

**Upstream Pheromone Signals and Downstream Putative Long Non-Coding RNAs
Work within the *STE12* Pathway to Regulate Mating-Type Switching in
*Ogataea polymorpha***

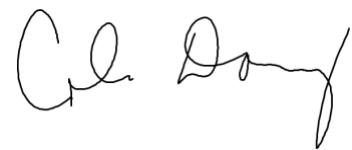
A Senior Thesis Presented to
The Faculty of the Department of Molecular Biology,
Colorado College

By
Juliana C. Olliff

Professor Sara J. Hanson
Primary Thesis Advisor



Professor Cole Dovey
Secondary Thesis Advisor



1. Abstract

Sexual reproduction is a risky and complex process that can lead to DNA damage or cell death. Thus, organisms who sexually reproduce use intricate signal cascades to ensure reproduction proceeds smoothly. Yeast are an optimal system for studying reproduction because scientists have elucidated many of the mechanisms responsible for the correct progression of reproduction. In yeast, mating is one of the best understood forms of reproduction and meiosis. In budding yeast, two mating types (MT), **a** and α , exist; these opposite MTs mate, go through meiosis and form spores. However, when a cell is isolated without an opposite MT partner, it is unable to mate and form spores.

To resolve this, yeast have evolved a mechanism, called mating-type switching (MTS), that allows an isolated cell to perform meiosis without a mating partner available. During MTS, yeast perform mitosis to form two identical cells. The mother cell then undergoes DNA rearrangement to switch to the opposite MT so the mother and daughter cell can mate. This rearrangement has significant risks and requires complex signal cascades to mitigate DNA damage. In some yeast species, the transcription factor *STE12* is one of the molecules responsible for this regulation, and is upregulated during both mating and MTS.

Our research focuses on MTS in the methylotrophic yeast *Ogataea polymorpha*, which is distantly related to the model yeast, *Saccharomyces cerevisiae*. While *STE12* has been identified as both necessary and sufficient for inducing MTS in *O. polymorpha*, less is known about how *STE12* controls the MTS signal cascade. Here, we

investigated two pathways related to *STE12*-mediated MTS in *O. polymorpha*; the upstream pheromone response pathway and potential regulatory downstream long non-coding (lnc)RNAs.

Previous research indicated that pheromones were not necessary to initiate the MTS signal cascade. However, it remains unclear whether pheromone exposure might suppress *STE12*-mediated MTS, because cells, sensing a mating partner was present, would divert their signal cascade towards mating rather than MTS. To test this hypothesis, we induced MTS in **a** cells and exposed the cultures to two variants of α pheromones. We saw qualitative suppression of MTS, leading us to develop a semi-quantitative PCR procedure to quantify suppression. Our semi-qPCR method correctly determined known cell mixture ratios, demonstrating the effectiveness of this method and making us confident in its use for future quantification.

Based on the presence of regulatory lncRNAs in sexual processes in other yeast species, we hypothesized that lncRNAs may help regulate the MTS signal cascade in *O. polymorpha*. We performed an RNA-seq analysis on switched cells to identify putative lncRNAs regulated by *STE12*. We used a bioinformatics pipeline to identify novel transcripts upregulated by *STE12*. We found 5 upregulated lncRNAs, one of which was proximally located near an important MTS gene and potentially involved in its regulation. Because of their correlation with *STE12* upregulation, these lncRNAs require further analysis to elucidate their potential roles in MTS. This study strengthens our general understanding of how sexual reproduction is transmitted, maintained and regulated.

2. Introduction

Sexual Reproduction is Complex and Heavily Regulated.

Sexual reproduction, an extremely complex process, often drives the behavior of many eukaryotic species. Despite the fact that a cell can pass only half of its genetic material to its offspring during sexual reproduction, environmental pressure has forced many species to rely on sexual, rather than asexual reproduction to keep genetic diversity in a population and ensure the species' survival (Wallen et al. 2018).

Additionally, meiosis comes with the risk of DNA damage and cell death, when double stranded (ds)DNA breaks are formed to mix chromosomes, known as crossing over, to increase diversity (Ginsburg et al. 2014).

The study of sexual reproduction is particularly interesting because of its connection to pathogenesis. Sexual reproduction is often triggered by stressful environmental conditions, which promote the expression of pathogenesis genes and the dispersal of pathogenic cells (Heitman 2006). Studies in *Candida albicans* have revealed one prominent example of the correlation between sexual reproduction and pathogenesis (Heitman 2006). *C. albicans* has two morphologies which are triggered by differences in the cell's environment: a white, mating-incompetent morphology and an opaque mating-competent morphology (Tao et al. 2014). While both morphologies are infectious, the ability to sexually reproduce coincides with a particular type of infectious behavior in the yeast (Tao et al. 2014). During the white phase of *C. albicans* infection, the cells can infect the blood stream; however, the opaque mating-competent cells can infect skin, allowing for better transmission to another host and potential exit from stressful environmental conditions (Si et al. 2013). This correlation between mating and

infectious behavior underscores the importance of studying sexual reproduction, not only to grasp the risks involved in reproduction, but to better characterize infection transmission and pathogenic behavior.

Signal Cascade Pathways are the Mechanism of Sexual Reproduction Regulation.

Signal transduction pathways, a fundamental cellular process that transmits messages throughout cells, plays a key role in the regulation of sexual reproduction and the transition to pathogenesis (Fissore et al. 2019). The signal cascade pathway that triggers a cell to undergo sexual reproduction helps mitigate the potential risks of sex, such as DNA damage or even cell death (Hancock 2007). If the signal cascade is interrupted or recognizes a mistake at any point, the cell can stop meiosis and fix the mistake before proceeding (Hancock 2007). It is through signal cascades that cells regulate sexual reproduction and protect their genome to pass on the best chance of survival to their progeny. By understanding how cells use complex molecular pathways to reduce this risk, we might garner more insight into how to avoid DNA damage in human cells or decrease susceptibility to infection.

Signal Cascades Regulate Transitions in the Yeast Sexual Lifecycle.

Yeast are an optimal model for studying the mechanisms behind sexual reproduction because many critical reproduction genes have been identified through sequencing, their genomes are easily manipulated, and their cellular pathways are similar to those found in human cells (Mohammadi et al. 2015). Yeast sexual pathways, such as entrance into G₁, chromosomal division and transition between meiosis phases, are regulated by signal transduction pathways (Fissore et al. 2019). An integral part of that cycle, highly dependent on correct signal transduction, is mating. In yeast in

Saccharomycotina, mating occurs between two mating types, **a** and α (Wolfe et al. 2017). Only cells of the opposite mating types can form diploids under specific conditions to undergo meiosis and form spores (Fig. 1; Haber 2012). Spores are hardy and can survive in harsh environments, as well as populate a variety of locations (Hanson et al. 2014). Thus, sporulation is an exceptionally important aspect of the yeast lifecycle and has been conserved in many yeasts in Ascomycota (Hanson et al. 2017).

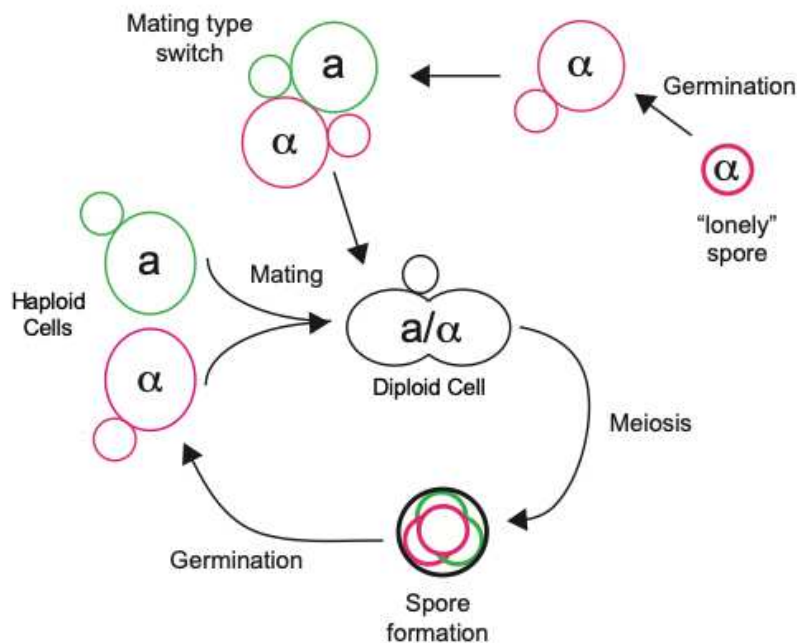


Figure 1. Haploid Cells Can Perform MTS to Reenter the Meiotic Lifecycle. Edited from Hanson et al. 2014 to show the yeast sexual lifecycles, which ultimately lead to spore production.

Pheromones Trigger and Mediate Sexual Reproductive Behavior in Yeast.

The whole yeast sexual lifecycle is regulated by a myriad of signals, both external and internal. Pheromone exposure is an important trigger that stimulates these sexual processes, particularly mating (Jones et al. 2011). Studies in *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and a variety of other yeast species have

demonstrated that mating is initiated by the binding of pheromones to the pheromone receptors on the cell membrane (Coria et al. 2005). This behavior is also present in our organism of interest, the methylotrophic budding yeast *Ogataea polymorpha*. Studies in *O. polymorpha* have shown that pheromone receptors are required for the progression of cell-type specific mating (Maekawa et al. 2014 and Yamamoto et al. 2017). Both **a** and α cells express pheromone receptors that bind pheromones on the outer cell membrane and interpolate the signal (Bardwell et al. 2004). In other yeasts such as *S. cerevisiae*, STE2 and STE3, expressed on **a** and α cells respectively, bind the pheromones from the opposite mating type and activate a pheromone signal cascade that ultimately leads to the upregulation of sexual reproduction specific genes (Madhani 2006). Depending on the type of pheromone present (either **a** or α), the cell can decide if a viable mating partner is available (Madhani et al. 2006).

The α pheromone, known as the α -factor, is encoded by the *Mating-Factor α* gene. In *S. cerevisiae*, the α gene contains multiple repeats of the same sequence which is eventually transcribed and translated into a polypeptide with four repeats of the α factor's peptide sequence (Madhani 2006). In *O. polymorpha*, the "precursor protein" containing four repeats is most likely post-translationally cleaved into four mature pheromones, containing thirteen amino acids each (Hanson Lab, unpublished). In *O. polymorpha*, some of the repeats within the α gene contain a base pair mutation that results in a substitution at the 8th amino acid, where a serine is replaced with an asparagine (Riley et al. 2016). This mutation may result in the production of two variants of the α factor, α and α -N, once the premature peptide is cleaved into independent proteins (Hanson Lab, unpublished). While both the wild-type and mutation variants are

present in the *O. polymorpha* MF- α gene, there has been no further investigation into whether both variants are produced by *O. polymorpha* and elicit the same responses (Riley et al. 2016). However, in *S. cerevisiae*, it is clear that MF α is released by α cells and binds to STE2 (the α pheromone receptor) on the surface of **a** cells, starting the cascade that leads to the upregulation of mating-specific genes (Madhani 2006). The loss of these same pheromone receptors in *O. polymorpha* inhibits the cells' ability to mate, allowing us to infer that the pheromone cascade is initiated by the same mechanism as *S. cerevisiae* (Maekawa et al. 2014).

Long Non-Coding RNAs are Important Signal Transduction Regulatory Elements.

In *O. polymorpha*, many protein-coding mating genes, homologous to those found in *S. cerevisiae*, have been identified (Hanson et al. 2017). However, there is relatively little information about the production of long non-coding (lnc)RNAs in *O. polymorpha* from mating-specific genes or their roles in sexual reproductive signal transduction. Studies have illustrated the regulatory capacities of lncRNAs in sexual reproduction in other related yeast species (van Werven et al. 2012). lncRNAs are RNAs that are greater than 200 nucleotides in length and do not encode for proteins (Yamashita et al. 2016); their functions vary depending on factors like sequence, secondary structure and enzymatic activity (Yotsukura et al. 2017). Many of these lncRNAs originate from the same or opposite orientation as a coding region and are even found within intergenic spaces (Yamashita et al. 2016). Their regulation of gene expression often comes from their position in relation to protein coding regions; while many lncRNAs act in *trans*, there are many lncRNAs that regulate the expression of protein coding regions through close proximity to genes themselves (Atkinson et al.

2011). Research in *Schizosaccharomyces pombe* has revealed that most, but not all lncRNAs in *S. pombe* are polyadenylated; however, they were most often found regulating the gene expression of coding regions and are not translated into proteins (Marguerat et al. 2012).

Known lncRNAs Play a Regulatory Role in Sexual Reproduction in Yeast.

An example of this lncRNA-mediated gene regulation of sexual reproduction occurs in *S. cerevisiae*. The lncRNA *IRT1* regulates the expression of *IME1*, a gametogenesis inducer (van Werven et al. 2012). *IRT1* is located very close to the promoter of *IME1* and regulates *IME1*'s gene expression in *cis* (van Werven et al. 2012). When *IRT1* is expressed, it recruits methyltransferases and deacetylases to *IME1*'s promoter, closing the chromatin structure and inhibiting the progression of gametogenesis through the suppression of *IME1* expression (van Werven et al. 2012). Thus, when *IRT1* is upregulated, *IME1* is downregulated, giving them an inverse relationship (Yamashita et al. 2016). The involvement of lncRNAs in the regulation of meiosis in *S. cerevisiae* led us to hypothesize that sexual reproduction in *O. polymorpha* may be mediated, in part, by lncRNAs.

***STE12* is Essential for Transmitting the Reproductive Signal Cascade Pathway.**

In *O. polymorpha*, a master regulator of the sexual reproduction signal transduction pathway is the transcription factor *STE12* (Hanson et al. 2017). *STE12* is an integral player in amplifying and mediating a yeast cell's response to reproductive cues (Madhani 2006). It is highly conserved across yeast in Saccharomycotina and is expressed in both haploid mating types as well as diploid mated cells, depending on the

species (Sorrells et al. 2015). Following pheromone exposure, *STE12* is phosphorylated as one of the last downstream targets of the signal transduction pathway and acts as a master regulator of the cell's response to pheromones (Madhani 2006 and Sorrells et al. 2015). This targeting frequently occurs through cooperative binding to transcription activators that recruit other transcription factors to promoter sites of mating and MTS-specific genes (Sorrells et al. 2015).

In *S. cerevisiae*, some of the most notable downstream targets of *STE12*-mediated pheromone response are *STE4* and *FUS3*, a G-protein involved in pheromone reception and a MAP kinase involved in mating, respectively (Madhani 2006). Furthermore, when knocked out, mating is unable to proceed, and diploid cells are unable to express numerous mating-specific genes (Sherwood et al. 2014). The variation in functionality of downstream genes targeted by *STE12* is indicative of the prolific impact the transcription factor has on the progression of mating. *STE12* is required during mating in numerous yeast species (Sherwood et al. 2014) and is necessary for mating in *O. polymorpha* (Hanson et al. 2017).

Mating-Type Switching Allows for Reentry into Meiosis without a Mating Partner.

In *O. polymorpha*, *STE12* is responsible for the regulation of a sexual reproduction phenomenon known as mating-type switching (MTS) (Hanson et al. 2017). As aforementioned, the ability to enter meiosis is an important part of the yeast lifecycle because it allows for the formation of spores (Herskowitz 1988). However, since yeast of the same mating type are unable to mate with each other, cells isolated in homogenous populations, or cells that germinate in so-called lonely environments are unable to enter the meiotic lifecycle and produce spores (Haber 2012).

The MTS process occurs automatically after one round of budding in *S. cerevisiae* (Lee et al. 2015) but requires additional nutritional and environmental cues to progress in other yeast like *O. polymorpha* (Hanson et al. 2014). Generally, in budding yeast, the initiation of this process triggers an “isolated” cell, whether in a homogenous population or physically alone, to begin mitosis (Herskowitz 1988). The resulting mother and daughter cells are the same mating type (either both **a** or α) and thus unable to mate with each other (Duina et al. 2014). To rectify this, the mother cell switches to the opposite mating-type in order to create a mating-partner for the daughter cell (Haber 2012). The mother and daughter can then mate and reproduce sexually (Haber 2012). Once the cells have had the chance to self-diploidize (Herskowitz 1988), they can then go through meiosis and eventually sporulate, ultimately allowing them to propagate and disperse (Hanson et al. 2017) (Fig. 1). In *S. cerevisiae*, mating and MTS are coordinated but are mechanistically independent of each other (Hanson et al. 2017).

***O. polymorpha* has a Unique MTS Mechanism that is Evolutionarily Related to MTS in *Saccharomycotina*.**

MTS is less well characterized in *O. polymorpha* than other types of yeast, specifically *S. cerevisiae* and *S. pombe*. *O. polymorpha*'s MTS mechanism is known as the Flip/Flop system and is homologous to the three-locus system found in *S. cerevisiae*, indicating a common evolutionary ancestor of the system (Fig. 2; Hanson et al. 2014).

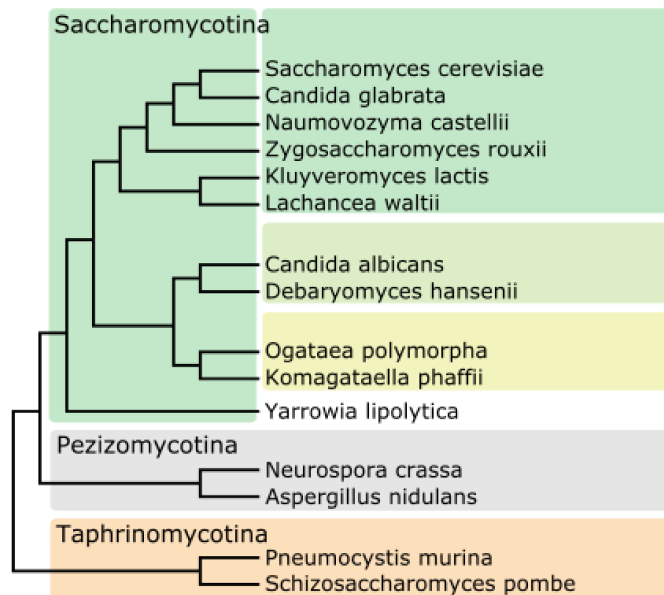


Figure 2. Yeast in Saccharomycotina have a Common MTS Locus Ancestor. Edited figure from Hanson et al. 2014 demonstrating the evolutionary relationship between different yeast species. In this figure, *K. phaffii* is *O. polymorpha*'s closest relative.

The *O. polymorpha* Mating-Type (MAT) Locus contains two distinct clusters of genes, MAT α and MAT \mathbf{a} (Fig. 3; Hanson et al. 2014). Proximity to the centromere on Chromosome 3 transcriptionally silences one of the loci, while the other locus is transcriptionally active, thus dictating the mating type of the cell (Hanson et al. 2014 and Maekawa et al. 2014). When nutritional (nitrogen starvation) signals initiate MTS, a double-stranded break (DSB) forms in the inverted repeat sequences that are adjacent to the MAT loci (Hanson et al. 2014). A flip flop occurs, and the MAT region reorients (Hanson et al. 2014). The opposite MAT locus is closer to the centromere and silenced while the other locus becomes transcriptionally active (Hanson et al. 2014). Thus, the mother cell transcribes the opposite mating genes from the daughter cell and becomes the opposite mating type, allowing the two to mate (Hanson et al. 2014).

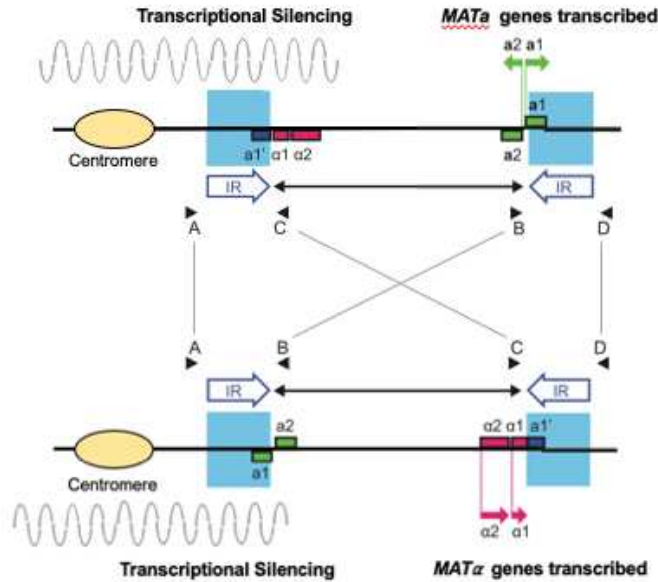


Figure 3. MTS Occurs Through the Formation of a Double-Stranded Break and the Inversion of the MAT Locus. Edited figure from Hanson et al. 2017 depicting the MAT locus surrounded by inverted repeats (IRs in blue). The MAT α locus is denoted in pink; the MATa is labelled in green. Primers Opol_MATa2/Opol_MATc1(A/C) and Opol_MATb1/Opol_MATd1 (B/D) amplify when the cell is in the **a** orientation (Table 2). Opol_MATa2/Opol_MATb2 (A/B) and Opol_MATc1/Opol_MATd1 (C/D) amplify when the cell is in the **α** orientation. When a dsDNA break forms at the IRs, the MAT locus reorients, and the opposite MAT locus is silenced.

STE12 Potentially Mediates Upstream Pheromone and Downstream Putative IncRNA Signals to Progress MTS.

STE12 plays an integral role in the MTS pathway: it is both necessary and sufficient for MTS in *O. polymorpha* and acts as a master regulator of the MTS signal transduction pathway (Hanson et al. 2017). Additionally, it has been shown that pheromone signaling is not necessary to kick start the MTS process (Yamamoto et al. 2017). Thus, we can infer that *STE12*'s role in the mediation of the MTS pathway must be triggered by nutritional or other environmental cues. However, researchers have not tested if pheromones inhibit the progression of MTS. MTS is an energetically expensive as well as a risky process for yeast cells, as the formation of a DSB can be lethal

(Hanson et al. 2014). We hypothesize, like Barsoum et al. 2011, that if pheromones from the opposite mating-type were present in a cell's environment, a cell would suppress MTS in favor of mating (Fig. 4). This suppression would help mitigate the risks of MTS, as well as divert the cell towards entering meiosis and producing hardy spores (Hanson et al. 2017). However, no research has been done on the potential inhibitory role pheromone exposure may have on the progression of MTS in *O. polymorpha*.

Additionally, we do not fully understand how *STE12* amplifies the signals leading to MTS; however, based on the regulation of sexual processes in other yeast by lncRNAs (Yamashita et al. 2016), we hypothesize that lncRNAs play a role in sexual reproduction in *O. polymorpha*. In this experiment, we investigated the upstream and downstream signals of the *STE12*-mediated MTS pathway to better characterize MTS from initiation to completion. Overall, we hypothesize that, upstream of *STE12* expression, exposure to pheromones from the opposite mating-type suppresses MTS in favor of mating. However, once *STE12*-mediated MTS has been initiated, we expect that lncRNAs are some of the important downstream effectors responsible for the transduction and completion of MTS (Fig. 4).

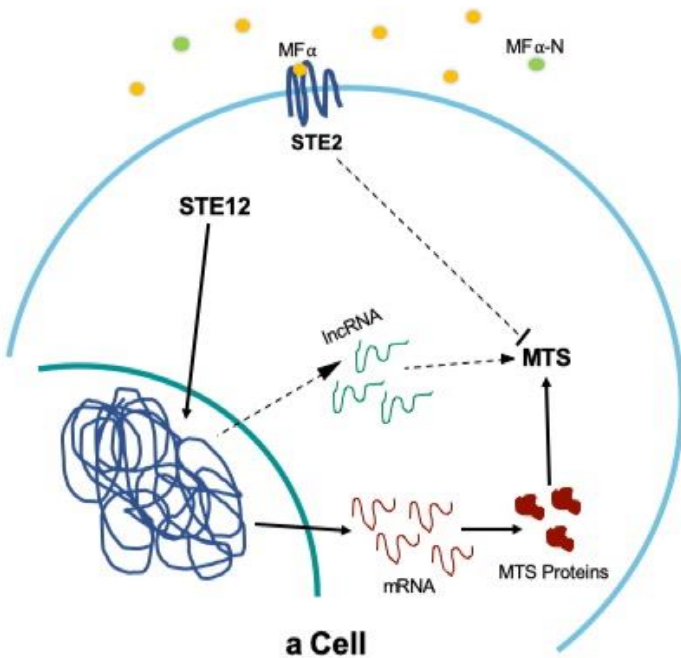


Figure 4. Hypothesized Regulatory Signal Cascade Pathway of MTS. MF α and MF α -N are labelled in yellow and green respectively. IncRNAs are denoted in green, while protein coding mRNA is labelled in red, along with the red proteins translated by the mRNA. The solid lines represent known components of the MTS signal cascade pathway, while the dashed lines represent all of the potential interactions we investigate in this research. All of these interactions occur within an **a cell**.

To perform this investigation, we relied on *STE12* overexpression to induce MTS in *O. polymorpha*. We qualitatively analyzed suppression of MTS after pheromone exposure, then designed a multiplexing semi-quantitative PCR technique to quantitatively measure MTS. We used an experimentally verified regression line to confirm the validity of our semi-qPCR technique. Additionally, we knocked out *STE2*, the receptor for α pheromones, to see if the suppression phenotype was rescued. To investigate *STE12*-dependent IncRNAs involved in the downstream MTS pathway, we performed RNA-seq and used a bioinformatics pipeline to find novel differentially expressed transcripts with no coding potential. Here we present the results of our

investigation into pheromone signaling and putative lncRNAs to help us better characterize MTS in *O. polymorpha*.

3. Materials and Methods

Yeast Strains for Qualitative and Quantitative Pheromone Exposure:

Strains were obtained from Hanson et al. 2017. Plasmids with *STE12* or empty vectors under the control of the *AOX* promoter (a methanol inducible promoter) were cloned into *E. coli* cells and extracted with a miniprep. The plasmids were transformed into the SHY199 strains. This procedure created the *pAOX-STE12* strains SHY202/1, SHY202/2 and SHY202/3, and the *pAOX* strains SHY199/2 and SHY199/3. All strains used were streaked out from freezer stocks, and then single colonies from the streak plates were inoculated in the appropriate media.

Table 1. All Strains Used in This Study.

Strain Name	Contents	Function	Source
SHY 199/2	MATa ade11 met6 pAOX-HPH ku80::ZEO	Negative Control for <i>STE12</i> Induction	Hanson et al. 2017
SHY 199/3	MATa ade11 met6 pAOX-HPH ku80::ZEO	Negative Control for <i>STE12</i> Induction	Hanson et al. 2017
SHY 202/1	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO	Induces <i>STE12</i> Overexpression	Hanson et al. 2017
SHY 202/2	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO	Induces <i>STE12</i> Overexpression	Hanson et al. 2017
SHY 202/3	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO	Induces <i>STE12</i> Overexpression	Hanson et al. 2017
NCYC 495	MATa ade11 met6	qPCR Validation	Kantcho Lahtchev, Bulgarian Academy of Sciences, Sofia, Bulgaria
NCYC 495s-1	MATalpha ade11 met6	qPCR Validation	Kantcho Lahtchev, Bulgarian Academy of

			Sciences, Sofia, Bulgaria
JOY04/1	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO ste2::nat	<i>STE2</i> Deletion	Created for this study
JOY04/2	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO ste2::nat	<i>STE2</i> Deletion	Created for this study
JOY04/3	MATa ade11 met6 pAOX-STE12-HPH ku80::ZEO ste2::nat	<i>STE2</i> Deletion	Created for this study

Creation of *STE2* Deleted Strain:

Strains from Hanson et al. 2017 were used as background strains for the deletion of *STE2*, a gene encoding for the α pheromone receptor (Table 1). Reference Table 2 for all primer sequences used to make the *STE2* knockout. OpSTE2p1/OpSTE2p3NAT created the *STE2* 5' flank. OpSTE2p4NAT/pSTEp6 created the 3' flank. PhusionTaq Polymerase was used for the flank PCR amplification, with 35 cycles, an annealing temperature of 55°C and an elongation time of 1 minute. Fusion PCR was used to stitch a Nourseothricin (NAT) resistance cassette into the flanking regions using nested primers OpSTE2p2/OpSTE2p5. The fusion PCR amplification used PhusionTaq Polymerase for 35 cycles with 55°C annealing temperature and 2 minutes of elongation.

The electrotransformation method was altered from Faber et al. 1994. Cells were incubated with TED (100mM Tris-HCl, 50mM EDTA, 25mM DTT [from a 1M Frozen Stock and added fresh to the TED each time] at pH = 8.0) for 15 minutes. They were washed with three times with STM (270mM Sucrose, 10mM Tris-HCl, 1mM MgCl₂ at pH = 8.0) and electrotransformed to introduce the NAT cassette into the cell to delete *STE2*. Double-stranded break repair was used by the cell to incorporate the NAT cassette into the genome, resulting in the *STE2* deletion. Electrotransformed cells were grown in a YPD-NAT liquid media and YPD-NAT plates to select for NAT resistant

colonies. A plasmid containing a NAT marker was transformed into cells as a positive control, while negative control cells received no DNA during electrotransformation.

Colony PCR checked for correct NAT orientation on the 5' side using

OpSTE2P1/NAT+103R and OpSTE2P6/NAT+516F on the 3' side (Fig. 9). NAT/HPH-379F/NAT+103R checked for the presence of the NAT cassette (Fig. 9).

OpSTE2F1/OpSTE2R1 checked for the deletion of *STE2* (not included in Fig. 9). DNA extraction and PCR analysis were performed on colonies with the deletion to confirm the deletion location was correct. PCR Primer information is listed in Table 2.

Table 2. Primers for All PCRs.

Primer Name	Primer Sequence (5'-3')	Purpose	Source
CDC28-11	AACACAACAACCGCGTAGTG	Multiplex qPCR	Yamamoto et al. 2017
CDC28-5	CTCCATCTTTGTGCTGTTGC	Multiplex qPCR	Yamamoto et al. 2017
HpoIMATa2	CCACTCATGGGAAATGATCCG	Check MAT Orientation	Hanson et al. 2014
HpoIMATb1	GAGTCATGGGGTCTGGTTTG	Check MAT Orientation	Hanson et al. 2014
HpoIMATb2	CTGCATGATATGACTACCAGCC	Check MAT Orientation	Hanson et al. 2014
HpoIMATc1	CTCAGATGATCCCACCACTAGG	Check MAT Orientation	Hanson et al. 2014
HpoIMATd1	CTGCGTCAGCTCAGGAATC	Check MAT Orientation	Hanson et al. 2014
NAT+103R	CGGTGTCGGTGGTGAAGG	NAT Internal Primers	Designed for this study
NAT+516F	CTGGACACCGCCCTGTAC	NAT Internal Primers	Designed for this study
NATHPH-379F	AGCTTGCCTCGTCCCCG	NAT Cassette Amplification Primers	Designed for this study
NAT +877R	TCGATTACAACAGGTGTTG	NAT Cassette Amplification Primers	Designed for this study
OpSTE2P1	CTGACGCAGAACGCAAGCTC	Amplify <i>STE2</i> Region	Designed for this study
OpSTE2P2	GGATAGTTCAGGAACACATCTG	Nested <i>STE2</i> Fusion Primers	Designed for this study
OpSTE2P3nat	CGGGGACGAGGCAAGCTCAA ACCTGAAGGAAGAGAGTATG	NAT/ <i>STE2</i> Fusion Primers	Designed for this study
OpSTE2P4nat	CAACACCTGTTGTAATCGAC GGTTTTTCGATCGTCCATAAC	NAT/ <i>STE2</i> Fusion Primers	Designed for this study

OpSTE2P5	CGTTCGAAAACAAGACGTCGG	Nested <i>STE2</i> Fusion Primers	Designed for this study
OpSTE2P6	CTAACTGGATTGCCTCCGG	Amplify <i>Ste2</i> Region	Designed for this study
OpSTE2F1	GTTTGTGGCCATTCGGACCAG	Check for <i>Ste2</i> Deletion	Designed for this study
OpSTE2R1	GTGTTGGAAGCTGAGCTGGAGG	Check for <i>Ste2</i> Deletion	Designed for this study

Yeast Growth Conditions for *STE12* Overexpression:

Yeast strains were streaked out from a -80°C freezer onto YPD plates and grown in a 37°C incubator for 48h. To induce MTS, single colonies were grown overnight in Mineral Media containing Glucose (MMG). The following day, the cells were grown to log phase twice in MMG, to an OD600 of 1.5 and 2.0, respectively. The cells grew for 12 hours and were diluted again in Mineral Media containing Methanol (MMM) to an OD600 of 0.02, after they had reached an OD600 of 2.0. For the timecourse, the cells were placed in the shaking incubator at 37°C for 8, 10, 12 and 14h. At each timepoint, a tube of *pAOX* and *pAOX-STE12* cells were removed from the incubator, spun down and stored for DNA extraction at -20°C.

Pheromone Exposure:

Strains were grown under the conditions required for *STE12* overexpression. To expose the cultures to pheromones, cells were diluted in 30mL of MMM and divided into 3 conical tubes (10mL of culture per tube). 10% DMSO, 50µg/mL of MFα or MFα-N were added to the tubes (so each strain was exposed to all three reagents individually). Cultures were grown for 10h, spun down and stored at -20°C for DNA extraction.

DNA Extraction

Phenol:chloroform:isoamyl alcohol and acid-washed glass beads were used to remove DNA from the yeast cells. The phenol:chloroform was mixed in a 1:1 ratio with

lysis buffer to lyse the cells. 100% and 70% EtOH were used to precipitate the DNA following phenol extraction.

PCR Amplification of the MAT Locus

HpoIMAT primers (Table 2) were used to amplify the MAT locus (Hanson et al. 2014). HpoIMATa2/HpoIMATb2 and HpoIMATd1/HpoIMATc1 amplify the α orientation. HpoIMATa2/HpoIMATc1 and HpoIMATd1/HpoIMATb1 amplify the **a** orientation. DreamTaq Polymerase Master Mix was used to amplify the DNA for 25 cycles with an annealing temperature of 57°C and an extension time of 2.5 minutes. A 1%(w/v) gel with 1X GelRed was run at 115V for 35 minutes. The gel was run in an electrophoresis chamber for 35 minutes at 115V. A 1kb+ (NEB) ladder was used for reference.

Semi-Quantitative PCR of the MAT Locus.

Multiplexing was used to quantify the amount of **a** to α cells in each sample. Only 1 primer combination was used to amplify each mating type. HpoIMATa2/HpoIMATb2 amplified MAT α and HpoIMATa2/HpoIMATc1 amplified the MAT**a** locus (Table 2). Using primers designed by Yamamoto et al. 2017, *CDC28*, a constitutively expressed gene, was amplified in the same samples in the same tube (Table 2). This resulted in the appearance of two bands in the same lane. We used PhusionTaq to perform a 20 cycle PCR amplification with a denaturation temperature of 95°C, annealing temperature of 57°C and extension time of 2.5 minutes. Samples were run in an electrophoresis chamber on a 1%w/v gel at 85V for 90 minutes. 1X GelRed was used to stain the DNA. The *CDC28* gene provided a standard to compare to the MAT locus bands, so the intensity of the MAT locus bands could be quantified. Images taken from an iBright were uploaded to ImageJ, which quantified the pixels of each MAT band and compared them

to the standard using the Analyze -> Gel function. The function showed density curves for each band, and the area under the curve (representing the density of the band) was quantified using the wand tool. The density of the MAT locus was then divided by the density of the CDC locus to create a standardized density for each band. The corrected MAT bands were compared to quantify the amount of MTS that had occurred.

Creating Cell Mixtures at Known Quantities.

NCYC495 and NCYC495s-1 (**a** and α strains respectively) were grown in a YPD media overnight (Table 1). Cell density was measured with a spectrophotometer and **a** and α cells were mixed together in known ratios based on their density. Genomic DNA was extracted, and the multiplexing semi-qPCR technique was used to quantify the amount of each cell type in the mixtures. Microsoft Excel was used to perform a log transformation on data points obtained from the cultures, as well as create our regression line. The data points were then transferred to RStudio, which gave us an R^2 value and confidence interval.

RNA Extraction:

The protocol developed by Hanson et al. 2017 was used. All workbenches and tools were sterilized by RNaseZap and lab work was performed under a fumehood to prevent contamination by RNases. Phenol:Chloroform:Isoamyl alcohol and chloroform were used to extract the RNA from 10mL of fresh overnight cultures. A nanodrop was used to quickly test concentration, then a Qubit was used to accurately test the quantity and quality following a DNase treatment. Samples were stored at -80°C while MTS was confirmed, then shipped to the University of Colorado Anschutz Medical Campus Genomics and Microarray Core Facility.

Library Prep and Sequencing:

Library Prep and Sequencing were done by lab technicians at using an Illumina NovaSeq at the University of Colorado Anschutz Medical Campus Genomics and Microarray Core Facility. A ribodepletion kit (normalized for *S. cerevisiae*) was used to remove all of the ribosomal RNA from the samples. Adaptors were ligated to the ends of the reads in order to amplify the sequences.

Sequencing Analysis Protocol:

The workflow for our RNA-seq analysis was developed from a combination of techniques listed in Hanson et al. 2017, Sun et al. 2019 and Pertea et al. 2016. Analysis was performed on the online University of Pennsylvania's Galaxy server using the programs provided by the server with the exception of the coding potential analysis.

Quality Control:

The quality control pipeline tools were designed by Babraham Bioinformatics (Wingett et al. 2018). Quality of the raw reads were checked with FastQC (Version 0.72) and adaptors were trimmed off with Trim Galore! (Version 0.4.3.1). For Trim Galore!, the library was set to paired-end and the Illumina Universal adaptor was trimmed. FastQC was used to check the quality again before alignment.

Transcript Assembly:

The John Hopkins University Center for Computational Biology created all of the programs used for our transcript assembly process (Pertea et al. 2016). HISAT2 (Version 2.1.0 + galaxy4) was used to align the trimmed reads with the *O. polymorpha* genome FASTA file provided by Hanson et. al 2017. The reads were set to paired end and the specified strand information was set as Forward. StringTie (Version 1.3.4) was

used to assemble the reads into full length transcripts. The specified strand information was set to Forward and a reference GFF genome provided by Hanson et al. 2017 was used to guide assembly. All of the samples were merged together using StringTie Merge (Version 1.3.4). This step accounted for samples that contained partially covered transcripts and created fully covered transcripts by merging the transcripts together. The Hanson et al. 2017 reference GFF genome was used to guide the merge. GFFCompare (Version 0.9.8) was run on the merged file to create statistics on transcript assembly and novel transcripts. Both the FASTA and GFF genome files were used to analyze the merged file.

Differential Expression Analysis:

htseq-count (Version 0.9.1) (Anders et al. 2015) was used to count the number of reads coming from each transcript. HISAT2 provided the reads while the StringTie Merge File was used as a reference transcriptome. The strandedness was set to reverse and the ID Attribute was changed to transcript_id. DESeq2 (Version 2.11.40.2) (Love et al. 2014) was used to compare the *pAOX-STE12* strains to the *pAOX* strains to find transcripts upregulated by *STE12*. Factor level #1 was set to *STE12* (strains containing the *pAOX-STE12* vector were selected) while factor level #2 was set to vector (strains containing the empty vector were selected). Files did not have a header. The DESeq2 results file were joined with a file containing transcript information and *S. cerevisiae* homologous genes. The StringTie Merge file was exported to RStudio to remove all data but gene_id, transcript_id and gene_name. Join (Version 1.1.1) was set to -a1 -a2 to keep all transcript information, even if no gene name was present for that

transcript. This prevented the removal of non-coding RNAs and novel transcripts. The data was sorted to show the most upregulated genes.

Long Non-Coding RNA Identification Protocol:

Using the sorted DESeq2 file, novel transcripts were identified; novel transcripts did not have annotated gene names and did not have a *S. cerevisiae* homologue. The *S. cerevisiae* homologue dataset was provided by Hanson et al. 2017. For this experiment, only novel transcripts with a log-fold change greater than or equal to 1.4 were examined (*BEM1*, a gene involved in schmoos formation, represented our least upregulated mating-related gene and therefore marked our lowest limit). The location of the novel transcripts was manually found in the StringTie Merge File, and the sequence of each transcript was identified using Artemis. The sequences were entered into the Coding Potential Calculator created by the Center for Bioinformatics (Kong et al. 2007). Transcripts with coding potentials less than 1 were identified as non-coding.

Prediction of lncRNA Targets in cis:

Transcripts with a coding potential less than one were identified and annotated in Artemis. The neighboring genes were found in the sorted DESeq2 file and their function was identified using *S. cerevisiae* homologue information.

4. Results

***STE12* Overexpression Induces Robust MTS After 10 Hours.**

In order to observe the impact of pheromone exposure on the progression of MTS in *O. polymorpha*, we overexpressed *STE12* as a mechanism of MTS induction. *STE12* overexpression induces MTS, regardless of environmental conditions (Hanson

et al. 2017); while we could have used nitrogen starvation, other metabolic processes would have been impacted, making it difficult to draw conclusions (Hanson et al. 2017). The timing of cell growth required to induce MTS in methanol was unknown. In order to precisely time the experiment, we performed a growth timecourse following *STE12* induction. We used a strain previously created by Hanson et al. 2017 that contained a *STE12* overexpression vector under the control of a methanol inducible promoter (*pAOX-STE12*) (Table 1). All cells were in the **a** orientation before MTS. When *pAOX-STE12* cells were grown in methanol, they underwent MTS. Mating types and MTS are shown through PCR amplification of the MAT regions in *O. polymorpha*. Four primer combinations amplify either the MAT \mathbf{a} or MAT α orientations, indicating if a population of yeast is **a**, α or heterogeneous (Table 2 and Fig. 5). Populations that perform MTS are heterogeneous for both **a** and α orientations. *pAOX-STE12* strains saw minimal MTS after 8 hours and robust MTS after 10 hours (Fig. 5). As expected, no MTS was observed in the vector-only negative control (Fig. 5). This data demonstrated that 10 hours of growth in methanol was sufficient to induce MTS in *O. polymorpha* and would be an appropriate growth timepoint following pheromone exposure.

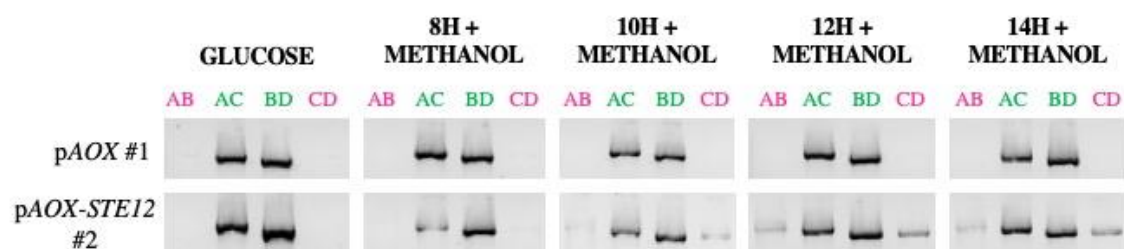


Figure 5. Robust MTS is Seen After 10H after *STE12* Induction. Cells containing empty (*pAOX*) and *STE12* overexpression (*pAOX-STE12*) vectors were grown in glucose, then transferred to methanol. Culture was removed and saved after 8, 10, 12 and 14 hours. A DNA extraction and PCR were performed to visualize the orientation of the MAT locus. MAT PCR Primer information can be found in Table 2. The green

primers at the top correspond to the **a** orientation; the pink correspond to the α orientation.

Overexpression of *STE12* During α Pheromone Exposure Results in Qualitative Suppression of MTS.

In order see if MTS is suppressed when cells detect pheromones of the opposite mating-type, we overexpressed *STE12* while exposing cells to pheromones. We used *pAOX-STE12* and vector-only strains, along with two synthetic pheromone variants, MF α and MF α -N. We used only α pheromones in our experiments; the mature **a** pheromone undergoes a series of posttranslational modifications, including the addition of a lipid to the end of the protein, making the synthesis of artificial **a** pheromones difficult (Madhani 2006). In order to expose cells to pheromones from the opposite mating-type, we only used **a** cells in combination with α pheromones. As aforementioned, the MF α gene encodes a premature peptide with multiple repeats that are eventually cleaved into mature pheromones (Madhani 2006). The two variants of the α pheromone, MF α and MF α -N vary by a single substitution of a serine for an asparagine at the 8th amino acid (Riley et al. 2016). We believe that the majority of the mature peptides produced are MF α , while a smaller number are MF α -N (Hanson lab, unpublished). In order to determine what variant elicits a stronger response during MTS, we used both variants individually for our experiments.

pAOX-STE12 and *pAOX* were exposed to 50 μ g/mL of MF α , 50 μ g/mL of MF α -N or 10% (v/v) DMSO (the control that should not have suppressed MTS) for 10 hours after *STE12* induction. The DNA from the cells was extracted and analyzed using PCR to look for effects of pheromone exposure on MTS. MTS was completely suppressed in

cells that received MF α treatment compared to cells that received the DMSO treatment, indicating that exposure to pheromones from the opposite mating type can suppress MTS in a cells (Fig. 6). Cells that received the MF α -N treatment also suppressed MTS, although a very faint MAT α band appeared in the gel even after exposure to the MF α -N pheromone (Fig. 6). Although quantification of the amount of MTS is necessary, this preliminary data suggests that exposure to pheromones produced by the opposite mating type can suppress MTS in *STE12* overexpressing cells. Furthermore, qualitative partial suppression of MTS by MF α -N could indicate that MF α -N elicits a smaller response in the MTS pathway. Overall, this visual analysis of MTS suppression implicated pheromones as an upstream regulator of MTS in *O. polymorpha* and led us to continue our analysis of MTS suppression in other replicates (Fig. 6).

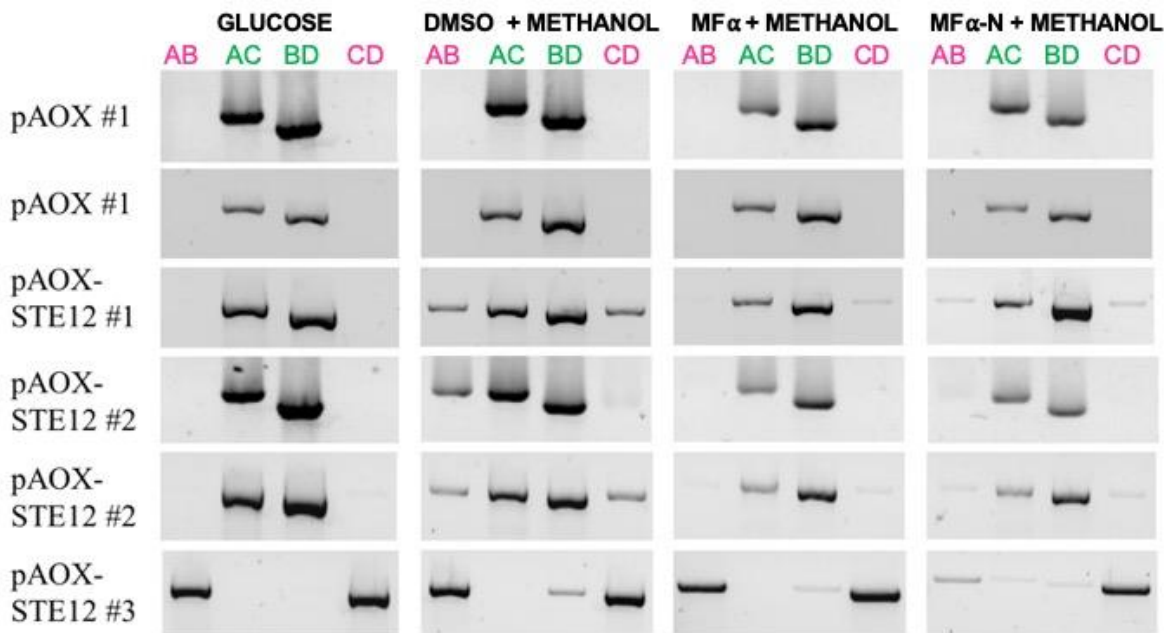


Figure 6. MTS is Suppressed by Exposure to Pheromones from the Opposite Mating Type. Cells containing empty (*pAOX*) and *STE12* overexpression (*pAOX-STE12*) vectors were grown in glucose, then transferred to methanol. At the beginning of the methanol induction, cells were exposed to DMSO, Mating Factor (MF) α or MF α -

N. A DNA extraction and PCR were performed to visualize the orientation of the MAT locus. MAT PCR Primer information can be found in Table 2. The green primers at the top correspond to the α orientation; the pink correspond to the α orientation. There was 1 biological replicate of *pAOX* and 3 biological replicates of *pAOX-STE12* (replicate #3 switched to α before the experiment). The two *pAOX-STE12* #2s are technical replicates.

MF α and MF α -N Exposure Does Not Suppress MTS in α Cells.

One of our *pAOX-STE12* biological replicates, streaked from a freezer stock and then inoculated from a single colony, had switched to the α Mating-Type prior to our pheromone exposure experiment. We exposed this α cell replicate (*pAOX-STE12* #3) to DMSO, MF- α and MF- α -N pheromones: the replicate was grown with the other replicates for 10 hours after *STE12* induction. Qualitative analysis using PCR revealed that, as expected, α pheromones do not effectively repress MTS in α cells (Fig. 6). This is most likely because α pheromones are derived from the same mating-type as α cells, and therefore do not indicate to the cells that viable mating partners are present. Without a mating partner present, the cell is forced to switch mating types. Although one-mating type band was missing, it is clear that *pAOX-STE12* #3 performed MTS, despite the presence of pheromones.

Semi-Quantitative PCR for the MAT Locus is Optimized Using Multiplexing.

Using previously induced samples, we performed various PCRs to optimize the conditions for semi-quantitative PCR. The product size necessary to detect MTS (around 3kb) is too large for quantitative PCR, so we had to develop a semi-quantitative method. Using primers designed by Yamamoto et al. 2017 for the amplification of the *CDC28* gene, we tested different concentrations of the primer to see what resulted in

the best visualization of the multiplexed PCR. The *CDC28* product acted as a control because it is constitutively expressed and would therefore be present in all of the lanes, regardless of mating-type. Both *MAT* and *CDC28* primer combinations were added to the same PCR tube. We found that using 0.2 μ M rather than 0.4 μ M in the final concentration of Master Mix of the *CDC28* primers resulted in the most consistent *CDC28* product between all 4 lanes (Fig. 7). Furthermore, the less amplified *MAT* locus combinations (lanes 1 and 4, which represent the α combination) were more visible with a lower *CDC28* primer combination. Since the *CDC28* product is smaller, it is more likely to be amplified by the PhusionTaq Polymerase. By decreasing the *CDC28* primer concentration, the faint *MAT* bands were more apparent, and the *CDC28* bands were approximately the same density in each lane. The same density made quantification across lanes easier because the control was more consistent. ImageJ was used for the quantification; boxes were drawn around the bands and the pixels were measured. Once ImageJ produced pixel data, MTS was quantified by measuring the pixels of the *MAT* product over the *CDC28* product. Approximately 12% of the culture switched in the 0.2 μ M *CDC* sample.

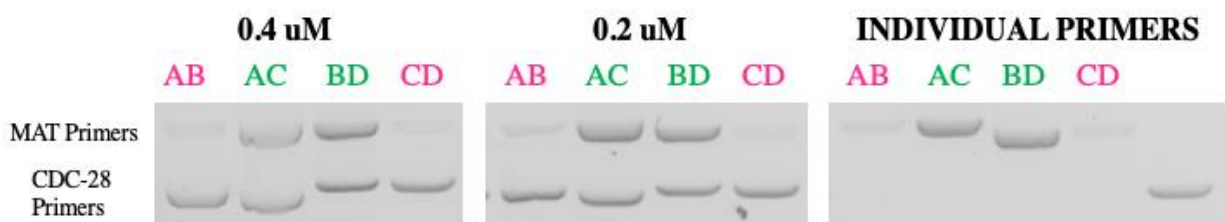


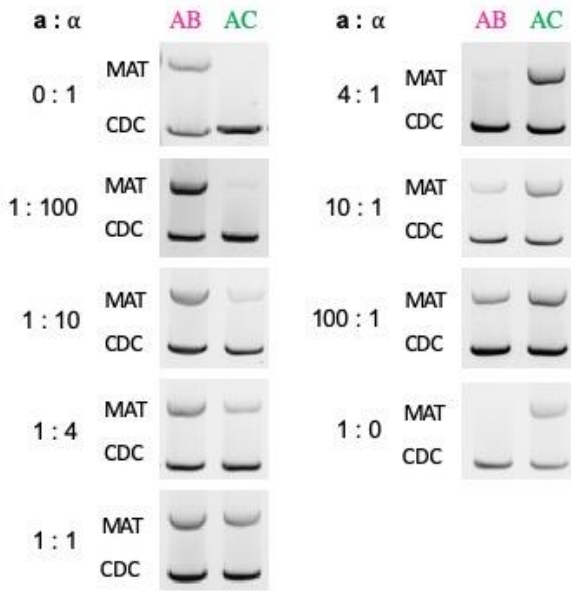
Figure 7. Multiplexing the *MAT* Locus with Constitutively Expressed Genes Allows for MTS Quantification. Cells containing *STE12* overexpression (*pAOX-STE12*) vectors were grown in glucose, then transferred to methanol in order to induce MTS. A DNA extraction and Multiplexing PCR was performed to visualize the orientation of the *MAT* locus. The *MAT* locus products are around 2.5kb long and appear as the top band. The *CDC28* product is 1kb long and appear as the lower band. *MAT* and *CDC28* Primer information can be found in Table 2. The different *CDC28* primer concentrations

are listed at the top. The green primers at the top correspond to the *MATa* orientation; the pink correspond to the *MAT α* orientation. ImageJ was used to quantify the band pixels in each lane.

A semi-qPCR MTS Regression Line was Created with Known Ratios of α :a Cells

In order to use semi-qPCR to quantify MTS, known quantities of **a** and α cells were mixed in specific ratios. Their DNA was extracted, and our semi-qPCR method was used to amplify the DNA (Fig 8a). ImageJ was then used to quantify the amount of each cell type in the mixture. The quantification from each cell mixture was log-transformed, and the ratio of expected **a**: α cells was plotted (x-axis) compared to the actual **a**: α values (y-axis). A linear regression line, $y = 0.7082x - 0.0768$, was obtained. The adjusted R^2 value was 0.9913, meaning that the line accounts for 99% of the variation between the actual cell ratios and the expected ratios (Fig. 8b). We performed a confidence interval analysis on the data and found, with 95% certainty, that the true slope of the regression line is between (0.6004, 0.8160). This small interval, along with the high adjusted R^2 value, convinced us that this regression line is valid for future use. The development of this equation was especially important for future quantification of MTS. ImageJ analysis of band density allows us calculate the ratio of **a**: α cells, then enter the data into the regression line to accurately estimate how many of each cell type exists, and how much MTS occurred under a variety of conditions.

a)



b)

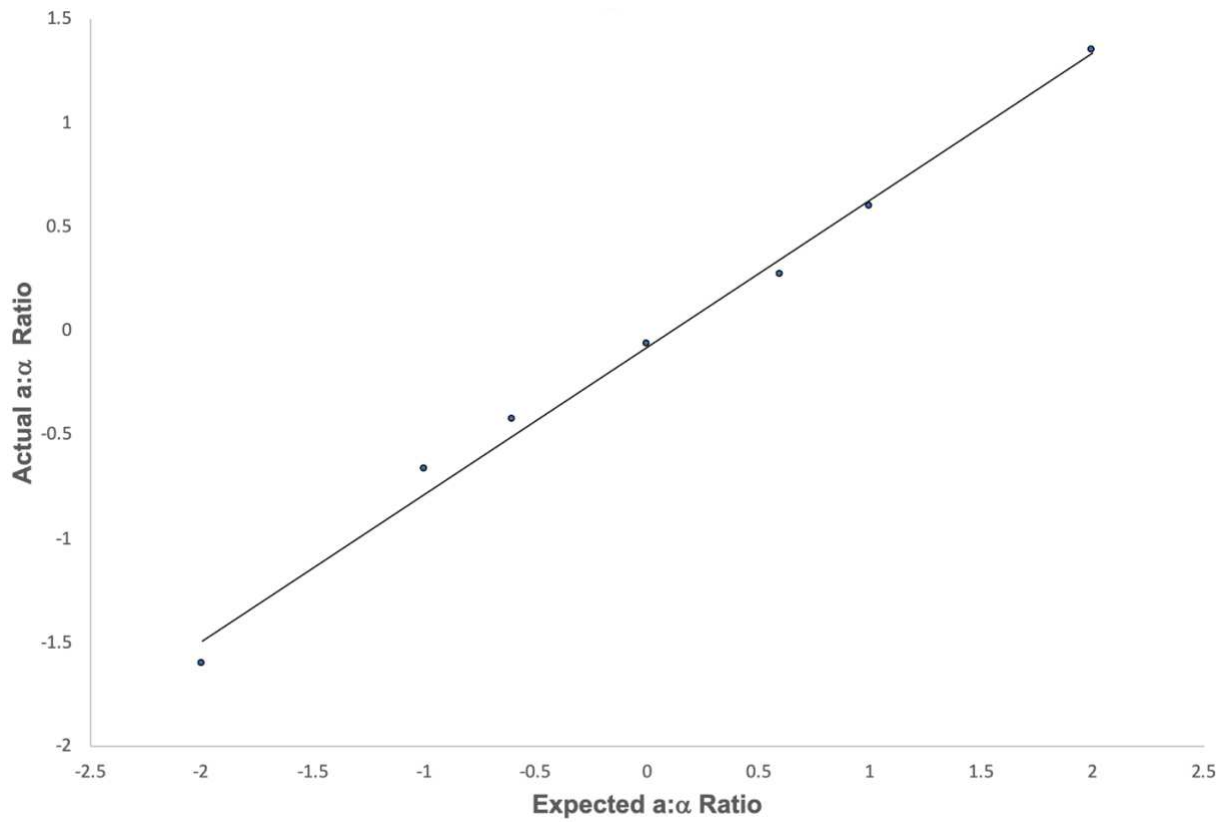


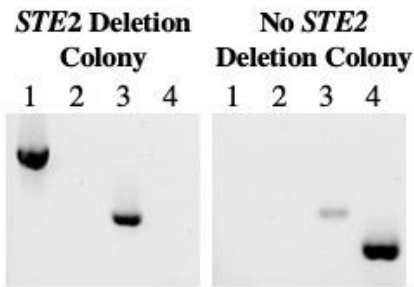
Figure 8. a) α and a Cells Were Mixed in Known Ratios to Create a Regression Line to Quantify MTS. NCYC495 (an a strain) and NCYC495s-1 (an α strain), created by Hanson et al. 2017, were grown in homogeneous cultures in YPD Broth. Their density was measured using a spectrophotometer and they were mixed in known ratios. DNA was extracted and multiplexing qPCR was used to visualize samples. *MAT* and *CDC28* bands are denoted on the left of each sample. Primer information is listed in Table 2. ImageJ was used to quantify the density of each *MAT* band compared to the *CDC28* controls. **b)** a and alpha cells were mixed in known ratios; their DNA was extracted and the amount of DNA from each cell type in the cultures were quantified using our multiplexing protocol. The actual values were plotted against the expected values and a regression line was produced: $y = 0.7082x - 0.0768$.

***STE2* Was Deleted to Create Control Strains Missing the α Pheromone Receptor.**

To confirm that pheromone signaling was responsible for the suppression phenotype we qualitatively observed, we deleted *STE2*. By deleting *STE2*, the receptor for the α pheromone, we removed the cell's ability to bind pheromones and trigger the suppression phenotype. We expected to see the suppression phenotype rescued if *STE2* was deleted. We used Fusion PCR to create a Nourseothricin (NAT) resistance cassette with ends that matched the *STE2* flanking regions. The cassette was electrotransformed into competent yeast strains and the colonies containing the deletion were selected for on Nourseothricin plates. PCR was used to visualize the genotypes of the colonies, and strains containing the correct deletion were saved (Figure 9a). Lanes 1 and 2 checked for the correct orientation of the NAT cassette in *STE2*. The Lane 2 product did not appear in either sample, indicating that something prevented the primers from amplifying that specific region; no conclusions could be drawn from that lane. In Lane 3, NAT/HPH-379F/NAT+103R amplified the NAT cassette, indicating that both types of colonies had gained NAT resistance, but not necessarily in the correct location. However, in Lane 4, the *STE2* internal primers OpSTE2F1/OpSTE2R1 did not appear in the deletion strains, indicating that *STE2* had been knocked out. Figure 9b shows the

locations of each primer on the *STE2-NAT* cassette. Three biological replicates containing deletions were obtained from this electrotransformation. In the future, we will use semi-qPCR on this strain, in tandem with wild-type α and α cells, to quantify the ability to switch in the presence of α pheromones without the α pheromone receptor.

a)



b)

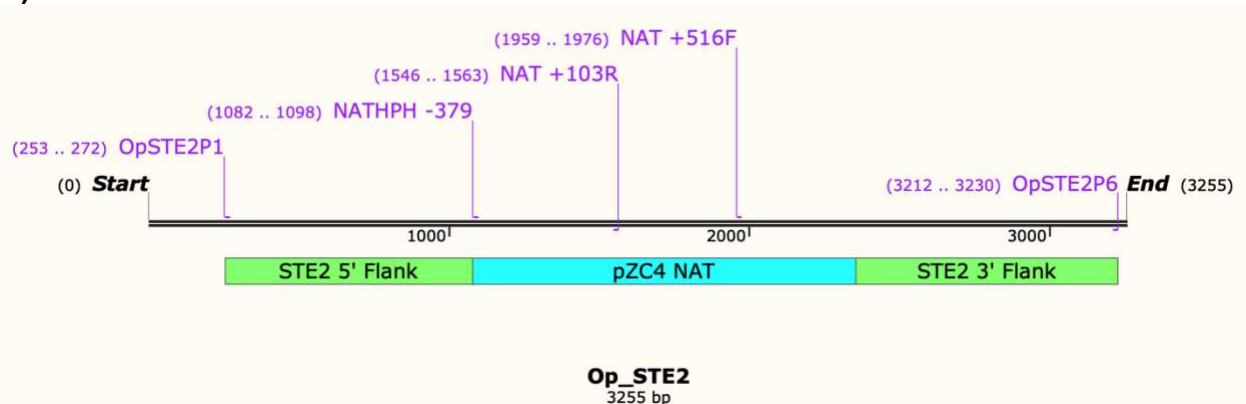


Figure 9. *STE2* Deleted Strains Replaced *STE2* with a *NAT* Cassette. a) Strains obtained from Hanson et al. 2017 were electrotransformed with a *NAT* cassette. Cells were grown on YPD-*NAT* plates to select for *STE2* deleted strains. Individual colonies were streaked out and tested for deletions. b) In Panel A, Lane 1 is OpSTE2P1/*NAT*+103R. Lane 2 is OpSTE2P6/*NATHPH*+516F. Lane 3 is *NATHPH*-379F/*NAT*+103R. Lane 4 primers are internal to *STE2* and not included in Panel B. 3 *STE2*-deleted strains were obtained but only 1 is shown to eliminate redundancy (left). These data are representative of 8 samples.

Cells overexpressing *STE12* performed MTS.

In order to examine the downstream regulatory elements of *STE12*-mediated MTS, we looked for putative lncRNAs that were regulated by *STE12* and might play a role in MTS. As aforementioned, the presence of lncRNAs in meiosis transduction pathways in other yeast species led us to hypothesize that lncRNAs might be a method of MTS regulation. To look for *STE12* regulated lncRNAs, we used methanol to induce MTS in α strains. Dr. Hanson's MB350 Laboratory and Genomics Block 6 2017/2018 class and Hanson et al. 2017 had previously done poly-A seqs for novel *STE12* upregulated mRNA transcripts, so our analysis solely focused on lncRNA discovery. To confirm that *STE12* had been overexpressed before our samples were sequenced, we looked for the MTS phenotype using PCR and gel electrophoresis. The presence of the MTS phenotype is very convincing evidence that *STE12* has been upregulated, because *STE12* is sufficient for inducing MTS even in the absence of nutritional cues (Hanson et al. 2017). While only 2 α bands were present in the pre-induction and vector cultures, 4 bands appeared in the cells overexpressing *STE12*, confirming the presence of **a** cells and that MTS had occurred. Thus, we were sufficiently convinced that *STE12* had been upregulated and that the samples were valid for RNA-sequencing (Fig. 10).

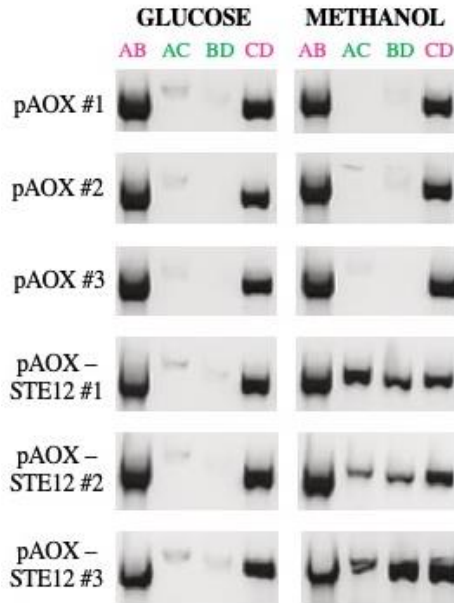


Fig. 10. α Cells Overexpressing *STE12* Performed MTS. Cells containing empty (*pAOX*) and *STE12* overexpression (*pAOX-STE12*) vectors were grown in glucose, then transferred to methanol. A DNA extraction and PCR was performed to visualize the orientation of the MAT locus. MAT PCR Primer information can be found in Table 2. The green primers at the top correspond to the **a** orientation; the pink correspond to the α orientation. There were 3 biological replicates of both *pAOX* and *pAOX-STE12*.

Differentially Expressed Transcripts Were Identified Using a Modified Galaxy

RNA-Seq Bioinformatics Pipeline.

The University of Colorado Anschutz Medical Campus Genomics and Microarray Core Facility performed the ribodepletion RNA library prep and sequencing for our lab. We performed the rest of the bioinformatic analysis on our data. We trimmed the raw reads and aligned them to the *O. polymorpha* genome, with approximately 70% of the reads aligning (Fig. 11).

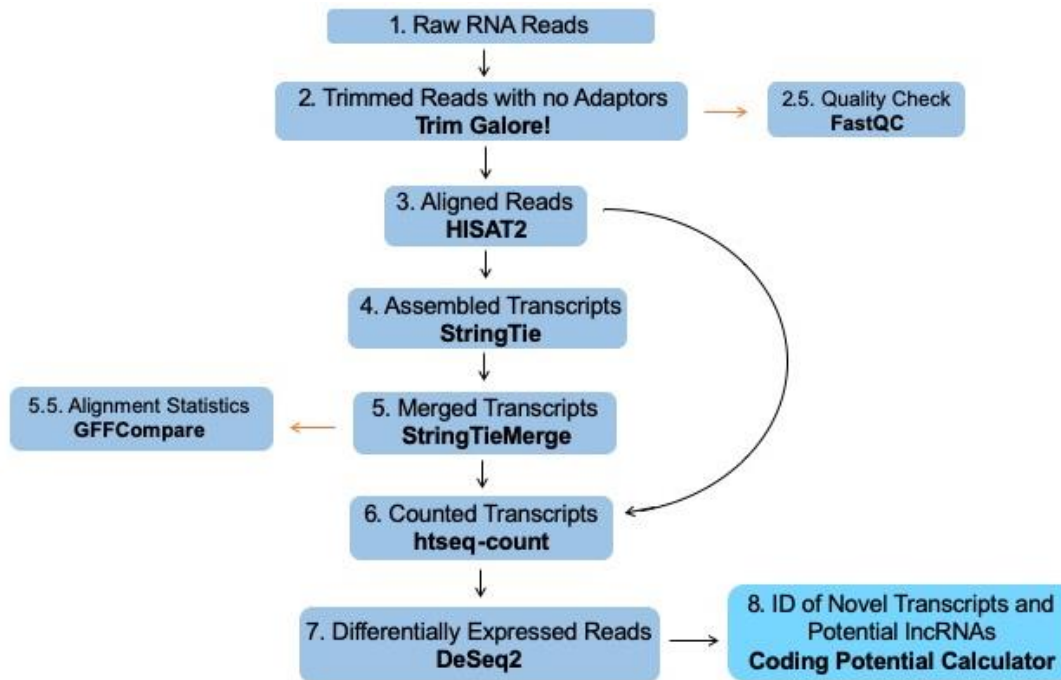


Figure 11. Workflow of Protocol Developed to Identify Long Non-Coding RNAs (lncRNAs). Raw RNA-Seq data was input into Galaxy and analyzed using programs on the Galaxy server. Programs are listed in bold, while outputs are listed in non-bold typeface. The Coding Potential Calculator was developed by the Center for Bioinformatics and is not found on the Galaxy server.

The aligned reads were then assembled into full length transcripts. Alignment statistical analysis (using GFFCompare) told us that our data had high sensitivity and average precision. With a high sensitivity of 92.7%, we are confident that almost all previously known annotated features were found in our samples. However, the average precision was 72.0%, indicating that many features in our dataset aligned with the genome but were unannotated. We believed that we were most likely to find non-coding RNAs within that group of unannotated features. Further information regarding our bioinformatics pipeline, modified from Pertea et al. 2016, can be found in Materials and Methods.

Using htseq-count, the expression of each assembled transcript was counted. DESeq2 was used to compare the overexpression cells to the normal cells to see how gene expression was impacted by the overexpression of *STE12*. The transcripts were sorted based on their log₂ fold change and the top 25 upregulated transcripts were compiled (Table 3). *STE12* had a log₂ fold change of 5, or 32 times the baseline expression level, which confirmed that *STE12* was overexpressed. Additionally, many of the genes with the highest upregulation play a role in the MTS and mating pathway, confirming both that MTS had occurred and that the upregulation of *STE12* was sufficient for inducing the upregulation of other mating related genes.

Hanson et. al 2017 previously performed a poly-A RNA-seq on MATa *STE12* overexpressing strains; we analyzed this data with our differential expression pipeline in conjunction with poly-adenylated RNA-seq data from Dr. Hanson's MB350 Block 6 2017-2018 class on MATα *STE12* overexpression strains. The expression levels for the top 25 upregulated genes from our dataset was compared to across the three datasets (ribo-depletion MATα, poly-A MATα and poly-A MATa) (Table 3). The top 25 genes all had similar levels of upregulation, confirming that *STE12* was upregulated in all datasets and that our data was consistent with past data sets. Table 3 highlights *STE12* and a novel transcript, MSTRG.6170.1, which was highly upregulated in all 3 data sets.

Table 3. *STE12* Upregulates Mating Specific Genes.¹

Gene Name	Function	MAT α total RNA AOX- <i>STE12</i> vs. AOX	MAT α mRNA AOX- <i>STE12</i> vs. AOX	MAT α mRNA AOX- <i>STE12</i> vs. AOX
MFalpha1	alpha-factor pheromone	8.53	4.03	8.04
BAR1	barrier protease that cleaves alpha-factor	6.09	6.75	6.94
FUS3	MAP kinase involved in mating	5.84	5.71	5.57
MFa	a-pheromone	5.59	5.56	5.28
GPA1	"G protein coupled to pheromone receptors, alpha subunit"	5.52	6.23	7.47
STE3	receptor for a-factor	5.36	5.74	5.49
STE12	"transcription factor, pheromone response pathway"	5.00	5.88	4.61
CDS:62574_63419	unknown	5.00	4.36	5.57
OPOL_16726	prenylcysteine lyase	4.62	4.43	4.90
FAR1	cell cycle arrest in response to pheromone	4.27	4.30	4.24
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
Ty5-like_s4_pseudo_retrotransposon	unknown	4.14	4.25	4.02
PGC1	Phosphatidyl glycerol phospholipase C	3.99	4.05	5.62
PGC1	Phosphatidyl glycerol phospholipase C	3.99	4.05	5.62
SST2	"GAP for Gpa1, regulates desensitization to alpha factor"	3.86	4.18	5.10
ABZ2	aminodeoxychorismate lyase	3.61	3.62	3.84
OPOL_93463+FUS2	unknown	3.57	3.19	4.53
YPL088W	aryl alcohol dehydrogenase; Tpa5 LTR insertion at 5' end	3.52	4.92	7.56
MSTRG.6170.1	unknown	3.51	3.69	5.12

While looking for novel transcripts, we assigned the cutoff of minimum upregulation to be a log₂ fold change of 1.40. *BEM1*, the gene responsible for the formation of buds and shmoos, is upregulated in our data set by a log₂ fold change of 1.39. Since this was the lowest upregulated mating related gene, it marked the cutoff in our search for upregulated non-coding RNAs involved in the MTS pathway. We identified 19 novel transcripts that were upregulated in our data set above a log₂ fold change of 1.40.

Five Novel Transcripts Upregulated by *STE12* Have Non-Coding Potential

Using Galaxy's MultiJoin tool, all of the htseq-count files were joined together so we could look at transcript counts rather than expression levels. Although many of the

¹ Our ribodepletion sample (1st column) compared to upregulation in mRNA data sets (2nd and 3rd columns). *STE12* is highlighted in red.

transcripts were upregulated, they had very low counts for each transcript (below 50 counts per read). The low counts demonstrated that the observed upregulation of those transcripts was artificially inflated. The low number of hits indicated that the low count transcripts were not likely to be important in the MTS pathway. Novel transcripts with less than 50 transcript counts were discarded from the data set, leaving behind 7 novel transcripts (Table 4). Using Artemis, the location of each of the transcripts was found on the correct scaffold and the nucleotide sequence from the correct strand was exported. The sequences were entered into the Coding Potential Calculator (CPC) developed by the Center for Bioinformatics (Kong et al. 2007). The program returned a coding potential score and the justification behind the score. A score lower than 0 was considered to be non-coding.

Out of the 7 novel transcripts with counts over 50, 5 were non-coding (Table 4). As our analysis focused on lncRNA upregulation, we disregarded two coding transcripts. All 5 transcripts appear in the other data sets as well, indicating that they are all polyadenylated. The most upregulated of the 5 genes, MSTRG.6170.1, is highly upregulated in the 2 other data sets as well. Furthermore, it is located next the *S. cerevisiae* homologue *AXL1*. *AXL1* is an endoprotease that makes one of two necessary N-terminal cleavages to create a mature a mating factor (Chen et al. 1997). *AXL1* works in tandem with *STE24p* to process the pheromone N-terminus and prepare the pheromone for export (Chen et al. 1997). The proximity to this mating-affiliated gene indicates that MSTRG.6170.1 might be involved in the *cis* regulation of the protease (Yamashita et al. 2016). However, for all 5 lncRNAs and especially MSTRG.6170.1, further bioinformatic analysis is required.

Table 4. The Total RNA-Seq Dataset Contains Seven Significant Novel Transcripts²

Transcript Name	Log ₂ (Fold Change)	Length	Coding Potential Score	Coding or Non-Coding
MSTRG.6170	3.51	3096	-1.03047	Non-Coding
MSTRG.74	2.77	17772	6.8913	Coding
MSTRG.1933	2.2	4869	-1.01701	Non-Coding
MSTRG.2246	1.97	12190	3.76732	Coding
MSTRG.2241	1.73	4250	-0.274634	Weak Non-Coding
MSTRG.1617	1.57	1917	-1.04476	Non-Coding
MSTRG.1283	1.54	952066	-32.5407	Non-Coding

5. Discussion

Over the course of this project, our central aim was to characterize the upstream and downstream pathways involved in *STE12*-mediated MTS in *O. polymorpha*. We investigated the upstream pathway, involving pheromone signal reception, by exposing cells to synthetic Mating-Factor α and α -N. We hypothesized that exposure to pheromones from the opposite mating type would deter a cell from performing MTS, since the process is energetically expensive and potentially lethal, if the double stranded breaks formed during MTS are not fixed correctly. Our results suggest that exposure to pheromones from the opposite mating-type does qualitatively suppress MTS. While our qualitative data indicated that there is a difference in the strength of suppression between MF α and MF α -N, further investigation is needed to quantify which variant elicits a stronger response. We successfully developed a semi-quantitative multiplexing PCR technique to quantify the amount of MTS suppression following

² A list of the transcripts, fold change, length, coding potential score and coding value. Scores less than 0 indicate the transcript is noncoding. 5 novel non-coding transcripts were found and are highlighted in blue.

pheromone exposure. With this technique, we created a regression line to calculate the amount of MTS in a sample, based on known cell ratios.

By combining our multiplexing semi-qPCR technique and our regression line, we will be able to quantify the amount of MTS in any sample, under a variety of conditions. Although the central aim of developing this semi-qPCR procedure was to quantify the amount of MTS suppression following pheromone exposure, this technique will allow us to analyze MTS under any condition. To confirm that MTS is truly suppressed following exposure to pheromones from the opposite mating-type, we will need to redo our previous qualitative experiments with our newly developed semi-qPCR protocol. This quantification will be essential to see how effective pheromones are at halting MTS.

While our data suggest that pheromone exposure suppresses MTS, further research is required to confirm that pheromone exposure promotes mating. Increased mating capacity following pheromone exposure has been demonstrated in *K. phaffii*, a close relative of *O. polymorpha* (Fig. 2), indicating that this behavior may be present in *O. polymorpha* as well (Heistingner et al. 2018). As Yamamoto et al. 2017 noted in their paper, overexpression of *RME1*, another transcription factor responsible for MTS regulation, resulted in shmoo formation (mating projections that allow two cells to fuse). While our data only indicated that MTS is suppressed by pheromone exposure, we hypothesize that exposure to pheromones from the opposite mating-type diverts a cell's signaling cascade towards mating. We noted that, following 10 hours of growth with pheromones, cells were sticking to the sides of their test tubes. This could possibly indicate that the cells were producing a sticky excrement in order to mate more effectively (data not shown). Shaking the cultures during growth typically prevents

mating (Banderas et al. 2016), perhaps lending evidence towards a need for sticky excrements allowing the cells to grow towards each other. Extensive microscopy work, including 3D Z-Stacked Images, is necessary to investigate this byproduct from the cells, as well as look for shmoo formation (an indicator of mating initiation).

Additional investigation into mRNA differential expression after pheromone exposure could reveal changes in gene expression due to the activation of the pheromone signaling cascade. By overexpressing *STE12* with and without pheromone exposure, followed by RNA-sequencing, we could examine differential expression of genes involved in the MTS and mating specific pathways. This research would hopefully reveal an upregulation in mating-specific genes and a downregulation in MTS-specific genes, implying that pheromones from the opposite mating-type force cells to halt MTS in favor of mating. Finally, to see if the suppression we visualized was due to pheromone exposure, we successfully knocked out *STE2*, the α pheromone receptor. While research has looked into the impacts of *STE2* knockouts on mating (Yamamoto et al. 2017), we will need to examine the impacts of the knockout on the ability of cells to switch. We hypothesize that without the capacity to interact with pheromones, there should be no MTS suppression because cells will not detect a mating partner.

While we understand that environmental stimuli can quickly change a cell's gene expression, the pathways responsible for mediating those responses are still poorly understood. By looking at *O. polymorpha*'s pheromone pathway, we can contribute information to how environmental responses rapidly interrupt or divert signal cascade pathways.

Our downstream investigation of the MTS signal cascade pathway focused on the identification of lncRNAs upregulated during MTS. We used ribo-depletion RNAseq and differential expression analysis to find novel transcripts upregulated in the presence of *STE12*, meaning that they are upregulated during MTS. We then analyzed the coding potential of our most upregulated novel features and identified 5 putative lncRNA transcripts. The discovery of these putative lncRNAs marks the first experiment to describe lncRNAs in any context in *O. polymorpha*.

Going forward, overexpressing the lncRNAs, perhaps with a *pAOX* promoter system (Hanson et al. 2017), would be an effective way to understand the roles of each lncRNA. A more laborious but precise characterization of the putative lncRNAs would be to knock them down. Since many of our lncRNAs are antisense to essential genes, normal homologous recombination would not be an effective method of genome modification. Instead, by using a dead (d)Cas9 complex that targets the promoter sequences of the lncRNAs, we would be able to knockdown the lncRNAs without damaging the sense strand (Cao et al. 2018). We could then examine phenotype, using PCR, to see if MTS progression was halted following the loss of the lncRNAs. Another RNA-seq experiment examining whole genome differential expression would be necessary to characterize the lncRNAs' exact functions. By knocking down the lncRNAs, we would hope to see changes in MTS and mating-specific gene expression.

Additional bioinformatic analysis of the lncRNAs could also help us understand their conservation and function. lncRNA sequences mutate much quicker than protein coding sequences, making sequence analysis difficult when looking for evolutionary linkages. However, Yotsukura et al. 2017 demonstrated that promoter sequence

analysis was an effective tool for identifying conserved lncRNAs. In combination with structural and scaffolding analysis, which hints at lncRNA function, it is possible to use promoter sequence analysis to find conserved lncRNAs and derive their function from annotation in related species, such as *K. phaffii*, *C. albicans* and *S. cerevisiae* (Fig. 2).

lncRNAs are extremely abundant in a variety of yeast species, including close relatives of *O. polymorpha* (Yamashita et al. 2016). Researchers have identified a variety of regulatory roles played by lncRNAs. *IRT1* is a particularly striking example of this non-enzyme mediated regulation; when *IRT1* is expressed, it blocks the promoter region of *IME1*, a master regulator of meiosis and prevents cell division (van Werven et al. 2013). Since MTS, mating and meiosis are initiated by one coordinated pathway in *O. polymorpha* (Hanson et al. 2014), the regulation of *IME1* by *IRT1* in *S. cerevisiae* leads us to hypothesize that similar lncRNAs are present in *O. polymorpha*. However, even if no lncRNA is identified as a regulator of MTS, this research will help us understand how lncRNAs are able to regulate gene expression, what mechanisms they use and how they are conserved across species. This ultimately adds to the growing body of knowledge about non-enzyme mediated cellular regulation.

Overall, this project helped us analyze more broadly how eukaryotic cells mediate responses to their environments and what molecules are responsible for the progression of those responses. Understanding how eukaryotic cells 'talk' to each other, how they learn from their environment, and how they rapidly adapt has large human health implications. Signal cascades regulate the mechanisms responsible for sexual reproduction and evolution, which coincides with other behaviors that impact human health, such as pathogenesis. This research helps illuminate the regulatory

mechanisms behind specific methods of sexual reproduction. This lends to a better understanding of basic science and potential avenues in human health intervention, such as preventing infection and understanding the reproduction of pathogens.

6. Acknowledgements

We would like to thank the Colorado College Department of Molecular Biology for their support in creating and finishing this project. Specifically, we would like to thank Dr. Jenn Garcia for her mentorship contributions, Dr. Carrie Moon for her technical support and strain creation, and the Hanson Lab Summer 2018 and 2019 for their never-ending support. Additionally, we would like to thank the Summer Collaborative Research Grant and the Alfred W. Alberts Grant for funding our research over the last two years. Finally, I would like to thank Dr. Sara Hanson for her mentorship, support and guidance.

7. Works Cited

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169.
- Atkinson, S.R., Marguerat, S., and Bähler, J. (2012). Exploring long non-coding RNAs through sequencing. *Seminars in Cell & Developmental Biology* 23, 200-205.
- Banderas, A., Koltai, M., Anders, A., and Sourjik, V. (2016). Sensory input attenuation allows predictive sexual response in yeast. *Nat. Commun.* 7, 12590.
- Bardwell, L. (2005). A walk-through of the yeast mating pheromone response pathway. *Peptides* 26, 339-350.
- Barsoum E., Rajaei N., and Åström S. U. (2011). RAS/Cyclic AMP and Transcription Factor Msn2 Regulate Mating and Mating-Type Switching in the Yeast *Kluyveromyces lactis*. *Eukaryotic Cell* 10, 1545-1552.
- Cao, M., Gao, M., Ploessl, D., Song, C., and Shao, Z. (2018). CRISPR–Mediated Genome Editing and Gene Repression in *Scheffersomyces stipitis*. *Biotechnology Journal* 13, e1700598-n/a.
- Chen, P., Sapperstein, S.K., Choi, J.D., and Michaelis, S. (1997). Biogenesis of the *Saccharomyces cerevisiae* mating pheromone a-factor. *J. Cell Biol.* 136, 251-269.
- Coria, R., Kawasaki, L., Torres-Quiroz, F., Ongay-Larios, L., Sanchez-Paredes, E., Velazquez-Zavala, N., Navarro-Olmos, R., Rodriguez-Gonzalez, M., Aguilar-Corachan, R., and Coello, G. (2006). The pheromone response pathway of *Kluyveromyces lactis*. *FEMS Yeast Res.* 6, 336-344.
- Duina, A.A., Miller, M.E., and Keeney, J.B. (2014). Budding Yeast for Budding Geneticists: A Primer on the *Saccharomyces cerevisiae* Model System. *Genetics* 197, 33-48.
- Faber, K.N., Haima, P., Harder, W., Veenhuis, M., and AB, G. (1994). Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*. *Curr. Genet.* 25, 305-310.
- Fissore, R.A., Burton, A., and Lykke-Hartmann, K. (2019). Editorial: Molecular and Cellular Mechanisms in Reproduction and Early Development. *Front. Cell. Dev. Biol.* 7, 36.
- Ginsburg, E.S., Racowsky, C. (2014). Chapter 31 – Assisted Reproduction. Barbieri, R.L., Strauss, J.F. *Yen & Jaffe's Reproductive Endocrinology*. W.B. Saunders. Philadelphia, Pennsylvania. 7th Edition. p. 734-773.
- Haber, J.E. (2012). Mating-Type Genes and MAT Switching in *Saccharomyces cerevisiae*. *Genetics* 191, 33-64.

- Hancock, J.T. (2007). 3 – The Principles of Cell Signalling. Bentley, P.J., Kumar, S. *On Growth Forms and Computers*. Academic Press. London, U.K. p. 64-81.
- Hanson, S.J., Byrne, K.P., and Wolfe, K.H. (2017). Flip/flop mating-type switching in the methylotrophic yeast *Ogataea polymorpha* is regulated by an Efg1-Rme1-Ste12 pathway. *PLoS Genetics* *13*, e1007092.
- Hanson, S.J., and Wolfe, K.H. (2017). An Evolutionary Perspective on Yeast Mating-Type Switching. *Genetics* *206*, 9-32.
- Heistingering, L., Moser, J., Tatto, N.E., Valli, M., Gasser, B., and Mattanovich, D. (2018). Identification and characterization of the *Komagataella phaffii* mating pheromone genes. *FEMS Yeast Res.* *18*, 10.1093/femsyr/foy051.
- Heitman, J. (2006). Sexual reproduction and the evolution of microbial pathogens. *Curr. Biol.* *16*, 711.
- Herskowitz, I. (1988). Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.* *52*, 536-553.
- Jones, S.K., and Bennett, R.J. (2011). Fungal mating pheromones: choreographing the dating game. *Fungal Genet. Biol.* *48*, 668-676.
- Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* *35*, 345.
- Lee, C.S., and Haber, J.E. (2015). Mating-type Gene Switching in *Saccharomyces cerevisiae*. *Microbiol. Spectr.* *3*, MDNA3-2014.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550-8.
- Madhani, H. (2006). *From a to α : Yeast as a Model for Cellular Differentiation*. Cold Springs Harbor Laboratory Press. New York.
- Maekawa, H., and Kaneko, Y. (2014). Inversion of the Chromosomal Region between Two Mating Type Loci Switches the Mating Type in *Hansenula polymorpha*. *PLoS Genetics* *10*, e1004796.
- Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bahler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* *151*, 671-683.
- Mohammadi, S., Saberidokht, B., Subramaniam, S., and Grama, A. (2015). Scope and limitations of yeast as a model organism for studying human tissue-specific pathways. *BMC Syst. Biol.* *9*, 96-0.
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* *11*, 1650-1667.

- Riley, R., Haridas, S., Wolfe, K.H., Lopes, M.R., Hittinger, C.T., Goker, M., Salamov, A.A., Wisecaver, J.H., Long, T.M., Calvey, C.H., *et al.* (2016). Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 9882-9887.
- Sherwood, R.K., Scaduto, C.M., Torres, S.E., and Bennett, R.J. (2014). Convergent evolution of a fused sexual cycle promotes the haploid lifestyle. *Nature* *506*, 387-390.
- Si, H., Hernday, A.D., Hiraakawa, M.P., Johnson, A.D., and Bennett, R.J. (2013). *Candida albicans* white and opaque cells undergo distinct programs of filamentous growth. *PLoS Pathog.* *9*, e1003210.
- Sorrells, T.R., Booth, L.N., Tuch, B.B., and Johnson, A.D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature* *523*, 361-365.
- Sun, W.H., Wang, Y.Z., Xu, Y., and Yu, X.W. (2019). Genome-wide analysis of long non-coding RNAs in *Pichia pastoris* during stress by RNA sequencing. *Genomics* *111*, 398-406.
- Tao, L., Cao, C., Liang, W., Guan, G., Zhang, Q., Nobile, C.J., and Huang, G. (2014). White cells facilitate opposite- and same-sex mating of opaque cells in *Candida albicans*. *PLoS Genet.* *10*, e1004737.
- van Werven, F., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. (2012). Transcription of Two Long Noncoding RNAs Mediates Mating-Type Control of Gametogenesis in Budding Yeast. *Cell* *150*, 1170-1181.
- Wallen, R.M., and Perlin, M.H. (2018). An Overview of the Function and Maintenance of Sexual Reproduction in Dikaryotic Fungi. *Front. Microbiol.* *9*, 503.
- Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* *7*, 1338.
- Wolfe, K.H., and Butler, G. (2017). Evolution of Mating in the Saccharomycotina. *Annu. Rev. Microbiol.* *71*, 197-214.
- Yamamoto, K., Tran, T.N.M., Takegawa, K., Kaneko, Y., and Maekawa, H. (2017). Regulation of mating type switching by the mating type genes and RME1 in *Ogataea polymorpha*. *Scientific Reports* *7*, 16318-12.
- Yamashita, A., Shichino, Y., and Yamamoto, M. (2016). The long non-coding RNA world in yeasts. *Biochim. Biophys. Acta-Genet. Regul. Mech.* *1859*, 147-154.
- Yotsukura, S., duVerle, D., Hancock, T., Natsume-Kitatani, Y., and Mamitsuka, H. (2017). Computational recognition for long non-coding RNA (lncRNA): Software and databases. *Briefings in Bioinformatics* *18*, 9-27.