I. Low Cost Pathogen Detection with Yeast

II. Tools for Synthetic Multicellular Systems

Miguel Jimenez

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

We can now manipulate the genetic material of living organism routinely and cheaply. This has inspired a burgeoning field of synthesis based on DNA as a building block. The development of this new synthetic field has mirrored the trajectory of synthetic organic chemistry from small molecular systems to complex macromolecular assemblies. At first, this field of synthetic biology delivered recombinant proteins that enhanced our understanding of the structure-function relationship of biological macromolecules. Now, as the synthetic tools and analysis methods have come of age, synthetic whole-cell and multicellular systems have come within reach. In Chapter 1 we review the significant advances in DNA synthesis and analysis that have brought us to this point.

In this work, we first ask what practical applications will benefit most from the unique qualities of synthetic whole-cell system, such as their ability to replicate, sense and respond with molecular specificity. In Chapter 2, we implement a pathogen detection platform based solely on genetically modified yeast. This approach holds the potential to deliver ultra low-cost sensors that can be used and produced at the point-of-care. In Chapter 3, we develop methods to target these yeast-based sensors for the detection of any peptide biomarker of choice.

We next look forward to the potential of synthetic multicellular systems. While natural multicellular systems can be directly manipulated, our ability to rationally build multicellular systems from the bottom-up is still in its infancy. There still remain gaps in the available tools to make and analyze such synthetic systems. In Chapter 4, we leverage the explosion of available genomic databases to uncover a highly extensible set of cell-cell signaling modules. In Chapter 5, we implement ratiometric fluorescent tags to track mixed cell populations in multiplex. Together these components will be useful in implementing and analyzing synthetic communication networks that will be key components of advanced synthetic multicellular systems.

TABLE OF CONTENTS

TABLE OF FIGURES

Papers

Portions of chapters 2 and 4 will be published in

Nili Ostrov*, Miguel Jimenez*, Sonja Billerbeck*, James Brisbois, Joseph Matragrano, Alastair Ager, Virginia W. Cornish. "A Modular Yeast Biosensor for Low-Cost Point-of-Care Pathogen Detection." *Submitted.*

Portions of chapter 5 will be published in

Andrew V. Anzalone*, Miguel Jimenez*, Virginia W. Cornish et al. "CellTags: Resolvable Cell Markers for Population and Multiplexed Gene Expression Analysis". *In preparation*.

Casey J. Brown*, Gabriella Sanguineti*, Miguel Jimenez, Joshua L. Avins, Ruben L. Gonzalez, Jr., Virginia W. Cornish "Single Site Incorporation of β-Amino Acids." *In preparation*.

Patent Applications

Portions of chapters 2 , 3, and 4 were submitted as part of

Virginia W. Cornish, Nili Ostrov, Miguel Jimenez, Sonja Billerbeck. "Detection of Analytes Using Live Cells." *PCT Int. Appl.* PCT/US15/61373 (2015).

Awarded Grants

Portions of chapters 2 and 3 were submitted as part of

NIH R01 AI110794-01A1. PI Cornish, Co-PI Ager. "A Household Yeast Biosensor for Cholera." 2015-2020.

Portions of chapter 4 were submitted as part of

DARPA HR0011-15-2-0032. PI Cornish, Co-PI Boeke. "A scalable peptide/GPCR communication language." 2015-2017.

ACKNOWLEDGEMENTS

My mentor, Virginia Cornish, has been instrumental in shaping the direction and scope of the work presented here. I am very thankful for her guidance both scientifically and personally. I leave her group even more bright-eyed than when I joined. It truly has been a pleasure and a unique opportunity to learn from and work with her. I will never forget why we do this.

Nili Ostrov conceived the yeast sensor work and laid the foundation that enabled me to realize our goal. More importantly, Nili taught me and continues to teach me how science is done. She has set the example as a scientist that inspired me throughout my graduate years.

Sonja Billerbeck carried out the in depth work in validating the yeast sensors against the *Candida* pathogens. She was also key in conceiving of the fungal receptors as communication modules. She has also deeply shaped my approach to science, and helped me realize the richness of the biological fleamarket.

Sydney Blattman played a significant role in the optimization of the lycopene reporter and was one of the first undergraduates that I mentored. I was extremely lucky. She allowed the project to proceed in double time. Ana Pascoini and Katie Kim carried out a substantial portion of the clone characterization for the directed evolution experiments and helped establish the orthogonality of the natural receptors. I am proud to have been their mentor and I look forward to their success.

Andrew Anzalone and his *in vitro* work to develop the frame shift motifs both enabled and inspired the CellTags. Together we let our imaginations extend far beyond what we could ever hope to accomplish on the bench.

Jamie Brisbois carried out the sensitivity characterization of the diverse set of fungal receptors and has pushed far beyond what we initially imagined. He is an impeccable experimentalist.

Betty Lafourcade and Martin Mengel at Africhol have been an inspiration of the large impact that even small advances can have on overcoming the many challenges faced around the world. Similarly, Alastair Ager has provided many fresh views on why, how and for whom we do science.

Fernando Rodrigues and Chad Rappleye provided the *Paracoccidioides* and *Histoplasma* samples. It was a pleasure communicating with such committed and responsive colleagues.

It has also been a pleasure working with the other members of the Cornish group on a daily basis. The many different trajectories provide me an endless source of intellectual and social joy. More recently Joe Matragrano has picked up were I have left off, almost instantly moving beyond learning to results. I am excited to see the yeast sensor projects continue forward.

Ruben Gonzalez and Luis Campos have helped guide my path at Columbia. I have learned the most from them about how to inspire groups.

On a personal note, Miki Hayano has provided invaluable support, happiness and momentum forward. Fernando Jimenez, his dreams are inspiring. Maria Helena Delgado made it all possible. She has set the clearest example, that really anything is possible.

To the women in my life, to whom I owe so much.

1 AN INTRODUCTION TO SYNTHETIC BIOLOGICAL SYSTEMS

1.1 Introduction

The discovery of the molecular nature of genes in the 1950's fundamentally changed our ability to manipulate biological systems [1]. This soon led to the construction of the first recombinant virus and the establishment of DNA as a synthetic building block [2]. Since then, advances in methods to construct and sequence DNA have enabled the synthesis of ever more complex biological systems. The manipulation of single genes has yielded synthetic proteins such as recombinant human insulin, therapeutic antibodies and even enzymes that catalyze entirely non-natural reactions [3–5]. In turn, the manipulation of multiple genes has led to the construction of synthetic protein pathways such as those to generate anti-malarial compounds and others that can compute digital logic [6,7]. Building on these advances, the manipulation of whole genomes has brought synthetic cells within reach [8]. Through careful consideration and design, synthetic cells built around synthetic proteins have the potential to transform our society as small molecules did before them.

By observing the trajectory of organic chemistry and developments in small molecule synthesis, it is clear that we are now at a stage where synthetic cells can be targeted for practical applications[9]. As an example, we can take the heroic synthesis of the small molecule vitamin B_{12} by Woodward and Eschenmoser [10]. This landmark achievement drove the development of methods now routinely applied for the synthesis of the relatively simpler yet powerful therapeutic small molecules that dominate our pharmaceutical industry [11]. Likewise, the heroic projects to sequence the human genome and build the first fully synthetic microbial genome drove the development of an array of technologies that have made the manipulation of a cell's DNA routine [12,13]. With these methods in place we can now turn to the synthesis of cells that can be used directly to treat and diagnose disease, control pollution, enhance agricultural production and even harvest clean energy [14]. Indeed, applications such as synthetic T-cell cancer therapy are already being put into practice [15].

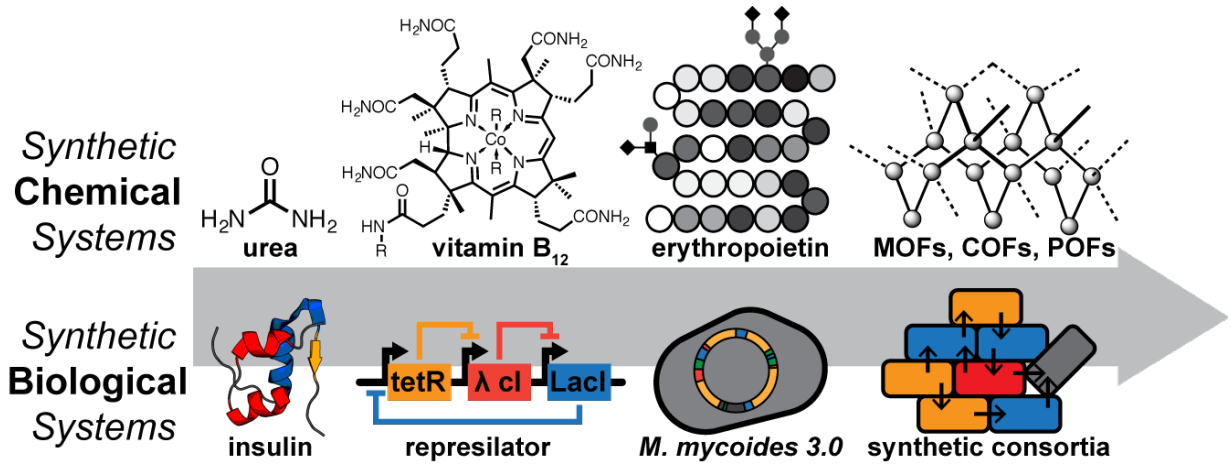


Figure 1.1 Synthesis of chemical and biological systems of increasing complexity
The synthetically accessible biological systems (built with DNA) have steadily increase in complexity, mirroring the progress made in the synthesis of chemical systems (built with covalent bonds).

Looking to the future, as the development of supramolecular chemistry has followed from advanced methods in small molecule synthesis, we can already anticipate that applications of synthetic multicellular systems are just around the corner [16,17].

The manipulation of DNA is however unique in several ways. Unlike small molecule synthesis, DNA manipulation directly yields self-replicating systems and can permanently modify existing species, including our own. While some of the greatest potential applications are possible because of these qualities, these synthetic capabilities require thoughtful consideration of possible ethical, legal and social implications. The scientific community set forth guiding principles immediately following the construction of the first recombinant DNA, and more recently has convened with the public at large to ratify research and development guidelines that will lower the risks and increase the benefits of synthetic biological systems [18,19]. Among them are (1) continued public international discourse to proportionally distribute risk while making resulting benefits widely available, (2) ethics education of the scientific community, (3) education of the public to increase scientific literacy, (4) prioritizing goals with the highest potential for public good, and (5) implementation of safeguards to monitor, contain and control synthetic biological systems. With these considerations in mind, we are eager to realize applications of synthetic cells, leveraging their unique properties in the hope of yielding substantial benefits for society.

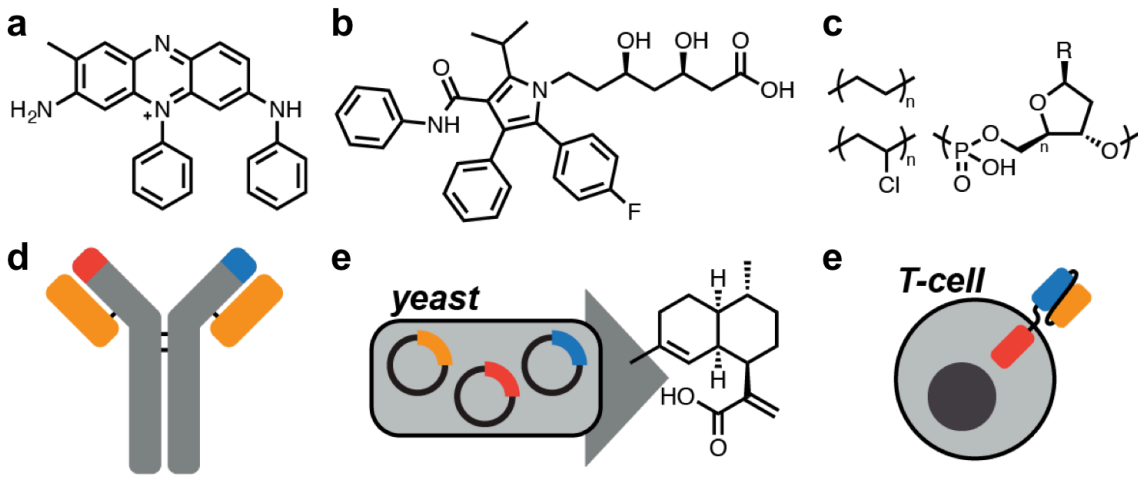


Figure 1.2 Practical applications of chemical and biological synthesis
Many of the most powerful applications of synthetic **(a-c)** chemical and **(d-e)** biological systems are often simple compared to the synthetically accessibible complexity (Figure 1.1). **(a)** mauvine, the first mass produced synthetic dye. **(b)** Atorvastatin, one of the best selling small molecule drugs. **(c)** Polyethelene, polyvinylchloride and oligonucleic acids, some of the most widely used synthetic polymers. **(d)** Therapeutic antibodies, one of the first biologics to come to market. **(e)** Yeast production of artemisinic acid, one of the successful applications of pathway engineering. **(f)** Chimeric receptor T-cells, one of the first commercial use of synthetic cells.

In this dissertation, we first describe the design, construction and evaluation of synthetic yeast cells for the detection of fungal pathogens (Chapter 2) and methods to develop synthetic receptors to expand the target range of these yeast-based sensors (Chapter 3). We then describe the development of tools for building (Chapter 4) and characterizing (Chapter 5) synthetic multicellular systems. As an introduction to this research, in the following sections we discuss the key milestones in the methods to construct and sequence DNA that have opened the door to this new synthetic science.

1.2 Building with DNA

Advances in our ability to use DNA as a synthetic building block have occurred at three levels: synthesis of short DNA oligonucleotides, assembly of these fragments into genes and incorporation of multi-gene pathways into genomes. In parallel with these target-oriented methods, a set of diversity-oriented methods has been developed that allow for the generation of large sets of variants that can be searched for the desired function. Through a combination of these two synthetic strategies several impressive results have already been achieved at the gene, pathway and genome level. The field of DNA-based synthesis is now turning to the development of tools for manipulating and building multicellular systems.

1.2.1 Oligonucleotide synthesis

By far the largest driver of progress has been the optimization of oligonucleotide synthesis that has drastically reduced the labor and cost of producing user-defined DNA sequences [20]. Early work established that phosphoramidites were the ideal building block for iterative construction of oligonucleotides [22]. Combinations of this coupling strategy with solid-phase synthesis and automation led to the commercialization of sequence-specific oligonucleotides in the 1980's [20]. These strategies have continued to be optimized, with oligonucleotides of up to 100 nucleotides in length now widely available for less than 10¢ US per nucleotide [21]. These methods are now being surpassed by microarray-based oligonucleotide synthesis. Microarray-based strategies take advantage of photolithography, ink-jet printing or semi-conductor electrochemistry for the multiplex synthesis of spatially localized oligonucleotides on a solid surface at high-densities [23–26]. Microarray-based synthesis is already being commercialized yielding oligonucleotides for less than 0.1¢ US per nucleotide with future enhancements through the use of microfluidics promising further reduction in cost to 0.001¢ US per nucleotide [21].

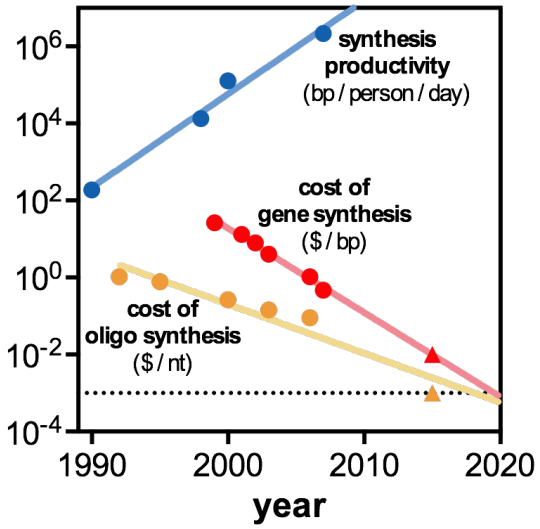


Figure 1.3 The cost of oligonuclotide and gene synthesis and synthesis productivity
The cost of oligonuclotide synthesis dropped from $1 per nucleotide (nt) in 1992 to $0.001 in 2015. This led to a drop in gene synthesis cost and increase in total base pairs (bp) that can be produced. Adapted from ref. [20] and recent points (triangles) taken from ref. [21].

1.2.2 Gene assembly

Even before cheap oligonucleotides were commonplace, the field started developing reliable methods for assembling gene-length fragments of DNA (200-5000 bp). These methods were originally designed to amplify and assemble DNA fragments derived from natural sources. The discovery and purification of DNA restriction and ligase enzymes allowed for the sequence-specific cutting and pasting of DNA [27–29]. Following this, the application of thermostable DNA polymerases and development of the polymerase chain reaction (PCR) led to an explosion in

methods to generate hybrid DNA molecules [30]. Since then, improvements have focused on optimizing reaction conditions and developing enzymes with improved properties in terms of fidelity, stability and speed. Now there is a formidable set of *in vitro* gene assembly methods, all based on some combination of DNA restriction, ligase and polymerase enzymes [31]. Among them, the isothermal assembly method developed by Gibson has become widely used due to its versatility in assembling any target sequence of choice [32]. Additionally, there has been a recent interest in defining gene "assembly standards" that are useful in automating and therefore commercializing gene assembly [33].

As the availability of cheap oligonucleotides has surged, many of these gene assembly methods have been modified to directly assemble batches of oligonucleotides into gene-length DNA. Here the focus has been on developing protocols that allow the correct fragments to assemble in "one-pot" reactions of many oligonucleotides. Small batches of oligonucleotides are routinely assembled using polymerase cycling assembly or isothermal assembly protocols [34,35]. The much larger pools of diverse oligonucleotides generated by microarray-based synthesis are assembled through the incorporation of auxiliary sequences. These "barcode" sequences are used to direct the PCR amplification of smaller subpools and are then enzymatically removed to yield small oligonucleotide batches that can be assembled by the standard methods [36].

A second challenge of assembling synthetic oligonucleotides has been the reduction of errors that are inevitably generated in the oligonucleotide synthesis step. Even the highest quality oligonucleotides have error rates of 1 in 200 nt which necessitates labor-intensive screening of many assembly products by transformation and clonal amplification in bacterial cells to find error-free products [21]. Therefore, mixtures of assembly products are first error-corrected through the use of enzyme cocktails that recognize base pair mismatches and then degrade or correct the DNA fragments with errors [37]. More recent approaches involve the application of high throughput sequencing methods to identify and remove oligonucleotides with errors [38]. A combination of these approaches can now lead to final product error rates as low as 1 in 10,000 nt [21].

1.2.3 Genome manipulation

Once generated, gene-length DNA can be further assembled into the genome of a host organism in order to enhance or supplement natural biological process. A majority early of methods focused on generating recombinant plasmids (circular DNA molecules of 3000 to 15,000 nt) that could be easily moved in and out of cells [39]. However, as synthetic targets have become more ambitious, the focus has turned to methods for directly manipulating the genomes of the host organism. Many techniques rely on natural DNA recombination processes, with many studies in yeast and mammalian cells simply delivering DNA fragments with homology to genomic loci [40,41]. These methods were later adapted to bacteria by expression of bacteriophage recombinases [42]. As an alternative, in bacteria and mammalian cells, bacteriophage and viral integration is also used [43,44]. These methods rely on natural integration processes, however for mammalian cells, viral delivery sets an upper limit in the size of the DNA fragment that can be integrated into the genome [44].

This has encouraged the improvement of recombination-based methods. In yeast and mammalian cells recombination can be made more efficient through the generation of double-stranded breaks to induce homologous recombination [45]. In yeast, application of homing endonucleases led to the development of methods such as Reiterative Recombination and Delitto Perfetto that allow for incorporation of large gene pathways and precise chromosomal deletions [46,47]. Efforts in mammalian cells have led to the development of enzyme systems such as TALENs, zinc finger nucleases and CRISPR-Cas9 that allow DNA cleavage at any user-specified location [48]. These systems, especially CRISPR-Cas9, are now being applied to a wide variety of organism including mammalian cells, yeast, bacteria, plants, fish and insects [49].

A parallel strategy aims to assemble chromosome-size DNA that can coexist with the natural genome of the host rather than integrating into it. This has led to the development of systems for bacteria (BACs), yeast (YACs and neochromosomes) and human cells (HACs) [50]. These methods are often used to study or assemble large pieces of DNA in a genetically tractable organism such as bacteria or yeast. These organisms are used as a sand box to build and produce these large DNA assemblies that can then be transferred to an alternate organism as a whole piece.

Yet another strategy involves the use mutagenic oligonucleotides to modify different locations in a genome. This approach has been adapted for the automated multiplex targeting of many genomic loci to enable large-scale modification of natural bacterial genomes [51].

1.2.4 Library approaches

Diversity-oriented methods aim to generate libraries of DNA molecules that can then be searched for the desired target function. These methods acknowledge that our ability to rationally design functional biological systems is limited by gaps in our understanding of the underlying biological processes. Therefore, many of the DNA manipulation methods described above have been adapted to generate libraries of molecules rather than single target molecules.

At the oligonucleotide level, mixtures of phosphoramidite building blocks can be used at defined steps in oligonucleotide synthesis. This generates libraries of oligonucleotides with randomized bases at user-defined positions [52]. Microarray-synthesis improves on this technique by individually synthesizing each desired variant oligonucleotide as just another member within the pool of generated oligonucleotides [53]. At the gene level, error-prone PCR techniques allow random miss-incorporation of nucleotides through the use of nucleotide analogues or altered buffer conditions [54,55]. Complementary to these techniques, gene libraries can also be generated through

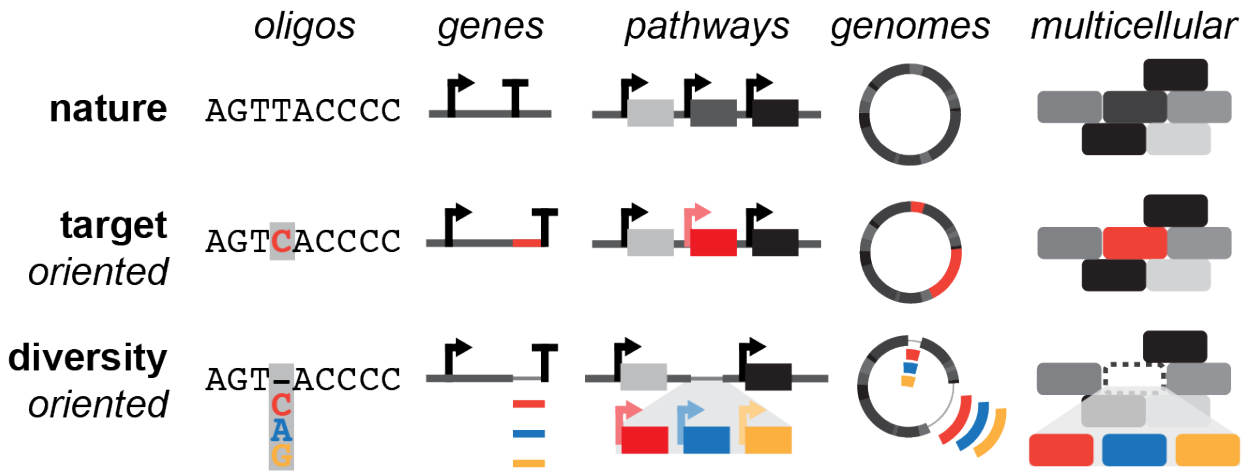


Figure 1.4 Synthesis of biological systems can be targeted or diversity oriented
Many aspects of natural biological systems remain unknown. Therfore, synthesis of new biological systems builds on natural ones (top row), modifying only small portions. These modifications can be targeted (middle rosw) or randomized to generate libraries of diverse variants (bottom row). There are now methods for carrying out these manipulations from the oligonuclotide through the multicellular level.

"shuffling" in which homologous "parent" DNA fragments are mixed into hybrid "daughter" DNA fragments by various PCR methods [56,57]. At the multi-gene level, mixtures of DNA fragments can be used in ligation and recombination based methods to generate DNA assemblies with random combinations of the individual components [58]. Finally, at the genome-level, mating and protoplast fusion can be used to shuffle genomes [59,60]. Alternatively, randomization systems can be built directly into fully synthetic genomes, such as the SCRaMbLE system in *Sc2.0* to generate libraries of synthetic cells [61].

An entirely different approach of generating diversity involves the use of environmental DNA. In these approaches random DNA is purified from environmental samples and subsequently inserted into microbial hosts [62]. These metagenomic libraries are then searched for a target function such as production of secondary metabolites. While these approaches are now primarily focused on biosynthetic pathways [63], it is possible that the rich source of environmental DNA can be searched for other complex functions, such as detection, energy storage, signaling and computation.

1.2.5 Building synthetic multicellular systems

As the tools for DNA manipulation continue to advance, soon practical applications based on synthetic multicellular systems will be possible. In a way, this has already occurred through the development of genetically modified agricultural plants [64]. While very useful, these examples are primarily single gene modifications. Advanced synthetic multicellular systems will be designed to carry out specific macroscopic functions through the collective actions of the individual synthetic cells. One simple method is to build such synthetic multicellular systems my mixing individually designed synthetic cells [65]. Examples of this have already been achieved in the laboratory. One example is a two-cell systems developed to oscillate only when both members are present [66]. A different approach involves the *in situ* manipulation of existing multicellular systems such as the microbiome [67]. A third approach involves co-opting natural developmental processes of multicellular organisms [68]. In the same way genome manipulations allow us to leverage existing cellular functions for new purposes, *in situ* manipulation will enable the redeployment of natural multicellular functions for novel applications.

1.3 Reading DNA and function

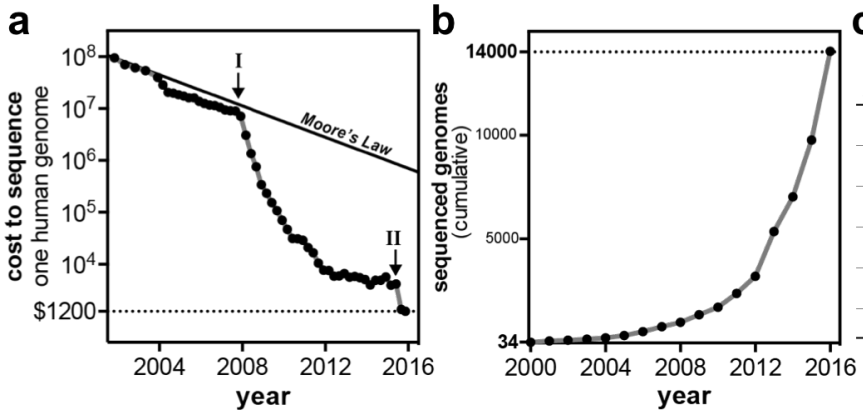
Our ability to sequence DNA has progressed in parallel to DNA manipulation technologies with many advances intertwined. These capabilities allow us to quickly validate and characterize synthetic biological systems build from DNA. In addition, DNA sequencing has allowed us to build immense databases of natural biological systems that can be redeployed as parts in synthetic systems. As with DNA manipulation, the field has developed diversity-oriented methods to characterize large DNA libraries by linking genotypes to selectable phenotypes. Now, with advanced synthetic methods in hand, the focus has turned to developing tools for characterizing and reading the phenotypes of synthetic multicellular systems.

1.3.1 Gene sequencing

Methods to sequence gene-length DNA have been the key enabling technology in the development of DNA as a synthetic building block. Maxam and Gilbert developed the first widely used sequencing method in 1977 [69]. This technique used four chemical degradation reactions of end-labeled DNA to generate characteristic fragment patterns from which the DNA sequence could be deduced. Almost at the same time, Sanger developed the chain-terminating inhibitor method. This method used a small fraction of labeled dideoxynuclnuclotides to prematurely terminate a polymerase extension [70]. The lengths of the resulting fragments could then be correlated to the nucleotide identity at that position. The Sanger method was subsequently improved through the use of fluorescent labels, capillary electrophoresis and automation [71]. The commercialization of Sanger sequencing supplanted the Maxam-Gilbert method and became the most widely used DNA sequencing method for 40 years and is still in wide use today. While this method is still ideal for sequencing targeted regions of DNA, the effective throughput is too low for the cost-effective sequencing of genomes.

1.3.2 Next generation sequencing and genomics

The next generation of sequencing technologies all achieved improvements in the throughput of DNA that could be sequenced through massive parallelization, reaction miniaturization and use of advanced imaging devices [72]. The key advance across all the technologies was the design of "one-pot" reactions that could amplify individual DNA molecules while keeping identical molecules



Genome size (Mbp)	Organism
150000	*P. japonica* (plant)
3300	*H. sapiens* (human)
2700	*M. musculus* (mouse)
12.1	*S. cerevisiae* (yeast)
4.6	*E. coli* (bacteria)
0.01	HIV (virus)

Figure 1.5 There has been an explosion in the available genomic data

(a) The cost to sequence a human genome has dropped at a pace far exeeding Moore's Law for the cost of computer chips. This super-exponential decrease was triggered by the commercialization of next generation sequencing (**I**) and more recently by the release of the Illumina X10 (**II**). **(b)** Due to this increase in sequencing power, there has been an explosion in the available fully sequenced genomes in public databases. **(c)** Genome sizes of some common organisms. The costs in panel a reflect human size genome. Data retireved from [75,76]

spatially co-localized. This effectively amplified the signal of single DNA molecules that could then be assayed in parallel. First among these technologies was pyrosequencing that applied emulsion PCR (emPCR) to amplify DNA on beads that could then be distributed to individual microwells for interrogation [73]. The most successful of the second-generation sequencing methods was Illumina sequencing [74]. This method uses a bridge amplification technique to amplify large batches of DNA molecules attached to a solid surface to generate > 10^8 clonal DNA spots. These are then read through the use of a polymerase extension and reversible fluorescent terminators.

Continued improvements in DNA sequencing technologies led to a reduction in the raw cost of DNA sequencing from US $2,000 per Mbp ($10^6$ base pairs) around the completion of the human genome project in 2003 to less than US 2¢ per Mbp in 2015 [75]. This dramatic drop in cost has sparked a "genomic revolution" leading to an explosion in the number of sequenced genomes from a total of 184 species by 2003 to 12,700 species by 2016 [77]. This database represents an immense amount of information and a rich source from which to derive parts for building synthetic biological systems.

More recent advances in DNA sequence have focused on generating useful sequence information from single molecules without amplification [78]. These new methods can enable single cell sequencing and will be useful in quickly uncovering heterogeneity in synthetic cellular systems.

1.3.3 Reading function through screening and selection

Advances in DNA construction and sequencing enable the rapid construction of synthetic genotypes, however our ultimate goal is to build biological systems with a target functional phenotype. Evaluation of function is straightforward for target-oriented approaches that yield a relatively small number of synthetic biological systems. However, as discussed in Section 1.2.4, many of the DNA construction methods can be used in a diversity-oriented manner to generate large libraries of genetic variants. Current methods can generate up to 10^8-10^9 synthetic genotypes. Therefore, while DNA sequencing methods can easily characterize this genetic diversity, phenotype analysis methods have lagged behind.

The majority of advances in phenotype screening have come from the field of metabolic engineering, which seeks to generate synthetic protein pathways that can biosynthesize a target small molecule. Direct methods use low through put screens ($<10^3$) based on gas and liquid chromatography coupled to mass spectrometry or NMR [79]. If the phenotype can be tied to a chromophore or fluorophore, micro titer plate-based or agar plate-based techniques can be applied. Micro titer plates can be used to screen 10^5 variants with high sensitivity. Agar plate-based colony screening methods can now cover 10^6 variants when coupled to robotics, however these methods have much lower sensitivity. In both cases, if the pheonotype does not generate a chromophore or fluorophore supplementary methods can be used such multiplexed extraction and chemical staining or incorporation of enzymes that can generate a chromophore from the target phenotype [79].

A complementary set of methods have been developed in the field of directed evolution, which seeks to apply Darwinian evolution on time-scales accessible in the laboratory to generate synthetic proteins [80]. In contrast to screening, the field of directed evolution has developed selection-based methods that effectively allow the interrogation on multiple phenotypes in "one-pot" reactions. The most powerful methods are growth-based selections that tie the target phenotype to a fitness advantage [80]. These methods can search libraries of $>10^{10}$ variants by allowing the fittest variants to enrich through exponential growth. Selections based on fluorescence-activated cell sorting (FACS) represent a second widely used selection method [81]. In these methods the target phenotype is tied to the production of a fluorescent signal, and desired variants

are sorted at high rates that can cover libraries of 10^9 in one day. An analogous method relies on microfluidics based sorting of water-in-oil emulsions [82]. Yet another set of methods involves the use of batch affinity-based purification to search libraries of variants whose genotype can be tied to the expression of affinity tags [83].

All of these methods are limited to phenotypes that have an easily measured output. Therefore much focus has been given to the development of methods to read a larger set of phenotypes. Many of these methods rely on tying the target pheonotype to a reporter gene such as a fluorescent protein or auxotrophic marker that can be used for screening or selection. Chemical complementation was an early example that tied enzymatic bond formation or cleavage to the transcriptional modulation of a reporter gene [84]. Other methods have used allosterically modulated transcription factors, protein dimerization, protein stability, protein conformational changes and riboswitches to tie the target phenotype to a measureable output [85].

More recent methods seek to leverage the extremely high through put of the DNA sequencing technologies discussed in Section 1.3.2 to instead read phenotypes. In these methods, the target phenotype is tied to an output that leads to changes in the sequence, abundance or co-localization of DNA molecules [86,87]. Subsequent sequencing can then reveal the presence of this output signal encoded in DNA.

1.3.4 Reading synthetic multicellular systems

As the tools for reading DNA and measuring phenotypes have advanced, they have also been applied to the characterization of multicellular systems. For example DNA sequencing technologies have led to the sequencing of the human microbiome, the collection of interacting microbial communities that live within and around our bodies [88,89]. In addition, the most advanced sequencing methods are now being used to sequence the genetic variation within individual cells in natural multicellular organism. For example, recent methods have been developed for sequencing the receptor repertoire of the immune cells within one individual [90]. These methods can in turn be used to read the genotypes of synthetic multicellular systems. As with more simple synthetic biological systems, the ability to read the phenotype of multicellular systems has also lagged behind. Methods such as differential growth on selective plates, rRNA sequencing, quantitative PCR tagging

and flow cytometry have been used to measure the underlying phenotypic structure of synthetic multicellular systems [91–94]. However, there remains a gap in high through put methods to track these complex systems in real time.

1.4 Conclusions

Since the middle of the 20th century, we have seen an unprecedented refinement in our ability to build synthetic biological systems by assembling DNA and characterizing the resulting function. As in the development of organic chemistry, DNA-based synthesis was at first used simply as a tool to gain greater insight into the underlying relationships between the structure and function of natural biological systems. In the last 20 years, these new synthetic capabilities have been increasingly used to develop synthetic biological systems with novel functions not found in their natural counterparts. These efforts have progressively increased in ambition, from synthetic proteins, to synthetic protein pathways, to synthetic cells and now to synthetic multicellular systems. At the first of these levels, synthetic proteins such as therapeutic antibodies have already become standard in clinical use. Efforts at the subsequent levels are close behind. We expect that this DNA-based synthetic science will yield significant solutions to the most pressing challenges facing our society.

1.5 References

1. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).

2. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2904–2909 (1972).

3. Goeddel, D. V. *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 106–110 (1979).

4. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: successes, limitations and hopes for the future. *Br. J. Pharmacol.* **157**, 220–233 (2009).

5. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).

6. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).

7. Weiss, R. *et al.* Genetic circuit building blocks for cellular computation, communications, and signal processing. *Nat. Comput.* **2**, 47–84

8. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).

9. Yeh, B. J. & Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **3**, 521–525 (2007).

10. Woodward, R. B. The total synthesis of vitamin B12. *Pure Appl. Chem.* **33**, (1973).

11. Lindsley, C. W. 2014 Prescription Medications in the United States: Tremendous Growth, Specialty/Orphan Drug Expansion, and Dispensed Prescriptions Continue to Increase. *ACS Chem. Neurosci.* **6**, 811–812 (2015).

12. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

13. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).

14. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* **15**, 289–294 (2014).

15. Dai, H., Wang, Y., Lu, X. & Han, W. Chimeric Antigen Receptors Modified T-Cells for Cancer Therapy. *J. Natl. Cancer Inst.* **108**, djv439 (2016).

16. Lehn, J. M. Supramolecular chemistry. *Science* **260**, 1762–1763 (1993).

17. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).

18. Berg, P. *et al.* Potential Biohazards of Recombinant DNA Molecules. *Science* **185**, 303–303 (1974).

19. *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. (Presidential Commission for the Study of Bioethical Issues, 2010).at <http://bioethics.gov/synthetic-biology-report>

20. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091–1094 (2009).

21. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).

22. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).

23. Singh-Gasson, S. *et al.* Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* **17**, 974–978 (1999).

24. Hughes, T. R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).

25. Saaem, I., Ma, K.-S., Marchi, A. N., LaBean, T. H. & Tian, J. In situ Synthesis of DNA Microarray on Functionalized Cyclic Olefin Copolymer Substrate. *ACS Appl. Mater. Interfaces* **2**, 491–497 (2010).

26. Egeland, R. D. & Southern, E. M. Electrochemically directed synthesis of oligonucleotides for DNA microarray fabrication. *Nucleic Acids Res* **33**, e125 (2005).

27. Lehnman, I. R. DNA Ligase: Structure, Mechanism, and Function. *Science* **186**, 790–797 (1974).

28. Smith, H. O. & Welcox, K. W. A Restriction enzyme from Hemophilus influenzae. *J. Mol. Biol.* **51**, 379–391 (1970).

29. Kelly, T. J. & Smith, H. O. A restriction enzyme from Hemophilus influenzae. *J. Mol. Biol.* **51**, 393–409 (1970).

30. Mullis, K. *et al.* Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).

31. Chao, R., Yuan, Y. & Zhao, H. Recent advances in DNA assembly technologies. *FEMS Yeast Res.* **15**, 1–9 (2015).

32. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

33. Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* **16**, 568–576 (2015).

34. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).

35. Dormitzer, P. R. *et al.* Synthetic Generation of Influenza Vaccine Viruses for Rapid Response to Pandemics. *Sci. Transl. Med.* **5**, 185ra68-185ra68 (2013).

36. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–1299 (2010).

37. Carr, P. A. *et al.* Protein-mediated error correction for de novo DNA synthesis. *Nucleic Acids Res.* **32**, e162–e162 (2004).

38. Schwartz, J. J., Lee, C. & Shendure, J. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat. Methods* **9**, 913–915 (2012).

39. Lodish, H. *et al.* DNA Cloning with Plasmid Vectors. (2000).at <https://www.ncbi.nlm.nih.gov/books/NBK21498/>

40. Thaler, D. S. & Stahl, F. W. DNA Double-Chain Breaks in Recombination of Phage lambda and of Yeast. *Annu. Rev. Genet.* **22**, 169–197 (1988).

41. Capecchi, M. R. The new mouse genetics: Altering the genome by gene targeting. *Trends Genet.* **5**, 70–76 (1989).

42. Zhang, Y., Buchholz, F., Muyrers, J. P. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123–128 (1998).

43. Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. M. New Applications for Phage Integrases. *J. Mol. Biol.* **426**, 2703–2716 (2014).

44. Gaj, T., Epstein, B. E. & Schaffer, D. V. Genome Engineering Using Adeno-associated Virus: Basic and Clinical Research Applications. *Mol. Ther.* **24**, 458–464 (2016).

45. Wyman, C. & Kanaar, R. DNA Double-Strand Break Repair: All's Well that Ends Well. *Annu. Rev. Genet.* **40**, 363–383 (2006).

46. Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U A* **108**, 15135–15140 (2011).

47. Storici, F. & Resnick, M. A.in *DNA Repair Part B* **Volume 409**, 329–345 (Academic Press, 2006).

48. Gaj, T., Gersbach, C. A. & Barbas III, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).

49. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).

50. Lufino, M. M., Edser, P. A. & Wade-Martins, R. Advances in High-capacity Extrachromosomal Vector Technology: Episomal Maintenance, Vector Delivery, and Transgene Expression. *Mol. Ther.* **16**, 1525–1538 (2008).

51. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

52. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* **9**, 1797–1804 (2008).

53. Larman, H. B. *et al.* Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **29**, 535–541 (2011).

54. McCullum, E., Williams, B. R., Zhang, J. & Chaput, J.in *Vitro Mutagen. Protoc.* (Braman, J.) 103–109 (Humana Press, 2010).doi:10.1007/978-1-60761-652-8_7

55. Zaccolo, M. & Gherardi, E. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 β-lactamase1. *J. Mol. Biol.* **285**, 775–783 (1999).

56. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).

57. Tee, K. L. & Wong, T. S. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnol. Adv.* **31**, 1707–1721 (2013).

58. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4**, e5553 (2009).

59. Zhang, Y.-X. *et al.* Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**, 644–646 (2002).

60. Hou, L. Novel methods of genome shuffling in Saccharomyces cerevisiae. *Biotechnol. Lett.* **31**, 671–677 (2009).

61. Shen, Y. *et al.* SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36–49 (2016).

62. Kim, J. H. *et al.* Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **93**, 833–844 (2010).

63. Milshteyn, A., Schneider, J. S. & Brady, S. F. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chem. Biol.* **21**, 1211–1223 (2014).

64. Key, S., Ma, J. K.-C. & Drake, P. M. Genetically modified plants and human health. *J. R. Soc. Med.* **101**, 290–298 (2008).

65. Song, H., Ding, M.-Z., Jia, X.-Q., Ma, Q. & Yuan, Y.-J. Synthetic microbial consortia: from systematic analysis to construction and applications. *Chem. Soc. Rev.* **43**, 6954–6981 (2014).

66. Chen, Y., Kim, J. K., Hirning, A. J., Josić, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Science* **349**, 986–989 (2015).

67. Sheth, R. U., Cabral, V., Chen, S. P. & Wang, H. H. Manipulating Bacterial Communities by in situ Microbiome Engineering. *Trends Genet.* **32**, 189–200 (2016).

68. Ismagilov, R. F. & Maharbiz, M. M. Can we build synthetic, multicellular systems by controlling developmental signaling in space and time? *Curr. Opin. Chem. Biol.* **11**, 604–611 (2007).

69. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci.* **74**, 560–564 (1977).

70. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463–5467 (1977).

71. Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. Large-scale and automated DNA sequence determination. *Science* **254**, 59–67 (1991).

72. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).

73. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).

74. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

75. Wetterstrand, K. A. DNA Sequencing Costs: Data. *NHGRI Genome Seq. Program GSP* at <www.genome.gov/sequencingcostsdata>

76. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

77. Kitts, P. *Genome Assembly and Annotation Process*. (National Center for Biotechnology Information (US), 2003).at <https://www.ncbi.nlm.nih.gov/books/NBK21086/>

78. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).

79. Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu. Rev. Biochem.* **79**, 563–590 (2010).

80. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).

81. Yang, G. & Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **10**, 2704–2715 (2009).

82. Agresti, J. J. *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci.* **107**, 4004–4009 (2010).

83. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).

84. Baker, K. *et al.* Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U A* **99**, 16537–16542 (2002).

85. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42**, 84–91 (2016).

86. Gu, L. *et al.* Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature* **515**, 554–557 (2014).

87. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* aag0511 (2016).doi:10.1126/science.aag0511

88. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).

89. Di Bella, J. M., Bao, Y., Gloor, G. B., Burton, J. P. & Reid, G. High throughput sequencing methods and analysis for microbiome research. *J. Microbiol. Methods* **95**, 401–414 (2013).

90. McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat. Protoc.* **11**, 429–442 (2016).

91. Morgan, X. C. & Huttenhower, C. Chapter 12: Human Microbiome Analysis. *PLOS Comput Biol* **8**, e1002808 (2012).

92. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).

93. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci.* **111**, E2149–E2156 (2014).

94. Dhoble, A. S. *et al.* A novel high-throughput multi-parameter flow cytometry based method for monitoring and rapid characterization of microbiome dynamics in anaerobic systems. *Bioresour. Technol.* **220**, 566–571 (2016).

2 A YEAST SENSOR FOR LOW-COST POINT-OF-CARE PATHOGEN DETECTION

2.1 Introduction

A scalable, cost-effective approach for pathogen detection outside the laboratory would greatly enhance our ability to track disease [1], ensure food safety [2], and neutralize bioterrorism threats [3]. The need for such an approach is particularly critical in the developing world [4]. While antibody and nucleic acid bioassays have been very effective in the laboratory and the clinic, these approaches are limited by the need for expensive reagents, equipment and techniques [5]. These challenges have spurred the development of many innovative pathogen biosensor devices, however, specialized reagents and matrix materials continue to be critical for function [6]. In fact, conventional culturing methods remain in widespread use *despite* being slow and requiring trained personnel [7].

The emerging field of synthetic biology provides a powerful toolset of parts for addressing this global health challenge [8,9]. Much like advances in molecular biology gave rise to antibody diagnostics [10], the ability to design and build synthetic organism holds the potential to fill gaps in existing networks for health care delivery and surveillance. While synthetic organisms have primarily been leveraged for economical fermentation of industrial and biomedical commodities via metabolic engineering [11], significant applications in public health, animal health and agriculture remain untapped.

In this chapter, we describe the development of a simple and highly specific colorimetric assay for detection of pathogen-derived peptides based on *Saccharomyces cerevisiae* - a genetically tractable model organism and household product. Integrating G-protein coupled receptors (GPCRs) with a visible, reagent-free lycopene readout, we demonstrate detection of major human fungal pathogens with nanomolar sensitivity (Figure 2.1). We further optimized a one-step rapid dipstick prototype that can be used in complex samples including urine and soil. This modular biosensor can be economically produced at large-scale, is not reliant on cold-chain storage, can be detected without additional equipment, and thus is a compelling platform scalable to global surveillance of pathogens.

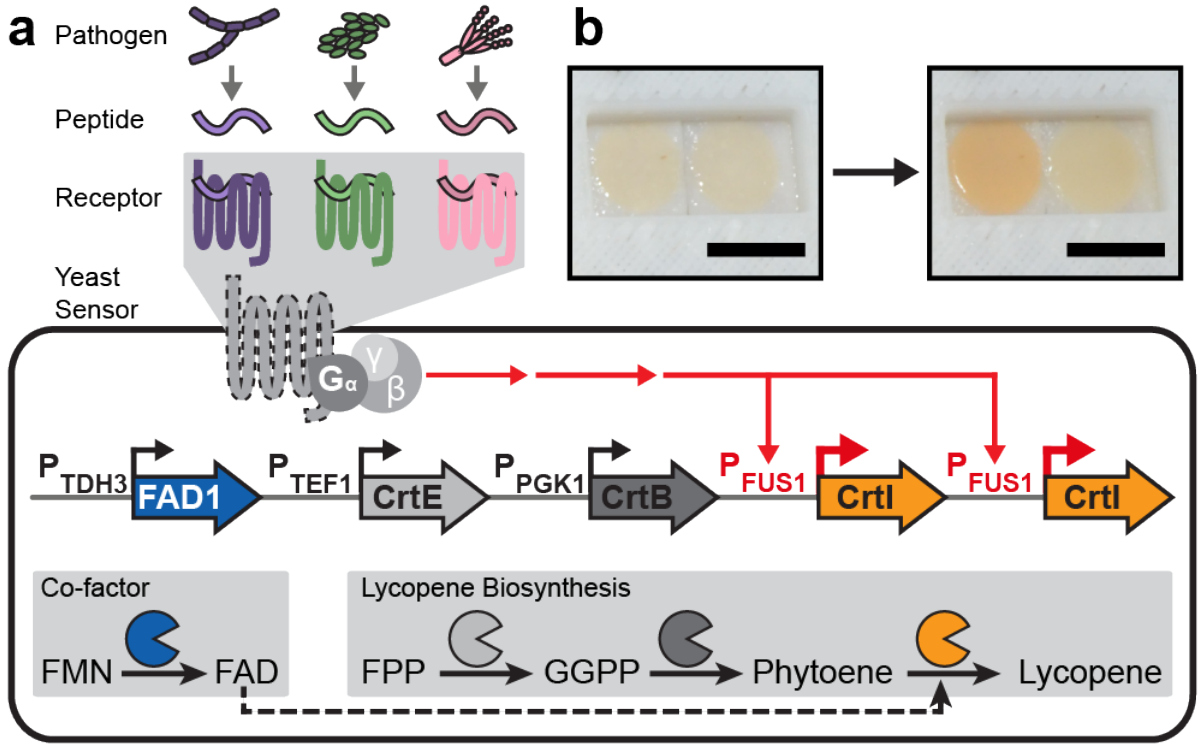


Figure 2.1 ***S. cerevisiae*** **biosensor for detection of fungal pathogens**
(a) Overview of biosensor components. Highly specific fungal receptors provide sensitive response to mating peptides secreted by pathogenic fungi. Activation of the downstream mating signaling pathway induces transcriptional activation of biosynthetic genes for production of red lycopene pigment visible to the naked eye. FMN: flavin-mononucleotide, FAD: flavin-adenine-dinucleotide, FPP: farnesyl-pyrophosphate, GGPP: geranylgeranyl-pyrophosphate. **(b)** Color signal as shown in paper-based dipstick assay (bar = 0.5cm)

2.1.1 Whole-cell sensors

The natural mechanisms used by cells to sense and react to changes in their environment have long been recognized as paradigm for building synthetic sensors out of whole cells. Indeed, several groups have established that microbial cells can be used for sensing toxic metals[12], vital parameters [e.g. glucose, NOx] [13], hormones [14], tumors [15], fine chemicals [16], quorum sensing molecules [17], antibiotics [18] and explosives [19]. Other have proposed the use of mammalian cells as pathogen sensors [20]. However, mammalian cell-based assays are not easily translated outside the laboratory since they can only survive in a narrow range of conditions. Furthermore, available microbial-based assays primarily rely on *E. coli*, a host that has the potential to become pathogenic. Finally, these previous approaches have not developed readouts that can be used by a non-specialist without equipment outside the laboratory.

As opposed to other cell types, yeast is an ideal choice for making a product for common use. In contrast to *E. coli* and mammalian cells, the yeast *S. cerevisiae* is impervious to a wide range of environments, can be freeze-dried for long-term storage, is a non-toxic organism that has existed in the human diet since the rise of agriculture, and, most importantly, it is a proven commercially viable product already distributed worldwide in the food industry [21–24].

2.1.2 G-protein coupled receptors in yeast

Previous sensors developed from yeast cells have relied on non-specific response mechanisms or intracellular receptors [22,25]. There have been several reports on the development of yeast sensors to cytotoxic compounds, however these systems use the global stress responses in yeast to generate a signal [25]. Therefore while useful, target compounds cannot be detected with molecular specificity. The second major class of yeast sensors has been developed around nuclear hormone receptors [26]. These receptors are allosterically regulated transcription factors that naturally respond to an array of small molecule hormones. While there has been great success in these systems, the detection is inherently limited to cell permeable compounds. Moreover, it is unclear how wide of a target scope can be reached by engineering these receptors.

G-protein couple receptors (GPCRs) represent an attractive alternative for pathogen detection. GPCRs are the ideal sensing element because these receptors have a natural ligand diversity rivaling antibodies [27]. Moreover, these receptors are expressed on the cell surface and naturally transform an extracellular binding event into an intracellular signal. Therefore, they have the potential to be

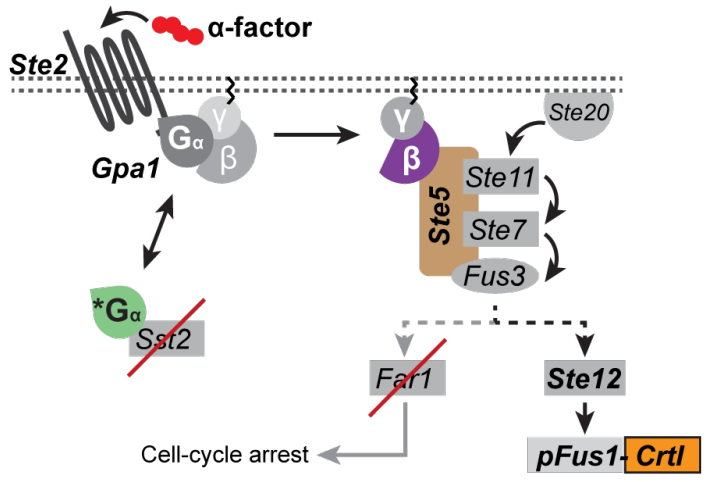


Figure 2.2 Pheromone response signaling pathway
Detailed view of red arrows in Figure 2.1. Activation of the pheromone response initiates a phosphorelay cascade that culminates in activation of *Far1* and *Ste12*. Deletion of *Sst2* and *Far1* allows efficient expression of reporter genes under the *Fus1* promoter. Double dashed lines represent lipid bilayer.

used to sense any target ligand outside the cell. In addition, GPCR function is well established in yeast. There has been substantial GPCR research in yeast to develop high-throughput assays for drug discovery based around mammalian receptors [28–30]. Motivated by previous results, we initially explored mammalian GPCRs as potential pathogen sensing elements [31,32]. However, though functional, we found these systems to be fragile and sensitive to strain background. In this chapter, we instead focus on the development of pathogen sensor based around fungal mating GPCRs. This GPCR family naturally binds a diverse set of peptides and thus has the potential to be applied for the detection of any peptide biomarker.

2.2 Results

2.2.1 Considerations of developing a yeast-based point-of-care sensor

Our goal was to design, build and test a sensor for pathogenic organisms that had the potential to be used directly outside the laboratory for point-of-care detection. Simply building a functional sensor could be accomplished by using one of several existing sensing and reporting modalities in common use for laboratory-based biosensors [33]. However, our aim was to build a sensing platform that could fill real unmet needs outside the laboratory by leveraging the unique capabilities of living cells. We thus engaged experts in pathogen surveillance, diagnostic development and field-testing and international implementation to establish key requirements for a yeast biosensor to have significant clinical and epidemiological utility.

Analysis of existing commercial diagnostic tools suggested that a key benefit that could be delivered by a whole-cell biosensor was an extreme reduction in the cost per test. While available nucleic acid-based diagnostics are becoming the gold standard detection method in clinical pathology laboratories, the requirement for expensive equipment and reagents prevents widespread use outside the laboratory for pathogen surveillance [34]. Similarly, antibody-based assays, while benefiting from high molecular specificity, still remain prohibitively expensive for widespread use outside the laboratory. Even products specifically designed for field-use such as the Pasteur Institute's rapid-diagnostic test for *V. cholerae* cost $10 per test and in practice are used only to validate highly suspect samples [35,36]. These considerations suggested that a yeast-based sensor

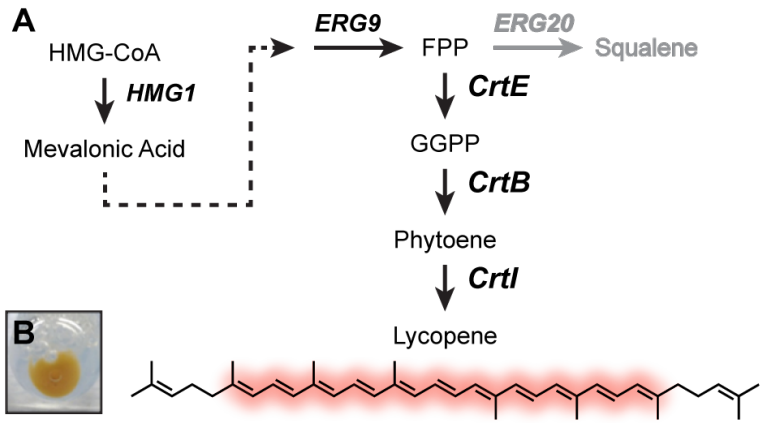


Figure 2.3 Lycopene biosynthetic pathway

(a) Introduction of *E. herbicola* carotenoid enzymes (*CrtEBI*) result in biosynthesis of lycopene from endogenous yeast farnesyl pyrophosphate. **(b)** A lycopene-producing yeast strain becomes visibly colored.

should be designed to minimize sample pre-treatment and completely eliminate the need for expensive equipment and reagents to generate a readout. Based on cost analysis of current yeast production, a reagent-less yeast-based assay could be predicted to cost less than 10 cents per test with the potential to cost as little as 1 cent per test.

This guiding principle further suggested that a yeast-based biosensor would stand to have the highest impact for applications in resource poor settings. Specifically, detection of pathogens afflicting low- and middle-income countries should be considered as primary targets for detection. Additionally, for high-income countries, detection of pathogens afflicting agricultural crops and life stock should also stand to benefit from low cost sensors. Furthermore, any application requiring direct environmental detection or continuous real-time surveillance would also become accessible with the development of ultra low cost sensors.

Ease of use by non-specialists in resource poor settings was determined to be a second key quality for a yeast-based sensor. A cheap and easy to use detection device would stand to enable community-based assessment and surveillance of disease. Due to disease characteristics and available infrastructure in low-income countries, substantial increases in the number of symptomatic individuals within a population are the first real signs of the presence of a disease outbreak.[37] Therefore, prompt assessment and even self-assessment of symptomatic individuals would potentially alert communities to the spread of a disease early in an outbreak. In addition, ease of use by non-specialists would enable educational applications for communities and state players to implement disease prevention campaigns [38]. These considerations suggested that a yeast-based sensor should be robust enough to function outside carefully controlled conditions only available to

specialized personnel. This pointed to the use of robust well-established components as well as building the components directly into the genome of yeast to avoid the use of plasmids that require specialized selective media.

This analysis led us to set the following design requirements: cost <10 cents, time to readout <12 hours, no reagents to generate readout, clear binary readout, amenable for "one-step" use. In addition, based on established rapid diagnostics we determined that detection should be based on a detection element with high molecular specificity and a sensitivity of < 1 uM and ideally <100 nM for the target analyte. Based on these requirements we designed a yeast sensor built around genome-integrated GPCRs as modular sensing elements, the yeast pheromone response pathway as a universal signal transduction element and lycopene biosynthesis as a readout element visible to the naked eye (Figure 2.1).

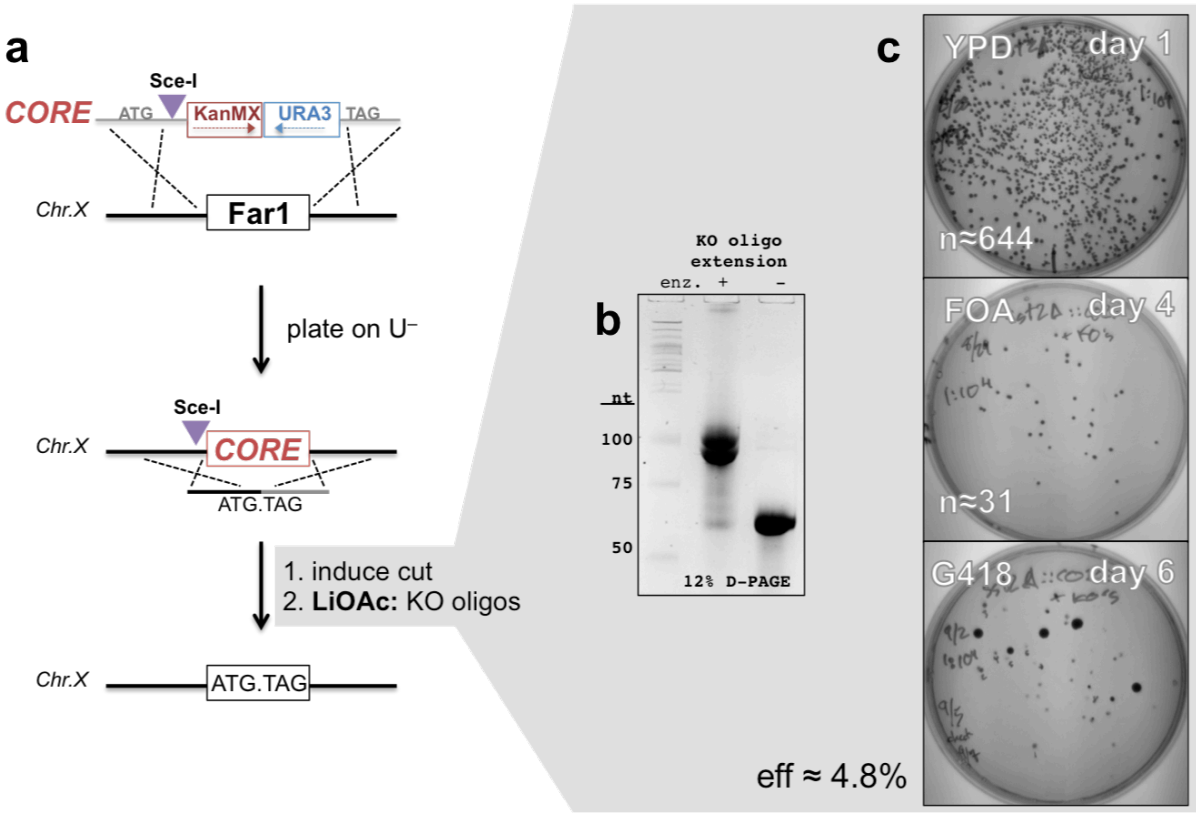


Figure 2.4 Genomic gene delition by Delitto Perfetto

(a) Genes are deleted in a two step process. First, the CORE casette is integrated into the target locus (FAR1). Then, a double strand break is induced at the Sce-I cut site and knock out (KO) oligonucleotides are transformed, repairing the break and completing the deletion. **(b)** KO oligos are extended to make a 100 nt duplex before use. **(c)** Typical replica plating results giving a deletion efficiency of 4.8%.

2.2.2 Construction of a first generation sensor

We envisioned that the natural plant pigment lycopene would be the ideal readout for a point-of-care sensor because it can be visualized by the naked eye, is widely validated in yeast metabolic engineering, and is non-toxic. Lycopene is the first intermediate in carotenoid biosynthesis that has a sufficiently conjugated π-system to absorb in the visible region (Figure 2.3)[39]. Thus, unlike standard laboratory reporters like *lacZ* that require exogenously added caged dyes (X-gal) or fluorescent proteins that require specialized equipment (fluorimeter), lycopene can be directly observed by a non-specialist. Additionally, the biosynthesis of lycopene from endogenous yeast farnesyl pyrophosphate is well established, requiring only three heterologous genes.[40] This simplicity suggested that a lycopene-based readout would be robust enough to function outside ideal laboratory conditions. Finally, not only is lycopene non-toxic, metabolic engineering of yeast aims to produce lycopene and its derivatives at industrial scales as beneficial food supplements.[41]

Similarly, we expected the *S. cerevisiae* GPCR-based pheromone response pathway to be ideally suited for signal transduction because it has fast kinetics and it is very well characterized. We expected that a simple, transcriptionally controlled reporter would give a fast readout, because transcripts increase in <15 min after binding the native ligand [42]. Also, all the component proteins of the MAP kinase pathway that mediate pheromone signaling have been well characterized and this pathway served as a model for elucidating other known MAP kinase pathways (Figure 2.2). Moreover, of the two GPCR pathways in yeast, the pheromone pathway leads to a very specific response: the up regulation of an established set of genes via the transcription factor, *Ste12*.[43,44] We expected the resulting sensor be robust outside ideal conditions since this endogenous signal transduction system has been perfected through natural selection for yeast mating in a wide range of environments. Finally, the demonstrated flexibility of this signaling pathway to couple to a range of heterologous GPCRs crucially enables the use of GPCRs as modular sensing elements to generate sensor to an array of targets.

We began by generating a first generation sensor: a yeast strain that produces lycopene in response to the natural *S. cerevisiae* pheromone, α-Factor (αF). We made a parental sensor strain by deleting the cyclin-dependent kinase inhibitor *Far1* to prevent cell-cycle arrest and deleted the G-

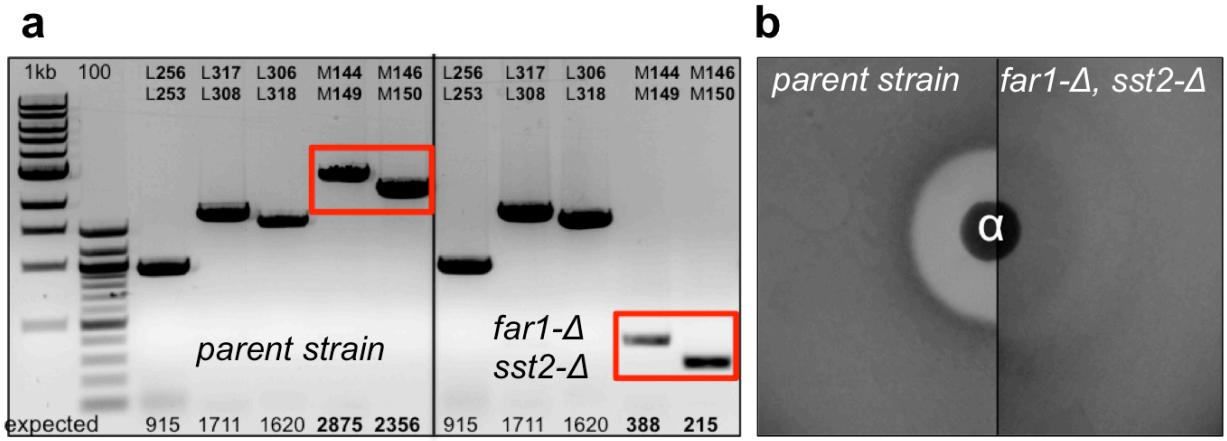


Figure 2.5 Deletion of FAR1 prevents cell cycle arrest

(a) PCR analysis of gene deletions. Red boxes show fragments for the target loci (far1, sst2) are smaller after deletion. **(b)** Halo assays showing far1 deletion prevents growth arrest. Dark areas have dense cell growth and light areas have reduced growth. α – 10 nmol of alpha factor spoted in both panels.

protein activating protein *Sst2* to prevent signal attenuation using Delitto Perfetto (Figure 2.4) [28,45]. As expected, a halo assay with against alpha factor demonstrated that the parental strain no longer had a growth arrest phenotype (Figure 2.5) [46]. However, a band of slightly lower cell density can be seen some distance from the inducing spot (Figure 2.5). It is clear that the underlying pheromone response is complex with secondary growth effects revealed by the deletion of FAR1 and SST2. Next, using Reiterative Recombination, we placed lycopene biosynthesis under the control of the pheromone response pathway [47]. We accomplished this by placing carotenoid genes derived from *E. herbicola* (CrtE, CrtB) under the control of constitutive promoters and placing the final biosynthetic gene *CrtI* under control of the promoter from FUS1, a downstream target of the pheromone response pathway (Figure 2.2) [43].

In shake flask conditions, this first generation sensor took over two days to produce lycopene levels discernible by the naked eye. FUS1 control of *CrtB* was also tested but this lead to high basal lycopene biosynthesis. This first generation sensor validated that an extracellular analyte could trigger the intracellular biosynthesis of a pigment visible to the naked eye. Nevertheless, it did not meet our design requirements for the readout both in terms of speed and strength of the signal.

2.2.3 A robust method for quantifying lycopene production

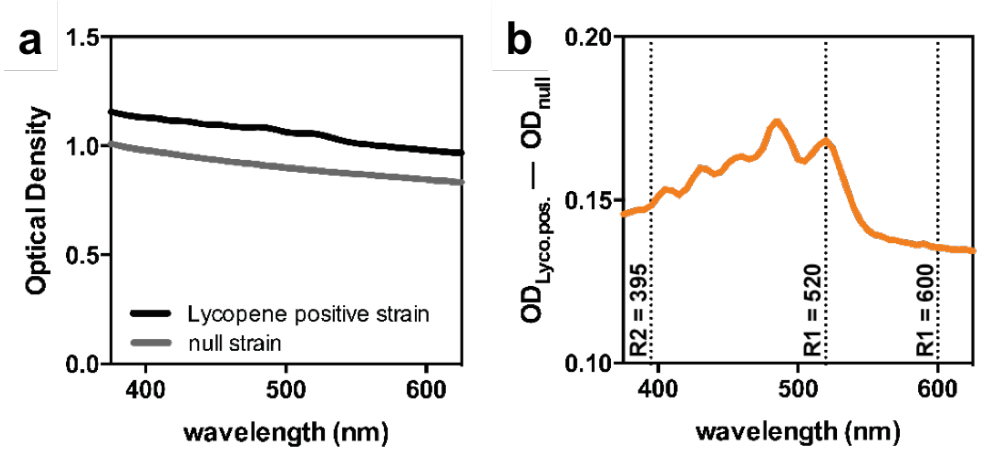


Figure 2.6 Optical density spectrum of lycopene in yeast cells

(a) Optical density spectrum of constitutive lycopene producing and lycopene null strains. **(b)** The spectrum of lycopene in yeast cells calculated from panel a. This spectrum allows selection of wavelengths for spectroscopic measurement of lycopene per cell.

To facilitate optimization of the lycopene signal, we developed a visible spectroscopy method for high throughput quantitative measurement of pigment content in cells. This method circumvents the low throughput chemical extraction traditionally used for measurement of lycopene yield [48]. Additionally, since the method is non-destructive we were able to perform time course analysis of the kinetics of lycopene production. This ability was integral to our goal, since unlike previous effort in metabolic engineering, induction speed rather than final yield was our primary target. Finally, due to its technical simplicity, we were able to robustly benchmark strains across different development years.

We adapted a method proposed by Myers to characterize pigmented cells through optical density measurements at multiple wavelengths [49]. We made several improvements in order to reduce the noise due to variations in cell growth phase, cell density and other sample irregularities. This enabled the precise evaluation of lycopene content in a high throughput micro titer plate format. First, we determined the spectrum of lycopene absorbance directly in cells in order to properly select sensitive and robust wavelengths for interrogation. Second, we measured the optical detector response curves at these wavelengths in order to make accurate lycopene measurements even at high cell densities. Third, we empirically determined a low-noise approximation for calculating cell scatter. Lastly, we determined a recording scheme to reduce well-to-well noise when

measuring a large number of samples in micro titer plates. Our method is described in detail in the Appendix.

2.2.4 An second generation lycopene reporter system

To optimize the speed of lycopene production, we used our spectroscopic method to characterize the effect of a variety of plasmid-borne enhancements to the carotenoid biosynthetic pathway (Figure 2.7). We determined that our first generation sensor (Lyco-1) generated a lycopene

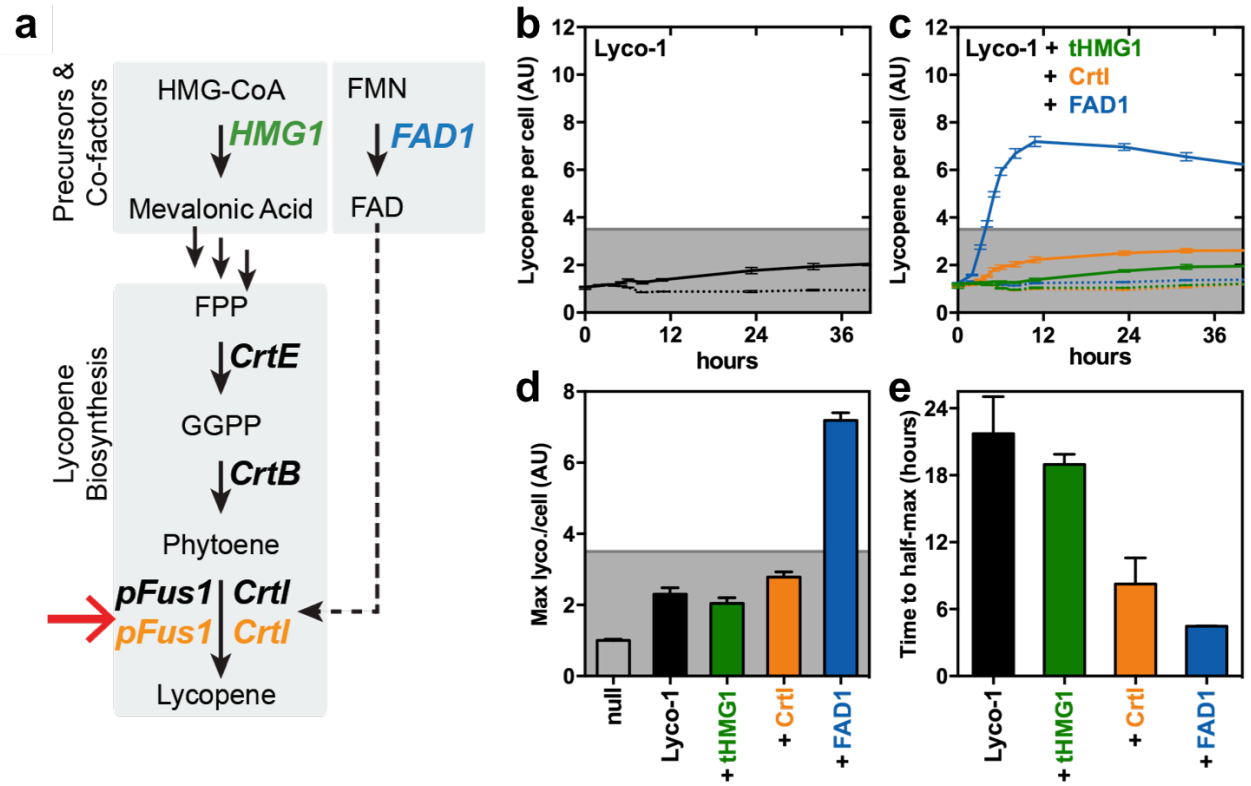
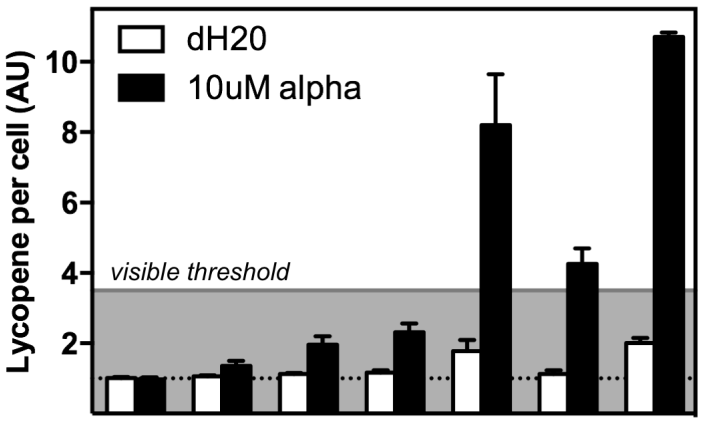


Figure 2.7 Optimization of lycopene readout with plasmid-borne enhancements
(a) Lycopene biosynthetic pathway. Lycopene production is induced (red arrow) by mating-signal dependent activation of the *FUS1* promoter. Biosynthetic enzymes shown in bold. Genes targeted for optimization shown in colors. HMG-CoA: 3-hydroxy-3-methylglutaryl-coenzyme A, FMN: flavin mononucleotide, FAD: flavin adenine dinucleotide, FPP: farnesyl pyrophosphate, GGPP: geranylgeranyl pyrophosphate. **(b,c)** Time course of lycopene strains induced with 10 μM of *S. cerevisiae* peptide (solid line) or water (dotted line). Grey line marks visible threshold of 3.5 LPC. Strains as in d. **(d)** Maximal lycopene per cell calculated from time course data in b and c. "Null" (grey) - strain with only CrtEB; "Lyco-1" (black) - parental strain with single copy CrtE, CrtB and CrtI; "+tHMG1" (green) - Lyco-1 with plasmid-borne truncated copy of HMG1; "+CrtI" (orange) - Lyco-1 with plasmid-borne copy of CrtI; "+Fad1" (blue) - Lyco-1 with plasmid-borne copy of Fad1. **(e)** The time to half-maximal lycopene yield was used to compare readout speed. Strains as in d.

signal of 2.3 LPC (lycopene per cell units). This was just over a two-fold increase in the apparent color signal compared to a lycopene null strain (1 LPC). By measuring a constitutive lycopene producing strain, we determined that an acceptable lycopene signal should be > 3.5 LPC.

We hypothesized that readout speed could be enhanced by either increasing the quantity of the phytoene substrate used by CrtI to make lycopene, by increasing the amount of CrtI produced upon pathway activation or by enhancing the catalytic activity of CrtI. Overexpression of the catalytic domain of the HMG-CoA reductase HMG1 has been shown to increase phytoene availability.[41] However, this led to no improvement in the kinetics of the lycopene readout (+tHMG1) and thus we reasoned that phytoene was not rate limiting. We then increased the CrtI amount with an additional copy of the *pFus1-CrtI* construct. This led to only a moderate increase in the maximum lycopene signal to 2.8 LPC (+CrtI). However, we observed a substantial reduction in the time to half-max from 22 hours down to 8.3 hours. Additionally, it has been shown that CrtI is an FAD-dependent enzyme, which relies on this co-factor for structural stability.[50] We thus sought to improve the catalytic activity of CrtI by increasing FAD content in the cell through the overexpression of the FAD synthetase FAD1 [51]. This led to a drastic increase of the max lycopene signal to 7.2 LPC as well as a further decrease in the time to half-max signal to 4.5 hours (+FAD1). Taken together this suggested that co-factor availability and therefore catalytically active CrtI was the primary rate-limiting component of the reaction.



CrtEB	1	1	1	1	1	1	1
pFUS1-CrtI	-	**1**	**2**	**3**	**2**	**2**	**2**
pTDH3-FAD1	-	-	-	-	**1**	-	**1***
pFUS1-FAD1	-	-	-	-	-	**1**	-

Figure 2.8 Optimization of lycopene output with genomic enhancements

Strains with varying lycopene output constructs were induced with either water (dH2O) or with alpha factor (10uM) for 6 hours. The copy number of each construct present in the genome of each strain is listed below the corresponding bars. A dash indicates the construct is not present. The asterisk indicates the construct is present at an alternated genomic locus.

We next characterized genome-integrated versions of the plasmid-born enhancements. We initially integrated one, two and three *pFus1-CrtI* (Lyco-1 = 1X *pFus1-CrtI*). As with the plasmid-borne variant the strain with two total *pFus1-CrtI* constructs also showed an improvement in lycopene yield relative to the single construct strain. However, a third copy of *pFus1-CrtI* did not significantly improve the lycopene yield further (Figure 2.8). We thus chose the 2X *pFus1-CrtI* strain to move forward.

Next we integrated an FAD1 overexpression construct (*pTDH3-FAD1*). We observed the drastic increase in performance as seen with the plasmid-born variant (Figure 2.8). However, we also observed an increase in the lycopene signal of the uninduced strain (dH2O). While this increase was well below the visible threshold of 3.5 LPC, we considered whether a pheromone inducible FAD1 construct might retain the lower uninduced level but still generate the desired enhancement upon induction. However, while integration of a *pFus1-FAD1* construct into the 2X *pFus1-CrtI* strain did have a lower basal level, it had a much weaker enhancement of the lycopene signal (Figure 2.8). Therefore we chose the *pTDH3-FAD1* overexpression construct to move forward.

Finally, we tested whether a different version of this FAD1 overexpression construct could provide a better enhancement. We observed that in fact integration of this construct at the HO locus

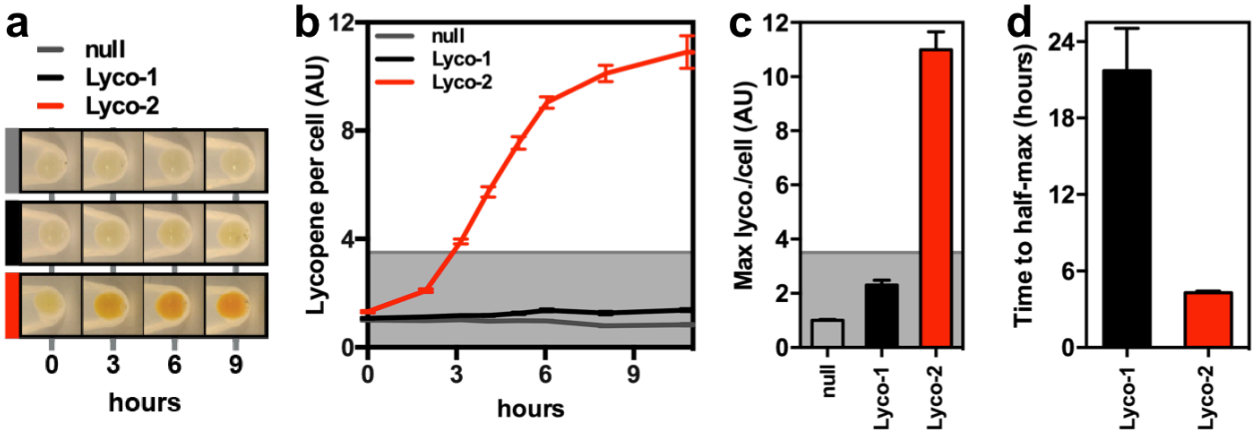


Figure 2.9 Comparison of first and second generation lycopene reporters
(a) Photos of cell pellets (5x10^7 cells) corresponding to strains in panel B. **(b)** Time course of lycopene production in reporters induced 10 μM alpha factor. "Null" (grey) - strain with only CrtEB; "Lyco-1" (black) – first generation reporter with CrtEB and single copy of *pFUS1-CrtI*; "Lyco-2" (red) – second generation reporter with CrtEB, two copies of *pFUS1-CrtI* and one copy of *pTDH3-FAD1*; Grey line marks the visible threshold of 3.5 LPC. **(c)** Maximal lycopene per cell calculated from time course data in panel b. **(d)** The time to half-maximal lycopene yield calculated from time course data in panel b.

(asterisk in Figure 2.8) led to a higher enhancement when compared to the same construct integrated at the FUS1 locus. This difference might be from the different genomic contexts or from the different terminators used in each case (tFUS1 v. tPGK1).

In summary, we made a second-generation reporter (Lyco-2) with a greatly enhanced lycopene signal (Figure 2.9). When characterized in micro titer plates conditions, this optimized lycopene readout yielded a maximum LPC of 11 relative to 2.3 for the original strain (Lyco-1). Additionally, the optimization led to a decrease in the time to half-max signal to 4.3 hours relative to 22 hours for the original strain. This enhanced reporter system met and exceeded our original design requirements and stood to meet them even in non-ideal conditions outside the laboratory.

2.2.5 Construction of a sensor for fungal pathogens

With an optimized lycopene reporter in hand, we chose to develop a second-generation sensor for invasive fungal pathogens. Fungal pathogens are a rising global public health burden particularly acute in low-income countries [52–54]. Fungal pathogens cause an estimated 2 million deaths annually and inflict devastating losses of plant crops and population decline in animal wildlife [52,54]. However, efforts to abate fungal infections prevalent in resource-limited areas are hampered by the dearth of fungal diagnostics [55]. We envisioned that the fungal GPCRs derived from these pathogens could be directly implemented as sensing elements for detection the peptide pheromones secreted by these human pathogens.

Fungal GPCRs were expected to couple robustly to the native pheromone response pathway due to high homology in regions known to mediate signal transduction between the endogenous receptor Ste2p and its G-protein Gpa1p (See Chapter 4). In fact, a handful of these receptors have been shown to rescue growth arrest, schmoo and mating phenotypes in *ste2* null mutants [56–59]. Additionally, we anticipated these receptors to be highly specific for their respective peptide ligands since they mediate species-specific mating reactions while preventing interspecies breeding [60]. Moreover, targeting peptides as the analyte held the potential to expand this set of GPCRs for the detection of unique peptide signature from any pathogen or disease [61,62] (See Chapter 3).

We chose to target *Paracoccidioides brasiliensis* (Pb) and *Candida albicans* (Ca) since they are among the top ten most significant invasive fungal pathogens worldwide, causing life-threatening

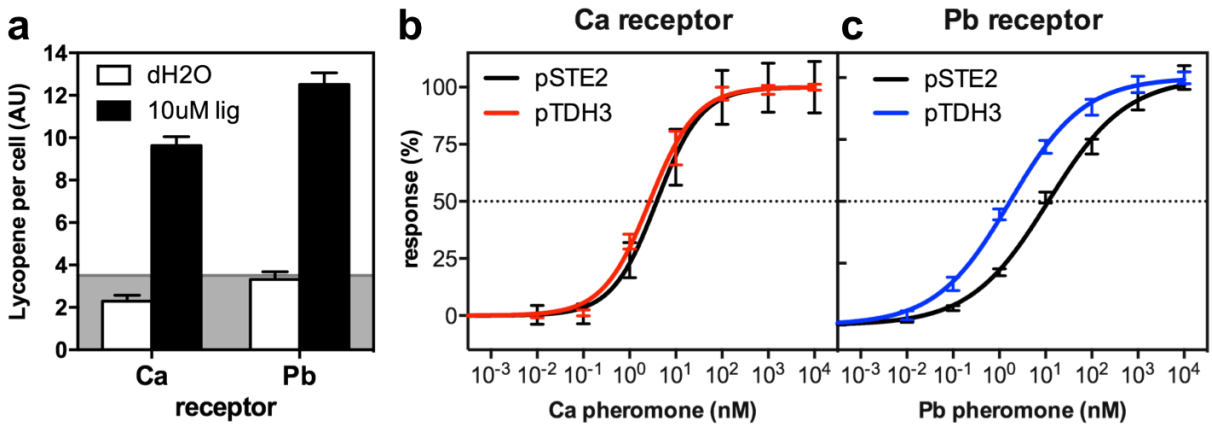


Figure 2.10 Optimization of fungal pathogen sensors

(a) Activation of sensors for *C. albicans* (Ca) and *P. brasiliensis* (Pb) with 10 μM of the cognate ligands. Equivalent to colors in panels b and c. **(b,c)** Dose response curve of Ca (left) and Pb (right) sensors expresing the receptors from different promoters. Lycopene measured at 7 hours.

infections in immunocompromised patients [53]. Furthermore, their GPCRs and respective peptide pheromones had been previously characterized in yeast [57,58]. Therefore, we integrated the Pb and Ca receptors into the STE2 locus of the optimized lycopene reporter. We observed that both of these sensor strains generated robust lycopene signal when induced with 10 μM of their respective fungal pathogen pheromones (Figure 2.10a). To further optimize these pathogen sensors we tested the effects on the lycopene signal when using a TDH3 promoter or the natural STE2 promoter to control the expression of the receptors. We found that strong overexpression from the TDH3 promoter led to a higher sensitivity for the Pb sensor and lower variability for the Ca sensor (Figure 2.10b,c). We therefore chose the *pTDH3-receptor* construct for the final optimized pathogen sensors.

Importantly, these optimized pathogen sensors showed sensitivities to the target pathogen pheromones in the low nanomolar range, 1.7 nM for the Pb pheromone and 2.7 nM for the Ca pheromone (Figure 2.10b,c). As measured in the laboratory, these sensitivities met our design requirements and had the potential to be robust in non-ideal conditions.

2.2.6 Characterization of sensor in non-ideal conditions

To validate the potential of our pathogens sensors to function outside the laboratory, we characterized their sensitivity and signal levels across a variety of media, pH and temperature. Furthermore, we validated sensor functionality in complex clinical samples such as urine and serum.

Functionality across this diverse range of conditions was a critical design goal. Since our sensors are based on yeast, a hardy microbe that has naturally evolved to cope with a range of environmental conditions, we expected that sensor functionality would be similarly resilient.

First, we tested if the lycopene readout required specialized media to function. While our sensors do not require expensive reagents to generate a readout, they will require nutrients to maintain the viability of the yeast cells during the assay. One concern was that this function was dependent on defined laboratory media that is composed from more expensive highly purified components [63,64]. We therefore determined which if any of these components were required for function. As expected, there was an absolute requirement for a carbon source with a steady increase in signal level with increased glucose concentration (Figure 2.11). However, as little as 27.5 mM glucose produced significant signal level. As might be expected from an organism prototrophic for amino acids, we found that purified amino acid supplements were not required for function and in fact growth without amino acids boosted the lycopene signal (Figure 2.11). In addition, we observed that while a nitrogen source (YNB = yeast nitrogen base) was also necessary, increases over the standard quantity were not beneficial (Figure 2.11). Finally, use of a cheaper undefined complex media (YPD = yeast extract-peptone-dextrose) actually resulted in the highest lycopene signal (Figure 2.11). These results validate that our sensors do not depend on any of the specialized components of laboratory media. This will enable our sensors to be produced through the use of standard industrial yeast feed stocks such as enriched black molasses, a cheap waste product of

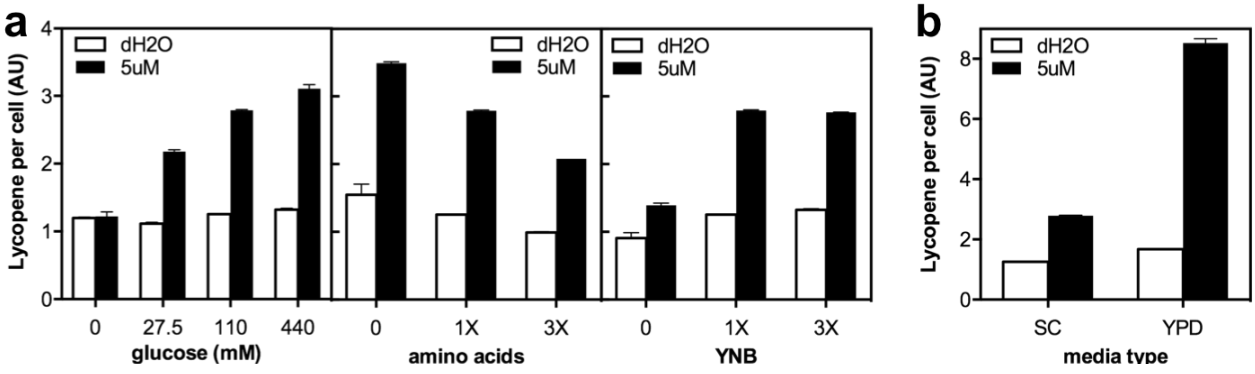


Figure 2.11 Effects of nutrient composition on lycopene readout

(a) The lycopene reporter was induced in synthetic media with components held constant while independently varying the amount of glucose (left), amino acids (middle) or yeast nitrogen base (right). Lycopene measured at 6 hours. **(b)** Lycopene reporter induced in synthetic media (SC) or complex media (YPD).

sugar refining.

Second, we characterized the sensitivity and signal strength of our sensors when exposed to a range of pH and temperatures (Figure 2.12). We observed that our sensors were functional from a pH range between 4.3 and 9, as well as a temperature range between 20°C and 40°C. However, we did observe that lycopene signal levels generally decreases at 20°C and 40°C as compared to the ideal yeast growth temperature of 30°C (Figure 2.12c,d). Interestingly, while the Ca sensor was nearly unaffected across the range of pH, the Pb sensor had worse sensitivity at both low and high pH (Figure 2.12a,b). This effect may be cause by the protonation state on key residues of the ligand-binding pocket or the ligand itself.

Third, we validated the functionality of the pathogen sensors in human serum and urine. These two body fluids represent a range of complex sample types that might be encountered in field use of the sensors. In each case the sample was diluted by 50% with media to provide the required carbon and nitrogen source. Surprisingly, we observed an *increase* in sensitivity in both urine and

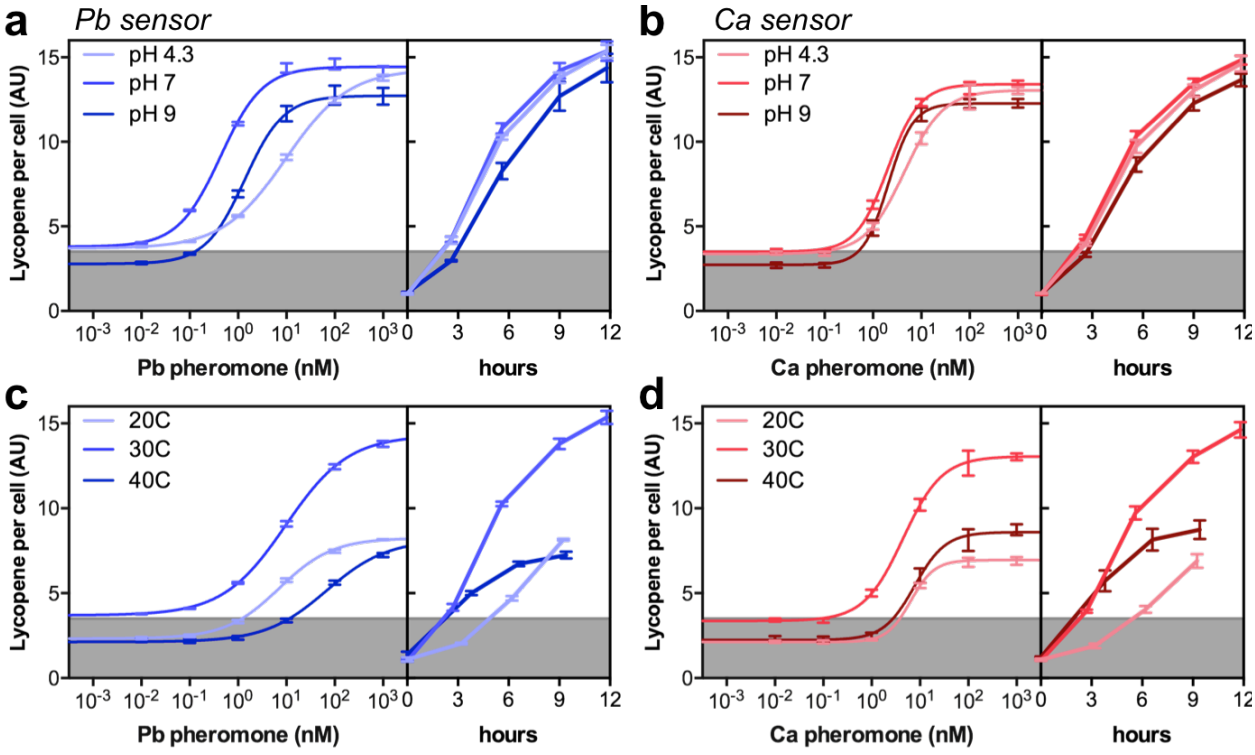


Figure 2.12 Fungal pathogen sensors function over a range of temperatures and pH
Sensitivity and speed of response for *P. brasiliensis* sensor (blues, Pb) and *C. albicans* sensor (reds, Ca) in different conditions: **(a,b)** pH and **(c,d)** temperature. Dose responses at 9 hours and time courses at 1 μM.

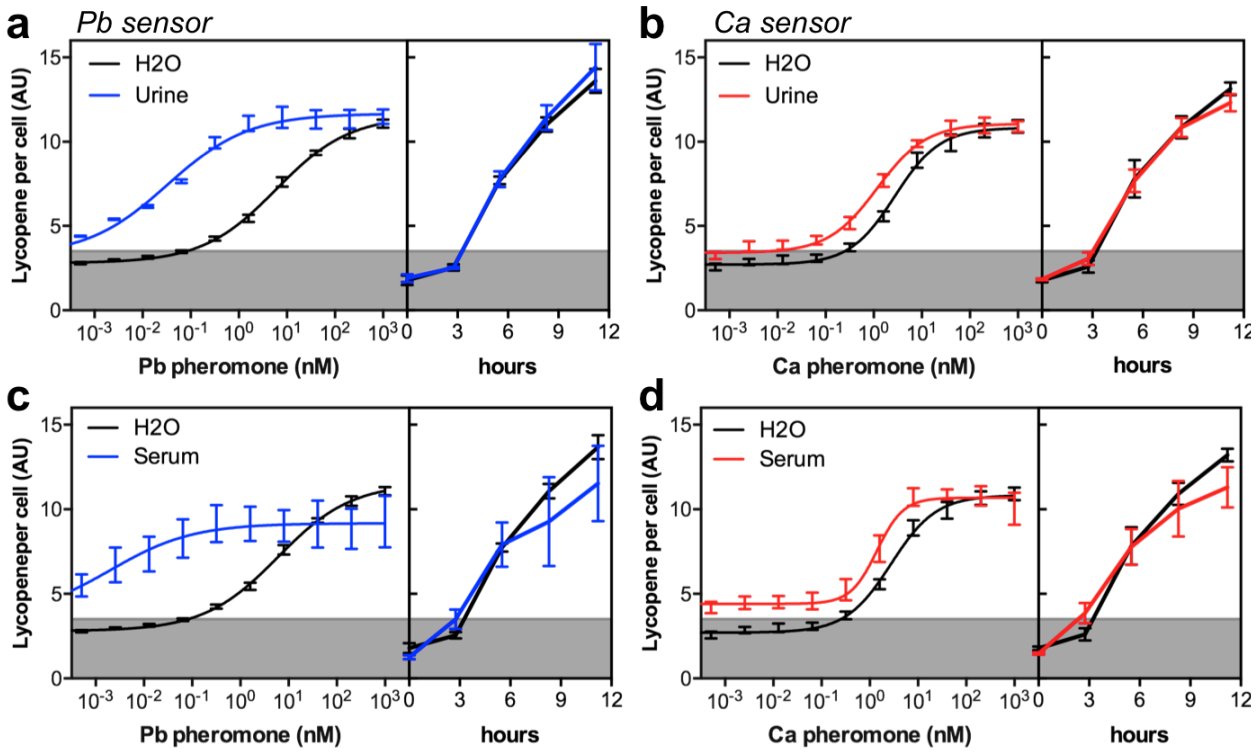


Figure 2.13 Fungal pathogen sensors function in human urine and serum

Sensitivity and speed of response for *P. brasiliensis* sensor (blues, Pb) and *C. albicans* sensor (reds, Ca) in different conditions: **(a,b)** 50% human urine and **(c,d)** 50% human serum. Dose responses at 9 hours and time courses at 1 μM

serum. This effect was magnified for the Pb sensor, reaching an EC_{50} of 20-60 pM (95%CI) in urine and 20-0.2 pM (95%CI) in serum. As observed with the pH sensitivity, the Ca sensor was more resilient to effect from urine and serum resulting in a smaller improvement in the EC_{50} values. In terms of lycopene signal level, the sensors were unaffected by urine and showed a small decrease in serum after 6 hours.

These results strongly support our hypothesis that our sensor platform is robust enough to function in non-ideal conditions. This quality is gained by embedding our sensor components within a living cell that can leverage a host of adaptive responses to protect and maintain the working environment of the sensor components.

2.2.7 Detection of fungal pathogens

We next challenged our sensor platform by testing it directly against clinically-isolated pathogen strains. Specifically, we focused on validating the Pb sensor against the target pathogen

Paracoccidioides brasiliensis [4]. In addition, we also tested the Pb sensor against strains of *Histoplasma capsulatum* . This second pathogenic fungus is also among the top ten leading causes of invasive fungal infections in humans [53]. Moreover, it is predicted to secrete an identical pheromone to *Paracoccidioides brasiliensis*, and so we expected to be able to also detect *Histoplasma capsulatum* with the Pb sensor [65]. Therefore, we aimed to validate whether the sensors had (1) a high enough sensitivity to successfully detect natural concentrations of these pheromones and (2) a high enough specificity to prevent false positive signals that could be caused by other fungal components secreted by the pathogens.

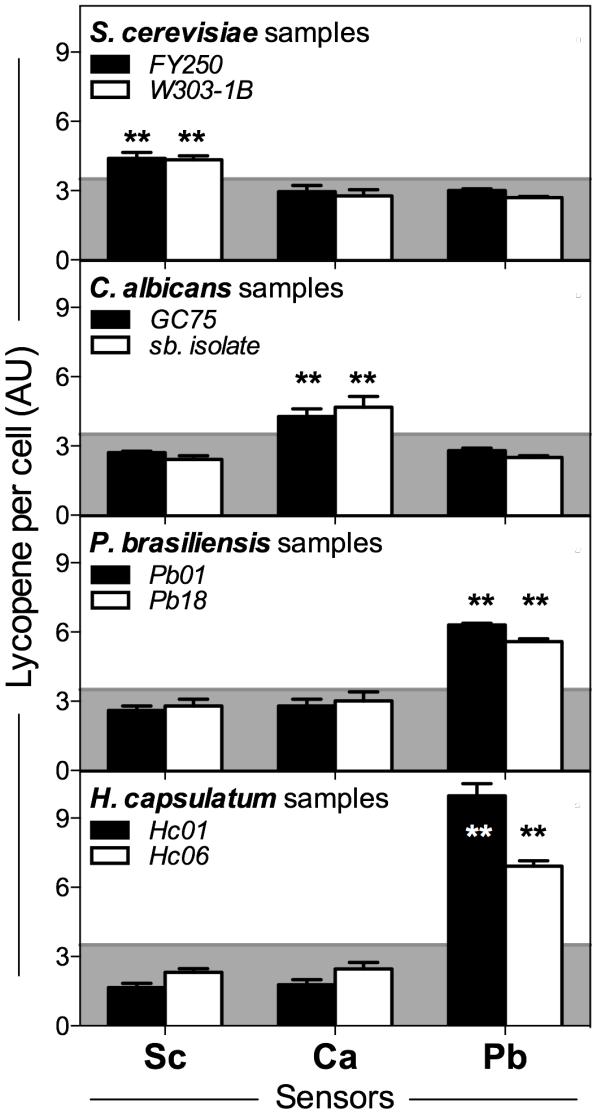


Figure 2.14 Detection of fungal pathogens
Samples were collected from three human fungal pathogens species (two representative strains each): *C. albicans*, *P. brasiliensis*, *H.capsulatum* and as a control *S. cerevisiae*. Pathogen samples were assayed with three yeast sensors (Sc, Ca and Pb) and lycopene signals were measured at 9 hours. ** $P \leq 0.001$

We acquired samples from four clinically-isolated strains, two *Paracoccidioides brasiliensis.* and two *Histoplasma capsulatum*. All the samples were of supernatants taken from the spent media remaining after growth of the pathogen strains. Importantly, since fungal species are often composed of a complex of distinct subspecies, we chose a diverse set of strains to capture potential differences between these different subtypes. For *Paracoccidioides brasiliensis*, we used samples from Pb18, a strain representing the major group S1, and from Pb01, a strain representing the *Paracoccidioides lutzii* subdivision of the species complex [66]. For *Histoplasma capsulatum*, we used samples from Hc06, representing the most common

subtype (North America class 2) and Hc01, representing a more rare subtype (North America class 1) [67].

We observed that indeed our sensors successfully detected the presence of all the pathogenic strains (Figure 2.14). Importantly, we used both the Ca and Sc sensors as well as control samples derived from laboratory strains of *S. cerevisiae* and *C. albicans* as a way to directly test the specificity of the assay. In no case did we observe a false positive signal from the off target sensors or the off target samples (Figure 2.14). Furthermore, a majority of the samples generated a lycopene signal well above the visible threshold (>3.5 LPC) by 9 hours. By comparing to the lycopene signals to those from the synthetic ligand, the pathogen samples likely contained the target pheromones in the range of 1-100 nM, with the *Pracoccidioides brasiliensis* isolates producing a signal that corresponded with the lower end of this range and the *Histoplasma capsulatum* isolates producing a signal that corresponded with at least 10 nM. These results support the potential of our Pb sensor for detection of the human pathogens *Paracoccidioides brasiliensis* and *Histoplasma capsulatum* and suggest the need for further investigation into the nature pheromone secretion by these pathogens.

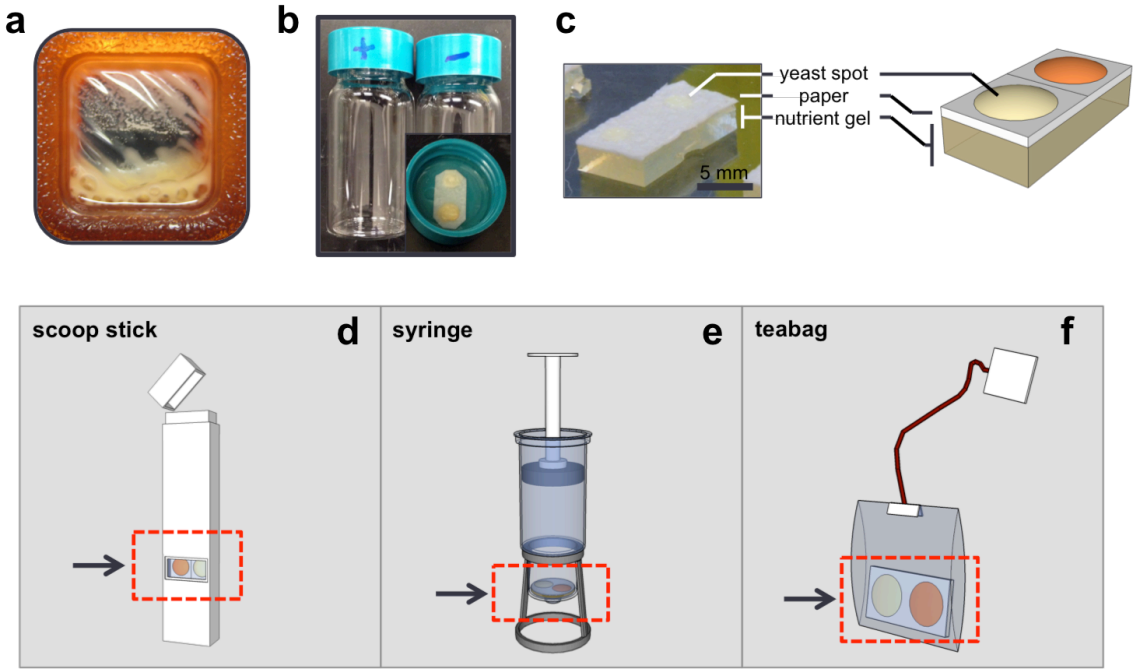


Figure 2.15 Initial designs for a point-of-care yeast sensor
A point-of-care assay could have yeast sensor cells confined in **(a)** pouches, **(b)** small containers or as a **(c)** "Yeast Block" made of a block of nuterient gel. **(d,e,f)** Possible modes of use for a "Yeast Block" design.

2.2.8 Development of a point-of-care dipstick assay

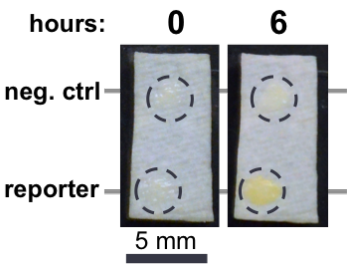


Figure 2.16 Function of first sensor prototype
Photographs of "Yeast Block" function after induction for 6 hours with 10 nmol ligand.

Having validated the robustness and functionality of our sensor, we next turned to developing an easy-to-use dipstick assay. We aimed to translate the core benefits of the yeast-based biosensors (cheap, sensitive, specific, robust) to a format that was usable by non-specialists. The key goals were to (1) eliminate the need for micro titer plates and pipets, (2) incorporate controls to generate a reliable assay, and (3) develop a format amenable to cheap mass production. A secondary goal was to show that a "one-step" assay was possible.

The initial challenge was to design a format that standardized the effective yeast cell concentration between each test and over time. While the lycopene color was clearly visible in our earlier characterization in micro titer plates, the apparent intensity of this colored depended strongly on the concentration of yeast cells. Well-suspended cultures of yeast had a lower apparent intensity, while pelleted cells had a high intensity color. Therefore, initial designs centered on a way of confining the yeast in a container that could sit undisturbed allowing the yeast to naturally settle and develop a strong apparent color (Figure 2.15). However, pouch-based design depended on potentially expensive dialysis membranes and bottle-based designs were expected to be easily disturbed. Inspired by our "spot blot" tests of early reporter strains, we developed an improved product profile that centered on a "Yeast Block" (Figure 2.15c). This design involved spotting the sensor strains on filter paper and subsequently placing them on a block of nutrient gel. We envisioned that the "Yeast Block" could then be incorporated into a range of final product casings depending on the specific requirements of the assay. Initial tests showed that this approach did in fact lead to a consistent generation of lycopene signal with high apparent color intensity (Figure 2.16). While promising, we found that producing these "Yeast Blocks" was labor intensive and the nutrient gel was prone to drying and was not easily rehydrated when placed in a sample.

We improved on our initial design by removing the nutrient gel and developing a stencil-based approach for making a high number of yeast spots. The design involved placing two or more of these "Sensor Spots" onto a strip of paper towel to act as wicking paper (Figure 2.17a).

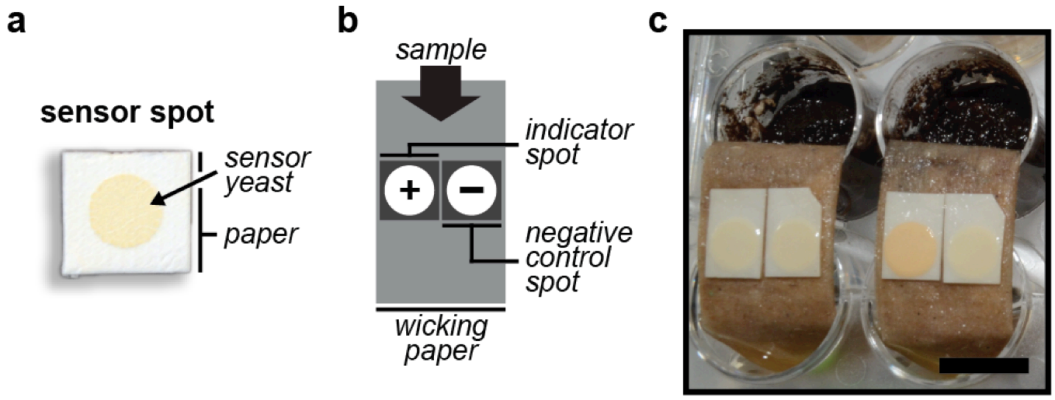


Figure 2.17 A paper-based dipstick device for point-of-care use of yeast sensors
(a) Engineered yeast sensor cells spotted on paper are the only active component required for the dipstick assay. Spot diameter - 5 mm. **(b)** Dipstick assay includes two spots, indicator sensor strain and control strain, placed on top of a strip of paper towel that acts as wicking paper. The indicator sensor spot detects the target ligand and the negative control spot contains a strain with an off-target receptor. This design enables easy visual interpretation of the results as well as quantification by calculating the difference in the pixel color values between the two spots. **(c)** Photograph of the dipstick for detection of the fungal pathogen *P. brasiliensis* in soil. Left - no mating peptide in soil. Right - mating peptide added to soil. Scale bar - 1 cm.

To run the assay the wicking paper was dipped into a sample pre-mixed with media to deliver both the nutrients and potential pheromones to the "Sensor Spots". This two-step assay largely met all three of our goals. There was no need for pipets or micro titer plates, only a small container was needed to mix the sample with media. We could incorporate controls by making "Sensor Spots" from different strains, one responsive to the target pheromone and the other responsive to an off-target pheromone. Finally, the components had the potential for cheap mass production since the stencil-based filter method could be scaled easily and both the "Sensor Spot" and the media could be dried using standard fluidized bed driers currently used in large-scale yeast production [68].

With this dipstick assay in hand, we next characterized the specificity, sensitivity and speed of the sensors in this assay context. In addition to visual inspection, we quantified lycopene accumulation with time-lapse photography and pixel color analysis (see Methods). Characterization in plain media showed that the assay functioned as expected leading to a strong color change apparent with the naked eye. Importantly, we found that the molecular specificity of the sensors was maintained and we did not observe any false negative signals from the off-target controls (Figure 2.19). We also measured the EC_{50} of the dipstick set up and observed that there was only a slight decrease in sensitivity for the Pb sensor from 1.7 nM to 14 nM (Figure 2.18). However, we observed a

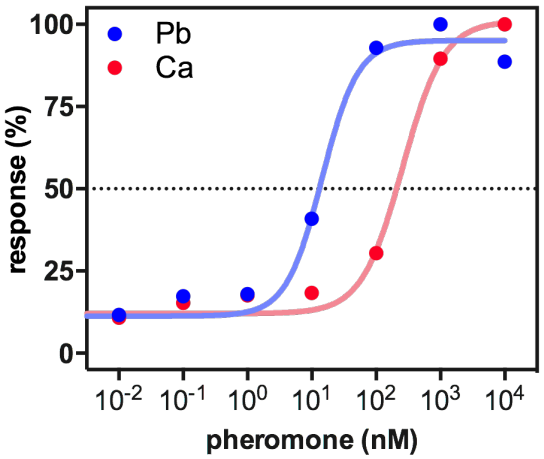


Figure 2.18 Dose response of paper-based dipstick

Paper dipsticks were made from the Pb and Ca sensor strains and each set was treated to a range of concentrations of their corresponding cognate pheropmone in rich media. Lycopene signal was measured at 12 hours and quantified by pixel color analysis (Methods).

substantial decrease in sensitivity for the Ca sensor from 2.7 nM to 260 nM. We expect that this may be due to adverse interactions between the target pheromone and on of the non-yeast components of the dipstick. As the target pheromone of the Ca sensor is predicted to be more hydrophobic, it is possible that the decrease in sensitivity is caused by adsorption of the target pheromone to the wicking or filter paper leading to a lower effective concentration around the yeast sensor cells.

In terms of the speed of signal development, we observed a decrease in the time to half-max signal from 4.3 hours in micro titer plates to 10 hours in the dipstick assay (Figure 2.19a,c). This decrease is likely due to the decrease in access to nutrients and oxygen in the crowded environment of the "Sensor Spot" relative to the well-mixed conditions in micro titer plates. Importantly, even with this decrease in speed the optimized lycopene system generated a strong visible signal in less than 12 hours.

We next validated whether the dipstick assay would be compatible with detection in complex samples relevant for clinical and environmental detection. We observed strong signals in 100% soil, 50% human urine and 50% human serum (Figure 2.19b,d). As with the previous characterization we observed a slight attenuation of the signal in serum samples. We also tested the assay with whole human blood, however because of the strong red color of whole blood high concentrations were not viable as a sample. However, we successfully observed a lycopene signal in 2% blood (Figure 2.19b,d). These results show that the robustness of our sensors is maintained in the dipstick format. Importantly, while soil samples are normally too heterogeneous for measurement in micro titer plates, the use of our dipstick format enabled direct assay of important environmental sample type.

Finally, to further increase the ease-of-use we then designed and built a simple plastic holder (Figure 2.20). This holder made the assay portable as it maintained all the components in proper register. We also found that by leaving only a small opening for the "Sensor Spots" they did not dry out as easily once the assay was started. We found that this plastic holder had no significant effect on the function of the sensors and generated a comparably visible signal (Figure 2.20c). While not necessary for assay function, this simple plastic holder is not expected to add substantial cost per

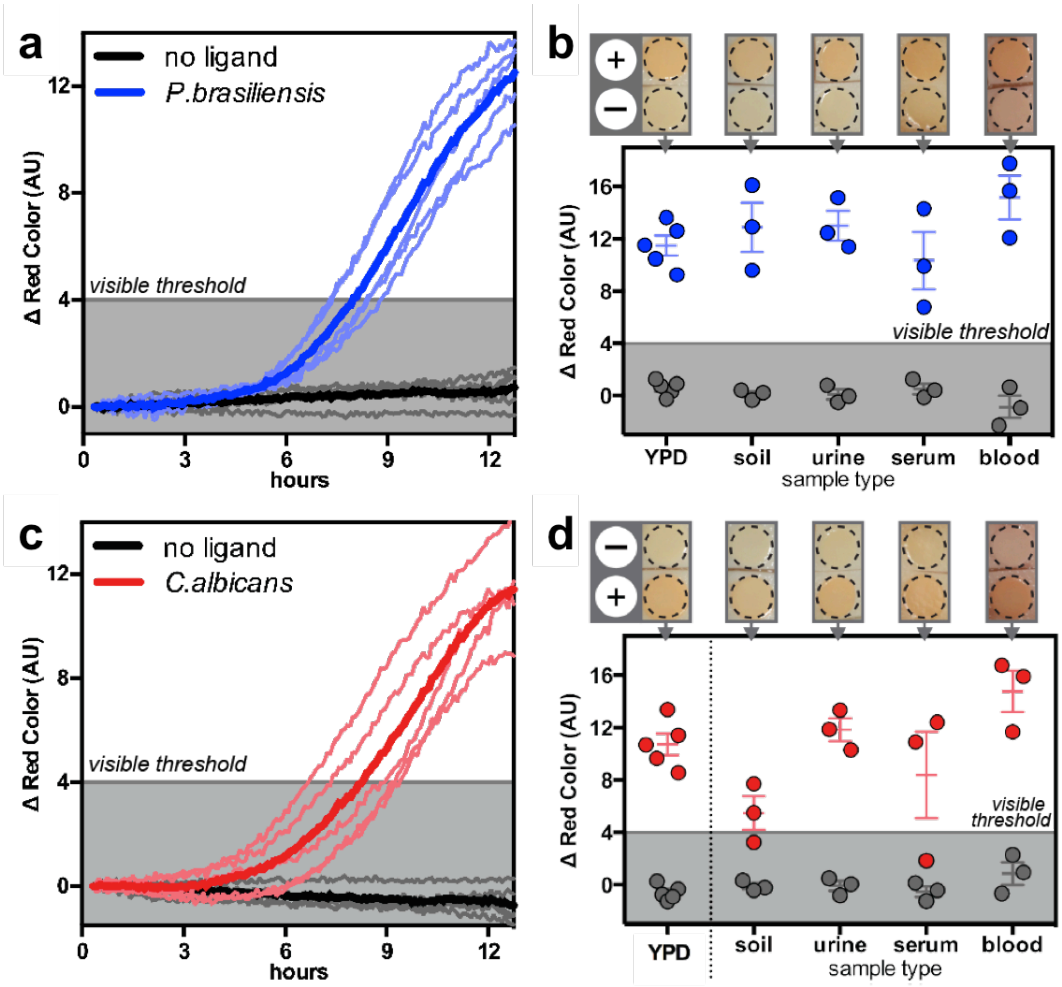


Figure 2.19 The paper-based dipstick functions in complex samples
Quantitative analysis of lycopene production using dipstick assay, as scored by time-lapse photography for detection of **(a,b)** *P. brasiliensis* mating pheromone by the Pb dipstick and **(c,d)** *C. albicans* mating pheromone by the Ca dipstick. **(a, c)** Time course of induction with pheromones in media. Individual runs shown in light color, average response shown in dark color. Shading indicates visible threshold. **(b,d)** Detection of mating pheromones in complex samples. Experiments were performed using 1 μM pheromone (color dots) or water (grey dots) as a negative control. YPD - media only; Soil - standard potting soil; Urine - 50% pooled human urine; Serum - 50% human serum; Blood - 2% whole blood. All samples were supplemented with media.

test. Important, it enables a true "one-step" assay rendering our yeast sensor platform extremely easy to use by a non-specialist.

2.3 Discussion

In this chapter we demonstrated the specificity and sensitivity of our yeast-based sensors in the detection of fungal pathogens in complex samples. This approach integrates all bioassay functions, eliminating the need for sophisticated equipment or materials for its use. In addition, the production of this yeast biosensor is as scalable and cost-effective as its use.

While mammalian cells [20], and whole-cell extracts [69] have been used to detect pathogens, these systems require costly and complex production techniques. In contrast yeast is already produced in massive quantities as a household product. Since our yeast platform is a one-component device, it can directly leverage existing yeast production infrastructure. Using current production costs of packaged yeast, this indicates an expected initial cost per test of 1¢ US. Furthermore, yeast production is so well established that our biosensor could be produced in resource poor settings directly by local communities.

Our results show that engineered yeast cells are uniquely poised to provide a practical and cost-effective platform for detecting pathogens outside the laboratory. As with any system based on

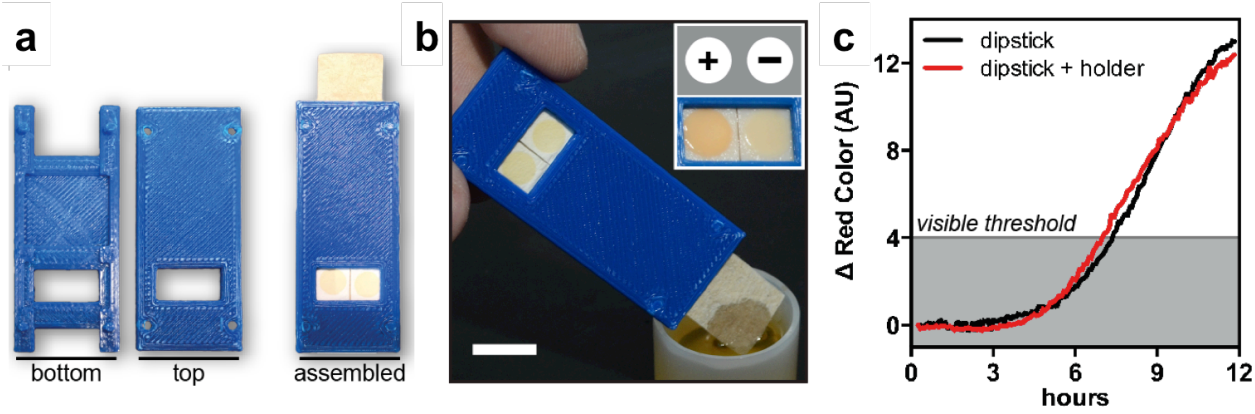


Figure 2.20 A plastic holder for improving dipstick portability
(a) A simple plastic holder was designed to enable portable use of the dipstick assay. Thin black bars - 2 cm. **(b)** Dipstick device in use. Inset: positive readout, "+" biosensor strain. "-" negative control strain **(c)** Dipstick holder does not affect biosensor performance as shown by time course measurement of the *P. brasiliensis* dipstick test response using 1 μM cognate peptide

engineered organisms, successful transfer of this technology to the field will require incorporation of established genetic safety mechanism [70]. Furthermore, the peptide-sensitive GPCRs described here lay the groundwork for generating sensors to other pathogen-specific peptides by directed evolution (see Chapter 3). Overall, by combining yeast, lycopene and GPCRs we developed an extremely cost-effective approach to pathogen detection and surveillance.

2.4 References

1. Caliendo, A. M. *et al.* Better Tests, Better Care: Improved Diagnostics for Infectious Diseases. *Clin. Infect. Dis.* **57**, S139–S170 (2013).

2. Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K. & Adley, C. An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnol. Adv.* **28**, 232–254 (2010).

3. Lim, D. V., Simpson, J. M., Kearns, E. A. & Kramer, M. F. Current and Developing Technologies for Monitoring Agents of Bioterrorism and Biowarfare. *Clin. Microbiol. Rev.* **18**, 583–607 (2005).

4. Mabey, D., Peeling, R. W., Ustianowski, A. & Perkins, M. D. Tropical infectious diseases: Diagnostics for the developing world. *Nat. Rev. Microbiol.* **2**, 231–240 (2004).

5. Giljohann, D. A. & Mirkin, C. A. Drivers of biodiagnostic development. *Nature* **462**, 461–464 (2009).

6. Singh, R. *et al.* Biosensors for pathogen detection: A smart approach towards clinical diagnosis. *Sens. Actuators B Chem.* **197**, 385–404 (2014).

7. Bissonnette, L. & Bergeron, M. G. Diagnosing infections—current and anticipated technologies for point-of-care diagnostics and home-based testing. *Clin. Microbiol. Infect.* **16**, 1044–1053 (2010).

8. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic Biology Moving into the Clinic. *Science* **333**, 1248–1252 (2011).

9. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* **15**, 289–294 (2014).

10. Lequin, R. M. Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clin. Chem.* **51**, 2415–2418 (2005).

11. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).

12. Ramanathan, S., Ensor, M. & Daunert, S. Bacterial biosensors for monitoring toxic metals. *Trends Biotechnol.* **15**, 500–506 (1997).

13. Courbet, A., Endy, D., Renard, E., Molina, F. & Bonnet, J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.* **7**, 289ra83-289ra83 (2015).

14. Michelini, E., Leskinen, P., Virta, M., Karp, M. & Roda, A. A new recombinant cell-based bioluminescent assay for sensitive androgen-like compound detection. *Biosens. Bioelectron.* **20**, 2261–2267 (2005).

15. Danino, T. *et al.* Programmable probiotics for detection of cancer in urine. *Sci. Transl. Med.* **7**, 289ra84-289ra84 (2015).

16. Mukherjee, K., Bhattacharyya, S. & Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth. Biol.* **4**, 1261–1269 (2015).

17. Gupta, S., Bram, E. E. & Weiss, R. Genetically Programmable Pathogen Sense and Destroy. *ACS Synth. Biol.* **2**, 715–723 (2013).

18. Kotula, J. W. *et al.* Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci.* **111**, 4838–4843 (2014).

19. Radhika, V. *et al.* Chemical sensing of DNT by engineered olfactory yeast strain. *Nat. Chem. Biol.* **3**, 325–330 (2007).

20. Rider, T. H. *et al.* A B Cell-Based Sensor for Rapid Identification of Pathogens. *Science* **301**, 213–215 (2003).

21. Walmsley, R. M. & Keenan, P. The eukaryote alternative: Advantages of using yeasts in place of bacteria in microbial biosensor development. *Biotechnol. Bioprocess Eng.* **5**, 387–394 (2000).

22. Baronian, K. H. R. The use of yeast and moulds as sensing elements in biosensors. *Biosens. Bioelectron.* **19**, 953–962 (2004).

23. Berny, J.-F. & Hennebert, G. L. Viability and Stability of Yeast Cells and Filamentous Fungus Spores during Freeze-Drying: Effects of Protectants and Cooling Rates. *Mycologia* **83**, 805–815 (1991).

24. Parry, J. M. Use of tests in yeasts and fungi in the detection and evaluation of carcinogens. *IARC Sci. Publ.* 471–485 (1999).

25. Jarque, S., Bittner, M., Blaha, L. & Hilscherova, K. Yeast Biosensors for Detection of Environmental Pollutants: Current State and Limitations. *Trends Biotechnol.* **34**, 408–419 (2016).

26. Evans, R. M. & Mangelsdorf, D. J. Nuclear Receptors, RXR, and the Big Bang. *Cell* **157**, 255–266 (2014).

27. Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).

28. Pausch, M. H. G-protein-coupled receptors in Saccharomyces cerevisiae: high-throughput screening assays for drug discovery. *Trends Biotechnol* **15**, 487–94 (1997).

29. O'Malley, M. A. *et al.* Progress toward heterologous expression of active G-protein-coupled receptors in Saccharomyces cerevisiae: Linking cellular stress response with translocation and trafficking. *Protein Sci. Publ. Protein Soc.* **18**, 2356–2370 (2009).

30. Dowell, S. J. & Brown, A. J.in *G Protein-Coupled Recept. Drug Discov.* (Leifert, W. R.) 213–229 (Humana Press, 2009).at <http://link.springer.com/protocol/10.1007/978-1-60327-317-6_15>

31. Erlenbach, I. *et al*. Functional expression of M1, M3 and M5 muscarinic acetylcholine receptors in yeast. *J. Neurochem.* **77**, 1327–1337 (2001).

32. Minic, J. *et al*. Functional expression of olfactory receptors in yeast and development of a bioassay for odorant screening. *FEBS J.* **272**, 524–537 (2005).

33. Naylor, L. H. Reporter gene technology: the future looks bright. *Biochem. Pharmacol.* **58**, 749–757 (1999).

34. Yang, S. & Rothman, R. E. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.* **4**, 337–348 (2004).

35. Ley, B. *et al*. Evaluation of a rapid dipstick (Crystal VC) for the diagnosis of cholera in Zanzibar and a comparison with previous studies. *PloS One* **7**, e36930 (2012).

36. Njanpop-Lafourcade, B. personal communication. (2013).at <Africhol>

37. Dick, M. H., Guillerm, M., Moussy, F. & Chaignat, C.-L. Review of two decades of cholera diagnostics--how far have we really come? *PLoS Negl. Trop. Dis.* **6**, e1845 (2012).

38. Bhopal, A., Callender, T., Knox, A. F. & Regmi, S. Strength in numbers? Grouping, fund allocation and coordination amongst the neglected tropical diseases. *J. Glob. Health* **3**, (2013).

39. Armstrong, G. A. Genetics of eubacterial carotenoid biosynthesis: a colorful tale. *Annu. Rev. Microbiol.* **51**, 629–659 (1997).

40. Chemler, J. A., Yan, Y. & Koffas, M. A. Biosynthesis of isoprenoids, polyunsaturated fatty acids and flavonoids in Saccharomyces cerevisiae. *Microb. Cell Factories* **5**, 20 (2006).

41. Verwaal, R. *et al*. High-Level Production of Beta-Carotene in Saccharomyces cerevisiae by Successive Transformation with Carotenogenic Genes from Xanthophyllomyces dendrorhous. *Appl. Environ. Microbiol.* **73**, 4342–4350 (2007).

42. Hackett, E. A., Esch, R. K., Maleri, S. & Errede, B. A family of destabilized cyan fluorescent proteins as transcriptional reporters in S. cerevisiae. *Yeast Chichester Engl.* **23**, 333–349 (2006).

43. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26**, 339–350 (2005).

44. Versele, M., Lemaire, K. & Thevelein, J. M. Sex and sugar in yeast: two distinct GPCR systems. *EMBO Rep.* **2**, 574–579 (2001).

45. Storici, F. & Resnick, M. A.in *DNA Repair Part B* **Volume 409**, 329–345 (Academic Press, 2006).

46. Hoffman, G. A., Garrison, T. R. & Dohlman, H. G.in *Methods Enzymol.* (Hildebrandt, R. I. and J. D.) **344**, 617–631 (Academic Press, 2002).

47. Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U A* **108**, 15135–15140 (2011).

48. Xie, W., Lv, X., Ye, L., Zhou, P. & Yu, H. Construction of lycopene-overproducing Saccharomyces cerevisiae by combining directed evolution and metabolic engineering. *Metab. Eng.* **30**, 69–78 (2015).

49. Myers, J. A., Curtis, B. S. & Curtis, W. R. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys.* **6**, 4 (2013).

50. Schaub, P. *et al.* On the Structure and Function of the Phytoene Desaturase CRTI from Pantoea ananatis, a Membrane-Peripheral and FAD-Dependent Oxidase/Isomerase. *PLoS ONE* **7**, e39550 (2012).

51. Wu, M., Repetto, B., Glerum, D. M. & Tzagoloff, A. Cloning and characterization of FAD1, the structural gene for flavin adenine dinucleotide synthetase of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **15**, 264–271 (1995).

52. Denning, D. W. & Bromley, M. J. How to bolster the antifungal pipeline. *Science* **347**, 1414–1416 (2015).

53. Brown, G. D. *et al.* Hidden Killers: Human Fungal Infections. *Sci. Transl. Med.* **4**, 165rv13-165rv13 (2012).

54. Fisher, M. C. *et al.* Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194 (2012).

55. Arvanitis, M., Anagnostou, T., Fuchs, B. B., Caliendo, A. M. & Mylonakis, E. Molecular and Nonmolecular Diagnostic Methods for Invasive Fungal Infections. *Clin. Microbiol. Rev.* **27**, 490–526 (2014).

56. Mayrhofer, S. & Pöggeler, S. Functional Characterization of an α-Factor-Like Sordaria macrospora Peptide Pheromone and Analysis of Its Interaction with Its Cognate Receptor in Saccharomyces cerevisiae. *Eukaryot. Cell* **4**, 661–672 (2005).

57. Gomes-Rezende, J. A. *et al.* Functionality of the Paracoccidioides Mating α-Pheromone-Receptor System. *PLoS ONE* **7**, e47033 (2012).

58. Janiak, A. M. *et al.* Functional expression of the Candida albicans α-factor receptor in Saccharomyces cerevisiae. *Fungal Genet. Biol.* **42**, 328–338 (2005).

59. Wendland, J., Dünkler, A. & Walther, A. Characterization of α-factor pheromone and pheromone receptor genes of Ashbya gossypii. *FEMS Yeast Res.* **11**, 418–429 (2011).

60. Martin, S. H., Steenkamp, E. T., Wingfield, M. J. & Wingfield, B. D. Mate-recognition and species boundaries in the ascomycetes. *Fungal Divers.* **58**, 1–12 (2013).

61. Xiang, Y., Kurokawa, M., Kanke, M., Takakuwa, Y. & Kato, T.in *Peptidomics* (Soloviev, M.) 259–271 (Humana Press, 2010).doi:10.1007/978-1-60761-535-4_20

62. Cordwell, S. J., Nouwens, A. S. & Walsh, B. J. Comparative proteomics of bacterial pathogens. *PROTEOMICS* **1**, 461–472 (2001).

63. Complete Minimal (CM) or Synthetic Complete (SC) and Drop-out Media. *Cold Spring Harb. Protoc.* **2006**, pdb.rec8190 (2006).

64. Dropout mix. *Cold Spring Harb. Protoc.* **2006**, pdb.rec8585 (2006).

65. Bubnick, M. & Smulian, A. G. The MAT1 Locus of Histoplasma capsulatum Is Responsive in a Mating Type-Specific Manner. *Eukaryot. Cell* **6**, 616–621 (2007).

66. Torres, I. *et al.* Presence and expression of the mating type locus in Paracoccidioides brasiliensis isolates. *Fungal Genet. Biol.* **47**, 373–380 (2010).

67. Goughenour, K. D., Balada-Llasat, J.-M. & Rappleye, C. A. Quantitative Microplate-Based Growth Assay for Determination of Antifungal Susceptibility of Histoplasma capsulatum Yeasts. *J. Clin. Microbiol.* **53**, 3286–3295 (2015).

68. Gómez-Pastor, R., Garre, E., Matallana, E. & Pérez-Torrado, R. *Recent advances in yeast biomass production*. (INTECH Open Access Publisher, 2011).at <http://www.intechopen.com/source/pdfs/19075/InTech-Recent-advances-in-yeast-biomass-production.pdf>

69. Pardee, K. *et al.* Paper-Based Synthetic Gene Networks. *Cell* **159**, 940–954 (2014).

70. Moe-Behrens, G. H. G., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4**, (2013).

3 DIRECTED EVOLUTION OF PEPTIDE ACTIVATED GPCRS

3.1 Introduction

G-protein coupled receptors have the potential to be used as sensing elements for the detection of any biomarker of choice. This receptor family binds a diverse set of ligands and represents a rich source of molecular recognition rivaling that of antibodies [1,2]. In the same way antibodies can be raised against novel ligands by immune selection, GPCRs can be modified to bind novel ligands through directed evolution (DE). In fact, several groups have established DE in yeast as tool for changing GPCR ligand specificity [3,4].

Among the diverse GPCR family, fungal peptide-activated GPCR represent a compelling starting point for directed evolution towards the detection of novel peptide ligands. First, fungal *Ste2*-type GPCRs are expected to couple robustly to the native pheromone pathway, and several have been expressly validated in *S. cerevisiae* with little to no further modifications.[5–8] Second, *Ste2*-type GPCRs from related fungi already recognize different peptides based on the natural evolution of this class of GPCR.[9] Third, *Ste2*-type receptors are highly specific for their respective peptides since they must mediate the species-specific mating reaction while preventing interspecies breeding.[10] Finally, previous groups have generated extensive biochemical characterization and mutagenesis data for the wild type yeast receptor that can be used as guide for DE of *Ste2*-type receptors.[11–16]

In this chapter, we develop an optimized directed evolution pipeline to generate synthetic receptors that can bind to novel peptide ligands (Figure 3.1a). We then apply this method to develop a receptor for a peptide derived from the Cholera Toxin. These new receptors can be introduced directly into the yeast-based detection platform developed in the previous chapter and therefore have the potential to be used as a household sensor for *Vibrio cholera* (Figure 3.1b).

3.1.1 Current cholera detection

Diarrheal diseases are one of the top ten causes of death in developing countries, with extremely high prevalence in children under 5 years of age.[17–19] *V. cholerae* is one of the principal bacterial causative agents of diarrhea annually infecting an estimated ~4 million and killing ~120,000

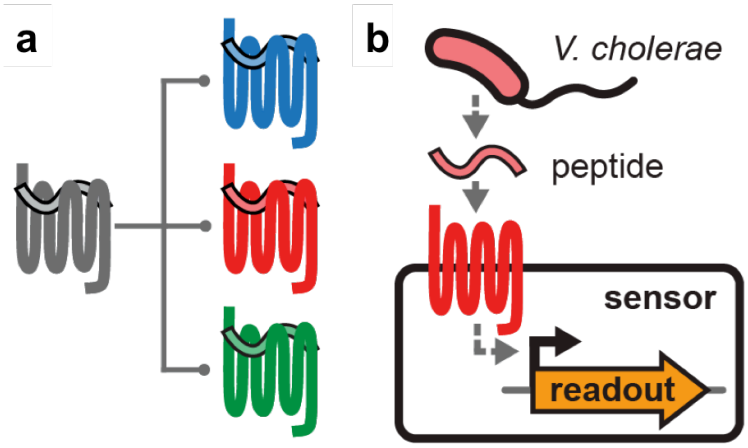


Figure 3.1 Development of receptors to new pathogen targets

(a) Receptors sensitive to new target ligands (colors) can be derived from a native receptor (grey) through directed evolution. **(b)** These new receptors are used as sensing elements in the yeast sensor platform developed in Chapter 2 to make sensor for any target pathogen of choice, such as *V. cholerae*.

people.[20,21] Difficulty of detection and the high virulence of *V. cholerae* lead to the epidemics characteristic of cholera. Moreover, *V. cholerae* can persist in water supplies for long periods adding to its ability to prevail on a yearly basis.[22] Though cholera can be easily treated by oral rehydration, there is no currently available clinical or environmental diagnostic cheap and simple enough for routine monitoring of this deadly pathogen by local communities.

Antibody-based diagnostics have been extremely useful in resource-rich settings, however they are inadequate for cholera that prevails in resource-poor settings. Though functional, current commercial antibody-based dip-stick tests for cholera require sample pre-treatment, exhibit batch inconsistency, and are prohibitive in cost (>$10 per test).[23,24] Other commercial detection in widespread use is quantitative PCR, which is very expensive and requires technically complex equipment.[25] Thus in practice, culturing remains the real test for the presence of *V. cholerae*.[26]

3.2 Results

3.2.1 Overview of DE pipeline

The design of our DE strategy is based on previously reported methods for DE of GPCRs in yeast [3,27–31]. We sought to enhance these methods in terms of throughput and control of selection thresholds. Central to our strategy is a more versatile reporter strain (see section 3.2.2) and a screening pipeline based on 384-well plates (Figure 3.2). Our reporter strain enables positive and negative growth selections as well as selections based on fluorescence-activated cell sorting (FACS). Use of FACS allows for user-defined selection thresholds. Additionally, this reporter strain enables fluorescence-based screening and validation of receptor variants, which reduces the time required

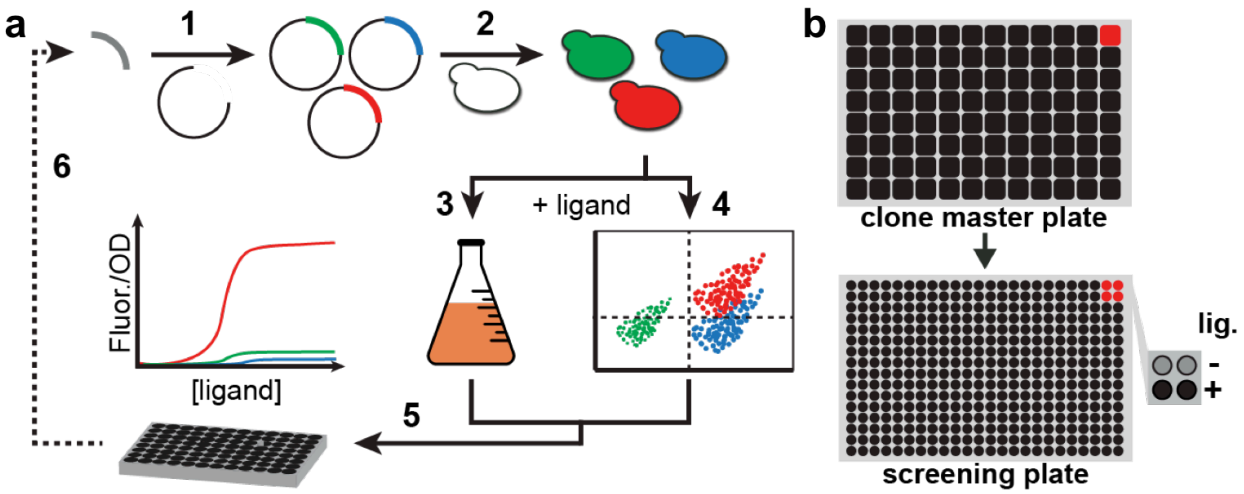


Figure 3.2 Directed evolution pipeline

(a) Fungal GPCRs were evolved by (1) mutagenizing receptor genes, (2) transforming into yeast that carry *pFUS1* reporter constructs and enriching responsive variants by (3) growth on selective media or (4) FACS. Screening in (5) micro titer plates yields hits for (6) further rounds of DE. Figure adapted from [3]. **(b)** In the screening step, individual clones are pre-grown in 96-square well plates that are replicated into 384-well plates, with each clone (red square) directly replicated into four screening conditions (red circles and inset).

for screening and characterization to 12 hours per assay compared to 3 days for previous growth-based reporters [3]. To further improve this screening step, we implemented a pipeline that combines liquid handling robotics, a pin replicator tool and 384-well screening plates (Figure 3.2b). A key feature of this method is the incorporation of duplicates in the screening plate with no added labor. This improves the confidence when potential hits are observed, further reducing labor lost to retransforming and sequencing spurious screening results. Finally, the choice of micro titer format lends itself to full automation of the DE cycle in the future.

3.2.2 Construction of a versatile reporter strain

We constructed a stable reporter strain to perform DE on plasmid-borne receptor variants (yMJ194). As with the yeast biosensor strain (See Chapter 2), the DE reporter strain is derived from a parental strain that has the SST2 and FAR1 genes deleted to prevent deactivation of the pheromone response pathway and avoid growth arrest [27]. Additionally, the STE2 gene was also deleted to avoid any undesired activation due to the endogenous pheromone receptor. A DE reporter strain should enable multiple selection and screening modalities, therefore we explored the following reporters: *pFUS1-HIS3* or *pFUS1-MET15* integrated at the FUS1 locus for positive growth selections;

pFUS2-URA3 or *pFUS2-TRP1* integrated at the FUS2 locus for counter-selections; *pFUS1-mCherry* integrated at the ReRec locus for fluorescence-based selection and screening.

For positive growth selections, we tested pheromone induction of either HIS3 or MET15. Our parental strain lacks both HIS3 and MET15 [32], therefore by introducing either a *pFUS1-HIS3* or *pFUS1-MET15* construct we generated strains that should be conditionally prototrophic for histidine or methionine. However, when either strain was grown in selective media both strains grew at similar rates in the presence or absence of activation of the pheromone pathway by alpha factor. In both cases, high basal transcription from the FUS1 promoter likely produced enough His3 and Met15 gene products to render the strains prototrophic for histidine and methionine. This has been observed previously for selections based on the FUS1 promoter [3]. Additionally, it is well established that 3-amino-1,2,4-triazole (3-AT), an inhibitor of the His3 gene product, can be used to reduce growth due to basal levels of His3 [33]. Therefore we chose *pFus1-His3* as the final positive selection construct. Through titration of 3-AT, we found that 40mM 3-AT reliably gave a large growth difference between a pheromone-induced strain and an uninduced strain (Figure 3.3a).

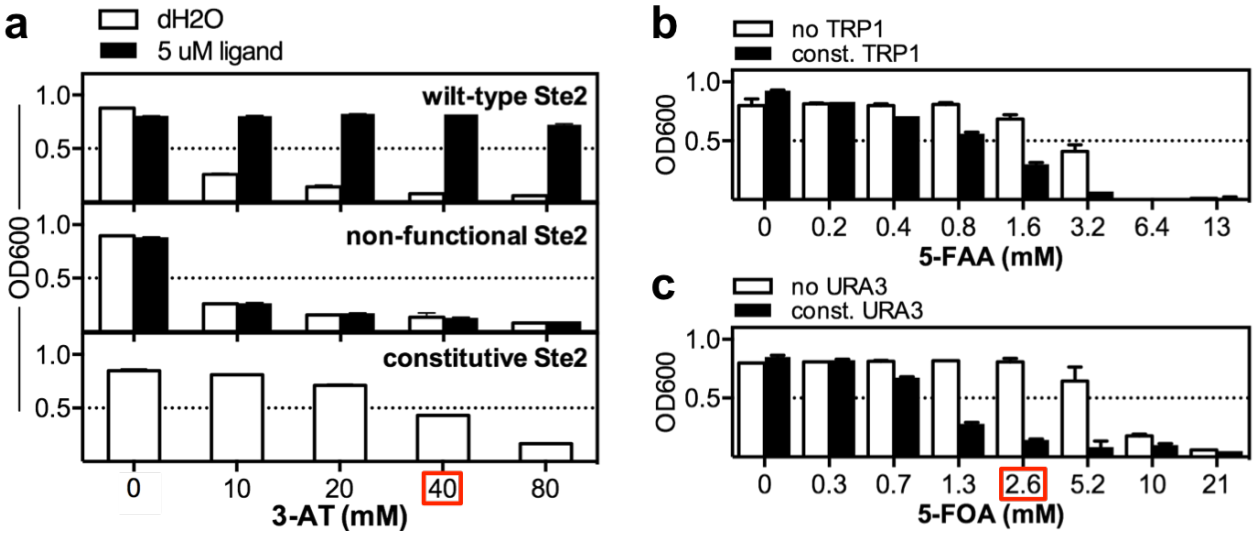


Figure 3.3 Determination of positive and negative selection conditions
(a) Strains expressing three Ste2 receptor variants were grown in positive selection media with varying amounts of 3-AT. 40 mM 3-AT (red square) allows substantial growth only for the induced wild-type receptor. **(b)** Strains with or without constitutive expression of TRP1 were grown in media with varying amounts of 5-FAA. **(c)** Strains with or without constitutive expression of URA3 were grwon in media with varying amounts of 5-FOA. 2.6 mM 5-FOA (red square) prevents growth of the constitutive URA3 strain with out substantially affecting the strain with no URA3 expression. Optical density measured at 48 hours.

For negative growth selections, there are two established counter selectable genes in yeast: URA3 with 5-fluoroorotic acid (5-FOA) and TRP1 with 5 fluoroanthrinilic (5-FAA) [34,35]. Initially we compared the growth difference between our parent DE strain either constitutively expressing or lacking URA3 and TRP1. Titrations of 5-FOA and 5-FAA, showed 3.2 mM 5-FAA and 2.6 mM 5-FOA gave the largest difference in growth for both negative selections (Figure 3.3b,c). However, 5-FAA led to lower growth for the strain lacking the counter-selectable gene. Therefore, we chose the URA3 negative selection. We integrated a *pFus2-Ura3* into the parent strain and showed that at the determined concentration of 5-FOA (2.6 mM) the strain had a large reduction in growth when induced with alpha factor relative to the growth of the uninduced strain (Figure 3.4). Furthermore, immediately after negative selection this strain, which also contained the positive selection construct, showed the appropriate growth behavior in the positive selection media.

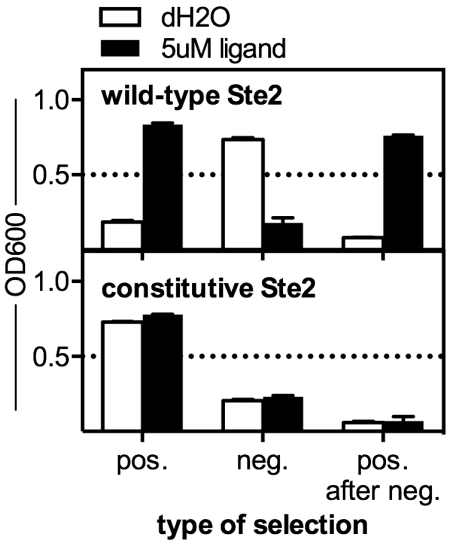


Figure 3.4 Behavior of reporter strain in the positive and negative selections
The final reporter strain shows large differences in growth with and without induction in both the positive (pos.) and negative (neg.) selection conditions. A constitutive Ste2 variant does not grow in the negative selection and can no longer grow in the positive selection following a negative selection.

We then tested whether the counter selection could successfully selected against strains expressing constitutive variants of the Ste2 receptor. As expected, a strain expressing a constitutive mutant of the Ste2 receptor did not grow in the counter selection (Figure 3.4). Interestingly, when immediately switched to the positive selection, this constitutive strain was no longer able to grow, suggesting that the counter-selection not only prevented growth but also caused cell death. This is unexpected since previous reports showed that strains could partially grow post negative selection [34]. However, in the present case the more severe cell death could be a result of concomitant activation of the pheromone response. This fungicidal effect of 5-FOA on cells with an activated pheromone response is expected to enhance the effectiveness of the counter-selection for DE. This is because unwanted constitutive receptor variants remaining in the population after counter-selection

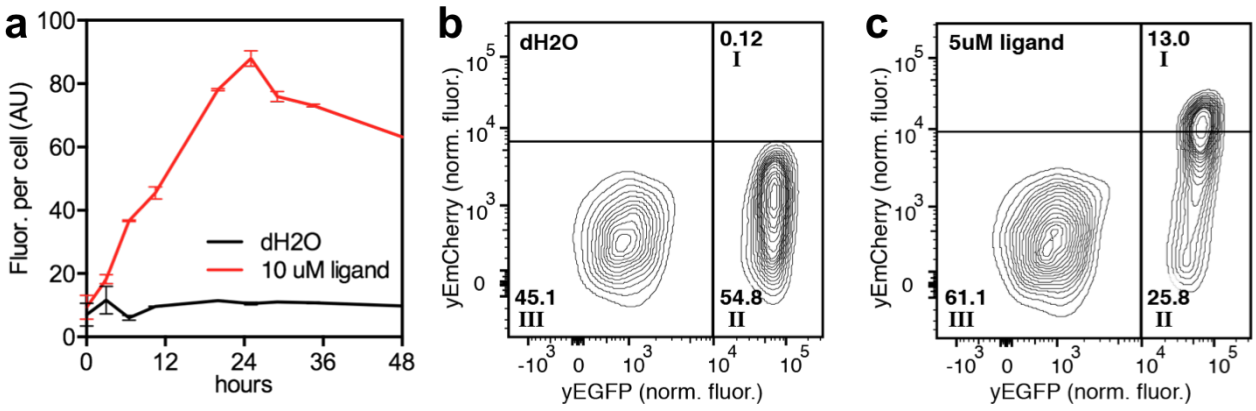


Figure 3.5 Validation of fluorescence reporter

Measurement of the *pFUS1-mCherry* reporter with a **(a)** fluorimeter or **(b, c)** flow cytometer. Contour lines at 5% density levels for b and c. **I** – GFP+/RFP+ population corresponting to induced cells; **II** – GFP+/RFP- population corresponding to uninduced cells; **III** – GFP-/RFP- population corresponding to dead cells.

should no longer be able to compete against the desired inducible receptor variants in a subsequent positive selection.

As with reporter strains constructed previously for the yeast two-hybrid system [36], we sought to include a third reporter construct not based on growth. Previously, this third reporter was the bacterial beta-galactosidase enzyme that can be assayed quantitatively in micro titer plates by visual spectroscopy upon addition chromogenic substrates [37]. Instead, we chose to use a yeast codon optimized variant of the mCherry fluorescent protein [38]. This fluorescent reporter can be assayed in micro titer plates by fluorescence spectroscopy without cell pre-treatment. Additionally, it enables single cell characterization of the variant population by flow cytometry and selections by fluorescence-activated cell sorting (FACS). We thus introduced a *pFus1-mCherry* construct in the DE reporter strain by Reiterative Recombination [39]. Upon induction with alpha factor this strain produced a significant fluorescence signal relative to an uninduced strain (Figure 3.5a). This signal could be distinguished from basal fluorescence by 6 hours and peaked at 18 to 24 hours after induction.

Importantly, our strain also contains a construct constitutively expressing a yeast codon optimized variant of the GFP fluorescent protein (eGFP) as part of the Reiterative Recombination locus [39]. This constitutive eGFP expression acts as an internal control for protein expression levels and can be used in flow cytometry and FACS to exclude dead, non-expressing cells. When this strain

was induced with alpha factor an activated population of cells could be clearly observed by flow cytometry (Figure 3.5b,c). After induction, we always observed a residual eGFP positive/mCherry negative population. We expect that this population represents live cells that no longer carry the plasmid with the receptor gene and are therefore not inducible. It is known that yeast populations carrying plasmids with prototrophic selectable markers can maintain auxotrophic subpopulations lacking the plasmid through cross-feeding [40]. Alternatively, this subpopulation may represent cells that have deactivated the pheromone response through mechanisms not prevented by the SST2 deletion such as receptor internalization [41]. Additionally, the constitutive eGFP expression also provides a second proxy for cell number in micro titer plates in addition to optical density.

In summary, we constructed and validated a versatile reporter strain for DE of receptors that activate the pheromone response pathway in yeast. This strain contains deletions of SST2, FAR1 and STE2 as wells the genome integrated reporters *pFus1-His3*, *pFus2-Ura3* and *pFus1-mCherry*.

3.2.3 Hybrid peptides as intermediate ligands for evolving fungal GPCRs

An important feature of our DE platform is the choice of fungal mating GPCRs as the parent receptors. This protein family is ideal for development of novel sensors since these receptors bind a diverse range of short peptide ligands (see Chapter 4) [9]. Furthermore, peptide ligands aid the DE process, since it is straightforward to design "intermediate" ligands between the native peptide and the target peptide. The use of such intermediate molecules for incremental DE has been used successfully for the evolution of novel substrate specificities of enzymes such as tRNA-synthetases, P450 monooxygenases and steroid receptors [42–45]. Additionally, peptide synthesis is now an outsourced technique making receptor fine-tuning through the use of several intermediate and near-cognate peptides very technically tractable. We therefore hypothesized that fungal GPCRs represent a rich repertoire of diversity that can be used to develop receptors to any peptide of choice.

To test this idea, we chose to develop a receptor for the detection of a cholera-specific peptide. Specifically, we chose a well-established protein biomarker for *V. cholerae*: cholera toxin (CTx). The presence of CTx has been established in stool samples and is a validated proxy for pathogenic *V. cholera*.[46,47] Moreover, there have been several efforts to make antibody-based diagnostics for

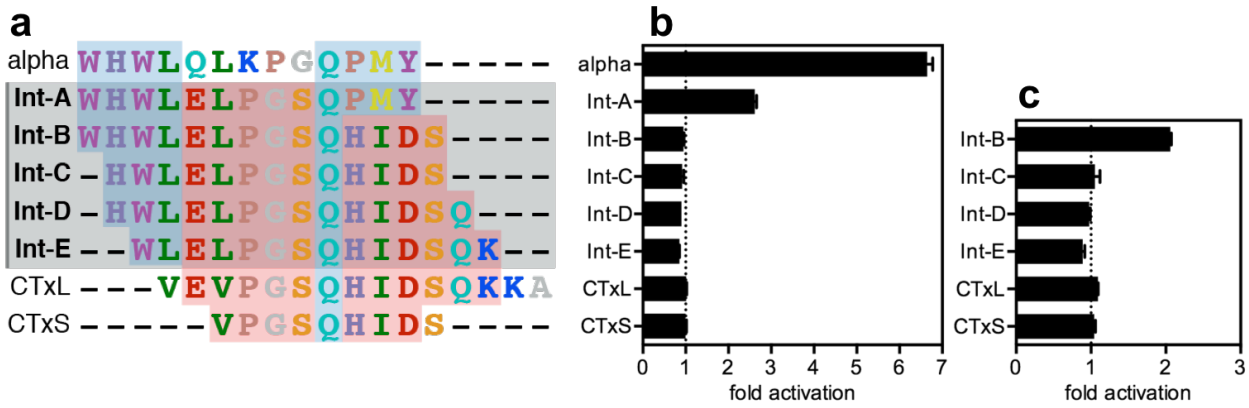


Figure 3.6 Design and characterization of intermediate peptide ligands

(a) Sequences of native ligand (alpha), intermediate ligands (Int-letters, grey shading) and target cholera toxin peptides (CTx). Regions homologous to native ligand highlighted in blue, and to the target peptides highlighted in red. **(b)** Fold activation of the parent Ste2 receptor with 10 μM (alpha, Int-A,B) or 40 μM (Int-C,D,E, CTx-L,S). **(c)** Fold activation with high concentration of the peptides (250 μM)

CTx, validating both its utility and providing a benchmark for the sensitivity required for detection.[46,48]

We expected that there is an optimal size range of peptides that can be detected with *Ste2*-type GPCRs that balance specificity and accessibility through the cell wall to the surface of the cell. Guided by the length of the native ligands of the fungal receptors [9] and a previously described CTx epitope [49] we chose two target peptides raging from 9 to 15 residues. Specifically we chose the peptides: VEVPGSQHIDSQKKA (CTxL) and VPGSQHIDS (CTxS), since they span a region on the bottom face of the cholera toxin beta subunit recognized by the GM1 ganglioside in the intestinal epithelium. Importantly, this region represents a conserved epitope that has been characterized as the target of several antibodies raised against CTx and has also been successfully used in CTx detection assays.[50,51] Furthermore, these peptides share some features with the *S. cerevisiae* alpha factor and related pheromones, such as a PG motif framed by a hydrophobic amino acid four residues from a glutamine. With these features in mind, we designed a set of intermediate peptides that march from alpha factor to the target CTx peptides (Figure 3.6a).

Before beginning DE, we characterized the response of *S. cerevisiae* receptor to these intermediate ligands (Figure 3.6b,c). As expected, the strength of the response decreased rapidly as the homology of the ligands decreased relative to the cognate ligand. Unexpectedly, the wild-type receptor showed a response to the first intermediate ligand (*Int-A*) at 10 μM and a minimal but

measurable response to the second intermediate ligand (*Int-B*) at very high concentrations (0.25 mM). This suggested that DE could begin with selections directly for *Int-B* and that *Int-A* could be used for neutral drift.

3.2.4 Evolution of *S. cerevisiae* Ste2 towards new peptide ligands

We undertook rounds of DE towards a CTx receptor starting from two parent fungal receptors from *S. cerevisiae* and an orthologue from *V. polyspora*. For library creation, we generated receptor variants by error-prone PCR (epPCR) [52] of the whole receptor genes or focused regions containing the putative ligand binding pockets. FACS was then used with the designed intermediate peptides in a series for neutral drift and positive selections. Screening of the selected variants was carried out in 384-well plates as described in Section 3.2.1 and top performing clones were retransformed into a naïve reporter strain for further characterization. The various routes taken during the initial phase of DE are shown schematically in Figure 3.7.

We chose to mutagenize the endogenous *S. cerevisiae Ste2* pheromone receptor (*ScSte2*) and an orthologue from *V. polyspora* (*VpSte2*) as the parental GPCRs since their native peptide pheromones resemble features of the CTx peptide targets and the designed intermediate hybrid peptides. For each, we generated whole gene libraries (*ORF*) and focused libraries spanning from the second to the third extracellular loops (*loops*). We chose this focused region since it has been implicated in ligand binding [14,15,53]. We generated the libraries through epPCR, with $MnCl_2$ concentrations to

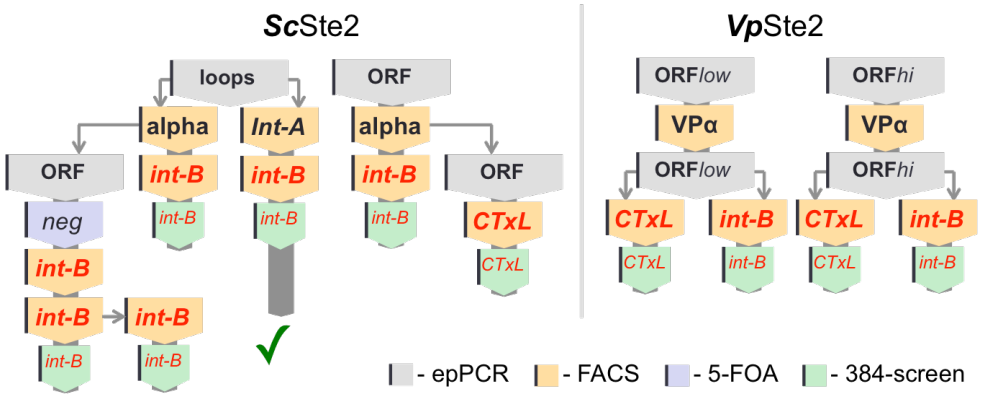


Figure 3.7 Summary of DE routes in phase 1

Two fungal GPCRs were used as parent receptors (ScSte2, VpSte2). Type of mutagenesis (grey boxes) and neutral drift ligand (orange boxes) marked in black font. Positive selection (orange boxes) and screening (green boxes) lignads marked in red font. Green check mark denotes roud leading to active mutant.

give either 1-3 (*low*) or 2-7 (*high*) mutations per gene as confirmed by sequencing 8 – 16 clones per library [28,52]. The final yeast libraries were made via gap repair by co-transforming amplified epPCR products and a linearized plasmid. Library size ranged from 10^5 – 10^6.

Furthermore, we implemented a neutral drift strategy that has been previously applied to GPCRs in yeast [54,55]. Neutral drift involves pre-selecting the library with high concentrations of a ligand that already has activity with the parent receptor of that DE cycle. This removes non-functional or misfolded receptor variants. The library DNA was then extracted and re-mutagenized for further rounds of neutral drift or selection. This iterative process accumulates receptor variants that are well behaved while maintaining sequence diversity.

For FACS selections we selected thresholds based on the fluorescence level of the uninduced library. We set this threshold to capture between 0.5% and 1% of the uninduced library (population **II** in Figure 3.5) and collected 10^5-10^6 cells from the induced library. For neutral drift selections, this threshold captured 30% to 40% of the population. For positive selections with the target ligand, this threshold captured 1% to 2% of the population. FACS-selected cells were amplified in plasmid-selective media and plated to recover individual clones for screening.

In the first phase of DE, we generated 11 separate mutagenesis and selection combinations representing a diversity of ~10^7 variants. After the final selections, we screened 1134 individual clones of which 892 gave replicate results with a CV of less than 15%. Of these clones, we sequenced 48 clones representing 3 high confidence clones and 43 low confidence clones (Figure 3.8a).

Of the low confidence clones, 59% were found to be strains containing empty plasmids not expressing any receptor variant (15% of sequenced clones) or wild type receptors (40% of sequenced clones). These wild type receptors were true library members and not contamination since most had mutations at the DNA level that were silent at the protein level. This was expected since the parent receptor is enriched in neutral drift selections and should represent a significant portion of the negative variants observed in screening. The remaining low confidence non-responsive clones were strains containing a diverse set of mutations ranging from 1 to 6 amino acid changes per variant. As expected, when these low confidence variants were reassayed individually, all were found to have

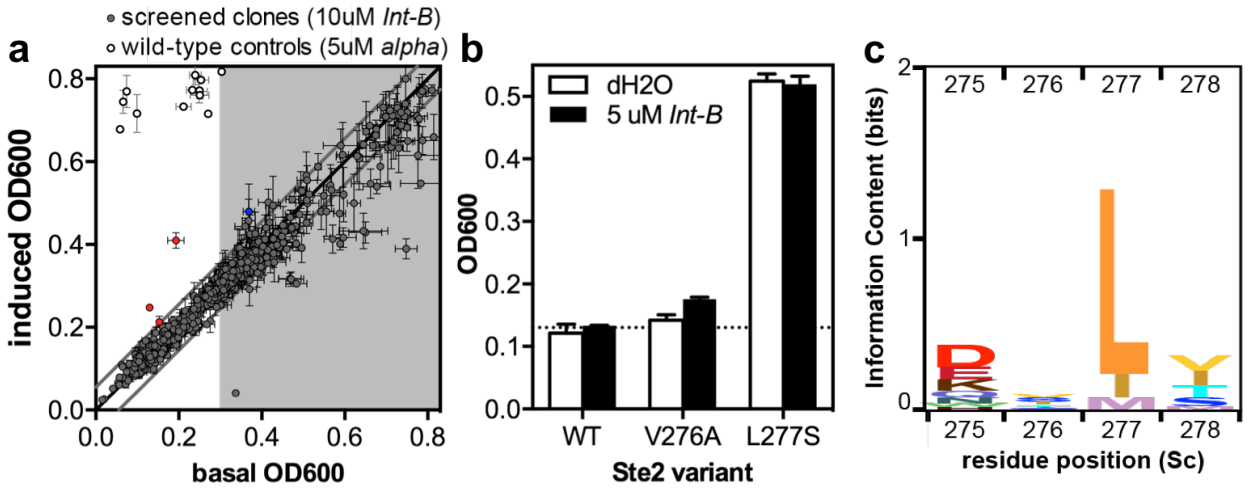


Figure 3.8 Characterization of screening results of phase 1

(a) Mean basal and induced signal values for each clone are used to determine which clones significantly deviate from the null line (black line, basal = induced). Those more than two standard deviations away (grey diagonals) are considered high confidence hits (marked in red). Grey area marks high basal levels. **(b)** Reassaying confirms that the V276A variant (red dots in panel a) is inducible by *Int-B* and the L277S variant (blue dot in panel a) is constitutive. **(c)** Analysis of close orthologues of the wild-type Sc receptor reveals that position 276 is highly variable, while poisition 277 is more highly conserved.

no change in activity to the target ligand relative to the wild type receptor. The exceptions was a variant with the mutation L277S that showed an increase in the basal level.

The three high confidence clones contained only one mutation: V276A. All of these clones resulted from the focused loop library of the *S. cerevisiae* receptor that was first neutrally selected with *Int-1* and then positively selected with *Int-2*. Reassaying confirmed that this mutation increased the activity of the receptor to the *Int-B* peptide (Figure 3.8b).

L277 and V276 are immediately adjacent to each other and both lie at the boundary between the third extracellular loop and seventh transmembrane helix of the receptor. Yet mutation of L277 led to a constitutive variant and mutation of V276 lead to a functional variant with a larger ligand scope. Among the closely related orthologues of *ScSte2* that bind different ligands, L277 is more highly conserved compared to V276 (Figure 3.8c). This reflects the functional difference observed in the variants uncovered in our DE experiment. Taken together, this suggests that these two residues are intimately involved in ligand recognition (V276) and signal transduction (L277). Additionally, it has been previously observed that a nearby residue (K269) makes contact with the N-terminal tryptophan of the cognate ligand [56]. Therefore the mutation of V276 to a smaller alanine residue

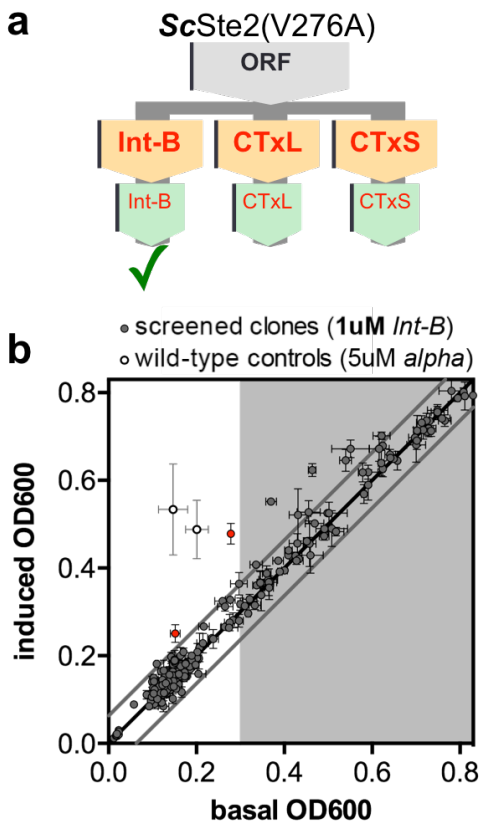


Figure 3.9 Phase 2 screening results
(a) Selections carried out in the second phase of directed evolution. Colors as in Figure 3.7 **(b)** Screening results showing two high confidence hits (red dots) for higher sensitivity receptors to *Int-B*. Colors and lines as in Figure 3.8a.

may serve to better accommodate the N-terminus of a larger peptide like *Int-2* that has 14 amino acids relative to 13 in the cognate ligand.

Encouraged by these results, we repeated the DE cycle to generate higher sensitivity receptors, using the V276A variant as a parent receptor (Figure 3.9a). However, we made some changes to our selection thresholds. To reduce the level of non-responding variants observed in the screens, we increased the stringency of the FACS selections to a level that captured only 0.1% to 0.2% of the uninduced library. Additionally for the selections with *Int-B*, we lowered the concentration from 10 uM to 1 uM, a level that does not activate the V276A variant. In this phase, we generated three library and selection combinations representing ~10^5 variants. After selection, 384 clones were screened of which 244 gave replicate results with CV less than 15%. Of these clones, 2 were high confidence non-constitutive clones (Figure 3.9b).

These two clones reproducibly had higher activity to *Int-B* relative to the parent variant with only the V276A mutation. One variant had the additional mutation Y193C and the other variant had two additional mutations, G129S and S325P. Both Y193 and G129 are in extracellular loops expected to interact with the peptide ligand, while S325 is in the intracellular C-terminal tail of the receptor. To further probe the effect of these mutations, we determined the EC_{50} values to the cognate ligand and *Int-B* for the wild-type receptor (grey), the V276A variant (red), the V276A/Y193C variant (green) and the V276A/G129S/S325P variant (blue) (Figure 3.10a,b,c).

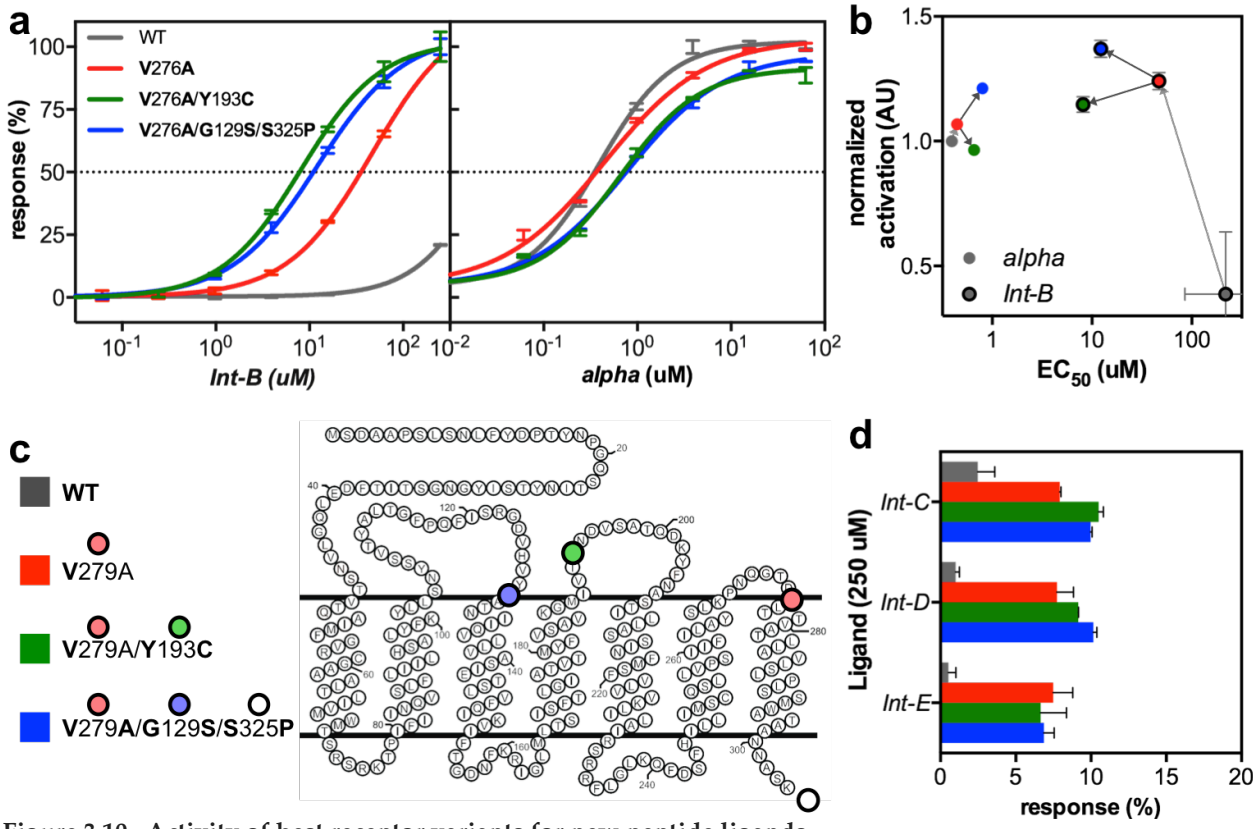


Figure 3.10 Activity of best receptor variants for new peptide ligands

(a) Dose response of wild-type and receptor variants for *Int-B* (left) and cognate ligand (*alpha*, right). Colors as in panel c. **(b)** Comparison of sensitivity and magnitude of activation (span) for the receptor variants towards *alpha* and *Int-B*. Grey arrows denote the ancestry of the receptors. Colors as in panel c. **(c)** Detailed view of mutations for each of the receptor vairants. Double black lines mark the plasma membrane with the extracellular side up and intracellular side down. **(d)** Activation of the receptor variants by high concentrations (250 uM) of the next intermediate ligands. Colors as in panel c.

The V276A/Y193C and the V276A/G129S/S325P variants showed a similar increase in sensitivity of to *Int-B*, however the V276A/G129S/S325P double mutant showed increased maximal activation (Figure 3.10b). This may be due to the S325 mutation that is near phosphorylation and ubiquitnation sites involved in receptor internalization [57,58]. Furthermore, while the original V276A variant showed no change in the EC_{50} value towards the cognate ligand, both phase 2 variants showed a decrease in sensitivity for the cognate ligand (alpha in Figure 3.10a,b). This suggests that these additional mutations have not only increased the ligand scope of the receptor but have also changed its specificity. Furthermore, when we exposed to high concentrations of *Int-C*, *Int-D* and *Int-E*, all receptor variants showed higher activity to compared to wild-type receptor.

In summary, we generated receptor variants that showed an increase in sensitivity of nearly two orders of magnitude to a peptide that has 67% identical residues to the target cholera toxin epitopes. Additionally, the receptor variants showed measureable activity to a peptide with 80% identity to the target ligand and only 23% identity to the cognate ligand.

3.3 Discussion

We successfully designed an optimized DE strategy based on the step-wise use of intermediate hybrid peptides to change the ligand specificity of the GPCRs. We then applied this strategy to engineer a receptor that binds a hybrid peptide 67% identical to a cholera toxin (CTx) peptide epitope with a sensitivity of 8±2 uM. Importantly, CTx is known to be present in clinical samples at a concentration of 350 nM and previous assays have sensitivities ranging from 100nM to 100pM.[47,48]. So while further optimization is required, these results largely support our hypothesis that fungal GPCRs can be engineered to detect any small peptide of choice.

Looking forward, our platform may also have the potential to detect *V. cholerae* in environmental samples. Cheap detection in environmental samples is an urgent problem that remains largely unsolved since the gold-standard culturing technique cannot detect the dormant form of the pathogen. The high sensitivity demonstrated by fungal GPCRs suggests that continued optimization of the receptors devised here might yield extremely sensitive sensors for cholera. Such sensor may open the door for a new way of conducting cholera surveillance.

Furthermore, this work raises the potential for the wider use of peptides as biomarkers in point-of-care diagnostics. Historically, most biomarkers have been proteins and or glycosides. However, there is an emerging field of "peptidomics" that seeks to determine peptide signatures for disease by mass-spectrometry[59,60]. This field is expected to generate a growing list of peptide biomarkers for various diseases. Therefore, our yeast sensors and DE strategy can serve as a complementary technology to generate point-of-care diagnostics to various diseases as their peptide biomarkers are validated.

3.4 References

1. Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).

2. Lequin, R. M. Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clin. Chem.* **51**, 2415–2418 (2005).

3. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5163–5168 (2007).

4. Conklin, B. R. *et al.* Engineering GPCR signaling pathways with RASSLs. *Nat. Methods* **5**, 673–678 (2008).

5. Mayrhofer, S. & Pöggeler, S. Functional Characterization of an α-Factor-Like Sordaria macrospora Peptide Pheromone and Analysis of Its Interaction with Its Cognate Receptor in Saccharomyces cerevisiae. *Eukaryot. Cell* **4**, 661–672 (2005).

6. Janiak, A. M. *et al.* Functional expression of the Candida albicans α-factor receptor in Saccharomyces cerevisiae. *Fungal Genet. Biol.* **42**, 328–338 (2005).

7. Gomes-Rezende, J. A. *et al.* Functionality of the Paracoccidioides Mating α-Pheromone-Receptor System. *PLoS ONE* **7**, e47033 (2012).

8. Wendland, J., Dünkler, A. & Walther, A. Characterization of α-factor pheromone and pheromone receptor genes of Ashbya gossypii. *FEMS Yeast Res.* **11**, 418–429 (2011).

9. Martin, S. H., Wingfield, B. D., Wingfield, M. J. & Steenkamp, E. T. Causes and Consequences of Variability in Peptide Mating Pheromones of Ascomycete Fungi. *Mol. Biol. Evol.* **28**, 1987–2003 (2011).

10. Martin, S. H., Steenkamp, E. T., Wingfield, M. J. & Wingfield, B. D. Mate-recognition and species boundaries in the ascomycetes. *Fungal Divers.* **58**, 1–12 (2013).

11. Martin, N. P., Ćelić, A. & Dumont, M. E. Mutagenic mapping of helical structures in the transmembrane segments of the yeast α-factor receptor. *J. Mol. Biol.* **317**, 765–788 (2002).

12. Leavitt, L. M., Macaluso, C. R., Kim, K. S., Martin, N. P. & Dumont, M. E. Dominant negative mutations in the α-factor receptor, a G protein-coupled receptor encoded by the STE2 gene of the yeast Saccharomyces cerevisiae. *Mol. Gen. Genet. MGG* **261**, 917–932 (1999).

13. Naider, F. & Becker, J. M. The α-factor mating pheromone of Saccharomyces cerevisiae: a model for studying the interaction of peptide hormones and G protein-coupled receptors. *Peptides* **25**, 1441–1463 (2004).

14. Mathew, E. *et al.* Differential Interactions of Fluorescent Agonists and Antagonists with the Yeast G Protein Coupled Receptor Ste2p. *J. Mol. Biol.* **409**, 513–528 (2011).

15. Hauser, M., Kauffman, S., Lee, B.-K., Naider, F. & Becker, J. M. The First Extracellular Loop of the Saccharomyces cerevisiae G Protein-coupled Receptor Ste2p Undergoes a Conformational Change upon Ligand Binding. *J. Biol. Chem.* **282**, 10387–10397 (2007).

16. Lin, J. C., Parrish, W., Eilers, M., Smith, S. O. & Konopka, J. B. Aromatic residues at the extracellular ends of transmembrane domains 5 and 6 promote ligand activation of the G protein-coupled alpha-factor receptor. *Biochemistry (Mosc.)* **42**, 293–301 (2003).

17. Kosek, M., Bern, C. & Guerrant, R. L. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull. World Health Organ.* **81**, 197–204 (2003).

18. WHO | The world health report 2004 - changing history. *WHO* at <http://www.who.int/whr/2004/en/>

19. United Nations *The millennium development goals report 2005*. (United Nations : UN Dept. of Public Information, 2005).

20. Cholera, 2011. *Wkly. Epidemiol. Rec. Health Sect. Secr. Leag. Nations* **87**, 289–304 (2012).

21. Sánchez, J. & Holmgren, J. Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea. *Curr. Opin. Immunol.* **17**, 388–398 (2005).

22. Drasar, B. S. & Forrest, B. D. *Cholera and the Ecology of Vibrio Cholerae*. (Springer, 1996).

23. Ley, B. *et al.* Evaluation of a rapid dipstick (Crystal VC) for the diagnosis of cholera in Zanzibar and a comparison with previous studies. *PloS One* **7**, e36930 (2012).

24. Njanpop-Lafourcade, B. personal communication. (2013).at <Africhol>

25. Lyon, W. J. TaqMan PCR for Detection of Vibrio cholerae O1, O139, Non-O1, and Non-O139 in Pure Cultures, Raw Oysters, and Synthetic Seawater. *Appl. Environ. Microbiol.* **67**, 4685–4693 (2001).

26. Dick, M. H., Guillerm, M., Moussy, F. & Chaignat, C.-L. Review of two decades of cholera diagnostics--how far have we really come? *PLoS Negl. Trop. Dis.* **6**, e1845 (2012).

27. Pausch, M. H. G-protein-coupled receptors in Saccharomyces cerevisiae: high-throughput screening assays for drug discovery. *Trends Biotechnol* **15**, 487–94 (1997).

28. Dong, S., Rogan, S. C. & Roth, B. L. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nat. Protoc.* **5**, 561–573 (2010).

29. Erlenbach, I. *et al.* Functional expression of M1, M3 and M5 muscarinic acetylcholine receptors in yeast. *J. Neurochem.* **77**, 1327–1337 (2001).

30. Price, L. A., Strnad, J., Pausch, M. H. & Hadcock, J. R. Pharmacological characterization of the rat A2a adenosine receptor functionally coupled to the yeast pheromone response pathway. *Mol. Pharmacol.* **50**, 829–837 (1996).

31. Price, L. A., Kajkowski, E. M., Hadcock, J. R., Ozenberger, B. A. & Pausch, M. H. Functional coupling of a mammalian somatostatin receptor to the yeast pheromone response pathway. *Mol. Cell. Biol.* **15**, 6188–6195 (1995).

32. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast Chichester Engl.* **14**, 115–132 (1998).

33. Bartel, P. L. & Fields, S. *The Yeast Two-hybrid System*. (Oxford University Press, 1997).

34. Boeke, J. D., LaCroute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet. MGG* **197**, 345–346 (1984).

35. Toyn, J. H., Gunyuzlu, P. L., Hunter White, W., Thompson, L. A. & Hollis, G. F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553–560 (2000).

36. Vidal, M., Braun, P., Chen, E., Boeke, J. D. & Harlow, E. Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci.* **93**, 10321–10326 (1996).

37. Naylor, L. H. Reporter gene technology: the future looks bright. *Biochem. Pharmacol.* **58**, 749–757 (1999).

38. Keppler-Ross, S., Noffz, C. & Dean, N. A new purple fluorescent color marker for genetic studies in Saccharomyces cerevisiae and Candida albicans. *Genetics* **179**, 705–710 (2008).

39. Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U A* **108**, 15135–15140 (2011).

40. Pronk, J. T. Auxotrophic Yeast Strains in Fundamental and Applied Research. *Appl. Environ. Microbiol.* **68**, 2095–2100 (2002).

41. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26**, 339–350 (2005).

42. Chen, Z. & Zhao, H. Rapid Creation of a Novel Protein Function by in Vitro Coevolution. *J. Mol. Biol.* **348**, 1273–1282 (2005).

43. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13**, 3–9 (2009).

44. Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D. & Schultz, P. G. Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **131**, 12921–12923 (2009).

45. Fasan, R., Chen, M. M., Crook, N. C. & Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome P450BM3 Exhibiting Nativelike Catalytic Properties. *Angew. Chem.* **119**, 8566–8570 (2007).

46. Almeida, R. J. *et al.* Comparison of a latex agglutination assay and an enzyme-linked immunosorbent assay for detecting cholera toxin. *J. Clin. Microbiol.* **28**, 128–130 (1990).

47. Jin, D. *et al.* Quantitative detection of Vibrio cholera toxin by real-time and dynamic cytotoxicity monitoring. *J. Clin. Microbiol.* **51**, 3968–3974 (2013).

48. Rowe-Taitt, C. A., Cras, J. J., Patterson, C. H., Golden, J. P. & Ligler, F. S. A Ganglioside-Based Assay for Cholera Toxin Using an Array Biosensor. *Anal. Biochem.* **281**, 123–133 (2000).

49. Jacob, C. O., Sela, M. & Arnon, R. Antibodies against synthetic peptides of the B subunit of cholera toxin: crossreaction and neutralization of the toxin. *Proc. Natl. Acad. Sci.* **80**, 7611–7615 (1983).

50. Rowe-Taitt, C. A., Cras, J. J., Patterson, C. H., Golden, J. P. & Ligler, F. S. A ganglioside-based assay for cholera toxin using an array biosensor. *Anal Biochem* **281**, 123–33 (2000).

51. Shoham, M. Crystal Structure of an Anticholera Toxin Peptide Complex at 2·3 Å. *J. Mol. Biol.* **232**, 1169–1175 (1993).

52. Cirino, P. C., Mayer, K. M. & Umeno, D.in *Dir. Evol. Libr. Creat.* (Arnold, F. H. & Georgiou, G.) 3–9 (Humana Press, 2003).at <http://link.springer.com/protocol/10.1385/1-59259-395-X%3A3>

53. Umanah, G. K. E., Huang, L.-Y., Maccarone, J. M., Naider, F. & Becker, J. M. Changes in conformation at the cytoplasmic ends of the fifth and sixth transmembrane helices of a yeast G protein-coupled receptor in response to ligand binding. *Biochemistry (Mosc.)* **50**, 6841–6854 (2011).

54. Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939–942 (2008).

55. Ault, A. D. & Broach, J. R. Creation of GPCR-based chemical sensors by directed evolution in yeast. *Protein Eng. Des. Sel.* **19**, 1–8 (2006).

56. Umanah, G. K. E. *et al.* Identification of Residue-to-residue Contact between a Peptide Ligand and Its G Protein-coupled Receptor Using Periodate-mediated Dihydroxyphenylalanine Cross-linking and Mass Spectrometry. *J. Biol. Chem.* **285**, 39425–39436 (2010).

57. Hicke, L. & Riezman, H. Ubiquitination of a Yeast Plasma Membrane Receptor Signals Its Ligand-Stimulated Endocytosis. *Cell* **84**, 277–287 (1996).

58. Jenness, D. D., Li, Y., Tipper, C. & Spatrick, P. Elimination of defective alpha-factor pheromone receptors. *Mol. Cell. Biol.* **17**, 6236–6245 (1997).

59. Sauer, S. & Kliem, M. Mass spectrometry tools for the classification and identification of bacteria. *Nat. Rev. Microbiol.* **8**, 74–82 (2010).

60. Mischak, H. *et al.* Capillary electrophoresis–mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: An update of recent developments. *Mass Spectrom. Rev.* **28**, 703–724 (2009).

4 FUNGAL RECEPTOR-PEPTIDE PAIRS: EXTENSIBLE COMMUNICATION MODULES FOR SYTHETIC BIOLOGY

4.1 Introduction

The construction of synthetic multicellular systems requires a new set of parts beyond the ones developed for single cell systems [1–3]. One key requirement is a communication language that can be used to coordinate and trigger the actions of the constituent cell types [4–6]. An ideal communication language is composed of orthogonal communication channels whose cell-to-cell connections are easily programmed by the researcher [7]. One approach used by natural systems is the secretion of diffusible small molecules, peptides or proteins that trigger specific receptors uniquely expressed by the target cell [8]. Some examples include the peptide hormones and small molecule steroid hormones that coordinate the intricate multicellular programs of development and the menstrual cycle [9,10]. While, some simple multicellular systems have already been demonstrated [11–13], more complex systems have remained out of reach. This has been mainly due to inherent limitations in the synthetic communication systems currently available. The development of a more versatile communication language would enable access to synthetic multicellular systems with properties such as division of labor, self-patterning and distributed computation.

In this chapter, we propose that fungal GPCRs and their peptide ligands are the ideal modules to develop a synthetic communication language for coordinating multicellular system. To evaluate this idea, we first mine fungal genomes to extract GPCR and their corresponding putative peptide ligands. We then express them in yeast and characterize their function, sensitivity and orthogonality. Our results show that these receptor-peptide pairs already far exceed what is possible with available systems and stand to become an easily extensible communication language.

4.1.1 Current systems for synthetic cell-cell communication

There has been an array of communication systems used for the implementation of synthetic multicellular systems in bacteria, yeast and mammalian cells [4–6]. However, almost all of these systems use signaling components that are not immediately extensible and therefore are inherently limited to the implementation of just a few communication channels.

By far the most used systems are based on bacterial quorum sensing. Many bacteria use diffusible acyl homoserine lactones as signaling molecules that trigger a population wide response at high concentrations [14,15]. Different bacterial species have slightly different signals with variation primarily in the length of the acyl chain. While some of the natural receptors are partially orthogonal, a majority of signals activate multiple receptors [16]. Even through protein engineering of the receptors, at most two fully orthogonal communication channels have been demonstrated simultaneously [17]. A second approach is the use of metabolic cross-feeding as type of communication [7,18]. In these systems the cells are made interdependent on each other for survival, each producing a metabolite for which the other is auxotrophic. A more recent approach, implemented the use of DNA as a communication channel [19]. The signal DNA was transferred from cell to cell via a virus that could escape from the sender cell and then infect the receiver cell.

Synthetic communication systems in mammalian cells have also used natural diffusible signals such as nitrous oxide (NO), amino acids and growth factors [20–22]. Alternatively, mammalian systems have also been made that implement contact based signaling mediated by two protein domains displayed on the surface of the interacting cells [23].

An artificial quorum sensing system has also been implemented in yeast by rewiring secretion of the native mating pheromone [24]. In another example, the citokinin pheromone system from plants was imported and successfully linked to yeast transcription factors [25]. A third example from metabolic engineering showed that yeast could be co-cultured with bacteria through a system of cross feeding and detoxification [26]. Finally, a system was built allowing the communication between bakers yeast and fission yeast by mutual secretion of their respective pheromones [27]. These initial examples demonstrate the portability of the yeast mating receptor and peptide pheromone, and thus set the stage for the exploration in this chapter of the potential of these systems to be used as synthetic communication modules.

4.2 Results

4.2.1 Genome mining for fungal mating GPCRs

While the Ste2 mating receptor of *S. cerevisiae* has received extensive focus for its role in the prototypic mitogen-activated protein kinase signal pathway in yeast [28], it has been redeployed

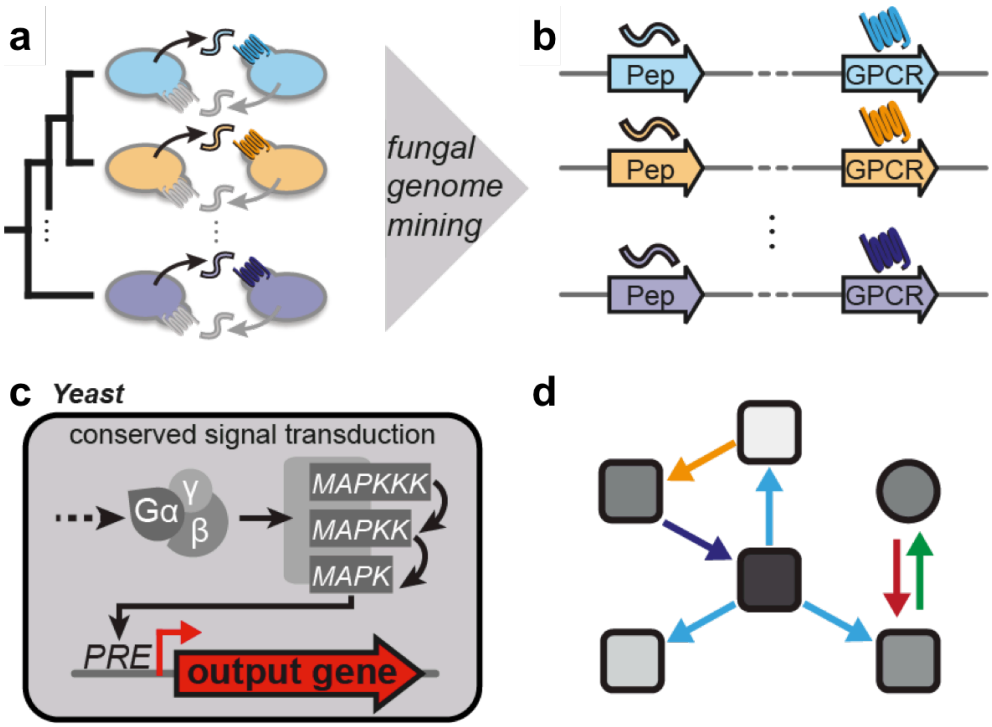


Figure 4.1 Fungal genomes can be mined for novel communcation modules

(a) Most members of the Ascomycete fungi mate via species-specific petide/receptor signaling. **(b)** The secreted peptides (Pep) and the receptors (GPCR) that sense them are homologous and can be found by genome mining. **(c)** The underlying signal transduction pathway conserved and so receptors/peptides can be used in the model organism (yeast) in a plug-and-play manner. **(d)** These peptide-receptor pairs can be used as extensible communication modules for building novel communication networks.

only recently as a signaling module in synthetic biology. Growing interest in this receptor has led to its use to implement multicellular Boolean logic and interspecies communication [27,29].

These results reflect the extensive exploration carried out in the 1980's that demonstrated the portability of the Ste2 and Ste3 receptors between different mating types [30,31], as well as more recent work using *S. cerevisiae* to explore the biological significance of receptors from other fungi [32–34]. Moreover, with the explosion in the number of sequenced fungal genomes [35,36], it has now become clear that the pheromone response pathway is highly conserved within the Ascomycete fungi (Figure 4.1a,b)[37]. These newly sequenced genomes have revealed that a majority of ascomycetes share orthologues for many of the proteins involved in this signaling pathway, from receptors to downstream effectors. Even fungi previously thought to be asexual, have now been demonstrated to have cryptic mating cycles [38]. Taken together, these previous results suggest that a very large set of fungal GPCRs exists that could be used as sensing elements in synthetic cell-cell

communication systems. Furthermore, the demonstrated portability of a handful of receptors to *S. cerevisiae* hints that in fact *most* fungal GPCRs might readily couple to the pheromone response pathway in this well understood model organism (Figure 4.1c).

To validate this hypothesis, we first sought to build an exhaustive data set of all orthologues in the Ste2 receptor family. We aimed to incorporate species information in this database in order to filter for receptors expected to bind unique peptide ligands (those from distinct species) while also maintaining receptors from different variants within the same species expected to provide sequence information on neutral mutations. We accomplished this by curating the InterPro group IPR000366 [39]. This is an automatically generated group of 617 sequences that match the InterPro Ste2 motif. We found 143 to be obsolete entries in the Uniprot datatbase, 93 to be non-fungal and 381 to be true Ascomycete sequences. We reduced the redundancy by removing collapsing pairs of sequences that were more than 98% identical (mostly removing sequences from different strains fo the same species). The remaining 330 sequences all contained conserved regions corresponding to the core seven transmembrane helix domain of the *S. cerevisiae* receptor (Y17 to N301). We further pruned the list to 303, removing members with large gaps (>20 residues) in this region. Analysis of the putative core receptor region revealed a highly diverse family with the lowest homology between receptor pairs of 10% (Figure 4.2). However on average each member was 80% homologous to its closest neighbor. It is interesting to note that the C-termini were highly diverse with some including additional predicted transmembrane domains.

To characterize the potential portability of these receptors as functional sensing elements in *S. cerevisiae* we also characterized the conservation of residues spanning from the fifth to sixth transmembrane helix and including the third intracellular loop. This region has been implicated in signal transduction and is though to directly contact the heterotrimeric G-protein [40,41]. Receptors with higher homology to the *S. cerevisiae* Ste2 in this region are therefore predicted to couple functionally to the endogenous G-protein. We found that indeed residues in this region were more highly conserved than residues in the remaining portions of the receptors (Figure 4.3). In particular we found that all receptors contained positively charged residues in the third intracellular loop.

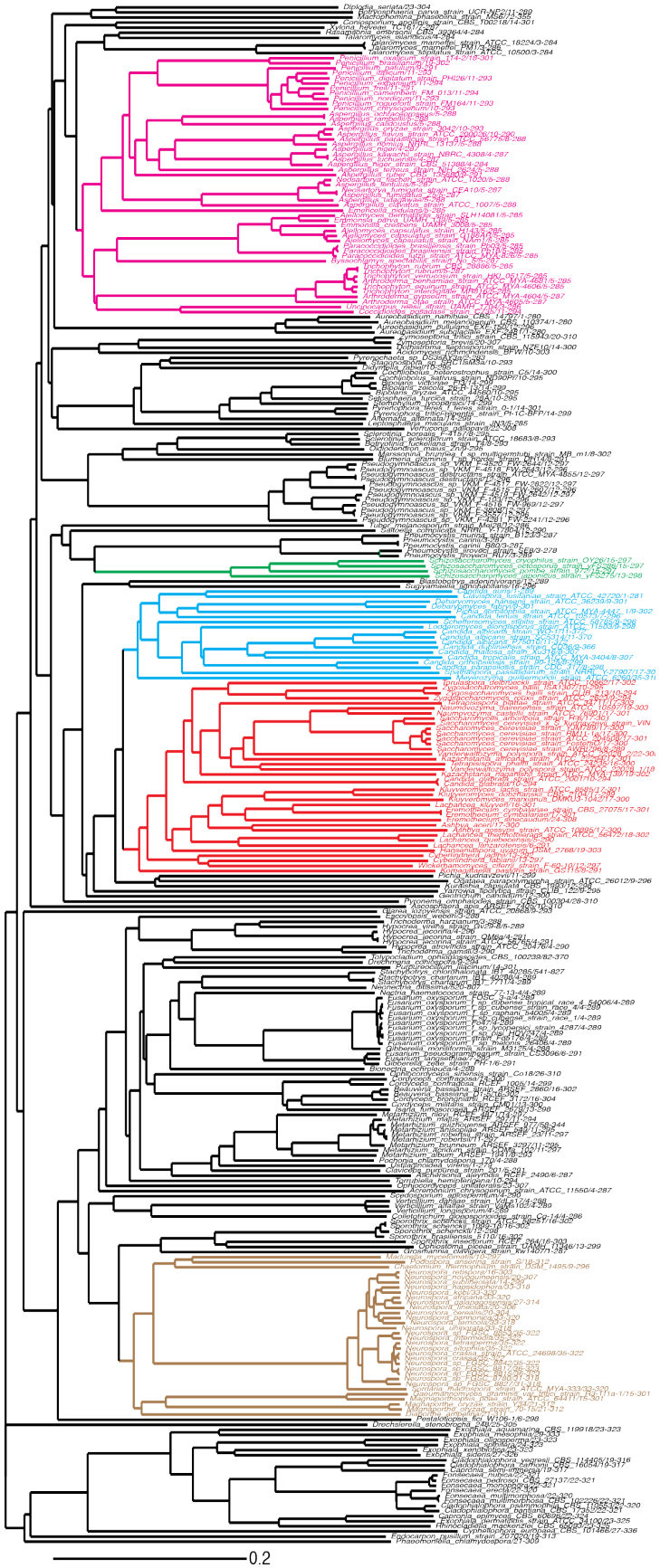


Figure 4.2 Phylogenetic tree of mined fungal receptors

Fungal GPCRs homologous to *S. cerevisiae* Ste2 were curated from the InterPro group IPR000366. A high quality set was aligned and the conserved receptor core was analysized (Y15 to N301 relatice to WT Ste2). Phylogenetic analysis of this region shows a diverse family of receptors that mirrors the expected species phylogeny. Pink – receptors from the fungi orders Eurotiales and Onygenales including genuses: *Penicillum*, *Aspergillus*, *Histoplasma* and *Paracocidioides*. Green – receptors from the genus *Schizosaccharomyces* including the model organism *S. pombe*. Blue – receptors from fungi related to the human pathogen *Candida albicans*. Red – receptors from fungi closely related to the baker's yeast (*S. cerevisiae*). Brown – receptors from the fungi order Sordariales including the model organism *Neurospora crassa*.

4.2.2 Genome mining for fungal peptide pheromones

With a set of putative fungal receptors in hand we next sought to generate a corresponding set of pheromone genes. This goal expands on a set of putative pheromone genes reported by Martin et al [42]. This existing set contains pheromone genes from 64 unique fungal species. As with *S. cerevisiae* some of these fungal species have multiple structural genes encoding highly homologous peptide pheromones. Using these initial genes as seeds we expanded the set of putative pheromone genes and matched them to their corresponding receptors by using the taxonomic information gathered previously (see Section 4.2.1). A simple BLAST routine yielded 119 pheromone genes

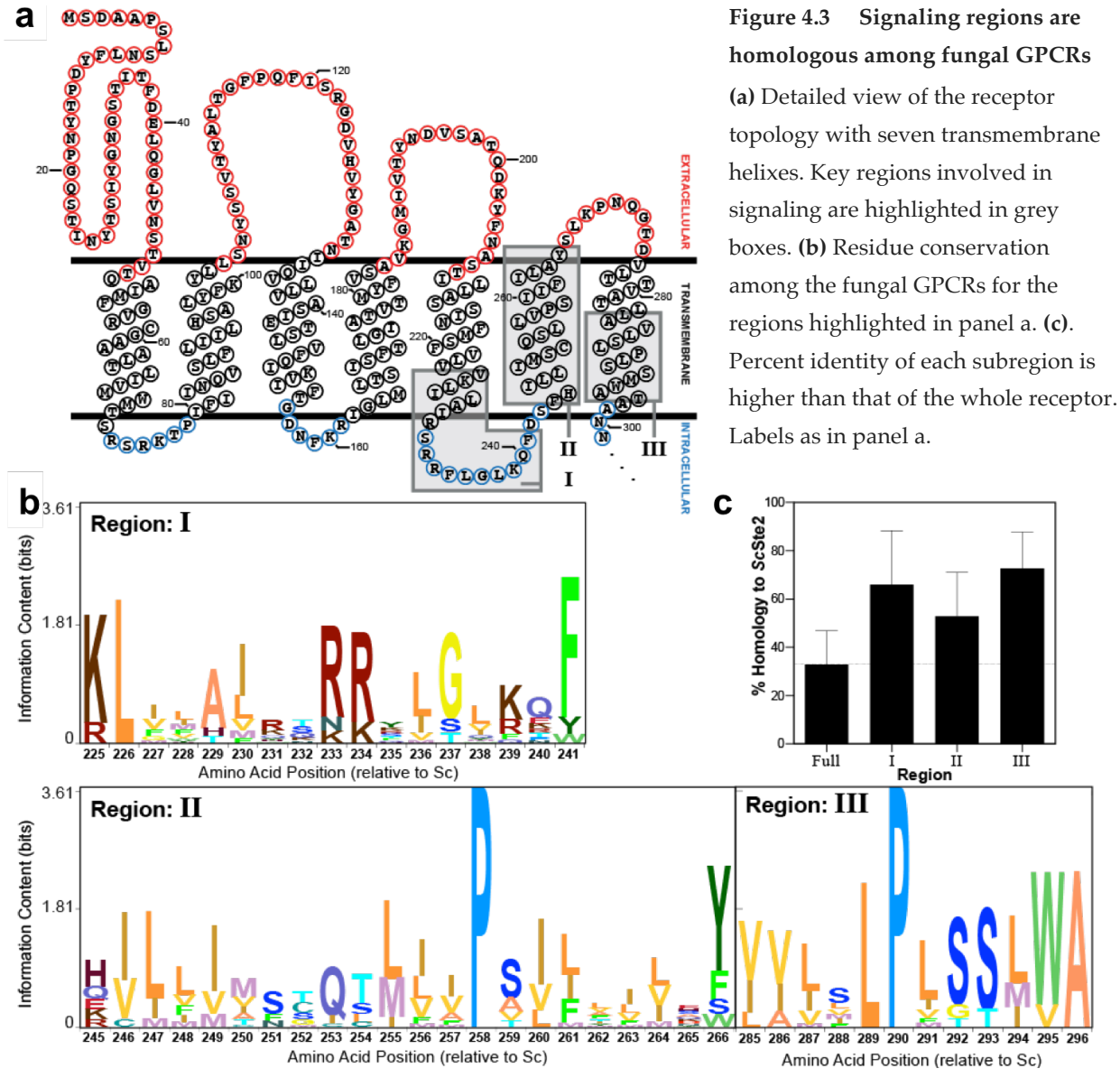


Figure 4.3 Signaling regions are homologous among fungal GPCRs **(a)** Detailed view of the receptor topology with seven transmembrane helixes. Key regions involved in signaling are highlighted in grey boxes. **(b)** Residue conservation among the fungal GPCRs for the regions highlighted in panel a. **(c)**. Percent identity of each subregion is higher than that of the whole receptor. Labels as in panel a.

derived from unique fungal species [43–45].

As described previously, putative secreted peptides were predicted from processing motifs present in the pheromone genes [42]. These pheromone genes have a signature architecture consisting of a hydrophobic prepro secretion signal followed by repeats of the putative secreted peptide flanked by proteolitic processing sites [46]. The repetitive nature of these genes was the most salient feature that enabled prediction of a majority of the mature peptides. The repeated regions contained short stretches of 10-15 residues that were perfectly conserved (Figure 4.4). When there were different residues between these repeats the mismatched residues were often from the same chemical class (e.g. K and R).

These alignments also revealed the nature and extent of variation of the flanking processing sites. As noted previously, these flanking processing sites were runs of X-A and X-P dipeptides, and Kex2-like cleavage sites (KR, QR, NR) [42,47–51]. For a subset of putative pheromone genes no repeats were found, however many of these members still contained the flanking processing sites, often with a single site preceding a short C-terminal peptide as has been observed with *H. capsulatum* and *P. brasiliensis* [34]. Finally, we note that the pheromone genes from the *Aspergillus* genus did not contain perfect repeats or single peptides flanked by well-defined processing sites. In these fungi there were often three imperfect repeats with differing termini and lengths as predicted from the processing sites. This variety in length stood out since other pheromone genes with repeated regions had near perfect conservation in predicted peptide length. This variation in the peptide repeats may suggest that these pheromone genes are vestigial in the historically asexual *Aspergillus* genus [38].

While there was great variation in the sequence of the predicted peptides there were some conserved features. Across all peptides, there was strong preference for aromatic amino acids at the termii, particularly tryptophan at the N-terminus. As with alpha-factor from *S. cerevisiae* these terminal aromatic residues may represent a feature required to engage hydrophobic regions in the respective receptors to induce the conformational changes required for signal transduction [52,53]. Additionally, a subset of peptides from related species contained paired cysteine residues near the termii, often at the second and penultimate positions. A possible disulfide bond suggests that these

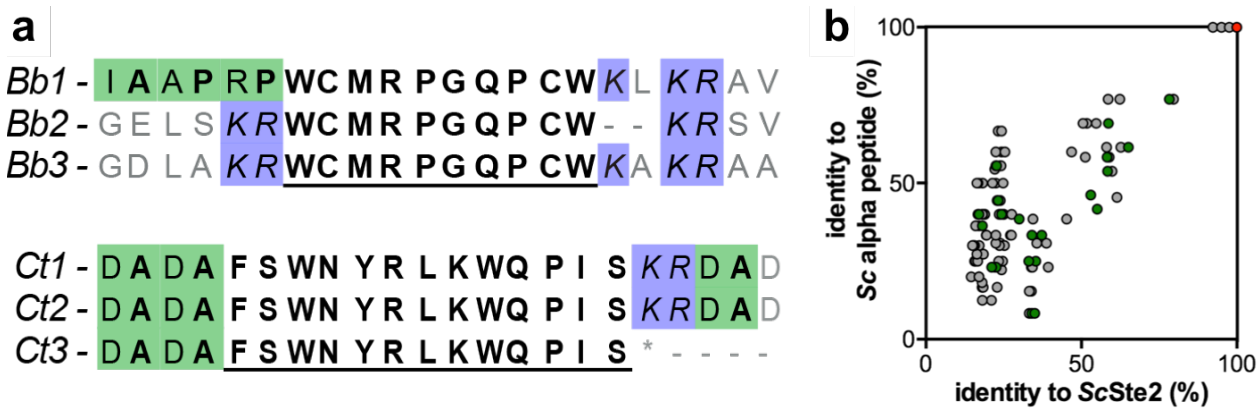


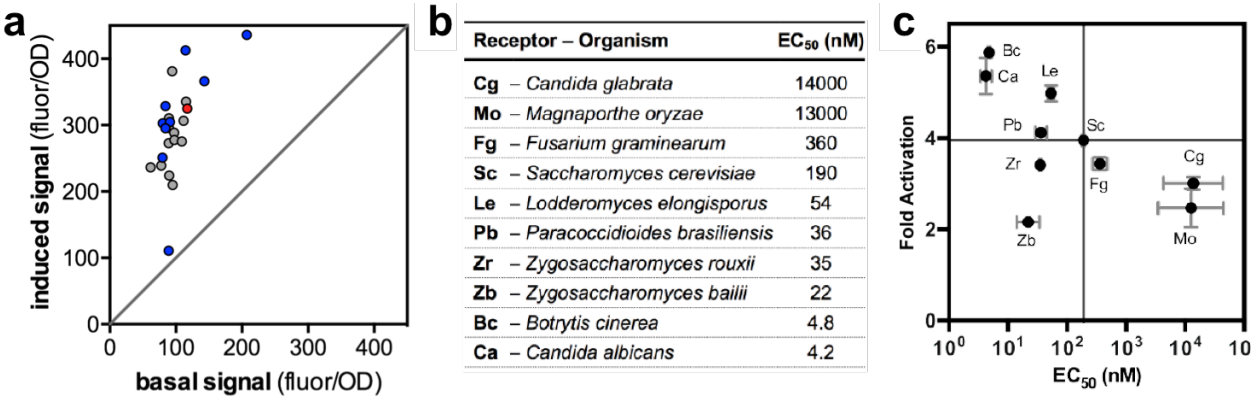
Figure 4.4 Secreted fungal peptides can be found by homology of repeasts and processing sites
(a) Two examples of determining the secreted peptide sequence from *Beauvaria bassiana* (*Bb*) and *Candida tenuis* (*Ct*). The sequence context is shown for the perfect repeats found in each peptide gene. Kex2-like processing sites in blue. Ste13-like processing sites in green. Predicted mature sequence underlined. **(b)** Percent identities of the peptide-receptor pairs to their *S. cervisiae* counterparts (red). Peptide-receptor pairs screened for function marked in Figure 4.5a marked green.

peptides may be active as conformationally locked cyclic peptides. This mirrors other small peptides that engage cell surface proteins, such as the gram-positive bacterial autoinducing peptides and the scorpions toxins [54,55].

In summary, through mining of the available ascomycete genomes we assembled a database of putative receptor-peptide pairs (100+) expected to be orthogonal and to functionally couple to the *S. cerevisiae* pheromone response pathway (Figure 4.4). The large ligand diversity may be immediately accessible for building communication modules in yeast.

4.2.3 Functional validation of GPCR-peptide pairs

To validate the set of receptor-peptide pairs mined from fungal genomes, we expressed a small subset of receptors in our fluorescent reporter strain (see Chapter 3) and tested their function with synthetic peptides. We chose this subset to span a wide range of receptor and peptide homologies relative to the wild-type *S. cerevisiae* counterparts (green dots in Figure 4.4b). Additionally, this set included receptors genes cloned directly from the source fungal genomic DNA that retained the original codon bias of the host as well as receptor genes cloned from synthetic yeast codon-optimized DNA. Reporter strains constitutively expressing these heterologous receptors were induced with 5 uM of the corresponding predicted peptide ligands. Of 25 initially screened receptors, 22 generated a fluorescent signal above baseline when induced (Figure 4.5a). Considering



Receptor – Organism	EC_{50} (nM)
Cg – *Candida glabrata*	14000
Mo – *Magnaporthe oryzae*	13000
Fg – *Fusarium graminearum*	360
Sc – *Saccharomyces cerevisiae*	190
Le – *Lodderomyces elongisporus*	54
Pb – *Paracoccidioides brasiliensis*	36
Zr – *Zygosaccharomyces rouxii*	35
Zb – *Zygosaccharomyces bailii*	22
Bc – *Botrytis cinerea*	4.8
Ca – *Candida albicans*	4.2

Figure 4.5 Functional validation of peptide-receptor pairs

(a) An initial screen revealed that a majority of predictd peptide-receptor pairs were functional in a yeast reporter strain. Each receptor was induced with 5 uM of its cognate ligand. Wild-type marked in red. Receptors chosen for further chracterization marked in blue. **(b)** The receptors showed a range of sensitivities with most in the nanomolar range. Calculated from data in Figure 4.6. **(c)** Most of the receptors perform as well of better than the wild-type receptor (*Sc*) with no need for aditional engineering.

the range of receptor and peptide diversity represented by this subset of receptors, this result supports our initial hypothesis: most fungal GPCRs readily couple to the pheromone response pathway of the model organism *S. cerevisiae*.

Next, to determine whether these receptor-peptide pairs could be used as synthetic communication modules, we chose to further characterize 9 receptors. We also included 3 receptors that did not initially show a response in the screen. We determined the EC_{50} and maximal fold-activation values for all the receptors (Figure 4.5). Of the three initially non-responsive receptors, one (Mo) could in fact be activated with high ligand concentrations, while the two other receptors (Afm, Afl) could not be activated even at high ligand concentrations. The nine receptors that were responsive in the screen could be reproducibly activated with their cognate ligands and showed maximal fold-activation values similar to the wild-type receptor. Surprisingly, several of these receptors showed sensitivities several orders of magnitude higher than the wild-type receptor (Figure 4.5b). However, since our reporter strain still expresses the alpha-factor protease Bar1, the EC_{50} value measured for the wild-type receptor does not represent the true sensitivity of this receptor since a substantial portion of alpha factor will be degraded. Nevertheless, a handful of receptors still

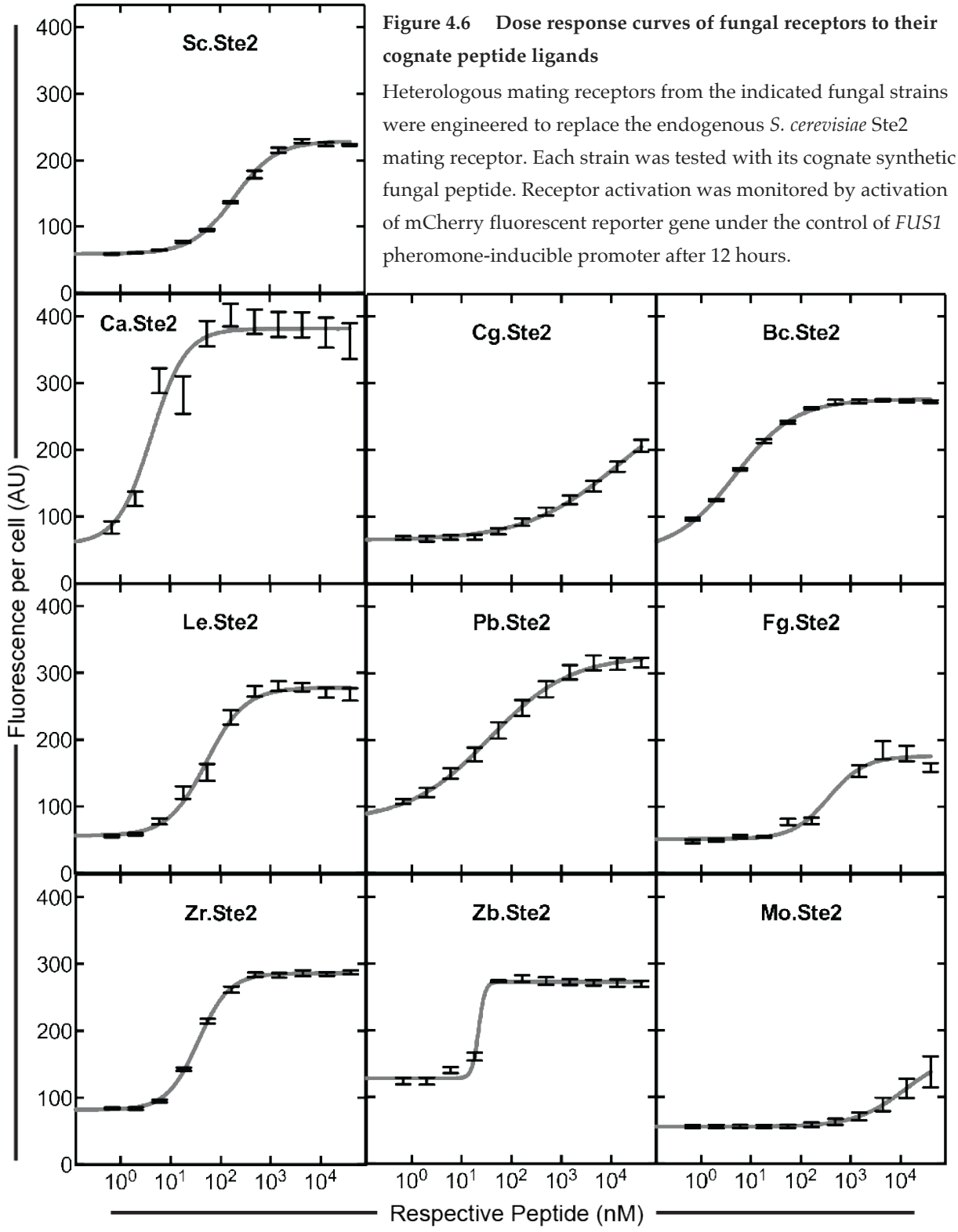


Figure 4.6 Dose response curves of fungal receptors to their cognate peptide ligands

Heterologous mating receptors from the indicated fungal strains were engineered to replace the endogenous *S. cerevisiae* Ste2 mating receptor. Each strain was tested with its cognate synthetic fungal peptide. Receptor activation was monitored by activation of mCherry fluorescent reporter gene under the control of *FUS1* pheromone-inducible promoter after 12 hours.

had higher sensitivities than that previously reported for the wild-type receptor in Bar1 negative strains [33].

To further probe the two non-functional receptors we tested their function against alternate peptides predicted from their putative pheromone genes. As noted in Section 4.2.2, the pheromone genes from these *Aspergillus* species showed imperfect conservation between repeats of the putative secreted peptides. Therefore, it was possible that the sequence of the synthetic peptides originally used to test the receptors from these fungi did not represent the true biologically active ligands. We thus attempted to induce strains carrying these two receptors with three alternate synthetic peptides. We also included the peptide ligand from *P. brasiliensis* since this peptide had high homology to the putative peptides from the *Aspergillus* species. We did not observe any response from these receptors across this array of synthetic peptides (Figure 4.7). This may be due to improper expression of these receptors in our receptor strain, or because the tested peptides do not represent the true ligands of these receptors. Alternatively, as noted earlier, these receptors and pheromone genes may be vestigial in the *Aspergillus* species.

Interestingly, the control receptor from *P. brasiliensis* showed small but measurable responses to some of the peptides included in the *Aspergillus* array. While it was only maximally activated by its cognate ligand, the Pb receptor responded to four of the five alternate peptides used. The Pb

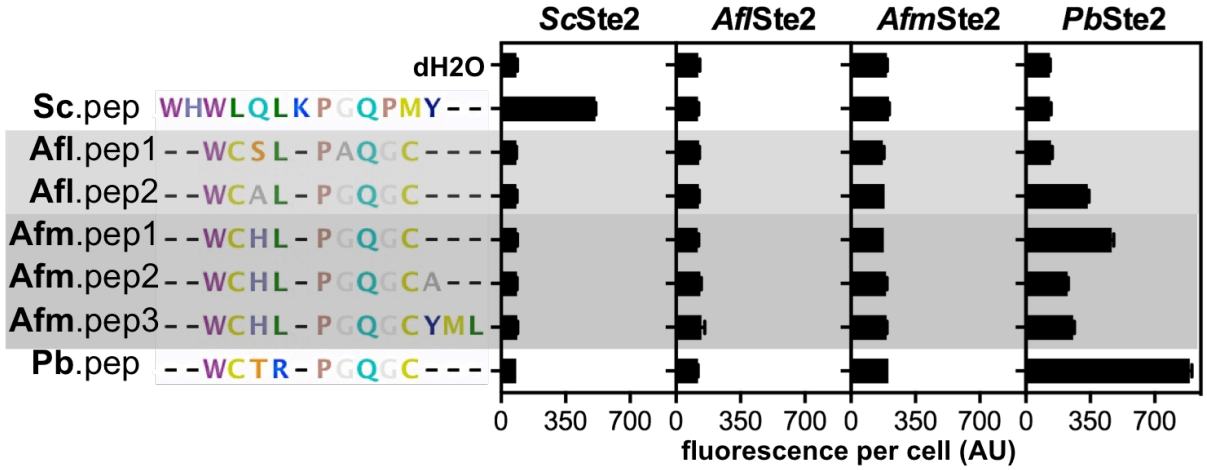


Figure 4.7 Characterization of non-functional receptors from *Aspergillus* species
Several putative peptides can be predicted from the genomes of *Aspergillus flavus* (Afl.pep1,2) and *Aspergillus fumigatus* (Afm.pep1,2,3). However, even at high concentrations the peptides do not activate the predicted receptors. Fluorescence measured at 24 hours after induction with 100 uM of each peptide.

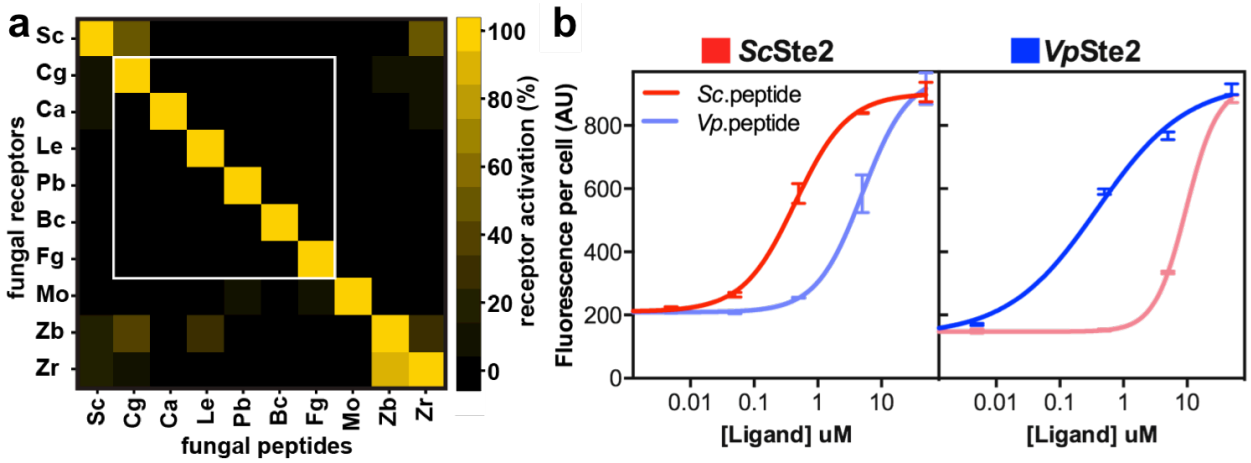


Figure 4.8 Fungal peptide-receptor pairs are exquisitely specific

(a) Specificity of heterologous fungal receptors. Receptors were activated by 5 μM of synthetic peptides and the response was measured by fluorescence and normalized to the maximum signal for each receptor. The white box marks a set of six mutually orthogonal receptors. **(b)** Analysis of cross reacting receptors from *S. cerevisiae* and *V. polyspora* reveals that cross reacting receptors retain preference for their cognate ligand (bold colors).

receptor even showed a response to a peptide with two internal substitutions and a three residue C-terminal extension relative to the cognate Pb peptide. However, the Pb receptor showed no response to a peptide with just three internal mutations and of the same length as the cognate Pb peptide.

The rich variety in the response of the Pb receptor to non-cognate peptides led us to ask whether the other fungal receptors also had overlapping ligand specificities. We therefore tested the cross-reactivity of set of eight receptors with unique peptide ligands. Surprisingly, this showed that a majority of the receptors were exquisitely specific for their own ligand showing no response to non-cognate ligands (Figure 4.8a). We further characterized the sensitivity of two receptors that showed cross-reactivity (Figure 4.8b). Even in these cases of cross-reactivity, the receptors showed a strong preference in sensitivity to their cognate peptides by at least one order of magnitude. In general, receptors that cross-reacted against non-cognate ligands had a sequence homology to the cognate receptors of the cross-reacting ligands of at least 30% or more (Figure 4.9).

Within the small set of receptor-peptide pairs characterized here, we uncovered six mutually orthogonal pairs that represent an extensible set of communication modules for synthetic biology (white box in Figure 4.8a). Significantly, five of these orthogonal receptors had ligand sensitivities better than 100 nM with no additional engineering required. Furthermore, the characteristic

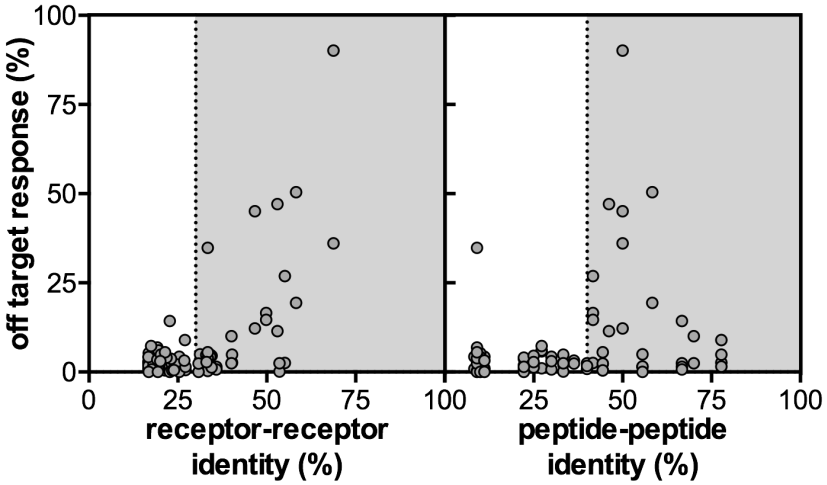


Figure 4.9 Sequence homology predicts response orthogonality
Receptors with higher homology to each other tend to respond to the respective off-target ligands (left). However, some peptides can be highly homologous and still not lead to off-target activation (right). Grey regions mark % identity above which cross-activation is expected.

sequence homology to orthogonality relationships observed suggest that receptors with sequence homologies lower than 30% to each other and ligands less than 40% homologous are likely to have orthogonal ligand specificities (unshaded regions in Figure 4.9). Additionally, even receptors with sequence homology of lower than 30% relative to the wild-type Ste2 receptor coupled robustly to the pheromone response in *S. cerevisiae*. Therefore, it is likely that many more mutually orthogonal receptor-peptide pairs can be found in the set of receptors described here.

4.3 Discussion

In summary, we catalogued a set of over 100 GPCR-peptide pairs. From a subset of 25 receptors we found that 22 were functionally expressed in yeast. Further characterization of 10 of these receptors showed that 7 of them were orthogonal, with a subset of 6 being exquisitely orthogonal. This set already represent a drastic increase over available communication systems such as bacterial quorum sensing which to date have only yielded two fully orthogonal communication modules [6]. Importantly, the set of GPCR-peptide pairs directly characterized here only represents a small portion of the available pairs available in the larger genome mining dataset. Furthermore as more fungal genomes are sequenced, the set of potential GPCR-peptide pairs will continue to grow.

While we have demonstrated that most of the fungal GPCRs can be functionally expressed in yeast, it still remains to be seen how many of the corresponding peptide ligands can be successfully secreted. We expect that simple replacement of the wild-type pheromone sequence in the MFalpha1/2 genes will result in secretion of a majority of peptides. In fact, other efforts to secrete heterologous protein from yeast have used the MFalpha1 prepro secretion signal successfully [56–

58]. Taken together, the results described here validate the idea that fungal mating GPCRs and their peptide ligands have the potential to become a flexible and extensible synthetic communication language.

4.4 References

1. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489 (2008).

2. Chuang, J. S. Engineering multicellular traits in synthetic microbial populations. *Curr. Opin. Chem. Biol.* **16**, 370–378 (2012).

3. Mee, M. T. & Wang, H. H. Engineering ecosystems and synthetic ecologies. *Mol. BioSyst.* **8**, 2470–2483 (2012).

4. Pai, A., Tanouchi, Y., Collins, C. H. & You, L. Engineering multicellular systems by cell–cell communication. *Curr. Opin. Biotechnol.* **20**, 461–470 (2009).

5. Bacchus, W. & Fussenegger, M. Engineering of synthetic intercellular communication systems. *Metab. Eng.* **16**, 33–41 (2013).

6. Hennig, S., Rödel, G. & Ostermann, K. Artificial cell-cell communication as an emerging tool in synthetic biology applications. *J. Biol. Eng.* **9**, 13 (2015).

7. Wintermute, E. H. & Silver, P. A. Dynamics in the mixed microbial concourse. *Genes Dev.* **24**, 2603–2614 (2010).

8. Alberts, B. *et al.* General Principles of Cell Communication. (2002).at <https://www.ncbi.nlm.nih.gov/books/NBK26813/>

9. Perrimon, N., Pitsouli, C. & Shilo, B.-Z. Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning. *Cold Spring Harb. Perspect. Biol.* **4**, a005975 (2012).

10. Jabbour, H. N., Kelly, R. W., Fraser, H. M. & Critchley, H. O. D. Endocrine Regulation of Menstruation. *Endocr. Rev.* **27**, 17–46 (2006).

11. Weber, W., Baba, M. D.-E. & Fussenegger, M. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl. Acad. Sci.* **104**, 10435–10440 (2007).

12. Kim, H. J., Boedicker, J. Q., Choi, J. W. & Ismagilov, R. F. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proc. Natl. Acad. Sci.* **105**, 18188–18193 (2008).

13. Chen, Y., Kim, J. K., Hirning, A. J., Josić, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Science* **349**, 986–989 (2015).

14. Hooshangi, S. & Bentley, W. E. From unicellular properties to multicellular behavior: bacteria quorum sensing circuitry and applications. *Curr. Opin. Biotechnol.* **19**, 550–555 (2008).

15. Bassler, B. L. & Losick, R. Bacterially Speaking. *Cell* **125**, 237–246 (2006).

16. Fuqua, C., Parsek, M. R. & Greenberg, E. P. Regulation of Gene Expression by Cell-to-Cell Communication: Acyl-Homoserine Lactone Quorum Sensing. *Annu. Rev. Genet.* **35**, 439–468 (2001).

17. Collins, C. H., Leadbetter, J. R. & Arnold, F. H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24**, 708–712 (2006).

18. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci.* **104**, 1877–1882 (2007).

19. Ortiz, M. E. & Endy, D. Engineered cell-cell communication via DNA messaging. *J. Biol. Eng.* **6**, 16 (2012).

20. Bacchus, W. *et al.* Synthetic two-way communication between mammalian cells. *Nat. Biotechnol.* **30**, 991–996 (2012).

21. Wang, W.-D., Chen, Z.-T., Kang, B.-G. & Li, R. Construction of an artificial intercellular communication network using the nitric oxide signaling elements in mammalian cells. *Exp. Cell Res.* **314**, 699–706 (2008).

22. Carvalho, A. *et al.* Genetically Encoded Sender–Receiver System in 3D Mammalian Cell Culture. *ACS Synth. Biol.* **3**, 264–272 (2014).

23. Matsuda, M., Koga, M., Nishida, E. & Ebisuya, M. Synthetic Signal Propagation Through Direct Cell-Cell Interaction. *Sci Signal* **5**, ra31-ra31 (2012).

24. Williams, T. C., Nielsen, L. K. & Vickers, C. E. Engineered Quorum Sensing Using Pheromone-Mediated Cell-to-Cell Communication in Saccharomyces cerevisiae. *ACS Synth. Biol.* **2**, 136–149 (2013).

25. Chen, M.-T. & Weiss, R. Artificial cell-cell communication in yeast Saccharomyces cerevisiae using signaling elements from Arabidopsis thaliana. *Nat. Biotechnol.* **23**, 1551–1555 (2005).

26. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33**, 377–383 (2015).

27. Hennig, S., Clemens, A., Rödel, G. & Ostermann, K. A yeast pheromone-based inter-species communication system. *Appl. Microbiol. Biotechnol.* **99**, 1299–1308 (2014).

28. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26**, 339–350 (2005).

29. Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207–211 (2011).

30. Bender, A. & Sprague Jr., G. F. Yeast peptide pheromones, a-factor and α-factor, activate a common response mechanism in their target cells. *Cell* **47**, 929–937 (1986).

31. Nakayama, N., Miyajima, A. & Arai, K. Common signal transduction system shared by STE2 and STE3 in haploid cells of Saccharomyces cerevisiae: autocrine cell-cycle arrest results from forced expression of STE2. *EMBO J.* **6**, 249–254 (1987).

32. Mayrhofer, S. & Pöggeler, S. Functional Characterization of an α-Factor-Like Sordaria macrospora Peptide Pheromone and Analysis of Its Interaction with Its Cognate Receptor in Saccharomyces cerevisiae. *Eukaryot. Cell* **4**, 661–672 (2005).

33. Janiak, A. M. *et al.* Functional expression of the Candida albicans α-factor receptor in Saccharomyces cerevisiae. *Fungal Genet. Biol.* **42**, 328–338 (2005).

34. Gomes-Rezende, J. A. *et al.* Functionality of the Paracoccidioides Mating α-Pheromone-Receptor System. *PLoS ONE* **7**, e47033 (2012).

35. Fitzpatrick, D. A., Logue, M. E., Stajich, J. E. & Butler, G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* **6**, 99 (2006).

36. Wang, H., Xu, Z., Gao, L. & Hao, B. A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol. Biol.* **9**, 195 (2009).

37. Xue, C., Hsueh, Y.-P. & Heitman, J. Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. *FEMS Microbiol. Rev.* **32**, 1010–1032 (2008).

38. Ene, I. V. & Bennett, R. J. The cryptic sexual strategies of human fungal pathogens. *Nat. Rev. Microbiol.* **12**, 239–251 (2014).

39. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).

40. Clark, C. D., Palzkill, T. & Botstein, D. Systematic mutagenesis of the yeast mating pheromone receptor third intracellular loop. *J. Biol. Chem.* **269**, 8831–8841 (1994).

41. Ćelić, A. *et al.* Sequences in the Intracellular Loops of the Yeast Pheromone Receptor Ste2p Required for G Protein Activation†. *Biochemistry (Mosc.)* **42**, 3004–3017 (2003).

42. Martin, S. H., Wingfield, B. D., Wingfield, M. J. & Steenkamp, E. T. Causes and Consequences of Variability in Peptide Mating Pheromones of Ascomycete Fungi. *Mol. Biol. Evol.* **28**, 1987–2003 (2011).

43. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

44. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

45. Madden, T. *The BLAST Sequence Analysis Tool*. (National Center for Biotechnology Information (US), 2013).at <https://www.ncbi.nlm.nih.gov/books/NBK153387/>

46. Carey, P. R. *Protein Engineering and Design*. (Academic Press, 1996).

47. Julius, D., Blair, L., Brake, A., Sprague, G. & Thorner, J. Yeast α factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* **32**, 839–852 (1983).

48. Julius, D., Schekman, R. & Thorner, J. Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell* **36**, 309–318 (1984).

49. Pöggeler, S. Two pheromone precursor genes are transcriptionally expressed in the homothallic ascomycete Sordaria macrospora. *Curr. Genet.* **37**, 403–411

50. Bobrowicz, P., Pawlak, R., Correa, A., Bell-Pedersen, D. & Ebbole, D. J. The Neurospora crassa pheromone precursor genes are regulated by the mating type locus and the circadian clock. *Mol. Microbiol.* **45**, 795–804 (2002).

51. Darby, N. J. & Smyth, D. G. Endopeptidases and prohormone processing. *Biosci. Rep.* **10**, 1–13 (1990).

52. Lee, B.-K., Khare, S., Naider, F. & Becker, J. M. Identification of Residues of the Saccharomyces cerevisiae G Protein-coupled Receptor Contributing to α-Factor Pheromone Binding. *J. Biol. Chem.* **276**, 37950–37961 (2001).

53. Lin, J. C., Parrish, W., Eilers, M., Smith, S. O. & Konopka, J. B. Aromatic Residues at the Extracellular Ends of Transmembrane Domains 5 and 6 Promote Ligand Activation of the G Protein-Coupled α-Factor Receptor†. *Biochemistry (Mosc.)* **42**, 293–301 (2003).

54. Sturme, M. H. J. *et al.* Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie Van Leeuwenhoek* **81**, 233–243

55. Tytgat, J. *et al.* A unified nomenclature for short-chain peptides isolated from scorpion venoms: α-KTx molecular subfamilies. *Trends Pharmacol. Sci.* **20**, 444–447 (1999).

56. Baldari, C., Murray, J. A., Ghiara, P., Cesareni, G. & Galeotti, C. L. A novel leader peptide which allows efficient secretion of a fragment of human interleukin 1 beta in Saccharomyces cerevisiae. *EMBO J.* **6**, 229–234 (1987).

57. Bourbonnais, Y., Bolin, D. & Shields, D. Secretion of somatostatin by Saccharomyces cerevisiae. Correct proteolytic processing of pro-alpha-factor-somatostatin hybrids requires the products of the KEX2 and STE13 genes. *J. Biol. Chem.* **263**, 15342–15347 (1988).

58. Brake, A. J. *et al.* Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4642–4646 (1984).

5 RATIOMETRIC FLUORESCENT CELLTAGS

5.1 Introduction

Methods to analyze synthetic multicellular systems have lagged behind our ability to build these systems [1,2]. Available analysis methods are limited in either information bandwidth or time resolution [3]. Therefore only simple multicellular systems have been successfully implemented and characterized. Methods with high information bandwidth such as next generation sequencing are severely limited in their ability to be applied with high time resolution. Therefore, while next generation sequencing can characterize a large number of co-existing cell types, only endpoint measurements are practical [4]. Conversely, methods with high time resolution such as microscopy are limited in their ability to track more than just a few constituent cell types [5,6]. In the mean time, theoretical models of multicellular systems have increased in sophistication and have generated tantalizing predictions of the potential capabilities of synthetic multicellular systems [7–9]. However, without better analysis methods, these models are difficult to validate and improve. An ideal method should enable straightforward interrogation of many constituent cell types (10+) in real time and through non-destructive means [1,7].

In this chapter we develop a new class of fluorescent tags (CellTags) for tracking mixed cell population by flow cytometry. CellTags are based on precise ratios of two fluorescent proteins and enable the simultaneous characterization of up to 20 different subpopulations. Importantly, CellTags use only two fluorescent protein colors, reserving a third color for use as a reporter. We apply these tags for tracking mixed populations of competing cells and for multiplex profiling of yeast transcriptional programs.

5.1.1 Current methods for characterizing multicellular systems

Several methods have been developed to characterize natural multicellular systems, with each differing in terms of through put, mode of interrogation, and time resolution. While powerful for characterizing the heterogeneity of natural systems, none of these methods can be immediately used for the non-destructive characterization of mixed populations in real time with out the need of

external tagging reagents. Furthermore a majority of studies on synthetic multicellular systems continue to use labor intensity colony counting on selective media to characterize population structure .

Methods based on next generation sequencing have incredibly high throughput and can be used to distinguish $>10^4$ cell types at once [4]. However, these methods suffer from low time resolution and the destructive nature of the interrogation. Quantitative PCR based methods can distinguish 10 cell types although this number depends on the ability to obtain robustly performing barcodes [10]. Quantitative PCR methods can be performed at higher time resolution, however they also suffer from the destructive nature of the interrogation.

As an alternative to DNA-based methods, microscopy and flow cytometry use fluorescence based tags to distinguish cell types [6,11]. These methods have very high time resolution; with microscopy able to record continuously and flow cytometry have a time resolution of 5 minutes [12]. However, the throughput of cell types that can be distinguished is limited by the availability of fluorescent tags [13]. When interrogating natural systems, these methods rely on fluorescent antibody-based stains that can distinguish 10 or more cell types through the combinatorial staining signature of each cell type [14]. More recent approaches combine flow cytometry with mass spectrometry and use of antibodies labeled with heavy atoms to increase the number of resolvable cell types [15].

However, these methods are hindered by the need to use expensive labeling reagents and lengthy labeling procedures. An alternative approach is to use expressed fluorescent proteins as tags, however currently available fluorescent proteins enable only the characterization of at most 4 cell types when used individually [16,17]. In the field of neurobiology, cell-tagging systems based on mixtures of fluorescent proteins have elegantly overcome this challenge [18,19]. However, these systems rely on random generation of the mixtures and use three or more fluorescent proteins. Therefore, they are not directly portable for the deterministic characterization of synthetic multicellular systems. Nevertheless, these stoichiometry-based tags provide a good intellectual starting point for designing a tag system with high information bandwidth and time resolution, while maintaining an economy of parts and labor.

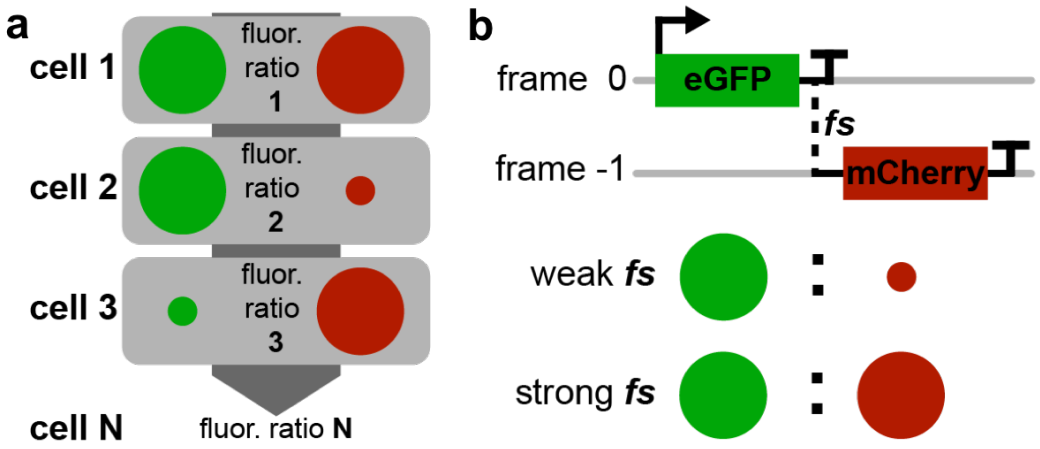


Figure 5.1 Mixed cell populations can be tracked using ratiometric fluorescent tags
(a) Fluorescent proteins (colored circles) with fixed ratios provide a way to track individual cells within a mixed population. Only two fluorescent colors are required to track N different cells as long as the ratios can be resolved from each other. **(b)** Programmed -1 frame shifting (*fs*) sets a specific ratio of two gene products such as red (mCherry) and green (eGFP) fluorescent proteins. Weak frame shifting gives a small ratio of mCherry to eGFP and strong frame shifting gives a larger ratio.

5.2 Results

5.2.1 Lowering noise in fluorescent protein stoichiometry

We envisioned that resolvable tags could be generated based on tightly controlled stoichiometries between two fluorescent proteins within the cell. Different relative stoichiometries of two proteins can be directly achieved by expressing these proteins from different strength promoters on a plasmid [20]. However, these systems suffer from high extrinsic and intrinsic noise and generally have wide distributions in the total protein levels as well as in the relative ratios between the two proteins [21,22]. Therefore, previous groups primarily rely simply on the presence or absence of a fluorescent protein to distinguish between mixed populations [12,16]. Thus, to make a large palette of ratiometric fluorescent tags we first sought to lower intrinsic noise by controlling relative protein levels with frame shift motifs and extrinsic noise by genome integration

To reduce intrinsic noise at the transcriptional and translational level we were inspired by a recently developed set of small RNA motives that induce programed -1 ribosomal frameshifting [23]. When placed between two open reading frames, these frameshift (*fs*) motifs were shown to produce a range of ratios between the two protein products [23]. By expressing both fluorescent proteins from a single mRNA we can avoid intrinsic noise introduced by stochastic differences in the

upstream activation pathway, transcription factors levels, rate of transcription initiation and mRNA stability which affect proteins expressed from different promoters [24]. Additionally, since these *fs* motifs set the relative ratio co-translationally we also avoid intrinsic noise due to differences in the rate of translation initiation observed between proteins controlled by different ribosome entry sites [25]. Indeed, when we placed one of these *fs* motifs between two distinct fluorescent proteins, we observed reduced variation in the relative ratio between the two proteins relative to a construct expressing these proteins from different promoters (yellow line in Figure 5.2 a,b).

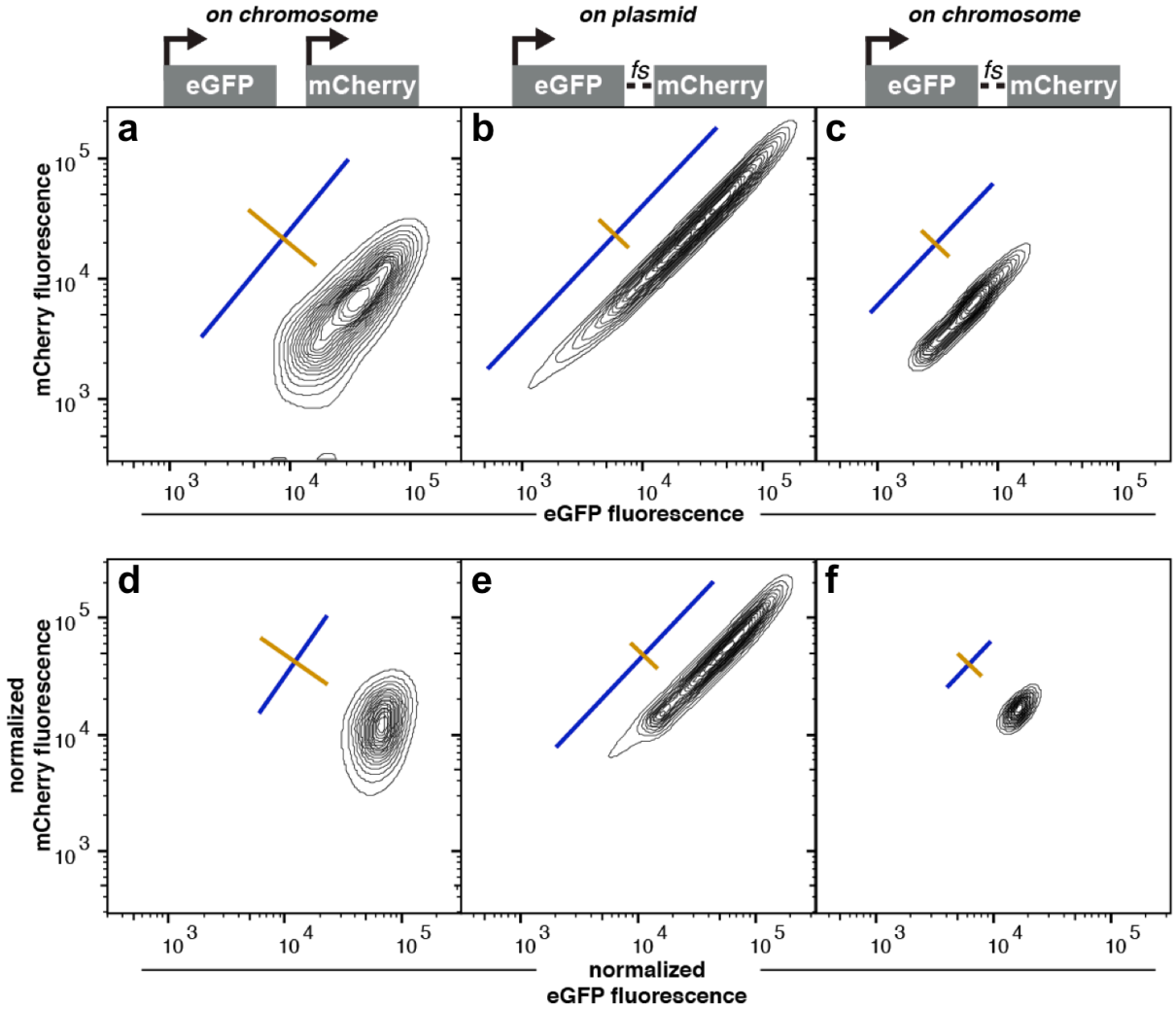


Figure 5.2 Intrinsic and extrinsic noise in fluorescent protein expression *in vivo*.
Flow cytometry analysis of the distribution of single cell expression of two fluorescent proteins (eGFP, mCherry). Black contour lines mark 5% density levels. **(a,d)** FPs expressed from two separate promoters on the chromosome. **(b,e)** FPs expressed from one frameshift (*fs*) construct on a plasmid. **(c,f)** FPs expressed from one *fs* construct on the chromoseme. **(a,b,c)** raw fluorescence or **(d,e,f)** fluorescence normalized to side scatter. Colored lines mark relative scale of intrinsic (yellow) and extrinsic (blue) noise.

To reduce extrinsic noise we designed the tag constructs to be readily integrated into the chromosome. Variation in the absolute protein levels between cells is caused primarily by differences in cell size and DNA copy number [22]. Variations in cell size can be easily corrected by normalizing to scatter parameters in flow cytometry that are known to correlate with cell size (blue lines, top vs. bottom rows in Figure 5.2) [12]. Variation in plasmid copy number is significant in yeast, even for centromeric plasmids that range from 2 to 5 copies per cell [26]. Thus, integration of the constructs into the genome resulted in a drastic reduction in the distribution of absolute fluorescence compared to the same construct expressed from a centromeric plasmid (blue lines in Figure 5.2 e,f).

5.2.2 Constructing a large palette of ratiometric fluorescent tags

With a low noise design in hand, we next screened a library of *fs* motifs to find a mutually resolvable set. The *fs* motif library was inserted between a green fluorescent protein (eGFP) and a red fluorescent protein (mCherry) and screened for a range of mCherry : eGFP ratios. We found a set of 9 *fs* motifs that that lead to mCherry fluorescence spanning more than two orders of magnitude and gave relative mCherry : eGFP ratios from 0.3 to 100 percent (Figure 5.3 b). We chose a subset of 5 *fs* motifs that gave baseline separation between populations both for total mCherry fluorescence and mCherry : eGFP ratios (Figure 5.3 c).

We next designed a modular dual *fs* dual FP construct to generate a large combinatorial palette of

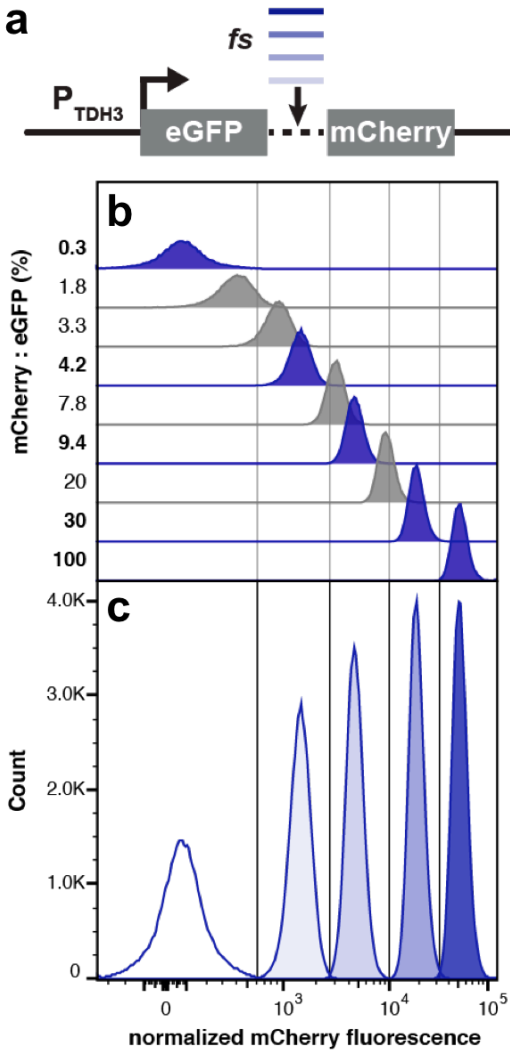


Figure 5.3 A set of resolvable frame shift constructs
(a) A set of dual FP constructs was made from a library of frameshift motifs (*fs*). **(b)** Flow cytometry distributions in absolute mCherry expression and ratios (%) relative to the upstream eGFP. **(c)** A subset is resolvable with baseline separation of the distributions.

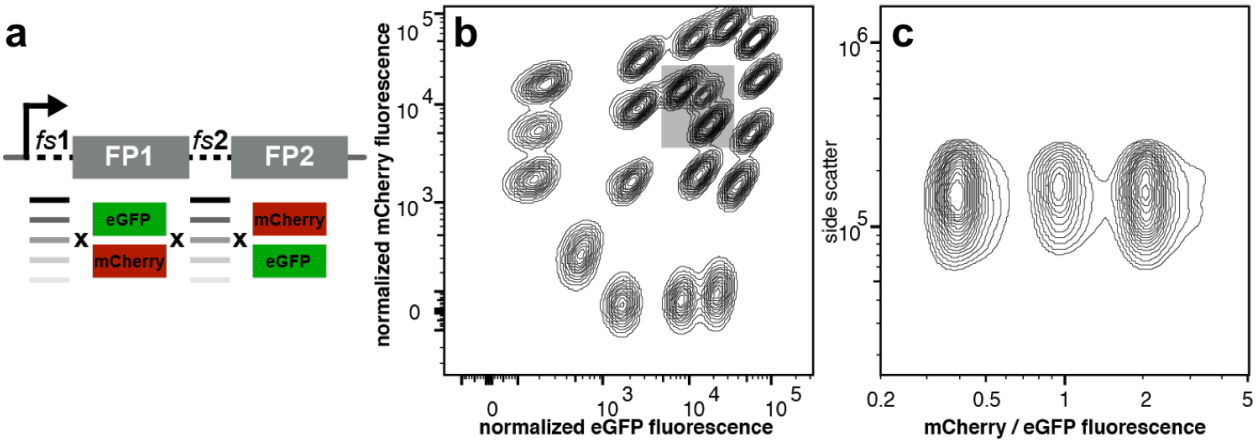


Figure 5.4 A large palette of fluorescent CellTags

(a) Design of a dual frameshift (*fs*) dual fluorescent protein (FP) construct to make a pallet of CellTags, combining 5 *fs* motifs and 2 FPs. **(b)** Single cell fluorescence distributions of a mixture of populations marked with 20 CellTag. Contour lines mark 5% density levels. **(c)** Populations in grey box in panel b plotted by ratio of mCherry to eGFP fluorescence. Side scatter is used as a second dimension to calculate 5% contour lines.

ratiometric fluorescent tags (Figure 5.4 a). Using 5 *fs* motifs combined with two possible *fs* arrangements, this construct can generate 45 theoretically distinct FP stoichiometries. However, since the absolute level of the second open reading frame is a result of the multiplied frameshift efficiencies of both *fs* motifs that precede it, we expected that combinations of two low efficiency *fs* motifs would result in a level indistinguishable from background cell fluorescence. Even with this limitation, we constructed a total of 20 resolvable ratiometric fluorescent CellTags (Figure 5.4 b). Even closely spaced CellTags had distinct 95% density contour levels when plotted against their mCherry : eGFP ratios (Figure 5.4 c). We note that the remaining spectral space could accommodate an additional 7 tags.

It is interesting to note that the *fs* motifs have unexpected effects beyond simple control of the downstream open reading frame. First, we observed that as the frameshift efficiency of the *fs* motif is decreased the absolute level of a FP *preceding* the *fs* motif also decreases (Figure 5.5 a). Nonsense mediated decay (NMD) of the mRNA is a likely cause, since constructs with lower efficiency *fs* motifs are expected to present a higher portion ribosomes at premature termination codons (PTCs) leading to NMD and an overall decrease in expression from the construct [27]. We also observed that an "early" *fs* motif close to the 5′ end of the mRNA leads to a reduction in expression relative to the expression level from the same *fs* motif placed "late" in the middle of the mRNA (Figure 5.5 b). This

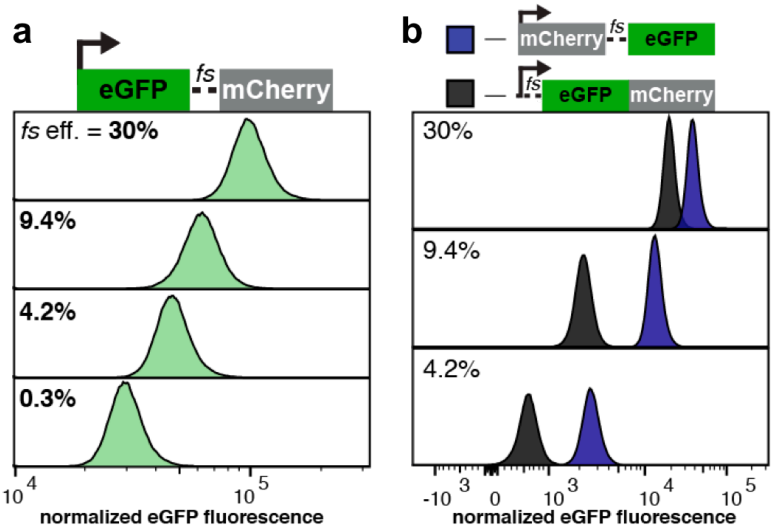


Figure 5.5 Frameshift motifs have secondary effects on protein levels
(a) Fluorescence distributions of eGFP show that frameshift motifs (*fs*) also affect the expression of the upstream open reading frame. Efficiency of *fs* is calculated as mCherry/eGFP signal. **(b)** Comparison of eGFP fluorescence controlled by late (purple) or early (black) *fs* motifs. The same *fs* motifs are used as in panel a, labeled by their efficiency.

effect is also likely caused by NMD. It is well established that PTCs close to the 5′ end of an mRNA cause more NMD than PTCs closer to the 3′ end of an mRNA [27]. CellTag constructs with early *fs* motifs have an early PTC that is expected to lower mRNA levels compared to constructs with late *fs* motifs leading to the effect observed. One possible confounding factor is that the expressed protein is different (eGFP fused after or before mCherry) and so may lead to different protein levels due to non-mRNA related effects such as protein stability. However, we observe a symmetry of effects regardless of which FP fluorescence is analyzed. This points to an mRNA-based mechanism. Overall, the combination of these effects leads to a puckered arrangement of the CellTags (Figure 5.4 b). Importantly, these potential NMD effects do not detract from our ability to robustly distinguish the CellTags from each other. In fact, these NMD effect may be exploited to generate additional tags. For example, insertion of a leading "dummy" open reading frame into all the constructs with an early *fs* may generate intermediate protein stoichiometries that result in CellTags that could be accommodated into the currently available spectral space.

An important feature of our system is that it is based on just two fluorescent proteins. Therefore, a third fluorescent protein can be used as an orthogonal reporter. To determine which third fluorescent protein would be most appropriate, we tested four additional FPs: mTagBFP2, mTurquoise2, mVenus and mKO2 [28–31]. Each additional FP was placed under control of a galactose-inducible promoter and expressed in conjunction with a CellTag with intermediate levels of eGFP and mCherry. By comparing the effect of the third FP on the CellTag between induced and

uninduced conditions we found that mTagBFP2 (BFP) was an ideal third FP with no impact on the CellTag signal.

We also note that our CellTag design could be easily extended in the future (Figure 5.6). In addition to its use as an orthogonal reporter, the third FP can instead be used to generate even more CellTags. By simply replacing either GFP or RFP in the validated dual *fs* dual FP constructs we could generate 29 additional tags for a total of 49 CellTags. Furthermore, by constructing triple *fs* triple FP constructs we could access the full three dimensional (RFP, GFP, BFP) space to generate at least 80 total CellTags. These triple FP tags could in turn be combined with recently reported infra red FPs that are expected to be orthogonal to the three FPs currently used in our constructs [32,33].

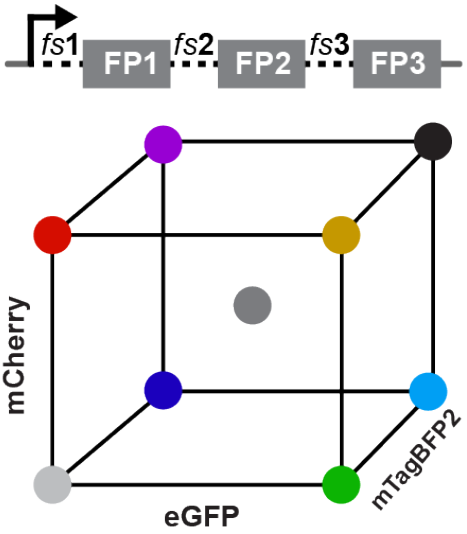


Figure 5.6 A triple FP tag
The three fluorescent proteins (FPs) mCherry, eGFP and mTagBFP2 can be used to make triple *fs* triple FP tags. This expands the possible number of resolvable CellTags.

5.2.3 An automated gating method for high throughput analysis of tagged cells

In order to take full advantage of the CellTag palette we designed an automated high throughput flow cytometry analysis pipeline. Flow cytometry data is commonly analyzed using programs such as FlowJo (Tree Star, http://www.flowjo.com). While they are powerful tools for discovering relationships in single cell data, they are cumbersome to apply for the analysis of multiplexed CellTag data since they require manual gating. Although we initially implemented fixed gates to deconvolute CellTag signals, we observed that the precise location of the CellTags shifted depending on growth conditions. These shifts were on the scale of the peak-to-peak distance between the CellTags and led to misaligned gates that often required manual adjustment. Importantly, these shifts always affected all the CellTags in concert leaving the relative spacing between CellTags preserved. We thus envisioned that peak-finding algorithms could be implemented to automatically and reproducible extract individual CellTag signals.

We built this automated gating pipeline around openCyto, an R package that enables programmatic processing and analysis of flow cytometry data [34]. Although there are two-

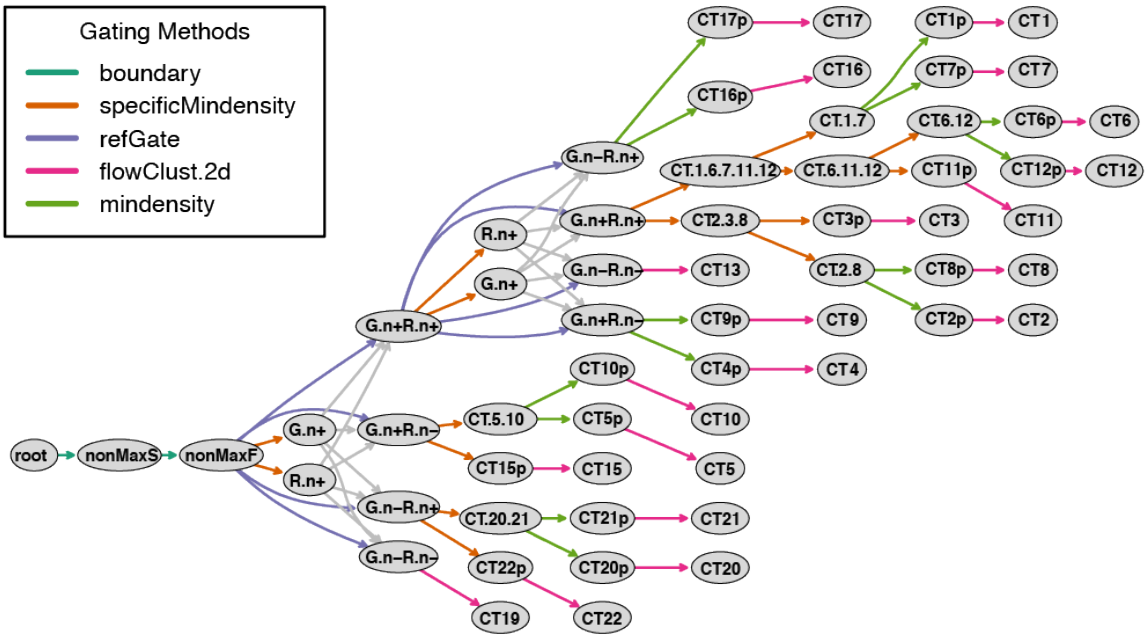


Figure 5.7 Automated hierarchical gating of the CellTags

An automated gating routine was developed base on the openCyto R package. Starting from a "root" population, each CellTag (CT#) is isolated through hirerarchical subdivisions (left to right) based on simple 1-dimensional gating methods (arrows). Each final population is gated in two dimensions (pink, flowClust.2d) to capture only cells in the 80% quantile of the peak and excluding cells in peak overlap regions.

dimensional multi-cluster prediction algorithms that can locate the CellTags in a single calculation [35], we found that these were computationally costly when applied to a 20-member cluster. Instead, we developed a hierarchical subdivision routine based around fast one-dimensional peak finding algorithms (Figure 5.7 and Appendix). This approach takes advantage of the sharp peaks and valleys characteristic of the CellTags as well as the predictable superposition between CellTags along one-dimensional histograms. Importantly, this analysis pipeline makes no assumption about the absolute fluorescence values in the data and is therefore unaffected by differences due to instrument parameters. Instead, the subdivision routine relies on the known peak/valley order.

Since not all experiments make use of the full set of 20 CellTags, we also implemented an automated gating template generator. In standard use, the openCyto tool requires an input "gating template" that is applied to gate the flow cytometry data [34]. To facilitate the use of our CellTag analysis pipeline, we included a subroutine to automatically generate the required gating template. This builds a gating template based on a user-provided list of the CellTags present in the samples.

By using this automated analysis tool in R, we were able to process high throughput samples containing over 4000 data points (220 samples x 20 CellTags) in just a few minutes. The data is output as graphs for visual inspection or in CSV format for further analysis.

5.2.4 Tracking mixed populations of competing cells

To experimentally validate the CellTags, we applied them to characterize mixed populations of cells competing for a shared resource. We expected that the CellTags would enable direct interrogation of how such mixed communities of cells evolve over time and how they are affected by initial conditions. We chose a simple system based around strains with different growth rates in a selective media. When grown in a mixed culture these cells have different abilities to compete for the shared carbon source (glucose) with fast growing strains expected to dominate over slow growing strains. Such a predictable outcome allowed us to validate the robustness of the CellTags.

We built this system of strains from the well-characterized yeast-three hybrid reporter strain, MaV203 [36]. This strain contains a URA3 reporter gene under the control of a Gal4-inducible

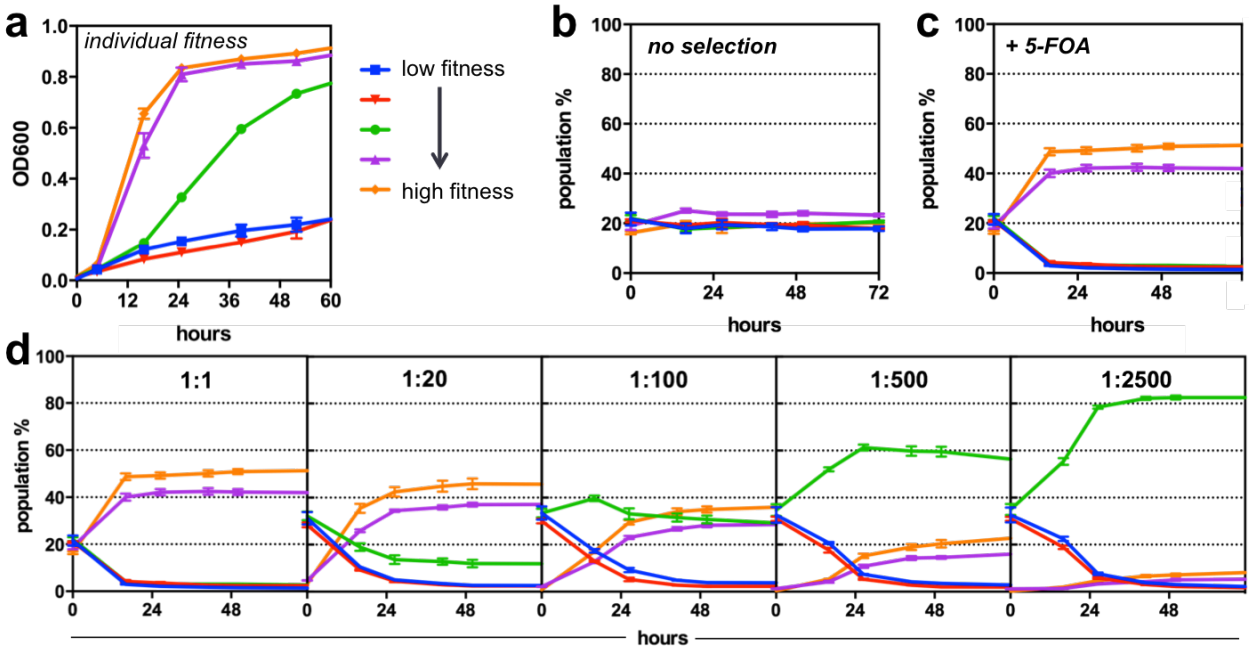


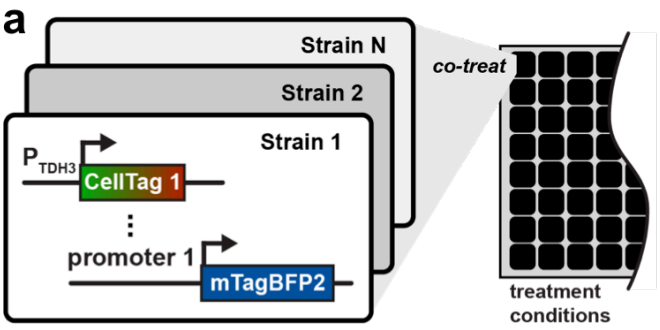
Figure 5.8 CellTags can be used to track competing populations of cells
(a) Individual growth curves of five strains of yeast with varying degrees of fitness in a 5-FOA selection. These strains were tagged and grown as a mixture in either **(b)** non-selective conditions or **(c)** in the 5-FOA selction. Population percent for each strain was determined by flow cytometry of the mixture and counting the number of cells in each CellTag cluster. **(d)** Titration of the two fittest strains relative to the total number of cells.

promoter. When grown in counter selective media containing 5-fluorouracil (5-FOA), the growth rate of the strain depends on the level of URA3 induction, with high induction leading to slow growing strains and vice versa. Thus we generated a set of strains with a variety of fitnesses by introducing constructs that produced different fixed levels of Gal4-mediated induction. These Gal4 activation constructs have been previously described and consist of translationally set ratios of active Gal4 (with an activation domain) and inactive Gal4 (lacking an activation domain) [23]. As with our CellTags, these ratios are set with frameshift motifs. We initially built a set of five MaV203 derived strains that were also tagged with the CellTags. We validated that these strains had different growth rates when grown in 5-FOA media, with two fast, one medium and two slow growing strains (Figure 5.8 a).

We next used the CellTags to track the evolution of a mixture of these strains over time. As expected, when grown in non-selective media, we observed maintenance of even population proportions over the entire growth period of the culture (Figure 5.8 b). When grown in selective media, the fast growing strains dominated the community even at early time points. Even though when individually grown the two fast growing strains had nearly identical growth rates, in the mixed competitive environment we observe a significant differentiation between these strains (Figure 5.8 c). This is expected as the slight differences in fitness are compounded over time in exponentially growing cultures.

To probe the characteristics of this system, we titrated the initial fraction of the two fast growing strains relative to the other three strains (Figure 5.8 d). Surprisingly even with a 1:20 starting disadvantage the fast-growing strains dominated the culture by 24 hours. Even with a 1:100 disadvantage the fastest growing strain still held the highest proportion of the population by 48 hours. We observed domination by the medium growth strain only when it had at least a 500:1 advantage over the fast growing cells. Even though the slow growing strains had measurable growth when grown individually, these still were not able to compete in our system at any dilution of the fast growing strains.

These results demonstrate that the CellTag system developed here enables simple tracking of mixed yeast populations with high time resolution. The proportions of the individual



b

CellTag	reporter	response	*continued*		
CT1	HSP12	general stress	CT11	ZRT1	zinc
CT2	SSA1	heat shock	CT12	FET3	iron
CT3	ERO1	ER stress	CT13	ARR3	arsenic
CT4	GPD1	osmostress	CT14	HIS4	starvation
CT5	FUS1	mating	CT15	PRM5	wall stress
CT6	TRX2	redox stress	CT16	DAL5	catabolite
CT7	RNR3	DNA damage	CT17	PMC1	calcineurir
CT8	GAL1	galactose	CT18	OLE1	hypoxia
CT9	HXT1	glucose	CT19	ACT1	control
CT10	CUP1	copper	CT20	TEF1	control

Figure 5.9 CellTags can be used to assay transcription reporters in multiplex
(a) Transcriptional reporters (promoter-BFP) can be assayed in multiplex by indexing the identity of the promoter with a CellTag. The mixture of reporters is co-cultured and co-treated across many conditions. The resulting BFP signals are deconvoluted by flow cytometry using the CellTags. **(b)** List of reporters used.

subpopulations can be extracted during different phases of growth and in different media growth conditions. While we only track 5 populations here, we used the most closely spaced CellTags so it will be possible to apply this system to more complex mixed population experiments.

5.2.5 Multiplex profiling of yeast transcriptional programs

By using a third fluorescent protein reporter, our CellTag system can also be applied to the multiplex profiling of transcriptional programs of the cell. A transcriptional response in the cell can be studied quantitatively by placing a fluorescent protein reporter under the control of a promoter responsive to a particular cellular program of interest [37]. We envisioned that we could multiplex this method by introducing multiple transcriptional reporters into the same strain and indexing them with CellTags. In this modality, each CellTag does not define a unique strain but rather serves to specify the identity of the reporter present in each particular subset of cells (Figure 5.9 a). A mixture of these tagged strains could then be used as a "reporter cocktail" to effectively report on multiple cellular responses simultaneously.

To test this approach, we chose to build a reporter cocktail for several stress and metabolic responses in *S. cerevisiae* (Figure 5.9 b). We chose 18 well-established promoter reporters for general stress (HSP12), heat shock (SSA1), unfolded proteins (ERO1), osmotic stress (GPD1), mating pheromone (FUS1), oxidative stress (TRX2), DNA damage (RNR3), galactose (GAL1), glucose (HXT1), copper toxicity (CUP1), low zinc (ZRT1), low iron (FET3), arsenic toxicity (ARR3), amino acid starvation (HIS4), cell wall stress (PRM5), the nitrogen catabolite response (DAL5), calcineurin

signaling (PMC1) and hypoxia/cobalt (OLE1) [38–55]. We also included 2 promoters from housekeeping genes (ACT1, TEF1) as controls [56,57]. Each of these promoters was put into a BFP fluorescent reporter construct and indexed by transforming into a corresponding CellTag derivative of the laboratory strain FY251.

We then mixed these reporters together to generate a reporter cocktail for characterizing the global stress and metabolic responses of yeast to a variety of conditions. We first tested whether our approach was reproducible by comparing the distributions from two samples of the reporter cocktail allowed to grow independently. While there was great variation in the shape of the distributions between different promoters, the distribution of each promoter across the two samples was nearly identical (Figure 5.10 a).

Next, we tested galactose induction, a system commonly used to control engineered systems in yeast. We observed that while galactose induced its target promoter (GAL1) it also had widespread effects on multiple transcriptional programs. Some were expected such as a decrease signal from the glucose reporter (HXT1) [46], however others were unexpected such as significant signals from the general stress (HSP12), heat shock (SSA1), osmostress (GPD1) and amino acid starvation (HIS4) response (Figure 5.11 a). Such a drastic change in the global transcriptional program might be expected from an organism like yeast for which carbon metabolism is a key part of its life style. When this galactose induction is performed in the laboratory, the media is often supplemented with raffinose a preferred carbon source that does not cause repression of GAL genes [58]. When we exposed the reporter cocktail to a mixture of galactose and raffinose, we did in fact observe that most of the transcriptional response returned to baseline levels while the galactose targets (GAL1, HXT1) were modulated as expected (Figure 5.11 b). Though this system has substantial off-target effects it is still widely used due to the large fold induction observed.

We next characterized the copper induction, another system commonly used to control engineered systems. When exposed to high copper concentrations our reporter cocktail showed the expected signal from the copper reporter (Figure 5.10 b) [47]. In contrast to galactose induction, our reporter cocktail revealed that this induction is quite specific showing almost no change for any of the other transcriptional programs. Surprisingly, we observed that exposure to hydrogen peroxide,

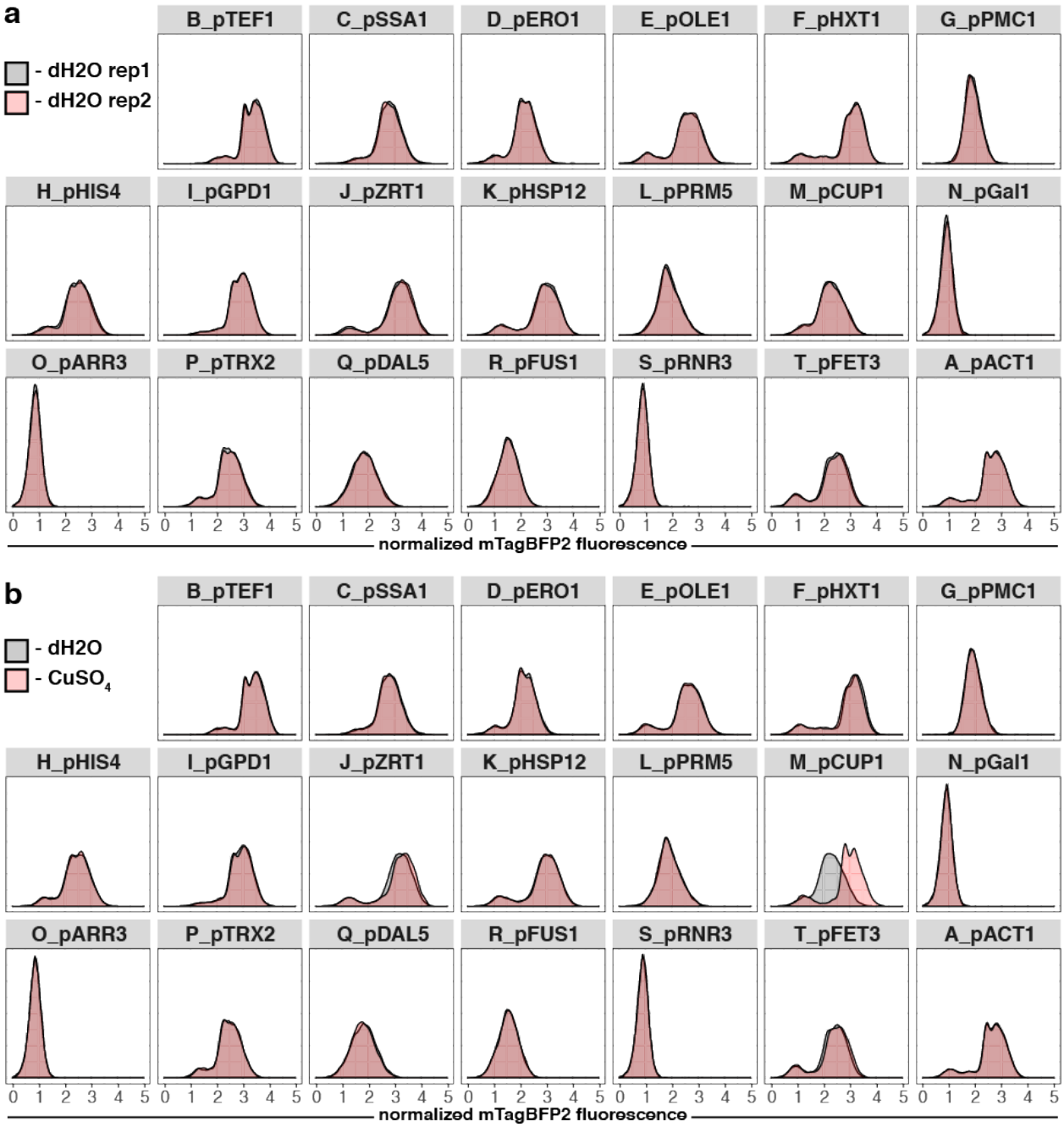


Figure 5.10 CellTags allow recording multiplex reporters robustly and specifically

(a) Reporter mixture was incubated in media for 6 hours. Two replicates show nearly identical BFP distributions for each of the 20 reporters. **(b)** Reporter mixture was treated with water or 400 uM $CuSO_4$ in media for 6 hours. The copper reporter (CUP1) shows a signal while the rest of the reporters remain at baseline.

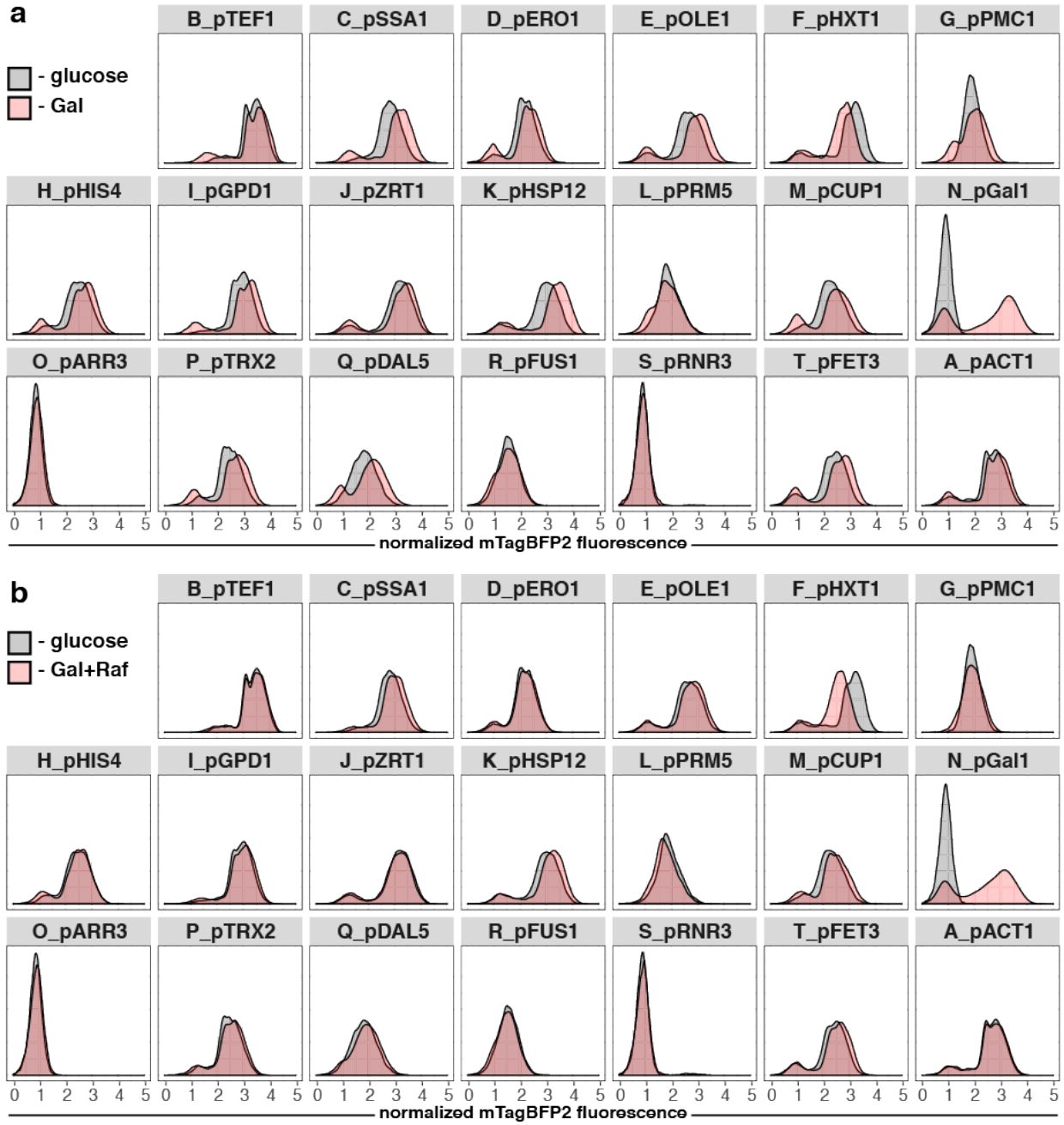


Figure 5.11 Multiplex reporters reveal yeast activates stress responses when grown in galactose
(a) Reporter mixture was incubated in 2% glucose or 2% galactose media for 6 hours. In galactose yeast upregulates pathways for general stress (HSP12), heat shock (SSA1), osmostress (GPD1) and amino acid starvation (HIS4). **(b)** All these stress responses return to baseline when yeast is grown in 2% galactose media supplemented with 2% raffinose. Galactose (GAL1) and low glucose (HXT1) responses are still observed.

also had very specific effect on the cells, leading only to a large increase in the redox stress reporter (TRX2) with some smaller increases noted in the low iron (FET3) and hypoxia/cobalt (OLE1) reporters. Our results with reporter cocktail suggest that the oxidative response in combination with low quantities of hydrogen peroxide could be used as an alternative low cost inducer [59].

When exposed to the alpha factor pheromone our reporter cocktail showed drastic changes across multiple transcriptional programs. This was expected from induction of the pheromone response pathway that causes cell cycle arrest, cell shape remodeling and other widespread phenotypic changes in preparation for mating [60]. In addition we also observed increases in BFP negative cells across all the reporter cell subsets. This is possibly due to the cell cycle arrest, which is expected to decrease the selective pressure for maintaining an auxotrophic growth marker. This phenomenon was also observed on exposure to 5-FOA that also caused growth arrest since all our cell subsets express URA3, which converts 5-FOA into a cytotoxic compound [61]. Therefore as expected from this effect we observed that 5-FOA also caused a large increase in the signal from the DNA damage reporter (RNR3).

Finally, when exposed to FK506 an inhibitor of calcineurin signaling in yeast, we observed several of the expected downstream effects. The calcineurin/calcium reporter (PMC1) showed a decrease in signal [54]. Additionally FK506 is known to inhibit amino acid transport in yeast [62], and as an expected downstream effect of this activity we observed increased signal from the amino acid starvation reporter (HIS4). In addition to these established effects, we also observed a reduction in signal from the cell wall damage (PRM5) and pheromone (FUS1) reporters. FK506 and its effect on calcium flux has in fact been previously linked to both the cell wall damage response and the pheromone response pathways that in turn also modulate calcium levels [63].

Taken together these results largely validate that the CellTag system can function as a robust and sensitive multiplex reporter for a wide array of cellular transcriptional programs. Furthermore, due to the ease of use, we were able to quickly recapitulate a diverse set of key results from yeast genetics in a single experiment. Importantly, our approach also captures the effects of a treatment across the entire panel of reporters and so it also reports on unexpected effects that may reveal previously unrecognized connections between transcriptional networks.

5.3 Discussion

In summary, we developed a palette of ratiometric fluorescent tags that can be used to track mixed cultures of yeast with up two 20 subpopulations. Furthermore, we validated that a third fluorescent reporter can be tracked in each of these subpopulations. Importantly, these CellTags can be measured directly, are non-destructive and require minimal hands-on time. Therefore, they enable high throughput, time resolved experiments. Additionally, since the information is available immediately from a small sample of cells, the experimental parameters can be updated in real time. We envision that the CellTags could be combined with continuous flow cytometers and chemo stats to generate, manipulate and tune complex microbial communities.

Furthermore, the underlying principle of translationally controlled protein ratios can be easily applied to generate equivalent CellTag systems for bacteria and mammalian cells. A range of bacterial CellTags can be immediately envisioned through the use of the well-characterized variants of the dnaX of RF2 frame shift motifs [64–66]. In mammalian cells, the initial reported library of frame shift motifs could be searched for variants with the appropriate properties.

Applications of the CellTags beyond the ones tested explicitly in this chapter include: profiling of drug-receptor interactions, use to develop a sensor cocktail based on mix of multiple sensor cells and mapping cell-cell communication networks.

5.4 References

1. Goers, L., Freemont, P. & Polizzi, K. M. Co-culture systems and technologies: taking synthetic biology to the next level. *J. R. Soc. Interface* **11**, 20140065 (2014).

2. Jia, X. *et al.* Design, analysis and application of synthetic microbial consortia. *Synth. Syst. Biotechnol.* **1**, 109–117 (2016).

3. *Engineering and Analyzing Multicellular Systems*. **1151**, (Springer New York, 2014).

4. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).

5. Brimacombe, K. R. *et al.* A Dual-Fluorescence High-Throughput Cell Line System for Probing Multidrug Resistance. *ASSAY Drug Dev. Technol.* **7**, 233–249 (2009).

6. Muzzey, D. & van Oudenaarden, A. Quantitative Time-Lapse Fluorescence Microscopy in Single Cells. *Annu. Rev. Cell Dev. Biol.* **25**, 301–327 (2009).

7. Mee, M. T. & Wang, H. H. Engineering ecosystems and synthetic ecologies. *Mol. BioSyst.* **8**, 2470–2483 (2012).

8. Rollié, S., Mangold, M. & Sundmacher, K. Designing biological systems: Systems Engineering meets Synthetic Biology. *Chem. Eng. Sci.* **69**, 1–29 (2012).

9. Großkopf, T. & Soyer, O. S. Synthetic microbial communities. *Curr. Opin. Microbiol.* **18**, 72–77 (2014).

10. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci.* **111**, E2149–E2156 (2014).

11. Picot, J., Guerin, C. L., Le Van Kim, C. & Boulanger, C. M. Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* **64**, 109–130 (2012).

12. Zuleta, I. A., Aranda-Díaz, A., Li, H. & El-Samad, H. Dynamic characterization of growth and gene expression using high-throughput automated flow cytometry. *Nat. Methods* **11**, 443–448 (2014).

13. Bagwell, C. B. & Adams, E. G. Fluorescence Spectral Overlap Compensation for Any Number of Flow Cytometry Parameters. *Ann. N. Y. Acad. Sci.* **677**, 167–184 (1993).

14. Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* **4**, 648–655 (2004).

15. Bandura, D. R. *et al.* Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).

16. Gore, J., Youk, H. & van Oudenaarden, A. Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**, 253–256 (2009).

17. Higuchi-Sanabria, R. *et al.* Characterization of Fluorescent Proteins for Three- and Four-Color Live-Cell Imaging in S . cerevisiae. *PLOS ONE* **11**, e0146120 (2016).

18. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).

19. Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W. & Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013).

20. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* (2015).doi:10.1021/sb500366v

21. Maheshri, N. & O'Shea, E. K. Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).

22. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).

23. Anzalone, A. V., Lin, A. J., Zairis, S., Rabadan, R. & Cornish, V. W. Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. *Nat. Methods* **13**, 453–458 (2016).

24. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).

25. Houdebine, L. M. & Attal, J. Internal ribosome entry sites (IRESs): reality and use. *Transgenic Res.* **8**, 157–177

26. Karim, A. S., Curran, K. A. & Alper, H. S. Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS Yeast Res.* **13**, 107–116 (2013).

27. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**, 107–113 (2009).

28. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLoS ONE* **6**, e28674 (2011).

29. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, 751 (2012).

30. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90 (2002).

31. Sakaue-Sawano, A. *et al.* Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132**, 487–498 (2008).

32. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nat. Methods* **10**, 751–754 (2013).

33. Yu, D. *et al.* A naturally monomeric infrared fluorescent protein for protein labeling in vivo. *Nat. Methods* **12**, 763–765 (2015).

34. Finak, G. *et al.* OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. *PLOS Comput Biol* **10**, e1003806 (2014).

35. Lo, K., Hahne, F., Brinkman, R. R. & Gottardo, R. flowClust: a Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics* **10**, 145 (2009).

36. Vidal, M., Braun, P., Chen, E., Boeke, J. D. & Harlow, E. Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci.* **93**, 10321–10326 (1996).

37. Soboleski, M. R., Oaks, J. & Halford, W. P. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *FASEB J.* (2005).doi:10.1096/fj.04-3180fje

38. García, R., Rodríguez-Peña, J. M., Bermejo, C., Nombela, C. & Arroyo, J. The High Osmotic Response and Cell Wall Integrity Pathways Cooperate to Regulate Transcriptional Responses to Zymolyase-induced Cell Wall Stress in Saccharomyces cerevisiae. *J. Biol. Chem.* **284**, 10901–10911 (2009).

39. Slater, M. R. & Craig, E. A. Transcriptional regulation of an hsp70 heat shock gene in the yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.* **7**, 1906–1916 (1987).

40. Tan, S.-X., Teo, M., Lam, Y. T., Dawes, I. W. & Perrone, G. G. Cu, Zn Superoxide Dismutase and NADP(H) Homeostasis Are Required for Tolerance of Endoplasmic Reticulum Stress in Saccharomyces cerevisiae. *Mol. Biol. Cell* **20**, 1493–1508 (2009).

41. Rep, M. *et al.* Osmotic Stress-Induced Gene Expression in Saccharomyces cerevisiae Requires Msn1p and the Novel Nuclear Factor Hot1p. *Mol. Cell. Biol.* **19**, 5474–5485 (1999).

42. Hagen, D. C., McCaffrey, G. & Sprague, G. F. Pheromone response elements are necessary and sufficient for basal and pheromone-induced transcription of the FUS1 gene of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **11**, 2952–2961 (1991).

43. Coleman, S. T., Epping, E. A., Steggerda, S. M. & Moye-Rowley, W. S. Yap1p Activates Gene Transcription in an Oxidant-Specific Fashion. *Mol. Cell. Biol.* **19**, 8302–8313 (1999).

44. Zhou, Z. & Elledge, S. J. Isolation of crt mutants constitutive for transcription of the DNA damage inducible gene RNR3 in Saccharomyces cerevisiae. *Genetics* **131**, 851–866 (1992).

45. Giniger, E., Varnum, S. M. & Ptashne, M. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**, 767–774 (1985).

46. Ozcan, S. & Johnston, M. Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. *Mol. Cell. Biol.* **15**, 1564–1572 (1995).

47. Huibregtse, J. M., Engelke, D. R. & Thiele, D. J. Copper-induced binding of cellular factors to yeast metallothionein upstream activation sequences. *Proc. Natl. Acad. Sci.* **86**, 65–69 (1989).

48. Zhao, H. & Eide, D. The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation. *Proc. Natl. Acad. Sci.* **93**, 2454–2458 (1996).

49. Waters, B. M. & Eide, D. J. Combinatorial Control of Yeast FET4 Gene Expression by Iron, Zinc, and Oxygen. *J. Biol. Chem.* **277**, 33749–33757 (2002).

50. Wysocki, R. *et al.* Transcriptional Activation of Metalloid Tolerance Genes in Saccharomyces cerevisiae Requires the AP-1–like Proteins Yap1p and Yap8p. *Mol. Biol. Cell* **15**, 2049–2060 (2004).

51. Nagawa, F. & Fink, G. R. The relationship between the 'TATA' sequence and transcription initiation sites at the HIS4 gene of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **82**, 8557–8561 (1985).

52. Jung, U. S., Sobering, A. K., Romeo, M. J. & Levin, D. E. Regulation of the yeast Rlm1 transcription factor by the Mpk1 cell wall integrity MAP kinase. *Mol. Microbiol.* **46**, 781–789 (2002).

53. Cooper, T. G., Rai, R. & Yoo, H. S. Requirement of upstream activation sequences for nitrogen catabolite repression of the allantoin system genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **9**, 5440–5444 (1989).

54. Cunningham, K. W. & Fink, G. R. Calcineurin inhibits VCX1-dependent H+/Ca2+ exchange and induces Ca2+ ATPases in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **16**, 2226–2237 (1996).

55. Vasconcelles, M. J. *et al.* Identification and Characterization of a Low Oxygen Response Element Involved in the Hypoxic Induction of a Family ofSaccharomyces cerevisiae Genes IMPLICATIONS FOR THE CONSERVATION OF OXYGEN SENSING IN EUKARYOTES. *J. Biol. Chem.* **276**, 14374–14384 (2001).

56. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

57. Teste, M.-A., Duquenne, M., François, J. M. & Parrou, J.-L. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC Mol. Biol.* **10**, 99 (2009).

58. Johnston, M., Flick, J. S. & Pexton, T. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **14**, 3834–3841 (1994).

59. Jamieson, D. J. Saccharomyces cerevisiae has distinct adaptive responses to both hydrogen peroxide and menadione. *J. Bacteriol.* **174**, 6678–6681 (1992).

60. MacKay, V. L. *et al.* Gene Expression Analyzed by High-resolution State Array Analysis and Quantitative Proteomics Response of Yeast to Mating Pheromone. *Mol. Cell. Proteomics* **3**, 478–489 (2004).

61. Ko, N., Nishihama, R. & Pringle, J. R. Control of 5-FOA and 5-FU resistance by Saccharomyces cerevisiae YJL055W. *Yeast* **25**, 155–160 (2008).

62. Heitman, J. *et al.* The immunosuppressant FK506 inhibits amino acid import in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **13**, 5010–5019 (1993).

63. Muller, E. M., Mackin, N. A., Erdman, S. E. & Cunningham, K. W. Fig1p Facilitates Ca2+ Influx and Cell Fusion during Mating of Saccharomyces cerevisiae. *J. Biol. Chem.* **278**, 38461–38469 (2003).

64. Farabaugh, P. J. Programmed Translational Frameshifting. *Annu. Rev. Genet.* **30**, 507–528 (1996).

65. Larsen, B., Gesteland, R. F. & Atkins, J. F. Structural probing and mutagenic analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting: programmed efficiency of 50%1. *J. Mol. Biol.* **271**, 47–60 (1997).

66. Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Release factor 2 frameshifting sites in different bacteria. *EMBO Rep.* **3**, 373–377 (2002).

6 MATERIALS AND METHODS

6.1 Materials

Chemical reagents were purchased from Sigma Aldrich (St. Louis, MO, USA). Synthetic mating peptides (≥ 95% purity) were ordered from GenScript (Piscataway, NJ, USA) or Zymo Research (Irvine, CA, USA). Polymerases, restriction enzymes and Gibson assembly mix were obtained from New England Biolabs (NEB) (Ipswich, MA, USA). Media components were obtained from BD Bioscience (Franklin Lakes, NJ, USA) and Sigma Aldrich (St. Luis, MO, USA). Primers and synthetic DNA were purchased from Integrated DNA Technologies (IDT) (Coralville, Iowa, USA). Plasmids were cloned and amplified in either *E. coli* TG1 (Lucigen, Madison, WI, USA) or C3040 (NEB). Human urine (Catalog No: IR100007P) and single donor human whole blood (Catalog No: IPLA-WB1) were purchased from Innovative Research (Novi, MI, USA). Human serum, normal off-the-clot (frozen) (Catalog No: HSER-2ML) was purchased from ZenBio (Research Triangle Park, NC, USA). Professional potting mix soil was purchased from Fafard (Agawam, MA, USA).

6.2 General cloning methods in S. cerevisiae

All strains were derived from parental Reiterative Recombination acceptor strain LW2591 (*MATa-inc* genotype), the common laboratory strain FY251 (ATCC 96098) or the yeast two-hybrid reporter strain MaV203 [1,2]. Cloning of expression modules into the *HO* locus was performed using Reiterative Recombination [1]. Scarless gene deletions and some gene replacements were carried out using Delitto Perfetto [3]. The remainder of genomic manipulations was carried out by homology-mediated recombination using PCR fragments [4]. Endogenous yeast promoters, terminators and open reading frames were obtained by PCR from strain FY251 or LW2591 [1]. Yeast transformations were carried out using the lithium acetate method [5]. All plasmids are derivatives of the pRS41x series of centromeric shuttle plasmids, cloned using standard molecular biology protocols and Gibson assembly [6]. See Appendix for list of sequences used.

6.3 Methods specific to Chapter 2: Yeast Sensors

6.3.1 Cloning of lycopene biosensor strains

The parent lycopene biosensor strain (Lyco-1; yMJ118) was constructed by cloning of lycopene pathway genes from *Erwinia herbicola* at reiterative recombination acceptor site in strain yMJ105. The *CrtE*, *CrtB* and *CrtI* ORFs were obtained from plasmid pSC203 [7] (kind gift from Gregory Stephanopolous). The CrtE (geranylgeranyl diphosphate synthase) and CrtB (phytoene synthase) ORFs were cloned into a constitutive expression module containing promoters from *TEF1* and *PGK1* and the bidirectional terminator *ADH1*. The CrtI (lycopene synthase) ORF was cloned into a pheromone-inducible expression module containing *FUS1* promoter the *ACT1* terminator. The enhanced parent lycopene biosensor strain (Lyco-2; yMJ251) carried an additional copy of the pFUS1-CrtI-tACT1 expression module integrated at the *MET15* locus and a pTDH3-FAD1-tCYC1 overexpression module integrated at the reiterative recombination acceptor site. All fungal biosensor strains described in this study were derived from yMJ251 by replacement of the endogenous *STE2* gene with the appropriate fungal receptor expression module (Table S4)

6.3.2 Characterization of the lycopene readout in liquid culture

Induction of lycopene was assayed using strains yMJ118 or yMJ251 in 96-well microtiter plates cultured at 30 °C and 800 RPM. Cells were seeded at an OD_{600} of 2 in standard complete synthetic media (2% dextrose) supplemented with 5% yeast extract peptone dextrose media (YPD) and with the indicated concentration of synthetic peptide. All measurements were performed in triplicates. Cell density and media conditions were chosen to more closely mimic conditions relevant for the yeast-based paper assay (i.e. high cell density and non-selective complex media) while enabling more precise spectroscopic determination of lycopene content (i.e. higher bulk signal at early time points and low-absorbance media). Under these conditions the cultures grew to maximal OD_{600} of 6. Relative lycopene content was calculated by spectroscopy [8] using Eq. S5 (Appendix). Optical densities were measured with an Infinite M200 plate reader (Tecan). Lycopene values were normalized by the culture $OD_{600}$to give a measure of lycopene per cell. For each strain, maximum yield of lycopene per OD_{600} was determined as the largest observed value over a 72-hour period for each biological replicate. Half-max lycopene per cell for each biological replicate was calculated as

the average of the largest and smallest lycopene per OD_{600} value observed over a 72-hour period. Time to half-max lycopene per cell for each strain was determined by linear interpolation between the two time points with lycopene per OD_{600} values that spanned the calculated half-max lycopene per OD_{600} for each biological replicate.

6.3.3 Characterization of biosensor strains in liquid culture (pH, temperature, and complex samples)

P. brasiliensis or *C. albicans* biosensor strains (yMJ258 and yMJ260, respectively) were characterized in 96-well microtiter plates cultured at 800 RPM. Cells were seeded at an OD_{600} of 2 in standard complete synthetic media (2% dextrose) supplemented with 5% yeast extract peptone dextrose media (YPD) and the indicated concentration of synthetic peptide. A 2x stock of media and a 10x stock of the ligand were diluted to reach the appropriate 1x concentration. All measurements were performed in triplicates. Lycopene production was measured by absorbance as described above. For temperature assays, the microtiter plate shaker was pre-equalized to the appropriate temperature for 1 hour before the start of the assay. For pH assays, the pH of the media stock was titrated to the appropriate value with sodium hydroxide. For complex sample assays, urine and serum were centrifuged to remove particulates and used at a final concentration of 50% (in YPD).

6.3.4 Preparation of culture supernatant from clinically isolated fungal pathogens.

H. capsulatum - supernatants from strains Hc01 and Hc06 cultures were a generous gift from Dr. Chad Rappleye, Department of Microbiology, Ohio State University, Columbus, OH, USA. These strains are clinical isolates representing North America class 2 (NAm2) and North America class 1 (NAm1), respectively [9]. *H. capsulatum* strains were added to liquid SDA medium (40 g/L glucose, 10 g/L peptone) at 10^5 cells/mL and incubated for 10 days at 26 °C without agitation to induce conversion to mycelia. Conversion to mycelia was confirmed by phase-contrast microscopy. Mycelia were then transferred to HMM media [10] and the cultures incubated at 26°C. After 3 weeks of growth, mycelia were separated from the supernatant by filtration through a cellulose filter (Whatman qualitative filter paper #2, 8 μm-diameter pores) and the filtrate subsequently filtered through a polyethersulfone membrane (0.45 μm diameter pores) to obtain the

final culture filtrate. The supernatants were lyophilized, resuspended in 0.1 volume of H_2O (10x concentration) and kept at -20°C.

Paracoccidioides - supernatants from *P. brasiliensis* Pb18 and *P. lutzii* Pb01 cultures were a generous gift from Dr. Fernando Rodrigues, Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal. These strains are clinical isolates containing mating loci MAT1-2 and MAT1-1, respectively [11]. The mycelium form was grown at 24°C at 150 rpm in synthetic McVeigh Morton (MMvM) liquid medium [12]. Supernatants were collected by filtration 10 days after the yeast-mycelium transition. The supernatants were lyophilized, resuspended in 0.1 volume of H_2O (10x concentration) and kept at -20°C.

C. albicans - Human isolate GC75 with MTLα/MTLα [13], was a generous gift from Dr. Richard Bennett, Department of Molecular Microbiology and Immunology, Brown University, Providence, RI, USA. A second human isolate of *C. albicans* was generously provided by Dr. Anne-Catrin Uhlemann, Department of Medicine, Division of Infectious Diseases, Columbia University, New York, NY, USA. This isolate, referred to as ySB36, was genotyped by PCR for its mating loci using genomic DNA as template as described previously[14] and found to be heterozygous for its mating loci, MTLa/MTLα (for primers see Supplementary Table 4). Homozygous MTLα/MTLα derivatives of ySB36 were obtained by selection on sorbose as previously described[15]. In brief: ySB36 was cultured for 16 hours in YPD liquid media at 30°C, washed once with water and ~ 10^5 cells were plated on 2% sorbose media (0.67% yeast nitrogen base without amino acids, 2% sorbose). Colonies were visible after 4 days incubation at 30°C. Several colonies were re-streaked on 2% sorbose media, followed by re-streaking on YPD media and genotyping by colony PCR as above. One homozygous MTLα/MTLα isolate (ySB45) was used for supernatant preparation.Phenotypically switched opaque colonies of GC75 and ySB45 were isolated by Phloxine B staining as previously described [16]. In brief: A single colony of GC75 or ySB45 was incubated for 24 h at 25°C in liquid YPD media without agitation. In total ~ $2x10^3$ cells were plated on YPD agar supplemented with 5 μg/ml Phloxine B (Sigma Aldrich) and incubated at 25°C for 4 days. Opaque colonies stained pink on Phloxine B containing media. For supernatant preparation, a single opaque colony of *C. albicans* GC75 or ySB45 was cultured overnight in YPD media at 25°C, and used to

inoculate 50 ml of YPD liquid media. Cells were cultured for ~ 24 h at 25°C to a final OD_{600} of 9.5 (~2.8x10^8 cells/ml) and 7.9 (~2.3x10^8 cells/ml), respectively. Cells were pelleted by centrifugation, the supernatant was reduced to dryness by vacuum concentration and resuspended in 0.1 volume H_2O (10x concentration) and kept at -20°C.

S. cerevisiae - samples were obtained from *S. cerevisiae* strain FY250 with MTLα [17] and W303-1B with MTLα (ATCC 201238). Cells were cultured in 50 ml YPD media for 20 h at 30°C to a final OD_{600} of 9.8 (~2.9x10^8 cells/ml) and 8.5 (~2.5x10^8 cells/ml), respectively. Cells were pelleted by centrifugation, the supernatant of FY250 was reduced to dryness by vacuum concentration and resuspended in 0.1 volume H_2O (10x concentration) and kept at -20°C. The supernatant of W303-1B was kept at 1x concentration at -20°C.

6.3.5 Detection of mating peptides in supernatants of clinically isolated fungal strains.

P. brasiliensis or *C. albicans* biosensor strains (yMJ258 and yMJ260, respectively) and a control S. cerevisiae strain (yMJ251) were used to test for the presence of the respective mating peptides in supernatants derived from clinically isolated pathogenic fungi or *S. cerevisiae* (supernatants preparation described above). Cells were seeded at an OD_{600} of 2 in the indicated supernatant mixed with standard complete synthetic media (2% dextrose) supplemented with 5% YPD in 96-well microtiter plates, cultured at 30°C and 800 RPM, and lycopene production was measured by absorbance as described above. A 2x stock of media and a 10x stock of the supernatant were used and diluted to reach the appropriate 1x concentration. The control supernatant for W303-1B was diluted to 50% in the final assay. Statistical significance of signal (i.e. biosensor strain treated with its cognate-supernatant) over noise (same biosensor strain treated with non-cognate supernatants) was determined by performing a paired parametric t-test in Prism (GraphPad). All measurements were performed in triplicates.

6.3.6 Paper-based dipstick assay for detection of fungal peptides in complex samples.

A simple, low-cost dipstick assay was designed with the biosensor strains spotted on paper as the only required active component. To assemble the dipstick, the biosensor strains were pre-cultured in 50 mL of yeast extract peptone dextrose media (YPD) at 30 °C at 300 RPM for 72 hours. The culture was diluted with water to an OD_{600} of 2.5 and vacuum filtered onto a glass fiber filter

paper (Thermo Scientific, DS0281-7500) using a plastic stencil to generate spots with a diameter of 5 mm. An appropriate culture volume was used to give about $5x10^7$ cells per spot. The filter paper with biosensor spots was cut into small squares (8x8 mm, 1 biosensor spot) and placed onto a strip of wicking paper made of a standard brown paper towel. Each paper-based dipstick assay contained two different spots - an indicator (biosensor) spot and a control spot composed of S. cerevisiae carrying off-target receptor as a negative control.

To characterize its functionality, the dipstick was dipped into 1 mL of liquid sample and incubated at 30°C. The lycopene readout was inspected visually and quantitatively measured using time-lapse photography analyzed with ImageJ (see next section) [18]. A 24-well plate was used to easily array several dipsticks in the field of view of the camera. For all assays, a 10x stock of media was used and diluted to reach the appropriate 1x concentration. All measurements were performed in three or more replicates. For YPD assays, the dipstick was dipped into YPD media supplemented with 1 μM of the indicated fungal pathogen peptide. For soil assays, 0.5 g of soil was pre-conditioned with 2 nmol (in 200 μL of water) of the indicated fungal pathogen peptide and allowed to air dry for 1 hour. Then the dipstick was inserted into the soil added with 2 mL of YPD media to give a concentration of 1 μM of fungal peptide. For urine and serum assays, the samples were vortexed briefly to resuspend particles, supplemented with YPD media and diluted to give a concentration of 50%. For blood assays, the sample was supplemented with YPD media and diluted to give a final concentration of 2% blood.

Additionally, we designed a small plastic holder to facilitate the ease of use of this dipstick assay. This plastic holder was 3D printed out of acrylonitrile butadiene styrene (ABS). We validated the holder it did not negatively impact the assay functionality.

6.3.7 Determination of lycopene content by time-lapse photography

To enable quantitative characterization of the paper-based dipstick assay we developed a method to measure lycopene production based on time-lapse photography and pixel color value analysis. Specifically, dipsticks dipped in samples and a tripod-mounted digital single-lens reflex camera (DSLR, Nikon D7000) were placed in a dark box kept at 30 °C. Flash photographs were taken automatically every 5 minutes. The resulting sequence of photographs was analyzed using ImageJ[49].

For each time point, the average pixel color values were measured for each of the two dipstick spots using constant measurement areas. The apparent level of red color of each spot was first calculated by the following:

$$R_{apparent} = \frac{R - \left(\frac{G+B}{2}\right)}{R} \quad (7)$$

where R, G, B are the measured red, green and blue color values, respectively. Since the color of the biosensor spots ranges from off-white to red-orange the color values are such that R>G>B is always true. Therefore, $R_{apparent}$ is a value that scores the level of red from 0 to 1. We then calculated the total level of positive lycopene readout produced by the dipstick by the following:

$$\Delta\, Red\ Color = R_{app,indicator} - R_{app,negative} \quad (8)$$

where $R_{app,\ indicator}$ and $R_{app,\ negative}$ are the apparent red color values of the indicator biosensor spot and the negative control yeast spot, respectively given by Equation 7. Importantly, since the two yeast spots of the dipstick assay are always in close proximity to each other, the Δ Red Color value is not sensitive to variations in light levels and can be used to compare dipsticks placed anywhere in the field of view of the camera.

Using these sequences of photographs we also generated time-lapse clips showing that the lycopene color change can be visualized by the naked eye. These clips are motion and exposure equalized to remove flicker between frames.

6.3.8 Visibility threshold of lycopene readout.

When measured spectroscopically (Appendix), we determined a visible threshold of 3.5 lycopene per cell units (LPC) in liquid cultures. This threshold was determined by visually inspecting pellets of $5x10^7$ cells.

We also determined a visibility threshold for paper-based dipstick assay when measured by time-lapse photography and pixel color analysis (see previous section). This was done by visually inspecting time-lapse clips. The visible threshold for the dipstick assay was determined to be 4 Δ Red Color units.

6.4 Methods specific to Chapter 3: GPCR directed evolution

6.4.1 Construction of directed evolution reporter strain

A codon-optimized copy of the mCherry fluorescent reporter gene [19] was cloned with *S. cerevisiae* pheromone-inducible promoter *FUS1* and *ACT1* terminator. This expression module was integrated at the reiterative recombination acceptor site in strain yMJ105. The *pFUS1-HIS3-tHIS3* construct was integrated into the *FUS1* locus by selection on histidine minus media. The *pFUS2-URA3-tFUS2* was integrated into the *FUS2* locus by selection on uracil minus media and subsequent selection on 5-FOA media. The *STE2* gene was deleted to yield the fluorescent reporter strain yMJ194.

6.4.2 GPCR expression plasmids and library construction

Receptor variants were incorporated into expression modules containing the *S. cerevisiae TDH3* promoter and *STE2* terminator. These expression modules were cloned by gap repair into low copy plasmids derived from pRS414. For library generation, parental expression modules were amplified in an error-prone PCR with $MnCl_2$ and then reamplified for transformation into strain yMJ194. Transformations were performed on a large scale of 10^9 cells and transformants were selected in batch liquid culture (1 L) allowing an enrichment of at least 10x over the initial cell density. Diversity and transformation yield of the library was analyzed by isolating individual transformants and comparing the number to the number of colonies on non selective plates. The receptor ORF was amplified by colony PCR and sequenced for 10 – 20 clones. A portion of the plasmid-enriched libraries were kept as glycerol stocks and combined with later transformations at a proportional diversity.

6.4.3 Positive and negative growth selection

Transformed libraries were selected in 1 to 10 mL volume cultures incubated at 30°C and 230 RPM. Cells were seeded at an OD_{600} of 0.01 in standard synthetic dropout media (2% dextrose). For positive selections the media lacked histidine and was supplemented with 40 mM 3-AT and the indicated concentration of selection ligand. For negative selections the media was supplemented with 2.6 mM 5-FOA. The cultures were grown until substantial growth above baseline occurred

(ususually 2-3 days). The selected libraries were then plated to isolate individual clones for screening.

6.4.4 Fluorescence-activated cell sorting (FACS) selection

Transformed libraries were induced in 1 mL volume cultures with the indicated concentration of selection ligand and incubated for 12 – 18 hours at 30°C and 230 RPM. Clonal control cultures and uninduced samples were grown in parallel. Cells were seeded at an OD_{600} of 0.1 in standard synthetic dropout media (2% dextrose) selecting for the plasmid. Flow cytometry analysis of the control samples was used to determine gating thresholds to capture the indicated residual population percent. Induced libraries were diluted in PBS buffer and FACS was used to collect 10^5-10^6 cells falling within the target gates resulting in 0.1 to 1 mL samples. The sorted cells were pelleted at 3K RMP for 30 minutes, 90% of the supernatant was removed and replaced with 0.5 volume of fresh media. After substantial growth occurred (1-2 days) the cells were plated to isolate individual clones.

6.4.5 High throughput screening

Isolated colonies from the selections were inoculated into 96-square well plates prefilled with plasmid selective media. At this stage, control strains were also inoculated into reserved wells of the micro titer plate. These master clone plates were incubated at 30°C without shaking for 24 hours, then a portion was used for screening and the remaining cultures were supplemented with glycerol and stocked at -80°C directly in the 96-square well plates.

384-well screening plates were pre-filled with 45uL of growth media using a robotic liquid handling system (BioMek FX, Beckman Coulter). 5 uL of induction ligands were transferred from 384-well 10X ligand stock plates arrayed in the desired configuration of plus and minus ligand. Lgand stock plates were kept at -80°C and the prefilled screening plates were kept at 4°C until use (<24 hours).

To initiate screening, the master clone plates were resuspended by shaking at 800 RPM for 5 minutes. Then the clones were transferred from the 96-square well master clone plates to the 384-well screening plates using a 384-pin replicator with 2uL transfer notches (VP384S2, V & P

Scientific). The screening plates were incubated at 30°C without shaking. Fluorescence and culture turbidity (OD_{600}) were measured using an Infinite M200 plate reader (Tecan).

6.4.6 Characterization of selected variants

Clones chosen for further characterization were regrown from the stocked master clone plates and then assayed in 96-well microtiter plates cultured at 30°C and 800 RPM. Cells were seeded at an OD_{600} of 0.1 in standard synthetic dropout media (2% dextrose) lacking the appropriate selective component with the indicated concentration of ligand or water (control). Measurements were performed in duplicates using an Infinite M200 plate reader (Tecan).

6.5 Methods specific to Chapter 4: Fungal genome mining

6.5.1 Determination and analysis of fungal GPCRs sequences

A database of fungal receptors was curated from the InterPro (IPR000366) [20] and PFAM (PF02116) families [21]. Sequence identifiers were standardized using the UniProt ID mapping tool (http://www.uniprot.org/uploadlists/). UniProt IDs were used to programmatically retrieve associated taxonomic information (Appendix). Taxonomic information was used to filter out non-fungal sequences and fragments. Sequences homology relationships and percent identity was calculated by multiple sequence alignment using MUSCLE [22]. Sequence alignments were curated using the Jalview alignment editor [23]. Sequences were curated by trimming N- and C-terminal residues corresponding to those before tyrosine 17 and those after asparagine 301 in the *S. cerevisiae* Ste2 protein. The redundancy was reduced by removing sequences with greater than 98% identity. Sequences with large gaps (>20 residues) in highly conserved regions across the group were also removed. Phylogenetic and conservation analysis of the resulting receptor list was carried out with the Phylogeny.fr [24] and Skylign [25] tools. Positions of the extracellular loops, cytoplasmic loops and treansmembrane helices were predicted using the TMHMM2.0 tool [26]. See Appendix for list of receptors.

6.5.2 Determination of fungal mating peptide sequences

The corresponding peptide ligands were taken determined by using the method of Martin et al [27]. Using these pheromone genes as a starting point, additional putative pheromone genes were gathered by BLAST searches against the UniProt KB database. This process was repeated excluding

species for which the pheromone gene was already gathered. The resulting set of pheromone genes was manually curated to determine the presence of repeated peptide sequences and putative processing sites. Putative repeats and surrounding residues for each pheromone gene were aligned using MUSCLE to reveal core conserved regions and present aligned processing sites for manual determination of predicted secreted peptide sequences. The final set of secreted peptide sequences was realigned with MUSCLE to determine homology relationships and percent identity. See Appendix for list peptides.

6.5.3 Construction of GPCR expression plasmids

Mating receptor ORFs were synthesized as codon optimized genes for *S. cerevisiae* and others were cloned directly from the appropriate fungal genomic DNA (American Type Culture Collection) or plasmid pLPreB. All receptor ORFs were incorporated into expression modules containing the *S. cerevisiae TDH3* promoter and *STE2* terminator. For fluorescent assays using reporter strain yMJ183, receptor expression modules were cloned into low copy plasmids derived from pRS416.

6.5.4 Functional characterization of GPCR-peptide pairs

Fungal mating receptor Ste2 activity was measured in strain yMJ183 using the fluorescent reporter mCherry. The fluorescent reporter strains carrying the appropriate fungal Ste2 expression plasmid were assayed in 96-well microtiter plates cultured at 30°C and 800 RPM. Cells were seeded at an OD_{600} of 1 in standard synthetic dropout media (2% dextrose) lacking the appropriate selective component with either synthetic fungal mating peptide or water (control) as indicated. All measurements were performed in triplicates. mCherry fluorescence (excitation: 588nm, emission: 620nm) and culture turbidity (OD_{600}) were measured using an Infinite M200 plate reader (Tecan). For determination of EC_{50} and fold-activation values, the fluorescence response of strain yMJ183 carrying the appropriate Ste2 receptor was measured at different concentrations of the appropriate synthetic mating peptide. All raw fluorescence values were normalized by the OD_{600}, plotted against the corresponding peptide concentration and the points fit with a four-parameter logistic curve using Prism (GraphPad). Fold-activation was calculated for each receptor strain as the maximum OD_{600}-normalized fluorescence of peptide-treated cells divided by the OD_{600} normalized fluorescence value of water-treated cells. For cross-reactivity measurements, receptor activation was individually

measured using each of the synthetic fungal mating peptides (5 μM). Percent receptor activation was calculated by setting the OD_{600}-normalized fluorescence value of cognate-peptide activation to 100% and the value of water treated-cells to 0% (see Appendix).

6.6 Methods specific to Chapter 5: CellTags

6.6.1 Construction of CellTag integration constructs and tagged strains

The integration construct was derived from the pNH600 series of vectors [28]. These vectors contain integration constructs with a multiple cloning site, an ADH1 terminator from *Candida albicans*, selectable auxotrophic markers from *Candida glabrata* and flanking 500 bp homology regions to the target locus (pNH605: LEU2). These integration constructs were cloned onto a pRS416 backbone to allow comparison of plasmid-borne and genome-integrated constructs from the same vector.

The CellTag constructs are expressed from a TDH3 promoter and contain fluorescent open reading frames spaced by a framsshift motif consisting of an intervening linker peptide (FLAG tag), a slippery site, a stop codon, and a pseudoknot region as described previously [29]. For frame shift motifs placed at the 5′ end of the open reading frame an alternate short leader peptide is used instead of the linker peptide.

A library of framshift motifs was cloned into the integration vector by gap repair in yeast strain FY251. Individual clones were isolated by plating and the ratios of eGFP and mCherry fluorescence were assayed in 96-well plates using an Infinite M200 plate reader (Tecan). Variants representing a range of fluoresnce protein ratios were selected, the plasmids were extracted and amplified in *E. coli*, and the sequence of the frameshift motif was determined. These selected vectors were linearized and integrated into the LEU2 locus of a fresh FY251 strain. The resulting fluorescence stoichiometries were then characterized by flow cytometry.

The expanded palette of CellTags was constructed by constructing all possible combinations of 5 frame shift motifs at early and late positions, with two fluorescent proteins. The resulting vectors were linearized and integrated individually into strain FY251. The resulting strains were characterized by flow cytometry and a well resolved set of 20 CellTags was chosen.

6.6.2 Characterization of CellTags by flow cytometry

Cells were precultured overnight to high density in 96-well plates and then inoculated into fresh mediat (2% glucose) at an OD600 of 0.03 and allowed to grow for less than 12 hours at 30°C and 800 RPM. Samples were diluted with PBS and analyzed by flow cytometry. The resulting fluorescence values were normalized by dividing by the side scatter and scaled to match the initial value range. Individual CellTags were inspected manually using FlowJo and mixtures of CellTags were analyzed programmatically using the openCyto R package (see Appendix) [30].

6.6.3 Tagged MaV203 strains and Gal4 induction constructs

The CellTag vectors were purified from *E. coli* and linearized as a mixture. This CellTag library was integrated into the LEU2 locus of the yeast two-hybrid reporter strain MaV203. The transformed library was plated to isolate individual clones, which were then characterized by flow cytometry to yield a set of tagged MaV203 strains.

Well characterized frameshift motives were introduced between the Gal4 DNA binding domain and activation domain on a pRS414 plasmid as previously described [29]. This set of Gal4 constructs up regulates the expression of genes under the control of the GAL1 promoter at a level set by the efficiency of the frame shift motif. These Gal4 induction construct were individually transformed into separated CellTag MaV203 strains, and the individual fitness in 5-FOA media was characterized in 96-well plates.

These strains were grown as a mixture and inoculated into 5-FOA selective media at an OD600 of 0.01. The growth was measured by optical density measurements and the proportion of the subpopulations was analyzed by flow cytometry.

6.6.4 Yeast stress reporter cocktail

A set of native promoters was cloned into an mTagBFP2 expression construct on pRS416 plasmids (see Appendix for promoter sequences). These reporter plasmids were transformed individually into separate CellTag FY251 strains. These strains were grown over night as a mixture and inoculated into 96-well plates at and OD600 of 0.1 in growth media supplemented with the indicated inducing agents. The mixed cultures were incubated for 4 – 6 hours at 30°C and 800 RPM

and subsequently analyzed by flow cytometry. The bulk signal from the mTagBFP2 reporters was subdivided programmatically using the CellTag populations for deconvolution (see Appendix).

6.7 References

1. Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U A* **108**, 15135–15140 (2011).

2. Vidal, M., Braun, P., Chen, E., Boeke, J. D. & Harlow, E. Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci.* **93**, 10321–10326 (1996).

3. Storici, F. & Resnick, M. A.in *DNA Repair Part B* **Volume 409**, 329–345 (Academic Press, 2006).

4. Moqtaderi, Z. & Geisberg, J. V.in *Curr. Protoc. Mol. Biol.* (John Wiley & Sons, Inc., 2001).at <http://onlinelibrary.wiley.com.ezproxy.cul.columbia.edu/doi/10.1002/0471142727.mb1310cs104/abstract>

5. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).

6. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

7. Wingler, L. M. Harnessing Saccharomyces cerevisiae Genetics for Cell Engineering. (2011).at <http://academiccommons.columbia.edu/catalog/ac:162778>

8. Myers, J. A., Curtis, B. S. & Curtis, W. R. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys.* **6**, 4 (2013).

9. Goughenour, K. D., Balada-Llasat, J.-M. & Rappleye, C. A. Quantitative Microplate-Based Growth Assay for Determination of Antifungal Susceptibility of Histoplasma capsulatum Yeasts. *J. Clin. Microbiol.* **53**, 3286–3295 (2015).

10. Worsham, P. L. & Goldman, W. E. Quantitative plating of Histoplasma capsulatum without addition of conditioned medium or siderophores. *J. Med. Vet. Mycol. Bi-Mon. Publ. Int. Soc. Hum. Anim. Mycol.* **26**, 137–143 (1988).

11. Torres, I. *et al.* Presence and expression of the mating type locus in Paracoccidioides brasiliensis isolates. *Fungal Genet. Biol.* **47**, 373–380 (2010).

12. Restrepo, A. & Jiménez, B. E. Growth of Paracoccidioides brasiliensis yeast phase in a chemically defined culture medium. *J. Clin. Microbiol.* **12**, 279–281 (1980).

13. Pujol, C., Pfaller, M. & Soll, D. R. Ca3 Fingerprinting of Candida albicans Bloodstream Isolates from the United States, Canada, South America, and Europe Reveals a European Clade. *J. Clin. Microbiol.* **40**, 2729–2740 (2002).

14. Magee, B. B. & Magee, P. T. Induction of Mating in Candida albicans by Construction of MTLa and MTLα Strains. *Science* **289**, 310–313 (2000).

15. Janbon, G., Sherman, F. & Rustchenko, E. Monosomy of a specific chromosome determines l-sorbose utilization: A novel regulatory mechanism in Candida albicans. *Proc. Natl. Acad. Sci.* **95**, 5150–5155 (1998).

16. Anderson, J. M. & Soll, D. R. Unique phenotype of opaque cells in the white-opaque transition of Candida albicans. *J. Bacteriol.* **169**, 5579–5588 (1987).

17. Sherwood, P. W. (Columbia U. & Carlson, M. Mutations in GSF1 and GSF2 alter glucose signaling in Saccharomyces cerevisiae. *Genet. USA* (1997).at <http://agris.fao.org/agris-search/search.do?recordID=US1997081878>

18. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).

19. Keppler-Ross, S., Noffz, C. & Dean, N. A new purple fluorescent color marker for genetic studies in Saccharomyces cerevisiae and Candida albicans. *Genetics* **179**, 705–710 (2008).

20. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).

21. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).

22. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

23. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).

24. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).

25. Wheeler, T. J., Clements, J. & Finn, R. D. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**, 7 (2014).

26. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. *J. Mol. Biol.* **305**, 567–580 (2001).

27. Martin, S. H., Wingfield, B. D., Wingfield, M. J. & Steenkamp, E. T. Causes and Consequences of Variability in Peptide Mating Pheromones of Ascomycete Fungi. *Mol. Biol. Evol.* **28**, 1987–2003 (2011).

28. Zalatan, J. G., Coyle, S. M., Rajan, S., Sidhu, S. S. & Lim, W. A. Conformational Control of the Ste5 Scaffold Protein Insulates Against MAP Kinase Misactivation. *Science* **337**, 1218–1222 (2012).

29. Anzalone, A. V., Lin, A. J., Zairis, S., Rabadan, R. & Cornish, V. W. Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. *Nat. Methods* **13**, 453–458 (2016).

30. Finak, G. *et al.* OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. *PLOS Comput Biol* **10**, e1003806 (2014).

7 APPENDIX

7.1 A high throughput method for quantifying pigment production in cells

As described by Myers et al. [8] density of the cell suspension measured at a sensitive wavelength (i.e. corresponding to an absorption maxima of the pigment) is approximately composed of two additive components: scatter due to cells and absorbance due to the pigment. Therefore the pigment content in a cell suspension is proportional to the measured optical density corrected for the scattering component as follows:

$$[\boldsymbol{pigment}] \propto \boldsymbol{Abs_{S,P} = OD_S - OD_{S,scat}} \quad \textbf{(1)}$$

where $Abs_{S,P}$ is the absorbance due to the pigment at the sensitive wavelength S, OD_S is the measured optical density at the sensitive wavelength *S*, and $OD_{S,scat}$ is a calculated scattering component at the sensitive wavelength *S*. Since there was noticeable Raleigh-like wavelength dependence in the scatter of lycopene null strains we chose the following functional form to approximate scatter at a particular wavelength λ:

$$\boldsymbol{OD_{\lambda,scat} = B - \log_{10}\left(1 - \frac{A}{\lambda}\right)} \quad \textbf{(2)}$$

where *A* and *B* are constants that reflect changes in cell density and other sample irregularities. At each time point and for each sample, we can calculate the corresponding values of *A* and *B* by using the optical density values measured at two robust wavelengths (i.e. corresponding to wavelengths where scatter is the only or dominant component). Substituting these additional scatter-only optical density measurements into Eq. S2 and solving for *A* and *B* we get:

$$\boldsymbol{A = R1\left(\frac{1-T}{\frac{R1}{R2}-T}\right)}, \text{ where } \boldsymbol{T = 10^{OD_{R1} - OD_{R2}}} \quad \textbf{(3)}$$

$$\boldsymbol{B = OD_{R1} + \log_{10}\left(1 - \frac{A}{R1}\right)} \quad \textbf{(4)}$$

where OD_{R1} and OD_{R2} are the measured optical densities at the robust wavelengths *R1* and *R2*. Therefore, by setting $\lambda = S$ and substituting Eq. S2 into Eq. S1, the relative content of lycopene in a cell suspension is given by:

$$[\boldsymbol{pigment}] \propto \boldsymbol{Abs_{S,P} = OD_S + \log_{10}\left(1 - \frac{A}{S}\right) - B} \quad \textbf{(5)}$$

To apply this method to lycopene in yeast, we determined the appropriate sensitive and robust wavelengths by obtaining the absorbance spectrum of lycopene directly in yeast cells. The spectrum was determined by subtracting the optical density spectrum of a lycopene null strain yMJ105 from that of a constitutive lycopene producing strain LW2671 (Extended Data Fig. 1b). This spectrum showed the characteristic profile of lycopene absorbance and had two major absorbance maxima at 485 nm and 520 nm (Extended Data Fig. 1c). Based on this spectrum, 520 nm was chosen as the sensitive wavelength (S = 520) since it is furthest away from other natural chromophores in yeast that absorb below 500 nm (e.g. flavins). 395 nm and 600 nm were chosen as the two robust wavelengths ($R1$ = 600 and $R2$ = 395) with low absorbance from lycopene and other natural chromophores.

Three additional considerations were crucial to yield reproducible lycopene measurements in a microtiter plate format. First, all three optical density measurements (at 395 nm, 520 nm and 600 nm) were taken at the same time for each well to reduce errors due to the settling of cells during the measurement of a whole microtiter plate. Second, assay wells were blanked using a reference well on the same microtiter plate containing identical media conditions as the assay wells but with no cells. This was particularly important when colored media was used. Finally, high cell densities ($OD_{600} \geq 2$) were used to yield larger bulk lycopene signals even with the short path length of micro titer plates (~3 mm). Since these high optical density values were outside the linear range of the photodetector, all optical density values were first corrected using the following formula to give true optical density values:

$$\boldsymbol{OD_{true} = \frac{k \cdot OD_{meas}}{OD_{sat} - OD_{meas}}} \quad \textbf{(6)}$$

where OD_{meas} is the measured optical density, OD_{sat} is the saturation value of the photodetector and k is the true optical density at which the detector reaches half saturation of the measured optical density. Appropriate values for OD_{sat} and k were determined by plotting direct optical density

measurements of a range of cultures of several strains, against the true optical densities determined by dilution to the linear range. Optical densities were taken at 395 nm, 520 nm and 600 nm. All points were fit once with Equation 6 using Prism (GraphPad) to give $OD_{sat} = 3.57$ and $k = 3.16$. These values were used to correct all optical density measurements in this study.

7.2 Fungal receptor and peptide sequences

7.2.1 Fungal mating receptors in the Ste2 family

Species name	Receptor UniProt ID	Species Taxonomic ID	Family	Order
Acidomyces_richmondensis_BFW	A0A150VDK8	766039	Dothideomycetes incertae	Dothideomycetes
Acremonium_chrysogenum_strain_ATCC_11550	A0A086SWK6	857340	Hypocreales incertae sedis	Hypocreales
Ajellomyces_capsulatus_strain_G186AR	C0NQ16	447093	Ajellomycetaceae	Onygenales
Ajellomyces_capsulatus_strain_H143	C6HLQ1	544712	Ajellomycetaceae	Onygenales
Ajellomyces_capsulatus_strain_NAm1	A6QUU6	339724	Ajellomycetaceae	Onygenales
Ajellomyces_dermatitidis_strain_SLH14081	A0A179UUK7	559298	Ajellomycetaceae	Onygenales
Alternaria_alternata	A0A177DMP1	5599	Pleosporaceae	Pleosporales
Arthrobotrys_oligospora_strain_ATCC_24927	G1X8M4	756982	Orbiliaceae	Orbiliales
Arthroderma_benhamiae_strain_ATCC_MYA-4681	D4AND1	663331	Arthrodermataceae	Onygenales
Arthroderma_gypseum_strain_ATCC_MYA-4604	E5R1C9	535722	Arthrodermataceae	Onygenales
Arthroderma_otae_strain_ATCC_MYA-4605	C5FBT2	554155	Arthrodermataceae	Onygenales
Aschersonia_aleyrodis_RCEF_2490	A0A168AUR9	1081109	Clavicipitaceae	Hypocreales
Ascosphaera_apis_ARSEF_7405	A0A167VMP9	392613	Ascosphaeraceae	Onygenales
Ashbya_aceri	R9XEV1	566037	Saccharomycetaceae	Saccharomycetales
Ashbya_gossypii_strain_ATCC_10895	Q752Q1	284811	Saccharomycetaceae	Saccharomycetales
Aspergillus_calidoustus	A0A0U5CD47	454130	Aspergillaceae	Eurotiales
Aspergillus_clavatus_strain_ATCC_1007	A1CLD3	344612	Aspergillaceae	Eurotiales
Aspergillus_flavus_strain_ATCC_200026	B8NF30	332952	Aspergillaceae	Eurotiales
Aspergillus_fumigatus_Z5	A0A0J5PTK8	1437362	Aspergillaceae	Eurotiales
Aspergillus_kawachii_strain_NBRC_4308	G7XMN4	1033177	Aspergillaceae	Eurotiales
Aspergillus_lentulus	A0A0S7DJF6	293939	Aspergillaceae	Eurotiales
Aspergillus_luchuensis	A0A146FQ34	1069201	Aspergillaceae	Eurotiales
Aspergillus_niger	A0A100IM28	5061	Aspergillaceae	Eurotiales
Aspergillus_niger_strain_CBS_51388	A2QU32	425011	Aspergillaceae	Eurotiales
Aspergillus_nomius_NRRL_13137	A0A0L1J1T8	1509407	Aspergillaceae	Eurotiales
Aspergillus_ochraceoroseus	A0A0F8U8N5	138278	Aspergillaceae	Eurotiales
Aspergillus_oryzae_strain_3042	I8U4V3	1160506	Aspergillaceae	Eurotiales
Aspergillus_parasiticus_strain_ATCC_56775	A0A0F0I7R7	1403190	Aspergillaceae	Eurotiales
Aspergillus_rambellii	A0A0F8U3T7	308745	Aspergillaceae	Eurotiales
Aspergillus_ruber_CBS_135680	A0A017S298	1388766	Aspergillaceae	Eurotiales
Aspergillus_terreus_strain_NIH_2624	Q0CS34	341663	Aspergillaceae	Eurotiales
Aspergillus_udagawae	A0A0K8L9B1	91492	Aspergillaceae	Eurotiales
Aureobasidium_melanogenum_CBS_110374	A0A074VLE7	1043003	Aureobasidiaceae	Dothideales
Aureobasidium_namibiae_CBS_14797	A0A074XMD1	1043004	Aureobasidiaceae	Dothideales
Aureobasidium_pullulans_EXF-150	A0A074XT98	1043002	Aureobasidiaceae	Dothideales
Aureobasidium_subglaciale_EXF-2481	A0A074YTM0	1043005	Aureobasidiaceae	Dothideales
Baudoinia_compniacensis_strain_UAMH_10762	M2LX19	717646	Teratosphaeriaceae	Capnodiales
Beauveria_bassiana_D1-5	A0A0A2VS91	1245745	Cordycipitaceae	Hypocreales
Beauveria_bassiana_strain_ARSEF_2860	J5JMP7	655819	Cordycipitaceae	Hypocreales
Bionectria_ochroleuca	A0A0B7KEZ6	29856	Bionectriaceae	Hypocreales
Bipolaris_oryzae_ATCC_44560	W6Z6J4	930090	Pleosporaceae	Pleosporineae
Bipolaris_victoriae_FI3	W7EF59	930091	Pleosporaceae	Pleosporineae
Bipolaris_zeicola_26-R-13	W6YNK7	930089	Pleosporaceae	Pleosporineae
Blastobotrys_adeninivorans	A0A060T2K3	409370	Trichomonascaceae	Saccharomycetales
Blumeria_graminis_f_sp_hordei_strain_DH14	N1J7M2	546991	Erysiphaceae	Erysiphales
Botryosphaeria_parva_strain_UCR-NP2	R1GET9	1287680	Botryosphaeriaceae	Botryosphaeriales
Botryotinia_fuckeliana_strain_T4	G2YE05	999810	Sclerotiniaceae	Helotiales
Byssochlamys_spectabilis_strain_No_5	V5GA62	1356009	Thermoascaceae	Eurotiales
Candida_albicans_P75010	A0A0A6JZS6	1094994	Debaryomycetaceae	Saccharomycetales
Candida_albicans_strain_SC5314	Q59Q04	237561	Debaryomycetaceae	Saccharomycetales
Candida_albicans_strain_WO-1	C4YM83	294748	Debaryomycetaceae	Saccharomycetales
Candida_auris	A0A0L0P8C9	498019	Metschnikowiaceae	Saccharomycetales
Candida_dubliniensis_strain_CD36	B9WM67	573826	Debaryomycetaceae	Saccharomycetales
Candida_glabrata	A0A0W0DD93	5478	Saccharomycetaceae	Saccharomycetales
Candida_glabrata_strain_ATCC_2001	Q6FLY8	284593	Saccharomycetaceae	Saccharomycetales
Candida_maltosa_strain_Xu316	M3K0H9	1245528	Debaryomycetaceae	Saccharomycetales
Candida_orthopsilosis_strain_90-125	H8X566	1136231	Debaryomycetaceae	Saccharomycetales
Candida_parapsilosis_strain_CDC_317	G8BFM9	578454	Debaryomycetaceae	Saccharomycetales
Candida_tenuis_strain_ATCC_10573	G3BD19	590646	Debaryomycetaceae	Saccharomycetales
Candida_tropicalis_strain_ATCC_MYA-3404	C5M3P6	294747	Debaryomycetaceae	Saccharomycetales
Capronia_epimyces_CBS_60696	W9X9V4	1182542	Herpotrichiellaceae	Chaetothyriales
Capronia_semi-immersa	A0A0D2CB06	5601	Herpotrichiellaceae	Chaetothyriales
Ceratocystis_fimbriata_f_sp_platani	A0A0F8B357	88771	Ceratocystidaceae	Microascales
Chaetomium_globosum_strain_ATCC_6205	Q2GU85	306901	Chaetomiaceae	Sordariales
Chaetomium_thermophilum_strain_DSM_1495	G0S9F6	759272	Chaetomiaceae	Sordariales
Cladophialophora_bantiana_CBS_17352	A0A0D2H164	1442370	Herpotrichiellaceae	Chaetothyriales
Cladophialophora_carrionii_CBS_16054	V9D2C4	1279043	Herpotrichiellaceae	Chaetothyriales
Cladophialophora_psammophila_CBS_110553	W9VYJ4	1182543	Herpotrichiellaceae	Chaetothyriales
Cladophialophora_yegresii_CBS_114405	W9VGJ2	1182544	Herpotrichiellaceae	Chaetothyriales
Claviceps_purpurea_strain_201	M1WDR5	1111077	Clavicipitaceae	Hypocreales
Clavispora_lusitaniae_strain_ATCC_42720	C4Y9B0	306902	Metschnikowiaceae	Saccharomycetales
Coccidioides_posadasii_strain_C735	C5PF60	222929	Onygenales incertae sedis	Onygenales
Cochliobolus_heterostrophus_strain_C5	M2URM4	701091	Pleosporaceae	Pleosporineae
Cochliobolus_sativus_strain_ND90Pr	M2QUN4	665912	Pleosporaceae	Pleosporineae
Colletotrichum_fioriniae_PJ7	A0A010Q0K6	1445577	Glomerellaceae	Glomerellales
Colletotrichum_gloeosporioides_strain_Cg-14	T0K3N5	1237896	Glomerellaceae	Glomerellales
Colletotrichum_gloeosporioides_strain_Nara_gc5	L2FCZ0	1213859	Glomerellaceae	Glomerellales
Coniosporium_apollinis_strain_CBS_100218	R7YPZ5	1168221	Herpotrichiellaceae	Chaetothyriales

Cordyceps_brongniartii_RCEF_3172	A0A167IHY8	1081107	Cordycipitaceae	Hypocreales
Cordyceps_confragosa	A0A179ILG3	1105325	Cordycipitaceae	Hypocreales
Cordyceps_confragosa_RCEF_1005	A0A168IZL0	1081108	Cordycipitaceae	Hypocreales
Cordyceps_militaris_strain_CM01	G3JKW0	983644	Cordycipitaceae	Hypocreales
Cyberlindnera_fabianii	A0A061AJE3	36022	Phaffomycetaceae	Saccharomycetales
Cyberlindnera_jadinii	A0A0H5BZE0	4903	Phaffomycetaceae	Saccharomycetales
Cyphellophora_europaea_CBS_101466	W2S4E2	1220924	Cyphellophoraceae	Chaetothyriales
Debaryomyces_fabryi	A0A0V1PSR1	58627	Debaryomycetaceae	Saccharomycetales
Debaryomyces_hansenii_strain_ATCC_36239	Q6BYC0	284592	Debaryomycetaceae	Saccharomycetales
Diaporthe_ampelina	A0A0G2FGT3	1214573	Diaporthaceae	Diaporthales
Didymella_rabiei	A0A163BXA9	5454	Didymellaceae	Pleosporineae
Diplodia_seriata	A0A0G2E461	420778	Botryosphaeriaceae	Botryosphaeriales
Dothistroma_septosporum_strain_NZE10	N1Q4Q2	675120	Mycosphaerellaceae	Capnodiales
Drechmeria_coniospora	A0A151GM17	98403	Ophiocordycipitaceae	Hypocreales
Drechslerella_stenobrocha_248	W7I376	1043628	Orbiliaceae	Orbiliales
Emericella_nidulans	Q7SI72	162425	Aspergillaceae	Eurotiales
Emmonsia_crescens_UAMH_3008	A0A0G2J9S8	1247875	Ajellomycetaceae	Onygenales
Emmonsia_parva_UAMH_139	A0A0H1BAF5	1246674	Ajellomycetaceae	Onygenales
Endocarpon_pusillum_strain_Z07020	U1HY26	1263415	Verrucariaceae	Verrucariales
Eremothecium_cymbalariae	G0XP51	45285	Saccharomycetaceae	Saccharomycetales
Eremothecium_cymbalariae_strain_CBS_27075	G8JMH5	931890	Saccharomycetaceae	Saccharomycetales
Eremothecium_sinecaudum	A0A0X8HRQ0	45286	Saccharomycetaceae	Saccharomycetales
Escovopsis_weberi	A0A0M8MV01	150374	Hypocreaceae	Hypocreales
Eutypa_lata_strain_UCR-EL1	M7T4F8	1287681	Diatrypaceae	Xylariales
Exophiala_aquamarina_CBS_119918	A0A072PDE7	1182545	Herpotrichiellaceae	Chaetothyriales
Exophiala_dermatitidis_strain_ATCC_34100	H6BSM7	858893	Herpotrichiellaceae	Chaetothyriales
Exophiala_mesophila	A0A0D1X796	212818	Herpotrichiellaceae	Chaetothyriales
Exophiala_oligosperma	A0A0D2DBN2	215243	Herpotrichiellaceae	Chaetothyriales
Exophiala_sideris	A0A0D1YM75	1016849	Herpotrichiellaceae	Chaetothyriales
Exophiala_spinifera	A0A0D1YGB1	91928	Herpotrichiellaceae	Chaetothyriales
Exophiala_xenobiotica	A0A0D2C0F9	348802	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_erecta	A0A178Z6Z0	1367422	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_monophora	A0A177F142	254056	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_multimorphosa	A0A178BUX8	979981	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_multimorphosa_CBS_102226	A0A0D2JMN8	1442371	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_nubica	A0A178DBT6	856822	Herpotrichiellaceae	Chaetothyriales
Fonsecaea_pedrosoi_CBS_27137	A0A0D2EJA9	1442368	Herpotrichiellaceae	Chaetothyriales
Fusarium_langsethiae	A0A0N0DGM2	179993	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_cubense_strain_race_1	N4UWI3	1229664	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_cubense_strain_race_4	N1RVA8	1229665	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_cubense_tropical_race_4	X0KQL5	1089451	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_lycopersici_strain_4287	A0A0D2Y2Y4	426428	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_melonis_26406	X0AAF8	1089452	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_pisi_HDV247	W9PM09	1080344	Nectriaceae	Hypocreales
Fusarium_oxysporum_f_sp_raphani_54005	X0CCQ3	1089458	Nectriaceae	Hypocreales
Fusarium_oxysporum_Fo47	W9K2M0	660027	Nectriaceae	Hypocreales
Fusarium_oxysporum_FOSC_3-a	W9IAH9	909455	Nectriaceae	Hypocreales
Fusarium_oxysporum_strain_Fo5176	F9F4J6	660025	Nectriaceae	Hypocreales
Fusarium_pseudograminearum_strain_CS3096	K3V2E5	1028729	Nectriaceae	Hypocreales
Gaeumannomyces_graminis_var_tritici_strain_R3-	J3P889	644352	Magnaporthaceae	Magnaporthales
Geotrichum_candidum	A0A0J9X829	1173061	Dipodascaceae	Saccharomycetales
Gibberella_fujikuroi	A0A0J0BY83	5127	Nectriaceae	Hypocreales
Gibberella_fujikuroi_strain_CBS_19534	S0E2K7	1279085	Nectriaceae	Hypocreales
Gibberella_moniliformis_strain_M3125	W7MQM8	334819	Nectriaceae	Hypocreales
Gibberella_zeae_strain_PH-1	I1RG07	229533	Nectriaceae	Hypocreales
Glarea_lozoyensis_strain_ATCC_20868	S3DBU4	1116229	Helotiaceae	Helotiales
Grosmannia_clavigera_strain_kw1407	F0XDY3	655863	Ophiostomataceae	Ophiostomatales
Hanseniaspora_uvarum_DSM_2768	A0A0F4XDF5	1246595	Saccharomycodaceae	Saccharomycetales
Hypocrea_atroviridis_strain_ATCC_20476	G9NY94	452589	Hypocreaceae	Hypocreales
Hypocrea_jecorina	G9IJ58	51453	Hypocreaceae	Hypocreales
Hypocrea_jecorina_strain_ATCC_56765	A0A024S6P5	1344414	Hypocreaceae	Hypocreales
Hypocrea_jecorina_strain_QM6a	G0RMK2	431241	Hypocreaceae	Hypocreales
Hypocrea_virens_strain_Gv29-8	G9MQ44	413071	Hypocreaceae	Hypocreales
Hypocrella_siamensis	A0A172Q4C2	696354	Clavicipitaceae	Hypocreales
Isaria_fumosorosea_ARSEF_2679	A0A167XIR1	1081104	Cordycipitaceae	Hypocreales
Kazachstania_africana_strain_ATCC_22294	H2ASI7	1071382	Saccharomycetaceae	Saccharomycetales
Kazachstania_naganishii_strain_ATCC_MYA-139	J7RM21	1071383	Saccharomycetaceae	Saccharomycetales
Kluyveromyces_dobzhanskii_CBS_2104	A0A0A8LC24	1427455	Saccharomycetaceae	Saccharomycetales
Kluyveromyces_lactis_strain_ATCC_8585	Q6CIP0	284590	Saccharomycetaceae	Saccharomycetales
Kluyveromyces_marxianus_DMKU3-1042	W0TFI2	1003335	Saccharomycetaceae	Saccharomycetales
Komagataella_pastoris_strain_GS115	C4R6X5	644223	Phaffomycetaceae	Saccharomycetales
Kuraishia_capsulata_CBS_1993	W6MJ91	1382522	Saccharomycetales incertae	Saccharomycetales
Lachancea_kluyveri	P12384	4934	Saccharomycetaceae	Saccharomycetales
Lachancea_lanzarotensis	A0A0C7N6G7	1245769	Saccharomycetaceae	Saccharomycetales
Lachancea_quebecensis	A0A0P1KZX7	1654605	Saccharomycetaceae	Saccharomycetales
Lachancea_thermotolerans_strain_ATCC_56472	C5DBK0	559295	Saccharomycetaceae	Saccharomycetales
Leptosphaeria_maculans_strain_JN3	E5A529	985895	Leptosphaeria	Pleosporineae
Lodderomyces_elongisporus_strain_ATCC_11503	A5E1D9	379508	Debaryomycetaceae	Saccharomycetales
Macrophomina_phaseolina_strain_MS6	K2S5Z6	1126212	Botryosphaeriaceae	Botryosphaeriales
Madurella_mycetomatis	A0A175W3I2	100816	mitosporic Sordariales	Sordariales
Magnaporthe_oryzae_strain_70-15	G4MR89	242507	Magnaporthaceae	Magnaporthales
Magnaporthe_oryzae_strain_Y34	L7HVB4	1143189	Magnaporthaceae	Magnaporthales
Magnaporthiopsis_poae_strain_ATCC_64411	A0A0C4DS73	644358	Magnaporthaceae	Magnaporthales
Marssonina_brunnea_f_sp_multigermtubi_strain_MB	K1X8D8	1072389	Dermateaceae	Helotiales
Metarhizium_acridum_strain_CQMa_102	E9DXW9	655827	Clavicipitaceae	Hypocreales
Metarhizium_album_ARSEF_1941	A0A0B2WQA5	1081103	Clavicipitaceae	Hypocreales
Metarhizium_anisopliae_ARSEF_549	A0A0B4EKU5	1276135	Clavicipitaceae	Hypocreales
Metarhizium_anisopliae_BRIP_53293	A0A0D9NQS0	1291518	Clavicipitaceae	Hypocreales
Metarhizium_brunneum_ARSEF_3297	A0A0B4FKS3	1276141	Clavicipitaceae	Hypocreales
Metarhizium_guizhouense_ARSEF_977	A0A0B4H8M1	1276136	Clavicipitaceae	Hypocreales

Metarhizium_majus_ARSEF_297	A0A0B4HXD6	1276143	Clavicipitaceae	Hypocreales
Metarhizium_rileyi_RCEF_4871	A0A167AMF2	1081105	Clavicipitaceae	Hypocreales
Metarhizium_robertsii	A0A014PAK1	568076	Clavicipitaceae	Hypocreales
Metarhizium_robertsii_strain_ARSEF_23	E9EMS3	655844	Clavicipitaceae	Hypocreales
Meyerozyma_guilliermondii_strain_ATCC_6260	A5DFC0	294746	Debaryomycetaceae	Saccharomycetales
Naumovozyma_castellii_strain_ATCC_76901	G0VD13	1064592	Saccharomycetaceae	Saccharomycetales
Naumovozyma_dairenensis_strain_ATCC_10597	G0WE84	1071378	Saccharomycetaceae	Saccharomycetales
Nectria_haematococca_strain_77-13-4	C7ZA34	660122	Nectriaceae	Hypocreales
Neonectria_ditissima	A0A0P7AWF2	78410	Nectriaceae	Hypocreales
Neosartorya_fischeri_strain_ATCC_1020	A1D5Z2	331117	Aspergillaceae	Eurotiales
Neosartorya_fumigata_strain_CEA10	B0XZZ4	451804	Aspergillaceae	Eurotiales
Neurospora_africana	K7ZVW9	5143	Sordariaceae	Sordariales
Neurospora_calospora	K7ZWV9	165411	Sordariaceae	Sordariales
Neurospora_cerealis	K7ZW01	29881	Sordariaceae	Sordariales
Neurospora_crassa	D2N2E0	5141	Sordariaceae	Sordariales
Neurospora_crassa_strain_ATCC_24698	Q1K6I3	367110	Sordariaceae	Sordariales
Neurospora_galapagosensis	K7ZWN2	88769	Sordariaceae	Sordariales
Neurospora_hapsidophora	K7ZW48	176947	Sordariaceae	Sordariales
Neurospora_intermedia	D2N2E7	5142	Sordariaceae	Sordariales
Neurospora_kobi	K7ZVX0	241062	Sordariaceae	Sordariales
Neurospora_lineolata	K7ZWW0	88717	Sordariaceae	Sordariales
Neurospora_novoguineensis	K7ZW03	241060	Sordariaceae	Sordariales
Neurospora_pannonica	K7ZWN3	83678	Sordariaceae	Sordariales
Neurospora_retispora	K7ZW49	241054	Sordariaceae	Sordariales
Neurospora_santi-florii	K7ZVX1	176682	Sordariaceae	Sordariales
Neurospora_sitophila	D2N2F3	40126	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8780	D2N2G4	482004	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8815	D2N2F6	228687	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8817	D2N2F7	481997	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8827	D2N2G3	482003	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8842	D2N2G2	482002	Sordariaceae	Sordariales
Neurospora_sp_FGSC_8853	D2N2F9	481999	Sordariaceae	Sordariales
Neurospora_sublineolata	K7ZWW1	165293	Sordariaceae	Sordariales
Neurospora_terricola	K7ZWN4	88718	Sordariaceae	Sordariales
Neurospora_tetrasperma	D2N2F4	40127	Sordariaceae	Sordariales
Neurospora_uniporata	K7ZW50	241063	Sordariaceae	Sordariales
Ogataea_parapolymorpha_strain_ATCC_26012	W1QE65	871575	Pichiaceae	Saccharomycetales
Oidiodendron_maius_Zn	A0A0C3HTW3	913774	mitosporic Myxotrichaceae	Leotiomycetes incertae sedis
Ophiocordyceps_sinensis_strain_Co18	T5A148	911162	Ophiocordycipitaceae	Hypocreales
Ophiocordyceps_unilateralis	A0A0L9SIN1	268505	Ophiocordycipitaceae	Hypocreales
Ophiostoma_piceae_strain_UAMH_11346	S3C5N9	1262450	Ophiostomataceae	Ophiostomatales
Paracoccidioides_brasiliensis_strain_Pb03	C0SDN9	482561	Onygenales incertae sedis	Onygenales
Paracoccidioides_brasiliensis_strain_Pb18	C1GFU7	502780	Onygenales incertae sedis	Onygenales
Paracoccidioides_lutzii_strain_ATCC_MYA-826	C1H517	502779	Onygenales incertae sedis	Onygenales
Paraphaeosphaeria_sporulosa	A0A177CPX6	1460663	Didymosphaeriaceae	Massarineae
Penicillium_brasilianum	A0A0F7TPZ2	104259	Aspergillaceae	Eurotiales
Penicillium_camemberti_FM_013	A0A0G4P840	1429867	Aspergillaceae	Eurotiales
Penicillium_chrysogenum	B1GVB8	5076	Aspergillaceae	Eurotiales
Penicillium_digitatum_strain_PHI26	K9G3Z6	1170229	Aspergillaceae	Eurotiales
Penicillium_expansum	A0A0A2K1S7	27334	Aspergillaceae	Eurotiales
Penicillium_freii	A0A101MNI9	48697	Aspergillaceae	Eurotiales
Penicillium_italicum	A0A0A2LAS4	40296	Aspergillaceae	Eurotiales
Penicillium_nordicum	A0A0M8PFN9	229535	Aspergillaceae	Eurotiales
Penicillium_oxalicum_strain_114-2	S7Z940	933388	Aspergillaceae	Eurotiales
Penicillium_patulum	A0A135LCC8	5078	Aspergillaceae	Eurotiales
Penicillium_roqueforti_strain_FM164	W6PVN7	1365484	Aspergillaceae	Eurotiales
Pestalotiopsis_fici_W106-1	W3XDQ7	1229662	Sporocadaceae	Xylariales
Phaeomoniella_chlamydospora	A0A0G2HF89	158046	Phaeomoniellales incertae	Phaeomoniellales
Phaeosphaeria_nodorum_strain_SN15	Q0UCT8	321614	Phaeosphaeriaceae	Pleosporineae
Pichia_kudriavzevii	A0A099NXR5	4909	Pichiaceae	Saccharomycetales
Pichia_sorbitophila_strain_ATCC_MYA-4447	G8YMJ7	559304	Debaryomycetaceae	Saccharomycetales
Pichia_sorbitophila_strain_ATCC_MYA-4447	G8YMZ0	559304	Debaryomycetaceae	Saccharomycetales
Pneumocystis_carinii	A2TJ26	4754	Pneumocystidaceae	Pneumocystidomycetes
Pneumocystis_carinii_B80	A0A0W4ZHE5	1408658	Pneumocystidaceae	Pneumocystidomycetes
Pneumocystis_jiroveci_strain_SE8	L0PDU6	1209962	Pneumocystidaceae	Pneumocystidomycetes
Pneumocystis_jirovecii_RU7	A0A0W4ZVY3	1408657	Pneumocystidaceae	Pneumocystidomycetes
Pneumocystis_murina_strain_B123	M7P3B3	1069680	Pneumocystidaceae	Pneumocystidomycetes
Pochonia_chlamydosporia_170	A0A179FF27	1380566	Clavicipitaceae	Hypocreales
Podospora_anserina_strain_S	B2ADL1	515849	Lasiosphaeriaceae	Sordariales
Pseudocercospora_fijiensis_strain_CIRAD86	N1Q996	383855	Mycosphaerellaceae	Capnodiales
Pseudogymnoascus_destructans	A0A177ADM2	655981	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_destructans_strain_ATCC_MYA	L8G637	658429	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-103	A0A094E1R1	1420912	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-3557	A0A093XIK8	1437433	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-3775	A0A094AA23	1420901	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-3808	A0A093YGI7	1391699	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4246	A0A093Z5B5	1420902	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4281_FW-2241	A0A094CRD8	1420906	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4513_FW-928	A0A094BQ07	1420907	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4515_FW-2607	A0A094FEM7	1420909	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4516_FW-969	A0A094CTP6	1420910	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4517_FW-2822	A0A094FK10	1420911	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4518_FW-2643	A0A094ET92	1420913	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4519_FW-2642	A0A094K4N9	1420914	Pseudeurotiaceae	Leotiomycetes incertae sedis
Pseudogymnoascus_sp_VKM_F-4520_FW-2644	A0A094JHH7	1420915	Pseudeurotiaceae	Leotiomycetes incertae sedis
Purpureocillium_lilacinum	A0A179GB12	33203	Ophiocordycipitaceae	Hypocreales
Pyrenochaeta_sp_DS3sAY3a	A0A178DZ21	765867	Cucurbitariaceae	Pleosporineae
Pyrenophora_teres_f_teres_strain_0-1	E3RI43	861557	Pleosporaceae	Pleosporineae
Pyrenophora_tritici-repentis_strain_Pt-1C-BFP	B2WIP5	426418	Pleosporaceae	Pleosporineae
Pyronema_omphalodes_strain_CBS_100304	U4LPJ5	1076935	Pyronemataceae	Pezizales

Rasamsonia_emersonii_CBS_39364	A0A0F4YHC8	1408163	Trichocomaceae	Eurotiales
Rhinocladiella_mackenziei_CBS_65093	A0A0D2H556	1442369	Herpotrichiellaceae	Chaetothyriales
Saccharomyces_arboricola_strain_H-6	J8Q5L6	1160507	Saccharomycetaceae	Saccharomycetales
Saccharomyces_bayanus	Q8J1R6	4931	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_strain_ATCC_204508	D6VTK4	559292	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_strain_AWRI796	E7KC22	764097	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_strain_FostersO	E7NH73	764101	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_strain_RM11-1a	B3LUI5	285006	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_strain_YJM789	A7A213	307796	Saccharomycetaceae	Saccharomycetales
Saccharomyces_cerevisiae_x_Saccharomyces_kudri	H0GU93	1095631	Saccharomycetaceae	Saccharomycetales
Saccharomyces_paradoxus	Q8J080	27291	Saccharomycetaceae	Saccharomycetales
Saccharomyces_pastorianus	Q8J1Q4	27292	Saccharomycetaceae	Saccharomycetales
Saccharomyces_sp_'boulardii'	A0A0L8VRV2	252598	Saccharomycetaceae	Saccharomycetales
Saitoella_complicata_NRRL_Y-17804	A0A0E9NKH5	698492	Protomycetaceae	Taphrinales
Scedosporium_apiospermum	A0A084FZY6	563466	Microascaceae	Microascales
Scheffersomyces_stipitis_strain_ATCC_58785	A3LXU7	322104	Debaryomycetaceae	Saccharomycetales
Schizosaccharomyces_cryophilus_strain_OY26	S9VVX5	653667	Schizosaccharomycetaceae	Schizosaccharomycetales
Schizosaccharomyces_japonicus_strain_yFS275	B6JZE2	402676	Schizosaccharomycetaceae	Schizosaccharomycetales
Schizosaccharomyces_octosporus_strain_yFS286	S9PVP9	483514	Schizosaccharomycetaceae	Schizosaccharomycetales
Schizosaccharomyces_pombe_strain_972	Q00619	284812	Schizosaccharomycetaceae	Schizosaccharomycetales
Sclerotinia_borealis_F-4157	W9C8T9	1432307	Sclerotiniaceae	Helotiales
Sclerotinia_sclerotiorum_strain_ATCC_18683	A7EY95	665079	Sclerotiniaceae	Helotiales
Setosphaeria_turcica_strain_28A	R0KC11	671987	Pleosporaceae	Pleosporineae
Sordaria_macrospora_strain_ATCC_MYA-333	F7W5S1	771870	Sordariaceae	Sordariales
Spathaspora_passalidarum_strain_NRRL_Y-27907	G3AJU2	619300	Debaryomycetaceae	Saccharomycetales
Sphaerulina_musiva_strain_SO2202	N1QN82	692275	Mycosphaerellaceae	Capnodiales
Sporothrix_brasiliensis_5110	A0A0C2IIS5	1398154	Ophiostomataceae	Ophiostomatales
Sporothrix_insectorum_RCEF_264	A0A162MTF1	1081102	Ophiostomataceae	Ophiostomatales
Sporothrix_schenckii	H9XTI1	29908	Ophiostomataceae	Ophiostomatales
Sporothrix_schenckii_1099-18	A0A0F2M7E2	1397361	Ophiostomataceae	Ophiostomatales
Sporothrix_schenckii_strain_ATCC_58251	U7Q511	1391915	Ophiostomataceae	Ophiostomatales
Stachybotrys_chartarum_IBT_40288	A0A084RP20	1283842	Stachybotriaceae	Hypocreales
Stachybotrys_chartarum_IBT_7711	A0A084ASH4	1280523	Stachybotriaceae	Hypocreales
Stachybotrys_chlorohalonata_IBT_40285	A0A084QT65	1283841	Stachybotriaceae	Hypocreales
Stagonospora_sp_SRC1lsM3a	A0A178ACM9	765868	Massarinaceae	Massarineae
Stemphylium_lycopersici	A0A0L1HGK2	183478	Pleosporaceae	Pleosporineae
Sugiyamaella_lignohabitans	A0A161HL65	796027	Trichomonascaceae	Saccharomycetales
Talaromyces_islandicus	A0A0U1LRR7	28573	Trichocomaceae	Eurotiales
Talaromyces_marneffei_PM1	A0A093XYN6	1077442	Trichocomaceae	Eurotiales
Talaromyces_marneffei_strain_ATCC_18224	B6Q4A9	441960	Trichocomaceae	Eurotiales
Talaromyces_stipitatus_strain_ATCC_10500	B8M557	441959	Trichocomaceae	Eurotiales
Tetrapisispora_blattae_strain_ATCC_34711	I2H305	1071380	Saccharomycetaceae	Saccharomycetales
Tetrapisispora_phaffii_strain_ATCC_24235	G8C206	1071381	Saccharomycetaceae	Saccharomycetales
Togninia_minima_strain_UCR-PA7	R8BGY4	1286976	Togniniaceae	Togniniales
Tolypocladium_ophioglossoides_CBS_100239	A0A0L0N0N3	1163406	Ophiocordycipitaceae	Hypocreales
Torrubiella_hemipterigena	A0A0A1SZJ6	1531966	Clavicipitaceae	Hypocreales
Torulaspora_delbrueckii_strain_ATCC_10662	G8ZR18	1076872	Saccharomycetaceae	Saccharomycetales
Trichoderma_gamsii	A0A0W7VR33	398673	Hypocreaceae	Hypocreales
Trichoderma_harzianum	A0A0F9XI50	5544	Hypocreaceae	Hypocreales
Trichophyton_equinum_strain_ATCC_MYA-4606	F2PNP9	559882	Arthrodermataceae	Onygenales
Trichophyton_interdigitale_MR816	A0A059J435	1215338	Arthrodermataceae	Onygenales
Trichophyton_rubrum	A0A178ETN9	5551	Arthrodermataceae	Onygenales
Trichophyton_rubrum_CBS_28886	A0A022VRI2	1215330	Arthrodermataceae	Onygenales
Trichophyton_verrucosum_strain_HKI_0517	D4DBK6	663202	Arthrodermataceae	Onygenales
Trichophyton_violaceum	A0A178FB33	34388	Arthrodermataceae	Onygenales
Tuber_melanosporum_strain_Mel28	D5GJK5	656061	Tuberaceae	Pezizales
Uncinocarpus_reesii_strain_UAMH_1704	C4JL18	336963	Onygenaceae	Onygenales
Uncinula_necator	A0A0B1P9N6	52586	Erysiphaceae	Erysiphales
Ustilaginoidea_virens	A0A063BN49	1159556	Hypocreales incertae sedis	Hypocreales
Vanderwaltozyma_polyspora_strain_ATCC_22028	A7TJQ6	436907	Saccharomycetaceae	Saccharomycetales
Vanderwaltozyma_polyspora_strain_ATCC_22028	A7TQX4	436907	Saccharomycetaceae	Saccharomycetales
Verruconis_gallopava	A0A0D2AMB2	253628	Sympoventuriaceae	Venturiales
Verticillium_alfalfae_strain_VaMs102	C9SGY3	526221	Plectosphaerellaceae	Glomerellales
Verticillium_dahliae_strain_VdLs17	G2X5W7	498257	Plectosphaerellaceae	Glomerellales
Verticillium_longisporum	A0A0G4M417	100787	Plectosphaerellaceae	Glomerellales
Wickerhamomyces_ciferrii_strain_F-60-10	K0KPE3	1206466	Phaffomycetaceae	Saccharomycetales
Xylona_heveae_TC161	A0A165HIN9	1328760	Xylonomycetaceae	Xylonomycetales
Yarrowia_lipolytica_strain_CLIB_122	Q6C2Z3	284591	Dipodascaceae	Saccharomycetales
Zygosaccharomyces_bailii_ISA1307	W0VI75	1355161	Saccharomycetaceae	Saccharomycetales
Zygosaccharomyces_bailii_strain_CLIB_213	S6EXB4	1333698	Saccharomycetaceae	Saccharomycetales
Zygosaccharomyces_rouxii_strain_ATCC_2623	C5DX97	559307	Saccharomycetaceae	Saccharomycetales
Zymoseptoria_brevis	A0A0F4GDL4	1047168	Mycosphaerellaceae	Capnodiales
Zymoseptoria_tritici_strain_CBS_115943	F9X131	336722	Mycosphaerellaceae	Capnodiales

7.2.2 Fungal secreted peptide pheromones

Species	Gene ID	Predicted Peptide Sequence	Previously predicted in [27]
Alternaria_brasicicola	ACIW01002317	WSFTQKRPYGLPIG	*
Arthrobotrys_oligospora	G1X8M4	WCPYNSCP	
Ashbya_aceri	R9XEV1	WHWLRFGDGQSM	
Ashbya_gossypii	Q752Q1	WFRLSLHHGQSM	
Aspergillus_clavatus	A1CLD3	QWCELPGQGCYMI	*
Aspergillus_flavus	B8NF30	WCSLPAQGCYML	*
Aspergillus_fumigata	Q4WYU8	WCHLPGQGCYML	*
Aspergillus_kawachii	G7XMN4	WCHLPGQPCNMI	
Aspergillus_nidulans	Q5BAB0	WCRFAGRICPPT	*
Aspergillus_niger	G3XMV3	WCVLPGQPCNMI	*
Aspergillus_oryzae	Q2U819	WCALPGQGC	*
Aspergillus_ruber	A0A017S298	WCALPGQICS	
Aspergillus_terreus	Q0CS34	WCWLPGQGCYML	*
Baudoinia_compniacensis	M2LX19	GWIGRCGVPGSSC	
Beauveria_bassiana	J5JMP7	WCMRPGQPCW	
Botryosphaeria_parva	R1GET9	WCRWKGQPCS	
Botrytis_cinerea	G2YE05	WCGRPGQPC	*
Candida_albicans	Q59Q04	GFRLTNFGYFEPG	*
Candida_dubliniensis	B9WM67	KFKLTNFGYFEPG	*
Candida_glabrata	Q6FLY8	WHWVRLRKGQGLF	*
Candida_guilliermondii	A5DFC0	KKNSRFLTYWFFQPIM	
Candida_lusitaniae	C4Y9B0	WKWIKFRNTDVIG	
Candida_parapsilosis	G8BFM9	KPHWTTYGYYEPQ	*
Candida_tenuis	G3BD19	FSWNYRLKWQPIS	
Candida_tropicalis	C5M3P6	KFKFRLTRYGWFSPN	*
Capronia_coronata	W9Y1I9	LSYWKGVNDGGSS	
Capronia_epimyces	W9X9V4	LSYWAGVNDGGSS	
Chaetomium_globosum	Q2GU85	WCKQFLGMPCW	*
Chaetomium_thermophilum	G0S9F6	SWCTRFPGQPCW	
Chryphonectria_parasitica	O14431	WCLFHGEGCW	*
Claviceps_purpurea	M1WDR5	WCWRPGQGCW	
Coccidioides_immitis	J3KG99	WCQRPGEPC	
Colletotrichum_gloeosporioides	T0K3N5	WCTKPGQPCW	
Coniosporium_apollinis	R7YPZ5	WGSRFCHKTGQGCP	
Dactylellina_haptotyla	S8AWC4	WCVYNSCP	
Debaryomyces_hansenii	Q6BYC0	KFHWMTYRFFQPNL	*
Endocarpon_pusillum	U1HY26	WWGFRWSRHGTSSW	
Eremothecium_cymbalariae	G8JMH5	WHWLRFDRGQPIH	
Fusarium_oxysporum	F9F4J6	WCTWRGQPCW	*
Fusarium_pseudograminearum	K3V2E5	WCTWKGQPCW	
Gaeumannomyces_graminis	J3P889	QNGCQYRGQSCW	
Geotrichum_candidum	A0A024JBH3	DWGWFWYVPRPGDPAM	
Gibberella_fujikuroi	S0E2K7	WCTWRGQPCW	
Gibberella_moniliformis	W7MQM8	WCTWRGQPCW	
Gibberella_zeae	I1RG07	WCWWKGQPCW	*
Glarea_lozoyensis	S3DBU4	QCIRHGQPCW	
Grosmannia_clavigera	F0XDY3	QWCQWYGQACW	
Kazachstania_africana	H2ASI7	WHWLSIAPGQPMYI	
Kazachstania_naganishii	J7RM21	WHWLRLSYGQPIY	
Kluyveromyces_lactis	Q6CIP0	WSWITLRPGQPIF	*
Kluyveromyces_marxianus	W0TFI2	WKWLSLRVGQPIY	
Kluyveromyces_waltii	AADM01000052	WRWLSLARGQPMY	*
Komagataella_pastoris	F2R066	FRWRNNEKNQPFG	
Kuraishia_capsulata	W6MJ91	RLGARIYAKGQPIY	
Lachancea_kluyveri	P12384	WHWLSFSKGEPMY	
Lachancea_thermotolerans	C5DBK0	WRWLSLSRGQPMY	*
Lodderomyces_elongisporus	A5E1D9	WMWTRYGRFSPV	*
Magnaporthe_oryzae	G4MR89	QWCPRRGQPCW	*
Magnaporthe_poae	M4FRS1	QNGCPYPGQSCW	
Marssonina_brunnea	K1X8D8	CGYRGQPCP	
Metarhizium_acridum	E9DXW9	WCWQPGQPCW	
Metarhizium_anisopliae	E9EMS3	WCWRPGQPCW	
Mycosphaerella_graminicola	F9X131	GNSFVGWCGAIGAPCA	

Mycosphaerella_pini	N1Q4Q2	GVLTRCTVPGLACG	
Nectria_haematococca	C7ZA34	WCFYPGQPCW	
Neosartorya_fischeri	A1D5Z2	WCHLPGQGCYML	*
Neurospora_crassa	Q1K6I3	QWCRIHGQSCW	*
Neurospora_tetrasperma	F8MS57	QWCRIHGQSCW	
Ogataea_parapolymorpha	W1QE65	WGWHRVNRNEVIF	
Ophiostoma_piceae	S3C5N9	QWCPMVGQPCW	
Paracoccidioides_lutzii	C1H517	WCTRPGQGC	
Penicillium_chrysogenum	B6H2Y5	WCGHIGQGCY	*
Penicillium_digitatum	K9GDZ2	WCGHIGQGCY	
Penicillium_oxalicum	S7Z940	WCAHPGQGCA	
Penicillium_roqueforti	W6PVN7	WCGHIGQGCY	
Phaeosphaeria_nodorum	Q0UCT8	YNGWRYRPYGLPVG	*
Pichia_sorbitophila	G8YMJ7	FHWFKYNKYDPIT	
Podospora_anserina	B2ADL1	QWCLRFVGQSCW	*
Pseudogymnoascus_destructans	L8G637	FCWRPGQPCG	
Pyrenophora_teres_f_teres	E3RI43	VTWTQKRPYGMPVG	
Pyrenophora_tritici-repentis	B2WIP5	SWTQKRPYGMPVG	*
Saccharomyces_bayanus	Q8J1R6	WHWLQLKPGQPMY	*
Saccharomyces_castellii	G0VD13	NWHWLRLDPGQPLY	*
Saccharomyces_cerevisiae	P0CI39	WHWLQLKPGQPMY	*
Saccharomyces_dairenensis	G0WE84	WHWLRLDPGQPLY	*
Saccharomyces_mikatae	AACH01001097	WHWLQLKPGQPMY	*
Saccharomyces_paradoxis	Q8J094	WHWLQLKPGQPMY	*
Scheffersomyces_stipitis	A3LXU7	WHWTSYGVFEPG	*
Schizosaccharomyces_japonicus	B6JZE2	VSDRVKQMLSHWWNFRNPDTANL	*
Schizosaccharomyces_octosporus	S9PVP9	KTYEDFLRVYKNWWSFQNPDRPDL	*
Schizosaccharomyces_pombe	Q00619	KTYADFLRAYQSWNTFVNPDRPNL	*
Sclerotinia_borealis	W9C8T9	WCGRPGQPC	
Sclerotinia_sclerotiorum	A7EY95	WCGRPGQPC	*
Sordaria_macrospora	F7W5S1	QWCRIHGQSCW	*
Sporothrix_schenckii	H9XTI1	YCPLKGQSCW	
Tetrapisispora_blattae	I2H305	HWLRLGRGEPLY	
Tetrapisispora_phaffii	G8C206	WHWLRLDPGQPLY	
Thielavia_heterothallica	G2QGA8	WCVQFLGMPCW	
Togninia_minima	R8BGY4	WCTKHGQSCW	
Torulaspora_delbrueckii	G8ZR18	GWMRLRLGQPL	
Trichoderma_atroviridis	G9NY94	WCWRVGESCW	*
Trichoderma_jecorina	G0RMK2	WCYRIGEPCW	*
Trichoderma_virens	G9MQ44	WCYRVGMTCGW	*
Tuber_melanosporum	D5GJK5	WTPRPGRGAY	
Vanderwaltozyma_polyspora_1	A7TJQ6	WHWLELDNGQPIY	*
Vanderwaltozyma_polyspora_2	A7TQX4	WHWLRLRYGEPIY	*
Verticillium_alfalfae	C9SGY3	PCPRPGQGCW	*
Verticillium_dahliae	G2X5W7	PCPRPGQGCW	*
Wickerhamomyces_ciferrii	K0KPE3	WQWRKYLNGSPNY	
Yarrowia_lipolytica	Q6C2Z3	WRWFWLPGYGEPNW	*
Zygosaccharomyces_bailii	S6EXB4	HLVRLSPGAAMF	
Zygosaccharomyces_rouxii	C5DX97	HFIELDPGQPMF	*

7.3 Computer scripts

7.3.1 Automated gating with the R package openCyto

```
library(flowWorkspace)
library(openCyto) #needs to be >=1.11.8, from Bioc Devel as of Aug 2016
library(ggcyto)
library(flowIncubator)  #install from GitHub
library(parallel)

numCores <- detectCores()-1

# ask for input folder
inFolder <- readline(prompt="Enter path of folder FCS files: ")

#detect/ask for CellTag index file
ctFile <- list.files(path = inFolder, pattern = "CellTag-Strain", full = TRUE)
if (length(ctFile)==0) ctFile <- readline(prompt="Enter path of CellTag-Strain.csv index file: ")
CTindex = read.csv(ctFile, stringsAsFactors = FALSE)

#extract CT names and corresponding strain identities
CellTagName <- CTindex[["CellTag"]]
StrainName <- CTindex[["Strain"]]

#detect/ask for tube identities file
tubeFile <- list.files(path = inFolder, pattern = "Tube-Treatment", full = TRUE)
if (length(tubeFile)==0) tubeFile <- readline(prompt="Enter path of Tube-Treatment.csv index file: ")
Tubeindex = read.csv(tubeFile, stringsAsFactors = FALSE)
Tubeindex$gsIndex <- 0 #gtIndex column

# get names of csv files to analyze, path defaulting to current working directory
fcsFilePaths <- list.files(path = inFolder, pattern = "*.fcs", full = TRUE)

# load FCS files
fs  <- read.flowSet(fcsFilePaths, alter.names=TRUE)

# create folder to output files
dir.create(file.path(inFolder, "output"))

#extract parameter names and display
channelNames <- colnames(fs)
cat("The detected parameters are:", channelNames,sep="\n")


##### set normalization parameters
# ask for correct channels
gfpName <- readline(prompt="Enter name of GFP channel: ")
rfpName <- readline(prompt="Enter name of RFP channel: ")
bfpName <- readline(prompt="Enter name of BFP channel: ")
sscName <- readline(prompt="Enter name of SSC Area channel: ")
fscName <- readline(prompt="Enter name of FSC Area channel: ")
#fschName <- readline(prompt="Enter name of FSC Height channel: ")

scaleBy1 <- 100000 #factor to scale ssc normalized fluorescence to match original scale of fluorescence
minNormFluor <- -1000 #minimum G.n,R.n,B.n cutoff to avoid extreme neg values generated by scaling, only
important for logicle estimation

ratioRange <- 10 #expected range of log10(R.n/G.n), only needed for auto plotting
logicleMin <- -0.5 #expected minimum value of logicle(G.n,R.n,B.n), only needed for auto plotting


cat("Starting analysis...\n")
if(gfpName == "") gfpName <- "GFP.A"
if(rfpName == "") rfpName <- "mCherry.A"
if(bfpName == "") bfpName <- "Cascade.Blue.A"
if(sscName == "") sscName <- "SSC.A"
if(fscName == "") fscName <- "FSC.A"
```

```
gfpIndex <- which(colnames(fs)==gfpName)
rfpIndex <- which(colnames(fs)==rfpName)
bfpIndex <- which(colnames(fs)==bfpName)
sscIndex <- which(colnames(fs)==sscName)
fscIndex <- which(colnames(fs)==fscName)
#fschIndex <- which(colnames(fs)==fschName)


# rename columns for custom transform
colnames(fs)[gfpIndex] <- "G"
colnames(fs)[rfpIndex] <- "R"
colnames(fs)[bfpIndex] <- "B"
colnames(fs)[sscIndex] <- "S.A"
colnames(fs)[fscIndex] <- "F.A" # can't use just "F", reserved for false
#colnames(fs)[fschIndex] <- "F.H"


# normalize by SSC
cat("Normalizing by SSC...")
fs <- transform(fs, G.n = scaleBy1*(G/S.A), R.n = scaleBy1*(R/S.A), B.n = scaleBy1*(B/S.A))
cat("Done\n")

#truncate for extreme negs produced by SSC normalization/scaling
truncTrans <- truncateTransform(transformationId="truncTrans", a=minNormFluor)
fs <- transform(fs, G.n=truncTrans(G.n), R.n=truncTrans(R.n), B.n=truncTrans(B.n))


# extract fluor total and R:G fluor ratio as log
cat("Calculating R:G ratio and total...")
fs <- transform(fs, fTotal = log10(G.n+R.n), fRatio = log10(R.n/G.n))
cat("Done\n")


# clean up min max for better plotting and logicle transform
p <- parameters(fs[[1]])
G.nIndex <- which(colnames(fs)=="G.n")
R.nIndex <- which(colnames(fs)=="R.n")
B.nIndex <- which(colnames(fs)=="B.n")
fTotalIndex <- which(colnames(fs)=="fTotal")
fRatioIndex <- which(colnames(fs)=="fRatio")
p[['maxRange']][G.nIndex] <- p[['maxRange']][gfpIndex]
p[['maxRange']][R.nIndex] <- p[['maxRange']][rfpIndex]
p[['maxRange']][B.nIndex] <- p[['maxRange']][bfpIndex]
p[['maxRange']][fTotalIndex] <- log10(p[['maxRange']][gfpIndex] + p[['maxRange']][rfpIndex])
p[['maxRange']][fRatioIndex] <- ratioRange/2
p[['minRange']][G.nIndex] <- p[['minRange']][gfpIndex]
p[['minRange']][R.nIndex] <- p[['minRange']][rfpIndex]
p[['minRange']][B.nIndex] <- p[['minRange']][bfpIndex]
p[['minRange']][fTotalIndex] <- 0
p[['minRange']][fRatioIndex] <- -ratioRange/2
L <- length(fs)
res <- mapply(1:L, FUN = function(x){parameters(fs[[x]]) <- p})

# transform with Logicle transform
cat("Logicle transform of G,R,B...")
transFuncts <- estimateLogicle(fs[[1]], channels = c("G.n", "R.n", "B.n"))
fs <- transform(fs, transFuncts)
cat("Done\n")

#clean up minRange after logical transform
p <- parameters(fs[[1]])
p[['minRange']][G.nIndex] <- logicleMin
p[['minRange']][R.nIndex] <- logicleMin
p[['minRange']][B.nIndex] <- logicleMin
L <- length(fs)
res <- mapply(1:L, FUN = function(x){parameters(fs[[x]]) <- p})
```

```
# clean up desc column (marker names) automatically generated by transform()
chnls <- colnames(fs)
markers <- rep(NA_character_,length(chnls)) # use R's NA character value
names(markers) <- chnls
markernames(fs) <- markers

### create custom gating methods
#mindensity method were peaks to use are selected by left to right rank order
.specificMindensity <- function(fr, pp_res, channels=NA, filterId="specificMindensity", selectedPeak=1,
totalPeaks=NULL, selectFromRight=FALSE, findAdjust=2, peakMeanDensityThreshold=0.1, minAdjust=1, ...){

        x <- exprs(fr)[, channels]
        peaks <- openCyto:::.find_peaks(x, num_peaks=totalPeaks, adjust=findAdjust, ...)
        meanPeaks <- mean(peaks[,"y"])
        peaks <- peaks[peaks[,"y"] > (meanPeaks*peakMeanDensityThreshold), "x"] #reject minor peaks below
average
        peaks <- sort(peaks, decreasing=selectFromRight)
        min <- peaks[selectedPeak]
        max <- peaks[selectedPeak + 1]


    my_gate <- openCyto:::.mindensity(fr, channel=channels, filterId = filterId, min=min, max=max,
adjust=minAdjust, ...)
    return(my_gate)
}
registerPlugins(fun=.specificMindensity,methodName='specificMindensity',dep=NA)


#register gating method for when a CellTag is missing and want to gate whole range
.dummySubgate <- function(fr, pp_res, channels=NA, filterId="dummySubgate",minAtmin=T, ...){

        x <- exprs(fr)[, channels]
        if (minAtmin) {
                min <- min(x, na.rm=T) # ignore NaN as they can be produced by normalization
                max <- NULL
        } else {
                min <- max(x, na.rm=T)
                max <- NULL
        }
        my_gate <- openCyto:::.boundary(fr, channels=channels, min=min, max=max,...)
        return(my_gate)
}
registerPlugins(fun=.dummySubgate,methodName='dummySubgate',dep=NA)


### automatically make gating template based on CTs in index file
# function for adding rows to gating template data frame to reduce syntax
addGtRow <- function(df, alias=NA, pop=NA, parent=NA, dims=NA, gmethod=NA, gargs=NA) {
        row <- list(alias=alias
                        , pop=pop
                        , parent=parent
                        , dims=dims
                        , gating_method=gmethod
                        , gating_args=gargs)
        df[nrow(df)+1, names(row)] <- row
        return(df)
}

#detect which CellTags are present in primary and secondary quadrants
Q1 <- c("CT1","CT2","CT3","CT4"
        ,"CT6","CT7","CT8","CT9"
        ,"CT11","CT12","CT13","CT16","CT17")
Q1 <- Q1 %in% CellTagName

Q2 <- c("CT20","CT21", "CT22") %in% CellTagName
Q3 <- c("CT19") %in% CellTagName
Q4 <- c("CT5","CT10", "CT15") %in% CellTagName

if(sum(Q1,Q2,Q3,Q4)<1) { stop("No recognized CellTags in index file!")}
```

```
q1 <- c("CT1","CT2","CT3"
        ,"CT6","CT7","CT8"
        ,"CT11","CT12")
q1 <- q1 %in% CellTagName

q2 <- c("CT16","CT17") %in% CellTagName
q3 <- c("CT13") %in% CellTagName
q4 <- c("CT4","CT9") %in% CellTagName


#ask for scatter cutoffs, default by no answer
fscMIN <- readline(prompt="Enter min FSC.A cutoff (default=3e4): ")
fscMAX <- readline(prompt="Enter max FSC.A cutoff (default=2.6e5): ")
sscMIN <- readline(prompt="Enter min SSC.A cutoff (default=1.5e4): ")
sscMAX <- readline(prompt="Enter max SSC.A cutoff (default=2.6e5): ")

if(fscMIN == "") fscMIN <- "3e4"
if(fscMAX == "") fscMAX <- "2.6e5"
if(sscMIN == "") sscMIN <- "1.5e4"
if(sscMAX == "") sscMAX <- "2.6e5"

#add scatter gate
scGateArgs <- paste("min = c("
                        ,fscMIN
                        ,","
                        ,sscMIN
                        ,"), max=c("
                        ,fscMAX
                        ,","
                        ,sscMAX
                        ,")"
                        , sep="")

#initialize gating template data frame
df <- data.frame(alias=character(),
                 pop=character(),
                 parent=character(),
                 dims=character(),
                 gating_method=character(),
                 gating_args=character(),
                 collapseDataForGating=character(),
                 groupBy=character(),
                 preprocessing_method=character(),
                 preprocessing_args=character(),
                 stringsAsFactors=FALSE)


df <- addGtRow(df,"nonMaxS", "nonMaxS+", "root",  "F.A,S.A", "boundary", scGateArgs)

### primary quadrants (Q) ###
#G.n Q ref split
alias <- "Q1Q4"
pop <- "G.n+"
parent <- "nonMaxS"
dims <- "G.n"
if (sum(Q1,Q4)>0 & sum(Q2,Q3)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "specificMindensity", "selectedPeak=1,
findAdjust=2.5")
} else if (sum(Q1,Q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate")
} else if (sum(Q2,Q3)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate", "minAtmin=F")
}


#R.n Q ref split
alias <- "Q1Q2"
pop <- "R.n+"
```

```
parent <- "nonMaxS"
dims <- "R.n"
if (sum(Q1,Q2)>0 & sum(Q3,Q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "specificMindensity", "selectedPeak=1,
findAdjust=2.6, totalPeaks=3")
} else if (sum(Q1,Q2)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate")
} else if (sum(Q3,Q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate", "minAtmin=F")
}

#Q gates
parent <- "nonMaxS"
dims <- "G.n,R.n"
method <- "refGate"
args <- "Q1Q4:Q1Q2"
#Q1 gate
if (sum(Q1)>0) df <- addGtRow(df, "Q1", "G.n+R.n+", parent, dims, method, args)
#Q2 gate
if (sum(Q2)>0) df <- addGtRow(df, "Q2", "G.n-R.n+", parent, dims, method, args)
#Q3 gate
if (sum(Q3)>0) df <- addGtRow(df, "Q3", "G.n-R.n-", parent, dims, method, args)
#Q4 gate
if (sum(Q4)>0) df <- addGtRow(df, "Q4", "G.n+R.n-", parent, dims, method, args)


### secondary quadrants in Q1 (q) ###
#G.n q ref split
alias <- "q1q4"
pop <- "G.n+"
parent <- "Q1"
dims <- "G.n"
if (sum(q1,q4)>0 & sum(q2,q3)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "specificMindensity", "selectedPeak=1, findAdjust=1,
minAdjust=1.5")
} else if (sum(q1,q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate")
} else if (sum(q2,q3)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate", "minAtmin=F")
}


#R.n q ref split
alias <- "q1q2"
pop <- "R.n+"
parent <- "Q1"
dims <- "R.n"
if (sum(q1,q2)>0 & sum(Q3,Q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "specificMindensity", "selectedPeak=1, findAdjust=1,
minAdjust=1.5")
} else if (sum(q1,q2)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate")
} else if (sum(q3,q4)>0) {
        df <- addGtRow(df, alias, pop, parent,  dims, "dummySubgate", "minAtmin=F")
}

#q gates
parent <- "Q1"
dims <- "G.n,R.n"
method <- "refGate"
args <- "q1q4:q1q2"
#q1 gate
if (sum(q1)>0) df <- addGtRow(df, "q1", "G.n+R.n+", parent, dims, method, args)
#q2 gate
if (sum(q2)>0) df <- addGtRow(df, "q2", "G.n-R.n+", parent, dims, method, args)
#q3 gate
if (sum(q3)>0) df <- addGtRow(df, "q3", "G.n-R.n-", parent, dims, method, args)
#q4 gate
if (sum(q4)>0) df <- addGtRow(df, "q4", "G.n+R.n-", parent, dims, method, args)
```

```
### subgates in Q4, Q2 ###
## gates in Q4 ##
low <- "CT15" %in% CellTagName
hi <- isTRUE(sum(c("CT5","CT10") %in% CellTagName) > 0)
if (low & hi) {
        df <- addGtRow(df, "CT15p", "G.n-", "Q4", "G.n", "specificMindensity", "selectedPeak=1")
        df <- addGtRow(df, "CT.5.10", "G.n+", "Q4", "G.n", "specificMindensity", "selectedPeak=1")
} else if (low) {
        df <- addGtRow(df, "CT15p", "G.n+", "Q4", "G.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.5.10", "G.n+", "Q4", "G.n", "dummySubgate")
}

low <- "CT10" %in% CellTagName
hi <- "CT5" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT10p", "G.n-", "CT.5.10", "G.n", "mindensity")
        df <- addGtRow(df, "CT5p", "G.n+", "CT.5.10", "G.n", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT10p", "G.n+", "CT.5.10", "G.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT5p", "G.n+", "CT.5.10", "G.n", "dummySubgate")
}


## gates in Q2 ##
low <- "CT22" %in% CellTagName
hi <- isTRUE(sum(c("CT20","CT21") %in% CellTagName) > 0)
if (low & hi) {
        df <- addGtRow(df, "CT22p", "R.n-", "Q2", "R.n", "specificMindensity", "selectedPeak=1")
        df <- addGtRow(df, "CT.20.21", "R.n+", "Q2", "R.n", "specificMindensity", "selectedPeak=1")
} else if (low) {
        df <- addGtRow(df, "CT22p", "R.n+", "Q2", "R.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.20.21", "R.n+", "Q2", "R.n", "dummySubgate")
}

low <- "CT21" %in% CellTagName
hi <- "CT20" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT21p", "R.n-", "CT.20.21", "R.n", "mindensity")
        df <- addGtRow(df, "CT20p", "R.n+", "CT.20.21", "R.n", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT21p", "R.n+", "CT.20.21", "R.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT20p", "R.n+", "CT.20.21", "R.n", "dummySubgate")
}


### subgates in q4, q2 ###
## gates in q4 ##
low <- "CT9" %in% CellTagName
hi <- "CT4" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT9p", "G.n-", "q4", "G.n", "mindensity")
        df <- addGtRow(df, "CT4p", "G.n+", "q4", "G.n", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT9p", "G.n+", "q4", "G.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT4p", "G.n+", "q4", "G.n", "dummySubgate")
}

## gates in q2 ##
low <- "CT17" %in% CellTagName
hi <- "CT16" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT17p", "R.n-", "q2", "R.n", "mindensity")
```

```
        df <- addGtRow(df, "CT16p", "R.n+", "q2", "R.n", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT17p", "R.n+", "q2", "R.n", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT16p", "R.n+", "q2", "R.n", "dummySubgate")
}


### subgates in q1 ###
## gate q1 low wing ##
low <- "CT3" %in% CellTagName
mid <- isTRUE(sum(c("CT2","CT8") %in% CellTagName) > 0)
hi <- isTRUE(sum(c("CT1","CT6","CT7","CT11","CT12") %in% CellTagName) > 0)
if ((low | mid) & hi) {
        if (low & mid) {
                args <- paste("selectedPeak=", "2", ", totalPeaks=5", sep="")
        } else {
                args <- paste("selectedPeak=", "1", ", totalPeaks=5", sep="")
        }
        df <- addGtRow(df, "CT.2.3.8", "fRatio-", "q1", "fRatio", "specificMindensity", args)
        df <- addGtRow(df, "CT.1.6.7.11.12", "fRatio+", "q1", "fRatio", "specificMindensity", args)
} else if (low | mid) {
        df <- addGtRow(df, "CT.2.3.8", "fRatio+", "q1", "fRatio", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.1.6.7.11.12", "fRatio+", "q1", "fRatio", "dummySubgate")
}


# subgates in q1 low wing #
low <- "CT3" %in% CellTagName
hi <- isTRUE(sum(c("CT2","CT8") %in% CellTagName) > 0)
if (low & hi) {
        df <- addGtRow(df, "CT3p", "fRatio-", "CT.2.3.8", "fRatio", "specificMindensity",
"selectedPeak=1")
        df <- addGtRow(df, "CT.2.8", "fRatio+", "CT.2.3.8", "fRatio", "specificMindensity",
"selectedPeak=1")
} else if (low) {
        df <- addGtRow(df, "CT3p", "fRatio+", "CT.2.3.8", "fRatio", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.2.8", "fRatio+", "CT.2.3.8", "fRatio", "dummySubgate")
}

low <- "CT8" %in% CellTagName
hi <- "CT2" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT8p", "fTotal-", "CT.2.8", "fTotal", "mindensity")
        df <- addGtRow(df, "CT2p", "fTotal+", "CT.2.8", "fTotal", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT8p", "fTotal+", "CT.2.8", "fTotal", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT2p", "fTotal+", "CT.2.8", "fTotal", "dummySubgate")
}

## gate q1 high wing ##
low <- isTRUE(sum(c("CT1","CT7") %in% CellTagName) > 0)
hi <- isTRUE(sum(c("CT6","CT11","CT12") %in% CellTagName) > 0)
if (low & hi) {
        df <- addGtRow(df, "CT.1.7", "fRatio-", "CT.1.6.7.11.12", "fRatio", "specificMindensity",
"selectedPeak=1")
        df <- addGtRow(df, "CT.6.11.12", "fRatio+", "CT.1.6.7.11.12", "fRatio", "specificMindensity",
"selectedPeak=1")
} else if (low) {
        df <- addGtRow(df, "CT.1.7", "fRatio+", "CT.1.6.7.11.12", "fRatio", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.6.11.12", "fRatio+", "CT.1.6.7.11.12", "fRatio", "dummySubgate")
}

# subgates in q1 high wing #
low <- isTRUE(sum(c("CT6","CT12") %in% CellTagName) > 0)
```

```
hi <- "CT11" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT.6.12", "fRatio-", "CT.6.11.12", "fRatio", "specificMindensity",
"selectedPeak=1, selectFromRight=TRUE")
        df <- addGtRow(df, "CT11p", "fRatio+", "CT.6.11.12", "fRatio", "specificMindensity",
"selectedPeak=1, selectFromRight=TRUE")
} else if (low) {
        df <- addGtRow(df, "CT.6.12", "fRatio+", "CT.6.11.12", "fRatio", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT.11p", "fRatio+", "CT.6.11.12", "fRatio", "dummySubgate")
}


low <- "CT12" %in% CellTagName
hi <- "CT6" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT12p", "fTotal-", "CT.6.12", "fTotal", "mindensity")
        df <- addGtRow(df, "CT6p", "fTotal+", "CT.6.12", "fTotal", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT12p", "fTotal+", "CT.6.12", "fTotal", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT6p", "fTotal+", "CT.6.12", "fTotal", "dummySubgate")
}

## subgates in q1 mid wing ##
low <- "CT7" %in% CellTagName
hi <- "CT1" %in% CellTagName
if (low & hi) {
        df <- addGtRow(df, "CT7p", "fTotal-", "CT.1.7", "fTotal", "mindensity")
        df <- addGtRow(df, "CT1p", "fTotal+", "CT.1.7", "fTotal", "mindensity")
} else if (low) {
        df <- addGtRow(df, "CT7p", "fTotal+", "CT.1.7", "fTotal", "dummySubgate")
} else if (hi) {
        df <- addGtRow(df, "CT1p", "fTotal+", "CT.1.7", "fTotal", "dummySubgate")
}



### all final population gates ###
dims <- "G.n,R.n"
method <- "flowClust"
args <- "K=1, quantile=0.8"

if ("CT1" %in% CellTagName) df <- addGtRow(df, "CT1", "CT1", "CT1p", dims, method, args)
if ("CT2" %in% CellTagName) df <- addGtRow(df, "CT2", "CT2", "CT2p", dims, method, args)
if ("CT3" %in% CellTagName) df <- addGtRow(df, "CT3", "CT3", "CT3p", dims, method, args)
if ("CT4" %in% CellTagName) df <- addGtRow(df, "CT4", "CT4", "CT4p", dims, method, args)
if ("CT5" %in% CellTagName) df <- addGtRow(df, "CT5", "CT5", "CT5p", dims, method, args)
if ("CT6" %in% CellTagName) df <- addGtRow(df, "CT6", "CT6", "CT6p", dims, method, args)
if ("CT7" %in% CellTagName) df <- addGtRow(df, "CT7", "CT7", "CT7p", dims, method, args)
if ("CT8" %in% CellTagName) df <- addGtRow(df, "CT8", "CT8", "CT8p", dims, method, args)
if ("CT9" %in% CellTagName) df <- addGtRow(df, "CT9", "CT9", "CT9p", dims, method, args)
if ("CT10" %in% CellTagName) df <- addGtRow(df, "CT10", "CT10", "CT10p", dims, method, args)
if ("CT11" %in% CellTagName) df <- addGtRow(df, "CT11", "CT11", "CT11p", dims, method, args)
if ("CT12" %in% CellTagName) df <- addGtRow(df, "CT12", "CT12", "CT12p", dims, method, args)
if ("CT13" %in% CellTagName) df <- addGtRow(df, "CT13", "CT13", "q3", dims, method, args)
if ("CT15" %in% CellTagName) df <- addGtRow(df, "CT15", "CT15", "CT15p", dims, method, args)
if ("CT16" %in% CellTagName) df <- addGtRow(df, "CT16", "CT16", "CT16p", dims, method, args)
if ("CT17" %in% CellTagName) df <- addGtRow(df, "CT17", "CT17", "CT17p", dims, method, args)
if ("CT19" %in% CellTagName) df <- addGtRow(df, "CT19", "CT19", "Q3", dims, method, args)
if ("CT20" %in% CellTagName) df <- addGtRow(df, "CT20", "CT20", "CT20p", dims, method, args)
if ("CT21" %in% CellTagName) df <- addGtRow(df, "CT21", "CT21", "CT21p", dims, method, args)
if ("CT22" %in% CellTagName) df <- addGtRow(df, "CT22", "CT22", "CT22p", dims, method, args)


## make temp cvs from df
tmp = tempfile(fileext=".csv")
write.csv(df,tmp,row.names=FALSE)
```

```
## apply gating template cvs to fs
gt <- gatingTemplate(tmp, autostart = 1L)
gs <- GatingSet(fs)
gating(gt, gs, parallel_type="multicore",mc.cores=numCores)

# adjust param for plotGate()
flowWorkspace.par.set("plotGate", list(default.y = "S.A")) # set default Y-axis from "SSC-A" to "S.A"

#extract gated CT populations, add keywords corresponding to CT/strain identities
cat("Extracting final gated CellTag populations...")
res <- vector("list", 0)
for (i in 1:length(gs)) {
        fFrames <- lapply(X=1:length(CellTagName), FUN=function(X){

                tmp <- getData(gs[[i]], CellTagName[X])
                keyword(tmp) <- c(keyword(tmp), CellTag=CellTagName[X], Strain=StrainName[X])
                return(tmp)
        })
        res[[i]] <- flowSet(fFrames)
        pData(res[[i]])$Sample <- keyword(res[[i]], "TUBE NAME")[,1]
        pData(res[[i]])$Strain <- keyword(res[[i]], "Strain")[,1]

        # extract tube number from tube name
        n <- strsplit(pData(res[[i]])[1,"Sample"], "[^0-9]+") #split string to get ""  and tube numer as
list
        n <- as.numeric(unlist(n))[2] #unlist and take only second entry, not first ""

        # add gs index to row in tube index dataframe that matches tube number
        Tubeindex[Tubeindex$Tube == n,]$gsIndex <- i
}
cat("Done\n")


#find control
ctrlIndex <- Tubeindex[tolower(Tubeindex$Treatment) == "control",]
ctrlIndex <- ctrlIndex$gsIndex
if (length(ctrlIndex)<1) {
        cat("The detected treatments are:", Tubeindex[,"Treatment"],sep="\n")
        ctrlIndex <- readline(prompt="Enter name of control treatment: ")
        ctrlIndex <- Tubeindex[Tubeindex$Treatment == ctrlIndex,]
        ctrlIndex <- ctrlIndex$gsIndex
}

#plot blue fluorescence results write file, use sample 1 to set min/max
cat("Creating summary png files\n")
min <- parameters(res[[1]][[1]])[["minRange"]][B.nIndex]
max <- parameters(res[[1]][[1]])[["maxRange"]][B.nIndex]

for (i in 1:length(gs)) {
        p <- ggcyto(res[[ctrlIndex]], aes(x = B.n))
        p <- p + geom_density(alpha = 0.2, fill="black")
        p <- p + geom_density(data=res[[i]], alpha = 0.2, fill="red")
        p <- p + scale_x_continuous(limits = c(min, max))
        p <- p + facet_wrap(~Strain)

        name <- Tubeindex[Tubeindex$gsIndex == i,]
        name <- paste("Tube", name$Tube, "_", name$Treatment, sep="")
        p <- p + labs(title=name)
        p <- p + theme(strip.text.x = element_text(size = 16))

        outName <- paste(name, ".png", sep="")
        png(filename=file.path(inFolder, "output", outName)
                , width = 1024, height = 768)
        plot(p)
        dev.off()
        cat(paste(outName, "\n", sep=""))
}

cat("Finished!\n")
```

7.3.2 Automated retrieval of taxonomic fata from UniProt using python

```
# Adapted from https://github.com/Ensembl/ensembl-rest/wiki/Example-Python-Client
import sys
import urllib
import urllib2
import json
import time
import requests
import socket
import csv
import code
import untangle #sudo easy_install pip #sudo pip install untangle

class EnsemblRestClient(object):
    def __init__(self, server='http://www.uniprot.org/uniprot/', reqs_per_sec=15):
        self.server = server
        self.reqs_per_sec = reqs_per_sec
        self.req_count = 0
        self.last_req = 0

    def perform_rest_action(self, id, targetProp, endpoint, hdrs=None):
        if hdrs is None:
            hdrs = {}

        if 'Content-Type' not in hdrs:
            hdrs['Content-Type'] = 'application/json'

        data = None

        # check if we need to rate limit ourselves
        if self.req_count >= self.reqs_per_sec:
            delta = time.time() - self.last_req
            if delta < 1:
                time.sleep(1 - delta)
            self.last_req = time.time()
            self.req_count = 0

        try:
            request = urllib2.Request(self.server + endpoint, headers=hdrs)
            response = urllib2.urlopen(request, timeout=4)
            content = response.read()
            if content:
                data = untangle.parse(content)
            self.req_count += 1

        except urllib2.HTTPError, e:
            # check if we are being rate limited by the server
            if e.code == 429:
                if 'Retry-After' in e.headers:
                    retry = e.headers['Retry-After']
                    time.sleep(float(retry))
                    self.perform_rest_action(endpoint, hdrs, params)
            else:
                sys.stderr.write('Request failed for {0}: Status code: {1.code} Reason:
{1.reason}\n'.format(endpoint, e))
                return {'id': id, targetProp: e}

        except socket.timeout, e:
            # For Python 2.7
            sys.stderr.write('Request failed for {0}: Reason: {1}\n'.format(id, e))
            return {'id': id, targetProp: e}

        return data

    def get_data(self, id, targetProp):
        seq = self.perform_rest_action(
```

```
            id=id,
            targetProp=targetProp,
            endpoint='{0}.xml'.format(id)
            #params={'db': 'taxonomy', 'id': id }

        )
        if seq:
            return seq
        return None


def run(idsFile):
    targetProp = 'info' #1 'seq'
    geneIDs = []
    with open(idsFile) as f:
        csv_f = csv.reader(f)
        for row in csv_f:
            geneIDs.append(row[0])

    with open('out.csv', 'w') as outfile:
        writer = csv.writer(outfile)
        writer.writerow(['ID', 'txid', 'scientific name', 'common name', 'synonym', 'taxonomy'])

        for ID in geneIDs:
            client = EnsemblRestClient()
            data = client.get_data(ID, targetProp)
            #code.interact(local=locals())
            if data:
                if data.uniprot.entry.organism.name['type'=='scientific']:
                    sciName = data.uniprot.entry.organism.name['type'=='scientific'].cdata
                else:
                    sciName = data.uniprot.entry.organism.name.cdata
                if data.uniprot.entry.organism.name['type'=='common']:
                    commonName = data.uniprot.entry.organism.name['type'=='common'].cdata
                else:
                    commonName = 'none'
                if data.uniprot.entry.organism.name['type'=='synonym']:
                    synName = data.uniprot.entry.organism.name['type'=='synonym'].cdata
                else:
                    synName = 'none'
                txid = data.uniprot.entry.organism.dbReference['id'] #get target id
                l = len(data.uniprot.entry.organism.lineage.taxon)
                i = 1
                tax = []
                while i <= l:
                    if data.uniprot.entry.organism.lineage.taxon[l-i]:
                        tax.append(data.uniprot.entry.organism.lineage.taxon[l-i].cdata)
                    else:
                        tax.append('no tax{0} found'.format(i))
                    i += 1
                writer.writerow([ ID, txid, sciName, commonName, synName] + tax)
                print targetProp + ' retrieved for {1}: {0}'.format(sciName, ID)

#efetch.fcgi?db=taxonomy&id=1182543


if __name__ == '__main__':
    if len(sys.argv) == 2:
        idsFile = sys.argv[1]
    else:
        idsFile = 'in.csv'

    run(idsFile)
```

7.4 Additional Figures

7.4.1 For Chapter 4

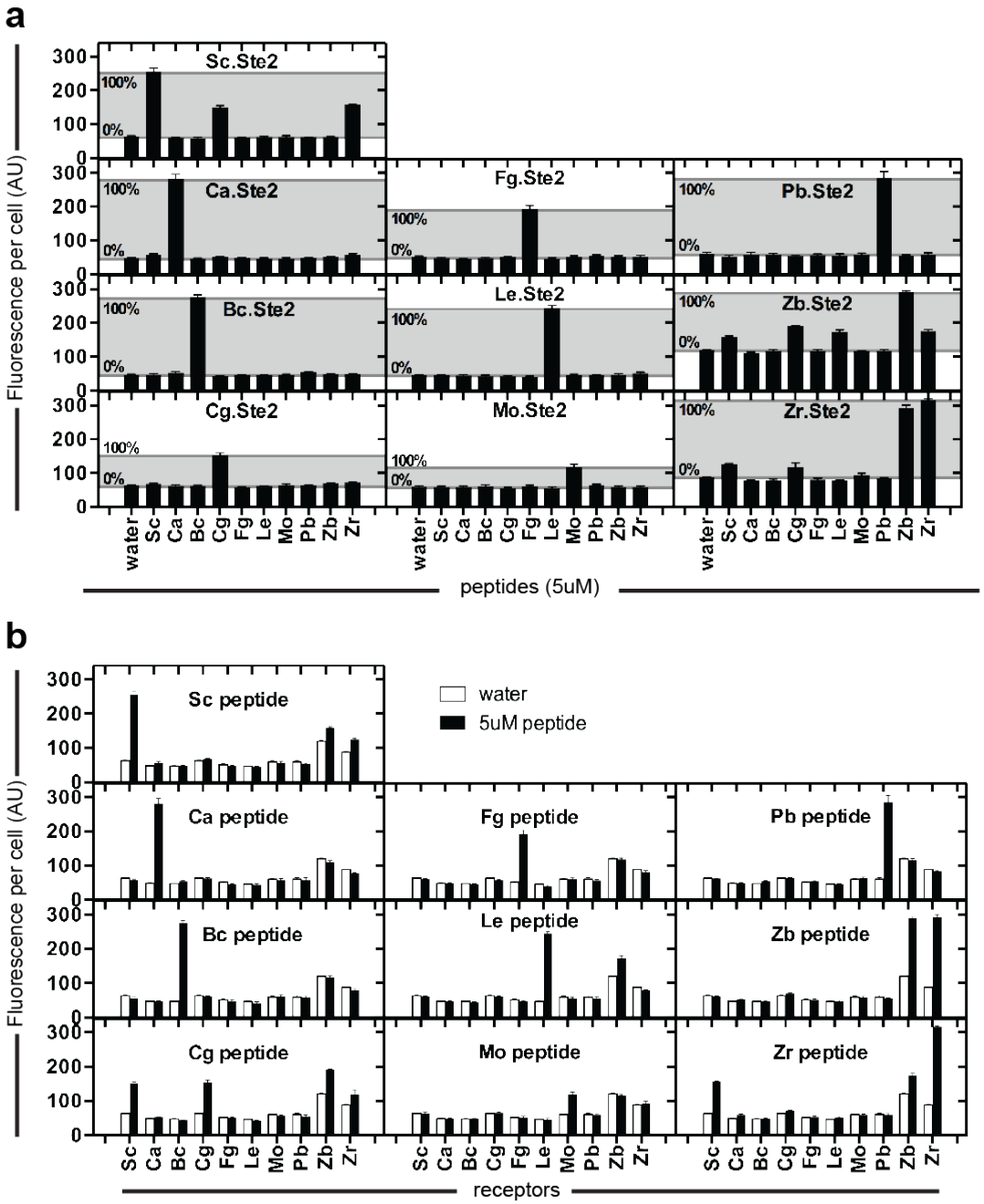


Figure 7.1 Specificity of fungal mating receptors

(a) Heterologous receptors (‘*species*.Ste2’) were induced with 5 μM of the indicated fungal mating peptide. mCherry fluorescence was measured after 9 hours. Basal (0%) and maximal (100%) fluorescence used to generate orthogonality figure in Chapter 4. **(b)** Data as in a. Activation of heterologous mating receptors shown here grouped by mating peptide.

7.4.2 For Chapter 5

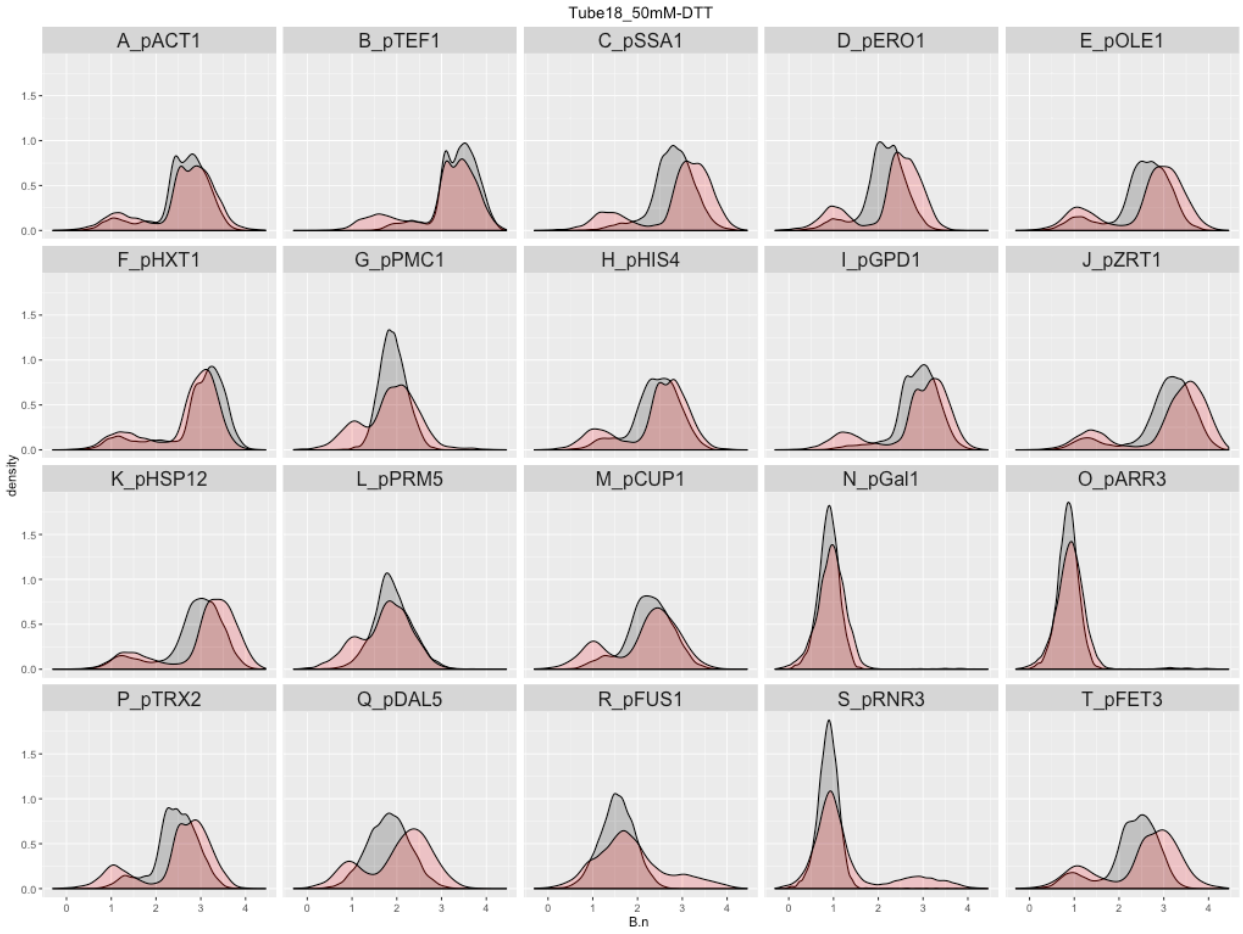


Figure 7.2 Reporter responses to DTT

Reporter cocktail was exposed to either standard media (grey histograms) or media with 50mM dithiothreitol (DTT, red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

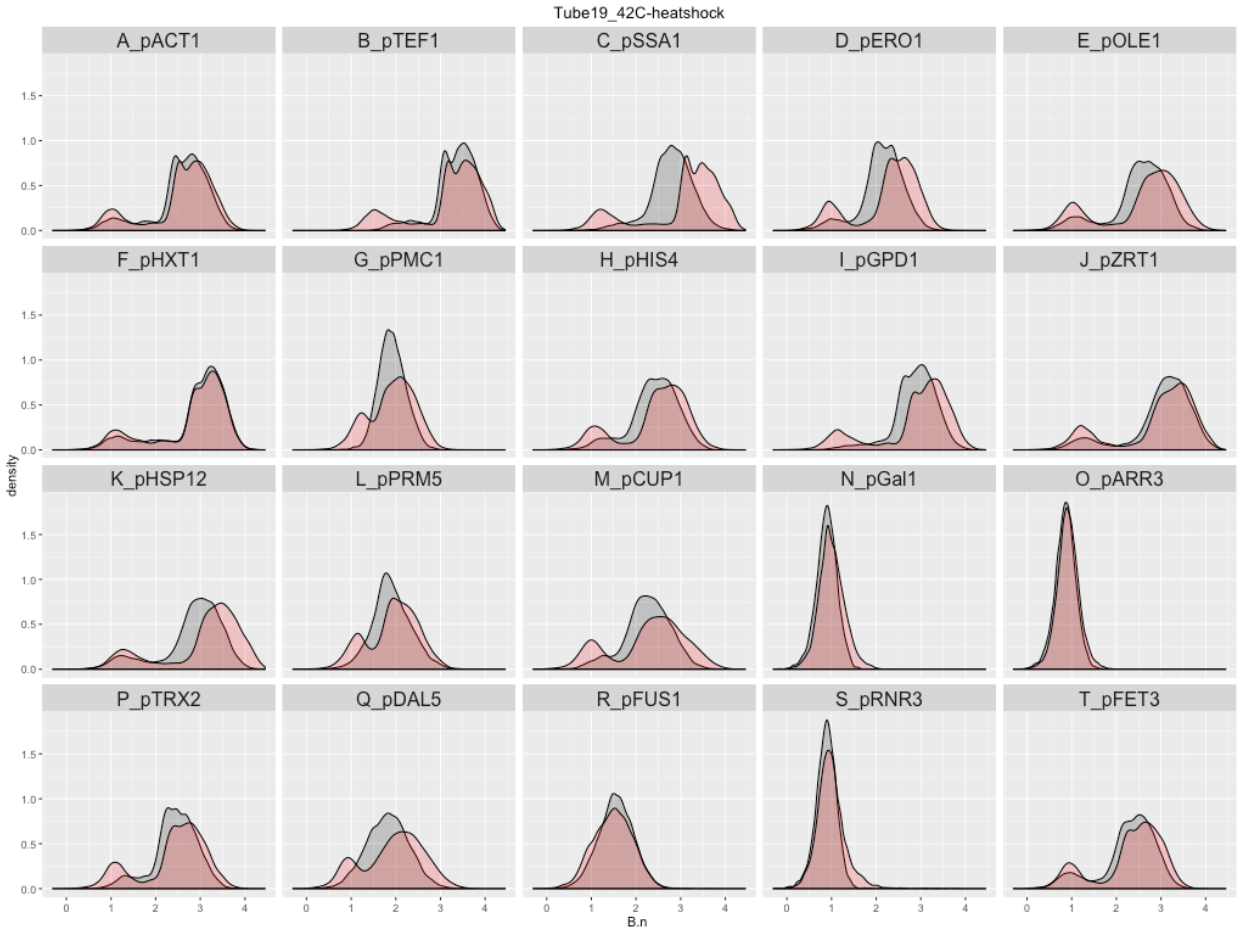


Figure 7.3 Reporter responses to heatshock

Reporter cocktail was exposed to either standard temperature 23C (grey histograms) or 42C (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

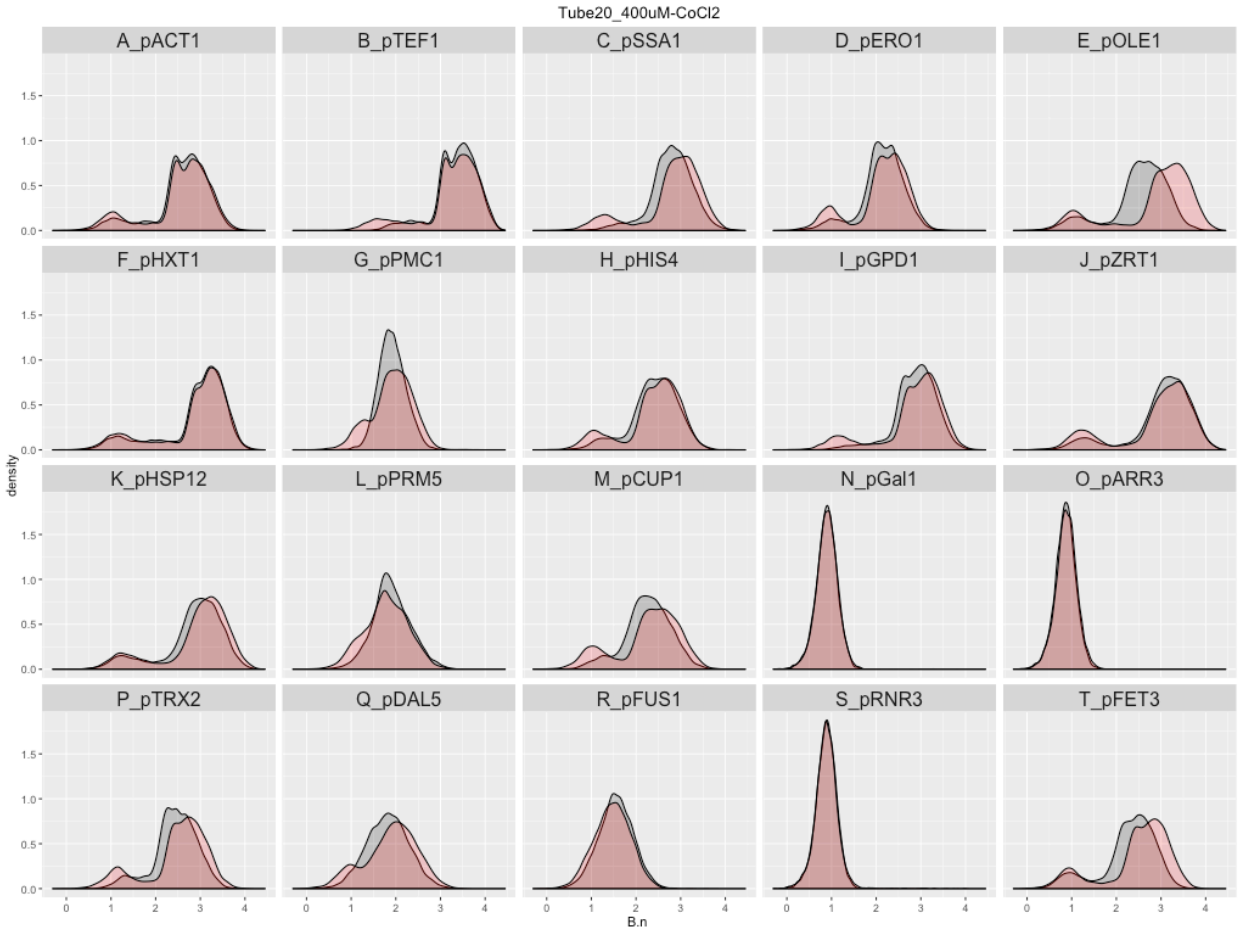


Figure 7.4 Reporter responses to cobalt

Reporter cocktail was exposed to either standard media (grey histograms) or media with 400uM $CoCl_2$ (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

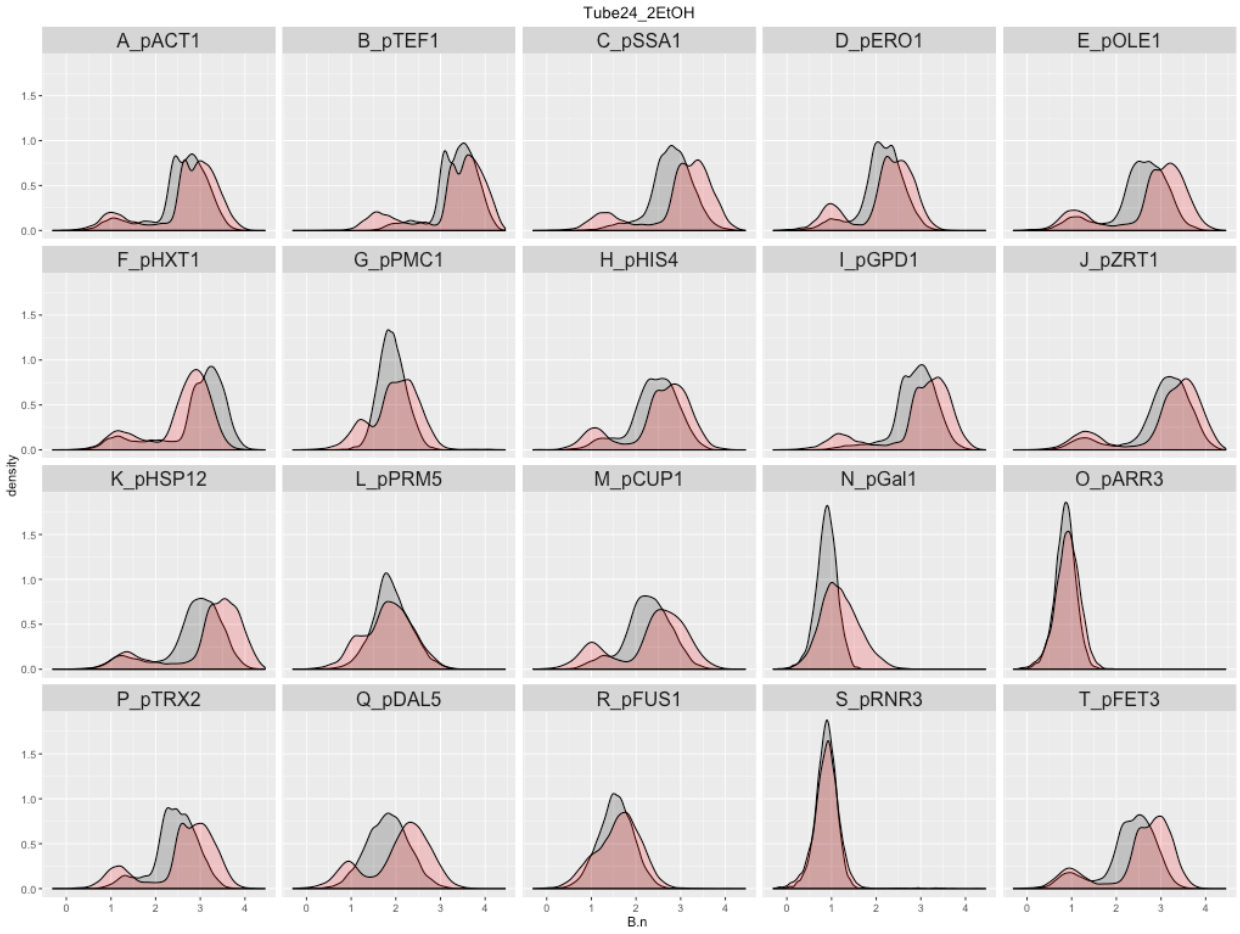


Figure 7.5 Reporter responses to ethanol

Reporter cocktail was exposed to either standard media with 2% glucose (grey histograms) or media with 2% ethanol as a sole carbon source (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

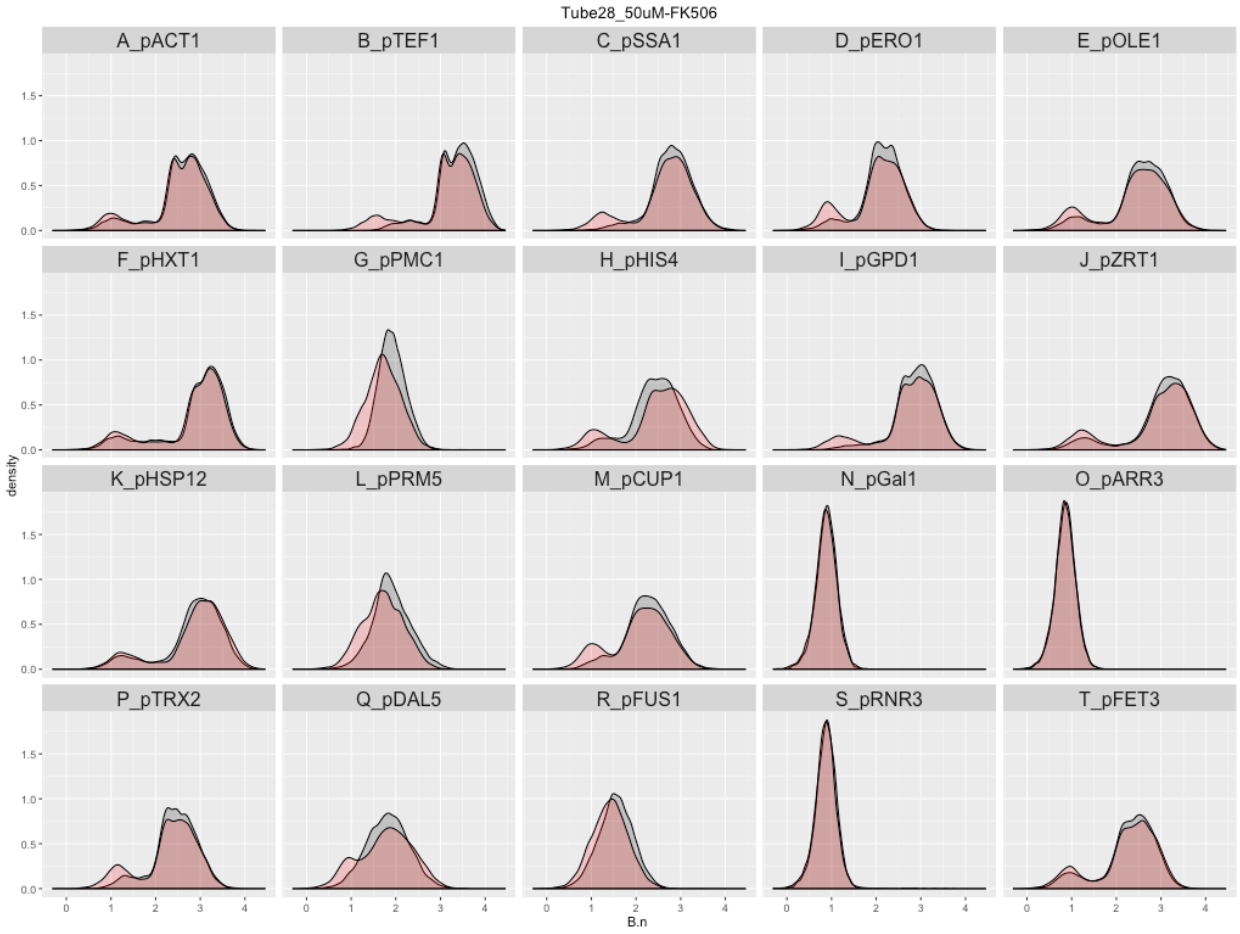


Figure 7.6 Reporter responses to FK506

Reporter cocktail was exposed to either standard media (grey histograms) or media with 50uM FK506 (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

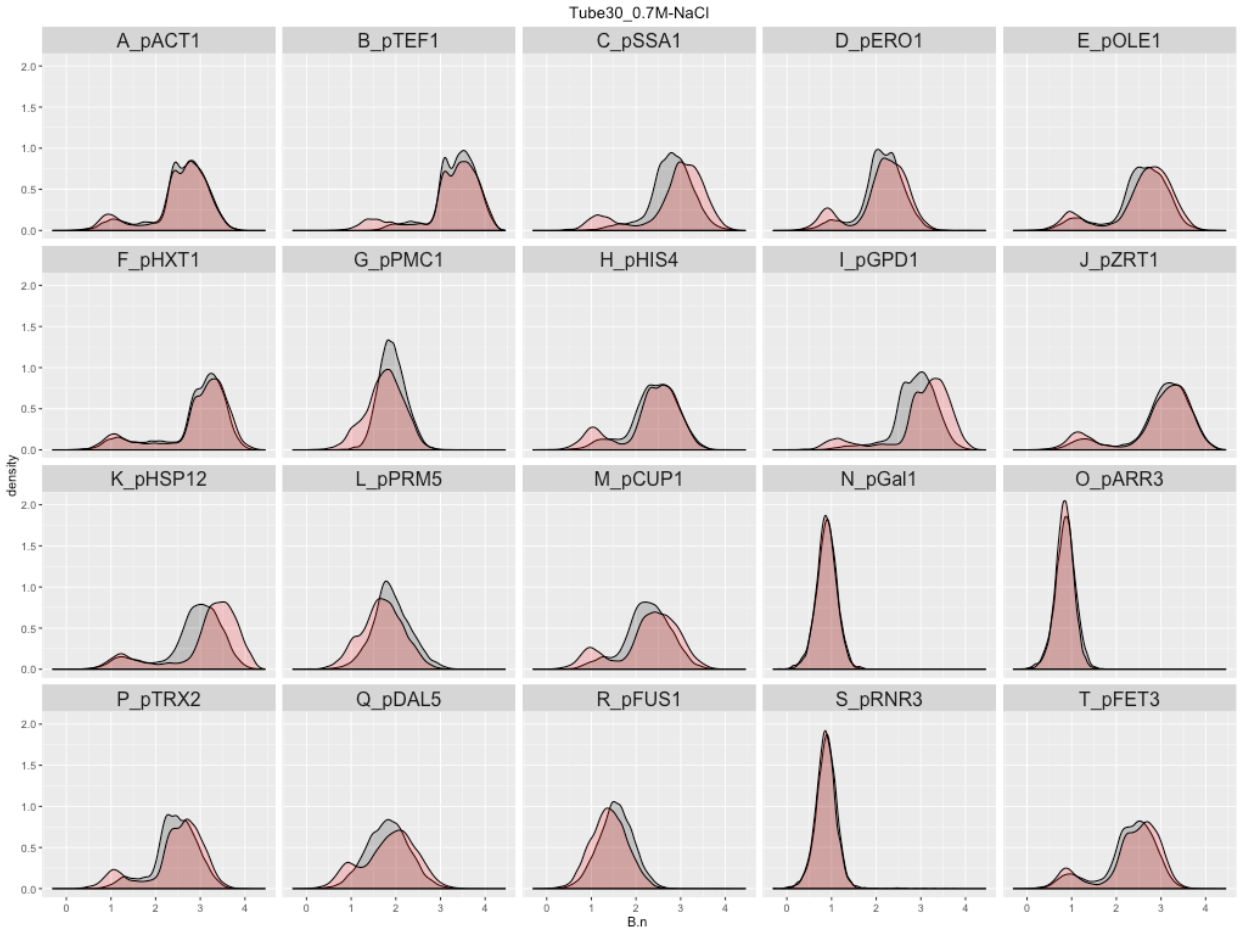


Figure 7.7 Reporter responses to osmotic shock

Reporter cocktail was exposed to either standard media (grey histograms) or media with 0.7M NaCl (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

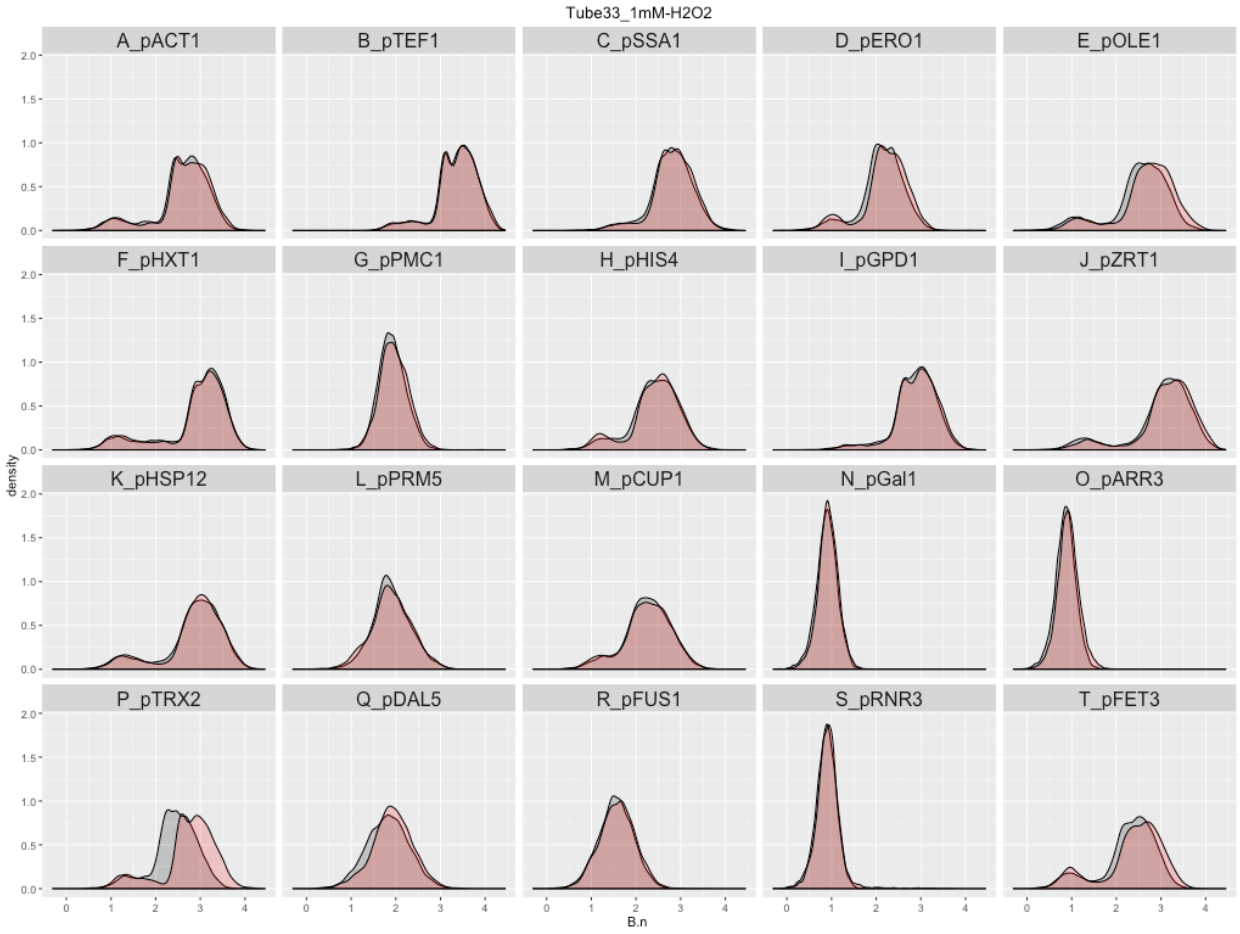


Figure 7.8 Reporter responses to hydrogen peroxide

Reporter cocktail was exposed to either standard media (grey histograms) or media with 1mM hydrogen peroxide (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

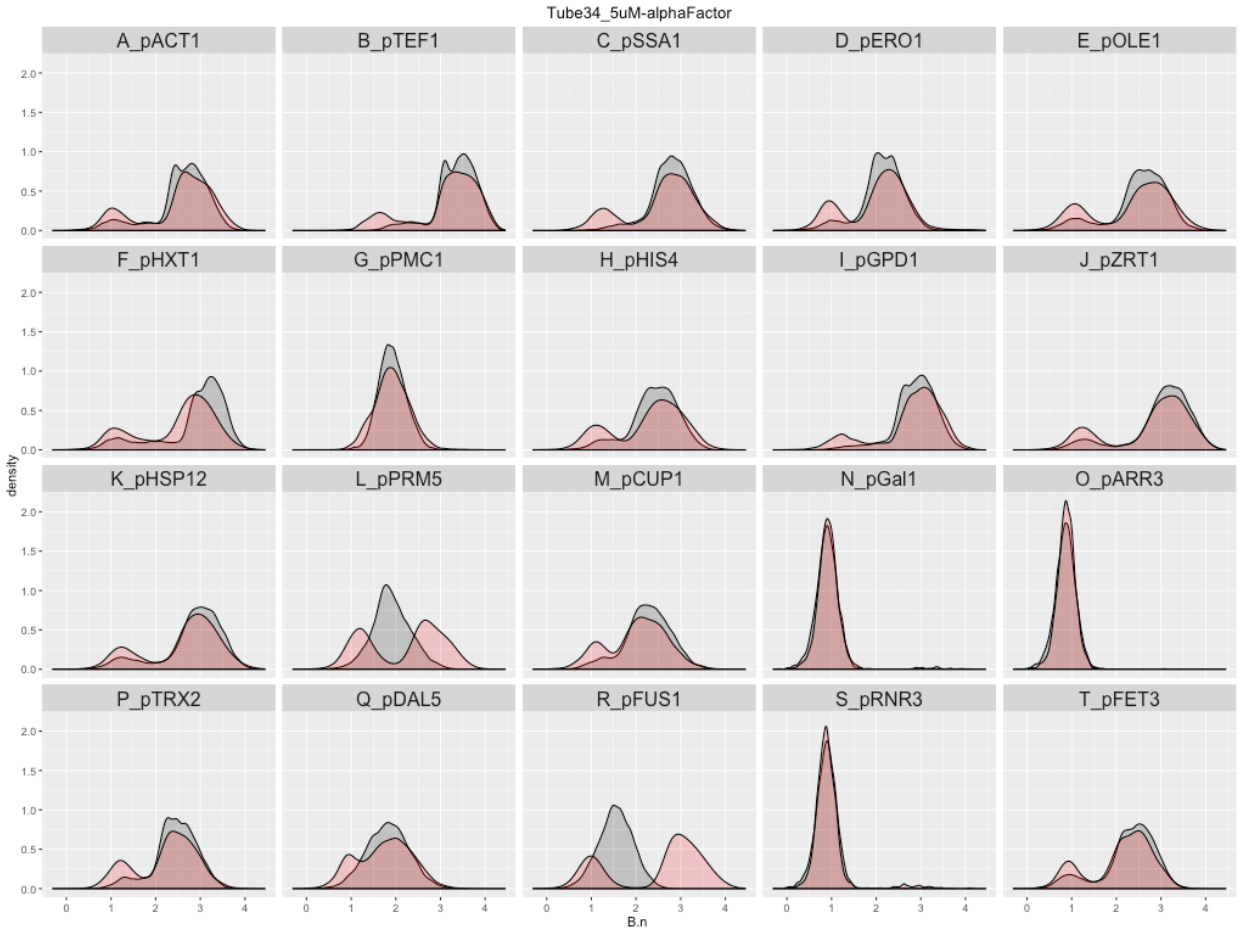


Figure 7.9 Reporter responses to mating pheromone

Reporter cocktail was exposed to either standard media (grey histograms) or media with 5uM alpha factor mating pheromone (red histograms) for 6 hours. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

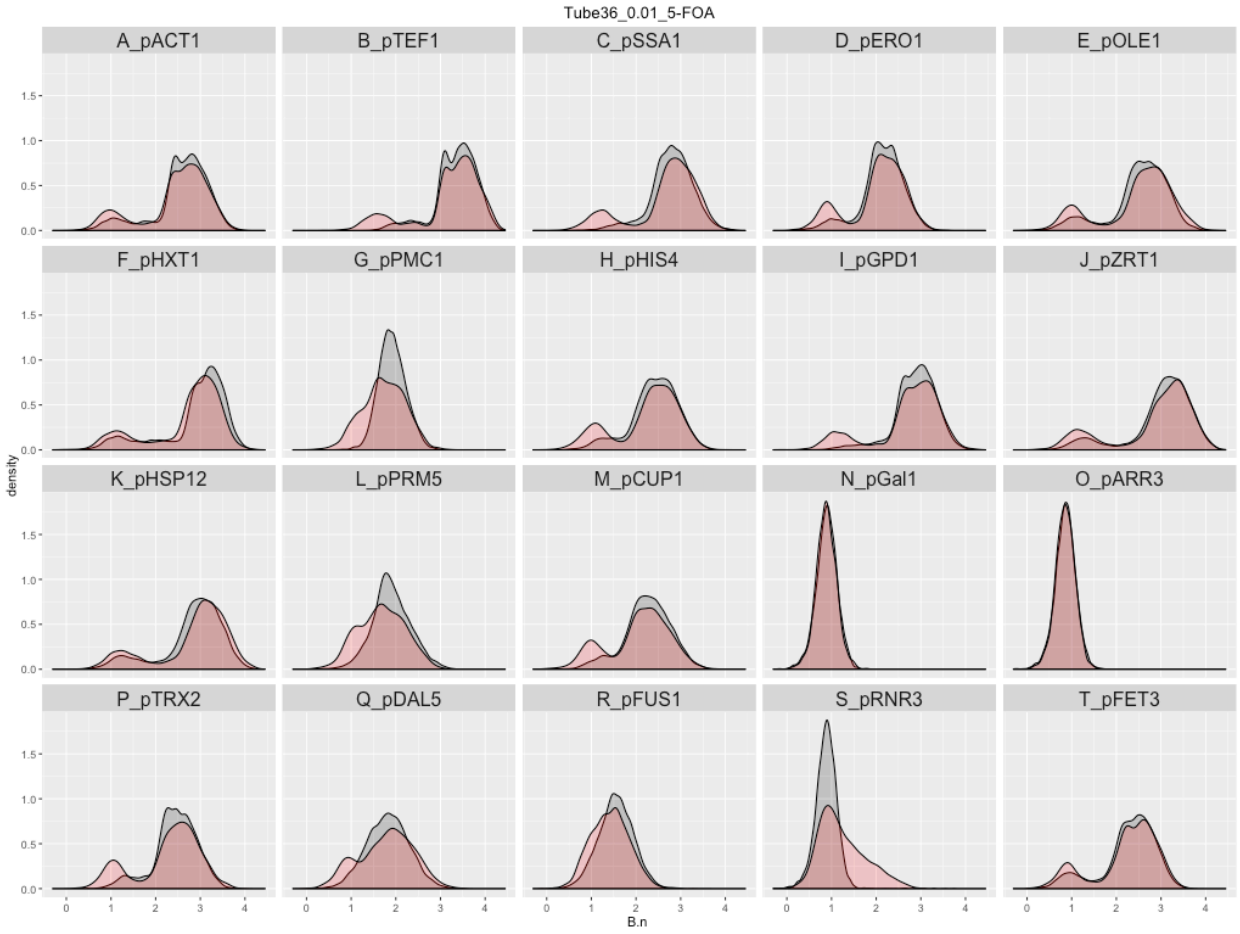


Figure 7.10 Reporter responses to 5-FOA

Reporter cocktail was exposed to either standard media (grey histograms) or media with 500uM 5-FOA (red histograms) for 6 hours. All the strains express the Ura3 gene product that converts 5-FOA to the cytotoxic compound 5-FU. The normalized fluorescece from the mTagBFP2 reporter (B.n) was deconvoluted using the CellTag signals.

7.5 DNA Sequences and strains

7.5.1 Strains.

All strains are *S. cerevisiae* unless otherwise noted. Strains were generated in this study except where a source is noted. The nomenclature "ReRec[N]::" refers to expression modules inserted in the N^{th} round of reiterative recombination at the acceptor site located in the HO locus (*31*).

Strain	Genotype	Comments
FY251	*MAT*a *his3-Δ200, leu2-Δ1 trp1-Δ63*, *ura3-52*	ATCC 96098
BY4733	*MATa his3Δ200 leu2Δ0 met15Δ0 trp1Δ63 ura3Δ0*	ATCC 200895
LW2591	BY4733 *MATa-inc HOΔ::ReRec*	Reiterative Recombination acceptor strain (*31*)
LW2671	BY4733 derivative overexpressing *CrtEBI*	Constitutive lycopene producing strain (*38*)
yMJ105	LW2591 *sst2-Δ far1-Δ*	Parental biosensor strain
Fluorescence Readout Strains		
yMJ183	yMJ105 *ste2-Δ fus1Δ::pFUS1-HIS3-tHIS3* ReRec[1]::*pFUS1-yCherry-tACT1*	Receptor-less fluorescence reporter strain
yMJ281	yMJ183 + pMJ093	*S. cerevisiae* Ste2
yMJ282	yMJ183 + pMJ090	*C. albicans* Ste2
yMJ284	yMJ183 + pMJ095	*B. cinerea* Ste2
yMJ285	yMJ183 + pMJ096	*C. glabrata* Ste2
yMJ286	yMJ183 + pMJ097	*F. graminearum* Ste2
yMJ288	yMJ183 + pMJ099	*L. elongisporous* Ste2
yMJ289	yMJ183 + pMJ100	*M. oryzea* Ste2
yMJ290	yMJ183 + pMJ101	*P. brasiliensis* Ste2
yMJ294	yMJ183 + pMJ105	*Z. bailii* Ste2
yMJ295	yMJ183 + pMJ106	*Z. rouxii* Ste2
yMJ312	yMJ183 + pMJ117	*H. capsulatum* Ste2
yJM06	yMJ183 + pJM13	Codon-optimized *C. glabrata* Ste2
Lycopene Sensor Strains		
yMJ116	yMJ105 ReRec[1]::*pTEF1-CrtE-tADH1*-(*CrtB-pPGK1*,rev)	Lycopene null strain
yMJ118	yMJ105 ReRec[1]::*pTEF1-CrtE-tADH1*-(*CrtB-pPGK1*,rev) ReRec[2]::*pFUS1-CrtI-tACT1*	Unoptimized lycopene biosensor Lyco-1
yMJ151	yMJ118 + pMJ006	"+ 2X CrtI" intermediate
yMJ152	yMJ118 + pMJ009	"+ tHMG1" intermediate
yMJ165	yMJ118 + pMJ012	"+ FAD1" intermediate
yMJ251	yMJ105 *met15Δ::pFUS1-CrtI-tACT1-MET15* ReRec[1]::*pTEF1-CrtE-tADH1*-(*CrtB-pPGK1*,rev) ReRec[2]::*pFUS1-CrtI-tACT1* ReRec[3]::*pTDH3-FAD1-tPGK1*	Optimized lycopene biosensor Lyco-2 (Sc biosensor)
yMJ258	yMJ251 ste2Δ::pTDH3-Pb.Ste2-tSTE2	Pb biosensor
yMJ260	yMJ251 ste2Δ::pTDH3-Ca.Ste2-tSTE2	Ca biosensor

Strains Used to Generate Pathogen and Control Supernatants		
W303-1B	*MAT*α *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*	ATCC 201238
FY250	*MAT*α *his3-Δ200, leu2-Δ1 trp1-Δ63*, *ura3-52*	(*48*)
GC75	*Candida albicans,* MTLα/MTLα	Genebank assembly number GCA_000773735.1 (*44*)
ySB36	*Candida albicans*, MTLa/MTLα	Clinical isolate obtained from A-C. Uhlemann, mating loci (MTL) were genotyped by PCR
ySB45	*Candida albicans,* MTLα/MTLα	sorbose selected isolate, derivative of isolate ySB36, MTL were genotyped by PCR
Pb01	*Paracoccidioides lutzii*, MAT1-1	Supernatant prepared by Prof. Fernando Rodrigues (*42*)
Pb18	*Paracoccidioides brasiliensis*, MAT1-2	Supernatant prepared by Prof. Fernando Rodrigues (*42*)
Hc01	*Histoplasma capsulatum*, NAm2	Supernatant prepared by Prof. Chad Rappleye (*40*)
Hc06	*Histoplasma capsulatum*, NAm1	Supernatant prepared by Prof. Chad Rappleye (*40*)

7.5.2 Plasmids.

Plasmids were generated in this study except where a source is noted.

Plasmid	Construct Details	Comments
pSC203	*Erwinia herbicola* CrtEBI	Kind gift from Gregory Stephanopoulos
yEpGAP-Cherry	Yeast codon-optimized mCherry	Kind gift from Neta Dean (*37*)
pLPreB	*P. brasiliensis* mating receptor	Kind gift from Fernando Rodrigues (*30*)
pMJ006	pRS416, pFUS1-CrtI-tACT1	Pheromone inducible CrtI
pMJ009	pRS416, pTDH3-tHMG1-tCYC1	Overexpressed truncated HMG1
pMJ012	pRS416, pTDH3-FAD1-tCYC1	Overexpressed FAD1
pMJ090	pRS416, pTDH3-Ca.Ste2-tSTE2	Overexpressed *C. albicans* Ste2 homologue
pMJ093	pRS416, pTDH3-Sc.Ste2-tSTE2	Overexpressed wild type Ste2
pMJ095	pRS416, pTDH3-Bc.Ste2-tSTE2	Overexpressed *B. cinerea* Ste2 homologue
pMJ096	pRS416, pTDH3-Cg.Ste2-tSTE2	Overexpressed *C. glabrata* Ste2 homologue
pMJ097	pRS416, pTDH3-Fg.Ste2-tSTE2	Overexpressed *F. graminearum* Ste2 homologue
pMJ099	pRS416, pTDH3-Le.Ste2-tSTE2	Overexpressed *L. elongisporus* Ste2 homologue
pMJ100	pRS416, pTDH3-Mo.Ste2-tSTE2	Overexpressed *M. oryzea* Ste2 homologue
pMJ101	pRS416, pTDH3-Pb.Ste2-tSTE2	Overexpressed *P. brasiliensis* Ste2 homologue
pMJ105	pRS416, pTDH3-Zb.Ste2-tSTE2	Overexpressed *Z. bailii* Ste2 homologue
pMJ106	pRS416, pTDH3-Zr.Ste2-tSTE2	Overexpressed *Z. rouxii* Ste2 homologue
pMJ117	pRS416, pTDH3-Hc.Ste2-tSTE2	Overexpressed *H. capsulatum* Ste2 homologue
pJM13	pRS416, pTDH3-Cg.Ste2opt-tSTE2	Overexpressed codon-optimized *C. glabrata* Ste2 homologue

7.5.3 DNA sequences of expression modules constructed

Promoters and terminators in upper case, open reading frames (ORFs) in lower case.

Description	Sequences
Receptor Expression Module	
pTDH3-*species*.Ste2-tSTE2	AGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTTA TTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACACA GAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCGC TTTTTAAGCTGGCATCCAGAAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAGTT CATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAAACGGGCACAACCTC AATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTATCTATCTCA TTTTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGT TCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAAT CTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACCAAGAACTTAGT TTCGACGGATACTAGTAAA***[receptor ORF]***CTCGAGACGGCTTTGAAAAAGTAATTTCGTGACCTTCGGTATA AGGTTACTACTAGATTCAGGTGCTCATCAGATGCACCACATTCTCTATAAAAAAAAATGGTATCTTTCTTAT TTGATAATATTTAAACTCCTTTACATAATAAACATCTCGTAAGTAGTGGTAGAAACCACCTTTGCTTTTACG AGTTCAAGCTTTTTTCTTGCCATGATCTAGAACTCTCAGGCAATATATACAGTTAATCTTTTTTTACTGGGTT GTAGTTCTAATGTATTGTTTCGAAAAATAGCAACCAGGCACA
Fluorescent Reporter Module (yMJ183,yMJ194)	
pFUS1-yCherry-tACT1	TACCATGTGGACCCTTTCAAAACAGAGTTGTATCTCTGCAGGATGCCCTTTTTGACGTATTGAATGGCATAA TTGCACTGTCACTTTTCGCGCTGTCTCATTTTGGTGCGATGATGAAACAAACATGAAACGTCTGTAATTTGA AACAAATAACGTAATTCTCGGGATTGGTTTTATTTAAATGACAATGTAAGAGTGGCTTTGTAAGGTATGTGT TGCTCTTAAAATATTTGGATACGACATCCTTTATCTTTTTTCCTTTAAGAGCAGGATATAAGCCATCAAGTTT CTGAAAATCAAAatggtttcaaaaggtgaagaagataatatggctattattaaagaatttatgagatttaaagttcatatggaaggttcagttaatggtcatgaatttgaaatt gaaggtgaaggtgaaggtagaccatatgaaggtactcaaactgctaaattgaaagttactaaaggtggtccattaccatttgcttgggatattttgtcaccacaatttatgtatggttca aaagcttatgttaaacatccagctgatattccagattatttaaaattgtcatttccagaaggttttaaatgggaaagagttatgaattttgaagatggtggtgttgttactgttactcaagatt catcattacaagatggtgaatttatttataaagttaaattgagaggtactaattttccatcagatggtccagttatgcaaaaaaaaactatgggttgggaagcttcatcagaaagaatgt atccagaagatggtgctttaaaaggtgaaattaaacaaagattgaaattaaaagatggtggtcattatgatgctgaagttaaaactacttataaagctaaaaaaccagttcaattacca ggtgcttatatgttaatattaaattggatattacttcacataatgaagattatactattgttgaacaatatgaaagagctgaaggtagacattcaactggtggtatggatgaattatataa ataaTCTCTGCTTTTGTGCGCGTATGTTTATGTATGTACCTCTCTCTCTATTTCTATTTTTAAACCACCCTCTCA ATAAAATAAAAATAATAAAGTATTTTTAAGGAAAAGACGTGTTTAAGCACTGACTTTATCTACTTTTTGTAC GTTTTCATTGATATAATGTGTTTTGTCTCTCCCTTTTCTACGAAAATTTCAAAAATTGACCAAAAAAAGGAA TATATATACGAAAAACTATTATATTTATATATCATAGTGT
Positive Selection Reporter Module (yMJ183,yMJ194)	
pFUS1-HIS3-tHIS3 @FUS1	TACCATGTGGACCCTTTCAAAACAGAGTTGTATCTCTGCAGGATGCCCTTTTTGACGTATTGAATGGCATAA TTGCACTGTCACTTTTCGCGCTGTCTCATTTTGGTGCGATGATGAAACAAACATGAAACGTCTGTAATTTGA AACAAATAACGTAATTCTCGGGATTGGTTTTATTTAAATGACAATGTAAGAGTGGCTTTGTAAGGTATGTGT TGCTCTTAAAATATTTGGATACGACATCCTTTATCTTTTTTCCTTTAAGAGCAGGATATAAGCCATCAAGTTT CTGAAAATCAAAatgacagagcagaaagccctagtaaagcgtattacaaatgaaaccaagattcagattgcgatctctttaaagggtggtcccctagcgatagagcac tcgatcttcccagaaaaagaggcagaagcagtagcagaacaggccacacaatcgcaagtgattaacgtccacacaggtatagggtttctggaccatatgatacatgctctggcc aagcattccggctggtcgctaatcgttgagtgcattggtgacttacacatagacgaccatcacaccactgaagactgcgggattgctctcggtcaagcttttaaagaggccctactg gcgcgtggagtaaaaaggtttggatcaggatttgcgcctttggatgaggcactttccagagcggtggtagatctttcgaacaggccgtacgcagttgtcgaacttggtttgcaaag ggagaaagtaggagatctctcttgcgagatgatcccgcattttcttgaaagctttgcagaggctagcagaattaccctccacgttgattgtctgcgaggcaagaatgatcatcaccg tagtgagagtgcgttcaaggctcttgcggttgccataagagaagccacctcgcccaatggtaccaacgatgttccctccaccaaaggtgttcttatgtagTGACACCGAT TATTTAAAGCTGCAGCATACGATATATATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTA TACGAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCG CTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTCGA*CGGATCTGAAAATAATATTGACGTTCGCA* *TTTAATCTATACCTATAATTCTGTACTTATATACTGTT*
Negative Selection Reporter Module (yMJ194)	
pFUS2-URA3-tFUS2 @FUS2	TTTTTCTTTGTGAAACCAATTTTAGGTTTTCTTGTTATAGTAAGTTCTTAAGAAAAAGACAAGAAAACCCCT TGCGatgtcgaaagctacatataaggaacgtgctgctactcatcctagtcctgttgctgccaagctatttaatatcatgcacgaaaagcaaacaaacttgtgtgcttcattggatgt tcgtaccaccaaggaattactggagttagttgaagcattaggtcccaaaatttgtttactaaaaacacatgtggatatcttgactgatttttccatggagggcacagttaagccgctaaa ggcattatccgccaagtacaatttttactcttcgaagacagaaaatttgctgacattggtaatacagtcaaattgcagtactctgcgggtgtatacagaatagcagaatgggcagac attacgaatgcacacggtgtggtgggcccaggtattgttagcggtttgaagcaggcggcagaagaagtaacaaaggaacctagaggccttttgatgttagcagaattgtcatgca agggctccctatctactggagaatatactaagggtactgttgacattgcgaagagcgacaaagattttgttatcggctttattgctcaaagagacatgggtggaagagatgaaggtt acgattggttgattatgacacccggtgtgggtttagatgacaagggagacgcattgggtcaacagtatagaaccgtggatgatgtggtctctacaggatctgacattattattgttgg aagaggactatttgcaaagggaagggatgctaaggtagagggtgaacgttacagaaaagcaggctgggaagcatatttgagaagatgcggccagcaaaactaaCAAAA GAGTATATTTAGCTTATAGTTTTTAGAATGTTTTGTTTTCTTTTTTATTAAAGTAGTACTACTGCGCTG
Lycopene Readout Modules (yMJ116, yMJ118, yMJ251)	
pTEF1-CrtE-tADH1-	ATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACT TCAAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACT AAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTT

(CrtB-pPGK1, rev)	ATCACGTTTCTTTTTCTTGAAAATTTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGT TAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGC TCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTACAAAatggtttctggttcgaaagcaggagtatcacctcataggg aaatcgaagtcatgagacagtccattgatgaccacttagcaggattgttgccagaaacagattcccaggatatcgttagccttgctatgagagaaggtgttatggcacctggtaaac gtatcagacctttgctgatgttacttgctgcaagagacctgagatatcagggttctatgcctacactactggatctagcttgtgctgttgaactgacacatactgcttccttgatgctgga tgacatgccttgtatggacaatgcggaacttagaagaggtcaaccaacaacccacaagaaattcggagaatctgttgccatttggcttctgtaggtctgttgtcgaaagcttttggct tgattgctgcaactggtgatcttccaggtgaaaggagagcacaagctgtaaacgagctatctactgcagttggtgttcaaggtctagtcttaggacagttcagagatttgaatgacg cagctttggacagaactcctgatgctatcctgtctacgaaccatctgaagactggcatcttgttctcagctatgttgcaaatcgtagccattgcttctgcttcttcaccatctactaggga aacgttacacgcattcgcattggacttggtcaagcctttcaactgctagacgatttgagggatgatcatccagagacaggtaaagaccgtaacaaagacgctggtaaaagcactc tagtcaacagattgggtgctgatgcagctagacagaaactgagagagcacattgactctgctgacaaacacctgacatttgcatgtccacaaggaggtgctataaggcagtttatg cacctatggtttggacaccatcttgctgattggtctccagtgatgaagatcgcctaaGCGAATTTCTTATGATTTATGATTTTTATTATTAAATA AGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTT GAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCttagacaggtctttgccataaaccagc aggtcttggtgtaactctggttgttttggcacgaatgacttgtcctggtgcagccataagcatcgcaatcttctctcctttgctagtgtgttgacgtctatcccaagcagaaccacctgc agcttttaccttgatgccaatctctctgtagacagatcttgcagtagctatagcccaagcacatctaggtggtagatcatgcaatccagcttgacttgagatgtaatagggttcagcag cgtctatcagcctttcagcaactcttgctaatgcagctctgttctctcttgcagcgtaattctcaggagttagaccagcatcttgcaaccattcagcaggtagatagcatctgtcaatag ctgcatcgtcgataatatctctcgcgatgtttgtcagctgaaaagccaaacctagatcacaagctctgtccaaaaccctttcgtctcttacacccattactcttgccatcatcaaaccaa ctactccagcaacatggtagcagtatctcaaggtgtcttcaaaggtcacgtaacgagtttgagcaacatccattgcgaaaccatccaagtgatcaagtgccattcttggcgtaatac cgtgtgttagtgcaacttcttggaatgcagcaaaagcaggatcttgcatctcagcaccttcaaaagctgcaagtgtaagcgttctcaatctagccaatctctgagtagcctcttcttctg cagcagcttcagatgcgaaaccatgtgtctggtcatctataacgtcatcacagtgtctacaccaagtgtacagcataagcactgatcttctagtagctgggtcaaacagtttagctgc tgtagcaaaggacttggaaccattagccatcgtttgagtagcatgatccaacaaaggtggttgactcatTGTTTTATATTTGTTGTAAAAAGTAGATAA TTACTTCCTTGATGATCTGTAAAAAAGAGAAAAAGAAAGCATCTAAGAACTTGAAAAACTACGAATTAGA AAAGACCAAATATGTATTTCTTGCATTGACCAATTTATGCAAGTTTATATATATGTAAATGTAAGTTTCACG AGGTTCTACTAAACTAAACCACCCCCTTGGTTAGAAGAAAAGAGTGTGTGAGAACAGGCTGTTGTTGTCAC ACGATTCGGACAATTCTGTTTGAAAGAGAGAGAGTAACAGTACGATCGAACGAACTTTGCTCTGGAGATCA CAGTGGGCATCATAGCATGTGGTACTAAACCCTTTCCCGCCATTCCAGAACCTTCGATTGCTTGTTACAAAA CCTGTGAGCCGTCGCTAGGACCTTGTTGTGTGACGAAATTGGAAGCTGCAATCAATAGGAAGACAGGAAGT CGAGCGTGTCTGGGTTTTTTCAGTTTTGTTCTTTTTGCAAACAAATCACGAGCGACGGTAATTTCTTTCTCGA TAAGAGGCCACGTGCTTTATGAGGGTAACATCAATTCAAGAAGGAGGGAAACACTTCCTTTTTCTGGCCCT GATAATAGTATGAGGGTGAAGCCAAAATAAAGGATTCGCGCCCAAATCGGCATCTTTAAATGCAGGTATGC GATAGTTCCTCACTCTTTCCTTACTCACGAGTAATTCTTGCAAATGCCTATTATGCAGATGTTATAATATCTG TGCGTCTTGAGTTGAAGTCAGGAATCTAAAATAAAAATTAAGGTTAATAAAAAGAGGAAAGAAAAAAAAA TTAATCGATTTACAGAAACTTGCACACTAAAAATACACAACTAAAAGCAATTACAGTATGGGAAGTCATCG ACGTTATCTCTACTATAGTATATTATCATTTCTATTATTATCCTGCTCAGTGGTACTTGCAAAACAAGATAA GACCCCATTCTTTGAAGGTACTTCTTCGAAAAATTCGCGTCT
pFUS1-CrtI-tACT1	TACCATGTGGACCCTTTCAAAACAGAGTTGTATCTCTGCAGGATGCCCTTTTTGACGTATTGAATGGCATAA TTGCACTGTCACTTTTCGCGCTGTCTCATTTTGGTGCGATGATGAAACAAACATGAAACGTCTGTAATTTGA AACAAATAACGTAATTCTCGGGATTGGTTTTATTTAAATGACAATGTAAGAGTGGCTTTGTAAGGTATGTGT TGCTCTTAAAATATTTGGATACGACATCCTTTATCTTTTTTCCTTTAAGAGCAGGATATAAGCCATCAAGTTT CTGAAAATCAAAatgaagaaaaccgtagtgattggtgcaggttttggtggtttagctttggctatacgtctacaagctgcaggtattcctacagtgctattggagcaaaga gacaaaccaggaggaagagcttatgtttggcacgatcaaggctttaccttgatgctggtcctacagtcatcactgatcctactgcattggaagctttgttcaccttagctggtagaag aatggaagattatgtccgtcttattgcctgtcaagccgttttacagattgtgttgggaatctggtaaaaccctagattacgccaatgacagtgctgaactagaagctcagattacgcagt ttaatcccagagatgtcgaaggttacaggagattccttgcctattcccaagctgttttccaagagggttatcttcgtttgggttcagttccattcctgtcctttagggatatgcttagagca ggtcctcagttgttgaagctacaagcatggcaaagtgtgtatcagtctgtttcgagatttatcgaggatgaacatctgagacaagcattctcattccacagtcttctagttggaggtaat ccctttaccacatcgagcatatatacgttgattcacgctttggaaagagaatggggagtttggtttcctgaaggtggaacaggtgctttggttaatggtatggtgaagctattcacgga tttgggtggagaaatagagctgaatgcaagagtggaagaacttgttgtagcagacaacagagtctcacaagttagacttgctgatggtaggatcttcgatacagatgctgtagcttc aaacgcagatgtagtgaacacttataaaaagttgttgggacatcatcctgttggacaaaagagagcagctgctttggagaggaaatctatgagcaactcgttgtttgtcctttacttg ggctgaatcaaccacactcacaactagctcatcacacaatctgctttggtcctagatacagagagctgatagatgaaattttcactggatctgctttagcagacgatttttccctgtact tgcattcaccatgtgttactgatccctctttagcaccacctggttgtgctagcttctatgtactagcacctgtaccacatttgggtaatgctccattagattgggcacaagaaggaccga aattgagggataggatcttcgactatttggaagaacgttacatgccaggtttgagatctcagttggttacacagaggatattcacaccagctgattttcatgatactctagatgcgcatt taggtagcgcttttccattgagccacttttgacgcaaagtgcttggtttagaccacacaacagagattctgacattgccaatctgtacctagtaggtgcaggaactcatccaggagct ggtattcctggagttgtagcttctgctaaagctactgctagtctgatgatcgaggatttgcagtaaTCTCTGCTTTTGTGCGCGTATGTTTATGTATGT ACCTCTCTCTCTATTTCTATTTTTAAACCACCCTCTCAATAAAATAAAAATAATAAAGTATTTTTAAGGAAA AGACGTGTTTAAGCACTGACTTTATCTACTTTTTGTACGTTTTCATTGATATAATGTGTTTTGTCTCTCCCTTT TCTACGAAAATTTCAAAAATTGACCAAAAAAAGGAATATATATACGAAAAACTATTATATTTATATATCAT AGTGT
pTDH3-FAD1-tPGK1	AGTTTATCATTATCAATACTCGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTT ATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCG CTTTTTAAGCTGGCATCCAGAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAG TTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAAACGGGCACAACC TCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTATCTATCT CATTTTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCA GTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAA ATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTTAGTTTTAAAACACCAAGAACTTA GTTTCGAAGGATTCTAGAACTAGTAACatgcagttgagcaaggctgctgagatgtgttatgagataacaaactcttacttacacatagaccagaaatctc agataatagcaagtacacaagaagcgatacggttgacaagaaaatacttactaagtgaaatttttgtacgttggagtccactgaatggggaaatatcattctcgtacaacggaggaa aagattgccaggtattactactgttatatctgagttgcttatgggaatatttcttcattaaggctcaaaattcccaattcgatttcgagtttcaaagcttccccatgcaaagacttccaactg ttttcattgatcaagaagaaacttttccctacattagagaattttgtactggaaacctcagagcgatattgcctttccttatacgaatcacaaaggcaatctggtgcatcggtcaatatggc

	agacgcatttagagattttataaagatataccctgagaccgaagctatagtgataggtattagacacacagacccatttggtgaagcattaaagcctattcaaagaacagattctaac tggcctgattttatgaggttgcaacctctcttacactgggacttaaccaatatatggagtttcttactgtattctaatgagccaatttgtggactatatggtaaaggtttcacatcaatcggc ggaattaacaactcattgcctaacccacacttgagaaaggactccaataatccagccttgcattttgaatgggaaatcattcatgcatttggcaaggacgcagaaggcgaacgtagt tccgctataaacacgtcacctatttccgtggtggataaggaaagattcagcaaataccatgacaattactatcctggctggtatttggttgatgacactttagagagagcaggcagga tcaagaattaaATTGAATTGAATTGAAATCGATAGATCAATTTTTTTCTTTTCTCTTTCCCCATCCTTTACGCTAAA ATAATAGTTTATTTTATTTTTTGAATATTTTTTATTTATATACGTATATATAGACTATTATTTATCTTTTAATG ATTATTAAGATTTTTATTAAAAAAAAATTCGCTCCTCTTTTAATGCCTTTATGCAGTTTTTTTTTCCCATTCG ATATTTCTATGTTCGGGTTCAGCGTATTTTAAGTTTAATAACTCGAAAATTCTGCGTTCGTTAAAGCTTTCGA GAAGGATATTATTTCGAAATAAACCGTGTTGTGTAAGCTTGAAGCCTTTTTGCGCTGCCAATATTCTTATCC ATCTATTGTACTCTTTAGATCCAGTATAGTGTATTCTTCCTG
Expression Modules for Lycopene Readout Optimization (pMJ006, pMJ009, pMJ012)	
pFUS1-CrtI-tACT1	Same as pFUS1-CrtI-tACT1 expression module in *Lycopene Readout Modules* above.
pTDH3-tHMG1-tCYC1	AGTTTATCATTATCAATACTCGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTT ATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCG CTTTTTAAGCTGGCATCCAGAAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAG TTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAAACGGGCACAACC TCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTATCTATCT CATTTTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCA GTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAA ATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTTAGTTTTAAAACACCAGAACTTAG TTTCGACGGATTCTAGAACTAGTatggaccaattggtgaaaactgaagtcaccaagaagtcttttactgctcctgtacaaaaggcttctacaccagttttaacc aataaaacagtcatttctggatcgaaagtcaaaagtttatcatctgcgcaatcgagctcatcaggaccttcatcatctagtgaggaagatgattcccgcgatattgaaagcttggataa gaaaatacgtcctttagaagaattagaagcattattaagtagtggaaatacaaaacaattgaagaacaaagaggtcgctgccttggttattcacggtaagttacctttgtacgctttgg agaaaaaattaggtgatactacgagagcggttgcggtacgtaggaaggctctttcaattttggcagaagctcctgtattagcatctgatcgtttaccatataaaaattatgactacgac cgcgtatttggcgcttgttgtgaaaatgttataggttacatgcctttgcccgttggtgttataggccccttggttatcgatggtacatcttatcatataccaatggcaactacagagggtt gtttggtagcttctgccatgcgtggctgtaaggcaatcaatgctggcggtggtgcaacaactgttttaactaaggatggtatgacaagaggcccagtagtccgtttcccaacttttgaa aagatctggtgcctgtaagatatggttagactcagaagagggacaaaacgcaattaaaaaagcttttaactctacatcaagatttgcacgtctgcaacatattcaaacttgtctagca ggagatttactcttcatgagatttagaacaactactggtgacgcaatgggtatgaatatgatttctaaaggtgtcgaatactcattaaagcaaatggtagaagagtatggctgggaag atatggaggttgtctccgtttctggtaactactgtaccgacaaaaaaccagctgccatcaactggatcgaaggtcgtggtaagagtgtcgtcgcagaagctactattcctggtgatgt tgtcagaaaagtgttaaaaagtgatgtttccgcattggttgagttgaacattgctaagaatttggttggatctgcaatggctgggtctgttggtggatttaacgcacatgcagctaattta gtgacagctgttttcttggcattaggacaagatcctgcacaaaatgttgaaagttccaactgtataacattgatgaaagaagtggacggtgatttgagaatttccgtatccatgccatc catcgaagtaggtaccatcggtggtggtactgttctagaaccacaaggtgccatgttggacttattaggtgtaagaggcccgcatgctaccgctcctggtaccaacgcacgtcaatt agcaagaatagttgcctgtgccgtcttggcaggtgaattatccttatgtgctgccctagcagccggccatttggttcaaagtcatatgacccacaacaggaaacctgctgaaccaac aaaacctaacaatttggacgccactgatataaatcgtttgaaagatgggtccgtcacctgcattaaatcctaaCTCGAGTCATGTAATTAGTTATGTCACG CTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGG TCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGT ACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCT TTAATTTGCGGCC
pTDH3-FAD1-tCYC1	AGTTTATCATTATCAATACTCGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTT ATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCG CTTTTTAAGCTGGCATCCAGAAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAG TTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAAACGGGCACAACC TCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTATCTATCT CATTTTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCA GTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAA ATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTTAGTTTTAAAACACCAGAACTTAG TTTCGACGGATTCTAGAACTAGTAACatgcagttgagcaaggctgctgagatgtgttatgagataacaaactcttacttacacatagaccagaaatctcag ataatagcaagtacacaagaagcgatacggttgacaagaaaatacttactaagtgaaatttttgtacgttggagtccactgaatggggaaatatcattctcgtacaacggaggaaaa gattgccaggtattactactgttatatctgagttgcttatgggaatatttcttcattaaggctcaaaattcccaattcgatttcgagtttcaaagcttccccatgcaaagacttccaactgttt tcattgatcaagaagaaacttt ccctacattagagaattttgtactggaaacctcagagcgatattgcctttccttatacgaatcacaaaggcaatctggtgcatcggtcaatatggca gacgcatttagagattttataaagatataccctgagaccgaagctatagtgataggtattagacacacagacccatttggtgaagcattaaagcctattcaaagaacagattctaact ggcctgattttatgaggttgcaacctctcttacactgggacttaaccaatatatggagtttcttactgtattctaatgagccaatttgtggactatatggtaaaggtttcacatcaatcggc ggaattaacaactcattgcctaacccacacttgagaaaggactccaataatccagccttgcattttgaatgggaaatcattcatgcatttggcaaggacgcagaaggcgaacgtagt tccgctataaacacgtcacctatttccgtggtggataaggaaagattcagcaaataccatgacaattactatcctggctggtatttggttgatgacactttagagagagcaggcagga tcaagaattaaCTCGAGTCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCG AAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAA CGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAAC CTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCC

CellTag Expression Modules	
main CellTag expression pTDH3-(CellTag)-t*Ca*ADH1	CAGTTCGAGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCT AACTTTATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGT TACACAGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAG CCCGCTTTTTAAGCTGGCATCCAGAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCAT CAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAAACGGGCACA ACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGCATGTATCTA TCTCATTTTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAAC CAGTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGT AAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACCAAGAACT TAGTTTCGAATAAACACACATAAACAAACAAA.(***CellTag ORF***).tagCCGCGGTAAGCAAATAGCTAAATTATA TACGAATTAATATTATGATTAAGTGTTTACGTGAGTGCGATATTTTTATTACTATCTTATACAGTTGTATATA CTCTATAAAATGAGTTGTCTATTAATTAACGCGATGAATGCTTTCTGGGTTTACCTCTCCAACAACTCTAGTT TACTTCTCAATACATTCAATTGTATTTGATTTGTCAATACTTCATCATTAATCAATTCTATAGTTTTGTTTTTC TCGTTTATTTCCAAATTTAATGCATCAATTTTATTATTCAATTTGTCGTTGATTTTGGTTAATGATTTTATGGT TTGATCTCTGGCATTGATTGTTTGTGTTAGTTTTTCATTATTGATAATTAAATTATTTAAGTTAGTTATCAACT CGGTGTTTTCAAGTTTCAAGTTTTCAATTTCTTTAGAGTTTATTAGATTTGTCAAAGTTTCTGAATTGCTTGAT TGGTCCTGTAGAAGAGTATTTGTTGTTGTGGATAATTGATTCAATTTTTGAGACAATTGCTGGAAGGCGTTG AAATATCTAGCATCAATCTCATGGTTTTTTTCCCGAGAGTCTCGTAGATTCAATTGTTTTAATATATCTTGGG ACCACTCTTGATTTGAACTCATGGAAATTAAACTGGGTGTTGTGTTGTGGTGTAATGATTGTACCCCCTTTGC TTATAATTGTGTGG
ORF type 1 (FP1-***fs***-FP2) eGFP-***fs***-mCherry FPs in lower case, slippery site underlined, internal stop in bold	Atgtctaaaggtgaagaattattcactggtgttgtcccaattttggttgaattagatggtgatgttaatggtcacaaattttctgtctccggtgaaggtgaaggtgatgctacttacggta aattgaccttaaaatttatttgtactactggtaaattgccagttccatggccaaccttagtcactactttcggttatggtgttcaatgttttgctagatacccagatcatatgaaacaacatg actttttcaagtctgccatgccagaaggttatgttcaagaaagaactatttttttcaaagatgacggtaactacaagaccagagctgaagtcaagtttgaaggtgataccttagttaata gaatcgaattaaaaggtattgattttaaagaagatggtaacattttaggtcacaaattggaatacaactataactctcacaatgtttacatcatggctgacaaacaaaagaatggtatca aagttaacttcaaaattagacacaacattgaagatggttctgttcaattagctgaccattatcaacaaaatactccaattggtgatggtccagtcttgttaccagacaaccattacttatc cactcaatctgccttatccaaagatccaaacgaaaagagagaccacatggtcttgttagaatttgttactgctgctggtattacccatggtatggatgaattgtacaaaGCTAGC GGCAGCGGCGACTACAAGGACGACGACGACAAGACTTTAAACTAGTTGA[*pseudoknot variable region*]GGCG GTTCTATGGGAATGTCTGGAGACGTCatggtttcaaaaggtgaagaagataatatggctattattaaagaatttatgagatttaaagttcatatggaaggttc agttaatggtcatgaatttgaaattgaaggtgaaggtgaaggtagaccatatgaaggtactcaaactgctaaattgaaagttactaaaggtggtccattaccatttgcttgggatatttt gtcaccacaatttatgtatggttcaaaagcttatgttaaacatccagctgatattccagattatttaaaattgtcatttccagaaggttttaaatgggaaagagttatgaattttgaagatgg tggtgttgttactgttactcaagattcatcattacaagatggtgaatttatttataaagttaaattgagaggtactaattttccatcagatggtccagttatgcaaaaaaaaactatgggttg ggaagcttcatcagaaagaatgtatccagaagatggtgctttaaaaggtgaaattaaacaaagattgaaattaaaagatggtggtcattatgatgctgaagttaaaactacttataaa gctaaaaaaccagttcaattaccaggtgcttataatgttaatattaaattggatattacttcacataatgaagattatactattgttgaacaatatgaaagagctgaaggtagacattcaa ctggtggtatggatgaattatataaa
ORF type 2 (FP2-***fs***-FP1) mCherry-***fs***-eGFP	Same as eGFP-***fs***-mCherry module with portions in lower case exactly swapped.
ORF type 3 (***fs***-FP1-***fs***-FP2) ***fs***-eGFP-***fs***-mCherry	ATGGGTTCAGGTGAACAATCAAAGACTTTAAACTAGTTGA[*pseudoknot variable region*]GGCGGTTCTATGGG AATGTCTGGAGTCGAC[*eGFP ORF from above*]GCTAGCGGCAGCGGCGACTACAAGGACGACGACGACAAG ACTTTAAACTAGTTGA[*pseudoknot variable region*]GGCGGTTCTATGGGAATGTCTGGAGACGTC[m*Cherry ORFfrom above*]
ORF type 4 (***fs***-FP2-***fs***-FP1) ***fs***-mCherry-***fs***-eGFP	Same as ***fs***-eGFP-***fs***-mCherry module with FP ORFs exactly swapped.
Pseudoknot A (0.3%)	CGCGCAACTAATTCAGGCCGCGTTAAACGTTCTAGAA
Pseudoknot B (4.2%)	CGCGTCGCTAACACGTGGCGCGTTAAACCATCTAGAA
Pseudoknot C (9.4%)	CGCGGATCTAGCTTGTAACGCGTTAAACCAGCTAGAA
Pseudoknot D (30%)	CGCGTCGCTACCGCCCGGCGCGTTAAACACACTAGAA

7.5.4 Key primers

Gibson assembly was used for receptor cloning except where restriction sites are underlined.

Primers used for cloning fungal receptors from genomic DNA and pLPreB	
Sc.Ste2	MJ492 ACCAAGAACTTAGTTTCGACGGATACTAGTAAAATGTCTGATGCGGCTCCTTC MJ493 ACGAAATTACTTTTTCAAAGCCGTCTCGAGCTATAAATTATTATTATCTTCAGTCCAGAA
Ca.Ste2	MJ440 acgtcaaggagaaaaaaccccggaaactagtaAAATGAATATCAATTCAACTTTCATACC MJ362 gcaagtctcgagCTACACTCTTTTGATGGTGATTTG
Cg.Ste2	MJ498 ACCAAGAACTTAGTTTCGACGGATACTAGTAAAATGGAGATGGGCTACGATCC MJ499 ACGAAATTACTTTTTCAAAGCCGTCTCGAGCTATTTGTCACACTGACTTTGTTG
Le.Ste2	MJ504 ACCAAGAACTTAGTTTCGACGGATACTAGTAAAATGGACGAAGCAATCAATGCAAAC MJ505 ACGAAATTACTTTTTCAAAGCCGTCTCGAGCTATTTTTTCAACATAGTCACTTC
Pb.Ste2	MJ508 ACCAAGAACTTAGTTTCGACGGATACTAGTAAAATGGCACCCTCATTCGACC MJ509 ACGAAATTACTTTTTCAAAGCCGTCTCGAGCTAGGCCTTTGTGCCAGCTTC
Zr.Ste2	MJ518 ACCAAGAACTTAGTTTCGACGGATACTAGTAAAATGAGTGAGATTAACAATTCTACCTAC MJ519 ACGAAATTACTTTTTCAAAGCCGTCTCGAGCTATAATTTCTTTAGGATAATTTTTTTACT
Primers used for genotyping MTL loci of C. albicans	
MTLα	SB469 TGTAAACATCCTCAATTGTACCCGA SB470 TTCGAGTACATTCTGGTCGCG
MTLa1	SB471 TTCGAGTACATTCTGGTCGCG SB472 ATCAATTCCCTTTCTCTTCGATTAGG
Primers used for cloning stress promoter reporters	
pACT1	MJ615 ctcactaaagggaacaaaagctggagctctagtCCTTAAAAACATATGCCTCACCCT MJ616 GCATATTTTCTTTAATCAATTCAGACATtttctagaCAGTAAATTTTCGATCTTGGGAAG
pTEF1	MJ617 ctcactaaagggaacaaaagctggagctctagtATAGCTTCAAAATGTTTCTACTCCT MJ618 GCATATTTTCTTTAATCAATTCAGACATtttctagaTTAGATTGCTATGCTTTCTTTC
pHSP12	MJ619 ctcactaaagggaacaaaagctggagctctagtAGTGAAAATCTCCGGGAGCG MJ620 GCATATTTTCTTTAATCAATTCAGACATtttctagaTGAGTTGTTTGTTTGAGATTATCG
pSSA1	MJ621 ctcactaaagggaacaaaagctggagctctagtGGCATTTTCGTTCTTGTGGA MJ622 GCATATTTTCTTTAATCAATTCAGACATtttctagaATTTTTGTTTCTTGTAATACTTGA
pERO1	MJ623 ctcactaaagggaacaaaagctggagctctagtAAAGAACACGGCGGTAAGAA MJ624 GCATATTTTCTTTAATCAATTCAGACATtttctagaTTTACCTGCACGTTACTGTGG
pGPD1	MJ625 ctcactaaagggaacaaaagctggagctctagtCTGGGGTTTGAGCAAGTCTA MJ626 GCATATTTTCTTTAATCAATTCAGACATtttctagaTTATCAATATTTGTGTTTGTGGAG
pFUS1	MJ627 ctcactaaagggaacaaaagctggagctctagtTGCCTCAATCCTTCTTTTGCTT MJ628 GCATATTTTCTTTAATCAATTCAGACATtttctagACTTGATGGCTTATATCCTGCTCT
pTRX2	MJ629 ctcactaaagggaacaaaagctggagctctagtACTTTTACGGGTGGCAACG MJ630 GCATATTTTCTTTAATCAATTCAGACATtttctagaTCGTAGACTCTCGTGTATGTGTGC
pRNR3	MJ631 ctcactaaagggaacaaaagctggagctctagtGTAATAACAAGCAGGTGGGCG MJ632 GCATATTTTCTTTAATCAATTCAGACATtttctagaTTATTGCTGCTGCTATTCTTGCTT
pHXT1	MJ633 ctcactaaagggaacaaaagctggagctctagtTGCAAAAAGCTTCCGATCCT MJ634 GCATATTTTCTTTAATCAATTCAGACATtttctagACGTATATCAACTAGTTGACGATTA
pCUP1	MJ635 ctcactaaagggaacaaaagctggagctctagtTCACCACCCTTTATTTCAGGC MJ636 GCATATTTTCTTTAATCAATTCAGACATtttctagaTGTGATGATTGATTGATTGATTGT
pZRT1	MJ637 ctcactaaagggaacaaaagctggagctctagtGGCAAGAGTATTTCAGACTTTCCT MJ638 GCATATTTTCTTTAATCAATTCAGACATtttctagATTTGTGCTGTTGTTTTATTGTCT
pFET3	MJ666 ctcactaaagggaacaaaagctggagctctagtGATAATGCCTTGGCTTGCCT MJ640 GCATATTTTCTTTAATCAATTCAGACATtttctagaTACTCTTCCTTACACTGGGGTCC
pARR3	MJ641 ctcactaaagggaacaaaagctggagctctagtCACGTGCAAAATCTTCTCTTCG MJ642 GCATATTTTCTTTAATCAATTCAGACATtttctagaCCTGATGATTTGTTGGTTGGGT
pHIS4	MJ643 ctcactaaagggaacaaaagctggagctctagtAAACCCATGCACAGTGACTC MJ644 GCATATTTTCTTTAATCAATTCAGACATtttctagaATTGTATTACTATTACACAGCGCA
pPRM5	MJ645 ctcactaaagggaacaaaagctggagctctagtCTCACCCGGATCGTAGTCAC MJ646 GCATATTTTCTTTAATCAATTCAGACATtttctagaTCTTGCGTTTTGAGTGTCAATTT
pDAL5	MJ649 ctcactaaagggaacaaaagctggagctctagtAGCGTTCTCATCAGTCACTTG MJ650 GCATATTTTCTTTAATCAATTCAGACATtttctagaATCCTTGTTTTGTTGTTTTCTTCA
pPMC1	MJ651 ctcactaaagggaacaaaagctggagctctagtGTTTTTACCCGGCAAAGAAGC MJ652 GCATATTTTCTTTAATCAATTCAGACATtttctagaTATTTTTTTTGTTACGCACACAGT

pOLE1	MJ653 ctcactaaagggaacaaaagctggagctctagtCATGTCCCGGGGTTAGCG MJ654 GCATATTTTCTTTAATCAATTCAGACATtttctagaTTGTTGTAATGTTTTAGTGCTGT

7.5.5 Open reading frames of fungal receptors cloned

Description	Sequence
Yeast codon-optimized fungal receptor ORFs	
B. cinerea STE2 homologue	ATGGCTTCTAACTCTTCTAACTTCGACCCATTGACTCAATCTATCACTATCTTGATGGCTGACGGTA TCACTACTGTTTCTTTCACTCCATTGGACATCGACTTCTTCTACTACTACAACGTTGCTTGTTGTATC AACTACGGTGCTCAAGCTGGTGCTTGTTTGTTGATGTTCTTCGTTGTTGTTGTTTTGACTAAGGCTGT TAAGAGAAAGACTTTGTTGTTCGTTTTGAACGTTTTGTCTTTGATCTTCGGTTTCTTGAGAGCTATGT TGTACGCTATCTACTTCTTGCAAGGTTTCAACGACTTCTACGCTGCTTTCACTTTCGACTTCTCTAGA GTTCCAAGATCTTCTTACGCTTCTTCTGTTGCTGGTTCTGTTATCCCATTGTGTATGACTATCACTGT TAACATGTCTTTGTACTTGCAAGCTTACACTGTTTGTAAGAACTTGGACGACATCAAGAGAATCAT CTTGACTACTTTGTCTGCTATCGTTGCTTTGTTGGCTATCGGTTTCAGATTCGCTGCTACTGTTGTTA ACTCTGTTGCTATCTTGGCTACTTCTGCTTCTTCTGTTCCAATGCAATGGTTGGTTAAGGGTACTTTG GTTACTGAAACTATCTCTATCTGGTTCTTCTCTTTGATCTTCACTGGTAAGTTGGTTTGGACTTTGTA CAACAGAAGAAGAAACGGTTGGAGACAATGGTCTGCTGTTAGAATCTTGGCTGCTATGGGTGGTTG TACTATGGTTATCCCATCTATCTTCGCTATCTTGGAATACGTTACTCCAGTTTCTTTCCCAGAAGCTG GTTCTATCGCTTTGACTTCTGTTGCTTTGTTGTTGCCAATCTCTTCTTTGTGGGCTGGTATGGTTACT GACGAAGAAACTTCTGCTATCGACGTTTCTAACTTGACTGGTTCTAGAACTATGTTGGGTTCTCAAT CTGGTAACTTCTCTAGAAAGACTCACGCTTCTGACATCACTGCTCAATCTTCTCACTTGGACTTCTC TTCTAGAAAGGGTTCTAACGCTACTATGATGAGAAAGGGTTCTAACGCTATGGACCAAGTTACTAC TATCGACTGTGTTGTTGAAGACAACCAAGCTAACAGAGGTTTGAGAGACTCTACTGAAATGGACTT GGAAGCTATGGGTGTTAGAGTTAACAAGTCTTACGGTGTTCAAAAGGCTTAG
F. graminearum STE2 homologue	ATGTCTAAGGAAGTTTTCGACCCATTCACTCAAAACGTTACTTTCTTCGCTCCAGACGGTAAGACTG AAATCTCTATCCCAGTTGCTGCTATCGACCAAGTTAGAAGAATGATGGTTAACACTACTATCAACT ACGCTACTCAATTGGGTGCTTGTTTGATCATGTTGGTTGTTTTGTTGGTTATGGTTCCAAAGGAAAA GTTCAGAAGACCATTCATGATCTTGCAAATCACTTCTTTGGTTATCTCTTGTTGTAGAATGTTGTTGT TGTCTATCTTCCACTCTTCTCAATTCTTGGACTTCTACGTTTTCTGGGGTGACGACCACTCTAGAATC CCAAGATCTGCTTACGCTCCATCTGTTGCTGGTAACACTATGTCTTTGTGTTTGGTTATCTCTGTTGA AACTATGTTGATGTCTCAAGCTTGGACTATGGTTAGATTGTGGCCAAACGTTTGGAAGTACATCAT CGCTGGTGTTTCTTTGATCGTTTCTATCATGGCTATCTCTGTTAGATTGGCTTACACTATCATCCAAA ACAACGCTGTTTTGAAGTTGGAACCAGCTTTCCACATGTTCTGGTTGATCAAGTGGACTGTTATCAT GAACGTTGCTTCTATCTCTTGGTGGTGTGCTATCTTCAACATCAAGTTGGTTTGGCACTTGATCTCT AACAGAGGTATCTTGCCATCTTACAAGACTTTCACTCCAATGGAAGTTTTGATCATGACTAACGGT ATCTTGATGATCATCCCAGTTATCTTCGCTTCTTTGGAATGGGCTCACTTCGTTAACTTCGAATCTGC TTCTTTGACTTTGACTTCTGTTGCTGTTATCTTGCCATTGGGTACTTTGGCTGCTCAAAGAATCGCTT CTTCTGCTCCATCTTCTGCTAACTCTACTGGTGCTTCTTCTGGTATCAGATACGGTGTTTCTGGTCCA TCTTCTTTCACTGGTTTCAAGGCTCCATCTTTCTCTACTGGTACTACTGACAGACCACACGTTTCTAT CTACGCTAGATGTGAAGCTGGTACTTCTTCTAGAGAACACATCAACCCACAAGGTGTTGAATTGGC TAAGTTGGACCCAGAAACTGACCACCACGTTAGAGTTGACAGAGCTTTCTTGCAAAGAGAAGAAA GAATCAGAGCTCCATTGTAG
M. oryzea STE2 homologue	ATGGACCAAACTTTGTCTGCTACTGGTACTGCTACTTCTCCACCAGGTCCAGCTTTGACTGTTGACC CAAGATTCCAAACTATCACTATGTTGACTCCAGCTTTGATGGGTCAAGGTTTCGAAGAAGTTCAAA CTACTCCAGCTGAAATCAACGACGTTTACTTCTTGGCTTTCAACACTGCTATCGGTTACTCTACTCA AATCGGTGCTTGTTTCATCATGTTGTTGGTTTTGTTGACTATGACTGCTAAGGCTAGATTCGCTAGA ATCCCAACTATCATCAACACTGCTGCTTTGGTTGTTTCTATCATCAGATGTACTTTGTTGGTTATCTT CTTCACTTCTACTATGATGGAATTCTACACTATCTTCTCTGACGACTTCTCTTTCGTTCACCCAAACG ACATCAGAAGATCTGTTGCTGCTACTGTTTTCGCTCCATTGCAATTGGCTTTGGTTGAAGCTGCTTT GATGGTTCAAGCTTGGGCTATGGTTGAATTGTGGCCAAGAGCTTGGAAGGTTTCTGGTATCGCTTTC TCTTTGATCTTGGCTACTGTTACTGTTGCTTTCAAGTGTGCTTCTGCTGCTGTTACTGTTAAGTCTGC TTTGGAACCATTGGACCCAAGACCATACTTGTGGATCAGACAAACTGACTTGGCTTTCACTACTGC TATGGTTACTTGGTTCTGTTTCTTGTTCAACGTTAGATTGATCATGCACATGTGGCAAAACAGATCT ATCTTGCCAACTGTTAAGGGTTTGTCTCCAATGGAAGTTTTGGTTATGGCTAACGGTTTGTTGATGG TTTTCCCAGTTTTGTTCGCTGGTTTGTACTACGGTAACTTCGGTCAATTCGAATCTGCTTCTTTGACT ATCACTTCTGTTGTTTTGGTTTTGCCATTGGGTACTTTGGTTGCTCAAAGATTGGCTGTTAACAACA CTGTTGCTGGTTCTTCTGCTAACACTGACATGGACGACAAGTTGGCTTTCTTGGGTAACGCTACTAC TGTTACTTCTTCTGCTGCTGGTTTCGCTGGTTCTTCTGCTTCTGCTACTAGATCTAGATTGGCTTCTC CAAGACAAAACTCTCAATTGTCTACTTCTGTTTCTGCTGGTAAGCCAAGAGCTGACCCAATCGACTT GGAATTGCAAAGAATCGACGACGAAGACGACGACTTCTCTAGATCTGGTTCTGCTGGTGGTGTTAG AGTTGAAAGATCTATCGAAAGAAGAGAAGAAAGATTGTAG
Z. bailii STE2 homologue	ATGTCTGGTTTGGCTAACAACACCTCTTACAACCCATTGGAATCTTTCATTATTTTCACTTCTGTTTA CGGTGGTGATACCATGGTTAAGTTCGAAGACTTGCAATTAGTCTTCACCAAGCGTATTACTGAAGG TATTTTGTTCGGTGTCAAGGTTGGTGCCGCTTCTTTGACTATGATTGTTATGTGGATGATTTCCAGA AGAAGAACCTCCCCAATCTTCATCATGAACCAATTGTCTTTGGTTTTCACCATCTTGCACGCTTCTT TTTACTTTAAGTACTTATTGGACGGTTTCGGTTCTATTGTCTACACTTTGACCTTGTTCCCACAATTA ATTACTTCCTCTGACTTGCACGTTTTCGCTACTGCTAACGTTGTTGAAGTCTTATTGGTTTCTTCCAT CGAAGCCTCTTTGGTTTTCCAAGTCAACGTCATGTTCGCTGGTTCTAACCACAGAAAGTTCGCTTGG

	TTGTTGGTCGGTTTCTCTTTGGGTTTGGCTTTGGCCACTGTCGCTTTGTACTTCGTTACTGCTGTCAA GATGATCGCTTCCGCTTACGCTTCTCAACCACCAACTAACCCAATCTACTTCAACGTTTCCTTGTTC TTGTTGGCTGCCTCCGTTTTCTTGATGACTTTAATGTTGACCGTCAAGTTGATCTTGGCTATCAGATC CAGAAGATTCTTGGGTTTGAAGCAATTCGACTCTTTCCACATTTTGTTGATTATGTCTTGTCAAACT TTGATCGCTCCATCTGTTTTGTACATCTTGGGTTTTATTTTGGATCACAGAAAGGGTAACGACTACT TGATTACCGTCGCTCAATTGTTGGTCGTTTTGTCTTTGCCATTGTCCTCCATGTGGGCCACTACTGCT AACGATGCTTCCTCCGGTACTTCTATGTCTTCCAAGGAATCCGTCTACGGTTCTGATTCCTTATACT CTAAGTCTAAGTGTTCCCAATTCACCAGAACCTTCATGAACAGATTCTCTACTAAGCCAACTAAGA ACGACGAAATTTCTGATTCCGCTTTCGTCGCTGTTGATTCCTTGGAAAAGAACGCTCCACAAGGTAT CTCTGAACACGTTTGTGAATTCCCACAATCTGACTTATCTGATCAAGCTACTTCCATCTCCTCCAGA AAAAAGGAAGCTGTTGTTTACGCTTCCACTGTTGATGAAGATAAGGGTTCTTTCTCCTCTGACATCA ACGGTTACACTGTTACCAACATGCCATTGGCTTCCGCTGCTTCTGCTAACTGTGAAAACTCCCCATG TCACGTTCCAAGACCATACGAAGAAAACGAAGGTGTCGTCGAAACCAGAAAAATTATTTTGAAGA AGAACGTCAAATGGTAG
H. capsulatum STE2 homologue	ATGTCTTCCTCTTTCGACCCATTCGACCAAAACGTTGTCTTCCACAAGGCTGATGGTACTCCATTCA ACGTTTCCATTCACGACTTGGACGAATTCGTCCAATACGGTATTAGAGTCTGTATCAACTACGCTGC TCAATTGGGTGCCACTGTCATTGCTATTGTCATGTTGGCTTTGTTGACTCAATCCGATAAGAGAAGA ACCCCAGTCTTCTTCTTGAACACTTCTGCTTTGACTATGAACTTCGCTAGATTGTTGTGTATGACTAT TTACTTCACCACTGGTTTCAACTCTACCTACGCTTTCTTCTCTTTGGACTACTCCAGAGTCCCAGGTT CTGCCTACGCTGATTCTATCTTGGGTATCGCTTTCGCTACTATTTTGGTCATTTGTATGGAAATGTCT TTGGTTATTCAAACTCAAGTTGTTTGTGCCACCTTGTCTGAAATCCAAAGAAGATTGTTGTTGGTCG TCTCCATTTTGATCGCTTTGTTGGCTATTGGTTTCAGAATGGGTTTGATGGTTGAAAACTGTATCGC TATCATGAACGCCTCTAACTTCAGACCATTCATCTGGTTGCAATCCGCTTCTAACATTGCTATTACC ATCTCTACTTGTTTCTTCTCTGCTGTTTTCGTCACCAAGTTGGGTTACGCTTTGGTTACCAGAAGAAG ATTGGGTATGACTAGATTCGGTGCTATGCAAGTCATGTTCATTATGTCCTTCCAAACTATGGTCATC CCAGCTATTTTCTCCATTATCCAATACCCAATCCCATTGTACGAAATGAACTCTAACGTCTTCACTT TGGTCGCTATTTTCTTGCCATTGTCTTCCTTGTGGGCCGCTGCTGCTACTAAGCACTCCTTCGAAACT TTGACCTCTGGTCCACACCAATACTTGTGGTCCTCTGAAAGATCCAACTCTACCTCTTCCGCTACCG GTCACCAAGGTTCTTTGTGTCAAAACCAATCTACTATCAGATCTGGTGGTTCTGTTGCTACTTCCTT GTCCCCAGACCAATTGGACAGATTGTACACTGGTTTGGACTTCGACGCCTGTGCCAAGGCTTAG
C. glabrata STE2 homologue	ATGGAAATGGGTTACGACCCAAGAATGTACAACCCAAGAAACGAATACTTGAACTTCACTTCTGTT TACGACGTTAACGACACTATCAGATTCTCTACTTTGGACGCTATCGTTAAGGGTTTGTTGAGAATCG CTATCGTTCACGGTGTTAGATTGGGTGCTATCTTCATGACTTTGATCATCATGTTCATCTCTTCTAAC ACTTGGAAGAAGCCAATCTTCATCATCAACATGGTTTCTTTGATGTTGGTTATGATCCACTCTGCTT TGTCTTTCCACTACTTGTTGTCTAACTACTCTTCTATCTCTTACATCTTGACTGGTTTCCCACAATTG ATCACTTCTAACAACAAGAGAATCCAAGACGCTGCTTCTATCGTTCAAGTTTTGTTGGTTGCTGCTA TCGAAGCTTCTTTGGTTTTCCAAATCCACGTTATGTTCACTATCGAAAACATCAAGTTGATCAGAGA AATCGTTTTGTCTATCTCTATCGCTATGGGTTTGGCTACTGTTGCTACTTACTTGGCTGCTGCTATCA AGTTGATCAGAGGTTTGCACGACGAAGTTATGCCACAAACTCACTTGATCTTCAACTTGTCTATCAT CTTATTGGCTTCTTCTATCAACTTCATGACTTTCATCTTAGTTATCAAGTTGTTCTTCGCTATCAGAT CTAGAAGATACTTAGGTTTGAGACAATTCGACGCTTTCCACATCTTGTTGATCATGTTCTGTCAATC TTTGTTGATCCCATCTGTTTTGTACATCATCGTTTACGCTGTTGACTCTAGATCTAACCAAGACTACT TGATCCCAATCGCTAACTTGTTCGTTGTTTTGTCTTTGCCATTGTCTTCTATCTGGGCTAACACTTCT AACAACTCTTCTAGATCTCCAAAGTACTGGAAGAACTCTCAAACTAACAAGTCTAACGGTTCTTTC GTTTCTTCTATCTCTGTTAACTCTGACTCTCAAAACCCATTGTACAAGAAGATCGTTAGATTCACTT CTAAGGGTGACACTACTAGATCTATCGTTTCTGACTCTACTTTGGCTGAAGTTGGTAAGTACTCTAT GCAAGACGTTTCTAACTCTAACTTCGAATGTAGAGACTTGGACTTCGAAAAGGTTAAGCACACTTG TGAAAACTTCGGTAGAATCTCTGAAACTTACTCTGAATTGTCTACTTTGGACACTACTGCTTTGAAC GAAACTAGATTGTTCTGGAAGCAACAATCTCAATGTGACAAGTAG
Fungal receptor ORFs from genomic DNA (stop codons changed to TAG)	
Wildtype *S. cerevisiae* STE2	ATGTCTGATGCGGCTCCTTCATTGAGCAATCTATTTTATGATCCAACGTATAATCCTGGTCAAAGCA CCATTAACTACACTTCCATATATGGGAATGGATCTACCATCACTTTCGATGAGTTGCAAGGTTTAGT TAACAGTACTGTTACTCAGGCCATTATGTTTGGTGTCAGATGTGGTGCAGCTGCTTTGACTTTGATT GTCATGTGGATGACATCGAGAAGCAGAAAAAACGCCGATTTTCATTATCAACCAAGTTTCATTGTTT TTAATCATTTTGCATTCTGCACTCTATTTTAAATATTTACTGTCTAATTACTCTTCAGTGACTTACGC TCTCACCGGATTTCCTCAGTTCATCAGTAGAGGTGACGTTCATGTTTATGGTGCTACAAATATAATT CAAGTCCTTCTTGTGGCTTCTATTGAGACTTCACTGGTGTTTCAGATAAAAGTTATTTTCACAGGCG ACAACTTCAAAAGGATAGGTTTGATGCTGACGTCGATATCTTTCACTTTAGGGATTGCTACAGTTAC CATGTATTTTGTAAGCGCTGTTAAAGGTATGATTGTGACTTATAATGATGTTAGTGCCACCCAAGAT AAATACTTCAATGCATCCACAATTTTACTTGCATCCTCAATAAACTTTATGTCATTTGTCCTGGTAG TTAAATTGATTTTAGCTATTAGATCAAGAAGATTCCTTGGTCTCAAGCAGTTCGATAGTTTCCATAT TTTACTCATAATGTCATGTCAATCTTTGTTGGTTCCATCGATAATATTCATCCTCGCATACAGTTTGA AACCAAACCAGGGAACAGATGTCTTGACTACTGTTGCAACATTACTTGCTGTATTGTCTTTACCATT ATCATCAATGTGGGCCACGGCTGCTAATAATGCATCCAAAACAAACACAATTACTTCAGACTTTAC AACATCCACAGATAGGTTTTATCCAGGCACGCTGTCTAGCTTTCAAACTGATAGTATCAACAACGA TGCTAAAAGCAGTCTCAGAAGTAGATTATATGACCTATATCCTAGAAGGAAGGAAACAACATCGG ATAAACATTCGGAAAGAACTTTTGTTTCTGAGACTGCAGATGATATAGAGAAAAATCAGTTTTATC AGTTGCCCACACCTACGAGTTCAAAAAATACTAGGATAGGACCGTTTGCTGATGCAAGTTACAAAG

	AGGGAGAAGTTGAACCCGTCGACATGTACACTCCCGATACGGCAGCTGATGAGGAAGCCAGAAAG TTCTGGACTGAAGATAATAATAATTTATAG
C. glabrata STE2 homologue	ATGGAGATGGGCTACGATCCAAGAATGTATAATCCAAGAAATGAATACTTGAATTTCACGTCGGTA TATGATGTAAATGACACAATCAGATTTTCGACTCTGGACGCCATTGTAAAAGGATTGCTTAGAATT GCCATTGTTCATGGAGTTAGATTGGGAGCAATATTCATGACGTTAATAATAATGTTTATCTCATCAA ATACATGGAAAAAACCCATATTTATAATTAACATGGTGTCGTTGATGTTAGTTATGATTCATTCCGC ACTTAGCTTCCATTACCTTTTATCGAATTATTCTTCAATTTCTTATATACTGACAGGGTTTCCTCAGT TGATTACAAGCAATAATAAACGAATTCAAGATGCAGCGAGTATAGTCCAAGTTTTATTGGTTGCTG CGATAGAAGCATCATTGGTATTTCAGATTCATGTTATGTTTACGATTGAAAACATTAAGCTTATTAG AGAAATAGTACTCTCTATATCGATAGCAATGGGATTGGCAACAGTGGCTACATATCTTGCTGCAGC AATAAAGCTGATAAGAGGACTGCATGATGAGGTAATGCCACAAACACATCTTATTTTCAATTTATC TATAATATTGCTTGCATCCTCCATAAATTTTATGACATTTATATTGGTCATTAAACTTTTCTTCGCTA TTAGATCTAGAAGATATCTCGGTCTTCGTCAATTCGATGCTTTTCATATTTTATTAATCATGTTCTGC CAGTCATTATTGATACCCTCAGTATTATATATTATAGTTTACGCGGTTGATAGCAGATCTAATCAGG ATTATCTGATTCCAATTGCCAATTTATTTGTTGTTTTATCTTTGCCATTATCCTCTATCTGGGCTAAC ACATCAAATAACTCATCCAGATCTCCAAAATATTGGAAAAACTCTCAAACGAATAAGAGCAATGG GTCTTTTGTCTCTTCAATATCTGTCAATAGTGACTCACAAAACCCTTTGTACAAAAAGATTGTACGT TTTACATCAAAAGGCGACACTACCCGTAGTATTGTAAGTGATTCAACATTAGCAGAGGTGGGAAAA TACTCTATGCAAGACGTTAGCAATTCAAACTTTGAATGTCGAGACCTTGATTTTGAGAAGGTAAAA CATACTTGCGAAAATTTTGGCAGAATATCTGAAACATATAGTGAGTTAAGTACTTTAGATACCACT GCCCTCAATGAGACTCGGTTGTTTTGGAAACAACAAAGTCAGTGTGACAAATAG
C. albicans STE2 homologue	ATGAATATCAATTCAACTTTCATACCTGATAAACCAGGCGATATAATTATTAGTTATTCAATTCCAG GATTAGATCAACCAATTCAAATTCCTTTCCATTCATTAGATTCATTTCAAACCGATCAAGCTAAAAT AGCTTTAGTCATGGGGATAACTATTGGGAGTTGTTCAATGACATTAATTTTTTTGATTTCTATAATG TATAAAACTAATAAATTAACAAATTTAAAATTAAAATTAAAATTAAAATATATCTTGCAATGGATA AATCAAAAAATCTTCACCAAAAAAAGGAATGACAACAAACAACAACAACAACAACAACAACAAC AAATTGAATCATCATCATATAACAATACTACTACTACGCTGGGGGGTTATAAATTATTTTTATTTTA TCTTAATTCATTGATTTTATTAATTGGTATTATTCGATCAGGTTGTTATTTAAATTATAATTTAGGTC CATTAAATTCACTTAGTTTTGTATTTACTGGTTGGTATGATGGATCATCATTTATATCATCCGATGTA ACTAATGGATTTAAATGTATTTTATATGCTTTAGTGGAAATTTCATTAGGTTTCCAAGTTTATGTGA TGTTCAAAACTTCAAATTTAAAAATTTGGGGGATAATGGCATCATTATTATCAATTGGTTTAGGATT GATTGTTGTTGCCTTTCAAATCAATTTAACAATTTTATCTCATATTCGATTTTCCCGGGCTATATCAA CTAACAGAAGTGAAGAAGAATCATCATCATTATCATCTGATTCGGTTGGGTATGTGATTAATT CAATATGGATGGATTTACCAACAATATTATTTTCCATTAGTATTAATATAATGACAATATTATTGAT TGGTAAACTTATAATTGCTATTAGAACAAGACGTTATTTAGGATTGAAACAATTTGATAGTTTCCAT ATTTTATTAATTGGTTTCAGTCAAACATTAATTATTCCTTCAATTATTTTGGTGGTTCATTATTTTTA TTTATCACAAAATAAAGATTCTTTATTACAACAAATTAGTCTTTTATTGATTATTTTAATGTTACCAT TAAGTTCTTTATGGGCTCAAACTGCTAATAATACTCATAATATTAATTCATCTCCAAGTTTATCATT CATATCTCGTCATCATCTGTCTGATAGTAGTCGTAGTGGTGGTTCCAATACAATTGTTAGTAATGGT GGTAGTAATGGTGGTGGTGGTGGTGGGAATTTCCCTGTTTCAGGTATTGATGCACAATTACCA CCTGATATTGAAAAAATCTTACATGAAGATAATAATTATAAATTACTTAATAGTAATAATGAAAGT GTAAATGATGGAGATATTATCATTAATGATGAAGGTATGATTACTAAACAAATCACCATCAAAAGA GTGTAG
L. elongisporus STE2 homologue	ATGGACGAAGCAATCAATGCAAACCTTGTTTCTGGAGATATTATAGTCTCTTTTAACATTCCTGGTT TGCCAGAACCGGTACAAGTGCCATTCAGCGAATTTGATTCGTTTCATAAAGACCAGCTCATTGGAG TCATCATTCTTGGAGTCACTATTGGAGCATGCTCGCTTTTGTTGATATTGCTACTTGGAATGTTATA CAAGAGCCGTGAAAAGTATTGGAAATCACTATTATTTATGCTCAATGTATGCATCTTGGCTGCCAC AATCTTAAGGAGCGGTTGCTTCTTAGACTATTATCTAAGTGATTTGGCCAGTATCAGTTATACATTT ACTGGAGTATACAATGGTACCAGCTTTGCTAGCTCTGACGCGGCAAATGTGTTCAAGACTATTATG TTTGCCTTGATTGAAACTTCGTTAACCTTTCAAGTGTATGTCATGTTTCAAGGGACCACTTGGAAAA ATTGGGGCCATGCTGTCACTGCATTATCGGGTCTCTTGTCTGTTGCCTCAGTGGCGTTCCAGATCTA CACCACGATTTTATCCCACAATAATTTCAATGCTACAATCTCGGGAACCGGTACATTAACTTCAGGT GTTTGGATGGACTTACCAACACTCTTGTTTGCCGCAAGTATCAATTTTATGACCATTTTGTTGTTATT TAAGTTGGGAATGGCCATTAGACAAAGAAGGTATTTAGGTTTAAAACAGTTTGATGGGTTCCATAT CTTATTCATCATGTTTACCCAAACATTGTTCATACCCTCGATTTTGCTTGTGATCCACTACTTTTACC AGGCAATGTCTGGACCATTCATCATCAACATGGCGTTGTTCTTGGTGGTGGCATTCTTGCCATTGAG TTCATTATGGGCACAAACTGCAAACACTACTAAAAAGATTGAATCTTCGCCAAGTATGAGCTTTAT TACTAGACGAAAATCAGAGGATGAGTCACCACTGGCTGCTAACGACGAGGATAGGTTACGAAAAT TCACCACAACTTTGGATTTGTCGGGCAACAAGAACAATACAACAAACAATAATAACAATAGCAAC AACATTAACAACAATATGAGCAACATCAACTACCCTTCTACAGGACTGGGAGAAGACGATAAATC CTTTATATTTGAGATGGAACCCAGTCGGGAAAGAGCTGCAATAGAAGAGATTGATCTTGGAGCAA GGATCGATACCGGTTTGCCCAGAGATTTAGAGAAATTTCTAGTTGATGGGTTTGACGATAGTGATG ACGGAGAAGGAATGATAGCCAGAGAAGTGACTATGTTGAAAAAATAG
P. brasiliensis (lutzii) STE2 homologue	ATGGCACCCTCATTCGACCCCTTCAACCAAAGCGTGGTCTTCCACAAGGCCGACGGAACTCCATTC AACGTCTCAATCCATGAACTAGACGACTTCGTGCAGTACAACACCAAAGTCTGCATCAACTACTCT TCCCAGCTCGGAGCATCTGTCATTGCAGGACTCATGCTTGCCATGCTGACACACTCAGAAAAGCGT CGTCTGCCAGTTTTCTTCCTAAACACATTCGCACTGGCCATGAACTTTGCCCGCCTGCTCTGCATGA CCATCTACTTCACCACGGGCTTCAACAAGTCCTATGCCTACTTTGGTCAGGATTACTCCCAGGTGCC

	TGGGAGCGCCTACGCAGCCTCTGTCTTGGGCGTTGTCTTCACCACTCTCCTGGTAATCAGCATGGAA ATGTCCCTCCTGATCCAAACAAGGGTTGTCTGCACGACCCTTCCGGATATCCAACGTTATCTACTCA TGGCAGTTTCCTCCGCGATTTCCCTGATGGCCATCGGGTTCCGCCTTGGCTTAATGGTTGAGAACTG CATTGCCATTGTGCAGGCGTCGAATTTCGCCCCTTTTATCTGGCTTCAAAGCGCCTCGAACATCACC ATTACGATCAGCACATGTTTCTTCAGTGCCGTCTTTGTTACGAAATTGGCATATGCACTCGTCACTC GTATACGACTAGGCTTGACGAGGTTTGGTGCTATGCAGGTTATGTTCATCATGTCCTGCCAGACTAT GGTGATTCCAGCCATCTTCTCAATTCTCCAATACCCACTCCCCAAGTACGAAATGAACTCCAACCTC TTTACGCTGGTGGCCATTTTCCTCCCTCTTTCCTCGCTATGGGCTTCAGTTGCTACGAGATCCAGTTT CGAGACGTCTTCTTCCGGCCGCCATCAGTATCTTTGGCCAAGCGAACAGAGCAATAACGTCACCAA TTCGGAAATTAAGTATCAGGTCAGCTTCTCTCAGAACCACACTACGTTGCGGTCTGGAGGGTCTGT GGCCACGACACTCTCCCCGGACCGGCTCGACCCGGTTTATTGTGAAGTTGAAGCTGGCACAAAGGC CTAG
Z. rouxii STE2 homologue	ATGAGTGAGATTAACAATTCTACCTACAATCCAATGAATGCATATGTAACGTTTACATCAATATAT GGTGATGATACTATGGTACGTTTCAAAGATGTGGAATTGGTAGTTAACAAAAGGGTTACAGAAGCC ATTATGTTCGGCGTCAAAGTTGGTGCAGCTTCGTTGACACTCATCATCATGTGGATGATCTCTAAGA AAAGAACAACACCGATATTTATCATAAATCAGTCTTCGCTTGTATTTACCATAATACATGCTTCGCT TTATTTTGGGTACCTTTTGTCAGGATTTGGTAGTATAGTTTACAATATGACATCGTTCCCGCAGTTA ATAAGCTCCAATGACGTTCGTGTGTACGCAGCTACAAATATTTTTGAGGTCCTGTTGGTAGCATCTA TCGAAATCTCTCTGGTTTTTCAGGTCAAAGTTATGTTTGCCAACAATAATGGTCGAAGATGGACTTG GTGTTTGATGGTAGTTTCCATAGGGATGGCACTAGCTACTGTAGGACTTTATTTTGCCACTGCCGTT GAGTTGATCAGAGCTGCTTACAGCAATGATACTGTTAGCCGCCATGTTTTTTACAATGTTTCTCTGA TCTTACTAGCGTCATCTGTCAATCTAATGACACTAATGCTAGTGGTAAAATTAGTATTAGCGATCAG ATCAAGAAGATTTTTGGGGTTAAAACAGTTTGACAGTTTCCACATATTACTTATAATGTCTTGCCAG ACTCTAATAGCACCTTCCATTCTATTCATTTTGGGTTGGACCTTAGACCCTCATACTGGTAATGAGG TTTTAATTACAGTTGGTCAATTGCTAATAGTACTGTCATTACCGCTGTCATCTATGTGGGCTACAAC CGCTAACAATACCAGTTCATCTAGTAGTTCGGTGTCCTGTAATGACAGCTCTTTTGGTAATGACAAT CTCTGTTCCAAGAGTTCGCAATTTAGAAGAACTTTTATGAATAGATTCCGTCCCAAGTCGGTTAATG GTGACGGTAATTCTGAAAATACCTTTGTTACAATTGATGATTTGGAAAAAAGCGTTTTTCAAGAAT TATCAACACCTGTTAGCGGAGAATCAAAGATAGATCATGATCATGCAAGTAGTATTTCATGTCAAA AGACATGTAATCATGTTCATGCTTCGACAGTGAATTCAGATAAGGGATCTTGGTCCTCTGATGGTA GTTGTGGCAGTTCTCCGTTAAGAAAGACTTCCACCGTTAATTCTGAAGATTTACCTCCACATATATT GAGCGCCTACGATGACGATCGAGGTATAGTAGAAAGTAAAAAAATTATCCTAAAGAAATTATAG