12-2015

Genomic and transcriptomic analyses of *Microbotryum lychnidis-dioicae* provide insights into the biology of a fascinating fungal phytopathogen.

Su San Toh

University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Part of the [Other Genetics and Genomics Commons](#)

Recommended Citation

Toh, Su San, "Genomic and transcriptomic analyses of *Microbotryum lychnidis-dioicae* provide insights into the biology of a fascinating fungal phytopathogen." (2015). *Electronic Theses and Dissertations*. Paper 2327.
<https://doi.org/10.18297/etd/2327>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

GENOMIC AND TRANSCRIPTOMIC ANALYSES OF
MICROBOTYRUM LYCHNIDIS-DIOICAE
PROVIDE INSIGHTS INTO THE BIOLOGY OF
A FASCINATING FUNGAL PHYTOPATHOGEN

By

Su San Toh
B.Sc., National University of Singapore, 2003

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

Department of Biology
University of Louisville
Louisville, Kentucky

December 2015

GENOMIC AND TRANSCRIPTOMIC ANALYSES OF
MICROBOTYRUM LYCHNIDIS-DIOICAE
PROVIDE INSIGHTS INTO THE BIOLOGY OF
A FASCINATING FUNGAL PHYTOPATHOGEN

By

Su San Toh
B.Sc., National University of Singapore, 2003

A Dissertation Approved on

November 17, 2015

by the following Dissertation Committee:

Dr. Michael H. Perlin

Dr. David J. Schultz

Dr. Kenneth E. Palmer

Dr. Mark P. Running

Dr. Ted Kalbfleisch

DEDICATION

This dissertation is dedicated to my most beloved husband

Mr. Eugene Quek

for his unconditional love and support through this bumpy journey,

and for taking care of our children,

Zacharie Jayden 予乐 & Zaphanie Joy 予欣

in more ways any husband ever would.

ACKNOWLEDGMENTS

My first acknowledgement goes to the funding sources. Without these, none of this work would have been possible. These funding include NSF Grant No. 0947963 to Dr. Michael Perlin, Dr. David Schultz and Dr. Christina Cuomo, the ANR-09-BLAN-064 and ERC GenomeFun 309403 grants to Dr. Tatiana Giraud, and BIO2012-37161 and NPRP 5-298-3-086 to Dr. Toni Gabaldón.

We received a lot of help in the projects, many through these personal favors. We thank Dr. Michael E. Hood from Amherst College for providing us with *Silene latifolia* seeds and the starter fungal strains. We acknowledge the Broad Institute Sequencing Platform for generating all DNA and RNA sequences described here, and Sinéad Chapman for coordinating the sequencing. We thank Dr. Mark Lawrence from Mississippi State University, for sharing the *Rhodotorula glutinis* ATCC 204091 genome and annotation methods prior to publication. We would also like to thank Vincent Lombard, Elodie Drula from Architecture et Fonction des Macromolécules Biologiques and Anthony Levasseur from INRA, for their help with the day-to-day development of the CAZy database. We thank Dr. Alexander Idnurm from University of Missouri for generously providing us with the hygromycin resistance cassette, the pPZP-201BK plasmid backbone and *A. tumefaciens* EHA105 strains, as well as a detailed protocol to help us get started on the work. We are grateful to Dr. Cindy Corbitt from University of Louisville (UofL) for the use the dissecting scope setup. I want to thank Vennila

Arumugam from UofL Genomics Core for accommodating my schedule in receiving my precious RNA samples and getting the Bioanalyzer results back to me very promptly. Last but not least, I thank Dr. David Schultz from UofL for being patient in training me in some very important aspects of the project and also for the use of various equipment and chemicals in his lab.

The floral buds used in the project could not have been available if not for large team of undergraduates and fellow colleagues who have helped out in caring for the plants, regardless of the day of the week. They are Sarah Kuykendall, SeongYeon Park, Brittany Hubert, Katelyn Reilly, Jared Andrews, Dominique Razeeq, Dhruvil Brahmbhatt, Kyung Ji Lee, Lauren Grant, Tascha Hader, Ann-Claude Rakotoniaina, Hung Vuong, Swathi Kuppireddy and Ming-Chang Tsai.

Although many friendships and encounters in my graduate career are transient, they are all the more precious. I want to extend my most heartfelt appreciation to these past and present members of the Goat Lab and many beyond the confines of the Goat lab, namely, Dr. Charu Argawal, Kavita Aulakh, Seth Adams, Margaret Wallen, Michael Copper, Ming-Chang Tsai, Parul Singh, Nick Short, Doris Meadows, Judy Felli, Sheryl Meier, Terri Norris, Sabine Waigel and Dr. Wolfgang Zacharias, for their very precious companionship and friendship. I also send my special love to my closest comrades and confidantes, Richa Singhal, Sanaya Bamji, Dr. Jinny Paul and Swathi Kuppireddy. Thank you for being there for me when I needed it. I also will not forget the professors who have extended their helping hands in my graduate career in very tangible ways: Dr. Ronald Fell, Dr. Perri Eason, Dr. Ronald Atlas and my committee members: Dr. David Schultz, Dr. Ted Kalbfleisch, Dr. Kenneth Palmer and Dr. Mark Running.

We are forever indebted to friends in the US who have adopted us and treated us like family, namely the Larks, the Yoders, the Goeppers and the Voits, as well as the church family of Highland United Methodist Church. You have made our stay here a lot less lonely and scary; on the contrary, much warmer than we would dare imagined.

I cannot forget my family and friends from home. I want to express my deepest gratitude for my parents and parents-in-law for supporting my husband and me on this crazy journey, morally, spiritually, physically and financially. You have made this a little easier on us. Friends who kept in contact and those who flew here to visit us, I would not list by name, but thanks for missing us and keeping us in your thoughts.

Lastly, I want to express my deepest gratitude for my mentor and advisor, Dr. Michael Perlin. His personal care and concern for both my graduate career and my personal life were above and beyond what any student should expect. The only reason I can pull this off is no doubt because of him. So thank you for being my mentor superstar, Dr. Perlin!

ABSTRACT

GENOMIC AND TRANSCRIPTOMIC ANALYSES OF *MICROBOTYRUM LYCHNIDIS-DIOICAE* PROVIDE INSIGHTS INTO THE BIOLOGY OF A FASCINATING FUNGAL PHYTOPATHOGEN

Su San Toh

December 4, 2015

This study made use of the *Silene latifolia*/*Microbotryum lychnidis-dioicae* phytopathogen system as the *focal* system to establish the first reference genome for *Microbotryum violaceum sensu lato*. *In silico* analysis was performed on the genome assembly to identify various characteristics of the genome. Using RNA-Sequencing technologies on the Illumina platform, we collected transcriptomic data for both *in vitro* and *in planta* life stages of the fungus, providing the most comprehensive look at the gene expression and regulation of this fungus. Due to a lack of identifiable domains on the predicted genes, gene set enrichment analysis was done in context, by including gene sets like “secreted proteins”, “small secreted proteins” and “unique proteins”, to aid discovery of the features in the different datasets. To further research into *Microbotryum* species in general, we developed, for the first time, a robust and repeatable *Agrobacterium*-mediated transformation system. Using genomic and transcriptomic data,

we were able to select native promoters that drive transcription in specific conditions, making it a highly versatile and controllable system.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iv
ABSTRACT	vii
LIST OF TABLES	xi
LIST OF FIGURES.....	xiv
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: SEX AND PARASITES: GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF <i>MICROBOTRYUM LYCHNIDIS- DIOICAE</i> , THE BIOTROPHIC AND PLANT-CASTRATING ANTHR SMUT FUNGUS.....	7
Introduction.....	7
Results.....	12
Discussion.....	55
Materials and Methods.....	60
Conclusion	75
CHAPTER 3: RNA-SEQ ANALYSIS OF <i>MICROBOTRYUM LYCHNIDIS- DIOICAE</i> IN ARTIFICIAL MEDIA	76
Introduction.....	76
Materials and Methods.....	78
Results and Discussion	81
Conclusion	103

CHAPTER 4: SIZE DOES MATTER: STAGING OF <i>SILENE LATIFOLIA</i> FLORAL BUDS FOR RNA-SEQUENCING PURPOSES'	104
Introduction.....	104
Results and Discussion	109
Materials and Methods.....	125
Conclusion	128
CHAPTER 5: TRANSCRIPTOMIC ANALYSIS OF <i>MICROBOTRYUM</i> <i>LYCHNIDIS-DIOICAE</i> IN YOUNG FLORAL TISSUE OF HOST	130
Introduction.....	130
Materials and Methods.....	133
Results and Discussion	136
Conclusions.....	153
CHAPTER 6: RELIABLE TRANSFORMATION SYSTEM FOR <i>MICROBOTRYUM LYCHNIDIS-DIOICAE</i> INFORMED BY GENOME AND TRANSCRIPTOME PROJECT	155
Introduction.....	155
Materials and Methods.....	158
Results.....	173
Discussion	189
CHAPTER 7: SUMMARY AND CONCLUSIONS	193
REFERENCES.....	195
APPENDICES.....	220
Appendix 1: RNA Processing in the Goat Lab.....	220
Appendix 2: Ligation of pPZP-MV-HYG with Fusion Amplicon	234
Appendix 3: Supplementary Tables and Figures	235
Appendix 4: Germination and maintenance of <i>Silene latifolia</i>	254
Appendix 5: Recipes for transformation of <i>Microbotryum lychnidis-dioicae</i>	258
Appendix 6: Command Line BLAST+ & Unix.....	260
CURRICULUM VITA.....	265

LIST OF TABLES

TABLE	PAGE
Table 2-1. Genome sequencing statistics.	12
Table 2-2. Genome statistics of nuclear genome and mating-type chromosome regions.	13
Table 2-3. RNA read statistics	15
Table 2-4. Genome coverage of TE families (10,283 copies of 286 REPET consensus sequences).	18
Table 2-5. TE used in RIPCAL analysis	23
Table 2-6. Expanded and depleted PFAM domain in <i>M. lychnidis-dioicae</i> . 32	32
Table 2-7. qRT-PCR validation of secretory lipase expression.	37
Table 2-8. Selected CaZy expansions and depletions.	39
Table 2-9. Properties of predicted secreted proteins.	46
Table 2-10. PFAM domains enriched in MI-late induced genes.	52
Table 3-1. List of enriched gene sets that were up-regulated in p1A1 NL when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of nominal enrichment score (NES).	86
Table 3-2. List of enriched gene sets that were up-regulated in p1A2 NL when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of NES.	86

Table 3-3. List of enriched gene sets that were up-regulated in Mated when compared to p1A1 Rich and p1A2 Rich, ranked by decreasing magnitude of NES.	87
Table 3-4. List of significantly enriched gene sets that were down-regulated in Mated when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of normalized enrichment score (NES).	88
Table 3-5. List of enriched gene sets that were down-regulated in NL when compared to p1A1 Rich, ranked by decreasing magnitude of normalized enrichment score (NES).	89
Table 3-6. List of enriched gene sets that were down-regulated in NL when compared to p1A2 Rich, ranked by decreasing magnitude of NES.	90
Table 3-7. Summary of trends observed in genes of <i>M. lychnidis-dioicae</i> over time when treated with phytol.	93
Table 3-8. List of enriched gene sets that were up-regulated in Pmated12, Pmated24 and Pmated48 when compared to Mated, ranked by decreasing magnitude of NES.	94
Table 3-9. List of enriched gene sets that were up-regulated in Pmated12, Pmated24 and Pmated48 when compared to each other, ranked by decreasing magnitude of NES.	95
Table 4-1. Size observations and staging of female floral buds from two plants	117
Table 4-2. Size observations and staging of female uninfected floral buds from four plants.	118
Table 4-3. Summary of RNA extracted from floral bud tissue based on the stages of interest.	122
Table 5-1. List of gene sets that were enriched in genes up-regulated in plant compared to Mated, arranged in order of decreasing magnitude of NES. ...	144
Table 5-2. List of gene sets that were enriched in the genes down-regulated in plants when compared with Mated with no order of normalized enrichment score (NES).	145

Table 5-3. Transcripts per million of three copper oxidases detected to be up-regulated in male but not in female, to illustrate the importance of biological replicates.	146
Table 5-4. Number of common proteins in each core enriched gene set up-regulated in all plant stages when compared to Mated	147
Table 5-5. List of gene sets that were enriched among genes that were down-regulated in late stage of floral development compared to earlier plant stages.	151
Table 6-1. List of potential gene promoters used for MEME analysis	158
Table 6-2. Parameters used for the selection of motifs on MEME	159
Table 6-3. Primers used in overlap PCR	160
Table 6-4. Primers used for verification of <i>A. tumefaciens</i> transformants..	165
Table 6-5. Primers used for verification of <i>M. lychnidis-dioicae</i> transformants.....	167
Table 6-6. Results of electroporation of <i>A. tumefaciens</i> EHA105	176
Table 6-7. Verification of presence of inserts in pPZP vectors and <i>A. tumefaciens</i> transformants.....	177
Table 6-8. Yield of <i>M. lychnidis-dioicae</i> transformation by <i>A. tumefaciens</i>	179
Table 6-9. Number of verified transformants at genomic level	180
Table A3-1. Source of genome data for comparative analysis.	235
Table A3-2. CAZymes present in <i>M. lychnidis-dioicae</i>	237
Table A3-3. Predicted secreted proteins.	243
Table A3-4. Summary of PCR results of selected isolates using cDNA template.	252

LIST OF FIGURES

FIGURE	PAGE
<u>Figure 2-1.</u> Lifecycle of <i>Microbotrym lychnidis-dioicae</i>	9
<u>Figure 2-2.</u> Conservation of core eukaryotic genes (CEGs) set across <i>M. lychnidis-dioicae</i> and other fungal genomes.....	14
<u>Figure 2-3.</u> Correlation between GC content and gene density.	16
<u>Figure 2-4.</u> Preferred codons for each amino acids.....	17
<u>Figure 2-5.</u> Distribution of transposable elements in <i>M. lychnidis-dioicae</i> genome according to their TE classification.	19
<u>Figure 2-6.</u> Proximity of TE copies for main orders (Class I LTR and LINE, Class II TIR and Helitrons) to the closest gene.....	20
<u>Figure 2-7.</u> Comparison of TE proximity for secreted proteins compared to all other genes.	21
<u>Figure 2-8.</u> Frequency of mutation transition types in different TE classes.	22
<u>Figure 2-9.</u> Comparisons of sequence characteristics in non-recombining regions (NRR) of the mating-type chromosome, pseudoautosomal regions (PAR), and autosomes.....	27
<u>Figure 2-10.</u> Codon adaptation of autosomal and sex-specific genome regions.	28
<u>Figure 2-11.</u> Phylogenetic relationship and gene conservation of <i>Microbotrym lychnidis-dioicae</i> and 18 compared fungi.	30
<u>Figure 2-12.</u> Phylogenetic tree of DUF1034 proteins.....	34

<u>Figure 2-13.</u> Expansion of lipase gene family in <i>M. lychnidis-dioicae</i> and <i>M. globosa</i>	36
<u>Figure 2-14.</u> Growth of <i>M. lychnidis-dioicae</i> on different sole carbon sources, including dextrose, cellulose, xylan, and pectin.	42
<u>Figure 2-15.</u> Model of <i>M. lychnidis-dioicae</i> interactions with its host.	51
<u>Figure 3-1.</u> Venn diagram of number of predicted genes in each condition that were significantly (A) up- and (B) down-regulated compared to p1A1 Rich.	83
<u>Figure 3-2.</u> Morphology of haploid and mated cells in the absence or presence of phytol.	92
<u>Figure 3-3.</u> Venn diagram of number of predicted genes that were (A) up- and (B) down-regulated in Phytol-mated (Pmated) compared to Mated condition.....	93
<u>Figure 3-4.</u> Venn diagrams illustrating the number of secreted protein genes that were up-regulated and how they overlap in (A) Pmated conditions compared to Mated (B) Mated and NL compared to Rich and Pmated compared to NL.....	98
<u>Figure 3-5.</u> Venn diagrams illustrating the number of SSP genes that were up-regulated and how they overlap in (A) Pmated conditions compared to Mated (B) Mated and NL compared to Rich and Pmated compared to NL. 99	
<u>Figure 3-6.</u> Venn diagrams illustrating the number of unique protein genes that were up-regulated and how they overlap in (A) Pmated conditions compared to Mated (B) Mated and NL compared to Rich and Pmated compared to NL.....	100
<u>Figure 4-1.</u> Illustration of how the measurements of floral bud (A) and anthers (B) were taken using a glass stage micrometer.	112
<u>Figure 4-2.</u> Stage 6 of 0.2-mm (A) and 0.25-mm (B) male infected floral bud.....	113
<u>Figure 4-3.</u> Stage 7 of 0.3-mm male infected floral bud.....	113
<u>Figure 4-4.</u> Stage 8 (A) and 9 (B) male infected floral buds.....	114

<u>Figure 4-5.</u> Scatterplot of the external and internal anther sizes of the infected and uninfected floral buds against the size of the floral buds.	116
<u>Figure 4-6.</u> Scatterplot of male infected bud floral stage against the size of the floral buds.....	117
<u>Figure 4-7.</u> Stage 7 female uninfected floral bud.....	120
<u>Figure 4-8.</u> Stage 8 (A) and 9 (B) of female infected floral buds.	120
<u>Figure 4-9.</u> Stage 10 of female infected (A) and uninfected (B) floral buds where the styles start to appear.	121
<u>Figure 4-10.</u> Unusual floral bud morphologies.	124
<u>Figure 5-1.</u> Venn diagram of the number of genes that are (A) up- and (B) down-regulated compared to Mated.....	136
<u>Figure 5-2.</u> Venn diagram of number of genes that were (A) up- and (B) down-regulated compared to infected male floral stems (MIFS).	137
<u>Figure 5-3.</u> Venn diagram of the number of genes that are (A) up- and (B) down-regulated compared to Mated.....	138
<u>Figure 5-4.</u> Venn diagram of number of fungal genes that were (A) up- and (B) down-regulated compared to infected female floral stems (FIFS).	138
<u>Figure 5-5.</u> Venn diagram of number of unique and overlapping genes that were (A) up- and (B) down-regulated compared to male and female floral stems (MIFIFS).	140
<u>Figure 5-6.</u> edgeR smear plots of Log fold change against average Log counts per million for male (MI) against female (FI) at each stage: FS, 8, 9 10 and Late respectively. Only a very small number of genes were differentially regulated, hence combining them stage-wise for further comparison should be acceptable. The black dots are non differentially expressed genes and red dots are differentially expressed genes.....	141
<u>Figure 5-7.</u> Venn diagrams showing how the up-regulated (A) SSPs, (B) SPs and (C) Unique proteins that contributed to core enrichment in each category overlap.....	148

<u>Figure 5-8.</u> Venn diagrams showing how the up-regulated (A) SSPs, (B) SPs and (C) Unique proteins that contributed to core enrichment in each category overlap.....	149
<u>Figure 6-1.</u> Most common motif found upstream of genes upregulated in "Rich" condition.....	173
<u>Figure 6-2.</u> Most common motif found among regions upstream of genes upregulated in the "Nutrient-Limited" condition.	174
<u>Figure 6-3.</u> Most common motifs found among regions upstream of genes upregulated in "MI Late".....	174
<u>Figure 6-4.</u> Verification of difference in product size between pPZP-MV-HYG with and without new fusion inserts for expression of <i>eGFP</i>	176
<u>Figure 6-5.</u> Electrophoresis of <i>eGFP</i> PCR products fro cDNA template through 2% agarose gels.	181
<u>Figure 6-6.</u> End sequence of larger <i>eGFP</i> amplicon from 18W cDNA.....	183
<u>Figure 6-8.</u> Part of the alignment of the <i>HYG</i> PCR amplicons with cDNA template (17R).....	185
<u>Figure 6-9.</u> Northern blot of RNA samples isolated from "Rich" and "Nutrient-Limited" conditions.....	187
<u>Figure A3-1.</u> Phylome for <i>Microbotryum lychnidis-dioicae</i>	237

CHAPTER 1: INTRODUCTION

The anther smut disease caused by the *Microbotryum violaceum sensu lato* (formerly known as *Ustilago violacea*), which evolved within the Microbotryales, in the Pucciniomycotina phylum (Bauer et al., 2006; Begerow et al., 1997; Weiss et al., 2004), has been characterized as a sexually transmitted disease of the plant hosts (Antonovics, 2005). It is a biotrophic disease that causes sterility in the host (Baker, 1947), which is touted to be a better strategy for a pathogen than causing mortality (Sloan et al., 2008). In the male, the fungus disguises itself as the pollen so well that Linneaus mistook the infected flower as a different species (Linné, 1737; Liro, 1924), and that mistake was perpetuated into modern times (Bocquet, 1967; Lu et al., 2008; Oxelman et al., 2001; Zhou et al., 2008) until it was discovered in 2003 (Hood & Antonovics, 2003). In fact, most herbarium collectors did not notice that their specimens were diseased (Hood et al., 2010). Since the teliospores occupy the space of the pollen in the anthers, infection in the female then seems futile, as there is no stamen. To overcome that, the fungus alters the female reproductive organs by suppressing the gynoecium and makes the supposedly rudimentary stamen continue its development into anthers that can house the teliospores, effecting a sex change in the female host (Ruddat et al., 1991; Uchida et al., 2005; Uchida et al., 2003).

Anther smut is known to infect over 200 perennial host species of the Caryophyllaceae (pinks) family (Hood et al., 2010; Thrall et al., 1993). The question of

broad or narrow host range has been extensively investigated. It has been shown via molecular phylogenetic methods that *Microbotryum violaceum sensu lato* consists of multiple host-specific siblings, some having evolved sufficiently (with the host) to be recognized as a distinct species (Bucheli et al., 2000; Denchev et al., 2009; Freeman et al., 2002; Garr et al., 1997; Le Gac et al., 2007a; Lutz et al., 2005; Perlin et al., 1997; Refrégier et al., 2008; van Putten et al., 2005), although generalists have also been found (de Vienne et al., 2009a). In fact, incomplete sterilization of the female had been observed in a novel host-pathogen combinations in which the two did not co-evolve (Sloan et al., 2008), suggesting that specific resistance and pathogenicity (de Vienne et al., 2009a) can be expected.

In this dissertation, we focus on the species *Microbotryum lychnidis-dioicae* isolated from *Silene latifolia*. *S. latifolia* is one of the most studied species among vascular plants that display sexual dimorphism and is also a model system of increasing scientific importance for the study of evolution of plant sex chromosomes (Mrackova et al., 2008). The sex chromosomes are similar in many ways to the animal sex chromosome system, but these plant sex chromosomes are relatively early in the evolutionary strata, *i.e.*, in which the Y chromosome is in its early stages of degeneration (Filatov, 2005; Filatov et al., 2000; Guttman & Charlesworth, 1998; Miller & Kesseli, 2011; Nicolas et al., 2005). It was determined that gynodioecy and hermaphroditism are ancestral to the *Silene* genus (Marais et al., 2011) and that a Y chromosome mutation can reverse the dioecy (Miller & Kesseli, 2011). While it is known that steroids do not affect sex determination (Ruddat et al., 1991), it is still unclear what does.

Side effects of the disease

In field studies of this system, diseased plants were found to flower earlier than healthy plants, and neither the sex of the plant nor the plant and fungal population had an influence over it (Shykoff & Kaltz, 1998). However, it was concluded in the studies that the earlier flowering phenotype might be an innate attribute of the host population, which relates to the higher vulnerability to being infected. Various studies also found that the diseased plants appeared to produce lower root/shoot ratio (Shykoff & Kaltz, 1998), have larger asymmetry of floral petals (Shykoff & Kaltz, 1998) and a reduction in inflorescence height, flower size and nectar production of the flowers (Biere & Honders, 1996). The paper suggested various explanations for the effects, either in favor of defense by the plants or offense by the fungus.

Genomic landscape of *M. lychnidis-dioicae*

The first time the anther smut community was able to get a glimpse at the genomic landscape of *M. lychnidis-dioicae* was when an EST library (Yockteng et al., 2007) of the mated fungus prepared in the presence of α -tocopherol (which stimulated dikaryon formation like phytol) was made available via an online database, MICROBASE (<http://genome.jouy.inra.fr/microbase>). In this study, a high frequency of genes without significant BLAST hits was encountered. The dissimilarity of pheromone receptors compared to other Basidiomycetes explained their failure to clone the gene using degenerate primers designed with sequences from *Ustilago maydis* and *U. hordei*. There was also a lack of cell wall degrading enzymes commonly found in plant pathogens. This was attributed to the biotrophic lifestyle of the fungus. This study

predicted 594 secreted proteins, and 86.4% of these had no significant BLAST hits, which suggests specific interactions between the host and its pathogen.

The a_1 mating type of the pathogen was first sequenced using Illumina technology ((Perlin et al., 2015), Chapter 2). When Single Molecule Real Time (SMRT) sequencing became available, the a_2 mating type was sequenced and yielded 33.3 Mb of assembly, because the technology was able to incorporate repetitive DNA elements more completely by using long read chemistry (Badouin et al., 2015). This was almost 7 Mb larger than the a_1 assembly made on the Illumina platform. The findings concurred with the first EST study, with a lack of cell wall degrading enzymes but an expansion of CAZyme and protein domains that identify the pathogenicity factors.

Even before a whole genome sequence was available, the mating type chromosomes of *M. lychnidis-dioicae* have served as a model for investigating the evolution of sex chromosome (Fontanillas et al., 2015; Hood et al., 2004; Hood et al., 2013; Petit et al., 2012; Whittle et al., 2015). It was the first fungus found to have heteromorphic mating type chromosomes (a_1 , ~2.8 - 3.1 Mb, and a_2 , 3.4 - 4.2 Mb) (Hood, 2002); moreover, the a_1 mating type chromosome is more gene-dense than the a_2 mating type chromosome (Hood et al., 2004). There is also an agreement on the homogeneity of the autosomes and a large non-recombining region (NRR) on the mating type chromosome (Fontanillas et al., 2015; Hood et al., 2013). Gene diversity identification of mating chromosomes is based primarily on optical mapping (Hood et al., 2013), and yet there are genes in the NRR that are very similar in terms of genomic sequence (Fontanillas et al., 2015). At the same time, comparison of the a_2 pheromone of different *Microbotryum* species (Badouin et al., 2015) revealed high sequence variation.

In order to discover the function of any gene, molecular genetics that render the organism tractable is highly desired. Several attempts at developing such a system have been made (Bej & Perlin, 1989a; Bej & Perlin, 1989b; Bej & Perlin, 1991; Perlin et al., 1990), mostly modeled after other transformation systems in tractable organisms like *Saccharomyces cerevisiae* and *Ustilago maydis*. However, these techniques were not reproducible by other researchers; even biolistic approaches proved ineffective (M. Perlin, personal communication). Thus, none of these approaches were pursued further. With the large body of genomic and transcriptomic data available, the motivation to develop a reproducible transformation system is higher than ever before.

Overview of the dissertation

In this dissertation, we attempted to obtain as much genomic and transcriptomic data as possible on the *S. latifolia*/*M. lychnidis-dioicae* pathosystem. The genome paper was published (Perlin et al., 2015) and presented in Chapter 2 of this dissertation and the genome has served as the reference genome for other studies since (Badouin et al., 2015; Fontanillas et al., 2015).

For transcriptomic data, we performed RNA-Seq on as many life stages of *M. lychnidis-dioicae* as we could simulate. *In vitro* samples that represented the lifestyle of the fungus in nutrient-rich and nutrient-limited environment as haploids potentially reveal the physiological differences of the two mating types, despite the homogeneity of the autosomes. When the two mating types were placed together in the nutrient-limited environment, it also stimulates mating. Further, on exposure to phytol (Castle & Day, 1984), stable dikaryons like those that would “travel” in the host plants were formed. Transcriptomic information from these stages was analyzed and reported in Chapter 3.

Gene expression data in the host plants had been limited and was never discrete to the early stages of floral development nor to the sex of the host. We first attempted to break down the early stages of floral development by examining the floral morphology and correlating it to bud sizes to aid with rapid tissue collection to preserve the integrity of the RNA. This work has been published (Toh & Perlin, 2015) and is documented in Chapter 4. The findings from the RNASeq data that explored gene expression differences between male and female host in various early stages of floral development is reported in Chapter 5.

To fulfill the need for a reliable transformation system, we adapted and developed one that takes advantage of *Agrobacterium tumefaciens*. The detail of the system is reported in Chapter 6.

Overall, we hope that this dissertation serves to provide a body of data that will be useful in guiding future hypothesis and research, both to the fungal and the host community.

CHAPTER 2: SEX AND PARASITES: GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF *MICROBOTRYUM LYCHNIDIS-DIOICAE*, THE BIOTROPHIC AND PLANT-CASTRATING ANTHER SMUT FUNGUS¹

Introduction

Members of the genus *Microbotryum* are pathogenic fungi that have a global distribution and infect over nine families of host plants, with most replacing floral structures with fungal spores (Kemler et al., 2009). Within this genus, the anther-smut fungi of plants in the Caryophyllaceae consist of many recognized and cryptic species, most possessing a very narrow host range, although some are more generalist (de Vienne et al., 2009a). Hybrid incompatibilities for the species examined in this complex leads to post-zygotic isolation in the forms of inviability and sterility (de Vienne et al., 2009b; Le Gac et al., 2007a; Le Gac et al., 2007b). The high rates of selfing and ecological specialization on different host plants are factors that should promote speciation in *Microbotryum* (Gibson et al., 2012; Giraud et al., 2008a). The anther smut fungi thus allow examination of the ecology and evolution of host/pathogen interactions in “wild”, non-agricultural environments (Alexander, 1989; Antonovics et al., 2002), in which the

¹ Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, Duplessis S, Henrissat B, Young S, Zeng Q *et al.* (2015) Sex and parasites: Genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics* 16:461. doi: 10.1186/s12864-015-1660-8

genetically variable hosts provides an important contrast to the heavily studied, more monocultural hosts of agricultural systems. *Microbotryum* species also serve as a model for emerging infectious disease through host shifts (Antonovics et al., 2002; Refrégier et al., 2008), for studying the evolution of mating systems, non-recombining mating-type chromosomes and sex chromosomes (Hood et al., 2004; Hood et al., 2013), and for examining pathogens that alter the development of the host (Hughes & Perlin, 2005).

As obligate parasites, *Microbotryum* anther smut fungi must complete their life cycle in association with a host plant. Their fungal diploid teliospore masses give the flowers a dark, powdery appearance, thus the name “anther smut”. Teliospores of *Microbotryum* are transported from diseased to healthy plants by direct transmission when plants are in close proximity (Roche et al., 1995) or by pollinating insects (Jennersten, 1983), where once deposited the diploid fungus germinates and undergoes meiosis to give rise to four haploid cells (Giraud et al., 2008a). Each of these cells can bud off yeast-like sporidia on the plant surface. New infectious dikaryotic hyphae are rapidly produced by conjugation of two cells of opposite mating-type (a_1 and a_2) and enter the host tissue to grow endophytically until they reach the bud meristems and anthers (Akhter & Antonovics, 1999). Here the nuclei fuse (karyogamy) and teliosporogenesis occurs, thus completing the life cycle (**Fig. 2-1**; (Giraud et al., 2008a; Kokontis & Ruddat, 1986; Kokontis & Ruddat, 1989; Scutt et al., 1997)).

The commandeering of insect pollinators for disease transmission is associated with significant pathogenic alterations of the host’s floral morphology, particularly in relation to male and female structures. Although diseased plants are only slightly affected in vegetative morphology and survival (Alexander & Antonovics, 1995), infection often

results in complete host sterility, as no pollen is produced in the anthers and the ovary becomes rudimentary. Interestingly, in dioecious or gynodioecious species of *Silene* (e.g., *S. latifolia* and *S. vulgaris*, respectively), ovaries of diseased female plants are aborted, while a male morphology develops with spore-bearing anthers. The basis for the changes in female plants is not fully understood. Studies have identified host genes expressed during infection of females that are also normally expressed in uninfected males (*MEN* genes (Robertson et al., 1997); *SLM2* (Kazama et al., 2005)).

Figure 2-1. Lifecycle of *Microbotrym lychnidis-dioicae*.

The infection cycle for *M. lychnidis-dioicae* is shown. Infection begins when diploid teliospores germinate on a suitable plant surface, and after meiosis, produce linear tetrads of haploid basidiospores. These cells may mate with cells of opposite mating-type (*i.e.*, a_1 or a_2), either directly within the tetrad or after budding to yield free yeast-like sporidia.

Mated cells form conjugation bridges and, upon receiving suitable cues from the host plant, develop into dikaryotic hyphae that penetrate the host tissue. The hyphae progress systemically through the plant and migrate to the flower primordia. There nuclear fusion (karyogamy) occurs, as the hyphae break up and develop into diploid teliospores that replace the pollen in the anthers of the developing flowers. Pollinator species are then able to transmit the spores to other uninfected flowers, thus completing the cycle.

The formation of a hyphal dikaryon by mating between haploid cells is a prerequisite for infection in *Microbotryum*, and thus each newly diseased plant represents the completion of a sexual reproductive cycle. In heterothallic fungi, mating can occur only between haploid cells carrying different mating-types, which are controlled by alleles at one or two loci (Billiard et al., 2011; Billiard et al., 2012). In a few fungi, recombination suppression has evolved around the mating-type locus/loci (Fraser & Heitman, 2004; Fraser et al., 2007; Menkis et al., 2008), sometimes leading to structurally dimorphic chromosomes similar to the sex chromosomes in plants and animals (Hood, 2002; Hood & Antonovics, 2000). The suppression of recombination on mating-type chromosomes is expected to lead to degeneration; this can be manifested as transposable element accumulation (Hood, 2005), codon usage degeneration (Whittle et al., 2011), accumulation of deleterious mutations (Whittle & Johannesson, 2011), and eventually to chromosomal dimorphism, and thus the emergence of allosomes (Hood et al., 2013). Such degeneration has in fact been observed on the mating-type chromosomes of some fungi, including the ascomycete *Neurospora tetrasperma* (Whittle et al., 2011)

and the basidiomycete *Microbotryum lychnidis-dioicae* (Fontanillas et al., 2015; Hood, 2005; Hood et al., 2013), the most well-studied representative of the anther-smut fungi.

We sequenced the genome of haploid isolate Lamole p1A1 (Hood et al., 2004), of the a₁ mating type, to represent the *M. lychnidis-dioicae* species found in association with the perennial, dioecious host, *Silene latifolia*. RNA-Sequencing of distinct life cycle stages was incorporated to validate gene content and measure expression changes during infection. We identified gene family expansions that could play a role in plant infection by comparing *M. lychnidis-dioicae* to other basidiomycetes, taking advantage of increasing genome coverage of the Pucciniomycotina subphylum. The identification of genes that are induced or repressed during infection highlighted carbohydrate active enzymes (CAZymes) that may be involved in host cell degradation or manipulation of host development. Additionally, the *M. lychnidis-dioicae* genome is riddled with a diverse array of transposable elements (TEs), including a higher proportion of Helitron elements than found in the much larger and more highly-repetitive genomes of related rust fungi. Genome regions corresponding to the mating-type chromosomes of *M. lychnidis-dioicae* (Hood et al., 2004) are enriched for repetitive sequence. Further analysis of the sequence of the entire a₁ mating-type chromosome identified more than 300 genes linked with mating-type. Together, these findings provide an in-depth portrait of genetic architecture and adaptation in a specialized fungal plant pathogen.

Results

Genomic sequence and content

The 25.2 Mb haploid genome of *M. lychnidis-dioicae* was sequenced using 454 technology, generating high coverage of three different-sized libraries (**Table 2-1**), and assembled using Newbler (**Table 2-1**). The assembly was comprised of 1,231 scaffolds where the average base was present in a scaffold of 185 kb and a contig of 50 kb (N50 measure, **Table 2-2**). Despite the large number of contigs, the assembly was a nearly complete representation of the sequenced genome, comprising 97% of sequenced bases. The assembly included five scaffold ends with the typical fungal telomere repeat (TTAGGG), though three of these scaffolds were smaller than 1kb in size.

Table 2-1. Genome sequencing statistics.

454 Library Insert	Sequencing Type	Molecular Barcode	Reads (count)	Average Read Length (bases)	PF Bases (count)	Sequence Depth (Fold)
Fragment	Single read	Madrid	1,145,609	371.3	425,393,315	16.36
2.7 kb	Paired read	Bratislava	510,499	280.2	143,013,765	5.50
5.5 kb	Paired read	Cairo	686,469	162.2	111,354,758	4.28
5.5 kb	Paired read	Budapest	233,405	189.1	44,129,203	1.70
Total			2,575,982		723,891,041	27.84

High coverage strand-specific RNA-Seq, generated from three biological conditions (**Table 2-3**) assisted with the prediction of 7,364 protein coding genes and identified expression changes potentially important for the pathogenic lifecycle. Sampled conditions included two *in vitro* conditions, haploid cells grown on yeast peptone dextrose media (YPD) agar (referred to as rich) or on 2% water agar (referred to as

nutrient limited). These were compared to a sample from infected male plant tissue during the late stages of fungal development, when teliospores form on partially and fully opened smutted flowers (Farbos et al., 1997; Grant et al., 1994) (referred to as “MI-late”). Incorporation of RNA-Seq into predicted gene structures (Methods) defined UTRs for the vast majority (more than 6,100) of predicted genes; the average length of 5’ and 3’ UTRs was 183 bases and 253 respectively. Coding sequences average 1,614 bases (median of 1,338 bases) in length and contain 5.6 exons; genes are separated by intergenic regions 502 bases in length on average. The *M. lychnidis-dioicae* gene set has high coverage of a core eukaryotic gene set (Parra et al., 2007), highest in fact of any of the fungal gene sets used in comparative analysis (**Fig. 2-2**), suggesting the assembly includes a highly complete gene set.

Table 2-2. Genome statistics of nuclear genome and mating-type chromosome regions.

	Nuclear genome	NRR* regions	PAR** regions
Assembly size (Mb)	26.1	1.86	0.38
Scaffolds (count)	1,231	85	2
Scaffold N50 (kb)	185	48	381
Contigs (count)	2,104	229	16
Contig N50 (kb)	50	13	45
GC content (%)	55.4	54.6	53.9
TE content (%)	14	41	13
Protein coding genes	7,364	350	99
Mean coding length	1,614	1,344	1,408
Median coding length	1,338	954	1,302
Mean exons/gene (count)	5.6	4.6	5.1
Mean intercds length (bp) [#]	1,181	2,600	1,861
tRNAs	134	5	2

[#]average length between coding sequence (cds) start and stop. *Non-recombining regions (NRR).

**Pseudo-autosomal regions (PAR).

The mitochondrial genome consisted of a single finished contig of 97 kb and included the canonical set of genes. These were the respiratory-related proteins of the

NADH dehydrogenase family (*nad1-6* and *nad4L*), apocytochrome b (*cob*), cytochrome oxidases (*cox1-3*), the proteins related to ATP synthesis (*atp6*, *atp8* and *atp9*); ribosomal RNAs (*rns*, and sequence similar to *rnl*); ribosomal proteins (*rps3*); DNA polymerase (*dpo*) and 25 tRNAs. Homing endonucleases of the LAGLIDADG and GIY-YIG families and a maturase protein were located in mitochondrial intronic regions, including within three introns of *cox1*.

Table 2-3. RNA read statistics

Sample	Reads	PF* Reads	PF Reads aligned**	PF Reads aligned (%)**
Haploid rich	17,025,508	16,239,792	14,613,444	89.99%
Haploid nutrient limited	19,275,930	18,335,352	16,366,386	89.26%
MI-late	20,165,846	19,575,594	4,441,268	22.69%
Total	56,467,284	54,150,738	35,421,098	

*Illumina Passing Filter high quality criteria.

**Number and percent of PF reads aligned to *M. lychnidis-dioicae* genome

To examine the presence of AT-rich isochores and more generally the genome structure in GC composition in *M. lychnidis-dioicae*, we measured the fluctuation of GC percent along the assembly. Although the genome is not organized in discrete isochores as in some fungal genomes (Rouxel et al., 2011), we observe some large-scale variation (> 100 kb) in base composition within chromosomes, as well as finer-scale fluctuations (**Fig. 2-3A**). The GC content was positively correlated with coding density, with the most significant correlations for 5 and 10 kb windows explaining 16.8 % of the variance (**Fig 2-3B**). This correlation has been explained in other systems by biased codon usage toward GC-rich codons or alternatively biased gene conversion occurring more frequently in coding than in non-coding sequences (Duret & Galtier, 2009). In fact, an analysis of the preferred codons (*i.e.*, the most frequently used codons in the predicted genes) showed that 17 out of 18 had a GC base in the third position, which is the most degenerate position and therefore primarily influences the GC composition of genes (**Fig. 2-4**).

Figure 2-3. Correlation between GC content and gene density.

A. Using the assembly generated from 454 sequencing, GC percent was computed in non-overlapping windows of 10 kbp and plotted against genomics coordinates. Vertical blue lines represent the limits between scaffolds.

B. Gene density was also measured in the same windows as in A. There was a significant positive correlation ($p\text{-value} < 2.3\text{e-}05$) between density in (CDS) and GC percent, explaining 16.8 % of the variance.

Shifts in transposable element type, location and impact of RIP

The genome of *M. lychnidis-dioicae* contained diverse transposable elements (TEs), represented by 286 consensus elements, covering 14% of the total assembly. The overall TE content was lower than of other species in Pucciniomycota; in *Puccinia*

graminis f. sp. *tritici* and *Melampsora larici-populina*, in which TEs account for nearly 45% of the assembled genomes, contributing to the expanded genome size of these fungi (Duplessis et al., 2011). Among class I retrotransposon elements (36% of TE sequences), Long Terminal Repeat (LTR) elements were the most common (28% of TE sequences), in particular *Copia*-like elements (20% of TE sequences), in agreement with prior studies of genome sampling and expressed sequence profiles (Hood, 2005; Yockteng et al., 2007). The remainder of class I elements consisted of Long Interspersed Nuclear Element (LINE) (7% of TE sequences) and Dictyostelium Intermediate Repeat Sequence (DIRS) elements (1% of TE sequences). Among class II DNA transposons (23% of TE sequences), Terminal Inverted Repeat (TIR) and *Helitron* elements, which transpose by rolling-circle replication (Kapitonov & Jurka, 2001), account for 12% and 10% of TE sequences, respectively (**Fig. 2-5, Table 2-4**). The *Helitron* proportion was an order of magnitude higher than in the more repetitive genomes of other Pucciniomycota fungi *P. graminis* f. sp. *tritici* and *M. larici-populina*, for which a similar analysis characterized only 1% of TE elements as *Helitrons* (Duplessis et al., 2011).

Table 2-4. Genome coverage of TE families (10,283 copies of 286 REPET consensus sequences).

TE copies	Copy number	Coverage relative to TE space	Coverage relative to assembly size		
LTR	1018	27.70	Class I 35.64	3.89	Class I 5.01
DIRS	57	1.34		0.19	
LINE	392	6.61		0.93	
SINE	0	0		0	
TIR	438	12.18	Class II 22.76	1.71	Class II 3.20
MITE	39	0.39		0.06	
Helitron	373	10.18		1.43	
Unknown	1942	41.60	41.60	5.85	5.85
Total	4259	100	100	14.06	14.06

The TE categories varied significantly in their proximity to genes. A chi-squared test of heterogeneity found a significant difference in the TE content of regions nearby genes, comparing regions less than versus greater than 1kb upstream and downstream of genes (upstream region p-value < 2.2e-16; downstream region p-value < 2.2e-08, **Fig. 2-6**). In particular, class II elements (TIR and *Helitrons*) were closer to genes than class I elements (LTR and LINE), with a greater enrichment upstream of genes compared to downstream (**Fig. 2-6**). In addition, there appeared to be an association between TE-rich regions in *M. lychnidis-dioicae* and genes for Small Secreted Proteins (SSPs), which can include effector proteins involved in host-pathogen interactions as suggested in some pathogen genomes (Gladieux et al., 2014; Haas et al., 2009). SSPs were indeed located

closer to TEs than the set of all other non-SSP genes (Chi-squared p-value < 5e-4; **Fig. 2-7**).

Hypermethylation in TEs that resembles the genome defense, Repeat-Induced Point mutation (RIP), has previously been observed in the LTR elements (*copia*-like and *gypsy*-like elements) and *Helitron* transposons of *M. lychnidis-dioicae* (Hood et al., 2005; Horns et al., 2012). Some genes that appeared similar to those necessary for RIP in *Neurospora crassa* (Cambareri et al., 1989; Selker et al., 1987) were found in the *M. lychnidis-dioicae* genome. These include a cytosine methyltransferase (MVLG_04160), but establishing orthology with the RIP-essential *rid* gene from *N. crassa* versus other cytosine methyltransferases (e.g. Dim-2) would require further investigation. *M. lychnidis-dioicae* sequences similar to the Dim-5 H3 histone methyltransferase that is essential for marking RIP regions in *N. crassa* (MVLG_02125, MVLG_05378) were also found.

Evaluation of dinucleotide signature at transition mutations in 179 TE families (2,298 genome copies, **Table 2-5**) revealed that 40% of these TE copies exhibited elevated substitution rates that were particular to which nucleotide was 3' to the cytosine (Methods, **Fig. 2-8**). Eighty per cent of TE copies with high frequencies of cytosine mutation showed a bias toward CpG dinucleotides, consistent with the "CpG effects" (Walser & Furano, 2010) of maintenance methylation known in eukaryotes (Fryxell & Moon, 2005; Jiang & Zhao, 2006; Morton et al., 2006), including fungi (Amselem et al., 2015; Horns et al., 2012) where CpG methylation of TEs have been shown in ascomycete and basidiomycete fungi (Zemach et al., 2010). Notably, the rate of CpG mutations varied according to TE order and superfamily examined: TE copies exhibiting a pattern of frequent transition at CpG sites appeared lower for class II DNA transposons

(*Helitron*-type and TIR elements) than that of class I Retrotransposons (21% and 58% respectively) (**Fig. 2-8**). In addition, the GC content in TEs (54.4%) was very close to GC content in genes (56.0%) and the genome (55.4%).

Table 2-5. TE used in RIPCAL analysis

2,298 copies from 179 TE families were used in RIPCAL calculations. GC% is calculated on whole sequences used.

LTR families	LTR copies	DIRS families	DIRS copies	LINE families	LINE copies		
37	569	2	24	11	203		
TIR families	TIR copies	MITE families	MITE copies	Helitron families	Helitron copies	Unknown families	Unknown copies
18	300	1	10	15	264	95	928
GC% copies	GC% genes	GC% genome					
54.4	56.04	55.43					

In addition, the *M. lychnidis-dioicae* genome was found to contain the core RNAi machinery components, which may act to constrain proliferation of transposable elements, contrasting with some other Basidiomycetes that have lost this pathway (Nicolas et al., 2013). The RNAi pathway components were identified based on similarity to known RNAi genes in fungi and validated by examining predicted functional domains (Methods). The genome of *M. lychnidis-dioicae* was also found to contain one copy of a RNA-dependent RNA polymerase (MVLG_02137), two copies of Argonaute (MVLG_06823 and MVLG_06899), and one copy of Dicer (MVLG_01202). All of these components of RNAi machinery were expressed under each of the conditions examined, suggesting the pathway is active across life cycle states, although one copy of Argonaute (MVLG_06823) was significantly more highly expressed (corrected p-value <0.01) during infection compared to nutrient limited and rich agar media.

Mating-type locus and chromosome

The central proteins involved in mating-type determination in Basidiomycetes were found in *M. lychnidis-dioicae*. Orthologs of the two-component homeodomain transcription factor that functions in post-mating compatibility, HD1 (MVLG_07149) and HD2 (MVLG_07150), were assembled in a 14.1 kb region; as previously described HD1 and HD2 are adjacent and divergently transcribed in *M. lychnidis-dioicae* (Petit et al., 2012), similar to other fungi (Gillissen et al., 1992). Both the mating pheromone receptor (Devier et al., 2008) and the homeodomain compatibility factor identified above are located at the ends of their respective supercontigs, such that the genomic proximity of these two essential mating-type-determining loci is unclear. The chromosomes bearing these two mating-type loci show suppressed recombination across most of their length, with only two small recombining regions at their ends, *i.e.*, pseudo-autosomal regions (PARs) (Fontanillas et al., 2015; Hood et al., 2004; Hood et al., 2013; Petit et al., 2012; Votintseva & Filatov, 2009). The assembly scaffolds corresponding to the non-recombining regions (NRRs) and to the PARs were identified based on alignment to an optical map of the mating-type chromosomes (Fontanillas et al., 2015; Hood et al., 2013) and by performing additional sequencing of gel purified chromosomes (Methods). A total of 449 genes mapped to the a₁ mating-type chromosome, including 350 genes found on the NRRs and 99 genes found on the PARs (**Table 2-2, Additional file 11**²). Other than the genes for the pheromone receptor, the homeodomain transcription factors, and the

² Available for download at
<http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s11.xlsx>

STE20 protein kinase, no other genes on the mating-type chromosome have a predicted function linked to mating in other systems.

Consistent with the expectation in regions of suppressed recombination (*e.g.*, (Bachtrog, 2003)), the TE density (**Fig. 2-9A**) was several fold higher in the non-recombining region (NRR) of the a_1 mating-type chromosome (41%) relative to the autosomes (9%), confirming prior studies of a TE accumulation on the mating-type chromosomes as a whole (Fontanillas et al., 2015; Hood et al., 2004; Hood et al., 2013). The pseudoautosomal regions (PARs, supercontigs 37 and 43) displayed a TE content (~13%) more similar to the estimate for autosomal regions than to the NRR of the mating-type chromosome. Gene density in the NRR of the mating-type chromosome (23%), estimated as CDS density, was less than half the gene density of the autosomes (49%) (**Fig. 2-9B**). The number of genes predicted per 10 kb positions also indicated a lower density in the NRR of the mating-type chromosome (1.7 genes) than in the autosomal partition (2.9 genes) (**Fig. 2-9C**). As with TE content, the PARs displayed gene density values (~40% for CDS density and ~2.5 for genes per 10 kb) closer to the autosomal estimates than the NRR of the mating-type chromosome. The NRR of the mating-type chromosomes contained genes of the same distribution, with the exception of 2-fold elevated density of small secreted proteins (SSPs) and, in particular, a 5-fold enrichment of Cys-rich SSPs compared to the autosomes.

With regard to base pair composition, again the NRR exhibited a pattern distinct from the PARs or autosomes in GC content. The GC content in protein coding genes, irrespective of codon position, were similar among the NRR of the mating-type chromosome, PAR, and autosomes (**Fig. 2-9D**), as were the contents represented by TEs

(**Fig. 2-9E**). However, in other sequences, representing inter-genic regions not consisting of TEs, the NRR displayed markedly reduced GC content (**Fig. 2-9F**) while PARs and autosomes had similar GC levels. GC content of codon positions within protein-coding genes did not vary among the NRR, PARs, and autosomes (**Fig. 2-9G-I**), notably in that the third-codon position patterns were not reflective of intergenic GC composition variation across regions. This pattern as well as the lower GC content observed for the second-codon position compared to first or third positions is consistent with some prior research (Elhaik et al., 2009). The correspondence analysis of codon usage (**Fig. 2-10**) among the non-recombining region of the mating-type chromosome, PAR and autosomal did not indicate obvious differences in codon usage.

Figure 2-9. Comparisons of sequence characteristics in non-recombining regions (NRR)

of the mating-type chromosome, pseudoautosomal regions (PAR), and autosomes.

The genomic regions are shown, with results for the two supercontigs (“37” and “43”) corresponding to the two PARs presented separately. **A)** Transposable element (TE)

density is shown as the total length of TE sequences over the total length of DNA analyzed. **B, C)** Gene density is shown as the proportion of the total length of coding

region (CDS) over the total length of DNA analyzed, and number of putative genes identified per 10000 nucleotides, respectively. **D-F)** Proportion GC base pair contents are

shown for CDS, TEs, and for the remaining, predominantly intergenic regions. **G-I)** Proportion GC base pair content for protein-coding genes are shown relative to first-,

second-, and third-codon positions (“GC1”, “GC2” and “GC3,” respectively).

Figure 2-10. Codon adaptation of autosomal and sex-specific genome regions.

Correspondence analysis of synonymous codon usage, which compares sets of genes belonging to non-recombining mating type regions, pseudoautosomal regions (“PAR”) and autosomes.

Gene conservation and lineage-specific changes

The comparison of 7,364 predicted proteins of *M. lychnidis-dioicae* to those of diverse basidiomycetes revealed gene loss and gain patterns relevant in terms of the growth and pathogenesis of this organism. We included representatives of the three subphyla of Basidiomycetes (Pucciniomycotina, Agaricomycotina, and Ustilaginomycotina), as well as three Ascomycota species as outgroups (**Fig. 2-11, Table A3-1**). Within the Pucciniomycotina, species compared included *M. lychnidis-dioicae* and two closely related Microbotryomycetes (*Sporobolomyces roseus* and *Rhodotorula glutinis*) and three other more distantly related species; within this group the two other plant pathogens (*P. graminis* f. sp. *tritici* and *M. larici-populina*) are biotrophic, like *M. lychnidis-dioicae*. These comparisons revealed 2,451 *Microbotryum* gene clusters representing 2,613 genes that were broadly conserved in the Basidiomycota (present in at least 13 of the examined 15 other Basidiomycete genomes). The gene families specific to Pucciniomycotina, with orthologs present only in *M. lychnidis-dioicae* and/or the other Pucciniomycotina species, (**Fig. 2-11**, orange boxes) was composed of a small set of 224 predicted proteins from *M. lychnidis-dioicae*; the two rusts (*P. graminis* and *M. larici-populina*) share a larger number of Pucciniomycotina-specific proteins in part due to their expanded genome size. Examining proteins conserved across the Microbotryomycetes, orthologs of 4,844 *M. lychnidis-dioicae* proteins were present in at least one other species, and of these 233 were specific to the Microbotryomycetes. A set of 190 gene duplications occurring specifically along the *M. lychnidis-dioicae* lineage was identified

using phylogenetic analysis (**Additional file 14³**, **Fig. A3-1**, phylomedb.org); these did not display functional enrichment for gene ontology (GO) term assignments, as GO terms were only assigned for 19 of the 190 genes. While most genes (70%) were shared with occurrence in at least one other species, the remaining set of 1,534 genes appear specific to *M. lychnidis-dioicae*.

Figure 2-11. Phylogenetic relationship and gene conservation of *Microbotryum lychnidis-dioicae* and 18 compared fungi.

³ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s14.xlsx>

Phylogeny (left panel) is based on concatenated MUSCLE alignments of 80 single copy genes. Species phylogeny was inferred using RAxML (PROTCATWAG model) with 1,000 bootstrap replicates; all nodes were supported by at least 99% of replicates. Ortholog conservation (right panel) highlights genes conserved in all species (core, green), genes conserved in at least two species (conserved, blue), genes unique to the Pucciniomycotina or Microbotryomycetes (orange and red, respectively) and genes unique to a given species (species-specific, yellow).

The identification of enriched or depleted PFAM domains for *M. lychnidis-dioicae* compared to other fungi revealed significant differences in functional categories between these genomes. A total of six protein domains were significantly enriched or depleted ($q\text{-value} < 0.05$) in *M. lychnidis-dioicae* compared to the other Basidiomycetes examined (**Table 2-6, Additional file 16⁴**). Enriched domains include secretory lipases and two domains of unknown function, DUF23 (glycosyl transferase 92) and DUF1034 (Fn3-like). DUF23 (PF01697) was present in five copies in *M. lychnidis-dioicae* and was only present otherwise in *Mixia osmundae* and *S. roseus* in our comparison; three of these five genes were mapped to the GT2 CAZY family expanded in *M. lychnidis-dioicae* (see below). Both *M. lychnidis-dioicae* and the rusts contain a large number of proteins with the DUF1034 domain; 8 of the 10 *M. lychnidis-dioicae* proteins with a DUF1034 domain also contained a subtilase family protease domain. Phylogenetic analysis of proteins with the DUF1034 domain suggested that independent gene family expansions occurred in

⁴ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s16.xlsx>

different species; the rusts formed a separate clade from *Microbotryum* that is further subdivided, mostly along species lines (**Fig. 2-12**). Domains depleted in *M. lychnidis-dioicae* relative to other basidiomycetes were also identified, including Cytochrome p450, NACHT, and F-box domains (**Table 2-6, Additional file 16⁴**). A more narrow comparison of the three Microbotryomycetes to the other species within the Pucciniomycotina identified additional expansions and depletions common to the species in this lineage. A fungal trichothecene efflux pump (PF06609) gene family was enriched in *M. lychnidis-dioicae* and *S. roseus*, with 16 copies in each genome (**Table 2-6**). By contrast, the alpha-kinase family that is highly expanded in the rusts (Duplessis et al., 2011) was absent in the Microbotryomycetes. The cutinase domain shows a similar conservation pattern; while multiple cutinase genes are found in other biotrophic pathogens, no copies were detected in *M. lychnidis-dioicae*. At lower levels of significance, the CBM1 cellulose binding domain was detected as absent from all species in the Pucciniomycotina with the exception of *R. glutinis* (**Table 2-6, Additional file 16⁵**). More specific analysis of these enriched and depleted domains is presented below.

Table 2-6. Expanded and depleted PFAM domain in *M. lychnidis-dioicae*.

PFAM domain	<i>M. lychnidis-dioicae</i>	<i>S. roseus</i>	<i>R. glutinis</i>	<i>M. osmundae</i>	<i>M. larici-populina</i>	<i>P. graminis-tritici</i>	<i>S. reilianum</i>	<i>U. maydis</i>	<i>M. globosa</i>	Agaricomycetes ⁵ (7)
PF00067.15 Cytochrome P450	10	7	5	14	29	17	15	20	6	95
PF05729.5 NACHT	1	2	1	1	1	1	1	1	0	44
PF01697.20 Glycosyltransferase family 92	5	1	0	1	0	0	0	0	0	0
PF06280.5 Fn3-like (DUF1034)	10	0	0	0	5	9	1	1	0	1
PF00646.26 F-box	7	18	12	12	9	11	11	10	3	43
PF03583.7 Secretory lipase	7	0	0	0	0	0	3	2	6	0

⁵ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s16.xlsx>

PF00734.11 CBM1 Fungal cellulose binding	0	0	3	0	0	0	0	0	0	21
PF02816.11 Alpha kinase	0	0	0	0	79	39	0	0	0	5
PF07690.9 MFS1 Major Facilitator Superfamily	119	110	26	64	90	72	104	98	30	129
PF01753.11 zf-MYND finger	7	11	26	3	7	4	2	2	1	41
PF01083.15 Cutinase	0	0	1	4	21	9	3	4	0	2
PF01670.9 Glycosyl hydrolase family 12	0	0	0	10	14	3	0	0	0	1
PF11327.1 DUF3129	0	0	0	4	13	10	0	0	0	1
PF00097.18 Zinc finger, C3HC4 type	23	11	9	26	22	93	28	24	12	24
PF00080.13 Copper/zinc superoxide dismutase	0	0	0	2	6	18	0	0	0	1
PF00098.16 Zinc knuckle	10	6	1	6	11	59	8	7	5	11
PF06609.6 Fungal trichothecene efflux pump	16	16	5	6	6	3	8	9	2	12
PF12013.1 DUF3505	0	0	0	1	4	16	0	10	0	0
PF03101.8 FAR1 DNA-binding domain	0	0	0	1	14	4	1	1	0	1
PF00083.17 Sugar (and other) transporter	54	56	13	31	46	34	56	46	14	61
PF07738.6 Sad1 / UNC-like	2	1	2	3	24	7	2	1	2	2

PFAM domain	Basidiomycete comparison		Pucciniales comparison [#]	
	p-value	q-value	p-value	q-value
PF00067.15 Cytochrome P450	1.43E-11	6.49E-08	1.28E-02	1
PF05729.5 NACHT	6.20E-08	1.41E-04	4.51E-01	1
PF01697.20 Glycosyltransferase family 92	1.65E-05	2.50E-02	1.99E-02	1
PF06280.5 Fn3-like (DUF1034)	2.54E-05	2.88E-02	1.00E+00	1
PF00646.26 F-box	5.20E-05	4.33E-02	2.72E-02	1
PF03583.7 Secretory lipase	5.72E-05	4.33E-02	1.76E-03	3.07E-01
PF00734.11 CBM1 Fungal cellulose binding	1.21E-04	7.10E-02	6.60E-02	1
PF02816.11 Alpha kinase	1.25E-04	7.10E-02	2.80E-27	1.27E-23
PF07690.9 MFS1 Major Facilitator Superfamily	4.11E-02	1	2.06E-08	4.67E-05
PF01753.11 zf-MYND finger	4.61E-04	2.33E-01	5.16E-08	7.82E-05
PF01083.15 Cutinase	7.90E-02	1	6.49E-07	5.89E-04
PF01670.9 Glycosyl hydrolase family 12	1.70E-01	1	1.08E-06	6.56E-04
PF11327.1 DUF3129	2.62E-01	1	1.08E-06	6.56E-04
PF00097.18 Zinc finger, C3HC4 type	6.82E-01	1	1.16E-06	6.56E-04
PF00080.13 Copper/zinc superoxide dismutase	2.62E-01	1	1.94E-06	9.79E-04
PF00098.16 Zinc knuckle	7.63E-01	1	6.31E-06	2.87E-03
PF06609.6 Fungal trichothecene efflux pump	4.10E-02	1	1.18E-05	4.86E-03
PF12013.1 DUF3505	2.65E-01	1	2.15E-05	8.14E-03
PF03101.8 FAR1 DNA-binding domain	4.09E-01	1	6.65E-05	2.32E-02
PF00083.17 Sugar (and other) transporter	3.28E-01	1	1.71E-04	5.54E-02
PF07738.6 Sad1 / UNC-like	5.87E-01	1	2.45E-04	7.42E-02

^{*}*M. lychnidis-dioicae* compared to all other Basidiomycetes; [#] Microbotryales (*M. lychnidis-dioicae*, *S. roseus*, *R. glutinis*) compared to other Pucciniales (*M. larici-populina*, *P. graminis-tritici*, *M. osmundae*); [^]Agaricomycetes represent average of the 7 species in this group; see Additional file 11⁶ for counts per species

⁶ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s11.xlsx>

Figure 2-12. Phylogenetic tree of DUF1034 proteins.

Protein sequences were aligned with MUSCLE, and a phylogeny was inferred from this alignment using RAxML with the following settings: substitution model: PROTCAT; Matrix name: DAYHOFF; algorithm: (d) Hill-climbing-default and *N. crassa* as the

outgroup. A total of 1,000 bootstrap replicates were performed, with the option Bootstrap random seed (b) 12345, and the percent of replicates shown on the tree nodes. All *M. lychnidis-dioicae* representatives are indicated by the MVLG_ prefix; other prefixes correspond to other species as follows: NCU, *N. crassa*; Srei, *S. reilianum*; UM, *U. maydis*; CNA, *C. neoformans*; CC1G, *C. cinereus*; Pchr, *P. chrysosporium*; LBIC, *L. bicolor*; PIIN, *P. indica*; PGTG, *P. graminis*; Mlp, *M. larici-populina*.

The expansion of the secretory lipases appeared specific to *M. lychnidis-dioicae* within the Pucciniomycotina (**Table 2-6**). Among all other Basidiomycete genomes compared, secretory lipases are also highly represented in *Malassezia globosa* (Ustilaginomycotina), a skin fungus associated with human dandruff and dependent on its host for lipids. *Malassezia globosa* has an additional gene family expansion associated with lipid acquisition; this species has 6 copies of phospholipase C, whereas *M. lychnidis-dioicae* contains only a single phospholipase C protein. Unlike *M. globosa*, *M. lychnidis-dioicae* does not depend on lipids for growth, and contains a predicted fatty acid synthase (MVLG_04698). The secretory lipase family is also present at lower copy number in the two Ustilaginomycotina corn smuts, *Sporisorium reilianum* and *Ustilago maydis*, of which the latter responds to lipids, including corn oils, as part of a developmental switch (Klose et al., 2004). A phylogenetic analysis of the secretory lipases in this comparison revealed that the *M. lychnidis-dioicae* lipases have undergone a lineage specific expansion, as in *M. globosa* (**Fig. 2-13A**). Most lipases were predicted to be secreted, including three of the seven *M. lychnidis-dioicae* lipases. However, four of the seven genes appeared partial based on alignment of the protein sequences; the

missing 5' end from two genes deleted the region containing a secretion signal in paralogous copies. Further refinement of the assembly or transcripts is needed to identify the full length version of these genes or confirm if they are perhaps pseudogenes (see below), and establish their relative location in the genome.

UM, *Ustilago maydis*; sr, *Sporisorium relianum*; MGL, *Malassezia globosa*. Scale corresponds to substitutions per site. B-E. Treatment of mated *M. lychnidis-dioicae* cells with various lipids. B, water-treated mated cell control (Inset: higher magnification showing cells with conjugation bridge, black arrows); C, mated cells treated with commercially-available corn oil (Inset: higher magnification of filamentation); D, mated cells treated with α -tocopherol; E, mated cells treated with phytol. The green arrows show areas of filamentation emanating from mated cells after treatment with each specific type of lipid. B and C, size bars, 25 μ m; D and E, size bars, 50 μ m.

Table 2-7. qRT-PCR validation of secretory lipase expression.

Strain(s)	Condition	Gene	Log fold Change vs. p1A1 Rich	Direction
p1A1	Water	MVLG_07291	0.77	Up
		MLVG_00914	0.327	Up
		MLVG_05549	1.922	Up
		MLVG_07229 & 07284 ^a	1.899	Up
p1A2	Water	MVLG_07291	1.84	Up
		MLVG_00914	-1.308	Down
		MLVG_05549	0.327	Up
		MLVG_07229 & 07284	2.034	Up
Mated (p1A1 x p1A2)	Water	MVLG_07291	1.424	Up
		MLVG_00914	-1.098	Down
		MLVG_05549	0.672	Up
		MLVG_07229 & 07284	1.828	Up
Mated (p1A1 x p1A2)	Phytol	MVLG_07291	0.913	Up
		MLVG_00914	1.2	Up
		MLVG_05549	1.499	Up
		MLVG_07229 & 07284	2.098	Up
Mated (p1A1 x p1A2)	MI-late	MVLG_07291	-1.897	Down
		MLVG_00914	-1.772	Down
		MLVG_05549	0.77	Up
		MLVG_07229 & 07284	-0.564	Down

^aThe primers used could not distinguish between MVLG_07229 & 07284; so results are presented here for both, assuming their expression is similar.

Previous work in *M. lychnidis-dioicae* has shown that mating mixtures of haploid cells produce hyphae in response to phytols and to tocopherols (Castle & Day, 1984). We

therefore exposed haploid sporidial cells (p1A1 or p1A2 strains) or mated cells (p1A1 and p1A2 mixed together) to various lipids or oils, including commercially available corn oil, (+/-)- α -tocopherol, and phytol. Each stimulated filamentous growth of mated mixtures (**Fig. 2-13B-E**), but had no observable effect on haploid cells (unmated). The observation that mated cells also respond to corn oil supports the hypothesis that lipid response may be important for the development of this species. This is possible, as lipids are likely present on the host meristem (Suh et al., 2005). Moreover, we find that at least three lipase genes are differentially expressed when exposed to phytols (a constituent of chlorophyll), compared to similarly treated cells in the absence of phytols (see below; **Table 2-7**).

An important group of proteins identified as being expanded or depleted in *M. lychnidis-dioicae* are those predicted to be involved in cell wall modifications, both in terms of fungal cell walls, but also as might affect host plants. Families of structurally-related carbohydrate catalytic and carbohydrate-binding modules (or functional domains) are described in the CAZy database (www.cazy.org) (Lombard et al., 2014). Such enzymes break down, modify, or build glycosidic bonds. The assignment of the predicted proteins derived from a genome to CAZy families helps to shed light on the particular glycobiological features of an organism (Cantarel et al., 2009). A total of 236 *M. lychnidis-dioicae* protein models were mapped to protein families in the CAZy database based on sequence conservation (percentage identity over CAZy domain length) (**Table A3-2**). The CAZy family profile of *M. lychnidis-dioicae* was then compared to that

recently published for 33 Basidiomycetes (Riley et al., 2014), in order to identify expanded and reduced families (**Table 2-8, Additional file 20⁷**).

With 98 candidate glycosyltransferases (GTs), *M. lychnidis-dioicae* has more than any of the 33 Basidiomycetes used in the comparison (average=69.5; min=41; max=95). Both *M. lychnidis-dioicae* and *Puccinia graminis* contain a candidate fucosyltransferase, which is not present in the other basidiomycetes surveyed. This is compatible with the known presence of fucose in the cell wall of *Microbotryum* (Prillinger et al., 1991; Roelijmans et al., 1998). Two other families expanded in *M. lychnidis-dioicae* include alpha-mannosyltransferases (GT32 and GT62), suggesting that the cell wall includes a larger fraction of alpha-mannan than in other species. In fungi, the synthesized cell wall carbohydrates are frequently remodeled by the action of dedicated glycoside hydrolases and transglycosidases that are found in distinct CAZy glycosyl hydrolase (GH) families. A notable feature of *M. lychnidis-dioicae* is that it has a reduced number β -1,3-glucan cleaving or modifying enzymes of families (GH16, GH72, GH81, and GH128).

Table 2-8. Selected CaZy expansions and depletions.

	GT total	GH total	Beta glucan modification*	Cellulose related**	Xylan related***	GH26 beta-mannanases	Pectin/pectate lyases
<i>Microbotryum lychnidis-dioicae</i>	98	82	4	0	0	4	0
Average basidiomycete#	69.5	179.7	28.9	37.3	14.0	0.5	2.9
<i>Piriformospora indica</i>	65	194	31	98	41	1	12
<i>Ustilago maydis</i>	58	100	26	4	8	1	1
<i>Puccinia graminis f. tritici</i>	81	154	11	15	5	4	4
<i>Melampsora larici-populina</i>	84	169	14	23	9	3	4

*GH16, GH72, GH81, GH128. **GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH131, AA9, and CBM1.

***GH10, GH11, GH30, GH29, GH95, GH51, GH115, GH74. #Average count for 33 Basidiomycete genomes; see Additional file 20⁸.

⁷ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s20.xlsx>

Examination of the other GH families and of the other categories involved in carbohydrate breakdown (PL, CE, AA and ancillary CBM) revealed that the *M. lychnidis-dioicae* genome completely lacks cellulases (no GH6, GH7, GH8, GH9, GH12, GH44, GH45, nor any GH5 with highest sequence similarity to characterized cellulases) (**Table 2-8, Additional file 20^{***}**). In addition no cellulose-targeting lytic polysaccharide monooxygenase of family AA9, nor any broad specificity β -glucanase of family GH131, nor any cellulose-binding module of family CBM1 could be found. This clearly shows that *M. lychnidis-dioicae* does not interact with nor digests cellulose during its interaction with plants, a finding confirmed by the failure of *M. lychnidis-dioicae* to grow on cellulose as a sole carbon source (**Fig. 2-14**). *Microbotryum lychnidis-dioicae* also completely lacks xylanase (no GH10, GH11, GH30), xyloglucanase (no GH74), and the enzymes for the cleavage of side-chains of xylan, xyloglucan and rhamnogalacturonan (no GH29, GH95, GH51, GH115), indicating that these cell wall polymers are not a carbon source for the fungus (**Table 2-8, Additional file 20⁹**). Consistent with this prediction, *M. lychnidis-dioicae* failed to grow on xylan as a sole carbon source (**Fig. 2-14**).

Microbotryum lychnidis-dioicae was able to grow on pectin as a sole carbon source (**Fig. 2-14**), although it does not break down pectin by the action of pectin/pectate lyases, as these enzymes are also absent from the genome (**Table 2-8, Additional file 20⁹**). Instead, the genome harbors a suite of six family GH28 enzymes, which cleave

⁸ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s20.xlsx>

⁹ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s20.xlsx>

polygalacturonic acid after its methylester groups have been removed by the action of six family CE8 pectin methylesterases. This CE8 family is present at high numbers in the two rust fungi and *M. lychnidis-dioicae*; the copy number amplification in *M. lychnidis-dioicae* appears to be due to tandem duplication, with one array of two genes and a second array of four genes. Four of the six CE8 copies have a predicted secretion signal, supporting a potential role in interacting with the host plant. Compared to 33 other basidiomycetes, *M. lychnidis-dioicae* stands out in having a significant expansion of its enzymatic arsenal for the breakdown of β -mannan, a polysaccharide present throughout plants but more abundant in flowers, siliques and stems (Liepman et al., 2007). *M. lychnidis-dioicae* encodes four candidate β -mannanases of family GH26 and a comparison with biochemically characterized enzymes shows that 10 out of its 19 GH5 enzymes also target β -mannan (the other *M. lychnidis-dioicae* GH5 enzymes target β -1,3-glucans (5 proteins), glucocerebrosides (3 proteins) and β -1,6-glucan (1 protein)). This β -mannan digestion arsenal is augmented by the presence of a GH2 enzyme, which shows a strong relatedness to characterized β -mannosidases.

Figure 2-14. Growth of *M. lychnidis-dioicae* on different sole carbon sources, including dextrose, cellulose, xylan, and pectin.

Strains p1A1 and p1A2 were grown overnight in HSS liquid medium (Holliday, 1974) containing 50 mM $(\text{NH}_4)_2\text{SO}_4$ and 2% dextrose. Cells were collected by centrifugation and washed in sterile distilled H_2O , then resuspended in sterile distilled H_2O to an A_{600} of 1. Onto each type of agar, 10 μl samples were spotted of undiluted, and 10-fold, 100-fold, and 1000-fold dilutions of each strain. Growth tests were conducted in duplicate for each strain on each type of medium and plates were incubated at 26 °C for 96 h. Each agar plate contained HSS medium with 50 mM $(\text{NH}_4)_2\text{SO}_4$ and 2% of the indicated carbon source: D, dextrose; Cf, cellulose, fibrous (Spectrum Chemical); Cm, cellulose, microcrystalline (Spectrum Chemical); X, xylan (TCI America); P, pectin (TCI America).

Homogalacturonan is a major component (60%) of plant pectin, and the homogalacturonan degradation pathway is required in several stages of plant development. One of these stages is anther dehiscence, when pollen grains are released; this process requires pectinesterases and polygalacturonases. As indicated above, a total of six CE8 family pectin methylesterases are found in *M. lychnidis-dioicae*, of which four are predicted secreted proteins (MVLG_02682, 04072, 04073, 4074). Part two of the pathway requires polygalacturonase; a total of six *M. lychnidis-dioicae* proteins contain the polygalacturonase GH28 (PF00295) domain, of which MVLG_02498 is highly induced (over 1,000 fold) in MI-late. The homogalacturonan degradation pathway of *M. lychnidis-dioicae* may thus perform a similar role as pollen in anther dehiscence when the flowers bloom, since, during teliospore formation of *M. lychnidis-dioicae*, host pollen is no longer available to perform that function.

Multiple classes of transporters are expanded in *M. lychnidis-dioicae* (**Table 2-6**), enabling uptake of diverse substrates. Major facilitator transporters, sugar transporters, and the fungal trichothecene efflux pump are present at high copy number relative to other Pucciniomycotina. A fungal trichothecene efflux pump, *TRI12*, was first described in *Fusarium sporotrichioides* as part of the gene cluster involved in trichothecene biosynthesis (Alexander et al., 1999); trichothecenes are a group of mycotoxins produced by various species of fungi. As the TRI12 domain is present at high copy number in *M. lychnidis-dioicae* and other Basidiomycetes are not known to produce trichothecenes, this suggests that this domain may have a role in transporting other small molecules.

Sugar transporters also play an important role in virulence of biotrophic plant pathogens, such as *Ustilago maydis* and several species of rust fungi. Specifically, a

plasma membrane-localized sucrose transporter (Srt1) in *U. maydis* facilitated direct utilization of sucrose, thus eluding the plant defense mechanism (Wahl et al., 2010). The HeXose Transporter 1 (Hxt1) gene in the rust fungus *Uromyces fabae* is localized to haustoria to take advantage of that structure for sugar uptake (Voegelé et al., 2001). The *M. lychnidis-dioicae* genome contains a total of 26 potential sugar transporters, with multiple high identity matches to Srt1 and Hxt1, and may fulfill similar roles.

One additional contrast between *M. lychnidis-dioicae* and the two plant pathogenic rust fungi examined suggests a difference in the relative importance of response to superoxides. Reactive oxygen species (ROS), including superoxides or H₂O₂ produced by the host plant, are a canonical part of the defense response to pathogens. *Microbotryum lychnidis-dioicae* is depleted in domains for peroxidase (PF01328) and copper/zinc superoxide dismutase (PF00080); the two other Microbotryomycetes also lack proteins with these domains. Despite the reduced repertoire of such predicted proteins in *M. lychnidis-dioicae* relative to the rust fungi, six proteins (MVLG_00980, MVLG_03089, MVLG_03931, MVLG_02439, MVLG_03568, and MVLG_04684) were identified as containing peroxidase 2 (PF01328), peroxidase (PF00141), redoxin (PF08534), or glutathione peroxidase (PF00255) domains. Of these predicted proteins, only MVLG_03089 was differentially expressed under the conditions examined and was up-regulated in MI-late relative to growth *in vitro* in rich medium. In addition, four predicted proteins with iron/manganese superoxide dismutase domains (PF02777 and PF00081), glutaredoxin (PF00462) or catalase (PF00199) domains were found (MVLG_00659, MVLG_06630, MVLG_06939, MVLG_04131). Finally, pathway analysis via MetaCyc predictions (<http://fungicyc.broadinstitute.org>) suggests that *M.*

lychnidis-dioicae contains components of the glutathione-mediated detoxification pathway: Glutathione transferase (EC 2.5.1.18: MVLG_05985, MVLG_04790) and membrane alanyl aminopeptidase (EC 3.4.11.2: MVLG_03673). However, there appears to be a missing component (3.4.19.9) in this pathway to facilitate formation of an intermediate of a glutathione-toxin conjugate. Biochemical and functional analyses will be required to determine the importance of these predicted enzymes in the ability of the pathogen to survive and flourish in its host.

Secreted proteins (SP) and candidate effectors

A total of 279 secreted proteins (SPs) were predicted in *M. lychnidis-dioicae* and their expression and conservation examined to identify candidates for interacting with the host (**Table 2-9**, **Table A3-3**). Among the 71 SPs that were smaller than 250 amino acids (small secreted proteins, SSPs), 46 were species specific in our comparative set and further do not share sequence similarity (e-value <1e-3) with any protein in the NCBI protein database. SSPs indeed often appear species-specific, likely because they co-evolve rapidly in an arms race with their hosts (Gladieux et al., 2014). Notably, 48 of the SSPs were significantly up-regulated during plant infection (MI-late compared to rich media), but were not differentially expressed when comparing expression on rich and nutrient limited agar (**Table 2-9**), suggesting that these SSPs may play a specific role during plant infection.

Several cysteine-rich multigene families were identified among predicted secreted proteins. In some cases these families include tandemly duplicated genes; the MVLG_04105 family contains 4 members predicted to be SSPs (MVLG_04105, 04106, 04107, and 04096), three of which are adjacent in the genome on the mating-type

chromosome (see below). Although these proteins lack PFAM domains, two of these are induced during infection. An additional family of cysteine-rich proteins with nine members has a subset of six clustered in the genome (MVLG_05513, MVLG_05514, MVLG_05515, MVLG_05533, MVLG_05534, MVLG_05538). Seven of the nine proteins in this family were predicted to be secreted, yet their expression was highly variable, with two up-regulated in nutrient limited conditions and two down-regulated during infection. A small subset of cysteine-rich proteins contains known protein domains. Two proteins (MVLG_02283 and MVLG_02288) contain the cysteine-rich secretory protein family domain (PF00188). In addition, a total of 9 proteins contain the fungal-specific cysteine-rich CFEM domain (PF05730). All of these CFEM proteins were predicted to be secreted; four of these were significantly induced and three were repressed in MI-late relative to rich and nutrient limited agar.

Table 2-9. Properties of predicted secreted proteins.

Protein length*	SP count	Induced in MI-late	Repressed in MI-late	Induced in Water	Repressed in Water	FPKM>1	Highly Expressed only in MI-late
100	15	4	0	0	0	11	3
150	23	5	0	1	0	20	4
200	15	5	0	2	1	14	
250	18	5	1	1	0	14	3
400	62	8	10	8	3	59	2
500	56	8	7	5	2	55	
600	33	4	2	7	0	33	
700	23	7	1	0	0	23	
800	7	0	2	1	0	7	
900	6	1	0	0	0	6	1
1000	13	2	2	0	1	13	
2000	6	0	1	1	0	6	
3000	2	0	0	0	0	2	
Total	279	49	26	26	7	263	

*Protein length is rounded up to the nearest 100 amino acid

To identify genes that could provide a mechanism for linking flower development to fungal development, we compared expression of the predicted secreted proteins with the *S. latifolia* EST library produced from flowers (Moccia et al., 2009). A total of 37 genes share sequence similarity with *S. latifolia* ESTs; ten secreted proteins matched plant ESTs with at least an e-value of e^{-19} . One *S. latifolia* EST (09F02) showed similarity (BLASTx, e-value $< 7e^{-24}$) with two *M. lychnidis-dioicae* proteins (MVLG_02043 and MVLG_02936); the best match, MVLG_02043, encodes a predicted secreted gamma-glutamyltranspeptidase (GGT). Another EST (33C05) shared sequence similarity with two *M. lychnidis-dioicae* proteins (MVLG_00083, MVLG_02276); these in turn share similarity with the expansin family of “ripening related” proteins and were down-regulated during either growth on nutrient limited agar or late in infection (MI-late). The precise function of plant expansins is poorly defined at the molecular level, and the predicted function of the similar fungal proteins is even less well established. However, one such expansin-related protein in *Laccaria bicolor* was recently found to be present specifically in the extracellular matrix (ECM) of symbiotic tissues and localized within the fungal cell wall (Veneault-Fourrey et al., 2014).

Two cysteine-rich secreted proteins (MVLG_02288 and MVLG_02283) matched a *S. latifolia* flower EST annotated as having similarity with the plant PR-1 class of pathogenesis related proteins (PRs); these are proteins defined as encoded by the host plant but induced only in pathological or related situations (possibly of non-pathogenic origin). To be included among the PRs, a protein must be induced upon infection but not necessarily in all pathological conditions (Antoniw et al., 1980). Another *S. latifolia* gene of interest, *SLM2*, is expressed in the stamens of smut infected flowers but not uninfected

flowers (Kazama et al., 2005); four *M. lychnidis-dioicae* proteins showed BLAST similarity with e-value less than e-6. All of the hits were hypothetical proteins that contained the SRF-type transcription factor domain (PF00319), and three of the four were predicted to be targeted to the nucleus (MVLG_04297, MVLG_06278, and MVLG_07052).

Response to oxidative environments

Laccase-like multi-copper oxidase proteins are capable of degrading phenolic compounds like polymeric lignin and humic substances (Baldrian, 2006) and may be involved in the interaction of fungal pathogens with their host plants. Four proteins in *M. lychnidis-dioicae* contain the three multi-copper oxidase (MCO) domains (PF07731, PF07732, PF00394). Two MCO proteins were predicted to be secreted, and another MCO was predicted to have a GPI-anchor to the membrane. The fourth MCO was a membrane-anchored protein (MVLG_03092), with the N-terminus of the polypeptide outside the cell.

The glyoxal oxidase catalyzes the oxidation of aldehydes to carboxylic acid and is an essential component of the extracellular lignin degradation pathway of the root rot fungus, *Phanerochaete chrysosporium*. Of a total of seven *M. lychnidis-dioicae* proteins that contain the glyoxal oxidase N-terminal domain (PF07250), two are predicted to be secreted and four have a predicted GPI-anchor. While two of these glyoxal oxidase genes are adjacent in the genome, they do not form a gene cluster with any MCO as observed at the lignin peroxidase gene cluster in *P. chrysosporium*.

Genes similar to plant hormone synthesis genes

Since *M. lychnidis-dioicae* infection of female *S. latifolia* hosts can alter normal flower development so as to produce pseudomale flowers, we investigated whether the genome might contain genes for pathways that could be associated with such changes. The predicted *M. lychnidis-dioicae* protein database was examined for components of biosynthesis pathways of eight plant hormones (**Additional file 23¹⁰**), as well as for other signaling pathways that could have an impact on host gene expression or development. Based on sequence similarity to components for these pathways from plants and microbes, as well as additional confirmation of some complete pathways using MetaCyc predictions (<http://fungicyc.broadinstitute.org>), we found evidence that *M. lychnidis-dioicae* encodes enzymes that could participate in hormone biosynthetic pathways, such as polyamine biosynthesis pathways, which produces compounds known to play developmental and stress-response roles in plant physiology (Takahashi & Kakehi, 2010). If these enzymes are, in fact, promoting hormone biosynthesis pathways, one possible explanation is that precursors for these pathways are provided by the plant. Alternatively, the potential components of these pathways we have identified are used by the fungus for functions other than manipulating host development. For example, cytokinin degradation could be mediated by a predicted FAD-oxidase (MVLG_04134), if this enzyme can function as a cytokinin dehydrogenase (EC 1.5.99.12); however, this predicted protein most closely resembles other fungal D-lactate dehydrogenases based on sequence similarity. Similarly, a predicted 2 β -dioxygenase (EC1.14.11.13) (MVLG_00840) could

¹⁰ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s23.xlsx>

be involved in gibberellin inactivation via hydroxylation, although this protein falls more generally into the 2OG-Fe(II) oxygenase superfamily.

Another potential set of pathways for plant signaling involves the production of glycerol lipids, such as diacylglycerol (DAG) or triacylglycerol (TAG). Many studies have demonstrated the importance of compounds like DAG in mammalian signaling (Brose et al., 2004), and DAG is important in pollen tube elongation in some plant species (Dong et al., 2012). Such pathways often involve the action of phospholipase C; a predicted phospholipase C gene (MLVG_07108) and potential phospholipase A (MVLG_03207, MVLG_03384, MVLG_4789) and D (MLVG_01917, MVLG_03610) homologues are found in the genome. Therefore, the organism contains the proteins necessary for synthesizing 1,2-diacylglycerol, 1,2-diacyl-*sn*-glycerol-3-phosphate, 2-lysophosphatidylcholine, and 1-lysophosphatidylcholine. Based on MetaCyc prediction, *M. lychnidis-dioicae* possesses the requisite components of the CDP-diacylglycerol biosynthesis I pathway. In addition, a number of the secretory lipases whose PFAM domain is enriched in *M. lychnidis-dioicae* (PF03583, see above) are implicated in these phospholipase pathways.

Figure 2-15. Model of *M. lychnidis-dioicae* interactions with its host.

The potential pathways identified in *M. lychnidis-dioicae* based on inventory of the genome were used to predict products potentially secreted or taken up by the fungus that could affect host development (see text for more detailed description). GA, gibberellic acid.

Gene expression changes during infection and in response to lipids

To focus on genes potentially important for infection and interaction with the host, we identified genes whose expression was altered in MI-late. A total of 1,254 genes were differentially expressed in MI-late compared to either rich or nutrient limited agar; of these, a common set of 307 genes were induced in MI-late and 126 were repressed in MI-late compared to both other conditions (corrected p-value <0.001, **Additional file**

24¹¹, Methods). Of the 138 genes in both comparisons with a predicted PFAM domain, transporter domains were most frequently observed; a total of 20 of the 138 functionally assigned proteins corresponded to transporters, with MFS, sugar, OPT, and amino transporters each represented by two or more genes (**Table 2-10**). Carbohydrate active enzymes, kinases and transcription factors were also highly represented in these MI-late induced genes.

Table 2-10. PFAM domains enriched in MI-late induced genes.

PFAM domain	MI-late Induced genes	Other genes	Fisher pvalue	Corrected pvalue
PF07690.9 Major Facilitator Superfamily	15	113	8.42E-07	0.0011
TIGR00879 MFS transporter, sugar porter (SP) family	6	14	7.96E-06	0.0051
PF07250.4 Glyoxal oxidase N-terminus	4	4	2.83E-05	0.0091
PF09118.4 Domain of unknown function (DUF1929)	4	4	2.83E-05	0.0091
PF00295.10 Glycosyl hydrolases family 28	3	3	0.0003	0.0838
PF00128.17 Alpha amylase, catalytic domain	3	4	0.0006	0.0962
PF01590.19 GAF domain	3	4	0.0006	0.0962
PF03169.8 OPT oligopeptide transporter protein	4	12	0.0006	0.0962
PF05978.9 Ion channel regulatory protein UNC-93	2	0	0.0007	0.0962

Several different classes of transporters were transcriptionally induced during infection, potentially promoting the uptake of small molecules and nutrients from the host plant. Significant enrichments include domains found in MFS transporters (Fisher's exact test, corrected p-value < 0.001) and sugar transporters (p < 0.005), the largest classes of transporters in *M. lychnidis-dioicae* and many fungi. The oligopeptide transporter protein was also enriched at a lower level of significance (p<0.1) and four predicted proteins with

¹¹ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s24.xlsx>

this domain (MVLG_03106, MVLG_07217, MVLG_03161, and MVLG_00149) were up-regulated in MI-late. An MFS domain-containing protein of note up-regulated in MI-late was a nitrite transporter (TIGR00886.2; MVLG_00642); this gene is linked to two other genes associated with nitrate assimilation, a nitrite reductase (MVLG_00638) and a nitrate reductase (MVLG_00637). All three genes involved in nitrate assimilation were significantly up-regulated during infection.

In evaluating the expression of the 279 proteins predicted to be secreted, 48 were induced *in planta*. Several cysteine-rich secreted proteins were induced during infection. A pair of linked cysteine rich small secreted proteins (MVLG_04106 and MVLG_04107) was induced during infection. Four other proteins (MVLG_00115, MVLG_00802, MVLG_00815, and MVLG_00859) with the Cys-rich CFEM domain-containing family (PF05730) were significantly induced in infection. These proteins are good candidates for being potential effectors, based on proteins with similar properties that have been shown to be effectors in other systems (Stergiopoulos & de Wit, 2009).

Cell wall degrading enzymes may play a role in the infection, particularly during the stage where the fungus causes necrosis of host plant tissue. A total of nine glycoside hydrolases were up-regulated in MI-late; among GH families, GH28 polygalacturonase proteins are mostly highly enriched during infection ($p < 0.08$). Polygalacturonase is required for the second part of the homogalacturonan pathway implicated in pollen dehiscence. MVLG proteins that contain a glyoxal oxidase domain (PF07250) were also significantly enriched ($p < 0.009$) in genes induced during MI-late.

Mated cells of *M. lychnidis-dioicae* respond to corn oil, in addition to the other lipids previously reported (Castle & Day, 1984). This supports the hypothesis that lipid

response may be important for the development of this species. Although we observed no alteration of phenotype for un-mated haploid cells treated with lipids, three predicted cytoplasmic proteins with a secretory lipase domain (MVLG_07229, 07284, and 07291) were highly induced in haploids grown on nutrient limited agar compared to rich media or MI-late infections. To validate the expression levels from RNA-Seq data, relative levels for lipase genes were measured for mated and unmated cells, grown in nutrient limited, rich media, or treated with phytol by qRT-PCR (**Table 2-7**). Notably, when mated haploid cells were treated with phytol, after 12 h treatment, two genes (MVLG_00914, MVLG_05549) displayed substantial up-regulation, while the remaining three either increased slightly or decreased (**Table 2-7**). This could reflect priming of cells that are ready to mate for plant cues that would ultimately lead to stable dikaryon formation after successful mating.

Discussion

This analysis highlights genomic features of *M. lychnidis-dioicae* that reflect its particular lifecycle. *M. lychnidis-dioicae* grows as a biotroph, similar to the rust fungi, for most of the plant infection cycle. Notably in its capacity as a castrating pathogen, *M. lychnidis-dioicae* also causes necrosis that appears to be limited to developing flowers. Our analysis revealed that gene content contains a different profile than purely biotrophic or necrotrophic plant pathogens. This included a global number of CAZymes that is larger than other biotrophic fungi (**Additional file 20**¹²); however, at the same time it shows complete loss of many CAZyme families that target the plant cell wall. This is consistent with a primarily intercellular colonization pattern of host apical meristems (Bauer et al., 1997), though it raises questions about how appressorium penetration is accomplished.

Necrotic growth stages may be enabled by subtilases, laccases, and copper radical oxidases, which are ligninolytic enzymes (Palmieri et al., 2001; Palmieri et al., 2003; Whittaker et al., 1999); subtilases in particular may play a key role in regulating the activities of laccase (Baldrian, 2006). By contrast, certain CAZymes present in the biotrophic rust fungi are absent in *M. lychnidis-dioicae*; in particular the absence of cutinases suggests that penetration of the plant surface by *M. lychnidis-dioicae* does not require cutin degradation. This may reflect the fact that the normal portal of entry for infection is via the flower, a tissue that poses less of a barrier to the fungus. Of the plant polysaccharides that constitute a carbon source for many fungi, *M. lychnidis-dioicae* has

¹² Available for download at
<http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s20.xlsx>

lost the ability to digest cellulose, xylan, xyloglucan, and the highly substituted forms of pectin (rhamnogalacturonan). Retention of enzymes that break down polygalacturonic acid and β -mannan, components of pollen tubes and flowers, respectively, illustrates the high degree of specialization of this fungus.

Whereas smut fungi generally cause necrosis as a space-making process during sporulation (Luttrell, 1981), *Microbotryum* anther smuts are more aptly characterized as a growth-altering parasite (Cashion & Luttrell, 1988). At the stage of anther development, *M. lychnidis-dioicae* causes abortion of pollen production, and replacement by the diploid teliospores for dispersal by pollinator species (Schäfer et al., 2010). Moreover, in the female, atrophy of pistils occurs during infection. There are a number of candidate pathways that might participate in pollen tube elongation or in blocking this process. Aspartyl proteases are involved in pollen tube elongation or prevention by *Metschnikowia reukaufii* (Eisikowitch et al., 1990); seven candidate aspartyl proteases are predicted in *M. lychnidis-dioicae*. Additionally, consistent with transcriptional up-regulation of some component enzymes for the homogalacturonan degradation pathway of *M. lychnidis-dioicae*, this pathway may take over the role of pollen in anther dehiscence when the flowers bloom, since during teliospore formation of *M. lychnidis-dioicae*, host pollen is no longer available to perform that function (**Fig. 2-15**).

Finally, since *M. lychnidis-dioicae* appears to lack a large repertoire for dealing with host-generated defenses that utilize reactive oxygen species (ROS) one could expect that the interaction with its host normally does not elicit such host responses or the fungus actively down-regulates them. Paradoxically, *M. lychnidis-dioicae* also appears to secrete some proteins that would serve to bring on this plant response by serving as

prooxidants (*e.g.*, secreted gamma-glutamyltranspeptidases). Further elucidation of the roles of the predicted peroxidases superoxide dismutases and glutathione-mediated detoxification pathway components will require functional analyses to evaluate their biological importance, if any, in the interaction of the pathogen with its hosts.

Secretory lipases represent one of the most significantly expanded gene families in *M. lychnidis-dioicae* compared to the other fungi examined. While haploid cells show no outward phenotype alteration in response to lipids like phytol, several secretory lipases of *M. lychnidis-dioicae* in haploid cells are up-regulated on nutrient limited agar. We propose that regulation of secretory lipases primes haploid cells so that, after mating, they can respond to the appropriate plant-derived cues (including lipids) to progress to the next developmental stage, stable dikaryotic hyphae. In fact, most of the secretory lipases we investigated were up-regulated in mated cells and when such cells were exposed to phytol (**Table 2-7**).

The identification of genes induced during infection (MI-late) suggests their involvement in host invasion and evasion of physical and chemical defense systems of the plant. A role for CAZymes, including pectin methylesterases and GHs, during plant infection in *M. lychnidis-dioicae* is further supported by increased transcription in the MI-late sample. Analysis of the gene expression profile of both wheat stem and poplar rust (*P. graminis* and *M. larici-populina*) also found that many CAZyme genes related to cell wall degradation were up-regulated during plant infection (Duplessis et al., 2011). In addition to the CAZymes, *M. lychnidis-dioicae* shows significant induction of diverse transporters during plant infection, which may be critical for uptake of small molecules during biotrophic growth.

As in other plant pathogenic fungi, candidate effectors in *M. lychnidis-dioicae* were predicted based on predicted localization, expression during infection, and sequence conservation. Notably, SSPs are located closer to TEs than other protein coding genes, suggesting that this could impact SSP expression or duplication. TEs probably play a role in the expansion of such a family. Indeed they contribute to genome rearrangements and gene duplications (Biémont, 2010). In *Fusarium oxysporum* f. sp. *lycopersici*, effector genes are present on chromosomes or regions enriched for DNA transposons (Schmidt et al., 2013) [90]. Some secreted proteins are predicted to act on host cell walls and proteins, either for the remodeling of host development in the flower or for acquisition of additional nutrients by the fungus.

In assembling sequence of the a_1 mating-type chromosome, we characterized how the content of this allosome differs from autosomal regions, contrasting the non-recombining regions of the mating-type chromosome with the pseudo-autosomal regions (PARs) capable of recombination and with autosomes. Both the lower gene density and higher transposable element content in the non-recombining region of the mating-type chromosome relative to autosomal regions are consistent with a reduced efficiency of purifying selection due to the suppression of recombination, as occur on non-recombining sex chromosomes (Bachtrog, 2003). The two recombining PARs of the mating-type chromosome displayed TE content and gene density more similar to autosomes than the non-recombining part of the mating-type chromosome. By contrast, we observed no difference in codon usage nor in the GC content at third codon positions between autosomes and the mating-type chromosome, though this has been observed in the non-recombining part of the fungal mating-type chromosome of *Neurospora tetrasperma*

(Whittle et al., 2011). Overall the maintenance of homologous meiotic pairing and recombination in PAR regions may render them more similar to autosomes than allosomes with respect to evolutionary forces of selection and drift. However, their physical linkage to the non-recombining region of mating-type chromosomes suggests intermediate modes of evolution (Otto et al., 2011).

Materials and Methods

***Microbotryum lychnidis-dioicae* lineage(s) and *Silene latifolia* host(s)**

The *focal* lineage of *M. lychnidis-dioicae* for this work is the most studied in the context of disease ecology (“Lamole strain”: GenBank I00-15Lamole.1; (Antonovics et al., 2002; Hood et al., 2004)) and belongs to the recently-refined species designation *M. lychnidis-dioicae*, parasitizing *Silene latifolia*. From the original isolate, haploid sporidial strains were generated via micromanipulation of the meiotic products from a single tetrad, yielding the strains Lamole p1A1 and p1A2 that differ in electrophoretic karyotypes only in the mating-specific chromosome. For the work in this report, the haploid p1A1 strain (mating-type a_1) was used as the source of the *focal* genome. Additionally, the *focal* genome contains size-heteromorphic sex chromosomes that share many features with sex chromosomes in plant and animal systems (Hood et al., 2004). The corresponding a_2 strain, p1A2 was used together with its partner strain p1A1 in plant infections and in RNA-Seq analysis.

High molecular weight DNA preparation

Microbotryum lychnidis-dioicae Lamole p1A1 was grown on yeast peptone dextrose media (YPD; 1% yeast extract, 10% dextrose, 2% peptone, 1.5% agar) at room temperature for 5 days and ultimately extracted using a phenol chloroform isoamyl extraction method (Luo & Perlin, 1993). Harvested fungal cells were ground into fine powder using liquid nitrogen and resuspended in OmniPrep Genomic Lysis Buffer (G-Biosciences, cat no: 786-136) according to manufacturer’s recommended tissue to reagent ratio. The sample was heated in a 55-60 °C water bath for 15 min after extensive vortexing. Chloroform was added to the sample after allowing it to cool to room

temperature. Using wide bore tips thereafter, 3-4 extractions using phenol chloroform isoamyl (25:24:1) solution were performed, followed by a final extraction with chloroform isoamyl (24:1) solution. Nucleic acid was then precipitated and the pellet was rinsed twice with ice-cold 70% ethanol and then air-dried. The pellet was rehydrated using Tris-EDTA buffer (pH 8.0) (100 µl per 100 mg of ground tissue powder used) and treated with RNase (*Longlife* RNase, 5 mg/ml; G-Biosciences); 1 µl of RNase was added for every 100 µl of TE buffer used.

RNA isolation

Haploid cells

Haploid fungal cells of either Lamole p1A1 or p1A2 grown separately under rich conditions for 5 days on yeast peptone dextrose media (YPD; 1% yeast extract, 10% dextrose, 2% peptone, 2% agar) at room temperature were harvested for RNA extraction. RNAs were checked for quality using a Bioanalyzer (Agilent). The RNAs were then pooled in equal quantity (in terms of mass) based on the Bioanalyzer quantification. The same procedure was also performed for the haploid cells grown separately on 2% water agar for 2 days, to compare the gene expression when haploid cells were subjected to nutrient free environment without the mating partner. Again, haploid cell samples, p1A1 and p1A2, were extracted as independent samples, and then mixed in equal proportion for RNA sequencing.

Host plant infection for RNA-Seq (MI-late stage)

Silene latifolia seeds (harvested in Summer 2009 from Lamole, Italy) were sterilized and hydrated by soaking them in a sterilizing solution (40% household bleach, 20% absolute ethanol and 1 drop of Triton X-100 as surfactant per 50 ml of solution) and

washing five times in sterile distilled water, for 2 min per wash with constant agitation. Each seed was then individually planted in closed milk jars on sterile 0.3% phytagar (Life Technologies), 0.5x MS (Murashige and Skoog) salts (Sigma-Aldrich) and 0.05% MES (2-(N-morpholino)ethanesulfonic acid) buffer (FisherBiotech). Each jar was placed at 4 °C for 5 days to synchronize germination. The jars were then transferred to a 20 °C growth chamber with 13 h of fluorescent light daily. Germination starts within 3 days with the appearance of the radicle. When the seedlings were 15 days old, they were transplanted into 2" square pots filled with Sunshine MVP Professional Growing Mix (Sun Gro Horticulture Canada Ltd, cat no. 02392868) soil and replaced into the growth chamber. Humidity was kept high initially using dome covers and flood trays. Seedlings were gradually exposed to chamber environment for increasing amounts of time daily in order for the seedling to harden and adapt to the lower humidity. The plants were transplanted to 4" round pots when they began to bolt at about 30 days old. They were further transplanted into 7" round pots when they had almost attained maximum height or when the volume of soil was not sufficient to provide hydration requirement for the plant. The plants were watered every other day with 100-ppm fertilizer (Peters Professional® 15-16-17 Peat-Lite Special, Formula no: S12893).

To infect the host plants, mated cells were prepared as follows. Haploid cells grown on rich media were harvested and resuspended in distilled water, adjusted to a concentration of 1×10^9 cells/ml in equal proportion before being spotted onto nutrient-free solid agar media (2% agar) in 50 µl spots. The plates were allowed to dry and then incubated at 14 °C for about 48 h. Cells were inspected for conjugation tubes under the microscope and then 5 µl of 1×10^6 cells/ml resuspended in distilled water with anionic

surfactant was pipetted onto the floral meristem when the cotyledon was fully developed (11-12 days). Infection was determined by the consistent blooming of fully smutted flowers. The floral buds were staged (Toh & Perlin, 2015) according to previous literature (Grant et al., 1994) under a dissecting scope (Nikon, Model: SMZ-U), and parts of the floral buds were measured with a glass stage micrometer (Imaging Research, Inc.).

Tissue originating from host plants was collected in RNAlater RNA stabilizing reagent (QIAGEN, cat no: 76106) and left at 4 °C overnight until sufficient tissue had been collected for the RNA extraction. The solution was removed before storing the sample at -80 °C. For infected male plants, we collected floral tissue from buds ranging in size from 4 mm to fully open smutted flowers. These tissue samples were pooled to yield the source for ‘MI-late’ RNA used in RNA-Seq analysis. Thus they provide a pooled average picture of gene expression for this size range of infected tissue.

All RNA samples were extracted using the RNeasy Plant Mini Kit (QIAGEN, cat no: 74904) according to the manufacturer’s instructions. DNase treatment was performed using Ambion’s TURBO DNA-free (Applied Biosystems, cat no: AM1907), also according to manufacturer’s instruction. For quality assessment before Illumina sequencing, 5 µg of DNase-treated RNA was reverse transcribed with SuperScript III First Strand Synthesis System for RT-PCR (Life Technologies, cat no: 18080-051). PCR was performed using TaKaRa Ex Taq Hot-Start DNA Polymerase (Takara, cat. no: RR001B) using 25µl reaction volume. To check for DNA contamination and possible inhibitory substances in the RNA, three sets of housekeeping primers (Eurofins/MWG/Operon) were used. Amplification of a region of the *S. latifolia* partial *wdrlx* gene for a putative WD-repeat protein (GenBank IDs Y18519, Aj310656) was

used to assess host cDNA and contaminating genomic DNA; the forward primer, 5'-CTCTGCTGGAGGTGGAACAT-3' and reverse primer, 5'-AGCACTGAACACCCCAACTT-3'; in this case a 253 bp fragment would be produced for cDNA, vs. a 335 bp fragment for genomic DNA. Targeting the *M. lychnidis-dioicae mepA* gene, we used as forward primer, 5'-CTTTTGCGTAGGAAGAATGC-3' and as reverse primer, 5'-AGCACTGAACACCCCAACTT-3'; this combination yielded a 532 bp fragment from cDNA, compared with a 1039 bp fragment from genomic DNA. The other primer combination targeted the *M. lychnidis-dioicae* beta-tubulin gene, with forward primer, 5'-CGGACACCGTTGTCGAGCCT-3', and reverse primer, 5'-TGAGGTCGCCGTGAGTCGGT-3', yielding a 150 bp fragment from cDNA compared with a 215 bp fragment from genomic DNA. The PCR program was 30 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C for 35 cycles. RNA quality was also evaluated using an Agilent BioAnalyzer; all samples had RNA integrity number scores of at least 7.8, indicating highly intact RNA.

Treatment of cells with lipids

Haploid fungal cells of Lamole p1A1 and p1A2 were grown separately under rich conditions for 5 days on YPD at room temperature, then harvested into sterile distilled water. The concentration was adjusted and resuspended in equal proportions in each type of medium, to achieve a final concentration of 1×10^9 cells/ml.

To allow better solubility of the lipids, 50% ethanol was used as the solvent for the lipids. We used 1% corn oil (Carlini), ($\pm$)- α -tocopherol (Sigma-Aldrich, cat no: T3251-5G), and phytol (Sigma-Aldrich, cat no: P3647), dissolved in the solvent and used as the resuspension media for the fungal cells. The mixtures were then spotted in 50 μ l

spots onto 2% water agar and allowed to mate for 2 days at 14 °C. The cells were then observed under the microscope for conjugation tubes and filamentous structures. The solvent served as the control media to ensure that changes in phenotype were not due to the ethanol present.

Genome and transcriptome sequencing, assembly, and annotation

For genome sequencing, we constructed three libraries (**Table 3-1**) with different insert sizes and sequenced each using 454 Technology. The reads were assembled with Newbler (version MapAsmResearch-04/19/2010-patch-08/17/2010). The total assembly size of 26.1 Mb in scaffolds includes 99.6% of bases of at least Q40 quality; gaps encompass 3.45% of the total scaffold length.

For RNA-Seq, we purified polyA RNA and constructed a strand-specific library for each sample as previously described (Levin et al., 2010; Parkhomchuk et al., 2009) and sequenced each with Illumina technology generating 76 base paired reads. Across the three libraries, 96% of reads met the Illumina Passing Filter (PF) quality threshold. Read alignment rates to the genome varied between three libraries; for the rich and nutrient limited samples, 90% or 89% of reads aligned respectively; for the MI-late sample only 23% of reads aligned. This was expected, as these samples also contain the host *Silene* RNAs. To assemble transcripts for use in annotation, RNA-Seq reads were aligned to the assembly with BLAT (Kent, 2002), and then assembled using Inchworm (Grabherr et al., 2011) in the genome-guided mode.

To predict genes, we first generated a high confidence training set of 775 transcripts of at least 900 nucleotides (nt) using Genemark (Borodovsky & Lomsadze, 2011) and the assembled RNA-Seq data. This was used to train Augustus (Stanke et al.,

2004) and GlimmerHMM (Majoros et al., 2004). RNA-Seq data was processed by PASA (Haas et al., 2008) to generate longer transcripts, and ORFs of at least 600 nt were predicted. Available ESTs from Genbank and Microbase were also utilized. EVM (Haas et al., 2008) was then used to select a preliminary gene set from the *ab initio* gene calls (Augustus, Genemark, GlimmerHMM, and SNAP), Genewise (Birney et al., 2004), ESTs, PASA ORFs, and the training set, with the highest weight given to the RNA-Seq based PASA ORFs. The EVM gene set was compared to the PASA ORFs, and non-repeat genes found only in the PASA set were added to the gene set. Finally, PASA was run on the final gene set to all updates of all gene structures with the RNA-Seq and incorporate alternatively spliced transcripts. Genes likely corresponding to repeats were filtered out using TransposonPSI (requiring 1e-10 and 30% overlap), PFAM domains, similarity to repetitive elements and 7 or more BLAST hits to other genes in the set. Genes with flagged features (proteins ≤ 50 aa, internal exons ≤ 6 nt, introns ≤ 20 nt, introns ≥ 1500 nt, exons spanning gaps in the assembly, internal stop codons, overlapping other coding sequences, overlapping ncRNAs (tRNAs, rRNA, or other)) were manually reviewed and corrected where supported by the evidence. Gene names were assigned with the locus prefix MVLG.

The completeness of the gene set was evaluated by examining the conservation and completeness of core eukaryotic genes (CEGs, (Parra et al., 2007)). We compared the gene set of *M. lychnidis-dioicae* and of the 18 other fungal genomes used in comparisons to the CEGMA set, and identified BLAST hits above and below the recommended 70% coverage threshold (**Fig. 3-2**). A tool for streamlined analysis and visualization of

conservation of CEGs is available on SourceForge
(<http://sourceforge.net/projects/corealyze/>).

Differential Expression Analysis

We used differential expression analysis scripts in the Trinity pipeline (Grabherr et al., 2011; Haas et al., 2013) to process RNA-Seq data generated from the three conditions (nutrient limited, rich, and MI-late). Briefly, we first extracted protein coding gene sequences from the *M. lychnidis-dioicae* genome sequence based on coordinates of gene models, and added 100 bases of flanking sequence on each side to approximate UTRs. Then the RNA-Seq reads from each of the three samples were aligned to the extracted coding sequences using bowtie (Langmead et al., 2009). The alignment files were used to quantify transcript abundances by RSEM (Li & Dewey, 2011). Differential gene expression analysis was conducted using edgeR with TMM normalization (Kadota et al., 2012; Robinson et al., 2010) using a corrected p-value (Benjamini & Hochberg, 1995) cutoff of 1e-3. In comparing all pairs of the three conditions, a total of 1,413 genes were differentially expressed across the comparisons (**Additional file 24**¹³).

TE detection and annotation

Two pipelines from REPET package (<http://urgi.versailles.inra.fr/tools/REPET>) were run on the *M. lychnidis-dioicae* contigs. The TEdenovo pipeline (Flutre et al., 2011) was used to search for repeats in the genome. The first step uses Blaster with the following parameters [identity>90%, HSP (High Scoring segments Pairs) length >100b & <20Kb, e-value ≤ 1e-300]. HSPs found were clustered by 3 different methods: Piler

¹³ Available for download at
<http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s24.xlsx>

(Edgar & Myers, 2005), Grouper (Quesneville et al., 2005) and Recon (Bao & Eddy, 2002). Multiple alignments (MAP) of 20 longest members of each cluster (918 clusters) containing at least 3 members were used to derive a consensus. Consensus sequences were then classified based on their structure and similarities against Repbase Update (v15.11) (Jurka et al., 2005) before removing redundancy (Blaster + Matcher). Consensus sequences without any known structure or similarity were classified as "Unknown".

The library of 425 classified consensus sequences provided by the TEdenovo pipeline was used to annotate TE copies in the whole genome using TEannot pipeline (Quesneville et al., 2005). Annotation is based on 3 methods (Blaster, Censor, RepeatMasker). HSPs provided were filtered and combined. Three methods (TRF, Mreps and RepeatMasker) were also used to annotate SSR. TE duplicates and SSR were then removed. Finally a “long join procedure” (Flutre et al., 2011) was used to address the problem of nested TEs. This procedure finds and connects fragments of TEs interrupted by other TEs inserted more recently to build a TE copy. The nesting patterns of such insertion must respect the three constraints: fragments must be co-linear (both on the genome and the same TE consensus reference), have the same age and separated by younger TE insertion. The identity percentage with the reference consensus is used to estimate the age of a copy. Using results of this first TEannot pipeline, we filtered out 111 consensus sequences without full-length copy in the genome. A copy may be built using one or more fragments joined by the TEannot long join procedure. We ran a second TEannot using the 306 consensus elements remaining after filtering out TE consensus without any full-length copy.

We also used gene prediction and proceeded to manual curation in order to improve TE annotation. We removed TE copies of consensus that were identified as host genes. Indeed, these consensus built from family of repeats containing at least 3 members and classified as unknown by the TEdenovo pipeline has been predicted as host genes belonging to multigenic families. We also filtered out TE copies not satisfying the criteria (identity > 0.8 & length>150 & identity*length>150) and those corresponding to low complexity region of the consensus Mivi-B-R219-Map20_classI-LINE-incomp very highly represented in the genome included in predicted genes. The few copies just over these thresholds were manually removed, depending on their location in genes and evidence of the gene (PFAM domain not related to TEs).

Search for signature of transition type (C to T) mutation bias

We performed pairwise alignments between each copy and respective consensus to finally provide multiple alignments for each family (consensus) using in-house scripts. TE copies with less than 80% of identity with consensus and smaller than 400 bp were filtered out. We also filtered out TE families with less than 5 sequences in the multiple alignments. RIPCAL (Hane & Oliver, 2008) was run on each multiple alignment to count both potential single mutations (transitions and transversions) and di-nucleotide target used in all possible transitions. Results were analysed using in-house R scripts to select most reliable mutated copies (if transition rate > 2 * transversion rate). For 40% of copies exhibiting a transition mutation bias (of 2298 total copies, 179 consensus families (**Fig. 3-8**)), we considered that dinucleotide targets (CA+TG¹, CC+GG¹, CG+CG¹, CT+AG¹; ¹ for reverse complement) were preferentially used if they represent a minimum of 30% of the addition of the four possible. We expect 25% of each if they are equiprobable.

Measurement of distance between genes and TEs

We computed the distance from each gene to the closest TE (case 1), or from each TE to the closest gene (case 2) using `getDistance.py` from S-MART package (Zytnicki & Quesneville, 2011). Only distances up to 10 kb were considered. For case 1, we also compared the subset of genes encoding predicted secreted proteins with the set of all other genes for different classes of distance intervals. For the case 2, we compared different TE categories in two classes of distance intervals (<1 kbp and >1 kbp). A χ^2 test of homogeneity (Pearson's chi-squared) was computed to test that the observed difference between the sets did not occur by chance (p-value < 0.05). The graphics and statistical test were performed using in-house R-Scripts.

Identification of the mating-type chromosome supercontigs

Using the same haploid genotype from which the whole genome was sequenced, DNA enriched for the a_1 mating-type chromosome was isolated from agarose gels after pulsed-field electrophoresis. With this technique, the isolated bands could include of small amounts of autosomal fragments that co-migrate with mating-type chromosomes, though these preparations have been shown to be strongly enriched for mating-type chromosome DNA (Hood et al., 2013). The isolated DNA was amplified by whole genome amplification (REPLI-g kit, QIAGEN). The DNA was sequenced using 2- and 5 kb-insert size mate-paired libraries and 454 technology version Titanium (www.rockefeller.edu). Assembly of non-duplicated reads and excluding autosomal contamination yielded ~20-fold coverage.

The assembly was compared to the a_1 whole haploid genome sequence using NUCmer (<http://mummer.sourceforge.net/>) to validate assemblies and identify scaffolds

corresponding to the mating-type chromosome in the whole genome assembly. The regions corresponding to autosomal contamination were identified by uneven and low read coverage and were excluded from further analyses; mitochondrial DNA was also excluded (GenBank NC_020353). All whole-genome scaffolds with more than 20-fold depth of the enriched mating-type chromosome sequence were confidently assigned to the mating-type chromosomes (**Additional file 11**¹⁴). It was not possible to anchor the scaffolds onto the available optical map of the a₁ mating-type chromosome (Hood et al., 2013) due to the small sizes of the contigs relative to spacing of the restriction enzyme cut sites in the map. Annotation for mapped supercontigs regions was parsed from the genome-level annotation.

TE content was assessed using de novo TE annotation as described above. TE content was compared between the nonrecombining part of the mating-type chromosome (**Table 2-3**), the PARs, and the autosomes. GC content was compared between the non-recombining region of the mating-type chromosome, the PARs and the autosomes. Significance of the mean difference was assessed using a t-test and a nonparametric Wilcoxon test. In a second step, the GC content at the 3rd codon positions was inspected separately in identified coding regions. Mean GC contents at the 3rd codon positions were compared between the CDS on the non-recombining part of the mating-type chromosome, the PARs and the autosomal coding regions. All GC content analyses were conducted using in-house python and R scripts.

¹⁴ Available for download at <http://www.biomedcentral.com/content/supplementary/s12864-015-1660-8-s11.xlsx>

Prediction of the secretome

To predict a high confidence set of secreted proteins, results from several different software tools were integrated. These include TargetP1.1 (Emanuelsson et al., 2007), SignalP3.0 (Bendtsen et al., 2004), SignalP4.0 (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen et al., 2011), TMHMM2.0 (Krogh et al., 2001), PredGPI (Pierleoni et al., 2008), Phobius (Käll et al., 2007), NucPred (Brameier et al., 2007), Prosite (Sigrist et al., 2010), and WoLF PSORT (Horton et al., 2007). A subset of these tools were used to first exclude proteins as not secreted if they had transmembrane domains (two or more, from TMHMM or Phobius), an ER retention signal (0.00014 from Prosite), GPI anchor (specificity of >99.5% using the general model of PredGPI), or nuclear localization (>0.8 threshold in NucPred). Secreted proteins were then predicted based on passing four of the six thresholds examined (TargetP secreted localization, SignalP3.0 NN Dscore>0.43, SignalP3.0 HMM Sprob>0.8, SignalP4.0 D-score>0.45, WoLFPSort 'Extr' listed as major neighbor, or Phobius secreted localization).

Additional criteria were used for ambiguous predictions. If both TMHMM and Phobius agreed on the existence of 1-2 transmembrane (TM) domains in the protein, the protein was excluded from the probable secretome pool. If a protein was predicted to have a lowly probable GPI linkage (PredGPI specificity >99.0% and <99.5%) and a TM predicted by TMHMM and/or Phobius around the same region, this served as corroborating evidence for GPI anchorage to the membrane.

Where evidence conflicted or was insufficient for determining secretome status, BLASTp and Pfam domains were used to establish probable orthologs, followed by referencing the UniProtKB (Consortium, 2011; Consortium, 2012) and FunSecKB (Lum

& Min, 2011) database for confirmation of localization of the orthologs, where available. Out of 7,364 predicted proteins, 6,899 proteins were excluded from the pool of the secretome based on the criteria described above. Of the remainder, 189 proteins had no contradictory calls in the positive prediction for SP. Another 272 went through further confirmation, of which 182 of these were confirmed to be non-SP and 63 were SP. Of the rest, 27 of them could not be finalized due to lack of ortholog matches in the NCBI database and lack of conserved domain for reference.

Gene clustering and comparative analysis

We compared *M. lychnidis-dioicae* to 18 other fungi (**Table A3-1**) that sample the three subphyla in Basidiomycota, including 5 other Pucciniomycotina, 7 Agaricomycotina, 3 Ustilaginomycotina, as well as 3 Ascomycota outgroups. For *M. lychnidis-dioicae* and the 18 other fungal genomes, we identified ortholog clusters using OrthoMCL (Li et al., 2003) version 1.4 with a Markov inflation index of 1.5 and a maximum e-value of 1×10^{-5} . Two genomes, *R. glutinis* and *P. placenta*, are missing more broadly conserved orthologs than the other genomes; examining the 961 *Microbotryum* gene clusters with an ortholog missing in just one other genome, the number of missing clusters in any one Basidiomycete genome ranged from 1 to 34 with the exception of *R. glutinis* and *P. placenta*, missing 410 and 393 of these highly conserved clusters, respectively. PFAM domains within each gene were identified using Hmmer3 (Eddy, 2011), and gene ontology terms were assigned using BLAST2GO (Conesa et al., 2005).

To examine gene duplication history, the phylome, or complete collection of phylogenetic trees for each gene in a genome, was reconstructed for *Microbotryum lychnidis-dioicae* and 19 other fungi, including those used for OrthoMCL (**Table A3-1**)

and *Serpula lacrymans*. Phylomes were reconstructed using the previously described pipeline (Huerta-Cepas et al., 2011). All trees and alignments have been deposited in PhylomeDB (Huerta-Cepas et al., 2011) and can be browsed on-line (www.phylomedb.org, phylome code 180). Trees were scanned to detect and date duplication events (Huerta-Cepas & Gabaldon, 2011).

RNAi components from other fungi were used as BLAST queries to find homologs in *M. lychnidis-dioicae*; the queries used include *U. hordei* RdRp (CCF48827.1), *C. neoformans* Ago1 (XP_003194007), and *N. crassa* Dcl2 (Q75CC1.3) and Dcl1 (Q758J7.1). The putative function was confirmed by examining protein domains. The identified domains for each protein include: Piwi, PAZ and DUF1785 found in both copies of Argonaute (MVLG_06823, MVLG_06899); DEAD/DEAH helicase, double-stranded RNA binding, and RNaseIII (MVLG_01202). Sugar transporters were identified based on homology to the *Ustilago maydis* Srt1t transporter (Genbank: XP_758521) and the *Uromyces viciae-fabae* Hxt1 (Genbank: CAC41332).

The *M. lychnidis-dioicae* protein models corresponding to carbohydrate-active enzymes were assigned to families of glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), carbohydrate-binding modules (CBM), auxiliary activities (AA) and glycosyltransferases (GT) listed by the CAZy database (Lombard et al., 2014), exactly as previously done for the analyses of dozens of fungal genomes (Duplessis et al., 2011; Floudas et al., 2012; Ohm et al., 2012; Riley et al., 2014).

Conclusion

Altogether, this study provides an in-depth genomic portrait of a fungal castrating, biotrophic plant pathogen reflecting its unique life cycle. In particular, the unique absence of enzyme classes for plant cell wall degradation and maintenance of enzymes that break down components of pollen tubes and flowers provides a striking example of biotrophic host adaptation. In addition, while there are fewer enzymes to digest cellulose, xylan, xyloglucan, and highly substituted forms of pectin, as well as proteins that could protect the fungus from oxidative stress, the repertoire of predicted cell wall modifying enzymes and those that could manipulate host development has expanded. Given the place of *M. lychnidis-dioicae* in a large species-complex with a vast host species pool, the insights from this genomic and transcriptomic analysis combined with comparative approaches with other members of the *Microbotryum* species complex will be most informative on the evolutionary processes involved in a radiation and specialization on a wide array of plant species from different genera.

CHAPTER 3: RNA-SEQ ANALYSIS OF *MICROBOTRYUM LYCHNIDIS-DIOICAE* IN ARTIFICIAL MEDIA

Introduction

The growth of haploid *Microbotryum violaceum* species has been examined in various *in vitro* conditions (Hood & Antonovics, 1998). No difference was observed for growth on water agar between 48 hr from cells exposed to these conditions for 6 days at 15 °C, which would be considered a stress condition. For mated cells, conjugation was observed from linear tetrad germination, and effects of temperature and nutrient availability on the development and growth of the fungus have also been examined (Hood & Antonovics, 1998). In such observations, infectious hyphae would only emerge from one of the mated cells. Penetration of the host plants by *Microbotryum* species seems not to be due to high turgor pressure in the appressoria, but likely is due to lytic enzymes (Mendgen et al., 1996; Schäfer et al., 2010). Once inside the plants, it is unclear how the dikaryon gets nutrients from the host plant because there was no evidence of structural interaction between the fungus and the host from the time of formation of infectious hyphae through its travel in the xylem tracheary elements to shoot meristems (Bauer et al., 1997; Schäfer et al., 2010).

Most studies have focused on the mating type loci and reported degeneration in the form of transposable elements and the extent of the non-recombining regions

(Fontanillas et al., 2015; Hood et al., 2004). Until now, there have been no studies that explore the physiological differences between the two mating types when they are in the haploid state. This is probably because the prevailing understanding of sex chromosomes of *M. lychnidis-dioicae* is that they are completely identical except at the mating loci, with the notion that the mating types function as a non-self recognition mechanism during mating (Badouin et al., 2015). A high level of homozygosity was confirmed recently with the Illumina sequencing of the a₁ mating type (Perlin et al., 2015), supported by RNA-Seq evidence of various fungal stages, followed by the Single Molecule Real Time Sequencing (SMRT) of the a₂ mating type (Badouin et al., 2015).

It had been previously observed (Day, 1976; Xu et al., 2015) that conjugation tube initiation often starts and extends from the a₂ mating type partner more often than from the a₁ mating type strain. It was hypothesized that the presence of 2 homologs of STE12 in the a₂ mating type and the absence of an ortholog in the a₁ mating type might be the cause of this phenomenon (Badouin et al., 2015).

In the current study, we examine differences in global gene expression in cells under different conditions *in vitro*, including haploids grown separately on different media, as well as populations of mated cells, and mated cells exposed to phytol, as a signal to induce stable dikaryon formation (Castle & Day, 1984; Perlin et al., 2015).

Materials and Methods

Preparation of fungal cells

Haploid fungal strains (p1A1 and p1A2) (“Lamole” strains: Genbank 100-15Lamole; (Antonovics et al., 2002; Hood et al., 2004; Perlin et al., 2015)) grown under rich conditions (referred to as Rich) were allowed to grow for 4-5 days on yeast peptone dextrose media (YPD; 1% yeast extract, 10% dextrose, 2% peptone, 2% agar) at room temperature before RNA extraction. Two independently extracted RNA samples of each haploid strain were submitted to the Broad Institute and analyzed as biological replicates.

For the nutrient-limited (referred to as NL) condition, the same isolates were resuspended in sterile distilled water after growing for 4-5 days on YPD at room temperature, and adjusted to 10^9 cells/ml by A_{600} (absorbance at 600 nm) determination. Then, 800 μ l of the suspension was dispensed to cover the entire agar surface of a “Water” agar plate (composition: 2% Bacto agar) in an even layer and allowed to dry. The plates were incubated at 14 °C for about 48 hours before RNA was extracted. Two independently extracted RNA samples of each haploid strain were submitted and analyzed as biological replicates. The same was done for mated samples (referred to as Mated) except the two strains were combined in equal proportions to make the 10^9 cells/ml suspension. Two Mated samples were submitted and analyzed as biological replicates.

Samples in which mating occurred in the presence of phytol (referred to as Pmated) were prepared in the same way as Mated except the solution in which the suspension was prepared contained 5% ethanol and 1% phytol (Sigma-Aldrich, cat no: P3647). Ethanol and phytol were combined first in order to resuspend the phytol, and

then the mixture was added to sterile distilled water containing a total of 10^9 haploid cells/ml final concentrations. RNA was extracted from Pmated samples at 12 h, 24 h and 48 h based on observations made in previous pilot experiments. Four Pmated12, three Pmated24 and two Pmated48 samples were submitted to the Broad Institute and analyzed as biological replicates.

RNA extraction and sequencing

RNA was extracted and processed as previously published ((Perlin et al., 2015), Chapter 2, see Appendix 1 for details). A minimum of 5 µg of DNA-free RNA was prepared for each desired stage and sent to the Broad Institute for RNA-Sequencing via Illumina. RNA-Seq data from different conditions were processed using Trinity pipeline (Grabherr et al., 2011; Haas et al., 2013), and RNA-Seq reads from each sample were aligned using bowtie (Langmead et al., 2009) against protein coding sequences extracted based on annotation. Fungal gene expression levels were estimated by RSEM (Li & Dewey, 2011), and edgeR with TMM normalization (Kadota et al., 2012; Robinson et al., 2010) was used to identify differentially expressed genes between each pair of conditions with a corrected p-value cutoff of $1e-5$. RNA-Seq reads from all infected host samples were assembled using Trinity. Assembled transcripts were filtered based on the alignment with the *M. lychnidis-dioicae* fungal genome (Fontanillas et al., 2015; Perlin et al., 2015).

To manage the large set of data, the samples were divided into 2 main groups with Group 1 consisting of Mated, Rich and NL samples, and p1A1 Rich was used as the reference for detection of gene expression changes. Group 2 consisted of the Mated, Pmated12, Pmated24 and Pmated48, where Mated was used as the reference for detection of gene expression changes.

Venn Diagrams

Venn diagrams were generated using the online tool, Venny (Oliveros, 2007-2015). The edgeR data was sorted to extract the list of the predicted genes (Perlin et al., 2015) that were significantly up-regulated or down-regulated (false discovery rate, FDR < 1E-5) against a reference condition in the pairwise analysis.

Gene Set Enrichment Analysis (GSEA) Preranked

Gene set enrichment analysis (Mootha et al., 2003; Subramanian et al., 2005) was performed to discover the possible enrichment of PFAM domains, gene ontology terms (Conesa et al., 2005) and KEGG pathway maps (Kanehisa et al., 2015) domains in the conditions studied using the GenePattern software (Reich et al., 2006). In addition, the predicted small secreted proteins (SSPs, less than 250 amino acids long), secreted proteins (SPs, more than 250 amino acids long) and unique proteins based on orthomcl analysis (Li et al., 2003; Perlin et al., 2015) were included as gene sets. The resulting gene set list (.gmt) consisted of 3651 gene sets, with a range of 1-1116 members. Ranked lists (.rnk) of every pair-wise comparison made were processed by the GSEAPreranked module software. When running the software, the algorithm was set to allow minimum gene set of one and maximum of 1200. Gene sets that were significantly enriched at FDR<25% or nominal p-value<1% were listed and examined for further hypothesis-generation.

Results and Discussion

Differential gene expression in NL and Mated treatments compared to Rich

Part of the *in vitro* RNA-Seq analysis has already been published (Fontanillas et al., 2015). By re-sequencing the mating type chromosome, the genes located in the non-recombining region (NRR) and the mating type chromosome recombining region or pseudoautosomal region (PAR) were identified. Supercontigs 37 and 43 assembled in the original reference genome based on the a₁ mating type ((Perlin et al., 2015), see Chapter 2) were identified to be the PAR, harboring 99 predicted genes. The NRR consists of 349 identified predicted genes, found on various supercontigs. The NRR was filled with transposable element and repetitive sequences (Fontanillas et al., 2015; Perlin et al., 2015; Petit et al., 2012), hence was not assembled as larger contiguous sequences. It was determined that there were unique genes in the NRR that were highly similar between the a₁ and a₂ mating types. Moreover, genes in the PAR had less than 0.1% nucleotide divergence.

A heatmap of the differential expression data, comparing Rich, NL and Mated treatments, can be found in **Supplementary Figure S4A**¹⁵ ((Fontanillas et al., 2015); NL is referred to as Water in the figure). Based on the analysis, 177 genes were differentially expressed between the a₁ and a₂ mating type, and 93 of them were differentially expressed in both NL and Rich. 84 of them were located in the NRR and two in the PAR.

Fig. 3-1 shows the Venn diagrams of the number of predicted genes in p1A1 NL, p1A2 NL, Mated and p1A2 Rich that were significantly up- and down- regulated

¹⁵ Available for download at http://mbe.oxfordjournals.org/content/suppl/2014/12/21/msu396.DC1/FigS4_expression.pdf

compared to the p1A1 Rich treatment. A total of 1233 (16.7% of all predicted genes) and 1158 (15.7% of all predicted genes) were up- and down- regulated, respectively, when compared to p1A1 Rich. Among these, a large proportion of them, 462 (37.5%) and 546 (47.1%) were commonly up- and down-regulated among both NL and Mated conditions.

p1A1 NL and p1A2 NL conditions did share more common differential expression (148 up- and 96 down-regulated) than individual differential expression (79 up- and 27 down-regulated for p1A1 NL; 118 up- and 93 down-regulated for p1A2 NL), but interestingly, Mated appeared to share a much larger subset of common differential expression with p1A2 NL (96 up- and 89 down-regulated) than with p1A1 NL (29 up- and 15 down-regulated). This may suggest that the a_2 mating type “contributes” more in the initiation of the mating relationship, since the NL treatment is conducive for mating but the mating partner is absent. This is in line with the previous reports that the a_2 mating type produces the conjugation tube earlier and to a greater length (Day, 1976; Xu et al., 2015). This is further supported by the presence of two homologs of STE12 on the a_2 mating chromosome (Badouin et al., 2015).

As expected, p1A2 Rich had the least number of genes that were differentially regulated when compared to p1A1 Rich. We noted that among the 79 genes that were down-regulated in both p1A2 Rich and p1A2 NL, 65 of them belonged to the non-recombining region (NRR) of the mating type chromosome. Among the 19 that were also down-regulated in Mated, 16 of them also belonged to the NRR. On the other hand, there were 184 up- and 109 down-regulated genes in Mated condition only. Based on previous analysis (Fontanillas et al., 2015), there wasn’t more differential expression for genes located on the mating type chromosome in the Mated condition, nor were the

differentially regulated genes between a_1 and a_2 mating type up-regulated in Mated condition. The hypothesis that these genes have mating type specific roles was addressed, and it was concluded that most of the genes that have weak or no expression in one mating type is more probable to be due to deleterious mutations (Fontanillas et al., 2015). There was also further evidence that the mating type chromosome is extensively rearranged and degenerated (Petit et al., 2012).

However, if we consider the conditions under which the fungal cells were treated, it may be more logical to expect genes involved in initiation of mating to be detectable in NL rather than Mated. This is because the fungal cells were in a conducive environment (*i.e.* stressful both in terms of temperature and nutrient availability) for mating but lack the mating partner; whereas in Mated condition, mating has already taken place, and the conjugation tube has already been formed; thus transcripts related to initiation of mating may already have been lost. We investigated this possibility with further gene set enrichment analysis.

Gene set enrichment in NL and Mated up-regulated genes

From the gene set analysis, the RNA recognition motif (PF00076) was the only domain enriched in Mated when compared to p1A2 NL, and there were no other enrichments in the comparison between Mated and NL. It is difficult to access if this is an effect of nutrient limitation, or if the differential gene expression was not detectable between those two conditions and hence no gene enrichment were detected.

When both NL and Mated were compared to Rich (**Tables 3-1 – 3-3**), the common gene set enrichments among the up-regulated genes were oligopeptide transporters, secretory lipases, peptidase M43, alpha/beta hydrolase fold, secreted proteins and proteins “unique” to *M. lychnidis-dioicae*. The flavin-binding monooxygenase-like domain was enriched only in genes up-regulated in p1A1 NL when compared to p1A1 Rich. There was enrichment of cation transporting ATPase, major facilitator superfamily, DUF2424 (PF10340), classical zinc finger domain, carboxylesterase, and major intrinsic protein among the genes up-regulated in p1A2 NL when compared to Rich. Mated showed exclusive gene set enrichment of GMC oxidoreductase (PF05199), finger domains of DNA polymerase lambda (PF10391) and Rho GEF domain (PF00621).

The oligopeptide transporter (PF03169) was designated *OPT1* from *Candida albicans* (Lubkowitz et al., 1997) and was later found to be an extremely versatile peptide transporter with an extended substrate spectrum (Dunkel et al., 2013). Two of these transporters in *M. lychnidis-dioicae* (MVLG_00149 and 03161) remained at a very low, non-differentially expressed state for all the *in vitro* stages examined here but were found

to be expressed when interacting with the host plant (see Chapter 5). Three of them (MVLG_03162, 03488 and 04254) were expressed, but there was no significant differential expression; hence we believe they are housekeeping transporters. Five of them (MVLG_02729, 04545, 04546, 05201, 07217 (very low-level expression, but significant differential expression)) were up-regulated in Mated and NL compared to Rich, and they were not different in Pmated when compared to Mated (see below). These expression levels may be related to nutrient availability. Three of them (MVLG_00150, 03106 (very low expression levels, but significant differential expression) and 05200) were up-regulated in NL and Mated compared to Rich but reduced in expression in Pmated; three others (MVLG_02728, 04056 and 04057) were further up-regulated in Pmated (see below). Interestingly, five of these transporters (MVLG_04056, 04057, 04545, 05201 and 07217) were significantly up-regulated in p1A2 NL even when compared to Mated (although not significantly higher than p1A1 NL). Some if not all of these five transporters may be related to pheromone sensing, since the a_2 seems take more “initiative” in mating, and they may continue to be transcribed, whereas in Mated, their activity would no longer be required, with the completion of the mating process.

The identified homeodomain DNA binding proteins, MVLG_07149 and 07150, which determine post-mating compatibility, were almost completely switched off in Rich and also in p1A2 NL. p1A1 NL, on the other hand, showed 8.50 and 4.74 log₂ fold increases, respectively, compared to p1A2 NL, while Mated showed 2.21 and 5.91 log₂ fold increases compared to p1A1 NL. So there is transcription of the homeodomain in p1A1 NL even in the absence of a mating partner, but it appears the gene expression level increases post mating, which is consistent with its function.

Considering the possibility of greater contribution to initiation of mating by the a_2 mating type, and a_1 mating type playing a greater role in post-mating expression, this may indicate a division of labor between the two mating types in sexual reproduction, which is not a novel idea (Hood et al., 2004). Preference for the a_2 mating type progeny to inherit the a_2 parent mitochondrion DNA has also been observed and thought to be part of the mating type function (Wilch et al., 1992).

Table 3-1. List of enriched gene sets that were up-regulated in p1A1 NL when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of nominal enrichment score (NES).

Enriched in p1A1 NL vs. p1A1 Rich	Common enrichment	Enriched in p1A1 NL vs. p1A2 Rich
<ul style="list-style-type: none"> PF07859.6 ABHYDROLASE_3 [ALPHA/BETA HYDROLASE FOLD] PF00743.12 FMO-LIKE[FLAVIN-BINDING MONOOXYGENASE-LIKE] 	<ul style="list-style-type: none"> PF03169.8 OPT[OPT OLIGOPEPTIDE TRANSPORTER PROTEIN] TIGR00728[OPT_SFAM: OLIGOPEPTIDE TRANSPORTERS, OPT SUPERFAMILY] PF03583.7 LIP[SECRETORY LIPASE] UNIQUE TO <i>M. lychnidis-dioicae</i> 	<ul style="list-style-type: none"> PF05572.6 PEPTIDASE _M43[PREGNANCY-ASSOCIATED PLASMA PROTEIN-A]

Table 3-2. List of enriched gene sets that were up-regulated in p1A2 NL when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of NES.

Enriched in "p1A2 NL vs. p1A1 Rich	Common enrichment	Enriched in p1A2 NL vs. p1A2 Rich
<ul style="list-style-type: none"> PF00689.14 CATION_ATPASE_C[CATION TRANSPORTING ATPASE, C-TERMINUS] 	<ul style="list-style-type: none"> PF03169.8 OPT[OPT OLIGOPEPTIDE TRANSPORTER PROTEIN] PF07859.6 ABHYDROLASE_3[ALPHA/BETA HYDROLASE FOLD] TIGR00728[OPT_SFAM: OLIGOPEPTIDE TRANSPORTERS, OPT SUPERFAMILY] PF03583.7 LIP[SECRETORY LIPASE] SECRETED PROTEIN PF07690.9 MFS_1[MAJOR FACILITATOR SUPERFAMILY] PF05572.6 PEPTIDASE_M43[PREGNANCY-ASSOCIATED PLASMA PROTEIN-A] 	<ul style="list-style-type: none"> UNIQUE TO <i>M. lychnidis-dioicae</i> PF10340.2 DUF2424[PROTEIN OF UNKNOWN FUNCTION (DUF2424)] PF00096.19 ZF-C2H2[ZINC FINGER, C2H2 TYPE] PF00135.21 COESTERASE[CARBOXYLESTERASE] PF00230.13 MIP[MAJOR INTRINSIC PROTEIN]

Table 3-3. List of enriched gene sets that were up-regulated in Mated when compared to p1A1 Rich and p1A2 Rich, ranked by decreasing magnitude of NES.

Enriched in Mated vs. p1A1 Rich	Common enrichment	Enriched in Mated vs. p1A2 Rich
<ul style="list-style-type: none"> • PF07859.6 ABHYDROLASE_3 [ALPHA/BETA HYDROLASE FOLD] • PF05572.6 PEPTIDASE_M43[PREGNANCY-ASSOCIATED PLASMA PROTEIN-A] • TIGR00728[OPT_SFAM: OLIGOPEPTIDE TRANSPORTERS, OPT SUPERFAMILY] • PF05199.6 GMC_OXRED_C[G MC OXIDOREDUCTASE] • SECRETED PROTEIN • PF10391.2 DNA_POL_LAMB D_F[FINGERS DOMAIN OF DNA POLYMERASE LAMBDA] 	<ul style="list-style-type: none"> • PF03169.8 OPT[OPT OLIGOPEPTIDE TRANSPORTER PROTEIN] • PF03583.7 LIP[SECRETORY LIPASE] • UNIQUE TO <i>M. lychnidis-dioicae</i> 	<ul style="list-style-type: none"> • PF00621.13 RHOGF[R HOGF DOMAIN]

Gene set enrichment in NL and Mated down-regulated genes

Table 3-4 lists significantly enriched gene sets that were down-regulated in Mated when compared to p1A1 and p1A2 Rich, and **Tables 3-5** and **3-6** further list the enriched gene sets that were down-regulated in NL when compared to Rich.

In the Rich condition, we expect the fungal cells to be growing and dividing mitotically, whereas on the water agar where mating can take place, lack of nutrient and the low temperature probably arrest these processes. The gene set enrichment is consistent with this prediction in which related cellular components, biological processes, molecular functions and metabolism were among those gene sets that were down-regulated in NL and Mated. The KEGG pathway for Huntington's disease (KO05016), Parkinson's disease (KO05012) and Alzheimer's disease (KO05010) appeared to be enriched, because the annotations were found on the same genes as many of these core metabolic processes; hence they should not be viewed as different gene set enrichment.

The three multicopper oxidase domains were annotated on the same four genes and were significantly enriched among the genes down-regulated in the NL and Mated condition. A review of copper oxidases (Kües & Rühl, 2011) revealed that Basidiomycetes tend to have multiple copies of multicopper oxidases, which have been implicated in many different types of processes and regulation. KEGG BLAST of these proteins revealed that MVLG_03092 is likely a laccase (EC: 1.10.3.2), predicted to be secreted but was unique to *M. lychnidis-dioicae* and was up-regulated in NL. MVLG_00670 and MVLG_01868 were likely to be ferroxidases and the specific activities may vary from organism to organism; hence it is hard to predict their real functions. Together with MVLG_1284, they were downregulated in NL and Mated conditions.

Table 3-4. List of significantly enriched gene sets that were down-regulated in Mated when compared to p1A1 and p1A2 Rich, ranked by decreasing magnitude of normalized enrichment score (NES).

Enriched in p1A1 Rich vs. Mated	Common enrichment	Enriched in p1A2 Rich vs. Mated
<ul style="list-style-type: none"> GO:0019843[RRNA BINDING>MF] GO:0007346[REGULATION OF MITOTIC CELL CYCLE>BP] 	<ul style="list-style-type: none"> GO:0003735[STRUCTURAL CONSTITUENT OF RIBOSOME>MF] GO:0006412[TRANSLATION>BP] GO:0005840[RIBOSOME>CC] GO:0005634[NUCLEUS>CC] PF00125.17[HISTONE[CORE HISTONE H2A/H2B/H3/H4] GO:0003677[DNA BINDING>MF] GO:0006334[NUCLEOSOME ASSEMBLY>BP] GO:0000786[NUCLEOSOME>CC] GO:0006281[DNA REPAIR>BP] GO:0006096[GLYCOLYTIC PROCESS>BP] KO05016[HUNTINGTON'S DISEASE] KO05012[PARKINSON'S DISEASE] KO05010[ALZHEIMER'S DISEASE] SMALL SECRETED PROTEIN 	<ul style="list-style-type: none"> GO:0003723[RNA BINDING>MF] GO:0006414[TRANSLATIONAL ELONGATION>BP] GO:0006457[PROTEIN FOLDING>BP] KO00240[PYRIMIDINE METABOLISM] PF06723.6[MREB_MBL[MREB/M BL PROTEIN] KO04110[CELL CYCLE] KO04111[CELL CYCLE - YEAST] PF00493.16[MCM[MCM2/3/5 FAMILY] GO:0051287[NAD BINDING>MF] KO04113[MEIOSIS - YEAST] GO:0005730[NUCLEOLUS>CC] GO:0006270[DNA REPLICATION INITIATION>BP] GO:0055114[OXIDATION-REDUCTION PROCESS>BP] GO:0006260[DNA REPLICATION>BP]

	<ul style="list-style-type: none"> • PF03330.11 DPBB_1[RARE LIPOPROTEIN A (RLPA)-LIKE DOUBLE-PSI BETA-BARREL] • GO:0005737[CYTOPLASM>CC] • GO:0005829[CYTOSOL>CC] • GO:0005524 [ATP BINDING>MF] • GO:0015935[SMALL RIBOSOMAL SUBUNIT>CC] • GO:0006099[TRICARBOXYLIC ACID CYCLE>BP] • PF01522.14 POLYSACC_DEAC_1[POLYSACCHARIDE DEACETYLASE] • PF00394.15 CU- OXIDASE[MULTICOPPER OXIDASE] • PF07731.7 CU- OXIDASE_2[MULTICOPPER OXIDASE] 	<ul style="list-style-type: none"> • PF06280.5 DUF1034[FN3-LIKE DOMAIN (DUF1034)] • GO:0005488[BINDING>MF] • KO03420[NUCLEOTIDE EXCISION REPAIR] • GO:0051301[CELL DIVISION>BP] • PF01078.14 MG_CHELATASE[M AGNESIUM CHELATASE, SUBUNIT CHLI] • PF07732.8 CU- OXIDASE_3[MULTICOPPER OXIDASE] • KO03030[DNA REPLICATION] • PF01776.10 RIBOSOMAL_L22E[R IBOSOMAL L22E PROTEIN FAMILY] • GO:0022627[CYTOSOLIC SMALL RIBOSOMAL SUBUNIT>CC]
--	---	--

Abbreviation:- CC = cellular component; BP = biological process; MF = molecular function

Table 3-5. List of enriched gene sets that were down-regulated in NL when compared to p1A1 Rich, ranked by decreasing magnitude of normalized enrichment score (NES).

Enriched in p1A1 Rich vs. p1A1 NL	Common enrichment	Enriched in p1A1 Rich vs. p1A2 NL
<ul style="list-style-type: none"> • GO:0003677[DNA BINDING>MF] • GO:0015935[SMALL RIBOSMAL SUBUNTI>CC] • KO00520[AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM] • KO00190[OXIDATIVE PHOSPHORYLATION] • PF09468.3 RNASE_H2-YDR279[YDR279P PROTEIN FAMILY (RNASE H2 COMPLEX COMPONENT)] • PF00394.15 CU- OXIDASE[MULTICOPPER OXIDASE] • PF07732.8 CU- OXIDASE_3[MULTICOPPER OXIDASE] • PF07731.7 CU- OXIDASE_2[MULTICOPPER OXIDASE] 	<ul style="list-style-type: none"> • GO:0003735[STRUCTURAL CONSTITUENT OF RIBOSOME>MF] • GO:0006412[TRANSLATION>BP] • GO:0005840[RIBOSOME>CC] • GO:0005634[NUCLEUS>CC] • GO:0006096[TRICARBOXYLIC ACID CYCLE>BP] • PF00125.17 HISTONE[CORE HISTONE H2A/H2B/H3/H4] • GO:0006334[NUCLEOSME ASSEMBLY>BP] • GO:0006281[DNA REPAIR>BP] • KO05012[PARKINSON'S DISEASE] • GO:0000786[NUCLEOSOME>CC] • PF03330.11 DPBB_1[RARE LIPOPROTEIN A (RLPA)-LIKE DOUBLE-PSI BETA-BARREL] • KO05016[HUNTINGTON'S DISEASE] • GO:0006099[TRICARBOXYLIC ACID CYCLE>BP] • SMALL SECRETED PROTEIN • KO05010[ALZHEIMER'S DISEASE] 	<ul style="list-style-type: none"> • None

Abbreviation:- CC = cellular component; BP = biological process; MF = molecular function

Table 3-6. List of enriched gene sets that were down-regulated in NL when compared to p1A2 Rich, ranked by decreasing magnitude of NES.

Enriched in p1A2 Rich vs. p1A1 NL	Common enrichment	Enriched in p1A2 Rich vs. p1A2 NL
<ul style="list-style-type: none"> • SMALL SECRETED PROTEIN • KO03420[NUCLEOTIDE EXCISION REPAIR] • KO00710[CARBON FIXATION IN PHOTOSYNTHETIC ORGANISMS] • PF06280.5 DUF1034[FN3-LIKE DOMAIN (DUF1034)] • PF00808.16 CBFD_NFYB_H MF[HISTONE-LIKE TRANSCRIPTION FACTOR (CBF/NF-Y) AND ARCHAEOAL HISTONE] • GO:0050662[COENZYME BINDING>MF] • KO00190[OXIDATIVE PHOSPHORYLATION] • GO:0019673[GDP-MANNOSE METABOLIC PROCESS>BP] 	<ul style="list-style-type: none"> • GO:0003735[STRUCTURAL CONSTITUENT OF RIBOSOME>MF] • GO:0006412[TRANSLATION>BP] • GO:0005840[RIBOSOME>CC] • GO:0000786[NUCLEOSOME>CC] • GO:0005634[NUCLEUS>CC] • GO:0003677[DNA BINDING>MF] • PF00125.17 HISTONE[CORE HISTONE H2A/H2B/H3/H4] • GO:0006334[NUCLEOSOME ASSEMBLY>BP] • GO:0006281[DNA REPAIR>BP] • KO05016[HUNTINGTON'S DISEASE] • KO05012[PARKINSON'S DISEASE] • GO:0006096[TRICARBOLIC ACID CYCLE>BP] • KO05010[ALZHEIMER'S DISEASE] • GO:0005524[ATP BINDING>MF] • PF03330.11 DPBB_1[RLPARE LIPOPROTEIN A (RLPA)-LIKE DOUBLE-PSI BETA-BARREL] • KO04113[MEIOSIS - YEAST] • KO04111[CELL CYCLE - YEAST] • GO:0005737[CYTOPLASM>CC] • KO00520[AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM] • KO04110[CELL CYCLE] 	<ul style="list-style-type: none"> • GO:0005515[PROTEIN BINDING>MF] • GO:0003723[RNA BINDING>MF] • PF07728.7 AAA_5[AAA DOMAIN (DYNEIN-RELATED SUBFAMILY)] • KO00240[PYRIMIDINE METABOLISM] • GO:0005739[MITOCHONDRION>CC] • PF01078.14 MG_CHELATASE [MAGNESIUM CHELATASE, SUBUNIT CHL1] • PF00012.13 HSP70[HSP70 PROTEIN] • GO:0005622[INTRACELLULAR>CC] • GO:0004386[HELICASE ACTIVITY>MF] • KO03410[BASE EXCISION REPAIR] • PF08449.4 UAA[UAA TRANSPORTER FAMILY] • EC:1.2.4.1[PYRUVATE DEHYDROGENASE (ACETYL-TRANSFERRING)] • GO:0004739[PYRUVATE DEHYDROGENASE (ACETYL-TRANSFERRING ACTIVITY>MF)] • PF00046.22 HOMEODOMAIN[HOMEODOMAIN] • PF02629.12 COA_BINDING[COA BINDING DOMAIN] • GO:0003878[ATP CITRATE SYNTHASE ACTIVITY>MF] • TIGR01472[GMD: GDP-MANNOSE 4,6-DEHYDRATASE]

	<ul style="list-style-type: none"> • PF00493.16[MCM[MCM 2/3/5 FAMILY] • GO:0015935[SMALL RIBOSOMAL SUBUNIT>CC] • GO:0006260[DNA REPLICATION>BP] • GO:0006270[DNA REPLICATION INITIATION>BP] • GO:0006414[TRANSLATIONAL ELONGATION>BP] • KO03430[MISMATCH REPAIR] • PF06723.6[MREB_MBL [MREB/MBL PROTEIN] • GO:0005488[BINDING>MF] • GO:0005829[CYTOSOL >CC] • PF09468.3[RNASE_H2-YDR279[YDR279P PROTEIN FAMILY (RNASE H2 COMPLEX COMPONENT)] 	
--	--	--

Abbreviation:- CC = cellular component; BP = biological process; MF = molecular function

Phytol-mated samples

M. lychnidis-dioicae has been known to react to certain chemicals like phytol and α -tocopherol (Castle & Day, 1984) to produce hyphae on artificial medium. This is viewed as simulation of dikaryon formation required for entry into plant host. **Fig 3-2D** shows a filamentous appearance following formation of conjugation tube; compare to **Fig 3-2C** where no phytol was present.

Fig. 3-3 shows the Venn diagram of the number of genes that were differentially regulated in mated cells in the presence of phytol over 12 (Pmated12), 24 (Pmated24) and 48 (Pmated48) hours compared to mated cells in the absence of phytol at 48 hours (Mated). A total of 964 (13.1%) predicted genes were up-regulated compared to Mated and 445 (6.0%) were down-regulated compared to Mated. Of these, a smaller proportion

was up- (22.1%) and down- (2.7%) regulated throughout the entire course of treatment compared to Mated. The majority of the gene expression changes happened within the first 24 hours of the treatment.

Figure 3-2. Morphology of haploid and mated cells in the absence or presence of phytol.

A & B, phytol-treated haploid cell control (p1A1 and p1A2 respectively); C, water-treated mated cell control (Inset: higher magnification showing cells with conjugation bridge, black arrows); D, mated cells treated with phytol showing filamentation (white arrows). Size bars, 50 µm.

Table 3-7. Summary of trends observed in genes of *M. lychnidis-dioicae* over time when treated with phytol.

Trend ^a	Description	Number of genes	Percentage ^b
12h<24h<48h	Increasing from 12 hr to 48 hr	147	2.0%
12h>24h>48h	Decreasing from 12 hr to 48 hr	123	1.7%
12h<24h=48h	Increasing from 12-24 hr and plateaued	215	2.9%
12h>24h=48h	Decreasing from 12-24 hr and plateaued	127	1.7%
12h<24h>48h	Highest at 24 hr	223	3.0%
12h>24h<48h	Lowest at 24 hr	158	2.1%
12h=24h<48h	Constant from 12-24 hr then increase	24	0.3%
12h=24h>48h	Constant from 12-24 hr then decrease	150	2.0%

^aThe numbers represent the hours at which the phytol-treated sample were extracted. < represents an increase; > represents a decrease; = represents no detectable changes.

^bPercentages were calculated based on total gene set of 7364 genes.

Table 3-7 shows a summary of the number of genes that presented a detectable trend over the time course. This was determined by examining all the pairwise comparisons and taking all significant gene expression changes into account. However, it should be noted that these trends may or may not be exclusive to mating cells in the presence of phytol since the data for corresponding mating cell in the absence of phytol were not available.

Gene enrichments in Pmated during the time-course

Table 3-8 lists the gene sets that were enriched in the genes up-regulated in Pmated when compared to Mated and **Table 3-9** lists the gene set enrichment among the genes up-regulated at different time points. No gene set enrichment was detected in the comparison between Pmated24 and Pmated48.

Table 3-8. List of enriched gene sets that were up-regulated in Pmated12, Pmated24 and Pmated48 when compared to Mated, ranked by decreasing magnitude of NES.

Enriched in all Pmated vs. Mated	
<ul style="list-style-type: none"> SECRETED PROTEIN UNIQUE TO <i>M. lychnidis-dioicae</i> 	
Common enrichment in Pmated12 and Pmated24 vs. Mated	Common enrichment in Pmated24 and Pmated48 vs. Mated
<ul style="list-style-type: none"> SMALL SECRETED PROTEIN 	<ul style="list-style-type: none"> PF03583.7 LIP[SECRETORY LIPASE]
Enriched in Pmated12 only vs. Mated	Enriched in Pmated48 only vs. Mated
<ul style="list-style-type: none"> PF03330.11 DPBB_1[RARE LIPOPROTEIN A (RLPA)-LIKE DOUBLE-PSI BETA-BARREL] PF00026.16 ASP[EUKARYOTIC ASPARTYL PROTEASE] PF06280.5 DUF1034[FN3-LIKE DOMAIN (DUF1034)] PF00295.10 GLYCO_HYDRO_28[GLYCOSYL HYDROLASES FAMILY 28] 	<ul style="list-style-type: none"> PF00704.21 GLYCO_HYDRO_18[GLYCOSYL HYDROLASES FAMILY 18]

Table 3-9. List of enriched gene sets that were up-regulated in Pmated12, Pmated24 and Pmated48 when compared to each other, ranked by decreasing magnitude of NES.

Enriched in Pmated12 vs. Pmated24	Enriched in Pmated24 vs. Pmated12
<ul style="list-style-type: none"> • PF07728.7 AAA_5[AAA DOMAIN (DYNEIN-RELATED SUBFAMILY)] • PF00011.14 HSP20[HSP20/ALPHA CRYSTALLIN FAMILY] 	<ul style="list-style-type: none"> • SECRETED PROTEIN • SMALL SECRETED PROTEIN • PF03583.7 LIP[SECRETORY LIPASE]
Enriched in Pmated12 vs. Pmated48	Enriched in Pmated48 vs. Pmated12
<ul style="list-style-type: none"> • PF00125.17 HISTONE[CORE HISTONE H2A/H2B/H3/H4] • GO:0006281[DNA REPAIR>BP] • GO:0000786[NUCLEOSOME>CC] • GO:0006334[NUCLEOSOME ASSEMBLY>BP] • PF02225.15 PA[PA DOMAIN] 	<ul style="list-style-type: none"> • PF03583.7 LIP[SECRETORY LIPASE] • SECRETED PROTEIN

Abbreviation:- CC = cellular component; BP = biological process; MF = molecular function

Major facilitator superfamily

A cluster of genes (MVLG_06790-MVLG_06794) that consist of only a major facilitator domain (MFS, PF07690), except for a sugar transporter domain (PF00083) in MVLG_06794, was identified. They were unique but belong together as one ortholog family, suggesting similar functions, yet they had varying expression patterns. Three of these transporters were increasingly expressed from 12 hr to 48 hrs phytol treatment, and Mated expressed them at the lower, if not the same level as at 12 hr. These may be involved in the filamentation process. MVLG_06793 had the lowest expression at 24 hr, similar to Mated but overall had very low expression level (less than 6 transcripts per million (TPM)) and MVLG_06794 decreased from 12 hr to 24 hr and then plateaued with Mated being slightly lower than that at 48 hr. MVLG_06794 was significantly up-regulated in Rich.

Secretory lipase (PF03583)

In previous work (Perlin et al., 2015), the secretory lipase gene family expression presented some interesting patterns. With the new data, we have evidence of enrichment

of up-regulated secretory lipases both in Mated and NL when compared to Rich and also in Pmated24 and Pmated48 when compared to Mated.

The three cytoplasmic proteins (MVLG_07229, 07284 and 07291) were found to be highly induced in the NL condition as found previously (Perlin et al., 2015), and in addition, MVLG_00914 and MVLG_05549, which were predicted to be secreted, were also up-regulated in NL when compared to Rich, with a marginal log₂ fold change of 3 and 2, respectively. Upon closer examination, we found that the transcript abundance of MVLG_00914 remained below 1 TPM and that of MVLG_05549 remained below 10 TPM in all the *in vitro* conditions, which is very low. Therefore, only the three cytoplasmic proteins were significantly up-regulated in Mated when compared to Rich, and the expression levels of MVLG_00914 and 05549 in Mated were not very different from in NL. It should be noted that MVLG_07229 and 07284 have 92% protein similarity; hence they probably perform the same functions.

On the other hand, MVLG_00914 showed 9 and 11 log₂ fold increase in Pmated24 and Pmated48, respectively, when compared to Mated. MVLG_05549 showed about a 3 log₂ fold increase throughout the 48 hr period when compared to Mated, and the three cytoplasmic proteins were not differently expressed than in Mated. Thus, we believe that expression of the three cytoplasmic proteins may be related to nutrient availability.

In addition, two other secretory lipases, MVLG_00933 and 00934, were also up-regulated in Pmated. Previous qRT-PCR experiments (D. Razeeq, unpublished) were not able to establish the expression level of these two secretory lipases, and we had assumed they were pseudogenes. From the new data, we found that significant up-regulation was

only detectable at 24 hours in RNA-Seq, and we do not know if this is detectable with qRT-PCR, since the transcript abundance was only about 20 TPM, compared to more than 100 with the other secretory lipases that were detected in the qRT-PCR experiment. Interestingly, these two secretory lipases showed more than 9 and 11 log₂ fold increases in Pmated24, and further increases to 12 and 15 log₂ fold increases, respectively, in Pmated48. Hence, we propose that MVLG_00914, 00933, 00934 and 05549 are directly involved in the response to phytol and filamentation of the fungus. It is also interesting to note that MVLG_00914, 00933 and 00934 are located on the same supercontig, and only MVLG_00934 was not predicted to be secreted.

MVLG_04698, which was predicted to encode a fatty acid synthase, was also up-regulated in NL and Mated when compared to Rich, but the expression was the same for this gene in Pmated compared to Mated. At the same time, the phospholipase C protein (MVLG_07108) which is implicated in lipid acquisition, was also down-regulated in Rich, and Pmated was overall lower than Mated. For this gene, there was a weak trend of increased expression between Pmated24 and Pmated48.

SPs, SSPs and genes unique to *M. lychnidis-dioicae*

By setting up genes encoding secreted (SP), small and secreted (SSP) and products unique to *M. lychnidis-dioicae* as gene sets, we were able to discover their enrichment, which would not have otherwise been possible due to the lack of known domains. To understand whether the genes were up-regulated in Mated and NL and then further up-regulated in Pmated, we extracted the list of secreted proteins that contributed to core enrichment of each of the three gene sets in each treatment and made a comparison to see how much they overlapped.

Fig. 3-4A shows that out of the 65 secreted proteins that were up-regulated in Pmated compared to Mated, 21 (32.3%) of them were the same proteins. It is interesting that Pmated12 and Pmated48 did not share any up-regulated secreted proteins other than those that were up-regulated through Pmated24, suggesting that there were specific roles for the secreted protein at different time points. **Fig. 3-4B** shows that a total of 103 secreted proteins were involved in the mating and filamentation process. NL and Mated shared 30 (29.1%) of them and only 17 (16.5%) overlapped with what was also up-regulated in Pmated. Out of the 65 proteins that were up-regulated in Pmated, 48 (73.8%) were used exclusively in Pmated.

Fig. 3-5A shows that out of the 26 SSPs that were up-regulated in Pmated compared to Mated, 10 (38.5%) of them were the same proteins. The SSPs used in early treatment appeared to be different from those used in the last stage of Pmated treatment.

Fig. 3-5B shows that a total of 29 SSPs were involved in the mating and filamentation process. NL and Mated shared 3 (10.3%) of these, and 9 (31.0%) overlapped with what was also up-regulated in Pmated. Out of the 26 SSPs that were up-regulated in Pmated, 17 (65.4%) were used exclusively found as up-regulated in Pmated.

Fig. 3-6A shows that out of the 141 unique proteins that were up-regulated in Pmated compared to Mated, 42 (29.8%) of them were the same proteins. Only two of them were common between Pmated12 and Pmated48, and there appear to be substantial discrete and overlapping gene expression throughout the course of the treatment. **Fig. 3-6B** shows that a total of 393 unique proteins were involved in the mating and filamentation process. NL and Mated shared 148 (37.7%) of them, and only 29 (7.4%) overlapped with what was up-regulated in Pmated too. Out of the 141 proteins that were up-regulated in Pmated, 112 (79.4%) were used exclusively in Pmated.

SSP is the only gene set that had more overlap among NL, Mated and Pmated than when comparing NL and Mated only. Genes up-regulated in common by NL and Mated and exclusively used by Pmated were the major categories in the three gene sets. There is also an impression of tight control over which genes were used at each stage of the filamentation, in this case, defined by time. From these observations, we deduce that discrete sets of genes are responsible for initiation of mating and the hyphae formation that follows, perhaps in a very systematic manner.

A few novel observations about the mating type loci of *M. lychnidis-dioicae* have been made (Badouin et al., 2015). Unlike most Basidiomycetes that bear the loci encoding mating pheromone receptor determining premating recognition and the homeodomains on the same chromosome, there is a relatively huge 0.6 Mb distance

between the two loci in *M. lychnidis-dioicae*, compared to other fungi where they are found close together on the edge of NRR. Furthermore, the pheromone receptor and pheromone loci were also far apart (1.4 Mb), and the sequence is not similar to other Basidiomycetes (Yockteng et al., 2007). This suggests that the mating mechanism in *M. lychnidis-dioicae* may be novel, or atypical of other Basidiomycetes, and we have shown that it involves proteins for which functional domains have not yet been determined.

Components of the RNAi pathway

RNAi components ARGONAUTE and DICER were found among the predicted proteins (Perlin et al., 2015). Both MVLG_06823 (ARGONAUTE) and MVLG_06899 (DICER) were up-regulated in p1A2 Rich compared to all other haploids and Mated. The other gene predicted to encode ARGONAUTE, MVLG_06899, did not show a significant change in any of the *in vitro* stages but was up-regulated in the early stages of teliosporogenesis (see Chapter 5). These RNAi components perhaps play specific roles in the different stages.

Survival of fungal cells

Based on unpublished data (S.S. Toh), RNA extracted from NL treatment beyond 48 hours was often of very poor quality, suggesting that the fungal cells were no longer active and perhaps dying. In the comparison between Pmated12 and Pmated48, there was an enrichment of gene sets that suggest active cell growth, e.g., histone, DNA repair and other cellular components among the genes up-regulated in Pmated12. These are also the genes that were down-regulated in Pmated48; hence we hypothesize that the fungal cells may have a 48 hour window under mating conditions, whether or not mating partner is available, otherwise they may not survive. This is the first time a timeline is placed on the

survival of fungal cells under mating conditions (stressful environment) before infection takes place.

Conclusion

By applying gene set enrichment analysis in the context of our organism of interest, where known functional domains were scarce, we were able to determine the presence of enrichment of groups of proteins that may be of interest. In order to elucidate the mating pathway of *M. lychnidis-dioicae*, it will be necessary to move beyond computational analysis and this analysis serves to guide further experiments that may lead to elucidation of novel proteins.

CHAPTER 4: SIZE DOES MATTER: STAGING OF *SILENE LATIFOLIA* FLORAL BUDS FOR RNA-SEQUENCING PURPOSES^{16, 17}

Introduction

Silene latifolia, formerly known as *Melandrium album*, is a dioecious flowering plant, a member of the Caryophyllaceae family. It is one of the most studied species among vascular plants that display sexual dimorphism. *S. latifolia* is also a model system for the study of the evolution of plant sex chromosomes. The sex chromosomes are similar in many ways to the older animal sex chromosome system, but these plant sex chromosomes are relatively early in the evolutionary strata, *i.e.*, where the Y chromosome is in its early stages of degeneration (Filatov, 2005; Filatov et al., 2000; Guttman & Charlesworth, 1998; Miller & Kesseli, 2011; Nicolas et al., 2005). Hence, it is a useful model in the elucidation of sex chromosome evolution.

In natural populations of *S. latifolia*, the genus is associated with the fungal phytopathogen *Microbotryum lychnidis-dioicae*, a member of the *Microbotryum violaceum* fungal complex commonly known as anther smuts due to the dark teliospores found in the infected flowers. The fungus is biotrophic and can only complete its sexual

¹⁶ Toh SS, Perlin MH. (2015) Size does matter: Staging of *Silene latifolia* floral buds for transcriptome studies. *Int J Mol Sci* 16: 22027-45. doi: 10.3390/ijms160922027

¹⁷ © 2015 by the authors; licensee MDPI AG, Basel, Switzerland. This work is licensed under the Creative Commons Attribution 4.0 license. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

life cycle in the host. The healthy host plants display strong sexual dimorphism in vegetative (Zluvova et al., 2010) and reproductive traits (Delph et al., 2010). When infected, the fungus replaces the pollen in a male plant; in contrast, the gynoecium is suppressed in infected female flowers, and pseudo-anthers develop to house the teliospores, much like in the male plant (Ruddat et al., 1991; Uchida et al., 2005; Uchida et al., 2003). Infection of females may thus produce at least a partial sex reversal (Ruddat et al., 1991; Uchida et al., 2005; Uchida et al., 2003). The infection renders both the male and female plants sterile, in the former due to the lack of pollen, while in the latter the rudimentary ovary that can no longer set seeds.

When pollinators visit both infected and uninfected flowers for the nectar reward (Shykoff & Bucheli, 1995), the teliospores are transferred along with pollen (Jennersten, 1983). Partial infection may take place in the first year of infection, but systemic infection tends to develop in subsequent years. The deposited teliospores germinate and undergo meiosis, leading to a linear tetrad that produces haploid yeast-like sporidia of opposite mating types, a_1 and a_2 (Garber & Day, 1985), analogous to gametes of the fungus. Conjugation between sporidia of opposite mating types is required to initiate the growth of dikaryotic infectious hyphae (Cummins & Day, 1977), which “travel” through the intercellular spaces of the apical meristem of infected plants where they reside and subsequently infect the plant systematically. However, the exact location and mechanism of infection is still unclear (Schäfer et al., 2010).

Although the Y-chromosome is the sex-determining factor in the host (Warmke, 1946), the genes that enable the development of both the gynoecium and androecium are found on the X-chromosome (Warmke & Blakeslee, 1939; Westergaard, 1946), which is

why all of the floral buds are “bisexual” or hermaphroditic at the earliest stages of development (Grant et al., 1994). The fungus appears to co-develop with the floral buds, such that the developmental stage of the flower bud is an accurate indication of the stage of fungal teliosporogenesis (Uchida et al., 2005; Uchida et al., 2003). Previous expressed sequence tag (EST) library and RNA-Seq transcriptome studies were conducted, for the most part, from healthy plant floral tissue at less than Stage 10 (Grant et al., 1994) and mainly for the purpose of finding expression differences between the male and the female plants (Hardenack et al., 1994; Moccia et al., 2009; Muyle et al., 2012; Zemp et al., 2014). One study that examined infected flowers was done with the aim of elucidating specific expression differences in male plants using a cDNA subtraction approach using pooled tissue for Stages 8-11 (Ageez et al., 2005). Preliminary transcriptome studies identified genes differentially expressed in the fungus during late stage infection of male flowers (Perlin et al., 2015). With the current next-Generation RNA-sequencing technology, we can potentially correlate the interaction of gene expression between the fungus and its host at each stage of their co-development. Such a study ((Toh et al., In Preparation), see Chapter 5) would be a first in terms of stage-by-stage analysis of early flower development and early stages of teliosporogenesis for the *S. latifolia*/*M. lychnidis-dioicae* pathosystem. The accurate staging of the floral buds is therefore crucial to the effective discovery of the gene expression interaction between stages of development and the comparison among male and female, infected and uninfected plants. Floral development stages (Farbos et al., 1997; Grant et al., 1994), in relation to bud size, and the corresponding fungal stages in infected floral tissue (Uchida et al., 2003) had previously been documented [but also see (Schäfer et al., 2010), for SEM images].

However, we recognized that, in addition to light, temperature and other confounding factors that may affect the development of flowers, infection with *M. lychnidis-dioicae* itself may alter the relationship between bud size and floral developmental stage, especially for the very young floral bud. The overall reduction in inflorescence height, flower size and nectar production has previously been reported (Biere & Honders, 1996), indicating that infection by the fungus can indeed alter physical development of the host. Moreover, previous studies found that diseased plants produce smaller flowers with smaller reproductive structures than healthy plants (Biere & Honders, 1996; Shykoff & Kaltz, 1997) and more flowers, especially in females. Thus, after achieving stable infection of the plants serving as sources of tissue for RNA-Seq studies, it was necessary to examine the buds explicitly to stage them according to the laboratory conditions used for propagation. This process was meant to facilitate the rapid and accurate collection of tissue, so as to preserve the integrity of the RNA and to have confidence in the resulting RNA-Seq data.

In devising the infection method, we noted that previous studies showed full germination of teliospores on both seedlings and flowers; with the basidia of such germinations producing promycelia upon meiosis. On seedlings, most promycelia were two-celled compared to four-celled promycelia in the flower. In the two-celled basidia, conjugation hyphae formed rather than basidiospores (*i.e.*, sporidia), and, subsequently, an infectious hypha with a swollen appressorial structure at the tip of the hypha was observed (Schäfer et al., 2010), although the penetration appears to be due to enzymatic processes, rather than via turgor pressure. In contrast, infection in the flower resulted in the basidiospores reproducing via budding, probably because of the higher nutrient

availability. Therefore, inoculation of seedlings was preferred. It also eliminated the need to grow the plants to maturity before infection could take place.

We explored various delivery methods reported in the literature for laboratory infection. Cells were either pre-mated (Biere & Honders, 1996) or provided as unmated mixtures of the two haploid partners of opposite mating type (Fujita et al., 2012). Two main modes of delivery were also commonly employed: injection (*i.e.*, inoculation accompanied by puncture with a needle) (Biere & Honders, 1996; Fujita et al., 2012) and inoculation via application of inoculum as a droplet to undamaged tissue (Hood, 2003; Shykoff & Kaltz, 1997). Some reports employed teliospores as inoculum (Hood, 2003), but since we did not have infected plants to start with, we did not try this until much later (not reported in the data here). As the original aim of this work was to obtain robustly-infected plants, each infection experiment was an attempt to obtain more robustly-infected plants, rather than a systematic study of what was the most effective mode of infection.

In the work described in this paper, we attempted to document the various bud sizes of both male and female, infected and uninfected floral buds, first size categorized by hand (which will be the method of bud collection for RNA-sequencing samples), and verifying the size and their floral development under the dissecting microscope. The definition of the stages was informed by graphical and text illustration of previously-published studies (Farbos et al., 1997; Uchida et al., 2005; Uchida et al., 2003).

Results and Discussion

Infection method

It was difficult to determine the most successful infection method, since there was no one method that worked for every plant in a single trial in more than 20 batches of plants. Each infection experiment was performed with the aim of obtaining more robustly-infected plants. Some infected plants also showed different “degrees” of infection, which made it hard to determine the infection status.

For example, since smut infection can only be observed clearly when a plant flowers, those that remained as rosettes could not be labeled visually as infected until or unless flowering occurred, unless PCR amplification of diagnostic genes for *M.*

lychnidis-dioicae (Akhter & Antonovics, 1999) was employed. Such diagnostics were not of practical application in these studies, since the purpose was to accurately determine flower bud stages for potential tissue collection. There were also some plants that bore buds that never bloomed, turned brown or withered before they were ready to bloom, which again made it hard to determine the infection status. Moreover, some plants bore a few bolts that were smutted and a few that were not, and some “lost” the smut infection after some time. Some even simultaneously bore many bolts that never flowered. All of these could be some degree of disease presentation, since uninfected plants almost never displayed similar symptoms. Such symptoms had previously been observed and reported anecdotally (Fujita et al., 2012; Gibson et al., 2013). Various phenotypic changes were also observed with infection by *M. lychnidis-dioicae* in which the infection percentage was rather low (Shykoff & Kaltz, 1998). In that particular study, only 912 plants (33.6%) were diseased out of the 2717 that were inoculated and flowered (a total of 3808 were

inoculated, and 12% died in the inoculation experiment). Therefore, difficulty with achieving infection may not be uncommon; on the other hand, an acceptable infection rate is usually more than 60% (Gibson et al., 2013).

In our infection experiments (about 20 independent experiments performed over more than two years), 20 plants were injected and 111 plants were inoculated via the application of mated mixtures of the fungus without damage to the shoot apical meristem (treatments were not exclusive). Of these, there was some mortality not attributable to infection (*i.e.*, not significantly more than uninfected controls). Ten plants remained in rosette for as long as we had them; all of these had been inoculated, half with pre-mated cells, and the other half received a mixture of unmated cells. None of the control plants remained in the rosette stage. The batch of seeds used in this study was from progeny of healthy individuals in an infected population, grown in the greenhouse. Although the original population had smut infection, we cannot determine whether “innate” immunity plays a role in the difficulty we experienced in infecting the plants. The infection rate of this combination of host and fungus was originally found to be about 50% (Hood, 2015). Strain-specific resistance has also been suggested (Granberg, 2007), but we are unclear whether the difficulties were due to host-pathogen interactions (Antonovics et al., 2002) or because of imperfections in the artificial infection method. From all of these observations, we found that pre-mated cells were overall more effective as inocula than unmated cells. The total amount of cells used for inocula ranged from 1×10^4 – 5×10^7 cells, and there was no correlation in the optimal amount to use (taking into consideration that the infective suspension volume was always in excess, it is reasonable that we were unable to determine an optimal cell count for infections).

Although field studies have showed earlier onset of flowering between healthy male and female plants, no obvious difference was seen between diseased and healthy plants (Shykoff & Kaltz, 1997). On the other hand, the same study (Shykoff & Kaltz, 1997) found host sex-specific differences, with significantly higher infection rates among *S. latifolia* females than males. Similarly, differences in infection rate between male and female plants were observed in field populations of *S. dioica* (Carlsson-Graner & Pettersson, 2005). In our experiments, there were a total of eight female plants and 13 male plants that were infected (showing smutted flowers at some point), but not all were robustly infected (*i.e.*, with long-term, systemic infection and an otherwise “healthy” host), a requirement for the eventual RNA-Seq experiments. Including the plants that did not bloom at all, we estimate an infection rate of less than 25%. This includes plants that never flowered, but that might have been infected and presented symptoms not observed in uninoculated plants. Also included in this group were four plants that presented with unusual flower morphologies (see below). This lower overall infection rate probably reflects the variety of methods we tested before an optimal approach was reached.

Floral size and staging

The purpose of staging the floral buds was to develop a quick way of determining the floral buds that correspond to the desired floral developmental stage to be collected for RNA-sequencing, especially when examining the interaction of the fungus and host. Therefore, only the robustly-infected plants and corresponding uninfected plants were used. **Fig. 4-1** illustrates how measurements were taken under the dissecting scope on the glass micrometer. The goal for these studies was, thus, to assure or to determine the correspondence between infected and uninfected bud sizes and stages of floral

development. Only plants used as controls in the infection experiments were subsequently taken as tissue sources for the uninfected plants in staging here and later, in the RNA-Seq experiments.

Figure 4-1. Illustration of how the measurements of floral bud (**A**) and anthers (**B**) were taken using a glass stage micrometer.

Each division is 1 mm in (**A**) and 0.1 mm in (**B**). These pictures were taken in the light field, but measurements were typically taken in the dark field.

The staging for male buds depended on the morphology and size of the anthers and the morphology of the petals. Once the stages were established in the male infected floral buds, the male uninfected floral buds were sampled to ensure that the staging would correspond appropriately to the uninfected plant. Morphologically, the stamen first looked like a ball, known as the unilobal anther (**Fig. 4-2**). The ball then “split” into two parts to form the bilobal anther (**Fig. 4-3**) and, finally, into four parts, to form the tetralobal anther (**Fig. 4-4**). The petals were observed as small molds alternating with the sepal at about Stage 7 (**Fig. 4-3**) and developed into tongue-like structures at Stage 8 (**Fig. 4-4A**). As it grew larger, a notch formed in the center yielding a heart-shaped structure or cordiform (**Fig. 4-4B**). Development of the gynoecium was arrested and

became a finger-like projection; thus, the gynoecium was not used as a structure for staging males.

Figure 4-2. Stage 6 of 0.2-mm **(A)** and 0.25-mm **(B)** male infected floral bud.

Unilobal anthers and petals were observed as small molds. “Mushroom”-shaped gynoecium can be observed here. Size bars, approximately 50 µm. g, gynoecium; st, stamen; p, petal.

Figure 4-3. Stage 7 of 0.3-mm male infected floral bud

Bilobal external anthers and petals like small molds. Size bars, approximately 50 µm. g, gynoecium; an, anther; p, petal.

Fig. 4-5 shows a scatter plot of the sizes of the external and internal anthers against the bud size. At Stage 8, external anthers should be 0.2 mm, and internal anthers should be 0.1 mm. At Stage 9, the external anthers measure 0.3-1 mm and the internal anthers 0.2-0.7 mm. Stage 10 anthers measure 1.1 mm (external) and 0.8 mm (internal), respectively. Based on the plot, we can estimate the sizes of buds for each stage of interest (the morphology of anthers and petals was also considered in making the determination). **Fig. 4-6** presents the data in a way that may further inform this strategy, by plotting floral stages against bud sizes. Although there is some overlap of stages, this scatterplot analysis clearly shows separation of Stages 10 and 11 (at a 3.5-mm bud size), whereas Stages 8 and 9 overlap (around 1 mm) as do Stages 9 and 10 (around 3 mm) at some point over the range of bud sizes collected. In summary, for the male plants, the

smallest bud that could be isolated from a cluster was 0.3 mm (determined under a dissecting scope), and such buds were usually hidden in the sepal. Therefore the sepals were collected as the <1-mm group after all buds larger than 1 mm were removed. Based on our measurements, these buds were clearly Stage 8 (Farbos et al., 1997; Uchida et al., 2005; Uchida et al., 2003) or earlier in floral development (**Fig. 4-2, 3 and 4A**). This is consistent with the reference literature (Farbos et al., 1997; Uchida et al., 2005; Uchida et al., 2003). Another quick way to determine the cutoff for Stage 8 floral buds during tissue collection was the appearance of trichome on the calyx. Stage 9 (**Fig. 4-4B**) of floral development was long and was determined in our population to be floral buds of a size larger than 1 mm but smaller than 3 mm. In the male plants we examined, the maturation of anthers from bilobal to tetralobal was the main change. The petal also started to develop a notch in the tongue-like petal to look like a cordiform, with the color changing to pale green. Stage 9 for these male plants lasted a little longer than in the reference literature, in which 2.8-mm buds were already at Stage 10. In our plants, Stage 10 only started around 3 mm and lasted a very short period, until just below 4 mm. The brevity of Stage 10 posed a challenge in tissue collection in having enough Stage 10 floral buds for RNA-Seq analyses. Floral buds larger than 4 mm through bloomed flowers were determined to be Stage 11 and beyond and were broadly categorized as late stage.

Tables 4-1 and 4-2 provide an inventory of staging for infected and uninfected female flowers, respectively. In the female plants, there was great variation in anther sizes in the infected plants, thus precluding the type of scatterplot analysis used for the male flowers. In the reference literature (Farbos et al., 1997; Uchida et al., 2005; Uchida et al., 2003), the appearance of the gynoecium was consistent at each stage regardless of

infection status; hence, this was used as the point of reference for staging. The primordia of the gynoecium appeared to be “mushroom”-shaped at Stage 6, as in the male bud (**Fig. 4-2**). The “barrel and star”-shaped gynoecium can be observed in **Fig. 4-7** at Stage 7. The “barrel and star” then elongated (**Fig. 4-8A**) at Stage 8 before it started to close up (**Fig. 4-8B**) at Stage 9. Styles then started to appear (**Fig. 4-9A**) at Stage 10. Although we attempted to observe the size of the anthers and petals, there was no correlation to the gynoecium stage.

Figure 4-5. Scatterplot of the external and internal anther sizes of the infected and uninfected floral buds against the size of the floral buds.

Ellipsoids indicate the range in bud sizes for flowers of the respective developmental stage groups indicated.

Table 4-1. Size observations and staging of female floral buds from two plants

Bud Size (Rough)	Bud Size (Scope)	Shape of Gynoecium ^a	External Anthers	Internal Anthers	Petals	Stage
<1 (in sepal of 1.5 mm)	-	Petal and stamen primordia start to develop				5
<1 (in sepal of 2 mm)	-	"mushroom" shaped with petal primordia separated from internal stamen primordia	-	-	Not visible	6
<1	0.8	-	Unilobal	-	Not visible	7
<1	0.7	Shallow "barrel and star"	0.1 mm bilobal	0.05 mm unilobal	Not visible	7
0.5	0.7	"barrel and star" shape	0.075 mm bilobal	0.05 mm unilobal	Not visible	7
1	1.4	elongated "barrel and star"	0.25 mm tetralobal	0.15 mm bilobal	Not observed	8
1	1.5	elongated "barrel and star"	0.1 mm bilobal	Not observed	Not visible	8
1.5	1.75	elongated "barrel and star"	0.45 mm tetralobal	0.35 mm tetralobal	0.2-0.3 mm cordiform	8
2	1.75	gynoecium beginning to close	0.35 mm tetralobal	0.25 mm tetralobal	small tongue-like	9
2	2.25	gynoecium beginning to close	0.2 mm tetralobal	Not observed	tongue-like	9
3	2.5	gynoecium beginning to close	0.45 mm tetralobal	0.3 mm tetralobal	0.4 mm tongue-like	9
3	2.5	Styles appearing	0.2 mm tetralobal	Not observed	Overlapping	10
3.5	3.25	elongated gynoecium and closing of "barrel"	0.65 mm tetralobal	0.65 mm tetralobal	1 mm slight cordiform	9-10
3.5	3.25	Styles appearing	0.65 mm tetralobal	0.55 mm tetralobal	0.7 mm cordiform	10

3	3	Styles appearing	0.7 mm tetralobal	0.6 mm tetralobal	0.7 mm cordiform	10
4	3.5	Style appearing	0.2 mm tetralobal	Not observed	0.75 mm	10
4	4	Visible styles	0.75 mm tetralobal	0.6 mm tetralobal	1 mm cordiform	10
5	5	Elongation of styles	0.35 mm tetralobal	Not observed	1.25 mm	>11
5	5.5	Short style and stunted gynoecium	1.2 mm tetralobal	0.8 mm tetralobal	2.5 mm full petal	-
7	7.5	Elongated gynoecium with stunted style	2 mm tetralobal	1.25 mm tetralobal	5.5 mm full petal	>11
8	8.5	Long style	0.75 mm tetralobal	0.25 mm tetralobal	4 mm full petal	>11

^a **Fig. 4-2** illustrates the “mushroom”-shaped gynoecium. **Fig. 4-7** illustrates the “barrel and star”-shaped gynoecium and **Fig. 4-8** illustrates the elongation and closing of the “barrel and star” gynoecium. **Fig. 4-9** shows the styles appearing. **Fig. 4-4A** illustrates the tongue-like petals and **Fig. 4-4B** shows the cordiform petals.

Table 4-2. Size observations and staging of female uninfected floral buds from four plants.

Bud Size (Rough)	Bud Size (Scope)	Shape of Gynoecium ^a	Petals	Stage
<1	0.7	“barrel and star” shape	Visible and not overlapping	7
<1	0.7	“barrel and star” shape	Hardly visible	7
1	1.1	“barrel and star” shape	Visible and larger than others	7
1	1.2	elongated “barrel and star”	Tongue-like	8
Slightly >1 mm	1.5	elongated “barrel and star”	0.7 mm	8
1	1.5	elongated “barrel and star”	Cordiform	8
1.5	1.5	elongated “barrel and star”	Hardly distinguishable from anthers	8
2	1.75	elongated “barrel and star”	Visible	8
2	2	elongated “barrel and star”	Visible and not overlapping	8
2	2	elongated “barrel and star”	No petal	8
2.5	2.5	elongated “barrel and star”	0.2 mm transparent cordiform	8
2.5	2.75	gynoecium beginning to close	0.3 mm cordiform	9
3	2.5	Styles appearing	Visible and not overlapping	10
3	3	Visible styles	0.35 mm cordiform	10.1
4	4	Styles appearing	Visible and overlapping	10
4	4	Visible styles	0.75 mm cordiform	Mid 10
4	4	Visible styles	0.75 mm cordiform	10.2
4	4.5	Visible styles	0.9 mm cordiform	10
4.5	4.5	Visible styles	1.5 mm fully formed	10
5	5	Long styles, petal half of ovary	1.5 mm half of ovary	10

6	6	Long styles, petal as long as ovary	2.5 mm, covers ovary	>11
7	7	Long styles, petal as long as ovary	3.5 mm, covers ovary	>11
7	6	Long styles, petal as long as ovary	1.75 mm fully formed	>11

^a **Fig. 4-2** illustrates the “mushroom”-shaped gynoecium. **Fig. 4-7** illustrates the “barrel and star”-shaped gynoecium and **Fig. 4-8** illustrates the elongation and closing of the “barrel and star” gynoecium. **Fig. 4-9** shows the styles appearing. **Fig. 4-4A** illustrates the tongue-like petals and **Fig. 4-4B** shows the cordiform petals.

In the female plants, the divergence from published literature (Farbos et al., 1997) was a little more pronounced, although the floral bud size for each stage was consistent between the infected and uninfected plants (**Tables 4-1** and **4-2**). For practicality in tissue collection, we collected the buds smaller than 1 mm, often found in the sepal, as one category. In the reference literature (Farbos et al., 1997), female buds of this size should be at Stage 6 and below. In our buds, however, they had already reached Stage 7 (**Fig. 4-7**).

Buds ranging from 1.0-2.9 mm appeared to be in Stages 8 (**Fig. 4-8A**) and 9 (**Fig. 4-8B**) of floral development. Again for practical reasons, it was not possible to separate the buds in Stage 8 from Stage 9 for tissue collection purposes. In the reference literature (Farbos et al., 1997), these buds should only have been developed to Stage 7. Although the petal size should be above 1 mm by Stage 8, the petals in our buds were much smaller. We would attribute this to the fact that the overall size of the bud was much smaller than in the reference literature (Farbos et al., 1997); hence a larger petal would not be possible.

Buds ranging from 3.0-4.9 mm appeared to be in Stage 10 (**Fig. 4-9**) of floral development. This is much smaller than expected compared to the reference literature

(Farbos et al., 1997), where Stage 10 buds range from 8-11 mm in size. Floral buds 5 mm and beyond were categorized as late stage.

Figure 4-7. Stage 7 female uninfected floral bud.

The gynoecium has a “barrel and star” appearance from the top. Stamen primordia can be observed. Size bar, approximately 50 μm. g, gynoecium; st, stamen.

Figure 4-8. Stage 8 (A) and 9 (B) of female infected floral buds.

Stage 8 had an elongated “barrel and star” appearance while the gynoecium of the Stage 9 bud was closing up on top. In addition, the bilobal anthers were visible, and the petals were tongue-like shaped in the Stage 9 bud. Size bars, approximately 100 μm and 200 μm, respectively, in (A) and (B). g, gynoecium; p, petal; an, anther.

Figure 4-9. Stage 10 of female infected (**A**) and uninfected (**B**) floral buds where the styles start to appear.

In the infected bud, the anthers continue to develop, while the stamen in the uninfected bud becomes rudimentary. Size bars, approximately 200 µm and 250 µm, respectively, in **A** and **B**. g, gynoecium; sty, style; p, petal; an, anther; st, stamen.

To confirm that our size range was valid, we looked into previous studies of this pathosystem that involved smutted floral buds. A comparison of healthy and infected male and female gene expression used floral buds of a size 1-3.5 mm for both infected and uninfected buds and determined these to be Stages 8-11 (Ageez et al., 2005). In another study to identify MADS box genes in the early stages of floral development up to Stage 10, floral buds of less than 4 mm were used in the generation of their cDNA library (Hardenack et al., 1994). These previous studies gave us confidence that we had staged the floral buds accurately.

RNA extraction

Collected floral bud tissues were processed using a previously-published protocol (Perlin et al., 2015). **Table 4-3** is a summary of the estimated weight of the collected tissue and amount of crude RNA extracted. The amount of RNA yield was dependent on

the amount of tissue used as well as the number of extraction columns used. While each column can extract RNA from up to 100 mg of tissue, not every sample had that much tissue for extraction. Moreover, since only about 40 µg of RNA were needed for all quality assessment and RNA-Sequencing, not all ground tissue was processed.

Although all of the tissue submitted for RNA-sequencing met the integrity requirements of the Broad Institute where RNA-Seq was performed, not all of the samples produced a good RNA integrity number (RIN). We hypothesize that the RIN tool may not have been optimized for the electropherogram of plant and fungal samples since it was trained with human, mouse and rat RNA (Schroeder et al., 2006). It should also be noted that the integrity of the RNA does not necessarily reflect the accuracy of staging.

Table 4-3. Summary of RNA extracted from floral bud tissue based on the stages of interest.

Tissue type	Stage	Weight of tissue (mg)	Amount of crude RNA (µg)	RNA Integrity number (RIN)^a
Male infected	8	91	64	N/A
Male infected	8	38.2	62	9.4
Male infected	9	150.6	61	N/A
Male infected	9	101.5	85	8.9
Male infected	10	116.2	180	9
Male infected	10	110.2	115	9
Female infected	7	34.7	51	6.4
Female infected	9	55.2	91	6.4
Female infected	10	187.8	143	9.3
Male uninfected	8	28	45	6.9
Male uninfected	9	34.4	40	6.1
Male uninfected	9	19.2	63	N/A
Female uninfected	9	30	104	6.3
Female uninfected	10	90	103	6.2
Female uninfected	10	109	120	N/A

^aRIN as determined for Eukaryotic RNA chip for BioAnalyzer, using mammalian systems as standard for rRNA positions and integrity.

Unusual floral buds

There were many difficulties achieving stable infection with plants consistently producing smutted flowers. In the course of monitoring the plants, we noticed some morphologies that differed from the “normal” infection phenotype. Some of these included plants that remained in the rosette stage or floral buds that appeared, but never bloomed (**Fig. 2-10A**). We do not know whether these variations in morphology were sex-dependent, but the infected female plants tend to develop more variable phenotypes, such as the early death of buds. There were also some gross distortions in flowers that made it impossible to tell the gender unless a genetic analysis was performed or a “normal” flower presented itself on the plant at some point. Some examples of these distortions include: exposed smutted anthers with no calyx and petals (assumed male); stunted ovaries with no obvious smutted anthers (assumed female); infected male flowers developing pseudo-gynoecium (**Fig. 4-10B** and **4-10C**); and a robust ovary that can set seeds with smutted anthers (assumed female).

At this point, we are uncertain if these unusual phenotypes were due to mutation(s) in the pathogen as in the petalless flower (Fujita et al., 2012) or if they reflect a physiological defensive mechanism of the host, as a trade-off of being diseased (Shykoff & Kaltz, 1998). Alternatively, perhaps they reflect the incompatibility of the host and the pathogen (Antonovics et al., 2002). There has not been a systematic inventory published of all possible phenotypes of disease in this pathosystem, especially where each “phenotype” was only found in one plant. It would be interesting to investigate the genetic/ gene expression basis of the variation as, for example, in the petal-less flower (Fujita et al., 2012).

Materials and Methods

Germination

Silene latifolia seeds were harvested in summer 2009 from greenhouse-grown plants; these were originally from seeds of a field population in Clover Hollow (37.328-80.488) near Mt. Lake Biological Station, Virginia, and were kindly provided by Michael E. Hood. Seeds were sterilized and hydrated by soaking in a sterilizing solution (40% household bleach, 20% absolute ethanol and 1 drop of triton X-100 as surfactant per 50 ml of solution) and washing five times in sterile distilled water, for 2 min per wash with constant agitation. Each seed was then individually planted in closed milk jars on sterile 0.3% phytagar (Thermo Fisher, Pittsburgh, Pennsylvania, USA), 0.5× MS (Murashige and Skoog) salt (Sigma-Aldrich, St. Louis, Missouri, USA) and 0.05% MES (2-(*N*-morpholino)ethanesulfonic acid) buffer (Thermo Fisher, Pittsburgh, Pennsylvania, USA). Each jar was incubated at 4 °C for 5 days to synchronize germination before transferring them to the growth chambers. In accordance with the growth condition and its effect (unpublished data) on the maintenance of flowering status and infection in its hosts, we maintained the growth chamber at 20°C with 13 h of daylight. Germination started within 3 days with the appearance of the radicle. When the seedlings were about 15 days old, they were transferred into 2 inch square pots filled with Sunshine MVP Professional Growing Mix (Sun Gro Horticulture Canada Ltd, Agawam, Massachusetts, USA, Cat No. 02392868) and replaced into the growth chamber. Humidity was kept high initially using dome covers and flood trays. Seedlings were gradually exposed to the chamber environment for increasing amounts of time daily in order for the seedlings to harden and adapt to the lower humidity. The plants were repotted to 4 inch diameter round pots when

they began to bolt, at about 30 days old. They were further repotted into 7 inch diameter round pots when they had almost attained maximum height or when the volume of soil was not sufficient to provide hydration for the plant. The plants were watered every other day with 100-ppm fertilizer (Peters Professional 15-16-17 Peat-Lite Special[®] Formula No: S12893, Dublin, Ohio, USA) to wet the soil thoroughly.

Infection with *Microbotryum lychnidis-dioicae*

Haploid fungal cells (p1A1 and p1A2) (“Lamole strain”: GenBank 100-15Lamole; (Antonovics et al., 2002; Hood et al., 2004; Perlin et al., 2015)) grown under “rich” conditions were allowed to grow for 4-5 days on yeast peptone dextrose media (YPD; 1% yeast extract, 10% dextrose, 2% peptone, 2% agar) at room temperature. To prepare pre-mated mixtures, the cells were harvested and resuspended in sterile distilled water, adjusted to a concentration of 1×10^9 cells/ml in equal proportion before being spotted onto nutrient-free solid agar media (2% agar) in 50- μ l spots. The plates were allowed to dry and then incubated at 14 °C for about 48 h. Cells were inspected for conjugation tubes under the microscope before being harvested for infection purposes. To prepare unmated mixtures, the cells were harvested and resuspended in sterile distilled water in equal proportion before being used directly for infection purposes. For infection, various permutations of treatments were explored, including pre-mated vs. unmated cells, infection on different days after germination (3, 10, 12, 21, 28 days after germination and dual-infection for some), injection vs. inoculation (placing the cell suspension on the central whorl) of meristem and various volumes and concentrations of fungal cells used. For staging purposes, the method of infection was not taken into account. Instead, robustly-infected and uninfected plants that could be used for tissue collection were

staged specifically for RNA-Seq purposes. The gender, infection status and the growth chamber where the plant was located were noted as possible factors that might affect the size of floral buds and their developmental stage. Infection status was determined by the consistent blooming of fully-smutted flowers, and for uninfected plants, we used the control plants that were not in any way inoculated with the fungus and that grew in a separate chamber from infected plants.

Floral bud collection and examination

At the time of collection, each inflorescence was separated using forceps and hand-categorized at 1-mm intervals based on a marked scale on the collection plate. Each collection was limited to 30 min on a single plant, to minimize dehydration of the bud before measurement under the dissecting scope. All floral buds were examined as fresh tissue.

The floral buds were then staged according to previous literature (Farbos et al., 1997; Uchida et al., 2003) under the dissecting scope (Nikon, Melville, New York, USA, Model: SMZ-U). For male buds, both infected and uninfected, the stages were determined by the petal and size of the external and internal anthers. For female plants, the shape of the gynoecium was the major determinant. These measurements were made with a glass stage micrometer (Imaging Research, Inc., St. Catherines, Ontario, Canada) under the dissecting scope. The results were tabulated to determine if there was a correlation of floral bud size to its developmental stages across different plants.

Conclusion

In this study, we were able to correlate stages of flower development to bud size in both uninfected and infected male and female *S. latifolia* plants, a critical aspect of conducting reproducible analyses of differential transcriptome studies for discrete stages of late infection by *M. lychnidis-dioicae*. This was critical to our later large study of differential gene expression during these stages of development, for both the host plant and the fungus (Toh et al., In Preparation). As an example, comparison of Stage 8 with Stage 9 for infected male flowers indicated differences for approximately 140 *M. lychnidis-dioicae* genes, with a false discovery rate (FDR) $<10^{-5}$. One of the more commonly observed changes was upregulation of genes in the MFS (major facilitator superfamily) family of transporters. In addition, in these preliminary analyses, we observed upregulation of some members of the glycosyl hydrolase family and other pectinesterases associated with host plant cell wall degradation. Such preliminary results support the efficacy of the type of staging we used in the current study.

Throughout this study, we encountered and overcame difficulties obtaining reliable systemic infection, another necessary aspect of the RNA-Seq work comparing infected to uninfected plants. As part of this study, a subset of host plants was not used for collection purposes for sizing and RNA-Seq, as they exhibited some symptoms that did not fit the classic characteristics reported previously for infections in the *S. latifolia*/*M. lychnidis-dioicae* system. We hypothesize that the unusual plant symptoms we observed may be part of the plant's physiological defense strategy against the fungus (Gibson et al., 2013) or possible incompatibility between the host and pathogen (Antonovics et al., 2002). By suppressing bolting (remaining as a rosette), flowering

(non-flowering bolts) or blooming (pre-mature death of buds), the host would present an environment where the fungus would be unable to propagate and complete its sexual life cycle through teliosporogenesis in the anthers. Another plausible hypothesis is that a non-optimal infection method resulted in a hyper-infection that interfered with the normal plant development. Although this would ultimately prevent the fungus from completing its lifecycle, such cases of apparently counter-productive hyper-virulence have been observed in other systems (Alizon et al., 2009; Smith, 2011). One possible way to address aspects of these unusual plant symptoms in the future will be to conduct RNA-Seq on tissues from such plants to examine changes in the transcription levels of potential target genes associated with development (*e.g.*, MADS-box genes).

The simple approach in staging the floral buds developed in this study allowed us to quickly determine the correct floral bud sizes to collect. While the male buds were not too different from the documented literature, the female buds differed considerably. Failure to perform the in-house staging would have resulted in missing the early stages of fungal development in the floral buds used for the extensive transcriptome analyses.

CHAPTER 5: TRANSCRIPTOMIC ANALYSIS OF *MICROBOTRYUM LYCHNIDIS-DIOICAE* IN YOUNG FLORAL TISSUE OF HOST

Introduction

Silene latifolia is the dioecious flowering plant found in the Caryophyllaceae family. It is also one of the host plants that can become infected by smut fungus, which evolve closely with their host species (Le Gac et al., 2007a). In this case, the smut fungus is *Microbotryum lychnidis-dioicae*. The dimorphic sex chromosome, mainly the presence or absence of the Y chromosome of the host plant, controls the appearance and suppression of the appropriate reproductive organs (Westergaard, 1958). In the event that a male host is infected, the teliospores take the place of the pollen; in the female host, infection performs the additional task of suppressing the gynoecium and instructing the androecium to continue its development instead, so that the mature anthers can house the teliospores in a manner similar to a male host (Ruddat et al., 1991; Uchida et al., 2005; Uchida et al., 2003).

The co-development of the fungus in the floral bud was established, such that the stage of the floral bud (Grant et al., 1994) is indicative of the stage at which the fungus is engaged in teliosporogenesis (Farbos et al., 1997; Uchida et al., 2003). Most of these actions take place in early floral bud development, between Stages 8-11. Hence, we embarked on an effort to tease apart the transcriptomic interaction of the pathogen and its

hosts, by performing RNA Sequencing on the discrete stages of the floral tissue based on careful size determination of the corresponding stages ((Toh & Perlin, 2015), See Chapter 3).

RNA Sequencing (RNA-Seq) has revolutionized the way genetic studies are done (Wang et al., 2009). Instead of approaching a system gene-by-gene, as in classical genetics or reverse genetics methods, we are now able to look at gene expression changes globally and discover which genes may have caused a phenotype. Dual RNA-Seq (Westermann et al., 2012) of host and pathogen, which allows examination of the transcriptomes in parallel, takes the concept further by allowing an *in vivo* view of what the pathogen does to the host and how the host defends itself, at least as evidenced by alteration of gene expression at the transcriptional level.

Such dual RNA-Seq has been used in compatible and incompatible plant-pathogen systems to uncover pathogenesis and defense-related genes, for example, in the rice (*Oryza sativa*) and blast fungus (*Magnaporthe oryzae*) (Kawahara et al., 2012) system, the poplar and rust fungus (*Melampsora larici-populina*) (Petre et al., 2012) system and the cotton (*Gossypium barbadense*) and wilt fungus (*Verticillium dahliae*) system (Xu et al., 2011). A study with canola (*Brassica napus*) and *Leptosphaeria* species (Lowe et al., 2014) focused on discovery of important CAZyme genes instead of analyzing the system broadly. However, the larger body of data is available for future scrutiny.

In this study, we compared the gene expression of *M. lychnidis-dioicae* in male and female hosts, across discrete early floral development stages, to the transcriptome of

the mated fungal cells, to provide a broad view of the gene expression in the floral buds of this fascinating phytopathogen system.

Materials and Methods

Host plants and fungal strains

Silene latifolia seeds were harvested in Summer 2009 from greenhouse grown plants; these were originally from seeds of a field population in Clover Hollow (37.328 – 80.488) near Mt. Lake Biological Station, Virginia, and were kindly provided by M. Hood. Germination and infection of the host plant was performed as published ((Perlin et al., 2015; Toh & Perlin, 2015), see Chapter 2 & 3).

The floral buds were staged according to the methods presented in Chapter 2 (Toh & Perlin, 2015). In addition, the pedicels of the floral cluster were collected as the floral stem tissue. From two male infected plants, three floral stem (MIFS), two Stage 8 (MI8), two Stage 9 (MI9), two Stage 10 (MI10) and five late stage (MILate) samples were submitted for RNA-Seq. Two MILate samples were of low quality and not used in the final analysis. From a single female infected plant, two floral stem (FIFS), one Stage 7 (FI7), one Stage 9 (FI9), one Stage 10 (FI10) and two late stage (FILate) samples were submitted.

In analyzing the differential expression data from infected plants, independent changes were considered first before considering changes that happened in both male and female, since they are biologically very different. However, since there is only one sample for most of the female stages, it is difficult to determine if a change is not present or simply not strong enough statistically to be reported. To overcome that, we also made combined analysis for each stage, where transcripts from male and female were treated as biological replicates. MI8 and FI7 were treated as the biological replicates. Hence, we analyzed them as MIFIFS, MIFI8, MIFI9, MIFI10 and MIFILate.

RNA extraction and sequencing

RNA was extracted and processed as previously published ((Perlin et al., 2015), Chapter 2, see Appendix 1 for details). A minimum of 5 µg of DNA-free RNA was prepared for each desired stage and sent to the Broad Institute for RNA-Sequencing via Illumina. RNA-Seq data from different conditions were processed using the Trinity pipeline (Grabherr et al., 2011; Haas et al., 2013), and RNA-Seq reads from each sample were aligned using bowtie (Langmead et al., 2009) against protein coding sequences extracted based on annotation. Fungal gene expression levels were estimated by RSEM (Li & Dewey, 2011), and edgeR with TMM normalization (Kadota et al., 2012; Robinson et al., 2010) was used to identify differentially expressed genes between each pair of conditions with a corrected p-value cutoff of 1e-5. RNA-Seq reads from all infected host samples were assembled using Trinity. Assembled transcripts were filtered based on the alignment with the *M. lychnidis-dioicae* fungal genome (Fontanillas et al., 2015; Perlin et al., 2015).

Venn Diagrams

Venn diagrams were generated using the online tool Venny (Oliveros, 2007-2015). The edgeR data was sorted to extract the list of the predicted genes (Perlin et al., 2015) that were significantly up-regulated or down-regulated (false discovery rate, FDR < 1E-5) against a reference condition in the pairwise analysis.

Gene Set Enrichment Analysis (GSEA) Preranked

Gene set enrichment analysis (Mootha et al., 2003; Subramanian et al., 2005) was performed to discover the possible enrichment of PFAM domains, gene ontology terms (Conesa et al., 2005) and KEGG pathway maps (Kanehisa et al., 2015) domains in the

conditions studied using the GenePattern software (Reich et al., 2006). In addition, the predicted small secreted proteins (SSPs, less than 250 amino acids long), secreted proteins (SPs, more than 250 amino acids long) and unique proteins based on orthoMCL analysis (Li et al., 2003; Perlin et al., 2015) were included as gene sets. The resulting gene set list (.gmt) consisted of 3651 gene sets, with a range of 1-1116 members. Ranked lists (.rnk) of every pair-wise comparison made were processed by the GSEAPreranked module software. When running the software, the algorithm was set to allow minimum gene set of one and maximum of 1200. Gene sets that were significantly enriched at $FDR < 25\%$ or nominal $pvalue < 1\%$ were listed and examined for further hypothesis-generation.

Results and Discussion

Differential expression in infected male floral buds

Fig. 5-1 shows the Venn diagram of the number of genes that were up- and down-regulated in infected male floral buds compared to Mated. A total of 1645 (22.3%) and 1756 (23.8%) were found to be differentially regulated, up and down, respectively. About 30% of both groups applied to all the floral stages. When compared to male infected floral stems (MIFS), **Fig. 5-2** shows that genes that were turned on were rather discrete to the specific floral stages examined, whereas the down-regulated genes had more overlap.

Differential expression in infected female floral buds

Fig. 5-3 shows the Venn diagram of the number of genes that were found to be up- and down-regulated in female floral buds compared to Mated. A total of 1322 (17.9%) and 1382 (18.8%) were found to be differentially regulated, up and down, respectively. About 30% of the up-regulated and 23.5% of the down-regulated applied to all the floral stages. When the female floral bud stages were compared with female infected floral stems (referred to as FIFS), **Fig. 5-4** shows that there were only 83 and 155 genes were up- and down-regulated, respectively. Most of the changes (71% and 80% up- and down-regulated) happened exclusively in FILate. This result needs to be interpreted with caution, because there were only one sample each of Stages 7, 9 and 10 female infected floral buds (FI7, FI9 and FI10). This may have reduced the sensitivity and reproducibility in picking up slightly differentially regulated genes.

Combined analysis

Due to limited female infected sample size, we performed a combined analysis in an attempt to pick up some of the differentially regulated genes that were missing. For convenience of analysis, FI7 was grouped with MI8 as the same condition, since they were of the same size buds, and we called them MIFI8.

Fig. 5-5 shows a summary Venn diagram of the number of unique genes that were identified as differentially regulated in only the male (MI) and female (FI) and when the data sets were combined (MIFI). We interpret the genes identified in the combined analysis cautiously as they will be prone to false positives, since the two biological replicates for male floral buds may bias the statistical analysis and they may not be truly significant in both sexes. Also, by combining the corresponding stages, we are assuming that every *M. lychnidis-dioicae* gene will behave similarly in the two sexes at every stage, which is probably not true. Trends that were only significantly differentially expressed in one sex will most likely be missing in the combined analysis, especially if the expression levels that were in opposition to each other. Therefore, the independent analysis is also warranted for these analyses.

From the figure, an additional 58 (18.4%) and 61 (12.8%) genes were found to be up- and down-regulated respectively when the male and female samples were combined for each stage (*i.e.* MIFI8, MIFI9, MIFI10 and MIFILate) and compared to MIFIFS.

Figure 5-6. edgeR smear plots of Log fold change against average Log counts per million for male (MI) against female (FI) at each stage: **(A)** FS, **(B)** 8, **(C)** 9, **(D)** 10 and **(E)** Late respectively. Only a very small number of genes were differentially regulated, hence combining them stage-wise for further comparison should be acceptable. The black dots are non-differentially expressed genes and red dots are differentially expressed genes.

Global gene set enrichment

In gene set enrichment analysis, we try to look at the broader picture of fungal gene expression in male and female independently and when these data are combined as biological replicates. The plant stages we are examining in this study are those of the fungus transitioning from “travelling” in the plant as dikaryon to entering the floral meristem and eventually locating the androecium and taking the place of pollen to become teliospores. This is a complicated process, and in the female, the fungal cells have the additional task of instructing the host to develop the androecium instead of the gynoecium. We attempted to break down the early stages of floral development, hoping to capture the gene expression changes when the fungus undergoes a morphological switch. It will be difficult to have the full picture without combined analysis of the host plants at this point, but we hope to identify similarities and differences in fungal gene expression in the different hosts.

Table 5-1 and **5-2** list the gene sets that were enriched among the genes that were up- and down-regulated, respectively, in plants compared to Mated, based on individual sex and combined analysis at each floral development stage. Enriched gene sets that are found in one sex and also in the combined analysis probably indicate that the sexes have a neutral or same directional regulation in those genes. If they were only found in one sex and not in the combined analyses, then the other sex may have regulation of those genes in the other direction, thus negating the significance of differential regulation. Gene sets that are found in the combined analysis and not in the individual sex may suggest that the change is subtle and only upon increase number of “biological replicates” did the change become detectable.

We observed an enrichment of some of the gene sets responsible for active cell growth among the up-regulated genes. This is expected, especially when compared to Mated where the medium was nutrient limited and conjugation was already completed. Some of these gene sets were only detected in the combined analysis and not in the individual sexes, which emphasizes the importance of the combined analysis.

Genes sets that were detected to be enriched only in the male were Nucleosome Assembly (GO:0006334) and Response to Stress (GO: 0006950), while the female had Oligopeptide transporters (PF3169) and glyoxal oxidase N-terminus (PF07250) in the early stage and Peptidase S8 (PF00082) and domains of unknown function (PF09118 and PF06280) in the later stages. Multicopper oxidases were detected in the combined analysis but it should be noted that all three domains referred to the same three genes (MVLG_01868, 02184 and 03092). Upon examination of the raw data as listed in **Table 5-3**, we discovered that the transcript per million of these three genes were just as high in female, but perhaps because there is only one biological replicate for most stages in female, differential expression could not be determined easily. This observation highlighted the importance of analyzing the combined data, and also emphasized the need for reliable biological replicates to make the best judgment of the gene expression data. It also points to the importance of manual examination of data when there are few replicates and where expression changes are subtle.

Table 5-1. List of gene sets that were enriched in genes up-regulated in plant compared to Mated, arranged in order of decreasing magnitude of NES.

Common enrichment in genes up-regulated in both male and female over more than one stage when compared to Mated		
<ul style="list-style-type: none"> • SMALL SECRETED PROTEIN • UNIQUE TO MVIO • SECRETED PROTEIN • PF00295.10 GLYCO_HYDRO_28[GLYCOSYL HYDROLASES FAMILY 28]* • GO:0006412[TRANSLATION>BP] • GO:0006281[DNA REPAIR>BP] • PF00125.17 HISTONE[CORE HISTONE H2A/H2B/H3/H4] • GO:0003735[STRUCTURAL CONSTITUENT OF RIBOSOME>MF] • GO:0005840[RIBOSOME>CC] • PF00012.13 HSP70[HSP70 PROTEIN] • PF06723.6 MREB_MBL[MREB/MBL PROTEIN] • GO:0000786[NUCLEOSOME>CC] • PF01095.12 PECTINESTERASE[PECTINESTERASE] 		
MIFS vs. Mated	FIFS vs. Mated	MIFIFS vs. Mated
• None	• None	• None
MI8 vs. Mated	FI7 vs. Mated	MIFI8 vs. Mated
<ul style="list-style-type: none"> • GO:0006334[NUCLEOSME ASSEMBLY>BP]# • PF07731.7 CU-OXIDASE_2[MULTICOPPER OXIDASE]* 	<ul style="list-style-type: none"> • PF03169.8 OPT[OPT OLIGOPEPTIDE TRANSPORTER PROTEIN] 	<ul style="list-style-type: none"> • PF07250.4 GLYOXAL_OXID_N[GLYOXAL OXIDASE N-TERMINUS]^ • PF07731.7 CU-OXIDASE_2[MULTICOPPER OXIDASE]* • PF07732.8 CU-OXIDASE_3[MULTICOPPER OXIDASE]# • PF00394.15 CU-OXIDASE[MULTICOPPER OXIDASE]#
MI9 vs. Mated	FI9 vs. Mated	MIFI9 vs. Mated
<ul style="list-style-type: none"> • GO:0006334[NUCLEOSME ASSEMBLY>BP]# • GO:0006950[RESPONSE TO STRESS>BP]# 	<ul style="list-style-type: none"> • PF07250.4 GLYOXAL_OXID_N[GLYOXAL OXIDASE N-TERMINUS]# 	<ul style="list-style-type: none"> • PF09118.4 DUF1929[DOMAIN OF UNKNOWN FUNCTION (DUF1929)] • PF07732.8 CU-OXIDASE_3[MULTICOPPER OXIDASE]#
MI10 vs. Mated	FI10 vs. Mated	MIFI10 vs. Mated
<ul style="list-style-type: none"> • GO:0006334[NUCLEOSME ASSEMBLY>BP]*# • GO:0005634[NUCLEUS>CC]* • GO:0003677[DNA BINDING>MF]* • GO:0006950[RESPONSE TO STRESS>BP]# 	<ul style="list-style-type: none"> • PF07250.4 GLYOXAL_OXID_N[GLYOXAL OXIDASE N-TERMINUS]# • PF09118.4 DUF1929[DOMAIN OF UNKNOWN FUNCTION (DUF1929)] 	<ul style="list-style-type: none"> • GO:0006334[NUCLEOSME ASSEMBLY>BP]* • GO:0005634[NUCLEUS>CC]*# • GO:0003677[DNA BINDING>MF]* • GO:0006414[TRANSLATIONAL ELONGATION>BP] • PF00394.15 CU-OXIDASE[MULTICOPPER OXIDASE]# • PF11754.1 VELVET[VELVET FACTOR]

MILate vs. Mated	FIlate vs. Mated	MIFILate vs. Mated
<ul style="list-style-type: none"> None 	<ul style="list-style-type: none"> PF00082.15 PEPTIDASE_S8[SUBTILASE FAMILY] PF06280.5 DUF1034[FN3-LIKE DOMAIN (DUF1034)] PF01490.11 AA_TRANS[TRANSMEMBRANE AMINO ACID TRANSPORTER PROTEIN] 	<ul style="list-style-type: none"> GO:0005634[NUCLEUS>C C]#

* denotes gene set found in more than one comparison in that stage.

denotes gene set found in more than one stage of that type of condition.

^denotes gene set not found among similar stage or condition, but in other stages or condition.

Whereas among the down-regulated genes, the gene sets related to peptidases and transporters appear to be enriched. One of the domains that we were familiar with and have been paying attention to is the secretory lipase domain (PF03583). It was found to be significantly up-regulated in mating cells exposed to phytol (See Chapter 5). Its down-regulation in plants suggests that these proteins have specific functions right after mating, supporting our initial hypothesis that they may be directly involved in hyphae formation, as in the response elicited by exposure to phytol.

Table 5-2. List of gene sets that were enriched in the genes down-regulated in plants when compared with Mated with no order of normalized enrichment score (NES).

Common enrichment in more than 2 of Mated vs. MI, FI and MIFI stages		
<ul style="list-style-type: none"> PF10340.2 DUF2424[PROTEIN OF UNKNOWN FUNCTION (DUF2424)] PF05572.6 PEPTIDASE_M43[PREGNANCY-ASSOCIATED PLASMA PROTEIN-A] PF00501.21 AMP-BINDING[AMP-BINDING ENZYME] PF05686.5 DUF821[ARABIDOPSIS THALIANA PROTEIN OF UNKNOWN FUNCTION (DUF821)] PF03583.7 LIP[SECRETORY LIPASE] PF02129.11 PEPTIDASE_S15[X-PRO DIPEPTIDYL-PEPTIDASE (S15 FAMILY)] TIGR00976[/NOND: HYDROLASE COCE/NOND FAMILY PROTEIN] PF02129.11 PEPTIDASE_S15[X-PRO DIPEPTIDYL-PEPTIDASE (S15 FAMILY)] PF08613.4 CYCLIN[CYCLIN] 		
Mated vs. MIFIFS	Mated vs. MIFI9	Mated vs. MI10
<ul style="list-style-type: none"> PF12464.1 MAC[MALTOSE ACETYLTRANSFERASE] PF00383.15 DCMP_CYT_DEAM_1[CYTIDINE AND DEOXYCYTIDYLATE DEAMINASE ZINC- 	<ul style="list-style-type: none"> PF00664.16 ABC_MEMBRANE[ABC TRANSPORTER TRANSMEMBRANE REGION] PF02815.12 MIR[MIR 	<ul style="list-style-type: none"> PF06609.6 TRI12[FUNGAL TRICHOTHECENE EFFLUX PUMP (TRI12)]

BINDING REGION] • PF00621.13[RHOGEF[RHOGE F DOMAIN] • PF11899.1[DUF3419[PROTEIN OF UNKNOWN FUNCTION (DUF3419)]	DOMAIN]	
Mated vs. MIF110	Mated vs. MILate	Mated vs. FILate
• PF06609.6[TRI12[FUNGAL TRICHOECENE EFFLUX PUMP (TRI12)] • PF08407.4[CHITIN_SYNTH_1N[CHITIN SYNTHASE N-TERMINAL]	• PF08530.3[PEPX_C[X-PRO DIPEPTIDYL-PEPTIDASE C-TERMINAL NON-CATALYTIC DOMAIN]	• PF07859.6[ABHYDROLASE_3[ALPHA/BETA HYDROLASE FOLD]

Table 5-3. Transcripts per million of three copper oxidases detected to be up-regulated in male but not in female, to illustrate the importance of biological replicates.

Sample_ID	MVLG_01868	MVLG_02184	MVLG_03092
Mated_1	2.2	0	0.47
Mated_2	3.04	0	0.49
MIFS_1	47.81	0*	0*
MIFS_2	83.33	161.01	36.54
MIFS_3	40.28	119.58	44.74
MI8_1	35.48	172.71	13.17
MI8_2	36.02	127.45	12.44
MI9_1	29.22	197.18	13.43
MI9_2	26.28	148.89	4.6
MI10_1	30.21	303.95	7.37
MI10_2	22.31	246.64	6.89
MILate_1	55.25	63.43	219.36
MILate_4	84.87	181.65	791.23
MILate_5	43.22	181.41	451.16
FIFS_1	70.86	45.76	11.89
FIFS_2	70.69	17.13	10.05
FI8_1	72.21	191.48	31.89
FI9_1	45.89	204.57	20.2
FI10_1	38.95	250.5	13.22
FILate_1	73.17	352.89	343.05
FILate_2	63.49	200.59	277.05

* Probably outliers

Subfix denotes the replicate number

The most prominent gene set enrichment among the genes up-regulated in plants were the small secreted protein (SSPs), secreted proteins (SPs) and proteins unique to *M. lychnidis-dioicae*. We extracted the list of these proteins that contributed to core

enrichment in the analysis and checked whether they are generally up-regulated in all stages or used in discrete stages. **Table 5-4** illustrates the number of common genes over the number of unique genes used in each category. For secreted proteins and unique proteins, there appeared to be a large group used exclusively in the Late stage and these numbers are found in brackets. This may suggest that the formation of teliospores involve more secreted proteins and unique proteins than other stages. It should be noted in **Table 5-1** that there are not as many gene sets pertaining to known domains found in the late stage. This is significant since we saw in **Fig. 5-1A** and **5-3A** that there were many genes that were up-regulated in the late stage and yet not many gene sets were identified.

Table 5-4. Number of common proteins in each core enriched gene set up-regulated in all plant stages when compared to Mated

Types of protein	Male	Female	Combined
Small secreted protein	15/32	17/30	17/33
Secreted protein	18/41 (5)	15/47 (10)	23/44 (7)
Unique proteins	74/184 (57)	63/145 (35)	77/167 (51)

In **Fig 5-7**, we compared the SSPs, SPs and unique proteins that contribute to the core enrichment in each sex at each stage. We found that almost all the SSPs (85.3% in common) and SPs (79.5% in common) were the same while the unique proteins in male plants tend to be more exclusive (22.4%). SPs found up-regulated in female only (20.4%) was also higher than in male (4%).

In order to understand whether the same genes were involved in mating and in teliosporogenesis, we compared the list of SSPs, SPs and unique proteins found up-regulated in Plants (all *in planta* stages), to those up-regulated in mated cells exposed to phytol (Pmated) for 12-48 hours and those up-regulated in Mated and Nutrient-limited (referred to as NL, haploid in conditions conducive for mating) when compared to p1A1 Rich (p1A1 strain on YPD medium) (see Chapter 3).

First of all, it appeared that the SSPs involved in Mated and NL were also used either in Pmated or in Plants. While the level of expression may be very different, it is interesting to note that all the SSPs required for subsequent development, whether in producing hyphae or for teliosporogenesis, were activated once the environment was conducive for mating. Deficiency in these genes may have profound impact on the ability of the fungal cells to complete their sexual life cycle.

There were a much larger arsenal of SPs (32.3%) and unique proteins (47.1%) used exclusively in Mated and NL than in Pmated or in the host plants. This reinforces the hypothesis that the mating system of *M. lychnidis-dioicae* is riddled with proteins

with unknown functions (see Chapter 3). Expression in host plants was more similar to Pmated than with Mated and NL in all three categories, suggesting that treatment with phytol was probably a close approximation of dikaryon extension in the host plants, rather than a different response mechanism.

There is a caveat in this analysis, in that the up-regulation of genes in plants and Pmated were determined by comparison to Mated while that of Mated and Water were determined by comparison to p1A1 Rich. This means that a gene expression in Plants or Pmated that is not different from Mated but was higher than p1A1 Rich will not be captured as an up-regulated gene in this analysis. However, since Mated was the logical step before Pmated and in plants, it was more sound to determine differential expression based on Mated rather than p1A1 Rich.

Gene set enrichment between stages

The main aim of breaking down the floral developmental stages was to capture the differential expression of the fungus when co-developing with the plants. Only by pair-wise comparison between the stages will we be able to uncover the differences. To

this end, we applied the GSEA Preranked analysis again and found significant enrichment of various gene sets among the down-regulated genes in MILate when compared to earlier floral stages as listed in **Table 5-5**.

First of all, we noticed that functions related to active cellular growth were enriched. This supports the idea that by late floral developmental stage, the pathogen is almost done with teliosporogenesis, just waiting to be dispersed when the anthers dehisce. It was also significant that this enrichment was more noted in the comparison between Stage 10 and late stage. This might indicate that the cellular activities were the strongest at Stage 10 and then they decrease at the late stage after the work is completed. The lack of enrichment in similar comparisons in the female may be accurate, or it may be due to the fact that there was only one FI10 sample; it does indicate that further analysis with additional replicates of this condition be needed.

The most significant gene set enrichments were SSPs, SPs, glycosyl hydrolase family 28, oligopeptide transporter family and heat shock protein 20. It was previously reported that *M. lychnidis-dioicae* is enriched in the number of glycosyl hydrolase genes ((Perlin et al., 2015); see Chapter 2). It is therefore interesting, that through this analysis presented in this chapter, glycosyl hydrolase family 28 may be the single most important hydrolase family in the teliosporogenesis of the fungus when interacting with its host.

An enrichment of SSPs and SPs in the down-regulated genes of late stage development also suggests that they were more important in the earlier stages of co-development. There was almost no other enrichment found in the comparison of other stages. In the original literature that found co-development of the fungus with its hosts, it was observed that when the teliospores start appearing at Stage 9, hyphae could still be

observed (Uchida et al., 2003). Hence the change may be subtle and gradual, which may require us to decrease the stringency of calling differential expression (*i.e.* a higher FDR). Such adjustment was used in a study of rice and blast fungus (Kawahara et al., 2012), where a FDR of 10^{-3} was used initially. In order for the pathogenesis-related protein, PR1 to be detected as differentially expressed, a FDR of 10^{-2} was used instead. We also take into account the fact that fungal cells represent a very small portion of the total tissue in the early stages of development, compared to the amount of plant tissue. Hence, more biological replicates may be necessary for accurate detection of gene expression changes.

A recent study found that the smut fungi infection affects the host in a sex-specific manner and that defense-related genes were surprisingly over-represented only in infected males but not in the female (Zemp et al., 2015). The authors do caution against interpretation of the GO (gene ontology) enrichment, since defense-related genes tend to evolve quickly with the pathogen, and thus may not be detectable using means of established databases and signature sequences. The fact that only 26.63% for non sex-linked to 57.30% of sex-linked contigs were annotated by the GO terms attest to this caveat of the analysis.

Table 5-5. List of gene sets that were enriched among genes that were down-regulated in late stage of floral development compared to earlier plant stages.

Common enrichment in more than one stage across more than one set of comparison			
<ul style="list-style-type: none"> • SMALL SECRETED PROTEIN • SECRETED PROTEIN • PF00295.10 GLYCO_HYDRO_28[GLYCOSYL HYDROLASES FAMILY 28] • PF03169.8 OPT[OPT OLIGOPEPTIDE TRANSPORTER PROTEIN] • PF00011.14 HSP20[HSP20/ALPHA CRYSTALLIN FAMILY] 			
MIFS vs. MILate	MI8 vs. MILate	MI9 vs. MILate	MI10 vs. MILate
<ul style="list-style-type: none"> • None 	<ul style="list-style-type: none"> • KO04141[PROTEIN PROCESSING IN ENDOPLASMIC RETICULUM]# • PF01095.12 PECTI 	<ul style="list-style-type: none"> • None 	<ul style="list-style-type: none"> • GO:0003677[DNA BINDING>MF]# • GO:0005737[CYTOPLASM >CC]# • PF07724.7 AAA_2[AAA DOMAIN (CDC48

	NESTERASE[PEC TINESTERASE]#		SUBFAMILY)]# <ul style="list-style-type: none"> PF10431.2 CLPB_D2-SMALL[C-TERMINAL, D2-SMALL DOMAIN, OF CLPB PROTEIN]# PF00004.22 AAA[ATPASE FAMILY ASSOCIATED WITH VARIOUS CELLULAR ACTIVITIES (AAA)]# PF00158.19 SIGMA54_ACTIVAT[SIGMA-54 INTERACTION DOMAIN]
FIFS vs. FILate	FI7 vs. FILate	FI9 vs. FILate	FI10 vs. FILate
• None	• None	• None	• None
MIFIFS vs. MIFILate	MIFI8 vs. MIFILate	MIFI9 vs. MIFILate	MIFI10 vs. MIFILate
• None	<ul style="list-style-type: none"> KO04141[PROTEIN PROCESSING IN ENDOPLASMIC RETICULUM]# PF01095.12 PECTINESTERASE[PECTINESTERASE]*# 	<ul style="list-style-type: none"> PF01095.12 PECTINESTERASE[PECTINESTERASE]* GO:0005737[CYTOPLASM>CC]* PF10431.2 CLPB_D2-SMALL[C-TERMINAL, D2-SMALL DOMAIN, OF CLPB PROTEIN]* PF07724.7 AAA_2[AAA DOMAIN (CDC48 SUBFAMILY)]* PF00158.19 SIGMA54_ACTIVAT[SIGMA-54 INTERACTION DOMAIN] PF00004.22 AAA[ATPASE FAMILY ASSOCIATED WITH VARIOUS CELLULAR ACTIVITIES (AAA)]* 	<ul style="list-style-type: none"> GO:0003677[DNA BINDING>MF]# GO:0005737[CYTOPLASM>CC]*# GO:0006334[NUCLEOSOME ASSEMBLY>BP] GO:0000786[NUCLEOSOME>CC] PF00125.17 HISTONE[CORE HISTONE H2A/H2B/H3/H4] PF00158.19 SIGMA54_ACTIVAT[SIGMA-54 INTERACTION DOMAIN] PF07724.7 AAA_2[AAA DOMAIN (CDC48 SUBFAMILY)]*# PF00004.22 AAA[ATPASE FAMILY ASSOCIATED WITH VARIOUS CELLULAR ACTIVITIES (AAA)]*# PF10431.2 CLPB_D2-SMALL[C-TERMINAL, D2-SMALL DOMAIN, OF CLPB PROTEIN]*#

* Common enrichment in comparison with another stage in the same dataset

Common enrichment in the comparison in the same stage in a different dataset

Conclusions

This is the first time an explicit gene expression study has been conducted among the early stages of *S. latifolia* floral development when infected with *M. lychnidis-dioicae*. We found that a lot of the genes involved in teliosporogenesis in the host may be SSPs, SPs and proteins unique to *M. lychnidis-dioicae*, which unfortunately do not have functional domains that can be analyzed computationally. However, given that this phytopathogen system evolves closely with its hosts, to the extent of having a narrow host range (Le Gac et al., 2007a), it makes sense that the genes that are involved in the interaction with its host may be novel compared to other phytopathogen systems.

This broad analysis serves to provide a big picture of the gene sets enrichment and will spur further hypothesis testing using these data. The corresponding host data (Toh et al., In Preparation) will also reveal more of the interaction between the fungus and its pathogen and further comparison with uninfected host may illuminate more intricate actions of the pathogen on its hosts.

We also hope to make use of this large body of data to update and correct the reference genome annotations (Perlin et al., 2015), in combination with re-sequencing data on Illumina (Fontanillas et al., 2015) and single molecule real time sequencing platform (Badouin et al., 2015), to produce a more accurate reference genome as was done with *Fusarium graminearum* (Zhao et al., 2013).

Overall, we have learned that biological replicates are very important for both sensitivity and accuracy of detection of gene expression changes, especially when subtle changes are expected. However, the difficulty of obtaining sufficient tissue cannot be

ignored. Despite advanced bioinformatics techniques available, basic infection and tissue collection remains fundamental to the success of such dual RNA-Seq studies.

CHAPTER 6: RELIABLE TRANSFORMATION SYSTEM FOR *MICROBOTRYUM LYCHNIDIS-DIOICAE* INFORMED BY GENOME AND TRANSCRIPTOME PROJECT

Introduction

Microbotryum lychnidis-dioicae (= *M. violaceum sensu lato* isolated from *Silene latifolia*; formerly known as *Ustilago violacea*) is a member of a species complex, which has evolved within the Microbotryales, in the Pucciniomycotina phylum (Bauer et al., 2006; Begerow et al., 1997; Weiss et al., 2004); the organism causes anther smut of the Caryophyllaceae (pinks). Upon infection, the teliospores of the fungus replace the pollen in male flower anthers. In female counterparts, the gynoeceium is suppressed and pseudo-anthers develop to house the teliospores (Baker, 1947; Uchida et al., 2005; Uchida et al., 2003). As such, it is a unique sexually transmitted disease in plants.

M. lychnidis-dioicae was the first fungus in which heteromorphic fungal mating-type chromosomes were described (Hood, 2002) and it has served as the model system for investigating the evolution of sex chromosomes (Fontanillas et al., 2015; Hood et al., 2004; Hood et al., 2013; Petit et al., 2012; Whittle et al., 2015). Transposable elements (TEs) and repeat-induced point mutation (RIPs) are also rampant in the genome of *M. lychnidis-dioicae*, a phenomenon that was recently examined and confirmed for some species in other subphyla of Pucciniomycotina (Horns et al., 2012). *M. lychnidis-dioicae* has also been used as a model organism in a wide variety of research areas in non-agricultural environment, allowing a much-needed alternative for analyzing host-

pathogen interactions in a necessarily more diverse host environment (Alexander, 1989; Antonovics et al., 2002; Refrégier et al., 2008).

Historically, *M. lychnidis-dioicae* was found to have at least 14 chromosomes in the haploid cells based on the linkage group analyses (Baird & Garber, 1981; Day & Jones, 1969). A long literature series entitled “Genetics of *Ustilago violacea*” (1975 – 2000) published in *Botanical Gazette* (now the *International Journal of Plant Sciences*, University of Chicago Press) used classical genetics methods to observe the physiological effects of mutants and map these to linkage groups. In order to facilitate the study of the *M. lychnidis-dioicae* complex in the context of population structure and epidemiology, highly polymorphic satellite markers are being discovered and used (Bucheli et al., 1998; Giraud et al., 2002; Giraud et al., 2008b). The genomes of *M. lychnidis-dioicae* isolate p1A1 ((Perlin et al., 2015); see Chapter 2) and p1A2 (Badouin et al., 2015) have recently been sequenced. An abundance of RNA-sequencing data from *in vitro* fungal stages and distinct floral bud stages of infected and uninfected plant tissue are also available (see Chapter 3 and 5). Therefore, there is a huge motivation to develop an effective transformation system to allow potential genes of interest to be investigated at the molecular genetics level.

Several attempts at developing such a system have been made (Bej & Perlin, 1989a; Bej & Perlin, 1989b; Bej & Perlin, 1991; Perlin et al., 1990), mostly modeled after other transformation systems in tractable organisms like *Saccharomyces cerevisiae* and *Ustilago maydis*. However, these techniques were not reproducible by other researchers; even biolistic approaches proved ineffective (M. Perlin, personal communication). Thus, none of these approaches were pursued further. *Agrobacterium*-

mediated transformation has been used in plants since the 1980s (Caplan et al., 1983). The use of *Agrobacterium*-mediated transformation in fungi (Abbott et al., 2013; Bundock et al., 1995; Covert et al., 2001; Ianiri et al., 2011; Maruthachalam et al., 2011), from the very tractable *Saccharomyces cerevisiae* to various recalcitrant Pucciniomycotina yeasts, seemed to be worth pursuing in developing a usable, reproducible transformation system for *M. lychnidis-dioicae*.

In this study, we developed an *Agrobacterium*-mediated transformation technique that is reproducible for the transformation of *M. lychnidis-dioicae* with exogenous DNA. We are also able to introduce desired gene cassettes for over-expression of a transgene, making this a tractable and potentially robust transformation system.

Materials and Methods

Selection of promoter regions

Genes were chosen whose transcription was found to be highly up-regulated in haploid cells grown on YPD (“Rich”) or on 2% water agar (“Nutrient-Limited”) or in infected buds of male *S. latifolia*, late in infection (“MI Late”). To select the appropriate promoter region, 1000 bp upstream of the genes belonging to the categories “Up in Rich”, “Up in Nutrient-Limited” and “Up in MI late” group were analyzed. These categories were determined based on a set of RNA-Seq data previously reported ((Perlin et al., 2015); see Chapters 3).

Table 6-1. List of potential gene promoters used for MEME analysis

Up in “Rich”	Up in Water	Up in MI Late
MVLG_01449	MVLG_00999	MVLG_00574; MVLG_01061; MVLG_01286;
MVLG_02648	MVLG_01677	MVLG_01766; MVLG_02024; MVLG_02288;
MVLG_02944	MVLG_01930	MVLG_03543; MVLG_04479; MVLG_06764;
MVLG_03520	MVLG_04092	MVLG_00149; MVLG_00677; MVLG_00784;
MVLG_03620	MVLG_05396	MVLG_01045; MVLG_01050; MVLG_01460;
MVLG_04226	MVLG_05533	MVLG_02165; MVLG_02379; MVLG_02464;
MVLG_04567	MVLG_05570	MVLG_02622; MVLG_02818; MVLG_03398;
MVLG_05229	MVLG_05746	MVLG_03707; MVLG_04190; MVLG_04237;
MVLG_05589	MVLG_06232	MVLG_04746; MVLG_04885; MVLG_04949;
MVLG_05703	MVLG_06475	MVLG_05108; MVLG_05180; MVLG_05508;
MVLG_06952	MVLG_06512	MVLG_05940; MVLG_06179; MVLG_06180;
	MVLG_06949	MVLG_06353; MVLG_06466; MVLG_07060

The criteria for the group of predicted transcripts in “Up in Rich” had shown more than 2 log₂ fold increase compared to the “Nutrient-Limited” and “MI Late” condition. The criteria for the group of predicted transcript in “Up in Nutrient-Limited” had shown more than 6 log₂ fold increase compared to the “Rich” and “MI Late” condition. For the “Up in MI Late” group, the transcripts showed more than 30 log₂ fold increase. The promoter region sequences were analyzed using the motif identification algorithm Multiple Em for Motif Elicitation (MEME) (Bailey et al., 2009). The list of gene

promoter used for each category is listed in **Table 4-1** and the parameters used in MEME is listed in **Table 4-2**.

Table 6-2. Parameters used for the selection of motifs on MEME

Sequence file	sequences
Distribution of motif occurrences	Any number of repetitions
Number of different motifs	10
Minimum number of sites	2
Maximum number of sites	300
Minimum motif width	2
Maximum motif width	300

The upstream putative promoter region that contained the highest number of one or more motifs in the group was chosen as the promoter for the group. The region upstream of MVLG_05589 was chosen as the “Rich” promoter, MVLG_06949 as the “Nutrient-Limited” promoter and MVLG_01766 as the “MI Late” promoter.

Primer Design

Primers were designed using web interface Primer 3 Plus (Untergasser et al., 2007) to amplify at least 800 bp of the selected upstream region chosen as a putative promoter. The source of green fluorescent protein (GFP) gene was plasmid pMF-5n (Becht et al., 2006). The *eGFP* gene used in *Ustilago maydis* is found within positions 303–1014 (Feldbrügge, 2010) of the plasmid sequence. The gene was not codon optimized for *M. lychnidis-dioicae*. The hygromycin cassette (*HYG*) was amplified from pPZP-HYG2 (Walton et al., 2005). In order to link the promoter to the *eGFP* or the *HYG* gene, an overlap PCR technique (Heckman & Pease, 2007) was employed.

Each primer set was checked for primer dimer and non-specific binding. The combined product (promoter with gene-of-interest) was also analyzed with NEBcutter (Vincze, 2003) to design restriction enzyme sites that would be suitable for cloning using

TOPO[®] TA cloning and subsequent cloning into the pPZP-201BK (Covert et al., 2001) or pPZP-MV-HYG binary plasmid. *PacI* sites and *EcoRV* sites were added as overhangs to the primers to facilitate these subsequent cloning processes into pPZP-MV-HYG plasmid.

Overlap PCR

Table 6-3. Primers used in overlap PCR

Primer	Sequence	Product size
MVLG_05589F	GCTGTATAGGGGGTTCGTGA	874 bp + 20 bp “overhang” ^a
HYGRC_MVLG05589R	TTCATAGACATGTTGGGCGA AGAGAGGATGCGATCGAGTG	
MVLG05589RC_HYGF	CACTCGATCGCATCCTCTCTCGCCCAACATGTCTATGAA	1,413 bp +20 bp overhang
HYGR	AACGTCGTGACTGGGAAAAC	
PacI_MVLG05589F	TTAATTAA GCTGTATAGGGGGTTCGTGA	874 bp + 8 bp + 20 bp overhang
GFPRC_MVLG05589R	TTTTGTTCCTTTAGTGAGGG AGAGAGGATGCGATCGAGTG	
PacI_MVLG06949F	TTAATTAA CGAGACCGGAGAGAGAGAGA	826 bp + 8 bp + 20 bp overhang
GFPRC_MVLG06949R	TTTTGTTCCTTTAGTGAGGG CAAGAGCGACATGAGAACGA	
PacI_MVLG01766F	TTAATTAA CCGTTTCGTATCCTCCTGAA	817 bp + 8 bp + 20 bp overhang
GFPRC_MVLG01766R	TTTTGTTCCTTTAGTGAGGG GAGGGCAGGTCAAACAAAAG	
MVLG05889RC_GFPF	CACTCGATCGCATCCTCTCT CCCTCACTAAAGGGAACAAAA	840 bp + 6 bp + 20 bp overhang
MVLG0649RC_GFPF	TCGTTCTCATGTCGCTCTTG CCCTCACTAAAGGGAACAAAA	
MVLG01766RC_GFPF	CTTTTGTGTTGACCTGCCCTC CCCTCACTAAAGGGAACAAAA	
EcoRV_GFPR	GATATC GATCTTCTTGAAGCGGTTGG	
GFPR	GATCTTCTTGAACGGTTGG	

^a Here “overhang” means that the amplicon will carry an additional sequence at either its 5’ or 3’ end that represents a portion of the gene for which overlap is desired to generate a recombinant construct.

Standard PCR was performed to amplify the region of interest using TaKaRa Hot Start *Ex Taq* DNA polymerase (cat no. RR001A, Clontech Laboratories, Inc) based on manufacturer's instructions with 1.0 μ M of primers/ overlap primers (**Table 4-3**) and 1 μ l of template (typically $\sim$ 10 ng DNA). The conditions for initial amplification of the gene product was 4 min at 94 °C for initial denaturation, followed by 34 cycles of 10 sec at 94 °C, 30 sec at 60 °C and 1 min at 72 °C. The final extension time was 5 min and product was held at 4 °C until use. For each promoter, amplification was attempted for both *M. lychnidis-dioicae* p1A1 and p1A2 strains, to determine whether there might be differences in the promoter region of the two strains. The primers were designed based on sequences from the p1A1 strain ((Perlin et al., 2015), See Chapter 2).

Each PCR reaction was visualized via gel electrophoresis through 0.8% agarose gels (containing Tris-acetate-EDTA (TAE)) stained with ethidium bromide. Hi-Lo™ DNA marker (Minnesota Molecular) was used as the DNA size standard. Bands that were of correct size were excised and purified using the GENECLAN® II kit (MP Biochemicals) protocol. Briefly, the excised gel was resuspended in 800 μ l of 6M pH 7.1 sodium iodide (NaI) in a sterile 1.7-ml microcentrifuge tube and incubated in a 50 °C water bath until all the gel has melted. Next, 20 μ l of GLASSMILK™ was added and vortexed to mix. The tube was then set horizontally on a shaker to mix for 10 min. The suspension was centrifuged at $14,000 \times g$ for 1 min in a microcentrifuge and the supernatant was discarded. The pellet was washed with 500 μ l of NEW-Wash solution twice by centrifugation and decanting the supernatant. The pellet was completely dried and resuspended in 23 μ l of sterile Tris-EDTA (TE) pH 8.0. The purified amplicon was

quantified on a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific). Finally, 65 ng of each amplicon was used for the subsequent overlap PCR protocol.

The amplicons from promoters of interest and genes-of-interest served as the templates for the overlap PCR and the primers were excluded in the initial reaction mix. The rest of the components were added based on manufacturer's instruction. Initial priming of the overlap region by the two amplicons allowed linear amplification of the fusion amplicon; this was accomplished with an initial denaturation set at 98 °C for 4 min, followed by 10 cycles of 94 °C for 30 s, 61 °C for 1 min and 72 °C for 2 min 30 s. The product was held at 4 °C until 1.0 µM final concentration of the end primers was added. Standard PCR cycles were then performed to amplify copies of the fusion amplicon exponentially, using initial denaturation of 94 °C for 4 min, and 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 2 min, and final extension time of 5 min at 72 °C. The product was held at 4 °C until it was visualized by standard agarose gel electrophoresis, excised and purified as described above. The resulting amplicon was cloned into pCR[®]2.1-TOPO[®] TA vector (Invitrogen) according to manufacturer's instruction. Transformants were selected with 100 µg/ml ampicillin on LB (1% tryptone, 0.5% yeast extract, 1% sodium chloride, 2% agar) media (LB/AP) and potential clones differentiated from self-ligated vector with 40 µl of 40 mg/ml X-gal (Gold Biotechnology) spread onto the media before spreading 120 µl of the transformation mixture.

Resulting white colonies were picked and grown on LB/AP plates. Plasmids were isolated from potential clones by using an alkaline lysis procedure (Sambrook & Russell, 2001).

Cloning into pPZP binary vector

Plasmids from cloning into pCR[®]2.1-TOPO[®] TA were compared to plasmids where no insert was added by examination after electrophoresis through agarose gels. Plasmids that appeared to be of the correct size (*i.e.*, larger than vector alone) were cut with restriction enzymes (New England Biolabs) that were selected based on analysis with NEBcutter (Vincze, 2003). For insertion of MVLG_05589-HYG into pPZP-201BK vector, *HindIII* was used to cut the vector and the insert. For insertion of promoter-GFP inserts into pPZP-MV-HYG vector, *PacI* and *EcoRV* were used.

Each product was visualized on 0.8% gel electrophoresis with Hi-Lo[™] DNA marker as the size standard. The bands corresponding to the fusion amplicon size were excised and purified from the gel as described above.

Ligation to the pPZP-201BK plasmid backbone bearing MVLG05589-HYG insert

The vector used for these experiments (pPZP-MV-HYG) was derived via cloning of the MVLG_05589-HYG insert in the pPZP-201BK plasmid backbone (conducted by D. Treves). Ligation of pPZP-MV-HYG and the fusion amplicon was performed using Ligate-IT Rapid Ligation kit (Affymetrix). All resulting colonies were picked and grown on LB/KM (LB with 50 µg/ml of kanamycin) before screening. Screening of the plasmid by size was performed as described above. Plasmids that appeared to be larger than the self-ligated vector were cut with *HindIII*, which would release the insert containing the hygromycin resistance cassette and by size estimation, determine if the insert was in the plasmid backbone.

For confirmation of insertion of the proper construct, some of the potential plasmids were sequenced using *eGFP* primers (GFPSeqout5:

5'TGGCCGCGTTGGCCTCGAG-3'; GFPSeqout3: 5'-

GCGATCACATGGTCCTGCTGG-3') that prime outwards from within the gene.

Sequencing reactions were prepared using the BigDye[®] Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) according to manufacturer's instruction and the sequencing products were analyzed using ABI PRISM[®] 3100.

Transformation of *Agrobacterium* competent cells by electroporation

Agrobacterium tumefaciens EHA105 competent cells were transformed with the isolated binary plasmids using the electroporation protocol (Mattanovich et al., 1989). Briefly, competent cells were prepared (McCormac et al., 1998) by growing the *Agrobacterium* in 5 ml of LB broth overnight at room temperature. Next, 3 ml of the overnight culture was inoculated into 300 ml of prewarmed LB broth and grown to A₆₀₀ of 0.5 at 29 °C. The culture was cooled on ice for 15 min and centrifuged in 2 pre-chilled centrifuge tubes at 4,500 rpm at 4 °C. The cells were resuspended in 10 ml of sterile water and transferred to a 35-ml centrifuge tube. The cells were washed twice in sterile water by centrifugation at 4,500 rpm at 4 °C for 15 min in a Sorvall refrigerated centrifuge and decanting the supernatant. The cells were then washed once with 10 % glycerol and finally resuspended in 3 ml of 10% glycerol. The competent cells were then aliquoted into pre-chilled microcentrifuge tubes with 200 µl each. The aliquots were frozen in liquid nitrogen and stored at -80 °C until use.

The sterile cuvettes with 0.2 cm electrode distance (Bio-Rad) and sterile tips to be used in the experiment were pre-chilled at -20 °C for at least one hour. Thawed competent cells (40 µl) were kept on ice and aliquoted into a pre-chilled cuvette. Then, 5 µl of plasmid miniprep was added to the cells and gently mixed with the cells so that

mixture wound up at the bottom of the cuvette. All sides of the cuvette were wiped dry with Kimwipe before loading onto the Gene Pulser (Bio-Rad Laboratories). The electric pulse was set to initial voltage of 2.5 kV using the 25 μ F capacitor and 400 ohm setting on the pulse controller. An electric pulse of up to 10 sec was applied to the cuvette. Then, 0.8 ml of S.O.C medium (Invitrogen) was added to the cuvette immediately and the content was transferred to a sterile 2-ml microcentrifuge tube. The transformation reaction was incubated for at least 3 hrs at 25 °C and 120 μ l of the reaction was spread-plated onto LB/KM (LB with 50 μ g/ml kanamycin) plates. Colonies were counted and 2 colonies were picked from each transformation reaction after incubation at 25 °C for 48-72 hrs.

Plasmid was isolated from potential clones and screened by standard PCR using the primers listed in **Table 6-4**. The candidate targets for PCR screening were the “Rich” promoter that was linked to the *HYG* resistance cassette, the *eGFP* gene and the *HYG* resistance gene. One clone of each type of transformed *A. tumefaciens* strain was chosen for the transformation of *M. lychnidis-dioicae*.

Table 6-4. Primers used for verification of *A. tumefaciens* transformants

Primer	Sequence	Product size
PacI_MVLG5589F GFPRC_MVLG05589R	TTAATTAA GCTGTATAGGGGGTTCGTGA TTTTGTTCCCTTTAGTGAGGG AGAGAGG ATGCGATCGAGTG	902 bp
MVLG05889RC_GFPPF GFPR	CACTCGATCGCATCCTCTCT CCCTCACT AAAGGGAACAAAA GATCTTCTTGAACGGTTGG	860 bp
HYGF HYGR	AAAAGCCTGAACTCACCGC TCGTCCATCACAGTTTGCC	498 bp

Transformation of *M. lychnidis-dioicae*

M. lychnidis-dioicae p1A1 and p1A2 strains were grown for four days on yeast peptone dextrose solid media (YPD plate; 1% yeast extract, 2% peptone, 10% dextrose, 2% agar) for four days at 25°C. *Agrobacterium* strains were grown overnight on LB/KM plates. A 10-µl loopful of each culture was harvested and resuspended in 1 ml of induction medium broth (IM broth, recipe in Appendix 4) (Bundock et al., 1995; Hooykaas et al., 1979). The absorbance at 600 nm (A_{600}) was determined for a 100-fold dilution of the suspension. Cell density of the suspension was adjusted to 10^8 cells/ml in IM broth, based on *M. lychnidis-dioicae* $OD_{600} 1 = 3.4 \times 10^7$ cells/ml and *Agrobacterium* $A_{600} 1 = 8 \times 10^8$ cells/ml. A total volume of 200 µl of *M. lychnidis-dioicae* suspension and *Agrobacterium* suspension were mixed based on desired combination, by dropping on the induction medium plate (IM plate, recipe in Appendix 3). Controls without the *M. lychnidis-dioicae* host or the *Agrobacterium* were included in each experiment. The plates were left covered facing up without parafilm at room temperature for 2-3 days.

Each mass of cell mixture was harvested off the plates and resuspended in 1.8 ml of yeast peptone dextrose broth (YPD broth; 1% yeast extract, 2% peptone, 10% dextrose). A 120 µl aliquot of each suspension was spread-plated onto YPD containing 150 µg/ml hygromycin (Life Technologies) and 300 µg/ml cefoxatime or cefoxitin (Amresco) (YPD+HYG/CEF) plates. The plates were incubated at 25 °C and checked daily for contamination and growth. True colonies should start appearing at about 8 days and new colonies may continue to arise. New colonies were picked every 2 days and transferred onto YPD+HYG/CEF plates.

Genomic verification of *M. lychnidis-dioicae* transformants

Colony PCR ((Brachmann et al., 2004) (**Table 6-5**), was performed on all potential transformants by picking a small amount of cells and resuspending in 100 µl of sterile water. From this, 1 µl of the suspension was used in a standard PCR reaction. The initial denaturation step was changed to 96 °C for 10 min, and the rest of the cycles were performed as described above. The PCR products were visualized after separation by electrophoresis in 0.8% agarose gels. Putative transformants were screened for the presence of bands of the same size as the control, produced using as template pPZP-MV-HYG-derived plasmid containing the genes-of-interest or the housekeeping gene *mepA*, to confirm the identity of the cells as *M. lychnidis-dioicae*.

Table 6-5. Primers used for verification of *M. lychnidis-dioicae* transformants.

Primer	Sequence	gDNA pdt	cDNA pdt
mepA 5'(3) mepA 3'(2)	CTTTTGCGTAGGAAGAATGC ACGGTGCCGAGGATGATTGGA	1039 bp	532 bp
GFPRC_MVLG05589 R GFPR	TTTTGTTCCCTTTAGTGAGGG A GAGAGGATGCGATCGAGTG GATCTTCTTGAACGGTTGG	860 bp	860 bp
HYGF HYGR	AAAAGCCTGAACCTACCGC TCGTCCATCACAGTTTGCC	498 bp	498 bp

Transcriptional verification of *M. lychnidis-dioicae* transformants

In order to verify that transgenes were expressed in putative transformants, two of every type of transformant with a different promoter, were subjected to the “Rich” and “Nutrient-Limited” condition. As additional controls, wild type untransformed haploid strains of each mating-type and strains confirmed to have been transformed with the MV-HYG vector alone were treated similarly to the putative transformants bearing the GFP construct. For the “Rich” condition, the strains were grown on YPD for 4-5 days at room temperature and RNA was extracted. For the “Nutrient-Limited” condition, the same

isolates were resuspended in water after growing for 4-5 days on YPD at room temperature, and adjusted to 10^9 cells/ml by A_{600} (absorbance at 600 nm) determination. Then, 800 μ l of the suspension was dispensed to cover the entire agar surface of a “Nutrient-Limited” plate (composition: 2% Bacto agar) and allowed to dry. Two plates for each transformant were prepared and incubated at 14 °C for about 48 hours before RNA was extracted.

RNA was extracted using the Direct-zol™ RNA MiniPrep kit (Zymo Research) and TRIzol® (Invitrogen). The cells were harvested from the plates into pre-chilled mortar and ground into powder using liquid nitrogen. About 100 mg of tissue was transferred into a pre-cooled 2-ml microcentrifuge tube and 500 μ l of TRIzol® was added before the ground tissue thawed. The tube was set on a vortex and mixed for about 5 min until the solution was clear. The rest of the extraction was done based on manufacturer’s instruction, including the on-column DNase I (Zymo Research, cat no. E1009) treatment.

Samples were quantified and checked for purity based on $A_{260/280}$ and $A_{260/230}$ ratio using the NanoDrop spectrophotometer and 2.5 μ g of each sample was subjected to RT-PCR using SuperScript® III Reverse Transcriptase (Invitrogen) according to manufacturer’s instruction. The same primers for genomic verification were used in the transcriptional verification. The cDNA samples were verified to be free of genomic DNA (gDNA) contamination by the absence of a gDNA amplicon with the *mepA* primer set. PCR products were visualized after electrophoresis through 2% agarose gels and some bands were selected for sequencing using BigDye® Terminator v3.1 Cycle Sequencing kit

Northern Blot

As terminal PCR was performed on the RNA, it was difficult to determine if the promoter varied in strength in the different media. The PCR results also suggested that some transcripts may have been modified. Hence, a Northern blot was performed to determine whether the different promoters affected the length and expression levels of the transcript. Northern blot was performed in accordance to the “DIG Application Manual for filter hybridization” from Roche Laboratories.

The RNA samples were first analyzed with Agilent 2100 Bioanalyzer (Agilent Technologies) for integrity and accurate quantification before use in the Northern blot assay. Five μg of total RNA (based on Bioanalyzer analysis) was separated on a 1.5% denaturing gel (composition: 1.5% agarose, 1x MOPS buffer and 2% formaldehyde). For loading purposes, each RNA sample was volume adjusted to 10 μl (in RNase-free water). Ten μl of 2x loading buffer (New England Biolabs, cat no. B0363S) was then added and mixed into each sample. The RNA samples were denatured at 65 °C for 10 min and placed in an ice bath immediately for at least 1 min. Two μl of ssRNA ladder (New England Biolabs, cat no. N0362S) was also treated based on manufacturer’s instruction. Briefly, 2 μl of the ladder was mixed with 8 μl of the loading buffer and denatured at 70°C for 10 min and chilled in the ice bath for 1-2 min before being loaded onto the gel.

The denaturing gel was electrophoresed overnight at about 1.5 V/cm in 1x MOPS buffer (20 mM MOPS, 5 mM pH 5.2 sodium acetate, 2 mM EDTA). The length of run was based on the gel front of the loading dye. The gel was stained with 0.5 $\mu\text{g/ml}$ of ethidium bromide for 30 min and then visualized with UV transilluminator to ensure that the rRNAs were visible and to assess whether the electrophoresis process had proceeded

for a sufficient length of time. Unfortunately, the RNA ladder did not appear with the ethidium bromide stain on any of the gels.

Each gel was then cut to fit the Biobond-Plus™ nylon transfer membrane (Sigma-Aldrich). The gel was soaked twice for 15 min in 20x SSC (composition: 3 M sodium chloride, 0.3 M sodium citrate) to remove formaldehyde that can inhibit the transfer. Capillary transfer was set up to blot the RNA from the gel onto the membrane overnight using 20x SSC as the transfer buffer. Upon completion of the blot transfer, the membrane was placed on Whatman 3MM paper that had been soaked in 2x SSC and cross-linked with the GS Gene Linker® UV chamber (Bio-Rad) at 125 mJ. The membrane was then washed briefly in sterile DEPC-treated water and allowed to air dry. The dried membrane was stored between two sheets of dry Whatman 3MM paper and wrapped in plastic wrap until probe hybridization was performed. The blotted gel was stained with ethidium bromide again to check that amount of rRNA has reduced significantly from before the blot, as an indication for the successful transfer of RNA onto the membrane.

eGFP probe was prepared with the PCR DIG Probe Synthesis kit (Roche, cat no. 11 636 090 910) according to manufacturer's instruction using 10x dilution of an *Agrobacterium* plasmid preparation that was used as the positive control in the verification assay. Five µl of the labeled probe (from 50 µl reaction) was analyzed by electrophoresis through a 2% agarose gel to ensure that the amplicon was successfully labeled according to manufacturer's guidelines. The amplicons were synthesized using the same verification primers and PCR cycling conditions as described above.

Prehybridization and hybridization were performed at 50°C in the Robbin Scientific hybridization oven, set to rotate at a setting of "7". Two membranes were pre-

hybridized in each hybridization bottle using 30 ml of pre-warmed high SDS hybridization buffer (composition: 50% deionized formamide, 5x SSC, 50 mM sodium phosphate (pH 7.0), 2% blocking reagent, 0.1% laurylsarcosine, 7% SDS) per membrane. Then, 45 µl of DIG-labeled probes were mixed with 50 µl of DEPC treated water, heated in the thermocycler at 100°C for 5 min to denature the probe and promptly added to 30 ml of pre-warmed hybridization buffer for each hybridization tube. Hybridization with the *eGFP* probe was performed overnight, followed by two washes with the Low Stringency Wash buffer (composition: 2x SSC, 0.1% SDS) at room temperature and two washes with the High Stringency Wash buffer (composition: 0.1x SSC, 0.1% SDS) at 50°C, both with shaking on an orbital shaker.

Detection was performed using NBT/ BCIP chromogenic alkaline phosphatase substrate (Roche, cat no. 11 681 451 001) according to manufacturer's instruction. All incubations and washings were performed at room temperature with shaking unless otherwise stated. Each membrane was washed in the Washing buffer (composition: 0.1 M maleic acid buffer pH 7.5, 150 mM NaCl, 0.3% Tween 20) for 2 min, followed by incubation in the Blocking solution (composition: 1% w/v blocking solution (Roche) in maleic acid buffer pH 7.5) for 30 min (up to 3 hr). Anti-Digoxigenin-alkaline phosphatase (Roche, cat no. 11 093 274 910) was centrifuged for 5 min and diluted 1: 10 000 in Blocking solution to make the Antibody solution. The membranes were incubated with the Antibody solution for 30 min. Each membrane was then washed twice in Washing buffer for 15 min each, followed by equilibrating in Detection buffer (composition: 0.1M Tris-HCl pH 9.5, 0.1M NaCl) for 3 min without shaking. Afterwards, 200 µl of NBT/BCIP stock solution was diluted in each 10 ml of Detection buffer

required to make the color substrate solution. Each membrane was covered in the solution and incubated overnight without shaking at room temperature in the dark.

Once the desired intensity of bands was obtained, the reaction was stopped by rinsing in TE buffer (composition: 10 mM Tris-HCl, 1 mM EDTA pH 8.0) for 5 min. An image of each membrane was documented with a Canon PowerShot SX120 digital camera. The membranes were then stripped immediately, first by decolorization with dimethylformamide heated to 50-60 °C and washing the membrane until the color had been removed. Each membrane was then rinsed thoroughly in DEPC-treated water. The probes were then stripped by incubation in Stripping buffer (composition: 0.2 M NaOH containing 0.1% SDS) twice for 20 min each at 37 °C. Each membrane was then rinsed in 2x SSC, followed by a rinse in DEPC-treated water and then allowed to dry. The membranes were then stored for future use at 4 °C between dry Whatman 3MM paper.

Results

Selection of promoter regions

Eleven promoter regions of genes that were upregulated in the “Rich” condition were analyzed to discover a motif that may represent a commonly used recognition site by the transcription factors under growth in this condition. The putative promoter region of MVLG_05589 contains the motif shown in **Fig. 6-1**. The motif was found 16 times among the eleven sequences analyzed and occurred 7 times in the region upstream of MVLG_05589 that may be considered as its promoter.

Figure 6-1. Most common motif found upstream of genes upregulated in "Rich" condition.

Twelve putative promoter regions of genes that were upregulated in the “Nutrient-Limited” condition were analyzed. The motif identified in **Fig. 6-2** was found 24 times among the sequences analyzed and occurred 5 times in the putative MVLG_06949 promoter.

Figure 6-2. Most common motif found among regions upstream of genes upregulated in the "Nutrient-Limited" condition.

We analyzed 36 putative promoter regions of genes that were upregulated in the “MI Late” set of genes. The results of some of these analyses are shown in **Fig. 6-3**. Motif A was found 29 times, Motif B 33 times and Motif C 21 times among the 36 sequences analyzed. In the MVLG_01766 putative promoter, both A and C occurred 3 times, whereas Motif B occurred once.

Figure 6-3. Most common motifs found among regions upstream of genes upregulated in "MI Late".

Retrieval of insert from TOPO clones

Digestion with *PacI* and *EcoRV* of clones containing the desired fusion amplicons cloned into pCR®2.1-TOPO® TA vector allowed recovery of the respective amplicons for the following promoters: MVLG05589 from p1A1 (A1: *PacI*-MVLG05589p1A1-GFP), MVLG05589 from p1A2 (A2: *PacI*-MVLG05589p1A2-GFP-*EcoRV*), MVLG06949 from p1A1 (B1: *PacI*-MVLG06949p1A1-GFP-*EcoRV*), MVLG06949 from p1A2 (B2: *PacI*-MVLG06949p1A2-GFP-*EcoRV*) and MVLG01766 from p1A1 (C1: *PacI*-MVLG01766p1A1-GFP-*EcoRV*). Despite repeated attempts to produce a fusion amplicon for MVLG01766 from p1A2, this was not successful.

Cloning into pPZP-MV-HYG plasmid

After ligation, clones that appeared to be larger than pPZP-MV-HYG were digested using the restriction enzyme, *HindIII*, to determine whether the pPZP-MV-HYG plasmid contained the fusion amplicon insert. The 2.3 kb *HYG* cassette was released and the size of the remaining plasmid was used to verify that the 1.7 kb insert has been ligated into the binary plasmid, such as illustrated in **Fig. 6-4**.

Sequencing of selected clones also verified that the *eGFP* was linked to the desired promoter on one end and the binary plasmid on the other (not shown).

Electroporation of *A. tumefaciens* EHA105

Electroporation of the *A. tumefaciens* competent cells yielded many colonies (Table 6-6). Only 10 colonies of each promoter type were picked and two from each type (Table 6-7) were further analyzed to verify transformation and presence of the desired pPZP-MV-HYG-fusion amplicon plasmid.

Table 6-6. Results of electroporation of *A. tumefaciens* EHA105

Promoter type	Fusion amplicon	Number of colonies yielded	Comments
A1	MVLG05589p1A1-GFP_1	9	Arced twice during electroporation
A1	MVLG05589p1A1-GFP_2	94	Arced twice during electroporation

A1	MVLG05589p1A1-GFP_3	102	
A2	MVLG05589p1A2-GFP_8-8	0	Probably insufficient plasmid and arced 5 times during pulse
A2	MVLG05589p1A2-GFP_8-21	134, 225	
B1	MVLG06949p1A1-GFP_4-7	59, 21	
B2	MVLG06949p1A2-GFP_4-17	68,106	
C1	MVLG01766p1A1-GFP_4-19	332, 363	

Table 6-7. Verification of presence of inserts in pPZP vectors and *A tumefaciens* transformants

* denotes the clones chosen for *M. lychnidis-dioicae* transformation.

Promoter type_Clone ID/ Plasmid	PacI-MVLG05589-GFPF	MVLG05589RC-GFP	HYG
A1_9-1	-	+	+
A1_9-2*	+	+	+
A2_9-11*	+	+	+
A2_9-12	+	+	+
B1_9-21*	+	+	+
B1_9-22	+	+	+
B2_10-6*	+	+	+
B2_10-7	Weak +	+	+
C1_10-16*	+	+	+
C1_10-17*	+	+	+
MVLG05589p1A1-GFP_1	+	+	+
MVLG05589p1A2-GFP_8-21	+	+	+
MVLG06949p1A1-GFP_4-7	+	+	+
MVLG06949p1A2-GFP_4-17	+	+	+
MVLG01766p1A1-GFP_4-19	+	+	+
pPZP-MV-HYG	+	-	+
PCR H ₂ O ctrl	-	-	-

* denotes the clones chosen for *M. lychnidis-dioicae* transformation.

Transformation of *M. lychnidis-dioicae* by *A. tumefaciens*

For each strain of *M. lychnidis-dioicae* and *A. tumefaciens* used, controls without a partner were included. In experiments 1 and 2 pinpoint colonies were observed in control plates with only *M. lychnidis-dioicae*. In experiment 3, 4 and 5, none of the controls yielded any colonies. Larger colonies appeared on transformation plates and these were tested to verify bona fide transformation with each desired construct. Putative transformant colonies of *M. lychnidis-dioicae* strain p1A1 with the MV-HYG vector appeared at a rate ranging from 2 to 145 (**Table 6-8**) per plate, depending on the day of the experiment and the lab in which it was conducted (*i.e.*, transformations with the pPZP-MV-HYG vector were carried out by S.S. Toh in the Perlin lab and separately, by D. Treves in both the Perlin lab and in his own lab). Colonies of *M. lychnidis-dioicae* strain p1A2 with the pPZP-MV-HYG vector appeared at a rate ranging from 13 to 308 (**Table 6-8**) per plate. For the constructs that included the *eGFP* driven by putative *M. lychnidis-dioicae* promoters yielded between 4 and 22 colonies per plate. C1_10-16 was verified to carry the pPZP-MV-HYG plasmid with the appropriate inserts but transformation with the same strain failed twice, while transformation with C1_10-17 was successful. Transformation of *M. lychnidis-dioicae* p1A2 strain with B2_10-6 also failed but was not repeated since the desired promoter is already available in the p1A1 transformants. The performance of the promoter in different *M. lychnidis-dioicae* host strains was not specifically examined and will not be addressed in this work.

Table 6-8. Yield of *M. lychnidis-dioicae* transformation by *A. tumefaciens*

Experiment number	Category	<i>M. lychnidis-dioicae</i> strain	<i>A. tumefaciens</i> strain	Number of transformants
1 ^a	MV-HYG-p1A1	p1A1	MV-HYG_19	132
	MV-HYG-p1A1	p1A1	MV-HYG_3-1	110
	MV-HYG-p1A1	p1A1 (10x less)	MV-HYG_19	56
	MV-HYG-p1A1	p1A1 (10x less)	MV-HYG_3-1	13
2 ^a	MV-HYG-p1A1	p1A1	MV-HYG_19	142
	MV-HYG-p1A1	p1A1	MV-HYG_3-1	105
3 ^a	MV-HYG-p1A1	p1A1	MV-HYG_19	145
	MV-HYG-p1A2	p1A2	MV-HYG_19	308
4	MV-HYG-p1A1	p1A1	MV-HYG	2
	MV-HYG-p1A2	p1A2	MV-HYG	13
	A1-p1A1	p1A1	A1_9-2	22
	A1-p1A2	p1A2	A1_9-2	7
	A2-p1A1	p1A1	A2_9-11	9
	A2-p1A2	p1A2	A2_9-11	4
	B1-p1A1	p1A1	B1_9-21	4
	B1-p1A2	p1A2	B1_9-21	4
	B2-p1A1	p1A1	B2_10-6	6
5	C1-p1A1	p1A1	C1_10-17	14
	C1-p1A2	p1A2	C1_10-17	8

^a Experiment performed independently by different investigators in our lab

Genomic verification of transformants

All the *M. lychnidis-dioicae* colonies that arose from the transformation experiment were screened by colony PCR and they were only considered true transformants if all three target PCR products were amplified. A total of 93 colonies were screened and 84 were verified to be true transformants. This translates to 91% of the

colonies being true transformants, as summarized in **Table 6-9**. Among those that failed the verification, some or all of the test amplicons failed to appear.

Table 6-9. Number of verified transformants at genomic level

Category	Number of transformants	Number of verified transformants
MV-HYG-p1A1	2	2
MV-HYG-p1A2	13	10
A1-p1A1	22	19
A1-p1A2	7	7
A2-p1A1	9	9
A2-p1A2	4	4
B1-p1A1	4	4
B1-p1A2	4	4
B2-p1A1	6	4
C1-p1A1	14	13
C1-p1A2	8	8

Transcriptional verification of transformants

A table of results for the PCR of the cDNA is summarized in **Table A3-4**. The *mepA* primers were used to verify that all the RNA samples were gDNA-free. PCR of the *eGFP* insert produced amplicons of slightly varying sizes. Interestingly, they were consistent across the promoter types. Amplicons from each promoter type were chosen for sequencing as indicated in **Fig. 6-5**.

Products from the labeled lanes were purified from the gel and sequenced. Panel A: All samples were grown in rich medium. Lanes 2 & 3 were untransformed strains; Lanes 3 & 4 were transformed with MV-HYG without *eGFP*; Lanes 6-13 were *eGFP* transformants with “Rich” (MVLG_05589) promoter; Lane 14-19 were *eGFP* transformants with “Nutrient-Limited” (MVLG_06949) promoter. Lane 6 (5R) and 18 (17R) were selected for sequencing; Panel B: Lanes 2-5 were *eGFP* transformants with “Up in MI Late” (MVLG_01766) promoter and was grown on rich medium. Lanes 6-19 were grown on nutrient-limited medium. Lanes 5 & 6 were untransformed strains; Lanes 7 & 8 were transformed with MV-HYG without *eGFP*; Lanes 9-16 were *eGFP* transformants with “Rich” (MVLG_05589) promoter; Lanes 17 & 18 were *eGFP* transformants with “Nutrient-Limited” (MVLG_06949) promoter. Lane 2 (19R) and 6 (1W) selected for sequencing; Panel C: Lanes 2-9 were grown in Nutrient-Limited medium. Lanes 2-5 were *eGFP* transformants with “Nutrient-Limited” (MVLG_06949) promoter; Lanes 6-9 were *eGFP* transformants with “Up in MI Late” (MVLG_01766) promoter. Lanes 10 was a negative control for reverse transcription using RNase-free water. Lane 11 was a PCR negative control. Lane 12 (8-21) was the positive control binary plasmid that carried MVLG05589p1A2-GFP fusion gene Lane 5 (18W) and Lane 12 (8-21) was selected for sequencing.). HL, Hi-Lo Size Standard.

The B1 and B2 promoters were chosen as representatives of genes up-regulated in the “Nutrient-Limited” condition. Isolates that were transformed with the B1 and B2 promoters produced two bands when grown and isolated from “Nutrient-Limited” conditions. The two amplicons from 18W were purified for sequencing. Based on

alignment on all sequences that were of acceptable quality, the central sequence of the GFP amplicon was consistent with the reference sequence. With the smaller amplicon, about 400 bp in the 3' end was consistent with the reference sequence but the 300 bp on the 5' end was not. This could also explain why sequencing from the 5' direction did not work well.

The larger amplicon appeared to have an additional tail of over 40 bp at the 3' end that consisted of repeats of part of the primer region (**Fig. 6-6**). This also caused a sequencing error from the 3' end.

CCG(CCAA[CCGCCTTCAAGAAGA]TC)|GA[CCGCTTCAAGAAGA](CCAACCGC
TTCAAGAAGATC)AGGAG

Figure 6-6. End sequence of larger *eGFP* amplicon from 18W cDNA.

| indicate theoretical end of sequence. [] and () mark the repeat motifs.

Additionally, 17R, which also was a result of B2 promoter transformation, seemed to have a stretch of 50 bp in the 5' end that may have caused non-specific binding to the sequencing primer too. We are unsure why the amplicons from B1 and B2 promoters are differing from the reference sequence in particular. The faint band in 1W also turned out to be largely in consensus with the reference. It is also difficult to explain, when the amplicon only showed up in “Nutrient-Limited” but not in “Rich” for all four non-*eGFP* transformants. We do not rule out contamination of the cDNA at this point. As such, we decided to pursue Northern Blots to obtain a clearer understanding of expression of the *eGFP* transcript.

For comparison, we used RT-PCR to examine expression of the *HYG* gene used as a selectable marker. Since such transformants had been selected on hygromycin plates and we did not observe hygromycin resistance otherwise for the *M. lychnidis-dioicae* strains used, this provided evidence that the hygromycin resistance gene was indeed being expressed at the protein level. Interestingly, when RT-PCR was conducted, the primers for amplification of the diagnostic *HYG* amplicon, produced two amplicons for all the cDNA samples, as seen in **Fig. 6-7**. There were products of ~500 bp and ~360 bp. The positive control from MVLG05589p1A2_8-21 produced an amplicon of ~500 bp. These were sequenced from both ends using the *HYG* primers. From the alignment (**Fig. 6-8**), it can be observed that a 141 bp region was missing from the smaller product. It is possible that *M. lychnidis-dioicae* processed the transcript with a mechanism yet unknown. BLASTn shows that it matches the hygromycin resistance cassette in two separated segments as well.

Figure 6-7. Electrophoresis of *HYG* amplicons from cDNA template through 2% agarose gel.

17R is the PCR product resulting from template of cDNA of *M. lychnidis-dioicae* p1A1 strain transformed with *A. tumefaciens* carrying MVLG06949p1A2-GFP fusion gene,

grown on “Rich” media. 6W is the PCR product resulting from template of cDNA of *M. lychnidis-dioicae* p1A1 strain transformed with *A. tumefaciens* carrying MVLG05589p1A1-GFP fusion gene, grown on “Nutrient-Limited” media. 8-21 was the positive control binary plasmid that carried MVLG05589p1A2-GFP fusion gene. The product of 17R and 8-21 were retrieved for sequencing.

MVLG05589p1A2-GFP_8-21_HYG_R	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA	150
17R_HYG_R_2	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGG--	146
17R_HYG_R_1	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA	148
HYG_seq	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA	145
17R_HYF_F_1	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA	113
17R_HYG_F_2	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGG--	110
MVLG05589p1A2-GFP_8-21_HYG_F	CTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA	115

MVLG05589p1A2-GFP_8-21_HYG_R	AATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTT	200
17R_HYG_R_2	-----	
17R_HYG_R_1	AATAGATGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTT	198
HYG_seq	AATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTT	195
17R_HYF_F_1	AATACCTGCGCCGATGGTTTCTACAAAGATCGGTATGTTTATCGGCACCTT	163
17R_HYG_F_2	-----	
MVLG05589p1A2-GFP_8-21_HYG_F	AATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTT	165
MVLG05589p1A2-GFP_8-21_HYG_R	TGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCA	250
17R_HYG_R_2	-----	
17R_HYG_R_1	TGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCA	248
HYG_seq	TGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCA	245
17R_HYF_F_1	TGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCA	213
17R_HYG_F_2	-----	
MVLG05589p1A2-GFP_8-21_HYG_F	TGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCA	215
MVLG05589p1A2-GFP_8-21_HYG_R	GCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAGTTG	300
17R_HYG_R_2	-----GTGTCACGTTG	157
17R_HYG_R_1	GCGAGAGCCTGACCTATTGCATCTCCCGCCGCGCACAGGGTGTACAGTTG	298
HYG_seq	GCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAGTTG	295
17R_HYF_F_1	GCGAGAGCCTGACCTATTGCCTCTCCCGCCGCGCACAGGGTGTACAGTTG	263
17R_HYG_F_2	-----GTGTCACGTTG	121
MVLG05589p1A2-GFP_8-21_HYG_F	GCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAGTTG	265

Figure 6-8. Part of the alignment of the *HYG* PCR amplicons with cDNA template (17R)

17R_HYG_F_1 and 17R_HYG_R_1 originated from the larger ~500 bp amplicon.

17R_HYG_F_2 and 17R_HYG_R_2 originated from the smaller ~360 bp amplicon.

MVLG05589p1A2-GFP_8-21_HYG_F and _R were sequences of the positive control amplicon and HYG-seq is the theoretical sequence.

Northern blot

Shown in **Fig. 6-9** are the results from Northern blots. Results should be interpreted with caution regarding amounts of total RNA loaded, as there was no internal loading control except for the quantification performed on the total RNA and ethidium bromide visualization of rRNA bands on the agarose gels used for blots, as well and the possibility that blot transfer efficiency may vary. Similarly, size determination was only possible crudely since the commercially-obtained size standard was not visible on the gels used for blotting. With these caveats, some observations and conclusions were still possible. On these Northern blots, RNA from negative control strains 1R through 4R and 1W through 4W were negative for *eGFP* transcripts. This is consistent with the fact that they were not transformed with the *eGFP* insert. Transformants 5-12 had a much darker band when strains were grown on “Rich” medium (5R through 12R) compared to when the same strains were grown on “Nutrient-Limited” medium (5W through 12W). On the other hand, for transformants 13-18, no bands of hybridization were observed on “Rich” medium (13R through 18R), while on “Nutrient-Limited” medium (13W through 18W), a faint band appeared (visible by eye, but not captured by digital imaging). Transformants 19-22, showed no visible bands of hybridization from either “Rich” (19R through 22R) or “Nutrient-Limited” (19W through 22W) media. All the bands that appeared were approximately of the same relative size.

Although the RNA gel with the samples isolated from “Rich” media stained darker with ethidium bromide both before and after the transfer, we did not detect *eGFP* transcripts for the isolates with “Nutrient-Limited” and “MI Late” promoters (13R-18R, and 19R-22R, respectively). The presence of a band from 12R, with similar intensity as

the 5R through 11R on another membrane eliminated the possibility of false negatives on this membrane.

On the other hand, 5W-12W (*i.e.*, “Nutrient-Limited” promoters driving transcription on “Nutrient-Limited” growth medium) presented a much fainter band than 5R-12R and the transcript bands for 13W-18W showed up, albeit faintly. This indicated that the “Nutrient-Limited” promoter was driving transcription in the “Nutrient-Limited” media (13W-18W) but not in the “Rich” media (13R-18R). The “Rich” promoter on the other hand, was stronger in “Rich” media (5R-12R) than in “Nutrient-Limited” media (5W-12W). With the confirmation that membrane and hybridization were as expected for the “Rich” and “Nutrient-Limited”, we could conclude that transformants with “MI Late” promoters indeed did not produce any detectable *eGFP* transcripts in either *in vitro* medium condition.

transformants) and lane 19-22 (C1-p1A1 and C1-p1A2) from both conditions had no visible bands. Lanes 5-12 (A1-p1A1, A1-p1A2, A2-p1A1 and A2-p1A2) from “Rich” were darker than those from “Nutrient-Limited” conditions. Lanes 13-18 (B1-p1A1, B1-p1A2 and B2-p1A1) from “Rich” medium had no visible band while the same transformant bands from “Nutrient-Limited” condition were visible by eye but cannot be captured on camera.

This is consistent with the expression level of the genes from which the promoters originated. MVLG_05589 is predicted to encode a hypothetical protein with an iron-sulfur cluster domain, while both MVLG_06949 and MVLG_01766 encode predicted hypothetical proteins with no identifiable domains. MVLG_05589 had an increase on the order of 4 log₂ fold change in “Rich” compared to “Nutrient-Limited”. MVLG_06949 was highly expressed in “Nutrient-Limited” but almost none in “Rich”. Lastly, MVLG_01766 was exclusively expressed in the “MI Late” stage of infection; hence it is reasonable that expression was undetectable in both “Rich” and “Nutrient-Limited” media.

Discussion

Reproducible transformation method

For several decades, transformation of *M. lychnidis-dioicae* has proven to be difficult, and such recalcitrance has impeded the use of molecular genetic approaches to study this organism. In this report, we demonstrate that through the use of *Agrobacterium*-mediated transformation, we have successfully and stably introduced the *HYG* selectable marker. Part of this success was due to the development of a pPZP plasmid derivative that contains a hygromycin selectable marker (pPZP-MV-HYG), driven by a putative promoter from an endogenous *M. lychnidis-dioicae* gene highly up-regulated under “Rich” conditions. Further, this newly developed vector provided the backbone for introduction of additional transgenes, initially in this work as proof-of-principle for expression/over-expression studies. It also provided a strategy for testing hypotheses derived from transcriptome analysis using RNA-Seq that identified genes as falling into particular expression classes based on the growth condition examined (e.g., “Rich”, “Nutrient-Limited”, “MI Late”).

As the transformation technique has been performed by at least two different investigators and repeated more than once, it appears reproducible for the strains tested. This is a breakthrough for further molecular genetic research in this model system. Although the number of resulting transformants was different for the different investigators who conducted initial experiments, this may reflect subtle differences in handling by the respective individuals and remains an area to optimize as we proceed in implementing this technique for future investigations.

Native promoter-driven expression

Further experiment with fusion amplicon tagged to different promoters was aimed at creating nutrient/growth environment-driven expression of genes under the control of native promoters. GFP was chosen as a potentially useful reporter gene. However the inserted gene was not codon optimized for use in *M. lychnidis-dioicae*. Although we were not able to convincingly detect GFP using confocal fluorescent microscopy (results not shown), evidence of transformation at genomic and transcriptional levels has clearly been demonstrated.

Based on Northern blot results, we could further assert that the promoters were driving the expression level as designed. The following observations about the expression of the genes used as sources for the promoters puts forward expected outcomes using their upstream regions in our expression constructs. MVLG_05589 was upregulated 4 log₂ fold change in “Rich” condition compared to in the “Nutrient-Limited” condition and 3 log₂ fold change compared to “MI Late” based on previous analysis. However, it was not “turned off” in either the “Nutrient-Limited” or the “MI Late” condition. MVLG_06949 and MVLG_01766 were exclusively expressed in the “Nutrient-Limited” and “MI Late” conditions, respectively. While we were not able to show expression at the protein level (e.g., by confocal fluorescence microscopy, or by western blots (not attempted), we were able to show that *eGFP* transcript levels were controlled by the attached “promoter” and not affected by the genomic insertion site of the T-DNA, which may be different in each transformant. Further experiments (e.g. Southern blot) will need to be performed on the transformants, to establish the characteristics and number of insertion sites in each transformant.

RT-PCR results presented some questions about the transcripts that were difficult to explain but the Northern blot results were able to show the absence and presence of the *eGFP* transcript, and the length of the transcripts clearly. In the RT-PCR screening for *HYG* transcript however, it appeared that alternative splicing by *Microbotryum* might have occurred, or it is possible that the transcripts were resulting from different insertion sites, hence controlled by different upstream and downstream transcription factors and RNA-processing mechanisms. Although this gene also was not codon optimized for *M. lychnidis-dioicae*, the fact that the transformants were selected based on hygromycin resistance indicated that the gene was being expressed effectively.

Potential use of the transformation method

Examination of the literature on transformation of fungi using the *Agrobacterium*-mediated approach suggests that the insertion of T-DNA is random but more often in the transcriptionally active sites (Alonso et al., 2003; Brunaud et al., 2002; Szabados et al., 2002). More recently, it was proven that the finding was due to the use of selection pressure when isolating positive transformants and that the T-DNA insertion sites are indeed random (Kim et al., 2007). Techniques for targeted gene replacement have been developed in *Fusarium* (Frandsen et al., 2012; Frandsen et al., 2008) and this is definitely worth investigating in the future for *M. lychnidis-dioicae*. Despite this, the method will still be tremendously useful for the over-expression of any gene of interest, in conjunction with a specific nutrient environment. Although we have yet to try targeted gene disruption, this has been successful in other recalcitrant fungi, including red yeasts using the system from which we derived our current protocol (Abbott et al., 2013), and

there is also the potential for using RNA silencing techniques to suppress gene of interest (Johansen & Carrington, 2001; Winterhagen et al., 2009).

In conclusion, we have successfully developed a robust and repeatable method for the transformation of *M. lychnidis-dioicae*, and at the same time, delivered a strategy for hypothesis testing that can involve over-expression in different nutrient/ growth conditions. Moreover, this methodology provides the opportunity to validate interpretations from transcriptome studies under these varied stages of fungal growth and development, as it allows testing how the putative promoters for differentially expressed genes drive reporter transgenes in transformed *M. lychnidis-dioicae*.

CHAPTER 7: SUMMARY AND CONCLUSIONS

Reference genome

Through this project, a reference genome for *Microbotryum lychnidis-dioicae* is now available. The sequences can be obtained from the database hosted by the Broad Institute:

http://www.broadinstitute.org/annotation/genome/Microbotryum_violaceum/MultiHome.html

We made various observations of the characteristics of the genome based on *in silico* analysis and found it to be different from other Basidiomycetes in many aspects. Most notably, there were many predicted genes that do not display identifiable domains. As a model organism in many aspects, the reference genome will serve as a starting point for discovery of diversity in other sibling species.

RNA-Sequencing of various life stages

RNA-Seq on the Illumina platform was the core technology used in this project to capture the transcriptomic information of this phytopathogen system, both *in vitro* and *in planta*. A lack of identifiable domains continued to pose a challenge in the analysis of the data, but we were able to infer the importance of secreted protein, small secreted proteins and proteins that are unique to *M. lychnidis-dioicae*. Observation of expected gene set enrichment provided assurance that the analysis yielded reliable results.

The host RNA-Seq data obtained, in addition to non-infected RNA-Seq data that is not mentioned here, will serve to further our understanding of this unique phytopathogen system.

Transformation system

Making use of *Agrobacterium*-mediated transformation and native promoters derived from the fungus, we have successfully developed a reliable and repeatable transformation system that can be manipulated based on life stage-driven transcription. This is an important milestone in the study of this model organism.

Future directions

Based on our analyses, the suite of genes that were found to be differentially regulated in the infected plants cannot be functionally resolved because of the lack of conservation of such sequences among other fungal-host systems that have been annotated. As such, their functions will be reliant on traditional over-expression, knockout or more advanced gene-silencing methods for examining genes. In this study, we have devised a usable over-expression system for *M. lychnidis-dioicae*, something that could not be achieved in the last 30 years. This will be one of the first steps of putting genes of interest under further scrutiny. Both the RNA-Seq data made available by this study and continued interaction-based discovery will help direct the selection of such genes of interest and refine the genome sequences and gene prediction models.

REFERENCES

- Abbott EP, Ianiri G, Castoria R, Idnurm A. (2013) Overcoming recalcitrant transformation and gene manipulation in Pucciniomycotina yeasts. *Appl Microbiol Biotechnol* 97: 283-95. doi: 10.1007/s00253-012-4561-7
- Ageez A, Kazama Y, Sugiyama R, Kawano S. (2005) Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. *Genes Genet Syst* 80: 403-13. doi: 10.1266/ggs.80.403
- Akhter S, Antonovics J. (1999) Use of internal transcribed spacer primers and fungicide treatments to study the anther-smut disease, *Microbotryum violaceum* (= *Ustilago violacea*), of White Campion *Silene alba* (= *Silene latifolia*). *Int J Plant Sci* 160: 1171-76. doi: 10.1086/314197
- Alexander HM. (1989) An experimental field-study of anther-smut disease of *Silene alba* caused by *Ustilago violacea* - genotypic variation and disease incidence. *Evolution* 43: 835-47. doi: 10.2307/2409311
- Alexander HM, Antonovics J. (1995) Spread of anther-smut disease (*Ustilago violacea*) and character correlations in a genetically variable experimental population of *Silene alba*. *J Ecol* 83: 783-94. doi: 10.2307/2261416
- Alexander NJ, McCormick SP, Hohn TM. (1999) TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. *Mol Gen Genetics* 261: 977-84. doi: 10.1007/s004380051046
- Alizon S, Hurford A, Mideo N, Van Baalen M. (2009) Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J Evol Biol* 22: 245-59. doi: 10.1111/j.1420-9101.2008.01658.x
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301: 653-57. doi: 10.1126/science.1086391
- Amselem J, Lebrun MH, Quesneville H. (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics* 16: 141. doi: 10.1186/s12864-015-1347-1

- Antoniw JF, Ritter CE, Pierpoint WS, Van Loon LC. (1980) Comparison of three pathogenesis-related proteins from plants of two cultivars of tobacco infected with TMV. *J Gen Vir* 47: 79-87. doi: 10.1099/0022-1317-47-1-79
- Antonovics J. (2005) Plant venereal diseases: insights from a messy metaphor. *New Phytol* 165: 71-80. doi: 10.1111/j.1469-8137.2004.01215.x
- Antonovics J, Hood M, Partain J. (2002) The ecology and genetics of a host shift: *Microbotryum* as a model system. *Amer Nat* 160 Suppl 4: S40-53. doi: 10.1086/342143
- Bachtrog D. (2003) Accumulation of *Spock* and *Worf*, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol Biol Evol* 20: 173-81. doi: 10.1093/molbev/msg035
- Badouin H, Hood ME, Gouzy J, Aguilera G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. (2015) Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* 200: 1275-84. doi: 10.1534/genetics.115.177709
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202-08. doi: 10.1093/nar/gkp335
- Baird ML, Garber ED. (1981) The genetics and biochemistry of urease in *Ustilago violacea*. *Biochem Genet* 19: 1101-14. doi: 10.1007/BF00484568
- Baker HG. (1947) *Melandrium*. *J Ecol* 35: 271-74. doi: 10.2307/2256515
- Baldrian P. (2006) Fungal laccases - occurrence and properties. *FEMS Microbiol Rev* 30: 215-42. doi: 10.1111/j.1574-4976.2005.00010.x
- Bao Z, Eddy SR. (2002) Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res* 12: 1269-76. doi: 10.1101/gr.88502
- Bauer R, Begerow D, Sampaio J, Weiß M, Oberwinkler F. (2006) The simple-septate basidiomycetes: a synopsis. *Mycol Progress* 5: 41-66. doi: 10.1007/s11557-006-0502-0
- Bauer R, Oberwinkler F, Vánky K. (1997) Ultrastructural markers and systematics in smut fungi and allied taxa. *Can J Bot* 75: 1273-314. doi: 10.1139/b97-842
- Becht P, König J, Feldbrugge M. (2006) The RNA-binding protein Rrm4 is essential for polarity in *Ustilago maydis* and shuttles along microtubules. *J Cell Sci* 119: 4964-73. doi: 10.1242/jcs.03287

- Begerow D, Bauer R, Oberwinkler F. (1997) Phylogenetic studies on nuclear large subunit ribosomal DNA sequences of smut fungi and related taxa. *Can J Bot* 75: 2045-56. doi: 10.1139/b97-916
- Bej AK, Perlin MH. (1989a) A high efficiency transformation system for the basidiomycete *Ustilago violacea* employing hygromycin resistance and lithium-acetate treatment. *Gene* 80: 171-76. doi: 10.1016/0378-1119(89)90263-1
- Bej AK, Perlin MH. (1989b) Preparation of protoplasts of *Ustilago violacea* for transformation and pulse-field electrophoresis. *BioTechniques* 7: 824-28. PMID: 2631789
- Bej AK, Perlin MH. (1991) Acquisition of mitochondrial DNA by a transformation vector for *Ustilago violacea*. *Gene* 98: 135-40. doi: 10.1016/0378-1119(91)90116-S
- Bendtsen DJ, Nielsen H, von Heijne G, Brunak S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783-95. doi: 10.1016/j.jmb.2004.05.028
- Benjamini Y, Hochberg Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc B* 57: 289-300.
- Biémont C. (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* 186: 1085-93. doi: 10.1534/genetics.110.124180
- Biere A, Honders S. (1996) Host adaptation in the anther smut fungus *Ustilago violacea* (*Microbotryum violaceum*): Infection success, spore production and alteration of floral traits on two host species and their F1-hybrid. *Oecologia* 107: 307-20. doi: 10.1007/BF00328447
- Billiard S, Lopez-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. (2011) Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev Camb Philos Soc* 86: 421-42. doi: 10.1111/j.1469-185X.2010.00153.x
- Billiard S, López-Villavicencio M, Hood ME, Giraud T. (2012) Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J Evol Biol* 25: 1020-38. doi: 10.1111/j.1420-9101.2012.02495.x
- Birney E, Clamp M, Durbin R. (2004) GeneWise and Genomewise. *Genome Res* 14: 988-95. doi: 10.1101/gr.1865504

- Bocquet GF. (1967) Physolychnidium olim Gastrolychnidium nomenclaturae fundamentum includens combinationes taxaque nova nonnulla Silenes generis. *Candollea* 22: 1-38.
- Borodovsky M, Lomsadze A. (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatic* 35: 4.6.1–4.6.10. doi: 10.1002/0471250953.bi0406s35
- Brachmann A, König J, Julius C, Feldbrugge M. (2004) A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis*. *Mol Genet Genomics* 272: 216-26. doi: 10.1007/s00438-004-1047-z
- Brameier M, Krings A, MacCallum RM. (2007) NucPred - Predicting nuclear localization of proteins. *Bioinformatics* 23: 1159-60. doi: 10.1093/bioinformatics/btm066
- Brose N, Betz A, Wegmeyer H. (2004) Divergent and convergent signaling by the diacylglycerol second messenger pathway in mammals. *Curr Opin Neurobiol* 14: 328-40. doi: 10.1016/j.conb.2004.05.006
- Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G *et al.* (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. *EMBO Rep* 3: 1152-57. doi: 10.1093/embo-reports/kvf237
- Bucheli, Gautschi, Shykoff. (2000) Host-specific differentiation in the anther smut fungus *Microbotryum violaceum* as revealed by microsatellites. *J Evol Biol* 13: 188-98. doi: 10.1046/j.1420-9101.2000.00160.x
- Bucheli E, Gautschi B, Shykoff JA. (1998) Primer notes: Isolation and characterization of microsatellite loci in the anther smut fungus *Microbotryum violaceum*. *Mol Ecol* 7: 665-66. doi: 10.1046/j.1365-294X.1998.00340.x
- Bundock P, den Dulk-Ras A, Beijersbergen A, Hooykaas PJ. (1995) Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. *EMBO J* 14: 3206-14. PMID: PMC394382
- Cambareri EB, Jensen BC, Schabtach E, Selker EU. (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* 244: 1571-75. doi: 10.1126/science.2544994
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37: D233-8. doi: 10.1093/nar/gkn663

- Caplan A, Herrera-Estrella L, Inze D, Van Haute E, Van Montagu M, Schell J, Zambryski P. (1983) Introduction of genetic material into plant cells. *Science* 222: 815-21. doi: 10.1126/science.222.4625.815
- Carlsson-Graner U, Pettersson TM. (2005) Host susceptibility and spread of disease in a metapopulation of *Silene dioica*. *Evol Ecol Res* 7: 353-69.
- Cashion NL, Luttrell ES. (1988) Host parasite relationships in Karnal bunt of wheat. *Phytopathology* 78: 75-84. doi: 10.1094/Phyto-78-75
- Castle AJ, Day AW. (1984) Isolation and identification of α -tocopherol as an inducer of the parasitic phase of *Ustilago violacea*. *Phytopathology* 74: 1194-200. doi: 10.1094/Phyto-74-1194
- Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, Miyasato SR, Binkley J, Orvis J, Shah P *et al.* (2014) The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Res* 42: D705-10. doi: 10.1093/nar/gkt1029
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR *et al.* (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40: D700-5. doi: 10.1093/nar/gkr1029
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-76. doi: 10.1093/bioinformatics/bti610
- Consortium TU. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* 39: D214-19. doi: 10.1093/nar/gkq1020
- Consortium TU. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40: D71-75. doi: 10.1093/nar/gkr981
- Covert SF, Kapoor P, Lee M-h, Briley A, Nairn CJ. (2001) *Agrobacterium tumefaciens*-mediated transformation of *Fusarium circinatum*. *Mycol Res* 105: 259-64. doi: 10.1017/s0953756201003872
- Cummins JE, Day AW. (1977) Genetic and cell cycle analysis of a smut fungus (*Ustilago violacea*). *Methods Cell Biol* 15: 445-69. doi: 10.1016/S0091-679X(08)60232-0
- Day AW. (1976) Communication through fimbriae during conjugation in a fungus. *Nature* 262: 583-84.
- Day AW, Jones JK. (1969) Sexual and parasexual analysis of *Ustilago violacea*. *Genet Res* 14: 195-221. doi: 10.1017/S0016672300002056

- de Vienne DM, Hood ME, Giraud T. (2009a) Phylogenetic determinants of potential host shifts in fungal pathogens. *J Evol Biol* 22: 2532-41. doi: 10.1111/j.1420-9101.2009.01878.x
- de Vienne DM, Refrégier G, Hood ME, Guigue A, Devier B, Vercken E, Smadja C, Deseille A, Giraud T. (2009b) Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. *J Evol Biol* 22: 683-98. doi: 10.1111/j.1420-9101.2009.01702.x
- Delph LF, Arntz AM, Scotti-Saintagne C, Scotti I. (2010) The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. *Evolution* 64: 2873-86. doi: 10.1111/j.1558-5646.2010.01048.x
- Denchev CM, Giraud T, Hood ME. (2009) Three new species of anthericolous smut fungi on *Caryophyllaceae*. *Mycol Balcanica* 6: 79-84.
- Devier B, Aguileta G, Hood ME, Giraud T. (2008) Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* 181: 209-23. doi: 10.1534/genetics.108.093708
- Dong W, Lv H, Xia G, Wang M. (2012) Does diacylglycerol serve as a signaling molecule in plants? *Plant Signal Behav* 7: 472-75. doi: 10.4161/psb.19644
- Dunkel N, Hertlein T, Franz R, Reuß O, Sasse C, Schäfer T, Ohlsen K, Morschhäuser J. (2013) Roles of different peptide transporters in nutrient acquisition in *Candida albicans*. *Eukaryot Cell* 12: 520-28. doi: 10.1128/EC.00008-13
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci USA* 108: 9166-71. doi: 10.1073/pnas.1019315108
- Duret L, Galtier N. (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 10: 285-311. doi: 10.1146/annurev-genom-082908-150001
- Eddy SR. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7: e1002195. doi: 10.1371/journal.pcbi.1002195
- Edgar RC, Myers EW. (2005) PILER: identification and classification of genomic repeats. *Bioinformatics* 21 Suppl 1: 152-58. doi: 10.1093/bioinformatics/bti1003
- Eisikowitch D, Lachance MA, Kevan PG, Willis S, Collins-Thompson DL. (1990) The effect of the natural assemblage of microorganisms and selected strains of the yeast *Metschnikowia reukaufii* in controlling the germination of pollen of the

- common milkweed *Asclepias syriaca*. *Can J Bot* 68: 1163-65. doi: 10.1139/b90-147
- Elhaik E, Landan G, Graur D. (2009) Can GC content at third-codon positions be used as a proxy for isochore composition? *Mol Biol Evol* 26: 1829-33. doi: 10.1093/molbev/msp100
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953-71. doi: 10.1038/nprot.2007.131
- Farbos I, Oliveira M, Negrutiu I, Mouras A. (1997) Sex organ determination and differentiation in the dioecious plant *Melandrium album* (*Silene latifolia*): a cytological and histological analysis. *Sex Plant Reprod* 10: 155-67. doi: 10.1007/s004970050083
- Feldbrügge M. **Universität Düsseldorf: pMF5-3n with nourseothricin resistance cassette.** In: 2010. <http://www.mikrobiologie.hhu.de/de/ustilago-community/types-of-genetic-engineering/53-c-terminal-fusion-with-gfp-taptag/pmf5-3n-with-nourseothricin-resistance-cassette.html>.
- Filatov DA. (2005) Substitution rates in a new *Silene latifolia* sex-linked gene, *SlssX/Y*. *Mol Biol Evol* 22: 402-08. doi: 10.1093/molbev/msi003
- Filatov DA, Moneger F, Negrutiu I, Charlesworth D. (2000) Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* 404: 388-90.
- Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS *et al.* (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336: 1715-19. doi: 10.1126/science.1221748
- Flutre T, Duprat E, Feuillet C, Quesneville H. (2011) Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE* 6: e16526. doi: 10.1371/journal.pone.0016526
- Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, de Vienne DM, Aguileta G, Poulain J, Wincker P *et al.* (2015) Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. *Mol Biol Evol* 32: 928-43. doi: 10.1093/molbev/msu396
- Frandsen R, Frandsen M, Giese H: **Targeted gene replacement in fungal pathogens via *Agrobacterium tumefaciens*-mediated transformation.** In: *Plant Fungal Pathogens*. Edited by Melvin D. Bolton, Thomma Bart P. H. J., vol. 835 2012. Chapter 2. Humana Press, pp. 17-45 doi: 10.1007/978-1-61779-501-5_2

- Frandsen RJ, Andersson JA, Kristensen MB, Giese H. (2008) Efficient four fragment cloning for the construction of vectors for targeted gene replacement in filamentous fungi. *BMC Mol Biol* 9: 70. doi: 10.1186/1471-2199-9-70
- Fraser JA, Heitman J. (2004) Evolution of fungal sex chromosomes. *Mol Microbiol* 51: 299-306. doi: 10.1046/j.1365-2958.2003.03874.x
- Fraser JA, Hsueh YP, Findley KM, Heitman J: **Evolution of the mating-type locus: the Basidiomycetes**. In: *Sex in Fungi: Molecular Determination and Evolutionary Implication*. Edited by J. Heitman, Kronstad J., Taylor J., Casselton L. 2007. Chapter 2. ASM Press, Washington, DC. pp. 19-34
- Freeman AB, Duong KK, Shi TL, Hughes CF, Perlin MH. (2002) Isolates of *Microbotryum violaceum* from North American host species are phylogenetically distinct from their European host-derived counterparts. *Mol Phylogenet Evol* 23: 158-70. doi: 10.1016/S1055-7903(02)00003-9
- Fryxell KJ, Moon W-J. (2005) CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol* 22: 650-58. doi: 10.1093/molbev/msi043
- Fujita N, Aonuma W, Shimizu Y, Yamanaka K, Hirata A, Hood ME, Kawano S. (2012) A petalless flower caused by a *Microbotryum violaceum* mutant. *Int J Plant Sci* 173: 464-73. doi: 10.1086/665587
- Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859-68. doi: http://www.nature.com/nature/journal/v422/n6934/supinfo/nature01554_S1.html
- Garber ED, Day AW. (1985) Genetic mapping of a phytopathogenic basidiomycete, *Ustilago violacea*. *Bot Gaz* 146: 449-59.
- Garr SS, Hughes C, Welch J, Brown SA, Perlin MH. (1997) Correlation of haplotypes of a fungal plant pathogen with their respective host species of origin. *Plant Disease* 81: 936-41. doi: 10.1094/PDIS.1997.81.8.936
- Gibson AK, Hood ME, Giraud T. (2012) Sibling competition arena: selfing and a competition arena can combine to constitute a barrier to gene flow in sympatry. *Evolution* 66: 1917-30. doi: 10.1111/j.1558-5646.2011.01563.x
- Gibson AK, Petit E, Mena-Ali J, Oxelman B, Hood ME. (2013) Life-history strategy defends against disease and may select against physiological resistance. *Ecol Evol* 3: 1741-50. doi: 10.1002/ece3.583

- Gillissen B, Bergemann J, Sandmann C, Schroeer B, Bolker M, Kahmann R. (1992) A two-component regulatory system for self/non-self recognition in *Ustilago maydis*. *Cell* 68: 647-57. doi: 10.1016/0092-8674(92)90141-X
- Giraud T, Fournier E, Vautrin D, Solignac M, Shykoff JA. (2002) Isolation of 44 polymorphic microsatellite loci in three host races of the phytopathogenic fungus *Microbotryum violaceum*. *Mol Ecol Notes* 2: 142-46. doi: 10.1046/j.1471-8286.2002.00172.x
- Giraud T, Yockteng R, Lopez-Villavicencio M, Refrégier G, Hood ME. (2008a) Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell* 7: 765-75. doi: 10.1128/EC.00440-07
- Giraud T, Yockteng R, Marthey S, Chiapello H, Jonot O, Lopez-Granados E, De Vienne DM, Hood ME, Refrégier G, Gendrault-JacQuemard A *et al.* (2008b) Isolation of 60 polymorphic microsatellite loci in EST libraries of four sibling species of the phytopathogenic fungal complex *Microbotryum*. *Mol Ecol Resources* 8: 387-92. doi: 10.1111/j.1471-8286.2007.01967.x
- Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, Rodriguez de la Vega RC, Branco S, Giraud T. (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* 23: 753-73. doi: 10.1111/mec.12631
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644-52. doi: 10.1038/nbt.1883
- Granberg Å: ***Microbotryum violaceum* on *Silene dioica* - understanding traits that influence plant-pathogen interactions.** Arkitektkopia, Umeå, Sweden: Umeå University; 2007.
- Grant S, Hunkirchen B, Saedler H. (1994) Developmental differences between male and female flowers in the dioecious plant *Silene latifolia*. *Plant J* 6: 471-80. doi: 10.1046/j.1365-313X.1994.6040471.x
- Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA *et al.* (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* 40: D26-32. doi: 10.1093/nar/gkr947
- Guttman DS, Charlesworth D. (1998) An X-linked gene with a degenerate Y-linked homologue in a dioecious plant. *Nature* 393: 263-66. doi: 10.1038/30492

- Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461: 393-98. doi: 10.1038/nature08358
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8: 1494-512. doi: 10.1038/nprot.2013.084
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* 9: R7. doi: 10.1186/gb-2008-9-1-r7
- Hane JK, Oliver RP. (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* 9: 478. doi: 10.1186/1471-2105-9-478
- Hardenack S, Ye D, Saedler H, Grant S. (1994) Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. *Plant Cell* 6: 1775-87. doi: 10.1105/tpc.6.12.1775
- Heckman KL, Pease LR. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* 2: 924-32. doi: 10.1038/nprot.2007.132
- Holliday R: *Ustilago maydis*. In: *Bacteria, Bacteriophages, and Fungi*. Edited by Robert C King 1974. Chapter 31. Springer US, pp. 575-95 doi: 10.1007/978-1-4899-1710-2_31
- Hood M: **Infection rate of *Microbotryum lychnidis-dioicae* Lamole strains**. In. Edited by Toh. Su San; 2015.
- Hood ME. (2002) Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* 160: 457-61. PMID: PMC1461988
- Hood ME. (2003) Dynamics of multiple infection and within-host competition by the anther-smut pathogen. *Amer Nat* 162: 122-33. doi: 10.1086/375539
- Hood ME. (2005) Repetitive DNA in the automictic fungus *Microbotryum violaceum*. *Genetica* 124: 1-10. doi: 10.1007/s10709-004-6615-y
- Hood ME, Antonovics J. (1998) Two-celled promycelia and mating-type segregation in *Ustilago violacea* (*Microbotryum violaceum*). *Int J Plant Sci* 159: 199-205.

- Hood ME, Antonovics J. (2000) Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (= *Ustilago violacea*). *Heredity* 85: 231-41. doi: 10.1046/j.1365-2540.2000.00748.x
- Hood ME, Antonovics J. (2003) Plant species descriptions show signs of disease. *P Roy Soc Lond B Bio* 270: S156-58. doi: 10.1098/rsbl.2003.0063
- Hood ME, Antonovics J, Koskella B. (2004) Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics* 168: 141-46. doi: 10.1534/genetics.104.029900
- Hood ME, Katawczik M, Giraud T. (2005) Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics* 170: 1081-89. doi: 10.1534/genetics.105.042564
- Hood ME, Mena-Ali JJ, Gibson AK, Oxelman B, Giraud T, Yockteng R, Arroyo MT, Conti F, Pedersen AB, Gladieux P *et al.* (2010) Distribution of the anther-smut pathogen *Microbotryum* on species of the Caryophyllaceae. *New Phytol* 187: 217-29. doi: 10.1111/j.1469-8137.2010.03268.x
- Hood ME, Petit E, Giraud T. (2013) Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics* 193: 309-15. doi: 10.1534/genetics.112.146266
- Hooykaas PJJ, Roobol C, Schilperoort RA. (1979) Regulation of the transfer of Ti plasmids of *Agrobacterium tumefaciens*. *J Gen Microbiol* 110: 99-109. doi: 10.1099/00221287-110-1-99
- Horns F, Petit E, Yockteng R, Hood ME. (2012) Patterns of repeat-induced point mutation in transposable elements of basidiomycete fungi. *Genome Biol* 4: 240-47. doi: 10.1186/gb-2008-9-5-r77
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35: W585-87. doi: 10.1093/nar/gkm259
- Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldon T. (2011) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res* 39: 556-60. doi: 10.1093/nar/gkq1109
- Huerta-Cepas J, Gabaldon T. (2011) Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics* 27: 38-45. doi: 10.1093/bioinformatics/btq609

- Hughes CF, Perlin MH. (2005) Differential expression of *mepA*, *mepC* and *smtE* during growth and development of *Microbotryum violaceum*. *Mycologia* 97: 605-11. doi: 10.3852/mycologia.97.3.605
- Ianiri G, Wright SA, Castoria R, Idnurm A. (2011) Development of resources for the analysis of gene function in Pucciniomycotina red yeasts. *Fungal Genet Biol* 48: 685-95. doi: 10.1016/j.fgb.2011.03.003
- Jennersten O. (1983) Butterfly visitors as vectors of *Ustilago violacea* spores between Caryophyllaceous plants. *Oikos* 40: 125-30. doi: 10.2307/3544207
- Jiang C, Zhao Z. (2006) Directionality of point mutation and 5-methylcytosine deamination rates in the chimpanzee genome. *BMC Genomics* 7: 316. doi: 10.1186/1471-2164-7-316
- Johansen LK, Carrington JC. (2001) Silencing on the spot. Induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. *Plant Physiol* 126: 930-38. doi: 10.1104/pp.126.3.930
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110: 462-67. doi: 10.1159/000084979
- Kadota K, Nishiyama T, Shimizu K. (2012) A normalization strategy for comparing tag count data. *Algorithms Mol Biol* 7: 5. doi: 10.1186/1748-7188-7-5
- Käll L, Krogh A, Sonnhammer EL. (2007) Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Res* 35: W429-32. doi: 10.1093/nar/gkm256
- Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444: 97-101. doi: 10.1038/nature05248
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*: doi: 10.1093/nar/gkv1070
- Kapitonov VV, Jurka J. (2001) Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci USA* 98: 8714-19. doi: 10.1073/pnas.151269298
- Kawahara Y, Oono Y, Kanamori H, Matsumoto T, Itoh T, Minami E. (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS ONE* 7: e49423. doi: 10.1371/journal.pone.0049423

- Kazama Y, Koizumi A, Uchida W, Ageez A, Kawano S. (2005) Expression of the floral B-function gene *SLM2* in female flowers of *Silene latifolia* infected with the smut fungus *Microbotryum violaceum*. *Plant Cell Physiol* 46: 806-11. doi: 10.1093/pcp/pci080
- Kemler M, Lutz M, Göker M, Oberwinkler F, Begerow D. (2009) Hidden diversity in the non-caryophyllaceous plant-parasitic members of *Microbotryum* (Pucciniomycotina: Microbotryales). *Syst Biodivers* 7: 297-306. doi: 10.1017/S1477200009990028
- Kent WJ. (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12: 656-64. doi: 10.1101/gr.229202. Article published online before March 2002
- Kim SI, Veena, Gelvin SB. (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. *Plant J* 51: 779-91. doi: 10.1111/j.1365-313X.2007.03183.x
- Klose J, de Sa MM, Kronstad JW. (2004) Lipid-induced filamentous growth in *Ustilago maydis*. *Mol Microbiol* 52: 823-35. doi: 10.1111/j.1365-2958.2004.04019.x
- Kokontis J, Ruddat M. (1986) Promotion of hyphal growth in *Ustilago violacea* by host factors from *Silene alba*. *Arch Microbiol* 144: 302-06. doi: 10.1007/BF00410967
- Kokontis JM, Ruddat M. (1989) Enzymatic hydrolysis of hyphal growth factors for *Ustilago violacea* isolated from the host plant *Silene alba*. *Bot Gaz* 150: 439-44.
- Krogh A, Larsson Br, von Heijne G, Sonnhammer ELL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 305: 567-80. doi: 10.1006/jmbi.2000.4315
- Kües U, Rühl M. (2011) Multiple multi-copper oxidase gene families in Basidiomycetes – What for? *Curr Genomics* 12: 72-94. doi: 10.2174/138920211795564377
- Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. doi: 10.1186/gb-2009-10-3-r25
- Le Gac M, Hood ME, Fournier E, Giraud T. (2007a) Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution* 61: 15-26. doi: 10.1111/j.1558-5646.2007.00002.x
- Le Gac M, Hood ME, Giraud T. (2007b) Evolution of reproductive isolation within a parasitic fungal species complex. *Evolution* 61: 1781-87. doi: 10.1111/j.1558-5646.2007.00144.x

- Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* 7: 709-15. doi: 10.1038/nmeth.1491
- Li B, Dewey CN. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323. doi: 10.1186/1471-2105-12-323
- Li L, Stoeckert CJ, Roos DS. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178-89. doi: 10.1101/gr.1224503
- Liepmann AH, Nairn CJ, Willats WGT, Sorensen I, Roberts AW, Keegstra K. (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. *Plant Physiol* 143: 1881-93. doi: 10.1104/pp.106.093989
- Linné Cv. 1737. *Hortus Cliffortianus*. Amsterdam
- Liro JJ. 1924. *Die ustilagineen Finnlands*, vol. 17A. Helsinki: Suomalaisen Tiedeakatemian Kustantama
- Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA *et al.* (2005) The genome of the basidiomycetous yeast and human pathogen *Science* 307: 1321-24. doi: 10.1126/science.1103773
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42: D490-95. doi: 10.1093/nar/gkt1178
- Lowe RG, Cassin A, Grandaubert J, Clark BL, Van de Wouw AP, Rouxel T, Howlett BJ. (2014) Genomes and transcriptomes of partners in plant-fungal-interactions between canola (*Brassica napus*) and two *Leptosphaeria* species. *PLoS ONE* 9: e103098. doi: 10.1371/journal.pone.0103098
- Lu DQ, Lidén M, Oxelman B. ***Lychnis***. In: efloras. Published on the Internet. 2008. <http://efloras.org>. Accessed 31 Oct 2015. Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA.
- Lubkowitz MA, Hauser L, Breslav M, Naider F, Becker JM. (1997) An oligopeptide transport gene from *Candida albicans*. *Microbiology* 143: 387-96. doi: 10.1099/00221287-143-2-387
- Lum G, Min XJ. (2011) FunSecKB: the Fungal Secretome KnowledgeBase. *Database (Oxford)* 2011: doi: 10.1093/database/bar001

- Luo H, Perlin MH. (1993) The gamma-tubulin-encoding gene from the basidiomycete fungus, *Ustilago violacea*, has a long 5'-untranslated region. *Gene* 137: 187-94. doi: 10.1016/0378-1119(93)90005-N
- Luttrell ES. (1981) Tissue replacement diseases caused by fungi. *Annu Rev Phytopathol* 19: 373-89. doi: 10.1146/annurev.py.19.090181.002105
- Lutz M, Göker M, Piatek M, Kemler M, Begerow D, Oberwinkler F. (2005) Anther smuts of Caryophyllaceae: Molecular characters indicate host-dependent species delimitation. *Mycol Progress* 4: 225-38. doi: 10.1007/s11557-006-0126-4
- Majoros WH, Pertea M, Salzberg SL. (2004) TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* 20: 2878-79. doi: 10.1093/bioinformatics/bth315
- Marais GA, Forrest A, Kamau E, Kafer J, Daubin V, Charlesworth D. (2011) Multiple nuclear gene phylogenetic analysis of the evolution of dioecy and sex chromosomes in the genus *Silene*. *PLoS ONE* 6: e21915. doi: 10.1371/journal.pone.0021915
- Martin F, Aerts A, Ahren D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452: 88-92. doi: 10.1038/nature06556
- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P *et al.* (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci USA* 106: 1954-9. doi: 10.1073/pnas.0809575106
- Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 22: 695-700. doi: 10.1038/nbt967
- Maruthachalam K, Klosterman SJ, Kang S, Hayes RJ, Subbarao KV. (2011) Identification of pathogenicity-related genes in the vascular wilt fungus *Verticillium dahliae* by *Agrobacterium tumefaciens*-mediated T-DNA insertional mutagenesis. *Mol Biotechnol* 49: 209-21. doi: 10.1007/s12033-011-9392-8
- Mattanovich D, Ruker F, Machado AC, Laimer M, Regner F, Steinkellner H, Himmler G, Katinger H. (1989) Efficient transformation of *Agrobacterium* spp. by electroporation. *Nucleic Acids Res* 17: 6747. doi: 10.1093/nar/17.16.6747

- McCormac AC, Elliott MC, Chen DF. (1998) A simple method for the production of highly competent cells of *Agrobacterium* for transformation via electroporation. *Mol Biotechnol* 9: 155-59. doi: 10.1007/BF02760816
- Mendgen K, Hahn M, Deising H. (1996) Morphogenesis and mechanisms of penetration by plant pathogenic fungi. *Annu Rev Phytopathol* 34: 367-86.
- Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. (2008) The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet* 4: e1000030. doi: 10.1371/journal.pgen.1000030
- Miller PM, Kesseli RV. (2011) A sex-chromosome mutation in *Silene latifolia*. *Sex Plant Reprod*: 1-7. doi: 10.1007/s00497-011-0163-2
- Moccia MD, Oger-Desfeux C, Marais GA, Widmer A. (2009) A White Champion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* 10: 243. doi: 10.1186/1471-2164-10-243
- Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E *et al.* (2003) PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267-73. doi: http://www.nature.com/ng/journal/v34/n3/supinfo/ng1180_S1.html
- Morton BR, Bi IV, McMullen MD, Gaut BS. (2006) Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. *Genetics* 172: 569-77. doi: 10.1534/genetics.105.049916
- Mrackova M, Nicolas M, Hobza R, Negruțiu I, Moneger F, Widmer A, Vyskot B, Janoušek B. (2008) Independent origin of sex chromosomes in two species of the genus *Silene*. *Genetics* 179: 1129-33. doi: 10.1534/genetics.107.085670
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA. (2012) Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol* 10: e1001308. doi: 10.1371/journal.pbio.1001308
- Nicolas FE, Torres-Martinez S, Ruiz-Vazquez RM. (2013) Loss and retention of RNA interference in fungi and parasites. *PLoS Pathog* 9: e1003089. doi: 10.1371/journal.ppat.1003089
- Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negruțiu I, Charlesworth D, Monéger F. (2005) A gradual process of

- recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol* 3: e4. doi: 10.1371/journal.pbio.0030004
- Nishida H, Nagatsuka Y, Sugiyama J. (2011) Draft genome sequencing of the enigmatic basidiomycete *Mixia osmundae*. *The Journal of General and Applied Microbiology* 57: 63-67. doi: 10.2323/jgam.57.63
- Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE *et al.* (2010) Genome sequence of the model mushroom *Schizophyllum commune*. *Nat Biotechnol* 28: 957-63. doi: 10.1038/nbt.1643
- Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F *et al.* (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathog* 8: e1003037. doi: 10.1371/journal.ppat.1003037
- Oliveros JC. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. version 2.0. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>
- Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. (2011) About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet* 27: 358-67. doi: 10.1016/j.tig.2011.05.001
- Oxelman B, Lidén M, Turland NJ. (2001) Taxonomic and Nomenclatural Notes on Chinese *Silene* (Caryophyllaceae). *Novon* 11: 322-24. doi: 10.2307/3393038
- Palmieri G, Bianco C, Cennamo G, Giardina P, Marino G, Monti M, Sannia G. (2001) Purification, characterization, and functional role of a novel extracellular protease from *Pleurotus ostreatus*. *Appl Environ Microbiol* 67: 2754-59. doi: 10.1128/AEM.67.6.2754-2759.2001
- Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P. (2003) Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. *Enzyme Microb Technol* 33: 220-30. doi: 10.1016/s0141-0229(03)00117-0
- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, Lehrach H, Soldatov A. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 37: e123. doi: 10.1093/nar/gkp596
- Parra G, Bradnam K, Korf I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061-7. doi: 10.1093/bioinformatics/btm071

- Paul D, Magbanua Z, Arick M, 2nd, French T, Bridges SM, Burgess SC, Lawrence ML. (2014) Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC 204091. *Genome Announc* 2: doi: 10.1128/genomeA.00046-14
- Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, Duplessis S, Henrissat B, Young S, Zeng Q *et al.* (2015) Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics* 16: 461. doi: 10.1186/s12864-015-1660-8
- Perlin MH, Bej AK, Will III OH, Jacob RJ. (1990) Introduction and maintenance of prokaryotic DNA in *Ustilago violacea*. *J Ind Microbiol* 5: 355-63. doi: 10.1007/BF01578094
- Perlin MH, Hughes C, Welch J, Akkaraju S, Steinecker D, Kumar A, Smith B, Garr SS, Brown SA, Ando T. (1997) Molecular approaches to differentiate subpopulations or formae speciales of the fungal phytopathogen *Microbotryum violaceum*. *Int J Plant Sci* 158: 568-74.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785-86. doi: 10.1038/nmeth.1701
- Petit E, Giraud T, de Vienne DM, Coelho MA, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory F, Wincker P *et al.* (2012) Linkage to the mating-type locus across the genus *Microbotryum*: insights into nonrecombining chromosomes. *Evolution* 66: 3519-33. doi: 10.1111/j.1558-5646.2012.01703.x
- Petre B, Morin E, Tisserant E, Hacquard S, Da Silva C, Poulain J, Delaruelle C, Martin F, Rouhier N, Kohler A *et al.* (2012) RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers *PtSultr3;5*, a fungal-induced host sulfate transporter. *PLoS ONE* 7: e44408. doi: 10.1371/journal.pone.0044408
- Pierleoni A, Martelli PL, Casadio R. (2008) PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* 9: 392. doi: 10.1186/1471-2105-9-392
- Prillinger H, Deml G, Dörfler C, Laaser G, Lockau W. (1991) A contribution to the systematics and evolution of higher fungi: Yeast-types in the Basidiomycetes. Part II: Microbotryum-type. *Bot Acta* 104: 5-17. doi: 10.1111/j.1438-8677.1991.tb00188.x
- Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol* 1: 166-75. doi: 10.1371/journal.pcbi.0010022

- Refrégier G, Le Gac M, Jabbour F, Widmer A, Shykoff JA, Yockteng R, Hood ME, Giraud T. (2008) Cophylogeny of the anther smut fungi and their Caryophyllaceae hosts: Prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evol Biol* 8: 100. doi: 10.1186/1471-2148-8-100
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. (2006) GenePattern 2.0. *Nat Genet* 38: 500-01. doi: 10.1038/ng0506-500
- Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R *et al.* (2014) Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci USA* 111: 9923-28. doi: 10.1073/pnas.1400592111
- Robertson SE, Li Y, Scutt CP, Willis ME, Gilmartin PM. (1997) Spatial expression dynamics of *Men-9* delineate the third floral whorl in male and female flowers of dioecious *Silene latifolia*. *Plant J* 12: 155-68. doi: 10.1046/j.1365-3113X.1997.12010155.x
- Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-40. doi: 10.1093/bioinformatics/btp616
- Roche BM, Alexander HM, Maltby AD. (1995) Dispersal and disease gradients of anther-smut infection of *Silene alba* at different life stages. *Ecology* 76: 1863-71. doi: 10.2307/1939343
- Roeijmans H, Prillinger H, Umile C, Sugiyama J, Nakase T, Boekhout T: **Analysis of carbohydrate composition of cell walls and extracellular carbohydrates.** In: *The Yeasts - A Taxonomic Study*. 1998. Chapter 13. Elsevier B.V., pp. 103-05
- Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S *et al.* (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun* 2: 202. doi: 10.1038/ncomms1189
- Ruddat M, Kokontis J, Birch L, Garber ED, Chiang KS, Campanella J, Dai H. (1991) Interactions of *Microbotryum violaceum* (*Ustilago violacea*) with its host plant *Silene alba*. *Plant Sci* 80: 157-65. doi: 10.1016/0168-9452(91)90280-L
- Sambrook J, Russell DW. 2001. *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press

- Schäfer AM, Kemler M, Bauer R, Begerow D. (2010) The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*. *Botany* 88: 875-85. doi: 10.1139/B10-061
- Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K, Doeblemann G, Di Stasio M, Rössel N, Mendoza-Mendoza A, Pester D *et al.* (2010) Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330: 1546-48. doi: 10.1126/science.1195330
- Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, Chellappan B, Boeren S, Takken FL, Rep M. (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* 14: 119. doi: 10.1186/1471-2164-14-119
- Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 7: 3. doi: 10.1186/1471-2199-7-3
- Scutt CP, Kamisugi Y, Sakai F, Gilmartin PM. (1997) Laser isolation of plant sex chromosomes: studies on the DNA composition of the X and Y sex chromosomes of *Silene latifolia*. *Genome* 40: 705-15. doi: 10.1139/g97-793
- Selker EU, Cambareri EB, Jensen BC, Haack KR. (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell* 51: 741-52. doi: 10.1016/0092-8674(87)90097-3
- Shykoff JA, Bucheli E. (1995) Pollinator visitation patterns, floral rewards and the probability of transmission of *Microbotryum violaceum*, a venereal disease of plants. *J Ecol* 83: 189-98. doi: 10.2307/2261557
- Shykoff JA, Kaltz O. (1997) Effects of the anther smut fungus *Microbotryum violaceum* on host life history patterns in *Silene latifolia* (Caryophyllaceae). *Int J Plant Sci* 158: 164-71.
- Shykoff JA, Kaltz O. (1998) Phenotypic changes in host plants diseased by *Microbotryum violaceum*: Parasite manipulation, side effects, and trade-offs. *Int J Plant Sci* 159: 236-43.
- Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38: D161-66. doi: 10.1093/nar/gkp885
- Sloan DB, Giraud T, Hood ME. (2008) Maximized virulence in a sterilizing pathogen: the anther-smut fungus and its co-evolved hosts. *J Evol Biol* 21: 1544-54. doi: 10.1111/j.1420-9101.2008.01604.x

- Smith J. (2011) Superinfection drives virulence evolution in experimental populations of bacteria and plasmids. *Evolution* 65: 831-41. doi: 10.1111/j.1558-5646.2010.01178.x
- Stajich JE, Wilke SK, Ahren D, Au CH, Birren BW, Borodovsky M, Burns C, Canback B, Casselton LA, Cheng CK *et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc Natl Acad Sci USA* 107: 11889-94. doi: 10.1073/pnas.1003391107
- Stanke M, Steinkamp R, Waack S, Morgenstern B. (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32: 309-12. doi: 10.1093/nar/gkh379
- Stergiopoulos I, de Wit PJ. (2009) Fungal effector proteins. *Annu Rev Phytopathol* 47: 233-63. doi: 10.1146/annurev.phyto.112408.132637
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545-50. doi: 10.1073/pnas.0506580102
- Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F. (2005) Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. *Plant Physiol* 139: 1649-65. doi: 10.1104/pp.105.070805
- Szabados L, Kovács I, Oberschall A, Ábrahám E, Kerekcs I, Zsigmond L, Nagy R, Alvarado M, Krasovskaja I, Gál M *et al.* (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J* 32: 233-42. doi: 10.1046/j.1365-313X.2002.01417.x
- Takahashi T, Kakehi J. (2010) Polyamines: ubiquitous polycations with unique roles in growth and stress responses. *Ann Bot* 105: 1-6. doi: 10.1093/aob/mcp259
- Thrall PH, Biere A, Antonovics J. (1993) Plant life-history and disease susceptibility - the occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *J Ecol* 81: 489-98. doi: 10.2307/2261527
- Toh SS, Chen Z, Cuomo C, Perlin MH. (In Preparation) *Pas de deux*: An intricate dance of anther smut and its host.
- Toh SS, Perlin MH. (2015) Size does matter: Staging of *Silene latifolia* floral buds for transcriptome studies. *Int J Mol Sci* 16: 22027-45. doi: 10.3390/ijms160922027

- Uchida W, Matsunaga S, Kawano S. (2005) Ultrastructural analysis of the behavior of the dimorphic fungus *Microbotryum violaceum* in fungus-induced anthers of female *Silene latifolia* flowers. *Protoplasma* 226: 207-16. doi: 10.1007/s00709-005-0113-7
- Uchida W, Matsunaga S, Sugiyama R, Kazama Y, Kawano S. (2003) Morphological development of anthers induced by the dimorphic smut fungus *Microbotryum violaceum* in female flowers of the dioecious plant *Silene latifolia*. *Planta* 218: 240-48. doi: 10.1007/s00425-003-1110-8
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35: W71-74. doi: 10.1093/nar/gkm306
- van Putten WF, Biere A, van Damme JM. (2005) Host-related genetic differentiation in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and allopatric populations of two host species *Silene latifolia* and *S. dioica*. *J Evol Biol* 18: 203-12. doi: 10.1111/j.1420-9101.2004.00771.x
- Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga A, de Vries RP *et al.* (2014) Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genet Biol* 72: 168-81. doi: 10.1016/j.fgb.2014.08.007
- Vincze T. (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31: 3688-91. doi: 10.1093/nar/gkg526
- Voegele RT, Struck C, Hahn M, Mendgen K. (2001) The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proc Natl Acad Sci USA* 98: 8133-38. doi: 10.1073/pnas.131186798
- Votintseva AA, Filatov DA. (2009) Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* 182: 1391-96. doi: 10.1534/genetics.109.103192
- Wahl R, Wippel K, Goos S, Kämper J, Sauer N. (2010) A novel high-affinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis*. *PLoS Biol* 8: e1000303. doi: 10.1371/journal.pbio.1000303
- Walser J-C, Furano AV. (2010) The mutational spectrum of non-CpG DNA varies with CpG content. *Genome Res* 20: 875-82. doi: 10.1101/gr.103283.109
- Walton FJ, Idnurm A, Heitman J. (2005) Novel gene functions required for melanization of the human pathogen *Cryptococcus neoformans*. *Mol Microbiol* 57: 1381-96. doi: 10.1111/j.1365-2958.2005.04779.x

- Wang Z, Gerstein M, Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63. doi: 10.1038/nrg2484
- Warmke HE. (1946) Sex determination and sex balance in *Melandrium*. *Am J Bot* 33: 648-60.
- Warmke HE, Blakeslee AF. (1939) Sex mechanism in polyploids of *Melandrium*. *Science* 89: 391-92. doi: 10.1126/science.89.2313.391
- Weiss M, Bauer R, Begerow D: **Spotlights on heterobasidiomycetes**. In: *Frontiers in Basidiomycote Mycology*. Edited by R. Agerer, Piepenbring M., Blanz P. 2004. IHW-Verlag, Echingen, Germany. pp. 7-48
- Westergaard M. (1946) Aberrant Y chromosomes and sex expression in *Melandrium album*. *Hereditas* 32: 419-43. doi: 10.1111/j.1601-5223.1946.tb02784.x
- Westergaard M. (1958) The mechanism of sex determination in dioecious flowering plants. *Adv Genet* 9: 217-81. doi: 10.1016/S0065-2660(08)60163-7
- Westermann AJ, Gorski SA, Vogel J. (2012) Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* 10: 618-30. doi: 10.1038/nrmicro2852
- Whittaker MM, Kersten PJ, Cullen D, Whittaker JW. (1999) Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. *J Biol Chem* 274: 36226-32. doi: 10.1074/jbc.274.51.36226
- Whittle CA, Johannesson H. (2011) Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of *Neurospora tetrasperma*. *Heredity* 107: 305-14. doi: 10.1038/hdy.2011.11
- Whittle CA, Sun Y, Johannesson H. (2011) Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*. *Eukaryot Cell* 10: 594-603. doi: 10.1128/EC.00284-10
- Whittle CA, Votintseva A, Ridout K, Filatov DA. (2015) Recent and massive expansion of the mating-type-specific region in the smut fungus *Microbotryum*. *Genetics* 199: 809-16. doi: 10.1534/genetics.114.171702
- Wilch G, Ward S, Castle A. (1992) Transmission of mitochondrial DNA in *Ustilago violacea*. *Curr Genomics* 22: 135-40. doi: 10.1007/BF00351473
- Winterhagen P, Dubois C, Sinn M, Wetzel T, Reustle GM. (2009) Gene silencing and virus resistance based on defective interfering constructs in transgenic *Nicotiana*

- benthamiana* is not linked to accumulation of siRNA. *Plant Physiol Biochem* 47: 739-42. doi: 10.1016/j.plaphy.2009.03.012
- Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, Kronstad JW, Deangelis YM, Reeder NL, Johnstone KR *et al.* (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci USA* 104: 18730-5. doi: 10.1073/pnas.0706756104
- Xu L, Petit E, Hood ME. (2015) Variation in mate-recognition pheromones of the fungal genus *Microbotryum*. *Heredity Advance Online*: doi: 10.1038/hdy.2015.68
- Xu L, Zhu L, Tu L, Liu L, Yuan D, Jin L, Long L, Zhang X. (2011) Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *J Exp Biol* 62: 5607-21. doi: 10.1093/jxb/err245
- Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, Rodolphe F, Devier B, Wincker P, Dossat C, Giraud T. (2007) Expressed sequence tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. *BMC Genomics* 8: 272. doi: 10.1186/1471-2164-8-272
- Zemach A, McDaniel IE, Silva P, Zilberman D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328: 916-19. doi: 10.1126/science.1186366
- Zemp N, Minder A, Widmer A. (2014) Identification of internal reference genes for gene expression normalization between the two sexes in dioecious White Campion. *PLoS ONE* 9: e92893. doi: 10.1371/journal.pone.0092893
- Zemp N, Tavares R, Widmer A. (2015) Fungal infection induces sex-specific transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene latifolia*. *PLoS Genet* 11: e1005536. doi: 10.1371/journal.pgen.1005536
- Zhao C, Waalwijk C, de Wit PJ, Tang D, van der Lee T. (2013) RNA-Seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. *BMC Genomics* 14: 21. doi: 10.1186/1471-2164-14-21
- Zhou LH, Lidén M, Oxelman B. *Silene*. In: efloras. Published on the Internet. 2008. <http://efloras.org>. Accessed 31 Oct 2015. Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA.
- Zluvova J, Zak J, Janoušek B, Vyskot B. (2010) Dioecious *Silene latifolia* plants show sexual dimorphism in the vegetative stage. *BMC Plant Biol* 10: 208. doi: 10.1186/1471-2229-10-208

Zuccaro A, Lahrmann U, Guldener U, Langen G, Pfiffi S, Biedenkopf D, Wong P, Samans B, Grimm C, Basiewicz M *et al.* (2011) Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathog* 7: e1002290. doi: 10.1371/journal.ppat.1002290

Zytnicki M, Quesneville H. (2011) S-MART, a software toolbox to aid RNA-Seq data analysis. *PLoS ONE* 6: e25988. doi: 10.1371/journal.pone.0025988

APPENDICES

Appendix 1: RNA Processing in the Goat Lab

Overview

RNA Extraction with RNeasy® Plant MiniKit

This protocol has been adapted from RNeasy Mini Handbook 06/2012

(Purification of Total RNA from Plant Cells and Tissues and Filamentous Fungi, Qiagen, cat no: 74904)

Materials

RNeasy Plant Mini Kit

RNeasy Mini Spin Columns (Pink)
QIAshredder Spin Columns (Lilac)
1.5-ml Collection Tubes
2-ml Collection Tubes
Buffer RLT
Buffer RLC
Buffer RW1
Buffer RPE* (concentrate)
RNase-Free Water

Reagent

95% ethanol
 β -mercaptoethanol
Liquid nitrogen

Apparatus

P1000, P200 and P20 micropipettors and tips
Microcentrifuge
2-ml microcentrifuge tubes
1.5-ml microcentrifuge tubes

Apparatus requiring baking at 290°C for at least four hours

Mortar and pestle (store at -80 °C until use)
Pre-cooled spatula (store at -80 °C until use)
50-ml measuring cylinders

* Add 4 volumes (44ml) of 95% ethanol as indicated on the bottle to obtain a working solution of Buffer RPE.

Procedure for RNA Extraction using RNeasy® Plant Mini Kit

Note: Centrifuging is performed at >8,000 x g (>10,000 rpm) unless stated otherwise.

1. Add 10 µl β-mercaptoethanol per 1 ml of Buffer RLT in the fume hood.

Note: Buffer RLT containing β-ME can be stored at room temperature (15–25 °C) for up to 1 month.

2. Scrap cells from plates using sterile plastic 10-µl inoculating loop/ transfer floral tissue in the pre-cooled mortar and pestle.
3. Grind sample with liquid nitrogen until fine powder is obtained.
4. Transfer the powder immediately using a pre-cooled spatula into a clean 2-ml microcentrifuge tube and fill only the conical bottom of the tube (~100mg).

Note: Do not allow tissue to thaw until Buffer RLT is added.

5. Add 450 µl of **Buffer RLT** immediately.
6. Vortex vigorously and incubate at ~56-65 °C for 3 min.

Note: Two samples can be processed in one 2-ml microcentrifuge tube step 4-6 if desired.

7. Transfer the lysate to a **QIAshredder spin column** (lilac) placed in a 2-ml collection tube, and centrifuge for 2 min at full speed.

Note: It may be necessary to use a wide bore tip to facilitate pipetting of the lysate into the column.

8. Transfer the supernatant of the flow-through to a new 2-ml microcentrifuge tube without disturbing the cell-debris pellet in the collection tube.

Note: The sample can be set aside for batch processing or be stored at -70 °C for several months.

9. Add 0.5 volume of 95% ethanol to the cleared lysate and mix immediately by pipetting. Transfer the sample (usually 650 µl) immediately, including any precipitate

that may have formed, to an **RNeasy spin column** (pink) placed in a 2-ml collection tube.

Note: Two samples can be processed in one 2-ml microcentrifuge tube step 8-9 if desired. If volume exceeds 700 µl, repeat the loading and discard the flow-through after each centrifugation.

10. Close the lid gently, and centrifuge for 15s and discard the flow-through.

11. Add 700 µl **Buffer RW1** to the column to wash the column. Close the lid gently, centrifuge for 15 s and discard the flow-through.

Note: After centrifugation, carefully remove the spin column from the collection tube so that the column does not contact the flow-through. Be sure to empty the collection tube completely.

12. Add 500 µl **Buffer RPE** to the RNeasy spin column. Close the lid gently, and centrifuge for 15s to wash the membrane and discard the flow-through.

13. Add 500 µl **Buffer RPE** to the RNeasy spin column. Close the lid gently, and centrifuge for 2 min to wash the membrane.

14. Transfer the **RNeasy spin column** to a new **2-ml collection tube** and discard the old collection tube with the flow-through. Centrifuge at full speed for 1 min.

15. Place the RNeasy spin column in a new **1.5-ml collection tube**.

16. Add 50 µl **RNase-free water** directly to the spin column membrane and close the lid.

17. Allow the water to sit on the column for 1 min and centrifuge for 1 min to elute the RNA.

18. Repeat Steps 16 and 17.

19. Transfer the eluent to a sterile 1.5-ml microcentrifuge tube for storage.

20. Assess concentration and quality of RNA using NanoDrop.

RNA Extraction with Direct-zol™ RNA MiniPrep Kit

This protocol has been adapted from the Instruction Manual Ver 1.1.2 available at <http://www.zymoresearch.com/downloads/dl/file/id/467/r2050i.pdf>. This method yields large quantity of high quality RNA efficiently and the on-column DNase treatment avoids a separate step that can potentially damage the RNA. While its quality has not been assessed for suitability for RNA-Sequencing, it is adequate for Northern Blot in our experiments.

Materials

Direct-zol™ RNA Miniprep Kit

Zymo-Spin™ IIC Columns

Collection Tubes

DNA Digestion Buffer

RNA PreWash¹ (concentrate)

RNA Wash Buffer² (concentrate)

DNase I³ (lyophilized)

RNase-Free Water

DNase I Reaction Mix (5 µl of DNase I (6 U/µl) + 75 µl of DNA Digestion Buffer)

Reagent

TRIzol®

95-100% ethanol

Liquid nitrogen

Apparatus

Vortex with tube holder

P1000, P200 and P20 micro-pipettors and tips

Microcentrifuge

2-ml microcentrifuge tubes

1.5-ml microcentrifuge tubes

Apparatus requiring baking at 290 °C for at least four hours

Mortar and pestle (store at -80 °C until use)

Pre-cooled spatula (store at -80 °C until use)

50-ml measuring cylinders

¹ Add 10 ml of 95-100% ethanol to the 40 ml concentrate, as indicated on the bottle to obtain a working solution of **RNA PreWash**.

² Add 48 ml of 100% ethanol (or 52 ml of 95% ethanol) to the 12 ml concentrate to obtain a working solution of **RNA Wash Buffer**.

³ Reconstitute the lyophilized DNase I as indicated on the vial prior to use and store in frozen aliquots if necessary.

Procedure for RNA Extraction using Direct-zol™ RNA

Note: Centrifuging is performed at $>12,000 \times g$ ($>14,000$ rpm) unless stated otherwise.

1. Scrap cells from plates using sterile plastic 10- μ l inoculating loop/ transfer floral tissue in the pre-cooled mortar and pestle.
2. Grind sample with liquid nitrogen until fine powder is obtained.
3. Transfer the powder immediately using a pre-cooled spatula into a clean 2-ml microcentrifuge tube and fill only the conical bottom of the tube (~100mg).

Note: Do not allow tissue to thaw until TRIzol® is added.

4. Add 500 μ l of **TRIzol®** immediately.
5. Vortex vigorously for 5-10 min.

Note: Vortex can be left to homogenize one sample while the next sample is being ground. After homogenization, the tubes can be set aside for batch processing.

6. Add one volume ethanol (500 μ l) to the sample homogenate and mix well by vortexing.
7. Load the mixture into a **Zymo-Spin™ IIC** Column in a **Collection Tube** (max 700 μ l) and centrifuge for 30 seconds. Discard the flow-through.

Note: Change the collection tube if not performing In-column DNase digestion.

8. Perform *In-column* DNaseI digestion by first washing the column with 400 μ l of **RNA Wash Buffer** and centrifuge for 30 seconds. Discard the flow-through.
9. Add 80 μ l of **DNase I Reaction Mix** directly to the column matrix.
10. Incubate the column at 20-30 °C for 15 min.

Add 400 μ l **RNA PreWash** to the column and centrifuge for 30 seconds. Discard the flow-through. Repeat this step.

11. Add 700 μ l **RNA Wash Buffer** to the column and centrifuge for 2 minutes to ensure complete removal of the wash buffer. Transfer the column carefully into a 1.5-ml RNase-free microcentrifuge tube.
12. Add 50 μ l of **DNase/RNase-Free Water** directly to the column matrix and centrifuge for 30 seconds. Elute a second time.
13. The eluted RNA can be used immediately or stored at -70 °C.
14. Assess concentration and quality of RNA using NanoDrop.

DNase Treatment with Turbo DNA-free™ Kit

This protocol has been adapted and modified from TURBO DNA-free™ Kit (Ambion, Cat No: AM1907) **Publication Number 1907M Revision G** by Su San Toh (15 Oct 2013). Please refer to original publication for more information.

Materials

TURBO DNA-free™ kit

TURBO DNase (2 units/μl)

10x TURBO DNase Buffer

DNase Inactivation Reagent

Nuclease-free water

Apparatus

0.5-ml PCR tubes

1.5-ml microcentrifuge tubes

P200, P20 and P10 micropipettes and tips

Thermocycler

Microcentrifuge

Procedure for DNase Treatment with TURBO DNA-free™ Kit

1. Adjust RNA sample to a concentration of 150-200 ng/μl in a volume rounded to the nearest 50 μl with the **10x TURBO DNase buffer** and **Nuclease-free water**.
2. Aliquot the sample into 100 μl aliquots into sterile 0.5-ml PCR tubes.
3. For each 100μl reaction, add 1 μl of **TURBO DNase** enzyme and incubate for 30 min at 37°C in the thermocycler.

Note: Do not vortex the enzyme. Use 0.5 μl in 50 μl reactions.

4. Add another 1 μl of **TURBO DNase** to each 100 μl reaction and incubate for another 30 min.
5. Add 0.1 volume of **DNase inactivation reagent** to each reaction and mix well.

Note: Resuspend the DNase inactivation reagent by flicking or vortexing well before use and in between dispensing.

6. Incubate at room temperature for 5 min and flick 2-3 times during the incubation period to redisperse the reagent.

7. Centrifuge at $10,000 \times g$ for 2 min and transfer the RNA carefully (without introducing the inactivating reagent into the sample) to a fresh 1.5-ml microcentrifuge tube for storage.
8. Assess concentration and quality with NanoDrop.

RNA Cleanup using Lithium Chloride Precipitation

Using lithium chloride for cleaning up RNA is not as well known as the use of sodium acetate and ethanol. LiCl appear to be a more efficient and effective way for purification of RNA. Technical bulletin from Life Technologies (<http://www.lifetechnologies.com/us/en/home/references/ambion-tech-support/rna-isolation/general-articles/the-use-of-licl-precipitation-for-rna-purification.html>) presents a comparison of the two methods and you are encouraged to read it.

Materials and Apparatus

RNA sample
DEPC treated water
RNase-free 10M LiCl
Microcentrifuge tubes
Micropipettes and tips
Refrigerated microcentrifuge

Procedure for LiCl Precipitation

1. Add one-third volume of 10M LiCl (or other available concentration) to the RNA sample to obtain a final concentration of 2.5M LiCl.

2. Incubate the sample overnight at 4 °C or -20 °C for 30 min.

Note: Note that sample will freeze if you incubate for too long at -20 °C.

3. Centrifuge the sample at 4 °C for 30 min at full speed in the microcentrifuge and decant the supernatant.

4. Wash the pellet twice with 500-700 µl of cold 70% ethanol and centrifuge for 15 min.

Note: Do not agitate or try to resuspend the pellet in between washes. Instead, face the hinge of the tube in opposite direction each time to force the pellet to migrate across the tube.

5. Air-dry the pellet completely.

6. Resuspend the pellet with RNase-free water and store at -80 °C.

Assessing RNA Concentration and Quality using NanoDrop

Assessment of RNA quality is based on $A_{260/280}$ and $A_{260/230}$ ratio. The ideal range is 2.0-2.2. Low $A_{260/280}$ ratio indicates possible protein contamination while low $A_{260/230}$ ratio may indicate organic contamination. LiCl cleanup may be able to resolve the issue.

Materials

NanoDrop 2000 UV-Vis Spectrophotometer
2 μ l of RNA sample
RNase free water used in the RNA sample
P10 micropipette and tips
Kimwipes

Procedure for RNA Assessment using NanoDrop

1. Remove the Kimwipe from the pedestal.
2. Clean the pedestal by loading 2 μ l of RNase-free water and wipe clean with Kimwipe immediately and replace the arm.
3. Open the software interface and select “Nuclei Acid”. Decide whether you would like to append the data to the existing worksheet.
4. Click “OK” to allow the software to perform routine.
5. Change the sample type to “RNA” on the interface.
6. Load 2 μ l of RNase-free water and click “Blank” on the interface.
7. Wipe off the water, load 2 μ l of sample and lower the arm.
8. Enter the Sample name and click “Measure”. The interface will prompt for a filename and location to save the file. Once done, the sample will be read.
9. Repeat step 7 and 8 until all samples have been read.
10. Clean as in step 2 twice and replace the Kimwipe on the pedestal.

cDNA Synthesis using SuperScript® III First-Strand Synthesis System

This protocol has been adapted from Pub. No. MAN0001346, Rev. 3.0 (Invitrogen, Cat no. 18080-051) and is specific for the reverse transcription of the mRNA from total RNA sample. The cDNA synthesized can determine if the RNA sample still has DNA contamination and can also be used for downstream qPCR analysis at 20x dilution. A no-enzyme control is necessary if the primers used do not span an intron that can differentiate DNA amplicons from cDNA amplicons.

Materials

SuperScript® III kit

50 µM Oligo(dT)₂₀
10x RT buffer
25 mM MgCl₂
0.1 M DTT
10 mM dNTP mix
200 U/µl SuperScript® III RT
40 U/µl RNaseOUT™
2 U/µl *E. coli* RNase H
DEPC-treated water

Apparatus

Speedvac
P200 and P10 Micropipettes and tips
0.5-ml PCR tubes
Thermocycler

Procedure for cDNA synthesis with SuperScript® III

1. Aliquot 5 µg of DNase-treated RNA sample into a fresh 0.5-ml PCR tube (**RT+**).
Prepare two tubes if a no-enzyme control (**RT-**) is required.
2. Dry the RNA sample in the speedvac on low or medium heat.
3. Resuspend each sample in 8 µl of water and 1 µl of Oligo (dT)₂₀. Prepare an empty tube with the same reagents, which will serve as **no template control**.
4. Incubate the samples at 65 °C for 5 min in the thermocycler and **immediately chill** on ice for at least 1 min.
5. Prepare cDNA synthesis mastermix as follows:

Mastermix	Vol per rxn (µl)
RT buffer	2 µl
25 mM MgCl ₂	4 µl
0.1M DTT	2 µl
RNaseOUT™ (40 U/µl)	1 µl
10 mM dNTP	1 µl
Total	10 µl

6. Add 10 µl of mastermix to each sample.
7. Add 1 µl of SSIII enzyme to each **RT+** sample and the **no template control**. Add 1 µl to each **RT-** sample.
8. Incubate samples at 50°C for 50 min, followed by 85 °C for 5 min.
9. Add 1 µl of RNase H to all samples and incubate at 37 °C for 20 min.
10. Store cDNA at -20 °C.

Assessment of RNA Quality Using Bioanalyser

This protocol only specifies the preparation of the sample for analysis by the UofL Bioanalyzer facility and not the use of the instrument. Bioanalyzer analysis is required if the integrity of the RNA is critical to your experiment. An example of such will be the RNASeq experiment. qPCR experiment generally does not require Bioanalyser analysis unless there is questions about why qPCR experiments fail.

There can be a maximum of 12 samples in each set of Bioanalyzer run and it cost the same even if you have less samples. Therefore, you should aim to fill the spaces as much as possible. Below is the information of the facility that performs the analysis.

Contact person: Venilla Arumugam/ Sabine Waigel
Contact number: 502-852-7991
Email add: v0arum01@louisville.edu/ sabine.waigel@louisville.edu
Location: CTR Building Room 227F
505 South Hancock Street

The samples should be between 25-500 ng/μl with a minimum volume of 2.5 μl. Prepare the samples in 0.5-ml PCR tubes and keep frozen at -80 °C until it is ready to be transported for analysis.

Prepare a spreadsheet with **sample name** that correspond to the label on the tube, approximate **concentration** based on NanoDrop and the **volume** provided. Include your name, contact number and email address in the spreadsheet you give to her.

Call or email Venilla to set up a time to bring the samples to her. One of the ways to transport the samples is to fill an empty P1000 tip box with ice and use the tip rack as a tube rack. You are responsible for maintaining the integrity of the samples until they get to her. She will email the results to you. It is advisable to send the sample for Bioanalyzer only after confirming that the sample is DNA-free.

Appendix 2: Ligation of pPZP-MV-HYG with Fusion Amplicon

Restriction enzyme digest is required to retrieve the fusion amplicon from the pCR®2.1-TOPO® TA vector and prepare the pPZP-MV-HYG for receiving the amplicon. *PacI* and *EcoRV* (New England Biolabs) are the restriction enzymes required. Since the two enzymes use different buffer, digest was performed with *PacI* for 6 hours and then held at 4 °C until the next enzyme can be added. Sodium chloride was then added to a final concentration of 100 mM (*i.e.* 0.8 µl of 5M NaCl) before adding 1.8 µl of *EcoRV*. The rest of the reaction was prepared as follows.

Sterile water	22 µl
10x Buffer	4 µl
BSA (100 µg/ml)	0.4 µl
DNA	10 µl
Restriction Enzyme(s)	1.8 + 1.8 µl
Total	40 µl

Upon purification from the gel, the plasmid and vector backbone are ready for ligation using Ligate-IT Rapid Ligation kit (Affymetrix). First, a 10 µl reaction was set up with 2 µl of cut and purified pPZP-MV-HYG vector and 6 µl of fusion amplicon cut from the pCR®2.1-TA cloning vector. 0.8 µl of T4 DNA ligase was used in each reaction. The reaction was incubated at room temperature for 10 min. 5 µl of the reaction mix was then used to transform into 25 µl of *Escherichia coli* One Shot® TOP10 competent cells (Invitrogen) according to manufacturer's instruction. The mixture was incubated for at least 3 hrs at 37 °C to allow enough time for replication of the larger plasmid. 120 µl of each transformation reaction was spread plated onto LB/KM plates and incubated overnight at 37 °C. A negative control without insert was included each time a new vial of vector was prepared to ensure that the vector was not self-ligating.

Appendix 3: Supplementary Tables and Figures

Table A3-1. Source of genome data for comparative analysis.

Species	Genes	Genbank Accession or version	Taxonomy
<i>Microbotryum violaceum</i> (Perlin et al., 2015)	7364	AEIJ01000000	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Microbotryomycetes; Microbotryales; Microbotryaceae; Microbotryum; Microbotryum violaceum
<i>Sporobolomyces roseus</i> (Grigoriev et al., 2012)	5536	v.1.0	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Microbotryomycetes; Sporidiobolales; mitosporic Sporidiobolales; Sporobolomyces; Sporobolomyces roseus
<i>Rhodotorula glutinis</i> (Paul et al., 2014)	2835	AEVR00000000	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Microbotryomycetes; Sporidiobolales; mitosporic Sporidiobolales; Rhodotorula; Rhodotorula glutinis
<i>Mixia osmundae</i> (Nishida et al., 2011)	6726	BABT00000000.2	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Mixiomycetes; Mixiales; Mixiaceae; Mixia; Mixia osmundae
<i>Melampsora larici-populina</i> (Duplessis et al., 2011)	16829	AECX00000000*	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Pucciniomycetes; Pucciniales; Melampsoraceae; Melampsora; Melampsora larici-populina
<i>Puccinia graminis f. sp. tritici</i> (Duplessis et al., 2011)	15800	AAWC01000000	Fungi; Dikarya; Basidiomycota; Pucciniomycotina; Pucciniomycetes; Pucciniales; Pucciniaceae; Puccinia; Puccinia graminis
<i>Coprinopsis cinereus</i> (Stajich et al., 2010)	13342	AACS02000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetidae; Agaricales; Psathyrellaceae; Coprinopsis; Coprinopsis cinerea
<i>Laccaria bicolor</i> (Martin et al., 2008)	18204	ABFE00000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetidae; Agaricales; Tricholomataceae; Laccaria; Laccaria bicolor
<i>Schizophyllum commune</i> (Ohm et al., 2010)	13190	ADMJ00000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetidae; Agaricales; Schizophyllaceae; Schizophyllum; Schizophyllum commune
<i>Phaenerochaete chrysosporium</i> (Martinez et al., 2004)	10048	AADS00000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetes incertae sedis; Corticiales; Corticiaceae; Phanerochaete; Phanerochaete chrysosporium
<i>Postia placenta</i> (Martinez et al., 2009)	9081	ABWF00000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetes incertae sedis; Polyporales; Coriolaceae; Postia; Postia placenta
<i>Piriformospora indica</i> (Zuccaro et al., 2011)	11766	CAFZ00000000	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetes incertae sedis; Sebacinales;

			Sebacinaceae; mitosporic Sebacinaceae; Piriformospora; Piriformospora indica
<i>Cryptococcus neoformans</i> var <i>neoformans</i> (Loftus et al., 2005)	6270	AE017341-AE017353,AE017356	Fungi; Dikarya; Basidiomycota; Agaricomycotina; Tremellomycetes; Tremellales; Tremellaceae; Filobasidiella; Filobasidiella/Cryptococcus neoformans species complex; Cryptococcus neoformans; Cryptococcus neoformans var. neoformans
<i>Ustilago maydis</i> (Kamper et al., 2006)	6522	AACP02000000	Fungi; Dikarya; Basidiomycota; Ustilaginomycotina; Ustilaginomycetes; Ustilaginales; Ustilaginaceae; Ustilago; Ustilago maydis
<i>Sporisorium reilianum</i> (Schirawski et al., 2010)	6673	GCA_000230245.1	Fungi; Dikarya; Basidiomycota; Ustilaginomycotina; Ustilaginomycetes; Ustilaginales; Ustilaginaceae; Sporisorium; Sporisorium reilianum
<i>Malassezia globosa</i> (Xu et al., 2007)	4286	AAYY01000001-AAYY01000067	Fungi; Dikarya; Basidiomycota; Ustilaginomycotina; Exobasidiomycetes; Malasseziales; Malasseziaceae; Malassezia; Malassezia globosa
<i>Aspergillus nidulans</i> (Cerqueira et al., 2014)	10717	s05-m01-r01_v2	Fungi; Dikarya; Ascomycota; Saccharomyceta; Pezizomycotina; leotiomyceta; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Trichocomaceae; Emericella
<i>Neurospora crassa</i> (Galagan et al., 2003)	9733	NC10	Fungi; Dikarya; Ascomycota; saccharomyceta; Pezizomycotina; Leotiomyceta; Sordariomyceta; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora; Neurospora crassa
<i>Saccharomyces cerevisiae</i> (Cherry et al., 2012)	6477	Downloaded March 5, 2008	Fungi; Dikarya; Ascomycota; Saccharomyceta; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces

Figure A3-1. Phylome for *Microbotryum lychnidis-dioicae*.

Maximum likelihood tree of 20 fungal species including *M. lychnidis-dioicae*, and other basidiomycete and ascomycete species. The tree has been rooted by the midpoint method for clarity. The alignments of 52 proteins with a single ortholog in each of the 20 studied species were concatenated into a single trimmed alignment of 48,538 positions. The phylome in this study was reconstructed using the pipeline described in (Huerta-Cepas et al., 2011) and can be browsed on-line (www.phylomedb.org, phylome code 180).

Table A3-2. CAZymes present in *M. lychnidis-dioicae*.

ProteinId	Description	Modelnotes	Defline
MVLG_00950	AA6		Benzoquinone reductase
MVLG_03384	CE16		Carbohydrate Esterase Family 16 protein
MVLG_06329	CE16		Carbohydrate Esterase Family 16 protein
MVLG_00132	CE16		Carbohydrate Esterase Family 16 protein
MVLG_00378	CE4		Carbohydrate Esterase Family 4 protein
MVLG_01114	CE4		Carbohydrate Esterase Family 4 protein
MVLG_03732	CE4		Carbohydrate Esterase Family 4 protein
MVLG_03734	CE4		Carbohydrate Esterase Family 4 protein
MVLG_03735	CE4		Carbohydrate Esterase Family 4 protein

MVLG_05107	CE4		Carbohydrate Esterase Family 4 protein
MVLG_06954	CE4		Carbohydrate Esterase Family 4 protein
MVLG_00379	CE4		Carbohydrate Esterase Family 4 protein
MVLG_00784	CE4		Carbohydrate Esterase Family 4 protein
MVLG_02681	CE8	fragment N-term;	Carbohydrate Esterase Family 8 protein
MVLG_02682	CE8		Carbohydrate Esterase Family 8 protein
MVLG_04072	CE8		Carbohydrate Esterase Family 8 protein
MVLG_04073	CE8		Carbohydrate Esterase Family 8 protein
MVLG_04074	CE8	fragment C-term;	Carbohydrate Esterase Family 8 protein
MVLG_04075	CE8	fragment N-term;	Carbohydrate Esterase Family 8 protein
MVLG_03197	CBM21		Carbohydrate-Binding Module Family 21 protein
MVLG_03197	CBM21		Carbohydrate-Binding Module Family 21 protein
MVLG_05984	CBM48-GH13_8		Carbohydrate-Binding Module Family 48 / Glycoside Hydrolase Family 13 protein
MVLG_00623	CBM48		Carbohydrate-Binding Module Family 48 protein
MVLG_07075	CBM48		Carbohydrate-Binding Module Family 48 protein
MVLG_02917	CBM50		Carbohydrate-Binding Module Family 50 protein
MVLG_05819	CBM50		Carbohydrate-Binding Module Family 50 protein
MVLG_03568	AA2_cyt		Class II peroxidase
MVLG_02439	AA2		Class II peroxidase
MVLG_03774	AA5_1		Copper radical oxidase
MVLG_05878	AA5_1		Copper radical oxidase
MVLG_06807	AA5_1		Copper radical oxidase
MVLG_06808	AA5_1	fragment N-term;	Copper radical oxidase
MVLG_06808	AA5_1		Copper radical oxidase
MVLG_00829	AA5_1	fragment N-term/C-term ;	Copper radical oxidase
MVLG_01824	AA5_1	splicing problem	Copper radical oxidase
MVLG_03027	AA5_1		Copper radical oxidase
MVLG_00083	EXPAN		Distantly related to plant expansins
MVLG_00085	EXPAN-EXPAN		Distantly related to plant expansins
MVLG_02276	EXPAN-EXPAN		Distantly related to plant expansins
MVLG_02293	EXPAN		Distantly related to plant expansins
MVLG_02726	EXPAN		Distantly related to plant expansins
MVLG_02726	EXPAN		Distantly related to plant expansins
MVLG_04138	EXPAN		Distantly related to plant expansins
MVLG_04404	EXPAN		Distantly related to plant expansins
MVLG_05825	EXPAN		Distantly related to plant expansins
MVLG_06891	EXPAN		Distantly related to plant expansins
MVLG_02640	AA3_3		GMC oxidoreductase

MVLG_03692	AA3_2		GMC oxidoreductase
MVLG_06036	AA3_2		GMC oxidoreductase
MVLG_02641	AA3_3	fragment C-term;	GMC oxidoreductase
MVLG_02642	AA3_3	fragment N-term;	GMC oxidoreductase
MVLG_03242	AA7		Glucosyltransferase
MVLG_03711	AA7		Glucosyltransferase
MVLG_03243	AA7		Glucosyltransferase
MVLG_07092	GH1		Glycoside Hydrolase Family 1 protein
MVLG_04582	GH125	splicing problem	Glycoside Hydrolase Family 125 protein
MVLG_05595	GH128		Glycoside Hydrolase Family 128 protein
MVLG_02767	GH13_25- GH133		Glycoside Hydrolase Family 13 / Glycoside Hydrolase Family 133 protein
MVLG_02969	GH13_5		Glycoside Hydrolase Family 13 protein
MVLG_03286	GH13_40		Glycoside Hydrolase Family 13 protein
MVLG_04827	GH13_40		Glycoside Hydrolase Family 13 protein
MVLG_05424	GH13_40		Glycoside Hydrolase Family 13 protein
MVLG_05598	GH13_40		Glycoside Hydrolase Family 13 protein
MVLG_05598	GH13_40		Glycoside Hydrolase Family 13 protein
MVLG_02179	GH15	fragment N-term;	Glycoside Hydrolase Family 15 protein
MVLG_02179	GH15		Glycoside Hydrolase Family 15 protein
MVLG_05216	GH15		Glycoside Hydrolase Family 15 protein
MVLG_01046	GH16		Glycoside Hydrolase Family 16 protein
MVLG_05172	GH16		Glycoside Hydrolase Family 16 protein
MVLG_06424	GH16	fragment N-term;	Glycoside Hydrolase Family 16 protein
MVLG_06451	GH17		Glycoside Hydrolase Family 17 protein
MVLG_00243	GH17		Glycoside Hydrolase Family 17 protein
MVLG_07306	GH17	short fragment	Glycoside Hydrolase Family 17 protein
MVLG_00826	GH18		Glycoside Hydrolase Family 18 protein
MVLG_00887	GH18		Glycoside Hydrolase Family 18 protein
MVLG_01845	GH18	too long at N-term;	Glycoside Hydrolase Family 18 protein
MVLG_04090	GH18		Glycoside Hydrolase Family 18 protein
MVLG_05737	GH18		Glycoside Hydrolase Family 18 protein
MVLG_01897	GH18		Glycoside Hydrolase Family 18 protein
MVLG_04748	GH18	short fragment	Glycoside Hydrolase Family 18 protein
MVLG_05728	GH18	fragment N-term/C-term;	Glycoside Hydrolase Family 18 protein
MVLG_05728	GH18	fragment N-term/C-term;	Glycoside Hydrolase Family 18 protein
MVLG_07093	GH2		Glycoside Hydrolase Family 2 protein
MVLG_00002	GH20		Glycoside Hydrolase Family 20 protein
MVLG_03477	GH20		Glycoside Hydrolase Family 20 protein
MVLG_01713	GH23		Glycoside Hydrolase Family 23 protein
MVLG_01823	GH23		Glycoside Hydrolase Family 23 protein
MVLG_00042	GH26	too long at N-term;	Glycoside Hydrolase Family 26 protein
MVLG_00478	GH26		Glycoside Hydrolase Family 26 protein

MVLG_01285	GH26	too long at N-term;	Glycoside Hydrolase Family 26 protein
MVLG_00477	GH26		Glycoside Hydrolase Family 26 protein
MVLG_00326	GH28		Glycoside Hydrolase Family 28 protein
MVLG_03747	GH28		Glycoside Hydrolase Family 28 protein
MVLG_05108	GH28		Glycoside Hydrolase Family 28 protein
MVLG_02498	GH28	too long at N-term;	Glycoside Hydrolase Family 28 protein
MVLG_02968	GH28	fragment N-term;	Glycoside Hydrolase Family 28 protein
MVLG_06986	GH28	fragment N-term/C-term;	Glycoside Hydrolase Family 28 protein
MVLG_03749	GH3	fragment N-term/C-term;	Glycoside Hydrolase Family 3 protein
MVLG_05128	GH3		Glycoside Hydrolase Family 3 protein
MVLG_06984	GH3	short fragment	Glycoside Hydrolase Family 3 protein
MVLG_07032	GH3	fragment N-term;	Glycoside Hydrolase Family 3 protein
MVLG_00506	GH31		Glycoside Hydrolase Family 31 protein
MVLG_03097	GH31		Glycoside Hydrolase Family 31 protein
MVLG_05713	GH36		Glycoside Hydrolase Family 36 protein
MVLG_06060	GH36		Glycoside Hydrolase Family 36 protein
MVLG_04685	GH37		Glycoside Hydrolase Family 37 protein
MVLG_03104	GH38		Glycoside Hydrolase Family 38 protein
MVLG_03130	GH43		Glycoside Hydrolase Family 43 protein
MVLG_00013	GH47		Glycoside Hydrolase Family 47 protein
MVLG_00014	GH47		Glycoside Hydrolase Family 47 protein
MVLG_02500	GH47		Glycoside Hydrolase Family 47 protein
MVLG_00414	GH5		Glycoside Hydrolase Family 5 protein
MVLG_00809	GH5_9		Glycoside Hydrolase Family 5 protein
MVLG_00830	GH5_dist		Glycoside Hydrolase Family 5 protein
MVLG_01159	GH5_dist		Glycoside Hydrolase Family 5 protein
MVLG_01191	GH5		Glycoside Hydrolase Family 5 protein
MVLG_01192	GH5		Glycoside Hydrolase Family 5 protein
MVLG_02512	GH5		Glycoside Hydrolase Family 5 protein
MVLG_02704	GH5_dist		Glycoside Hydrolase Family 5 protein
MVLG_03124	GH5_9		Glycoside Hydrolase Family 5 protein
MVLG_03400	GH5_9		Glycoside Hydrolase Family 5 protein
MVLG_04497	GH5_12		Glycoside Hydrolase Family 5 protein
MVLG_04497	GH5_12		Glycoside Hydrolase Family 5 protein
MVLG_05052	GH5		Glycoside Hydrolase Family 5 protein
MVLG_05082	GH5_dist	fragment N-term;	Glycoside Hydrolase Family 5 protein
MVLG_05520	GH5		Glycoside Hydrolase Family 5 protein
MVLG_05754	GH5_9		Glycoside Hydrolase Family 5 protein
MVLG_05951	GH5_9		Glycoside Hydrolase Family 5 protein
MVLG_06689	GH5_12		Glycoside Hydrolase Family 5 protein
MVLG_06542	GH5	fragment C-term;	Glycoside Hydrolase Family 5 protein
MVLG_00906	GH63		Glycoside Hydrolase Family 63 protein
MVLG_00351	GH63	fragment C-term;	Glycoside Hydrolase Family 63 protein

MVLG_00352	GH63	fragment N-term;	Glycoside Hydrolase Family 63 protein
MVLG_05742	GH76		Glycoside Hydrolase Family 76 protein
MVLG_01911	GH85		Glycoside Hydrolase Family 85 protein
MVLG_01861	GT1		Glycosyltransferase Family 1 protein
MVLG_03868	GT1		Glycosyltransferase Family 1 protein
MVLG_05634	GT1		Glycosyltransferase Family 1 protein
MVLG_05795	GT10		Glycosyltransferase Family 10 protein
MVLG_05586	GT15		Glycosyltransferase Family 15 protein
MVLG_06423	GT15		Glycosyltransferase Family 15 protein
MVLG_06821	GT15		Glycosyltransferase Family 15 protein
MVLG_01701	GT17		Glycosyltransferase Family 17 protein
MVLG_01343	GT2-GT4		Glycosyltransferase Family 2 / Glycosyltransferase Family 4 protein
MVLG_04819	GT2-GT4		Glycosyltransferase Family 2 / Glycosyltransferase Family 4 protein
MVLG_00097	GT2		Glycosyltransferase Family 2 protein
MVLG_00449	GT2		Glycosyltransferase Family 2 protein
MVLG_00546	GT2		Glycosyltransferase Family 2 protein
MVLG_01319	GT2		Glycosyltransferase Family 2 protein
MVLG_01840	GT2		Glycosyltransferase Family 2 protein
MVLG_01926	GT2		Glycosyltransferase Family 2 protein
MVLG_02107	GT2		Glycosyltransferase Family 2 protein
MVLG_04049	Myosin_motor-GT2		Glycosyltransferase Family 2 protein
MVLG_04136	GT2		Glycosyltransferase Family 2 protein
MVLG_04450	GT2		Glycosyltransferase Family 2 protein
MVLG_04926	GT2		Glycosyltransferase Family 2 protein
MVLG_05947	Myosin_motor-GT2		Glycosyltransferase Family 2 protein
MVLG_06472	GT2		Glycosyltransferase Family 2 protein
MVLG_07193	GT2		Glycosyltransferase Family 2 protein
MVLG_01341	GT2		Glycosyltransferase Family 2 protein
MVLG_03832	GT2		Glycosyltransferase Family 2 protein
MVLG_01977	GT20		Glycosyltransferase Family 20 protein
MVLG_05573	GT20		Glycosyltransferase Family 20 protein
MVLG_01977	GT20	fragment N-term;	Glycosyltransferase Family 20 protein
MVLG_01704	GT21		Glycosyltransferase Family 21 protein
MVLG_02344	GT22	fragment N-term;	Glycosyltransferase Family 22 protein
MVLG_04246	GT22		Glycosyltransferase Family 22 protein
MVLG_04526	GT22	fragment C-term;	Glycosyltransferase Family 22 protein
MVLG_01438	GT24		Glycosyltransferase Family 24 protein
MVLG_01580	GT3		Glycosyltransferase Family 3 protein
MVLG_02438	GT31-GT32		Glycosyltransferase Family 31 / Glycosyltransferase Family 32 protein
MVLG_06385	GT31		Glycosyltransferase Family 31 protein
MVLG_06558	GT31		Glycosyltransferase Family 31 protein

MVLG_06827	GT31		Glycosyltransferase Family 31 protein
MVLG_06827	GT31		Glycosyltransferase Family 31 protein
MVLG_00142	GT32		Glycosyltransferase Family 32 protein
MVLG_03406	GT32		Glycosyltransferase Family 32 protein
MVLG_06963	GT32		Glycosyltransferase Family 32 protein
MVLG_03760	GT32	fragment C-term;	Glycosyltransferase Family 32 protein
MVLG_03816	GT32	fragment C-term;	Glycosyltransferase Family 32 protein
MVLG_05524	GT33		Glycosyltransferase Family 33 protein
MVLG_00751	GT35		Glycosyltransferase Family 35 protein
MVLG_02655	GT39		Glycosyltransferase Family 39 protein
MVLG_03404	GT39		Glycosyltransferase Family 39 protein
MVLG_02340	GT39		Glycosyltransferase Family 39 protein
MVLG_02106	GT4-GT2- GT8		Glycosyltransferase Family 4 / Glycosyltransferase Family 2 / Glycosyltransferase Family 8 protein
MVLG_02183	GT4		Glycosyltransferase Family 4 protein
MVLG_02876	GT4		Glycosyltransferase Family 4 protein
MVLG_03485	GT4		Glycosyltransferase Family 4 protein
MVLG_04819	GT4		Glycosyltransferase Family 4 protein
MVLG_05791	GT4		Glycosyltransferase Family 4 protein
MVLG_04460	GT41		Glycosyltransferase Family 41 protein
MVLG_01173	GT47		Glycosyltransferase Family 47 protein
MVLG_01603	GT47		Glycosyltransferase Family 47 protein
MVLG_03052	GT47		Glycosyltransferase Family 47 protein
MVLG_01218	GT48		Glycosyltransferase Family 48 protein
MVLG_04235	GT49		Glycosyltransferase Family 49 protein
MVLG_00765	GT49	fragment C-term;	Glycosyltransferase Family 49 protein
MVLG_04693	GT5		Glycosyltransferase Family 5 protein
MVLG_04156	GT50		Glycosyltransferase Family 50 protein
MVLG_00632	GT57		Glycosyltransferase Family 57 protein
MVLG_01325	GT57		Glycosyltransferase Family 57 protein
MVLG_06325	GT58		Glycosyltransferase Family 58 protein
MVLG_00779	GT59		Glycosyltransferase Family 59 protein
MVLG_02942	GT62		Glycosyltransferase Family 62 protein
MVLG_04399	GT62		Glycosyltransferase Family 62 protein
MVLG_06154	GT62		Glycosyltransferase Family 62 protein
MVLG_04420	GT64		Glycosyltransferase Family 64 protein
MVLG_02191	GT66		Glycosyltransferase Family 66 protein
MVLG_00490	GT69		Glycosyltransferase Family 69 protein
MVLG_03450	GT69		Glycosyltransferase Family 69 protein
MVLG_04576	GT71		Glycosyltransferase Family 71 protein
MVLG_06196	GT71		Glycosyltransferase Family 71 protein
MVLG_06114	GT76		Glycosyltransferase Family 76 protein
MVLG_00823	GT8		Glycosyltransferase Family 8 protein
MVLG_00849	GT8		Glycosyltransferase Family 8 protein

MVLG_00562	GT90		Glycosyltransferase Family 90 protein
MVLG_00571	GT90		Glycosyltransferase Family 90 protein
MVLG_00599	GT90		Glycosyltransferase Family 90 protein
MVLG_00627	GT90		Glycosyltransferase Family 90 protein
MVLG_04641	GT90		Glycosyltransferase Family 90 protein
MVLG_06868	GT90		Glycosyltransferase Family 90 protein
MVLG_06871	GT90		Glycosyltransferase Family 90 protein
MVLG_06873	GT90	fragment C-term;	Glycosyltransferase Family 90 protein
MVLG_06999	GT90		Glycosyltransferase Family 90 protein
MVLG_07257	GT90	short fragment	Glycosyltransferase Family 90 protein
MVLG_00047	AA11		Lytic polysaccharide monooxygenase
MVLG_01868	AA1_2		Multicopper oxidase
MVLG_02184	AA1		Multicopper oxidase
MVLG_00670	AA1		Multicopper oxidase
MVLG_03092	AA1_dist		Multicopper oxidase
MVLG_04973	PL14_3		Polysaccharide Lyase Family 14 protein
MVLG_01770	PL14		Polysaccharide Lyase Family 14 protein

Table A3-3. Predicted secreted proteins.

Gene	PFAM/TIGRfam	GO
MVLG_00002	(PF00728.15,"Glycosyl hydrolase family 20, catalytic domain"), (PF02838.8,"Glycosyl hydrolase family 20, domain 2")	none
MVLG_00033	none	none
MVLG_00047	none	none
MVLG_00083	(PF00967.10,Barwin family), (PF03330.11,Rare lipoprotein A (RlpA)-like double-psi beta-barrel)	none
MVLG_00132	(PF00657.15,GDSL-like Lipase/Acylhydrolase)	(GO:0003824,lipolytic protein g-d-s-l family)
MVLG_00174	(PF05922.9,Peptidase inhibitor I9), (PF00082.15,Subtilase family)	none
MVLG_00179	(PF02128.8,Fungalysin metallopeptidase (M36))	none
MVLG_00210	none	(GO:0005737,ubiquitin ligase subunit), (GO:0019788,ubiquitin ligase subunit), (GO:0045116,ubiquitin ligase subunit), (GO:0019941,ubiquitin ligase subunit), (GO:0005515,ubiquitin ligase subunit), (GO:0006950,ubiquitin ligase subunit), (GO:0005634,ubiquitin
MVLG_00243	none	none
MVLG_00257	none	none
MVLG_00258	none	none
MVLG_00259	none	none

MVLG_00260	none	none
MVLG_00290	none	none
MVLG_00326	(PF00295.10,Glycosyl hydrolases family 28), (TIGR03804,para_beta_helix: parallel beta-helix repeat)	none
MVLG_00369	none	none
MVLG_00385	none	none
MVLG_00398	none	none
MVLG_00419	(PF08737.3,Rgp1)	none
MVLG_00443	none	none
MVLG_00477	none	none
MVLG_00478	none	none
MVLG_00497	none	none
MVLG_00525	(PF06280.5,Fn3-like domain (DUF1034)), (PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_00533	(PF04389.10,Peptidase family M28)	none
MVLG_00551	(PF08520.3,Fungal protein of unknown function (DUF1748))	none
MVLG_00566	none	none
MVLG_00572	none	none
MVLG_00601	none	none
MVLG_00647	(PF06280.5,Fn3-like domain (DUF1034)), (PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_00648	(PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_00677	none	none
MVLG_00690	none	none
MVLG_00784	none	none
MVLG_00789	none	none
MVLG_00810	(PF09423.3,PhoD-like phosphatase)	none
MVLG_00826	(PF00704.21,Glycosyl hydrolases family 18)	none
MVLG_00885	none	none
MVLG_00887	(PF00704.21,Glycosyl hydrolases family 18)	none
MVLG_00906	(PF06202.7,"Amylo-alpha-1,6-glucosidase"), (PF03200.9,Mannosyl oligosaccharide glucosidase), (PF01204.11,Trehalase)	(GO:0003824,glucosidase i)
MVLG_00910	none	none
MVLG_00914	(PF03583.7,Secretory lipase)	none
MVLG_00933	(PF03583.7,Secretory lipase)	none
MVLG_00965	(PF03029.10,Conserved hypothetical ATP binding protein), (PF02373.15,JmjC domain)	none

MVLG_00974	none	none
MVLG_01005	none	none
MVLG_01053	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_01089	none	none
MVLG_01138	none	none
MVLG_01142	none	none
MVLG_01159	none	none
MVLG_01175	none	none
MVLG_01191	none	none
MVLG_01192	none	none
MVLG_01254	none	none
MVLG_01258	none	none
MVLG_01263	(PF03097.11,BRO1-like domain)	none
MVLG_01284	none	none
MVLG_01294	(PF02585.10,GlcNAc-PI de-N-acetylase)	none
MVLG_01310	none	none
MVLG_01427	none	(GO:0009987,tenascin xb), (GO:0008152,tenascin xb)
MVLG_01474	(PF06516.4,Purine nucleoside permease (NUP))	none
MVLG_01520	(PF06280.5,Fn3-like domain (DUF1034)), (PF02225.15,PA domain)	none
MVLG_01581	(PF04389.10,Peptidase family M28)	none
MVLG_01590	none	none
MVLG_01591	none	none
MVLG_01603	(PF03016.8,Exostosin family)	none
MVLG_01651	none	none
MVLG_01652	none	none
MVLG_01666	(PF00450.15,Serine carboxypeptidase)	none
MVLG_01690	(PF05730.4,CFEM domain)	none
MVLG_01713	(PF01464.13,Transglycosylase SLT domain)	none
MVLG_01717	(PF01556.11,DnaJ C terminal region), (PF00684.12,DnaJ central domain (4 repeats)), (PF00226.24,DnaJ domain)	(GO:0005488,chaperone regulator)
MVLG_01732	none	none
MVLG_01770	none	none
MVLG_01823	(PF01464.13,Transglycosylase SLT domain)	none
MVLG_01824	(PF09118.4,Domain of unknown function (DUF1929)), (PF07250.4,Glyoxal oxidase N-terminus)	none
MVLG_01844	none	none
MVLG_01845	(PF00704.21,Glycosyl hydrolases family 18)	(GO:0016787,glycoside hydrolase family 18 protein)
MVLG_01849	none	none

MVLG_01897	(PF00704.21,Glycosyl hydrolases family 18)	none
MVLG_01951	(PF10282.2,"3-carboxy-cis,cis-muconate lactonizing enzyme")	none
MVLG_01957	none	none
MVLG_01958	none	none
MVLG_02018	none	none
MVLG_02043	(TIGR00066,g_glut_trans: gamma-glutamyltransferase), (PF01019.14,Gamma-glutamyltranspeptidase)	none
MVLG_02071	none	none
MVLG_02083	(PF06280.5,Fn3-like domain (DUF1034)), (PF00082.15,Subtilase family)	none
MVLG_02088	(PF00328.15,Histidine acid phosphatase)	none
MVLG_02171	none	none
MVLG_02173	none	none
MVLG_02184	(PF07731.7,Multicopper oxidase)	none
MVLG_02245	none	none
MVLG_02276	(PF03330.11,Rare lipoprotein A (RlpA)-like double-psi beta-barrel)	none
MVLG_02283	(PF00188.19,Cysteine-rich secretory protein family)	none
MVLG_02288	(PF00188.19,Cysteine-rich secretory protein family)	none
MVLG_02293	(PF03330.11,Rare lipoprotein A (RlpA)-like double-psi beta-barrel)	none
MVLG_02306	none	none
MVLG_02331	(PF00082.15,Subtilase family)	none
MVLG_02393	(PF01764.18,Lipase (class 3))	(GO:0016236,lipase), (GO:0005768,lipase), (GO:0030437,lipase), (GO:0044446,lipase), (GO:0031090,lipase), (GO:0006995,lipase), (GO:0005783,lipase)
MVLG_02420	(PF00890.17,FAD binding domain)	none
MVLG_02455	(PF07749.5,"Endoplasmic reticulum protein ERp29, C-terminal domain"), (TIGR01126,pdi_dom: protein disulfide-isomerase domain), (PF00085.13,Thioredoxin)	none
MVLG_02469	(PF00328.15,Histidine acid phosphatase)	none
MVLG_02471	(PF00328.15,Histidine acid phosphatase)	none
MVLG_02500	(PF01532.13,Glycosyl hydrolase family 47), (PF02225.15,PA domain)	none
MVLG_02513	none	none
MVLG_02540	none	none
MVLG_02588	(PF07992.7,Pyridine nucleotide-disulphide oxidoreductase)	none
MVLG_02597	(PF07859.6,alpha/beta hydrolase fold), (PF00149.21,Calcineurin-like phosphoesterase), (PF07819.6,PGAP1-	(GO:0003824,alpha beta fold family)

	like protein)	
MVLG_02675	none	none
MVLG_02676	none	none
MVLG_02682	(PF01095.12,Pectinesterase)	none
MVLG_02726	(PF03330.11,Rare lipoprotein A (RlpA)-like double-psi beta-barrel)	none
MVLG_02749	(PF00245.13,Alkaline phosphatase), (PF00884.16,Sulfatase), (PF01663.15,Type I phosphodiesterase / nucleotide pyrophosphatase)	none
MVLG_02763	(PF06280.5,Fn3-like domain (DUF1034)), (PF00082.15,Subtilase family)	none
MVLG_02806	(PF00447.10,HSF-type DNA-binding)	none
MVLG_02872	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_02877	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_02882	(PF01368.13,DHH family), (PF02833.7,DHHA2 domain)	none
MVLG_02890	none	none
MVLG_02900	none	none
MVLG_03027	(PF09118.4,Domain of unknown function (DUF1929)), (PF07250.4,Glyoxal oxidase N-terminus)	none
MVLG_03058	(PF00884.16,Sulfatase)	(GO:0003824,arylsulfatase)
MVLG_03089	(PF01328.10,"Peroxidase, family 2")	none
MVLG_03092	(PF07731.7,Multicopper oxidase)	none
MVLG_03160	(PF02265.9,S1/P1 Nuclease)	none
MVLG_03242	(PF08031.5,Berberine and berberine like), (PF01565.16,FAD binding domain)	none
MVLG_03243	(PF08031.5,Berberine and berberine like), (PF01565.16,FAD binding domain)	none
MVLG_03292	(PF05572.6,Pregnancy-associated plasma protein-A)	none
MVLG_03293	(PF05572.6,Pregnancy-associated plasma protein-A)	none
MVLG_03351	(PF10681.2,"Chaperone for protein-folding within the ER, fungal")	(GO:0044464,rot1 protein)
MVLG_03356	none	none
MVLG_03367	none	none
MVLG_03368	none	none
MVLG_03398	none	none
MVLG_03457	none	none
MVLG_03478	(PF00168.23,C2 domain)	none
MVLG_03524	(PF00069.18,Protein kinase domain), (PF07714.10,Protein tyrosine kinase)	(GO:0005488,clathrin-coated vesicle protein)
MVLG_03562	none	none
MVLG_03638	none	none
MVLG_03674	none	none

MVLG_03707	(PF07470.6,Glycosyl Hydrolase Family 88)	none
MVLG_03715	none	none
MVLG_03732	(PF01522.14,Polysaccharide deacetylase)	(GO:0003824,family 4 carbohydrate esterase)
MVLG_03734	(PF01522.14,Polysaccharide deacetylase)	(GO:0003824,family 4 carbohydrate esterase)
MVLG_03737	none	none
MVLG_03747	(PF00295.10,Glycosyl hydrolases family 28)	none
MVLG_03832	(PF01697.20,Domain of unknown function)	none
MVLG_03888	none	none
MVLG_03903	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_03919	(PF00160.14,Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD)	none
MVLG_03923	(PF02221.8,ML domain)	none
MVLG_03931	(PF01328.10,"Peroxidase, family 2")	none
MVLG_03945	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_03994	none	none
MVLG_03995	none	none
MVLG_04028	none	none
MVLG_04072	(PF01095.12,Pectinesterase)	none
MVLG_04073	(PF01095.12,Pectinesterase)	none
MVLG_04074	none	none
MVLG_04083	(PF06427.4,UDP-glucose:Glycoprotein Glucosyltransferase)	none
MVLG_04096	none	none
MVLG_04105	none	none
MVLG_04106	none	none
MVLG_04107	none	none
MVLG_04112	(PF00226.24,DnaJ domain)	none
MVLG_04167	none	none
MVLG_04168	none	none
MVLG_04274	none	none
MVLG_04353	(PF07519.4,Tannase and feruloyl esterase)	none
MVLG_04366	(PF05577.5,Serine carboxypeptidase S28)	none
MVLG_04367	(PF05577.5,Serine carboxypeptidase S28)	none
MVLG_04379	none	none
MVLG_04381	none	none
MVLG_04415	none	none
MVLG_04442	(PF00450.15,Serine carboxypeptidase)	none
MVLG_04463	(PF00646.26,F-box domain)	(GO:0044424,f-box and wd40 domain)
MVLG_04465	none	none
MVLG_04474	none	none

MVLG_04505	none	none
MVLG_04547	none	none
MVLG_04566	(PF00775.14,Dioxygenase)	none
MVLG_04780	(PF00117.21,Glutamine amidotransferase class-I)	none
MVLG_04782	(PF01546.21,Peptidase family M20/M25/M40), (PF04389.10,Peptidase family M28)	none
MVLG_04783	none	none
MVLG_04796	none	none
MVLG_04797	none	none
MVLG_04806	none	none
MVLG_04864	none	none
MVLG_04910	(PF00248.14,Aldo/keto reductase family)	none
MVLG_04923	(PF07217.4,Heterokaryon incompatibility protein Het-C)	none
MVLG_04962	none	none
MVLG_05017	(PF08760.4,Domain of unknown function (DUF1793))	none
MVLG_05034	none	none
MVLG_05052	none	(GO:0003824,mannan endo- -beta-mannosidase)
MVLG_05060	(PF00450.15,Serine carboxypeptidase)	none
MVLG_05061	(PF00226.24,DnaJ domain)	none
MVLG_05108	(PF00295.10,Glycosyl hydrolases family 28), (TIGR03804,para_beta_helix: parallel beta-helix repeat)	none
MVLG_05119	none	none
MVLG_05120	none	none
MVLG_05122	none	none
MVLG_05125	none	none
MVLG_05170	none	none
MVLG_05176	(PF00271.24,Helicase conserved C-terminal domain), (PF00176.16,SNF2 family N-terminal domain), (PF00097.18,"Zinc finger, C3HC4 type (RING finger)")	(GO:0005488,snf2 family helicase)
MVLG_05203	none	none
MVLG_05275	none	none
MVLG_05317	(PF00135.21,Carboxylesterase)	none
MVLG_05376	none	none
MVLG_05398	none	none
MVLG_05416	none	none
MVLG_05513	none	none
MVLG_05514	none	none
MVLG_05520	none	(GO:0003824,fibronectin type iii domain-

		containing protein)
MVLG_05522	none	none
MVLG_05525	none	none
MVLG_05528	none	none
MVLG_05533	none	none
MVLG_05534	none	none
MVLG_05538	none	none
MVLG_05549	(PF03583.7,Secretory lipase)	none
MVLG_05557	none	none
MVLG_05565	(PF03666.6,Nitrogen Permease regulator of amino acid transport activity 3)	none
MVLG_05570	none	none
MVLG_05591	none	none
MVLG_05594	none	none
MVLG_05595	(PF11790.1,Glycosyl hydrolase catalytic core)	none
MVLG_05613	none	none
MVLG_05674	(PF01650.11,Peptidase C13 family)	none
MVLG_05716	none	none
MVLG_05717	none	none
MVLG_05720	none	none
MVLG_05723	none	none
MVLG_05737	(PF00704.21,Glycosyl hydrolases family 18)	(GO:0003824,family 18 glycoside hydrolase)
MVLG_05746	(PF05572.6,Pregnancy-associated plasma protein-A)	none
MVLG_05754	(PF00150.11,Cellulase (glycosyl hydrolase family 5))	none
MVLG_05804	none	none
MVLG_05822	none	none
MVLG_05825	none	none
MVLG_05835	none	none
MVLG_05836	none	none
MVLG_05864	(PF00111.20,2Fe-2S iron-sulfur cluster binding domain), (PF00667.13,FAD binding domain), (PF00258.18,Flavodoxin), (PF00175.14,Oxidoreductase NAD-binding domain), (PF06999.5,Sucrase/ferredoxin-like)	(GO:0005488,cytochrome p450 oxidoreductase), (GO:0016491,cytochrome p450 oxidoreductase), (GO:0016020,cytochrome p450 oxidoreductase)
MVLG_05876	(PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_05877	(PF06280.5,Fn3-like domain (DUF1034)), (PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_05924	none	none
MVLG_06016	(PF02089.8,Palmitoyl protein	none

	thioesterase)	
MVLG_06036	(PF05199.6,GMC oxidoreductase)	(GO:0009847,choline dehydrogenase)
MVLG_06128	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_06138	none	none
MVLG_06166	none	none
MVLG_06175	none	none
MVLG_06222	(PF00561.13,alpha/beta hydrolase fold)	none
MVLG_06223	(PF00026.16,Eukaryotic aspartyl protease)	none
MVLG_06251	none	(GO:0005488,precursor to secretory protein ssp120)
MVLG_06336	none	none
MVLG_06337	none	none
MVLG_06344	none	none
MVLG_06474	(PF07859.6,alpha/beta hydrolase fold), (PF00135.21,Carboxylesterase)	none
MVLG_06475	(PF07859.6,alpha/beta hydrolase fold), (PF00135.21,Carboxylesterase)	none
MVLG_06485	none	none
MVLG_06512	none	none
MVLG_06534	(PF02469.15,Fasciclin domain)	none
MVLG_06535	none	(GO:0003824,protein serine threonine kinase)
MVLG_06541	none	none
MVLG_06542	none	none
MVLG_06580	none	none
MVLG_06723	(PF00884.16,Sulfatase)	none
MVLG_06764	none	none
MVLG_06772	none	none
MVLG_06804	(PF06280.5,Fn3-like domain (DUF1034)), (PF02225.15,PA domain), (PF00082.15,Subtilase family)	none
MVLG_06814	(PF06280.5,Fn3-like domain (DUF1034)), (PF00082.15,Subtilase family)	none
MVLG_06949	none	none
MVLG_06952	(PF00149.21,Calcineurin-like phosphoesterase)	none
MVLG_06973	(PF01735.11,Lysophospholipase catalytic domain)	none
MVLG_07010	none	none
MVLG_07060	(PF00450.15,Serine carboxypeptidase)	(GO:0004185,carboxypeptidase c), (GO:0000324,carboxypeptidase c), (GO:0046982,carboxypeptidase c)
MVLG_07122	(PF03992.9,Antibiotic biosynthesis monooxygenase)	none
MVLG_07172	(PF07519.4,Tannase and feruloyl esterase)	none
MVLG_07199	none	none
MVLG_07323	none	none

Table A3-4. Summary of PCR results of selected isolates using cDNA template.

ID	Isolate type	Medium	<i>mepA</i> (gDNA/cDN A)	MVLG0558 9RC-GFP	HYG (500 bp/360 bp)
1R	Wild type p1A1	“Rich”	-/+	-	-/-
2R	Wild type p1A2		-/+	-	Faint/-
3R	MV-HYG-p1A1		-/+	-	+/+
4R	MV-HYG-p1A2		-/+	-	+/+
5R	A1-p1A1_1		-/+	+	+/+
6R	A1-p1A1_2		-/+	+	+/+
7R	A1-p1A2_1		-/+	+	+/+
8R	A1-p1A2_2		-/+	+	+/+
9R	A2-p1A1_1		-/+	+	+/+
10R	A2-p1A1_2		-/+	+	+/+
11R	A2-p1A2_1		-/+	+	+/+
12R	A2-p1A2_2		-/+	+	+/+
13R	B1-p1A1_1		-/+	Faint +	+/+
14R	B1-p1A1_2		-/+	Faint +	+/+
15R	B1-p1A2_1		-/+	Faint +	+/+
16R	B1-p1A2_2		-/+	Faint +	+/+
17R	B2-p1A1_1		-/+	Faint +	+/+
18R	B2-p1A1_2		-/+	Very faint +	+/+
19R	C1-p1A1_1		-/+	+	+/+
20R	C1-p1A1_2		-/+	+	+/+
21R	C1-p1A2_1		-/+	+	+/+
22R	C1-p1A2_2		-/+	+	+/+
1W	Wild type p1A1	“Nutrient- Limited”	-/+	Very faint +	Faint/faint
2W	Wild type p1A2		-/+	Very faint +	Faint/faint
3W	MV-HYG-p1A1		-/+	Very faint +	+/+
4W	MV-HYG-p1A2		-/+	Very faint +	+/+
5W	A1-p1A1_1		-/+	+	+/+
6W	A1-p1A1_2		-/+	+	+/+
7W	A1-p1A2_1		-/+	+	+/+
8W	A1-p1A2_2		-/+	+	+/+
9W	A2-p1A1_1		-/+	+	+/+
10W	A2-p1A1_2		-/+	+	+/+
11W	A2-p1A2_1		-/+	+	+/+
12W	A2-p1A2_2		-/+	+	+/+
13W	B1-p1A1_1		-/+	2 bands	+/+
14W	B1-p1A1_2		-/+	2 bands	+/+
15W	B1-p1A2_1		-/+	2 bands	+/+
16W	B1-p1A2_2		-/+	2 bands	+/+
17W	B2-p1A1_1		-/+	2 bands	+/+
18W	B2-p1A1_2		-/+	2 bands	+/+

19W	C1-p1A1_1		-/+	2 bands	+/+
20W	C1-p1A1_2		-/+	2 bands	+/+
21W	C1-p1A2_1		-/+	2 bands	+/+
22W	C1-p1A2_2		-/+	2 bands	+/+
-ve control	RT PCR no enzyme control	-	-	2 bands	-/-
-ve control	PCR H ₂ O control	-	-	-	-/-
+ve control	p1A1 gDNA	-	+/-	ND	ND
+ve control	p1A2 cDNA	-	-/+	ND	ND
+ve control	8-21 plasmid	-	ND	2 bands	+/-

ND = Not Done.

Appendix 4: Germination and maintenance of *Silene latifolia*

This protocol has been modified from what was originally done to reduce the logistic burden. The phytagar percentage was adjusted empirically to be firm enough but not be stifling to the seedlings. To maintain infected plants, set the growth chamber at 20 °C with 13 hours of daylight a day. Water the plants with 100-ppm 15-16-17 fertilizer every other day.

Media preparation

1. Prepare 0.5x MS salt (0.22g per 100 ml) and 0.05% of MES and adjust to pH 5.7 using 1M NaOH or HCl in a beaker that be set on a heater. Each milk jar will require 20 ml of phytagar.

Tip: Prepare extras for seeds that doesn't germinate well and in case the agar does not set right.

2. Add 0.3% of of phytagar to the pH-adjusted solution and start heating to dissolve the agar. Top off the volume with distilled water.
3. Pipette 20 ml of the medium into each milk jar and place lid loosely.
4. Autoclave the milk jars and when done, wear thermal gloves and give each jar a very good swirl. Close the lid completely and bring them to hood to cool down. They can be used immediately or stored at 4 °C for later use.

Note: Do not touch them while cooling otherwise the agar will not set properly. If using a clean hood, leaving the lid ajar or open will help the agar to cool down and dry off condensation on the lid a lot faster.

Sterilization of seeds

5. Prepare sterilization solution (*abbr.* SS) as follows:

Component	Proportion	Per 50 ml
Bleach	40% v/v	20 ml
95% alcohol	20% v/v	10 ml
Distilled water	-	20 ml
Triton X-100	-	1 drop

6. For each vial of seeds (Batch B), add 800 µl of SS and agitate by turning the tube for 2 min.
7. Remove the solution using the pipette and repeat once with SS and five more times with ddH₂O.

Note. To remove the most amount of liquid, agitate the liquid by pipetting and stick the tip of the pipette to the bottom of the tube quickly.

Transferring of seeds

8. Leave a tiny amount of liquid at the last withdraw. Using clean wide bore pipette tip, pick up a seed by water tension. Let the seed stay very near the tip, not inside the tip.

Note: Cut the tips of P200 pipette tips and autoclave them. There is no need to change the tips for each seed as long as you keep the tip clean.

9. Insert the tip with a seed into a milk jar and touch the seed very gently on the agar and allow water tension to pull the seed away from the tip.
10. Close the milk jars and place them at 4 °C for 5 days before shifting into a growth chamber.

Transferring of seedlings

Note: Infection should be performed in the milk jar before transferring the seedlings out of the sterile environment. As long as first pair of true leaves have appeared, the seedlings should be strong enough for transfer.

11. Prepare 2 in square pots by placing a piece of Kimwipe below to block the holes.

12. Add soil and tap to pack lightly. Continue until it levels off. Make a hole in the soil with a stick.
13. Using a hose on running tap, add some water to the milk jar to dislodge the agar.
14. Pour the agar onto hand and use the water to break all the agar away from the seedling.
15. Supporting the seedling by the hypocotyl (between the root and the cotyledon), place the seedling into the hole. Guide and roots into the soil and cover with soil.
16. Keeping the water running gently from the hose, wet the soil thoroughly by soaking through several times.
17. Place pot on trough filled with about 1 cm of water. Place a magenta box or any clear plastic cover tilted over the pot, but ensuring that air exchange is still possible.
18. Place the trough in the growth chamber.
19. Remove the cover for increasing amount of time over the next few days (suggestion: 1 hour, 2 hour, 4 hours, remove) and eventually remove the cover completely. The trough does not need to be flooded once the cover is removed.
20. Keep soil wetted but do not overwater.

Transferring plant to bigger pot

Note: When the pot dries out sooner than expected, or when the plant start needing support, it is time to transfer it.

21. Use Kimwipe to cover the holes and add lightly packed soil to leave slightly less than the current depth of soil in which the plant is in.
22. Squeeze the pot with the plant to release the soil from the pot.
23. Place the plant in new soil and add more soil around the plant and keep tapping the pot to pack the soil down.

24. Add a stick even if the plant does not need support immediately.
25. With a hose on running tap, wet the soil thoroughly several times.

Note: Remember to keep the labels on the pot consistent and legible.

Appendix 5: Recipes for transformation of *Microbotryum lychnidis-dioicae*

MM salts for induction media (Hooykaas et al., 1979)

	1L of 2.5x stock
KH ₂ PO ₄	3.625 g
K ₂ HPO ₄	5.125 g
NaCl	0.375 g
MgSO ₄ ·7H ₂ O	1.250 g
CaCl ₂ ·2H ₂ O	0.165 g
FeSO ₄ ·7H ₂ O	6.5 mg (0.0065 g)
(NH ₄) ₂ SO ₄	1.250 g

Dissolve one salt at a time. Usually some precipitates after the addition of CaCl₂·2H₂O. Store at RT (no need to autoclave).

Induction media, liquid (IM broth) (Bundock et al., 1995)

- Autoclave the following and allow to cool to less than 50°C.

	To make total volume of 200 ml
1x MM salts	80 ml of 2.5x stock
10 mM glucose	0.36 g
0.5% glycerol	1 ml
dH ₂ O	to 192 ml final vol

- Add 8 ml of 1M MES (pH 5.3) to the cooled media.
- Prepare at least 100 µl of 40 mg/ml acetosyringone in DMSO and add 100 µl to the cooled broth.
- Store liquid IM at 4°C.

Induction media, agar (IM agar) (Bundock et al., 1995)

- Autoclave the following and cool to less than 58 °C.

	To make total volume of 600 ml
1x MM salts	240 ml of 2.5x stock
10 mM glucose	0.54 g
0.5% glycerol	3 ml
Agar	12 g
dH ₂ O	330 ml

- Add 24 ml of 1M MES (pH 5.3) to the cooled media.
- Prepare at least 250 µl of 76 mg/ml in DMSO and add directly to the cooled media.
- Pour plates and use within a month or two

1M MES (2-(N-morpholino)ethanesulfonic acid), pH 5.3

- Molecular weight MES is 195.2 g/mol
- Weigh enough MES to be used for the media to be prepared and account for loss in filter sterilizing with 0.22 µm syringe filter.

YPD/HYG/CEF (Yeast peptone dextrose with hygromycin and cefoxitin)

- Autoclave the following and allow to cool to less than 50°C.

	600 ml
Yeast extract (1%)	6 g
Peptone (2%)	12 g
Dextrose (10%)	60 g
Agar (2%)	12 g

- Add 1.8 ml of 50 mg/ml hygromycin (150 µg/ml final) and 600 µl of 300 mg/ml cefoxitine (300 µg/ml final) to cooled medium before pouring.
- Try not to pour the plates too thin especially if the plate is to be used for patching due to longer incubation time needed at room temperature. When incubating, always place in bags to prevent dehydration of media and accidental contamination.

Appendix 6: Command Line BLAST+ & Unix

1. Getting started

1.1. Please go to the BLAST+ page to download and install BLAST+ to your computer.

Follow the instructions on <http://www.ncbi.nlm.nih.gov/books/NBK52640/> for specific instructions for your platform.

1.2. To check whether installation was successful, type in terminal:

```
which blastp
```

1.3. If the return reads `/usr/local/bin`, the installation was successful.

2. Creating and using custom database

2.1. Read the user manual for <http://www.ncbi.nlm.nih.gov/books/NBK279690/> for details that are not specified here.

2.2. Change directory (`cd`) to where the database file is located. Use “tab” key to help complete file or directory name to avoid typo.

```
cd filedirectory
```

To back out of the directory, use `cd . . .`. To list the documents in the folder, use `ls`.

2.3. Create database file with the *makeblastdb* application (4.6.8 in User Manual), where “FastaDatabase.fasta” is the input (`-in`) fasta file with the sequences you want to blast against and you need to specify whether the database type (`-dbtype`) is protein (`prot`) or nucleotide (`nuc1`).

```
makeblastdb -in FastaDatabase.fasta -parse_seqids -dbtype prot/nuc1  
-out DatabaseName
```

2.4. If done correctly, a few files with the same filename but different extensions should be created: `*.phr`, `*.pin`, `*.pog`, `*.psd`, `*.psi` and `*.psq` for protein database **or** `*.nhr`, `*.nin`, `*.nog`, `*.nsd`, `*.nsi` and `*.nsq` for nucleotide database.

2.5.To combine a few fasta files as a database, do step 2.3 for every database file you want to use before proceeding.

2.6.Use the *blastdb_aliastool* application (5.7 in User manual) where you list database files (`-dblist`) to be included. The database (`-dbtype`) can only be one type and specify the alias name (`-out`) for the database. Quotation marks to be used as shown.

```
$ blastdb_aliastool -dblist "FastaDatabase1 FastaDatabase2
FastaDatabase3" -dbtype prot/nucl -out AliasNameforDatabase -title
"DatabaseTitle"
```

If done correctly, "AliasNameforDatabase.pal" should be created.

2.7.To run a search, create a file for the query sequences (QueryFasta.fasta).

2.8.If using combined database, use only the AliasNameforDatabase and leave out the ".pal". *i.e.* `-db AliasNameforDatabase`. If using one fasta file, then the complete filename with extension should be used in the command.

2.9.For standard alignment view of the results, leave out output format (`-outfmt`) option.

`-outfmt 6` gives the results in tabular form. To save the results in a text file, use the output option (`-out`). Use `blastn -help` or `blastp -help` for more details.

Example script:

```
$ blastn -query QueryFasta.fasta -db
AliasNameforDatabase/FastaDatabase.fasta -outfmt 6 -max_target_seqs
num -out ResultsFilename
```

If done correctly, the output text file will be created and get bigger until the run is completed. You can leave the database file where they are or transfer them to the central *BLASTDB* that you create with NCBI databases like below.

3. Using NCBI databases

3.1. To include NCBI database in the search, create a folder named “ncbi-blast-2.2.version/db” in a location of your preference. Download the database you need by connecting to <ftp://ftp.ncbi.nlm.gov/> using a ftp application, e.g. FileZilla (<https://filezilla-project.org>). Browse to /blast/db for the database files (nr.*.tar.gz).

3.2. To unpack the databases:

```
ls *.gz |xargs -n1 tar -xvzf
```

If the files were unpacked correctly, they will appear in the db folder, not as individual folder. “nr.pal” acts as the alias that ties up the 40 volumes of database. It will not work if one is missing. Hence to use this database, you will specify `-db nr` when you run the blast search.

3.3. To remove the original zipped files:

```
rm *.gz.*
```

3.4. BLASTDB is the name of the database that the blast applications will use. Hence, you need to specify where BLASTDB is:

```
export BLASTDB = /path/to/ncbi-blast-2.2.version/db
```

3.5. To verify that you have specified the path correctly, exit to the home directory and type:

```
cd $BLASTDB
```

If you are taken back into the db folder you created, you have done it right. You have to do this every time you open a new terminal window unless you change the PATH to make this action run automatically.

3.6.To update the database, download the script from

http://www.ncbi.nlm.nih.gov/BLAST/docs/update/_blastdb.pl and place it in the db

folder. Browse to the db folder with terminal and type:

```
./update_blastdb.pl -passive -timeout 300 -force -verbose nr
```

3.7.Bear in mind that the nr database files takes up close to 100 GB of space and may take a while to download depending on your connection.

4. Perform HMMsearch on amino acid sequences

4.1.Install HMMER: <http://hmmer.janelia.org>

4.2.Download Pfam-A.hmm.gz from

ftp://ftp.ebi.ac.uk/pub/databases/Pfam/current_release and unpack it. You can also

build your own profile using hmmbuild.

4.3.Prepare query sequences in fasta format.

4.4.Run hmmsearch as such:

```
hmmsearch Pfam-A.hmm input.fasta > results.out
```

5. Useful expressions

5.1.To remove unwanted fasta sequence with specific header.

```
awk '/^>/ {P=index($0,"bad_header")==0}{if(P) print}' in.fasta >
out.fasta
```

5.2.To keep the fasta sequence with specific header.

```
awk '/^>/ {P=index($0,"good_header")}{if(P) print $0}' in.fasta >
out.fasta
```

5.3.To count number of fasta sequence you have in the file:

```
grep \> Test.fasta | wc -l
```


6. Side notes

- 6.1.If you are connecting an external hard disk with the database (lucky you!), all you have to do is follow the steps for installing BLAST+ on your computer, and perform steps 3.4-3.5.
- 6.2.If you believe that the database is outdated, perform 3.8, and then 3.2.

CURRICULUM VITA

NAME: Su San TOH
ADDRESS: 107 Fenley Ave Apt W9
Louisville, KY 40207
DOB: June 27, 1981

EDUCATION

2009 – 2015 Ph.D candidate in Biology, University of Louisville, Louisville, KY, USA (UofL)
Advisor: Dr. Michael H. Perlin
2000 – 2003 B.Sc in Microbiology, National University of Singapore, Singapore

PROFESSIONAL HISTORY

2013 – 2015 Remote Associated Researcher, Broad Institute of MIT and Harvard, Boston, MA
2012 – 2014 Graduate Teaching Assistant, UofL
2011 – 2012, Graduate Research Assistant, UofL
Fall 2013
2005 – 2009 Deputy Quality Manager, DSO National Laboratories, Singapore
2003 – 2011 Member of Technical Staff, DSO National Laboratories, Singapore
(Currently on no-pay leave)
2003 Teacher Aide (Part-time), Agency for Science, Technology and Research (A*Star), Singapore
2000 Patient Service Assistant, Raffles Medical Group, Singapore

SELECTED SKILLS & TECHNIQUES

Cell and Molecular Biology

- High molecular weight DNA isolation
- RNA isolation for Illumina RNASeq and Yeast Two-Hybrid cDNA library
- cDNA synthesis and qPCR
- Cloning
- *Agrobacterium* transformation in fungus
- DNA sequencing
- Northern Blot

Bioinformatics Experience

- GC FAME database on Sherlock® Microbial ID system (MIDI Inc, DE)
- VNTR and FAME database management using BioNumerics (Applied Maths, NV)
- Genome-wide DNA and protein sequence analysis
- RNASeq data analysis (Perl and Unix)
- Predicted proteome analysis

Protein & Biochemistry Techniques

- Bacterial Flow Cytometry
- Bacterial fatty acid methyl ester preparation
- Gas chromatography instrumentation and analysis
- Protein isolation
- ELISA
- Western blot

Immunology Techniques

- PBMC isolation and cell culture
- Phage display antibodies
- Bacterial and toxin detection using antibodies

Environmental Microbiology Techniques

- Bio-detection in environmental samples
- Large scale tangential flow filtration
- Aerosol concentration
- Soil surveillance (*Burkholderia pseudomallei*)
- Luminex technology

Non-Technical Skills

- Microsoft Office including Access
- R
- Quality management (ISO/IEC 17025 & ISO/IEC 15189)
- Internal audit

HONORS, AWARDS and SPECIAL RECOGNITION

2015	2014-2015 Faculty Favorite, UofL
2015	Dean's Citation Award, UofL
2015	Dissertation Completion Award, UofL
2012 – 2014	Graduate Teaching Assistantship, Department of Biology, UofL
2011 – 2012, Fall 2013	Graduate Research Assistantship, Department of Biology, UofL
2013	Bill Furnish Award for Excellence in Teaching of Biology (2012-2013), Department of Biology, UofL
2009 – 2011	Postgraduate scholarship, DSO National Laboratories, Singapore
2007	KINETIC award for Development of Real-time Molecular Diagnostics, DSO National Laboratories, Singapore

PROFESSIONAL SOCIETIES

2011 – present	Golden Key International Honor Society, University of Louisville Chapter
2007 – present	Singapore Society of Microbiology and Biotechnology
2007	Asia-Pacific Biosafety Association

COLLABORATIVE INTERACTIONS

Christina A. Cuomo (Broad Institute of MIT and Harvard)
David S. Treves (Assistant Professor, Indiana University Southeast)
David J. Schultz (Associate Professor, UofL)
Sébastien Duplessis (Centre INRA de Nancy, France)
Zehua Chen (Broad Institute of MIT and Harvard)

MASTER'S STUDENTS TRAINED

2008 Chaay Sokdalis (University of Health and Sciences, Cambodia)
Dr. Monidarin Chou was Sokdalis' advisor. Sokdalis was using the tangential flow system that I had customized for sampling large volume of commonly used water sources in Phnom Penh, Cambodia. She also used a customized high throughput PCR detection panel that I designed for the detection of the presence of 20 bacterial, protozoal and viral targets in the concentrated water sample.

UNDERGRADUATE STUDENTS MENTORED

2012-	Kyung Ji Lee	Summer attachment, UofL
2013	Dhruvil Brahmhatt	Summer attachment, UofL
	Jared Andrews	Undergraduate Research Fellow; Undergraduate Research Fellow, UofL
	Dominique Razeeq	Undergraduate Research Fellow; NSF funded – Research Experience for Undergraduates (REU) program in Plant Biology, UofL
2011	Katelyn Reilly	Bioengineering undergraduate, UofL
	Brittany Hubert	James Graham Brown Fellowship scholar, UofL
	Seongyeon Park	Summer attachment, Carnegie Mellon University, Pittsburg, PA
	Terum Hopper	Summer attachment, UofL
2010	Sarah Kuykendall	Summer Research Opportunities Program, UofL

TEACHING ACTIVITIES

2012-2014 Graduate Teaching Assistant in “Introduction to Microbiology” lab class to non-Biology major students, UofL. Served as head TA in Fall’ 2014.

2003 Conducted two 3-day Basic Microbiology Workshops for Secondary School students (Ages 13-16), River Valley High School, Singapore

NON-PROFESSIONAL AND VOLUNTEER EXPERIENCE

2014 Gel electrophoresis station mentor, NSF funded program - Microbes in your school and home (Seneca High School and Dr. Susanna Remold), UofL

2012 Judge in Cell and Molecular Biology, 48th Annual Louisville Regional Science Fair, Louisville, KY

2011 Catering coordinator, Public Symposium “Crossing the divide: On the adventure of getting Science across to the public”, UofL

2009-2013 Server and cleanup crew once a month at weekly community meal, Nicholasville United Methodist Church, Nicholasville, KY

2008-2009 Founded and trained dance worship team in church for both children and adults, Agape Methodist Church, Singapore

2005 Medical supplies coordinator and translator for mission trip to Medan, Indonesia, Faith Methodist Church, Singapore

2003 Participated in mission trip to Medan, Indonesia, Faith Methodist Church, Singapore

2001-2009 Sunday school teacher for 5th and 6th graders and Children's church worship leader, Agape Methodist Church, Singapore

ORAL PRESENTATIONS

1. **Toh SS**, Schultz DJ and Perlin MH. "RNA expression profile of *Microbotryum violaceum* during infection of *Silene latifolia*". Graduate Award Ceremony, 2012, Biology Department, UofL
2. **Toh SS**. "Genome of *Microbotryum violaceum* and analysis of its predicted genes". 2nd International Microbotryum Meeting, 2013, Amherst College, Amherst, MA

PUBLICATIONS

Peer Reviewed Manuscripts

1. Wong MSY, Lisanti O, Thibault F, **Toh SS**, Loh GK, Hilaire V, Lim J, Neubauer H, Vergnaud G, Ramisse V (2009). Validation of ten new polymorphic tandem repeat loci and application to the MLVA typing of *Burkholderia pseudomallei* isolates collected in Singapore from 1988 to 2004. *Journal of Microbiology Methods* 77(3): 297-301.
2. Perlin MH, Andrews J, **Toh SS**: Chapter Four - Essential Letters in the Fungal Alphabet: ABC and MFS Transporters and Their Roles in Survival and Pathogenicity. In: *Advances in Genetics*. Edited by Theodore Friedmann, Jay C. Dunlap, Stephen F. Goodwin, vol. 85 (2014). Academic Press, pp. 201-53. doi: 10.1016/B978-0-12-800271-1.00004-4
3. Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, de Vienne DM, Aguileta G, Poulain J, Wincker P, Chen Z, **Toh SS**, Cuomo CA, Perlin MH, Gladieux P, Giraud T (2015). Degeneration of the non-recombining regions in the mating type chromosomes of the anther-smut fungi. *Molecular Biology and Evolution* 32(4): 928-43. doi: 10.1093/molbev/msu396.
4. Perlin MH, Amselem J, Fontanillas E, **Toh SS**, Chen Z, Goldberg J, Duplessis S, Henrissat B, Young S, Zeng Q, Aguileta G, Petit E, Badouin H, Andrews J, Razeeq D, Gabaldón T, Quesneville H, Giraud T, Hood ME, Schultz DJ, Cuomo CA (2015). Sex and Parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics* 16: 461. doi: 10.1186/s12864-015-1660-8
5. **Toh SS**, Perlin MH (2015). Size does matter: Staging of *Silene latifolia* floral buds for transcriptome studies. *International Journal of Molecular Sciences* 16: 22027-45. doi: 10.3390/ijms160922027

Abstracts

1. **Toh SS** and Seah GT. "Immunomodulatory effects of *Mycobacteria*". Undergraduate Research Opportunities Program in Science (UROPS), 2003, National University of Singapore, Singapore

2. **Toh SS**, Loh GK, Tan YK, Woo KW, Teo KB and Ng LC. “Fatty acid methyl esters profile analysis for the identification of *Burkholderia* species”. 4th World Melioidosis Congress, 2004, Novotel Apollo, Singapore
3. Kuykendall SE, **Toh SS**, Treves DS, Schultz DJ and Perlin MH. “Infection model of *Microbotryum violaceum* in *Silene latifolia*: finding the route to success”. Summer Research Opportunities Program (SROP) Poster Day, 2010, UofL
4. **Toh SS**, Andrews JM, Cuomo CA, Schultz DJ and Perlin MH. “The *Microbotryum violaceum* genome and transcriptome provide insights into the lifestyles of a fascinating plant pathogen”. Graduate Research Symposium, 2012, UofL
5. **Toh SS**, Andrews JM, Cuomo CA, Schultz DJ and Perlin MH. “The *Microbotryum violaceum* genome and transcriptome provide insights into the lifestyles of a fascinating plant pathogen”. Gordon Research Conference on Cellular and Molecular Fungal Biology, 2012, Holderness, NH
6. **Toh SS**, Andrews JM, Duplessis S, Treves DS, Cuomo CA, Schultz DJ and Perlin MH. “Letters from the front: The *Microbotryum violaceum* genome and transcriptome project”. 27th Fungal Genetics Conference, 2012, Asilomar, Pacific Grove, CA
7. Razeeq D, **Toh SS**, Andrews JM and Perlin MH. “Relative expression analysis of secretory lipases in *Microbotryum violaceum*”. SROP Poster Day, 2013, UofL
8. Andrews JM, **Toh SS**, Razeeq D and Perlin MH. “Relative expression levels of sugar transporters I *Microbotryum violaceum*”. SROP Poster Day, 2013, UofL
9. **Toh SS**, Treves D and Perlin MH. “Reliable transformation system for *Microbotryum lychnidis-dioicae* informed by genome and transcriptome project”. 28th Fungal Genetics Conference, 2015, Asilomar, Pacific Grove, CA

Public Resource

Microbotryum violaceum Database [hosted by Broad Institute of MIT and Harvard] - http://www.broadinstitute.org/annotation/genome/Microbotryum_violaceum/MultiHome.html

References Upon Request