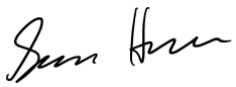
Mating-Type Specific Gene Expression in the Methylotrophic Yeast *Ogataea polymorpha*

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

By
Kristie Shirley
May 2020

Prof. Sara J. Hanson
Primary Thesis Advisor



_____

Prof. Ralph Garcia-Bertrand
Secondary Thesis Advisor

Ralph Garcia-Bertrand

Abstract

The mating types in *Ogataea polymorpha* are dictated by the expression of transcription factors *MATa2* or *MATα1* in a and α-type cells, respectively and the mating response is activated in response to environmental stress. Although it is known that these transcription factors are essential for mating-type identity, their regulatory targets are unknown in *O. polymorpha*. Since using environmental stressors to induce mating does not effectively show differential gene expression between mating types, we used a more direct approach by inducing a key regulator of mating, *STE12*. *STE12* is a sequence-specific transcription factor that is activated by environmental stressors (nitrogen deprivation) in *O. polymorpha* and as such is responsible for inducing the expression of mating genes.

In order to manipulate the mating response for experimentation, the expression of *STE12* was put under the control of an inducible promoter and overexpressed in both wildtype strains and strains that had either *MATa2* or *MATα1* deletions. Total RNA-seq analyses compared the RNA expression of deletion strains to control strains and the differential gene expression between the two revealed the regulatory targets of the MAT transcription factors. Many of the mating-type specific genes were involved in the mating-type specific pheromone response pathway and the haploid specific genes were a part of a more general pheromone response pathway.

By studying the mating-type specific gene expression in *O. polymorpha*, we can better understand not only how cells are different within a species but also between species through comparison of their mating-type specific genes. Since speciation relies on sexual barriers to separate species, comparing how mating-type specific genes are different and used differently

between yeast species helps us better understand how these species became sexually isolated and how new species are created as a result.

Introduction

Cellular Differentiation: One Genome for Many Cell-Types

Cellular differentiation is essential to multicellular beings and unicellular beings alike. In canonical differentiation, differentiated cells typically cannot dedifferentiate and do not differ in DNA content from other differentiated cells (Hanson and Wolfe 2017.) Cellular differentiation is especially important for the complexity within life because it allows for a single genome to differentially express genes, giving rise to many unique cell-types with specialized functions. Cellular differentiation is usually accomplished by the activation of a specific cohort of transcription factors which then go on to affect these changes within the wider regulatory network of the cell through combinatorial and cooperative activity (Hobert 2008.)

Generally, transcription factors act by binding DNA at sequence-specific, *cis*-regulatory sequences after which it can then either positively or negatively regulate transcription of its target gene (Hobert 2008.) The regulatory effects of transcription factors can be wide reaching since they need only a short series of nucleotides for sequence specificity and thus regulatory activity (Hobert 2008, Brivanlou and Darnell 2002.) Considering their profound effects within the cell, transcription factors are often very tightly regulated to ensure that they are expressed at the proper times such as during differentiation or in response to stimuli (Brivanlou and Darnell 2002.)

The aberrant expression of transcription factors can greatly affect health at both the cellular and organismal level (Zheng and Blobel 2011.) In fact, cancers often times rely on irregular expression of transcription factors, particularly those that regulate cellular differentiation and the cell-cycle, in order to survive and can even adopt an abnormal dedifferentiated state which can help them do so (Zheng and Blobel 2011.) The GATA

transcription factors are essential regulators of cellular differentiation and also play a role in stem cells where they regulate their expansion, maturation, and cell-cycle control (Zheng and Blobel 2011.) In cancerous cells, the GATA transcription factors can be structurally mutated or completely silenced which can help in the progression of the cancerous phenotype. For example, downregulation of GATA3 in breast cancer cells allows the tumor cells to modify the cell's surface to appear like a mammary epithelial stem cell and also allows the cancer cells to move more easily (metastasize) and is thus correlated with a worse prognosis (Zheng and Blobel 2011.)

Cellular differentiation relies on transcription networks to differentially activate subsets of genes and therefore allows for a diverse subset of cell-types that originate from a single genome (Hobert 2008.) Studying cellular differentiation is essential to understanding basic cell biology because transcription factors dictate the identity of a cell (Brivanlou and Darnell 2002.) By understanding how transcription factors enact this cellular identity, we can develop better ways of mitigating the effects of when its complex regulatory networks go awry.

Considering the importance of cellular differentiation to the health of an organism, intentional deviance from its canonical "rules" seems risky yet yeast seem to do so with ease (Hanson and Wolfe 2017.) Yeast are single-celled fungi, which like many other organisms, can also differentiate their cells. Interestingly though, the differentiated cells in yeast differ in their DNA content and can also dedifferentiate (Hanson and Wolfe 2017.) These attributes, in eukaryotic cells, usually indicate some abnormality within the cell but these unusual aspects are actually essential to the yeast lifecycle (Hanson and Wolfe 2017.)

Yeast can Reproduce Both Asexually and Sexually

Differentiated cells are the crux of the yeast reproductive cycle. Sexual cells, called mating types, are haploid and come in two forms: a and alpha. These haploid mating types can either reproduce asexually through mitosis or sexually through the syngamy of opposite mating types to form diploid a/alpha cells (Hanson and Wolfe 2017.) In many yeast, a/alpha diploid cells then immediately undergo meiosis to form haploid spores which are then encased in a protective ascus until the environment becomes more habitable (**Figure 1**) (Hanson et. al 2014.)

Sexual reproduction is highly regulated within yeast because they are immobile and therefore expend a lot of energy shmooing (growing) towards the source of the opposite mating type's pheromones in order to sexually reproduce (Bardwell 2005.) Once the cell's shmoos (mating projections) are touching, the cells fuse their cell walls and nuclear envelopes to form a diploid cell. Diploid a/alpha cells can then either undergo meiosis to return to haploidy or remain as a diploid, this preference is dependent on the species (Lee et. al 2010.) Given that the physical process of mating is directed almost entirely by the chemotropic sensing of the pheromone-gradient surrounding the cell, properly expressing mating-type specific pheromones and receptors is essential for effective mating partner recognition (Bardwell 2005, Maekawa and Kaneko 2014.). In fact, in the presence of high concentrations of pheromone, many yeasts often excrete proteases to cleave the opposite mating type's pheromones in order to re-sensitize themselves to the pheromone-gradient which ensures proper mating behavior (Bardwell 2005, Barkai et. al 1998.)

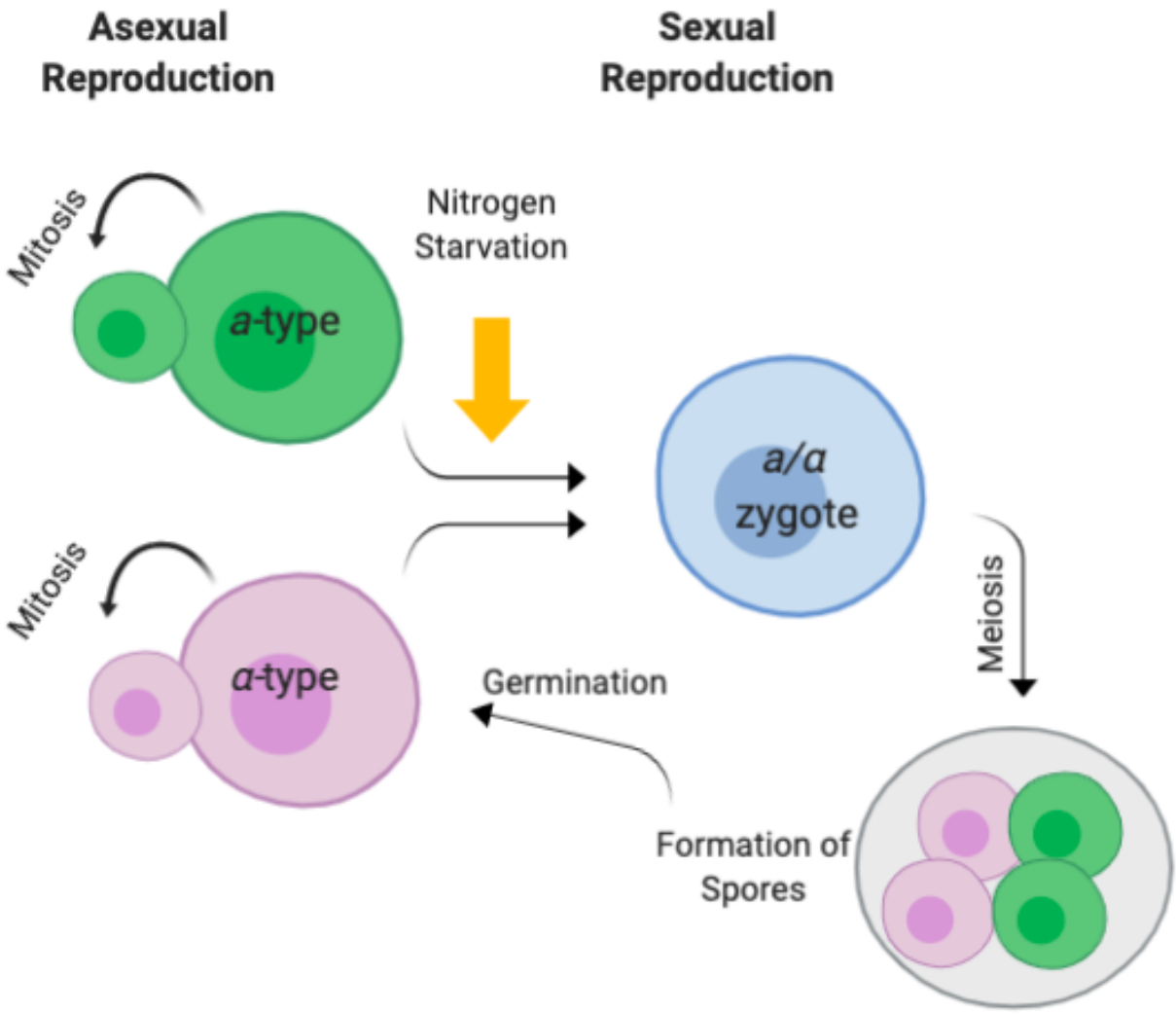


Figure 1: Reproductive Cycle in *O. polymorpha*. Haploid cells can reproduce both sexually and asexually. Sexual reproduction in *O. polymorpha* is regulated by nitrogen-starvation (yellow arrow) at which point they can both switch mating types (not shown here) and/or mate sexually. Zygotes immediately undergo meiosis and are enclosed within a protective ascus until which point the environment becomes suitable for germination. Cell-types are color-coded as is consistent throughout the paper.

Differentiation of Mating Types is Essential for Sexual Reproduction

The differentiation of mating-types, which is directed by the expression of the MAT transcription factors, is extremely important for sexual reproduction within yeast because they depend on mating-type specific pheromone and receptor production to find mating partners (Lee et. al 2010.) Canonically, a-type cells produce a-factor pheromone and alpha-factor receptors and

alpha-type cells produce alpha-factor pheromone and a-factor receptors (**Figure 2**) (Lee et. al 2010.) In yeast, the alpha and a pheromone receptors (*STE2* and *STE3*, respectively) are seven-transmembrane receptors, also known as G protein-coupled receptors (GPCRs,) which are a highly conserved family of cell-surface receptor proteins (Madhani 2007.) Binding of the appropriate ligand, the opposite mating-type's pheromone in this case, causes a conformational change that in turn activates a kinase cascade within the cell which ultimately arrests the cell cycle and activates the transcription of genes relevant to the mating response (Madhani 2007.)

Mating-type directed cellular differentiation, and therefore identity, is dependent on the expression of mating-type specific transcription factors which can be found at the mating-type (MAT) locus (Hanson et. al 2017.) In the methylotrophic yeast, *O. polymorpha*, these transcription factors are *MATa2*, which is essential for a-type identity, and *MATα1*, which is essential for alpha-type identity (Yamamoto et. al 2017.) The expression of the genes encoding the transcription factors is dependent on their position in relation to the centromere where the gene closest to the centromere is silenced and the gene farthest from the centromere is expressed (Hanson et. al 2014.) Contrary to most differentiated cells, yeast haploids can switch between mating types (differentiated states) through a physical DNA change at the MAT locus (Hanson and Wolfe 2017.) In *O. polymorpha*, this switch occurs through an inversion of the MAT locus which lies between two inverted repeats, but the mechanism through which this inversion occurs is still unknown (**Figure 3**) (Hanson et. al 2014, Maekawa and Kaneko 2014.)

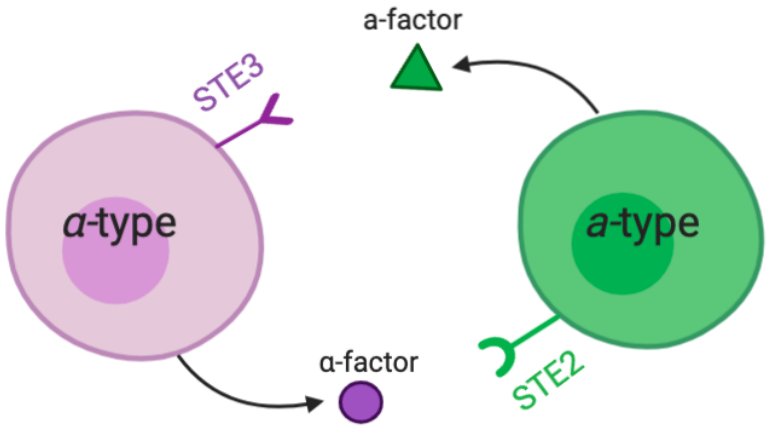


Figure 2: Mating-type Specific Pheromones and Pheromone Receptors Allow for Recognition of Sexual Partners. Canonical pheromone-based mating-partner identification where STE2 is the alpha-factor receptor and STE3 is the a-factor receptor.

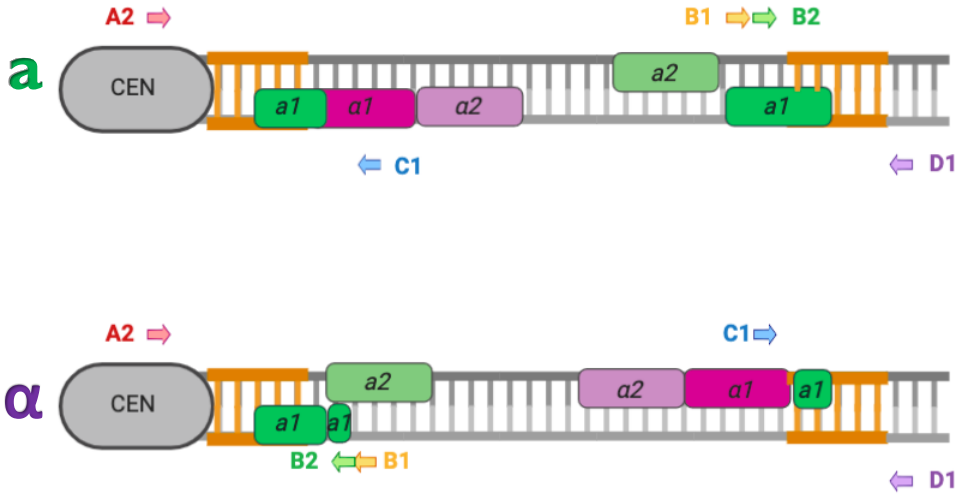


Figure 3: Mating-type (MAT) Locus Structure in *O. polymorpha*. Cell type has been specified to the left of the schematic. A-specific transcription factors are shown in green and alpha-specific transcription factors have been shown in purple. CEN denotes the position of centromeric sequence and gold-colored sequences are the inverted repeats between which the chromosomal inversion happens in mating-type switching. Colored arrows depict the primer combinations used to check for mating-type orientation where bands for A2/C1 and B1/D1 indicate an a-oriented cell and bands for A2/B2 and C1/D1 indicate an alpha-oriented cell.

Mating-Type Gene Regulatory Circuits

MATa2 and *MATα1* have been described as being essential for a and alpha-type cell identities, respectively, because deletion of either one prevents the cell from mating properly (Yamamoto et. al 2017.) While the mechanisms through which these transcription factors work has not been directly characterized in *O. polymorpha,* it has been extensively studied in *S. cerevisiae.* Using *S. cerevisiae*'s paradigm, it is inferred that these transcription factors, when expressed, bind to enhancer elements upstream of mating-type specific gene promoters and activate their transcription (**Figure 4**) (Hanson and Wolfe 2017.) In order to properly regulate mating-type specific gene expression, the MAT transcription factors require the Mcm1 dimer as a cofactor. The Mcm1 dimer confers their ability to regulate gene expression because it can form activator or repressor complexes, depending on which MAT transcription factors are being expressed (Christ and Tye 1991.) This ensures that mating-type specific genes are only transcribed in the presence of its transcription factors (Hanson and Wolfe 2017.)

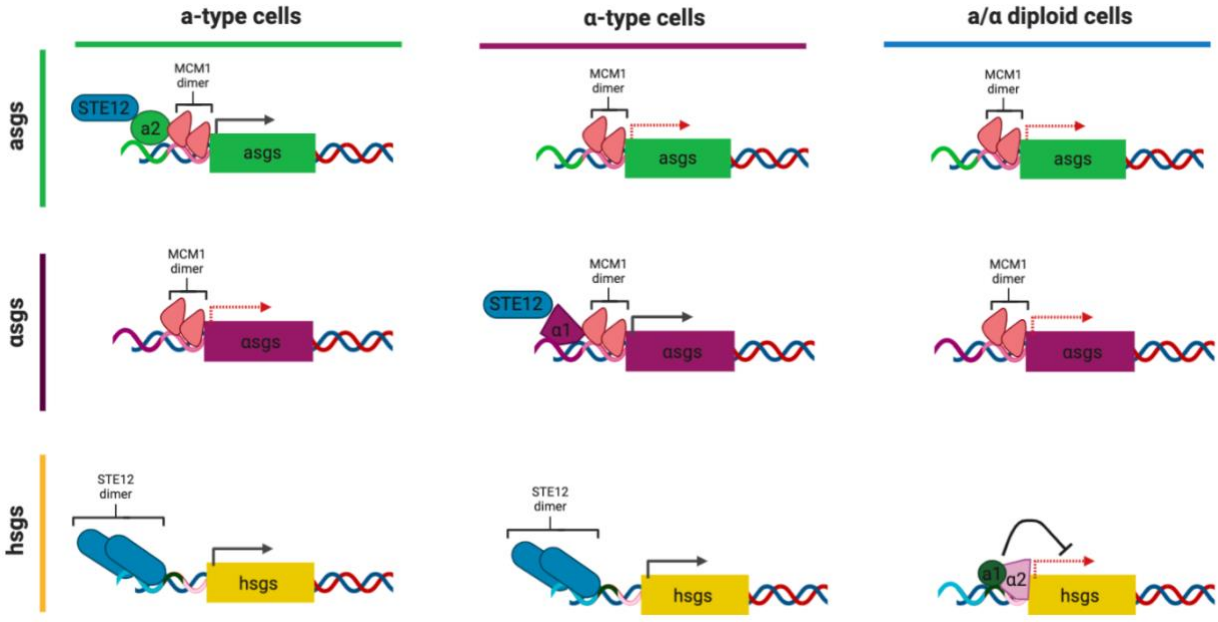


Figure 4: Regulatory Control of A-Specific (asg,) Alpha-Specific (αsg,) and Haploid-Specific (hsg) Genes in A, Alpha, and Diploid Cells. Red, dashed promoter lines indicate no expression whereas black lines indicate expressed genes. Adapted from Madhani 2007.

Ste12 is the key-regulator of the mating-response because it is the downstream-most transcription factor within the nutritional starvation and pheromone response pathways and is what ultimately activates the mating response (Hanson et. al 2017.) Unlike *S. cerevisiae,* which has Ste12 binding sites upstream of its a-specific gene (asg) promoters, work in other yeast species indicate that *O. polymorpha* likely indirectly recruits Ste12 to asg promoters through interactions with MATa2 (Sorrells et. al 2015, Hanson and Wolfe 2017.) The lack of Ste12 binding sites upstream of mating-type specific genes in *O. polymorpha* ensures that their expression is turned off by default whereas *S. cerevisiae* requires MATα2 repression of asgs (**Figure 4**) (Sorrells et. al 2015.)

Despite knowing that *MATa2* and *MATα1* are needed for proper mating behaviors in a and alpha-type cells, respectively, the regulatory targets of these transcription factors is unknown in *O. polymorpha* (Yamamoto et. al 2017.) Other yeast species such as *S. cerevisiae, Kluyveromyces lactis,* and *Candida albicans* have better defined mating-type specific genes, many of which are involved in ensuring proper expression of pheromone pathway genes (**Figure 2**) (Booth et. al 2010.)

To explore the effectors of the mating-type specific transcription factors in *O. polymorpha*, we used strains that overexpressed *STE12* in the presence of methanol and deleted *MATa2* and *MATα1* in a-type and alpha-type cells respectively. We then performed total RNA-seq and compared the difference in RNA expression between deletion and non-deletion strains within each mating type. This revealed the genes that were affected by the activation of the mating-type specific transcription factors in *O. polymorpha* and furthermore, allowed us to compare these findings to those in other yeast. Gaining a better understanding of how *O. polymorpha* defines its mating-types as compared to other species can contextualize how such a

diversity of yeast species can mate so accurately despite using such similar machinery to accomplish mating.

Results

Deletion of *MATα1* and replacement with a selectable marker

In order to identify mating-type specific genes in their respective cell types, the *MAT* transcription factors responsible for mating-type specific gene expression were deleted so that their gene expression during the *STE12*-induced mating response could be compared to that of control strains. Since the *MATa2* deletion strain had already been created, only the *MATα1* deletion strain needed to be constructed.

To delete *MATα1*, I created a fusion construct with 5' upstream and 3' downstream fragments of *MATα1* so that upon electroporation, the construct could be incorporated into the genome using homologous recombination. The 5' and 3' fragments were amplified along with the nourseothricin resistance gene (*NAT)* and fused using PCR (**Figure 5**.) This construct was then transformed into alpha-oriented cells and incorporated into the genome through homologous recombination with the fragments upstream and downstream of *MATα1,* ensuring a total deletion of the gene (**Figure 6**.)

In total, seven transformants were collected from the selective media plate and were quality checked for the proper elements. PCR amplification and analysis with gel electrophoresis confirmed the deletion of *MATα1*, MAT locus in alpha-orientation, and presence of *NAT* which gives them resistance to nourseothricin within six of the seven total transformants (**Figure 7**.) The primers used to ensure *MATα1* deletion amplified the interior of the *MATα1* gene. Similarly, primers used to confirm MAT locus orientation amplified only when the chromosomes were oriented in such a way that allows for the proper 5' and 3' attachments of the primers (primer combinations can be seen in (**Figure 3.**) Each orientation had two primer combinations that could yield a product yet, when confirming for alpha-orientation in the deletion strains, the

second primer combination failed to yield a product despite the cells being in alpha-orientation. This was because the C1/D1 primer combination amplified part of the *MATα1* which was effectively deleted from the genome as indicated by the lack of a band in the gel (**Figure 5D**.) Confirmation of all the necessary elements in the *MATα1* deletion strain made us confident in their quality, so they were used for the rest of the study.

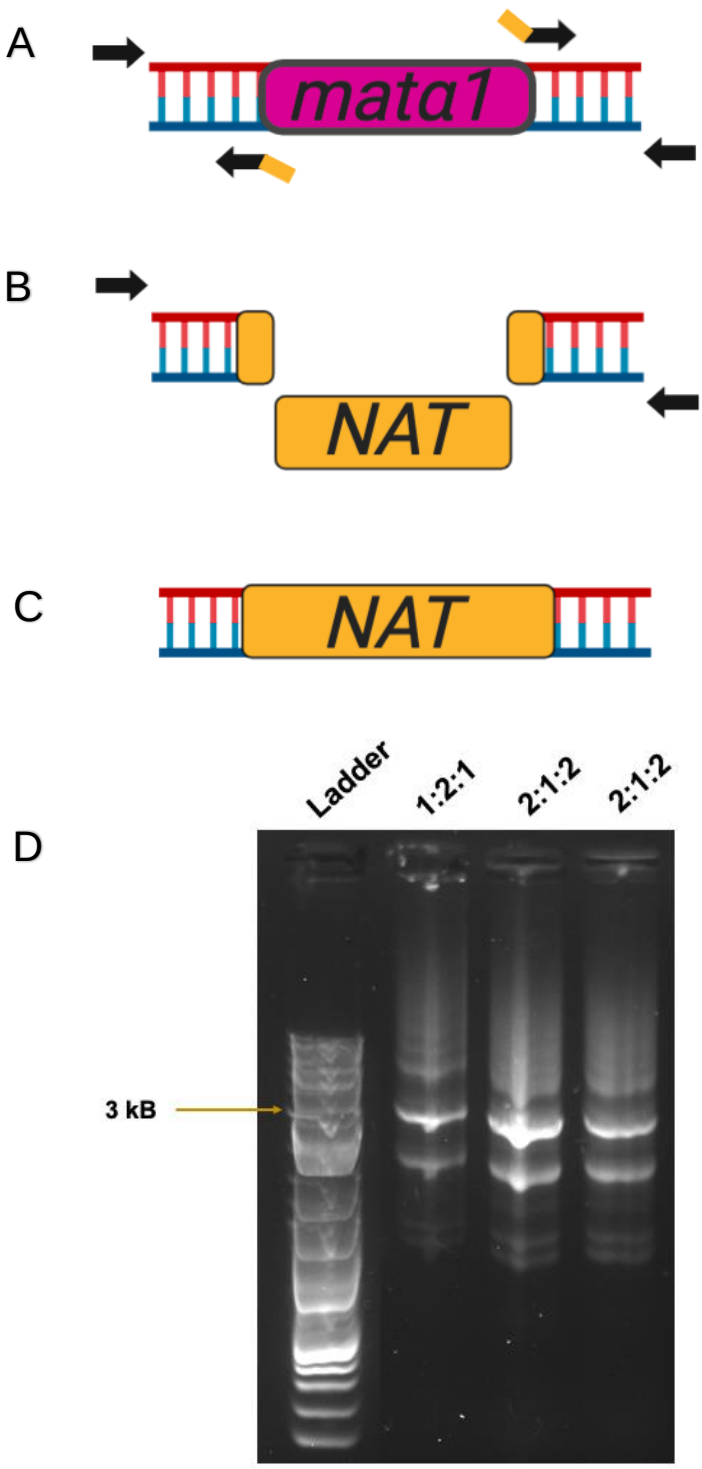


Figure 5: Deletion Construct Design. (A) Fragments 5' upstream and 3' downstream of *MATα1* were amplified from alpha-oriented yeast with primers that had fragments of the *NAT* gene as indicated by the yellow-tailed primer arrows. **(B)** The 5' and 3' fragments as well as the functional NAT gene (isolated from a plasmid) were **(C)** fused using PCR to produce the final fusion construct. **(D)** The 3kB construct was then visualized using gel electrophoresis. Lanes are labelled with molar ratios of 5' fragment : *NAT* gene : 3' fragment used in the fusion construct.

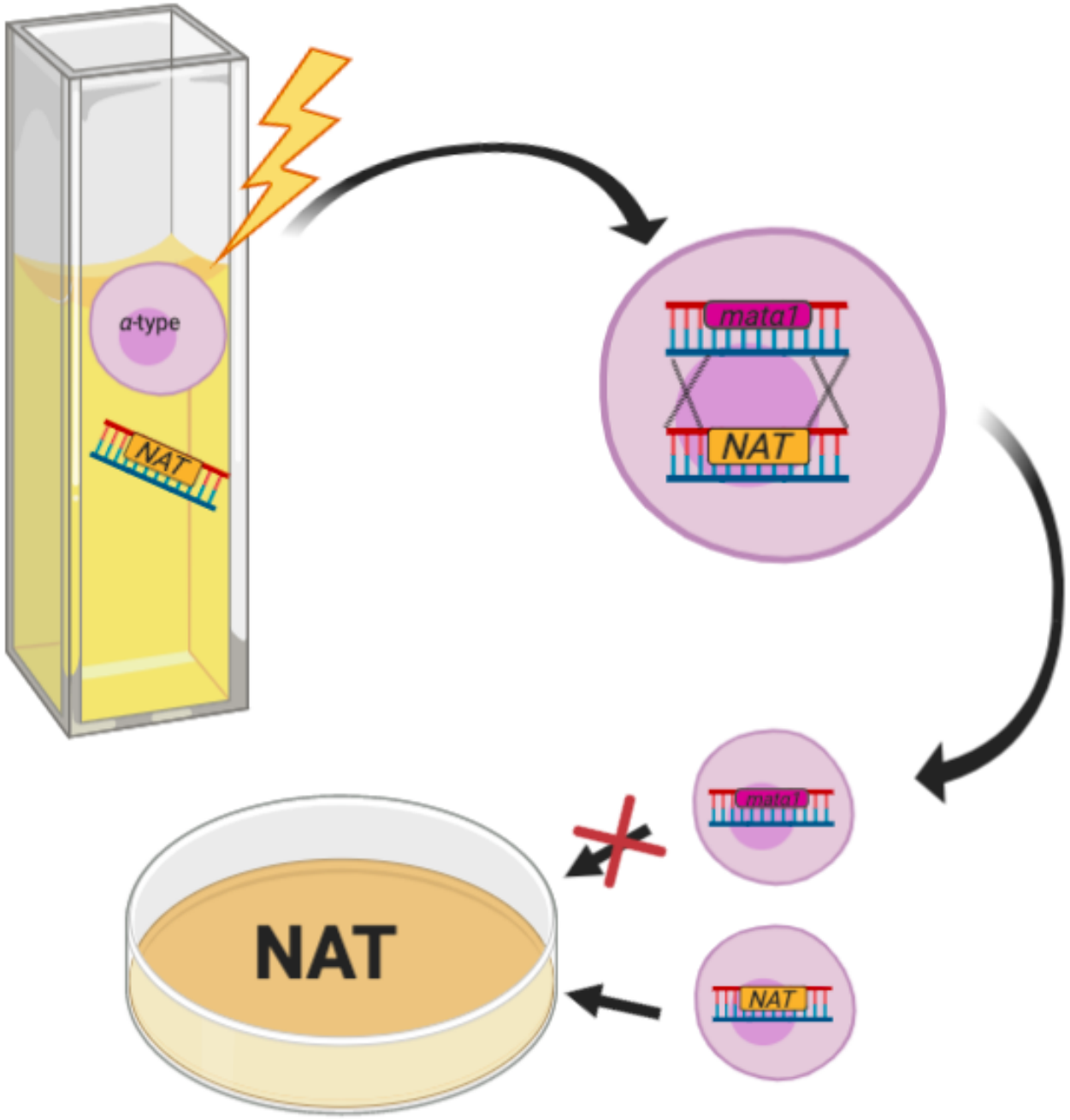


Figure 6: Transformation of Construct into Alpha-Oriented *O. polymorpha* Cells. Alpha-oriented *O. polymorpha cells* were electroporated in a cuvette containing the deletion construct. This allowed for the incorporation of the construct into the genome using homologous recombination which in turn deleted the entirety of the *MATα1* gene. These transformants were then selected for using YPD+NAT plates.

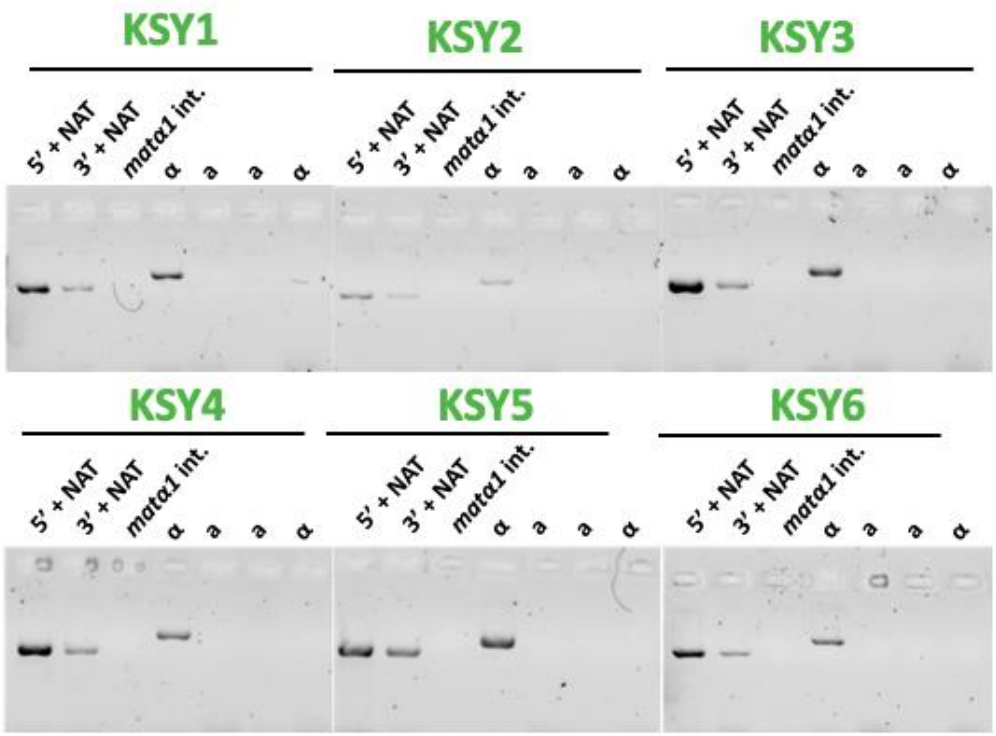


Figure 7: Transformants are Confirmed to be Deletion Mutants. Using gel electrophoresis, six of the seven transformants were quality checked to ensure that both the 5'/NAT (2.3kB) and 3'/NAT (2.3kB)fragments were present, that the interior of *MATα1* was deleted (300bp,) and that they were in alpha-orientation (2.7kB, 2.3kB) and not in a-orientation (2.7kB, 2.3kB.) The expected 2.3kB product for the second alpha-orientation primter combination failed to appear in the gel because it amplified a portion of the deleted *MATα1* gene.

***MATa2* and *MATα1* are not needed for *STE12*-induced mating-type switching**

As indicated by Hanson et. al 2017, *STE12* overexpression leads to a very strong mating response as indicated by upregulation of pheromone pathway genes and mating-type switching in *O. polymorpha*. The strains of *O. polymorpha* used in this study have *STE12* under the control of the alcohol oxidase promoter (P_{AOX}) which allowed us to induce the expression of *STE12* through incubation with a methanol solution (see Table 2 in appendix for strain genotypes.) We chose to induce *STE12* directly instead of using nitrogen deprivation to induce mating because *O. polymorpha* grows poorly in these conditions and we needed enough sample for RNA sequencing. Also, the nutritional starvation response induces so many genes in *O. polymorpha*

that it is impossible to separate which were related to mating and which were general to the starvation stress response of the cell (Hanson et. al 2017.) Because it was known that mating-type switching is not affected by the deletion of the mating-type transcription factors, we used switching as an indicator that *STE12*, and therefore the mating response, was properly and robustly induced (Yamamoto et. al 2017.)

Deletion and control strains were grown up to the log-phase in YPD and then in methanol to induce *STE12* and the mating pathway. Samples of culture were taken before and after methanol induction and were used for DNA extractions. Primers specific to mating-type orientation were used to visualize the orientation of the strains on an agarose gel (see Figure 2 for primer positions.) As seen in the gel in **Figure 8**, all strains switched mating types after the induction of *STE12* and the mating pathway. Mating-type switching indicates that the strains are responding to *STE12* effectively and expressing the expected mating-response genes and mating-type specific genes as well. After confirmation of switching, a portion of the post-methanol induction culture was taken and used for RNA extractions which were then sent for high-throughput sequencing.

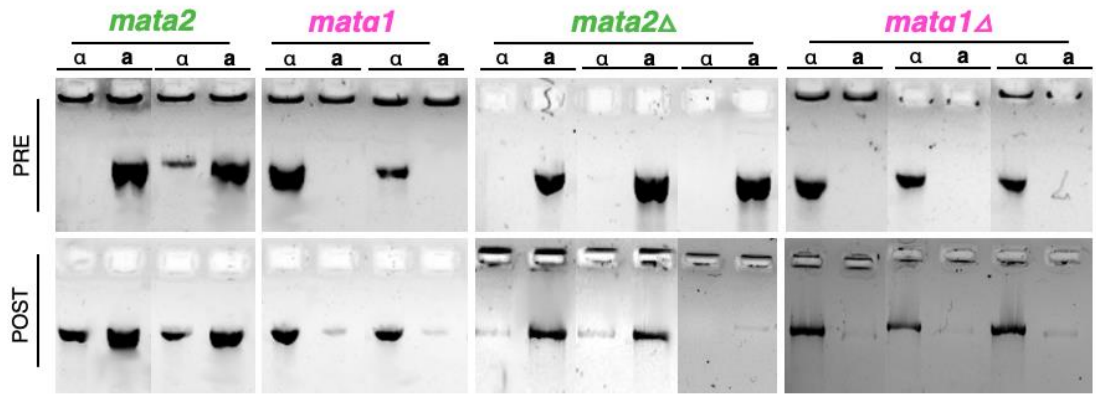


Figure 8: Mating-Type Switching Can Proceed Without *MATa2* and *MATα1*. Gels comparing the mating-type locus orientation in cells grown in YPD pe-induction (-STE12) and those grown in Mineral Media + Methanol (MMM) post-induction (STE12+.) Primers for a-oriented loci included A2/C1 and B1/D1 and primers for alpha-oriented loci included A2/B2 and C1/D1, primer sequences can be found in Table 1.

Deletion of *MATα1* Dramatically Decreases Expression of Pheromone Pathway Genes

The RNA collected from post-methanol incubation samples was sequenced using ribosomal RNA-depleted total RNA-sequencing, and we compared RNA expression between deletion and control strains. Computer analysis mapped sequence reads to an annotated *O. polymorpha* genome and counted the number of hits for each annotated region. These expression counts were used to then compare deletion and control strains the form of ratios where the deletion strain is the numerator and normalized to the control strain, the denominator. The numerical representation of this ratio is the $\log_2$ fold change values which indicate the $\log_2$ based fold-change in gene expression between the deletion and control conditions. Since the gene expression of the deletion strains are in the numerator and the control in the denominator, if *MATα1* positively regulates a gene's expression then the gene's expression will be higher in the denominator than in the numerator and this difference will appear as a negative $\log_2$ fold change and vice versa for negatively regulated genes.

The top 30 genes with the most negative $\log_2$ fold change between the *MATα1* deletion and control strains in the *MATα1* dataset (meaning the genes most positively regulated by *MATα1*) were made into a heatmap that includes the differential expression of those genes in the *MATa2* deletion and *STE12* overexpression dataset (**Figure 9**.) As expected, many of the genes most positively regulated by *MATα1* were involved in the pheromone pathway. Considering the linked cycle of switching and mating in *O. polymorpha*, it is unsurprising that the pheromones they use as a chemical compass in order shmoo towards each other for mating are incredibly important in defining a cell's mating-type identity (Hanson and Wolfe 2017, Bardwell 2005.)

Among these pheromone pathway genes were *MFalpha* pheromone (-4.32 $\log_2$ (FC,) $1.63\text{x}10^{-77}$ adj p-value,) *STE3* a-pheromone receptor (-2.99 $\log_2$ (FC,) $1.31\text{x}10^{-36}$ adj p-value,) *AFB1* a-factor blocker (-1.11 $\log_2$ (FC,) NA adj p-value,) and *KEX2* which is a protease involved in activation of proproteins in secretory pathway (-0.60 $\log_2$ (FC,) NA adj p-value) (Jenness et. al 1983, Mackay and Manney 1974, Huberman and Murray 2013, Leibowitz and Wickner 1976.)

When comparing the *MATα1* dataset to the *MATa2* dataset, it is noticeable that there is an overlap in genes that are positively regulated by the two as indicated by the similar coloring in the heatmap (**Figure 9**). While some amount unintended expression (i.e.promoter leakage) is expected for genes, the effects of these transcripts should be relatively widespread and therefore should not affect the overall read counts (Huang et. al 2015.) Notably, there was an upregulation of pheromone pathway genes that were unexpected for the mating-type such as expression of the receptor for its own pheromone as well as expression of the opposite mating-type's pheromone (STE2 -0.37 $\log_2$ (FC,) 0.10 adj p-value and MFa -1.01 $\log_2$ (FC,) $1.43\text{x}10^{-8}$ adj p-value) (Jenness et. al 1983, Bender and Sprague 1986.) This suggests that mating-type identity is not conferred

simply by the expression of these pheromone pathway genes since both mating-types appear to be expressing both types of pheromones and receptors.

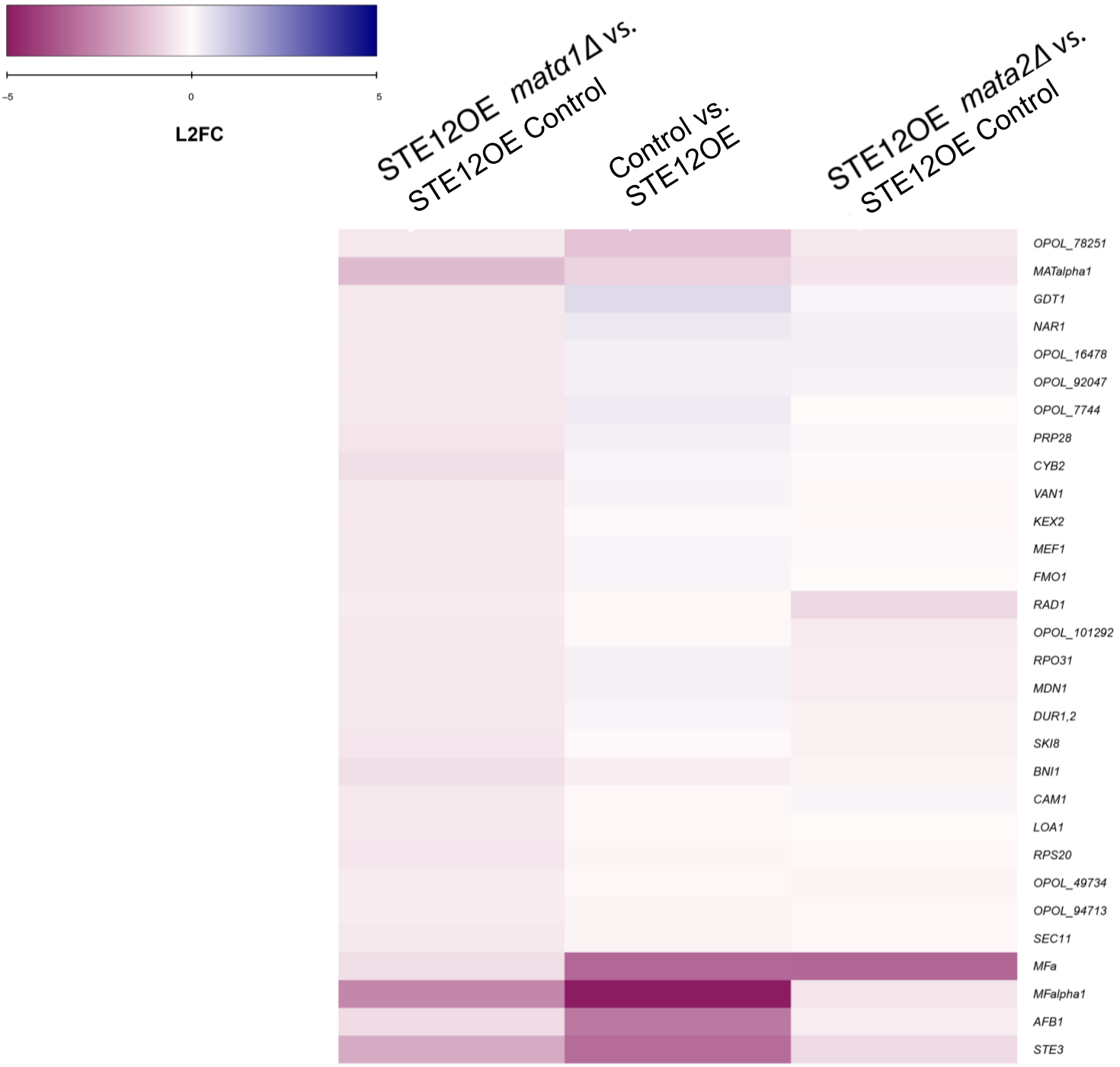


Figure 9: Genes Most Positively Regulated by *MATα1*. Top 30 genes sorted according to most negative $\log_2$ fold change (L2FC) in the *MATα1* dataset. Ratio order of strains are indicated above each column where the more negative the $\log_2$ fold change, the higher the expression of the gene in the control strain. The STE12OE column served as a control for our experimental groups since all strains used have the STE12OE construct.

***MATa2* most positively regulates pheromone pathway genes and DNA-repair genes**

Much like with the *MATα1* dataset, the top 30 genes with the most negative $\log_2$ fold change between the *MATa2* deletion and control strains in the *MATa2* dataset (meaning the genes most positively regulated by *MATa2*) were made into a heatmap that includes the differential expression of those genes in the *MATα1* deletion and *STE12* overexpression dataset (Figure 10.) *MATa2* also most strongly upregulates pheromone pathway genes (**Figure 10**.) Among these genes were the MFa pheromone (-5.53 $\log_2$ (FC,) 5.89x10^{-136} adj p-value,) *STE2* alpha-receptor (-4.80 $\log_2$ (FC,) 2.76x10^{-85} adj p-value,) *STE6* ABC a-factor transporter (-2.63 $\log_2$ (FC,) 6.93x10^{-38} adj p-value,) *AXL1* protease for a-factor maturation (-2.10 $\log_2$ (FC,) 4.01x10^{-11} adj p-value,) and *ASG7* regulator of signaling from *STE4* upstream of *STE3* (-2.07 $\log_2$ (FC,) 2.57x10^{-8} adj p-value) (Hagen et. al 1986, Wilson and Herskowitz 1984, Michaelis and Barrowman 2012, Zhong et. al 1999.)

MATa2 also strongly regulated DNA repair genes such as *SCC2*, a subunit of Scc2-Scc4 cohesin loading factor which establishes sister chromatid cohesion during double-stranded break repair (-1.58 $\log_2$ (FC,) 1.77x10^{-4} adj p-value) (Michaelis et. al 1997, Wong 2010.) *OPOL_83219* and *OPOL_12884*, both of which are putative transcription factors that localize to the cytoplasm and nucleus during double-stranded break repair (-1.22 $\log_2$ (FC,) 0.01 adj p-value and -1.42 $\log_2$ (FC,) 0.002 adj p-value) (localization was indicated in annotated *O. polymorpha* genome with *S. cerevisiae* orthologues provided by Dr. Hanson.) As well as *SRS2*, a DNA helicase and DNA dependent ATP-ase involved in DNA repair and checkpoint recovery, it is needed for proper commitment to meiotic recombination and transition from meiosis I to II (-1.08 $\log_2$ (FC,) 0.05 adj p-value) (Rong et. al 1991, Hegde and Klein 2000, Palladino and Klein 1992.)

Similar to the pattern of expression in the *MATα1* dataset, it was noted that *MATa2* positively regulated *STE3* receptor for a-factor which is its own pheromone (-1.19 $\log_2$ (FC,) 0.01 adj p-value) and MFalpha, the opposite mating-type's pheromone (-0.77 $\log_2$ (FC,) $7x10^{-3}$ adj p-value) (Hagen et. al 1986, Jenness et. al 1983.) Taken together, the data suggest that *MATα1* and *MATa2* upregulate both pheromones and receptors but to different extents as indicated by the difference in $\log_2$ (FC) values.

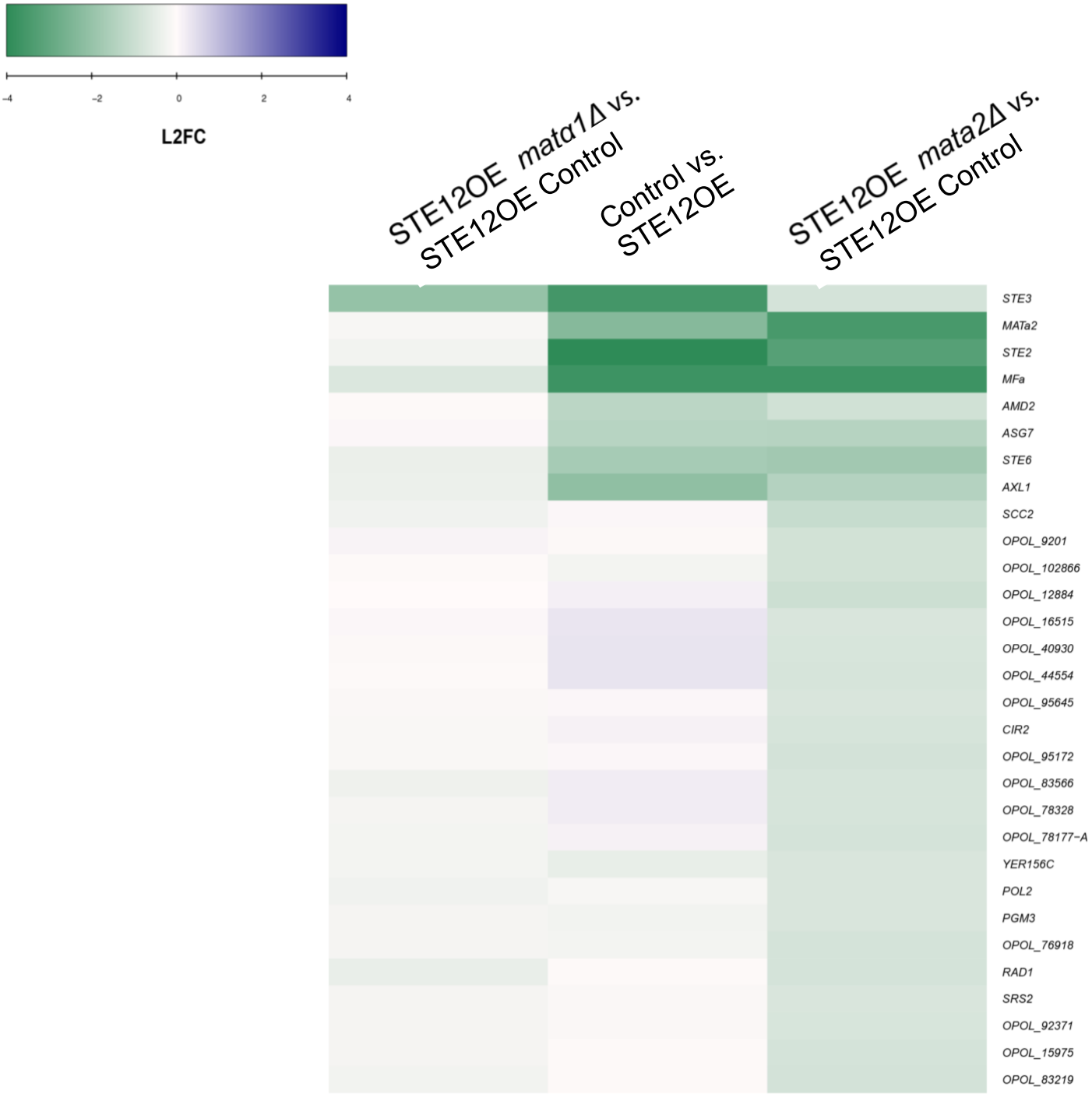


Figure 10: Genes Most Positively Regulated by *MATa2*. Top 30 genes sorted according to most negative $\log_2$ fold change (L2FC) in the *MATa2* dataset. Ratio order of strains are indicated above each column where the more negative the $\log_2$ fold change, the higher the expression of the gene in the control strain. The STE12OE column served as a control for our experimental groups since all strains used have the STE12OE construct.

STE12 most positively regulates genes regulating pheromone production and response pathways

All of the deletion strains and their controls had the PAOX-*STE12* construct so that the mating response could be induced experimentally. Due to this experimental design, our comparisons were comprised of the gene expression of *MAT* deletion STE12OE strains to control STE12OE strains. Ste12 not only regulates mating-type specific genes alongside the mating-type transcription factors but it also can act as a homodimer that regulates haploid-specific genes (**Figure 4**.) In comparing the STE12OE gene expression to cells with no construct we could address which genes need only Ste12 for regulation. These genes are the haploid-specific genes within *O. polymorpha* which are common to both a and alpha-type cells and do not require mating-type transcription factors for their expression (control dataset from Hanson et. al 2017.)

The way the ratios were set up are opposite to that of the *MATα1/MATa2* datasets in that the control strain is in the numerator and the experimental condition, *STE12* over expression in this case, is in the denominator. Using these ratios, any genes positively regulated by *STE12* over expression are indicated by a negative $\log_2$ (FC,) this way, interpreting the sign of the $\log_2$ (FC) value of any sample indicates a similar form of regulation.

When compared to the control strain, the *STE12* overexpression (STE12OE) strain overexpressed many of the genes also found to be positively regulated by *MATα1*and *MATa2* (**Figure 11**.) Among these were *MFalpha* and *MFa* pheromones (-8.37 $\log_2$ (FC,) -5.51 $\log_2$ (FC),) *STE2* alpha-receptor and *STE3* a-receptor (-5.98 $\log_2$ (FC,) -5.31 $\log_2$ (FC)) and MAT locus transcription factors *MATa1* and *MATa2* (-3.32 $\log_2$ (FC,) -3.44 $\log_2$ (FC)) (Jenness et. al 1983, Hagen et. al 1986.) *STE12* also positively regulates genes relating to chemosensitivity such

as *BAR1* which cleaves alpha-factor, *AXL1* which cleaves a-factor, and *SST2* a GTP-ase activating protein for *GPA1* which also regulates desensitization to alpha-factor (-6.02 $\log_2$ (FC,) -3.15 $\log_2$ (FC,) and -3.83 $\log_2$ (FC)) (Sprague and Herskowitz 1981, Michaelis and Barrowman 2012, Steden et.al 1989

The haploid-specific genes regulated by *STE12* were typically involved in the pheromone/mating response pathways. The *STE12* regulated proteins related to the pheromone response pathway included: *FUS3* MAP kinase for G-protein signaling (-5.76 $\log_2$ (FC)) and G-protein subunits: *GPA1* alpha-subunit *STE4* beta-subunit, and *STE18* gamma-subunit (-5.48 $\log_2$ (FC,) -2.59 $\log_2$ (FC,) -3.00 $\log_2$ (FC)) (Neiman and Herskowitz 1994, Whiteway et. al 1989, Miyajima et. al 1987.).) *STE12* regulated genes involved more directly in the mating response included: *DSE1* daughter-cell specific gene and *RIM4* required for expression of sporulation genes (-2.85 $\log_2$ (FC,) -3.29 $\log_2$ (FC)) (Colman-Lerner et. al 2001, Su and Mitchell 1993.)

Overall, the *STE12*-regulated haploid specific genes function within the general mating/pheromone response pathway. Due to the fact that *STE12* is expressed in both mating-types, it is rather surprising that some genes that would be assumed to be mating-type specific, such as the *BAR1* and *AXL1* proteases that help orient the cell's shmoo by cleaving pheromones and thus sensitizing the cell to the pheromone-gradient, can be regulated solely by *STE12* overexpression (Barkai et. al 1998, Bardwell 2005.)

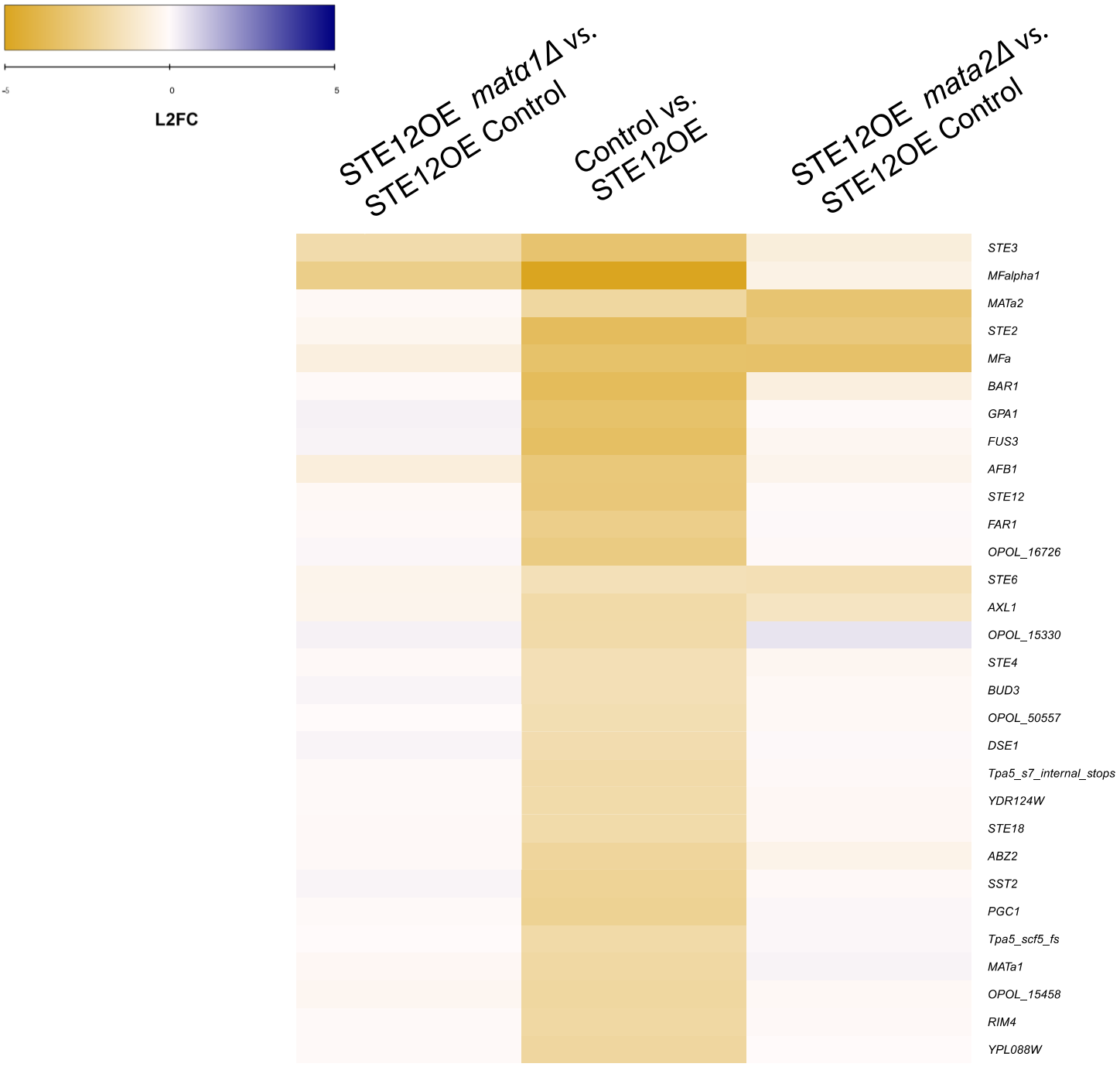


Figure 11: Genes Most Positively Regulated by *STE12*. Top 30 genes sorted according to most negative $\log_2$ fold change (L2FC) in the *STE12* dataset. Ratio order of strains are indicated above each column where the more negative the $\log_2$ fold change, the higher the expression of the gene in the control strain. The STE12OE column served as a control for our experimental groups since all strains used have the STE12OE construct.

Discussion

As their name suggests, mating types are defined by their mating behavior. In this paradigm, a-type cells only express MFa pheromone and Ste2 and alpha-type cells only express MFalpha pheromone and Ste3 (Hanson and Wolfe 2017.) This pattern of expression is canon among yeast where haploid specific genes are primarily involved in the pheromone response pathway and mating-type specific genes are often pheromones and their receptors (Yamamoto et. al 2017.) As such, it is unsurprising that the genes most positively regulated by the MAT transcription factors were involved in the pheromone pathway. However, it is noteworthy that rather than discrete expression, both mating types expressed both pheromones and receptors but in different proportions (**Figure 12**.)

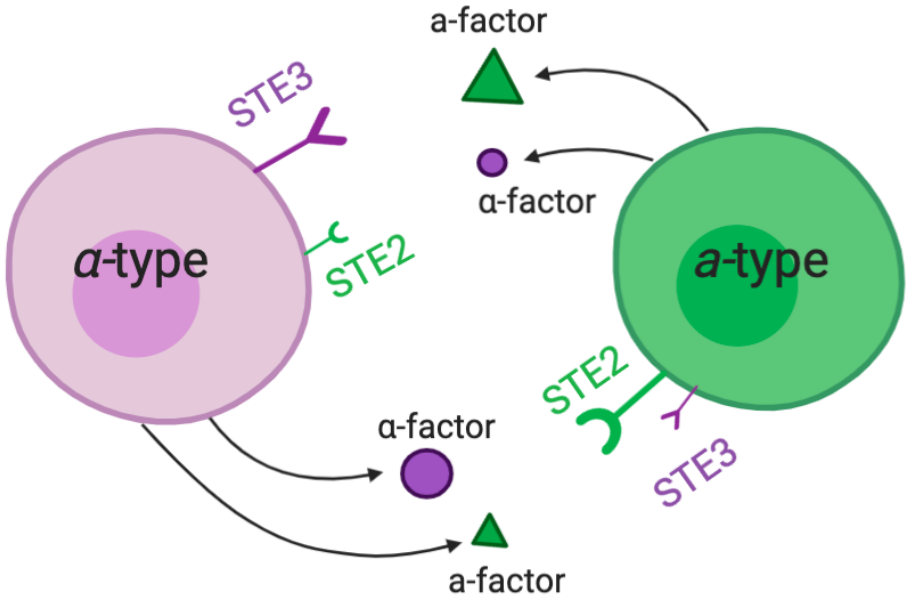


Figure 12: Both A and Alpha-Factor Pheromones and Receptors are Produced by Both Mating Types. Graphic showing scheme made from RNA-seq data that demonstrated the regulatory targets of *MATa2* and *MATa1*.

While both mating types expressed the expected receptors, their own pheromones, and even the opposite mating type's pheromones similarly, they differentially expressed the receptors

for their own pheromones. For example, in alpha-type cells, deletion of *MATα1* yielded a $\log_2$ FC value of -.0.37 for *STE2* alpha-factor receptor while deletion of *MATa2* yielded a $\log_2$ FC value of -1.19 for *STE3* a-factor receptor. As indicated by the negative $\log_2$ FC values, the deletion of the *MAT* transcription factor caused its expression to decrease, meaning that the transcription factors positively regulated the expression of the receptors for its own pheromone.

This pattern of expression was unexpected because in order to have a proper mating response it is important for cells to only be stimulated by the opposite mating type. This specificity is necessitated by the immense amount of energy needed to shmoo towards a mating partner and because the kinase cascade activated by the pheromone receptors is the same regardless of the pheromone that prompted it (Bardwell 2005, Madhani 2007.) Especially in the context under which *O. polymorpha* mates, a stressful nitrogen-depleted environment, mating is a self-protective event that ends in ascus-protected spores that only germinate in more suitable conditions, so the expression of both cell types' pheromones and receptors seems extremely risky (Yamamoto et. al 2017, Madhani 2007.) This non-discrete expression of pheromones and receptors may indicate a more graduated form of pheromone response where a certain concentration is needed before the cell commits to mating. This concept of chemotropic-gradient sensing is not unheard of in yeast where cells actively secrete proteases in order to better sense the gradient surrounding them (Barkai et. al 1998.)

Alternatively, it is also possible that these genes can be activated in the absence of their *MAT* transcription factor but are upregulated in their presence. Especially since the cells are in conditions where *STE12* is dramatically overexpressed, it is possible that *STE12* alone could have activated their transcription. While these findings are extremely interesting, they are limited in that they are based on solely transcriptional data and as such we cannot be sure that these

transcripts were translated into functional proteins. This is especially true of the *STE2/STE3* pheromone receptors which are known to be expressed at low levels, even in rich media (Maekawa and Kaneko 2014.)

In fact, regulating pheromone expression post-transcriptionally is not unheard of within yeast as there is an example in *S. cerevisiae* where *STE2* transcripts are produced in both mating-types but is ultimately only translated in a-type cells (Di Segni et. al 2011.) In *S. cerevisiae*, *STE2* is only actively transcribed in a-type cells so the upstream, cryptic poly-A site is skipped and the canonical 3' UTR poly-A site is poly-adenylated. Alternatively, in alpha-type cells, *STE2* transcripts are rarely transcribed so the process is slower, and the false, upstream poly-A site is used resulting in an aborted transcript. This mode of *STE2* regulation is drastically different than that of *STE3* which is only expressed in alpha-type cells and must be induced since alpha-specific genes are not constitutively active like a-specific genes (Di Segni et. al 2011.) Since the data analyzed in my study rely exclusively on transcriptional data, it is possible that *O. polymorpha* has a similarly complex system for ensuring mating-type genes are expressed specifically in their proper cell type. This is especially true since *O. polymorpha* has a more complex system of activating mating-type genes, requiring *MATa2* for activation of a-specific genes (Hanson and Wolfe 2017.) This possibility is relatively unlikely because the distribution of reads within the datasets would hopefully indicate the presence of such abrogated transcripts, but this can also easily be explored by searching for the presence of a cryptic poly-A site within the gene in question.

Kluyveromyces lactis, like *O. polymorpha,* is haplontic and also requires the activation of a-specific genes through *STE12-MATa2* interactions at a-specific gene promoters, this method of cell-type regulation is thought to be the ancestral state since *MATa2* was lost after a yeast whole

genome duplication event occurred (**Figure 13**; Sorrells et. al 2015.) Despite this similarity, *K. lactis* differs dramatically in terms of its pheromone receptors. While its receptors Ste2 and Ste3 are homologs of S. cerevisiae Ste2 alpha-receptor and Ste3 a-receptor, they are functionally opposite (Torres-Quiroz 2006.) In *K. lactis*, Ste2 is expressed in alpha-type cells and is an a-pheromone receptor and Ste3 is expressed in a-type cells and is an alpha pheromone receptor. Similar to *O. poymorpha*, deletion of the pheromone receptors in *K. lactis* completely abolish the cell's ability to mate (Torres-Quiroz 2006, Maekawa and Kaneko 2014.) This is interesting because even though these yeasts share immense sequence homology, their regulatory patterns can be dramatically different.

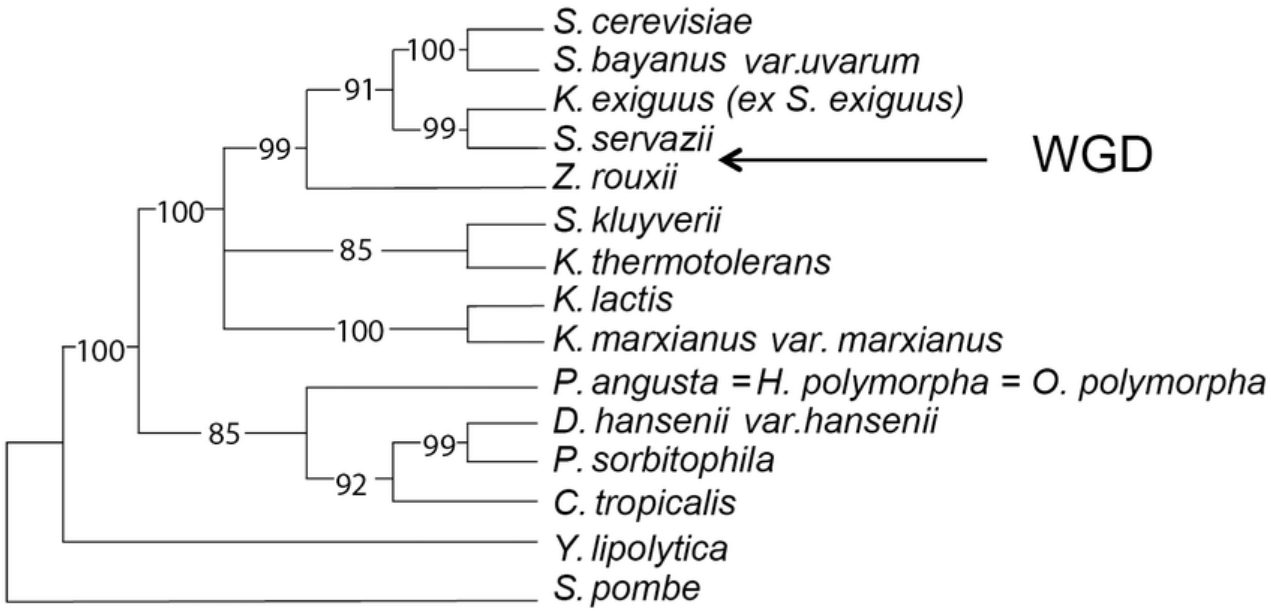


Figure 13: Evolutionary Relationships among Budding Yeast Species. Phylogenetic tree showing the timing of the whole genome duplication (WGD) event in relation to evolution of various species of yeast (Petryk et. al 2014.)

In *Schizosaccharomyces pombe*, the *MAT* locus looks very similar to that of *S. cerevisiae* but it operates through completely different mechanisms because they acquired unicellularity independently (**Figure 13**; Hanson and Wolfe 2017.) Instead of a and alpha cells, *S. pombe*

denotes cell types as minus (M) and plus (P.) *MATa2* and *MATα1* are not used at all and instead it is *MAP1-MAP4* which specify P-specific genes and *MAM1-MAM4* which specific M-specific genes (Yabana and Yamamoto 1996.) Almost all of the mating-type specific genes in *S. pombe* are related to pheromone production and receptors for P or M-factor (similar to alpha and a-factor respectively.)

Considering the fact that most yeast mating-types use pheromones, many of them even using variants of the canonical a and alpha-factor, reflects their importance in the yeast life cycle. Due to its large role in defining mating types and thus affecting sexual barriers between species, there has been much research done on pheromone evolution within yeast. A common way to study these interactions is by seeing if a cell will respond to the pheromone of another yeast species. Interestingly enough, there is a difference in how cells responded depending on whether they were exposed to a-factor or alpha-factor. While a-factor has an "all or nothing" response, alpha-factor had a more graduated response in terms of mating-efficiency (Rogers et. al 2015.)

This pattern of asymmetric pheromone usage was also seen in *S. pombe* when studying the ability of *S. octosporus* P and M-factor pheromones to cause a response in *S. pombe* receptors. They found that while *S. octosporus* P-factor was able to elicit a receptor response in *S. pombe*, M-factor could not. This along with the fact that M-factor was conserved within *S. pombe* strains and P-factor is extremely divergent gives evidence to the hypothesis that these pheromones diverged asymmetrically because they are used differently between mating-types. They pose the hypothesis that M-factor (akin to a-factor) defines the species while P-factor allows "flexible adaptation" and can be helpful in prezygotic isolation. Since P-factor is hydrophilic as opposed to the hydrophobic M-factor, it can diffuse more easily in aqueous

environments and can therefore transmit the mating-signal farther than M-factor can (Seike et. al 2019.)

In light of these studies, my data is very interesting because there appears to be a similar pattern in pheromone use by *O. polymorpha*. In *O. polymorpha*, the MAT transcription factors are upregulating the expression of pheromones similarly, but the regulation of the receptors differs more significantly. While *STE3* is upregulated 2-fold by *MATa2, STE2* is only upregulated 1.29-fold by *MATa1* which suggests that being able to identify and respond to a-factor may be more important than just mating-partner identification in alpha-type cells. This is a hypothesis that should be studied further in *O. polymorpha* because understanding how cell-types evolve can help us further understand how so many species can identify their proper mates despite using much of the same machinery to mate.

Understanding how yeast species differentiate their sexual identities is essential in understanding how they evolved because oftentimes it is prezygotic barriers such as mutations in an individual's ability to recognize its own species as a mating partner that causes the formation sexual barriers and thus divergence between species (Seike et. al 2019.) Comparative analyses of the polypeptide sequences of both a and α-factor within the *O. polymorpha* species as well as comparisons within and between close and distant evolutionary relatives should be done to explore the extent of this asymmetrical pattern. I hypothesize that in most yeast species, there will be a shared asymmetry in preference for a-factor recognition in both mating-types and that the α-factor will be more flexible in its ability to stray from the canonical sequence and still be recognized.

Methods

Cell Culture

The PCZ4 plasmid was grown in LB+AMP media and *O. polymorpha* strains (SHY202-1, SHY202-2, and SHY202-3) were grown in YPD overnight at 37°C with shaking for use in the fusion construct. To induce cells to express *STE12* the *O. polymorpha* cells were first grown on Mineral Media+Glucose (MMG) overnight. To grow the cultures to the log phase, they were diluted to an OD of 0.2 in fresh MMG with a total volume of 10mL (after an overnight incubation of at least 10 hours) and left to incubate with shaking at 37°C, the samples were then diluted every 6 hours or when the OD reached 1.5. In total, two dilutions were done in MMG and then the final dilution was done in MMM and left to incubate at 37°C with shaking overnight. A 1mL sample of each strain was collected after the last MMG incubation and after the final MMM incubation, these samples were pelleted and stored at -20°C to be used for DNA extraction. The rest of the MMM culture was used for RNA extractions.

PCR and Gel Electrophoresis

To check the orientation of the MAT locus, *O. polymorpha* DNA was PCR amplified with four different primer combinations, two for each mating-type orientation (Figure 2). The Master Mixes (1-4) were made with 1X Dreamtaq Master Mix, $4x10^{-7}$ M of each of the appropriate primers, 100ng of our DNA samples, with the rest of the solution being nuclease-free water. The PCR settings were as follows: initial denaturation at 95°C for 2min, a 25 cycle loop consisting of: further denaturation at 95°C for 30s, elongation at 57°C for 30s, and annealing at 72°C for 2min50s, and a final, single annealing cycle at 72°C for 5min, after which the samples were held indefinitely at 4°C. PCR products were then analyzed on a 1% w/v agarose gel with

1X GelRed stain by electrophoresis at 120V. All other DNA was visualized with the same settings.

Creation of the Deletion Construct

Once construct fragments of the proper size were obtained, they were fused using PCR at an annealing temperature of 65°C and DNA ratios of 5':NAT:3' of 1:2:1 or 2:1:2. The PCR mix contained Q5 DNA polymerase, 25X Q5 buffer, and $2.5x10^{-4}$ mM dNTPs. Used the following PCR settings: initial denaturation for one 30s cycle at 98°C, then for 29 cycles the sample was denatured again at 98°C for 10s, annealed at 65°C for 30s, and elongated at 72°C for 1.5min, the final elongation step occurred for 10min at 72°C, after which, the PCR products were held infinitely at 4°C and checked via electrophoresis to ensure the proper size construct (3kb) was obtained.

Transformation of *O. polymorpha* cells

The transformations were done using an electroporator taken from the procedure in Faber et al. 1994 (as *Hansenula polymorpha*.) Using the SHY202-1 strain that was previously quality-checked, the cells were treated with either no DNA as a negative control, PHIPN4 plasmid that contains NAT resistance as a positive control, and the fusion. Once the yeast were in solution with their respective treatments, they were electroporated at 1400V, immediately placed in fresh YPD, and incubated for an hour at 37°C in a shaking incubator to help them recover. Afterwards, the transformations were plated on YPD+NAT and left in a 37°C incubator for two days. Seven transformants were obtained and tested for quality (confirmed the presence of the NAT gene,

alpha-1 deletion, and alpha-orientation.) These strains were labelled as KSY1-KSY7 and stocks were created and stored in the -80°C freezer.

DNA Extraction

Plasmid DNA was collected using a Zyppy Plasmid miniprep kit and *O. polymorpha* genomic DNA was extracted using phenol:chloroform and analyzed by PCR and gel electrophoresis. *O. polymorpha* cell samples (stored at -20°C) were thawed and resuspended in lysis buffer and 25:24:1 phenol:chloroform:isoamyl solution which was added in a 1:1 ratio. About 300mg of acid washed glass beads were added to the samples and vortexed for 8min. Samples were then treated with TE (10mM Tris-HCl and 1mM EDTA,) mixed gently, and spun in a microcentrifuge for 5min after which the aqueous layer was transferred to a new 1.5mL tube. To precipitate the DNA, the samples were mixed with 100% ethanol, centrifuged for 5 min, and then the supernatant was discarded. This step was repeated with 70% ethanol. The pellets were left to dry for 10min and then resuspended in nuclease-free water and pipetted to mix. A Nanodrop was used to determine concentration of DNA in the plasmid and *O. polymorpha* samples. DNA samples were then stored at -20°C.

RNA Extraction and Purification

MMM incubated *O. polymorpha* cell cultures were pelleted and resuspended in TES (10mM Tris-Cl, pH 7.5 10mM EDTA, and0.5% SDS.) An acidic phenol:chloroform:isoamyl alcohol solution was added in a 1:1 ratio and then incubated on a 65°C heat block with vortexing 10 seconds every 10 minutes for an hour. Tubes were then placed on ice for 5 minutes and centrifuged for another 5 at 4°C. Aqueous phase was transferred to a clean tube and the RNA

was washed again with acidic phenol:chloroform:isoamyl alcohol, centrifuged at 4°C, and aqueous phase was transferred to a clean tube. This step was repeated with chloroform. Then, 0.05M sodium acetate and cold 100% ethanol were added to a final concentration of 38% ethanol, the sample was inverted to mix, centrifuged for 5min at 4°C, and supernatant was discarded. Pellet was then washed in cold 70% ethanol and centrifuged once more for 5min at 4°C, supernatant was removed, and pellet was allowed to air dry for 10min. The pellet was then resuspended in nuclease-free water and incubated at 65°C for 1min.

Nanodrop was used to calculate RNA concentrations so samples of 10μg could be measured for use in the DNAse digestion. In the digest: 10 units DNAseI per 10 μg RNA with 1X Reaction buffer and diluted with water to a final volume of 100 microliters. After a 15min incubation at room temperature, 25Mm EDTA Stop solution was added and the solution was incubated at 70°C for 10min. Zymo Research RNA Clean and Clear concentrator kit was used to purify the RNA. The sample was transferred to a Zymo spin column and collection tube and RNA prep buffer and 100% ethanol was added in a 1:2:3 fashion, centrifuged for 30s and discarded flow-through. Sample was washed with RNA binding buffer and centrifuged for 30s, the flow-through was then discarded. This step was repeated twice more with RNA wash buffer, with the last wash being centrifuged for 2min. The spin column was transferred to a clean 1.5mL tube, RNAse-free water was added, and the sample was centrifuged for 30s. The sample concentrations were then measured using a Qubit RNA Broad Range kit.

RNA Sequencing

RNA samples were sent to the University of Colorado Anschutz Genomic Microarray Facility where they were converted into cDNA libraries and sequenced using Illumina ribo-

depleted total-RNA sequencing. The FASTA files obtained were then analyzed using the usegalaxy.org server at Penn State University.

FASTA Adaptor Trimming and Quality Checks

The FASTQC tool version 0.11.9 (Andrews 2019) was used to ensure sequencing quality was comparable between the reads as well as between the different lanes. Ultimately, this step was incorporated into the Trim Galore! Version 0.6.3 (Kreuger 2019) workflow where the Illumina adaptors were trimmed off the reads and a Phred quality score of 20 (99% base call accuracy) and above was required. The following settings were used: paired end reads, changed overlap with adaptor sequence to trim a sequence to a more stringent 2bp rather than 1bp default. FASTQC (Andrews 2019) was used again after the trimming step to confirm sequences were adapter free and of quality.

Read Feature Mapping and Counting

The HISAT2 2.1.0 (Kim et. al 2015) tool was used to pair FASTA reads to the *O. polymorpha* genome. The settings were as follows: paired end, strand information is -fr (this option is essential for correct mapping, -fr option means the HTSEQ-COUNT setting must be stranded yes, using the -rf option means the stranded option must be set to reverse later on) and printed the alignment summary to a file. The alignment summary was analyzed to check for an overall alignment rate of 90% and above.

In order to count the annotated reads, HTSEQ-COUNT Galaxy Version 0.9.1 (Anders et. al 2014) (tool was used on the BAM files from the HISAT2 output. The settings were as follows:

uploaded a GFF file for *O. polymorpha*, union mode, minimum alignment quality of 10, stranded to yes, feature type to CDS, and ID attribute to ID.

Differential Gene Expression Analysis

DESEQ2 Galaxy Version 2.11.40.6 (Love et. al 2014) tool was used to compare the gene expression between control and experimental conditions. The first factor (the numerator in the comparison) was always the deletion strain and the control strain was always the second factor (the denominator,) so that it could be normalized to. The default settings were used, only added that there were no headers were present and also to output a normalized counts table. The plots output from this tool were compared between lanes before they were ultimately joined for future analysis. Only one sample, KSY-5, was deleted from the pool because it failed to group with the other deletion strains during variance analysis.

The lanes were combined using the column join tool to add up the counts for each annotated feature and the cut tool was used to get rid of the individual lane count columns. Genes with the highest difference in expression (log2 fold change) were compiled into a list that was then used to create a heatmap using an R-script that used tools in the gplots package version 3.0.1.1 (Warnes et. al 2020, Gregory et. al 2020) and the RcolorBrewer package version 1.1-2 (Neuwirth 2014.) The r-script used can be found in the appendix and was implemented using RStudio (RStudioTeam 2015.)

Works Cited

Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq — A Python framework to work with high-throughput sequencing data Bioinformatics, in print, online at doi:10.1093/bioinformatics/btu638

Andrews, S. (2019). FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Accessed August 2019.

Bardwell, L. (2005). A walk-through of the yeast mating pheromone response pathway. Peptides *2*, 339-350.

Barkai, N., Rose, M. D. and Wingreen, N. S. (1998). Protease helps yeast find mating partners. Nature *6710*, 422-423.

Bender, A. and Sprague, G. F. (1986). Yeast peptide pheromones, a-factor and α-factor, activate a common response mechanism in their target cells. Cell *6*, 929-937.

Booth, L. N., Tuch, B. B. and Johnson, A. D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. Nature *7326*, 959-963.

Brivanlou, A. H. and Darnell, J. E. (2002). Signal Transduction and the Control of Gene Expression. Science *5556*, 813.

Christ, C. and Tye, B. K. (1991). Functional domains of the yeast transcription/replication factor MCM1. Genes & Development *5*, 751-763.

Colman-Lerner, A., Chin, T. E. and Brent, R. (2001). Yeast Cbk1 and Mob2 Activate Daughter-Specific Genetic Programs to Induce Asymmetric Cell Fates. Cell *6*, 739-750.

Di Segni, G., Gastaldi, S., Zamboni, M. and Tocchini-Valentini, G. P. (2011). Yeast pheromone receptor genes *STE2* and *STE3* are differently regulated at the transcription and polyadenylation level. Proc. Natl. Acad. Sci. USA *41*, 17082-17086.

Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables (2020). gplots: Various R Programming Tools for Plotting Data.

Hagen, D. C., McCaffrey, G. and Sprague, G. F. (1986). Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone a factor: gene sequence and implications for the structure of the presumed receptor. Proc. Natl. Acad. Sci. USA *5*, 1418.

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 *11*, e1007092.

Hanson, S. J. and Wolfe, K. H. (2017). An Evolutionary Perspective on Yeast Mating-Type Switching. Genetics *1*, 9-32.

Hanson, S. J., Byrne, K. P. and Wolfe, K. H. (2014). Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. Proc. Natl. Acad. Sci. USA *45*, E4851.

Hegde, V. and Klein, H. (2000). Requirement for the SRS2 DNA helicase gene in non-homologous end joining in yeast. Nucleic Acids Res. *14*, 2779-2783.

Hobert, O. (2008). Gene Regulation by Transcription Factors and MicroRNAs. Science *5871*, 1785.

Huang, L., Yuan, Z., Liu, P. and Zhou, T. (2015). Effects of promoter leakage on dynamics of gene expression. BMC Systems Biology *1*, 16.

Huberman, L. B. and Murray, A. W. (2013). Genetically Engineered Transvestites Reveal Novel Mating Genes in Budding Yeast. Genetics *4*, 1277-1290.

Jenness, D. D., Burkholder, A. C. and Hartwell, L. H. (1983). Binding of α-factor pheromone to

yeast a cells: Chemical and genetic evidence for an α-factor receptor. Cell *2, Part 1*, 521-529.

Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods *4*, 357-360.

Kreuger, F. (2019). Trim Galore! Version 0.6.3 UseGalaxy tool, Barbraham Bioinformatics Institute. https://github.com/FelixKrueger/TrimGalore Accessed August 2019.

Latchman, D. S. (1997). Transcription factors: An overview. The International Journal of Biochemistry & Cell Biology *12*, 1305-1312.

Lee, S. C., Ni, M., Li, W., Shertz, C. and Heitman, J. (2010). The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiol. Mol. Biol. Rev. *2*, 298.

Leibowitz, M. J. and Wickner, R. B. (1976). A chromosomal gene required for killer plasmid expression, mating, and spore maturation in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *6*, 2061-2065.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *12*, 550.

Lucarelli, P. *et al.* (2018). Resolving the Combinatorial Complexity of Smad Protein Complex Formation and Its Link to Gene Expression. Cell Systems *1*, 75-89.e11.

Mackay, V. and Manney, T. R. (1974). MUTATIONS AFFECTING SEXUAL CONJUGATION AND RELATED PROCESSES IN *SACCHAROMYCES CEREVISIAE*. II. GENETIC ANALYSIS OF NONMATING MUTANTS. Genetics *2*, 273-288.

Madhani, H. (2007). From a to alpha: Yeast as a Model for Cellular Differentiation: Cold Spring Harbor Laboratory Press.

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 *11*, e1004796

Michaelis, S. and Barrowman, J. (2012). Biogenesis of the Saccharomyces cerevisiae Pheromone **a**-Factor, from Yeast Mating to Human Disease. Microbiol. Mol. Biol. Rev. *3*, 626.

Michaelis, C., Ciosk, R. and Nasmyth, K. (1997). Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. Cell *1*, 35-45.

Miyajima, I., Nakafuku, M., Nakayama, N., Brenner, C., Miyajima, A., Kaibuchi, K., Arai, K., Kaziro, Y. and Matsumoto, K. (1987). GPA1, a haploid-specific essential gene, encodes a yeast homolog of mammalian G protein which may be involved in mating factor signal transduction. Cell *7*, 1011-1019.

Neiman, A. M. and Herskowitz, I. (1994). Reconstitution of a yeast protein kinase cascade in vitro: activation of the yeast MEK homologue STE7 by STE11. Proc. Natl. Acad. Sci. U. S. A. *8*, 3398-3402.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes.

Palladino, F. and Klein, H. L. (1992). Analysis of mitotic and meiotic defects in Saccharomyces cerevisiae SRS2 DNA helicase mutants. Genetics *1*, 23-37.

Petryk, N. *et al.* (2014). Functional Study of the Hap4-Like Genes Suggests That the Key Regulators of Carbon Metabolism HAP4 and Oxidative Stress Response YAP1 in Yeast Diverged from a Common Ancestor. PloS one, e112263.

Rogers, D., Denton, J., McConnell, E. and Greig, D. (2015). Experimental Evolution of Species Recognition. Current Biology *13*, 1753-1758.

Rong, L., Palladino, F., Aguilera, A. and Klein, H. L. (1991). The hyper-gene conversion hpr5-1 mutation of Saccharomyces cerevisiae is an allele of the SRS2/RADH gene. Genetics *1*, 75-85.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/. Accessed January 2020.

Seike, T., Shimoda, C. and Niki, H. (2019). Asymmetric diversification of mating pheromones in fission yeast. PLoS biology *1*, e3000101.

Sorrells, T. R., Booth, L. N., Tuch, B. B. and Johnson, A. D. (2015). Intersecting transcription networks constrain gene regulatory evolution. Nature *7560*, 361-365.

Sprague, G. F. and Herskowitz, I. (1981). Control of yeast cell type by the mating type locus: I. Identification and control of expression of the a-specific gene BAR1. Journal of Molecular Biology *2*, 305-321.

Steden, M., Betz, R. and Duntze, W. (1989). Isolation and characterization of Saccharomyces cerevisiae mutants supersensitive to G1 arrest by the mating hormone a-factor. Molecular and General Genetics MGG *3*, 439-444.

Su, S. S. and Mitchell, A. P. (1993). Identification of functionally related genes that stimulate early meiotic gene expression in yeast. Genetics *1*, 67-77.

Torres-Quiroz, F., Kawasaki, L., Rodríguez-González, M., Patrón-Soberano, A. and Coria, R. (2006). The KlSTE2 and KlSTE3 genes encode MATα- and MATa-specific G-protein-coupled receptors, respectively, which are required for mating of Kluyveromyces lactis haploid cells. Yeast *1*, 17-25.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T.,

Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables B. (2020). gplots: Various R Programming Tools for Plotting Data.

Whiteway, M., Hougan, L., Dignard, D., Thomas, D. Y., Bell, L., Saari, G. C., Grant, F. J., O'Hara, P. and MacKay, V. L. (1989). The STE4 and STE18 genes of yeast encode potential β and γ subunits of the mating factor receptor-coupled G protein. Cell *3*, 467-477.

Wilson, K. L. and Herskowitz, I. (1984). Negative regulation of STE6 gene expression by the alpha 2 product of Saccharomyces cerevisiae. Mol. Cell. Biol. *11*, 2420.

Wong, R. W. (2010). An update on cohesin function as a 'molecular glue' on chromosomes and spindles. Cell Cycle *9*, 1754-1758.

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 *1*, 16318-12.

Yabana, N. and Yamamoto, M. (1996). Schizosaccharomyces pombe map1+ encodes a MADS-box-family protein required for cell-type-specific gene expression. Mol. Cell. Biol. *7*, 3420-3428.

Zheng, R. and Blobel, G. A. (2011). GATA Transcription Factors and Cancer. Genes & Cancer *12*, 1178-1188.

Zhong, H., McCord, R. and Vershon, A. K. (1999). Identification of Target Sites of the α2–Mcm1 Repressor Complex in the Yeast Genome. Genome Res. *11*, 1040-1047.

Appendix

Table 1: Primer Names and Sequences

Primer Name	Sequence 5'-3'
HpolMATa2	CCACTCATGGGAAATGATCCG
HpolMATb1	GAGTCATGGGGTCTGGTTTG
HpolMATb2	CTGCATGATATGACTACCAGCC
HpolMATc1	CTCAGATGATCCCACCACTAGG
HpolMATd1	CTGCGTCAGCTCAGGAATC
NAT/HPH-379F	AGCTTGCCTCGTCCCCG
NAT+877R	TCGATTACAACAGGTGTTG
OPMATALPHA1P1	ACGATAGATCCGCGGTACC
OPMATALPHA1P3NAT_2	CGGGGACGAGGCAAGCTTCTTCGGTGAATCATTATGAGGAG
OPMATALPHA1P4NAT	CAACACCTGTTGTAATCGAGAGGATGTATTTCAGGAATGCAGT
OPMATALPHA1P6	CGGCCTCAAAACAAGTTGC

Table 2: Strains and Plasmids Used

Species	Strain Name	Background	MAT locus orientation	Genotype	Source
Ogataea polymorpha	SHY202-1	NCYC495	A	*MATa2 ade11 met6 ku80::ZEO pAOX-STE12*	This study
Ogataea polymorpha	SHY202-2	NCYC495	Alpha	*MATα1 ade11 met6 ku80::ZEO pAOX-STE12*	This study
Ogataea polymorpha	SHY202-3	NCYC495	A	*MATa2 ade11 met6 ku80::ZEO pAOX-STE12*	This study
Ogataea polymorpha	SHY202-1S2	NCYC495	Alpha	*MATα1 ade11 met6 ku80::ZEO pAOX-STE12*	This study
Ogataea polymorpha	SHY202-2S1	NCYC495	Alpha	*MATα1 ade11 met6 ku80::ZEO pAOX-STE12*	This study
Ogataea polymorpha	KSY1	NCYC495	Alpha	*ade11 met6 ku80::ZEO matα1Δ::NAT pAOX-STE12*	This study
Ogataea polymorpha	KSY5	NCYC495	Alpha	*ade11 met6 ku80::ZEO matα1Δ::NAT pAOX-STE12*	This study
Ogataea polymorpha	KSY6	NCYC495	Alpha	*ade11 met6 ku80::ZEO matα1Δ::NAT pAOX-STE12*	This study
Ogataea polymorpha	CMY3-1	NCYC495	A	*ade11 met6 ku80::ZEO mata2Δ::NAT pAOX-STE12*	This study
Ogataea polymorpha	CMY4	NCYC495	A	*ade11 met6 ku80::ZEO mata2Δ::NAT pAOX-STE12*	This study
Ogataea polymorpha	JOY1-1	NCYC495	A	*ade11 met6 ku80::ZEO mata2Δ::NAT pAOX-STE12*	This study
Escherichia coli	pHIPN4	DH5alpha	-	pHIP containing nourseothricin marker, ampR	This study
Escherichia coli	pZC4	DH5alpha	-		Carter Z., Faculty of Life Sciences, Univeristy of Manchester, Oxford Road, Manchester, M13 9PT, UNITED KINGDOM.

Table 3: Alpha-Specific Genes of Interest Represented in Heatmap (Figure 8)

Scer ortho	Opol gene name	Basemean	L2FC	Std error	Wald-stats	p-value	adj p-value
MFalpha1	MFalpha_pheromone	385.9701526	4.32420283	0.226962576	19.05249273	6.26E-81	1.63E-77
STE3	OPOL_17446_STE3	259.8138413	2.99366484	0.226973494	13.1894909	1.01E-39	1.31E-36
MATalpha1	MATalpha1	106.9860333	2.3452428	0.239341686	9.798722653	1.14E-22	NA
AFB1	OPOL_77997	123.0963935	1.10787657	0.2394043	4.627638541	3.70E-06	NA
MFa	OPOL_MFa_pheromone	725.183996	1.00631176	0.177503518	5.66924969	1.43E-08	7.48E-06
CYB2	OPOL_17558	195.016069	0.98093367	0.232051597	4.22722222	2.37E-05	0.004404028
BNI1	OPOL_47977	260.9511112	0.95721096	0.218169704	4.387460495	1.15E-05	0.003038484
PRP28	OPOL_100787	87.24724922	0.76812662	0.239088948	3.212723252	0.00131483	NA
SKI8	OPOL_48400	114.7806835	0.75123941	0.238559339	3.149067296	0.00163792	NA
RPS20	OPOL_9037	466.1496953	0.74343551	0.192516829	3.861665065	0.00011262	0.01467398
NAR1	OPOL_16325	92.92997481	0.696437	0.23942987	2.908730622	0.00362899	NA
MEF1	OPOL_16320	110.2477865	0.69261458	0.238912229	2.89903361	0.00374315	NA
FMO1	OPOL_16654	61.70943243	0.69043123	0.233614373	2.955431273	0.00312232	NA
RPO31	OPOL_17249	149.9421385	0.68608362	0.235946769	2.907789849	0.00363993	NA
SEC11	OPOL_98907	110.4934482	0.65608518	0.238655397	2.749090071	0.0059761	NA
DUR1,2	OPOL_93142	44.59102088	0.65320114	0.222821452	2.931500264	0.00337329	NA
.	OPOL_7744	39.74410012	0.65270737	0.220975218	2.953758222	0.0031393	NA
MDN1	OPOL_95282	293.6494249	0.6499668	0.215701428	3.013270732	0.00258448	0.098601032
VAN1	OPOL_86755	88.19884228	0.64598745	0.238566381	2.707789109	0.0067733	NA
.	OPOL_16478	179.579323	0.62792723	0.23043005	2.725023175	0.0064297	NA
.	OPOL_101292	62.75965457	0.62584768	0.234276316	2.67140824	0.00755337	NA
.	OPOL_92047	164.3352787	0.61566433	0.231999688	2.653729126	0.00796077	NA
GDT1	OPOL_90462	257.3922698	0.61088529	0.228931618	2.668418185	0.00762093	0.178920299
KEX2	OPOL_47808	163.9708862	0.59859138	0.235445185	2.54238105	0.01101001	NA
.	OPOL_78251	230.527184	0.59590646	0.224320663	2.65649384	0.00789579	0.179471993
CAM1	OPOL_77217	825.5051494	0.59454144	0.191443247	3.1055754	0.00189909	0.081131699
LOA1	OPOL_94649	44.72240051	0.59364416	0.223621545	2.654682325	0.00793831	NA
RAD1	OPOL_36814_RAD1	94.67660638	0.58477186	0.239604188	2.440574469	0.01466392	NA
.	OPOL_94713	175.5470687	0.58153022	0.233414938	2.49140106	0.01272404	NA
.	OPOL_49734	61.21320978	0.5716857	0.235142017	2.431235843	0.01504742	NA

Table 4: A-Specific Genes of Interest Represented in Heatmap (Figure 9)

Scer gene name	Opol gene name	Basemean	L2FC	Standard error	wald-stats	p-value	adj p-value
MFa	OPOL_MFa_pheromone	4163.28646	5.52672746	0.21965612	25.1608167	1.08E-139	5.89E-136
MATa2	MATa2	236.200429	5.1827493	0.28585729	18.130548	1.83E-73	3.34E-70
STE2	OPOL_40804	1811.98679	4.79682984	0.24020426	19.969795	1.01E-88	2.76E-85
STE6	OPOL_88856	962.711792	2.63054781	0.19611616	13.4132128	5.06E-41	6.93E-38
AXL1	OPOL_95797	1348.77609	2.10188344	0.27756587	7.57255739	3.66E-14	4.01E-11
ASG7	OPOL_10419_ASG7	68.7876755	2.07406462	0.31161949	6.65576029	2.82E-11	2.57E-08
SCC2	OPOL_94200	70.2171214	1.58371816	0.30742347	5.15158501	2.58E-07	0.00017687
.	OPOL_12884	82.626976	1.42856292	0.31150467	4.586008	4.52E-06	0.00164998
AMD2	OPOL_44505	62.9771387	1.32576919	0.30777529	4.30758811	1.65E-05	0.00502285
.	OPOL_9201	33.4416761	1.29052352	0.31160451	4.14154308	3.45E-05	0.00853907
.	OPOL_102866	92.8968015	1.27666178	0.30891693	4.13270257	3.59E-05	0.00853907
.	OPOL_95172	37.068442	1.23367219	0.30355739	4.06404922	4.82E-05	0.00979698
.	OPOL_83219	82.1505329	1.22116337	0.31090074	3.9278239	8.57E-05	0.01467384
.	OPOL_15975	73.3223025	1.21972169	0.31171567	3.91293036	9.12E-05	0.01513635
.	OPOL_78177-A	29.2853566	1.21797866	0.29456487	4.13484019	3.55E-05	0.00853907
RAD1	OPOL_36814_RAD1	148.228957	1.21228485	0.29663746	4.08675583	4.37E-05	0.00958534
.	OPOL_76918	91.6829041	1.21129691	0.30016402	4.03544998	5.45E-05	0.01029444
STE3	OPOL_17446_STE3	359.872697	1.19037808	0.29430913	4.0446523	5.24E-05	0.01025185
.	OPOL_83566	125.725937	1.16943269	0.31146003	3.75467982	0.00017356	0.02331721
CIR2	OPOL_16793	44.9881023	1.16542911	0.31050731	3.75330653	0.00017452	0.02331721
.	OPOL_44554	56.9571267	1.15520498	0.31140661	3.70963535	0.00020756	0.02584097
.	OPOL_78328	69.2781694	1.14111912	0.31040712	3.67620146	0.00023673	0.02881825
.	OPOL_92371	89.251924	1.13683745	0.31136144	3.65118256	0.00026104	0.03042453
.	OPOL_40930	64.773029	1.11799831	0.29539586	3.78474601	0.00015387	0.0216122
YER156C	OPOL_16343	161.881499	1.09341105	0.29450444	3.712715	0.00020505	0.02584097
PGM3	OPOL_75847	61.7448618	1.09133077	0.30905995	3.5311297	0.00041379	0.04142211
SRS2	OPOL_75843	47.465493	1.07648766	0.31133444	3.4576569	0.0005449	0.04875352
.	OPOL_16515	28.3927018	1.061555	0.3011484	3.5250229	0.00042345	0.04142211
.	OPOL_95645	60.1494032	1.06136739	0.30915468	3.43312736	0.00059666	0.05028482
POL2	OPOL_85071	121.900203	1.06107724	0.30558898	3.47223661	0.00051614	0.04712369

R-script for Heatmaps

```
#r script for generating heatmaps
#All CAPS must be edited with each use
#opens packages necessary for heatmap
library(RColorBrewer)
library(gplots)

#defines the data as a variable able to be called, columns of file can be chosen using []
NAMEDATA <- FILE [,1:9]

#if not input when file is imported, rows can be defined by calling on the name of the column containing the rows
row.names(NAMEDATA) <- NAMEDATA$ROWNAME

#sometimes the function above makes a new column, the duplicated gene name column can be deleted by doing the following
NAMEDATA[1]<-NULL

#changes the data into a matrix if it is not already in the correct format
NAMEMATRIXDATA <- as.matrix(NAMEDATA[,2:9])

#chooses colors for heatmap, n defines the range of colors 299 is max yet 1000 works too
my_heatmap_palette <- colorRampPalette(c("navy", "snow",
"NEWCOLORFORSAMPLE"))(n =
1000)

#creates an empty PDF file for your heatmap to move onto your desktop from  Rstudio
pdf("~/Desktop/NAME_heatmap.pdf", width
= 8, height = 11, pointsize = 8, family = "sans", bg = "white")

#creates the heatmap, can input the title of the plot and dendrogram option clusters the genes
heatmap.2(NAMEdata_matrix, main = "Title: NAME", density.info =
"none", trace = "none", margins = c(20,10), col=my_heatmap_palette,
dendrogram = "row", Colv = "NA")

#closes the file, allowing the pdf to be viewed
dev.off()
```