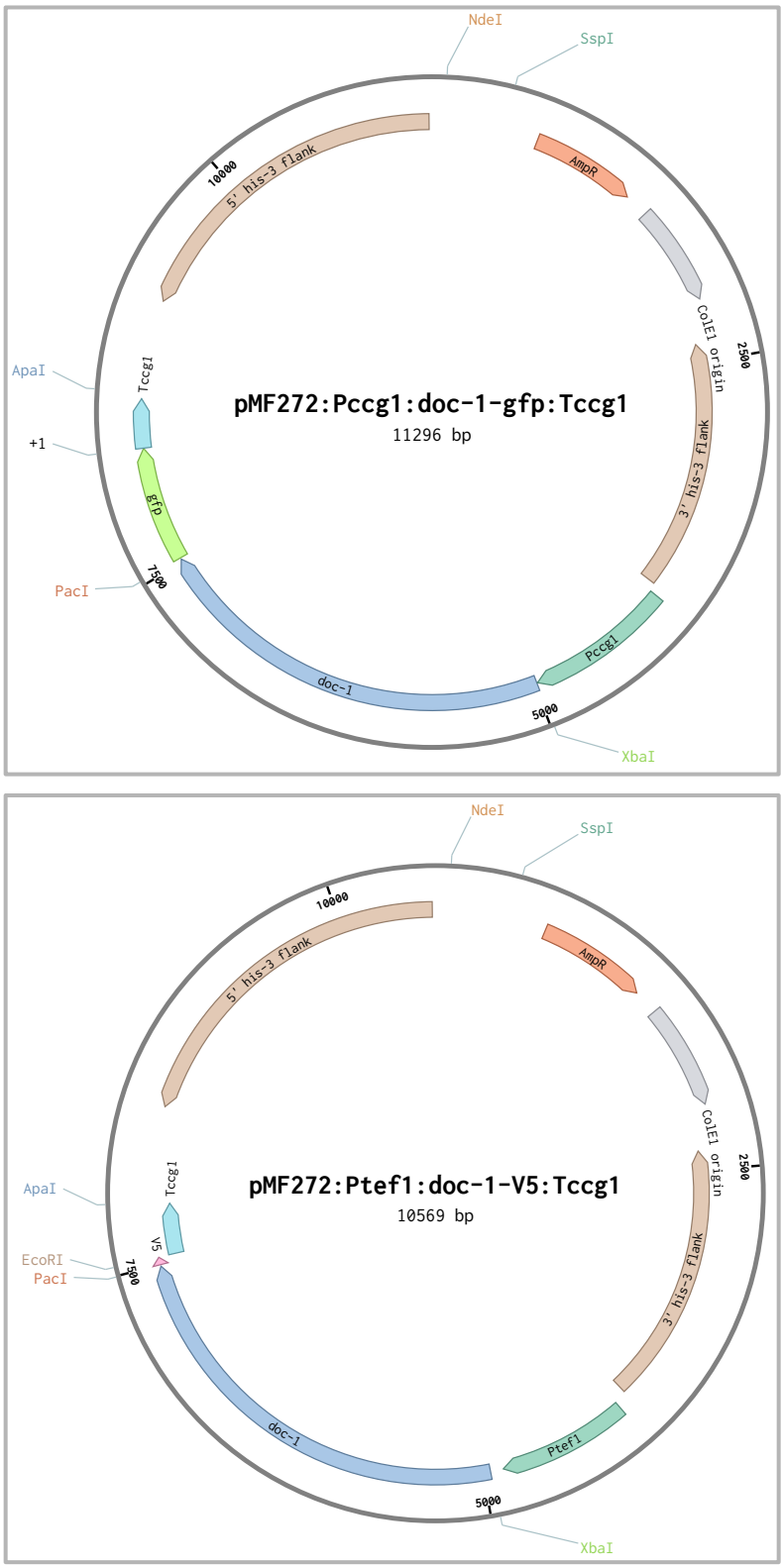


Figure 2.5.2-1. Plasmid maps of vectors used in this publication
Vectors are described in chapter 2, and depicted with *doc-1*^{CGH1} cloned into them. Other alleles were cloned into them using *Xba*I and *Pac*I restriction sites.

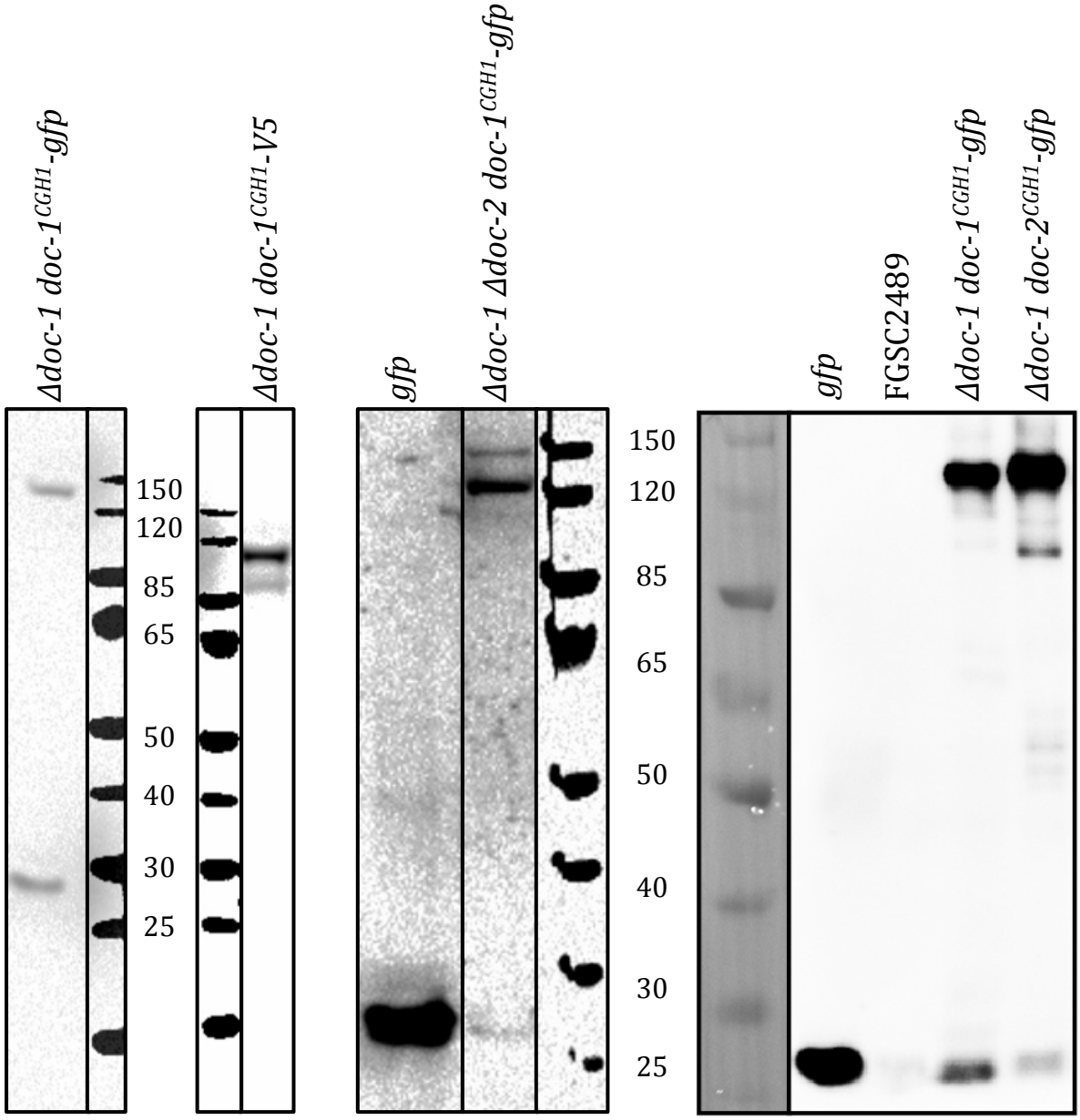


Figure 2.5.2-2. Western blots showing expression of GFP and V5-tagged DOC proteins

Western blots showing detection of GFP and V5-tagged DOC-1 from FGSC2489 in *Δdoc-1* and *Δdoc-1 Δdoc-2* genetic backgrounds. Samples were loaded to normalize total protein content across the lanes in each blot. See section 3.2 for more details. For clarity, separate blots are outlined in gray boxes. Expected band sizes: DOC-1^{CGH1}-GFP = 118839.1 Da, DOC-1^{CGH1}-V5= 93285.4 Da, DOC-2^{CGH1}-GFP = 118903.2 Da, free GFP = 26975.4 Da. Numbers indicate marker bands' kDa sizes. Right-most blot taken from panel C of "S5 Fig. Oscillation dynamics of DOC-1-GFP in hyphae and germlings" from Heller et al., 2016. DOC-2^{CGH1}-GFP could not be detected in the *Δdoc-1 Δdoc-2* genetic background.

Table 2.5.3-1. Summary of CG phenotypes by microscopy from section 2.5.3

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/ CG5	Comm w/ CG3	Comm w/ CG2	Comm w/ CG1	Self-comm
FGSC2489					
Δdoc-1	Not tested				Lower
Δdoc-1 doc-1-gfp	Not tested	Lower			Higher
Δdoc-2	Not tested				Lower
Δdoc-1 doc-2-gfp	Not tested				Higher
Δdoc-1 Δdoc-2					
CG3-swap strain					

Data from Heller et al., 2016. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483. CG5 = JW220 for assays of FGSC2489 and *Δdoc-1 Δdoc-2*, and CG5 = *Δdoc-1 Δdoc-2* for assays of the CG3-swap strain. For self-communication comparisons, high communication was defined as equivalent to FGSC2489's self-communication rate. For non-self-communication comparisons, high communication was defined as equivalent to the self-communication rate of the tester strain (e.g. the CG3-swap strain exhibited high communication with CG3 because it communicated with the CG3 tester strain as well as the CG3 tester strain communicated with itself). For all comparisons, low communication was defined as equivalent to the communication rate between CG1 and CG3, and intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of 0.05 or better. "Lower" or "Higher" are written in some boxes to compare deletions with complementation strains. Boxes of the same color may be statistically different with $p<0.05$: see p-value tables for all pairwise comparisons.

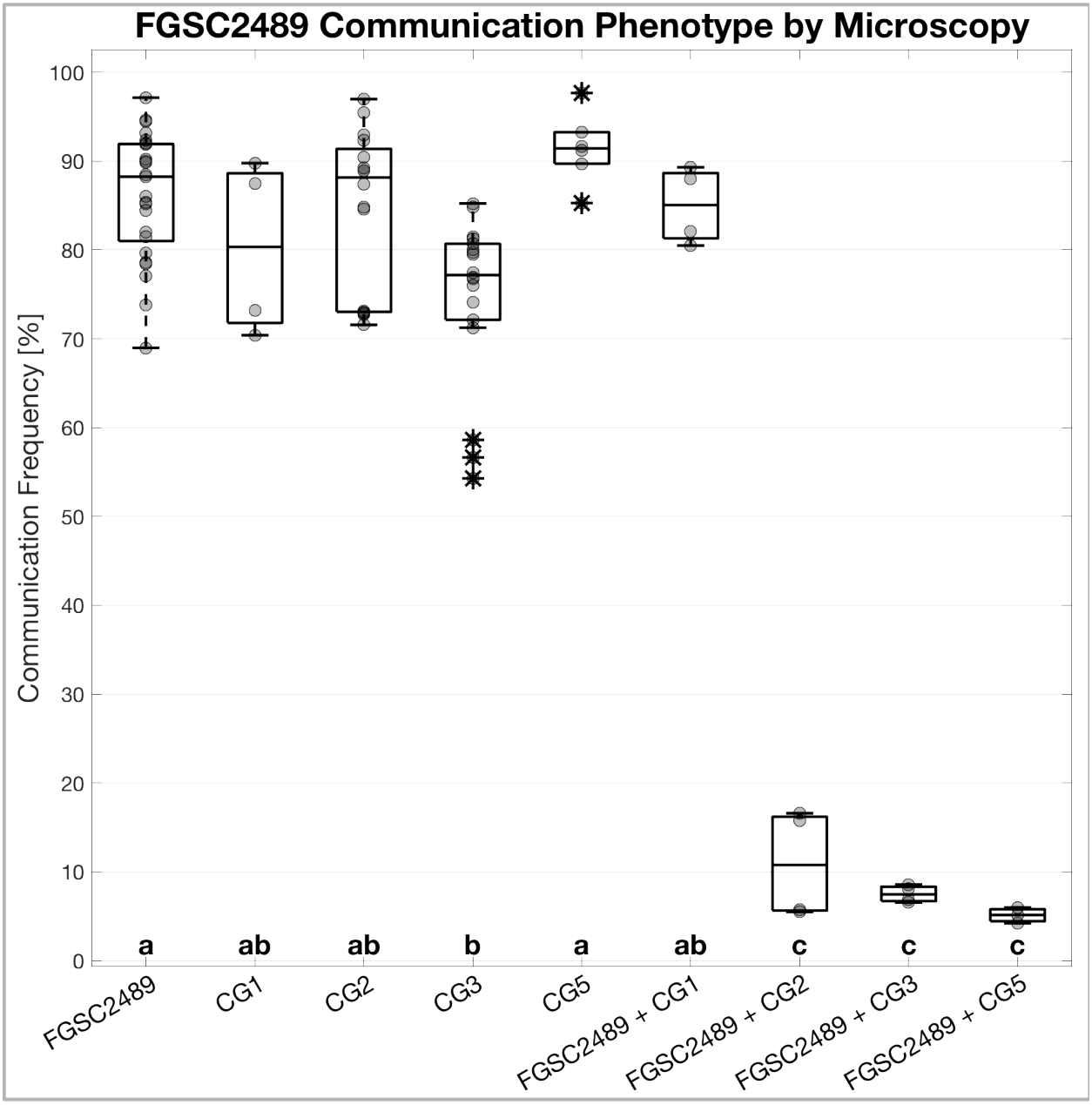


Figure 2.5.3-1. Microscopy communication assay results from FGSC2489 Boxplot summary of communication data from FGSC2489, collated from S1 Data in Heller et al., 2016. Microscopy communication assay data from all relevant strains were compiled and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05. CG1 = JW184, CG2 = JW262, CG3 = P4483, CG5 = JW220. Number of replicates vary between 25 and 3.

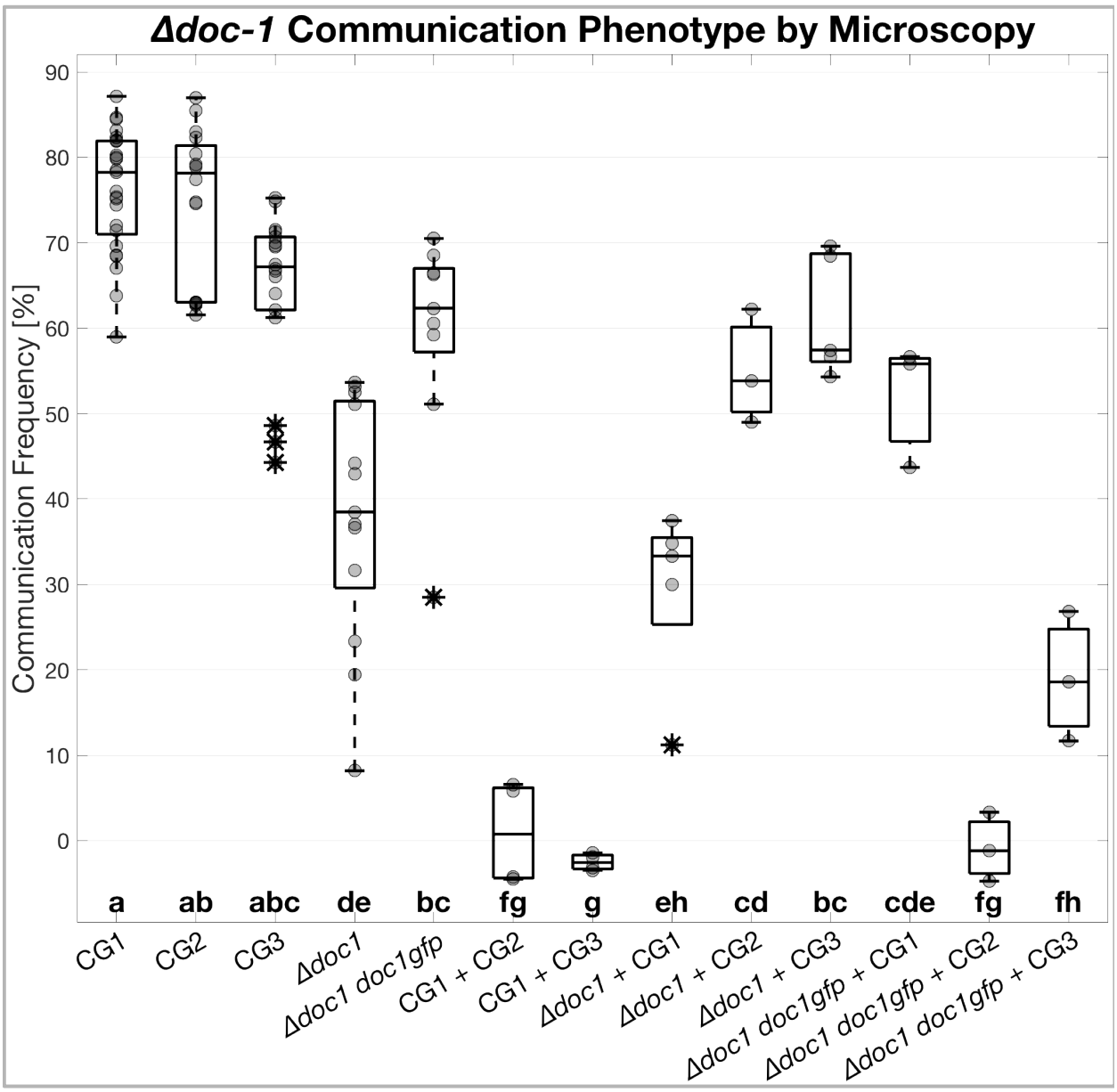


Figure 2.5.3-2. Microscopy communication results from *Δdoc-1* and *Δdoc-1 doc-1-gfp* strains

Boxplot summary of communication data from *Δdoc-1* strains, collated from S1 Data in Heller et al., 2016. Microscopy communication assay data from all relevant strains were compiled and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483, CG5 = JW220. Number of replicates vary between 25 and 3.

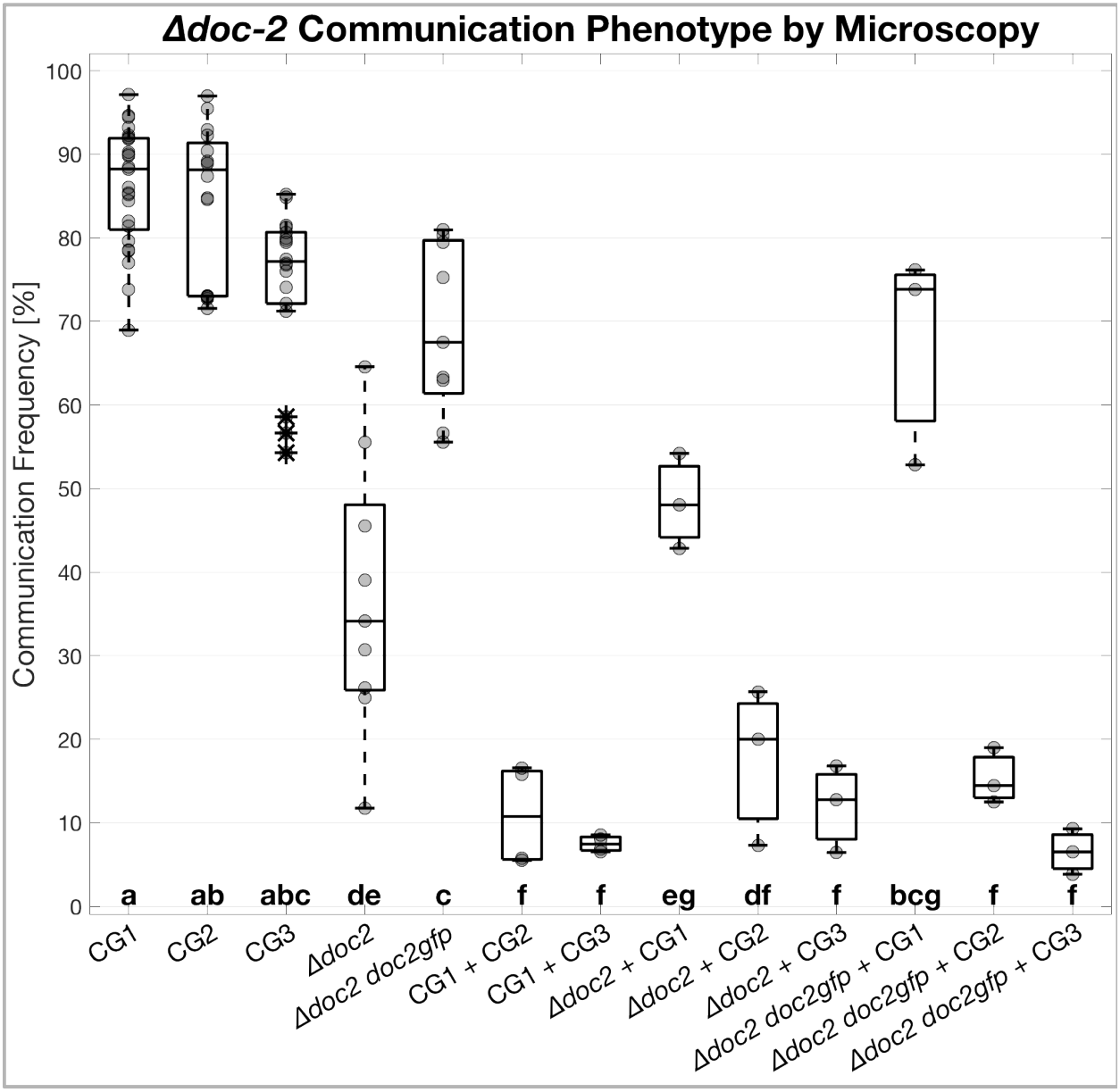


Figure 2.5.3-3. Microscopy communication results from *Δdoc-2* and *Δdoc-2 doc-2-gfp* strains

Boxplot summary of communication data from *Δdoc-2* strains collated from S1 Data in Heller et al., 2016. Microscopy communication assay data from all relevant strains were compiled and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483, CG5 = JW220. Number of replicates vary between 25 and 3.

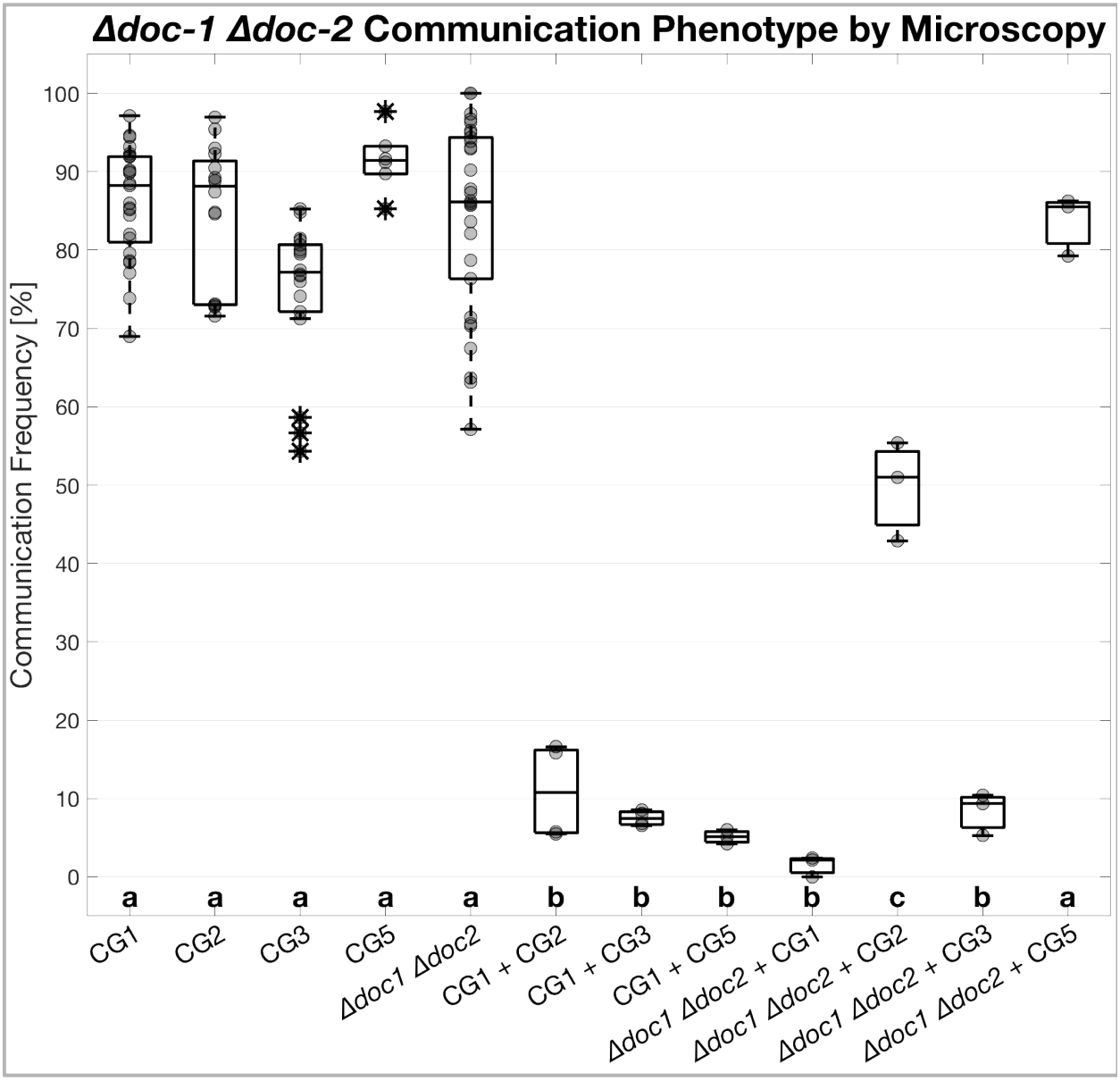


Figure 2.5.3-4. Microscopy communication results from *Δdoc-1 Δdoc-2*
Boxplot summary of communication data from *Δdoc-1 Δdoc-2*, collated from S1 Data in Heller et al., 2016. Microscopy communication assay data from all relevant strains were compiled and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483, CG5 = JW220. Number of replicates vary between 30 and 3.

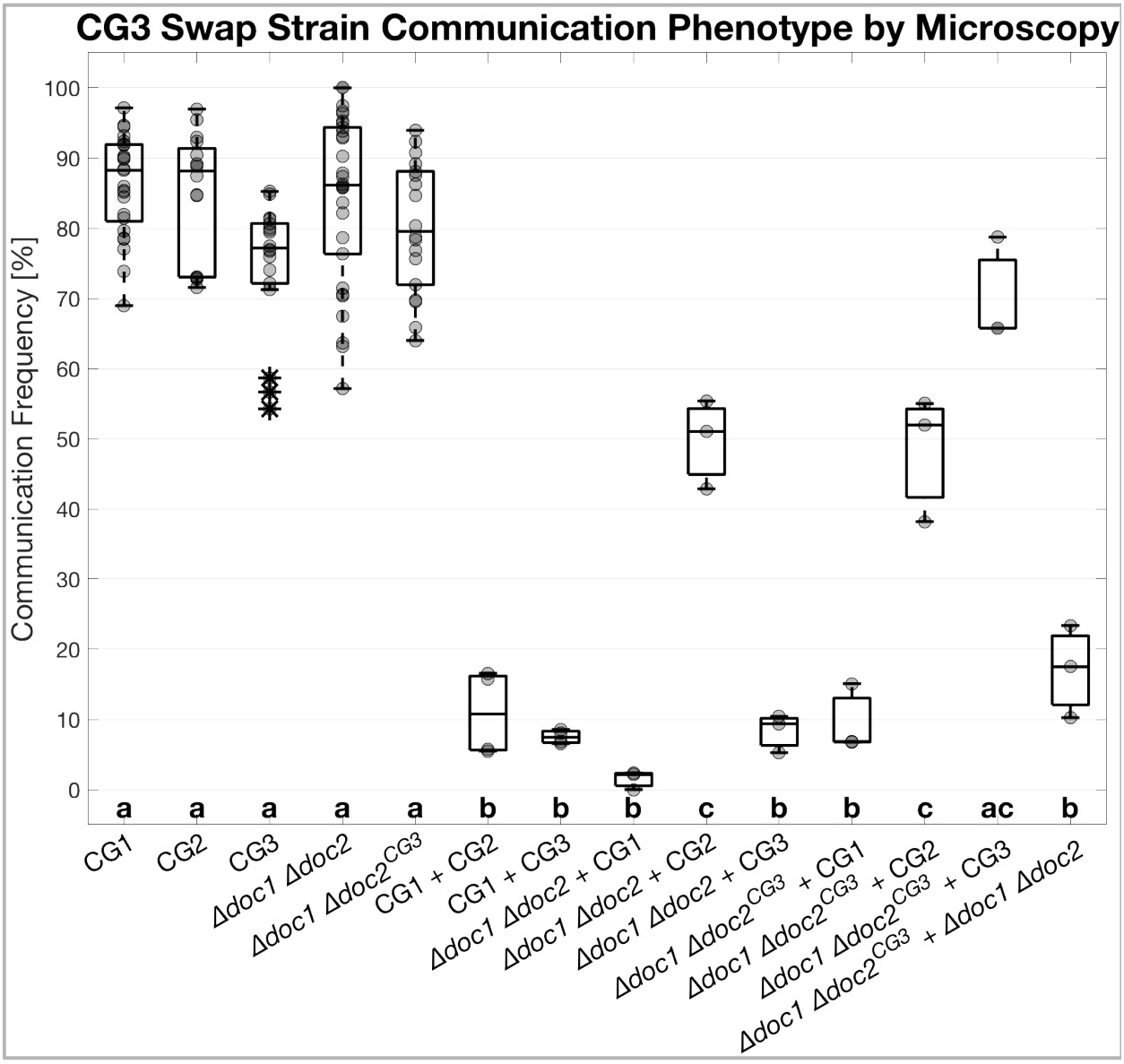


Figure 2.5.3-5. Microscopy communication assay results from the CG3-swap strain

Boxplot summary of communication data from the CG3-swap strain, collated from S1 Data in Heller et al., 2016. Microscopy communication assay data from all relevant strains were compiled and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05. CG1 = FGSC2489, CG2 = JW262, CG3 = P4483. Number of replicates vary between 30 and 3.

Table 2.5.4. Summary of CG phenotypes by IDFC from section 2.5.4

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1	Self-comm
FGSC2489				
Δdoc-1				Interpreted as intermediate
Δdoc-1 doc-1-gfp				Not tested
Δdoc-2				
Δdoc-1 Δdoc-2				
CG3-swap strain		*		*

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. Self-communication measurements of the *Δdoc-1* mutant could not be distinguished from either control due to variability in positive control measurements, but appeared to be low; I interpreted the *Δdoc-1* mutant's self-communication rate as intermediate. *The CG3-swap strain exhibited high communication with one CG3 DI backcross, and intermediate communication with another.

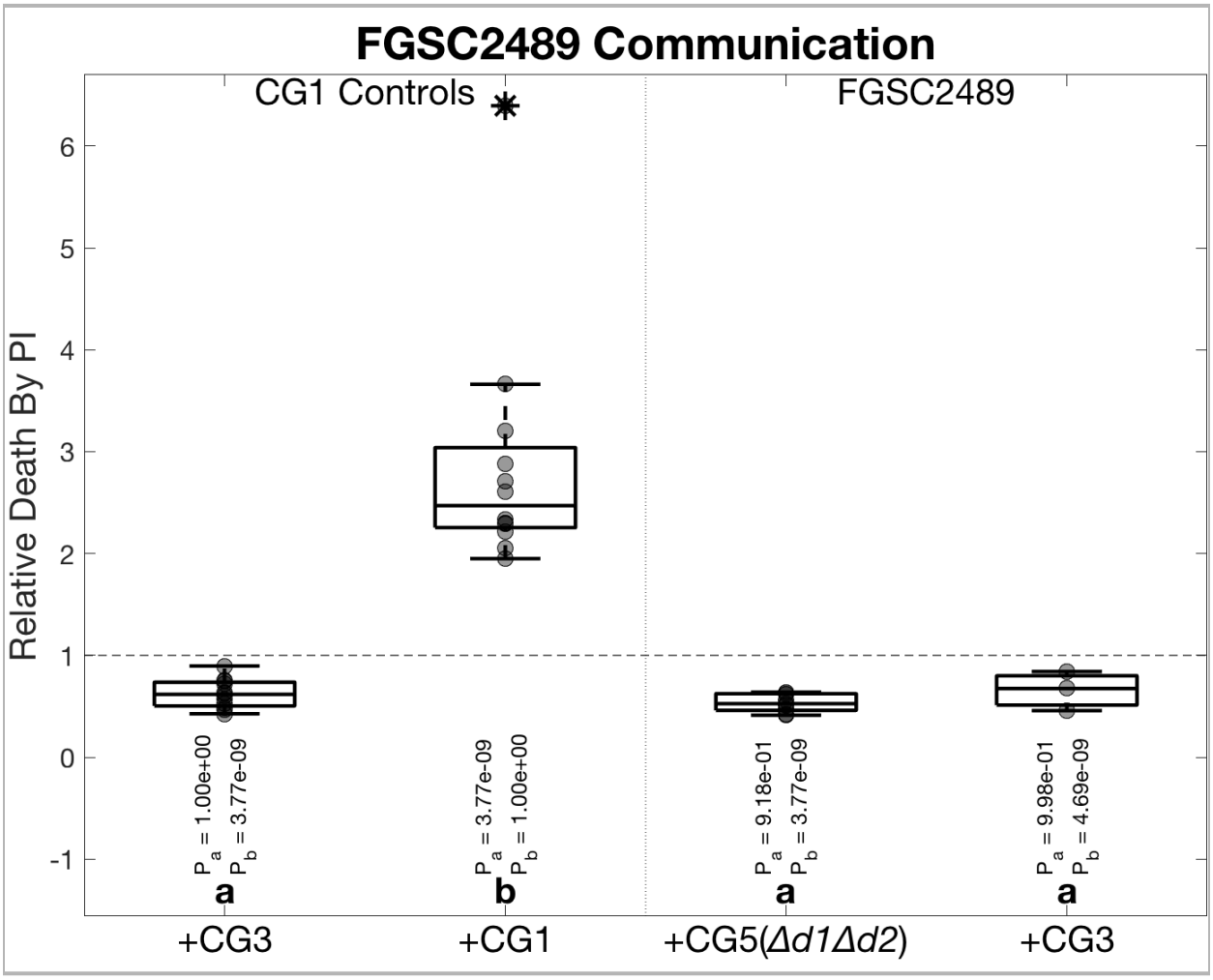


Figure 2.5.4-1. IDFC results from FGSC2489

Boxplot summary of communication phenotypes of FGSC2489 analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (outside 99% of replicate data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Number of replicates varies between 12 and 3 for different strain-pairings.

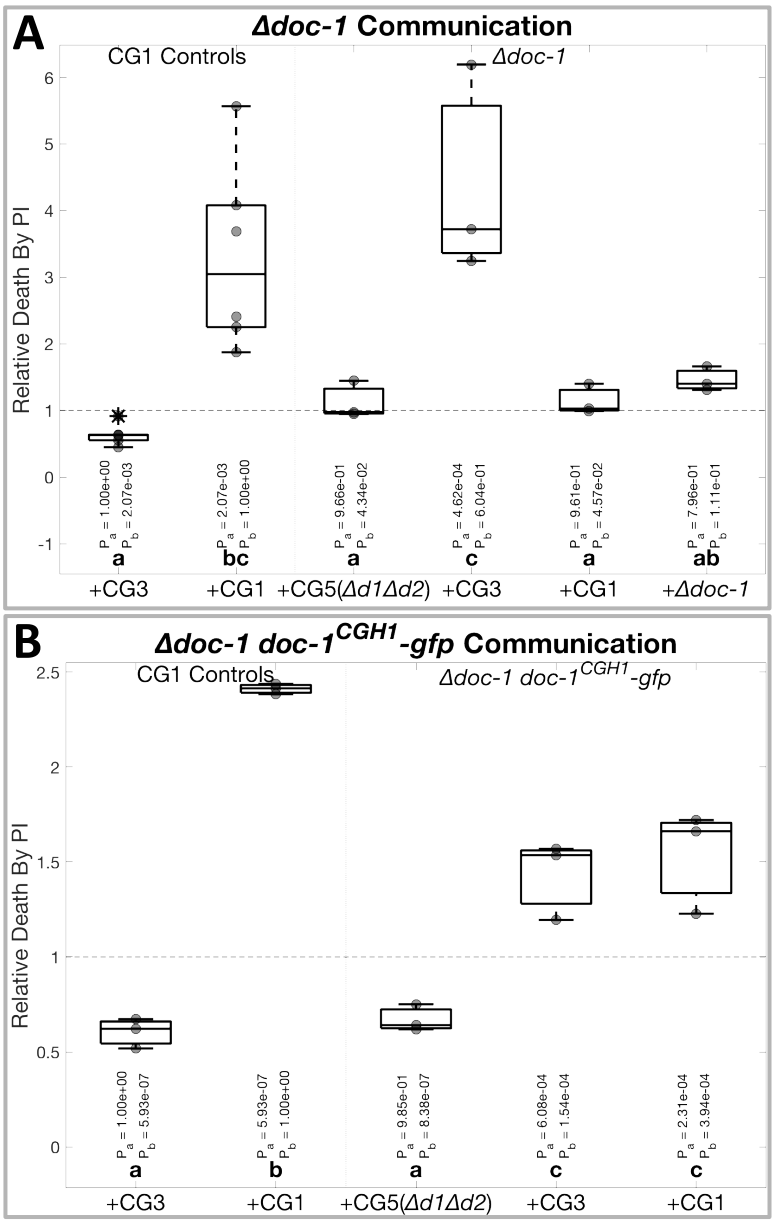


Figure 2.5.4-2. IDFC results for *Δdoc-1* and *Δdoc-1 doc-1-gfp* strains
Communication phenotypes of a *Δdoc-1* mutant and *Δdoc-1 doc-1-gfp* complementation strain analyzed with one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1*. Number of replicates vary between 7 and 3 for different pairs of strains. **B)** Triplicate IDFC results for *Δdoc-1 doc-1-gfp*.

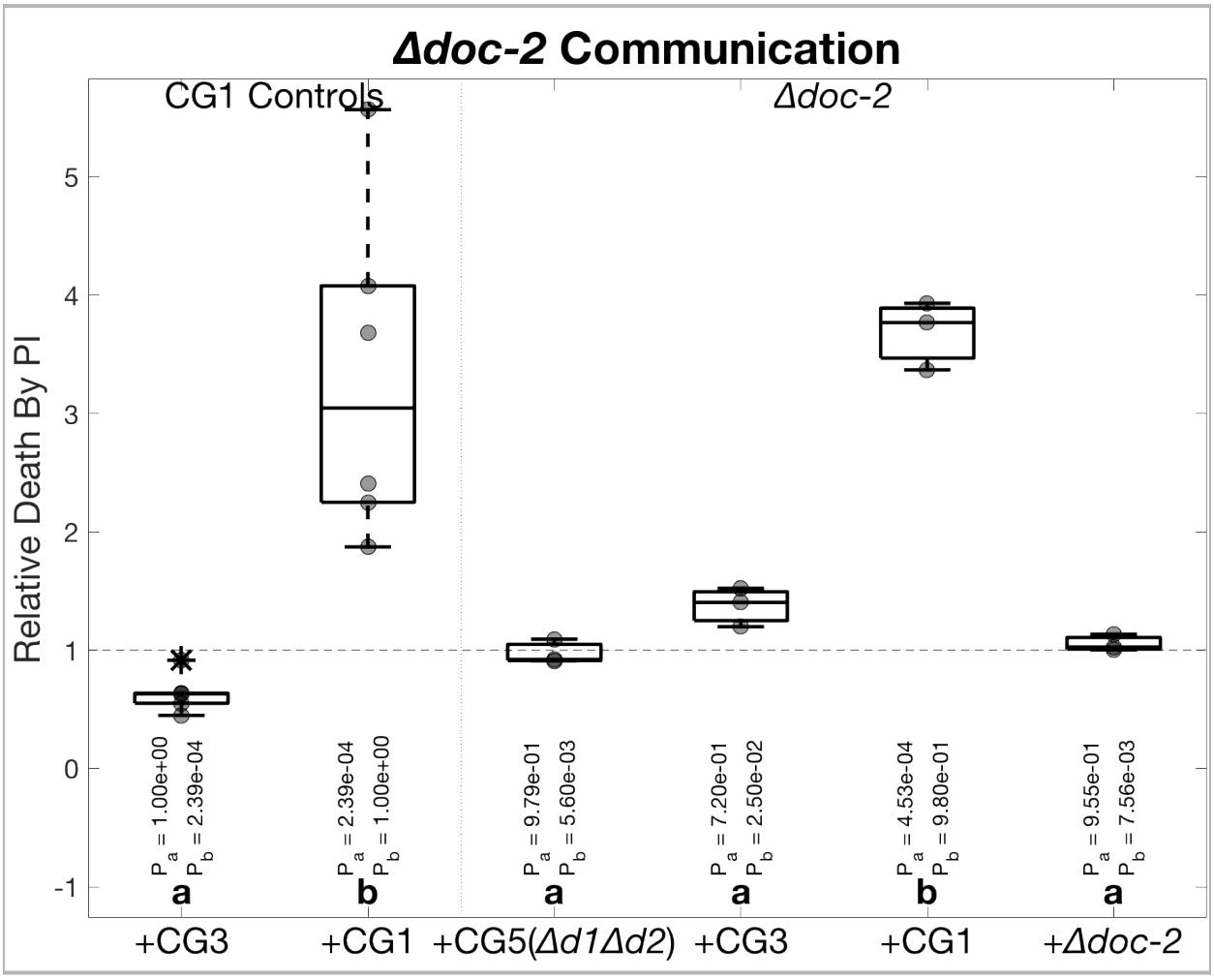


Figure 2.5.4-3. IDFC results from the *Δdoc-2* mutant

Boxplot summary of IDFC communication phenotypes of the *Δdoc-2* mutant. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (outside 99% of replicate data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Number of replicates vary between 7 and 3 for different pairs of strains.

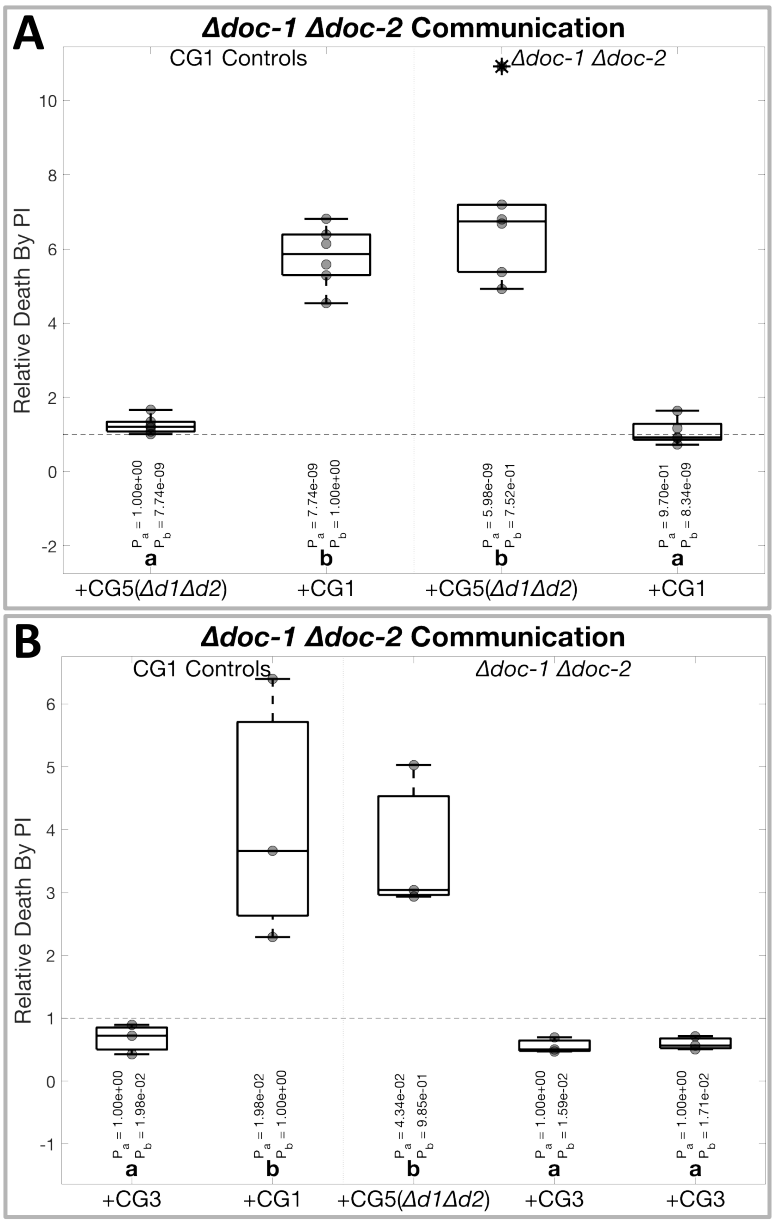


Figure 2.5.4-4. IDFC results from the *Δdoc-1 Δdoc-2* mutant

Comm. phenotypes of a *Δdoc-1 Δdoc-2* mutant analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** Quadruplicate IDFC results for *Δdoc-1 Δdoc-2* self-comm. and comm. with a CG1 DI. **B)** Triplicate IDFC results for *Δdoc-1 Δdoc-2* self-comm. and comm. with two CG3 DI strains.

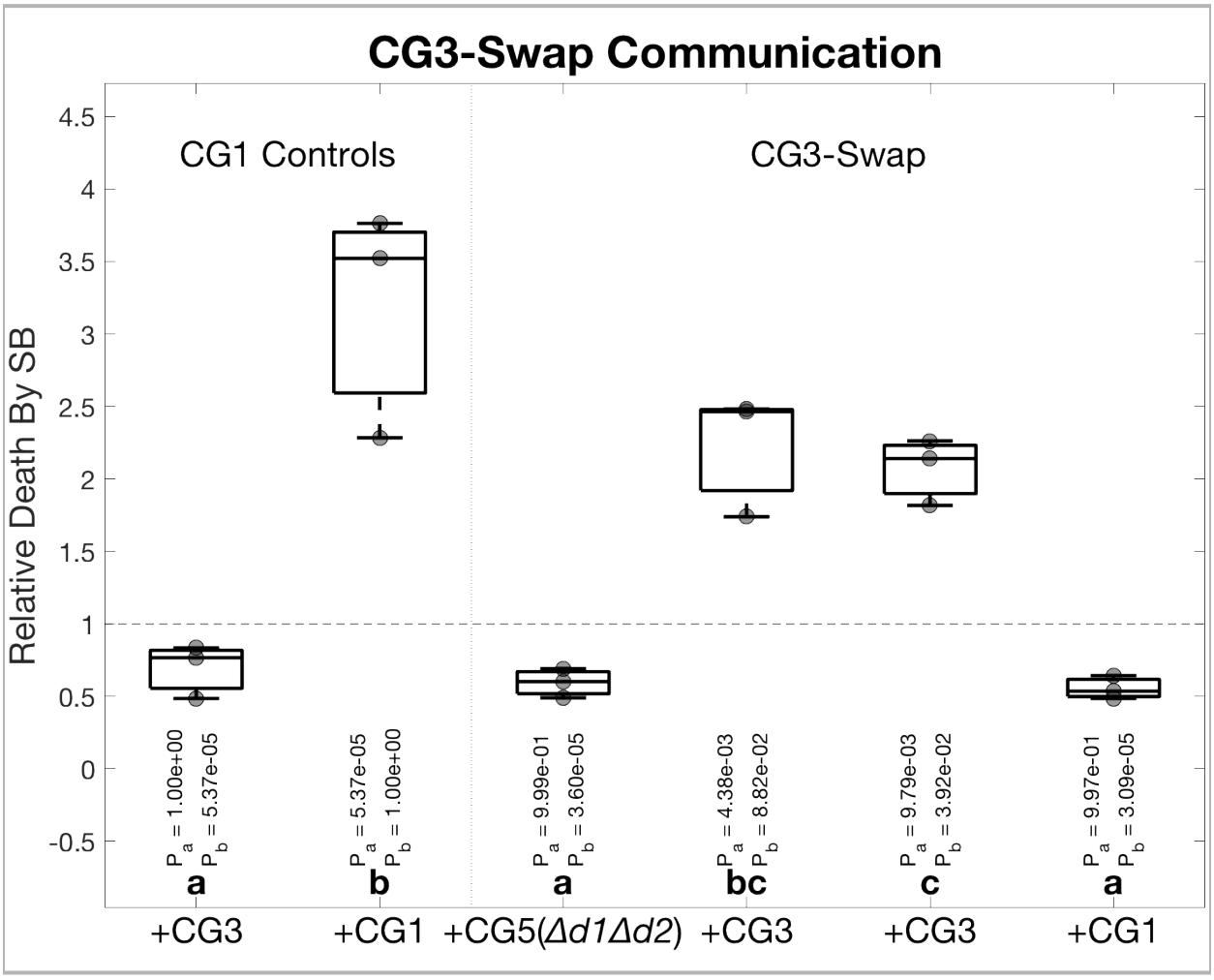


Figure 2.5.4-5. IDFC results from the CG3-swap strain

Boxplot summary of IDFC communication phenotypes of the CG3-swap strain. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Data from 3 replicates are presented.

Table 2.5.5. Summary of CG phenotypes by IDFC from section 2.5.5

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1
FGSC2489			
Δdoc-1			
Δdoc-1 doc-1-gfp			
Δdoc-1 doc-1-V5			
Δdoc-2			
Δdoc-1 Δdoc-2			
Δdoc-1 Δdoc-2 doc-1-gfp	Higher	Lower	
*Δdoc-1 Δdoc-2 doc-2-gfp**			

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. The *Δdoc-1 Δdoc-2 doc-1-gfp* strain communicated with CG3 and CG5 at intermediate but statistically different rates.
*DOC-2-GFP could not be detected via western blot in this strain.

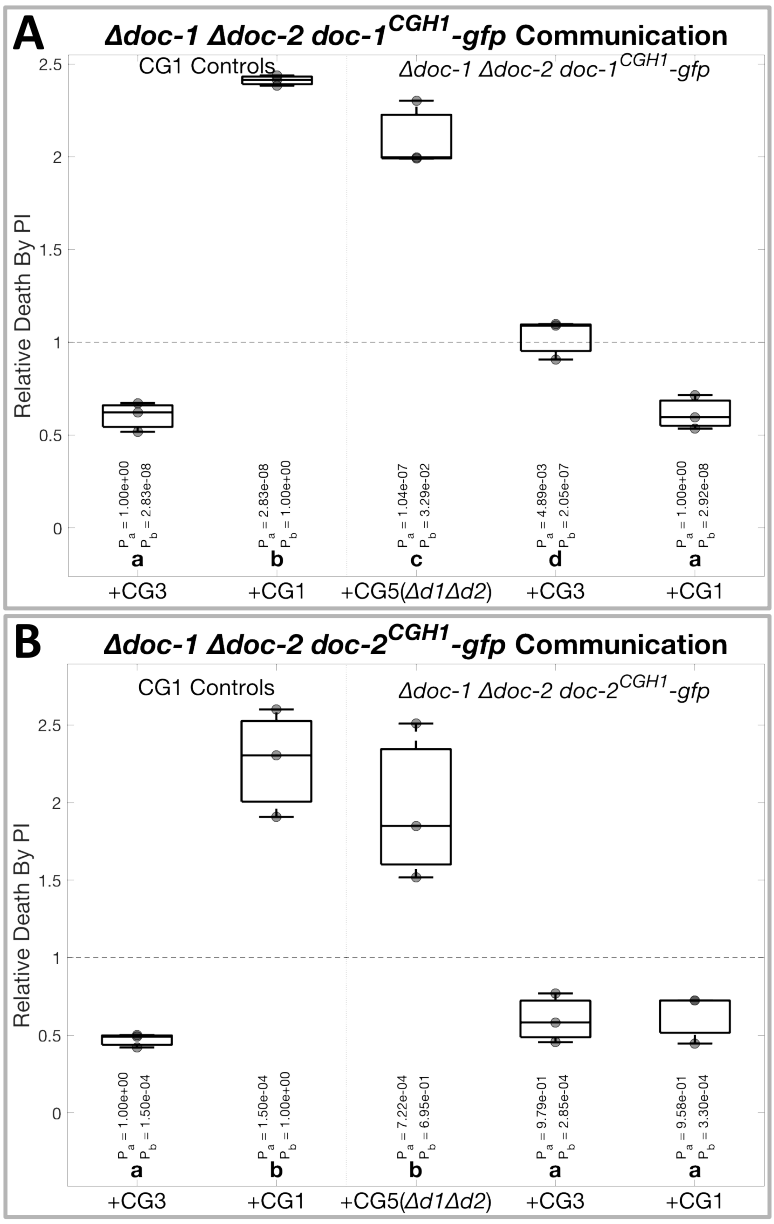


Figure 2.5.5-1. IDFC results from *Δdoc-1 Δdoc-2 doc-1^{CGH1}-gfp* and *Δdoc-1 Δdoc-2 doc-2^{CGH1}-gfp*

Boxplot summary of communication phenotypes of *Δdoc-1 Δdoc-2 doc-1^{CGH1}-gfp* and *Δdoc-1 Δdoc-2 doc-2^{CGH1}-gfp*, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show results from three replicates. **A)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{CGH1}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-2^{CGH1}-gfp*.

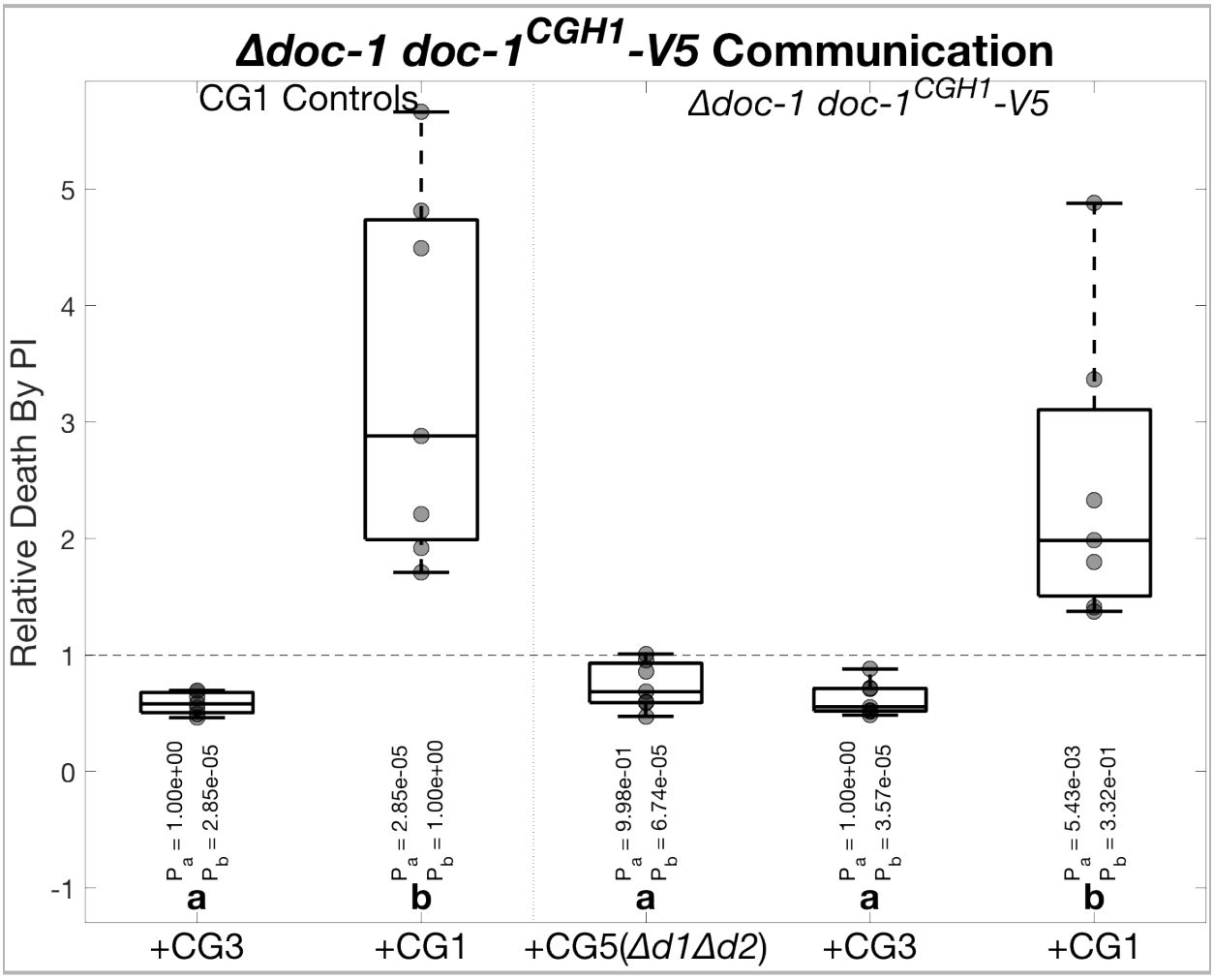


Figure 2.5.5-2. IDFC results from *Δdoc-1 doc-1^{CGH1}-V5*

Boxplot summary of IDFC communication phenotypes of *Δdoc-1 doc-1^{CGH1}-V5*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Outliers (defined as outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Data from 7 replicates are presented.

Table 2.6. Comparing communication and fusion assays

Assay	Pros	Cons	Notes
Microscopy	Versatile, only requires a microscope, requires no strain construction	Time-consuming, difficult, prone to bias, often gives ambiguous results	Many microscopy cons may be solved by automation
Quantitative Heterokaryons	Requires no specialized equipment, simple, easy	Requires the most strain construction and fusion between cells, takes a week to get results	
IDFC	Low sensitivity to bias, unambiguous, based on data from many cells, fast	Requires a flow cytometer, complicated data analysis, fusion between cells, and more strain construction than microscopy	May be improved by better vital dyes
dyeFC	Versatile, low sensitivity to bias, unambiguous, based on data from many cells, fast, requires no strain construction	Requires a flow cytometer and complicated data analysis, stains used are less reliable than the vital dyes used in IDFC, RSM metric is indirect	Better cell wall dyes may improve this assay, and it can be adapted for use in other organisms. dyeFC may yield less statistical power than IDFC.

Chapter 3. Communication between strains engineered to express combinations of incompatible *doc* alleles

3.1. Introduction to "confusion" strains

The model for DOC-mediated CG specificity presented by Heller et al., 2016, suggests that interactions between DOC proteins are important for controlling communication[42]. Furthermore, we found no evidence of recombination between CGHs, implying that interactions between unmatched DOC variants cause inappropriate communication behavior that reduces fitness. I tested these predictions by assessing the communication phenotypes of "confusion" strains. I put "confusion" in quotes because many of these strains exhibit clear CG preferences.

Experiments with "confusion" strains allowed me test three predictions of the DOC-system model presented in chapter one section 1.7: the non-allelic interaction hypothesis, the competition hypothesis, and the CGH5 non-functionality hypothesis. The non-allelic interaction hypothesis predicts that interactions between matched alleles of *doc-1* and *doc-2* are required for proper communication behavior. According to this hypothesis, coexpressing incompatible alleles in the same strain should cause promiscuous, reduced, or otherwise aberrant communication behavior; hence the moniker "confusion" strains. The competition hypothesis predicts that different variants of the same DOC protein should compete for interactions with and control of the cellular machinery involved in intercellular communication. Because the DOC-system represses communication by default, such competition should reduce communication with all CGs. Finally, the CGH5 non-functionality hypothesis predicts that *doc-1*^{CGH5} has lost its communication functions. This would explain why *Δdoc-1 Δdoc-2* phenocopied CG5 strains in Heller et al., 2016[42]. If true, expressing the CGH5 allele of *doc-1* should not affect communication behavior.

I used IDFC to test the CG phenotypes of "confusion" strains expressing most possible combinations of *doc-1*^{CGH1}, *doc-2*^{CGH1}, *doc-1*^{CGH3}, *doc-2*^{CGH3}, and *doc-1*^{CGH5}. Control strain phenotypes against which "confusion" strains must be compared were obtained via experiments described in chapter 2. "Confusion" strain phenotypes supported the non-allelic interaction and competition hypotheses, but rejected the CGH5 non-functionality hypothesis. However, the coexpression of incompatible alleles had unpredictable effects on communication and revealed an epistatic hierarchy among *doc* alleles.

Comparing CG phenotypes of the "confusion" strains I tested presents two problems. First, some strains expressed CGH1 *doc* alleles from their native loci, while others expressed *doc* alleles from the *his-3* locus. Moreover, native promoters drove the expression of CGH1 and CGH3 alleles in some strains, while other strains used the *ccg-1* promoter. Differences between the genomic locations and promoters of *doc* alleles probably resulted in unequal transcription; tanscriptome sequencing experiments have established expression-level differences among the *doc* genes, *his-3*, and *ccg-1*[91,92].

The second problem relates to epitope tags: CGH1 alleles in all "confusion" strains and CGH3 alleles in some strains weren't tagged, while other variants were fused to GFP. Epitope tags were required to confirm expression of non-native alleles via western blot. However, experiments described in chapter 2 demonstrated that GFP-tags interfere with DOC protein function; this strongly implies that "confusion" strains may behave differently

without large epitope tags. These problems limit the conclusions one can draw from my "confusion" strain data. Nonetheless, some "confusion" strains exhibited interesting phenotypes with implications for our model of DOC-mediated CG specificity.

3.2. Materials and methods

3.2.1. Strains, cloning, and growth conditions

Basic protocols for cultivation and manipulation of *N. crassa* can be found on the Neurospora homepage at the FGSC (www.fgsc.net/Neurospora/NeurosporaProtocolGuide.htm). Strains were grown on Vogel's minimal medium[85] (MM) or on Westergaard's synthetic cross medium[86] for crosses. FGSC2489 was the parent of all strains used in this study and served as a CG1 control for all experiments.

Table 3.2.1-1 lists the strains used in experiments discussed in this chapter. FGSC2489, wild isolates, classical mutants, and single gene deletion strains are available through the Fungal Genetics Stock Center (FGSC). Construction details for strains published in Heller et al., 2016, or Heller et al., 2018, can be found therein[42,64]. Construction details for plasmids used in this study are described in chapter 2 section 2.5.2. Strains generated for this study were produced as follows.

doc alleles were amplified from wild isolates and cloned into a modified pMF272 plasmid, with *ccg-1* promoter and terminator and C-terminal GFP tag, using *Xba*I and *Pac*I restriction sites[42]. This vector is designed to recombine into *N. crassa*'s *his-3* locus, removing a premature stop codon from an auxotrophic point-mutant and restoring histidine prototrophy[87]. See Fig. 3.2.1 for a plasmid map. Sequences of primers used for cloning are in Table 3.2.1-2.

Before transformation, all constructs were linearized using *Nde*I and/or *Ssp*I. Prepared conidia of the *his-3* strains FGSC6103 (*mat A*) or FGSC9716 (*mat a*) were then transformed using electroporation following standard protocols. Prototrophic transformants were obtained and backcrossed to His$^-$ versions of FGSC2489, *Δdoc-1*, *Δdoc-2*, or *Δdoc-1 Δdoc-2*, as required. Histidine auxotrophic versions of *doc* deletion strains were generated via backcrosses to FGSC6103 or FGSC9716. The CG3-swap strain (*Δdoc-1 Δdoc-2* with *doc-1*^{CGH3} and *doc-2*^{CGH3} under native promoters expressed at the *his-3* locus) was created by Heller et al., 2016[42]. Other strains expressing CGH3 alleles of *doc-1* and *doc-2* were generated by crossing the CG3-swap strain with His$^-$ versions of FGSC2489, *Δdoc-1*, or *Δdoc-2*. The CG1 DI strain (*Δplp-1 Δplp-2 sec-9*^{GRD3}) was created by Heller et al., 2018[64]. Crossing the CG1 DI strain with the CG3-swap strain and *his-3*; *Δdoc-1 Δdoc-2* produced the CG3 and CG5 DI strains, respectively.

Throughout this chapter, CG1 and CGH1 will refer to the strain FGSC2489 and its *doc* alleles, CG3 and CGH3 will refer to the strain P4471[89] and its *doc* alleles, CG5 will refer to the *Δdoc-1 Δdoc-2* mutant, and CGH5 will refer to the allele of *doc-1* from strain JW220[93]. All GFP-tagged alleles were integrated into the genome at *his-3*.

3.2.2. Western blots

Protein extractions from germlings and western blots were performed as described in Jonkers et al., 2014[75]. 250 mL Erlenmeyer flasks with 100 mL liquid MM were inoculated

to $1x10^6$ conidia per mL. Flasks were incubated at 30°C shaking at 220 rpm for 2.5 hours, then incubated another 2.5 hours at 30°C without shaking. Germlings were harvested by vacuum filtration over a nitrocellulose membrane and frozen in liquid nitrogen. Frozen germlings were then bead beaten with 0.5 mm glass beads at liquid nitrogen temperatures for one minute. 300 µL protein extraction buffer (described in Pandey et al., 2004[90]) was added to each sample and they were bead beaten at room temperature for 15 seconds. Samples were then centrifuged for 30 minutes at 4°C and the protein extract was separated from the cell debris.

Protein concentrations in the extracts were estimated using a NanoDrop™ spectrophotometer (Thermo Fischer Scientific™). Volumes of extracts were adjusted such that all samples run on a single gel contained similar total protein. Extracts were denatured at 70°C in 1x NuPAGE™ LDS sample buffer (Invitrogen™) with 5% β-mercaptoethanol (by volume) for 10 minutes, then run on 7% tris-acetate-SDS polyacrylamide gels (NuPAGE™, Invitrogen™) with PageRuler™ pre-stained protein ladder (10 to 180 kDa, Thermo Fischer Scientific™). Gels were blotted onto PVDF membranes with transfer buffer containing 20 mM tris base, 150 mM glycine, 20% methanol (by volume).

Blots were incubated in TBST (1x TBS (VWR), 0.5% Tween-20 (Sigma Aldrich)) with 5% milk for one hour at room temperature, with a buffer change after 30 minutes. Then blots were probed overnight at 4°C with primary monoclonal antibodies in TBST with 0.5% milk. Blots were washed three times with TBST, and then probed for one hour at room temperature with secondary polyclonal antibodies linked to HRP in TBST. After three more washes with TBST, blots were developed with SuperSignal™ West Pico developer (Thermo Fischer Scientific™) and imaged on a ChemiDoc™ XRS+ with ImageLab™ software (Bio-Rad). Western blots showing expression of GFP-tagged DOC proteins are pictured in Fig. 3.2.2.

3.2.3. Flow cytometry analyses

Most of the data presented in this chapter was obtained using induced death flow cytometry (IDFC). Detailed procedures for performing IDFC and dye flow cytometry (dyeFC) assays, and analyzing the resulting data can be found in Chapter 2 section 2.4. A brief overview of my IDFC protocol follows: strains were incubated in MM agar slants at 30°C in dark for three days, then 25°C in light for seven days. Conidia were then suspended in water and filtered through cheesecloth to remove hyphae. Conidial suspensions were diluted to $3x10^7$ conidia per mL, and spread on MM plates solidified with Pluronic F-127 (Sigma-Aldrich), either alone or in an equal mix with a death-inducing (DI) strain. For each mixture of strains, the constituent individual strains were also plated alone. Plates were incubated at 30°C in dark for 4 hours before harvesting and washing the germlings. Germlings were then stained with propidium iodide (PI, Sigma-Aldrich) and SYTOX Blue (SB, Life Technologies) before running them and conidial samples on a BD LSR Fortessa X-20 flow cytometer using BD FACSDiva™ software (BD Biosciences).

The run speed of the cytometer was adjusted for each sample such that data acquisition rates were between 500 and 1000 events per second for germling samples, and no more than 2000 events per second for conidial samples. The following parameters were recorded: forward scatter area and height, Pacific Blue area, and PerCP-Cy5-5 area. At least 20,000 events were recorded from conidial samples, and enough events to yield at least

10,000 germinated cells after gating out conidia were recorded from remaining samples (~30,000 events was usually sufficient).

In each experiment, a sample of FGSC2489 mixed with the CG1 DI strain was used as a positive control and a mixture of FGSC2489 and the CG3 or CG5 DI strain was used as a negative control. These controls were used during data analysis and to validate the experimental and data acquisition conditions.

dyeFC experiments were performed following the IDFC protocol with the following differences: conidia were stained with either calcofluor white (CFW, also called Fluorescence Brightener 28, Sigma-Aldrich) or concanavalin A, Alexa Fluor™ 488 conjugate (CAF, Invitrogen™) before plating, viability dyes were not used, and the fluorescence parameters BUV 496 area and FITC area were recorded instead of Pacific Blue and PerCP-Cy5-5.

3.2.4. Flow cytometry data analysis

A detailed explanation of how flow cytometry communication assay data was analyzed is available in Chapter 2 sections 2.4. Data was analyzed using a custom MATLAB™ (version R2018b, MathWorks) script. Briefly, data from ungerminated cells were removed from germling data by comparison with data from conidial samples. Fluorescence gates were then defined using the experiment's positive control fluorescence distributions in the SB and PI channels. Germinated cells more fluorescent than the gate in either channel were identified as dead. Relative death rates were calculated by dividing the cell-death percentage in a mixture of two strains by the average cell-death percentage of the constituent strains by themselves.

Relative death rates from three or more replicate experiments were combined and analyzed using one-way ANOVA and Tukey-Kramer multipler comparison tests with a significance threshold of 0.05. Usually SB and PI results agreed; in these cases, figures present only the PI data. Occasionally the positive control gave anomalously low or variable relative death rates in the SB or PI channel, resulting in a loss of statistical power. In these cases, figures present data from the channel retaining high quality controls.

3.3. Support for the non-allelic interaction hypothesis and rejecting the CGH5 non-functionality hypothesis

3.3.1. Interactions between DOC-1^{CGH3} and DOC-2^{CGH1} produce an intermediate CG phenotype

If the non-allelic interaction hypothesis is correct, co-expressing incompatible variants of DOC-1 and DOC-2 should result in aberrant CG phenotypes. I used IDFC to assess the CG phenotype of a strain expressing DOC-1^{CGH3}-GFP in a *Δdoc-1* genetic background. As a control, I also assayed the communication behavior of a *Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp* strain. For easy reference, Table 3.3 summarizes the CG phenotypes of all strains tested in section 3.3, along with relevant controls assayed in chapter 2.

Relative to the *Δdoc-1 Δdoc-2* double mutant, expressing DOC-1^{CGH3}-GFP in this background caused a reduction in communication with CG5, and no affect on communication with CG1 or CG3 (Fig. 3.3.1 panel B). This result demonstrates that the

CGH3 variant is at least as functional as DOC-1^{CGH1}-GFP in terms of communication suppression.

Although expressing DOC-1^{CGH3}-GFP in the *Δdoc-1* background did not statistically alter the mutant's preference for CG3, it lowered the strains communication with CG3 and increased its communication with CG1 enough to make communication rates with CG1 and CG3 comparable (Fig. 3.3.1 panel A). This phenotype is intermediate between those observed for a *Δdoc-1 doc-1^{CGH1}-gfp* strain and the CG3-swap strain, suggesting interactions between DOC-1^{CGH3} and DOC-2^{CGH1} yield ambiguous CG preferences.

3.3.2. Expressing DOC-2^{CGH3} does not affect the *Δdoc-2* mutant's CG phenotype

Next, I tested how interactions between DOC-1^{CGH1} and DOC-2^{CGH3} affect CG phenotype. As a control, I also tested how expressing DOC-2^{CGH3}-GFP alone altered communication preferences in the *Δdoc-1 Δdoc-2* double mutant.

As with the CGH3 allele of *doc-1*, expressing DOC-2^{CGH3}-GFP reduced communication with CG5 in the double mutant background (Fig. 3.3.2 panel B). This indicates the CGH3 allele of *doc-2* is capable of suppressing communication.

However, DOC-2^{CGH3}-GFP had no effect on the *Δdoc-2* mutant's CG preferences (Fig. 3.3.2 panel A). This result could indicate that DOC-1^{CGH1} and DOC-2^{CGH3} interact to specify CG1, or that DOC-1^{CGH1} is dominant over the CGH3 variant of DOC-2.

3.3.3. Expressing DOC-1^{CGH5} reduces CG3 communication in the *Δdoc-1* mutant

Based on the CG5 phenotype observed for the *Δdoc-1 Δdoc-2* double mutant[42], the CGH5 non-functionality hypothesis predicts that expressing DOC-1^{CGH5} in strains from other CGs will not affect communication behavior. I used IDFC to test the effects of expressing DOC-1^{CGH5}-GFP in *Δdoc-1* and *Δdoc-1 Δdoc-2* mutant backgrounds.

In the either background, expressing DOC-1^{CGH5}-GFP reduced communication with the mutant strains' preferred CG, lowering communication with CG3 in the *Δdoc-1* background and communication with CG5 in the double mutant background to intermediate levels (Fig 3.3.3). Communication with other CGs was unaffected. These results suggest the CGH5 variant of DOC-1 retains some ability to suppress communication, and may interact with DOC-2^{CGH1}. However, I cannot explain the reduction in communication with CG5 caused by expressing DOC-1^{CGH5}-GFP in the *Δdoc-1 Δdoc-2* background. Perhaps the GFP tag produces a gain-of-function in the CGH5 DOC-1 variant. Further tests are necessary to rule out tag effects, but these results tentatively reject the CGH5 non-functionality hypothesis.

Table 3.3 summarizes results from this section. Overall, interactions between DOC-1^{CGH3} or DOC-1^{CGH5} and DOC-2^{CGH1} produced somewhat aberrant CG phenotypes, supporting the non-allelic interaction hypothesis. Whether these phenotypes result from physical interactions between DOC proteins or from a combination of the incompatible proteins' mutually exclusive CG preferences cannot be determined from phenotypic data.

3.4. Confirming the competition hypothesis

3.4.1. DOC-1^{CGH1} is dominant over CGH3 and CGH5 variants, while CGH1 and CGH3 variants of DOC-2 compete

The competition hypothesis predicts that incompatible variants of the same DOC protein should compete for control over communication if they are co-expressed. I tested this hypothesis extensively, beginning with strains co-expressing incompatible alleles of *doc-1* in a *Δdoc-2* background.

Expressing DOC-1^{CGH3}-GFP did not alter the CG1 preference of the *Δdoc-2* mutant (Fig. 3.4.1-1 panel A). Although a strain expressing DOC-1^{CGH1} and DOC-1^{CGH5}-GFP exhibited super communication with CG1, this cannot be the result of competition between DOC-1 variants (Fig. 3.4.1-1 panel B). These results indicate DOC-1^{CGH1} is dominant over the CGH3 and CGH5 variants.

Next I tested whether CGH1 and CGH3 variants of DOC-2 compete in a *Δdoc-1* background. Expressing DOC-2^{CGH3}-GFP reduced communication with CG3 to intermediate levels, suggesting the two variants compete for communication control and reduce communication rates (Fig. 3.4.1-2). Results from these experiments and others in section 3.4, as well as controls from chapter 2, are summarized in Table 3.4 for easy reference.

3.4.2. Competition between incompatible DOC systems confirms the competition hypothesis and supports the non-allelic interaction hypothesis

Because results from "confusion" strains expressing incompatible alleles of a *doc-1* were somewhat ambiguous, I tested the CG phenotypes of strains expressing the complete CGH1 DOC system (CGH1 variants of DOC-1 and DOC-2) along with DOC-1^{CGH3}-GFP or DOC-1^{CGH5}-GFP. IDFC results for a strain expressing DOC-1^{CGH3}-GFP in an FGSC2489 background indicate the CGH1 DOC system was not affected by the CGH3 variant of DOC-1 (Fig. 3.4.2-1 panel A). However, the positive control in this series of experiments was unusually variable, resulting in a loss of statistical power. Expression of the gfp-tagged CGH3 allele may have slightly reduced communication with CG1, but this was not statistically distinguishable from the variable positive control. In contrast, expressing DOC-1^{CGH5}-GFP reduced FGSC2489's communication with CG1 to intermediate levels (Fig. 3.4.2-1 panel B). These results suggest CGH5 and CGH1 DOC systems interfere with each other, supporting the competition hypothesis.

Next, I assayed the CG phenotypes of strains expressing the complete CGH3 DOC system (CGH3 variants of DOC-1 and DOC-2) along with DOC-1^{CGH1} or DOC-2^{CGH1}. Unlike the "confusion" strains described earlier in this chapter, the DOC variants expressed in these strains were untagged. The *Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}* strain did not communicate with CG3 or CG5, and exhibited intermediate communication with CG1 (Fig. 3.4.2-2 panel A). These results indicate DOC-1^{CGH1} remains partially dominant over the CGH3 DOC system, but the CGH3 system competes for communication control and reduces communication with CG1. Among the "confusion" strains I tested, *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* is unique because it co-expresses alleles from incompatible haplotypes that specify the same CG. Alone, DOC-2^{CGH1} specifies CG3, as does the combination of DOC-1^{CGH3} and DOC-2^{CGH3}. Therefore, I was surprised to find the *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* strain communicates at intermediate levels with both CG1 and CG3 (Fig. 3.4.2-2 panel B). This suggests a non-allelic interaction between *doc-1^{CGH3}* and *doc-2^{CGH1}* allows communication with CG1, while the CG3 DOC system and DOC-2^{CGH1} acting alone continue to specify CG3. A slight shift towards CG1

specificity from CG3 was observed in the *Δdoc-1 doc-1^{CGH3}-gfp* strain (Fig. 3.3.3 panel A), but the switch is much clearer in the *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* strain despite the presence of DOC-2^{CGH3}.

Finally, I tested the CG phenotype of a double DOC system strain, expressing untagged CGH1 and CGH3 variants of both DOC-1 and DOC-2. This strain phenocopied the *Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}* strain, communicating only with CG1 at intermediate levels (Fig. 3.4.2-3). These results clearly demonstrate that DOC-1^{CGH1} is partially dominant over the CGH3 DOC system, and that incompatible DOC systems must compete for control over CG specificity. Table 3.4 summarizes results from experiments in section 3.4.

3.5. Summary and discussion

The experiments described in this chapter were intended to test how CGH1, CGH3, and CGH5 alleles of the *doc* genes interact *in vivo*. The results allowed me to evaluate the competition hypothesis, the non-allelic interaction hypothesis, and the CGH5 non-functionality hypothesis.

The CG3 specificity observed for the *Δdoc-1* mutant in chapter 2 already suggested that genetic interactions between *doc-1* and *doc-2* are essential for proper regulation of communication by the DOC-system. Results presented in section 3.3 imply that interactions between *doc-2^{CGH1}* and any allele of *doc-1* reduce communication with CG3, supporting the non-allelic interaction hypothesis. Interestingly, expression of *doc-1^{CGH1}* was required for robust communication with CG1, and completely prevented communication with CG3 regardless of co-expressed CG3 or CG5 DOC variants. The dominant CG1 preference imparted by DOC-1^{CGH1} shows that not all DOC variants specify a single CG with equal veracity, and helps explain why CG1 strains were the only isolates tested by Heller et al., 2016, that completely maintained their CG exclusivity[42]. However, the most striking support for the non-allelic interaction hypothesis was provided by the *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* strain: while *doc-2^{CGH1}* alone and *doc-1^{CGH3} doc-2^{CGH3}* (the CG3-swap strain) both independently specify CG3, expressing all three alleles at once both reduces communication with CG3 and increases communication with CG1 to intermediate levels. The reduction in communication with CG3 in the combination strain can be explained by competition between the CGH1 and CGH3 alleles of *doc-2* (Fig. 3.4.1-2), but the gain in communication with CG1 can only be explained by non-allelic interactions between *doc-1^{CGH3}* and *doc-2^{CGH1}*. A small increase in communication with CG1 was also seen in *doc-1^{CGH3}-gfp doc-2^{CGH1}*, and I suspect this increase would have been greater were it not for the gfp-tag. These results demonstrate that genetic interactions between alleles of *doc-1* and *doc-2* control CG phenotype, and that only alleles from the same haplotype produce phenotypes approximating those of wild isolates.

Obvious support for the competition hypothesis can be obtained by examining the CG phenotype of the *Δdoc-1 doc-2^{CGH3}-gfp* strain. Coexpression of the CGH3 and CGH1 alleles of *doc-2* were resulted in reduced communication with CG3, suggesting the variants of DOC-2 were competing for control over communication. The competition hypothesis was also supported by decreases in communication with CG1 observed in strains expressing *doc-1^{CGH1}* and other alleles of *doc-1* in conjunction with any allele of *doc-2*. However, non-allelic interactions may play some role in reducing communication with CG1 in these strains. The inability of DOC-1^{CGH3}-GFP and DOC-1^{CGH5}-GFP to interfere with CG1

communication in a *Δdoc-2* genetic background likely results from a combination of functional impairment by the GFP tags and the semi-dominance of DOC-1^{CGH1}. Moreover, the double DOC system strain clearly indicates that co-expressed, incompatible, complete DOC systems will compete for control over communication. These results confirm that incompatible DOC proteins interfere with each other, resulting in reduced communication with all CGs.

Finally, the CG phenotypes of strains expressing *doc-1^{CGH5}-gfp* demonstrate the CGH5 variant of DOC-1 retains some ability to influence communication. The reduction in communication with CG5 observed when *doc-1^{CGH5}-gfp* was expressed in the *Δdoc-1 Δdoc-2* background is perplexing because the JW220 isolate, from which *doc-1^{CGH5}* was sourced, was the CG5 type-strain used in Heller et al., 2016[42]. This suggests the gfp-tag may alter the CGH5 allele's functionality. However, the super communication with CG1 observed for the *Δdoc-2 doc-1^{CGH5}-gfp* strain demonstrates that the gfp-tag does not universally reduce communication. Despite possible artifacts induced by the gfp-tag, the results of "confusion" strain experiments lead me to reject the CGH5 non-functionality hypothesis. If *doc-1^{CGH5}* is functional, robust communication between CGH5 isolates and the *Δdoc-1 Δdoc-2* mutant suggests a CG5 signal must be sent in the absence of DOC proteins. This implies a default CG5 signal must be produced by other cellular components, and may be modified by DOC-1 and DOC-2 variants if they are present.

Results from experiments described in this chapter and chapter 2 support a model in which alleles of *doc-1* and *doc-2* interact to direct communication behavior (Fig. 3.5). Matched alleles work together to specify a CG and induce robust self-communication. Unmatched alleles still interact, genetically if not physically, but cannot induce full communication with a single CG. If two alleles of the same *doc* gene are coexpressed, they compete with each other and reduce communication rates. Finally, the CGH1 variant of DOC-1 is incompletely dominant over CGH3 and CGH5 variants: strains expressing DOC-1^{CGH1} will not communicate with CG3 or CG5 no matter what other variants are present. Aberrant communication behavior caused by *doc* allele competition or unmatched non-allelic interactions may reduce the fitness of any strain expressing a recombinant *doc* haplotype. If such fitness costs were substantial, one would not expect to find evidence of *doc* recombination in wild populations.

This work could be improved in several ways. First, "confusion" strains expressing gfp-tagged alleles should be replaced with untagged or small epitope-tagged versions. Second, self-communication rates from all strains would clarify the role of *doc* gene and allele interactions in directing self-communication. Third, including CGH2 and CGH4 alleles in "confusion" strain studies could help explain how *doc-2* gene duplication events affected communication in haplotypes containing *doc-3*. Finally, observing the localization and oscillation patterns of non-CGH1 *doc* alleles could increase confidence in this model. For example, do DOC-1^{CGH1}-mCherry and DOC-1^{CGH3}-GFP colocalize? Do they co-oscillate during an initial oscillation of the MAK-2 complex? Do they continue to colocalize if communication continues? Such observations could also confirm whether communication is binary (yes or no) in individual cells. Of course, mechanistic information about the DOC proteins and their interaction partners would greatly improve our model of the DOC-system, but biochemical functions of DOC proteins remain elusive.

Table 3.2.1-1: Strains referenced in chapter 3		
Wild isolates and classical mutants		
Strain	**Genotype**	**Reference**
FGSC2489 (CG1)	Laboratory Wild Type, *mat A*	FGSC
P4471 (CG3)	Wild isolate from the Louisiana Population	Bhat & Vyas, 2003
JW220 (CG5)	Wild isolate from the Louisiana Population	Dettman et. al., 2006
FGSC4200	Laboratory Wild Type, *mat a*	FGSC
FGSC6103	Laboratory Wild Type, *his-3 mat A*	FGSC
FGSC9716	Laboratory Wild Type, *his-3 mat a*	FGSC

Table 3.2.1-1: Strains referenced in chapter 3 continued		
Manipulated strains (all in FGSC2489 genetic background, all deletions marked with hygR)		
Strain	**Genotype**	**Reference**
Δdoc-1	*Δdoc-1 mat a*	Dunlap et. al., 2007, & Heller et. al., 2016
Δdoc-1 His- A	*his-3; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 His- a	*his-3; Δdoc-1 mat a*	This study
Δdoc-2	*Δdoc-2 mat a*	Dunlap et. al., 2007, & Heller et. al., 2016
Δdoc-2 His-	*his-3; Δdoc-2 mat A*	Heller et. al., 2016
Δdoc-1 Δdoc-2 (CG5)	*Δdoc-1 Δdoc-2 mat A*	Heller et. al., 2016
Δdoc-1 Δdoc-2 His- A	*his-3; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 His- a	*his-3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG3-swap strain	*his-3:: doc-1-CGH3 doc-2-CGH3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG1 DI strain	*his-3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat A*	Heller et. al., 2018
CG3 DI strain	*his-3:: doc-1-CGH3 doc-2-CGH3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat ?*	This study
CG5 DI strain	*Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+; Δdoc-1 Δdoc-2 mat ?*	This study
Δdoc-1 Δdoc-2 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp; Δdoc-1 Δdoc-2 mat a*	This study
Δdoc-1 Δdoc-2 doc-1-CGH5-gfp	*his-3:: doc-1-CGH5-gfp; Δdoc-1 Δdoc-2 mat ?*	This study
Δdoc-1 Δdoc-2 doc-2-CGH1-gfp	*his-3:: doc-2-CGH1-gfp; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 doc-2-CGH3-gfp	*his-3:: doc-2-CGH3-gfp; Δdoc-1 Δdoc-2 mat a*	This study
Δdoc-1 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp; Δdoc-1 mat a*	This study
Δdoc-1 doc-1-CGH5-gfp	*his-3:: doc-1-CGH5-gfp; Δdoc-1 mat ?*	This study
Δdoc-2 doc-2-CGH3-gfp	*his-3:: doc-2-CGH3-gfp; Δdoc-2 mat a*	This study
Δdoc-2 doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp; Δdoc-2 mat ?*	This study
Δdoc-2 doc-1-CGH5-gfp	*his-3:: doc-1-CGH5-gfp; Δdoc-2 mat ?*	This study
Δdoc-1 doc-2-CGH3-gfp	*his-3:: doc-2-CGH3-gfp; Δdoc-1 mat ?*	This study
doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp mat a*	This study
doc-1-CGH5-gfp	*his-3:: doc-1-CGH3-gfp mat ?*	This study
Δdoc-2 doc-1-CGH3 doc-2-CGH3	*his-3:: doc-1-CGH3 doc-2-CGH3; Δdoc-2 mat ?*	This study
Δdoc-1 doc-1-CGH3 doc-2-CGH3	*his-3:: doc-1-CGH3 doc-2-CGH3; Δdoc-1 mat ?*	This study
doc-1-CGH3 doc-2-CGH3	*his-3:: doc-1-CGH3 doc-2-CGH3 mat a*	This study

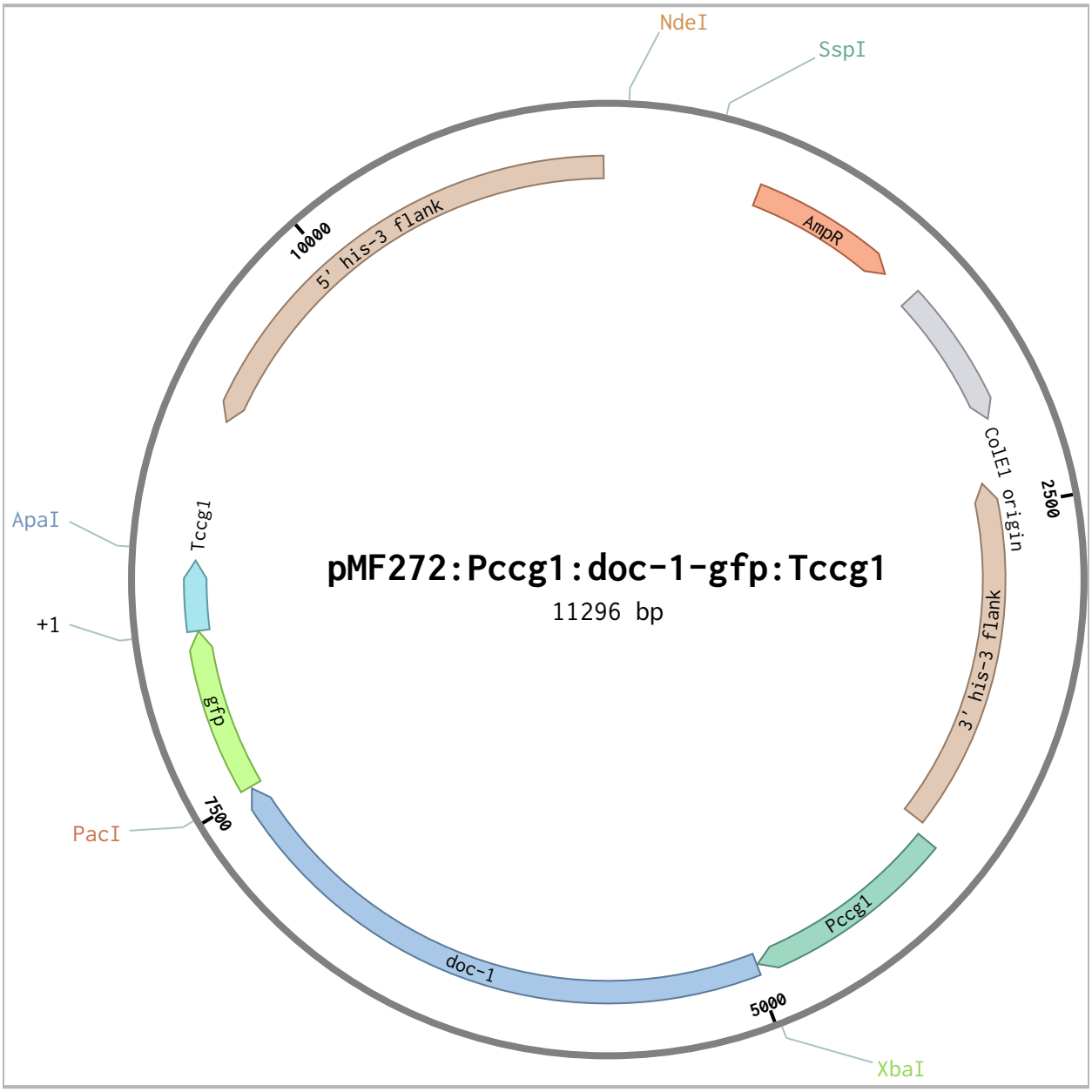


Figure 3.2.1. Plasmid map of pMF272:*Pccg1*:*doc-1*^{CGH1}*-gfp*:*Tccg1*
Plasmid map of the vector used in this study. Construction is described in chapter 2. Depicted with *doc-1*^{CGH1}; other genes and alleles were cloned into this vector using *Xba*I and *Pac*I restriction sites.

Table 3.2.1-2: Primers used for cloning in chapter 3		
Name	**Sequence (5'-3')**	**Purpose**
07191_start_JW220_XbaI	TCTAGAATGAGCTACGGCTCGCGCCAG	Forward primer to amplify *doc-1^{CGH5}* from JW220 adding an *Xba*I site
07191_end_JW220_wostop_PacI	TTAATTAAGCTCCATGCCATCTTGAGCC	Reverse primer to amplify *doc-1^{CGH5}* from *JW220* deleting the stop codon and adding a *Pac*I site
07191_start_P4471_XbaI	TCTAGAATGGGCACCGGTCTCC	Forward primer to amplify *doc-1^{CGH3}* from P4471 adding an *Xba*I site
07191_end_P4471_PacI	TTAATTAACGACGTCATGAACCCTAATTC	Reverse primer to amplify *doc-1^{CGH3}* from P4471 deleting the stop codon and adding a *Pac*I site
07192_start_P4471_XbaI	TCTAGAATGCCTGCTGTGTACCAGAG	Forward primer to amplify *doc-2^{CGH3}* from P4471 adding an *Xba*I site
07192_end_P4471_PacI	TTAATTAAGCTATCCGCCTCTAGCTTCTC	Reverse primer to amplify *doc-2^{CGH3}* from P4471 deleting the stop codon and adding a *Pac*I site

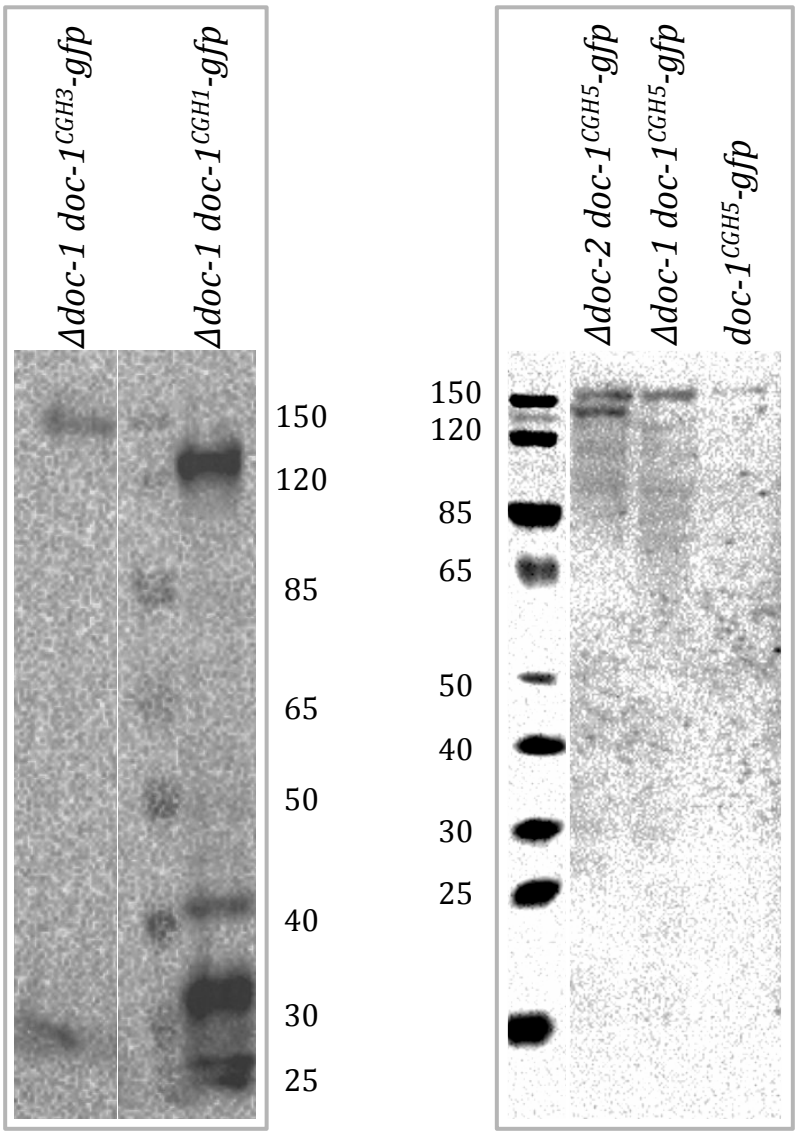


Figure 3.2.2. Western blots showing GFP-tagged CGH3 and CGH5 DOC proteins

Western blots showing detection of GFP-tagged CGH3 and CGH5 DOC-1 variants in FGSC2489, *Δdoc-1*, and *Δdoc-2* genetic backgrounds. Samples were loaded to normalize total protein content across the lanes in each blot. See section 3.2 for more details. For clarity, separate blots are outlined in gray boxes. Expected band sizes: DOC-1^{CGH1}-GFP = 118839.1 Da, DOC-1^{CGH3}-GFP = 126597.5 Da, DOC-1^{CGH5}-GFP = 131976.5 Da, free GFP = 26975.4 Da. Numbers indicate marker bands' kDa sizes. Expression of GFP-tagged proteins was not confirmed in the *Δdoc-1 Δdoc-2* genetic background. DOC-2^{CGH1}-GFP and DOC-2^{CGH3}-GFP could not be detected in "confusion" strains (data not shown).

Table 3.3. Summary of CG phenotypes by IDFC from section 3.3

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1
FGSC2489			
Δdoc-1			
Δdoc-1 doc-1^{CGH1}-gfp			
Δdoc-2			
Δdoc-1 Δdoc-2			
Δdoc-1 Δdoc-2 doc-1^{CGH1}-gfp	Higher	Lower	
*Δdoc-1 Δdoc-2 doc-2^{CGH1}-gfp**			
CG3-swap strain		✚	
Δdoc-1 doc-1^{CGH3}-gfp		Same as ->	<- Same as
Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp			
Δdoc-2 doc-2^{CGH3}-gfp			
Δdoc-1 Δdoc-2 doc-2^{CGH3}-gfp			
Δdoc-1 doc-1^{CGH5}-gfp			
Δdoc-1 Δdoc-2 doc-1^{CGH5}-gfp			

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. The *Δdoc-1 Δdoc-2 doc-1-gfp* strain communicated with CG3 and CG5 at intermediate but statistically different rates. *DOC-2-GFP could not be detected via western blot in this strain. ✚The CG3-swap strain exhibited high comm. with one CG3 DI backcross, and intermediate comm. with another.

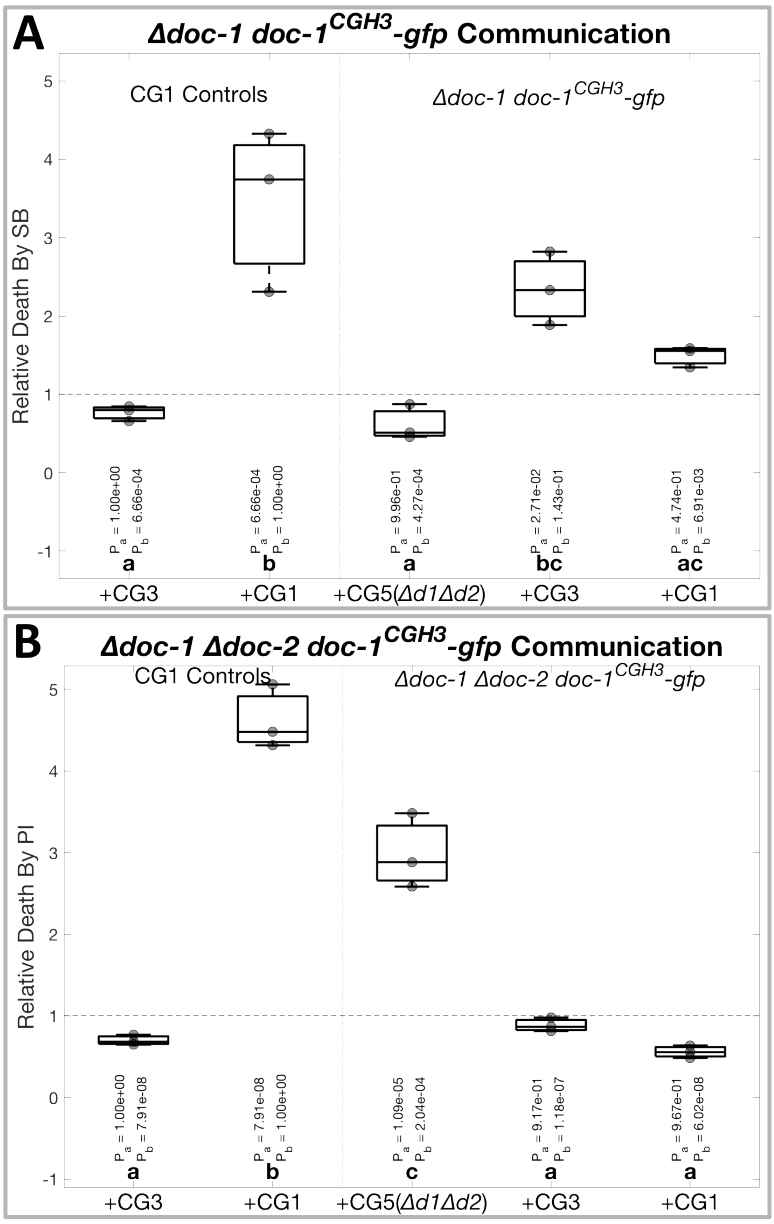


Figure 3.3.1. IDFC results from *Δdoc-1 doc-1^{CGH3}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp* strains

Boxplot summary of CG phenotypes of *Δdoc-1 doc-1^{CGH3}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp* strains, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide, SB = SYTOX Blue. Both panels show results from three replicates. **A)** IDFC results for *Δdoc-1 doc-1^{CGH3}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp*.

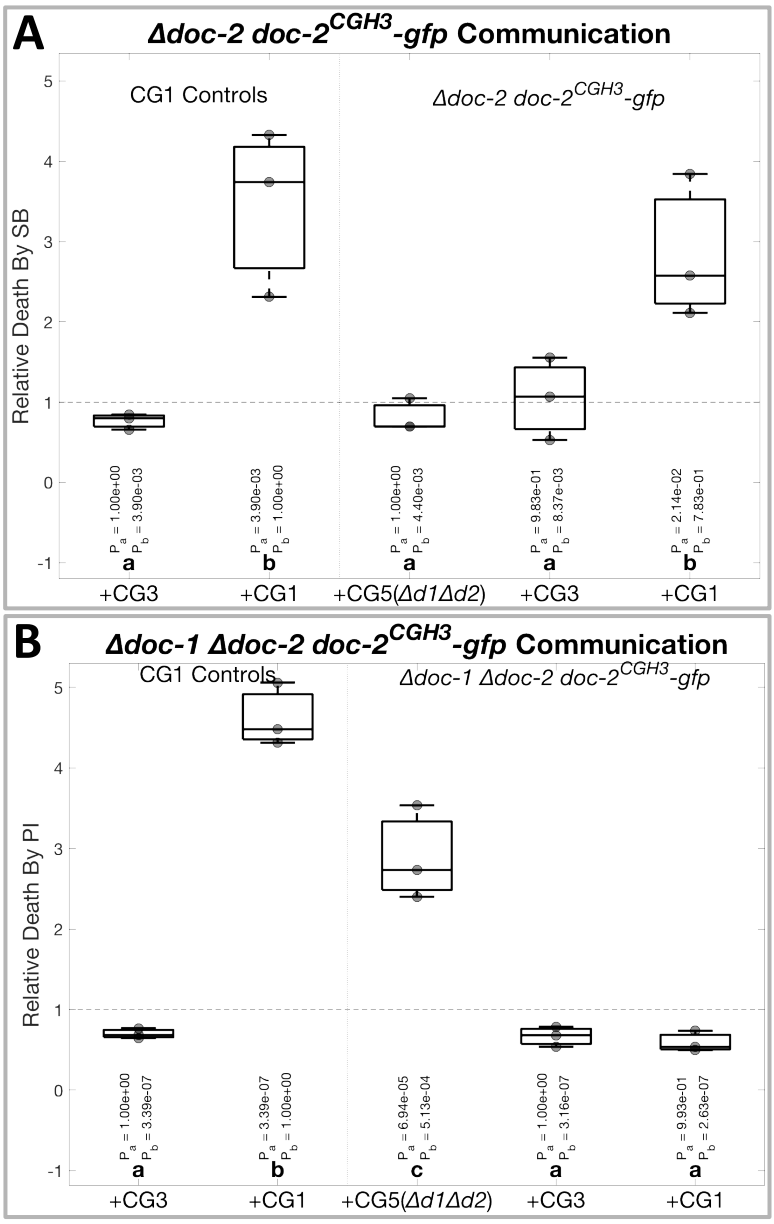


Figure 3.3.2. IDFC results from *Δdoc-2 doc-2^{CGH3}-gfp* and *Δdoc-1 Δdoc-2 doc-2^{CGH3}-gfp* strains

Boxplot summary of CG phenotypes of *Δdoc-2 doc-2^{CGH3}-gfp* and *Δdoc-1 Δdoc-2 doc-2^{CGH3}-gfp* strains, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide, SB = SYTOX Blue. Both panels show results from three replicates. **A)** IDFC results for *Δdoc-2 doc-2^{CGH3}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-2^{CGH3}-gfp*.

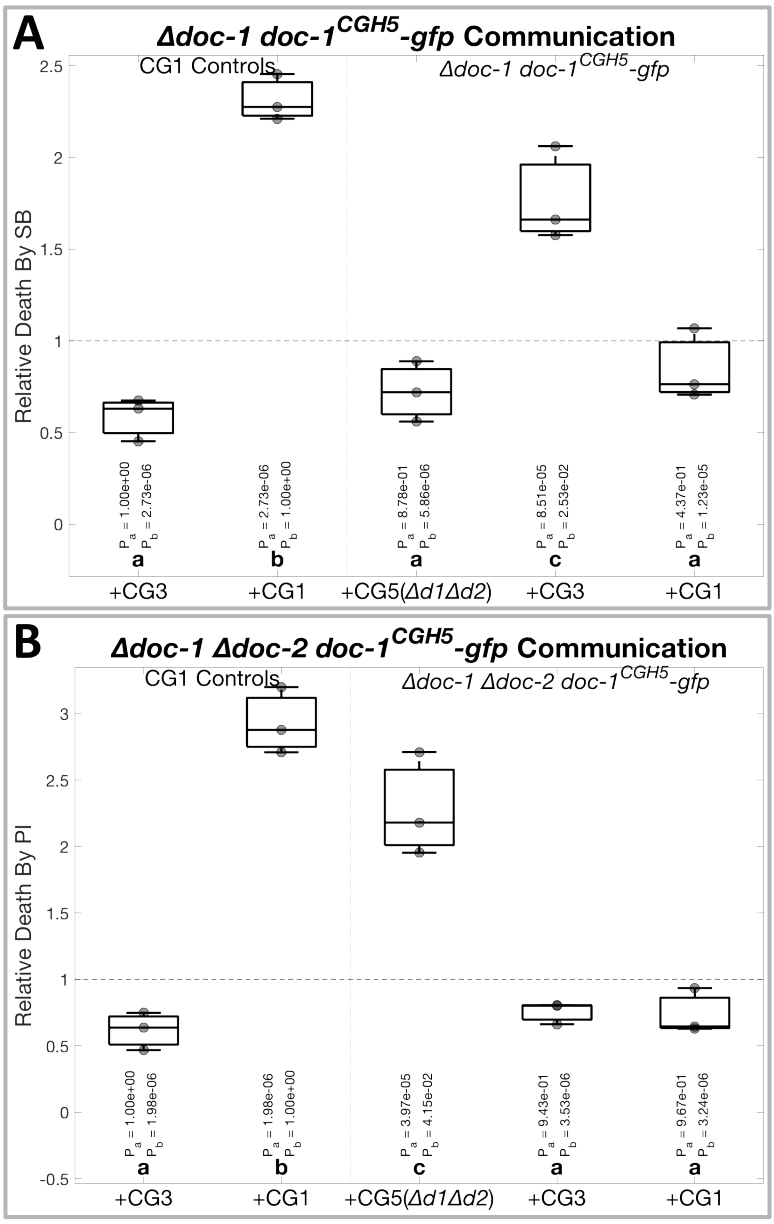


Figure 3.3.3. IDFC results from *Δdoc-1 doc-1^{CGH5}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{CGH5}-gfp* strains

Boxplot summary of CG phenotypes of *Δdoc-1 doc-1^{CGH5}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{CGH5}-gfp* strains, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide, SB = SYTOX Blue. Both panels show results from three replicates. **A)** IDFC results for *Δdoc-1 doc-1^{CGH5}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{CGH5}-gfp*.

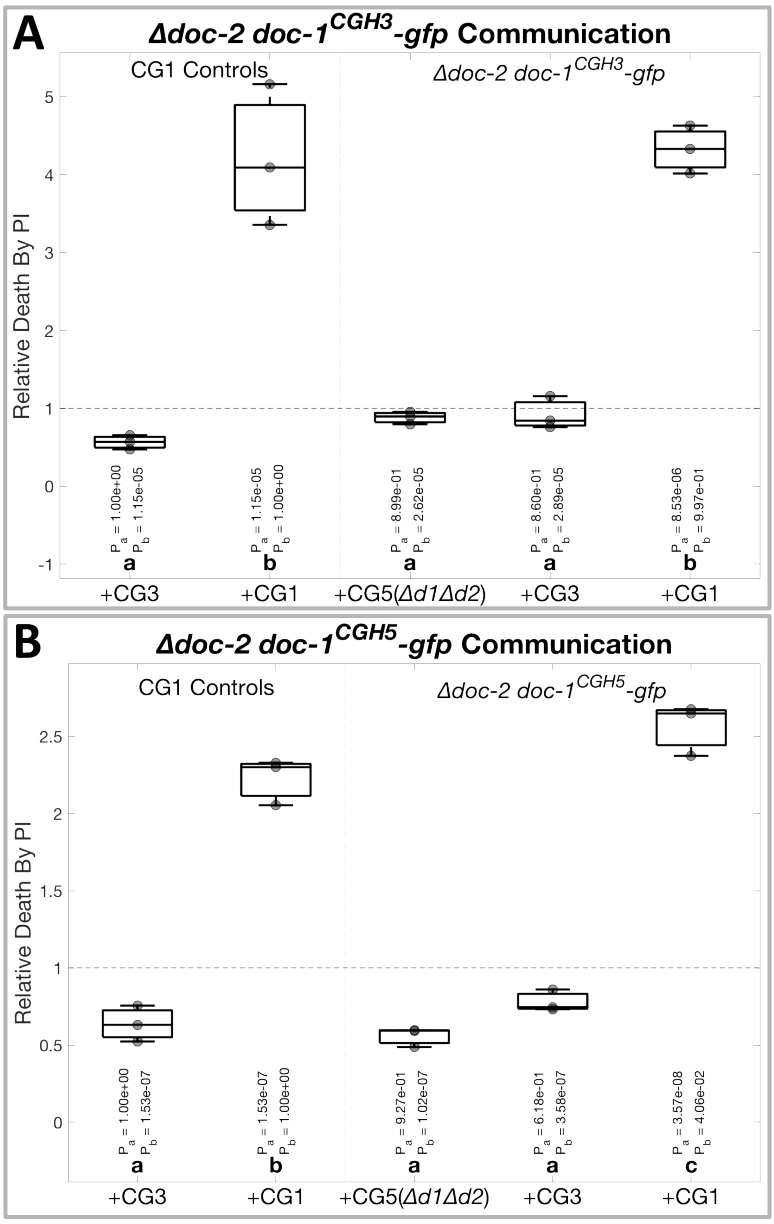


Figure 3.4.1-1. IDFC results from *Δdoc-2 doc-1^{CGH3}-gfp* and *Δdoc-2 doc-1^{CGH5}-gfp* strains

Boxplot summary of CG phenotypes from *Δdoc-2 doc-1^{CGH3}-gfp* and *Δdoc-2 doc-1^{CGH5}-gfp* strains, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show results from three replicates. **A)** IDFC results for *Δdoc-2 doc-1^{CGH3}-gfp*. **B)** IDFC results for *Δdoc-2 doc-1^{CGH5}-gfp*.

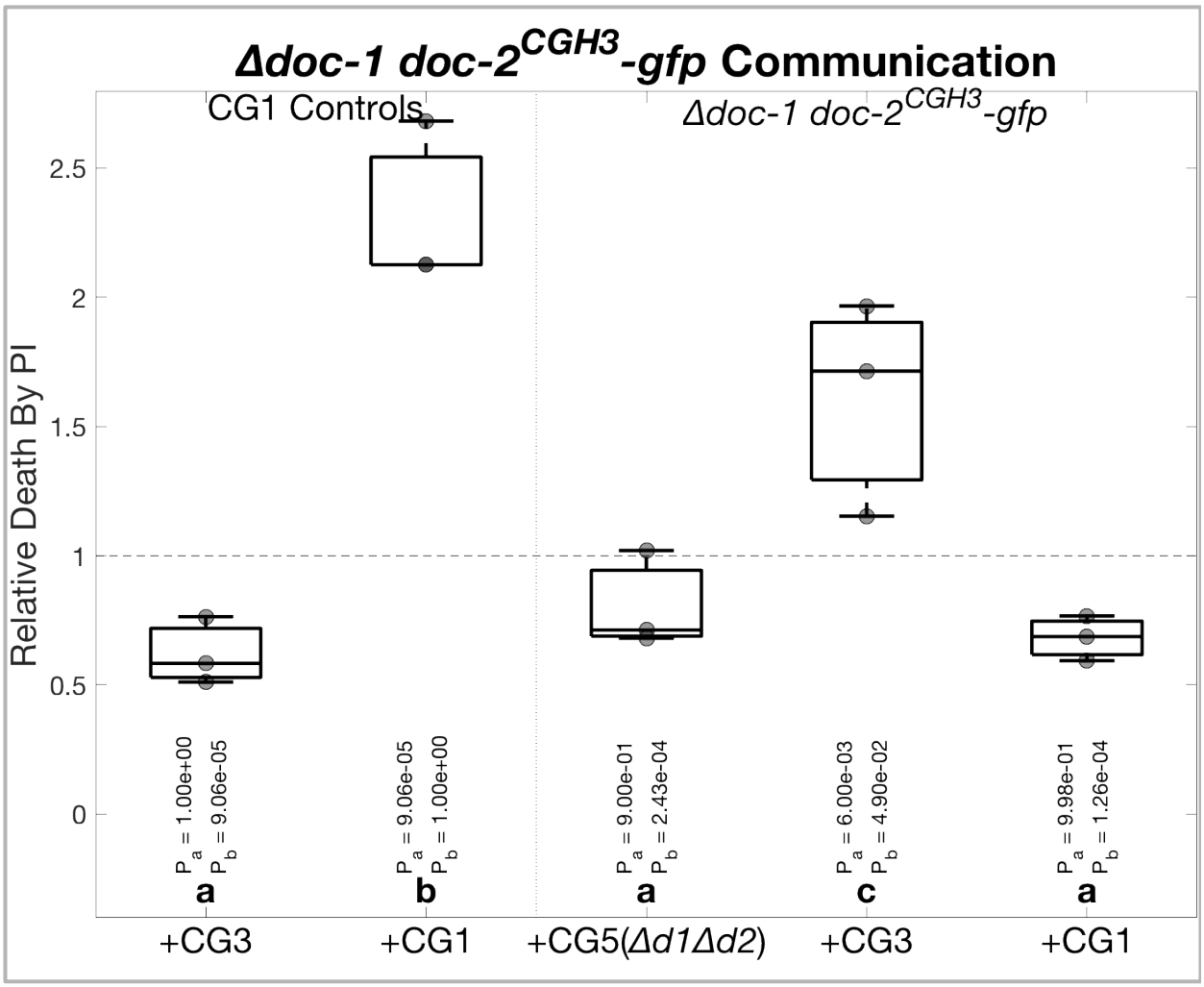


Figure 3.4.1-2. IDFC results from the *Δdoc-1 doc-2^{CGH3}-gfp* strain
Boxplot summary of IDFC CG phenotype of a strain in which CGH1 and CGH3 variants of DOC-2 compete. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide.

Table 3.4. Summary of CG phenotypes by IDFC from section 3.4

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1
FGSC2489			
Δdoc-1			
Δdoc-2			
Δdoc-1 Δdoc-2			
CG3-swap strain		✚	
Δdoc-2 doc-1^{CGH3}-gfp			
Δdoc-2 doc-1^{CGH5}-gfp			Super
Δdoc-1 doc-2^{CGH3}-gfp			
doc-1^{CGH3}-gfp			
doc-1^{CGH5}-gfp			
Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}			
Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}			
doc-1^{CGH3} doc-2^{CGH3}			

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. ✚The CG3-swap strain exhibited high comm. with one CG3 DI backcross, and intermediate comm. with another.

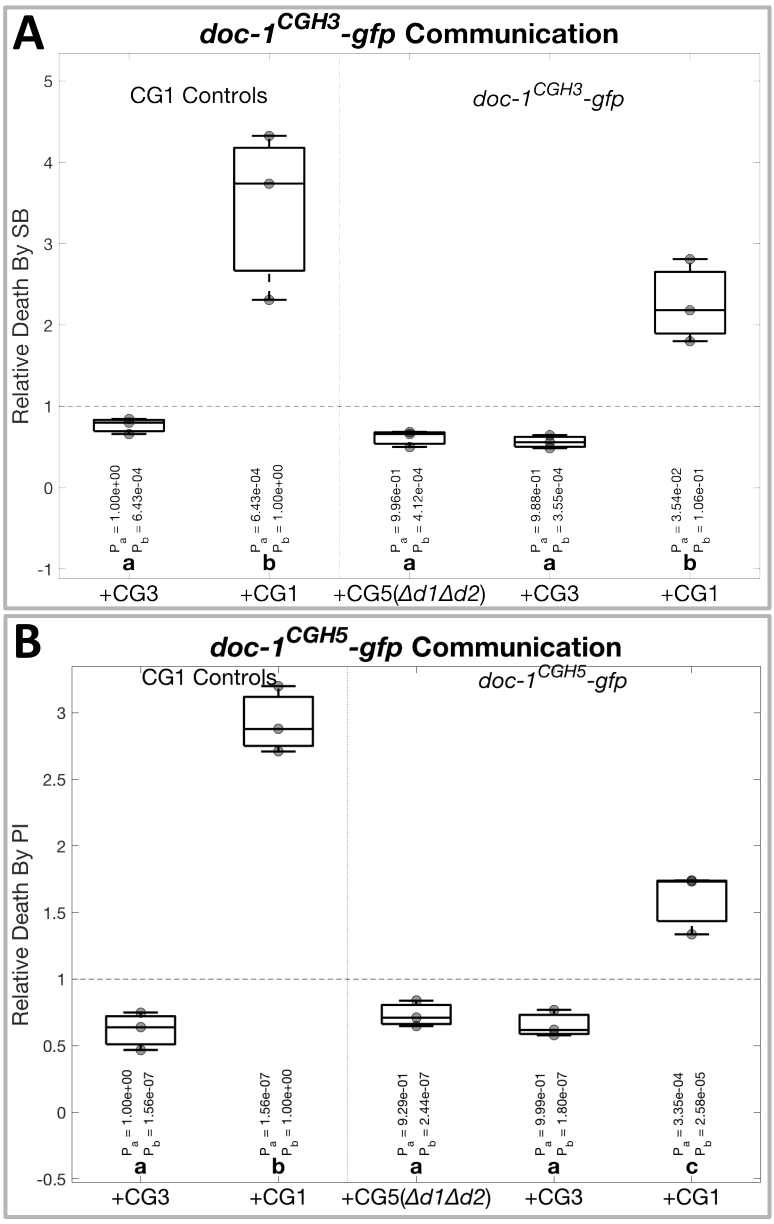


Figure 3.4.2-1. IDFC results from strains expressing *doc-1^{CGH3}-gfp* and *doc-1^{CGH5}-gfp*

Boxplot summary of communication phenotypes of strains expressing *doc-1^{CGH3}-gfp* or *doc-1^{CGH5}-gfp* in an FGSC2489 backgrond, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show results from three replicates. **A)** IDFC results for FGSC2489 expressing *doc-1^{CGH3}-gfp*. **B)** IDFC results for FGSC2489 expressing *doc-1^{CGH5}-gfp*.

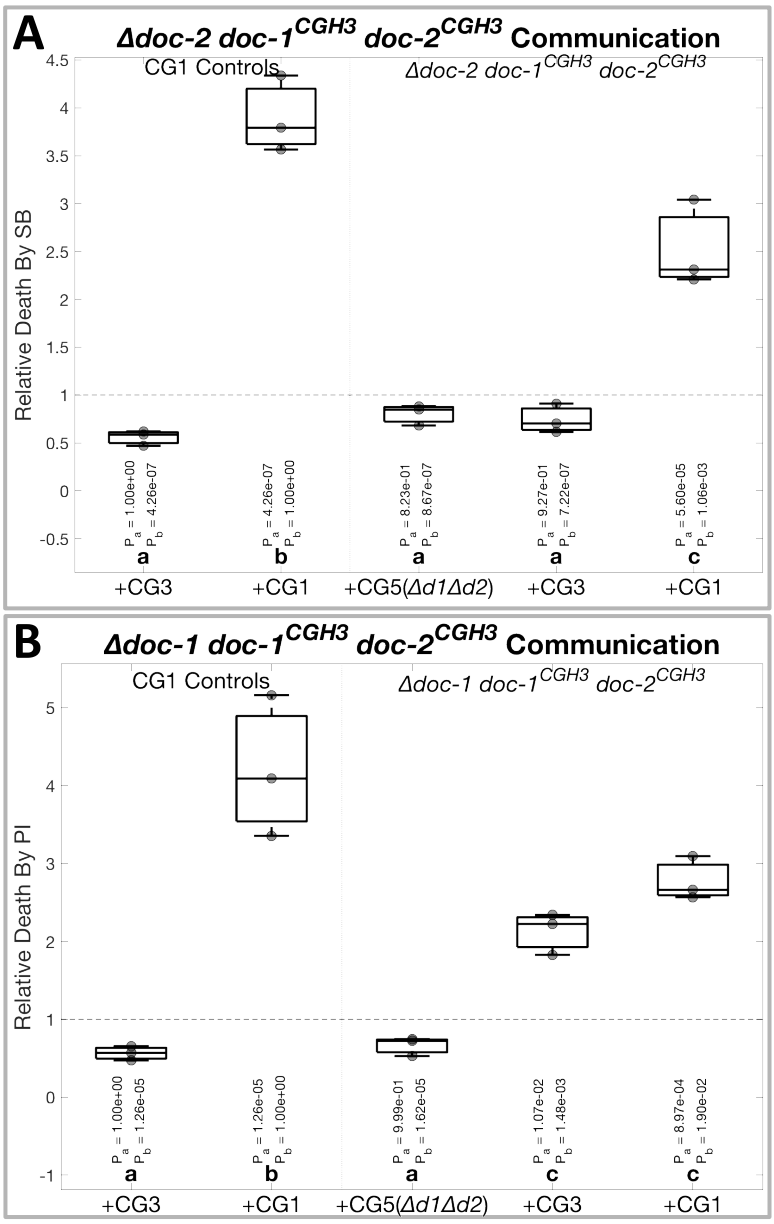


Figure 3.4.2-2. IDFC results from *Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}* and *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* strains

Boxplot summary of CG phenotypes from *Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}* and *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}* strains, analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI, CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap, PI = propidium iodide, SB = SYTOX Blue. Both panels show triplicate results. **A)** IDFC results for *Δdoc-2 doc-1^{CGH3} doc-2^{CGH3}*. **B)** IDFC results for *Δdoc-1 doc-1^{CGH3} doc-2^{CGH3}*.

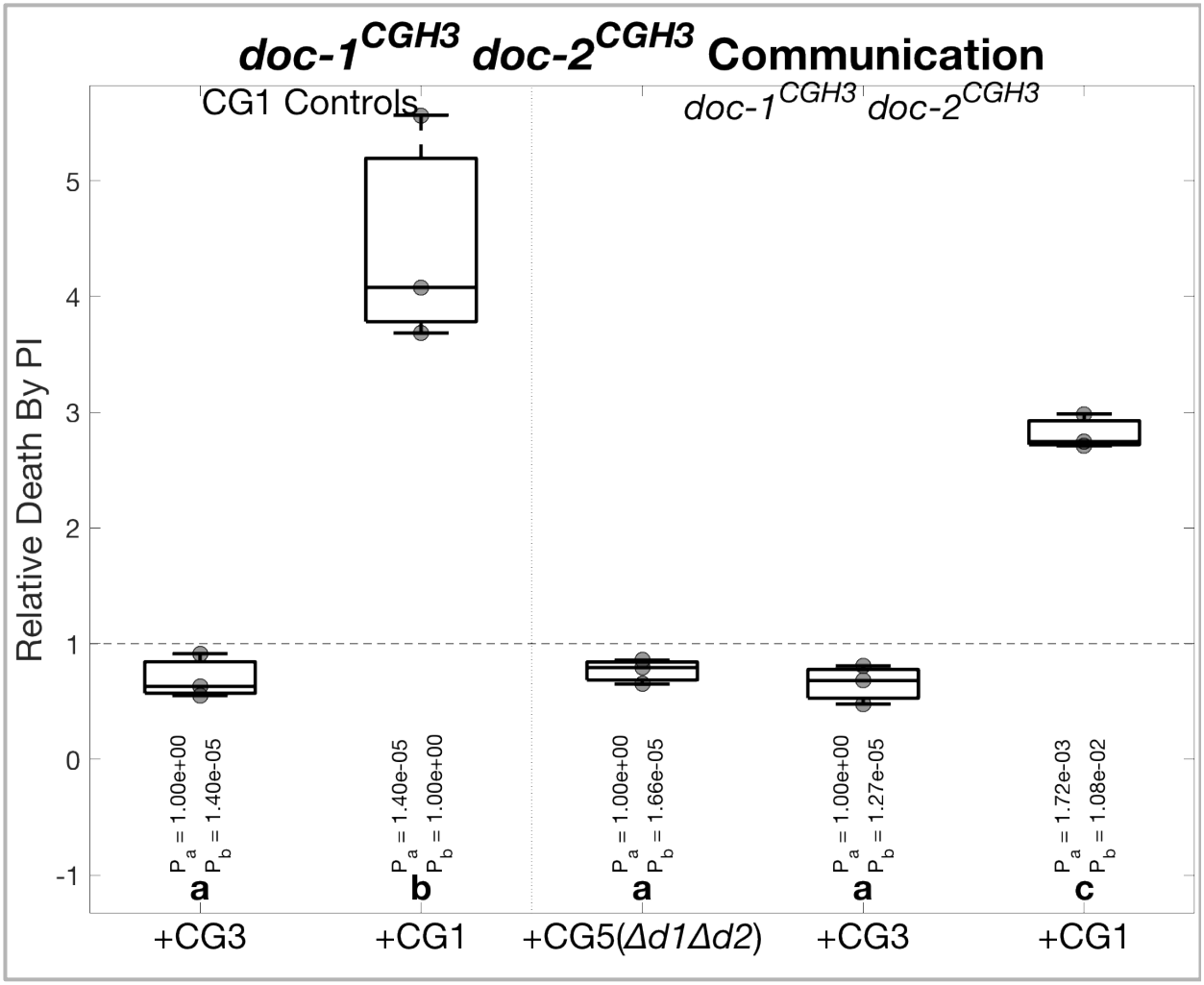


Figure 3.4.2-3. IDFC results from the double DOC strain: competing CGH1 and CGH3 DOC systems

Boxplot summary of IDFC CG phenotype of a strain expressing untagged CGH1 and CGH3 alleles of *doc-1* and *doc-2* with their native promoters in the FGSC2489 background (*his-3:: doc-1^{CGH3} doc-2^{CGH3}*). Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Data from three replicates are presented.

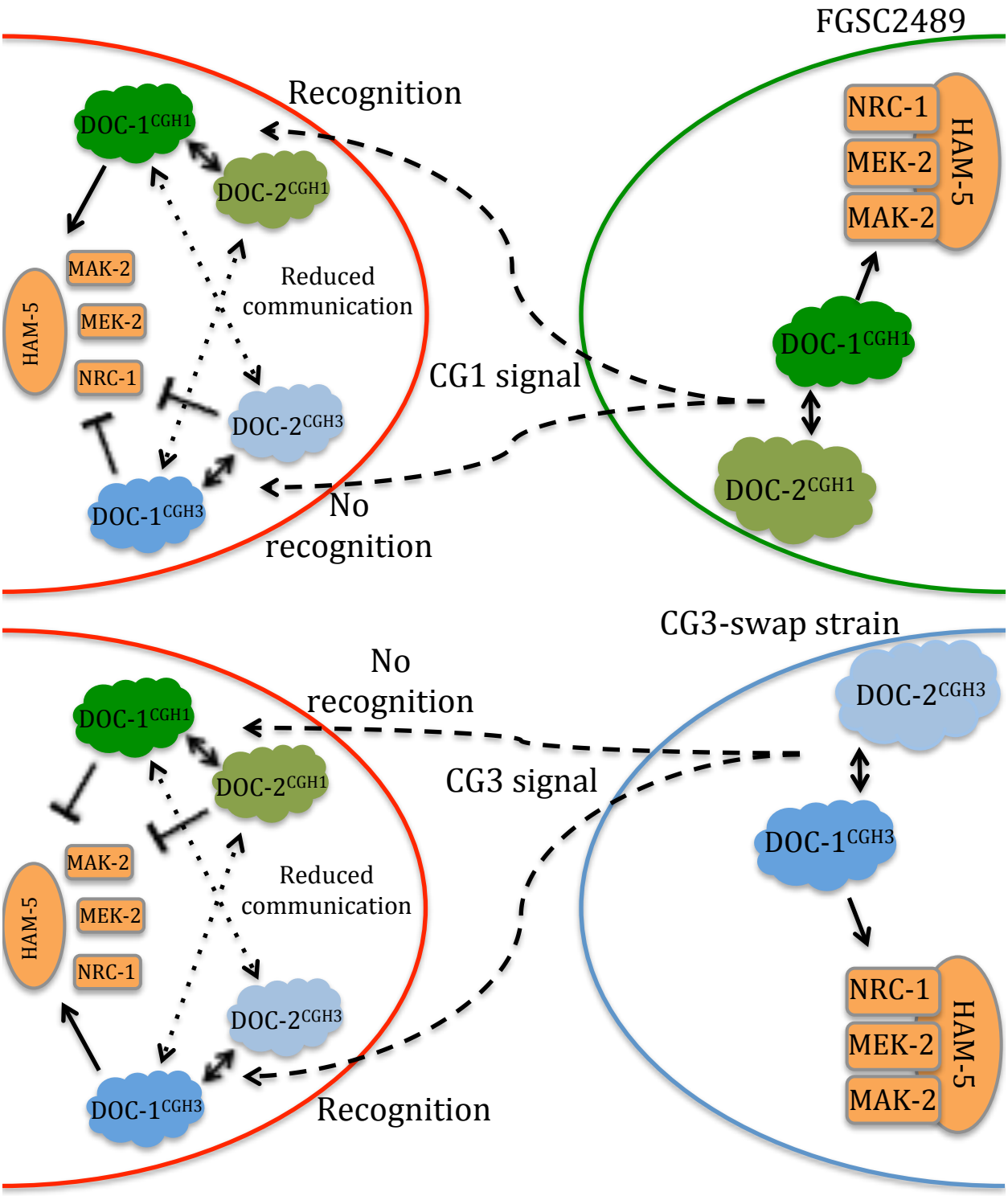


Figure 3.5. Model for interactions between DOC systems
When a strain expressing multiple DOC systems receives a CG signal, the DOC systems compete to control communication. Non-allelic interactions between the DOC systems produce aberrant communication behaviors and interference between the DOC systems yields an inconsistent response. This results in reduced communication rates with cells from either CG specified by the competing DOC systems.

Chapter 4. Functional analysis of the DOC proteins

4.1. Introduction to non-self recognition mechanisms

The biochemical functions of non-self recognition (NSR) proteins are as variable as the organisms that evolved them. This may be an understatement, as many organisms utilize multiple independent NSR modules encoding proteins with diverse functions. Even some individual NSR modules, like the contact-dependent inhibition and signaling system in *Burkholderia thailandensis*, use independent mechanisms to achieve multiple, separable NSR responses[3,23]. NSR proteins can be cytoplasmic, membrane anchored, or secreted. They may perform essential roles, or function only in NSR. They can be enzymes or non-enzymatic. However, these distinctions may be peripheral to core NSR functions. For example, *Escherichia coli*'s colicins have been classified as either enzymatic or non-enzymatic (pore-forming), but this does not appear to relate to the identities of colicin-producing or sensitive cells[94]. The myriad biochemical functions of NSR proteins may distract from the unique processes of NSR: identity broadcast, identity recognition, and response.

To accomplish those core processes, NSR systems must include an identifier and a recognizer. Identifiers and recognizers both carry identity information about the cell that produced them. However, recognizers interact with other cells' identifiers and directly or indirectly coordinate a cellular response. The key distinction between identifiers and recognizers is that recognizers transduce the self or non-self signal to the rest of the cell. I've used these terms instead of "ligand" and "receptor" because not all NSR systems follow traditional ligand-receptor paradigms. For example, self-infertility in the Solanaceae is mediated by interactions between pollen-encoded ubiquitin ligase substrate-recognition subunits (identifiers) and style-produced RNases (recognizers) in the male gametophyte's cytoplasm[4,29,30]. Male cells ubiquitinate and proteolyze female RNases from all but their own haplotype; self RNases are left to degrade rRNA in the pollen cell and prevent fertilization[95]. In this case, the RNases are recognizers because their actions go on to coordinate cellular behavior. The ubiquitin ligase subunits are only identifiers because they have no role in NSR beyond interaction with a recognizer. Recognizers and identifiers need not be separate molecules: *Podospora anserina*'s HET-s and *Neurospora crassa*'s RCD-1 VIC systems each use only one protein as identifier and recognizer. In the mixed cytoplasm of fused, non-identical *P. anserina* hyphae, interactions between incompatible variants of HET-s induce one version to form a prion, leading to amyloid formation and cell death[96,97]. Similarly, incompatible RCD-1 variants interact in heterokaryotic *N. crassa* cells to produce a pore-forming toxin that rapidly kills the fused cells[65].

Recognizers contain some common features: all variants of an NSR module's recognizer usually share conserved regions, while incompatible variants differ in some polymorphic regions. Variable regions are required for specific recognition of identifier variants. Conserved regions allow different recognizer variants to consistently trigger appropriate reactions to self and/or non-self recognition. The *het-c*/*pin-c* heterokaryon incompatibility system in *N. crassa* provides well-studied examples of conserved and variable regions in NSR proteins. *het-c* and *pin-c* are tightly linked components of an NSR module[98]. Three mutually incompatible haplotypes have been documented, and fusion between hyphae expressing incompatible alleles results in programmed cell death[36,99].

Recognition requires interactions between incompatible variants of HET-C and PIN-C, although allelic interactions between HET-C variants also contribute to the incompatibility response[98]. Both proteins contain conserved HET domains, which are common in filamentous ascomycete VIC genes and are required for cell-death induction[100]. Both proteins also contain inter-haplotype variable regions, which have been experimentally determined to confer allelic specificity in HET-C[36]. Whether HET-C and PIN-C are identifiers, recognizers, or both isn't entirely clear, but the presence of a conserved HET domain in each protein suggests they may both have some recognizer functions.

With no characterized homologs and no identifiable functional domains*, the DOC system is a black box. We know a CG identifier must be sent between cells before they touch, and we know the DOC system represses communication until the correct CG identifier is received[42]. Since *doc-1* and *doc-2* are necessary and sufficient to control CG phenotype, one or both must be recognizers. Additionally, the communication behavior of *Δdoc-1* and *Δdoc-2*, covered in chapter 2, rules out the possibility that either protein has exclusively identifier functions. Thus, both DOC-1 and DOC-2 must have recognizer functions. The specificity domain hypothesis predicts CGH-specific variations in *doc* alleles will contain specificity determinants. DOC-1's association with the MAK-2 kinase complex suggests all DOC-1 variants should also contain conserved domains that interact with some component of this complex (Fig. 1.6-3). The response domain hypothesis predicts regions that are conserved across incompatible DOC variants will contain domains required for core functions, including communication suppression and interactions with other cellular machinery.

In this chapter, I report my efforts to functionally characterize the DOC proteins. This work focused on DOC-1 because, relative to DOC-2, it has a more interesting localization pattern, its deletion has a more dramatic effect on CG phenotype, and all tested wild *N. crassa* isolates contain an allele of *doc-1*[42]. After searching for inter-haplotype variable and conserved regions in *doc-1* alleles, I was able to divide the DOC-1 proteins into variable N- and C-terminal regions, and a conserved middle region. These regions approximately aligned with predictions of unstructured and globular regions, respectively. Then I assayed the CG phenotypes of strains expressing manipulated DOC-1 variants to evaluate the specificity and response domain hypotheses, and whether the communication inhibition and CG recognition functions of DOC-1 could be separated. My results tentatively support the specificity domain hypothesis, partially reject the response domain hypothesis, and indicate that communication inhibition and CG recognition are separable.

Care must be taken when interpreting results from non-functional manipulated DOC variants. Without clear borders between domains, truncations and chimeras can easily cause misfolding, resulting in a non-functional protein. GFP-tags also affect DOC functions, as seen in chapter 2. Despite these caveats, I obtained interesting results that inform our model of the DOC system and reveal some functional features of DOC-1.

*Heller et al., 2016, putatively identified an OmpH-like outer membrane protein domain and a transmembrane domain in DOC-2 ^{CGH1}. However, I could not detect these features in other variants of DOC-2 (see section 4.3.1).

4.2. Materials and methods

4.2.1. Strains, cloning, and growth conditions

Basic protocols for cultivation and manipulation of *N. crassa* can be found on the Neurospora homepage at the FGSC (www.fgsc.net/Neurospora/NeurosporaProtocolGuide.htm). Strains were grown on Vogel's minimal medium[85] (MM) or on Westergaard's synthetic cross medium[86] for crosses. FGSC2489 was the parent of all strains used in this study and served as a CG1 control for all experiments.

Table 4.2.1-1 lists the strains used in experiments discussed in this chapter. FGSC2489, wild isolates, classical mutants, and single gene deletion strains are available through the Fungal Genetics Stock Center (FGSC). Construction details for strains published in Heller et al., 2016, or Heller et al., 2018, can be found therein[42,64]. Details for the construction of vectors used in this chapter may be found in chapter 2, with one exception, described below. Strains constructed for this study were produced as follows.

The pMF272:*Ptef1*:*gfp*:*Tccg1* vector was produced by swapping the *ccg-1* promoter in pMF272:*Pccg1*:*gfp*:*Tccg1* for the *tef-1* promoter in pMF272:*Ptef1*:*V5*:*Tccg1* using *Not*I and *Xba*I restriction sites. Plasmid maps for pMF272:*Pccg1*:*gfp*:*Tccg1* and pMF272:*Ptef1*:*V5*:*Tccg1* can be found in Fig 4.2.1. All the vectors used in this study were designed to recombine into *N. crassa*'s *his-3* locus, removing a premature stop codon from an auxotrophic point-mutant and restoring histidine prototrophy[87].

Terminally truncated *doc* alleles were subcloned from pMF272:*Ptef1*:*doc-1*-*gfp*:*Tccg1* by amplifying the truncated sequences with primers that added 5' *Xba*I and 3' *Pac*I sites, and then cutting and ligating into pMF272:*Pccg1*:*gfp*:*Tccg1*. *doc-1*$^{\Delta 221\text{-}623}$ was produced using fusion PCR[101], replacing *doc-1*'s sole intron with an identical oligonucleotide. *doc-2*$^{\Delta TM}$ was also generated using fusion PCR. Motif deletion alleles of *doc-1* were created in three steps: first, *doc-1* was subcloned into pCR-Blunt-II (Invitrogen™), next pCR-Blunt-II:*doc-1* was linearized using PCR with primers binding on either side of the motif being deleted, and finally, a kinase-ligase-*Dpn*I reaction was used to circularize the motif deletion allele construct (New England BioLabs®, see https://www.neb.com/products/m0554-kld-enzyme-mix for details). Chimeric alleles were also cloned via fusion PCR, using pMF272:*Pccg1*:*doc-1*^{CGH3}-*gfp*:*Tccg1* (see chapter 3) as a template for CGH3 regions. All alleles were inserted into their expression vectors using *Xba*I and *Pac*I restriction sites. Sequences of primers and other oligonucleotides used for cloning are in Table 4.2.1-2.

Before transformation, all constructs were linearized using *Nde*I and/or *Ssp*I. Prepared conidia of the *his-3* strains FGSC6103 (*mat A*) or FGSC9716 (*mat a*) were then transformed using electroporation following standard protocols. Prototrophic transformants were obtained and backcrossed to His$^-$ versions of FGSC2489, *Δdoc-1*, *Δdoc-2*, or *Δdoc-1 Δdoc-2*, as required. Histidine auxotrophic, *mat a* versions of *doc* deletion strains were generated via backcrosses to FGSC6103 or FGSC9716. The CG3-swap strain (*Δdoc-1 Δdoc-2* with *doc-1*^{CGH3} and *doc-2*^{CGH3} under native promoters expressed at the *his-3* locus) was created by Heller et. al., 2016[42]. Other strains expressing CGH3 alleles of *doc-1* and *doc-2* were generated by crossing the CG3-swap strain with His$^-$ versions of FGSC2489, *Δdoc-1*, or *Δdoc-2*. The CG1 DI strain (*Δplp-1 Δplp-2 sec-9*^{GRD3}) was created by Heller et. al., 2018[64]. Crossing the CG1 DI strain with the CG3-swap strain and *Δdoc-1 Δdoc-2; his-3* produced the CG3 and CG5 DI strains, respectively.

Throughout this chapter, *doc-1*, *doc-2*, DOC-1, and DOC-2 will refer to the FGSC2489 variants (CGH1), while CGH3 variants, denoted with superscript, will refer to variants from P4471[89]. All epitope-tagged alleles were integrated into the genome at the *his-3* locus.

4.2.2. Sequence analyses

Sequences of *doc-1* alleles from wild isolates were obtained via a BLAST[102] search using FGSC2489's *doc-1* and *doc-2* as a queries against de novo sequence assemblies from 27 wild isolates[47]. Codons were aligned and translated using MACSE[103], then visualized and processed with JalView[104]. Translated sequences were scanned for conserved domains and protein family signatures using HMMScan[105] and CDD searches[106]. PHMMER[105] and BLASTP[102] were used to identify proteins homologous to DOC-1 or DOC-2. After removing gapped and unaligned regions from the multiple sequence alignment, aligned codons were scanned for K_a/K_s and Tajima's D using the Selecton server[107,108]. GlobPlot[109] was used predict disordered and globular regions in DOC-1 and DOC-2 from FGSC2489. I used SPOCTOPUS[110] to scan for signal peptides and transmembrane domains in FGSC2489's variants DOC-1 and DOC-2, and in P4471's variant of DOC-2. Motifs were detected in DOC-1 variants from FGSC2489 and P4471 using the eukaryotic linear motif resource (ELM)[111]. All analyses were carried out using each program's default settings, except the JalView visualization was customized to show residue conservation as seen in Fig. 4.3.4-2.

4.2.3. Western blots

Protein extractions from germlings and western blots were performed as described in Jonkers et al., 2014[75]. 250 mL Erlenmeyer flasks with 100 mL liquid MM were inoculated to $1x10^6$ conidia per mL. fFlasks were incubated at 30°C shaking at 220 rpm for 2.5 hours, then incubated another 2.5 hours at 30°C without shaking. Germlings were harvested by vacuum filtration over a nitrocellulose membrane and frozen in liquid nitrogen. Frozen germlings were then bead beaten with 0.5 mm glass beads at liquid nitrogen temperatures for one minute. 300 µL protein extraction buffer (described in Pandey et al., 2004[90]) was added to each sample and they were bead beaten at room temperature for 15 seconds. Samples were then centrifuged for 30 minutes at 4°C and the protein extract was separated from the cell debris.

Protein concentrations in the extracts were estimated using a NanoDrop™ spectrophotometer (Thermo Fischer Scientific™). Volumes of extracts were adjusted such that all samples run on a single gel contained similar total protein. Extracts were denatured at 70°C in 1x NuPAGE™ LDS sample buffer (Invitrogen™) with 5% β-mercaptoethanol (by volume) for 10 minutes, then run on 7% tris-acetate-SDS polyacrylamide gels (NuPAGE™, Invitrogen™) with PageRuler™ pre-stained protein ladder (10 to 180 kDa, Thermo Fischer Scientific™). Gels were blotted onto PVDF membranes with transfer buffer containing 20 mM tris base, 150 mM glycine, 20% methanol (by volume).

Blots were incubated in TBST (1x TBS (VWR), 0.5% Tween-20 (Sigma Aldrich)) with 5% milk for one hour at room temperature, with a buffer change after 30 minutes. Then blots were probed overnight at 4°C with primary monoclonal antibodies in TBST with 0.5% milk. Blots were washed three times with TBST, and then probed for one hour at room temperature with secondary polyclonal antibodies linked to HRP in TBST. After three more washes with TBST, blots were developed with SuperSignal™ West Pico developer (Thermo

Fischer Scientific™) and imaged on a ChemiDoc™ XRS+ with ImageLab™ software (Bio-Rad).

4.2.4. Flow cytometry analyses

Most of the data presented in this chapter was obtained using induced death flow cytometry (IDFC). Detailed procedures for performing IDFC and dye flow cytometry (dyeFC) assays, and analyzing the resulting data can be found in Chapter 2, sections 2.4 through 2.8. A brief overview of my IDFC protocol follows: strains were incubated in MM agar slants at 30°C in dark for three days, then 25°C in light for seven days. Conidia were then suspended in water and filtered through cheesecloth to remove hyphae. Conidial suspensions were diluted to $3x10^7$ conidia per mL, and spread on MM plates solidified with Pluronic F-127, either alone or in an equal mix with a death-inducing (DI) strain. For each mixture of strains, the constituent individual strains were also plated alone. Plates were incubated at 30°C in dark for 4 hours before harvesting and washing the germlings. Germlings were then stained with propidium iodide (PI, Sigma-Aldrich) and SYTOX Blue (SB, Life Technologies) before running them and conidial samples on a BD LSR Fortessa X-20 flow cytometer using BD FACSDiva™ software (BD Biosciences).

The run speed of the cytometer was adjusted for each sample such that data acquisition rates were between 500 and 1000 events per second for germling samples, and no more than 2000 events per second for conidial samples. The following parameters were recorded: forward scatter area and height, Pacific Blue area, and PerCP-Cy5-5 area. At least 20,000 events were recorded from conidial samples, and enough events to yield at least 10,000 germinated cells after gating out conidia were recorded from remaining samples (~30,000 events was usually sufficient).

In each experiment, a sample of FGSC2489 mixed with the CG1 DI strain was used as a positive control and a mixture of FGSC2489 and the CG3 or CG5 DI strain was used as a negative control. These controls were used during data analysis and to validate the experimental and data acquisition conditions.

dyeFC experiments were performed following the IDFC protocol with the following differences: conidia were stained with either calcofluor white (CFW, also called Fluorescence Brightener 28, Sigma-Aldrich) or concanavalin A, Alexa Fluor™ 488 conjugate (CAF, Invitrogen™) before plating, viability dyes were not used, and the fluorescence parameters BUV 496 area and FITC area were recorded instead of Pacific Blue and PerCP-Cy5-5.

4.2.5. Flow cytometry data analysis

A detailed explanation of how flow cytometry communication assay data was analyzed is available in Chapter 2, sections 2.6 and 2.8. Data was analyzed using a custom MATLAB™ (version R2018b, MathWorks) script. Briefly, data from ungerminated cells was removed from germling data by comparison with data from conidial samples. Fluorescence gates were then defined using the experiment's positive control fluorescence distributions in the SB and PI channels. Germinated cells more fluorescent than the gate in either channel were identified as dead. Relative death rates were calculated by dividing the cell-death percentage in a mixture of two strains by the average cell-death percentage of the constituent strains by themselves.

Relative death rates from three or more replicate experiments were combined and analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests with a significance threshold of 0.05. Usually SB and PI results agreed; in these cases, figures present only the PI data. Occasionally the positive control gave anomalously low or variable relative death rates in the SB or PI channel, resulting in a loss of statistical power. In these cases, figures present data from the channel retaining high quality controls.

4.3. Bioinformatic predictions

4.3.1. Functional domains could not be identified in the DOC proteins

Twenty-seven alleles of *doc-1*, including multiple representatives from all five CGHs, were translated and scanned for domains and protein database matches using HMMScan, PHMMER, blastp, and CDD searches[105,106,112]. No significant domain-matches were found, and homology was detected only to uncharacterized proteins from close relatives of *N. crassa*. DOC-1 from FGSC2489 was also scanned for transmembrane domains and signal peptides using SPOCTOPUS[110]: nothing was detected.

A similar search was conducted on 28 alleles of *doc-2* and *doc-3*, including variants from CGH1-CGH4 (CGH5 contains only *doc-1*)[42]. Homology was detected only to uncharacterized proteins from *N. crassa*'s close relatives. Although a few domain-matches reached significance, none were consistently detected in all alleles from any CGH. However, several of the domain-matches that were found come from proteins that bind nucleotides (e.g. guanine exchange factors, DNA or RNA-binding proteins). These matches weren't pursued further because they weren't consistently detected. SPOCTOPUS identified a single-pass transmembrane domain in DOC-2 from FGSC2489. Assuming that such a significant feature would be present in all functional variants of DOC-2, I analyzed DOC-2^{CGH3} from P4471 using SPOCTOPUS. Nothing was detected in the CGH3 variant, suggesting the transmembrane domain identified in DOC-2^{CGH1} was probably spurious. This does not suggest the peripheral localization of DOC-2-GFP observed by Heller et al., 2016, was artifactual; DOC-2 may associate with the plasma membrane indirectly or through other means.

4.3.2. The DOC-1 consists of a globular core with N- and C-terminal unstructured regions

DOC-1 from FGSC2489 was scanned for globular and unstructured regions using GlobPlot[109]. This analysis revealed a large globular core covering over 50% of the protein's sequence (Fig. 4.3.5). DOC-1's N-terminal ~200 residues were predicted unstructured regions. DOC-1 also has a C-terminal ~200 amino acid region that was predicted to be unstructured. The unstructured and globular regions detected by GlobPlot approximately align with CGH-variable and conserved regions, respectively, identified in section 4.3.4. The size of the predicted globular domain also coincides with the stable DOC-1 degradation product I obtained while attempting to purify full-length DOC-1 from *E. coli* (data not shown).

4.3.3. DOC-1 may contain protein-protein interaction motifs

In addition to relatively large, structured domains, proteins may also use short, usually unstructured motifs to interact with other proteins or molecules[113]. These motifs are usually between 3 and 15 amino acids in length, making confident bioinformatic detection difficult and false-positive matches common. However, because I could not detect any functional domains in DOC-1, I used the Eukaryotic Linear Motif Resource (ELM) to search for interesting motifs in the CGH1 and CGH3 variants of DOC-1[111]. ELM returned 37 potential motifs for DOC-1^{CGH1} and 35 for DOC-1^{CGH3}. Of these, five motifs were identified with slightly higher confidence, were shared between CGH1 and CGH3 variants, and were involved in potentially relevant phenomena: one LIG_CaM_IQ_9 motif, two DOC_MAPK_DCC_7 motifs, one LIG_WW_1 motif, and one DOC_PP2B_PxIxI_1 motif. I found these names cumbersome, so I will refer to them as IQ, KIM, WW, and PP2B.

IQ motifs are recognized and bound by calmodulin, light-chain myosins, and other EF-hand family proteins[114]. Other proteins containing this motif include myosins, phosphatases, GTPase-activating proteins, channels, and receptors. However, IQ motifs are commonly identified in protein databases, and only a small subset of matches has been experimentally confirmed to direct protein-protein interactions. Nonetheless, the association of IQ motifs with signaling and cytoskeletal elements made this motif worth investigating. Interactions between DOC-1 and myosins could explain how the DOC system suppresses communication through manipulation of the cytoskeleton or intracellular transport.

KIM motifs are specific docking sites for members of the ERK1/2 and p38 MAPK subfamilies[115,116]. They are distinct from phosphorylation sites and are present on substrates, scaffolds, and kinase regulators. The potential presence of an KIM motif in DOC-1 is promising because *N. crassa*'s MAK-2 cascade is orthologous to the mammalian ERK1/2 pathway[76]. Given DOC-1's dynamic localization with the MAK-2 complex, identifying motifs that mediate interactions between DOC-1 and components of this complex would be valuable. However, DOC-1 is also phosphorylated by unknown kinases[117]: if real, this KIM motif may direct DOC-1's phosphorylation rather than allow DOC-1 to regulate any component of the MAK-2 complex.

WW motifs are docking sites recognized by WW domains. Named after a pair of tryptophan residues required for partner recognition, WW domains are widespread, modular protein-protein interaction domains[118]. WW motifs are proline-rich sequences found in diverse proteins. Although the presence of a WW motif in DOC-1 doesn't inspire specific hypotheses like the other motifs I've selected, WW domains and motifs are involved in a number of potentially relevant processes, such as cytoskeletal interactions with the extracellular matrix[119].

PP2B motifs are required for substrate recognition by calcineurin, a calmodulin and calcium-regulated protein phosphatase[120]. Calcineurin was previously shown to be involved in the maintenance of apical growth in *N. crassa*, so it may also affect communication and chemotropic growth[121]. While the presence of phosphorylation sites and a PP2B motif on DOC-1 suggests it may be regulated by calcium levels, the motif could also recruit calcineurin to the MAK-2 complex where its phosphatase activities could interrupt communication signaling.

The approximate locations of these motifs in DOC-1 are shown in Fig. 4.3.3. I deleted the sequences coding for each of these motifs from *doc-1* and attempted to test how the manipulated alleles affected CG phenotype. Unfortunately, no transformants could be obtained for strains expressing *doc-1* $^{\Delta IQ}$ or *doc-1* $^{\Delta KIM}$. This suggests deleting the IQ or KIM motifs somehow renders DOC-1 toxic. Results for the remaining motif deletion alleles are discussed in section 4.4.

4.3.4. Potential signatures of communication group specificity across *doc-1* alleles

The specificity domain hypothesis predicts regions of *doc-1* alleles that vary between CGHs, but are conserved within a CGH, are responsible for CG specificity. Following the example of Hall et al., 2010, I scanned a multiple sequence alignment of the coding sequences of 27 alleles of *doc-1*, including representatives from all five CGHs, for signatures of selection using the Selecton server[107,108]. The non-synonymous-to-synonymous mutation ratio (K_a/K_s) and Tajima's D statistic were calculated for each codon in the alignment[122,123]. K_a/K_s values under one at a codon suggest purifying selection on the amino acid it encodes. Values of K_a/K_s greater than one suggest directional or diversifying selection. Values near one imply no selection. For Tajima's D, negative values indicate directional or purifying selection, while values above zero indicate balancing selection. Values near zero suggest no selection. However, both Tajima's D and K_a/K_s can only be calculated for aligned regions: they do not tolerate gaps in the alignment, so all gapped codons were removed prior to analysis.

By scanning across *doc-1* alleles, codon by codon, I hoped to find some regions under diversifying and/or balancing selection and others under purifying selection. The regions under balancing or diversifying selection would likely contain specificity determinants, while the regions under purifying selection should contain conserved sequences required for interacting with other cellular components. Unfortunately, Tajima's D was greater than one for over 80% of the aligned codons, and no regions with obvious deviations could be found (Fig. 4.3.4-1). This indicates nearly the entire sequence of DOC-1 is subject to balancing selection. K_a/K_s spiked above one several times near the N- and C-termini, but remained low through most of DOC-1's center. These data suggest the N- and C-terminal regions are under diversifying selection while DOC-1's core is conserved.

Because Tajima's D and K_a/K_s cannot evaluate unaligned regions, these metrics overlook indels. Indels encoding NSR specificity determinants have previously been identified in *N. crassa*'s HET-C heterokaryon incompatibility protein[39], so I thought they could be important for DOC-1 as well. I visually examined an amino acid multiple sequence alignment of the same 27 DOC-1 variants and noticed a number of indels shared within CGHs, but not between CGHs (Fig. 4.3.4-2). These indels were clustered near the N- and C-termini of DOC-1, approximately corresponding to the regions GlobPlot predicted to be unstructured. These indels may control CG-specificity, and I tested this hypothesis in section 4.4.

4.3.5. A model of DOC-1

Integrating the results from bioinformatic predictions, the following model for DOC-1 emerges (Fig. 4.3.3):

- DOC-1 has a relatively conserved, globular, 403 amino acid core. This middle region

contains putative KIM motifs that could mediate interactions with kinases in the MAK-2 complex. The response domain hypothesis predicts that this core should be required for DOC-1 to influence communication.

- The N-terminal 220 amino acids of DOC-1 are predicted to be unstructured. This N-terminal region contains a putative IQ motif that may link the DOC system to cellular calcium levels. Most importantly, this region contains CGH-specific indels and amino acids under diversifying selection that the specificity hypothesis predicts should control CG specificity.
- Like the N-terminal region, the C-terminal 205 residue region is also unstructured and contains CGH-specific indels and amino acids under diversifying selection. As such, the specificity hypothesis also predicts this region should control CG phenotype. The C-terminal region contains a putative WW protein-protein interaction motif and a putative calcineurin substrate recognition motif, although both motifs are near the border with the conserved core.

With this model in mind, I set out to test the specificity and response domain hypotheses. In order to test the specificity domain hypothesis, I also divided DOC-1^{CGH3} into N-terminal, middle, and C-terminal regions, based on that variant's alignment with DOC-1^{CGH1}'s regions (Fig. 4.3.3). Strains expressing truncated and chimeric alleles of *doc-1* were constructed and their CG phenotypes were assayed, as described in sections 4.4 and 4.5.

4.4. Support for the specificity domain hypothesis

4.4.1. The N- and C-terminal regions of DOC-1 are involved in CG recognition, and the C-terminal region is required for independent suppression of communication

The model of DOC-1 outlined in section 4.3.5 predicts the CGH-specific indels in N- and C-terminal regions should control CG specificity. First, I used dyeFC and IDFC to test the CG phenotypes of strains expressing *doc-1* N- and C-terminal truncation alleles in *Δdoc-1* and *Δdoc-1 Δdoc-2* genetic backgrounds (see Fig. 4.4.1-1 for western blots showing detection of truncated proteins). The CG phenotypes of strains assayed in this section are summarized in Table 4.4.1 for easy reference.

DyeFC tests indicated *Δdoc-1 doc-*$1^{\Delta 1\text{-}220}$*-gfp* retains high self-communication (Fig. 4.4.1-2). However, the *Δdoc-1* strain's self-communication was never assayed via dyeFC, so this result must be interpreted cautiously. IDFC results showed DOC-$1^{\Delta 1\text{-}220}$-GFP suppresses *Δdoc-1*'s communication with CG3, indicating it retains some function (Fig. 4.4.1-3 panel A). However, unlike full-length DOC-1-GFP, the N-terminal truncation did not increase communication with CG1, suggesting the N-terminal region is responsible for CG1 recognition.

Δdoc-1 Δdoc-2 doc-$1^{\Delta 1\text{-}220}$*-gfp*'s self-communication could not be distinguished from positive or negative controls by dyeFC (Fig. 4.4.1-2). These data suggest the strain's self-communication is either somewhat reduced or highly variable. However, the non-self-communication phenotype generated by DOC-$1^{\Delta 1\text{-}220}$-GFP was striking: *Δdoc-1 Δdoc-2 doc-*$1^{\Delta 1\text{-}220}$*-gfp* did not communicate with CG1, CG3, or CG5 (Fig. 4.4.1-3 panel B). These data indicate that DOC-$1^{\Delta 1\text{-}220}$-GFP cannot respond to CG signals, but still suppresses

communication, suggesting the N-terminal region is required for DOC-1 to respond to CG signals, but is dispensable for communication suppression.

Δdoc-1 doc-1$^{Δ624-828}$-gfp displayed self-communication rates that could not be differentiated from positive or negative controls by dyeFC (Fig. 4.4.1-2), suggesting the C-terminal truncation allele does not complement the self-communication defect induced by *Δdoc-1* as well as the N-terminal or full-length variants. This strain's non-self communication behavior was similar to that of *Δdoc-1 doc-1$^{Δ1-220}$-gfp*: reduced communication with CG3, without increasing communication with CG1 or CG5 (Fig. 4.4.1-4 panel A). These data suggest the C-terminal region is also required for CG1 recognition.

DyeFC assays on *Δdoc-1 Δdoc-2 doc-1$^{Δ624-828}$-gfp* gave highly variable self-communication measures, with a mean similar to that of the positive control and a borderline significant difference from the negative control (P_a = 0.053, Fig. 4.4.2-1), suggesting that DOC-1$^{Δ624-828}$-GFP may not affect self-communication. IDFC indicated a non-self-communication phenotype similar to *Δdoc-1 Δdoc-2*, but with super communication with CG5 (Fig. 4.4.1-4 panel B). Together, the effects of DOC-1$^{Δ624-828}$-GFP expression in the *Δdoc-1* mutant, and the inability of this truncated protein to alter the communication behavior of the *Δdoc-1 Δdoc-2* mutant suggests DOC-1's C-terminal region is required for communication suppression without DOC-2 and may also be involved in CG recognition.

Results from truncation alleles analyzed in this section are summarized in Table 4.4.1. Taken together, it appears the N-terminal region of DOC-1 is required for CG recognition, but not for communication suppression. The C-terminal region may also play a role in CG recognition, and is required for communication suppression without DOC-2. Neither terminal region appears necessary for interactions with DOC-2, as either truncation allele still suppresses communication with CG3 when DOC-2 is present.

4.4.2. The N- and C-terminal regions of DOC-1 must coordinate with the middle region to properly specify CG

Having demonstrated that the N- and C-terminal regions of DOC-1 are involved in CG recognition, I tested the effects of CGH1/CGH3 chimeric *doc-1* alleles on the CG phenotypes of *Δdoc-1* and *Δdoc-1 Δdoc-2* mutants (see Fig. 4.4.2-1 for western blots showing detection of chimeric proteins). If the N- and C-terminal regions are fully responsible for CG specificity, chimeric alleles with matching N- and C-termini should mimic unaltered CGH1 and CGH3 variants tested in chapter 3. The CG phenotypes of strains analyzed in this section are summarized in Table 4.4.2 for easy reference. For brevity, I'll refer to chimeric alleles using a letter and a number referring to the part of DOC-1 and the CGH from which each part originated, respectively. For example, DOC-1^{N3M1C1} refers to a variant of DOC-1 with the N-terminal region from a P4471 (a CGH3 strain), and the middle and C-terminal regions from FGSC2489 (a CGH1 strain).

First, I tested the CGH1/CGH3 N-terminal chimeras. A *Δdoc-1 doc-1^{N3M1C1}-gfp* strain communicated at intermediate rates with CG1 and CG3 (Fig. 4.4.2-2 panel A). This phenotype is identical to that induced by expression of full-length DOC-1^{CGH1}-GFP in the *Δdoc-1* background, although DOC-1^{CGH3}-GFP produces a similar phenotype in the same background. DOC-1^{N3M1C1}-GFP in the *Δdoc-1 Δdoc-2* background reduced communication with CG5, increased communication with CG3, and marginally increased communication with CG1 (statistically comparable to negative control and communication with CG5, Fig.

4.4.2-2 panel B). This phenotype is distinct from those produced by the expression of either full-length variant, and suggests DOC-1^{N3M1C1}-GFP is capable of suppressing communication independently, and responds to both CG3 and CG1 signals.

Expression of the inverse N-terminal chimera, DOC-1^{N1M3C3}-GFP, did not alter CG phenotype in a *Δdoc-1 Δdoc-2* background (Fig. 4.4.2-3 panel B). However, this chimeric variant increased communication with CG1 and CG5, without decreasing communication with CG3, in a *Δdoc-1* background (Fig. 4.4.2-3 panel A). These data suggest DOC-1^{N1M3C3}-GFP interacts with DOC-2 but does not suppress communication with any CG.

Next, I tested the CGH1/CGH3 C-terminal chimeras. *Δdoc-1 doc-1^{N1M1C3}-gfp* did not communicate with CG1, CG3, or CG5, although communication with CG1 was insignificantly greater than communication with other CGs (Fig. 4.4.2-4 panel A). This phenotype is similar to that of a *Δdoc-1 doc-1^{CGH3}-gfp* strain, but the chimeric allele reduced communication with CG3 much further. In the *Δdoc-1 Δdoc-2* background, DOC-1^{N1M1C3}-GFP abolished communication with CG5, increased communication with CG3 to intermediate levels, and caused a marginal increase in communication with CG1 (Fig. 4.4.2-4 panel B). These results suggest DOC-1^{N1M1C3}-GFP suppresses communication and responds to both CG1 and CG3 signals by itself, but cannot properly respond to CG signals when coexpressed with DOC-2. Such behavior may indicate aberrant interactions between DOC-1^{N1M1C3}-GFP and DOC-2, or it may indicate a lack of coordination between the two proteins such that they compete to suppress communication with all CGs.

Expressing the inverse C-terminal chimera, DOC-1^{N3M3C1}-GFP, did not affect CG phenotype in the *Δdoc-1 Δdoc-2* background (Fig. 4.4.2-5 panel B). In the *Δdoc-1* background, expressing this chimeric variant gave the inverse phenotype from that induced by DOC-1^{N1M1C3}-GFP: low communication with all CGs, but insignificantly elevated communication with CG3 (Fig. 4.4.2-5 panel A). These data suggest DOC-1^{N3M3C1}-GFP still interacts with DOC-2, but cannot suppress communication on its own.

Finally, I tested the CGH1/CGH3 middle region chimeras. *Δdoc-1 Δdoc-2 doc-1^{N1M3C1}-gfp* phenocopied a *Δdoc-1 Δdoc-2* mutant, communicating only with CG5 (Fig. 4.4.2-6 panel B). However, the communication behavior of *Δdoc-1 doc-1^{N1M3C1}-gfp* was identical to that of *Δdoc-1 doc-1^{CGH3}-gfp*: low communication with CG5, insignificantly elevated communication with CG1, and high communication with CG3, but communication with CG1 and CG3 were statistically comparable (Fig. 4.4.2-6 panel A). These results could suggest DOC-1^{N1M3C1}-GFP interacts with DOC-2 as if it were a CGH3 variant, but it may also indicate this chimera weakly responds to CG1 signals and cannot inhibit DOC-2's preference for CG3. In either case, the chimeric allele cannot suppress communication independently of DOC-2.

The inverse middle region chimera, DOC-1^{N3M1C3}-GFP, suppressed communication with CG5 in the *Δdoc-1 Δdoc-2* background, but increased communication with CG1 to intermediate levels (Fig. 4.4.2-7 panel B). *Δdoc-1 doc-1^{N3M1C3}-gfp* communicated well with CG1 and at intermediate rates with CG3, although communication with CG1 and CG3 were statistically comparable (Fig. 4.4.2-7 panel A). The phenotype produced when DOC-1^{N3M1C3}-GFP interacts with DOC-2 is reminiscent of that produced by *Δdoc-1 doc-1^{CGH1}-gfp*, but the chimeric variant appears to specify CG1 more strongly than full length DOC-1-GFP in both *Δdoc-1* and *Δdoc-1 Δdoc-2* genetic backgrounds. However, interpreting these results is difficult because *Δdoc-1 doc-1^{CGH1}-gfp* and *Δdoc-1 doc-1^{CGH3}-gfp* have similar CG phenotypes.

Results from strains expressing chimeric alleles analyzed in this section are summarized in Table 4.4.2. Overall, the N- and C-terminal regions of DOC-1 clearly exert

control over CG specificity. However, the middle region probably also plays a role in specifying CG; otherwise, the middle region chimeras should have phenocopied the full-length DOC-1 variants more closely. This may be explained by hypothesizing inter-region interactions between matched terminal and middle region variants. In any case, it appears all three regions contain some specificity determinants.

4.4.3. The WW and PP2B motifs may contribute to DOC-1 functions, and suggest the border between the middle and C-terminal regions was misidentified

Some of DOC-1's functions may be mediated through the motifs putatively identified in section 4.3.3. I was only able to obtain transformants for *doc-1*$^{\Delta WW}$*-gfp* and *doc-1*$^{\Delta PP2B}$*-gfp*, and only free GFP could be detected from backcrossed homokaryons via western blot (data not shown), suggesting the motif-deleted proteins are unstable. Nonetheless, both the WW and PP2B motifs are in the C-terminal region, so I tested strains supposedly expressing these motif-deletion alleles to see if they phenocopy *doc-1*$^{\Delta 624\text{-}828}$*-gfp*. If the motif-deletion variants function like the C-terminal truncation variant, it would support the existence of these putative motifs. Such a result would also suggest the C-terminal region functions, at least in part, through these motifs.

In the *Δdoc-1* background, expressing DOC-1$^{\Delta WW}$-GFP did not affect CG phenotype, showing high communication with CG3 and low communication with CG1 and CG5 (Fig. 4.4.3-1 panel A). The WW motif-deletion variant did suppress communication with CG5 to intermediate levels in the *Δdoc-1 Δdoc-2* background (Fig. 4.4.3-1 panel B). Neither of these phenotypes resemble those induced by the expression of DOC-1$^{\Delta 624\text{-}828}$-GFP (Fig. 4.4.1-4), but they are identical to the phenotypes observed for strains expressing DOC-1$^{\Delta 221\text{-}623}$-GFP (discussed in section 4.5.1, Fig. 4.5.1). These results strongly suggest I misidentified the border between the middle and C-terminal regions of DOC-1, and the WW motif is actually within the conserved core. Furthermore, these results indicate DOC-1's middle region requires the WW motif for functionality.

Expression of the PP2B motif-deletion allele resulted in a more unique phenotype: total suppression of communication in the *Δdoc-1* background, and a reduction in communication with CG5 in the *Δdoc-1 Δdoc-2* background (Fig. 4.4.3-2). These results suggest DOC-1$^{\Delta PP2B}$-GFP partially suppresses communication, does not respond to CG signals, and interacts with DOC-2 to suppress its CG3 preference. Perhaps the PP2B motif mediates interactions between DOC-1 and DOC-2 via calcineurin phosphatase activity, although many more experiments are required to confirm this speculation.

4.5. Testing the response domain hypothesis

4.5.1 Deleting DOC-1's middle region impairs communication suppression

The model for DOC-1 presented in section 4.3.5 predicts the conserved middle region should be required for basic DOC-1 functionality, including DOC-1's default suppression of communication. To test this hypothesis, I assayed the CG phenotypes of strains expressing DOC-1$^{\Delta 221\text{-}623}$-GFP using dyeFC and IDFC (see Fig 4.4.1-1 for western blots showing expression). The CG phenotypes of strains examined throughout section 4.5 are summarized in Table 4.5 for easy reference.

DyeFC measurements of *Δdoc-1 doc-1$^{Δ221-623}$-gfp*'s self-communication could not be statistically distinguished from positive or negative controls (Fig. 4.4.1-2). However, this strain's self-communication rate was among the lowest of the mutants I tested via dyeFC. As measured via IDFC, *Δdoc-1 doc-1$^{Δ221-623}$-gfp* reproduced the *Δdoc-1* mutant's non-self-communication phenotype (Fig 4.5.1 panel A). These results are as expected if DOC-1's middle region is required for core functionality.

Expressing DOC-1$^{Δ221-623}$-GFP in the *Δdoc-1 Δdoc-2* background induced self-communication levels similar to those seen in *Δdoc-1 doc-1$^{Δ221-623}$-gfp* (Fig. 4.4.1-2). But results for *Δdoc-1 Δdoc-2 doc-1$^{Δ221-623}$-gfp*'s non-self-communication phenotype suggest DOC-1$^{Δ221-623}$-GFP retains some ability to limit communication, as communication with CG5 was reduced to intermediate levels (Fig 4.5.1 panel B). This phenotype is identical to that of *Δdoc-1 Δdoc-2 doc-1^{CGH5}-gfp*, which I interpreted as an artifact of the GFP tag. However, I doubt a GFP tag could rescue DOC-1's ability to suppress communication without the middle region unless the N- and/or C-terminal regions retained some communication-suppressing functions. Nonetheless, it appears the conserved middle region plays a major role in DOC-1's core function of suppressing communication.

4.5.2. DOC-1's middle and terminal regions may share communication-suppressing functions

Given that DOC-1$^{Δ1-220}$ could not derepress communication by itself (Fig. 4.4.1-3 panel B), and that DOC-1$^{Δ624-828}$ could not suppress communication by itself (Fig. 4.4.1-4 panel B), I wanted to test how deleting both terminal regions would affect CG phenotype. According to the response domain hypothesis, expressing DOC-1$^{221-623}$ should globally reduce communication. Because GFP tags interfere with DOC-1's functions, I constructed a *doc-1$^{221-623}$-V5* cassette to test the functionality of DOC-1$^{221-623}$ (see Fig 4.4.1-1 for western blots showing expression).

Expressing DOC-1$^{221-623}$-V5 in the *Δdoc-1* genetic background did not alter the mutant's CG phenotype (Fig. 4.5.2 panel A). Strangely, *Δdoc-1 Δdoc-2 doc-1$^{221-623}$-V5* phenocopied *Δdoc-1 Δdoc-2 doc-1$^{Δ221-623}$-gfp* (Fig. 4.5.2 panel B). This behavior may indicate some redundancy between the middle region and one or both of the terminal regions, but it may also be an artifact of the regional borders I defined in section 4.3.4.

4.5.3. Over-expressing DOC-1's middle region reduces communication with CG1

Presumably, DOC-1's colocalization with the MAK-2 complex relates to the communication suppressing functions of the DOC system[42]. If the middle region of DOC-1 mediates interactions between DOC-1 and other communication machinery (e.g. the MAK-2 complex), over-expressing DOC-1$^{221-623}$ should compete with full length DOC-1 for interaction sites with the MAK-2 complex. However, without the N-terminal region, DOC-1$^{221-623}$ should not be able to derepress communication. Thus, if DOC-1's middle region is responsible for communication suppression, over-expressing DOC-1$^{221-623}$ in a wild-type background should reduce, but not abolish, communication with CG1.

I used IDFC to test the CG phenotype of a strain expressing DOC-1$^{221-623}$-GFP (see Fig 4.4.1-1 for western blots showing expression). The over-expression strain did not communicate with CG3 or CG5, and communication with CG1 was reduced to intermediate levels (Fig. 4.5.3). These results support the response domain hypothesis, and suggest

coexpressed DOC-1 variants compete for interaction sites with other communication machinery.

Results from section 4.5 are summarized in Table 4.5. In contrast to N- and C-terminal truncations from section 4.4.1, DOC-1$^{\Delta 221\text{-}623}$-GFP could not suppress communication with CG3 in a *Δdoc-1* genetic background. Nonetheless, DOC-1$^{\Delta 221\text{-}623}$-GFP retains some ability to suppress communication, as indicated by suppression of communication with CG5 in a *Δdoc-1 Δdoc-2* background. These data suggest that DOC-1's middle region is responsible for interactions with DOC-2. However, DOC-1$^{221\text{-}623}$-V5 also failed to suppress communication with CG3 in a *Δdoc-1* background. Together, these results imply DOC-1's middle region and some portion of either the N-terminal or C-terminal regions are required for interactions with DOC-2.

4.6. Limited investigation of DOC-2

4.6.1. DOC-2's N-terminal region is not required to suppress communication

Although this work focuses on DOC-1, *doc-2* is paralogous to *doc-1* and also has important roles in specifying CG (see chapter 3)[42]. Therefore, I analyzed DOC-2 variants for CGH-specific indels and structural domains as described for DOC-1 variants in section 4.3.2 and 4.3.4 (data not shown). As for DOC-1, these analyses revealed CGH-specific indels concentrated in DOC-2's terminal regions. In contrast to DOC-1, GlobPlot only predicted a large unstructured region in the first ~200 residues of DOC-2. I tested the CG phenotypes of strains expressing DOC-2$^{\Delta 1\text{-}220}$-GFP in the *Δdoc-1 Δdoc-2* genetic background using IDFC. I also tested strains expressing DOC-2 with the putative transmembrane domain deleted. Although a transmembrane domain was identified only in CGH1 variants by SPOCTOPUS, DOC-2's localization at the cellular periphery merited a test of DOC-2$^{\Delta TM}$-GFP's functionality[42]. Strains expressing manipulated *doc-2* alleles in the *Δdoc-1* background were not tested because *doc-2* is dispensable for CG1 specificity in IDFC tests. Expression of both DOC-2$^{\Delta 1\text{-}220}$-GFP and DOC-2$^{\Delta TM}$-GFP were confirmed via western blot (Fig. 4.6.1-1).

Expression of DOC-2$^{\Delta 1\text{-}220}$-GFP reduced communication with CG5 to intermediate levels in the *Δdoc-1 Δdoc-2* background (Fig. 4.6.1-2 panel A). This suggests that DOC-2's N-terminal region is not required to suppress communication, although this suppression is not as extreme as seen for *Δdoc-1 Δdoc-2 doc-1$^{\Delta 1\text{-}220}$-gfp*.

Tests of strains expressing DOC-2$^{\Delta TM}$-GFP indicate this variant is non-functional (Fig. 4.6.1-2 panel B). However, unaltered *doc-2^{CGH1}-gfp* also failed to complement the *Δdoc-1 Δdoc-2* mutant, so one cannot determine the importance of DOC-2's putative transmembrane domain from these results.

4.7. Summary and discussion

In this chapter, I attempted to functionally characterize DOC-1. First, I identified potential specificity determinants and conserved regions in DOC-1. I then tested the functional relevance of these regions in specifying CG using truncated and CGH1/CGH3 chimeric alleles. My results suggest the more variable N- and C-terminal regions of DOC-1 play a major role in CG specificity, although the middle region is also involved. Either the middle or C-terminal regions are required for communication suppression. The N-terminal region is required to alleviate that suppression. The DOC-1's middle region and some

portion of either terminal region are needed for interactions with DOC-2. I also determined DOC-2's N-terminal region is not required for communication suppression.

The specificity domain hypothesis predicted regions of *doc-1* that vary between CGHs, but not within CGHs, should control CG specificity. I identified CGH-specific indels concentrated in the N- and C-terminal regions of DOC-1. Then I assayed the CG phenotypes of strains expressing alleles of *doc-1* in which the N- or C-terminal regions were deleted. These results demonstrated the N-terminal region is required for derepression of communication by DOC-1, suggesting this region may contain a major specificity determinant. The phenotypes of strains expressing CGH1/CGH3 chimeric alleles of *doc-1* showed the middle and C-terminal regions also contribute to CG specificity. However, the effects of deleting the WW motif from DOC-1 showed the border between the middle and C-terminal region was misidentified, calling into question my results for middle or C-terminal truncations and chimeras. The more variable N- and C-terminal regions might completely control CG specificity if proper domain borders could be defined. Overall, my results provide tentative support for the specificity domain hypothesis, although CG specificity may involve interactions between all three regions of DOC-1.

The response domain hypothesis is nearly the inverse of the specificity domain hypothesis: regions conserved between CGH variants of DOC-1 are probably involved in core functions, including localization and communication suppression. Deleting the C-terminal region abolishes DOC-1's ability to suppress communication without DOC-2. Strangely, suppression is restored if the N-terminus is also deleted. But deleting the middle region of DOC-1 does not abolish its suppression of communication with CG5 in the *Δdoc-1 Δdoc-2* background. These observations suggest that the middle and C-terminal regions of DOC-1 have communication suppressing functions, and the terminal regions regulate the middle region. The C-terminal region seems to promote communication suppression while the N-terminal region inhibits suppression. Finally, overexpressing *doc-1*$^{221\text{-}623}$*-gfp* in a wild-type background caused a drop in communication with CG1, confirming that the middle region must be involved in communication suppression and suggesting that coexpressed variants of DOC-1 compete for interaction partners during communication. Integrating these results paints an ambiguous picture regarding the response domain hypothesis. The conserved middle region isn't solely responsible for DOC-1's default suppression of communication. However, the middle region is involved in interactions with DOC-2 and communication suppression, supporting the response domain hypothesis. A parsimonious explanation is that the response domain hypothesis is correct: conserved regions are solely responsible for communication suppression and interactions with other cellular machinery. However, to demonstrate this clearly, one must include conserved parts of the N- and C-terminal regions as well as the middle region.

Accounting for the results presented in chapters 2-4, an updated model of DOC-mediated CG specificity is summarized in Fig. 4.7. As they begin to communicate, a CG-specific signal must be sent between cells. These signals are validated through interactions between the middle region of DOC-1 and DOC-2: matched CGH-variants of the DOC proteins are generally required for proper CG specificity. If the DOC system of a cell does not receive a compatible CG signal, DOC-1 and DOC-2 both prevent reinforcement of MAK-2 complex oscillations and suppress communication. This suppression involves the actions of the middle and C-terminal regions of DOC-1, and does not require the N-terminal region of DOC-2. If a compatible CG signal is detected, the DOC system allows MAK-2 complex

oscillations and chemotropic growth to continue. DOC-1's N-terminal region is required to properly derepress communication, and is probably involved in validating CG signals.

This work could be improved in at least four ways, listed below:

- The patterning of inter-CGH conserved and variable regions in DOC-1 are more complicated than can be captured by dividing the protein into three contiguous regions. Construction of alleles with many or all of the variable regions deleted or switched from CGH1 to CGH3 would probably validate the specificity domain hypothesis more clearly. Likewise, the response domain hypothesis could be unambiguously evaluated using alleles from which many or all conserved regions were deleted. This was not attempted here because cloning such alleles would be very difficult and the resulting proteins, especially if conserved regions were deleted, would probably be unstable. However, as costs of gene synthesis continue to fall, it may become feasible to order such alleles ready for use.
- Experimental results would almost certainly improve if untagged or small epitope-tagged alleles were used. Results from chapter 2 clearly demonstrated that GFP tags partially interfere with DOC-1's functions, and V5 tags were superior.
- The self-communication levels of most strains studied in this chapter were not evaluated. Testing self-communication via IDFC requires time-consuming strain construction for which no selectable markers are currently available. I attempted to evade this problem by developing and using dyeFC, but this assay requires further improvement and troubleshooting and might never yield the statistical power produced with IDFC assays.
- In the model I present for DOC-mediated CG recognition, I assume DOC-1 must interact with the MAK-2 complex (and possibly other communication machinery) to suppress communication with other strains from incompatible CGs. This aspect of the model could be evaluated by observing whether the GFP-tagged DOC-1 variants I tested here oscillate with MAK-2 during communication.

Despite these shortcomings, I believe this work has added important details to our understanding of the DOC system and to possible mechanisms underlying long-distance NSR in general.

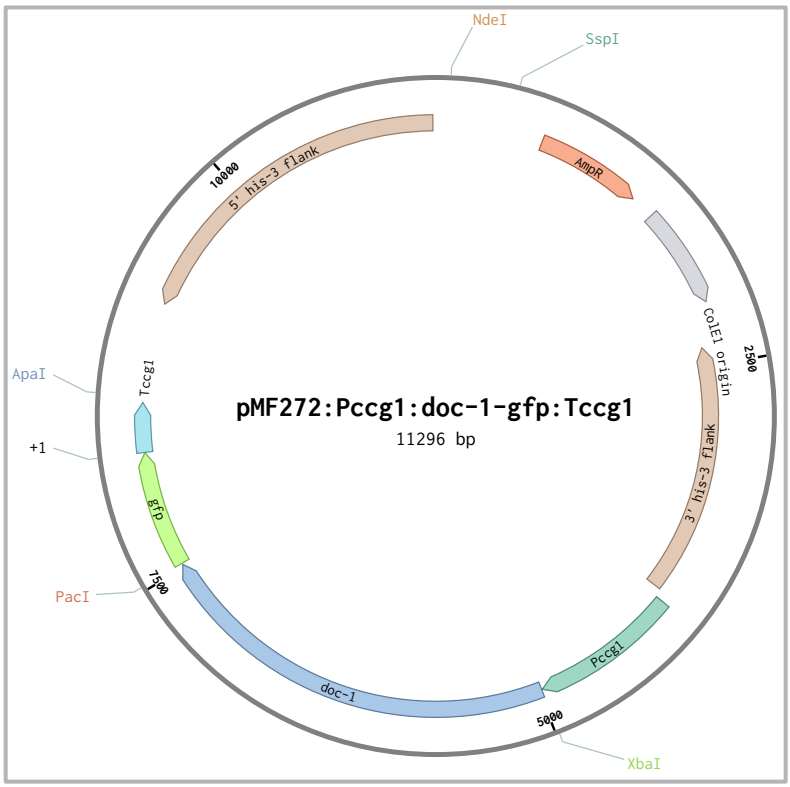
Table 4.2.1-1: Strains referenced in chapter 4

Wild isolates and classical mutants

Strain	Genotype	Reference
FGSC2489 (CG1)	Laboratory Wild Type, *mat A*	FGSC
P4471 (CG3)	Wild isolate from the Louisiana Population	Bhat & Vyas, 2003
FGSC4200	Laboratory Wild Type, *mat a*	FGSC
FGSC6103	Laboratory Wild Type, *his-3 mat A*	FGSC
FGSC9716	Laboratory Wild Type, *his-3 mat a*	FGSC

Manipulated strains (all in FGSC2489 genetic background, all deletions marked with hygR)

Strain	Genotype	Reference
Δdoc-1	*Δdoc-1 mat a*	Dunlap et. al., 2007, & Heller et. al., 2016
Δdoc-1 His- A	*his-3; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 His- a	*his-3; Δdoc-1 mat a*	This study
Δdoc-1 Δdoc-2 (CG5)	*Δdoc-1 Δdoc-2 mat A*	Heller et. al., 2016
Δdoc-1 Δdoc-2 His- A	*his-3; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 His- a	*his-3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG3-swap strain	*his-3:: doc-1-CGH3 doc-2-CGH3; Δdoc-1 Δdoc-2 mat a*	Heller et. al., 2016
CG1 DI strain	*his-3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat A*	Heller et. al., 2018
CG3 DI strain	*his-3:: doc-1-CGH3 doc-2-CGH3 Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+ mat ?*	This study
CG5 DI strain	*Δplp-1 Δplp-2 sec-9:: sec-9-GRD3 his-3+; Δdoc-1 Δdoc-2 mat ?*	This study
Δdoc-1 Δdoc-2 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 Δdoc-2 mat A*	This study
Δdoc-1 Δdoc-2 doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp; Δdoc-1 Δdoc-2 mat a*	This study
Δdoc-1 doc-1-CGH1-gfp	*his-3:: doc-1-CGH1-gfp; Δdoc-1 mat A*	Heller et. al., 2016
Δdoc-1 doc-1-CGH1-V5	*his-3:: doc-1-CGH1-V5; Δdoc-1 mat ?*	This study
Δdoc-1 doc-1-CGH3-gfp	*his-3:: doc-1-CGH3-gfp; Δdoc-1 mat a*	This study



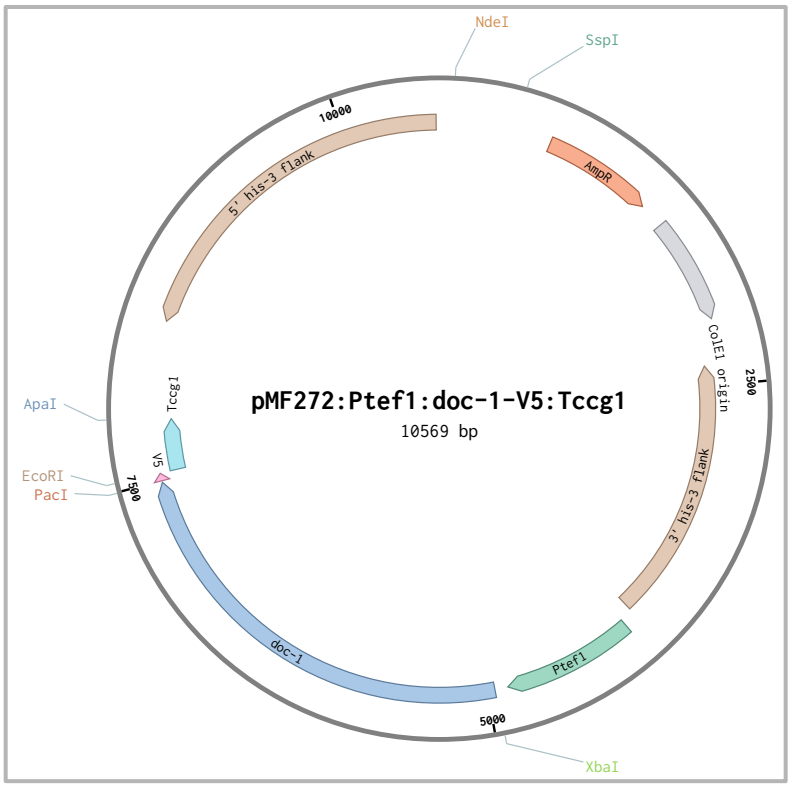


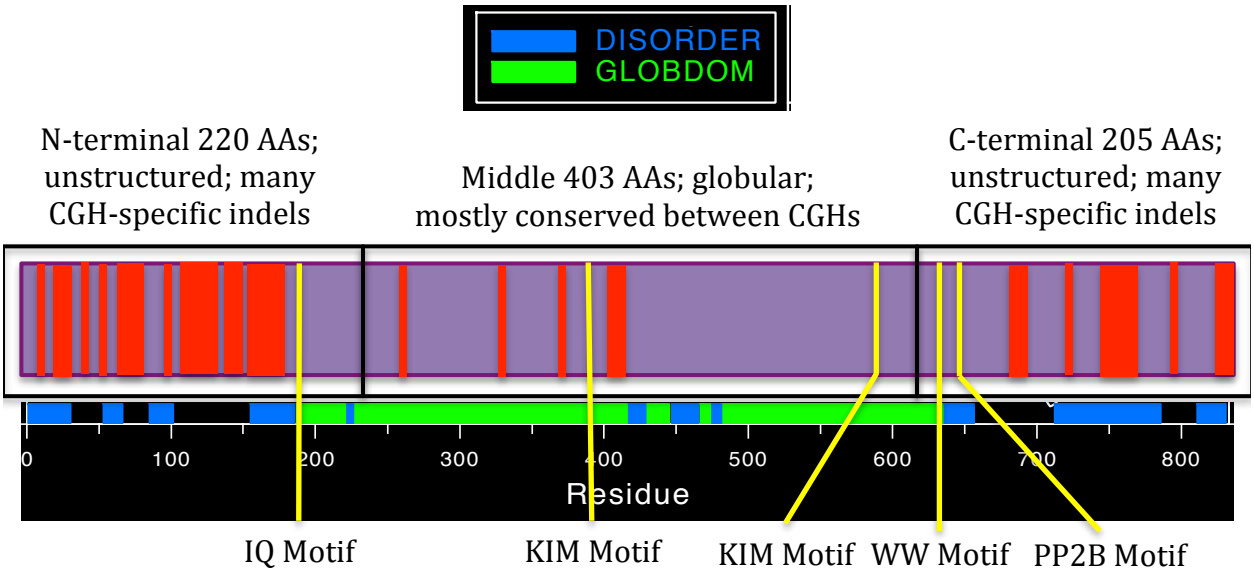
Figure 4.2.1. Plasmid maps of vectors used in this study
Vectors are described in chapter 2, and depicted with *doc-1*^{CGH1} cloned into them. Other alleles were cloned into them using XbaI and PacI restriction sites.

Table 4.2.1-2: Primers used for cloning in chapter 4			
Name	**Sequence(5'-3')**	**Paired with**	**Purpose**
Xba1-ATG-doc-1-ΔN220-f	GTCACTGTCTAGATGCCTGACAAGGAACTTCTATTGG	Pac1-doc1-r*	Forward primer to amplify *doc-1* without its N-terminal region, adding an *Xba*I site and start codon.
Pac1-doc1-r*	TTAATTAAAGCAATAGGCAAATCC	Xba1-doc-1-f or Xba1-ATG-doc-1-ΔN220-f	Reverse primer to amplify *doc-1* with a C-terminal *Pac*I site.
Xba1-doc-1-f	TCTAGATGAGTAGCGGCAAG	Pac1-doc1-r* or doc-1-ΔC205-r	Forward primer to amplify *doc-1* with an N-terminal *Xba*I site.
doc-1-ΔC205-r	CGCCCTTGCTCACCATGTTAATTAAGCGGGGGGCTTG	Xba1-doc-1-f	Reverse primer to amplify *doc-1* without its C-terminal region, adding a *Pac*I site.
pMF272-f	GCAAGCTTCAGCTGCTCGAGTTCTATAGTGTCACCTAAATCGTATGTG	doc-1-ΔM403-r, doc-1-N220-chim-r, doc-1-ΔC205-chim-r, doc-1-CG3-N320-r, or doc-1-CG3-M418-r	Forward primer used in fusion PCR cloning schemes for *doc-1*.
doc-1-ΔM403-r	CGGGTTAGGACGTACCACTCAGAAGGTCCGAGGGCTGCTG	pMF272-f	Reverse primer for amplifying pMF-doc-1-ΔM403 for fusion PCR to generate *doc-1-Δ221-623*.
doc-1 Intron	GTACGTCCTAACCCGTCACCCCTGATAGAATCTTGCTAACTGTGTCGTAG		Intronic sequence used for fusion PCR to generate *doc-1-Δ221-623*.
doc-1-ΔM403-f	TGCTAACTGTGTCGTAGCTCAAAAGGGAATGGAAGAGGTGCAAC	pMF272-r	Reverse primer for amplifying doc-1-ΔM403-pMF for fusion PCR to generate *doc-1-Δ221-623*.
pMF272-r	GAACTCGAGCAGCTGAAGC	doc-1-ΔM403-f, doc-1-ΔN220-chim-f, doc-1-C205-chim-f, doc-1-CG3-M418-f, or doc-1-CG3-C169-f	Reverse primer used in fusion PCR cloning schemes for *doc-1*.
HIS-3-r	CTCTCGAGTCCCGTTATTGC	pMF272-rev	Forward nested primer used in fusion PCR cloning schemes for *doc-1*.
pMF272-rev	CGTCCTTGAAGAAGATGGTGC	HIS-3-r	Reverse nested primer used in fusion PCR cloning schemes for *doc-1*.
doc-1-ΔN220-chim-f	CCTGACAAGGAACTTCTATTGG	pMF272-r or doc-1-ΔC205-chim-r	Forward primer to amplify *doc-1* without its N-terminal region or only the middle region, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-N220-chim-r	GTCCGAGGGCTGCTG	pMF272-f	Reverse primer to amplify *doc-1*'s N-terminal region, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-ΔC205-chim-r	TGCGCGGGGGGCTTG	pMF272-f or doc-1-ΔN220-chim-f	Reverse primer to amplify *doc-1* without its C-terminal region or only the middle region, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-C205-chim-f	GGAATGGAAGAGGTGCA	pMF272-r	Forward primer to amplify *doc-1*'s C-terminal region, used in fusion PCR cloning schemes for *doc-1* chimeras.

Table 4.2.1-2: Primers used for cloning in chapter 4 continued			
Name	**Sequence (5'-3')**	**Paired with**	**Purpose**
doc-1-CG3-N-f	CCCTCACATCAACCAAATCTAGATGGGCACCGGTCTC	doc-1-CG3-N320-r or doc-1-CG3-M418-r	Forward primer to amplify *doc-1-CGH3*'s N-terminal or N-terminal and middle regions, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-CG3-N320-r	TCTCCAATAGAAGTTCCTTGTCAGGCTCAGATGTGAATGACATCG	doc-1-CG3-N-f	Reverse primer to amplify *doc-1-CGH3*'s N-terminal region, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-CG3-M418-f	TCCCATTCAGCAGCAGCCCTCGGACGCCGATAAAGAGCTTCTGC	doc-1-CG3-M418-r or doc-1-CG3-C-r	Forward primer to amplify *doc-1-CGH3*'s middle or middle and C-terminal regions, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-CG3-M418-r	TTGGTTGTTGCACCTCTTCCATTCCTTGGGGGCTTTGTGC	doc-1-CG3-N-f or doc-1-CG3-M418-f	Reverse primer to amplify *doc-1-CGH3*'s N-terminal or N-terminal and middle regions, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-CG3-C169-f	CTGCCACGTCCAAGCCCCCCGCGCACCCAACGCTATTCC	doc-1-CG3-C-r	Forward primer to amplify *doc-1-CGH3*'s C-terminal region, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc-1-CG3-C-r	CGCCCTTGCTCACCATGTTAATTAACGACGTCATGAACCCTAA	doc-1-CG3-C169-f or doc-1-CG3-M418-f	Reverse primer to amplify *doc-1-CGH3*'s C-terminal or C-terminal and middle regions, used in fusion PCR cloning schemes for *doc-1* chimeras.
doc1-ΔWW-r	AGACTTTTCACGCTCGCTG	Xba1-doc-1-f	Reverse primer used to amplify *doc-1* without its WW motif.
doc1-ΔWW-f	CACGACGGCAGACCG	Pac1-doc1-r*	Forward primer used to amplify *doc-1* without its WW motif.
doc1-ΔPP2B-r	GCCGTCGTGGTAGCCG	Xba1-doc-1-f	Reverse primer used to amplify *doc-1* without its PP2B motif.
doc1-ΔPP2B-f	TCCCCCGCCATCGAA	Pac1-doc1-r*	Forward primer used to amplify *doc-1* without its PP2B motif.
doc-2-ΔN220-f	AACCCCTCACATCAACCAAATCTAGATGGGCCTTGCTCGCAAG	Pac1-doc-2-r*	Forward primer to amplify *doc-2* without its N-terminal region, adding an *Xba*I site and start codon.
Pac1-doc-2-r*	TTAATTAATGAACCCAGCTCAAG	doc-2-ΔN220-f or doc-2-fusion-f	Reverse primer to amplify *doc-2* with a C-terminal *Pac*I site; also reverse nested primer used in fusion PCR for *doc-2*.
doc-2-fusion-f	AAATCAACACAACACTCAAACC	Pac1-doc-2-r*	Forward nested primer used in fusion PCR cloning schemes for *doc-2*.
doc-2-ΔTM-r	GGCCCAGAAGGGGTTGGCGCCGCCGTTCTTGGCGGCGTC	HIS-3-r	Reverse primer to amplify *doc-2* without its putative transmembrane domain, used in fusion PCR cloning scheme.
doc-2-ΔTM-f	CGGCGGCGCCAACCCCTTCTGGGCCTTCAATGCCTTGCAGC	pMF272-rev	Forward primer to amplify *doc-2* without its putative transmembrane domain, used in fusion PCR cloning scheme.

DOC-1 Model

828 amino acids in FGSC2489 (CGH1)



907 amino acids in P4471 (CGH3)

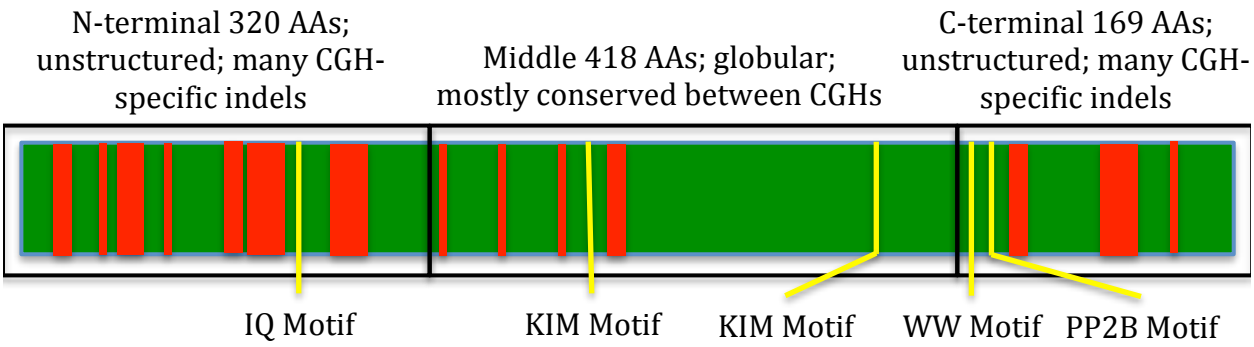


Figure 4.3.3. A speculative model for DOC-1

A model for DOC-1 based on bioinformatic predictions. A schematic of DOC-1^{CGH1} from FGSC2489 is shown in blue on top. N-terminal, middle, and C-terminal regions are boxed. Red bars depict CGH-specific indels (not to scale) and yellow lines indicate motifs. Globular domains and disordered regions are indicated beneath the schematic with green and blue boxes, respectively. A schematic for DOC-1^{CGH3} from P4471 is shown at the bottom in green with N-terminal, middle, and C-terminal regions boxed.

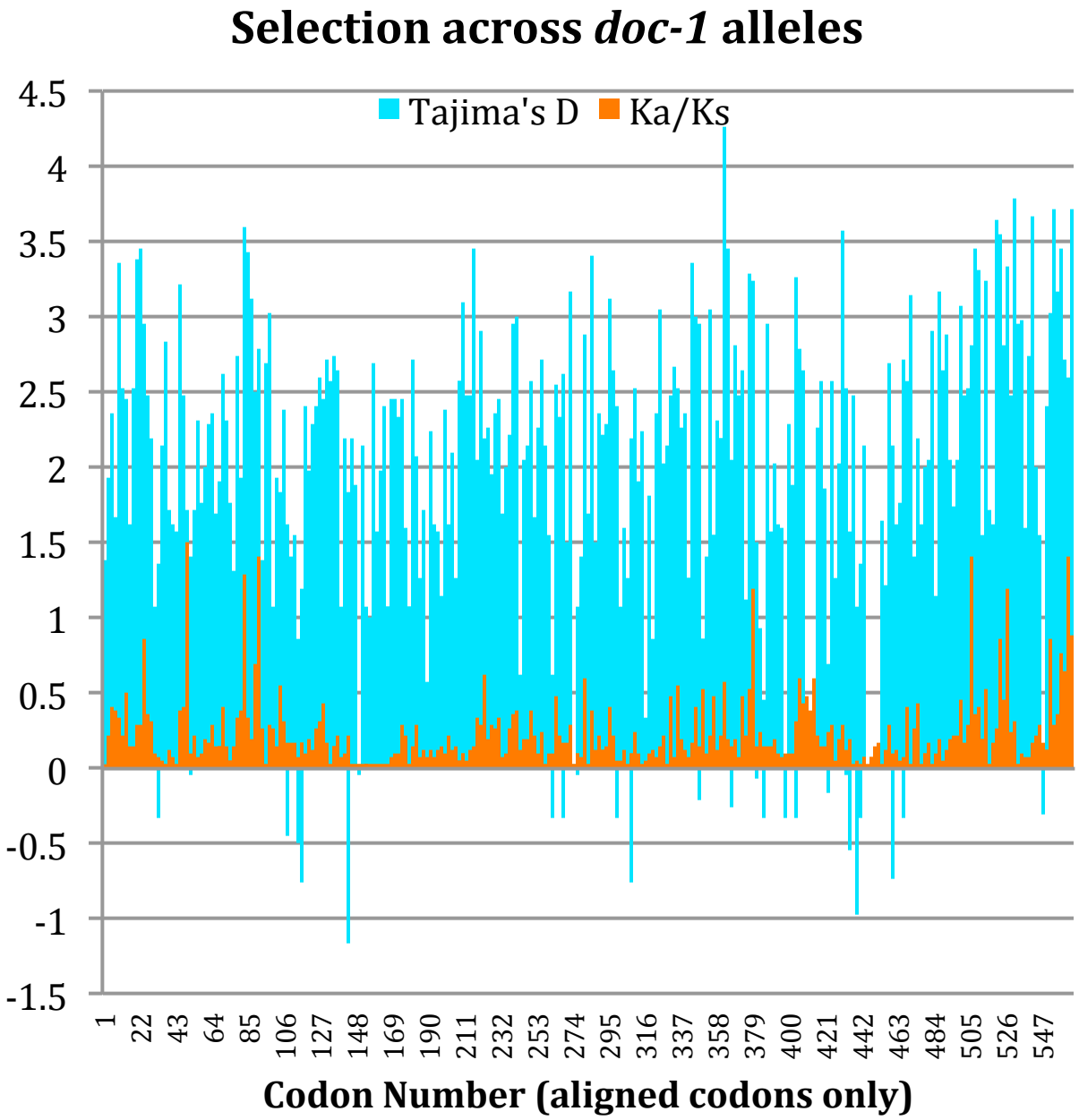


Figure 4.3.4-1. K_a/K_s and Tajima's D codon scan across *doc-1* alleles
A MSA of 27 *doc-1* alleles, including representatives from all five CGHs, had its gaps removed and was analyzed using the Selecton server. K_a/K_s values for each aligned codon are indicated with orange bars and Tajima's D values are indicated with cyan bars.

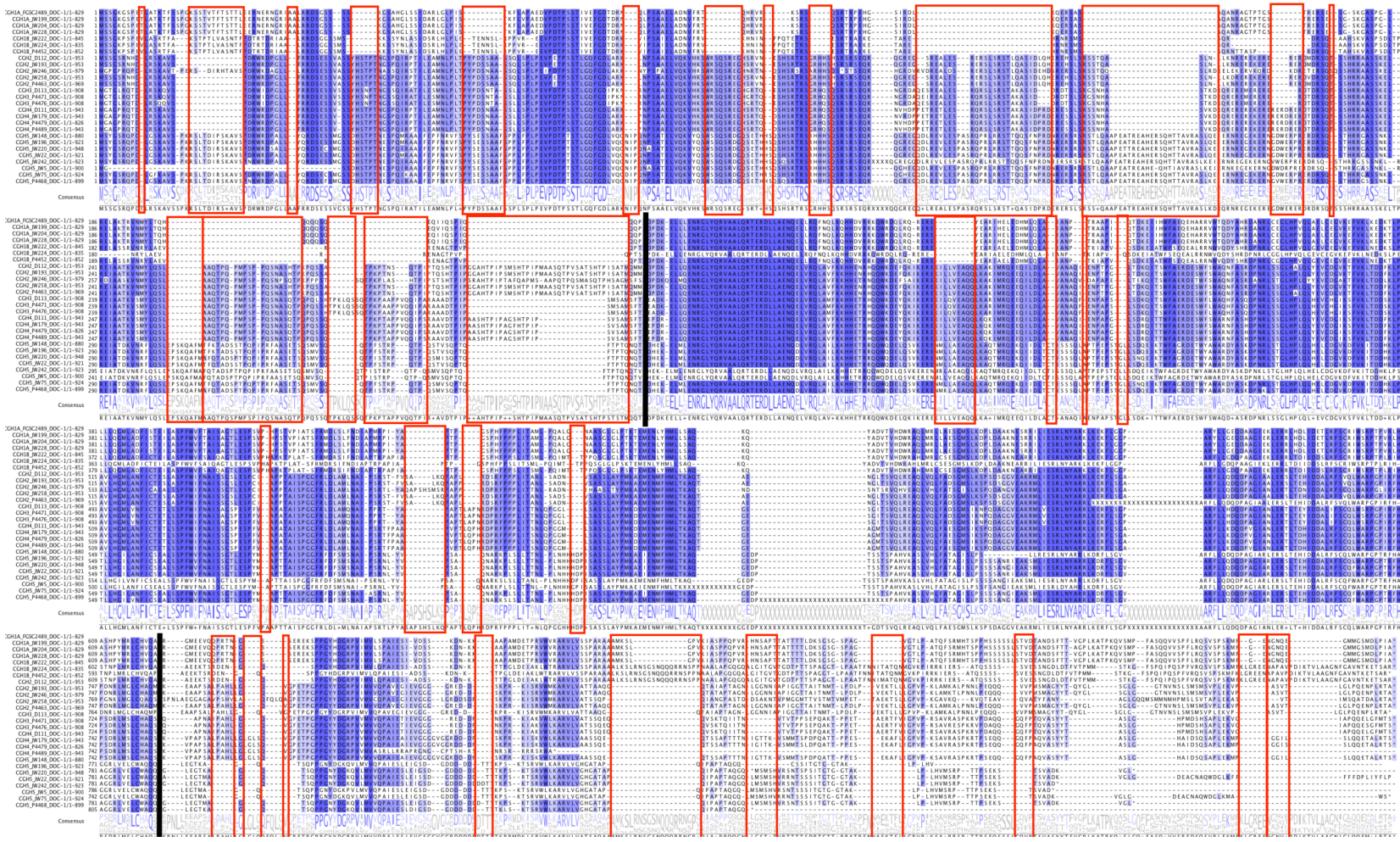


Figure 4.3.4-2. DOC-1 multiple sequence alignment

MSA of 27 DOC-1 variants, including representatives from all five CGHs. For the evolutionary relationships between strains, and alleles, see Heller et al., 2016. Sequences were translated and aligned using MACSE. MSA was visualized and colored using Jalview. Darker shades of blue indicate greater conservation. Red boxes indicate CGH-specific indels. Black lines demarcate borders between the N-terminal (220 residues in FGSC2489), middle (403 residues in FGSC2489), and C-terminal regions (205 residues in FGSC2489).

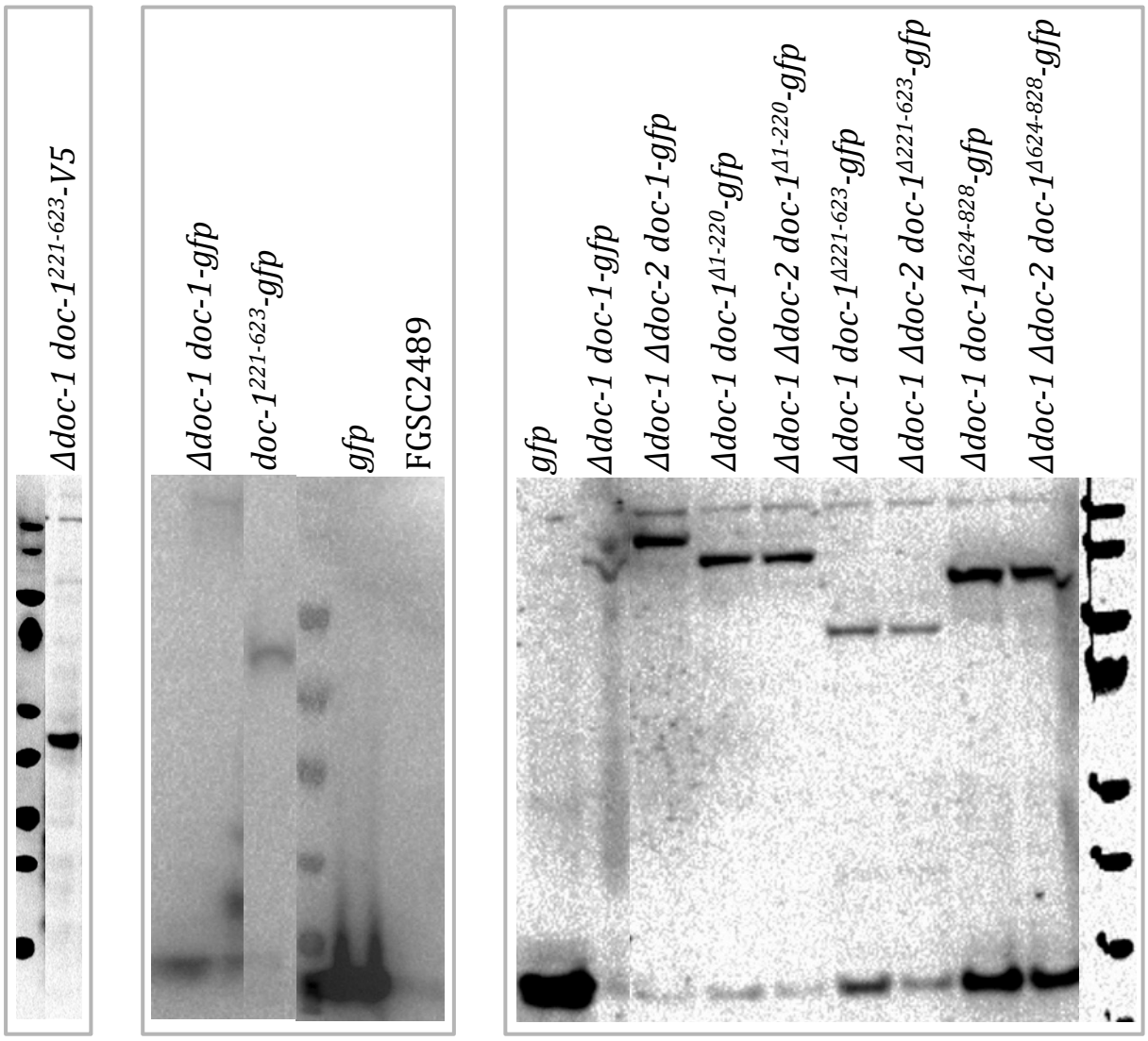


Figure 4.4.1-1. Western blots showing expression of DOC-1 truncation and internal deletion variants

Western blots showing detection of GFP and V5-tagged DOC-1^{CGH1} truncation alleles in *Δdoc-1* and *Δdoc-1 Δdoc-2* genetic backgrounds. Samples were loaded to normalize total protein content across the lanes in each blot. See section 4.2 for more details. Separate blots are outlined in gray boxes. Expected band sizes: full length DOC-1-GFP = ~119 kDa, DOC-1$^{\Delta 1\text{-}220}$-GFP = ~95 kDa, DOC-1$^{\Delta 221\text{-}623}$-GFP = ~73 kDa, DOC-1$^{\Delta 624\text{-}828}$-GFP = ~97 kDa, DOC-1$^{221\text{-}623}$-GFP = ~74 kDa, DOC-1$^{221\text{-}623}$-V5 = ~48 kDa, free GFP = ~27 kDa. Marker bands have the following sizes in kDa: 150, 120, 85, 65, 50, 40, 30, 25. For unknown reasons, the 150 and 120 kDa marker bands in the middle blot are very faint.

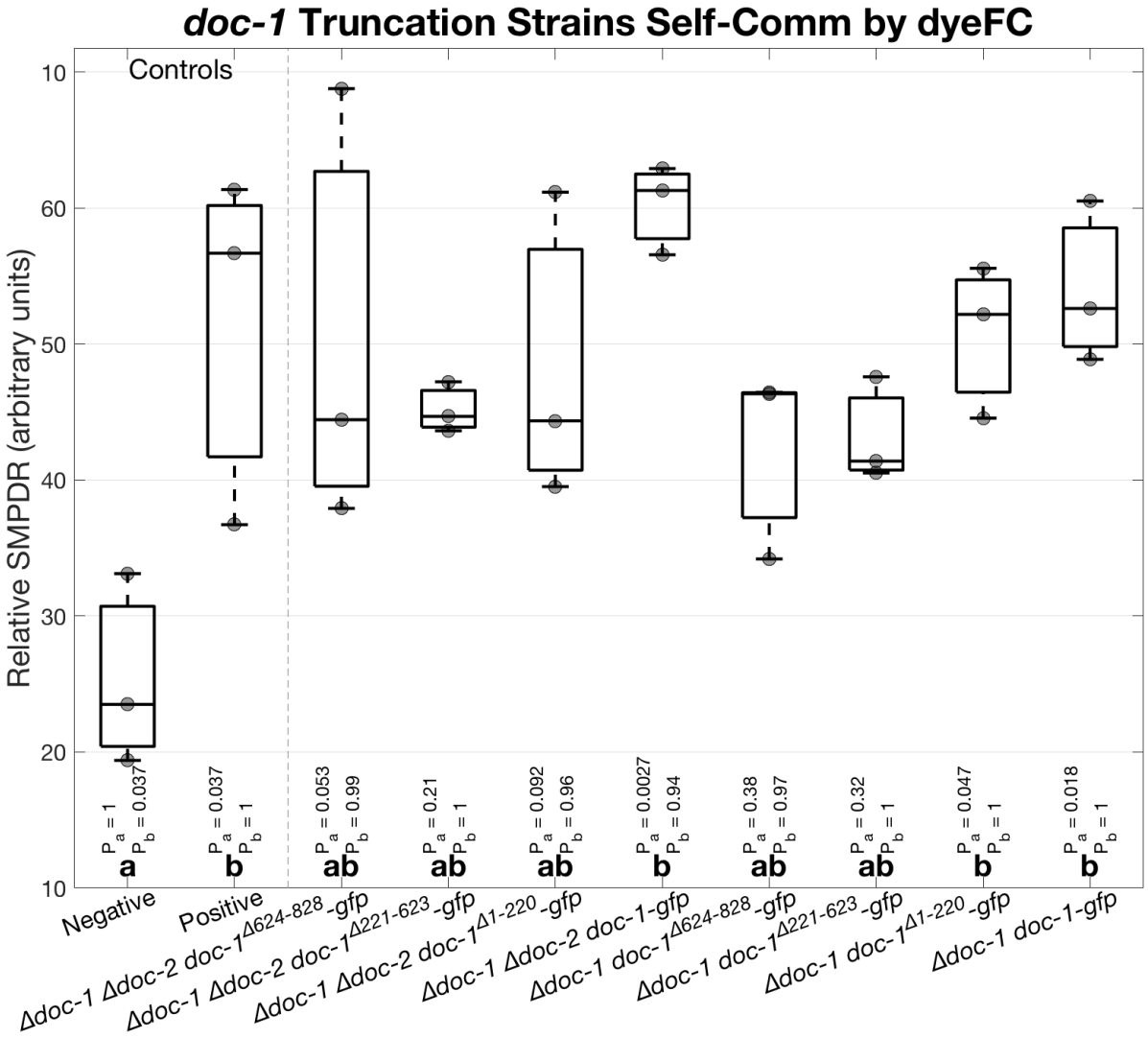


Figure 4.4.1-2. Self-communication phenotypes of *doc-1* truncation mutants via dyeFC

DyeFC boxplot summary of self-communication in *doc-1* truncation mutants. Details in the Results section. SMPDR = saddle-to-max probability density ratio; an explanation of the relative saddle-to-max probability density ratio can be found in Chapter 2. Results for the negative and positive controls are indicated on the left, separated from other samples by a gray dashed line. The negative control sample was FGSC2489 CFW + *Δdoc-1 Δdoc-2 doc-1*$^{\Delta 1\text{-}220}$*-gfp* CAF. The positive control sample was FGSC2489 CFW + FGSC2489 CAF. All other samples were mixtures of the indicated strains stained CFW and CAF. Data from triplicate experiments is shown. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold lowercase letters just above the x-axis indicate statistical groups with p-values less than 0.05, as determined via ANOVA and HSD multiple comparison tests. P-values for each strain's self-communication compared with the negative control (P_a) and the positive control (P_b) are displayed just above the statistical groups.

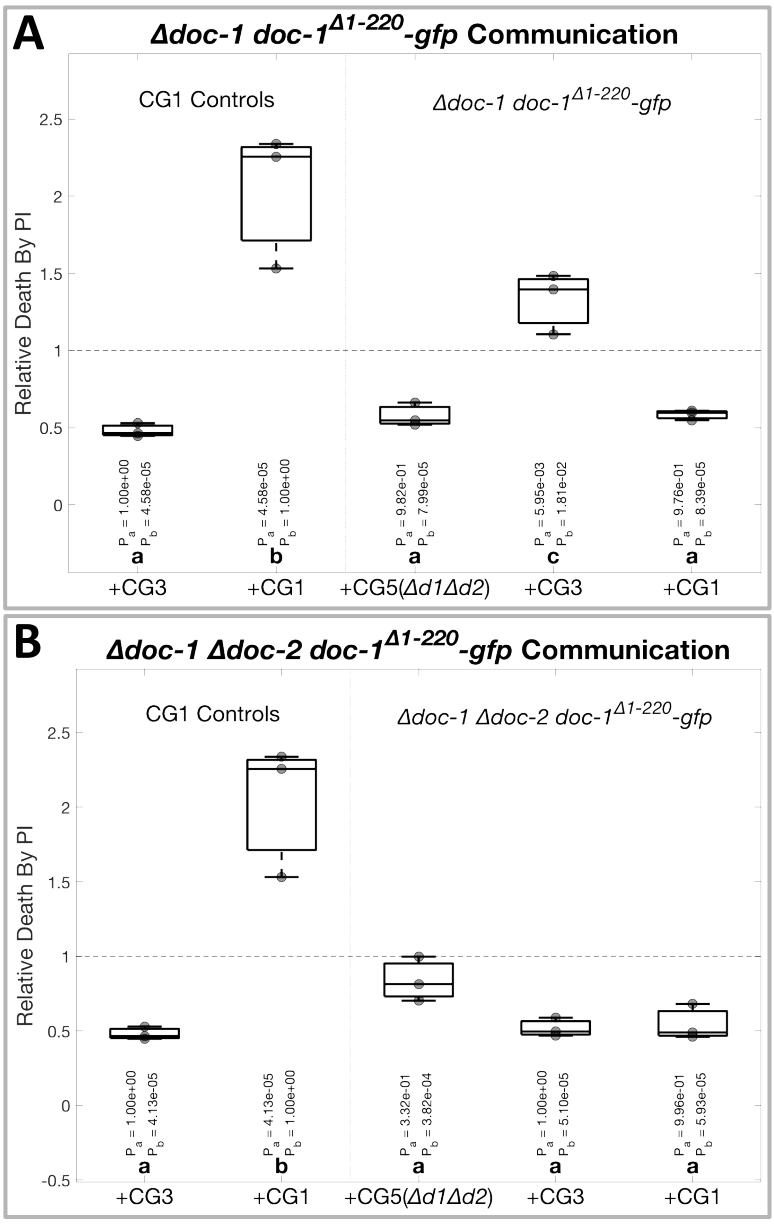


Figure 4.4.1-3. IDFC results from strains expressing N-terminal truncation alleles of *doc-1*

Boxplot summary of CG phenotypes of strains expressing *doc-1*$^{\Delta 1\text{-}220}$*-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1*$^{\Delta 1\text{-}220}$*-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1*$^{\Delta 1\text{-}220}$*-gfp*.

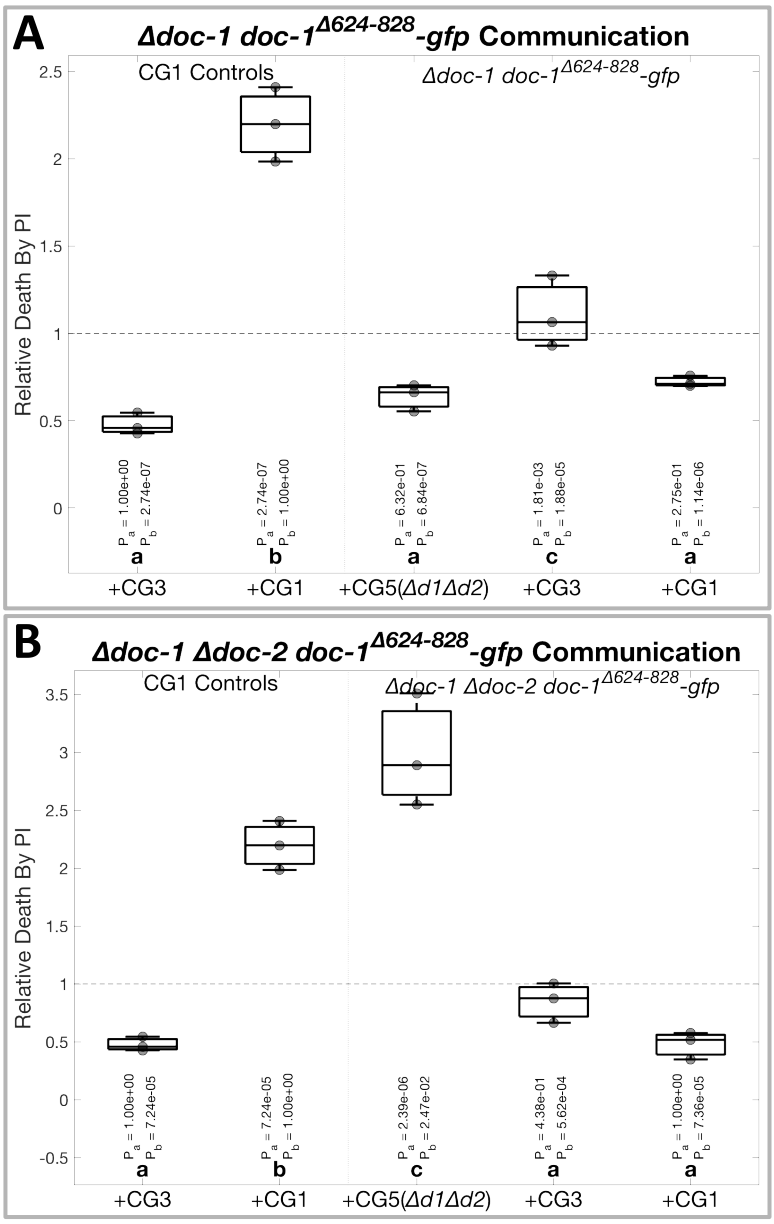


Figure 4.4.1-4. IDFC results from strains expressing C-terminal truncation alleles of *doc-1*

Boxplot summary of CG phenotypes of strains expressing *doc-1*$^{\Delta 624\text{-}828}$*-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1*$^{\Delta 624\text{-}828}$*-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1*$^{\Delta 624\text{-}828}$*-gfp*.

Table 4.4.1. Summary of CG phenotypes from section 4.4.1

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	**Comm w/CG5**	**Comm w/CG3**	**Comm w/CG1**
FGSC2489			
Δdoc-1			
Δdoc-1 doc-1-gfp			
Δdoc-1 Δdoc-2			
Δdoc-1 Δdoc-2 doc-1-gfp	Higher	Lower	
Δdoc-1 doc-1$^{\Delta 1\text{-}220}$-gfp			
Δdoc-1 Δdoc-2 doc-1$^{\Delta 1\text{-}220}$-gfp			
Δdoc-1 doc-1$^{\Delta 624\text{-}828}$-gfp			
Δdoc-1 Δdoc-2 doc-1$^{\Delta 624\text{-}828}$-gfp	Super		

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. Control data reproduced from chapter 2.

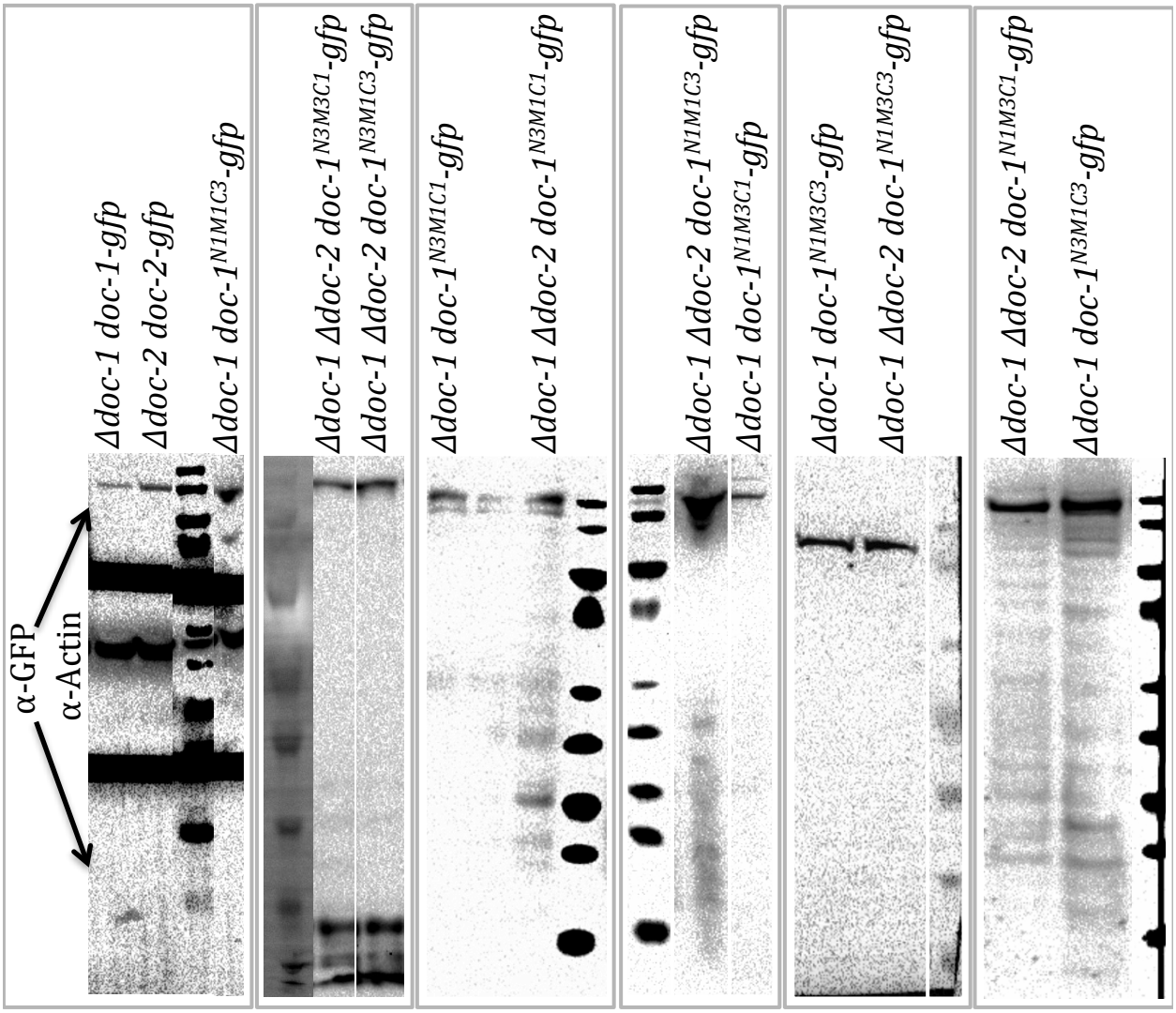


Figure 4.4.2-1. Western blots showing expression of DOC-1 CGH1/ CGH3 chimeras

Western blots showing detection of GFP-tagged DOC-1 CGH1/CGH3 chimeric alleles in *Δdoc-1* and *Δdoc-1 Δdoc-2* genetic backgrounds. Samples were loaded to normalize total protein content across the lanes in each blot. See section 4.2 for more details. For clarity, separate blots are outlined in gray boxes. The blot in the bottom panel was cut into three sections prior to detection, and the high and low weight sections were probed for GFP while the middle section was probed for actin. All other blots were probed only for GFP. Expected band sizes: DOC-1^{N3M1C1}-GFP = 130622.9 Da, DOC-1^{N1M3C3}-GFP = 114795.7 Da, DOC-1^{N1M3C1}-GFP = 118204.5 Da, DOC-1^{N3M1C3}-GFP = 127214.2 Da, DOC-1^{N1M1C3}-GFP = 115430.4 Da, DOC-1^{N3M3C1}-GFP = 129988.2 Da. DOC-1^{N3M3C1}-GFP could not be detected in the *Δdoc-1* background (blot not shown). Marker bands have the following sizes in kDa: 150, 120, 85, 65, 50, 40, 30, 25.

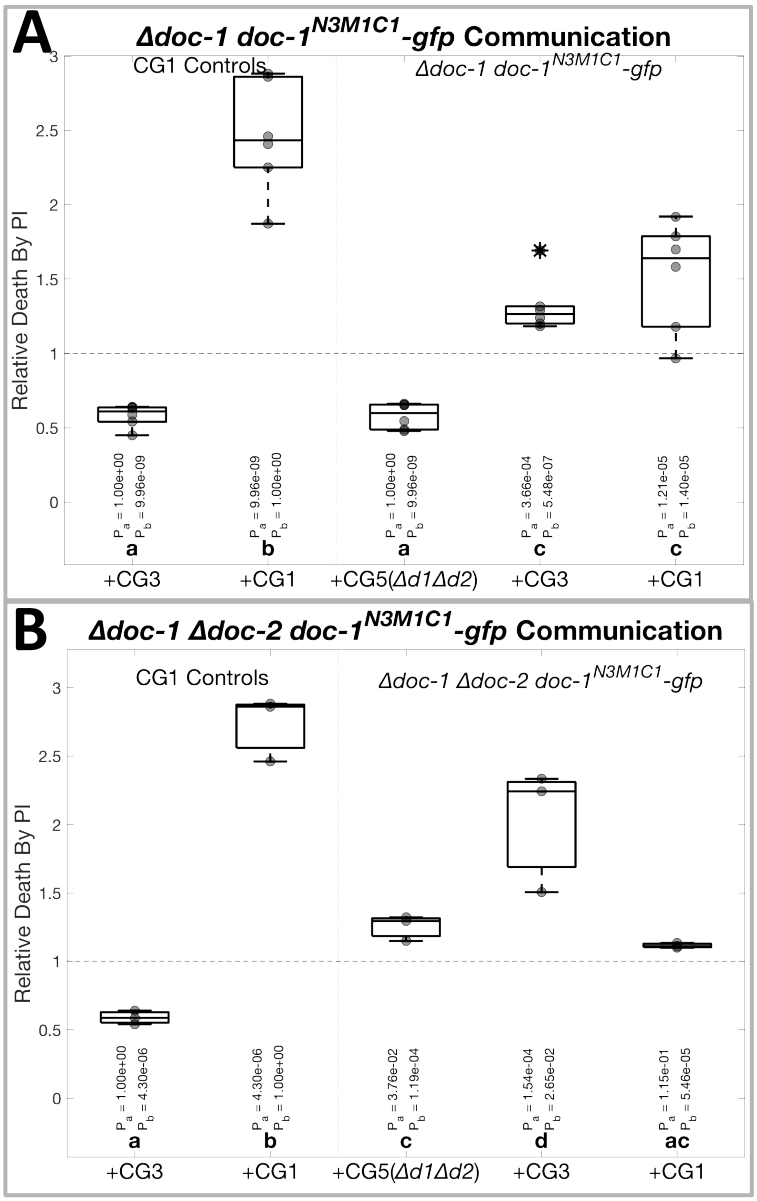


Figure 4.4.2-2. IDFC assay results from strains expressing *doc-1^{N3M1C1}* N-terminal region chimeric alleles

Boxplots showing CG phenotypes of strains expressing *doc-1^{N3M1C1}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show data points. Internal lines in boxes mark medians. Upper and lower box bounds mark quartiles. Capped dashed lines extend to extreme data. Outliers (outside 99% of the data range, assuming normality) are marked with asterisks and were not included in statistical analyses. Bold letters above the x-axis indicate statistical groups with p-values <0.05. P-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N3M1C1}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N3M1C1}-gfp*.

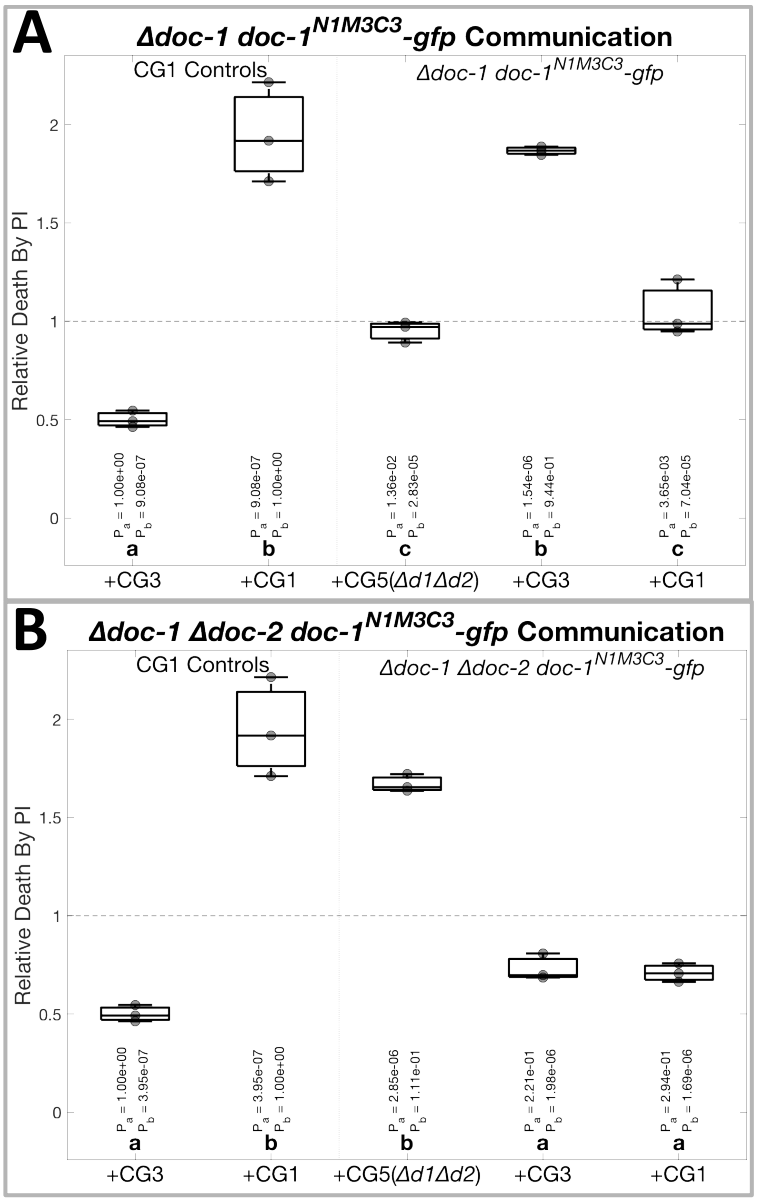


Figure 4.4.2-3. IDFC assay results from strains expressing *doc-1^{N1M3C3}* N-terminal region chimeric alleles

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-1^{N1M3C3}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N1M3C3}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N1M3C3}-gfp*.

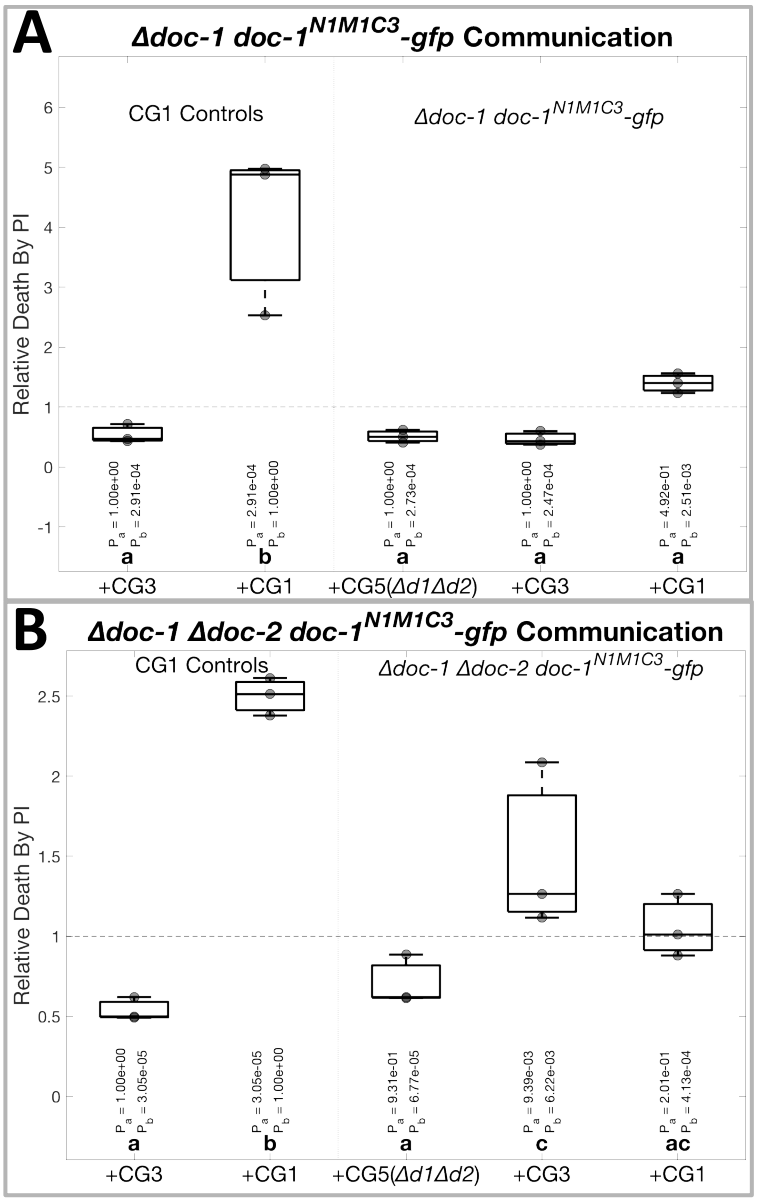


Figure 4.4.2-4. IDFC assay results from strains expressing *doc-1^{N1M1C3}* middle region chimeric alleles

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-1^{N1M1C3}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N1M1C3}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N1M1C3}-gfp*.

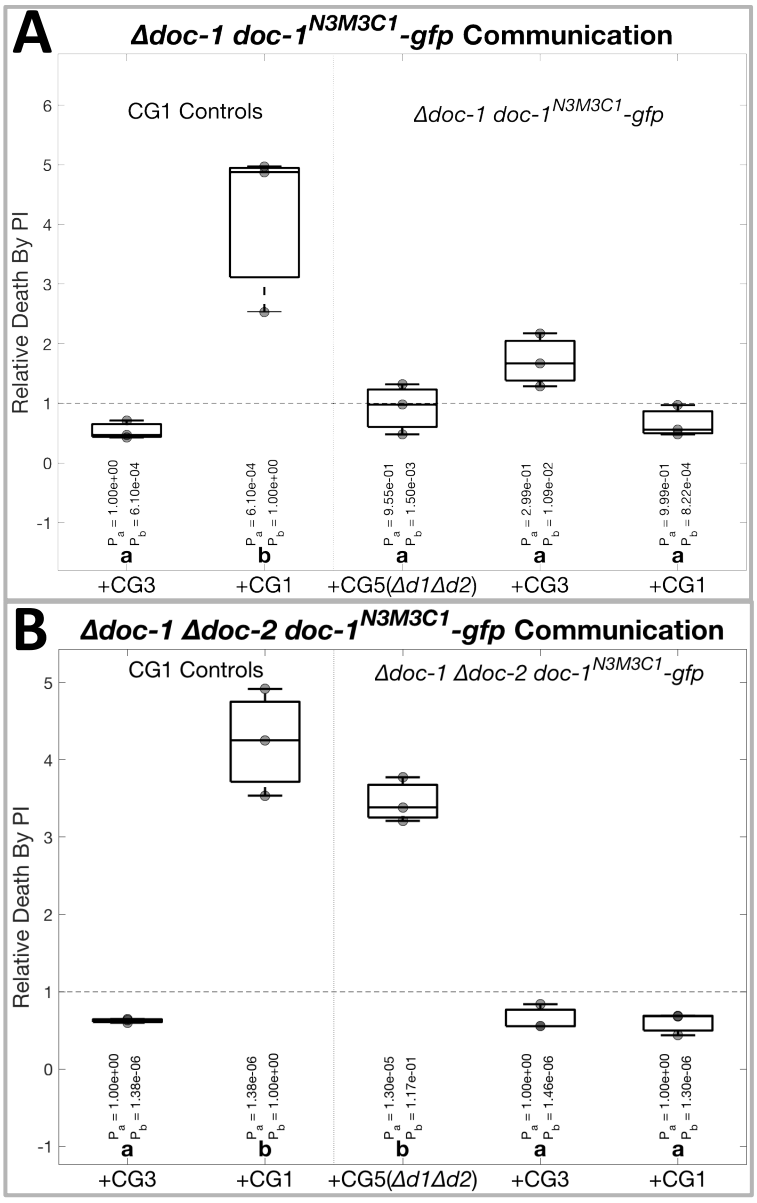


Figure 4.4.2-5. IDFC assay results from strains expressing *doc-1^{N3M3C1}* middle region chimeric alleles

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-1^{N3M3C1}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N3M3C1}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N3M3C1}-gfp*.

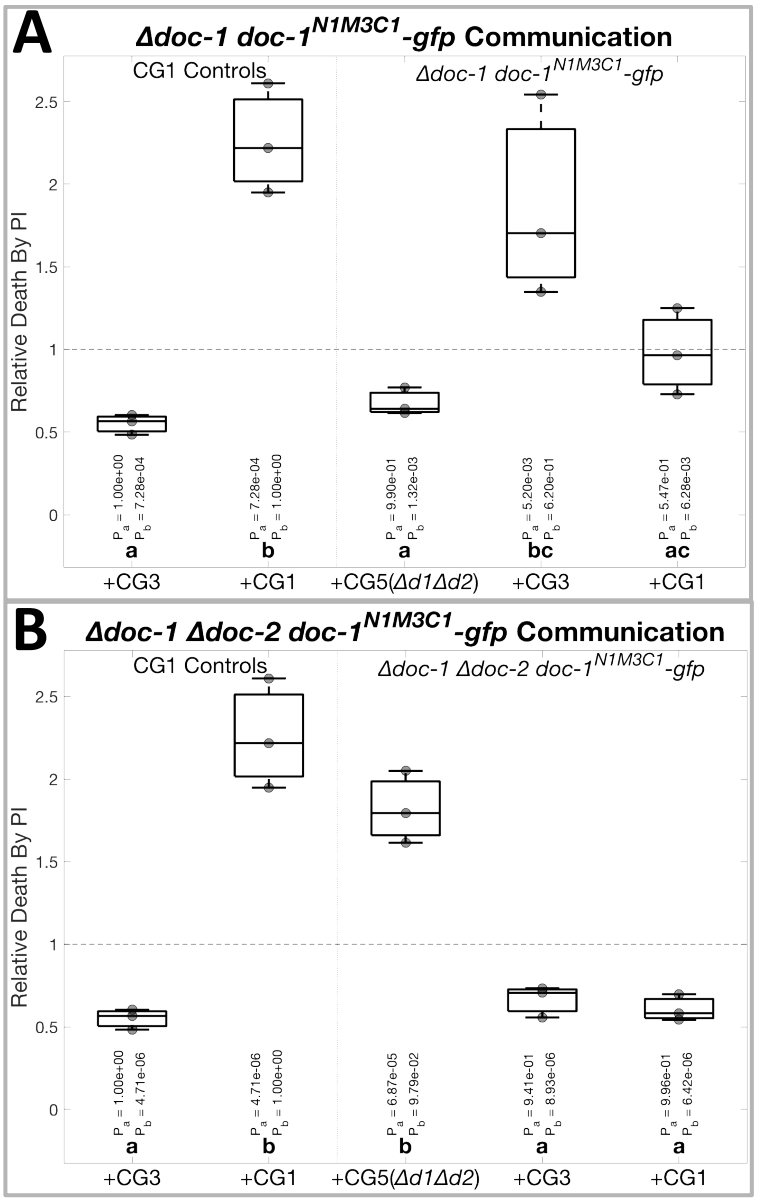


Figure 4.4.2-6. IDFC assay results from strains expressing *doc-1^{N1M3C1}* middle region chimeric alleles

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-1^{N1M3C1}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N1M3C1}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N1M3C1}-gfp*.

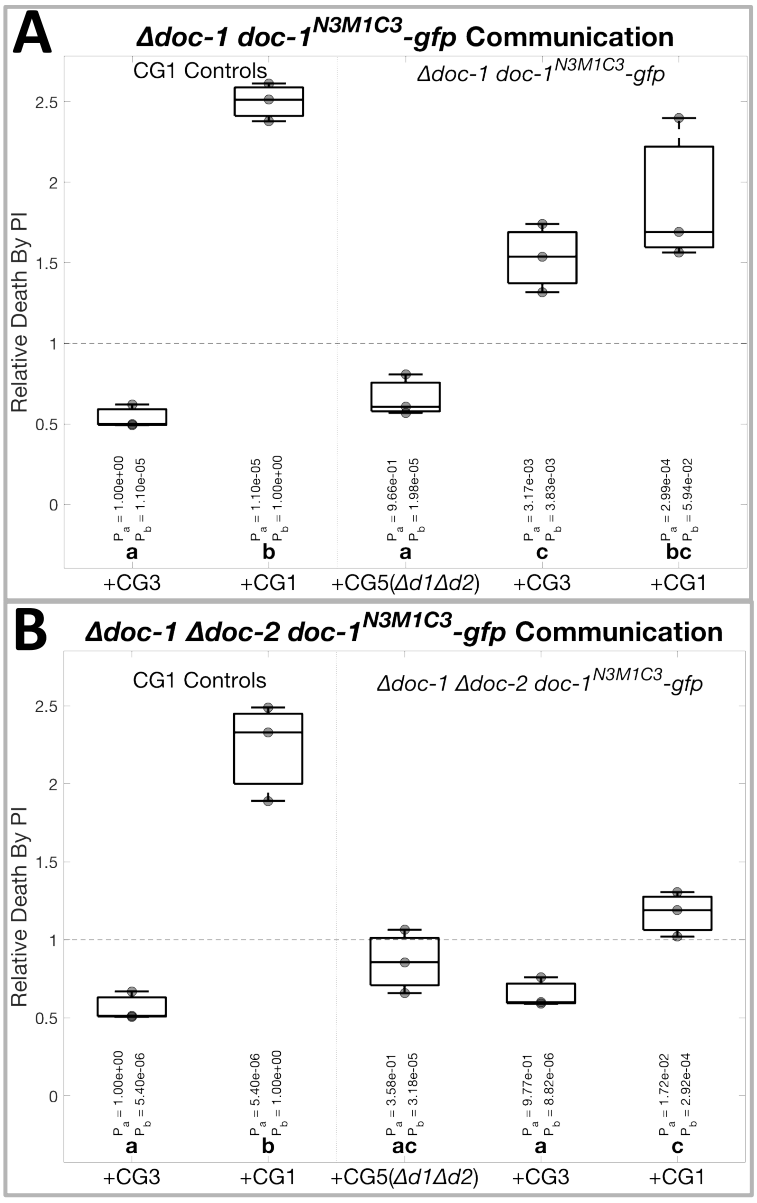


Figure 4.4.2-7. IDFC assay results from strains expressing *doc-1^{N3M1C3}* middle region chimeric alleles

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-1^{N3M1C3}-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. Both panels show triplicate data. **A)** IDFC results for *Δdoc-1 doc-1^{N3M1C3}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{N3M1C3}-gfp*.

Table 4.4.2. Summary of CG phenotypes from section 4.4.2

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1
FGSC2489			
Δdoc-1			
Δdoc-1 doc-1^{CGH1}-gfp			
Δdoc-1 doc-1^{CGH3}-gfp		Same as ->	<- Same as
Δdoc-1 Δdoc-2			
Δdoc-1 Δdoc-2 doc-1^{CGH1}-gfp	Higher	Lower	
Δdoc-1 Δdoc-2 doc-1^{CGH3}-gfp			
CG3-swap strain			
Δdoc-1 doc-1^{N3M1C1}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N3M1C1}-gfp	Lower	Higher	
Δdoc-1 doc-1^{N1M3C3}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N1M3C3}-gfp			
Δdoc-1 doc-1^{N1M1C3}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N1M1C3}-gfp			
Δdoc-1 doc-1^{N3M3C1}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N3M3C1}-gfp			
Δdoc-1 doc-1^{N1M3C1}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N1M3C1}-gfp			
Δdoc-1 doc-1^{N3M1C3}-gfp			
Δdoc-1 Δdoc-2 doc-1^{N3M1C3}-gfp			

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. Control data reproduced from chapters 2 and 3.

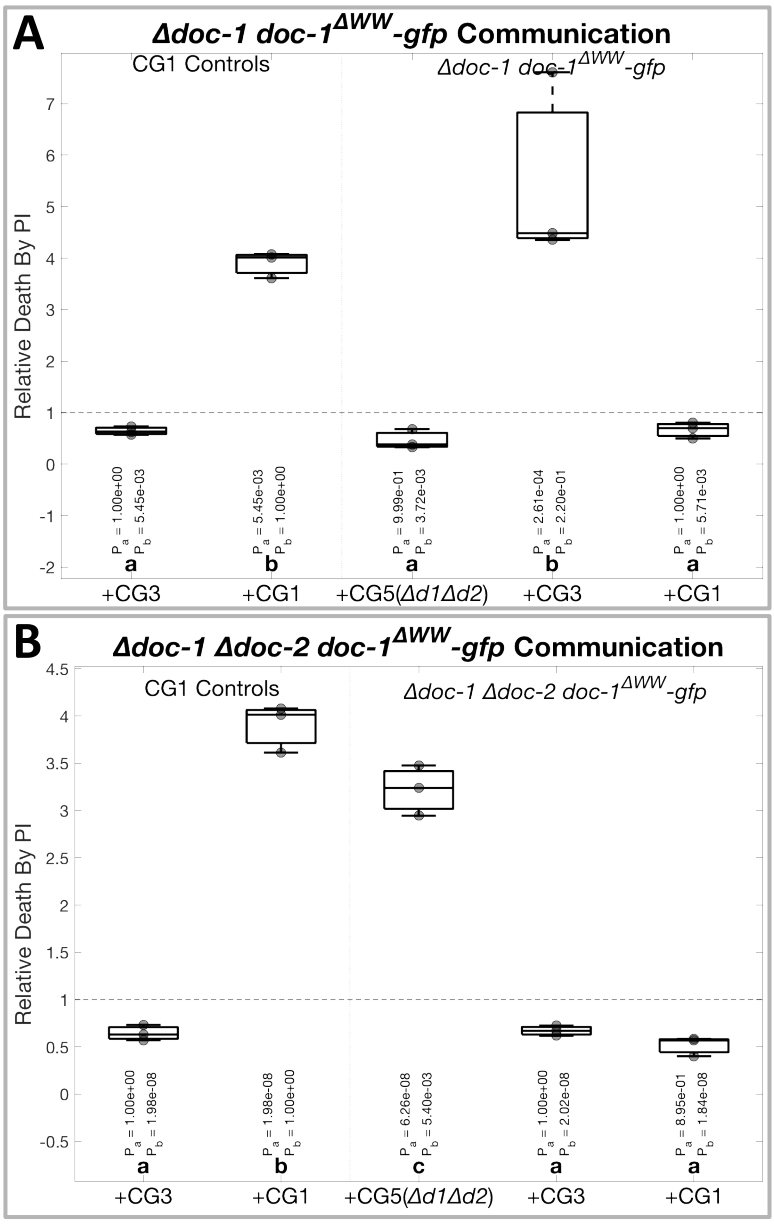


Figure 4.4.3-1. IDFC results from *Δdoc-1 doc-1^{ΔWW}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{ΔWW}-gfp*

Boxplot summary of CG phenotypes of strains expressing *doc-1^{CGH1}* with its putative WW motif deleted. Motifs were predicted using the ELM resource. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. WW = WW motif, CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1^{ΔWW}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{ΔWW}-gfp*.

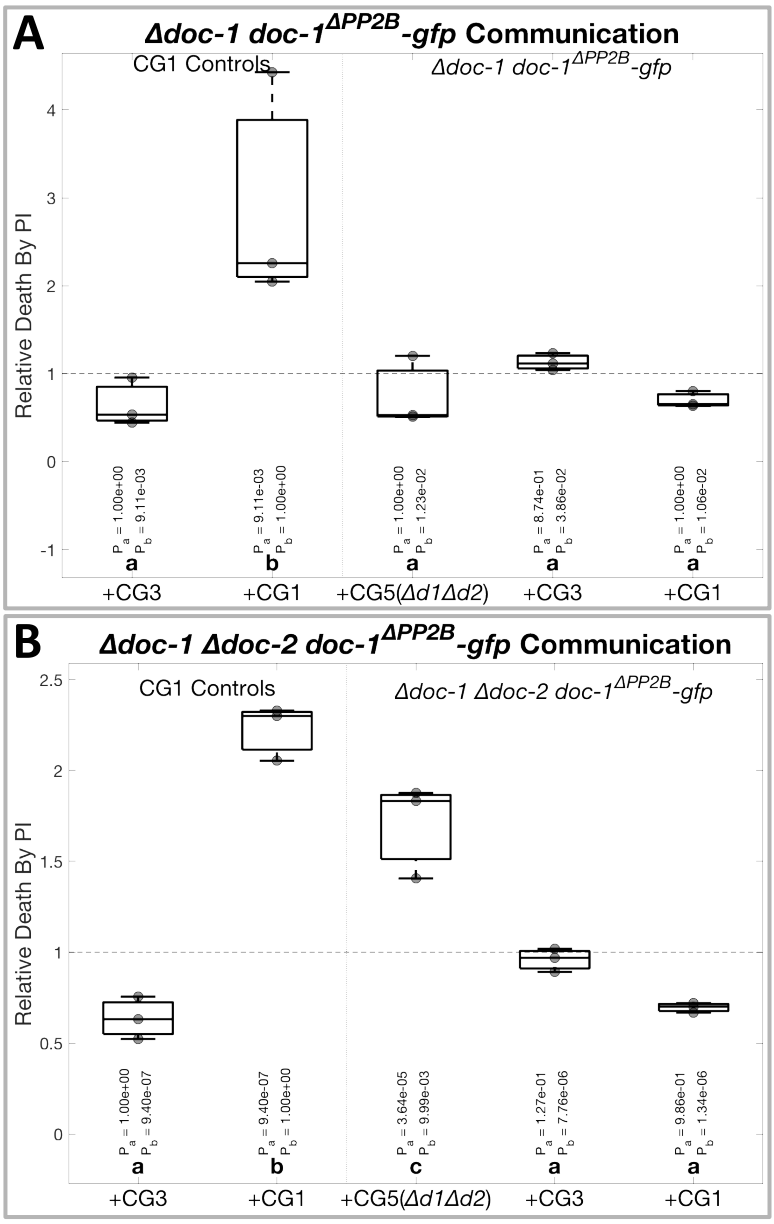


Figure 4.4.3-2. IDFC results from *Δdoc-1 doc-1^{ΔPP2B}-gfp* and *Δdoc-1 Δdoc-2 doc-1^{ΔPP2B}-gfp*

Boxplot summary of CG phenotypes of strains expressing *doc-1^{CGH1}* with its putative PP2B motif deleted. Motifs were predicted using the ELM resource. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown just above. PP2B = calcineurin docking motif, CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1^{ΔPP2B}-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1^{ΔPP2B}-gfp*.

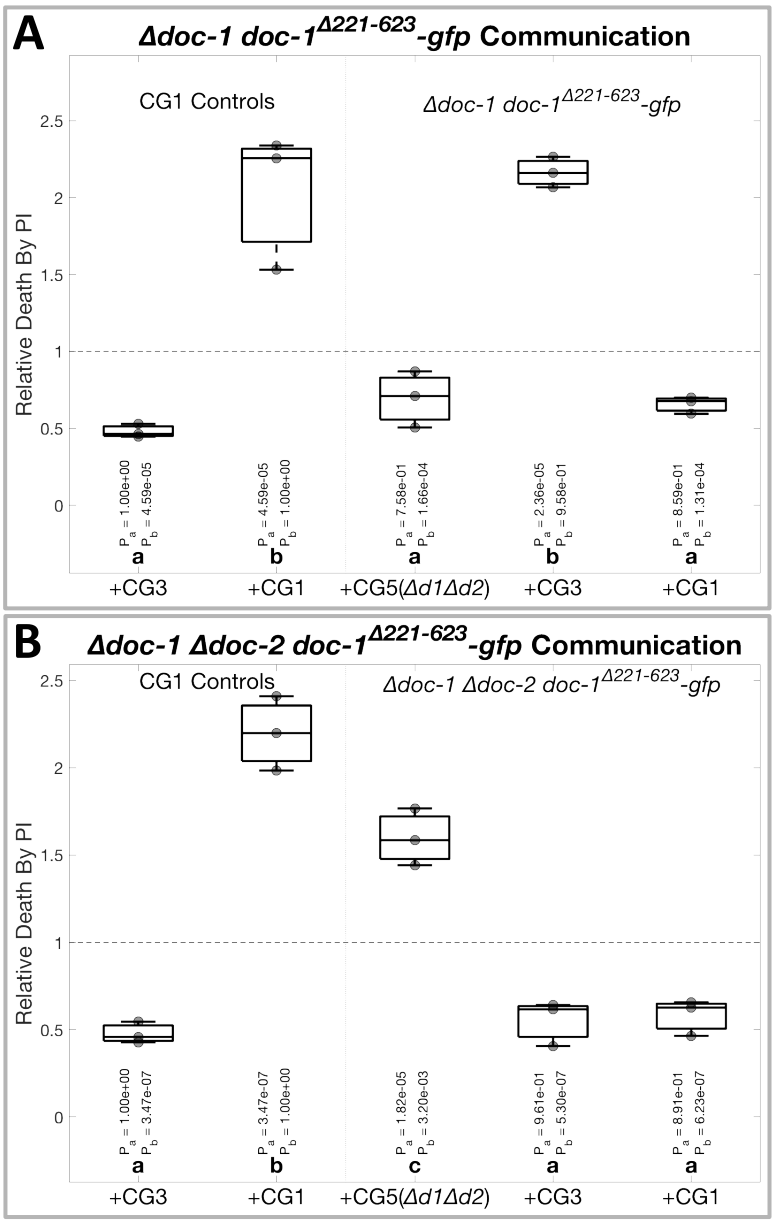


Figure 4.5.1. IDFC results from strains expressing middle region deletion alleles of *doc-1*

Boxplot summary of CG phenotypes of strains expressing *doc-1*$^{\Delta 221\text{-}623}$*-gfp*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1*$^{\Delta 221\text{-}623}$*-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1*$^{\Delta 221\text{-}623}$*-gfp*.

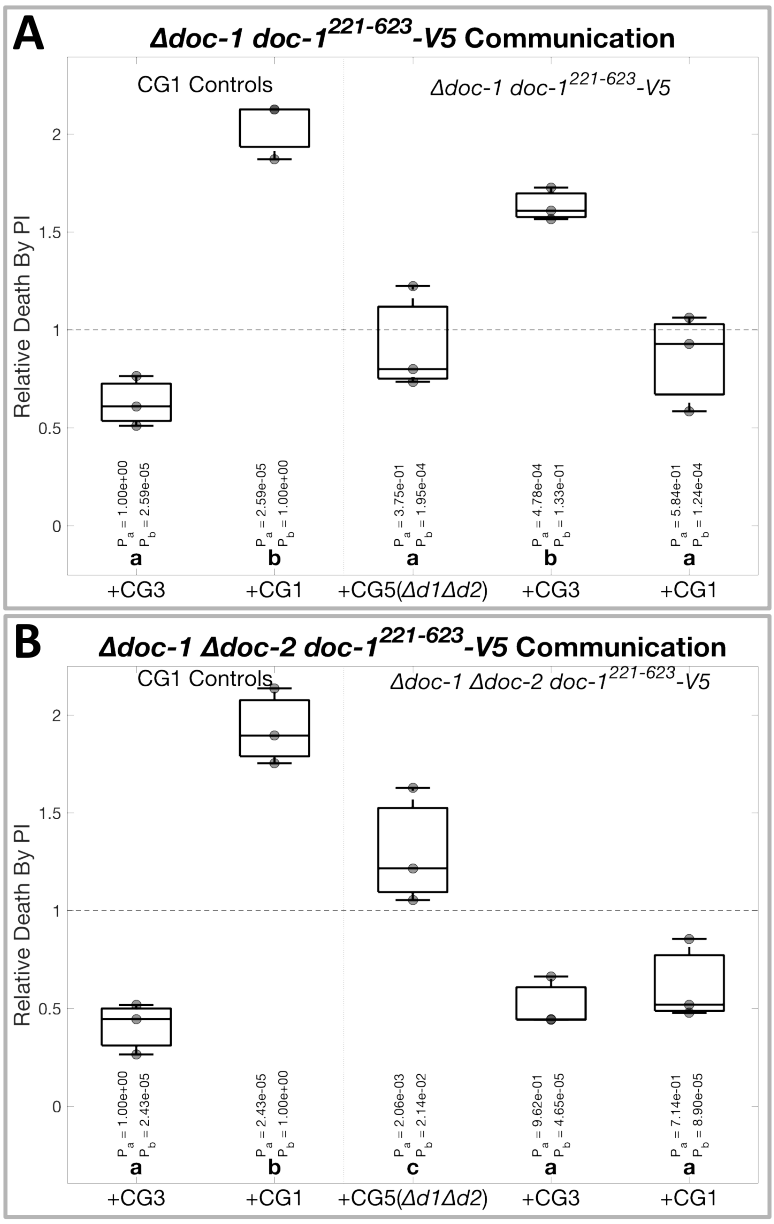


Figure 4.5.2. IDFC results from strains expressing DOC-1's middle region tagged with V5

Boxplot summary of CG phenotypes of strains expressing *doc-1$^{221-623}$-V5*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 doc-1$^{221-623}$-V5*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-1$^{221-623}$-V5*.

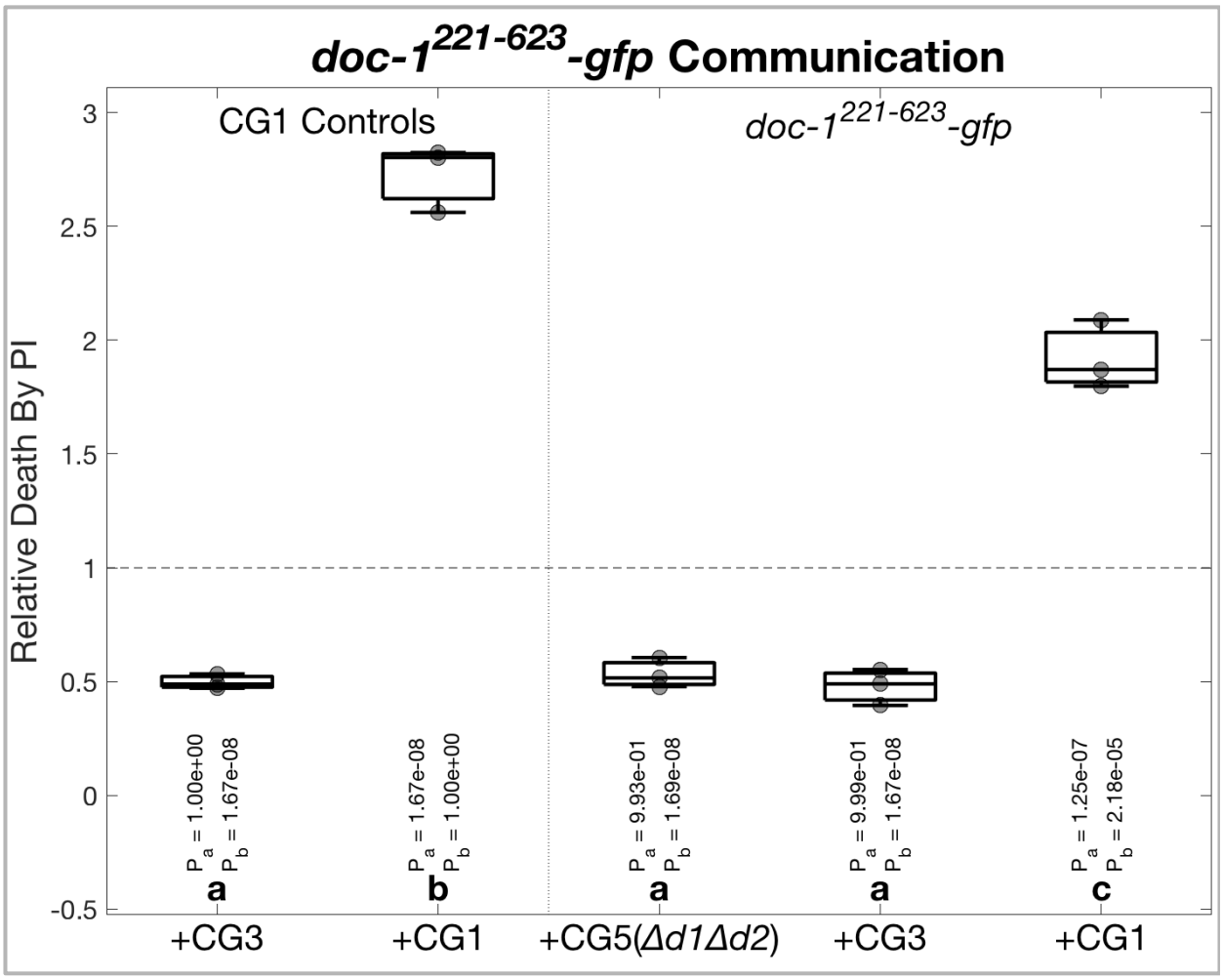


Figure 4.5.3. IDFC assay results from a strain expressing expressing DOC-1's middle region tagged with GFP

Boxplot summary of IDFC CG phenotype of a strain expressing *doc-1*$^{221\text{-}623}$*-gfp* with the CGH1 alleles of *doc-1* and *doc-2*. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide.

Table 4.5. Summary of CG phenotypes from section 4.5

Blue = high communication
Yellow = intermediate communication
Red = low communication

Strain	Comm w/CG5	Comm w/CG3	Comm w/CG1
FGSC2489			
Δdoc-1			
Δdoc-1 doc-1-gfp			
Δdoc-1 doc-1-V5			
Δdoc-1 Δdoc-2			
Δdoc-1 Δdoc-2 doc-1-gfp	Higher	Lower	
Δdoc-1 doc-1$^{\Delta 221\text{-}623}$-gfp			
Δdoc-1 Δdoc-2 doc-1$^{\Delta 221\text{-}623}$-gfp			
Δdoc-1 doc-1$^{221\text{-}623}$-V5			
Δdoc-1 Δdoc-2 doc-1$^{221\text{-}623}$-V5			
doc-1$^{221\text{-}623}$-gfp			

CG1 = FGSC2489 DI, CG3 = CG3-swap strain DI. CG5 = *Δdoc-1 Δdoc-2* DI, DI = death inducer via *sec-9*-swap. High communication was defined as equivalent to the CG1 self-communication positive control. Low communication was defined as equivalent to communication between CG1 and CG3. Intermediate communication was defined as above low communication and below high communication. All comparisons were evaluated at a significance level of p = 0.05 or better. Control data reproduced from chapter 2.

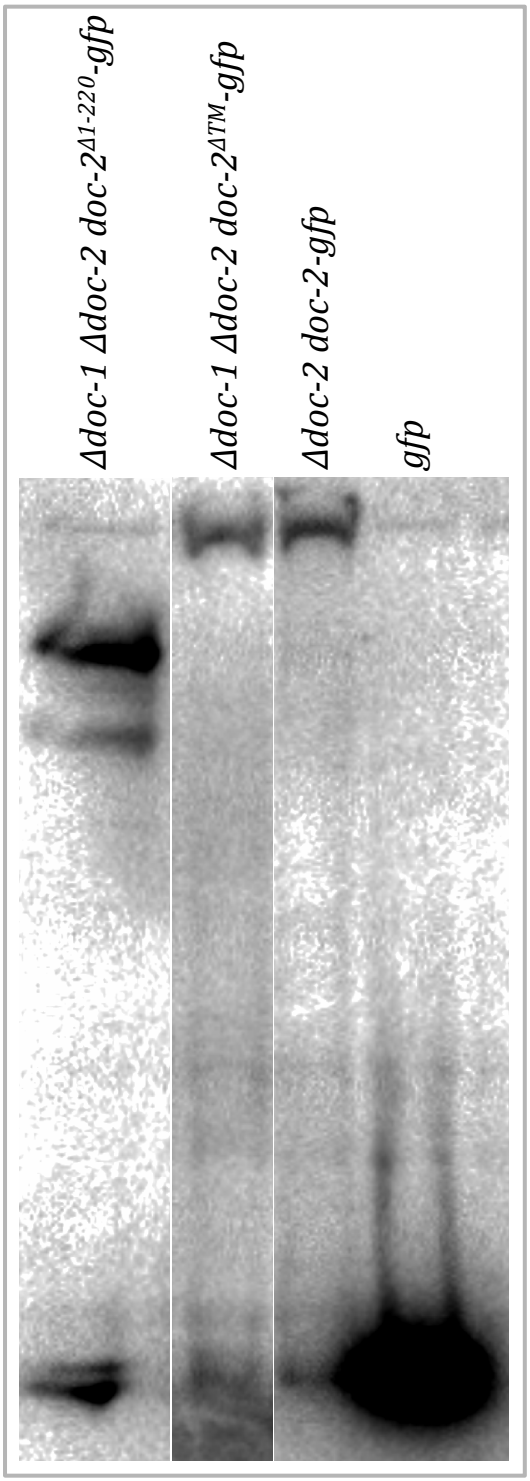


Figure 4.6.1-1. Western blots showing expression of DOC-2 truncations
Western blots showing detection of GFP-tagged DOC-2 truncation alleles in the *Δdoc-1 Δdoc-2* genetic background. Samples were loaded to normalize total protein content across the lanes. See section 4.2 for more details. Unfortunately, the white-light image of the ladder was lost. Expected band sizes: full length DOC-2-GFP = 118903.2 Da, DOC-$2^{\Delta 1\text{-}220}$-GFP = 95631 Da, DOC-$2^{\Delta TM}$-GFP = 115912.6 Da, free GFP = 26975.4 Da.

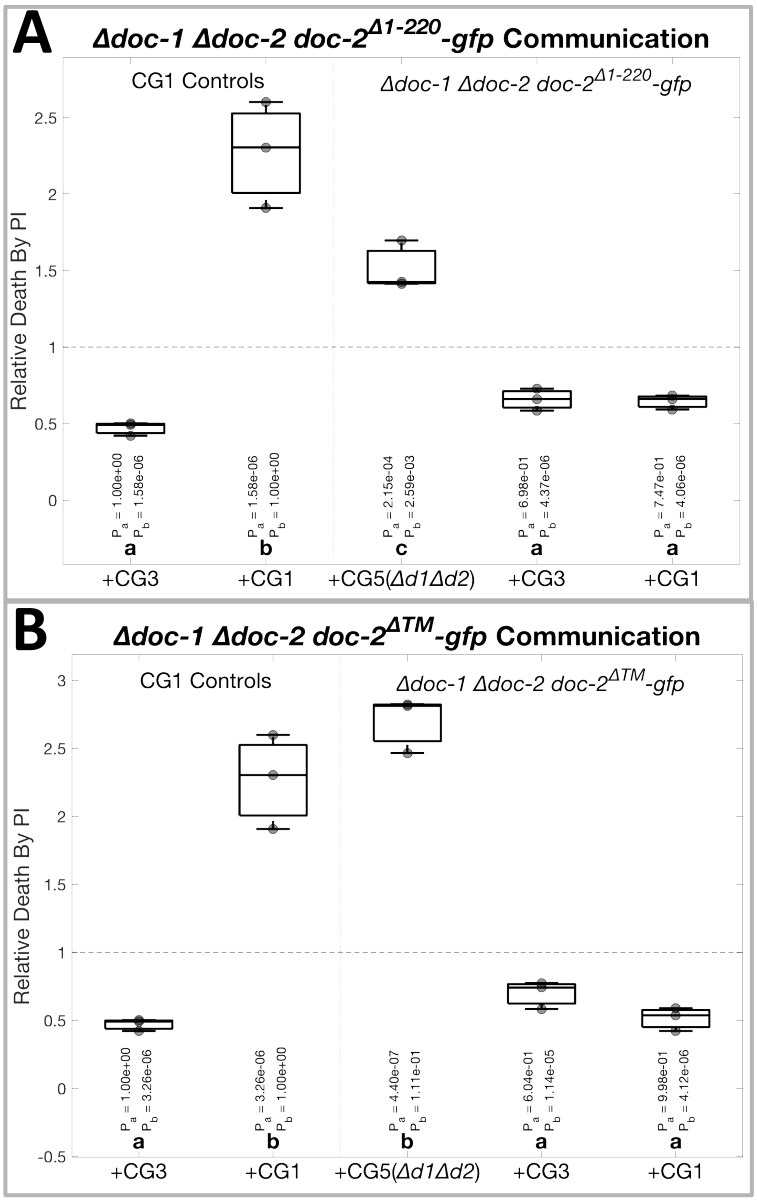


Figure 4.6.1-2. IDFC assay results from *doc-2$^{\Delta 1\text{-}220}$-gfp* and *doc-2$^{\Delta TM}$-gfp* strains

Boxplot summary of IDFC CG phenotypes of strains expressing *doc-2^{CGH1}* truncation alleles in a *Δdoc-1 Δdoc-2* genetic background. Data were analyzed using one-way ANOVA and Tukey-Kramer multiple comparison tests. Vertical dotted gray line separates negative and positive controls on the left from test data. Gray circles show individual data points. Internal lines in boxplots mark medians and upper and lower box bounds mark quartiles. Capped dashed lines extend to more extreme data. Bold letters just above the x-axis indicate statistical groups with p-values less than 0.05, and p-values for comparisons between each sample and the negative control (P_a) or positive control (P_b) are shown just above. CG1 = FGSC2489, CG3 = CG3-swap strain, CG5 = *Δdoc-1 Δdoc-2*, DI = death inducer via *sec-9*-swap, PI = propidium iodide. **A)** IDFC results for *Δdoc-1 Δdoc-2 doc-2$^{\Delta 1\text{-}220}$-gfp*. **B)** IDFC results for *Δdoc-1 Δdoc-2 doc-2$^{\Delta TM}$-gfp*.

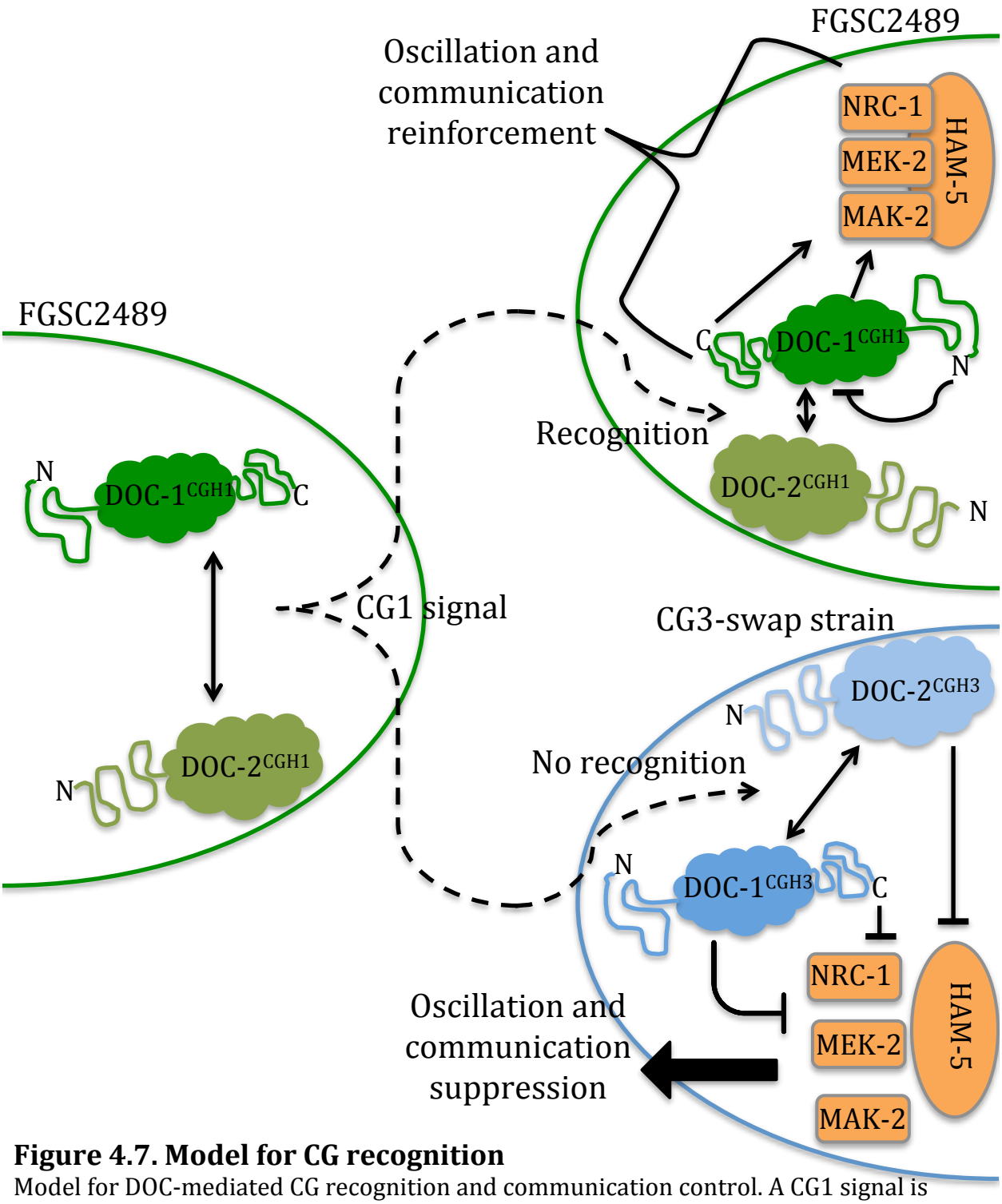


Figure 4.7. Model for CG recognition

Model for DOC-mediated CG recognition and communication control. A CG1 signal is generated by CGH1 DOC proteins in the left cell via unknown processes. This signal is recognized through the interactions of DOC-1^{CGH1}'s middle region with DOC-2^{CGH1} in the upper right cell. The N-terminal region of DOC-1 relieves its suppression of communication, resulting in reinforcement of oscillation and continued chemotropic growth. In the lower right cell, CGH3 DOC proteins do not recognize the CG1 signal: they suppress MAK-2 complex oscillations and communication.

Chapter 5. Discussion and conclusions

5.1. Summary and conclusions

In 2016, when we published our discovery that *Neurospora crassa* uses a long-distance non-self recognition (NSR) system to restrict communication between non-identical cells, we knew alleles at the *determinant of communication* (*doc*) locus were necessary and sufficient to specify which cells will be recognized as self[42]. We also reported that DOC-2 localized to the cellular periphery and DOC-1 oscillated between the cell bodies and growing tips of cells along with the MAK-2 complex during intercellular communication[73,75,124]. The DOC system was not required for self-communication, implying it suppresses communication until a compatible communication group (CG) signal is received. We confirmed that communication was not regulated by altering the oscillation frequencies of SOFT, another protein whose oscillatory intracellular movements are associated with intercellular communication and chemotropic growth[73]. Finally, we identified five distinct communication group haplotypes (CGHs) at the *doc* locus in an *N. crassa* population from Louisiana; *doc* alleles from different CGHs were less than 50% identical at the nucleotide level, and the CGHs showed no evidence of recombination and exhibited trans-species polymorphism among *N. crassa* and other members of its genus.

Little else was certain regarding the DOC system. With neither characterized homologues nor identifiable functional domains, the DOC proteins remained particularly opaque. The primary goal of my doctoral work was to advance our understanding of the DOC system. I focused on improving our model of DOC-mediated long-distance NSR by investigating how different *doc* genes and alleles interact *in vivo*, and evaluating the roles played by various regions of the DOC proteins in regulating communication. In pursuit of these aims, I spent considerable time and effort developing improved methods for quantifying intercellular communication.

In chapter 2, I refined and validated flow cytometry-based assays to quantify communication and fusion between *N. crassa* germlings[64]. I also developed computational pipelines to automatically analyze and present data from these assays. Using the assays and analyses I developed, I demonstrated that GFP-tags reduce DOC functionality. I used the flow cytometry assays I developed throughout the remainder of my doctoral work, and versions of this assay and analysis pipeline have been used in published work[63,65].

Chapter 3 covers my studies on how incompatible alleles of *doc-1* and *doc-2* interact to control communication phenotypes. I tested the communication preferences of strains expressing most possible combinations of *doc-1* and *doc-2* alleles from CGH1, CGH3, and CGH5. These experiments revealed that interactions between *doc* genes from the same haplotype are generally required for robust communication between strains sharing a specific CGH. The CGH1 variant of *doc-1* was an exception to this rule, producing vigorous communication with CG1 strains in the absence of *doc-2*. However, experiments in chapter 2 and Heller et al., 2016, demonstrated expression of either *doc* gene without the other reduces self-communication rates[42], further highlighting the importance of non-allelic interactions in the DOC system. Experiments in chapter 3 also suggest incompatible variants of DOC proteins compete for control over communication, resulting in reduced

communication rates overall. Again, the CGH1 variant of *doc-1* stood out by incompletely dominating its rivals: in all cases, expression of *doc-1* prevented communication with all but CG1 strains. I also confirmed the sole *doc* gene in CGH5 retains some ability to suppress communication. This finding was surprising because the *Δdoc-1 Δdoc-2* mutant specifies CG5[42], and it suggests a CG5 signal may be produced by cells without DOC proteins.

Finally, in chapter 4 I attempted to characterize the DOC proteins. Bioinformatic surveys suggested a possible link between calcium signaling and the DOC system, although further experiments will be necessary to confirm this connection. CGH-specific indels were also identified in DOC-1, and clustered in the putatively unstructured N- and C-terminal regions of the protein. I assayed the communication phenotypes of strains expressing truncated and otherwise manipulated versions of DOC-1^{CGH1}, as well as CGH1/CGH3 chimeric variants of DOC-1. These experiments revealed DOC-1's N-terminal region was necessary to allow communication with any tested CG, and its C-terminal region was required to suppress communication with the *Δdoc-1 Δdoc-2* mutant. Over-expressing the middle region, a predicted globular domain, showed this region competes with full-length DOC-1 for control over communication, implying the middle region is involved in interactions with DOC-2 and other communication machinery (e.g. the MAK-2 complex). I also tested the importance of DOC-2's N-terminal region and found it was not necessary for communication suppression. Further work on DOC-2 was hampered by the *Δdoc-2* mutant's lack of non-self communication deficits when assayed using flow cytometry. Chimeric allele experiments gave more ambiguous results, indicating all regions of DOC-1 play some role in CG signal recognition. However, various experiments throughout this chapter strongly suggested the borders I selected between N-terminal, middle, and C-terminal regions in DOC-1 were improperly defined. Furthermore, a simple three-part model of the protein cannot capture the pattern of CGH-specific indels across DOC-1 variants. A conclusive demonstration that CGH-specific indels encode CG specificity determinants will probably require a near-complete swap of indels between CGH1 and CGH3 variants across the length of DOC-1.

The impacts of my doctoral work are summarized in an updated model for DOC-mediated CG specificity presented in Fig. 4.7. As cells begin to communicate, general and CG-specific signals must be exchanged. CG signals are validated through interactions between the middle region of DOC-1 and DOC-2: matched CGH-variants of the DOC proteins are generally required for proper CG specificity. If the DOC system does not receive a compatible CG signal, DOC-1 and DOC-2 both prevent reinforcement of MAK-2 complex oscillations and suppress communication. This suppression involves the actions of the middle and C-terminal regions of DOC-1, and does not require the N-terminal region of DOC-2. If a compatible CG signal is detected, the DOC system allows MAK-2 complex oscillations and chemotropic growth to continue. DOC-1's N-terminal region is required to properly derepress communication, and is probably involved in validating CG signals.

5.2. Future Directions

In section 1.7 of my introductory chapter, I outlined some important questions and predictions about the DOC system. Although I've answered some of these questions and tested some predictions, two major aspects of the DOC system remain unaddressed: DOC

protein-protein interactions and the CG signal. Further investigations of the DOC system should begin with these features, which I've broken into several parts below:

DOC protein-protein interactions

- Does HAM-5 interact directly with DOC-1? Is HAM-5 the only MAK-2 complex component with which DOC-1 interacts? What parts of HAM-5 interact with what parts of DOC-1? When during communication do DOC-1 and HAM-5 interact? Do they still interact if the DOC system suppresses communication? Does HAM-5 also interact with DOC-2? Does HAM-5, or another MAK-2 complex component, interact with DOC variants from haplotypes other than CGH1?
    - I attempted to address some of these questions using co-immunoprecipitation experiments with strains expressing manipulated DOC-1 variants, described in chapter 4, and manipulated HAM-5 variants described in Jonkers et al., 2014[75]. However, HAM-5's large size (1686 residues, ~184 kDa) posed technical challenges that require further optimization to overcome (data not shown).
- Does DOC-1 directly interact with DOC-2? Do variants of DOC-1 and DOC-2 from incompatible CGHs interact? What parts of DOC-1 and DOC-2 interact? Do DOC-1 and DOC-2 interact only during communication? Do they continue to interact if they suppress communication?
    - Jens Heller and I both attempted to assess whether DOC-1 and DOC-2 interact in FGSC2489 with co-immunoprecipitation experiments (data not shown). While Jens found limited evidence that the proteins interact, I could not reproduce his results. Immunoprecipitation-mass-spectrometry (IPMS) experiments also failed to identify DOC-2 as a potential interaction partner with DOC-1 during intercellular communication (data not shown).
- Do the DOC proteins interact with other cellular components? Which ones? During chemotropic growth, or only while suppressing communication?
    - I attempted to find DOC-interaction partners using IPMS, but experiments with GFP-tagged DOCs failed to return quality data. Candidate interaction partners with DOC-1-V5 either could not be or were not validated (data not shown).

The CG signal

- What is the CG signal? In what ways is it distinct from the unidentified general communication signal?
    - In addition to SOFT oscillation frequency control, discussed in section 1.6 of the introductory chapter, I also investigated lipochitooligosaccharides as possible communication signaling molecules in *N. crassa*. I found no evidence that either of these play a role in communication.

    - The results from my experiments with DOC-1^{CGH5}, covered in chapter 4, suggest a CG5 signal is produced in the absence of DOC proteins. If true, perhaps *N. crassa* uses a modular communication signal, with a general part recognized by all genotypes and a CG-specific part recognized by the DOC system. In this case, DOC-1^{CGH5} must recognize the signal molecule without modifications by the DOC system.

- How is the CG signal produced? How do different DOC variants produce different CG signals?
- How and when do cells release the CG signal? How and when do cells detect it? What is the relationship between oscillations of the MAK-2 and SOFT complexes and CG and/or general signal release and detection? How do different DOC variants recognize different CG signals?
- The CG signal must be short-lived, so is it volatile, reactive, or actively degraded?

Improvements and extensions

The experiments I presented in this dissertation can also be improved and extended to yield clearer, more reliable data and answer more questions. My suggested improvements and extensions follow:

- Repeat my experiments without GFP-tags. Large tags reduce DOC functionality, as discussed in chapter 2. Use V5 or other small epitope-tags to confirm protein expression.
- Test the self-communication rates of all manipulated strains. This would not be technically challenging, but would require backcrosses and screening for each strain I tested. Alternatively, the dyeFC assay could return high-quality self-communication data without further strain construction, but the assay requires further development.
- Check for oscillatory behavior of all GFP-tagged DOC-1 variants. This would verify our model's assumption that DOC-1's oscillation is required to influence communication. Such observations could also help clarify what regions of DOC-1 are responsible for interacting with the MAK-2 complex and, therefore, what regions are responsible for suppressing communication.
- Although it was not mentioned previously in this dissertation, the communication group haplotype 1 contains two clades labeled A and B[42]. Strains with either haplotype show the same communication patterns and recognize each other as self. Apparently there is a recombination block between CGH1A and CGH1B, suggesting strong selection against crossovers between *doc-1* and *doc-2* in these haplotypes within CG1; more wild isolates should be examined to confirm the blockage of recombination between these sub-haplotypes. The existence of sub-haplotypes may represent an early stage in the emergence of a 6th CG. Comparing the differences between CGH1A and CGH1B alleles may shed light on the interactions between

matched alleles of *doc-1* and *doc-2*, and could also help narrow the search for CG specificity determinants.

To anyone reading this dissertation and continuing to work on the DOC system, I wish you luck. I hope you learned from my mistakes and successes, and can overcome the problems that limited my progress. My most important advice, cliché as it sounds, is to work smarter, not harder. Spend more time reading, planning, preparing, analyzing, thinking creatively, and writing than you do at the bench. Most of my time cloning, constructing strains, and running experiments returned nothing of value. Perhaps I could have saved time, avoided frustration, and accomplished more if someone had advised me as I am now advising you. This was a difficult project for me, and I'm confident it will continue to vex other researchers who attempt to unravel the mysteries of the DOC system.

5.3. References

1. Richman, A. Evolution of balanced genetic polymorphism. *Mol. Ecol.* **9**, 1953–1963 (2000).
2. Rosengarten, R. D. & Nicotra, M. L. Model Systems of Invertebrate Allorecognition. *Curr. Biol.* **21**, R82–R92 (2011).
3. Ocasio, A. B. & Cotter, P. A. CDI/CDS system-encoding genes of Burkholderia thailandensis are located in a mobile genetic element that defines a new class of transposon. *PLoS Genet.* **15**, (2019).
4. Fujii, S., Kubo, K. & Takayama, S. Non-self- and self-recognition models in plant self-incompatibility. *Nat. Plants* **2**, 16130 (2016).
5. Hirose, S., Benabentos, R., Ho, H.-I., Kuspa, A. & Shaulsky, G. Self-Recognition in Social Amoebae Is Mediated by Allelic Pairs of Tiger Genes. *Science* **333**, 467–470 (2011).
6. Brown, J. L. & Eklund, A. Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *Am. Nat.* **143**, 435–461 (1994).
7. Hutchison, E. A. & Glass, N. L. Programmed Cell Death and Heterokaryon Incompatibility in Filamentous Fungi. in *Biocommunication of Fungi* (ed. Witzany, G.) 115–138 (Springer Netherlands, 2012).
8. Vos, M. & Velicer, G. J. Evolution of social conflict in the bacterium Myxococcus xanthus: centimeter vs global scale populations. *Curr. Biol. CB* **19**, 1763–1767 (2009).
9. Gibbs, K. A. & Greenberg, E. P. Territoriality in Proteus: advertisement and aggression. *Chem. Rev.* **111**, 188–194 (2011).
10. Strassmann, J. E., Gilbert, O. M. & Queller, D. C. Kin Discrimination and Cooperation in Microbes. *Annu. Rev. Microbiol.* **65**, 349–367 (2011).
11. Czárán, T., Hoekstra, R. F. & Aanen, D. K. Selection against somatic parasitism can maintain allorecognition in fungi. *Fungal Genet. Biol.* **73**, 128–137 (2014).
12. Rankin, D. J., Bargum, K. & Kokko, H. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**, 643–651 (2007).

13. Bastiaans, E., Debets, A. J. M. & Aanen, D. K. Experimental evolution reveals that high relatedness protects multicellular cooperation from cheaters. *Nat. Commun.* **7**, 11435 (2016).

14. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature* **441**, 498 (2006).

15. Kuzdzal-Fick, J. J., Fox, S. A., Strassmann, J. E. & Queller, D. C. High Relatedness Is Necessary and Sufficient to Maintain Multicellularity in Dictyostelium. *Science* **334**, 1548–1551 (2011).

16. Hartfield, M. & Keightley, P. D. Current hypotheses for the evolution of sex and recombination. *Integr. Zool.* **7**, 192–209 (2012).

17. Lee, S. C., Ni, M., Li, W., Shertz, C. & Heitman, J. The Evolution of Sex: a Perspective from the Fungal Kingdom. *Microbiol. Mol. Biol. Rev.* **74**, 298–340 (2010).

18. Dawe, A. L. & Nuss, D. L. Chapter Five - Hypovirus Molecular Biology: From Koch's Postulates to Host Self-Recognition Genes that Restrict Virus Transmission. in *Advances in Virus Research* (ed. Ghabrial, S. A.) **86**, 109–147 (Academic Press, 2013).

19. Ghabrial, S. A. & Suzuki, N. Viruses of Plant Pathogenic Fungi. *Annu. Rev. Phytopathol.* **47**, 353–384 (2009).

20. Zhang, D.-X., Spiering, M. J., Dawe, A. L. & Nuss, D. L. Vegetative Incompatibility Loci with Dedicated Roles in Allorecognition Restrict Mycovirus Transmission in Chestnut Blight Fungus. *Genetics* **197**, 701–714 (2014).

21. Milgroom, M. G., Smith, M. L., Drott, M. T. & Nuss, D. L. Balancing selection at nonself recognition loci in the chestnut blight fungus, Cryphonectria parasitica , demonstrated by trans-species polymorphisms, positive selection, and even allele frequencies. *Heredity* **121**, 511 (2018).

22. Zhao, J. *et al.* Genomics and evolutionary approaches to identify nonself recognition loci in the filamentous fungus Neurospora crassa. *In Review.*

23. Garcia, E. C., Perault, A. I., Marlatt, S. A. & Cotter, P. A. Interbacterial signaling via Burkholderia contact-dependent growth inhibition system proteins. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8296–8301 (2016).

24. Saak, C. C. & Gibbs, K. A. The Self-Identity Protein IdsD Is Communicated between Cells in Swarming Proteus mirabilis Colonies. *J. Bacteriol.* **198**, 3278–3286 (2016).

25. Cardarelli, L., Saak, C. & Gibbs, K. A. Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *mBio* **6**, e00251-15 (2015).

26. Karadge, U. B., Gosto, M. & Nicotra, M. L. Allorecognition Proteins in an Invertebrate Exhibit Homophilic Interactions. *Curr. Biol.* **25**, 2845–2850 (2015).

27. Dishaw, L. J. & Litman, G. W. Invertebrate Allorecognition: The Origins of Histocompatibility. *Curr. Biol. CB* **19**, R286–R288 (2009).

28. Taketa, D. A. & De Tomaso, A. W. Botryllus schlosseri Allorecognition: Tackling the Enigma. *Dev. Comp. Immunol.* **48**, 254–265 (2015).

29. Li, J. *et al.* Electrostatic potentials of the S-locus F-box proteins contribute to the pollen S specificity in self-incompatibility in Petunia hybrida. *Plant J.* **89**, 45–57 (2017).

30. Kubo, K. *et al.* Collaborative Non-Self Recognition System in S-RNase–Based Self-Incompatibility. *Science* **330**, 796–799 (2010).

31. Nasrallah, J. B. Chapter Sixteen - Self-incompatibility in the Brassicaceae: Regulation and mechanism of self-recognition. in *Current Topics in Developmental Biology* (ed. Grossniklaus, U.) **131**, 435–452 (Academic Press, 2019).

32. Takayama, S. *et al.* The pollen determinant of self-incompatibility in Brassica campestris. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1920–1925 (2000).

33. Wheeler, M. J. *et al.* Identification of the pollen self-incompatibility determinant in Papaver rhoeas. *Nature* **459**, 992–995 (2009).

34. Benabentos, R. *et al.* Polymorphic Members of the lag Gene Family Mediate Kin Discrimination in Dictyostelium. *Curr. Biol.* **19**, 567–572 (2009).

35. Dementhon, K. *et al.* Rapamycin Mimics the Incompatibility Reaction in the Fungus Podospora anserina. *Eukaryot. Cell* **2**, 238–246 (2003).

36. Saupe, S. J. & Glass, N. L. Allelic Specificity at the Het-C Heterokaryon Incompatibility Locus of Neurospora Crassa Is Determined by a Highly Variable Domain. *Genetics* **146**, 1299–1309 (1997).

37. Shiu, P. K. & Glass, N. L. Molecular characterization of tol, a mediator of mating-type-associated vegetative incompatibility in Neurospora crassa. *Genetics* **151**, 545–555 (1999).

38. Wilson, J., Calligan, C. & Dempsey, J. A. A heterokaryon instability gene in the Rockefeller-Lindegren strains of Neurospora crassa and its possible relation to the het i gene in Oak Ridge-St. Lawrence strains. *Fungal Genet. Rep.* **46**, 23–24 (1999).

39. Wu, J. & Glass, N. L. Identification of Specificity Determinants and Generation of Alleles with Novel Specificity at the het-c Heterokaryon Incompatibility Locus of Neurospora crassa. *Mol. Cell. Biol.* **21**, 1045–1057 (2001).

40. Glass, N. L. & Kaneko, I. Fatal Attraction: Nonself Recognition and Heterokaryon Incompatibility in Filamentous Fungi. *Eukaryot. Cell* **2**, 1–8 (2003).

41. Boehm, T. Quality Control in Self/Nonself Discrimination. *Cell* **125**, 845–858 (2006).

42. Heller, J. *et al.* Characterization of Greenbeard Genes Involved in Long-Distance Kind Discrimination in a Microbial Eukaryote. *PLOS Biol* **14**, e1002431 (2016).

43. Queller, D. C. Expanded social fitness and Hamilton's rule for kin, kith, and kind. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10792–10799 (2011).

44. Gardner, A. & West, S. A. Greenbeards. *Evolution* **64**, 25–38 (2010).

45. West, S. A. & Gardner, A. Altruism, Spite, and Greenbeards. *Science* **327**, 1341–1344 (2010).

46. Grafen, A. Do animals really recognize kin? *Anim. Behav.* **39**, 42–54 (1990).

47. Zhao, J. *et al.* Identification of Allorecognition Loci in Neurospora crassa by Genomics and Evolutionary Approaches. *Mol. Biol. Evol.* **32**, 2417–2432 (2015).

48. Davis, R. H. *Neurospora: Contributions of a Model Organism*. (Oxford University Press, 2000).

49. Perkins, D. D., Radford, A. & Sachs, M. S. *The Neurospora Compendium: Chromosomal Loci*. (Academic Press, 2000).

50. Dunlap, J. C. *et al.* Enabling a Community to Dissect an Organism: Overview of the Neurospora Functional Genomics Project. *Adv. Genet.* **57**, 49–96 (2007).

51. Roche, C. M., Loros, J. J., McCluskey, K. & Glass, N. L. Neurospora crassa: Looking back and looking forward at a model microbe. *Am. J. Bot.* **101**, 2022–2035 (2014).

52. Morgan, W. T. & Watkins, W. M. Unravelling the biochemical basis of blood group ABO and Lewis antigenic specificity. *Glycoconj. J.* **17**, 501–530 (2000).

53. Franchini, M. & Bonfanti, C. Evolutionary aspects of ABO blood group in humans. *Clin. Chim. Acta* **444**, 66–71 (2015).

54. Ulloa, M., Hanlin, R. T. & Acosta, E. A. *Illustrated Dictionary of Mycology*. (APS Press, 2000).

55. Hann-Soden, C. Sex, Alternative Lifestyles, and a Graphic Study of a Model. (2019).

56. Hann-Soden, C. Geography and demography of clonal and sexual populations of Neurospora. (2018).

57. Springer, M. L. & Yanofsky, C. A morphological and genetic analysis of conidiophore development in Neurospora crassa. *Genes Dev.* **3**, 559–571 (1989).

58. Fleißner, A., Simonin, A. R. & Glass, N. L. Cell Fusion in the Filamentous Fungus, Neurospora crassa. in *Cell Fusion: Overviews and Methods* (ed. Chen, E. H.) 21–38 (Humana Press, 2008). doi:10.1007/978-1-59745-250-2_2

59. Simonin, A., Palma-Guerrero, J., Fricker, M. & Glass, N. L. Physiological Significance of Network Organization in Fungi. *Eukaryot. Cell* **11**, 1345–1352 (2012).

60. Richard, F., Glass, N. L. & Pringle, A. Cooperation among germinating spores facilitates the growth of the fungus, Neurospora crassa. *Biol. Lett.* **8**, 419–422 (2012).

61. Galagan, J. E. *et al.* The genome sequence of the filamentous fungus Neurospora crassa. *Nature* **422**, 859 (2003).

62. Fu, C. *et al.* Identification and Characterization of Genes Required for Cell-to-Cell Fusion in Neurospora crassa. *Eukaryot. Cell* **10**, 1100–1109 (2011).

63. Gonçalves, A. P. *et al.* Allorecognition upon cell-cell contact determines social cooperation and impacts the acquisition of multicellularity. *Curr. Biol.* (2019).

64. Heller, J., Clavé, C., Gladieux, P., Saupe, S. J. & Glass, N. L. NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc. Natl. Acad. Sci.* **115**, E2292–E2301 (2018).

65. Daskalov, A., Gladieux, P., Heller, J. & Glass, N. L. The allorecognition determinant rcd-1 controls programmed cell death in asexual spores of Neurospora crassa. *Prep.* (2019).

66. Gonçalves, A. P., Heller, J., Daskalov, A., Videira, A. & Glass, N. L. Regulated Forms of Cell Death in Fungi. *Front. Microbiol.* **8**, (2017).

67. Ishikawa, F. H. *et al.* Heterokaryon Incompatibility Is Suppressed Following Conidial Anastomosis Tube Fusion in a Fungal Plant Pathogen. *PLoS ONE* **7**, e31175 (2012).

68. Saupe, S. J. Molecular Genetics of Heterokaryon Incompatibility in Filamentous Ascomycetes. *Microbiol. Mol. Biol. Rev.* **64**, 489–502 (2000).

69. Glass, N. L. & Dementhon, K. Non-self recognition and programmed cell death in filamentous fungi. *Curr. Opin. Microbiol.* **9**, 553–558 (2006).

70. Metzenberg, R. L. & Glass, N. L. Mating type and mating strategies in Neurospora. *BioEssays* **12**, 53–59 (1990).

71. Shiu, P. K. & Glass, N. L. Cell and nuclear recognition mechanisms mediated by mating type in filamentous ascomycetes. *Curr. Opin. Microbiol.* **3**, 183–188 (2000).

72. Glass, N. L., Jacobson, D. J. & Shiu, P. K. T. The Genetics of Hyphal Fusion and Vegetative Incompatibility in Filamentous Ascomycete Fungi. *Annu. Rev. Genet.* **34**, 165–186 (2000).

73. Fleissner, A., Leeder, A. C., Roca, M. G., Read, N. D. & Glass, N. L. Oscillatory recruitment of signaling proteins to cell tips promotes coordinated behavior during cell fusion. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19387–19392 (2009).

74. Goryachev, A. B., Lichius, A., Wright, G. D. & Read, N. D. Excitable behavior can explain the "ping-pong" mode of communication between cells using the same chemoattractant. *BioEssays* **34**, 259–266 (2012).

75. Jonkers, W. *et al.* HAM-5 Functions As a MAP Kinase Scaffold during Cell Fusion in Neurospora crassa. *PLoS Genet.* **10**, (2014).

76. Fischer, M. S. Mechanisms of Communication between Genetically Identical Cells in Neurospora crassa. (UC Berkeley, 2018).

77. Fleißner, A. *et al.* The so Locus Is Required for Vegetative Cell Fusion and Postfertilization Events in Neurospora crassa. *Eukaryot. Cell* **4**, 920–930 (2005).

78. Fero, M. & Pogliano, K. Automated Quantitative Live Cell Fluorescence Microscopy. *Cold Spring Harb. Perspect. Biol.* **2**, (2010).

79. Doumane, M., Lionnet, C., Bayle, V., Jaillais, Y. & Caillaud, M.-C. Automated Tracking of Root for Confocal Time-lapse Imaging of Cellular Processes. *Bio-Protoc.* **7**, (2017).

80. Rabut, G. & Ellenberg, J. Automatic real-time three-dimensional cell tracking by fluorescence microscopy. *J. Microsc.* **216**, 131–137 (2004).

81. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

82. Chen, X., Velliste, M. & Murphy, R. F. Automated Interpretation of Subcellular Patterns in Fluorescence Microscope Images for Location Proteomics. *Cytom. Part J. Int. Soc. Anal. Cytol.* **69**, 631–640 (2006).

83. Wählby, C. *et al.* An image analysis toolbox for high-throughput *C. elegans* assays. *Nat. Methods* **9**, 714–716 (2012).

84. Shapiro, H. M. *Practical Flow Cytometry*. (John Wiley & Sons, 2005).

85. Vogel, H. J., Vogel, J. H., Vogel, H. J. & Vogel, H. A convenient growth medium for Neurospora crassa. (1956).

86. Westergaard, M. & Mitchell, H. K. Neurospora V. a Synthetic Medium Favoring Sexual Reproduction. *Am. J. Bot.* **34**, 573–577 (1947).

87. Freitag, M., Hickey, P. C., Raju, N. B., Selker, E. U. & Read, N. D. GFP as a tool to analyze the organization, dynamics and function of nuclei and microtubules in Neurospora crassa. *Fungal Genet. Biol.* **41**, 897–910 (2004).

88. Fischer, M. S., Wu, V. W., Lee, J. E., O'Malley, R. C. & Glass, N. L. Regulation of Cell-to-Cell Communication and Cell Wall Integrity by a Network of MAP Kinase Pathways and Transcription Factors in Neurospora crassa. *Genetics* **209**, 489–506 (2018).

89. Bhat, A., Noubissi, F. K., Vyas, M. & Kasbekar, D. P. Genetic analysis of wild-isolated Neurospora crassa strains identified as dominant suppressors of repeat-induced point mutation. *Genetics* **164**, 947–961 (2003).

90. Pandey, A., Roca, M. G., Read, N. D. & Glass, N. L. Role of a Mitogen-Activated Protein Kinase Pathway during Conidial Germination and Hyphal Fusion in Neurospora crassa. *Eukaryot. Cell* **3**, 348–358 (2004).

91. Ellison, C. E. *et al.* Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2831–2836 (2011).

92. Liu, H. *et al.* A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in Neurospora crassa. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7756–E7765 (2017).

93. Dettman, J. R., Jacobson, D. J. & Taylor, J. W. Multilocus sequence data reveal extensive phylogenetic species diversity within the Neurospora discreta complex. *Mycologia* **98**, 436–446 (2006).

94. Cascales, E. *et al.* Colicin Biology. *Microbiol. Mol. Biol. Rev.* **71**, 158–229 (2007).

95. McClure, B. A., Gray, J. E., Anderson, M. A. & Clarke, A. E. Self-incompatibility in Nicotiana alata involves degradation of pollen rRNA. *Nature* **347**, 757 (1990).

96. Saupe, S. J. The [Het-s] prion of Podospora anserina and its role in heterokaryon incompatibility. *Semin. Cell Dev. Biol.* **22**, 460–468 (2011).

97. Daskalov, A. *et al.* Contribution of specific residues of the β-solenoid fold to HET-s prion function, amyloid structure and stability. *PLoS Pathog.* **10**, e1004158 (2014).

98. Kaneko, I., Dementhon, K., Xiang, Q. & Glass, N. L. Nonallelic Interactions Between het-c and a Polymorphic Locus, pin-c, Are Essential for Nonself Recognition and Programmed Cell Death in Neurospora crassa. *Genetics* **172**, 1545–1555 (2006).

99. Hall, C., Welch, J., Kowbel, D. J. & Glass, N. L. Evolution and Diversity of a Fungal Self/Nonself Recognition Locus. *PLoS ONE* **5**, e14055 (2010).

100. Paoletti, M. & Clavé, C. The Fungus-Specific HET Domain Mediates Programmed Cell Death in Podospora anserina. *Eukaryot. Cell* **6**, 2001–2008 (2007).

101. Szewczyk, E. *et al.* Fusion PCR and gene targeting in Aspergillus nidulans. *Nat. Protoc.* **1**, 3111–3120 (2006).

102. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**, W20–W25 (2004).

103. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLOS ONE* **6**, e22594 (2011).

104. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

105. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).

106. Marchler-Bauer, A. *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).

107. Stern, A. *et al.* Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.* **35**, W506–W511 (2007).

108. Doron-Faigenboim, A., Stern, A., Mayrose, I., Bacharach, E. & Pupko, T. Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* **21**, 2101–2103 (2005).

109. Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708 (2003).

110. Viklund, H., Bernsel, A., Skwark, M. & Elofsson, A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinforma. Oxf. Engl.* **24**, 2928–2929 (2008).

111. Gouw, M. *et al.* The eukaryotic linear motif resource – 2018 update. *Nucleic Acids Res.* **46**, D428–D434 (2018).

112. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

113. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).

114. Bähler, M. & Rhoads, A. Calmodulin signaling via the IQ motif. *FEBS Lett.* **513**, 107–113 (2002).

115. Kusari, A. B., Molina, D. M., Sabbagh, W., Lau, C. S. & Bardwell, L. A conserved protein interaction network involving the yeast MAP kinases Fus3 and Kss1. *J. Cell Biol.* **164**, 267–277 (2004).

116. Garai, Á. *et al.* Specificity of Linear Motifs That Bind to a Common Mitogen-Activated Protein Kinase Docking Groove. *Sci. Signal.* **5**, ra74 (2012).

117. Xiong, Y. *et al.* The proteome and phosphoproteome of Neurospora crassa in response to cellulose, sucrose and carbon starvation. *Fungal Genet. Biol. FG B* **72**, 21–33 (2014).

118. Zarrinpar, A. & Lim, W. A. Converging on proline: the mechanism of WW domain peptide recognition. *Nat. Struct. Biol.* **7**, 611 (2000).

119. Huang, X. *et al.* Structure of a WW domain containing fragment of dystrophin in complex with β-dystroglycan. *Nat. Struct. Biol.* **7**, 634 (2000).

120. Roy, J., Li, H., Hogan, P. G. & Cyert, M. S. A Conserved Docking Site Modulates Substrate Affinity for Calcineurin, Signaling Output, and In Vivo Function. *Mol. Cell* **25**, 889–901 (2007).

121. Prokisch, H., Yarden, O., Dieminger, M., Tropschug, M. & Barthelmess, I. B. Impairment of calcineurin function in Neurospora crassa reveals its essential role in hyphal growth, morphology and maintenance of the apical Ca2+ gradient. *Mol. Gen. Genet. MGG* **256**, 104–114 (1997).

122. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

123. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).

124. Dettmann, A., Heilig, Y., Valerius, O. & Ludwig, S. Fungal Communication Requires the MAK-2 Pathway Elements STE-20 and RAS-2, the NRC-1 Adapter STE-50 and the MAP Kinase Scaffold HAM-5. *PLoS Genet.* **10**, (2014).