8-2020

Comparative secretomics and functional analysis of effectors utilized by the Microbotryum genus of anther-smut fungal pathogens, and their role in host-specificity.

William Christopher Beckerson
University of Louisville

COMPARATIVE SECRETOMICS AND FUNCTIONAL ANALYSIS OF

EFFECTORS UTILIZED BY THE *MICROBOTRYUM* GENUS OF

ANTHER-SMUT FUNGAL PATHOGENS, AND THEIR ROLE IN

HOST-SPECIFICITY

By

William Christopher Beckerson, M.Sc.

A Dissertation

Submitted to the Faculty of the University of Louisville in Partial Fulfillment of the Requirements

for the Degree of:

Doctor of Philosophy in Biology

Department of Biology: Program on Disease Evolution,

Molecular, Cellular, and Developmental Biology Division,

College of Arts and Sciences

University of Louisville

Louisville, Kentucky, United States

August 2020

COMPARATIVE SECRETOMICS AND FUNCTIONAL ANALYSIS OF EFFECTORS

UTILIZED BY THE MICROBOTRYUM GENUS OF ANTHER-SMUT FUNGAL

PATHOGENS, AND THEIR ROLE IN HOST-SPECIFICITY

By

William Christopher Beckerson, M.Sc.

Dissertation Approved on:

July 17th, 2020

By the following members of the Dissertation Committee:

_____

Dr. Michael H. Perlin, **Principal Advisor**

_____

Dr. David Schultz

_____

Dr. Susanna Remold

_____

Dr. Lee Dugatkin

_____

Dr. Scott Gold

ACKNOWLEDMENTS

"If I have seen further than others, it is by standing on the shoulders of giants"   ~Isaac Newton~

The success I have experienced throughout my research at the University of Louisville is due in great part to the many collaborations I have been a part of, both at the University of Louisville and Universities across the world. I would like to express my deepest gratitude towards the following groups for their guidance and openness towards research, practices that embody the essence of science and have allowed us to accomplish more together than I ever could have on my own.

The Giraud Lab at the Université Paris-Sud:

I would first like to thank Dr. Tatiana Giraud for accepting me into her lab early on in my dissertation and for her guidance across a wide variety of research projects during the span of my dissertation. Dr. Giraud is one of the most effective researchers I have ever had the privilege of working with and has been an excellent role model for me. I would also like to thank Dr. Ricardo de la Vega for his invaluable bioinformatic contributions, as well as Dr. Fanny Hartmann and Marine Duhamel for their contributions to our collaborative secretome research.

The Begerow Lab at Ruhr Universität Bochum:

Dr. Dominik Begerow is another individual I would like to think for his incredibly gracious hospitality throughout my research with his lab in Germany, and for his guidance regarding collaborative efforts between the Perlin and Begerow labs. I would also like to extend a special thank you to Sebastian Klenner, one of the first individuals I had the pleasure of working with during my dissertation, a great scientist, and a great friend. I would also like to thank Lucas Engelhardt, another excellent researcher from the Begerow lab, as well as Martin Kemler; an astute taxonomist, avid bird watcher, and a wonderful drinking buddy!

The Gold Lab at the USDA:

I would like to thank Dr. Scott Gold for his generous contributions to our research, particularly for his help in establishing the first reliable transformation system in *Microbotryum* and for his contributions to our

CRISPR research project. Dr. Gold's guidance and insight into fungal transformation has been an essential for establishing transformation methods in *Microbotryum*.

Dr. Yoder-Himes at the University of Louisville:

While this dissertation presents my research into the evolution of disease at the University of Louisville, I have also had the privilege to work with Dr. Yoder-Himes to analyze pedagogical practices in Biology. I would like to thank her for the opportunity to join her research team, including Dr. Jen Anderson and Dr. John Perpich, and for the opportunity to peruse teaching questions in effort to be the best instructor I can be.

The Menze Lab at the University of Louisville:

I would also like to thank Dr. Michael Menze, another collaborator with whom I worked on a variety of small research projects, as well as Clinton Belot, Rob Skolik, and David Grimm, and the rest of the Menze lab for their comradery and encouragement throughout my dissertation.

The Perlin Lab "aka the GOAT Lab" at the University of Louisville:

I would further like to thank my colleagues in the GOAT lab; Hector Mendoza, Sunita Khanal, Ming-chang "Nelson" Tsai, Joseph Paul Ham, Roxanne Leiter, and Swathi Kuppireddy, all of whom have made working in the GOAT lab a wonderful experience. Their willingness to work as a team, even on projects different from their own, facilitated a cooperative environment that made research fun and effective. I would also like to thank the many undergraduates in the GOAT lab who assisted directly with the completion of this work. Thank you to Catarina Cahill, Brittany Carman, Adney Rakotoniaina, Lloyd Bartley, Grace Long, Phillip Sullivan, and Rebecca Turney for your wiliness to learn and in turn teach me how to be a better mentor.

I would also like to extend a special thank you to the late SuSan Toh, whose work in the Perlin lab laid the foundation for this entire dissertation. While I did not know SuSan personally, her impact on the *Microbotryum* community through her research on *Microbotryum lychnidis-dioiace* opened the door for many research opportunities and enabled me to peruse the kinds of research questions that I have.

Graduate Committee Members:

I would like to thank my graduate committee; Dr. Michael Perlin, Dr. David Schultz, Dr. Lee Dugatkin, Dr. Susanna Remold, and Dr. Scott Gold. When selecting members for this committee, I chose individuals who are experts in fields that I had the least experience in, and whose passion for science was contagious. Doing so allowed me to surround myself with a team that I could readily seek advice from in areas where I needed it the most. I am therefore deeply appreciative for their patience with my many questions over the years, and for their guidance that has made me a much more well-rounded scientist today.

Faculty and Staff of the University of Louisville:

In addition to those who contributed directly to my research, I would also like to thank the staff and administration at the University of Louisville, whose hard work and dedication to the department has been paramount to my success. Thank you to Dorris Meadows, Terri Norris, and Charice Johnson for frequent assistance with travel, grants, scheduling for the Biology Graduate Student Association, and a plethora of random questions I've had over the years. I would also like to thank two faculty members in particular for their amazing work as Director of Graduate Studies during my dissertation. Thank you Dr. Sarah Emery for being there for students in need during a particularly difficult period for the UofL Biology Department, and a special thank you for your assistance with navigating university policy so that I could accept a wonderful opportunity to teach as an adjunct faculty at my alma mater, Georgetown College. And thank you Dr. Perri Eason for your frequent help as Director of Graduate Studies in the early part of my dissertation work and for your excellent work as the Chair of the Biology Department, particularly your willingness to assist the Biology Graduate Student Association. I would also like to thank the rest of the UofL faculty who have always been available to talk with regarding research question, even if they had no stake in the mater. Their collective cordial and collaborative nature make the University of Louisville a welcoming environment.

My Wonderful Family:

No one plans to defend their dissertation during a global pandemic, and it is no overstatement to say that without the support of my family I would have been unable to do so. Continued encouragement and reassurance from my parents, Darla and Chris Beckerson, were paramount in these uncertain times. Through

anxiety and hardship, they have always been there for me, along with my three brothers, Jon, Alex, and Nick, and of course Brooklyn and Meagan as well. Thank you all so much for being there for me when I needed it the most.

My loving wife:

And finally, I would like to thank my wife Mary Beckerson for her support through the entirety of this long and challenging process. During my dissertation, I spent a total of 8 months abroad working with collaborators across Europe, opportunities that would not have been possible without the support, emotional and otherwise, of Mary. In addition to supporting me during long periods of travel, her love and understanding towards late nights in the lab and weekend research has allowed me to be both successful in my endeavors and achieve some essence of work-life balance. Without her love and support, none of this would have been possible.

ABSTRACT

COMPARATIVE SECRETOMICS AND FUNCTIONAL ANALYSIS OF EFFECTORS

UTILIZED BY THE MICROBOTRYUM GENUS OF ANTHER-SMUT FUNGAL

PATHOGENS, AND THEIR ROLE IN HOST-SPECIFICITY

William Christopher Beckerson

July 14th, 2020

Understanding how pathogens evolve in response to changes in their host is paramount to combating the spread of emergent strains of disease. This is particularly true for plant pathogens that cause billions of dollars of damages to crops globally, every year. Understanding the molecular interactions between pathogens and their hosts therefore sheds light on the coevolutionary arms race that can result in host-specificity and host-shifts in plant pathogens. This research approaches the question of how fungal pathogens interact with their plant hosts utilizing both unique and shared arsenals of secreted proteins (SPs) during infection, and addresses the question of whether alterations to shared SPs or species-specific SPs play a more important role in host-specificity. To answer these questions, we annotated and compared the secretomes of three species from the *Microbotryum* genus of anther smuts, two closely related sister species that are able to infect each other's hosts, albeit to reduced degrees, *M. lychnidis-dioicae* and *M. silenes-dioicae*, and one distantly related species that is unable to infect either of the other two species' host plants and *vice versa*, *M. violaceum var. paradoxa*. We then characterized the function of the core SP MVLG_02245, an SP found in all three species with differing levels of conservation at the amino acid sequence level, and tested the importance of two species-specific SPs in host specificity, MvSl_01693 and MvSd_09295, via heterologous expression in each sister species. Finally, for future research into the role of SPs in host pathogenicity, we established a site-specific knockout system in *Microbotryum* using CRISPR Cas9 technology. Our results demonstrate that while host specificity in the *Microbotryum* genus is likely the result of alterations to the amino acid sequence of several core SPs, expression of novel SPs can have dramatic effects on pathogenicity. The research is therefore the first to identify key proteins involved in host specificity of the *Microbotryum* genus, and the first to establish a means of site-specific gene modification and knockout in the *Microbotryum* system using a CRISPR Cas9.

TABLE OF CONTENTS

CHAPTER 1: An introduction to pathogen/host coevolution and *Microbotryum,* a model genus for its study

CHAPTER 2: Identification of conserved and species-specific secreted proteins via comparative secretomics

CHAPTER 3: Functional characterization of the conserved small secreted protein MVLG_02245

LIST OF TABLES

LIST OF FIGURES

Figure	Page

CHAPTER 1

AN INTRODUCTION TO PATHOGEN/HOST COEVOLUTION AND A MODEL
GENUS FOR ITS STUDY, *MICROBOTRYUM*

1.1 The symbiotic relationship between plants hosts and their fungal pathogens

The symbiotic relationships between fungi and plants are ancient, originating at the dawn of terrestrial life.

The transition of plants from an aquatic environment onto land was facilitated by endophytic fungi that acted

as trade partners in the rhizosphere, increasing water and mineral uptake in roots of the plant host in exchange

for carbon sources (Strobel 2018). Over the following hundreds of millions of years, plants and fungi both

diversified; and in turn, so did their symbiotic relationships. Today there are 4 Divisions of terrestrial plants,

all of which include members known to interact with fungi in some way shape or form; however, not all of

these symbioses are mutually beneficial.

Fungal plant pathogens are distributed globally. In the top five agricultural crops, rice, wheat, maize,

potatoes, and soybeans, fungal disease is responsible for more than 125 million tons of destroyed crops every

year, enough food to feed 600 million people (Fisher et. al., 2012). This makes the study of emergent fungal

pathogens and the way in which they manipulate their plant hosts, particularly important for food security on

a global scale. A prime example of the dangers associated with the evolution of fungal pathogens can be

observed in the recent Banana Wilt outbreak. The causative agent of this agricultural disaster is a strain of

Fusarium oxysporum f. sp. Cubense, also known as the tropical race 4 or "TR4", that is currently wreaking

havoc on the Cavendish banana cultivars, the variety that makes up 99% of global exports (Dita et al., 2018).

While TR4 is not the first strain of *Fusarium* to infect banana plants, it is a particularly virulent strain that

has evolved to persist in warmer climates compared to other strains that typically only emerge during cooler

conditions. Understanding how these types of fungal pathogens manipulate their hosts at the molecular level,

and how the two co-evolve over time, is thus imperative for targeting existing agricultural pathogens and

preventing the spread of emergent strains to provide food security.

Another prime example of the large-scale damages caused by unchecked pathogens can be see in the *Puccinia* genus of wheat pathogens, also known as "Wheat Rusts". These generalist fungal pathogens attack the plant and grow on the exterior, feeding on the decaying plant tissue. This, combined with their ability to overwinter within their hosts, makes the pathogen difficult to treat and often leads to massive crop loss for infected wheats, barley, and ryes. The three most common strains found in temperate region where wheats, barley, and rye crops are common include the "Stem or Black Rust", *P. graminis*, the "Leaf or Brown Rust", *P. triticina*, and the "Strip or Yellow Rust", *P. stiriiformis*. *P. graminis* is a common wheat pathogen with particularly disastrous consequences. Because wheat fields are typically clonal, once the rust establishes itself in its first host, it can quickly spread throughout the field leading to sever losses, often between 50%-70% and sometimes destroying the entire field (USDA, 2017). Furthermore, susceptible cultivars cannot be grown in areas with recent outbreaks, as the lifecycle of the rust fungi allows it to lay dormant in the soil and overwhelm young wheat plants as they grow, increasing the operational damage caused by these rusts to the agricultural industry (USDA, 2017). While these particular rust pathogens are exceptionally detrimental to wheat cultivars in the agricultural setting, they are also known to naturally infect other plants including Rye, Barley, Foxtail Barley, Little Barley, and Russian Wildrye. This wide breadth of hosts makes *P. graminis* far more mobile in terms of spread, and changes within different populations of the pathogen and intermediate hosts could lead to potentially disastrous host-shifts.

1.2 Pathogen/host coevolution and reproductive strategies

The competitive nature of parasitism drives an intimate relationship between pathogens and their hosts as they struggle to achieve their conflicting interests. Because changes in either the pathogen or host have a direct impact on the survival of the other, the two act as reciprocal selective pressures on one another. Their allelic frequencies for genes particularly important in the infection/defense response thus follow an inverse parabola, as the success of one group ultimately selects for more fit individuals in the other, resulting in a repetitive tradeoff. This evolutionary tug-of-war was perhaps most elegantly described by Leigh Van Valen in 1973 using an excerpt from Lewis Carroll's, *Through the Looking-Glass* (Van Valen, 1973). In the fantasy world of Carroll's novel, the Red Queen tells Alice, "*here*, you see, it takes all the running you can do, to keep in the same place" (Carroll, 1971). Van Valen applied this statement to describe how pathogens

and hosts are constantly co-evolving, a phenomenon that would eventually be referred to as the Red Queen Hypothesis.

The Red Queen Hypothesis states that in situations of uncertainty, e.g., environmental unpredictability and parasitic load, there are intense selective pressures for adaptations that provide the host with an ability to change rapidly (Bergstrom & Dugatkin, 2011). One of the fastest ways in which the host can obtain greater diversity in their offspring is to reproduce sexually (Bergstrom & Dugatkin, 2011). Sexual reproduction evolves in populations under the selective pressures of an unstable environment, especially pressures imposed by parasitic partners. While the origin of variation through random genetic mutation is the same for sexually and asexually reproducing organisms, sexual reproduction allows for the mixing and matching of new phenotypes, thus creating a faster spread of emergent advantageous traits and a greater depth of genetic possibilities to help escape parasitism (Auld, Tinkler, & Tinsley, 2016) (Figure C1-1).

Genetic Variation Within a Population

Asexual	Sexual	Sexual
Reproduction	Reproduction	Reproduction
	- Similar	- Different
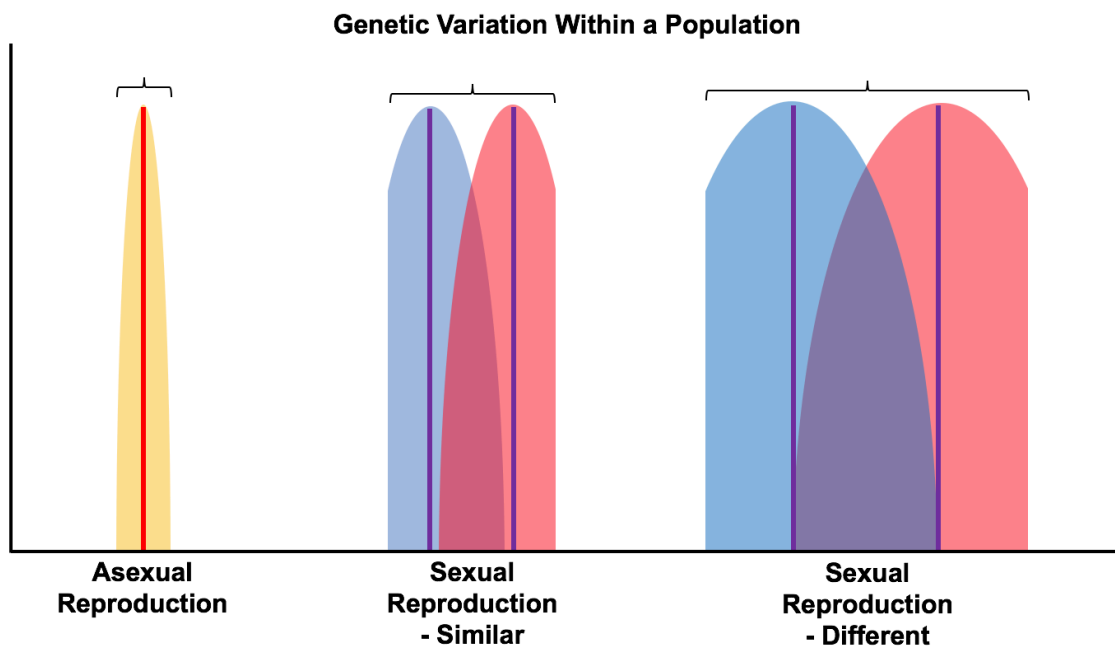
Figure C1-1 Genetic variation within a population by reproductive strategies. The y-axis represents a non-specific number of individuals within a population while the x-axis represents various phenotypic outcomes. Asexual organisms rely on random genetic mutations for their genetic diversity and are therefore much more similar. Sexually reproducing organisms can mix and match genetic variations to their genes, indicated by the overlapping color, and therefore cover a much wider range of phenotypic possibilities in a population.

Another advantage of sexually reproducing organisms involves the ability to purge deleterious mutation, including those that lead to physiological abnormalities that reduce the fitness of a population and those due to phenotypes that make them more susceptible to pathogens, through genetic recombination (Visser & Elena, 2007). When deleterious mutations arise in the genetic material of a sexual organism, having multiple copies of that gene can allow for elimination of the non-functioning copy in their offspring through the mixing and matching of genes in the offspring (Visser & Elena, 2007). Because sexually reproducing organisms each donate one copy of a gene, heterozygous individuals with non-functional copies can mate to form homozygous offspring with restored function of a beneficial gene. While this process is not guaranteed, and any other combination of offspring may arise with full or partial deletion of function in the same gene, selective pressures can drive the population back into the advantageous phenotype over time (Figure C1-2).
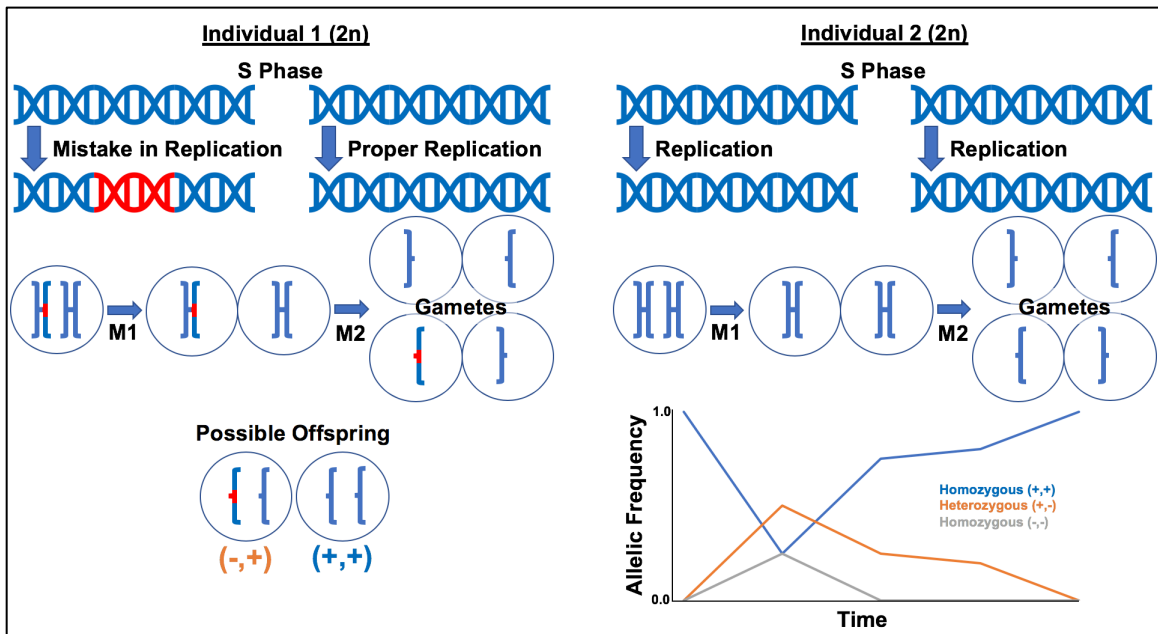
Figure C1-2 Elimination of emergent disadvantageous genes through sexually reproduction. As an emergent phenotype is selected against, genetic variability that is maintained in a population through sexual reproduction allows for return to the previous advantageous phenotype.

Furthermore, in addition to the elimination of novel disadvantageous genotypes, sexual reproduction allows for the reversal to a previous phenotype, even if that genotype should be completely lost in a population through the process of complementation (Perlin et al., 2020). This is particularly useful in creating a moving target for genes that play a primary role in host defense responses against pathogens. Suppose alterations to a particular defense response gene allows the host to avoid inhibitors secreted by the invading pathogen. This phenotype may be so advantageous that multiple different genes may mutate to provide the same phenotype. However, over time as the host acts as a selective pressure for adaptation in the pathogen, the inhibitors secreted by the pathogen may evolve to again recognize the new form of the defense response phenotype. In a sexually reproducing population where two different gene mutations exist, let us call them Mutation A and Mutation B, sexual reproduction and homologous recombination can restore the original wild type defense response in the plant that may now be more suited for the new generation of pathogens (Figure C1-3).
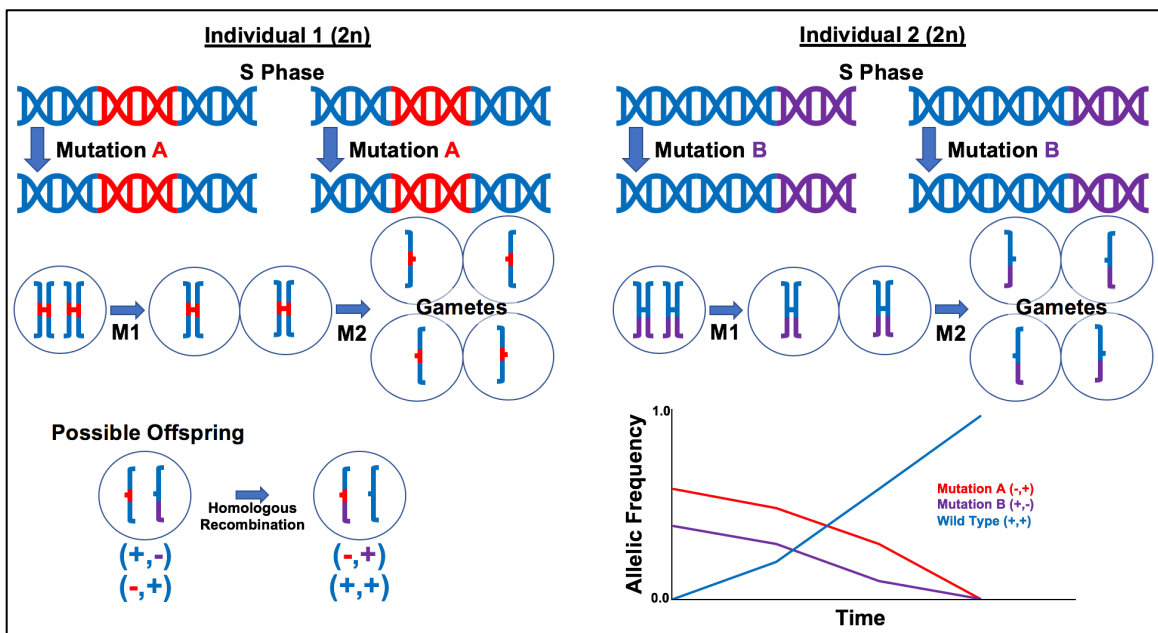
Figure C1-3 Restoration of wild type phenotype through complementation. Sexual selection can restore a population to a previous genotype through recombination of different regions of the genome previously changed in some lineages but not in others.

Sexual reproduction therefore acts as a mechanism for rapid change, an important feature for organisms that reproduce slower than their pathogens, as one generation of hosts may deal with millions of generations of asexually reproducing pathogens.

While organisms that reproduce at such high rates can rely on the natural variations that arise through mutation during mistakes in replication, asexual populations are more vulnerable to the pile-up of deleterious changes (Bergstrom & Dugatkin, 2011). Because many asexual populations are unable to perform genetic recombination, with the exception of horizontal gene transfer, as mutations arise in a population they cannot revert to their wild-type genotype. This phenomenon was proposed by Herman Muller and is widely coined as Muller's ratchet (Muller, 1964). Muller stated that in a population of asexual organisms, mistakes in replication lead to multiple lines of genotypes. As these genotypes all begin to accumulate mutations in a particular gene, eventually the populations will be made of several lines that all have at least 1 mutation. At this point the "ratchet" has clicked one step forward and now cannot be brought back to the original wild type phenotype, either through purifying selection through the elimination of lines with mutations or through horizontal gene transfer. The ratchet can continue to click forward in this manner, furthering the slow migration of the population towards mutation accumulation, or through the elimination of the remaining strains with fewer mutations than the rest, either through genetic drift or other random events (Bergstrom & Dugatkin, 2011). Sexual reproduction therefore offers an advantage in the evolutionary tug-of-war between pathogens and hosts.

It is worth noting that some pathogens can also reproduce sexually, particularly fungal pathogens. Most fungal pathogens possess the capacity to both reproduce sexually or asexually, depending on the conditions in their environment (Heitman et al., 2014). This allows the pathogen to spread quickly through asexual reproduction, forming clusters of clonal lineages before sexually recombining, usually before infection of a host (Heitman et al., 2014). This bimodal approach to reproduction allows fungal pathogens to adapt much more quickly to changes in their hosts, driving rapid evolution of virulence and host range.

But sexual reproduction is not without a cost, otherwise there would be no reason asexually reproducing organisms persist. One major drawback to sexual reproduction can be observed from the gene's-eye view of propagation. Richard Dawkins famously stated in his book "The Selfish Gene" that from a genes perspective, the primary function of life is to copy one's self (Dawkins, 1976). In the case of sexual vs asexual

reproduction, asexual organisms accomplish this goal at a rate of 2X that of sexual organisms (Bergstrom & Dugatkin, 2011). Because sexual reproduction requires the fusion of one gamete with a gamete from a partner, each parent is therefore only donating about half the genetic material compared to individuals from an asexual population. Furthermore, the presence of two separate mating types also imposes restrictions on population growth. In addition to a cost associated with finding a partner, the fact that a sexual population's offspring will contain a mixture of the two mating types at approximately 1:1 ratios, in most cases, means that the population will reproduce only half as fast as an asexual population in which all offspring can reproduce. This argument was made by John Maynard Smith and is referred to as the Twofold Cost of Sex (Bergstrom & Dugatkin, 2011). When one considers these two large costs, along with the many other disadvantages than can arise from sexual reproduction, including the possibility for sexual recombination to break favorable gene combinations and the costs associated with courting a mate (Bergstrom & Dugatkin, 2011) e.g., producing pheromones in fungal pathogens, it is a wonder that sexual reproduction exists at all. The bottom line however is tied back to the idea of the "Selfish Gene". In the presence of a persistent parasitic pathogen, by accepting the costs of sexual reproduction a population can spread more of its genetic information, albeit at a slower rate, than an asexual population completely consumed by disease.

In the case of pathogenic parasitism, the emergence of novel pathogens with a strong advantage on a host can lead to the rapid spread of a new strain through a host population with little diversity (Bergstrom & Dugatkin, 2011). We have seen many examples of this phenomenon in the world's agricultural industry, including the aforementioned Cavendish banana and wheat cultivars, as artificial selection practices favor monocultures of the largest. Because these crops are genetic replicas of one another, and therefore have little genetic diversity, once a pathogen evolves to successfully infect one there is no impediment to successful infection of the rest of the population as well. The same can be thought true for asexually reproducing populations. On the other hand, heirloom cultivars of crops which are often pollinated through open air pollination contain a much higher degree of genetic variation. While this diversity leads to a large variation in crop sizes, they are in turn are less affected by emergent disease (Dwivedi, Goldman, and Ortiz, 2019). Research into the advantage of heirloom practices demonstrated that crops were more resistant to soilborne insects, chronic disease, and exhibited greater stress tolerance compared to monoculture cultivars (Dwivedi, Goldman, and Ortiz, 2019). To continue the metaphor from *through the Looking-Glass*, although genetic

diversity usually does not allow a host to outrun their pathogens, the genetic diversity does allow them to create a moving target, or allows them to keep running to stay in the same place as Lewis Carroll would say, rather than being completely overwhelmed by the pathogen (Bergstrom & Dugatkin, 2011; Carroll, 1971). The Red Queen Hypothesis therefore accurately proposes that recombination of genetic material is selected for in hosts to increase their genetic diversity in order to help them out-pace their pathogens, and the genetic diversity that is created through sexual reproduction lays the ground work for an evolutionary tug-of-war between the pathogen and the host (Morran et al., 2011).

1.3 The role of secretory proteins in pathogen/host coevolution

While even sexually producing populations can share a wide degree of similarity in physical appearance, when we consider the types of changes that are going on at the molecular level within a population of pathogens and hosts we find that chemical compounds and proteins that play a role in cell communication and recognition/repression are constantly adapting to changes that arise in the new generations of hosts and pathogens, a phenomenon that over time can lead to host specialization. The success of a fungal pathogen infecting and otherwise manipulating their hosts depends heavily on their arsenal of secreted compounds. As saprophytic chemotrophs, fungi interact with the world around them through the secretion of compounds that play a variety of roles from external digestion to defense against predation (Urry et al., 2017). In the case of fungal pathogens, these compounds can play a role in host penetration, host manipulation, eliminating competing fungi or bacteria, breaking down carbon sources for food, and identification of potential mating partners. While some fungal pathogens are opportunistic and grow on dying plants, others live on or inside living hosts, utilizing secreted proteins that can dissolve cellulose to enter the cell wall of the plant host, as well as proteins that can moderate the plant host's defense response through suppression of signal transduction pathways or gene expression (Rep, 2005) and manipulate the host through modification to hormonal pathways (Rabe et al., 2013).

While all fungal pathogens manipulate their hosts to some degree through the secretion of various proteins and other compounds, their approach to doing so can vary dramatically. Of these different approaches, there are three main strategies; necrotrophic pathogens employ a particularly destructive approach of overwhelming the host and causing extensive necrotic damage to tissues killing the host in the

process, biotrophic pathogens on the other hand pursue a much more subtle strategy preferring to establish a long-term feeding relationship with their hosts by living in or on their hosts without killing them, and saprophytic pathogens forgo battling with hosts altogether, preferring to colonize and consume dead material instead. Regardless of strategy, each lifestyle requires the secretion of proteins to manage their hosts, perform external digestion, and/or fend off other opportunistic microorganisms.

Plants also utilize the secretion of proteins and other secondary metabolites in the arms race against their pathogens (Vincent, Rafiqui, & Job, 2019). Upon detection of a fungal pathogen, plants can utilize both the traditional ER-Golgi mediated secretion of upregulated proteins, and small extracellular vesicles for the secretion of leaderless secretory proteins and other secondary metabolites for combat in the extracellular space (Vincent, Rafiqui, & Job, 2019). If we consider the coevolution between the secreted proteins of the pathogen and the secreted proteins of the host an evolutionary arms race, then the battlefront is the apoplast of the host, where fungal pathogens try to break through and the plant host hunker down to defend the line. The apoplast of a plant is well fortified by the open extracellular space and its cell wall, comprised of complex networks of polysaccharide polymers and glycoproteins. In addition, the apoplastic space contains apoplastic fluid circulating throughout the cell wall and facilitating both fast communication between cells and delivering defense proteins (Delaunois et al., 2014). Signals facilitated by the apoplastic space allows the host to recognize microbial-associated molecular patterns (MAMPs) or damage-associated molecular patterns (DAMPs), two categories of Pattern Recognition Receptors (PPRs), to begin their defense response (Delaunois et al., 2014). This ultimately makes the apoplastic space the front line in the battle to determine who is likely to win the war. Because higher plants are constantly interacting with both mutualistic and parasitic microbes, they have evolved to recognize a large range of different MAMPs for a large variety of different pathogenic species, some examples of which are listed in Table C1-1.

Table C1-1. Examples of Microbe-Associated Molecular Patterns

MAMP Signal	Organisms Detected	Reference
Eicosapolyenoic acids	Oomycetes	(Savchenko et al., 2010)
β-glucans	Fungi	(Klarzynski et al., 2000)
peptidoglycans	Bacteria	(Willmann et al., 2011)
lipopolysaccharides	Gram Negative Bacteria	(Erbs and Newmann, 2012)
rhamnolipids	Bacteria	(Sanchez et al., 2007)
Chitin oligomers	Fungi	(Miya et al., 2007)

Because evolution selects for pathogens which are able to secrete effector proteins which specifically block the host proteins responsible for the recognition of MAMPs, plants have also evolved effector-triggered immunity, a fast-tracked version of defense response signaling that operates by directly recognizing these anti-MAMP effectors secreted by their pathogens (Delaunois et al., 2014). The response to recognition of virulence factors from a known pathogen elicits a much stronger immune response in the host, often triggering compartmentalization of infected cells through the closure of plasmodesmata and programmed localized cell death to prevent the spread of the pathogen (Tsuda and Katagiri, 2010). This process is accomplished through the activation of MAPK kinases, which release reactive oxygen species within minutes of detecting a foreign effector (Delaunois et al., 2014). It is here that we can begin to see the evolutionary arms race materialize at the molecular level as pathogens act on the host to evolve better recognition of pathogen secreted proteins, and in turn, hosts act as selective pressure for more discrete inhibition of host responses in the pathogens.

To win a battle in the ongoing evolutionary arms race, pathogens must secrete a variety of compounds or effectors to suppress PPR-mediated defense responses and otherwise manipulate their host. Many host-specific groups of pathogens have sets of proteins with yet unknown roles, roles that are vital to the infection of a particular species or genus of hosts. These fungal effector proteins are often small (<250 amino acids), have no know Pfam domain (regions of amino acid sequences shared amongst various protein families indicative of a particular known protein function), and are often limited in their phylogenetic distribution due to rapid evolutionary pressures imposed by the pathogen/host arms race (Rep, 2005). Interestingly, these small secreted proteins vital for infection are also highly likely to stimulate the host

immune system, further supporting the idea of coevolution between the pathogen and host (Rep, 2005). However, despite the absences of a Pfam domain or other known highly conserved function across the wide database of available pathogen genomes, small secreted proteins that play a role in host pathogenicity are more likely to be shared amongst closely related lineages that parasitize similar hosts. This combination of preservation of proteins amongst closely related pathogens with no known Pfam or GO terms a good place to start for the identification of host-specific effectors.

Through a decade of molecular genetics and bioinformatic analyses made possible by the ever-expanding accessibility and ease of genetic sequencing, many novel fungal effectors have been identified. These effectors either promote the virulence of fungal pathogens or suppress defense responses, allowing the fungus to colonize the host. Examples of well conserved effectors found across fungal pathogens include secretory lipases, effectors that inhibit plant-mediated immunity response through inhibition of callose formation (Marzin et al., 2016), and pectinesterases, which act to modify cell wall composition and allow for penetration into the plant host (Blümke et al., 2014). While these types of effectors are heavily conserved amongst plant pathogens, coevolution and genetic drift lead to a wide diversity of effector amino acid sequences. Therefore, using bioinformatics to identify regions of the amino acid sequences that may be more conserved amongst closely related groups, such as activation domains which may be under selective pressures due to their importance for protein function, can be useful to recognize conserved features among various effector families. Furthermore, a lack of identifiable Pfam domain or GO term can indicate a unique function for effectors in a system that result from the intimate coevolutionary relationship between the pathogen and its host.

Once established inside the plant, fungal pathogens can manipulate their hosts through the modification, or even the secretion of synthetic versions of plant hormones (Ma & Ma, 2016). Several recent studies have identified a variety of phytohormones that play a major role in regulating plant-microbe interaction (Vincent et al., 2020). These include the "Big Five" plant hormones, auxin, gibberellin, cytokinin, ethylene, and abscisic acid (Ma & Ma, 2016), as well as others such as salicylic acid, and jasmonic acid (Vincewnt et al., 2020). While some of these hormones are disrupted due to their direct role in combating the spread of the pathogen in the hosts, e.g., abscisic acid exhibits antifungal properties (Khedr et al., 2018), other plant hormones may be manipulated to induce more preferable conditions in the hosts, e.g., through the

induction of galls via modulation of indole-3-acetic acid and cytokinin (Davies et al., 2005; Mizoi et al., 2012). Furthermore, by manipulating the plant hormones pathogens can affect the development and growth of plant tissue, a process that can aid in the spread of disease to new hosts (Ma & Ma, 2016).

As the molecular arms race plays out over evolutionary time, slow divergence of hosts due to reproductive barriers can lead to slight changes in the host-specific secreted proteins of the fungal pathogen isolates. These changes can accumulate and lead to speciation events in the fungi themselves. As the amino acid sequence for a particular protein changes rapidly to keep up with evolving host defense response in a geographically isolated population, local adaptation can lead to different changes that over time can cause the fungi to be unable to infect hosts outside of their population. This can lead to interesting changes to conserved core secreted proteins or to the addition of entirely novel proteins to the secretome of these fungi. Over time, these changes amplify host-specificity of these fungi, and local adaptations can lead to post-zygotic barriers in fungi of different populations. There are then direct selective pressures for the plant to be able to quickly identify pathogens and direct inverse selective pressure for pathogens to be able to quickly mediate plant defense responses (Figure C1-4).
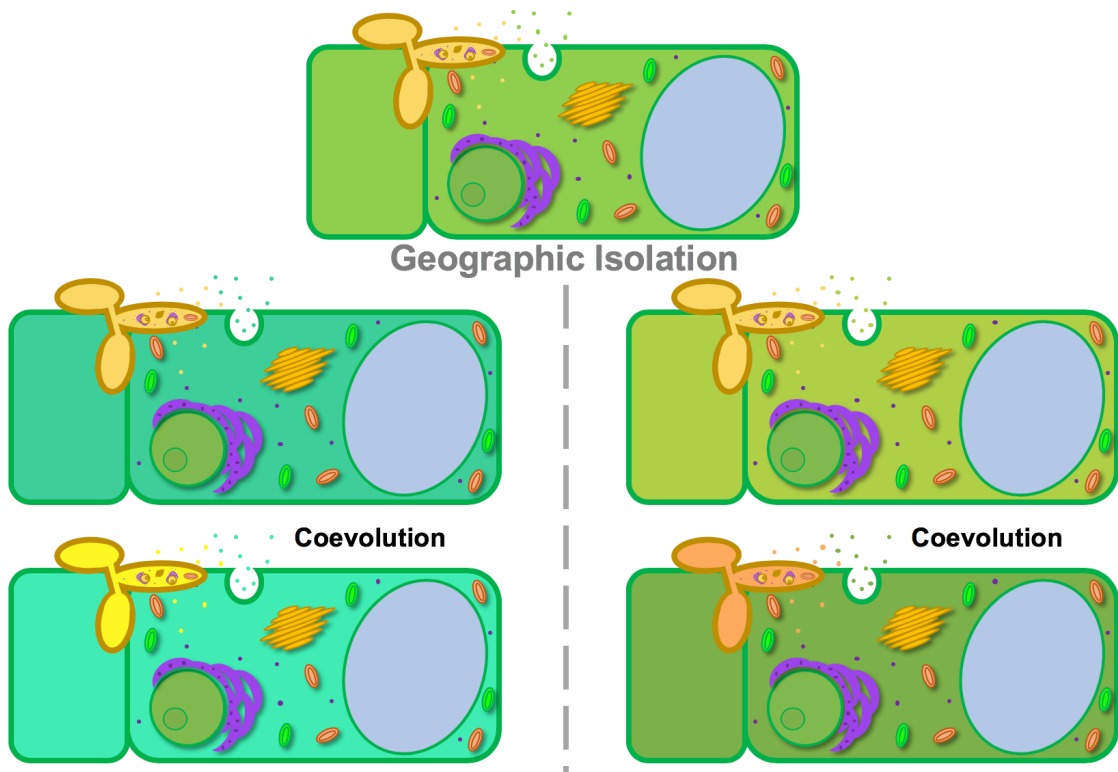
Figure C1-4 A Model for secreted protein molecular arms race. How geographic isolation of infected hosts can lead to coevolution and speciation in both the host and the pathogen.

1.4 The anther-smut fungi, *Microbotryum*

As pathogens become more and more specialized to the genetic diversity found within their population of hosts, limited gene flow between populations, either due to geographical or other reproductive barriers, can result in intense host-specificity. Therefore, speciation events in the host can subsequently lead to speciation events in their pathogens (Figure C1-4). In systems with pathogen-host coevolution, phylogeny between the two often represent a near mirror image. Such is the case for the anther-smut pathogen species complex, *Microbotryum violaceum* and their Caryophyllaceae hosts (Figure C1-5 from Hartmann et al., 2019).
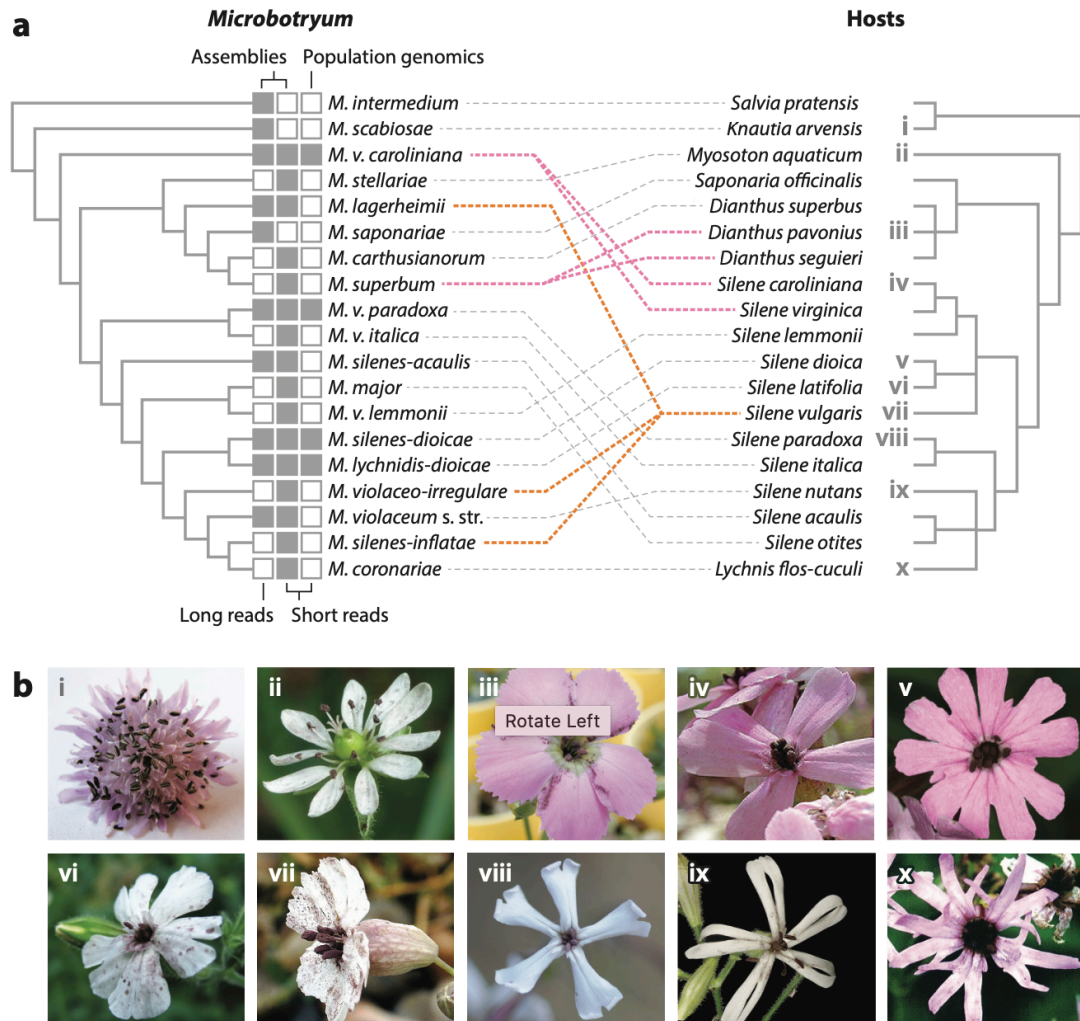
Figure C1-5 Comparison of plant and fungal phylogenies using one strain per host species. The phylogeny of *Microbotryum* species, right, and their Caryophyllaceae hosts, left demonstrate the coevolution and host-shifts observed in the pathogen/host pairings. Image is from BMC Evolutionary Biology with permission from Hartmann et al., 2019.

Upon its initial discovery, the *Microbotryum violaceum* species complex was originally described as a monophyletic generalist pathogen of the Ustilagoinomycotina lineage, *Ustilago violacea*, that parasitized various members of the Caryophyllaceae family of flowers (Baker, 1947; Fischer and Holton, 1957). However, morphological studies and infection assays have since demonstrated that the *Microbotryum* genus is a collection of separate species denoted by their intense host-specificity to one or two hosts. The life cycle of the fungus begins and ends in the anthers of their hosts. Many nocturnal moths and diurnal hoverflies are known to pollinate members of the Silene genus (Jürgens, Witt, & Gottsberger 1996), carrying the spores of their corresponding *Microbotryum* species to new hosts. In the particularly well studied interaction, that of *Microbotryum lychnidis-dioicae* on *Silene latifolia*, the nocturnal moth *Hadena bicruris* is one of the main pollinators driving the spread of *M. lychnidis-dioicae* (Jürgens, Witt, & Gottsberger 1996). Of interest, there does not appear to be discrimination by the *H. bicruris* for flowers uninfected with *Microbotryum* spores. While there is little research on the topic, it does beg the questions if the fungal spores are attracting the pollinators via chemical mimicry or otherwise fooling these insects in order to improve propagation. Once a teliospore of the fungus is deposited on a new host, germination occurs and meiosis results in the production of the yeast-like stage of the fungi, haploid cells known as sporidia. In nutrient rich conditions, such as those found in the nectar of the host, these fungal sporidia will continue to reproduce asexually through budding (Schäfer et al., 2010) (Figure C1-6-2). As the flowering season ends in the fall, the available carbon depletes and the fungal cells of opposite mating types form conjugation tubes, mating and forming the infectious dikaryotic filament. This dikaryotic hyphal structure penetrates the plant cell wall (Figure C1-6-4) and the fungus migrates to the roots where it will overwinter with the host until the spring. In the following spring, as the plant bolts and flowers begin to form, the fungus migrates back up to the anthers of developing flowers where the infection process is completed Separate fungal nuclei fuse to form diploid teliospores which are then deposited on the anther in of aborted pollen for transport by unsuspecting pollinators.
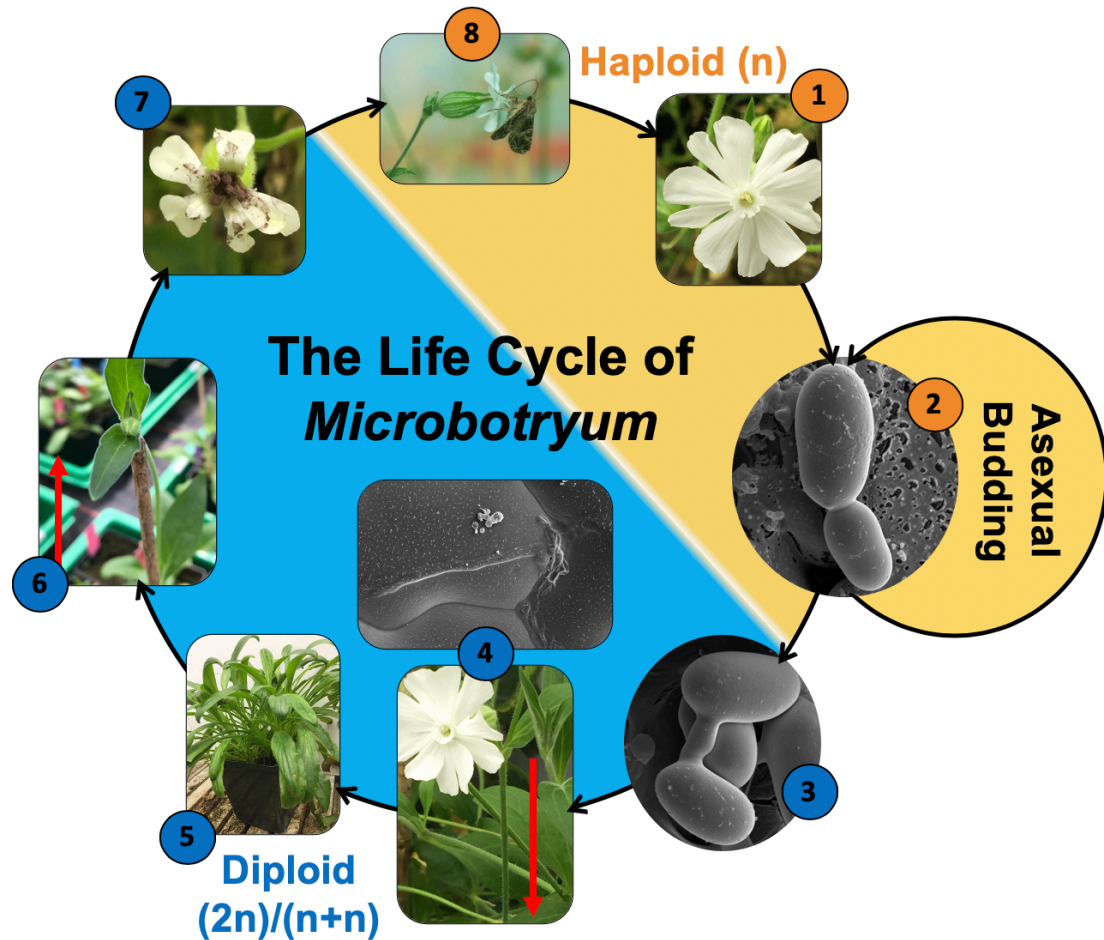
Figure C1-6 The life cycle of *Microbotryum*, demonstrated with *M. lychnidis-dioicae*. 1) Pollinators bring teliospores to new host flowers. 2) The teliospores germinate and undergo meiosis to produce haploid sporidia capable of asexual budding. 3) These yeast-like cells find compatible mating types and form a conjugation tube. 4) An infectious dikaryotic filament is formed and penetrates the plant. 5) The dikaryotic fungus travels to the roots of the hosts where they overwinter with the plant. 6) In the spring, the dikaryotic fungus travels to the developing anthers during bolting. 7) Karyogamy result in generation of fungal teliospores located on the anthers of the host. 8) A pollinator picks up fungal spores and transports them to a new host.

When infection studies are performed in a controlled lab setting via inoculation of seedlings with suspensions of both fungal mating types, infection is systemic and successfully infected plants will contain teliospores in all flowers; however, it has been observed in the natural setting that plants can have both infected and non-

infected flowers. This partial infection in nature may play an important role in ensuring propagation of the host species for future generations of *Microbotryum*, a sophisticated approach for any pathogenic life cycle, as systemic infection such as those seen in laboratory studies would even result in infected flowers produced by vegetative runners. When infected, *Silene* hosts have demonstrated an increased rate of fluorescence, a potentially advantageous host manipulation for the dissemination of more fungal spores. Furthermore, some species of *Microbotryum* have been observed displaying yet another adaptation for the infection of dioecious hosts, such as *S. latifolia*, where infection of a female flower leads to abortion of the flower ovary and induced production of pseudoanther for placement of fungal spores (Toh et al., 2018). While a mechanism for this unique phenomenon is not yet identified, the ability of the pathogen to manipulate the development and growth of its hosts is much in line with previously identified hormonal pathway manipulations in other fungal pathogen/plant host pairings (Ma & Ma, 2016).

The intense host specificity, together with the ability of *Microbotryum* species to infiltrate its host undetected and manipulate the reproductive chemistry of female flowers, suggests that each *Microbotryum* species utilizes a unique portfolio of secreted compounds during pathogenicity, a starting point to understanding pathogen/host coevolution between the many members of the *Microbotryum* genus and their many hosts over time. This dissertation will explore the role of *Microbotryum* effectors in host-specialization and manipulation by implementing both a bioinformatic approach to identify and compare the secretomes of closely related *Microbotryum* species and molecular genetics techniques to characterize the role of a few representatives of the small secreted proteins. Secretomic comparisons, discussed further in Chapter 2, were first performed on three closely related *Microbotryum* species, two sister species with capacity for cross infection of each other's host, *M. lychnidis-dioicae* and *M. silenes-dioicae*, and one more distantly related species for which infection in the hosts of the two sister-species is not observed, *M. violaceum* var *paradoxa*. From this, a list of conserved and species-specific small secreted proteins were identified. Next, in Chapter 3, this dissertation introduces the characterization of a core secreted protein through a series of molecular genetic tests to demonstrate secretion of the protein, identify physiological localization outside of the cell, and identify the target within the host. Chapter 4 then outlines the role of species-specific proteins in host specialization through infection studies utilizing heterologous expression of species-specific small secreted proteins from each sister species of *Microbotryum*. Finally, Chapter 5 of this dissertation addresses the work

done to implement a reliable CRISPR-Cas9 approach to generating site-specific gene knockouts in the *Microbotryum* genus in order to further characterize the identified list of both core and species specific effectors for future research.

CHAPTER 2

IDENTIFICATION OF CORE AND UNIQUE SECRETED PROTEINS THROUGH COMPARATIVE SECRETOMICS

Introduction

Host specialization is a phenomenon well documented in many fungal pathogen/plant host systems (Sánchez-Vallet et al. 2018), which most often occurs through host shifts (de Vienne et al. 2013). The ability to infect a new host is determined by the protein-protein interactions that occur at the pathogen/host interphase. For pathogens to be successful, they must not only be able to colonize the host, but must also work around a gauntlet of host defense responses, as well as manipulate the host to their advantage. Pathogens accomplish these ends through the deployment of many secreted effectors (Lanver et al. 2017; Anderson et al. 2010; Whisson et al. 2007).

It has been understood for several decades that plant pathogens utilize secreted effectors to infect their hosts (Albersheim and Anderson, 1971; Sánchez-Vallet et al., 2018), including the maize pathogen member of the "smut fungi", *Ustilago maydis* (Lanver et al. 2017). To defend against these pathogens, plants continuously evolve to recognize pathogen-associated molecular patterns and trigger a variety of immune responses (Jones and Dangl, 2006). Reciprocally, there is an ongoing selective pressure for plant pathogens to adapt to their host by developing new effectors, or otherwise alter the composition of their secretomes, to evade detection and find new ways to manipulate the host to their advantage. Secretomes can thus evolve rapidly, not only during host shift events but also due to intra-specific coevolution (Meile et al. 2018). It is, however, still unclear whether changes in secretomes leading to host specialization and local adaptation primarily involve effector gene gains/losses or changes in their sequences. Repeat-induced point mutations (RIP) is a fungal defense mechanism against transposable elements that has been suggested to play a role in effector diversification in fungi harboring effectors in regions rich in repetitive elements (Fudal et al. 2009; Van de Wouw et al. 2010). RIP indeed acts via mutations of repeated sequences at specific target sites and can "leak" on neighbor genes (Fudal et al. 2009; Van de Wouw et al. 2010).

Host specialization following host shift is particularly common in the fungal pathogen species complex *Microbotryum violaceum* (Refrégier et al. 2008). *Microbotryum* species are basidiomycete smut fungi that complete their life cycle in the anthers of their respective host plants, replacing the pollen with their own fungal spores (Schäfer et al. 2010). Originally described as a single species, these "anther smuts" are now understood to represent a complex of species (Perlin et al. 1997; Le Gac et al. 2007a), most being highly specific to particular species of the Caryophyllaceae family, also known as "pinks" (Hood et al. 2010). Intra-specific coevolution has also been suggested to occur based on local adaptation patterns, where host plants were more resistant to their local sympatric anther-smut pathogen than to those from geographically distant populations of the same species (Kaltz et al. 1999, Feurtey et al. 2016).

To infect their hosts, *Microbotryum* fungi, like many other plant pathogens, employ an array of effector proteins to block plant immune response and otherwise manipulate the host during infection (Perlin et al. 2015; Kuppireddy et al. 2017). While the specificity of the various *Microbotryum* species to their corresponding host plants has been extensively described (Hood et al. 2010; Le Gac et al. 2007; de Vienne et al. 2009), the molecular basis for host specialization and coevolution within the complex has just recently begun to be explored (Hartmann et al. 2018; Badouin et al. 2017; Aguileta et al. 2010). Understanding the changes that have occurred in the secretomes of these host-specific species will broaden our understanding of the mechanisms behind coevolution, host-shifts and emergent diseases. Furthermore, *Microbotryum* species offer a unique model system to study host shifts and specialization, with multiple host-specific and closely related pathogens (Hartmann et al. 2019), which is not often the case in agriculturally propagated crops.

To test whether host-specific or locally-adapted closely-related pathogens mainly differed in their secretomes by gene gains/losses or by rapid evolution of shared effectors, we compared the secretomes of three *Microbotryum* species, two sister species, *M. lychnidis-dioicae* and *M. silenes-dioicae*, and a more distantly related relative, *M. violaceum* var *paradoxa*. We sought to identify sets of core secreted proteins (i.e., orthologous genes encoding secreted proteins shared by all species), that likely play a major role in the pathogenicity of the species complex as a whole. We also sought to identify species-specific effectors and effectors evolving under positive selection and highly expressed *in planta*, thus perhaps involved in host specificity. To further our understanding of coevolution and local adaptation, we compared the secretomes

of two *M. lychnidis-dioicae* strains collected from geographically distant populations belonging to distinct genetic clusters that have shown contrasted infection patterns consistent with plant local adaptation (Feurtey et al. 2016). We also investigated whether the most frequent changes among host-specific species or locally-adapted clusters involved mostly the gain/loss of secreted proteins or the diversification of shared proteins. As RIP-like footprints have been detected in *Microbotryum* fungi (Hood et al. 2005), we also tested whether sequence divergence in genes under positive selection and/or in genes encoding secreted proteins could have been facilitated by RIP.

Materials and Methods

2.1 Comparative genomics

To analyze the relationship between various predicted effectors, we performed genomic analyses on the following available genomes, obtained using Pacific Bioscience (PacBio) single molecule real time sequencing: GCA_900015465.1 for *M. lychnidis-dioicae* Lamole a_1 (Italy) (Branco et al., 2017), GCA_900015495.1 for *M. violaceum* var *paradoxa* from *Silene paradoxa* 1252 a_1 (Branco et al., 2018), and QPIF00000000 for *M. silenes-dioicae* 1303 a_2 (Branco et al. 2017). These genomes were selected for comparison due to their relationship to one another; *M. lychnidis-dioicae* strains and *M. silenes-dioicae* are sister species, able to infect one another's host in the greenhouse, although to a lesser degree than their natural host (Gibson et al. 2014) and very little in natural populations (Gladieux et al. 2011), while *M. violaceum* var *paradoxa* serves as an outgroup, unable to infect either of the sister species' hosts or *vice versa* (de Vienne et al. 2009).

In total, we used eight sequence-based prediction tools to identify potential effectors by searching each genome for genes with hallmarks for secretion and without conflicting cellular localization predictions. The initial list of putative secreted proteins (SPs) were generated by running the entire genomes through SignalP 4.0 (Petersen et al., 2011). In order to increase the stringency of this analysis, the SPs must then have passed the following criteria to rule out potential localization or retention in various membranes within or on the cell, similar to the previously published protocol for *M. lychnidis-dioicae* (Perlin et al. 2015). Potential transmembrane domains were predicted with TMHMM (Krogh et al., 2001) and Phobius (Käll et al., 2007). Only gene models with none or a single transmembrane domain prediction overlapping the signal peptide

prediction were considered further (Perlin et al., 2015; Petersen et al., 2011). Prosite was used to screen for predicted endoplasmic reticulum retention signals, while PredGPI (Pierloni et al., 2008) was used to screen for potential glycosylphosphatidylinositol anchors, and NucPred (Bramaier et al., 2008) was used to screen for nuclear localization signals in the predicted protein (Figure 1).

Gene models predicted to be secreted and without conflicting localization predictions (i.e., negative for transmembrane domains, endoplasmic reticulum retention, GPI-anchoring, and nuclear localization) were further screened using additional criteria to identify strong predictive footprints of secretion in the signal peptide region. To qualify as a SP, the candidates must also have passed stringent cutoff values for secretion, listed in Figure 1, for at least three of the following four tests: a predicted secretion signal by TargetP (Emanuelsson et al., 2000), a D-score of greater than 0.43 for the neural network [NN], a secretion probability of greater than 0.8 for the hidden Markov model [HMM] from SignalP3.0, and predicted secretion by Phobius.

We searched the resulting putative SPs among the orthologous groups reconstructed previously (Branco et al., 2018). Briefly, the orthologous groups were obtained using mcl (van Dongen, 2000) to cluster high-scoring blastp matches between all gene models predicted in 15 haploid genomes from eight *Microbotryum* species, previously parsed with orthAgogue (Ekseth et al., 2014). We classified a predicted SP as a species-specific SP if there was no ortholog in two of the species being considered. For predicted SP belonging to orthologous groups, we distinguished between species-specific, two- or three-way orthologous groups (i.e., predicted as SP in a single, in two or in three species, respectively) and between orthologous groups composed exclusively by predicted SP (SP-only) and those containing at least one gene model not predicted as SP (SP-mixed). We defined the "core secretome" as the full set of predicted SPs belonging to SP-only three-way orthologous groups (i.e., present and predicted as SPs in all three species). Conversely, we defined as "accessory secretome" the predicted SPs that were either species-specific or belonged to SP-mixed or two-way SP-only orthologous groups (i.e., were not present in all species or not predicted as SP in all species; Figure 2). Together, the core and accessory secretomes make up the "pan-secretome", i.e., the full set of predicted SP in all species considered.

2.2 Pfam domain annotation

We searched Pfam release 32 (El Gebali et al., 2019) against the translated gene models of all predicted SP and their homologs with hmmsearch from the hmmer 3.1b1 suite (http://hmmer.org). Hits with an E-value smaller than 1e-3 were considered significant. The results were then categorized by size as well as presence/absence of a predicted Pfam domain (supplemental file SF1 from Beckerson et al., 2019).

2.3 Signal peptide clustering and experimental validation

We clustered the predicted signal peptide sequences with CD-HIT (Huang et al., 2010) allowing for up to five amino acid differences (non default options: -c 0.75 -l 5). We tested if predicted signal peptides could direct the secretion of the Suc2 invertase employing a yeast-based secretion trap method (Lee and Rose, 2012; Kuppireddy et al. 2017). Six signal-peptide encoding sequences, as determined by SignalP 4.1 software, were amplified by PCR. Standard PCR cycle was used with initial denaturation set at 94 °C for 4 min and 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s and final extension time of 5 min at 72 °C. The purified fragments were then subcloned into a TOPO vector using an Invitrogen TOPO TA Cloning® kit, and subjected to restriction digestion with *Eco* RI and *Not* I enzymes. The digested fragments were then purified and cloned into the pYST-0 vector, upstream and in-frame with an invertase coding sequence, *SUC2*. The presence of each signal peptide encoded in-frame with the *SUC2* coding region was confirmed by DNA sequencing (Eurofins, Louisville, KY).

Invertase deficient (suc2⁻) *Sacchromyces cerevisiae* strain (SEY 6210 - MATαleu2-3, 112 ura3-52 his-Δ200 trp1- Δ901 lys2-801 suc2⁻ Δ9 GAL) cells were transformed with the constructs using the Frozen-EZ Yeast transformation II kit™ from Zymo Research. Cells were then suspended in water and spread onto synthetic drop (SD) out, SD/-Leu (Clontech) selection plates with either sucrose as the sole carbon source or glucose as a control. Resulting colonies from the sucrose plates were grown overnight in 3 ml of SD/-Leu broth with sucrose and 10 μL of undiluted, 10-fold dilutions, and 100-fold dilutions were spotted onto SD/-Leu with glucose or sucrose as the carbon source and incubated for 2 days at 30 °C. Clones harboring functional signal peptides with the reconstituted invertase activity were able to grow on sucrose as the sole carbon source. Untransformed mutant yeast strain SEY 6210 and transformed SEY 6210 cells with empty

pYST-0 vector were used as negative controls. Plasmid DNA was extracted from the positive clones and used to retransform *E. coli*. The constructs were again checked for the presence of signal peptide sequence by DNA sequencing (Eurofins, Louisville, KY).

2.4 Tests for positive selection

We focused our selection analysis on single-copy three-way orthologous groups with one or three predicted SP. We found 163 three-way SP-only orthologous groups, among which 150 were single-copy orthologous groups (i.e., single-copy three-way SP-only orthologous groups or single-copy core secretome). Furthermore, 118 single-copy orthologous groups retained a single predicted SP after annotation (i.e., single-copy three-way SP-mixed orthologous groups from the accessory-secretome, hereafter abbreviated as monoSP). As a first method to test for positive selection, we compared evolutionary codon models M8 and M8a (Yang et al., 2000) on 150 core and 118 monoSP single-copy orthologous groups using SELECTON (Doron-Faigenboim et al. 2005). To check whether positive selection was more or less frequent in SPs compared to other (non-SP) genes, we performed the same test in 314 randomly picked single-copy three-way orthologous groups without predicted SP and with the same length distribution as predicted SPs. The evolutionary model M8, in which a proportion of sites are drawn from a category with dN/dS ratio greater than one, i.e., allowing for sites undergoing positive selection, was tested against M8a, in which no site is allowed to have a dN/dS ratio larger than one, i.e., does not allow for positive selection, using a likelihood ratio test with one degree of freedom to determine the statistical probability that the genes evolve under positive selection (Stern et al. 2007). We adjusted chi-squared p-values using Bonferroni's correction for multiple testing in R considering 582 tests.

We also performed McDonald–Kreitman (MK) tests to infer the existence of positive selection (McDonald & Kreitman 1991). MK tests contrast levels of polymorphism and divergence to test for a departure from neutrality in terms of non-synonymous substitutions (i.e., rapid amino-acid changes) while controlling for gene-specific mutation rates. MK tests estimate α, the fraction of amino acid substitutions that were driven by positive selection. To analyze within-species polymorphism, we used genome sequences previously obtained with Illumina paired-end sequencing technology for populations of the three focal species *M. lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum* var *paradoxa* (Whittle et al. 2015;

23

Badouin et al. 2017; Branco et al. 2018). We downloaded raw data publicly available from the NCBI Short Read Archive (SRA) under the BioProject IDs PRJNA295022, PRJNA269361 and PRJEB16741. Four major genetic clusters were identified in Europe in *M. lychnidis-dioicae* (Badouin et al. 2017), and we only considered strains belonging to the largest cluster in North Western Europe so that population subdivision does not bias selection inferences. A list of the isolates used in the analysis is presented in supplemental table ST1. We processed the raw genome data of 18 *M. silenes-dioicae*, 20 *M. lychnidis-dioicae*, and four *M. violaceum* var *paradoxa* isolates to build pseudo-alignments sequences of gene coding sequences within each species using as reference genomes the assemblies reported in GCA_900015465.1 for *M. lychnidis-dioicae*, GCA_900120095.1 for *M. silenes-dioicae* and GCA_900015485.1 for *M. violaceum* var *paradoxa*. First, reads were trimmed for quality (length >50; quality base >10) using the Cutadapt v1.12 software (Martin 2011). We mapped Illumina reads against the reference genomes of each species using bowtie2 v2.1.0 (Langmead et al., 2009) and filtered for PCR duplicates using picard-tools (http://broadinstitute.github.io/picard). We realigned reads, called for SNPs and filtered them for quality, high genotyping rate (>90%) and minor allele frequency (>10%) using GATK version 3.7 (McKenna et al., 2010) and vcftools version 0.1.13 (Danecek et al., 2011) as described previously (Branco et al., 2018; Hartmann et al., 2018). We built pseudo-alignments sequences of gene coding sequences from the VCF file produced by GATK using a customized script. For each strain, reference nucleotides were replaced by their variants in the reference sequence. We used MUSCLE (Edgar, 2004) and translatorX (Abascal et al., 2010) to perform codon-based alignments of gene coding sequences among and between species. We used the MKT() and get.MKT() functions in the POPGENOME Rpackage (Pfeifer et al., 2014) to perform MK tests.

With these tools, we performed three comparisons. We tested for positive selection comparing polymorphism and divergence of 148 core secretome and 115 monoSP orthologous groups for (1) *M. violaceum* var *paradoxa* against *M. lychnidis-dioicae* and *M. silenes-dioicae* strains; (2) *M. silenes-dioicae* against *M. violaceum* var *paradoxa* strains; and (3) *M. lychnidis-dioicae* against *M. violaceum* var *paradoxa* strains. We excluded from the analyses genes having multiple (paralogous) copies. No neutrality index or α value could be computed for 27 orthologous groups in the pairwise species comparison (1), 67 orthologous groups in the pairwise species comparison (2) and 67 orthologous groups in the pairwise species comparison (3), due to lack of synonymous or non-synonymous polymorphism. We performed the same three pairwise

comparisons with 314 genes from the control group described above. No neutrality index or α value could be computed for 30, 99 and 84 in the control pairwise comparisons (1), (2) and (3), respectively. We assessed significance of positive selection for genes having a neutrality index inferior to 1 and a positive α value using a Fisher test (p-value < 0.05).

2.5 Footprints of RIP (repeat-induced point mutations)

We investigated the extent of RIP-like footprints in *Microbotryum* genomes with a per-gene RIP-index defined as the ratio of *t* over *n* (RIP-index=*t/n*), with *t* being the sum of TTG and CAA trinucleotides (forward and reverse potentially RIP-affected targets; Hood et al., 2005) divided by the sum TCG and CGA (forward and reverse non RIP-affected targets), and *n* being the sum of all other non-target trinucleotides [ACG]TG and CA[CGT] divided by the sum of [ACG]CG and CG[CGT], to control for contextual sequence composition. A RIP-index greater than one thus represents an excess of potentially RIPed sites controlling for the base composition. We compared the distribution of per-gene RIP-index values between genes predicted to encode SPs and those not predicted to encode SPs (non-SPs), and considering whether or not the genes belonged to orthologous groups undergoing positive selection.

2.6 Genomic landscape analyses

We used OcculterCut v1.1 (Testa et al., 2016) to determine if *Microbotryum* genomes harbored AT-rich regions. Contigs suspected to contain mitochondrial sequences were removed from the assemblies prior to the analysis using the mito_filter.sh script, available as part of the OcculterCut distribution (https://sourceforge.net/projects/occultercut). Transposable elements locations for *M. lychnidis-dioicae* and *M. silenes-dioicae* were retrieved from a previous study (Hartmann et al., 2018) and predicted in *M. violaceum* var *paradoxa* using the same TE centroid sequence database (Hartmann et al., 2018). Distance to TE was parsed with bedtools (Quinlan and Hall, 2010).

2.7 Intraspecific secretome comparison between M. lychnidis-dioicae isolates from differentiated populations

For analyzing the genome-wide intraspecific variation in secretomes, a second genome (assembly GCA_003121365.1) of *M. lychnidis-dioicae* isolated in Olomouc, Czech Republic, and abbreviated as *M. lychnidis-dioicae* 1318, was analyzed (Hartmann et al., 2018). We used blastp and orthAgogue to obtain high-scoring pairs between gene models of *M. lychnidis-dioicae* 1318 and the entire gene model set analyzed previously (Branco et al., 2018) and re-ran the mcl algorithm. We then parsed the extended orthologous groups to identify the *M. lychnidis-dioicae* 1318 gene models homologous to the *M. lychnidis-dioicae* Lamole SPs identified in this work. We compared the frequency of synonymous and non-synonymous single nucleotide substitutions in codon-based pairwise alignments of *M. lychnidis-dioicae* Lamole and *M. lychnidis-dioicae* 1318 genes corresponding to the core secretome or to the non-SP control single-copy orthologous groups. Per-site substitution numbers were calculated as the sum of substitutions divided by the length of the nucleotide alignment.

2.8 Analysis of gene expression level across infection stages and mating conditions

We retrieved gene expression data across *M. lychnidis-dioicae* Lamole infection stages on *Silene latifolia* and phytol-induced mating conditions from previous studies (Perlin et al., 2015; Toh et al., 2017; Toh et al., 2018) as average log2 fold change (log2FC) against the mated (non-infection) condition (n=2-4 for each of the eight conditions analyzed). We obtained the one-to-one gene model correspondences between long- and short-read sequencing-based assemblies of the same *M. lychnidis-dioicae* Lamole strain as best reciprocal hits with blastp. We focused our analysis on predicted SPs from the core and monoSP orthologs, using gene models from the control set described above for comparisons. Only genes with a Benjamini-Hochberg's adjusted p-value lower than 1e-5 in at least one condition were considered. Clustering and plotting was performed in R with the heatmap.2 function of the gplots package using 10 bins for colouring the log2FC values and clustering by mean values per row. Pie charts were generated with the pie function of R base.

2.9 Plotting, statistical tests, and figures

Unless otherwise stated all plots and statistical tests were performed in R version 3.6.1 (R Core Team, 2019).

Final layout of the figures was produced with Inkscape version 0.92.3.

Results

3.1 Overview of Microbotryum predicted secretomes

Analysis of the three *Microbotryum* secretomes revealed inventories of SPs of similar sizes in all three species. Initial prediction identified around 600 genes with signal peptides in each species (Figure C2-1). Utilizing sequence-based criteria of cellular localization and secretory signals, we kept 302, 371, and 418 SPs in *M. violaceum* var *paradoxa*, *M. silenes-dioicae* and *M. lychnidis-dioicae*, respectively, for further analysis.
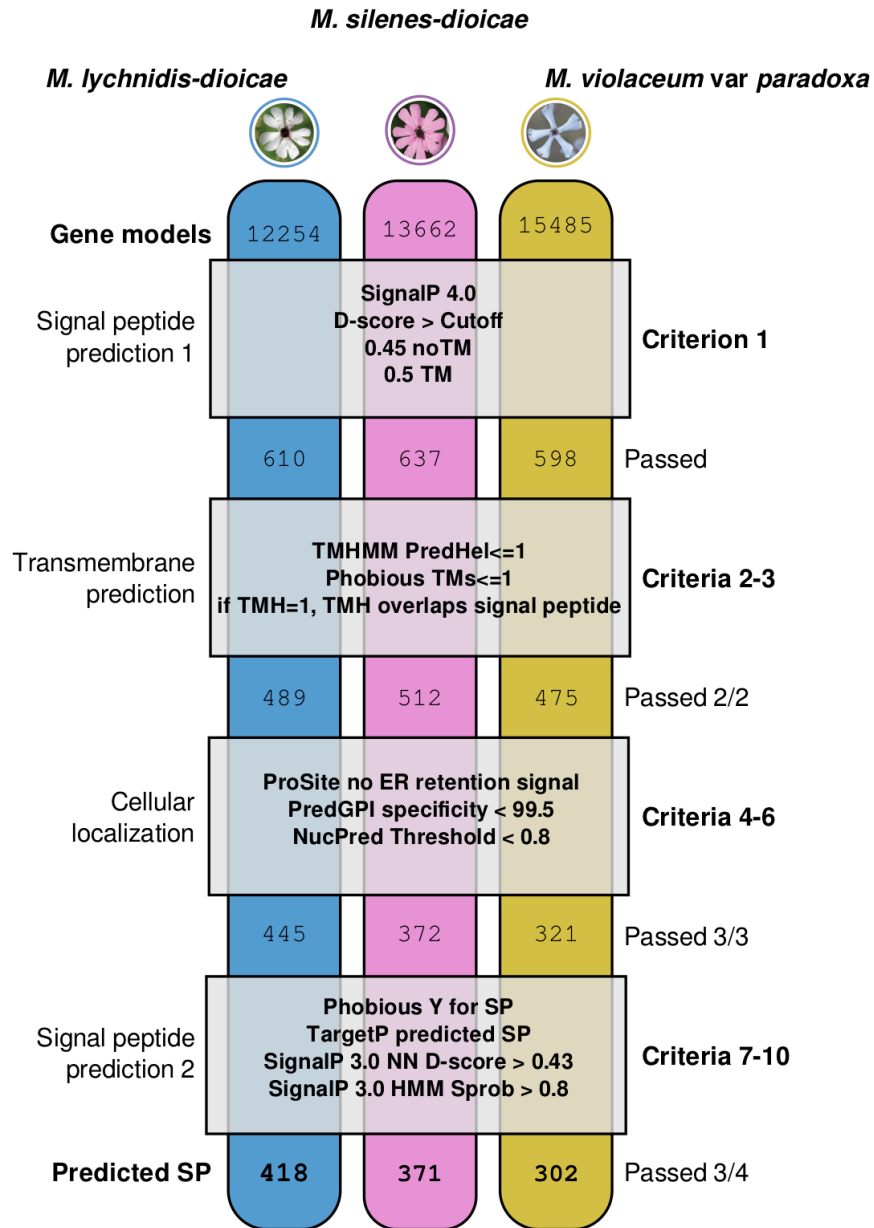
Figure C2-1 Procedural framework for predicting secreted proteins in three *Microbotryum* species.

The genomes for the three fungal species (*M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum* var *paradoxa*) were first screened to identify putative secreted proteins (criterion 1). The resulting proteins were then screened for transmembrane segments (criteria 2-3) and for conflicting cellular localization (criteria 4-6). Candidate secretory peptides were retained for further analysis if they passed all first six criteria (criteria

1-6) plus at least three out of four additional signal peptide prediction cutoffs (criteria 7-10). Each column corresponds to a species, each box to the criteria employed and the numbers to the translated gene models that passed the criteria above. Image from Beckerson et al., 2019.

Over 85% of the predicted SPs were clustered into 453 orthologous groups, 225 comprising exclusively predicted SPs (645 SPs), henceforth called "SP-only", and 239 in which at least one member was not predicted as SP (298 SPs), henceforth called "SP-mixed" (Figure C2-1). Over two thirds of the predicted SPs belonged to orthologous groups with genes in all three species (753 predicted SPs in 163 SP-only and 177 SP-mixed groups). Further, 190 predicted SPs belonged to orthologous groups shared by only two species. Only 148 SPs (i.e., 14% of the total) had no ortholog in two of the species and were therefore classified as species-specific SPs (62 in *M. violaceum* var *paradoxa*, 44 in *M. lychnidis-dioicae* and 42 in *M. silenes-dioicae*). Predicted SPs were significantly depleted in species-specific genes in all three species (Chi-square with Yates correction $p \leq 0.0002$). We classified as "core-secretome" 47% of the predicted SPs (513 genes belonging to 163 SP-only orthologous groups with members in all three species). In 118 SP-mixed orthologous groups with single-copy members in all three species, secretion signals were predicted in the orthologs of a single species, orthologs being non-SPs in the two other species; such orthologous groups will be referred to as "monoSP" hereafter (Figure C2-2 and Supplemental File SF1 from Beckerson et al., 2019).
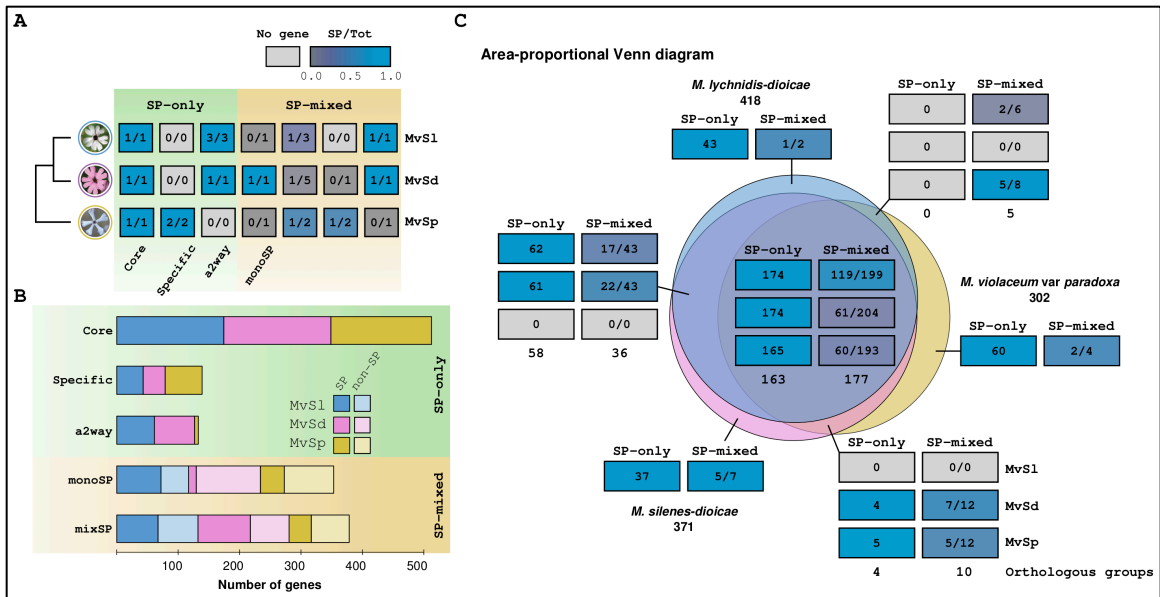
Figure C2-2 Comparison between the secretomes from three *Microbotryum* species. A) Key to the phylogenetic profile of predicted secreted protein (SP) and non-SP homologs with examples for the orthologous group terminology used in this study. Cladogram on the left shows the phylogenetic relationships of the three species. In the SP-only orthologous groups at left, with the light green background, all genes are predicted as secreted. In the core secretome, all three species have at least one predicted SP; in the species-specific orthologous groups, predicted SPs were represented in a single species (i.e., paralogous genes); in the accessory two-way (a2way) groups, one species did not have any ortholog in our reconstruction. In the SP-mixed orthologous groups at right, with the yellow background, not all orthologs were predicted as secreted; for example, in the monoSP group, a single species had predicted secreted proteins in the mono-copy orthologous group. The box color key corresponds to the ratio of predicted SPs over the total number of genes in a given orthologous group per species, with a gradient from blue when all orthologs in all three species are predicted as secreted to dark gray when no ortholog is predicted as secreted. Pale gray boxes represent missing genes in a given orthologous group. B) Stacked bar plots of gene counts in the different categories described in the panel A, with the same terminology, light colors correspond to non-SP homologs of predicted SPs. C) Area-proportional Venn diagram of predicted SP and non-SP homologs, also including species-specific genes. Each area is annotated with six-cell blocks with the number and proportion of predicted SPs in SP-only and SP-mixed orthologous groups, respectively, colored following the same gradient as in panel A. Numbers at the bottom of the blocks correspond to the number of SP-only (left) or

SP-mixed orthologous groups (right). Rows in the blocks correspond to *M. lychnidis-dioicae, M. silenes-dioicae*, *M. violaceum* var *paradoxa*, from top to bottom. Venn diagram was obtained with BioVenn (Hulsen et al., 2008). Abbreviations for all panels: a2way, accessory SP two-way orthologous groups; Core, orthologous groups in which all members are predicted as SP and with at least one gene in each species; mixSP, orthologous groups with both SP and non-SP genes not including monoSP; monoSP, orthologous groups with one gene in each species but with a single predicted SP; MvSl, *M. lychnidis-dioicae*; MvSd, *M. silenes-dioicae*; MvSp *M. violaceum* var *paradoxa*; SP-mixed, orthologous groups with at least one gene not predicted as encoding a SP; SP-only, orthologous groups in which all genes are predicted as encoding SPs. Image from Beckerson et al., 2019.

The majority of SPs for each species were smaller than the median length of all predicted proteins in the three species (57%, 68% and 65% of SPs were smaller than 361 amino acids for *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum* var *paradoxa,* respectively; Figure 3a and Supplemental File SF1 from Beckerson et al., 2019). Initial screening of secretomes showed a high percentages of SPs without known Pfam domains, i.e., 52.1% in *M. lychnidis-dioicae*, 67.9% in *M. silenes-dioicae*, and 62.3% in *M. violaceum var paradoxa*. The percentage of genes without identified Pfam domains was even higher for predicted SPs smaller than 250 amino acids, i.e., 81.7% in *M. lychnidis-dioicae*, 88.9% in *M. silenes-dioicae*, and 84.0% in *M. violaceum var paradoxa* (Figure C2-3b). This trend was further observed when analyzing the subset of core SPs (Figure C2-3b and supplemental file SF1 from Beckerson et al., 2019).
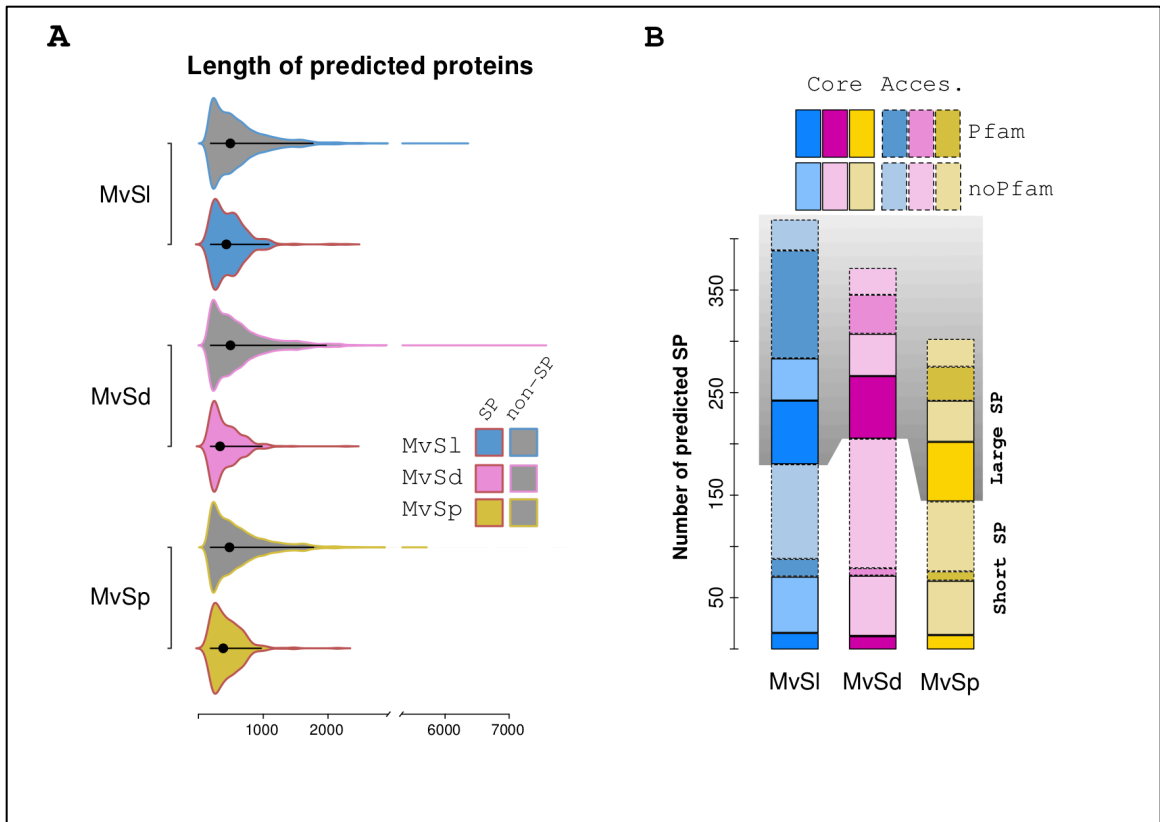
Figure C2-3 Overview of predicted SP (secreted protein) and non-SP homologs. A) Length distribution of predicted SPs (area colored by species) and non-SPs (gray area with outline colored by species) in the three species. Black bars and large black dots indicate the range containing 95% of the points and the median, respectively. B) Pfam screening results for predicted SP in each of the three species. Stacked bars show the number of predicted SPs with (dark colors) and without (light colors) hits among Pfam-A models. Predicted SPs from the core secretome are boxed with a continuous line and those from the accessory-secretome with broken lines. Shaded area corresponds to predicted SPs larger than 250 amino-acids (Large SP in the figure). *Microbotryum* species abbreviations are as in C2-2. Image from Beckerson et al., 2019.

3.2 Signal peptide clusters and yeast secretion trap results

The clustering of the signal peptides of predicted SPs resulted in 280 groups with two or more sequences at 75% sequence identity (823 sequences out of the 1091 predicted SPs). The signal peptides tested here together with the four previously tested (Kuppireddy et al., 2017) are representative of the signal peptides of 28 predicted SPs in the three *Microbotryum* species under study (Figure C2-4). To test whether the predicted

secretion signals can indeed direct secretion, we used an invertase-deficient mutant of *Saccharomyces cerevisiae*. Such mutants can grow on glucose but not on sucrose unless transformed with a plasmid containing the invertase gene with a functional secretion signal, which allows the invertase to cleave extracellular sucrose into glucose and fructose in the medium. Cells of the invertase-deficient mutant SEY6120 of *S. cerevisiae* were transformed with pYST-0 vectors containing each tested signal peptide region upstream and in-frame with the invertase gene. As evidenced by the ability of their respective secretion signals to allow SEY6120 to grow on medium containing sucrose as the sole carbon source, all 9 predicted secreted proteins that have been tested so far using yeast secretion trap have been confirmed to be secreted (Figure C2-4 and Kuppireddy et al., 2017). Interestingly, protein 12964 from *M. violaceum var paradoxa*, was originally filtered out of our list of predicted SPs, due to the prediction that it is GPI-anchored to the membrane. Nevertheless, in this assay using only the secretion signal of the protein, invertase was secreted, suggesting that our conservative approach to estimating secretion may initially filter out membrane proteins with potential functional components outside the fungal cell.
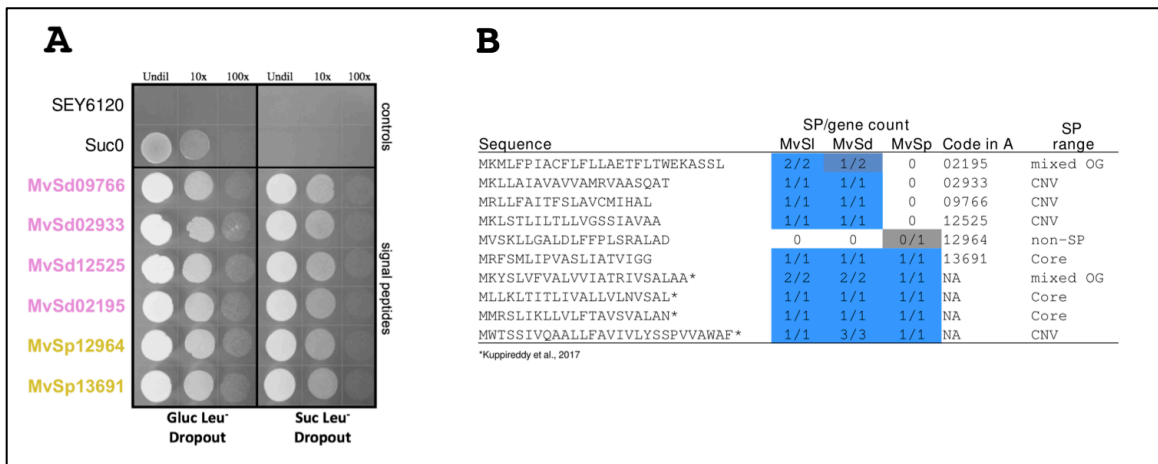
Figure C2-4 Experimental validation of predicted signal peptides. A) Yeast secretion trap analysis of a subset of putative secreted proteins from *Microbotryum silenes-dioicae* and *M. violaceum* var *paradoxa.* The invertase deficient mutant SEY6120 of *Saccharomyces cerevisiae* is shown in the top row and represents a negative control on medium containing sucrose as the sole carbon source. SEY6120 cells transformed with the pYST-0 vector without a signal peptide upstream of the invertase gene is shown in the second row. Such cells are able to grow on the glucose -leu dropout medium, but not when sucrose is the sole carbon source.

The SEY6120 cells in the following six rows are transformed with a construct in which the signal peptide region corresponding to the putative secreted protein ID listed on the left of the row is fused to the truncated *SUC2* gene. If the signal peptide allows secretion, then the transformed *S. cerevisiae* cells are able to grow on sucrose as the sole carbon source. Different dilutions of cells were made (undiluted, diluted 10x or 100x) to better distinguish differences, if any. B) Amino acid sequences and species range of signal peptides tested here and in a previous study (Kuppireddy et al., 2017). Cells under the "SP/gene count" columns follow the same color scheme as in Figure C2-2. *Microbotryum* species abbreviations are as in Figure C2-2. The signal peptide with the code 12964 in panel A corresponds to a protein from *M. violaceum* var *paradoxa* predicted to be GPI-anchored to the membrane. Image from Beckerson et al., 2019.

3.3 Interspecies comparison of Microbotryum predicted secretomes

As expected, due to their phylogenetic placement, the orthologous proteins of *M. silenes-dioicae* and *M. lychnidis-dioicae* were more similar (median identity 98.7%) than either of the two sister groups compared to *M. violaceum* var *paradoxa* (median 86.9% for *M. lychnidis-dioicae* / *M. violaceum* var *paradoxa* and 87.1% for *M. silenes-dioicae* / *M. violaceum* var *paradoxa*). Orthologous SPs, including those belonging to the core secretome, were significantly less similar to one another than control non-SPs from single-copy orthologous groups of similar lengths (Wilcoxon rank sum test with continuity correction $p < 7e-7$ for all three pairwise between-species comparisons, Figure C2-5). Out of the 150 single-copy orthologous groups with a SP predicted in each of the three species, i.e. most of what we call the core secretome (leaving out 13 single-copy orthologous groups with more than one gene in at least one species), we identified 92 groups with codons exhibiting more non-synonymous substitutions than synonymous substitutions. Likelihood ratio tests comparing models with or without positive selection indicated that the model with positive selection was significantly more likely in 18 of these groups (Bonferroni multiple test-corrected p-value <0.05, supplemental file SF2 from Beckerson et al., 2019). Similarly, we identified 74 out of 118 monoSP orthologous groups with codons exhibiting dN/dS values above one, among which multiple test-corrected likelihood ratio tests revealed 21 orthologous groups evolving under positive selection. Selection tests on the 314 control orthologous groups of similar lengths as SPs returned 20 groups evolving under positive selection. Core secretome and monoSP orthologous groups were found enriched in proteins with signs of

positive selection (Fisher's exact text p = 0.02505 for core versus control and p < 0.00048 for monoSP versus control; supplemental files SF1 and SF2 from Beckerson et al., 2019). We found nine core and fourteen monoSP orthologous groups under positive selection with hits in the Pfam-A database (supplemental file SF1 from Beckerson et al., 2019), among which pectinesterase (PF01095.19) and chitin deacetlyase (PF01522.21) have been implicated in fungal biotrophy, potentially for the manipulation of host development (Juge, 2006; Perlin et al., 2015). Glycosyl hydrolases (GHs) (PF00295.17 and PF00704.28) were found in the core and monoSP orthologous groups, despite an overall paucity of GHs represented among *M. lychnidis-dioicae* genes (Perlin et al., 2015). Enzymes of these particular families are interesting due to their ability to hydrolyze pectin, a process important in both pathogenic and saprophytic fungi life stages (Sprockett et al., 2011).

3.4 Intraspecific comparisons of Microbotryum predicted secretomes

We further investigated footprints of positive selection using McDonald–Kreitman (MK) tests that compare the amount of variation within a species (polymorphism) to the divergence between species (substitutions) at two types of sites, synonymous and non-synonymous. A ratio of nonsynonymous to synonymous polymorphism within species lower than the ratio of nonsynonymous to synonymous differences between species indicates positive selection (McDonald & Kreitman 1991). We performed three pairwise species comparisons between *M. violaceum* var *paradoxa*, *M. lychnidis-dioicae* and *M. silenes-dioicae,* using 148 core, 115 monoSP and 314 control orthologous groups. We used population genomics data from 20, 18, and 4 isolates from *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum* var *paradoxa*, respectively (Whittle et al. 2015; Badouin et al. 2017; Branco et al., 2018; supplemental table ST1 from Beckerson et al., 2019). Figure C2-5A shows the locations where the isolates were sampled. The MK tests indicated signatures of within-species positive selection in eight core secretome orthologous groups and fifteen monoSP orthologous groups (supplemental file SF3 from Beckerson et al., 2019). Out of the 23 orthologous groups with signatures of positive selection detected using MK tests, six were also detected to evolve under positive selection in the SELECTON analysis (supplemental file SF1 from Beckerson et al., 2019). Five orthologous groups were found undergoing intraspecific positive selection in all three comparisons. Intraspecific selection tests on control non-SP orthologous groups revealed that 11 underwent positive selection. While core SPs showed no excess of fixed non-synonymous polymorphisms, monoSPs were enriched in genes evolving

under within-species positive selection (15 out of 115 monoSPs versus 11 out of 314 non-SP genes, Fisher's exact test p = 0.0008147).

When we compared two well-assembled *M. lychnidis-dioicae* genomes, those of the Lamole and 1318 strains, originating from two differentiated populations maladapted to their sympatric hosts (Feurtey et al., 2016), we only found 29 Lamole *M. lychnidis-dioicae* SPs without a corresponding 1318 *M. lychnidis-dioicae* gene (12 predicted SPs in 10 orthologous groups and 17 species/strain-specific SPs). In addition, we found 11 orthologous groups for which gene model counts were different between the 1318 and Lamole *M. lychnidis-dioicae* strains. The ratio of SP-containing orthologous groups with gene count polymorphisms between *M. lychnidis-dioicae* strains was significantly smaller than the genome-wide ratio (21/357 SPs vs 2642/12277 all genes, Chi-square with Yates correction p < 1e-11). We found few predicted SPs within genome regions showing presence/absence polymorphism within species as analyzed previously (Hartmann et al., 2018) in both *M. lychnidis-dioicae* Lamole (five) and *M. silenes-dioicae* (two). Substitutions, on the other hand, were more frequent between *M. lychnidis-dioicae* Lamole and *M. lychnidis-dioicae* 1318 strains in predicted SPs than in control genes (Wilcox rank sum test with continuity correction p = 2.537e-05, Figure C2-5c and supplemental file SF4 from Beckerson et al., 2019).
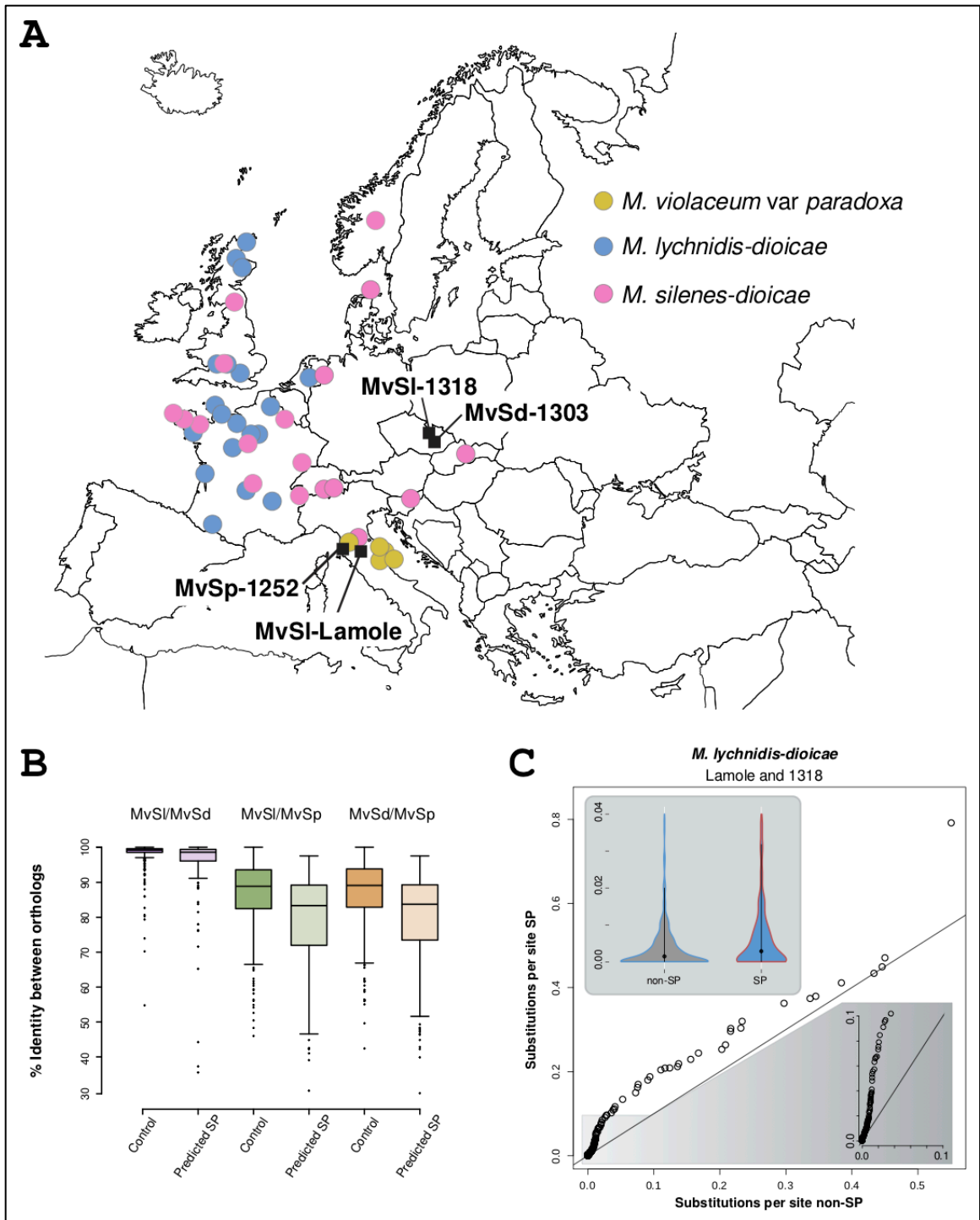
Figure C2-5 Inter- and intra-specific comparisons of *Microbotryum* secretomes. A) Sampling locations of the isolates used in this study. B) Distribution of pairwise percentage of amino-acid sequence identity between predicted SPs and background orthologous genes from *M. lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum* var *paradoxa*. C) Quantile-quantile (main) and violin (inset) plots of substitution numbers per

site between two strains of *M. lychnidis-dioicae* from Lamole, Italy (MvSl-Lamole), and from Olomouc, Czech Republic (MvSl-1318)*. The shaded area at the bottom right zooms into the low divergence zone of the quantile-quantile plot. The straight lines correspond to a 45 degree reference line (i.e., points would fall close to this line if the two data sets have the same distribution). *Microbotryum* species abbreviations in A and B are as in Figure C2-2.

3.5 Genomic context of predicted SPs

In contrast to some other plant pathogenic fungi with effectors frequently located in repeat-rich regions, we did not find genes encoding predicted SPs to be significantly closer to transposable elements than other genes (Figure C2-6) and found no evidence for genome compartmentalization into AT-rich or GC-rich regions in any of the three genomes analyzed, extending previous observations (Perlin et al., 2015). We nevertheless estimated the frequency of sites potentially affected by the RIP-like mechanism reported in *Microbotryum* fungi, targeting TTG and CAA trinucleotides. We calculated a RIP index that takes values above one when there is an excess of TTG and CAA trinucleotides over the corresponding target sites not affected by RIP (TCG and CGA), controlling for local sequence composition (see Methods). The coding regions of predicted SPs did not show any significant excess of RIP-affected trinucleotides, regardless of whether the orthologous groups showed signs of positive selection (Figure C2-6). Our RIP-index measure was negatively correlated with distance to transposable elements (TEs), indicating RIP leakage to TE-neighboring regions. The RIP index was not correlated with the ratio between non-synonymous and synonymous substitutions (Figure C2-6), indicating that the RIP-like mechanism does not play a significant role in the diversification of genes under positive selection in *Microbotryum* fungi.
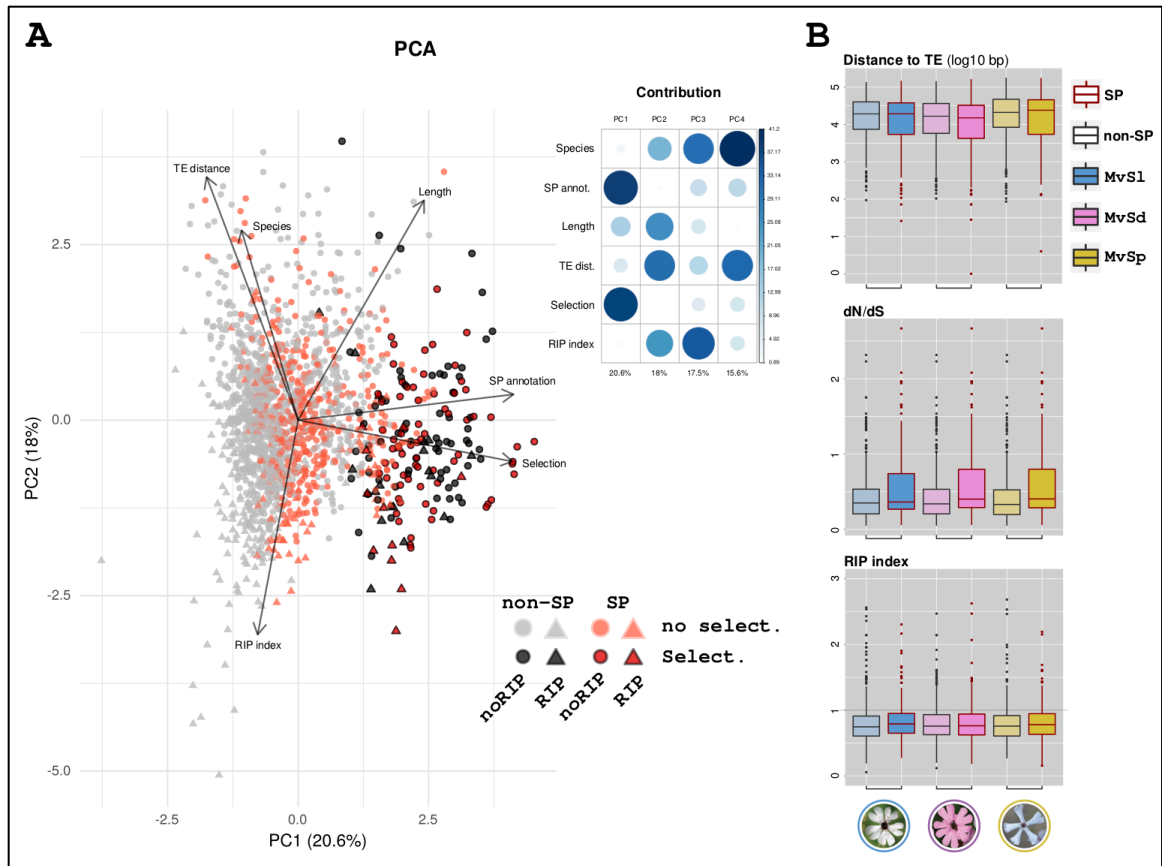
Figure C2-6 Investigation of the impact of RIP (repeat-induced point mutations) on gene diversification among species. A) Principal component analysis (PCA) of gene copies according to their trait value for six variables : (i) their annotation as binary variable, i.e. encoding secreted protein SP (genes colored in red) or non-SP (in grey), (ii) their length in bp as continuous variable, (iii) the species they belong to as category variable (MvSl: *Microbotryum lychnidis-dioicae,* MvSd: *M. silenes-dioicae,* MvSp: *M. violaceum* var *paradoxa*), (iv) their distance to the nearest transposable element as continuous variable (TE distance), (v) their RIP index as continuous variable (RIP-affected gene noted as triangles and non RIP-affected genes as circles) and (vi) the detection of positive selection (genes with dark colors) or the lack of positive selection (light colors) as binary variable. The projection of the variables is plotted as arrows in the space defined by the first (PC1) and second (PC2) components and the percentage of the total variance explained by each principal component is provided in brackets. The arrows representing the variable projection were scaled for better visualization (6-fold magnification). The contribution of the variables to principal components is shown in a correlation plot (upper right). B) TE distance, dN/dS (synonymous substitutions over non-synonymous substitutions) and RIP index distribution of predicted SPs (red contour)

39

or non-SPs (grey contour) in the three species (area colored according to species). Distance to TE was transformed as log10 bp distance; dN/dS was calculated within orthologous groups. The boxplots represent the median (center line), the 25th percentile and 75th percentiles (box bounds), 1.5 times the distance between the 25th and the 75th percentiles (whiskers), and points being the outliers. Image from Beckerson et al., 2019

3.6 Expression of predicted SPs across infection stages

We focused our analysis on *M. lychnidis-dioicae* Lamole genes expressed in at least one of the five infection stages or three mating conditions for which we retrieved expression data (Perlin et al., 2015; Toh et al., 2017; Toh et al., 2018). Among the 2,840 genes fulfilling this condition, we found 135 and 58 predicted SPs from the single-copy core and monoSP orthologous groups, respectively, and compared their expression profiles to 232 genes from the non-SP control group (same length distribution but not predicted as potential effectors). Hierarchical clustering of expression profiles across infection stages grouped the genes into low (31 genes, median log2FC range -7.35 – 4.15), medium (117 genes, median log2FC range 0.0 – 1.8), high (29 genes, median log2FC range 9.19 – 12.40), and no change (248 genes, median log2FC 0) average gene expression across infection stages. We found no major changes in expression of core, monoSP or non-SP genes across three mating conditions. Predicted SPs from the core orthologous groups were enriched among genes with high or low average expression across infection stages, respectively 19 and 18 out of 135 core SPs compared with 7 and 6 out of 232 control genes (Fisher's two tailed exact test p = 1.8E-3 and 1.1E-3, respectively; Figure C2-7). In line with the pattern observed across all predicted SPs, we could infer the function of only 14 core and 7 monoSP genes with either high or low average expression. Glycosyl hydrolases, often involved in pathogenesis (Sprockett et al., 2011), were among the most common hits (supplemental files SF1 and SF5 from Beckerson et al., 2019).
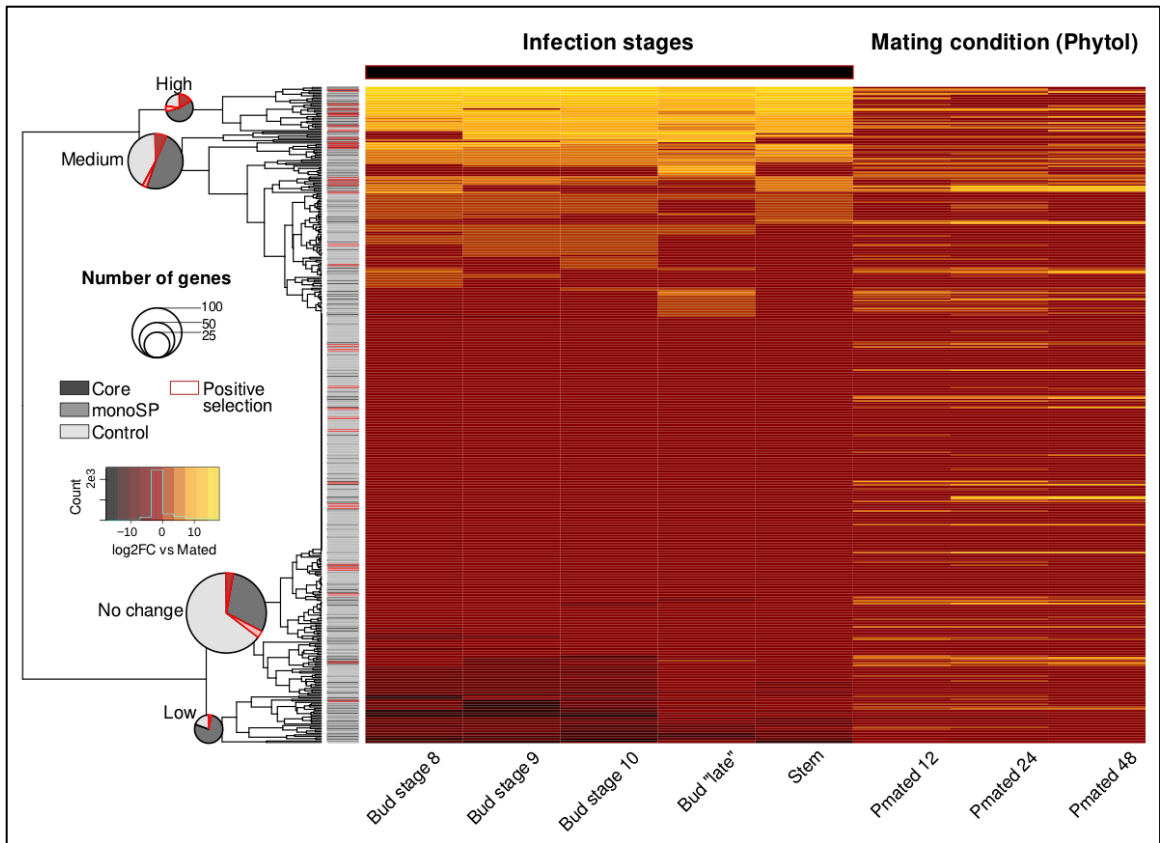
Figure C2-7 Relative expression of *Microbotryum lychnidis-dioicae* genes across infection stages on flower structures. Heatmap of average gene expression (n=2-4) across infection stages in flower structures (Toh et al., 2018) and mating conditions (Toh et al., 2017) as log2 fold change against a non-infection condition (mating on Phytol, "Pmated"). Hierarchical clustering based on mean row values across the infection stages (horizontal black bar) distinguish four expression profiles with average log2 fold change median values as follows: low, -6; no-change, 0; medium, 1.36; high, 12. Sidebar represents the annotation of the genes following the color scheme on the left. Pie charts detail the proportion of SP (core and monoSP) and non-SP (control) genes in each expression profile cluster. Pie chart area is proportional to the number of genes in each expression profile cluster. Red shades and outlines indicate genes with signatures of positive selection. Image from Beckerson et al., 2019.

Discussion

Microbotryum secretomes appeared as largely shared among species, i.e., with few gene gains/losses. Instead, we found SPs to be rapidly evolving as these were more differentiated among species and more often under positive selection compared to non-SP genes, indicating that many SPs likely evolved under diversifying selection among species parasitizing different hosts. Such rapid evolution was also indicated by the low percentage of SPs matching Pfam domains (31-47%), a percentage that decreased to less than 20% for the small secreted proteins. Such a finding regarding the lack of identifiable Pfam domains of a substantial proportion of SPs is consistent with previous reports in other smut pathogens and is a hallmark of secreted effectors involved in host-specificity (Jones et al., 2018). Diversifying selection in *Microbotryum* SPs is likely due to coevolution within species, local adaptation or specialization to different hosts, involving rapid changes in the sequences of secreted proteins to avoid detection in the plant and, more generally, to counteract evolving host defenses. Such a hypothesis is reinforced by the finding that SPs under positive selection were more often highly expressed *in planta* than non-SP genes. Although we found few species-specific SPs or with copy-number variation, these accessory SPs may also be involved in coevolution, local adaptation, and/or host specialization (Plissonneau et al., 2018; Schuster et al., 2018).

The results from the intraspecific comparison between the two *M. lychnidis-dioicae* strains shed further light on coevolution and local adaptation. We indeed found SPs to be more differentiated than non-SPs between two strains from genetically differentiated populations. These findings further support the idea that coevolutionary pressures may be causing divergence in effectors between differentiated populations of pathogens. In fact, the populations from South and Eastern Europe were genetically differentiated in both *M. lychnidis-dioicae* and its host plant *Silene latifolia*, and the plant showed local adaptation to the fungus (Feurtey et al., 2016), indicating the occurrence of coevolution. Gene presence-absence polymorphisms in *M. lychnidis-dioicae,* corresponding to the pathogen and host phylogeographic structure (Hartmann et al., 2018), and numerous selective sweeps across the genome (Badouin et al. 2017), further supported the existence of coevolution. In contrast with several crop pathogens (e.g., Plissonneau et al., 2016; Hartmann and Croll, 2017), neither presence-absence polymorphisms nor selective sweep regions were enriched in predicted SPs, even though nearly 10% of SPs were found located within recent selective sweeps in *M. lychnidis-dioicae*, which suggests recent adaptive events involving some SPs.

The identification of a set of shared and conserved SPs, i.e., the 126 core-secretome orthologous groups without positive selection, was also interesting, providing a starting point to search for effectors that play a central role in the common pathogenicity traits of these fungi, e.g., the effectors that allow the fungi to migrate to the plant anthers, to induce stunted ovary and pseudoanther development in female flowers, and to eliminate and replace host pollen with fungal spores. The observed differential expression of core secreted proteins further narrows the search for these central effectors and points to sets of genes within the secretome that may play other central roles in the fungal life cycle, including the secretion of extracellular enzymes for carbon source metabolism. Indeed, phosphatases, peptidases, lipases and glycosidases accounted for half of the Pfam annotations of core-secretome orthologous groups with no signs of positive selection (20 out of 38). While such enzymes are clearly associated with fungal pathogens (Brown et al., 2015; Monod et al., 2002; Keyhani, 2018), they are often found in animal (Monod et al., 2002; Keyhani, 2018), and necrotrophic plant pathogens (Sprockett et al., 2011; Reis et al., 2005; Gacura et al., 2016), rather than in biotrophic fungi. On the other hand, the up-regulation of many carbohydrate active enzyme genes related to cell wall degradation was also seen in both wheat stem and poplar rust, *P. graminis* and *M. larici-populina*, respectively (Duplessis et al., 2011). In the case of *M. lychnidis-dioicae*, GH28 polygalacturonase domain-containing proteins were up-regulated during infection and were among the proteins with signs of positive selection enriched in the core secretome and monoSP orthologous groups. Since polygalacturonase is required for the pathway implicated in pollen dehiscence (Wang et al., 2016), this is consistent with a fundamental role for such enzymes in the pathogenic lifestyle of anther-smut fungi.

Future research with *Microbotryum* will utilize these findings to better understand the function of the most promising SP candidates, by identifying their targets within each host. Such research geared towards identifying the targets of secreted effectors from *M. lychnidis-dioicae* in its corresponding host plant, *Silene latifolia,* has already made progress (Kuppireddy et al., 2017). For instance, we identified here MvSl-1064-A1-R4_MC02g04003 as part of the core secretome undergoing diversifying selection across species. We also found its transcript among the most highly expressed across infection stages. Its predicted protein product (residues 21-156) has been shown to interact with two host proteins in yeast two-hybrid assays (Kuppireddy et al., 2017). Extension of such work to analyze candidate effectors herein identified through *in silico* studies should add new insights into their relevance in host preference and the evolution of the *Microbotryum* species

complex. By narrowing down the genomes and identifying prime candidates that are likely to play a major role in the pathogen's life cycle, this work helps to bridge the gap between the quickly expanding availability of *Microbotryum* genomes (Branco et al., 2017, 2018; Hartmann et al., 2019) and the emerging cellular and molecular biology work being done to understand the role of effectors in this system (Kuppireddy et al., 2017).

More generally, this study showed that the molecular changes that lead to different host ranges between closely related plant pathogens, or different locally-adapted genetic clusters, involved little gene gains/losses in their secretome but instead rapid evolution of shared secreted proteins. This represents a significant advance in our understanding of pathogen evolution and may contribute to understanding host shifts and emergent diseases.

CHAPTER 3

FUNCTIONAL CHARACTERIZATION OF THE CONSERVED BUT
DIFFERENTIALLY ANNOTATED EFFECTOR, MVLG_02245

Introduction

As the capacity to perform genetic sequencing continues to gain more widespread accessibility, genomic comparisons between closely related organisms has become the standard for delineation of species, especially concerning asexual microorganisms. Robust genome testing is in fact used for rapid identification of clinically relevant microbial pathogens and is capable of both detecting specific strains of microbes as well as characterizing new species (Hasman et al., 2014). The same can be said about using genome sequencing combined with bioinformatic tools to evaluate the coevolutionary trajectories between these pathogens and their various hosts.

Such is the case for the *Microbotryum* genus of anther-smut fungi, for which nearly 20 genomes have been sequenced in the genus over the past few years (Hartmann et al., 2019), and bioinformatic comparisons between species have identified conserved and species-specific effectors amongst the group (Beckerson et al., 2019). While these predictive approaches can help identify rapid evolution of effectors that play a vital role in pathogen/host coevolution, recent studies have demonstrated that bioinformatics analyses, while effective for identification of candidate genes, are not alone sufficient for the identification of every predicted protein function (Pevsner, 2015; Eisenhaber, 2013; Droite, Poirier, and Hunter, 2005;), especially for small secreted proteins which lack Pfam domains and GO terms. Furthermore, while bioinformatic tools are essential for narrowing down the proteome to a list of candidate effectors, they are unable to predict the target molecules within the host for secreted proteins. Thus, molecular genetic analysis of effectors remains an important step in describing host/pathogen relationships.

Like many other fungal pathogens, *Microbotryum* fungi utilize an array of effectors to manipulate their Caryophyllaceae plant hosts; however, despite the rich scientific history and abundance of genomic data

available for this model system, only a handful of genes have characterized using molecular tools (Kuppireddy et al., 2017). In order to infect their hosts, *Microbotryum* possess an inventory of effectors to repress plant defense responses in order to infiltrate and reproduce inside their hosts, as well as to manipulate the host during their migration to the anthers and eventual replacement of the host's pollen with their own fungal spores (Schäfer, 2010). Based on this limited body of preliminary research, the *Microbotryum*/Caryophyllaceae complex does not exhibit a gene-for-gene relationship as seen in other species of phytopathogenic fungi, e.g., rusts (Liu, 2017; Thrall, 2016), which begs the question: what is the role of conserved small secreted proteins within the *Microbotryum* genus? While our previous studies have identified a handful of species-specific genes within the *Microbotryum* genus, bioinformatic comparisons of their secretomes has indicated that host-specialization in the genus is likely due instead to rapidly evolving shared sets of effectors (Beckerson et al., 2019). Of the secreted effectors identified by Beckerson et al., many had orthologous variants predicted to be non-secreted in other species (Beckerson et al., 2019). Therefore, host-specialization in the *Microbotryum* complex may be driven not only by stepwise changes to core-SPs, but also by the mobilization of effectors through changes to the signal peptide region of the gene.

In this research project, we analyze the molecular function of one such potential "mobilized" SP, MVLG_02245, a particularly conserved candidate SP with orthologs across the *Microbotryum* genus, but whose annotation is predicted differently between the first *Microbotryum* genome through JGI (https://mycocosm.jgi.doe.gov/Micld1/Micld1.home.html), and more recently sequenced genomes utilizing PacBio technology. Although more recent PacBio generated assemblies and annotation methods have not categorized MVLG_02245 as a protein, gene expression data suggest that MVLG_02245 is expressed, at least at the transcriptional level, in *M. lychnidis-dioicae*, and upregulated during infection. The changes observed in the signal peptide region of the conserved MVLG_02245 gene across species and the differences in its annotation between previous and more recent genome publications makes this particular putative effector an interesting candidate for both its functional analysis and its evaluation as an interesting case study for the difficulty of describing rapidly evolving effectors using bioinformatics alone.

Materials and Methods

2.1 Bioinformatic analysis of MVLG_02245 in *Microbotryum*

MVLG_02245 was initially predicted to be secreted by the annotation provided through the BROAD Institute (Perlin et al., 2015; currently maintained and updated at JGI); however, upon comparison to the Pacific Bioscience (PacBio) genome for *M. lychnidis-dioicae* (GCA_900015465.1) for *M. lychnidis-dioicae* Lamole, it was concluded that while the DNA coding sequence was found in both genomes the latter method did not predict it to be a protein.

Bioinformatic comparisons were carried out using the MVLG_02245 sequence of *M. lychnidis-dioicae*, retrieved from JGI (https://mycocosm.jgi.doe.gov/Micld1/Micld1.home.html) for BLAST analyses against the PacBio genome sequence of the same strain (GCA_900015465.1) *M. lychnidis-dioicae* Lamole, and the PacBio genomes QPIF00000000 *M. silenes-dioicae* 1303 a2 from *Silene dioica* and GCA_900015495.1 *M. violaceum var. paradoxa* from *Silene paradoxa* 1252 a1. Blastn and Blastp were performed using the NCBI local alignment tool (ncbi-blast-2.7.1+-x64-linux.tar.gz) Pfam 32 and HMMER 3.1b1 suite (https://hmmer.org) were used to screen the JGI translated gene model MVLG_02245 for any known protein families using a cutoff value of 1e-3 for significance. The coding DNA and translated protein sequences were also blasted against the NCBI online database to identify any similar proteins. SignalP4.1 was used to predict the secretion of coding sequences obtained from JGI, as well as to determine the signal peptide and functional protein regions of the corresponding translated protein sequence. PONDR and IUPred2A were used to screen for ordered protein folding of the protein sequence. Expression data were obtained from (Perlin et al., 2015; Toh et al., 2018) to verify production of the MVLG_02245 transcript.

2.2 Yeast Secretion Trap of MVLG_02245

SignalP4.1 was used to predict the signal peptide region of the translated protein for use in the Yeast Secretion Trap (YST) verification of protein secretion signals (Figure 1). Signal peptide regions that code for secretion of the following polypeptide can be used with the YST Suc2 plasmids, in combination with mutant yeast strains, to allow for secretion of an invertase protein capable of breaking down sucrose into its glucose and fructose monomers. Since these mutant yeast strains are not able to transport sucrose into the cell, secreting the invertase enzyme allows for growth of the strain on media where sucrose is the sole carbon source. The

signal peptide region for MVLG_02245 was cloned in-frame and upstream of the plasmid-encoded *Suc2* invertase gene. Standard PCR was used to amplify the signal peptide region using an initial denaturation phase at 96 °C for 5 min, followed by 35 cycles of: 1) denaturation at 96 °C for 30 sec, 2) annealing at 60 °C for 30 sec, and 3) elongation at 72 °C for 30 sec. The program concludes with one final extension period of 5 min at 72 °C before maintaining a 4°C temperature indefinitely. The resulting amplified signal peptide coding sequence was purified from agarose gel after gel electrophoresis and subcloned into pCR™2.1-TOPO™ vector using an Invitrogen TOPO™ TA Cloning™ Kit (Invitrogen, ThermoFisher Scientific, XX). The fragment was then digested from the TOPO vector, purified from agarose gel using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA), and cloned into the Suc2 vector, pYST0 (Lee and Rose, 2012). Proper in-frame placement of the signal peptide sequence was verified through DNA sequencing (Eurofins, Louisville, KY). The resulting plasmid was subsequently named Suc2_MVLG_02245sp.

>**MVLG_02245**   Signal Peptide   Protein Sequence
MTSQVRMQVESRAQRRAGAYASMRLLLALVFALCTLAHLPTTSAAPLASEQISSGLVFRQ EPPRWLQFSRPHEKVSHQGKDHLDWKNTSPSPFTSSEPSRRVKRDEMWEQYIEGDEIDGE KSEDVRAGDPDVAGDEVLTDTEIAGGADEAGEGSTGEKWWQARRRLRERRSATTRVVP

Figure C3-1 The signal peptide region of MVLG_02245 predicted using SignalP4.1.

The Suc2_MVLG_02245sp construct was transformed into the invertase-deficient (suc2-negative) *Saccharomyces cerevisiae* strain SEY 6210 (MATα leu2-3,112 ura3-52 his-Δ200 trp1-Δ901 lys2-801 suc2$^-$ Δ9 GAL), using the Frozen-EZ yeast transformation II kit from Zymo Research, and plated onto both glucose and sucrose synthetic dropout media lacking Leucine (SDO$_{Leu-}$). Colonies that grew on the Sucrose SDO$_{Leu-}$ plates were grown overnight in 3 mL of liquid Sucrose SDO$_{Leu-}$ media and plated onto both glucose and sucrose SDO$_{Leu-}$ assay plates in undiluted (10^7 cells/mL), 10X dilution (10^6 cells/mL), and 100X dilution (10^5 cells/mL) concentrations. These assay plates were incubated at 30°C for 2 days before images were taken. To verify transformation, plasmids were re-extracted from the yeast colonies formed during the assay using a general yeast miniprep protocol (Protocols: Yeast Miniprep), transformed back into *E. coli* cells, and purified for repeat DNA sequencing to confirm the proper in-frame cloning of the construct (Eurofins, Louisville, KY).

2.3 Yeast Two-Hybrid (Y2H) assay of MVLG_02245 targets

Using the NEBuilder HiFi DNA Assembly Master Mix kit, the fragment encoding the protein sequence for MVLG_02245 was cloned into the Y2H bait vector (pGBKT BD; Clontech, Mountain View, CA) without the signal peptide region (as determined by SignalP 4.1) and transformed into AH109 yeast cells using the Zymo Frozen EZ-Yeast Transformation II kit. The resulting transformants were mated against Y187 yeast cells containing the prey vectors bearing the cDNA library from *S. latifolia* infected with *M. lychnidis-dioicae* (Kuppireddy et al., 2017). The cDNA library in prey vector (pGAT7 AD, Clontech) was generated by CD Genomics (Shirley, NY, USA) as described previously (Kuppireddy et al., 2017). AH109 cells containing the MVLG_02245 bait vector were grown overnight in Trp⁻ Single Dropout (SDO$_{Trp-}$) liquid media. Cells were pelleted via centrifugation using a Labnet Hermle Z 233 M-2 centrifuge at 2,000 rpm and resuspended in 5 mL of 2X YPDA before being added to a 1 L flask along with 1 mL of Y187 cells containing the prey vectors and 45 mL of 2X YPDA containing 50 µg/mL Kanamycin. Cells were then gently shaken at 45 rpm on a platform shaker at 30 °C overnight to allow for mating. Successful mating was verified via microscopy (Figure C3-2) before cells were collected at 1,000 rpm, washed, and resuspended in 5 mL of sterile distilled water.
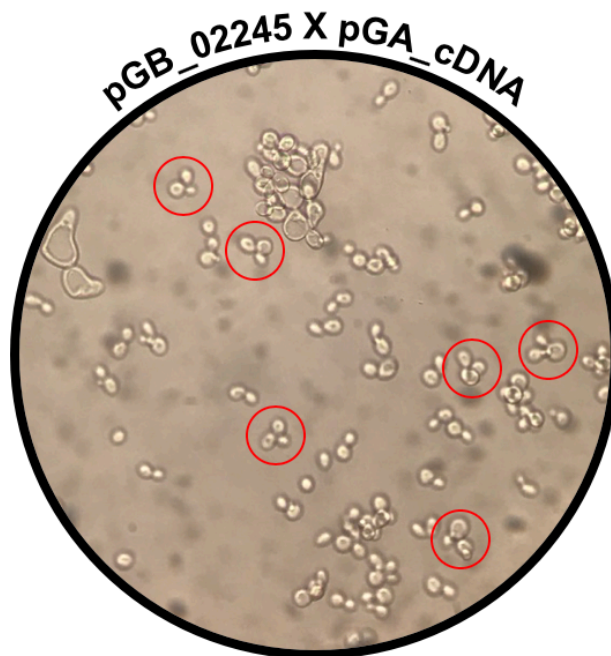
Figure C3-2 Mating verification between *Saccharomyces cerevisiae* cells. Expressing the pGB_02245 vector and cDNA library generated prey vectors.

Mated cells were gently centrifuged with a Labnet Z233M-2 centrifuge at 1,000 rpm for 10 min, and washed twice with diH2O before 1,000X and 10,000X dilutions were plated onto Trp⁻ Leu⁻ Double Dropout (DDO~Trp- Leu-~) plates and Trp⁻ Leu⁻ Ade⁻ His⁻ Quadruple Dropout (QDO~Trp- Leu- Ade- His-~) agar plates with 25 mM 3AT (Sigma-Aldrich), x-α-gal (Sigma-Aldrich), and 50 μg/mL Kanamycin. Both sets of plates were incubated at 30 °C for 3 days. Resulting blue colonies, indicating protein-protein interaction through the upregulation of galactase via localization of the DNA binding and activation domains, from the QDO~Trp- Leu- Ade- His-~ plates were re-streaked onto QDO~Trp- Leu- Ade- His-~ media containing 50 mM 3AT to reduce growth due to leaky expression of the *HIS3* gene that can confound screening for true interactions. Yeast minipreps to extract the bait and prey vectors from mated diploid cells were performed on colonies that grew on the 50 mM 3AT QDO~Trp- Leu- Ade- His-~ plates. The plasmid mixtures were transformed via heat shock into competent DH5α *E. coli* cells and plated onto LB agar containing 200 μg ampicillin per mL, to preferentially select for the prey vectors. These plates were incubated overnight at 37 °C and plasmid DNA was extracted from resulting bacterial colonies. Such plasmids were sequenced (Eurofins) using primers upstream of the cDNA insertion. Sequencing results were then used in Blast searches for orthologs against the NCBI database to identify the host target. To verify true interactions, the bait and prey vectors for the MVLG_02245 coding region and each of its putative interaction partners were swapped using the NEBuilder HiFi DNA Assembly Master Mix, and mating was repeated and again confirmed via light microscopy (Figure C3-3)
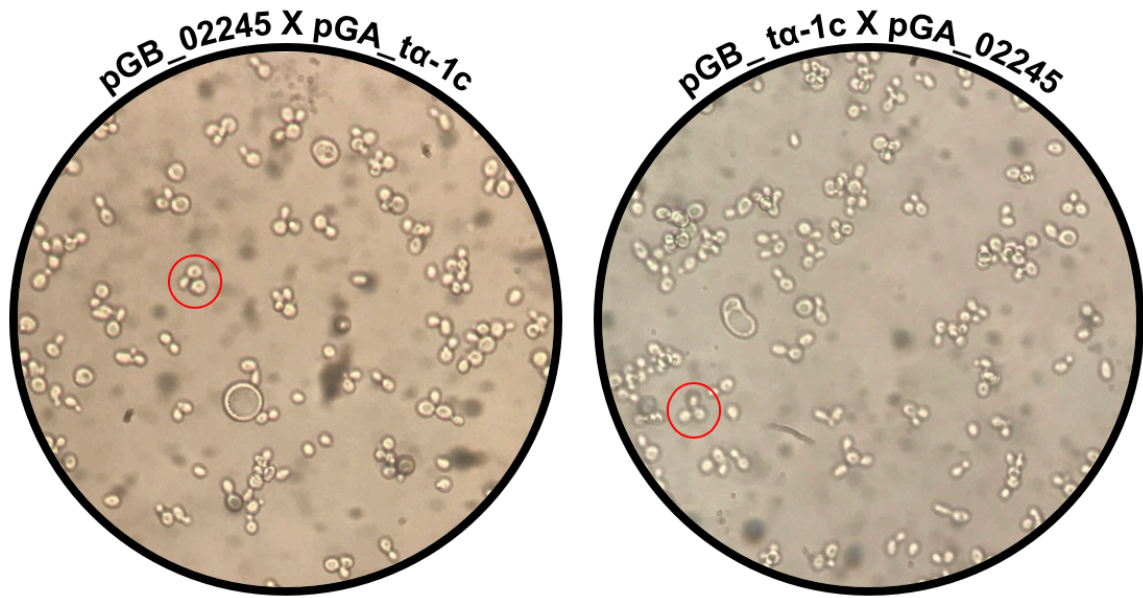
Figure C3-3 Repeat mating verification between *Saccharomyces cerevisiae* cells. Expressing the pGB_02245 vector and target match pGA_tα-1c from the cDNA prey library.

A spot assay for true interactors was prepared on both DDO$_{Trp- Leu-}$ and QDO$_{Trp- Leu- Ade- His-}$ media using undiluted (10E7 cells/mL), 10X dilution (10E6 cells/mL), and 100X dilution (10E5 cells/mL) concentrations of cell suspension. As a control, single vector colonies and mated controls were similarly prepared and spotted DDO$_{Trp- Leu-}$ assay plates were incubated at 30 °C for 2 days, while the QDO$_{Trp- Leu- Ade- His-}$ assays were incubated at 30 °C for 4 days before images were taken.

Results

3.1 Local BLAST results for predicted MVLG_02245 coding sequence

Local BLAST alignment of the DNA and protein sequence for MVLG_02245 demonstrated that the coding sequence for MVLG_02245 is conserved across multiple species of *Microbotryum* (Figure 3). While amino acid sequence was conserved between the sister species *M. lychnidis-dioicae* (MVLG/MvSl) and *M. silenes-dioicae* (MvSd), with only 4 amino-acid substitutions resulting from 6 nucleotide substitutions, the protein sequence was more different in the distantly related *M. violaceum var. paradoxa* (MvSp), with 29 amino acid substitutions resulting from 51 nucleotide substitutions (Figure C3-4, Supplemental Table C3-1), when

compared to *M. lychnidis-dioicae*. No insertions or deletions, and subsequently no frame shifts, were observed in the MVLG_02245 gene in any of the three species.

Figure C3-4 BlastP comparison between three species of *Microbotryum* for MVLG_02245. The sequence for MVLG_02245 retrieved from JGI was compared to that from the PacBio assembly for the same strain (MvSl), as well as the PacBio assemblies for its sister species (MvSd) and a more distantly related species of *Microbotryum* (MvSp). Amino acid substitutions, with reference to the MvSl strain, are shown in bolded letters color coded by their species (pink for MvSd, gold for MvSp).

When compared to the amino acid sequence for MVLG_02245 in *M. lychnidis-dioicae*, the *M. silenes-dioicae* and *M. violaceum var. paradoxa* orthologues only shared one amino acid substitution, G_{146} -> E_{146}. Interestingly, while there were no differences in the signal peptide region for MVLG_02245 found in *M. lychnidis-dioicae* relative to *M. silenes-dioicae*, the corresponding region of *M. violaceum var. paradoxa* had 6 substitutions out of 44 amino acids for this region (Figure C3-4).

3.2 Predicted secretion, function, and Expression of the MVLG_02245 effector

Screening of MVLG_02245 against SignalP 4.1 indicated that MVLG_02245 is predicted to be secreted (Supplemental Table C3-2), which is in accord with previous findings by Perlin et al., (2015); Toh et al., (2018). RNASeq further demonstrated that mRNA is present for MVLG_02245 and upregulated during infections in the floral and floral stem tissues of the host (Table 1; Perlin et al., 2015; Toh et al., 2018).

Table C3-1: Expression of MVLG_02245 on various media and *in planta*, presented as TPM[a]

Water	Rich Media	Mating on Nutrient-limited Medium	Male Infected Tissue: Stage 8	Male Infected Tissue: Stage 9	Male Infected Tissue: Stage 10	Male Infected Tissue: Late	Male Infected Stem Tissue
1	5	1.13	**73.23**	**53.65**	**41.88**	4.98	**30.06**

a – transcripts per million; Male-infected Stem tissue, male-infected floral stem tissue. 2-3 independent determinations via RSEM.

However, while the MVLG_02245 gene in *M. lychnidis-dioicae* and *M. silenes-dioicae* is predicted to be secreted, the amino acid substitutions observed in the signal peptide region of MVLG_02245 for *M. violaceum var. paradoxa* are not predicted to abrogate secretion of the protein (Supplemental Table C3-1). To predict whether the MVLG_02245 protein may play a role in manipulating the host, the amino acid sequences for all three species of *Microbotryum* were run against the online effector prediction software, EffectorP 1.0 (Sperschneider et al., 2015). The MVLG_02245 protein was predicted to be an effector in all three species, with a probability score of 0.865, 0.686, 0.645 for *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum var. paradoxa*, respectively (Supplemental Table C3-2). To screen for any shared sequence with known effectors, we used the Pfam 32 and HMMER 3.1b1 tools to screen for protein families; however, neither of the programs yielded any significant results (Supplemental Table C3-2).

The predicted effector function in all three species, the combination of predicted secretion in the sister species pair, and the lack of a Pfam domain indicates that MVLG_02245 is likely an effector with a unique function for the infection of *Silene* hosts, specifically that of the sister species given its lack of

mobilization in *M. violaceum var. paradoxa*. Evolutionarily speaking, this could either indicate genetic drift in the gene due to a lack of use in the *S. paradoxa* host, or represent positive selection of the effector in different evolutionary trajectories since divergence of *M. violaceum var. paradoxa* from the sister-species progenitor. Prediction that MVLG_02245 is a genus-specific effector for *Microbotyrum* is further supported by a lack of blastn and blastp hits to sequence outside the *Microbotryum* genus when screened against the general NCBI genome database. Furthermore, MVLG_02245 is predicted to be a disordered protein (Figure C3-5). Intrinsically disordered proteins allow for flexibility and have been described as another hallmark for small-secreted effectors for various pathogens (Marín et al., 2013), including in *Microbotryum* as outlined by Kuppireddy et al., (2015).
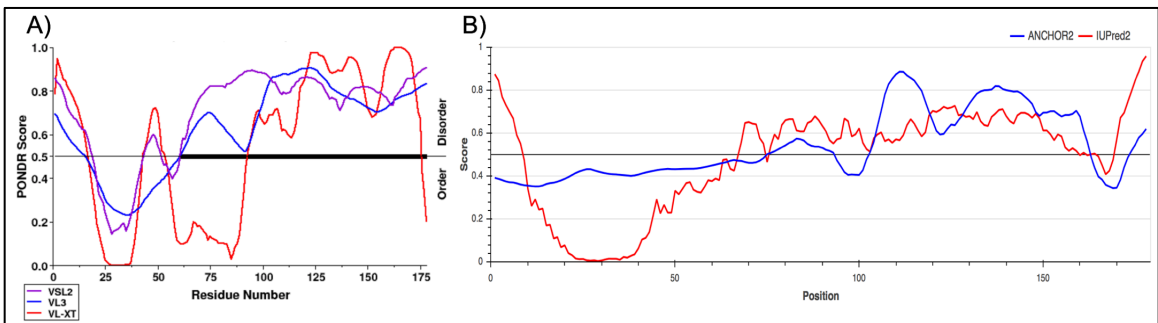
Figure C3-5 Prediction of disorder in the protein sequence for MVLG_02245. using A) PONDR and B) IUPred2A.

3.3 Yeast Secretion Trap results

The capacity for the *M. lychnidis-dioicae* and *M. silenes-dioicae* MVLG_02245 signal peptide to signal for protein secretion was confirmed through YST testing of the signal peptide region, predicted by SignalP 4.1 (Figure 5). SEY cells transformed with the Suc2_MVLG_02245sp vector were able to secrete the invertase enzyme to breakdown sucrose and survive on sucrose as a sole carbon source (Figure C3-6).
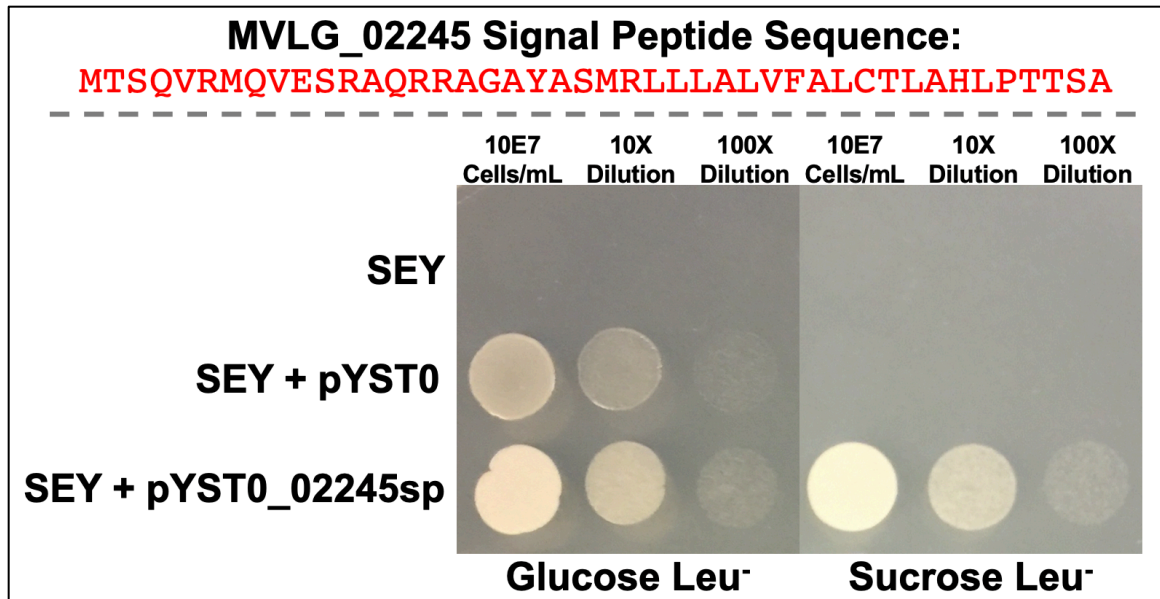
Figure C3-6 Yeast Secretion Trap results for the signal peptide region of MVLG_02245. The figure shows yeast colonies after 2 days of growth on Glucose Leucine Dropout Media, left, and Sucrose Leucine Dropout Media, right. In the top row are untransformed SEY strain cells of *Saccharomyces cerevisiae*. The second row contains SEY cells transformed with just the pYST0 vector. The third row contains SEY cells transformed with the pYST0 vector containing the signal peptide from MVLG_02245, as predicted by SignalP 4.0, cloned upstream and in-frame of the invertase enzyme. The signal peptide sequence is shown in red above the spotting assay.

3.4 Host targets of the MVLG_02245 effector in *Silene latifolia*

With an abundance of bioinformatic evidence suggesting that MVLG_02245 is a secreted effector in the *Microbotryum* genus, we tested the *M. lychnidis-dioiae* variant for a host target using a Y2H approach. Y2H mating assays between AH109 cells containing the MVLG_02245 bait vector with Y187 cells containing the infected plant tissue cDNA prey vector library yielded blue colonies on 50 mM 3AT QDO$_{Trp-\ Leu-\ Ade-\ His-}$ media. Plasmid extraction followed by sequencing for 50 of these diploid colonies demonstrated that four particular proteins were found more predominantly than others. The sequences for these identified plant proteins were compared against the NCBI database and yielded plant orthologs for a ferredoxin-thioredoxin reductase catalytic chain protein, a Photosystem II protein, a xyloglucan endotransglucosylase/hydrolase 4 protein, and a Tubulin α-1 chain protein (tα-1c). Only the fourth interaction, MVLG_02245 X tα-1c, yielded

blue colonies when the coding regions were swapped between the bait and prey vectors and mating was repeated (Figure 6). While the interaction between MVLG_02245 and tα-1c was confirmed in the vector swap, the resulting diploids grew slower in both the DDO and QDO media (Figure C3-7).
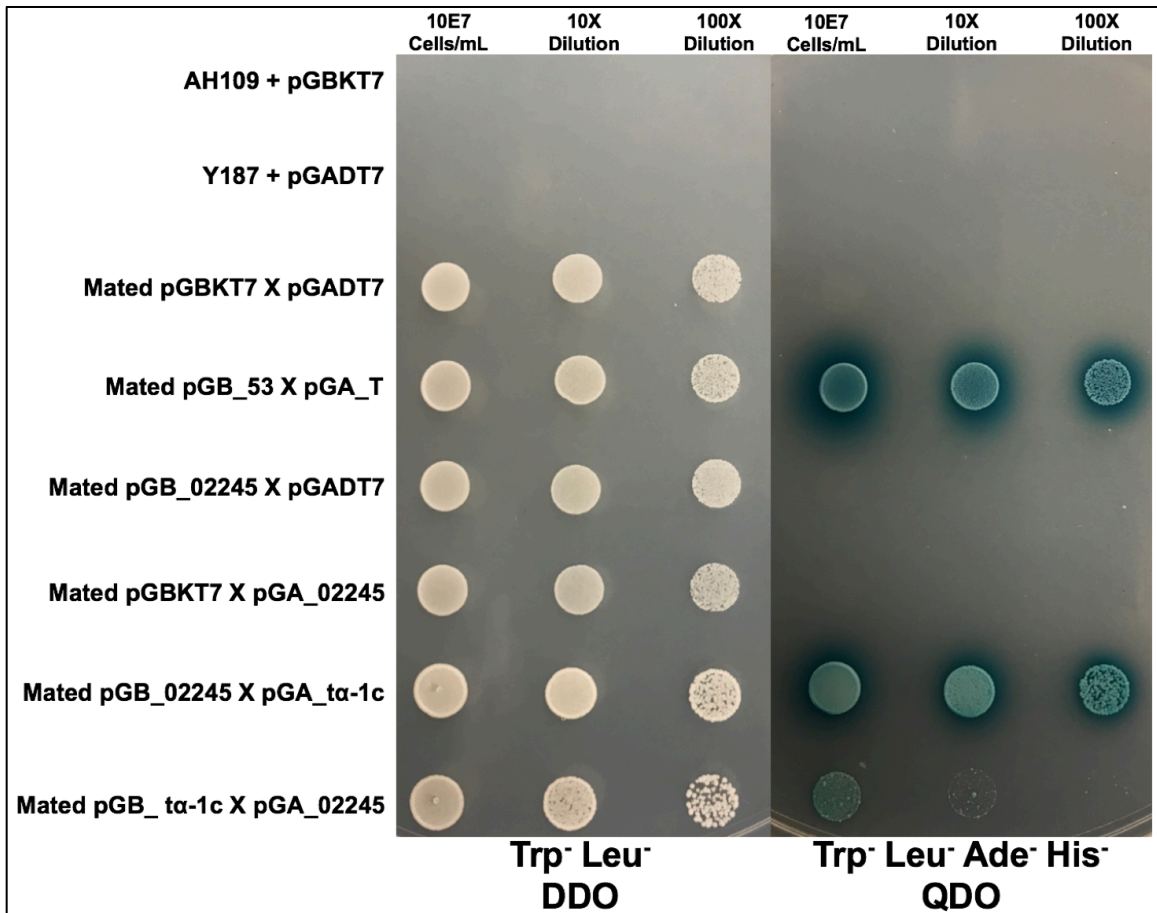
Figure C3-7 Yeast two-hybrid mating results between MVLG_02245 and tα-1c. Colonies are shown after two days of growth on DDO, left, and 4 days of growth on QDO, right. A series of negative controls were used including the AH109 yeast strain transformed with an empty bait vector (pGBKT7), top row, the Y187 strain transformed with an empty prey vector (pGADT7), second row, Diploid offspring of mated strains containing both the empty bait and empty prey vectors, third row, Diploid cells containing the MVLG_02245 bait vector and the empty prey vector, fifth row, and Diploid cells containing the empty bait vector with the tα-1c prey vector, sixth row. Diploid cells containing the bait and prey vectors for known strong interactors p53 and T-antigen were used as a positive control in the fourth row. Diploid cells containing

the MVLG_02245 bait vector and tα-1c prey vector are spotted in the seventh row, and diploid cells containing the swapped tα-1c bait vector and MVLG_02245 prey vector are spotted in the seventh row.

Discussion

Our findings demonstrate that, despite a difference in annotation, MVLG_02245 is indeed a secreted protein which likely acts as an effector utilized by *M. lychnidis-dioicae* during infection of its host, *Silene latifolia*. Interestingly however, the limited number of differences observed in the predicted secretion signal peptide sequences between the MVLG_02245 gene of the two sister species, *M. lychnidis-dioicae* and *M. silenes-dioicae,* compared with the more extensive changes for the corresponding region found in the more distantly related *M. violaceum var. paradoxa*, mirrors the host specificity that is observed between the three groups. While *M. lychnidis-dioicae* and *M. silenes-dioicae* perform much better at infecting their own hosts, *S. latifolia* and *S. dioica*, respectively, they are able to infect each other's hosts, albeit to a reduced degree compared to their natural hosts (de Vienne et al., 2009; Putten et al., 2003). This is in contrast to *M. violaceum var. paradoxa*, which is unable to infect either of the others' hosts, and *vice versa* (de Vienne et al., 2009). These observations are consistent with findings from Beckerson et al. (2019), where "SP-Mixed" groups were identified in which all three species contained orthologs for certain effectors, but only one or two species in the 3-way comparison were predicted to secrete the protein (Beckerson et al., 2019). These SP-Mixed groups are suspected to play a role in host-specificity through "mobilization" or loss of secretion in proteins (Beckerson et al., 2019).

While our Y2H results have identified an interaction partner for the MVLG_02245, a tubulin α-1 chain protein ortholog, its exact purpose for binding to the protein *in planta* still remains unclear. These tubulin proteins act as a linker in microtubule production in the host, and therefore MVLG_02245 may act to prevent tubulin α-1 chain proteins from linking with their β chain counterparts (Hashimoto, 2015). Doing so can disrupt structural arrangement of the plant cell cytoskeleton, mitosis, and other microtubule-based processes, which can have a wide variety of effects in the plant ranging from structural instability to disruption of microtubules that traffic vesicles vital for the plant immune response (Büttner, 2016). Experiments performed on effectors targeting tubulin chains by Lee et al., secreted by the bacterial phytopathogen *Pseudomonas syringae pv.* tomato, demonstrate an ability for these proteins to destabilize

microtubules and interfere with the host's secretion of reporter proteins to the apoplast (Lee et al., 2012). These microtubules also play a role in cell division, and their disruption may play a role in regulating growth of the host plant. Thus, it is conceivable that MVLG_02245 may ultimately weaken both the plant cell and the host immune response, making the host more susceptible to penetration by the fungi. The upregulation of MVLG_02245 in the floral stem tissue and its downregulation in late-stage infection of floral tissue of the host lends even more credence to the effector's use during primary infiltration and establishment in the plant.

To verify and further expand on these predictions, future studies will utilize GFP tagged MVLG_02245 for localization studies of the secreted protein within the plant host. To prepare for this work, the coding sequence for MVLG_02245 was cloned into the *Agrobacterium*-mediated transformation (ATMT) vector for *M. lychnidis-dioicae*, pMvHyg (Toh et al., 2017), using the NEBuilder HiFi DNA Assembly Master Mix kit, along with both the constitutively expressed promoter MVLG_05589-P and the coding sequence for GFP, fused in-frame to the MVLG_2245 coding region along with an intervening 2X Gly-Gly-Ser residue linker. The resulting construct, pMvHyg_MVLG_02245-GFP (Figure C3-8), was transformed into EHA105 *Agrobacterium* cells using electroporation (2.5 kV, 400 ohms and 25 μF) and transformants were selected on LB agar-containing 50 μg kanamycin/mL. Successful transformants were re-streaked onto LB containing 50 μg kanamycin and 100 μg spectinomycin per mL agar to ensure that both the pMvHyg and Helper plasmids were in the cells. Transformed EHA105 cells were further confirmed through colony PCR, before being used to transform *M. lychnidis-dioicae* using the ATMT protocol developed by Toh et al., (2016).
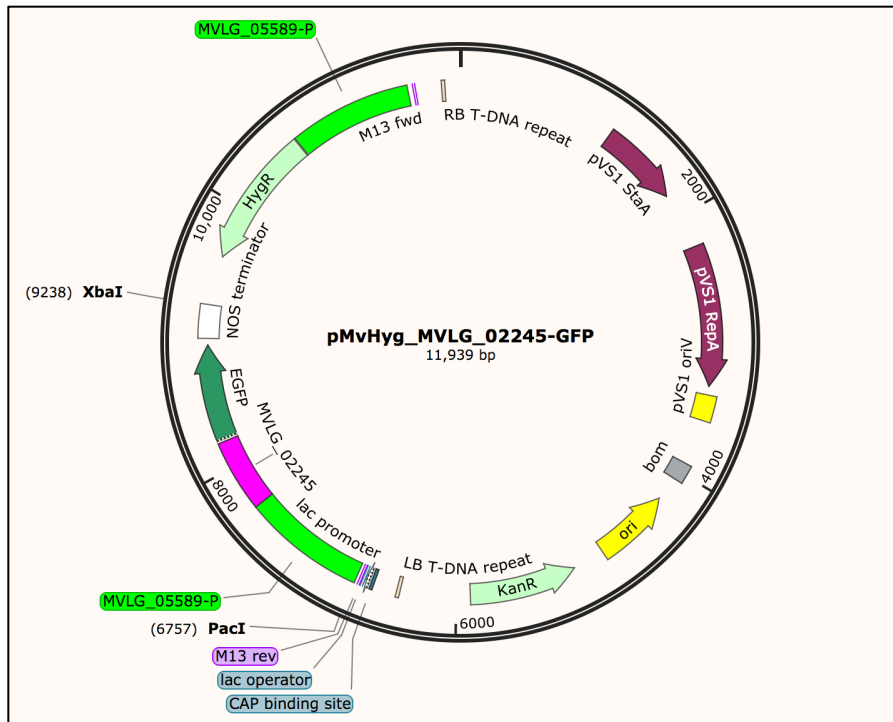
Figure C3-8 SnapGene image of pMvHyg_MVLG_02245-GFP. pMvHyg_MVLG_02245-GFP contains the *Microbotryum lychnidis-dioicae* gene MVLG_02245 fused in-frame at the 3' end with the GFP coding region. Expression of the MVLG_02245-GFP protein is driven by the constitutively expressed promoter. All of these components are cloned in-between the left and right border of the *Agrobacterium*-mediated transformation vector, pMvHyg. pMvHyg (Toh et al., 2017) also contains a hygromycin resistance gene included within the transfer region, also driven by the MVLG_05589-P promoter, to select for transformants after integration into the *M. lychnidis-dioicae* genome. (https://www.snapgene.com/snapgene-viewer/).

Prior to ATMT, the *M. lychnidis-dioicae* p1A1 strain was grown on yeast peptone dextrose agar with 10% dextrose (YPD-10%) for 2 days at 24°C. Cells were resuspended in 1 mL of induction medium broth and diluted to an OD_{600} of approximately 10E7 cells/mL. The same was done for the transformed EHA105 cells, approximately 10E8 cells/mL. A 1:1 ratio *Microbotryum* to *Agrobacterium* cells was mixed together and spotted onto IM plates using 100 μL of each cell suspension before being incubated at room temperature for 3 days, after which the resulting mass of cells were retrieved from the plates using sterile plastic loops and resuspended in 600 μL of YPD-10% liquid media. 200 μL of each suspension were then

spread onto YPD-10% containing 150 μg/mL Hygromycin B containing 100 μg/mL Cefoxitin plates. These plates were incubated at 24°C for 14 days. Resulting transformants were re-streaked onto YPD-10% containing 150 μg/mL Hygromycin B and used for genomic extractions and PCR screening for successful integration of the transfer plasmid. Verified transformants were used for confocal imaging of GPF and infection of *S. latifolia* to observe cellular localization of the MVLG_02245 in the host.

The findings in this research not only provide a plausible characterization for one of the many conserved core secreted proteins in the *Microbotryum* genus, but also provide an example of the necessity for molecular genetic studies to corroborate bioinformatic predictions. Bioinformatic predictions and functional analyses are therefore synergistic and intertwined, neither of which can be as productive without the other. Furthermore, because there are many atypical mechanisms at work in various species, especially those not as well defined as model organisms, caution should be taken when eliminating candidate genes from annotations. While predictive models are essential for initial work involving genome characterization, these models should be just that, a starting point for an eventual complete and more inclusive analysis.

CHAPTER 4

ASSESSING THE ROLE OF SPECIES-SPECIFIC EFFECTORS IN HOST-
SPECIALIZATION

Introduction

While gradual stepwise changes to the amino acid sequence of core secreted proteins (SPs) can lead to coevolution between a pathogen and its hosts, the emergence of novel SPs could lead to rapid evolutionary changes. The acquisition of novel secreted proteins, in this sense defined by an effector for which the host has no defensive response, can arise from a variety of biological processes including horizontal gene transfer (Casa-Esperón, 2012), sexual recombination (Grigg & Suzuki, 2003), mobilization of a previously non-secreted protein (Beckerson et al., 2019), and gene duplication (Andersson, Jerlström-Hultqvist, & Näsvall, 2015). While the emergence of a new effector can give the pathogen an immediate edge in the evolutionary tug-of-war with its host, we would expect the hosts to adapt over time to new effectors in a coevolution model. Therefore, unique SPs may act as evolutionary milestones and provide some insight into the divergence of species over time. Furthermore, these novel proteins may play a pivotal role in the pathogenicity of different species, contributing immensely to host specificity among pathogens separated by post-zygotic barrier, such as is seen in the *Microbotryum* genus. In this investigation, we identify secreted proteins unique to the two sister taxa of *Microbotryum*, *M. lychnidis-dioicae* and *M. silenes-dioicae*, and initiate experiments to evaluate their roles in pathogenicity on each host plant. To do so, we utilized a bioinformatic approach to compare eight annotated *Microbotryum* genomes for unique proteins with hallmarks of secretion, verified by parameters set in Beckerson et al., 2019. Finally, to test the role of species-specific effector in pathogenicity, a subset of these identified species-specific effector candidates were heterologous expressed in each sister species and infection assays were performed.

Materials and Methods

2.1 Identifying species-specific small secreted proteins

To identify small secreted proteins specific to individual species from the *Microbotryum* genus, a comparative secretomics analysis was performed on 8 species for which genomic annotation was available, provided by Dr. Ricardo de la Vega and Dr. Tatiana Giraud. To identify small secreted proteins in each of the 8 species, a pipeline following criteria from Beckerson et al., 2019 (see Chapter 2) was used to identify proteins with hallmarks of secretion as evidenced by in their protein leader sequence and to rule out potential cellular localization of proteins in organelles or the cellular membrane (Beckerson et al., 2019). Proteins that met these criteria were further screened against using 4 more secretomics tools to corroborate predicted secretion (Beckerson et al., 2019). To be considered for this study, proteins with predicted secretion had to have passed the initial selection with SignalP4.0, passed all criteria for non-cellular localization (TMHMM, Phobius, Prosite, PredGPI, NucPred), and pass 3/4 of the final 4 cutoff values for predicted secretion (Phobius, TargetP, SingalP3.0) (Figure C4-1). The resulting lists for each species were then blasted against the proteomes of the other 7 species to rule out any orthologous genes and identify species-specific genes. As demonstrated by Beckerson et al., 2019, some secreted proteins in one species may have non-secreted orthologs in another. Therefore, to identify truly unique species-specific secreted proteins, and to rule out those mobilized by selective pressures, only proteins with no orthologous match were used in this study. Furthermore, only amino acid sequences beginning with methionine were considered true candidate proteins for this study. For the sister species under analysis, *M. lychnidis-dioica* (MvSl) and *M. silenes-dioicae* (MvSd), the resulting list of unique secreted proteins was further screened for proteins smaller than 250 amino acids, resulting in a workable list of 3 unique proteins for MvSl and 5 for MvSd (Figure C4-1).

Computational Framework for Determining Species-Specific Small Secreted Proteins

Initial Selection
SignalP4.0 D-score >0.45 no TM, 0.5 TM

No Intracellular Localization Criteria:

Tools	Cutoff for SP
TMHMM	PredHel = 0
Phobius	TMs < 2
Prosite	No ER retention signal
PredGPI	Specificity <99.5
NucPred	Threshold <0.8

meets all criteria

Predicted Secretion Criteria:

Tools	Cutoff for SP
Phobius	"Y" for SP
TargetP	Predicted SP
SignalP3.0 NN	D-score >0.43
SignalP3.0 HMM	Sprob >0.8

meets at least 3/4 criteria

MvSn MvSl MvSd MvSp MvSv MvSc MvKa MvSpr

Non-Orthologous SPs

MvSl MvSd

<250aa

MvSl MvSd

Species-Specific SSPs

Gene	Length (aa)
MvSl_01693	172
MvSl_00814	51
MvSl_11239	42
MvSd_09295	59
MvSd_09648	72
MvSd_10910	85
MvSd_12336	88
MvSd_13216	123

Figure C4-1 Computational framework for identification of species-specific SSPs. The secretomes for *M. lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd) were determined in Beckerson et al., 2019 (see Chapter 2) using the 10 tools and cutoff values listed on the left of the diagram. SignalP4.0 was used to identify a list of secreted proteins (SPs) from the entire genome. These putative SPs were then analyzed for intercellular localization using TMHMM, Phobius, Prosite, PredGPI, and NucPred. To increase certainty that putative SPs are indeed secreted, the remaining list was run through Phobius TargetP1.0, and SignalP3.0, and must have passed at least 3 of the 4 cutoff values to be considered a candidate for this study. The resulting SPs for each sister species were then compared against each other, along with 6 additional distantly related species of *Microbotryum* (*M. violaceum sensu stricto* on *Silene nutans* [MvSn], *M. sensu lato* on *S. paradoxa* [MvSp], *M. lagarehemi* on *S. vulgaris* [MvSv], *M. violaceum sensu lato* on *S. carolineana* [MvSc], *M. lscabiosae* on *S. vulgaris* [MvKa], *M. intermedium* on *S. pratensis* [MvSpr]), to identify SPs without orthologs in other species, i.e., species-specific SPs. Any SPs with more than 250 amino acids were then removed to generate the final list of species-specific small secreted proteins.

2.2 EffectorP and gene expression

The list of species-specific small secreted proteins, 3 for *M. lychnidis-dioicae* and 5 for *M. silenes-dioicae*, were further analyzed using the EffectorP 1.0 program. EffectorP is a machine learning tool used to predict whether a secreted protein is an effector based on a pool of data for known plant effectors. The tool looks for patterns found among the sample set and compares these to the submitted amino acid sequence

(Sperschneider et al., 2015). The program then outputs a predictive number between 0 and 1 to represent a percent certainty that the secreted protein is an effector, as predicted by the program. A value over 50% is considered likely an effector. From the results for all 8 amino acid sequences tested, two candidates were selected, one from each species, with the highest likelihood of acting as an effector based on predictive measures (Table C4-1).

Table C4-1. EffectorP 1.0 Results for Species-Specific SPs

Protein Sequence	EffectorP 1.0 Likelihood	Amino Acid Sequence
MvSl_01693	Effector (0.585)	MQRAICHLPSPPHWVSAPRHRNTTGDHVTQSSHRSGELLQNNTDRYFF GCFDSRNVWIATRGVLCNALGCDTIDDADYTTLKRLLFCQKGITLKAK LSDQEGAGRTIKIFTGMGLEMISSGRGRMQKHGTRMESVERKRVVLEE GMIHRSFINSTKLQKRWRDREGQRHTLLEKRRNASKVEQEASAAERYP KRVMYGHFGAPEDEEKKKQGTVGGAIANGNANAD
MvSd_09295	Effector (0.935)	MVCATWFVFVSGRFTLTLVKKKCCGPPGRLWWMSQHTVVSAAPACG VHVPIRSHVQALW

From the initial screening, the *M. lychnidis-dioicae* gene MvSl_01693 with an effector likelihood of 0.585 was selected for heterologous expression in *M. silenes-dioicae*, and the *M. silenes-dioicae* gene MvSd_09295 with an effector likelihood of 0.935 was selected for heterologous expression in *M. lychnidis-dioicae*. While expression data is not available for the expression of *M. silenes-dioicae* genes, MvSl_01693 was aligned with its counterpart in the expression data available for *M. lychnidis-dioicae* at the BROAD institute (Perlin et al., 2015; Toh et al., 2017) to check for further evidence that the selected secreted protein is upregulated during infection and therefore a likely effector (Table C4-2).

Table C4-2: Expression of MvSl_01693 on various media and *in planta*, presented as TPM[a]

Water	Rich Media	Mating on Nutrient-limited Medium	Male Infected Tissue: Stage 8	Male Infected Tissue: Stage 9	Male Infected Tissue: Stage 10	Male Infected Tissue: Late	Male Infected Stem Tissue
0.04	0	0.02	0	7.5	0	489.16	397.24

a – transcripts per million; Male-infected Stem tissue, male-infected floral stem tissue. 2-3 independent determinations via RSEM.

Despite its marginal qualification in the predictive effector program EffectorP 1.0, significant upregulation in late infection in both floral and floral stem tissue along with no expression in rich media, water agar, or during mating, indicate that the MvSl_01693 secreted protein may play a role in infecting its *Silene latifolia* host, making it an ideal candidate for expression in *M. silenes-dioicae*.

2.3 *Agrobacterium*-mediated transformation of MvSd_09295 and MvSl_01693

The two genes encoding the species-specific small secreted proteins, MvSl_01693 and MvSd_09295, were selected for heterologous expression in the sister species. To express each gene in the other species, the reliable *Agrobacterium*-mediated transformation protocol by Toh et al., 2016 was used. Because there is no expression data available for *M. silenes-dioicae* during any stage of its lifecycle, and because the sister species are so closely related, the endogenous promoters for each gene, which we defined as the 1 kb sequence preceding the methionine codon, and the transcriptional termination sequence, which we defined as the 1kb sequence following the stop codon, were used to drive heterologous expression of the species-specific genes in effort to preserve proper time-specific transcription and to rule out any phenotypic differences that may be attributed to over expression of the gene if a constitutive promoter was used instead. Thus, the coding region along with 1 kb upstream and downstream elements were amplified from host genomic DNA and cloned into the pMvHyg (Toh et al., 2016) vector using the NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit for transformation into *Agrobacterium* (Figure C4-2).
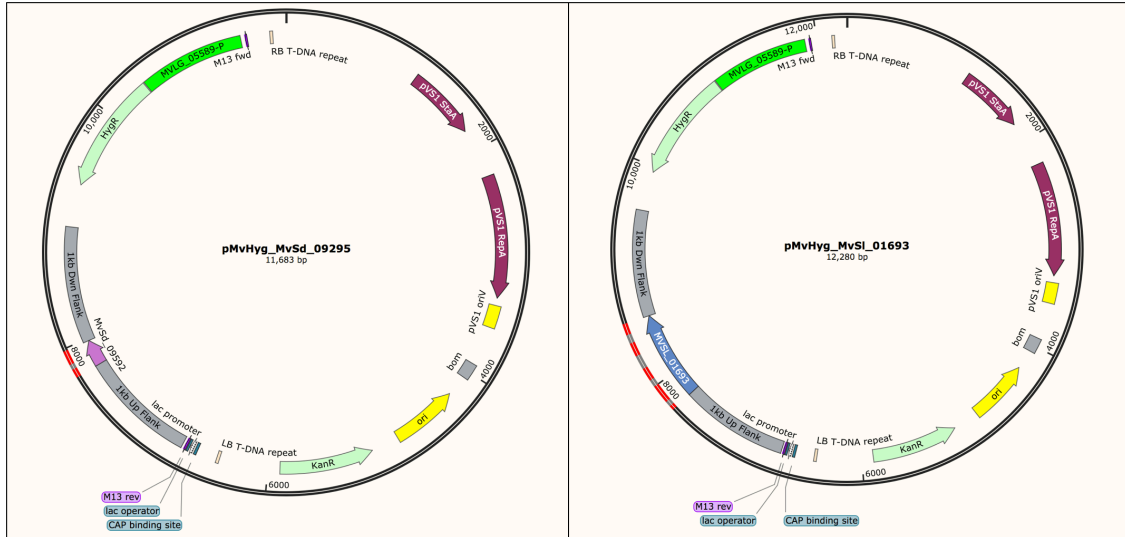
Figure C4-2 Snap gene image of the vectors for heterologous expression of species-specific genes. pMvHyg_MvSd_09295, left, was generated via cloning of the MvSd_09295 gene into pMvHyg from Toh et al., 2016 for expression in MvSl. Likewise, pMvHyg_MvSl_01693, right, was generated via cloning of the MvSl_01693 gene into pMvHyg for expression in MvSd.

Each respective plasmid was transformed into EHA105 *Agrobacterium* cells using electroporation (2.5 kV, 400 ohms and 25 μF) following the protocol outlined in Toh et al., (2016). Putative transformants were selected on LB agar-containing 50 μg kanamycin/mL and then re-streaked onto LB agar containing 50 μg kanamycin/ and 100 μg spectinomycin per mL agar to ensure that both the pMvHyg containing the respective species-specific gene and the helper plasmids were in the cells. Surviving EHA105 cells were further confirmed through colony PCR, before being used to transform the *M. lychnidis-dioicae* cells with pMvHyg_MvSd_09295 and the *M. silenes-dioicae* cells with pMvHyg_MvSl_01693. To transform their respective *Microbotryum* species, 1E7 cells/mL of each mating type for both *Microbotryum* species and each transformed *Agrobacterium* strain were measured using spectrophotometry (*Microbotryum*: OD_{600} 1 = 3.4E7 cells/mL; *Agrobacterium*: OD_{600} 1 = 8E8 cells/mL) and mixed in equal volumes. 200 μL of each suspension were spotted onto IM plates containing acetosyringone, and residual mixture was taken for confocal imaging to confirm adhesion of *Agrobacterium* to *Microbotryum* cells (Figure C4-3).
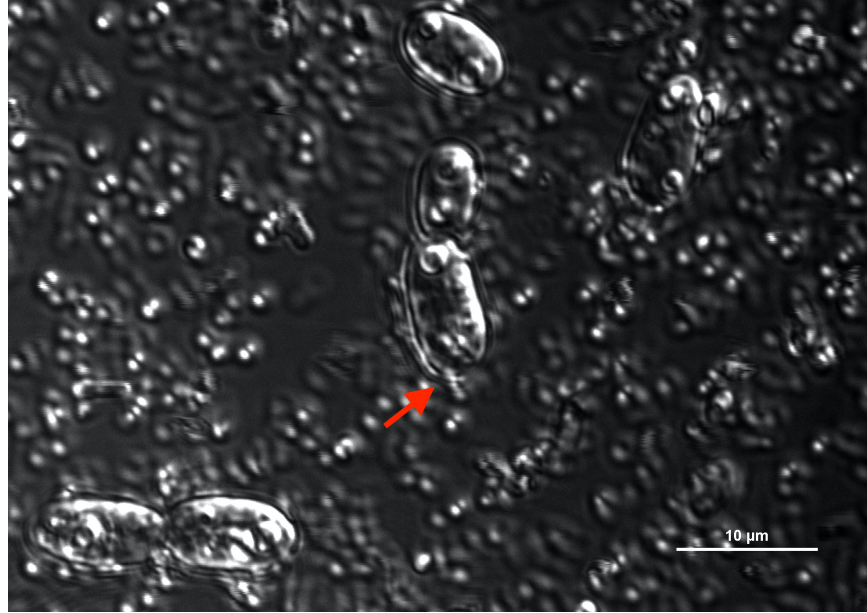
Figure C4-3 Confocal image of *Agrobacterium*-mediated transformation of *Microbotryum* cells.
Microbotryum cells appear as large oval structures approximately 10 microns in length. Attached agrobacterium cells are about 1 micron and labeled by red arrow.

Spotted plates were incubated at room temperature, ~25 °C, for 3 days, after which the resulting mass of cells was scraped from the plates, suspended in 600 μL of YPD-10% broth, and 200 μL of each suspension were spread onto YPD-10% containing 150 μg/mL Hygromycin and 100 μg/mL Cefotaxime plates. Each plate was then incubated for 12-15 days to select for transformed *Microbotryum* cells. Colonies were streaked onto fresh YPD-10% containing 150 μg/mL Hygromycin B plates for immediate storage. Putative transformants verified for successful species-specific gene insertions via PCR amplification, including the *hyg* gene and unique targets of each construct; DNA sequencing of amplified regions ensured the appropriate insertions/locations. Such verified transformants were then used for cross-infection studies in the respective host plant species.

2.4 Cross-infection analysis for *M. lychnidis-dioicae* expressing MvSd_09295

Cross-infection studies for the MvSd_09295 gene were performed in Bochum, DE using the growth facilities and electron microscope available at the Ruhr-Universität Bochum. a1 and a2 mating types for *M. lychnidis-dioicae* (p1A1 and p1A2 strains; Perlin et al., 2015; Toh et al., 2017), *M. silenes-dioicae*, and the *M. lychnidis-*

dioicae expressing the MvSd_09295 transgene were grown of YPD-10% plates, with added 150 μg/mL Hygromycin B for the transgenic strain, for 48h before suspension in water at an OD_{600nm} of 1. The corresponding mating types were then mixed at equal volumes and 10 μL of the cell suspensions were spotted onto *S. latifolia* and *S. dioica* seeds, totaling in 6 groups (Figure C4-4). For each group, 100 seeds for each group (300 *S. latifolia* and 300 *S. dioica* seeds in total) were evenly spaced on 2 large Minimal Salt (MS) 1.8% agar plates, 50 seeds on each plate. Each seed was spotted with 10 μL of cell suspension containing both mating types of the *Microbotryum* for that particular *Microbotyrum/Silene* grouping. The plates of spotted seeds were then sealed using parafilm and placed in the dark for 48h at 4°C for vernalization and to allow for fungal mating to occur. The plates were then moved to a growth chamber with the following program: On – 4am, 16h day length, 22 °C day, 18 °C night, 136 mmol/m^2/sec, fluorescent bulbs, 50-60% humidity.

Seeds were spotted in the chamber with sterile water every other day for two weeks. After 11 days, at least one sprouted seedling was taken from each of the 6 groups and observed via electron microscopy. Seedlings (defined in this study as sprouting seeds with cotyledon leaves) and sprouts (defined as sprouting seeds with only the emerging root tip) from each plate were transplanted into pots filled with dampened soil based on their overall growth progression, 11 days for *Silene latifolia* and 14 days for *Silene dioica* (Figure C4-4-A). Individuals from each group were transplanted, selecting first all of the seedlings followed by the sprouts to bring the total number for each group to 80. The number of seedlings and sprouts for each group is listed in Figure C4-4-A. After transplanting, *Hypoaspis miles* eggs, purchased from Sautter & Stepper, were sprinkled onto the top of the pots to act as a biological pest control. The pots were then arranged in randomized blocks, each block containing one pot for each of the 6 groups (Figure C4-4-B). Four blocks each were then randomly placed into 20 total watering trays (Figure C4-4-C) and moved to the greenhouse (Figure C4-4-D).
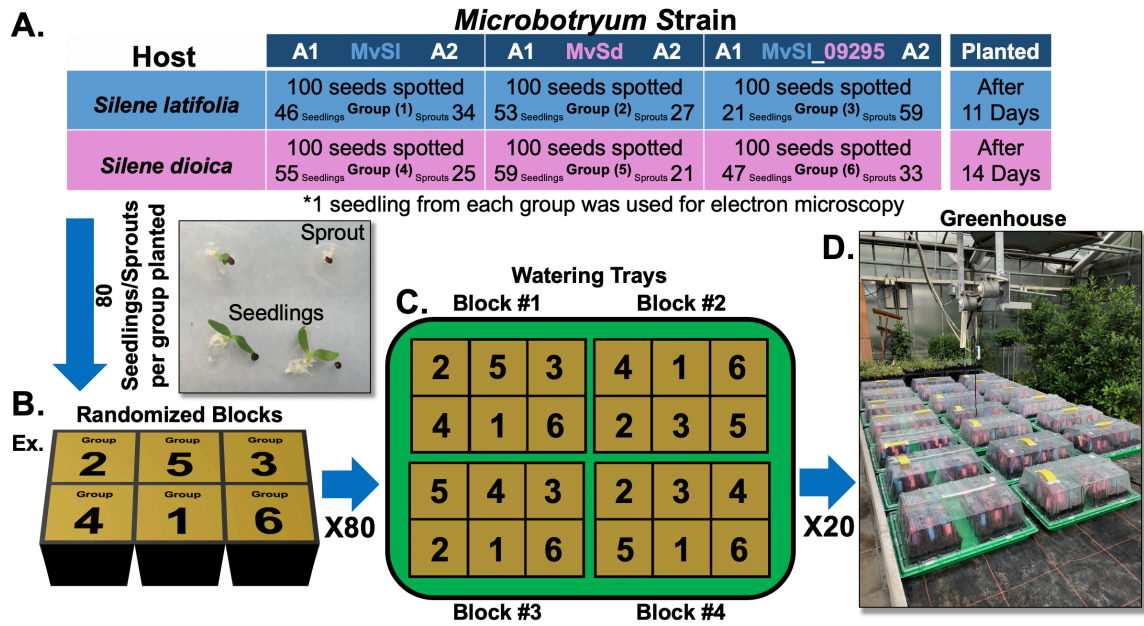
Figure C4-4 Outline of the planting procedures used in the Ruhr-Universität Bochum study. (A) demonstrates the total number of seeds spotted with a1 and a2 mating types of each *Microbotyrum* strain, as well as the number of seedlings and sprouts that were transplanted into soil after an incubation period (planted column). Each cell also contains the metadata number for each *Microbotryum*/*Silene* grouping. Seeded pots were then randomly organized in blocks of 6 (B), with one representative from each group in each block. Groups of four blocks were then randomly placed into 20 watering trays (C) and moved to the greenhouse (D).

The arrangement of each pot is illustrated in Figure C4-5. A lid was added to each tray to trap moisture from evaporation, and an insect strip was placed on the top of each tray near the opening to help trap insects (Figure C4-2-D). The lids for each tray were removed after 1 month of growth as the plant became tall enough to reach the lid. The experiment was run for a total of 1 year, with *Silene latifolia* hosts flowering after 2-3 months and the *Silene dioicae* hosts requiring overwintering before flowering between 10-12 months. Windows in the green house remained open for the duration of the experiment to allow for ambient temperatures, which were recorded and are listed in Figure C4-5.

Greenhouse Arrangement
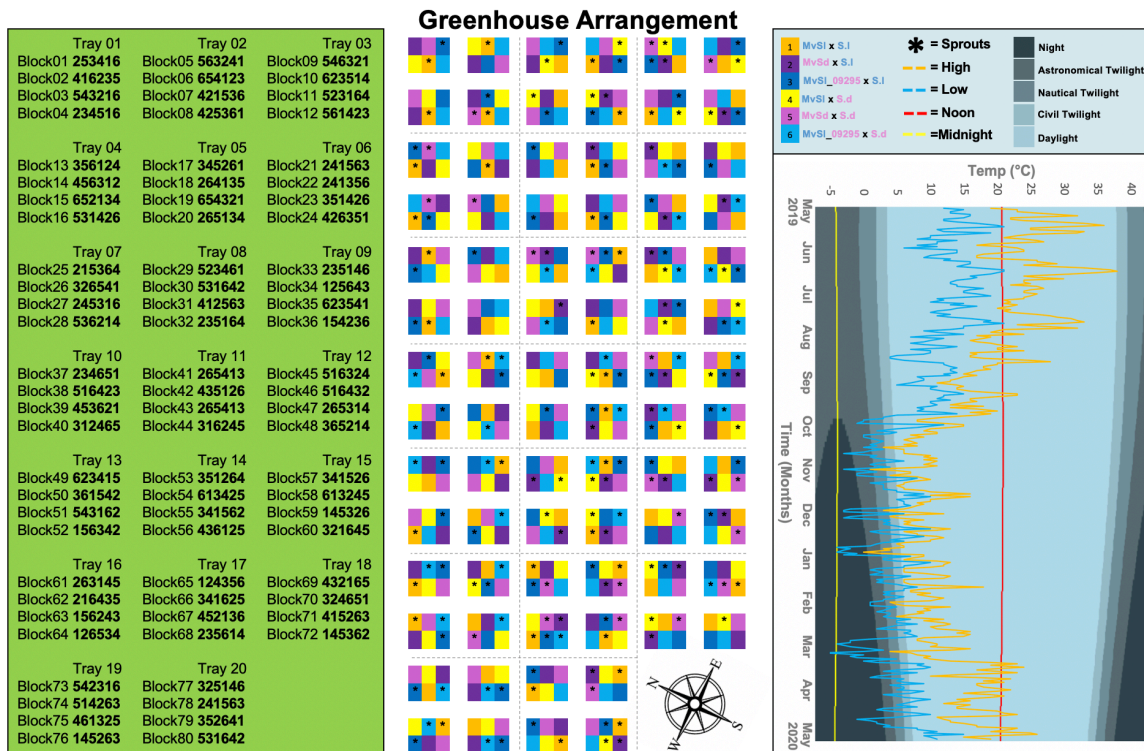
Figure C4-5 Greenhouse arrangement of plants in the Bochum study. demonstrates the arrangement for the randomized blocks and the conditions in the greenhouse. The order of groups for each block, and the order of blocks in each tray, was randomized using the List Randomizer tool at Random.org, results of which are shown in the green left column. Trays were organized in lines of three to fit the bench top of the greenhouse laying from NE to SW as shown in the white middle column. Each group is color coded by the plant and strain used, 1 (orange) represents MvSl on *S. latifolia*, 2 (yellow) represents MvSd on *S. latifolia*, 3 (light green) represents the transgenic MvSl expressing the MvSd_09295 gene on *S. latifolia*, 4 (dark green) represents MvSl on *S. dioica*, 5 (light blue) represents MvSd on *S. dioica*, and 6 (dark blue) represents the transgenic MvSl expressing the MvSd_09295 gene on *S. dioica*. Solid color squares represent pots that contained transplanted seedlings, while squares including a asterisk (*) symbol represent pots that contained transplanted sprouts. The daylight cycle for the greenhouse, divided into periods of international Night, Astronomical Twilight, Nautical Twilight, Civil Twilight, and Daylight, are displayed in the right column. Noon and midnight are represented by the red line and yellow line, respectively. The high (orange) and low (blue) temperature for each day throughout the year of the trial are superimposed over the daylight cycle.

2.5 Electron microscopy for MvSd_09295

SEM was performed on 1 seedling for each group with the exception of the MvSl control strain on *S. latifolia* grouping, which used 3 seedlings due to an inability to find dikaryotic filaments. Samples were manually opened, mounted on brass blocks with a mixture of Tissue Tek O.C.M. compound and colloidal graphite, and directly fixed using slushy nitrogen freezing and a cryo-transfer system (Quorum PT3000T) to prevent artifacts of chemical fixation and freeze drying. Once in the cryo-transfer system, water was sublimed at -105°C for 20 min, the specimens were sputtered with platinum at 30 mA for 90 sec and analyzed at -140°C at high vacuum with the SE2 detector using a Sigma VP at 8kV and an aperture of 30 μm.

2.6 Cross-infection analysis for MvSl_01693 in *M. silenes-dioicae*

Cross-infection studies for the MvSl_01693 gene were performed in Kentucky, US using the University of Louisville growth facilities. a1 and a2 mating types for *M. silenes-dioicae* and the *M. silenes-dioicae* expressing the MvSl_01693 transgene were grown on YPD-10% plates, with added 150 μg/mL Hygromycin B for the transgenic strain, for 48h before suspension in water at an $OD_{600nm} = 1.0$. The corresponding mating types were then mixed at equal volumes and seeds were spotted and cooled for vernalization and fungal mating in similar fashion to the study performed in Bochum, DE. The plates were then moved to a growth chamber with the following program to simulate the temperature and day cycle conditions in Bochum, DE during the beginning of flowering season for *Microbotryum* (May) with a reduction of 1 hour to conform with pervious growth experiments demonstrated to generate successfully infected hosts (Toh & Perlin, 2015): Dawn from 8:00-9:00 16°C 8x incandescent bulbs, Day from 9:00-21:00 20°C 8x incandescent bulbs with light intensity 136 μmol/m^2/sec at plant surface, Dusk from 21:00-22:00 20°C 8x incandescent bulbs, Night from 22:00-9:00 16°C no blubs on (Figure C4-6).

A.

Host	*Microbotryum* Strain						Planted
	A1	MvSd	A2	A1	MvSd_01693	A2	
Silene latifolia	100 seeds spotted 48 Seedlings			100 seeds spotted 48 Seedlings			After 11 Days
Silene dioica	100 seeds spotted 26 Seedlings			100 seeds spotted 48 Seedlings			After 14 Days

B.

Growth Chamber Settings

	°C	Flour.	Incan.
0:00	16	0	0
8:00	16	0	8
9:00	20	8	0
21:00	20	0	8
22:00	16	0	0
23:59	16	0	0

136 µmol/m²/sec

Night:
12:00 am - 3:53 am
11:43 pm - 12:00 am
Total: 04:10

Astronomical Twilight:
3:53 am - 4:52 am
10:43 pm - 11:43 pm
Total: 01:59

Nautical Twilight:
4:52 am - 5:39 am
9:55 pm - 10:43 pm
Total: 01:36

Civil Twilight:
5:39 am - 6:16 am
9:18 pm - 9:55 pm
Total: 01:14

Daylight:
6:16 am - 9:18 pm
Total: 15:01

Solar Noon/Midnight:
1:47 pm
1:47 am

May 9

Figure C4-6 Growth chamber conditions for the Louisville study. A) Breakdown of the four groups tested in the Louisville experiment by number of seeds spotted, number of seedlings transplanted, and after number of days of growth on water agar plates. B) Light cycle graph for Bochum, DE in May from Timanddate.com, left, and growth chamber conditions used for the Louisville experiment, right.

Because very few infected sprouts used in the Ruhr-Universität Bochum study survived to adult plants for analysis, only seedlings were planted for the Louisville study. *S. latifolia* seedlings were again transplanted after 11 days and *S. dioica* were transplanted after 14 days in similar fashion with the Bochum study. Without access to electron microscopy, dikaryotic filaments could not be observed; however, mating of the fungal strains were verified via light microscopy. Groups were again randomized into blocks; however, the number of blocks was limited to 40 for the Louisville study due to growth chamber size restrictions. Seedlings were transplanted into eggshell crates filled with dampened Star-Green Potting Mix Plus Fertilizer soil containing 0.1% Nitrogen, 0.08% Phosphate, and 0.06% Soluble Potash. Eggshell crates were fitted into trays to water the seedlings from the bottom up. After the seedlings reached the 8-leaf stage, they were transplanted into their own small pot with the same dampened Star-Green Potting Mix and placed in a tray to water from the bottom up.

Results

3.1 Microscopy results

With the access to electron microscopy in the Bochum study, we were able to verify the formation of a dikaryotic filament and penetration of the plant cell in all but the *MvSl on S. latifolia* control group (Figure C4-7). For each successful attempt, only one seedling was needed, indicating that successful infection was highly probable amongst the different groups. To rule out the chance that a lack of infection seen in the MvSl x *S. latifolia* pairing may have been due to random selection of non-infected seedlings, three seedlings were observed, all of which lacked the presence of dikaryotic filaments (Figure C4-7).
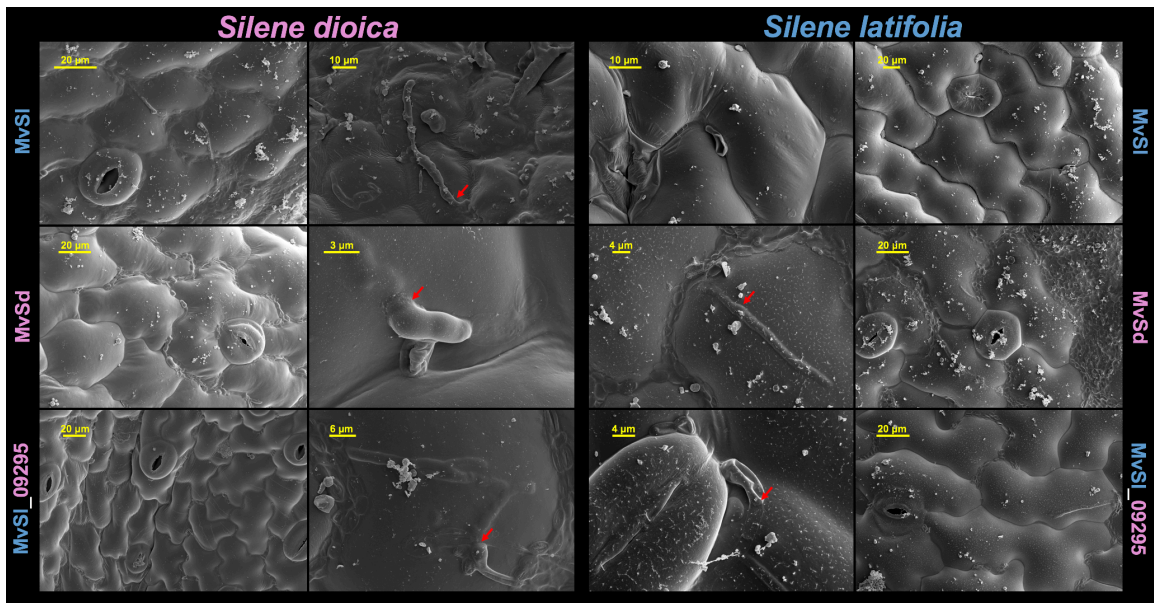
Figure C4-7 Electron micrographs of seedlings infected with *Microbotryum*. The images were taken of the cotyledon leaves of one seedling for each of the 6 groups, with the exception of MvSl on *Silene latifolia* for which three seedlings were observed. An overview is shown on the left of each paring, while a closeup of appressorium are shown on the right. The left column are images of infected *Silene dioica* seedlings, while the right column shows infected *Silene latifolia* seedlings. Each row indicates the strain/species of *Microbotryum* that was used for infection, MvSl (top), MvSd (middle), and transgenic MvSl expressing the MvSd_09295 gene (bottom).

To further rule out a lack of mating as the cause for a lack of dikaryotic filament formation in the MvSl x *S. latifolia* control group, samples were taken from the agar surrounding the spotted seedlings and mating was verified (Figure C4-8). These images demonstrate the MvSl strains were able to mate, forming conjugation tubes; however, they did not infect the seedlings on which they were spotted.
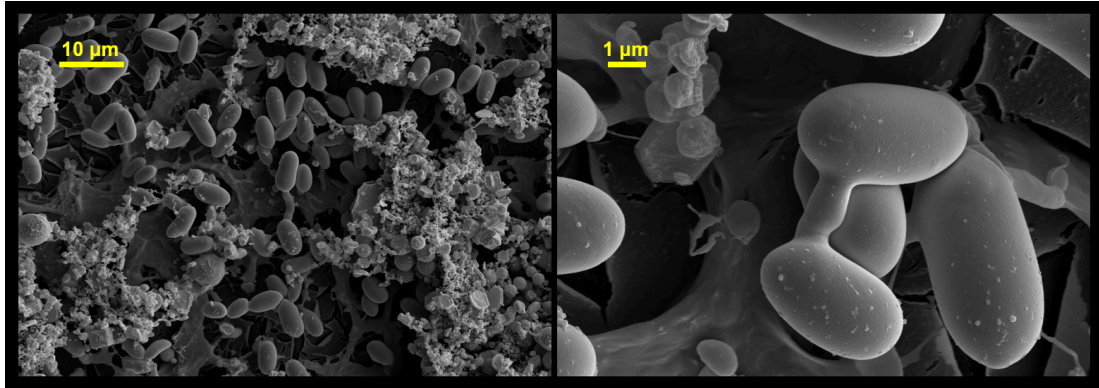
Figure C4-8 Electron micrographs of mating MvSl a1 and a2 cells. Cells were taken from MS agar around the locations where seedlings were removed.

While we were unable to observe dikaryotic filament formation on seedlings without electron microscopy in the Louisville study, mating was verified using light microscopy from cell suspensions located around spotted seeds (Figure C4-9).
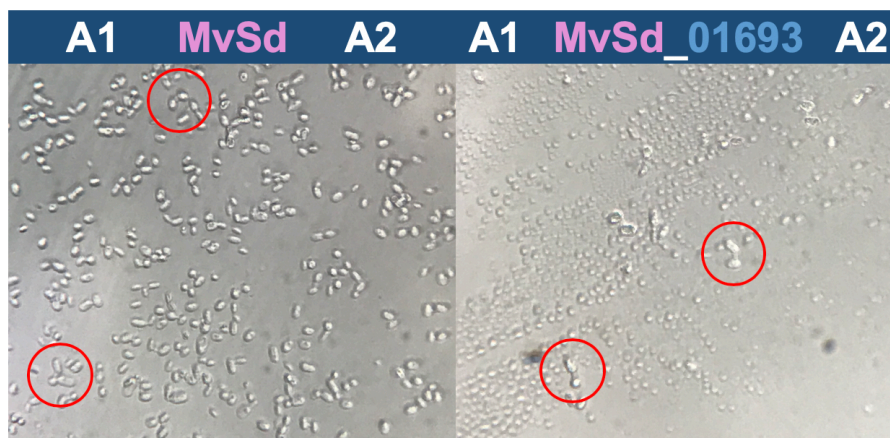
Figure C4-9 Confirmation of *Microbotryum* mating via light microscopy. Mating was observed between a1 and a2 mating types of the *M. silenes-dioicae* control (MvSd), left, and a1 and a2 mating types of the *M. silenes-dioicae* strain expressing the *M. lychnidis-dioicae* transgene MvSl_01693 (MvSd_01693), right.

Observation of conjugation tube formation in Figure C4-9 demonstrates that the proper mating types for each strain were mixed, and while they do not demonstrate the formation of the infectious dikaryotic filament the eventual observation of infected adult plants for each group corroborate successful mating an infection.

3.2 Infection results for the MvSd_09295 Ruhr-Universität Bochum study

From the six groups in this study, MvSl x *S. latifolia,* MvSd x *S. latifolia*, MvSl_09295 x *S. latifolia,* MvSl x *S. dioica*, MvSd x *S. dioica*, and MvSl_09295 x *S. dioica*, all three groups of *S. latifolia* plants flowered between 2-3 months. All 3 *S. dioica* groups required overwintering in the greenhouse and flowered between 13-14 months. Of the individuals planted as sprouts, only 10.55% survived into adult plants. This was compared to an overall survival rate of 74.73% for individuals transplanted as seedlings. For the groups infecting *S. latiofolia* hosts, 100% of the 93 adults survived to flowering; however, no infected flowers were observed (Table C4-3). Similarly, 100% of the 139 flowers that survived to adult plants flowered in the groups infecting *S. dioica* hosts, but only one infected host was observed in each group (Table C4-3).

Table C4-3. Tally of seedlings surviving, flowering, and infected flowers for the Bochum study

Spotted	Transplanted	Adult Plants	Flowered	Infected		
100 Seeds	46 Seedlings	35 (76.09%)	43 (100%)	0 (0%)	1	MvSl x S.l
	34 Sprouts	8 (25.23%)			2	MvSd x S.l
100 Seeds	53 Seedlings	26 (49.06%)	30 (100%)	0 (0%)	3	MvSl_09295 x S.l
	27 Sprouts	4 (14.81%)			4	MvSl x S.d
100 Seeds	21 Seedlings	15 (71.43%)	21 (100%)	0 (0%)	5	MvSd x S.d
	59 Sprouts	6 (10.17%)			6	MvSl_09295 x S.d
100 Seeds	55 Seedlings	43 (78.18%)	45 (100%)	1 (2.22%)		
	25 Sprouts	2 (8.00%)				
100 Seeds	59 Seedlings	55 (93.22%)	55 (100%)	1 (1.81%)		
	21 Sprouts	0 (0%)				
100 Seeds	47 Seedlings	36 (76.60%)	39 (100%)	1 (2.56%)		
	33 Sprouts	3 (9.09%)				

3.3 Infection results for the MvSl_01693 University of Louisville study

From the four groups in this study, MvSd_01693 x *S. dioica*, MvSd x *S. dioica*, MvSd_01693 x *S. latifolia*, and MvSd x *S. latifolia*, all transplanted seedlings survived to adult plants; however, only *S. dioica* hosts induced flowering (Table C4-4). Unlike in the MvSl_09295 study, the first flowers observed were from the MvSd_01693 x *S. dioica* group, only 56 days post transplantation of seedlings into soil. This was compared to 67 days for the first flowers of the control group, MvSd x *S. dioica* (Table C4-4). In total, nearly half (46.51%) of the flowers from the MvSd_01693 x *S. dioica* group were infected compared to only 16.66% in the control MvSd x *S. dioica* group (Table C4-4).

Table C4-4. Tally of seedlings surviving, bolting, flowering, and infected flowers for the Louisville study

Spotted	Transplanted	Adult Plants	Flowered	Infected		
					1	MvSd_01693 x S.l
100 Seeds	48 Seedlings	48 (100%)	0 (0%)	N/A	2	MvSd x S.l
100 Seeds	26 Seedlings	25 (100%)	0 (0%)	N/A	3	MvSd_01693 x S.d
100 Seeds	58 Seedlings	48 (100%)	43 (89.58%)	20 (46.51%)	4	MvSd x S.d
100 Seeds	48 Seedlings	48 (100%)	18 (72.00%)	3 (16.66%)		

While both male and female flowers were observed infected in the MvSd_01693 x *S. dioica* group at a nearly 1:1 ratio (11 females and 9 males), all three infected flowers observed in the control MvSd x *S. dioica* group were male flowers. The infected female flowers observed in the MvSd_01693 x *S. dioica* group had induced pseudo-anther production in the floral ovary at the base of the flower (Figure C4-10-B).
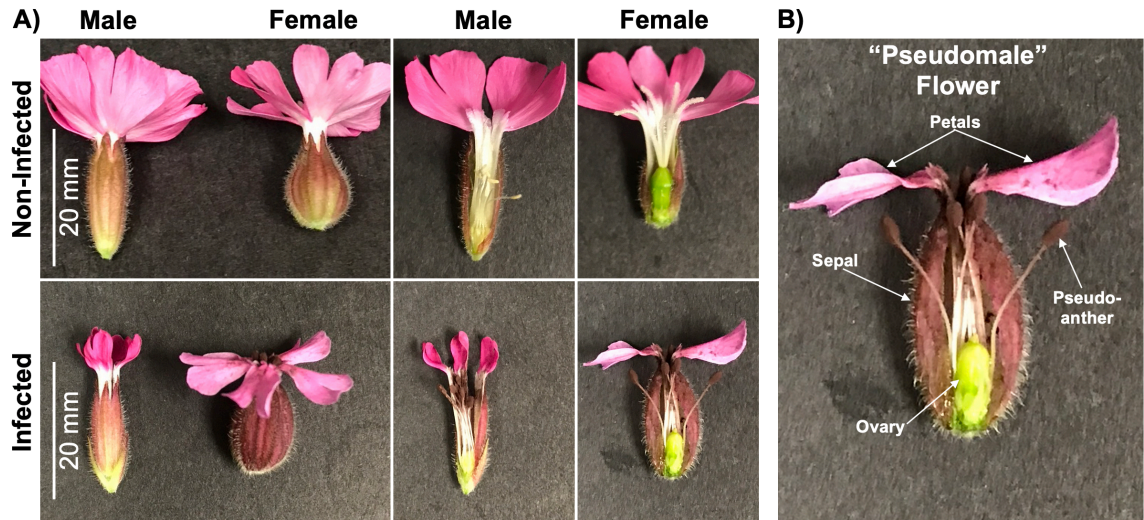
Figure C4-10 Infected and non-infected *Silene dioicae* flowers. Non-infected male and female flowers are shown in the top row of (A), while infected male flowers and infected female flowers with induced pseudo-anthers are shown in the bottom row of (A). B) demonstrates a closeup image, labeled with the morphological changes induced by infection of a female *S. dioica* flower, referred to as a "pseudomale" flower.

Discussion

This study examines the effect of species-specific genes on host-specificity in the *Microbotryum* genus through heterologous expression of genes unique to each of the sister species, *M. lychnidis-dioicae* and *M. silenes-dioicae*. The results for the two genes analyzed, MvSd_09295 and MvSl_01693, were mixed, with MvSd_09295 providing no difference towards infection on either host when expressed in *M. lychnidis-dioicae* and MvSl_01693 demonstrating a significantly increased capacity for infection of *S. dioica* when expressed in *M. silenes-dioicae*. While these results indicate that some species-specific genes may play a role in modulation of the host defense response, differences between the experimental setups may have contributed to observed differences in results, and therefore these studies must be repeated for further support of our hypotheses.

One major difference observed between the Bochum study and the Louisville study was the growth rate of *S. dioica* hosts. In the Bochum study, which used ambient greenhouse condition in Bochum, DE, *S. dioica* plants took over a year to flower (14 months). This dramatically differed from the *S. dioica* plants grown in the growth chambers in the Louisville study, which were able to flower after just 56 days. One

contributing factor may have been the growth chamber conditions used in Louisville, which mirrored the daylength and light intensity found during the flowering season for *S. dioica*, typically around May. Given that day length is a known major factor that contributes to signaling of flower production in plants, suspending plant growth under light conditions consistent with the flowering period, as opposed to our study in Bochum for which plants did not reach their adult stage until July, may explain the differences observed in flower timing. These results suggest that further studies performed in the greenhouse environment may be able to save time if started earlier in the year by aiming to have a majority of plants reach their adult stage in April. Alternatively, this paper highlights optimal conditions for growth of *S. dioica* plants in a chamber environment, which may prove to be a much more effective approach.

In addition to differences observed between studies, there is evidence to support that the control group MvSl x *S. latifolia* in the Bochum study may have been uninfected due to lab errors as well. This is shown by the lack of observable dikaryotic filament structure formation in the control group on all three seedlings observed, but clear dikaryotic filament formation in the two test groups for the *S. latifolia* host. One possible explanation for the lack of formation of dikaryotic infective structures in mated a1 and a2 strains of the MvSl control observed in the Bochum study may be due mislabeling of strains in the lab. While the MvSl strains transformed with the MvSd_09295 gene were generated in the lab at the University of Louisville, the control strains were taken from storage at Ruhr-Universität Bochum. Pervious work has demonstrated that post-zygotic barriers exist in hybridized strains of *Microbotryum*, which may explain why dikaryotic filament formation was not observed between if the matting assay was mistakenly set up as a mating event between two different species of *Microbotryum*. This may also explain why mating was still observed in the areas around where the seedlings were extracted from the agar.

While some species-specific SPs appear to play a role in pathogenicity of in their progenitor species, it is evident from this study that not all species-specific genes result in unlocking pathogenesis of their corresponding hosts. In fact, while a significant increase in ability for *M. silenes-dioicae* to infect its own host, *S. dioicae*, was observed in the transgenic strains expressing the MvSl_01693 gene, a greater ability for the transgenic strain to infect the host associated with the MvSl_01693 SP, *S. latifolia,* was not observed. Furthermore, the MvSd_09295 species-specific SP did not demonstrate a change in pathogenicity in either direction on either host. Despite these findings, the results from the heterologous expression of MvSl_01693

in *M. silenes-dioicae* are of interest. Increased ability of *M. silenes-dioicae* to infect its natural hosts, *S. dioica*, is outside of our original hypothesis which stated that species-specific genes were important for infection of their own associated hosts (i.e., an *M. lychnidis-dioicae* species-specific SP should have conferred a greater ability to infect the natural *M. lychnidis-dioicae* host, *S. latifolia*), but could help explain the genesis of such species-specific genes. One possible explanation for the increased pathogenicity observed in the transgenic MvSd_01693 strain on its own host, *S. dioica*, may be that the MvSl_01693 gene evolved initially in *M. lychnidis-dioicae* in response to coevolutionary pressures arising from changes in *S. latifolia* after the speciation event separating the two sister species/host pairings. If the novel MvSl_01693 gene coded for a SP that conferred a significant initial advantage for the pathogen during host pathogenicity, this gene would have spread rapidly through the *M. lychnidis-dioicae* population. Over time however, resistance to the effects of the MvSl_01693 SP would have been selected for in the *S. latifolia* population as a consequence of rapid spreading of this new strain of *M. lychnidis-dioicae*, resulting in recognition factors or other defense responses to the MvSl_01693 SP. This would explain why when the MvSl_01693 gene is introduced to *M. silenes-dioicae*/*S. dioicae* pathogen/host group, an SP for which the fungi does not otherwise express naturally, the host is more susceptible having not developed an immune response over years of reciprocal selective pressure. Therefore, we may expect that species-specific genes have arisen as a direct result of the selective pressures issued by the coevolutionary struggle between pathogen and host over time post speciation events, and not through a gene loss mechanisms in the other species. Furthermore, the demonstrated efficacy of species-specific SPs in alternative hosts seen in this study could help facilitate host shifts events withing the *Microbotryum* genus.

Understanding how these species-specific genes arise and the role they play in host-specialization sheds light on how coevolutionary pressures can change pathogens over time. While we were able to observe increase pathogenicity in the transgenic MvSd strains when infecting their own hosts, it is likely that modifications to core secreted proteins play a much larger role in coevolution and subsequent speciation between the two now separate species than unique proteins as outlined by Beckerson et al., 2019 and Chapter 2. While certain species-specific SPs may play a role in modulation of defense responses in the host, these core secreted proteins likely play a more mechanical role for entrance into the host tissue and manipulation of the plant, leading to a more stringent selective pressure. To further understand the role of the species-

specific SPs, future studies should analyze the importance of these species-specific genes in more distantly related pairings of *Microbotryum*/hosts to determine their capacity for affecting pathogenicity. Alternatively, examination of the "effectorization" of orthologues in different species that appear to have lost protein secretion ability due to the accumulation of mutations during speciation events may also help to understand the coevolutionary pressures impacting the fungi.

CHAPTER 5

IMPLEMENTATION OF CRISPR-CAS9 AS AN EFFECTIVE TARGET-SPECIFIC
KNOCKOUT TOOL FOR THE *MICROBOTRYUM* GENUS

Introduction

The first inkling of CRISPR dates back to the late 1980's, when Yoshizumi Ishino and his lab at Osaka

University discovered unusually repetitive sequences in the genome of *Escherichia coli* (Ishino et al., 1987).

These DNA repeats would later be called Clustered Regularly Interspaced Palindromic Repeats, or CRISPR.

While Yoshizumi's discovery ultimately led to a revolution in genetic modification of organisms, its full

significance wasn't appreciated until investigated more fully by Francisco Mojica, who identified a similar

arrangement of repeats in a very distant microbe, the halophile, *Haloferax mediterranei*, at the University of

Alicante in Spain. Although there was no sequence similarity between the regions seen by Ishino in *E. coli*

and those Mojica found in *H. mediterranei*, the latter realized that the arrangement of such patterns was

unlikely to have occurred by coincidence (Mojica et al., 1993; Lander, 2016). Mojica continued to investigate

and characterize these repeats, and by 2000 had found similar arrangements in a variety of prokaryotes,

including several bacteria associated with human disease (Mojica et al., 2000). The census of microbes

containing these expanded rapidly, so that by 2002 there were at least 40 species identified as having CRISPR

repeats (Jansen et al., 2002). Mojica continued to investigate the role of both the repeats and the spacers that

separate them, finally publishing a proposed role for them in a prokaryotic adaptive immune system (Mojica

et al., 2005).

Following its discovery, two main groups recognized the importance of CRSIPR and associated Cas

enzymes in genetic engineering and biotechnology. Virginijus Siksnys investigated whether the system could

be reconstituted functionally in *E. coli*, again a distantly-related microbe of his source organism,

Streptococcus thermophiles. Meanwhile, Emmanuelle Charpentier and Jennifer Doudna had formed a

collaboration that led to the use of recombinant Cas9 enzyme and the single guide RNA (sgRNA) now used

extensively in CRISPR transformations. Both groups published their work within two months of each other

(Jinek et al.,2012; Gasiunas et al., 2012); however, it wasn't until 2013 that CRISPR Cas9 was successfully adapted for use in modifying the genomes of other prokaryotic and eukaryotic cells. This innovative approach of using a prokaryotic system for genetic modification in eukaryotes was implemented by Feng Zhang and his lab at the Broad Institute of MIT and Harvard, and opened a floodgate of seemingly limitless genetic engineering applications (Cong et al., 2013; Mali et al., 2013). Since then, CRISPR Cas9 has been applied to many model systems across all 5 kingdoms and new innovative discoveries are still pushing the boundaries its application (Hovath and Barrangou, 2010; Char et al., 2017; Ochiai, 2015; Lu et al., 2017) (Table C5-1).

Table C5-1. List of Model Organisms with Established CRISPR Systems

Organism			CRISPR Transformation Method					
Classification			Plasmid	Purified Protein	RNA	Genomic Insertion	Endogenous	Reference
		Species						
PROKARYOTES	Archaea	Halophiles	*Haloferax mediterranei*					X	Li et al., 2013
			Haloferax volcanii			X		X	Stachler and Marchfelder, 2016
		Methanogens	*Methanococcus maripaludis*					X	Richter et al., 2012
			Methanosarcina acetivorans	X					Nayak and Metcalf, 2017
			Methanosarcina barkeri					X	Maeder et al., 2006
			Methanosarcina mazei		X	X		X	Nickel et al., 2013
		Sulfolobus	*Sulfolobus acidocaldarius*					X	Manica and Schleper, 2013
			Sulfolobus islandicus	X					Li et al., 2016
			Sulfolobus solfataricus	X	X	X			Zebec et al., 2014
		Thermococcales	*Pyrococcus furiosus*					X	Majumdar et al., 2015; Richter et al., 2012
			Thermococcus kodakarensis				X	X	Elmore et al., 2013
			Thermococcus onnurineus					X	Jung et al., 2016
	Eubacteria		*Azotobacter vinelandii*					X	Robson et al., 2015
			Bacillus subtilis	X		X	X		Altenbuchner, 2016; Price et al., 2019; Westbrook et al., 2018
			Clostridium thermocellum					X	Richter et al., 2012
			Escherichia coli	X		X	X		Jian et al., 2016

		Pseudomonas aeruginosa	X			X	X	Chen et al., 2018; Tan et al., 2018
		Streptomyces coelicolor	X			X		Tao et al., 2018
		Synechocystis	X				X	Xiao et al., 2018; Scholz et al., 2013
		Vibrio fischeri	X			X		Zeaiter et al., 2018

EUKARYOTES	**Protists**	*Chlamydomonas reinhardtii*		X	X			Gruzmán-Zapata et al., 2019
		Dictyostelium discoideum	X					Sekine et al., 2018
		Thalassiosira pseudonana	X					Hopes et al., 2016
	Fungi	*Ashbya gossypii*	X					Jiménez et al., 2019
		Aspergillus nidulans	X					Nødvig et al., 2015
		Coprinus cinereus	X		X	X		Chen et al., 2018
		Cryptococcus neoformans		X	X	X		Fan and Lin, 2018
		Neurospora crassa				X		Matsu-ura et al., 2015
		Saccharomyces cerevisiae	X			X		Akhmetov et al., 2018
		Schizophyllum commune		X	X			Vonk et al, 2019
		Schizosaccharomyces pombe	X			X		Ozaki et al., 2017
		Ustilago maydis	X					Schuster et al., 2016
	Plants	*Arabidopsis thaliana*	X			X		Wu et al., 2018; Miki et al., 2018
		Brachypodium distachyon				X		Abrash et al., 2018
		Lemna gibba				X		Liu et al., 2019
		Lotus japonicus				X		Wang et al., 2016
		Marchantia polymorpha	X			X		Sugano et al., 2018
		Medicago truncatula				X		Meng et al., 2017
		Nicotiana benthamiana				X		Jansing et al., 2018
		Oryza sativa	X			X		Fiza et al., 2019
		Physcomitrella patens	X					Lopez-Obando et al., 2016
		Populus tomentosa Carr.				X		Fan et al., 2015
		Setaria viridis				X		Demirci et al., 2017
		Zea mays	X		X	X		Doll et al., 2019
	Animals / **Invertebrates**	*Branchiostoma floridae*	X			X		Sanders et al., 2018
		Caenorhabditis elegans	X	X	X	X		Dickenson and Goldstein, 2016

Species						Reference
Coina intestinalis	X					Stolfi et al., 2014
Daphnia pulex		X	X			Hiruta et al., 2018
Drosophila melanogaster	X	X	X	X		Port et al., 2019; Gratz et al., 2015; Bier et al., 2018
Galleria mellonella			X			Wei et al., 2014
Gryllus bimaculatus			X			Sun et al., 2017
Hydra		X	X			Lommel et al., 2017
Mnemiopsis leidyi		X	X			Presnell and Browne, 2019
Nematostella vectensis	X		X	X		Ikmi et al., 2014
Oikopleura dioica						
Oscarella carmela						
Parhyale hawaiensis			X			Fan et al., 2015
Pristionchus pacificus		X	X			Witte et al., 2015
Tribolium castaneum	X		X			Fan et al., 2015
Ambystoma mexicanum		X	X			Fei et al., 2018
Canis lupus familiaris	X					Eun et al., 2019
Carolina anole		X	X			Rasys et al., 2019
Cavia porcellus	X					Bierle et al., 2016
Danio rerio		X	X			Sorlien et al., 2018
Gallus gallus domesticus	X					Zuo et al., 2016
Gasterosteus aculeatus		X	X			Erickson et al., 2016
Homo sapiens	X	X	X			Richardson et al., 2018; Zhang et al., 2017; Yang et al., 2014
Mesocricetus auratus	X	X	X			Fan et al., 2014
Mus musculus		X	X			Hirose et al., 2017
Nothobranchius furzeri				X		Harel et al., 2016
Orzais latipes		X	X			Liu et al., 2018
Petomyzon marius			X			Square et al., 2015
Rattus norvegicus		X	X			Kobayashi et al., 2018
Rhesus macaque			X			Xin et al., 2016
Takifugu rubripes		X	X			Kato-Unoki et al., 2018
Xenopus tropicalis		X	X			Nakayama et al., 2014

(Left margin label, rotated: **Vertebrates***)*

CRISPR operates in nature as an adaptive immunity to viral and/or plasmid invasion of many prokaryotic species with approximately 90% of Archaea and 40% of Bacteria containing CRISPR elements in their genome (Horvath and Barrangou, 2010). Upon viral invasion, CRISPR-associated (Cas) enzymes obtain short fragments of the invading viral DNA or RNA and incorporate the genetic information into the CRISPR locus, adjacent to Cas scaffold RNA regions; when transcribed together these form the guide RNA (gRNA). These gRNAs then attach to Cas enzymes to chaperone the DNA-endonuclease complexes to the virus targets. When bound to Cas enzymes, the 5' end of the gRNA binds to approximately 20 base pairs (bp) of complementary sequence and directs a conformational change, allowing the enzyme to cut the invading viral DNA or RNA (Hsu et al., 2014). To avoid cutting its own DNA, the short complementary sequences that become integrated into the host genome for CRISPR use are spaced between repeat sequences that allow the Cas enzymes to differentiate between self and non-self sequences.

There are three types of CRISPR systems: CRISPR Types I, II, and III, that differ from one another with regards to how the CRISPR array is processed and by the number of proteins that form the Cas DNA-endonucleases complex responsible for cleaving the target sequence. Each type of CRISPR system has a Cas DNA-endonuclease with a corresponding unique sequence required directly downstream of the 20 base pair target site called a Protospacer Adjacent Motif (PAM). These PAM sites vary in sequence and length by CRISPR type and by species, but are not included in the target sequence on the gRNA that chaperones the Cas endonuclease to the target site. Instead, PAM sites are located directly downstream of the target sequence and are required for activation of the DNA-endonuclease activity. Types I and III both use multi-protein complexes to recognize and cut foreign genetic material, while Type II uses a single enzyme to accomplish this task. Cas9 is one example of a single-sequence encoded Type II DNA-endonuclease. The Cas9 from *Streptococcus pyogenes* (SpCas9) is the predominantly used Cas9 enzyme for genetic engineering due to its short 3 bp PAM site sequence (NGG), compared to the Cas9 PAM sites for other species which have longer and less useful sequences (e.g., NNGRRT in *Staphylococcus aureus* and NNNNGATT in *Neisseria meningitidis*). For this reason, the Cas9 from *S. pyogenes* is simply referred to as "Cas9" in most CRISPR Cas9 work. Other Type II DNA-endonucleases, such as CRISPR-Cfp1, are smaller but share many of the same features as Cas9; however, the PAM site sequences, orientation, and target design vary. While there are certainly many interesting applications of other Type II enzymes, as well as those from Type I and III systems,

the simplicity, short PAM site, and comparatively small coding sequence for the Cas DNA-endonuclease of the Type II system makes Cas9 the model candidate for delivery of CRISPR for gene editing in new systems.

To utilize Type II CRISPR technology for genetic engineering, there are two main components that are required, the Cas9 DNA-endonuclease enzyme and a single guide RNA (sgRNA) that is comprised of the 20 base pair (bp) target sequence, and the 77 bp region encoding the RNA scaffolding flanked by five thymine residues coding for termination of transcription, totaling a 102 bp RNA fragment. The sequence for Cas9 is publicly available and can be found in Supplemental Material (Supplemental File C5-1). While the Cas9 amino acid sequence is largely the same regardless of the system in which it is to be used, codon optimization of the Cas9 gene for optimal translation in your particular organism improves efficiency when using a transformation that relies on translation of the supplied gene within the host. This can be accomplished if the codon bias is known for your organism; however, if codon bias information is not available, a plasmid for use of CRISPR Cas9 optimized for a closely related system can be obtained from one of the many plasmid repositories, such as Addgene.org (https://www.addgene.org/), and may work well enough in your organism. If you are using Cas9 in a eukaryotic system, it is important to include the addition of a nuclear localization signal (NLS) at the C or N terminus of the Cas9 protein to allow the Cas9 to enter the nucleus after translation (Supplemental File C5-1). Like Cas9, the gRNA scaffold code will also remain the same, unless you are working with a modified version of Cas9 designed for shorter sgRNA sequences (Xu et al., 2017).

While the Cas9 and gRNA scaffold sequences will remain relatively the same regardless of the system in which they are being used, their expression should be directed via endogenous or tested promoters for each new system. To drive the expression of the nuclear localization signal tagged Cas9, a strong promoter from your organism should be cloned upstream of the Cas9 coding region, thus yielding high-level expression. Alternatively, differentially expressed promoters can be used to express Cas9 under specific conditions. For expression of Cas9 in eukaryotes, an RNA Pol II promoter can be used; however, while Cas9 expression can utilize such promoters, a promoter for RNA Pol I or RNA Pol III must be used to drive transcription of the sgRNA component in order to avoid the addition of a 5'-cap and 3'-poly A tail. Promoters for U6 or U3 snRNAs are conventionally universal promoters for RNA Pol III that have been used with widespread success to transcribe sgRNAs in many, but not all, systems (Cong et al., 2013). For use in prokaryotes, the Cas9 encoding region can be cloned in between any conventional promoter and transcription

terminator site. The sgRNA however needs to be cloned in exactly 10 bp downstream of the -10 Pribnow box in order to ensure that no extra base pairs are added onto the 5' end of the target region of the sgRNA.

An alternative approach for expression of the sgRNA component in eukaryotes is to use the organism's natural tRNA promoter to create an sgRNA/tRNA chimera (Mefferd et al., 2015). This can be done by fusing the sgRNA sequence directly downstream of a functional tRNA, followed by a termination sequence. This will allow your organism to transcribe an sgRNA/tRNA fusion product that, after post-transcriptional modification of the tRNA, will yield both a functional tRNA and the sgRNA for use by Cas9. While use of tRNA promoters can allow for expression of sgRNAs in systems with limited genomic annotation, tRNA promoters can reduce efficiency of Cas9 activity when compared to the U6 or U3 promoters (Wei et al., 2017). Even so, U6 promoters have been demonstrated to be unsuitable for driving sgRNA in some organisms. For example, U6 promoters are not used in many yeasts. Instead, at least one group has used the promoter from the *Saccharomyces cerevisiae* small nucleolar RNA, snR52 (*Psc*-SNR52), shown to be effective in a variety of yeast species (Wang et al., 2018). Furthermore, while using RNA Pol III systems is the predominant option for initiating CRISPR Cas9 technology in a new system, inventive researchers have utilized artificial enzymes to remove the transcriptional additions of RNA Pol II and prevent the sgRNA from being transported to the cytosol, demonstrating that it is possible to express a functional sgRNA using Pol II systems, if necessary (Nødvig et al., 2015; Zhang et al., 2017). Examples of this alternative Cas9 expression approach can be found in several systems developed for a variety of filamentous fungi, where efficient promoters for RNA Pol III are not well characterized (Nødvig et al., 2015; Schuster et al., 2019).These researchers used a variant of the CRISPR-Cas9 system in which the sgRNA is embedded in the middle of a larger transcript synthesized by RNA polymerase II; the sgRNA is liberated from the larger transcript in the nucleus by the action of two ribozyme sequences, 5'-end hammerhead (HH) and 3'-end hepatitis delta virus (HDV), which flank the sgRNA (Nødvig et al., 2015). The RNA Pol II promoter driving expression of this RNA is the *Aspergillus nidulans gpdA* promoter, a strong constitutive promoter; this is combined with the *trpC* transcriptional terminator (Nødvig et al., 2015).

Once implemented in your system, CRISPR Cas9 can be easily adapted for new targets by simply swapping out the 20 bp target region of the sgRNA for a new target. However, even though CRISPR Cas9 is a highly robust system once implemented, it is not without its own limitations. The PAM site for Cas9

restricts the cutting activity of the enzyme to NGG sequences, which theoretically occur 1 in every 16 bp in the genome. Furthermore, off-target cuts are a major concern for the application of CRISPR Cas9. While the target region of the gRNA binds to 20 complementary bp, mismatch pairing is able to occur closer to the 5' end of the RNA (Hsu et al., 2014). With these limitations in mind, the rest of this introduction will serve as a guide through the process of implementing CRISPR Cas9 in a new system, after which, I will present data on our efforts to do so in the *Microbotryum* genus.

1.1 Things to consider before you begin

Before designing a CRISPR Cas9 construct for use in a new system, there are several parameters to consider: 1) what transformation systems are available for your organism? 2) What sort of genetic changes are your trying to make in your organism? 3) What is the target for your pilot study? 4) How do you plan to ensure transcription and translation of the Cas9 enzyme? And 5) How do you plan to provide transcription of the sgRNA?

1.2 Selecting a transformation method

Establishing a CRSIPR Cas9 protocol for a new organism will rely heavily on the methods of transformation available for that system. This section will provide an overview of several approaches that have been used with widespread success. Each of the following methods have their own pros and cons associated with implementing CRISPR Cas9, and should be considered in the context of your organism and the goals of your research. For example, some projects may tolerate off-target cutting while others may require strict fidelity of the enzyme. Furthermore, some organisms may be limited by their genetic toolboxes to a subset of the following approaches.

1.3 Single plasmid method

The conventional method of implementing CRISPR Cas9 is to use a plasmid that contains all the necessary coding regions and promoters to drive transcription and translation of the Cas9 DNA-endonuclease and the associated sgRNA within the organism itself, utilizing the endogenous cell machinery to do most of the work for you. Introduction of CRISPR Cas9 in this way requires, in addition to the coding regions for the sgRNA

and Cas9-NLS, a promoter for Cas9 expression, a promoter for the sgRNA, a selectable marker for successful transformation, and either an Origin of Replication (ORI) for prokaryotes or an Autonomous Replication Sequence (ARS) for eukaryotes to maintain the plasmid through cell division (Figure C5-1).
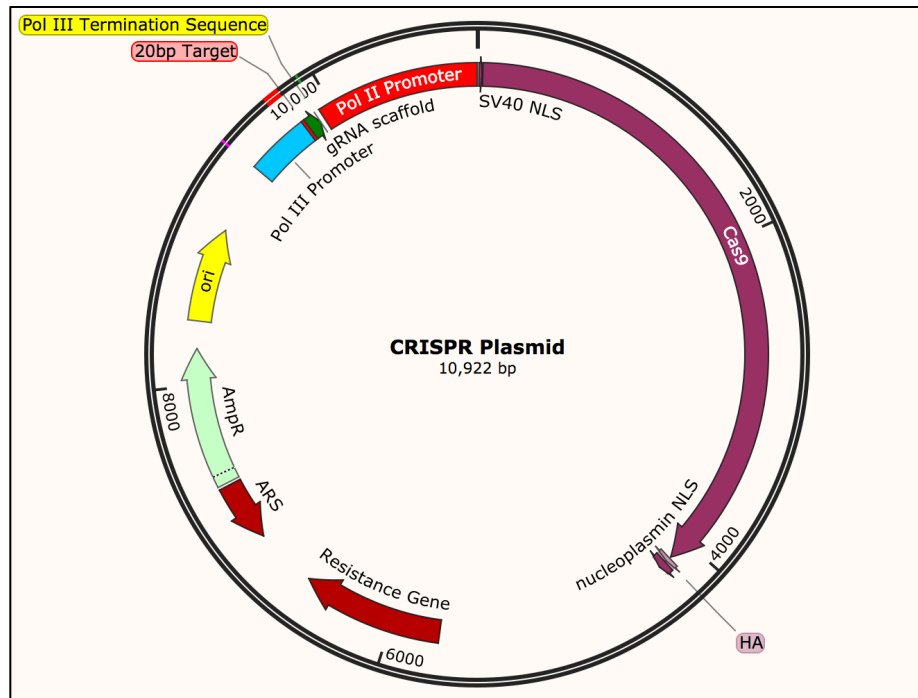
Figure C5-1 Diagram of the components required to implement CRISPR Cas9 via a single plasmid in a eukaryotic system. The components required are an RNA Polymerase III promoter, 20 base pair target, guide RNA scaffold, an RNA Polymerase II Promoter, Cas9 coding sequence with nuclear localization signal, selectable resistance gene for the transformed host, autonomously replicating system/ origin of replication. Image generated with SnapGene®.

Introduction of CRISPR Cas9 via plasmid can be performed using various transformation methods, including Heat Shock Transformation or Electroporation in prokaryotes (Froger and Hall, 2007; Tu et al., 2016), Polyethylene Glycol (PEG) facilitated transformation in fungi (Liu and Friesen, 2012), biolistics in plants (Char et al., 2017; Ismagul et al., 2018), and viral infection of animal cells (Kost et al., 2005). To identify successful transformants, cells carrying the plasmid are selected using media containing the antibiotic corresponding to the selectable resistance gene on your CRISPR plasmid; such initial transformants are then

subsequently passaged on non-selective media to allow cells to lose the plasmid during subsequent rounds of mitotic cell division. This will limit the exposure of the cells to Cas9 activity, thus lowering the risk for off-target cuts and unwanted modifications to the genome. Cells can then be screened for CRISPR-mediated changes.

1.4 *Agrobacterium*-mediated method

While plasmid introduction of CRISPR Cas9 is the most common and straightforward method, without an ORI or ARS, selection of transformed cells would require integration of the vector into the host genome. An alternative transformation method for organisms for which an ORI or ARS is unknown is *Agrobacterium*-mediated transformation (ATMT) (Char et al., 2017). To perform ATMT on eukaryotic cells, a Ti plasmid is used with the following components cloned between left and right T-DNA borders; an RNA Pol II promoter, the Cas9 encoding gene, a Pol III promoter, and the sequence for the sgRNA (Figure C5-2).
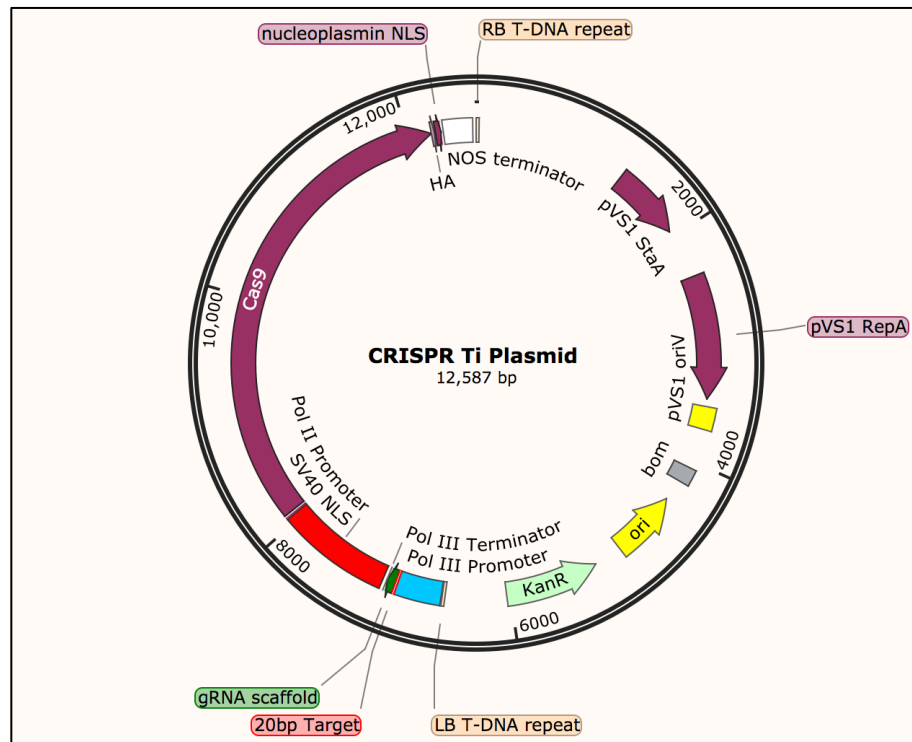
Figure C5-2 Diagram of the components required to implement CRISPR Cas9 via *Agrobacterium* mediated transformation in a eukaryotic system. The required components are: an RNA Polymerase III promoter, 20 base pair target, guide RNA scaffold, an RNA Polymerase II Promoter, Cas9 coding sequence

with nuclear localization signal. These components are cloned in-between the left and right T-DNA border regions and transformed into *Agrobacterium tumefaciens* for transformation of various cell types. Image generated with SnapGene®.

These Ti plasmids are then transformed into *Agrobacterium tumefaciens*, which will be used to shuttle the DNA between the T-DNA borders into the genome of your host. To do so, transformed *Agrobacterium* are plated together with the host cells on mating agar containing acetosyringone. The acetosyringone causes the *Agrobacterium* to form conjugative structures that will transfer the T-DNA cassette into essentially any cell type (Char et al., 2017; Kunik et al., 2001; Toh et al., 2016; Liu et al., 2017). While this method eliminates the need for an ORI or ARS, the CRISPR Cas9 components become permanently incorporated into the host genome, and thus removal of Cas9 via selection-free cell passaging is no longer possible. This ultimately increases the risk for off-target effects of Cas9; however, measures can be taken to reduce the risk of off-target hits when using ATMT. One solution is to add a second sgRNA, in addition to the sgRNA for your target of interest, one that targets the Cas9 gene itself, effectively acting as a kill switch and limiting the exposure that the functional Cas9 enzyme has with the host genome. While this method reduces off-target hits, it also reduces the efficiency of on-target cuts. Regulation of Cas9 with a differentially expressed promoter is another option for regulating Cas9 by controlling when the enzyme is being expressed through the use of particular media types. It is also important to consider that the random integration of the CRISPR cassette into transcriptionally active regions of the host genome may itself lead to gene interruptions and should be screened before attributing phenotypic changes to Cas9 action.

1.5 mRNA-encoded Cas9 method

Another alternative for systems without a viable means of plasmid replication is to transform cells with mRNA coding for the Cas9 enzyme. Like the plasmid method, the mRNA must also include the sequence for an NLS tag in order for the translated protein to be shuttled back into the nucleus; however, unlike the plasmid method, the mRNA itself does not require transportation into the nucleus and, of course, the mRNA is far less stable in the cell. While at first glance, a lack of stable expression of Cas9 may seem inefficient due to the limited window for expression and action of the Cas9 enzyme, the limited exposure of cellular

DNA to the Cas9 product reduces the risk of off-target cuts, therefore, making this method advantageous for stringent projects that cannot tolerate off-target cutting. This method is the predominant approach for animal systems and is often performed via *in vitro* production of mRNA followed by microinjections of a mixture of sgRNA and the Cas9 mRNA product into the cells (Figure C5-3).
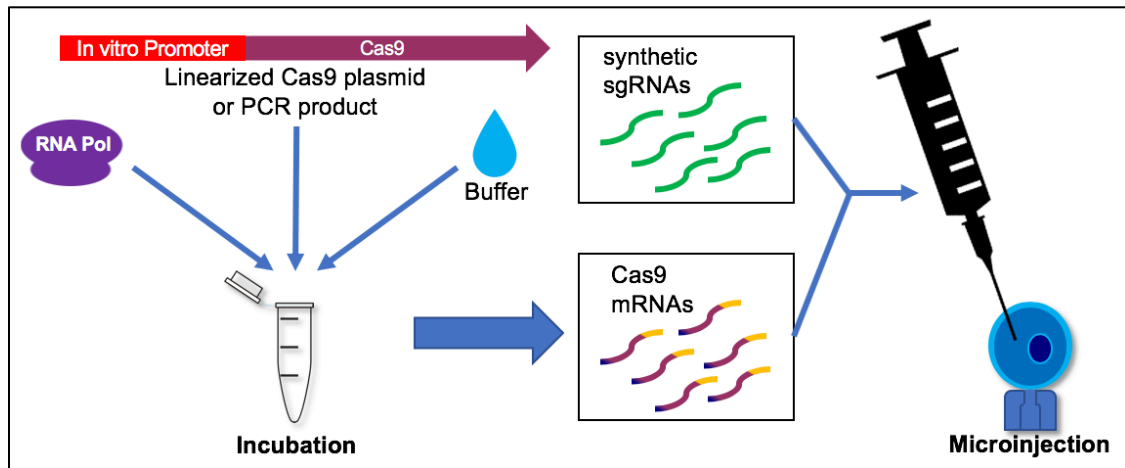
Figure C5-3 Introduction of CRISPR via mRNA-encoded Cas9. Cas9 plasmid with a promoter modified for use with *in vitro* mRNA synthesis kits can be linearized or used as template for PCR to generate the template for mRNA synthesis via RNA Polymerase. This process differs by kit, but usually includes an incubation period with a master mix or buffer, followed by an RNA purification step before use for microinjections into cells.

Transformation of cells using the mRNA-encoded Cas9 CRISPR method has the advantage of limiting the exposure of the cells to Cas9, since both mRNA and Cas9 protein lead to fewer opportunities for off-target cuts; however, genetic modification efficiency is also reduced when compared to more stable methods of Cas9 expression. In addition to reduced efficiency, this method also requires the tedious practice of isolating and handling RNA, which requires the use of kits including RNase-free buffers and lab equipment, as well as sterile RNA workspaces.

1.6 Purified Cas9 method

For systems where the genetic information to drive expression of engineered genes is unknown, the following approach can be used. Rather than driving expression of the sgRNA or Cas9 in the host via a plasmid or mRNA construct, one can now purchase commercially available recombinant Cas9 proteins from various sources. Alternatively, a protein expression plasmid utilizing a T7 promoter (pET plasmid) can be used to express and purify the Cas9 from bacteria, using a Histidine tag (Cas9-NLS-6xHis). *E. coli* DE3 cells can be easily transformed with these pET plasmids and grown with IPTG to induce production of the Cas9 enzyme. Cells can then be sonicated to obtain a Cas9-rich lysate for Ni column purification. The purified stock of enzyme can then be combined with synthetically produced sgRNAs and transformed into the host cells (Foster et al., 2018) (Figure C5-4).
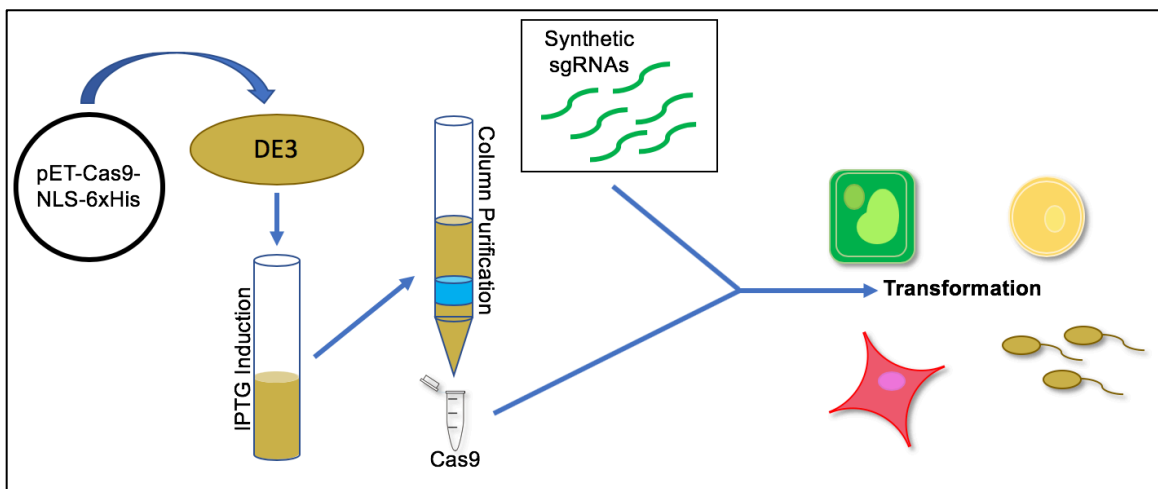
Figure C5-4 Protein purification of Cas9 for introduction to various cell types along with synthetic sgRNAs. IPTG-inducible Cas9 plasmids can be transformed into DE3 *Escherichia coli* cells and induced overnight in 2X growth media. A 6X-His tag fused to the end of the Cas9 DNA-endonuclease enzyme allows for its purification on nickel (Ni) columns. Synthetically produced sgRNAs can then be mixed with the Cas9 enzyme prior to introduction to various cells types.

Alternatively, as indicated earlier, for labs with the necessary funds, purified lyophilized Cas9 enzyme can be purchased from companies such as New England Biolabs, IDT, or System Biosciences© pre-made and ready for use. While this method requires the least amount of genetic information for the organism with

which you are working, and has one of the lowest chances for off-target effects, this method has comparatively low transformation efficiency compared to the other methods and the purchase of sgRNA oligos can be expensive. Newer approaches for the synthesis of RNAs *in vitro* have reduced the cost, making this a more viable approach for a larger breadth of labs, by taking advantage of the hydrogen bonding involved in the formation of loop structures to attach the sgRNA to the Cas9 enzyme to produce the sgRNA as two separate 60 base pair RNA oligos and relying on complementary sequences to hold them together before duplexing with the Cas9 enzyme (Figure C5-5). These two separate RNA strands are the crRNA, which includes the desired host target sequence, and tracrRNA, which contains the RNA scaffolding required to bind to the Cas9 endonuclease. By reducing the overall length of synthesis needed for each RNA oligo, price can be dramatically reduced. Furthermore, tracrRNAs are specific to the particular endonuclease to which they bind and therefore do not need to be resynthesized every time a researcher wishes to change targets. Instead, a new crRNA can be synthesized and complexed with the previous stock of tracrRNA, further reducing costs.
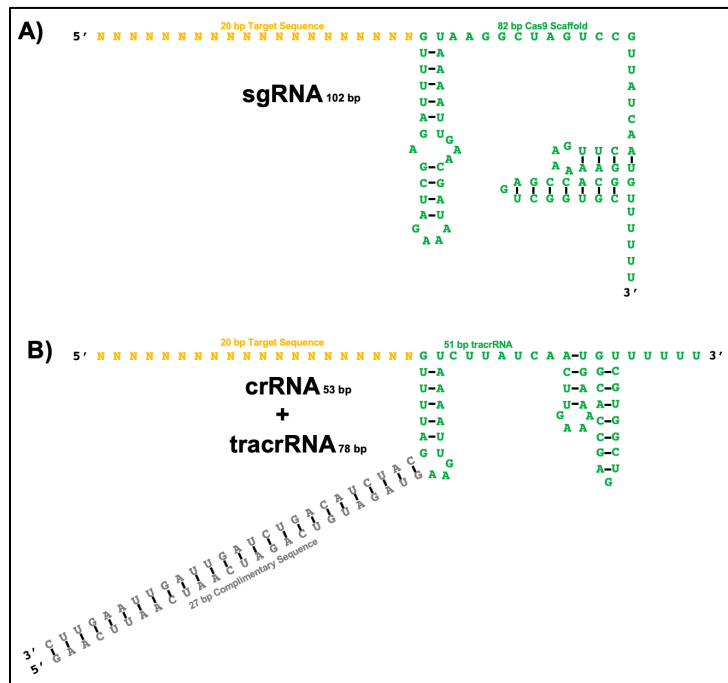
Figure C5-5 Diagram of the differences between sgRNA and crRNA/tracrRNA molecules. Hydrogen bonding arrangement of the traditional sgRNA used for CRISPR Cas9, A, and the more recent two oligo duplex using crRNA and tracrRNA, B.

If cost still remains an issue for the researchers, another option is to produce these sgRNAs on your own using an RNA synthesis kits, similar to the ones mentioned in the mRNA-Encoded Cas9 method.

1.7 Selecting a suitable Cas9 variant

The Cas9 DNA-endonuclease enzyme, originally isolated from *Streptococcus pyogenes* (SpCas9), has since been modified in several ways to alter its function and fidelity within the cell. Cas9 enzymes have two functional nuclease domains, the single nuclease domain HNH, which acts to cut the strand complementary to the gRNA, and RuvC, comprised of 3 subdomains, responsible for cutting the non-complement strand (i.e., the strand with the PAM site). Together, these active domains work together to create a blunt cut across the double stranded DNA target, 3-4 bp upstream of the PAM site. By changing the amino acid composition of these functional domains, Cas9 can be used for other types of genetic modification in your system (Figure C5-6).
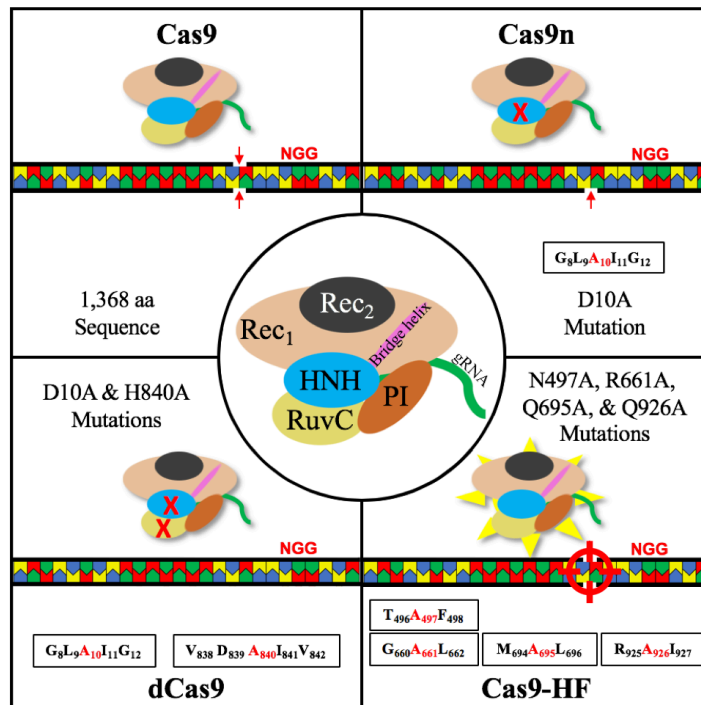
Figure C5-6 Illustration of various Cas9 mutants and their amino acid substitutions. The Cas9 DNA-endonuclease enzyme, color coded by protein domains, center. The traditional double stranded cut-inducing SpCas9 enzyme, top left. The HNH-nickase Cas9 mutant (Cas9n), top right. The HNH-RuvC double

knockout mutant for enzymatically dead Cas9 (dCas9), bottom left. And, the high-fidelity Cas9 variant (Cas9-HF), bottom right.

1.8 Cas9 endonuclease

Cas9 functions to generate double stranded cuts 3-4 bp upstream of the protospacer adjacent motif (PAM) site of your target (Cong et al., 2013). These double stranded breaks are then repaired through either non-homologous end joining or homology directed repair within the cell (Hsu et al., 2014), processes that are error prone and often lead to the removal or addition of nucleotides at the ends of the breaks. The addition or deletion of nucleotides before the strands are reconnected induces a frameshift that can lead to a change in reading frame and the introduction of premature stop codons, ultimately changing the peptide sequence and truncating the translated protein, giving Cas9 its characteristic knock-out function. In the event that the cell does manage to repair the double stranded break appropriately, the target site for the Cas9 is also repaired and further cuts can be made until there is a change to the sequence.

In addition to generating frameshifts, double stranded breaks induced by Cas9 have also been demonstrated to dramatically increase the rate of homologous recombination in cells (Miki et al., 2018). Selectable markers with upstream and downstream complementary regions to either side of the targeted double stranded cut in the organism's genome can be introduced along with Cas9 to facilitate homologous directed repair of the double-stranded break to insert a selectable marker into the coding region of the gene you wish to knockout (Miki et al., 2018) (Figure C5-7).
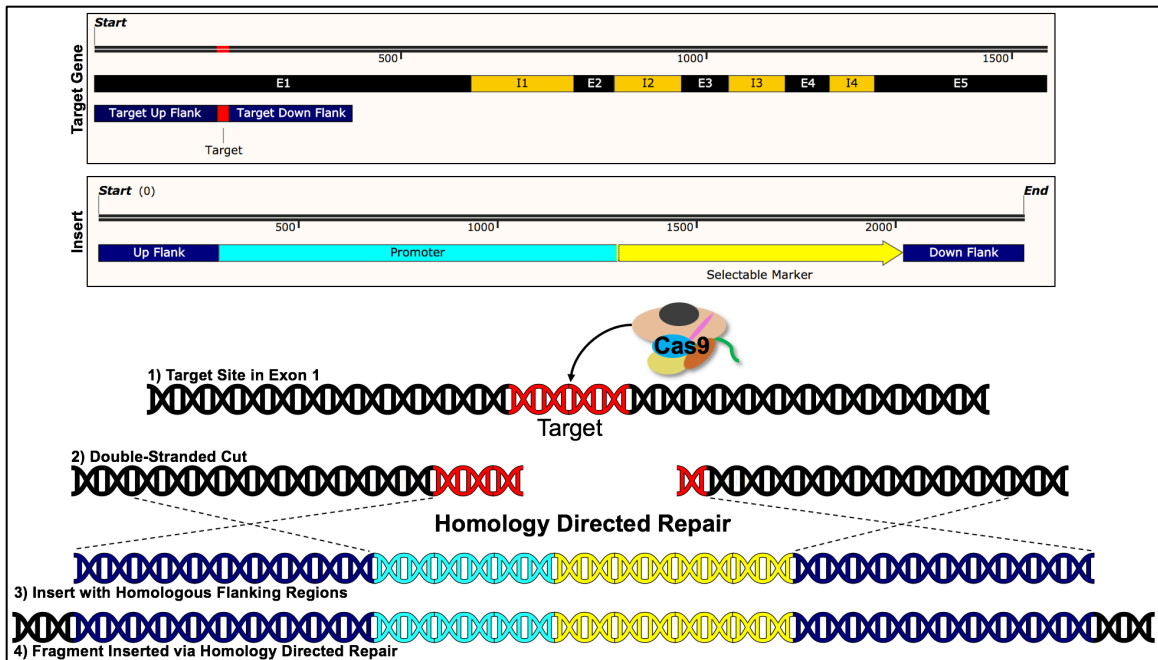
Figure C5-7 Homology directed repair of a Cas9-induced doubled-stranded break in the first exon of a target gene. 1) the targeting and binding of Cas9 to the first exon region of the target gene. 2) The endonuclease activity of Cas9 generates a double-stranded break 3-4 bp upstream of the PAM site of the target region. 3) The homologous flanks of the insert are used as a template for homology directed repair of the double-stranded break. 4) The repair results in the addition of the DNA between the up and down flank homologous regions, disrupting the target gene and allowing for expression of a selectable marker. The top two images were generated with SnapGene®.

1.9 Cas9n nickase

By changing the amino acid sequence of one of the nuclease domains of Cas9, function can be altered to that of a nickase enzyme, where one nuclease domain is still functional and cuts one side of the DNA, but the other domain does not (Chew et al., 2016). Nickase Cas9 mutants (Cas9n) can be used with a pair of sgRNAs to generate cuts on either side of a target DNA sequence several base pairs apart, effectively allowing for the strands to separate with regions of overhangs that can then be utilized for site specific insertions of DNA. Cas9n can thus be used for incorporation of transgenes into site-specific loci, or used to insert screenable markers, such as GFP, to disrupt function and screen for knockouts within a system. While Cas9n utilizes two sgRNAs, thus theoretically doubling the risk of off-target cuts, the lack of complete double stranded

breaks reduces the risk of off-target changes in the genome, as nicks are easily repaired without mutation. This makes Cas9n a great approach for groups seeking to minimize off-target effects in their system. Because this system utilizes two sgRNAs and relies on strand separation of regions held together by hydrogen bonds between the two nicks, the PAM sites of both targets should be on opposite strands and no more than 30-40 bp apart in order to effectively allow the two nicked strands of complementary DNA to separate. Increasing the distance between nicked strands increases the number and consequently the strength of the hydrogen bonds, ultimately decreasing the likelihood that the strands will separate.

1.10 dCas9 for CRISPRi

Another type of Cas9 mutant can be made by disrupting both nuclease domains, generating an enzymatically dead variant of Cas9 (dCas9) (O'Geen et al., 2017). While this Cas9 mutant is unable to cut the target, it is still able to bind specifically to the target site. This variant of Cas9 can be used to either interfere with cellular processes at the transcriptional level (CRISPRi), or in conjunction with activation domain attachments to act as an inducible activation delivery system. dCas9 has also been used as a vehicle for fluorescent tags for real-time DNA imaging in cells (Chen et al., 2013).

1.11 High fidelity Cas9 variants (Cas9-HF, eSPCas9, and xCas9)

In addition to enzymatic ability, another thing to consider when selecting a Cas9 variant is the importance of preventing off-target cuts in your system. High-Fidelity variants of Cas9 have been generated for use in systems that may need increased specificity, such as eSpCas9 and various forms of Cas9-HF (Kim et al., 2017; Kleinstiver et al., 2016). These Cas9 variants have amino acid substitutions that reduce the likelihood that annealing of the gRNA with mismatches at the 5' end of the target sequence will induce enzymatic activity in the Cas9 enzyme. In addition to the creation of high-fidelity Cas9 variants, recent researchers at Harvard and MIT have generated SpCas9 mutants through directed evolution that can utilize multiple different PAM site sequences (xCas9), further increasing their use for very site-specific purposes (Hu et al., 2018). Not only can these xCas9 variants recognize PAM sites other than NGG (e.g., NG, GAA, and GAT), but they also surprisingly demonstrated increased editing specificity (Hu et al., 2018). These High-Fidelity Cas9 variants can additionally be modified to operate with any of the aforementioned Cas9 enzymatic

abilities as well, although some of them have modified components of the sgRNAs that need to be taken into consideration when designing a target, such as alterations to the recognized PAM sequence in xCas9 (Hu et al., 2018). The sequences with appropriate substitutions for each of the aforementioned Cas9 variants can be viewed in the Supplemental Materials (Supplemental File C5-1).

1.12 Transposon Associated CRISPR Cas9

One group has created yet another form of CRISPR Cas9-mediated transformation that is able to insert DNA into a targeted area without the need for any cutting of the host genome at all (Klompe et al., 2019). This CRISPR approach utilized the specificity of the Cas9 sgRNA chaperone and a transposable element system to insert DNA without the risk of unintentional degradation to the host genome, making CRISPR applicable to systems with poor repair machinery, such as certain bacteria, where double-stranded breaks are fatal to the cell.

This idea of fusing CRISPR Cas9 with other gene editing tools has also been used to generate enzymes capable of site-specific nucleotide substitutions. By tethering cytidine or adenine deaminase to Cas enzymes, CRISPR can be used to introduce C-to-T or A-to-G changes to specific regions of a gene without the need for cutting the genome (Gaudelli et al., 2017; Zheng et al., 2018). The ability to make site-specific single nucleotide changes to DNA allows the user to change particular amino acids in proteins, a tremendous tool for those performing directed experimental evolution studies or investigating the function of proteins. While CRISPR-mediated base editing is currently limited to C-to-T changes and A-to-G changes, new CRISPR applications like the aforementioned Cas9-transposable element are expanding the capacity for genome editing without the need for cutting (Komor et al., 2016).

1.13 Selecting a target

One of the more challenging aspects to using CRISPR Cas9 for the first time is screening for cells with the intended genetic modifications. Cas9 DNA-endonuclease activity leads to double-stranded breaks in the primary DNA of an organism; however, not every cut leads to elimination of function. When repairs are made to the double-stranded breaks in the cell, some of the DNA will be successfully put back together with the addition or deletion of base pairs in triplicate, e.g., three nucleotides are deleted in a unimportant region

of the coding DNA. When this occurs, function of the targeted protein may persist. Furthermore, targeting of an intronic region in eukaryotes will also likely not lead to changes in protein function. Therefore, selection of an appropriate region of the DNA for the gene you want to disrupt is imperative. This can be accomplished by either targeting the active site for a protein, where even single amino acid insertions or deletions in the protein can cause disruption, or selecting a target as far upstream in the coding region as possible to increase the impact of induced frame shifts.

To determine whether CRISPR Cas9 has successfully altered the target in your system, a method for screening transformed cells will be needed. While genetic sequencing of the target region is a necessity to demonstrate successful genetic modification, screening thousands of colonies to find a successful knockout can be costly and time consuming. Therefore, pilot studies for implementing CRISPR Cas9 in a new system would do well to start by targeting a gene that leads to an easily discernible phenotypic alteration. Inducing a color change or changing an organism's ability to grow on certain media types is a great way to initially implement CRISPR Cas9 in your organism, which can then be altered later on for different targets. Targets that provide easily-identifiable phenotypic changes can also be used in tandem with future targets that may not lead to an observable change in order to select for cells in which the active CRISPR Cas9 has been successfully delivered, reducing the number of cells where DNA sequencing will be required to identify changes. If you are implementing CRISPR in a system that does not have the genetic information for an observable phenotypic change, the organism can first be transformed with a marker, such as GFP, that can then be subsequently targeted with Cas9 to test CRISPR in your system (Watters et al., 2018). Alternatively, Cas9n mutants can be used to insert markers like GFP into the coding region of the gene you wish to knockout.

1.14 Selecting a promoter for Cas9

Depending on the transformation method available to your system, different strategies can be used to drive Cas9 expression. To drive the transcription of plasmid-encoded Cas9, a constitutive promoter from a known gene from your organism can be cloned upstream of the coding region for a codon-optimized Cas9 variant. After selection of successful transformants, the Cas9 plasmid can be eliminated by repeated passaging using non-selective media (Schuster et al., 2016). When using a constitutive promoter, Cas9-HF is recommended

since the Cas9 will be constantly expressed during the passaging process and may lead to off-target cuts. An alternative to using a constitutive promoter for the Cas9 plasmid is to use a differentially expressed promoter that can upregulate expression of Cas9 on certain media types during transformation/selection and then be downregulated during the passaging step of the process. RNA-seq data can be used to identify promoters that drive the transcription of certain genes on rich media, but have significantly reduced transcription on minimal media, or *vice versa*.

Organisms for which the use of plasmids are not suitable, and more permanent means of transformation like homologous recombination or *Agrobacterium*-mediated transformation are used, differentially expressed promoters are recommended. Off-switches can also be used to conditionally inhibit Cas9 activity after its initial introduction into a system, giving the Cas9 time to make changes before it is turned off (Pawluk et al., 2016).

In systems where the Cas9 enzyme is purified from bacteria and added separately, commercially produced pET Cas9 plasmids are available that use an IPTG-inducible promoter for expression of Cas9 in large abundance within *E. coli* cells. For extraction of Cas9 in this manner, pET Cas9 plasmids with 6xHis tags can be obtained (e.g., Addgene.org) to purify Cas9 enzyme from bacterial lysate using a nickel flow column.

1.15 Selecting a promoter for the sgRNA

Like the Cas9 encoding region, a promoter is needed to drive transcription of the sgRNA that includes your targets and the scaffold for adherence to the Cas9 DNA-endonuclease. Transcription of sgRNAs in prokaryotic cells often uses a promoter modified with an SpeI restriction site to allow for easy exchange of the downstream target site later on. Transcription of the sgRNAs in eukaryotic systems needs to avoid adding a 5' cap or 3' poly-A tail to the RNA product, a feature of RNAs produced by RNA Polymerase II. The U6 promoter offers a somewhat universal opportunity for many, but not all, eukaryotic systems, although it requires the 5' end of the sgRNA target to begin with a G. Another potential drawback of the U6 promoter is that, in some cases, the U6 RNA itself contains regulatory sequences that can affect expression (Reich et al., 1990). For systems where the U6 promoter is not able to be utilized, use of a tRNA/sgRNA chimera is a suitable alternative to drive expression of the sgRNA by fusing it to the end of a highly expressed tRNA in

your system (Mefferd et al., 2015). Post transcriptional modifications to the tRNA will then separate the two, leaving the cell with a functional copy of both the tRNA and your sgRNA (Mefferd et al., 2015) (Figure C5-8). While uncommon, it is possible to drive expression of sgRNAs as an mRNA using polymerase II (Zhang et al., 2017) or as polycistronic miRNA (Xie et al., 2017).
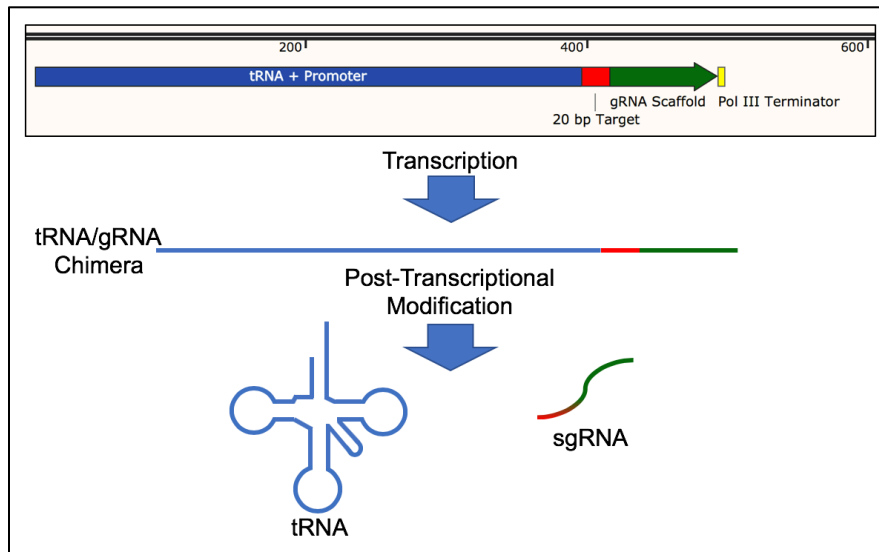
Figure C5-8 Endogenously driven synthesis of a sgRNA via fusion with a native tRNA. The sgRNA sequence flanked by an RNA Polymerase III terminator is cloned in place of the tRNA termination signal. Once RNA Polymerase III begins transcription of the tRNA coding sequence, it will continue through the 20 base pair target sequence and guide RNA scaffold, synthesizing a tRNA/gRNA chimera. Post-transcriptional modifications, including folding of the tRNA and processing via RNase Z, will generate a functional tRNA and result in the release of the sgRNA oligo.

1.16 Transformation and confirmation of alterations to target genes

Identification of successful mutations to the target region depends on the method of delivery of Cas9 into an organism, as well as which targets were used. For investigators who elect to utilize a plasmid containing the coding region for both the sgRNA and Cas9, the selectable marker on the plasmid is the first step in identifying transformants. Using a plasmid that has a resistance gene (e.g., conferring resistance to ampicillin or kanamycin in bacterial systems, hygromycin B or carboxin for fungal systems, glufosinate, glyphosate, and kanamycin for plants, blasticidin S and thymidine kinase in mammalian cells), will allow for selection

on media containing the drug for cells that have received the CRISPR Cas9 plasmid. Selectable markers are also possible for genomic transformations with ATMT and baculoviral transformation (Char et al., 2017; Hindriksen et al., 2017).

Because not all successful cuts with Cas9 disrupt the targeted protein, additional screening is required to identify cells with successful knockouts. For pilot studies, knocking out a gene that causes an observable phenotypic change without killing the organism is recommended. These sorts of changes can then be used in tandem with other sgRNA targets for future studies to select for cells with known functional Cas9 activity. However, in-frame deletions (i.e., in deletions in multiples of 3 bp), unless in coding regions for catalytically important amino acids, may not lead to disruption of the functional product. These in-frame deletions should theoretically be expected to occur randomly in 33.3% of cuts. Therefore, regardless of selection and screening, all Cas9-mediated changes should be verified via sequencing of the targeted region. Because double stranded cuts can lead to large segment deletions, to perform PCR at the target site, it is recommended to use primers that are least 100 base pairs upstream and downstream of the cut. When using Cas9n for inserting genes into a target region of the genome, screenable markers like GFP can be inserted to screen transformants for successful incorporation of the gene fragment. This can be followed up with sequencing to verify insertion of the marker into the proper location. In instances where a double mutant CRISPRi system is employed, qRT-PCR can be used to identify whether repression of the target gene has been achieved.

1.17 Changing targets and targeting multiple genes

One of the more exciting aspects of CRISPR Cas9 technology is the ability to target multiple genes using the same Cas9 enzyme. Because the sgRNA with an RNA Pol III termination sequence is only 102 bp in length, plus a Pol III promoter, several sgRNA cassettes can be included on a single CRISPR plasmid (Sekine et al., 2018), or simply added in groups together with mRNA encoded Cas9 or purified Cas9 enzyme without need to modify the enzyme. Another convenience of the system is the ease of changing sgRNA for new targets. Once CRISPR Cas9 has been established in your system, only the 20 bp target sequence needs to be changed in the sgRNA coding region to use the plasmid for a new gene. This can be easily and affordably accomplished with a 60 bp oligonucleotide consisting of the 20 bp target in the middle flanked by 20 bp of

complementary sequence to the Pol III promoter upstream and 20 bp of complementary sequence to the gRNA scaffold downstream; these pieces are then put together using Gibson overlap PCR, or variations thereof, and cloned into the CRISPR plasmid (Figure C5-9) (Gibson et al., 2009).
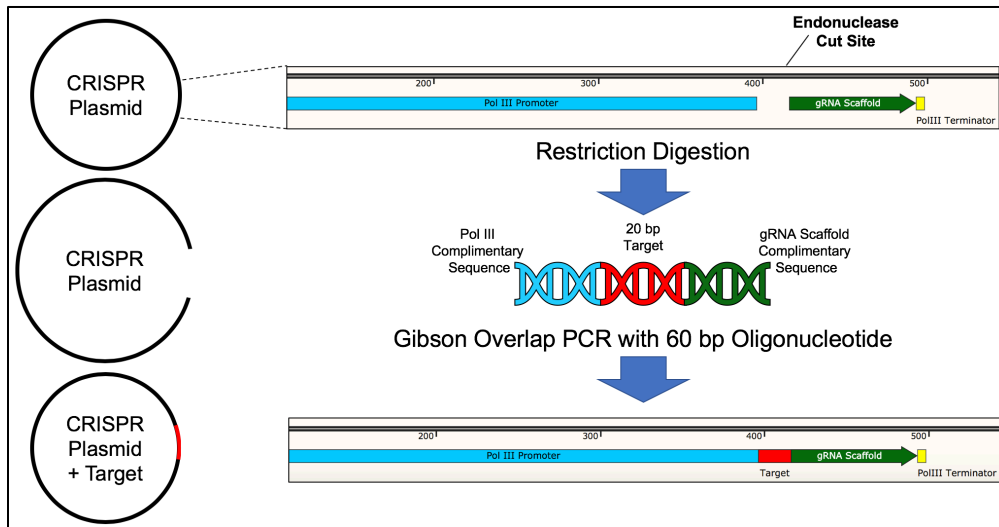
Figure C5-9 Illustration of easy target replacement in CRISPR Cas9 plasmid constructs. Insertion of a new target sequence via restriction digestion between the RNA Polymerase III promoter and guide RNA scaffold, followed by Gibson overlap PCR with a synthetically generated 60 bp oligonucleotide. The 60 bp oligo is comprised of three 20 bp components: 20 bp of complementary overlap sequence to the promoter, 20 bp of the desired target, and 20 bp of complementary overlap sequence to the gRNA sequence. Gibson overlap PCR of the digested CRISPR vector with the 60 bp oligo allows for easy cloning of new targets in-frame with the gRNA scaffold for use with Cas9.

For multiple targets, one can simply create several cassettes with different targets and amplify the product with primers that include appropriate cut sites for insertion into the plasmid. While modification of a CRISPR plasmid to target multiple genes can be done with ease, the likelihood of generating knockouts for all of the targets decreases with the number of targets being used. However, when considering the amount of work that would be necessary to sequentially introduce multiple genomic changes into a single strain of your organism, the additional screening to find cells with mutations in all targets is a much better alternative to other gene editing procedures. To estimate the likelihood that all targets would lead to knockouts in your system where

CRISPR Cas9 has been established, simply take the theoretical likelihood that a single cut will lead to a frameshift in the target gene (66.7%) and raise it to the power of how many targets you are using (e.g., for 3 targets the theoretical likelihood of finding a cell with all three genes knocked out is approximately 30%, while targeting 5 genes that would theoretically result in a quintuple knockout in roughly 13% of cells). Odds of success can be improved if the sgRNA target is in a coding region for the active site of a protein, where even in-frame deletions will lead to misfunction of the protein. Interestingly, because one would expect to have situations where some of the targeted genes are knocked out while other targets remain functional, the same single plasmid could additionally be used to test an array of combinations of functional genes followed by sequencing to determine which genes function under certain conditions, generating an extensive wealth of knowledge with a single transformation.

1.18 Checking for off-target cuts

While CRISPR Cas9 cuts with high specificity, especially the eSpCas9 and Cas9-HF variants, off-target cuts are possible and should be ruled out before associating phenotypic outcomes to a particular knockout in the cell. To understand the potential for off-target cuts, it helps to understand how Cas9 enzymatically operates. While the target sequence is used to bind to complementary sequence and chaperone Cas9 to the target region, one might expect the target to bind randomly in the genome once every 4^{20} bp (i.e., 1/1E12 bp), a level of certainty that for most organisms means a single site in the genome. This expectation is only true under the assumption that the genome is completely random and that there is no mis-annealing of the target sequence to the complementary sequence in the genome; however, research has demonstrated that the target sequence is much more specific at the 3' end, next to the PAM site, than at the 5' end (Hsu et al., 2014). While the base pairs at the 3' end of the target region are incredibly specific, the base pairs at the 5' end are not required for the conformational change of Cas9. Therefore, off-target binding is far more likely to occur in sequences with mismatches at the 5' end of the target. Additionally, Cas9 occasionally binds to sequences upstream of an N<u>A</u>G PAM site instead of the traditional N<u>G</u>G due to similarities between the hydrogen-bonding groups of purines (Hsu et al., 2014). To account for these potential mismatches, online tools such as E-CRISP Design created by the German Cancer Research Center are available to screen the genomes of organisms for potential

mismatches to help you create the optimal gRNA target for your sgRNA complex (http://www.e-crisp.org/E-CRISP/).

While these tools are helpful for reducing your risk of off-target cuts, whole genome sequencing for off-target cuts is the only way to be certain no off-target cuts have been made. If, like most researchers, the resources for screening multiple genomes for off-target cuts are not available to you, other methods can be used to either identify localization of the Cas9 within your cells, or to simply increase the certainty that observed phenotypic changes are due to the targeting of a particular gene, and not due to off-target cuts. To identify potential off-target cuts, a mismatch-detection nuclease assay can be performed (Wu et al., 2014). Cas9n could also be used to identify off-target regions via insertion of a bar code-like sequence that is not naturally found in your organism, followed by PCR to amplify regions where the short sequence is inserted. Multiple targets of the same gene can also be used sequentially to increase certainty that observed phenotypic changes are due to the targeted knockout. If the same phenotype is observed for all knockouts of the different areas of the target gene, using different target sequences, then the likelihood that the phenotypic changes are due to off-target interactions is greatly reduced.

1.19 CRISPR checklist

When it comes to implementing CRISRP Cas9 in difficult-to-work-with systems, a minimalist approach is recommended for pilot studies. Once a baseline functional CRISPR protocol is established, modifications can be made to make the CRISPR system more robust. While the first cut can be arduous, the capacity for applications of Cas9 are well worth the effort. To help design a CRISPR construct for use in non-model systems, the following checklist can be used as a roadmap for all the necessary components (Figure C5-10). While CRISPR Cas9 opens the door to genetic modification of many recalcitrant organisms, acting as a multi-tool for organisms with comparatively small genetic toolboxes, implementation of the system can be difficult. Many of the barriers to implementing the systems do not lie in the capacity of CRISPR Cas9 itself, but rather the delivery methods available to novel organisms.

Figure C5-10 Checklist, with examples, of components needed to generate a CRISPR Cas9 construct for use in a new system. Sequences for the example promoters and Cas9 variants can be found in supplemental files 1-3. Not shown are other species-specific RNA Pol II promoters used successfully to drive Cas9 expression in fungi (*Ustilago maydis,* p*hsp 70*, p*otef*; [Shuster et al., 2018]), protozoans (*Toxoplasma gondii*, *TUB1* promoter and *SAG1* 3' UTR; [Sidik et al., 2018]), insects (*Drosophila*, *hsp70Bb*; [Taning et al., 2017]), and zebrafish (tissue-specific promoters; [Albain et al., 2015; Hwang et al., 2013; Li et al., 2016]).

1.20 Implementing CRISPR Cas9 in the *Microbotryum* genus

The ability to manipulate an organism's primary genetic material depends heavily on the tools available to each system. Every established transformation approach can therefore be considered a new tool in an organism's "toolbox", the size of which varies among systems depending heavily on the difficulty of introducing material into cells and how the organism deals with the foreign material. Organisms that readily take up foreign DNA during transformation often draw the attention of researchers as easy to work with systems, and in turn the increase in their use results in the broadening of transformation approaches. Some of the most well studies organisms, e.g. *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*, are therefore the metaphorical "kings" of their respective kingdoms due to their ease of use, breadth of study, and consequently their comparatively large toolboxes.

Because of their extensive repertoire of approaches for molecular genetic analyses, model organisms are often used as proxies for studying genes from more recalcitrant organisms; however, studies that examine the unique relationships between species, e.g., exquisitely specific host-pathogen interactions, require at least a fundamental ability to generate knockouts in these species. Unique genomic milieus, reproductive cycles, co-evolution, or natural environmental factors are all variables that can have major effects on the ability of scientists to manipulate, species. The fungal pathogen species complex *Microbotryum violaceum* is one such recalcitrant group of organisms used to study the life cycle of dimorphic pathogens, the evolution of disease, and the emergence of host shifts (Schäfer et al*.,* 2010; Beckerson et al*.,* 2019); however, the *M. violaceum* species complex is difficult to transform and has a comparatively small molecular genetic toolbox.

While classical genetic approaches have been implemented in *Microbotryum* with widespread success, manipulating the primary genetic material of these fungi has proven difficult. Despite these difficulties, one

recent breakthrough study has demonstrated that *Agrobacterium*-mediated transformation (ATMT) can be utilized to randomly insert genetic material into transcriptionally active portions of the fungal genome (Toh et al., 2016), allowing for transgenic expression of genes in the *Microbotryum* species complex. However, although a reliable ATMT method has opened the door for over expression and transgenic expression of genetic material in this system, a means of site-specific gene targeting and knockouts have, until now, remained a major hurdle.

To establish CRISPR-Cas9 in *Microbotryum*, our group used a three-pronged approach of 1) modifying a *Ustilago maydis* CRISPR-Cas9 plasmid for delivery into *Microbotryum* via electroporation, 2) building upon the previously established ATMT approach to deliver the CRISPR-Cas9 components, and 3) utilizing a kit designed to create the Cas9-gRNA duplex *in vitro* and deliver it into the cells using PEG treatment of protoplasted cells.

Materials and Methods

2.1 Selecting a target

As a pilot study for demonstrating a successful knockout using CRISPR-Cas9 in the *Microbotryum* species complex, our group attempted to create an easily observable color phenotype change from the bright pink color of *M. lychnidis-dioicae* cells in their saprophytic yeast-like life stage by targeting the coding region for a putative oxidoreductase protein, MVLG_05585, with predicted activity in the carotene biosynthesis pathway (Garber and Day, 1985). To minimize off-target Cas9 activity, a high-fidelity variant of Cas9 (Cas9-HF) was used for the ATMT approach and target sequences least likely to cause off target cuts were selected for MVLG_05585 following guidelines in the CRISPR-Cas9 review published by Hsu, Lander, and Zhang (Hsu et al., 20174). Target sequences that met these criteria were used in Blastn searches against the available *Microbotryum lychnidis-dioicea* genome available at the Joint Genome Institute (JGI) Genome Portal (https://mycocosm.jgi.doe.gov/Micld1/Micld1.home.html), and any target with more than 1 hit was removed. For the ATMT method, a single target was selected within the first exon of the MVLG_05585 gene. For the PEG treatment approach, two targets were selected, one at the 5' beginning of the MVLG_05585 gene and

the other at the 3' end of the MVLG_05585 gene. These targets were used simultaneously with the Cas9 enzyme in attempt to completely excise the gene to ensure a knockout.

2.2 Constructing a plasmid for delivery of CRISPR-Cas9 via electroporation

The single plasmid construct for expression of a sgRNA and Cas9 in *Microbotryum*, pMvCC9, was created via modification of pMS10, provided by Dr. Regine Kahmann, which included the sequence for the Cas9 endonuclease optimized for *Ustilago maydis* expression, an poly A termination sequence, as well as an sgRNA cassette driven by the universal U6 Polymerase III promoter, along with an RNA Pol III terminator. To adapt this plasmid for use in *Microbotryum*, the U6 promoter was replaced with a tRNA/sgRNA chimera sequence (Wei et al., 2015) to drive expression of the sgRNA, for which the target sequence was replaced with 19 base pairs to target the MVLG_05585 gene. The Polymerase II promoter from the constitutively expressed MVLG_02523 gene, defined as the 1 kb region upstream of the start codon, was also cloned in place of the Potef promoter from pMS10 to drive expression of the Cas9 endonuclease in *Microbotryum* (Figure C5-11). Furthermore, to select for transformants, the Hygromycin B resistance cassette driven by the MVLG_05589 promoter from the *Microbotryum Agrobacterium*-mediated transformation plasmid, pMvHyg (Toh et al., 2016), was cloned in-between the sequence for Carboxin resistance present in pMS10 to allow for selection of *Microbotryum* cells on Hygromycin containing media and for quick excision if the plasmid need be reverted to a function Carboxin resistance gene (Figure C5-11).
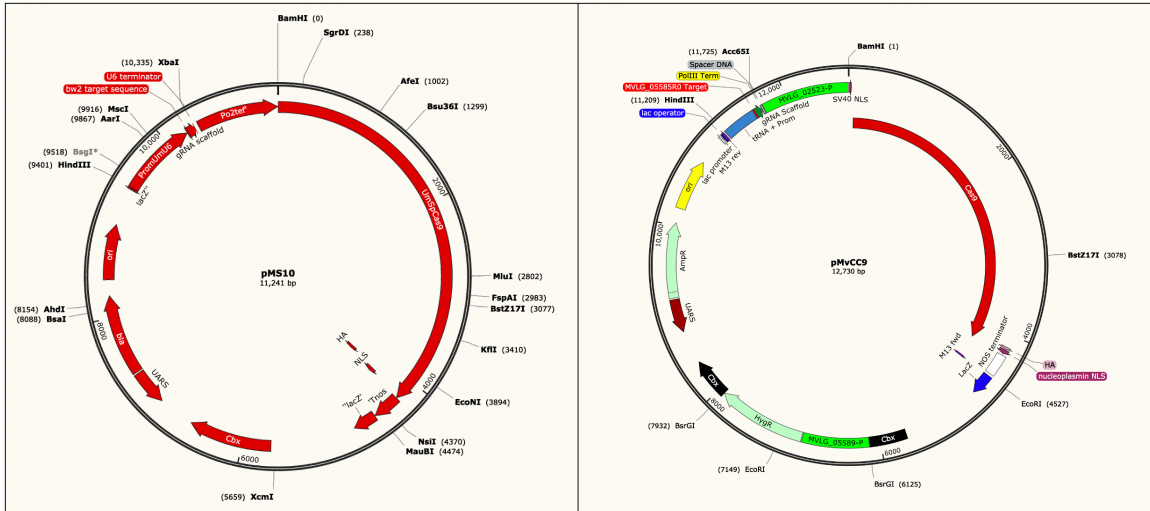
Figure C5-11 Plasmid maps for pMs10 and pMvCC9. The pMS10 plasmid is shown on the left, while the pMvCC9 plasmid which incorporates various components of pMS10 is shown on the right.

These substitutions were made using the restriction endonuclease cut sites available in the pMS10 vector, HindIII and Acc65I for the insertion of the tRNA/gRNA chimera sequence, and Acc65I and BamHI for insertion of the MVLG_02523 promoter sequence, followed by ligation and transformation of the modified plasmid into *E. coli*.

2.3 Electroporation of *Microbotryum lychnidis-dioicae* cells

An assay of various electroporation conditions was used to determine optimal transformation efficiency of *Microbotryum lychnidis-dioicae* cells using a Bulldog Bio© / Nepa Gene ELEP021 Electroporator. 2 mm gap cuvettes were used to electroporate *M. lychnidis-dioicae* cells suspended at an $OD_{600} = 1$ in sorbitol as a buffer. Two variables were assayed for the poring pulse, the voltage (V) and length of pulse in ms. The interval, number of pulses, and polarity for the poring pulse remained constant at 50 ms, 1 pulse, and positive polarity. Conditions were also kept constant for the transfer pulse voltage, length, interval, number of pulses, and polarity at 50 V, 50 ms, 50 ms, 5 pulses, and alternating positive/negative polarity. The ohms and volts, as well as the amps, and joules output for both the poring and transfer pulses, were recorded in each trial. Cells were then removed from the electroporator and immediately plated onto YPD-10% growth agar. Colonies were then counted to determine optimal survival rate with the highest shock values.

After conditions for optimal survivorship were determined, electroporation was repeated in attempt to transform *M. lychnidis-dioicae* cells with pMvCC9. Cells were immediately suspended in YPD-10% growth media and shaken at 27°C for 2 hours before being plated onto YPD-10% containing 150 μg/mL Hygromycin as a selective agar media.

Table C5-2. Optimal electroporation settings assay.

#	Set Parameters									
	Poring Pulse					Transfer Pulse				
	V	Length (ms)	Interval (ms)	No.	Polarity	V	Length (ms)	Interval (ms)	No.	Polarity
1	Control (cells and DNA without electroporation treatment)									
2	2000	1	50	1	+	50	50	50	5	+/-
3	2000	2	50	1	+	50	50	50	5	+/-
4	2000	3	50	1	+	50	50	50	5	+/-
5	2000	4	50	1	+	50	50	50	5	+/-
6	2000	5	50	1	+	50	50	50	5	+/-
7	2500	1	50	1	+	50	50	50	5	+/-
8	2500	2	50	1	+	50	50	50	5	+/-
9	2500	3	50	1	+	50	50	50	5	+/-
10	2500	4	50	1	+	50	50	50	5	+/-
11	2500	5	50	1	+	50	50	50	5	+/-
12	3000	1	50	1	+	50	50	50	5	+/-
13	3000	2	50	1	+	50	50	50	5	+/-
14	3000	4	50	1	+	50	50	50	5	+/-

2.4 Constructing a plasmid for ATMT delivery of CRISPR-Cas9

The CRISPR-Cas9 ATMT plasmid for use in *Agrobacterium*-mediated Transformation (ATMT) of *M. violaceum* species (pMvHyg_CRISPR) was created using scaffold from pMS8 (provided by Dr. Regine Kahmann) as a scaffold. The pMS8 plasmid orginally introduced the Cas9 gene from *S. pyogenes* to the *Ustilago maydis* transformation plasmid pNEBUC (Schuster, 2015). The U6 promoter attached to the sgRNA scaffold in pMS8 was replaced with a tRNA/sgRNA chimera sequence (Wei et al., 2017) generated using the TyrGTA tRNA from *M. lychnidis-dioicae*, including its promoter to drive Polymerase III transcription. TyrGTA was chosen due to demonstrated codon bias in *M. lychnidis-dioicae* (Perlin et al., 2015). The replacement of the U6 promoter with the tRNA/sgRNA chimera was performed via Gibson overlap extension

PCR utilizing the NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit to connect the TyrGTA tRNA with the MVLG_05585 target sequence, the gRNA scaffold from pMS8, and a tRNA transcriptional termination signal along with DNA spacer shown in Figure C5-11.
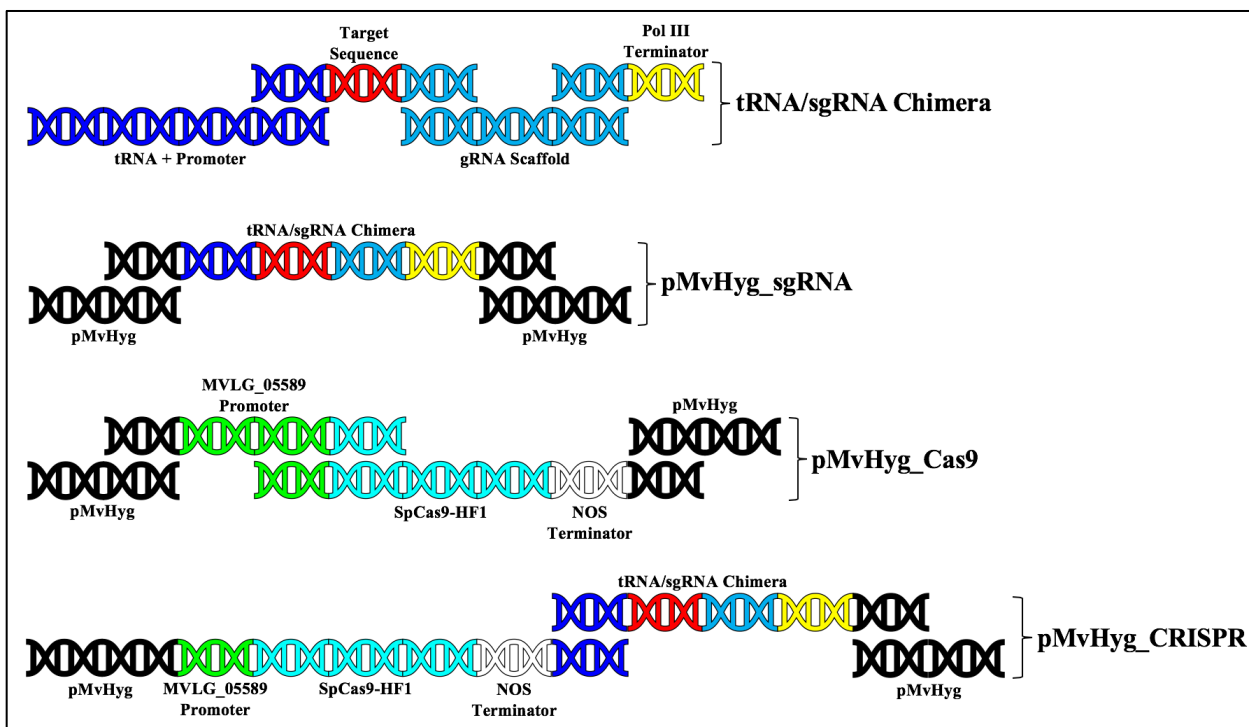
Figure C5-12 Stepwise overview of pMvHyg_CRISPR construction via Gibson Overlap PCR. Overlap PCR using double stranded DNA fragments with 20 bp complementary sequence to adjacent fragments.

The gRNA and the termination signal fragments were ordered as 60 bp oligonucleotides with 20 bp of overlap regions for Gibson overlap PCR. These oligonucleotides were converted into double stranded DNA by adding equimolar concentrations of each oligo and its corresponding complement oligo to a PCR tube and heating in a thermocycler at 94°C for 1 min to denature any secondary structures, followed by 60°C for 10 minutes to allow for annealing of the two complements. The tRNA sequence including the promoter were amplified from *M. lychnidis dioicae* genomic DNA, and gRNA scaffold was amplified from pMS8 via PCR. The primers used to amplify these two fragments also included a 5' 20 bp sequence overlap for Gibson overlap PCR (Figure 1). The 4 fragments were annealed using NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit, followed by another round of PCR using Ex Taq polymerase with a forward primer for the tRNA with a

20 bp complementary sequence to the pMvHyg XbaI restriction site, and a reverse primer for the 3' end of the fragment including a 20 bp complementary sequence for the StuI restriction sequence of pMvHyg (Figure 1). This generated a single fragment with XbaI and StuI cut sites as well as the overhangs necessary for Gibson overlap PCR. A double restriction digest was performed on the pMvHyg vector with XbaI and StuI, and both the vector and PCR product were separated by gel electrophoresis through a 0.8% agarose gel containing ethidium bromide alongside a DNA size standard via gel electrophoresis. The corresponding bands for the digested pMvHyg vector and tRNA/sgRNA fragment were excised from the gel, purified, and annealed using the NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit to create the vector pMvHyg_sgRNA (Figure 1)

A separate ATMT vector containing a high-fidelity mutant of the Cas9 gene (pMvHyg_Cas9) was also generated using Gibson overlap PCR. The coding region for the high fidelity Cas9 variant (also containing an NLS) was amplified using pCas9hf, obtained from Dr. Björn Sandrock. This high-fidelity Cas9 endonuclease was created using three missense substitutions; aa650 Arg->Ala, aa684 Gln->Ala, and aa915 Gln->Ala, and has been demonstrated to create fewer off-target cuts. To drive expression of this Cas9-HF endonuclease, the Po2tef promoter from the vector was replaced with a constitutively expressed promoter from *M. lychnidis-dioicae* gene, MVLG_05585. The MVLG_05589 promoter sequence was amplified from *M. lychnidis-dioicae* genomic DNA using 40 bp primers which contained 20 bp complementary sequence to the elements upstream of the PacI cut site on pMvHyg and Cas9-HF. The Cas9-HF coding region, including the NOS Termination sequence was amplified using similar 40 bp primers with 20 bp complementary sequence for the MVLG_05589 promoter and the elements downstream of the XbaI cut site in pMvHyg. These 2 fragments were then cloned into cut pMvHyg vector using the NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit to generate pMvHyg Cas9 (Figure C5-11).

To create the single transfer cassette containing both the tRNA/sgRNA chimera and the Cas9-HF, the tRNA/sgRNA fragment from pMvHyg_tRNA/sgRNA vector was amplified using 40 bp primers including 20bp overlap regions for the elements upstream and downstream of the XbaI and StuI cut sites, including the sequence to retain the cut sites themselves for future modifications, and the pMvHyg_Cas9 vector was double digested with XbaI and StuI. The fragment was then cloned into the cut vector using the

the NEBuilder® #E2621 HiFi DNA Assembly Master Mix Kit to form the pMvHyg_CRISPR Plasmid (Figure C5-11 and C5-12).
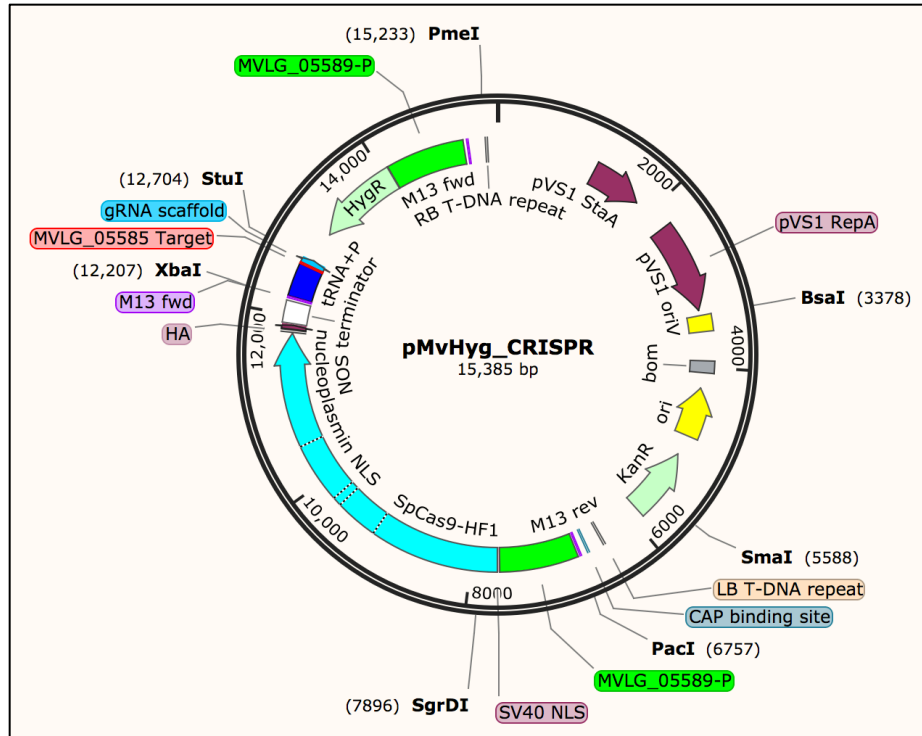
Figure C5-13 Snapgene image of the pMvHyg_CRISPR plasmid. The plasmid map for the ATMT vector containing the Cas9-HF gene driven by the constitutively expressed MVLG_05589 promoter, and the tRNA/gRNA chimera targeting the MVLG_05585 β-carotene synthesis gene.

When used in ATMT, the cassette cloned in-between the T-DNA Left Border and Right Border is incorporated randomly into transcriptionally active regions of the host genome.
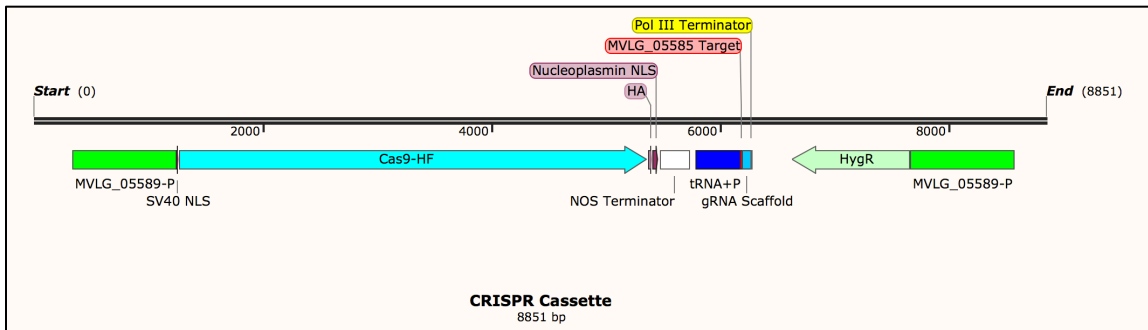
Figure C5-14 Snapgene image for the components of pMvHyg_CRISPR that are transferred by the left and right T-DNA borders. The T-DNA cassette that is excised from the ATMT vector and incorporated randomly into transcriptionally active areas of the host (*Microbotryum lychnidis-dioicae*) genome.

2.5 *Agrobacterium*-mediated transformation of *M. lychnidis-dioicae* cells with pMvHyg-CRISPR

pMvHyg_CRISPR was transformed into EHA105 *Agrobacterium* cells using electroporation (2.5 kV, 400 ohms and 25 μF) following the protocol outlined in Toh et al., (2016). Transformants were verified via streaking colonies onto LB agar-containing 50 μg kanamycin/mL. Putative transformants were then re-streaked onto LB containing 50 μg kanamycin and 100 μg spectinomycin per mL agar to ensure that both the pMvHyg containing the respective species-specific gene and the helper plasmids were in the cells. Surviving EHA105 cells were further confirmed as bearing the desired constructs through colony PCR, before being used to transform p1A1 mating type cells of *M. lychnidis-dioicae*. For these experiments, 10E7 of *Microbotryum* cells and 10E7 *Agrobacterium* cells were used, as measured spectrophotometrically (*Microbotryum*: OD_{600} 1 = 3.4E7 cells/mL; *Agrobacterium*: OD_{600} 1 = 8E8 cells/mL) and mixed in equal volumes. 200 μL of each suspension were spotted onto IM plates containing acetosyringone. Spotted plates were let sit at room temperature, ~25 °C, for 3 days, after which the resulting mass of cells was scraped from the plates, suspended in 600 μL of YPD-10% broth, and 200 μL of each suspension were spread onto YPD-10% containing 150 μg/mL Hygromycin and 100 μg/mL Cefotaxime plates. Each plate was then incubated for 12-15 days to select for transformed *Microbotryum* cells. Colonies appearing were streaked onto fresh YPD-10% containing 150 μg/mL Hygromycin B plates to ensure proper transformation and later verified for successful CRISPR cassette insertions via sequencing.

2.6 Protoplasting *Microbotryum* cells

Microbotryum cells were protoplasted using a protocol developed by Dr. Naoko Fujita. The enzyme solution was prepared using 2% lysing enzymes from *Trichoderma harzianum* and 2 % Driselase in 1M $MgSO_4$. The solution was mixed and stored overnight at 4°C, after which the tube was spun at 11,000 rpm for 10 minutes. The supernatant was filtered into a sterile 50 mL Falcon Tube using filters. To protoplast *M. lychnidis-dioicae* cells, p1A1 and p1a2 strains were grown on YPD with 10% dextrose (YPD-10%) agar plates for 2 days at 27°C. A generous loop of cells was suspended in 5 mL of the filtered enzyme solution in a 50 mL sterile tube and shaken using a platform shaker at medium-low speed overnight. Parafilm was placed around the lid to prevent any leakage. The following morning, 5 mL of 1.2M mannitol was carefully added top wise so as not to mix with the cell suspension. The layered mixture was then carefully spun at 2,000 rpm for 20 minutes. After 20 minutes, a layer of protoplasted cells formed between the cell suspension bottom layer and mannitol top layer. This middle layer was extracted using a P1000 micropipette tip. Of the 20 mL solution, approximately 5 mL of protoplasts were extracted. This layer of protoplasted cells were confirmed using compound microscopy and aliquoted into 100 μL aliquots in 600 μL microcentrifuge tubes. Aliquots were centrifuged at 14,000 rpm for 1 min, the supernatant was removed, and the protoplasted cells were resuspended in 100 μL of STC before storage long-term at -80°C. Protoplasted stored at -80°C were tested for viability on YPD-10% and demonstrated the capacity to regenerate their cell walls and grow even after 2 years of storage.
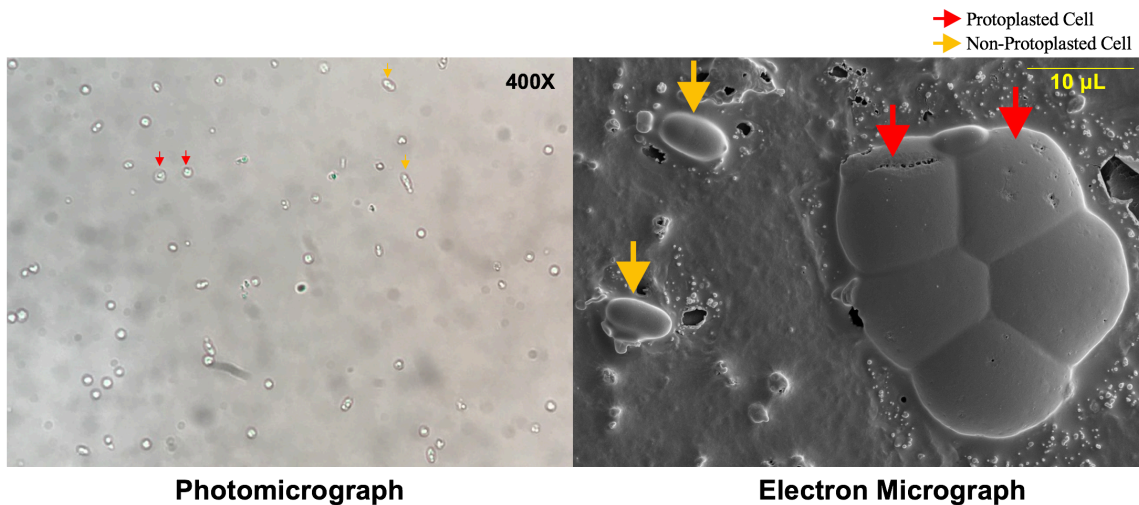
Figure C5-15 Microscopy images of protoplasted *Microbotryum* cells. Photomicrograph of spheroplasted (red arrow) and normal (orange arrow) *Microbotryum lychnidis-dioicae* cells, left, and electron micrograph of spheroplasted and normal *M. lychnidis-dioicae* cells, right.

2.7 PEG Transformation of *Microbotryum* with *in vitro* Cas9 Duplex

In vitro duplexing of S.p. Cas9 Nuclease was accomplished following the protocol for the IDT Alt-R™ kit. gRNAs duplexed with Cas9 were assembled from custom crRNAs annealed to Alt-R™ tracrRNAs by mixing the two at equimolar concentrations using 5 μL of each 100 μM working stock and 5 μL of Nuclease-Free Duplex Buffer provided with the Alt-R™ tracrRNAs, and incubating the mixture at 95°C for 5 minutes before allowing the tubes to cool to room temperature. To increase transformation efficiency, two gRNAs were designed for each knockout, one targeting an NGG PAM site at the 5' end of the gene and the other targeting an NGG PAM site at the 3' end of the gene, allowing for entire gene deletion and insertion of a selectable marker-encoding linear oligonucleotide with 40 bp overlap regions corresponding to the upstream and downstream regions of the excised target. Therefore, to direct Cas9 activity at both ends of the target gene, the Alt-R™ S.p. Cas9 Nuclease 3NLS was diluted 10x using 1 μL of the Cas9 and 9 μL of Cas9 Working Buffer (20 mM HEPES pH 7.5, 150 mM KCl in Nuclease-free water) and 1.5 μL of the enzyme solution was mixed with 1.5 μL of the 5' target gRNA and 1.5 of the μL 3' gRNA target in 22 μL of Cas9 Working Buffer. The duplex solution was incubated at room temperature for 5 minutes to allow for annealing of the Cas9 to gRNAs before use in PEG transformation.

M. lychnidis-dioicae protoplasts were retrieved from -80°C and thawed on ice. 100 μL of STC solution was added to an 8 mL snap cap tube and placed on ice. 7 μg of purified linear repair template containing the MVLG_05589 promoter and coding region for Hygromycin-B resistance in *M. lychnidis-dioicae* (amplified from the pMvHyg vector) was added to the STC solution along with the 26.5 μL of RNP duplex solution. 100 μL of protoplasts were then added to the tubes using wide-bore tips and gently mixed with the pipette to ensure even suspension without destroying the protoplasted cells. 50 μL of 30% PEG solution was added to the mixture and gently swirled before further incubation on ice for 50 min. After incubation on ice, the tube was placed at room temperature and 2 mL of 30% PEG was added to the solution and mixed via gentle repeated inversion. The tube was incubated at room temperature for 20 min. Protoplasted cells become clumped together during this time and were examined using a compound microscope. After verification of clumping, 2 mL of STC was added to the suspension and mixed via inversion. 500 μL of suspension was then top added to the top agar YPD-10% plates. Top agar plates were prepared by pouring 10 mL of YPD-10% media with 300 μg/mL Hygromycin-B, allowing the agar to solidify, and then adding 10 mL of YPD-10% non-drug top agar during the 20 min incubation step above. To increase efficiency, the bottom agar was prepared the morning of transformation. Plates were allowed to dry before storage at 25°C for two weeks. After 11 days, small colonies began to appear on the plates. More colonies appeared between days 12-16, and the original colonies were large enough to pick and streak onto a fresh YPD-10% containing 300 μg/mL Hygromycin-B plate by day 13. The newly streaked plates were incubated at 25°C for 3 more days before colonies with turbid growth, indicating true resistance to the Hybromycin-B, were selected and added to liquid YPD-10% containing 200 μg/mL Ampicillin tubes. These tubes were spun in an orbital shaker at a medium speed for 2 days before DNA extractions were performed and PCR products were sent for sequencing.

2.8 Sequencing to confirm successful knockouts

Primers for amplification of the target gene for PCR confirmation and sequencing were designed to begin 120 bp upstream of the 5' cut site and 120 bp downstream of the 3' cut site. PCR using Ex Taq DNA pol was performed on genomic DNA extracts from potential transformants and separated on an agarose gel to screen for insertion of the Hyg^R cassette. The PCR product for wildtype, untransformed cells is 2,351 bp for the

MVLG_05585 target, and 1,024 bp for the MvSl_01693 target. If the Hyg^R cassette is successfully inserted in place of the excised gene, the band size for both sets of primers should be approximately 2,140 bp. Bands that appeared at the appropriate sizes were purified using the Zymoclean^{TM} Gel DNA Recovery Kit and sent for sequencing using the 5' end forward primers.
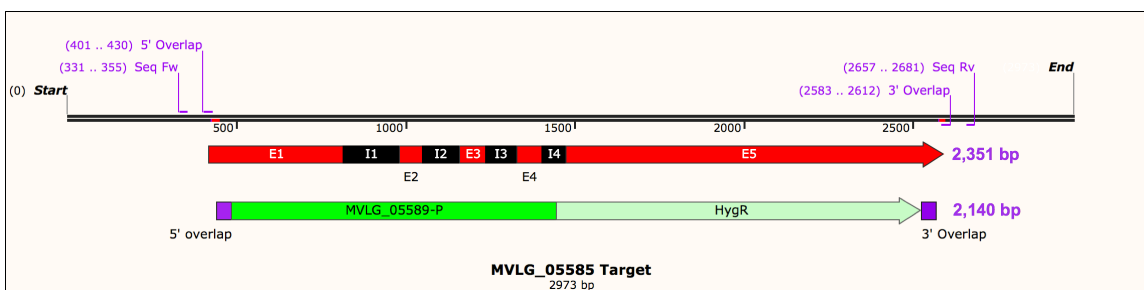
Figure C5-16 Snapgene image depicting the target regions and insertion construct for MVLG_05585 knockouts. MVLG_05585 gene, top, and Hyg^R cassette insert, bottom. Purple text on top of the DNA sequence indicates the location of the sequencing primers (Seq Fw and Seq Rv) 120 bp upstream and downstream of the 5' and 3' cut sites used for confirming knockouts, and the overlap locations for homology-directed repair utilizing the Hyg^R cassette insert. Thin red DNA bars indicate the target sequence, both on the reverse strand of MVLG_05585. Red feature boxes labeled E# indicate exonic regions of the gene while black feature boxes labeled I# indicate intronic regions of the gene. The purple feature boxes on the Hyg^R cassette insert indicate the 40 bp overlap sequences for use as homology-directed repair template. The bold purple text next to each set of features indicates the size of the PCR product when using the Seq Fw and Seq Rv primers.

Results

3.1 Electroporation of Single Plasmid CRISPR Construct

Optimal electroporation conditions for cell viability after the procedure were determined by observing the cultures plated from each of the trials in the electroporation assay. The control (untreated) demonstrated confluent growth, as did treatment conditions with 1-2 ms of 2000 V. Treatment with 3-4 ms of 2000 V reduced the survival rate of cells, resulting in cultures that had >200 colonies per plate, but not confluent

growth. Treatment conditions of 2000 V for 5 ms, as well as 2500 V for 1-5 ms resulted in a significant

decrease in survivability for cells. All trials using 3000 V of electricity resulted in no growth on culture plates.

Table C5-3. Electroporation results

#	kΩ	Poring Pulse A	Poring Pulse J	Transfer Pulse A	Transfer Pulse J	Voltage (V)	Comments
		Measurements				**Results**	
1							Confluent Growth
2	5.308	0.39	1.16	0.04	4.71	2000	Confluent Growth
3	5.129	0.40	2.62	0.05	5.64	2000	Confluent Growth
4	5.750	0.35	3.70	0.05	5.25	2001	Lots of Colonies
5	5.099	0.39	6.06	0.07	6.44	2001	Lots of Colonies *optimal conditions
6	5.722	0.35	6.63	0.06	5.61	2001	Few Colonies
7	4.981	0.53	1.96	0.05	5.45	2495	Few Colonies
8	4.666	0.55	4.77	0.07	6.49	2499	Few Colonies
9	4.668	0.55	8.17	0.08	7.44	2499	Few Colonies
10	4.868	0.83	17.70	0.14	10.22	2490	Few Colonies
11	3.883	0.66	22.40	0.17	11.70	2500	Few Colonies
12	4.303	0.73	3.25	0.07	6.50	2966	No Growth
13	4.083	0.77	8.96	0.10	9.16	2986	No Growth
14	3.993	0.79	29.36	0.01	2.21	2932	No Growth

Results from this preliminary survivorship assay indicates that the settings of 2000 V for 4 ms is likely to

result in the highest transformation efficiency for this system.

Using these settings, our we were able to verify that dyed molecules can be transformed into *M. lychnidis-dioicae* cells; however, transformation attempts with the pMvCC9 plasmid resulted in no colonies on selective media, indicating that while transfer of material into the cells is possible with electroporation problems with either the plasmid construct or its entry into the nuclease prevent the CRISPR components from being expressed via the single plasmid method.

3.2 *Agrobacterium*-mediated Transformation of *M. lychnidis-dioicae* with pMvHyg_CRISPR

Agrobacterium-medicated transformation of *Microbotryum* cells with the pMvHyg_CRISPR plasmid resulted in 5 colonies that were verified to contain the components between the left and right transfer borders

via PCR. Despite successful insertion of the both the tRNA/gRNA chimera and Cas9 coding regions, none of the 5 cell lines demonstrated any modification to the target region of MVLG_05585. Furthermore, cell color in each line remained the wild-type pink. However, despite an inability to knockout the MVLG_05585 gene, qrtPCR analysis of the transformed *Microbotryum* cells indicates that the pMvHyg_CRISPR plasmid does successfully induce the production of the Cas9 (Figure C5-17).
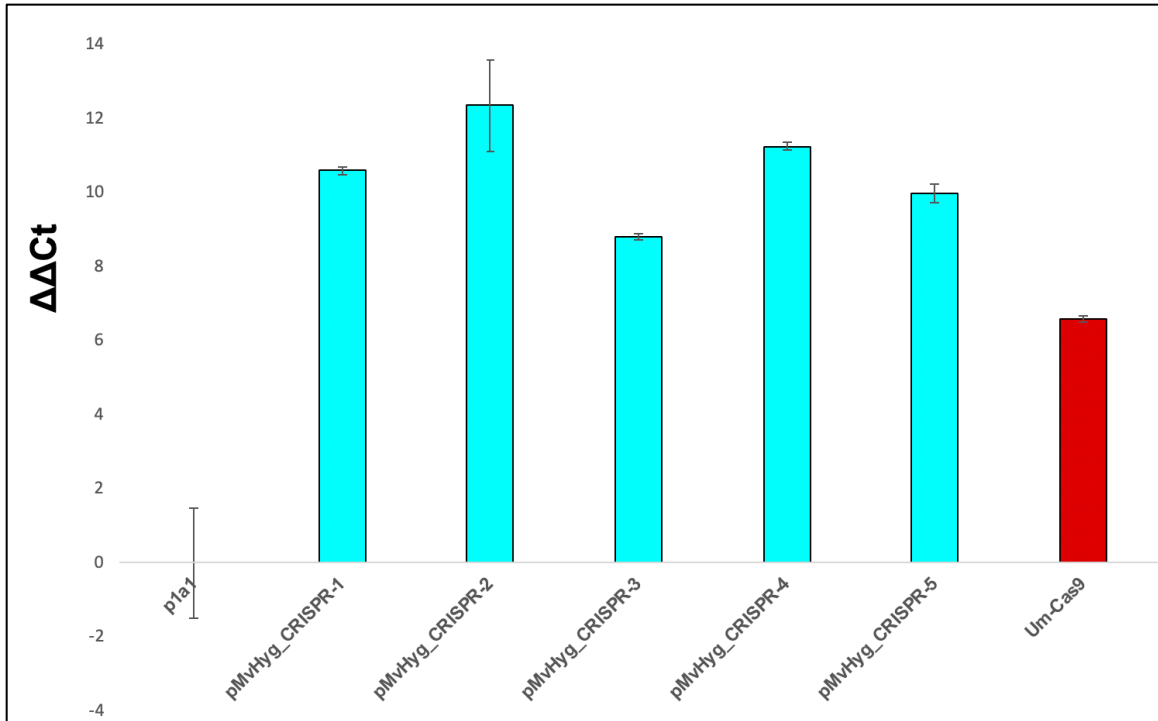
Figure C5-17 qrt PCR expression of Cas9 in *Microbotryum*. Gene expression was standardized against β-tubulin expression in wild type p1A1 cells, first lane. Lanes 2-6 represent different transformants collected from selection media. Lane 7 represents a positive control for Cas9 expression using *Ustilago maydis* transformed with pMS10.

3.3 PEG Transformation of *M. lychnidis-dioicae* with Cas9 duplex

When protoplasted cells were treated with PEG, the Cas9 duplexed with the crRNA/tracrRNA components and repair template, colonies were able to successfully recover on YPD-10% media containing Hygromycin B, indicating successful insertion of the repair template into these cells. Furthermore, PCR amplification of the target region demonstrated appropriate size reduction that would be expected to occur with the insertion

of the 1,801 bp Hygromycin B resistance cassette in-between the two Cas9 cut sites, while excising the 2,212 bp MVLG_05585 gene (Figure C5-18).
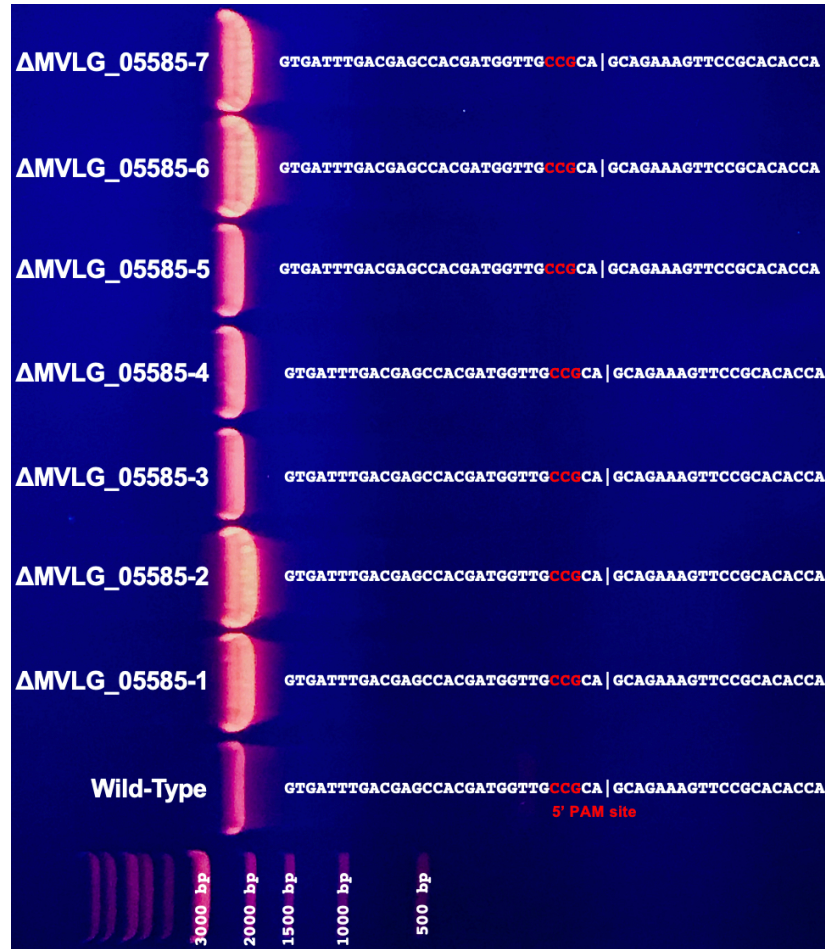
Figure C5-18 PCR verification of potential Cas9 transformants. PCR was performed using primers that amplify the MVLG_05585 coding region of *M. lychnidis-dioicae* starting 50 base pairs upstream and downstream of the target sites, resulting in a wild-type band of approximately 2,300 bp. Successful disruptions with homologous directed repair should appear as approximately 1,900 bp in length. 1 kb size standard, purchased from NEB, is shown and labeled on the bottom for size comparison, wild-type cells are shown in lane one, while putative transformants are shown in lanes 2-8. Sequencing results are shown to the right of each strain with the Cas9 PAM site of the target is shown in red.

However, despite growth on resistance media and seemingly successful PCR verification of MVLG_05585 disruption, follow-up sequencing demonstrates that the 5' and 3' target regions for the MVLG-05585 gene are still intact. This indicates that the Hygromycin B resistance cassette is being inserted into the *Microbotryum* genome but not in the Cas9 target region.

Discussion

Implementation of CRISPR-Cas9 in *Microbotryum* would provide a reliable means for transformation and generating target specific gene disruptions. Doing so would open the door for future analyses of the many novel genes identified by the rapidly expanding library of genomes available for this species complex, genes that are predicted to play a variety of roles in the fungal life cycle and pathogenicity. Additionally, the unique restriction sites on either side of the Cas9 coding gene in both the pMvCC9 and pMvHyg_CRISPR plasmids allow for the use of this systems with other variants of the Cas9 endonuclease, broadening the scope of application. For example, Cas9 nickase mutants can be used in conjunction with two gRNAs to excise regions of the genome with appropriate overhangs for insertions, and dCas9 double mutants can be used to downregulate expression of a gene by interfering with transcription, a process coined as CRISPRi. Furthermore, as new modifications to the Cas9 that increase the fidelity of the nuclease activity to reduce off-target cuts being are discovered, these new Cas9 mutants can be easily swapped into pMvCC9-ATMT vector for immediate use in this system.

While we were able to demonstrate the transformation potential of electroporation to introduce material into *Microbotryum* cells, the inability of cells to grow on selective media indicate either a problem with getting material into the nucleus or issues with maintenance of the pMvCC9 plasmid. One issue may be that the autonomous replication sequence (ARS) used in pMvCC9, which is the ARS used for *Ustilago*, may not be recognized by *Microbotryum*, and therefore successful transformants may be unable to propagate the plasmid in their progeny. One approach to addressing this potential problem would be to identify an endogenous ARS from *Microbotryum* and swapping it in place of the *Ustilago* ARS in pMvCC9. Currently, there are no known ARS sequences for the system; however, one might generate a library of *M. lychnidis-dioicae* genomic fragments using the Hygromycin resistance cassette as selection to randomly clone in digested fragments of genomic DNA from *Microbotryum* cells and generate a plasmid containing an ARS.

These plasmids from the library could then be electroporated into *Microbotryum* to select for regions of the genomic sequence that contain an endogenous ARS to use for the pMvCC9 plasmid.

Similarly, although we were able to successfully insert the tRNA/gRNA chimera and Cas9 coding regions from pMvHyg_CRISPR into the *Microbotryum* genome using ATMT, we were not able to observe changes to the MVLG_05585 target region. We were however able to detect expression of the Cas9 enodnuclease via qrtPCR (Figure C5-17). Expression of the Cas9 without proper target editing suggests that the tRNA/gRNA chimera is not properly separating during post transcriptional modification. If the tRNA and gRNA components are not cleaved properly, the tRNA would likely interfere with the binding of Cas9 to the target site and initiation of cutting, even if bound properly to the Cas9 endonuclease. While the current pMvHyg_CRISPR plasmid is not sufficient for targeted gene knockouts, establishment of a Cas9-producing strain of *Microbotryum* is a step forward and open the door for other transformation possibilities. One alternative would be to remove the tRNA component of the plasmid and replace the U6 promoter that has been demonstrated to drive sgRNA production in *Ustilago maydis*. Another approach may be to use a duel plasmid system, where the Cas9-producing *Microbotryum* strain is transformed with another either another vector containing the components for production of a sgRNA or simply the sgRNA or crRNA/tracrRNA itself.

Finally, while attempts to assemble the Cas9 duplex with crRNA/tracrRNA components *in vitro* before transformation into *Microbotryum* cells via PEG transformation were the closest thus far at achieving target specific knockouts in *M. lychnidis-dioicae*, discrepancies between PCR verification and sequencing cast doubts on whether or not modifications are being made appropriately via Cas9 excision of the target gene. Further work using this system should examine the potential for both modified cells and non-modified *Microbotryum* cells to exist in the same colony. This could be done through t-streaking colonies to assure that they represent a homogenous population of cells, allowing for a more robust identification of potential cells with the desired disruptions.

Given the breadth of its applications, CRISPR represents the addition of a multi-tool to the *Microbotryum* toolbox. The versatility of pMvCC9 provides the framework for future applications in *Microbotryum,* including site specific insertions, translational interference and DNA tagging, gene insertions, whole gene excision, and multiple gene knockouts (Hsu et al., 2014). This system will undoubtedly play a vital part in

future studies seeking to identify the role of many unique genes found amongst closely related species of the *Microbotryum* genus, shedding light on the molecular mechanisms behind speciation events resulting from co-evolution between the plant hosts and these fungal pathogens.

CHAPTER 6

CONCLUSIONS

The work presented in this dissertation was the first to identify and compare secretomes of various *Microbotryum* species at the molecular level, to understand the coevolutionary changes that have led to host-specificity within the *Microbotryum* genus. This work sets the groundwork for future research into the secreted proteins of the *Microbotryum* genus by identifying the secretomes portfolios of *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum var. paradoxa*. These lists of secreted proteins act as a starting point for future molecular genetics studies to understand their functions in the hosts. This dissertation also provides insight into the importance of different sets of secreted proteins, including those that are heavily conserved across the *Microbotryum* genus and those that are unique to each species.

While species-specific secreted proteins can play a role in overall pathogenicity, the more likely cause for adaptation and coevolution over time are small amino acid sequence changes to conserved *Microbotryum* effectors. Speciation events are therefore likely the result of diversifying selection in different populations of plants/host in which rapid changes to the proteins secreted to block plant defense responses are driven by subsequent rapid changes in mechanisms that detect the presence of the pathogen in the host. The secretomes utilized by different *Microbotryum* species are largely shared with few gene gains or losses. Instead, small stepwise changes in the amino acid sequences of core secreted proteins likely play a much larger role in host specificity. This is supported by the rapid evolution observed in the core SPs under positive selection compared to the non-secreted proteins observed in the genomes, and the comparatively small number of species-specific genes across the *Microbotryum* genus. In addition to changes that might alter the function of a secreted protein, this work also identified a mechanism for mobilization or de-mobilization of effectors by changes to the leader sequence of proteins, which could quickly change the arsenal of secreted proteins utilized by each species.

Molecular analysis of one such core secreted protein, MVLG_02245, provides further evidence to support that validity of the predictive measures used in our secretomes analyses. Through Yeast-Secretion-

Trap methods and Yeast-Two-Hybrid screening, we both demonstrated that the core protein was secreted and likely an effector by demonstrating that the leader sequence codes for secretion of other genes in a yeast system and targets a tubulin α-1 chain protein within the *Silene latifolia* hosts. While further analyses will be needed to visualize the localization of this secreted protein *in vivo*, interaction with the tubulin α-1 chain ortholog in *S. latifolia* demonstrates a mechanism in which the secreted protein can weaken the hosts in order for the fungal cells to penetrate the cell wall during pathogenesis. Thus, core secreted proteins are likely highly preserved due to their mechanistic role of infection in the host, i.e., contributing to the physical entrance and manipulation of the host.

Species-specific proteins on the other hand seem play a role in regulation and depression of specific host recognition factors and defense response. This research demonstrates that while not all species-specific genes confer an infectious advantage in the *Microbotryum* species that express them, the MvSl_01693 did significantly improve the pathogenicity of *M. silenes-dioicae* on its natural host, *S. dioica*. We hypothesize that the MvSd_01693 transgenic strain of *M. silenes-dioicae* may be more successful at infecting *S. dioica* due to the lack of exposure, and subsequent selective adaptation, to the MvSl_01693 effector. An increase in pathogenicity of its own host when expressing the species-specific gene of its sister species indicates that these species-specific genes may be artifacts of co-evolution post speciation events and may be useful for evaluating the evolutionary history and relatedness of members in the *Microbotryum* genus.

Finally, in effort to further lay the groundwork for future research analyzing the hundreds of secreted proteins identified by this research, this dissertation made progress in establishing a gene-specific knockout system using CRISPR Cas9. While we were unable to make changes to the target region used in our pilot study, this dissertation did make progress towards a reliable site-specific gene modification system by generating a strain of *Microbotryum* that expresses Cas9. This strain can be combined with a variety of other transformation practices to introduce guide RNA templates to chaperone the Cas9 to a desired target. Furthermore, slight modifications to the amino acid sequence of the plasmid vector used to introduce Cas9 to *M. lychnidis-dioicae* can be made to broaden the molecular genetic toolbox available for the *Microbotryum* genus by altering the function of Cas9 to act as a Nickase or for use in CRISPRi methods for gene knockdown studies.

REFERENCES

Abascal F, Zardoya R, Telford M. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*, 38: 7-13. https://doi.org/10.1093/nar/gkq291

Abrash E, Anleu Gil MX, Matos JL, Bergmann DC. (2018). Conservation and divergence of YODA MAPKKK function in regulation of grass epidermal patterning. *Development* 145(14). https://doi.org/10.1242/dev.165860

Akhmetov A, Laurent JM, Gollihar j, Gardner EC, Garge RK, Ellington AD, Kachroo AH, Marcotte EM. (2018). Single-step precision genome editing in yeast using CRISPR-Cas9. *Bio Protoc* 8(6): e2765. https://doi.org/10.21769/BioProtoc.2765

Albain J, Durand EM, Yang S, Zhou Y, Zon LI. (2015). A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Dev Cell* 32(6): 756-64. https://doi.org/10.1016/j.devcel.2015.01.032

Albersheim R, Anderson AJ. (1971). Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. *P Natl Acad Sci USA*. 68(8): 1815-1819 https://doi.org/10.1073/pnas.68.8.1815

Altenbuchner J. (2016). Editing of the *Bacillus subtilis* Genome by CRISPR-Cas9. *Appl Environ Microb* https://doi.org/10.1128/AEM.01453-16

Anderson JP, et al. (2010). Plants versus pathogens: an evolutionary arms race. *Funct Plant Biol*. 37(6): 499-512 https://doi.org/10.1071/FP09304

Andersson DI, Jerlström-Hultqvist J, Näsvall J. (2015). Evolution of new functions de novo and from preexisting genes. *Cold Spring Harbor Perspectives in Biology*. 7(6): a0-17996 https://doi.org/10.1101/cshperspect.1017996

Aguileta G, et al. (2010). Finding candidate genes under positive selection in non-model species: examples of genes involved in host specialization in pathogens. *Mol Ecol*. https://doi.org/10.1111/j.1365-294X.2009.04454.x

Auld SKJR, Tinkler SK, Tinsley MC. (2016). Sex as a strategy against rapidly evolving parasites. *Proc. R. Soc.* B 283: 20162226. http://dx.doi.org/10.1098/rspb.2016.2226

Badouin H, et al. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol Ecol*. 26(7): 2041–2062 doi: 10.1111/mec.13976.

Baker HG. (1947). Infection of species of *Melandrium* by *Ustilago violacea* (Pers.) Fuckel and the transmission of the resultant disease. *Ann. Bot. (Lond.),* 11: 333–348

Bebber PB, Gurr SJ. (2015). Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet Biol.* 74: 62-64 doi: 10.1016/j.fgb.2014.10.012

Beckerson WC, de la Vega RCR, Hartmann FE, Duhamel M, Giraud T, Perlin MH. (2019). Cause and Effectors: Whole genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi. *mBio*. mBio 10:e02391-19 https://doi.org/10.1128/mBio.02391-19

Begerow D, Nilsson H, Unterseher M, Maier W. (2010). Current state and perspectives of fungal DNA barcoding and rapid identification procedures*. Appl Microbiol biot*. 87(1): 99-108 doi: 10.1007/s00253-010- 2585-4

Bergstrom CT, Dugatkin LA. (2011). *Evolution*. NY & London: W. W. Norton & Company.

Bier E, Harrison MM, O'Connor-Giles KM, Wildonger J. (2018). Advances in engineering the fly genome with the CRISPR-Cas system. *Genetics* 208(1): 1-18. https://doi.org/10.1534/genetics.117.1113

Bierle CH, Anderholm KM, Wang JB, McVoy MA, Schleiss MR. (2016). Targeted mutagenesis of guinea pig cytomegalovirus using CRISPR/Cas9-mediated gene editing. *J Virol* 90(15): 6989-6998. https://doi.org/10.1128/JVI.oo139-16

Blümke A, Falter C, Herrfurth C, Sode B, Bode R, Schäfer W, Feussner I, Voigt CA. (2014). Secreted fungal effector lipase releases free fatty acids to inhibit innate immunity-related callose formation during wheat head infection. *Plant Physiology*. 165(1): 346-358 https://doi.org/10.1104/pp.114.236737

Brameier M, Krings A, MacCallum RM. (2007). NucPred – predicting nuclear localization of proteins. *Bioinformatics*. 23(9): 1159-60 DOI: 10.1093/bioinformatics/btm066

Branco S, et al. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature Communications*. 9: 2000 doi: 10.1038/s41467-018-04380-9

Brown NA, Hammond-Kosack KE. Secreted biomolecules in fungal plant pathogenesis. In: Gupta VK, Mach RL, Sreenivasaprasad S, editors. Fungal Biomolecules: Sources, Applications and Recent Developments. Oxford, UK: John Wiley & Sons, Ltd; 2015. pp. 263–310.

Casa-Esperón Edl. (2012). Horizontal transfer and the evolution of host-pathogen interactions. *International Journal of Evolutionary Biology*. 2012: 1-9 https://doi.org/10.1155/2012/679045

Char SN, Neelakandan AK, Nahampun H, Frame B, Main M, Spalding MH, Becraft PW, Meyers BC, Walbot V, Wang K, Yang B. (2017). An Agrobacterium-Delivered CRISPR/Cas9 System for High-Frequency Targeted Mutagenesis in Maize. *Plant Biotechnol J* 15(2): 257-268 https://doi.org/10.1111/pbi.12611

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* 155(7): 1479-91 https://doi.org/10.1016/j.cell.2013.12.001

Chen B-X, Wei T, Zhi-Wei Y, Yun F, Kang L-Z, Tang H-B, Guo L-Q, Lin J-F. (2018). Efficient CRISPR -Cas9 gene disruption system in edible-medicinal mushroom *Cordyceps militaris*. *Front Microbiol* 9: 1157. https://doi.org/10.3389/fmicb.2018.01157

Chen W, Zhang Y, Zhang Y, Pi Y, Gu T, Song L, Wang Y, Ji Q. (2018). CRISPR/Cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine deaminase-mediated base editing in *Pseudomonas* species. *iScience* 6: 222-231 https://doi.org/10.1016/j.isci.2018.07.024

Chew WL, Tabebordbar M, Cheng JKW, Mali P, Wu EY, Ng AHM, Zhu K, Wagers AJ, and Church GM. (2016). A Multifunctional AAV-CRISPR-Cas9 and Its Host Response. *Nat Met* 13(10): 868-74 https://doi.org/10.1038/nmeth.3993

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339(6121): 819-23 https://doi.org/10.1126/science.1231143

Danecek P, et al. (2011). The variant call format and VCFTOOLS. *Bioinformatics*, 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davies WJ, Kudoyarova G, Hartung W. (2005). Long-distance ABA signaling and its relation to other signaling pathways in the detection of soil drying and the mediation of the plant's response to drought. *Journal of Plant Growth and Reguationl*, 24: 285-295 DOI: 10.1007/s00344-005-0103-1

de Visser JAGM, Elena SF. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews: Genetics*. 8: 139-149 https://doi.org/10.1038/nrg1985

Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. (2012). The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. 13(4): 414-430 doi: 10.1111/j.1364-3703.2011. 00783.x

Delaunois B, Jeandet P, Clément C, Baillieul F, Dory S, Cordelier S. (2014). Uncovering plant-pathogen crosstalk through apoplastic proteomic studies. *Frontiers in Plant Science*. 5(249): 1-18 https://doi.org/10.3389/fpls.2014.00249

Demirci Y, Zhang B, Unver T. (2017). CRISPR/Cas9: An RNA-guided highly precise synthetic tool for plant genome editing. *J Cell Physiol* 233(3) https://doi.org/10.1002/jcp.25970

Dickinson DJ, Goldstein B. (2016). CRISPR-Based Methods for *Caenorhabditis elegans* genome engineering. *Genetics* 202(3) https://doi.org/10.1534/genetics.115.182162

Dita M, Barquero M, Heck D, Mizubuti ESG, Staver CP. (2018). *Fusarium* Wilt of Banana: Current Knowledge on Epidemiology and Research Needs Toward Sustainable Disease Management. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2018.01468

Doll NM, Gilles LM, Gérentes MF, Richard C, Just J, Fierlej Y, Borrelli VMG, Gendrot G, Ingram FC, Rogowsky PM, Widiez T. (2019). Single and multiple gene knockouts by CRISPR-Cas9 in maize. *Plant Cell Rep* 38(4):487-501 https://doi.org/10.1007/s00299-019-02378-1

Dongen S, Graph clustering by flow simulation. PhD thesis, University of Utrecht, May 2000. Available at https://micans.org/mcl/

Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. (2005). Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics*. 21: 2101–2103 DOI: 10.1093/bioinformatics/bti259

Droit A, Poirier GG, Hunter JM. (2005). Experimental and bioinformatic approaches for interrogating

protein-protein interactions to determine protein function. *Journal of Molecular Endocrinology*. 34(2) https://doi.org/101677/jme.1.01693

Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, Van de Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev IV, Szabo LJ, Martin F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci*. 108: 9166–71 doi: 10.1073/pnas.1019315108

Dwivedi S, Goldman I, Ortiz R. 2019. Pursuing the potential of heirloom cultivars to improve adaptation, nutritional, and culinary features of food crops. *MDPI Agronomy*. 9(8): 441 https://doi.org/10.3390/agronomy9080441

Eisenhaber F. (2013). Prediction of Protein Function Two Basic Concepts and One Practical Recipe. Austin TX: Landes Bioscience; 2000-2013

Edgar R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32 1792-1797 DOI: 10.1093/nar/gkh340

Ekseth OK, Kuiper M, Mironov V. (2014). orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics*. 30(5): 734-735 doi: 10.1093/bioinformatics/btt582

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47(D1): D427-D432 doi: 10.1093/nar/gky995

Elmore JR, Yokooji Y, Sato T, Olson S, Glover III CVC, Graveley BR, Atomi H, Terns RM, Terns MP. (2013). Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biol* 10(5): 828-840 https://doi.org/10.4161/rna.24084

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 300(4): 1005-1016

DOI: 10.1006/jmbi.2000.3903

Erbs G, Newman MA. (2012). The role of lipopolysaccharide and peptidoglycan, two glycosylated

bacterial microbe-associated molecular patterns (MAMPs), in plant innate immunity. *Mol. Plant*

Pathol. 13: 95-104 https://doi.org/10.1111/j.1364-3703.2011.00730.x

Erickson PA, Ellis NA, Miller CT. (2016). Microinjection for transgenesis and genome editing in

Threespine Sticklebacks. *JOVE-J Vis Exp* (111) https://doi.org/10.3791/54055

Eun K, Park MG, Jeong YW, Jeong YI, Hyun S-H, Hwang WS, Kim S-H, Kim H. (2019).

Establishment of TP53-knockout canine cells using optimized CRISPR/Cas9 vector system for

canine cancer research. *BMC Biotech* 19: 1 https://doi.org/10.1186/s12896-018-0491-5

Fan D, Liu R, Li C, Jiao B, Li S, Hou Y, Luo K. (2015). Efficient CRISPR/Cas9-mediated targeted

mutagenesis in *Populus* in the first generation. *Sci Rep-UK* 5:12217

https://doi.org/10.1038/srep12217

Fan Y, Lin X. (2018). Multiple application of transient CRISPR-Cas9 coupled with electroporation

(TRACE) system in the *Cryptococcus neoformans* species complex. *Genetics* 208(4): 1357-1372.

doi: 10.1534/genetics.117.300656

Fan Z, Li W, Lee SR, Meng Q, Shit B, Bunch TD, White KL, Kong I-K, Wang Z. (2014).

Efficient gene targeting in golden Syrian hamsters by the CRISPR/Cas9 system. *PLOS ONE*.

9(10): e109755 https://doi.org/10.1371/journal.pone.0109755

Fei J-F, Lou W P-K, Knapp D, Murawala P, Gerber T, Taniguchi Y, Nowoshilow S, Khattak S,

Tanaka EM. (2018). Application and optimization of CRISPR-Cas9-mediated genome engineering

in axolotl (*Ambystoma mexicanum*). *Nat Protoc* 13:2908-2943.

https://doi.org/10.1038/s41596-018-0071-0

Fischer GW and Holton CS. (1957). *Biology and control of the smut fungi*. New York, N.Y: The Ronald

Press Company

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. (2012). Emerging

fungal threats to animal, plants, and ecosystem health. *Nature*. 484(7393): 186

https://doi.org/10.1038/nature10947

Fiza S, Ahmad S, Noor MA, Wang X, Younas A, Riaz A, Riaz A, Ali F. (2019). Applications of the

CRISPR/Cas9 system for rice grain quality improvement: perspectives and opportunities. *Int J Mol Sci* 20(4) https://doi.org/10.3390/ijms20040888

Foster AJ, Martin-Urdiroz MM, Yan X, Wright S, Soanes DM, Talbot NJ. (2018). CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counter selection in rice blast fungus. *BioRxiv*. https://doi.org/10.1101/349134

Froger A, and Hall JE. (2007). Transformation of Plasmid DNA into E. Coli Using the Heat Shock Method. *JOVE-J Vis Exp* (6):253 https://doi.org/10.3791/253

Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, Balesdent MH, Rouxel T. (2009). Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Mol Plant Microbe Interact*. 22(8): 932-41. https://doi.org/10.1094/MPMI-22-8-0932

Gacura MD, Sprockett DD, Heidenreich B, Blackwood CB. (2016). Comparison of pectin-degrading fungal communities in temperate forests using glycosyl hydrolase family 28 pectinase primers targeting Ascomycete fungi. *J Microbiol Methods*. Apr;123:108-13. https://doi.org/10.1016/j.mimet.2016.02.013

Garber ED and Day AW. (1985). Genetic mapping of a phytopathogenic basidiomycete Ustilago violacea. *Bot Gaz*. 146(4): 449-459

Gasiunas G, Barrangou R, Horvath P, Siksnys V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive imunity in bacteria. *PNAS* 109(39)15539-15540. www.pnas.org/cgi/doi/10.1073/pnas.1208507109

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. (2017). Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. *Nature*. 551(7681): 464-471 doi: 10.1038/nature24644

Ghareeb H, et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*. 330(6010): 1546-1548 doi: 10.1126/science.1195330

Gibson AK, Refrégier G, Hood ME, Giraud T. (2014). Performance of a hybrid fungal pathogen on pure-species and hybrid host plants. *Int J Plant Sci*. 175(6): 724-730 DOI: 10.1086/676621

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, and Smith HO. (2009). Enzymatic

Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* 6(5):343-5

https://doi.org/10.1038/nmeth.1318

Gratz SJ, Dustin Rubinstein C, Harrison MM, Wildonger J, O'Connor-Giles KM. (2015). CRISPR-
Cas9 genome editing in *Drosophila*. *Curr Protoc Mol Biol*

https://doi.org/10.1002/0471142727.mb3102s111

Grigg ME, Suzuki Y. (2003). Sexual recombination and clonal virulence in *Toxoplasma*. *Microbes Infect.*
5(7): 685-90 https://doi.10.1016/s1286-4579(03)00088-1

Gruzmán-Zapata D, Sandoval-Vargas JM, Macedo-Osorio KS, Salgado-Manjarrez E, Castrejón-Flores JL,
Oliver-Salvador MdC, Durán-Figueroa NV, Nogué F, Badillo-Corona JA. (2019). Efficient editing
of the nuclear APT reporter gene in *Chlamydomonas reinhardtii* via expression of a CRISPR-Cas9
module. *Int J Mol Sci* 20(5):1247 https://doi.org/10.3390/ijms20051247

Hartmann FE, Croll D. (2018). Distinct trajectories of massive recent gene gains and losses in populations
of a microbial eukaryotic pathogen. *Mol Biol Evol*. 34(11): 2808-2822
doi: 10.1093/molbev/msx208

Hartmann FE, Rodríguez de la Vega RC, Carpentier F, Gladieux P, Cornille A, Hood ME, Giraud T.
(2019). Understanding Adaptation, Coevolution, Host Specialization, and Mating System in
Castrating Anther-Smut Fungi by Combining Population and Comparative Genomics. *Annu Rev
Phytopathol.* 57: 431-457 doi: 10.1146/annurev-phyto-082718-095947

Harel I, Valenzano DR, Brunet A. (2016). Efficient genome engineering approaches for the short-lived
African turquoise killifish. *Nat Protoc* 11:2010-2028 https://doi.org/10.1038/nprot.2016.103

Hartmann FE, Rodríguez de la Vega RC, Brandenbrug J-T, Carpentier F, Giraud T. (2018). Gene presence-
absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeorgraphic
structure. *Genome Biol Evol.* 10(5): 1298-1314 doi: 10.1093/gbe/evy089

Hashimoto T. (2015). Microtubules in Plants. *The Arabidopsis Book*. 13: e0179

https://doi.org/10.1199/tab.0179

Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Møller N, Aarestrup FM. Rapid
whole-genome sequencing for detection and characterization of microoganisms directly from
clinical samples. *Journal of Clinical Microbiology*. 52(1): 139-146

https://doi.org/10.1128/JCM.02452-13

Heitman J, Carter DA, Dyer PS, Soll DR. (2014). Sexual reproduction of human fungal pathogens. *Cold Spring Harbor Perspectives in Medicine*. 4(8): a019281
https://doi.org/10.1101/cshperspect.a019281

Hindriksen S, Bramer AJ, Truong MA, Vromans MJM, Post JB, Verlaan-Klink I, Snippert HJ, Lens SMA, Hadders MA. (2017). Baculoviral delivery of CRISPR/Cas9 facilitates efficient genome editing in human cells. *PLoS ONE* 12(6):e0179514
doi: 10.1371/journal.pone.0179514

Hirose M, Hasegawa A, Mochida K, Matoba S, Hatanaka Y, Inoue K, Goto T, Kaneda H, Yamada I, Furuse T, Abe K, Uenoyama Y, Tsukamure H, Wakana S, Honda A, Ogura A. (2017). CRISPR/Cas9-mediated genome editing in wild-derived mice: generation of tamed wild-derived strains by mutation of the a (*nonagouti*) gene. *Sci Rep-UK* 7:42476
https://doi.org/10.1038/srep42476

Hiruta C, Kakui D, Tollefsen KE, Iguchi T. (2018). Targeted gene disruption by use of CRISPR/Cas9 ribonucleoprotein complexes in the water flea *Daphnia pulex*. *Genes Cells* 23(6): 494-502.
https://doi.org/10.1111/gtc.12589

Hofmann NR. (2017). An emerging paradigm? RxLR cleavage before effector section. *Plant Cell*. 29: 1177-1178 doi: 10.1105/tpc.17.00454

Hood ME, Katawczik M, Giraud T. (2005). Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics*. 170(3): 1081-1089
DOI: 10.1534/genetics.105.042564

Hood ME, et al. (2010). Distribution of the anther-smut pathogen *Microbotryum* on species of the Caryophyllaceae. *New Phytol*. 187: 217-229 doi: 10.1111/j.1469-8137.2010.03268.x

Hopes A, Nekrasov V, Kamoun S, Mock T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant methods* 12:49.
https://doi.org/10.1186/s13007-0160148-0

Horns F, Petit E, Yockteng R, Hood ME. (2012). Patterns of repeat-induced point mutation in transposable

elements of basidiomycete fungi. *Genome Biol Evol*. 4(3): 240-7

https://doi.org/10.1093/gbe/evs005

Horvath P, and Barrangou R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* 327(5962):167-70 https://doi.org/10.1126/science.1179555

Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556(7699):57-63 https://doi.org/10.1038/nature26155

Huang Y, Niu B, Gao Y, Fu L, Li W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 26(5): 680-2 doi: 10.1093/bioinformatics/btq003

Hulsen T, de Vlieg J, Alkema W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics. 9: 488
doi: 10.1186/1471-2164-9-488

Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh JR. (2013). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS ONE* 9;8(7) e68708 https://doi.org/10.1371/journal.pone.0068708

Hsu PD, Lander ES, Zhang F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. 157(6):1262-78 *Cell* https://doi.org/10.1016/j.cell.2014.05.010

Ikmi A, McKinney SA, Delventhal KM, Gibson MC. (2014). TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan *Nematostella vectensis*. *Nat Commun* 5: 5486 https://doi.org/10.1038/ncomms6486

Ishino Y, Shinagawa H, Makino K, Amemura M, and Nakatura A. (1987). Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isoenzyme Conversion in Escherichia Coli, and Identification of the Gene Product. *J Bacteriol* 169(12): 5492-5433 https://doi.org/10.1128/jb.169.12.5429-5433.1987

Ismagul A, Yang N, Maltseva E, Iskakova G, Mazonka I, Skiba K, Bi H, Eliby S, Jatayev S, Yuri S, Borisjuk N, Langridge P. (2018). A Biolistic Method for High-Throughput Production of Transgenic Wheat Plants with Single Gene Insertions. *BMC Plant Biol* 18: 135 https://doi.org/10.1186/s12870-018-1326-1

Jansing J, Sack M, Augustin SM, Fischer R, Bortesi L. (2018). CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in *Nicotiana benthamiana* for the production of recombinant proteins lacking β-1,2-xylose and core a-1,3,-fucose. *Plant Biotechnol J* 17(2) https://doi.org/10.111/pbi.12981

Jansen R, Embden DD, Gaastra W, Schouls LM. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43(6): 1565-75. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jian Y, Chen B, Duan C, Sun B, Yang J, Yang S. (2016). Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microb* 81(7): 2506-2514. https://doi.org/10.1128/AEM.04023-14

Jiménez A, Muñoz-Fernández G, Ledesma-Amaro R, Buey RM, Revuelta JL. (2019). One-vector CRISPR/Cas9 genome engineering of the industrial fungus *Ashbya gossypii*. *Microb Biotechnol* https://doi.org/10.1111/1751-7915.13425

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive immunity. *Science* 337(6096): 816-821. https://doi.org/10.1126/science.1225829

Jones DAB, Bertazzoni S, Turo CJ, Syme RA, Hane JK. (2018). Bioinformatic prediction of plant-pathogenicity effector proteins of fungi. *Curr Opin Microbiol*. 46: 43-49. doi: 10.1016/j.mib.2018.01.017

Jones JDG, Dangl JL. (2006). The plant immune system. *Nature*. 444: 323-329 https://doi.org/10.1038/nature05286

Juge N. (2006). Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci*. 2006 Jul;11(7): 359-67 DOI: 10.1016/j.tplants.2006.05.006

Jung T-Y, Park K-H, An Y, Schulga A, Deyev S, Jung J-H, Woo E-J. (2016). Structural features of Cas2 from *Thermococcus onnurineus* in CRISPR-cas system type IV. *Protein Sci* 25(10): 1890-1897. https://doi.org/10.1002/pro.2981

Käll L, Krogh A, Sonnhammer EL. (2007). Advantages of combined transmembrane topology and signal peptide predition – the Phobius web server. *Nucleic Acids Res*. 35: W429-32

DOI: 10.1093/nar/gkm256

Kato-Unoki Y, Takai Y, Kinoshita M, Mochizuki T, Tatsuno T, Shimasaki Y, Oshima Y. (2018).
Genome editing of pufferfish saxitoxin- and tetrodotoxin-binding protein type 2 in *Takifugu rubripes*. *Toxicon*. 153:58-61. https://doi.org/10.1016/j.toxicon.2018.08.001

Keyhani NO. (2018). Lipid biology in fungal stress and virulence: Entomopathogenic fungi. *Fungal Biol*. 122(6): 420-429 https://doi.org/10.1016/j.funbio.2017.07.003

Khedr MA, Massarotti A, Mohamed ME. (2018). Rational discovery of (+) (S) abscisic acid as a potential antifungal agent: a repurposing approach. *Sci. Rep.* 8(1): 8565 https://doi.org/10.1038/s41598-018-26998-x

Kim S, Bae T, Hwang J, and Kim JS. (2017). Rescue of High-Specificity Cas9 Variants Using SgRNAs with Matched 5' Nucleotides. *Genome Biol*. 18:218 https://doi.org/10.1186/s13059-017-1355-3

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, and Joung JK. (2016). High-Fidelity CRISPR-Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects. *Nature* 529(7587): 468-9 https://doi.org/10.1038/nature16526

Klarzynski O, Plesse B, Joubert J-M, Yvin J-C, Kopp M, Kloareg B, Fritig B. (2000). Linear β-1,3 glucans are elicitors of defense responses in tobacco. *Plant Physiology*. https://doi.org/10.1104/pp.124.3.1027

Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature*. https://doi.org/10.1038/s41586-019-1323-z

Kobayashi T, Namba M, Koyano T, Fukushima M, Sato M, Ohtsuka M, Matsuyama M. (2018). Successful production of genome-edited rats by the rGONAD method. *BMC Biotech* 18:19. https://doi.org/10.1186/s12896-018-0430-5

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA Cleavage. 533: 420-424 https://doi.org/10.1038/nature17946

Korber B. (2000). HIV signature and sequence variation analysis. Computational analysis of HIV molecular sequences, chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learns, eds. Dordrecht, Netherlands: Kluwer Academic Publishers

Kost TA, Condreay JP, and Jarvis DL. (2005). Baculovirus as Versatile Vectors for Protein Expression in Insect and Mammalian Cells. *Nat Biotechnol* 23(5): 567-75 https://doi.org/10.1038/nbt1095

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with hidden Markov model: application to complete genomes. *J Mol Biol.* 305(3): 567-580 DOI: 10.1006/jmbi.2000.4315

Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, and Citovsky V. (2001). Genetic Transformation of HeLa Cells by Agrobacterium. *P Natl A of Sci* 98(4): 1871-6 https://doi.org/10.1073/pnas.98.4.1871

Kuppireddy VS, Uversky VN, Toh SS, Tsai MC, Beckerson WC, Cahill C, Carman B, Perlin MH. (2017). Identification and initial characterization of effectors of an anther smut fungus and the potential host target proteins. *Int J Mol Sci,* 18: 2489 doi: 10.3390/ijms18112489

Lander E. (2016). The heroes of CRISPR. *Cell*. https://doi.org/10.1016/j.cell.2015.12.041

Langmead B, Trapnell C, PopM, Salzberg SL. (2009). Ultrafast and memory efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3):R25 https://doi.org/10.1186/gb-2009-10-3-r25

Lanver D, Tollot M, Schweizer G, Lo Presti L, Reissmann S, Ma LS, Schuster M, Tanaka S, Liang L, Ludwig N, Kahmann R. (2017). *Ustilago maydis* effectors and their impact on virulence. *Nat Rev Microbiol*. 15: 409-421 doi: 10.1038/nrmicro.2017.33

Le Gac M, Hood ME, Giraud T. (2007). Evolution of reproductive isolation within a parasitic fungal species complex. *Evolution*. 61-7: 1781-1787 DOI: 10.1111/j.1558-5646.2007.00144.x

Lee AH, Hurley B, Felsensteiner C, Yea C, Wenzislava C, Bartetzko V, Wang PW, Quach V, Lewis JD, Liu YC, Börnke F, Angers S, Wilde A, Guttman D, Desveaux D. 2012. A bacterial acetyltransferase destroys plant microtubule networks and blocks secretion. *PLOS Pathogens*. 8(2): e1002523 https://doi.org/10.1371/journal.ppat.1002523

Lee SJ and Rose JK. (2012). A yeast secretion trap assay for identification of secreted proteins from eukaryotic phytopathogens and their plant hosts. *Methods Mol Biol,* 835: 519-30 doi: 10.1007/978-1-61779-501-5_32

Li M, Liu H, Han J, Liu J, Wang R, Zhao D, Zhou J, Xiang H. (2013). Characterization of CRISPR

RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon *Haloferax mediterranei*. *J Bacteriol* 195(4): 867-75 https://doi.org/10.1128/JB.01688-12

Li M, Zhao L, Page-McCaw PS, Chen W. (2016). Zebrafish genome engineering using the CRISPR-Cas9 system. *Trends Genet* 32(12):815-827 https://doi.org/10.1016/j.tig.2016.10.005

Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res* 44(4). https://doi.org/10.1093/nar/gkv1044

Liu J, Wang Y, Lu Y, Zheng P, Sun J, Ma Y. (2017). Development of a CRISPR/Cas9 genome editing toolbox for Corynebacterium glutamicum. *Microbial Cell Factories* 16:205 https://doi.org/10.1186/s12934-017-0815-5

Liu Q, Yuan Y, Zhu F, Hong Y, Ge R. (2018). Efficient genome editing using CRISPR/Cas9 ribonucleoprotein approach in cultured Medaka fish cells. *Biol. Open.* 7(8) https://doi.org/10.1242/bio.035170

Liu Y, Wang Y, Xu S, Tang X, Zhao J, Yu C, He G, Xu H, Wang S, Tang Y, Fu C, Ma Y, Zhou G. (2019). Efficient genetic transformation of CRISPR/Cas9-mediated genome editing in *Lemna aequinoctialis*. *Plant Biotechnol J.* 1-10 https://doi.org/10.1111/pbi.13128

Liu Z. (2017). Inverse gene-for-gene interactions contribute additively to tan spot susceptibility in wheat. *Theor Appl Genet*. 130(6): 1267–1276 https://doi.org/10.1007/s00122-017-2886-4

Liu Z and Friesen TL. (2012). Polyethylene Glycol (PEG)-Mediated Transformation in Filamentous Fungal Pathogens. *Met Mol Biol* 835:365-75 https://doi.org/10.1007/978-1-61779-501-5_21

Lommel M, Tursch A, Rustarazo-Calvo L, Trageser B, Holstein TW. (2017). Genetic knockdown and knockout approaches in Hydra. *Semantic Scholar*. https://doi.org/10.1101/230300

Lopez-Obando M, Hoffmann B, Géry C, Guyon-Debast A, Téoulé E, Rameau C, Bonhomme S, Nogué, F. (2016). Simple and efficient targeting of multiple genes through CRISPR-Cas9 in *Physcomitrella patens*. *G3 (Bethesda)* 6(11):3647-3653. https://doi.org/10.1534/g3.116.033266

Louis Carroll. (1871). *Through the Looking-Glass and What Alice Found There*. Chicago, IL: W.B. Conkey Co 1900

Lu S, Shen X, and Chen B. (2017). Development of an Efficient Vector System for Gene Knock-out and

near in-Cis Gene Complementation in the Sugarcane Smut Fungus. *Sci Rep-UK* 7(1):3113

https://doi.org/10.1038/s41598-017-03233-7

Ma K-W, Ma W. (2016). Phytohormone pathways as targets of pathogens to facilitate infection. *Plant Mol. Biol.*91: 713-725. doi: 10.1007/s11103-016-0452-0

Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS, Lapidus A, Metcalf MW, Saunders E, Tapia R, Sowers KR. (2006). The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extansive rearrangement within Methanosarcinal genomes. *J Bacteriol* 188: 7922-7931

https://doi.org/10.1128/JB.00810-06

Majumdar S, Zhao P, Pfister NT, Compton M, Olson S, Glover III CVC, Wells L, Graveley BR, Terns RM, Terns MP. (2015). Three CRISPR-Cas9 immune effector complexes coexist in *Pyrococcus furiosus*. *RNA* 21(6): 1147-1158 https://doi.org/10.1261/rna.049130.114

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, and Church GM. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science* 339(6121):823-6

https://doi.org/10.1126/science.1232033

Manica A, Schleper C. (2013). CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus *Sulfolobus*. *RNA Biol* 10(5):671-678 https://doi.org/10.4161/rna.24154

Martin M. (2011). Cutadapt removes adapter sequences from highthroughput sequencing reads. *EMBnet.journal* 17(1): 10–12 https://doi.org/10.14806/ej.17.1.200

Marzin S, Hanemann A, Sharma S, Hensel G, Kumlehn J, Schweizer G, Marion SR. (2016). Are PECTIN ESTERASE INHIBITOR genes involved in mediating resistance to *Rhynochosporium commune* in Barley? *PLOS ONE*. 11(3): e0150485

https://doi.org/10.1371/journal.pone.0150485

Matsu-ura T, Baek M, Kwon J, Hong C. (2015). Efficient gene editing in *Neurospora crassa* with CRISPR technology. *Fungal Biol Biotech* 2:4. https://doi.org/10.1186/s40694-015-0015-1

McDonald JH, Kreitman M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, 351:652–654

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,

Gabriel S, Daly M, DePristo MA. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9): 1297–1303

Mefferd AL, Kornepati AVR, Bogerd HP, Kennedy EM, and Cullen BR. (2015). Expression of CRISPR/Cas Single Guide RNAs Using Small TRNA Promoters. *RNA* 21(9): 1683-9 https://doi.org/10.1261/rna.051631.115

Meile L, Croll D, Brunner PC, Plissonneau C, Hartmann FE, McDonald BA, Sánchez-allet A. (2018). A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to *Septoria tritici* blotch. *New Phytologist*. 219(3): 1048-1061

Meng Y, Hou Y, Wang H, Ji R, Liu B, Wen J, Niu L, Lin H. (2017). Targeted mutagenesis by CRISPR/Cas9 system in the model legume *Medicago truncatula*. *Plant Cell Rep* 36(2):371-374. https://doi.org/10.1007/s00299-016-2069-9

Miki D, Zhang W, Zeng W, Feng Z, Jian-Kang Z. (2018). CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. *Nat Commun* 9: 1967 https://doi.org/10.1038/s41467-018-04416-0

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 104: 19613-19618 https://doi.org/10.1073/pnas.0705147104

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta.* 1819: 86-96.

Mojica FJ, Díez-Vilaseñor C, García-Martínez J, Soria E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60(2):174-182. https://doi.org/10.1007/s/00239-004-0046-3

Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* 36(1):244-246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica FJM, Juez G, Rodriguez-Valera F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified *Pst*I sites. *Mol Microbiol* 9(3).

https://doi.org/10.1111/j.1365-2958.1993.tb01721.x

Morran LT, Schmidt OG, Gelarden IA, Parrish II RC, Lively CM. (2011). Running with the Red Queen: Host-parasite coevolution selects for biparental sex. *Science*. 333(6039): 216-218

Monod M, Capoccia S, Léchenne B, Zaugg C, Holdom M, Jousson O. (2002). Secreted proteases from pathogenic fungi. *International Journal of Medical Microbiology* 292(5-6): 405-19 https://doi.org/10.1078/1438-4221-00223

Mueller D. (2013). Corn disease loss estimates from the United States and Ontario, Canada. 2012 Purdue University BP-96-12—W

Muller HJ. (1964). The relation of recombination to mutational advance. Mutat. Res. 1: 2-9.

Nakayama T, Blitz IL, Fish MB, Odeleye AO, Manohar S, Cho KWY, Grainger RM. (2014). Chapter Seventeen – Cas9-based genome editing in *Xenopus tropicalis*. *Met Enzymology*. 546: 355-375 https://doi.org/10.1016/B978-0-12-801185-0.00017-9

Nayak DD and Metcalf MW. (2017). Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *PNAS* 114(11): 2976-2981 https://doi.org/10.1073/pnas.1618596114

Newman M-A, Sundelin T, Nielsen JT, Erbs G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers of Plant Science*. https://doi.org/10.3389/fpls.2013.00139

Nickel L, Weidenbach K, Jäger D, Backofen R, Lange SJ, Heidrich N, Schmitz RA. (2013). Two CRISPR-Cas9 systems in *Methanosarcina mazei* strain Gö1 display common processing features despite belonging to different types I and III. *RNA Biol.* 10(5): 779-791 https://doi.org/10.4161/rna.23928

Nødvig CS, Nielsen JB, Kogle ME, Mortensen UH. (2015). A CRISPR-Cas9 system for genetic engineering of filamentous fungi. *PLOS ONE* https://doi.org/10.1371/journal.pone.0133085

Ochiai H. (2015). Single-Base Pair Genome Editing in Human Cells by Using Site-Specific Endonucleases. *Int J Mol Sci* 16(9):21128-37 https://doi.org/10.3390/ijms160921128.

O'Geen H, Ren C, Nicolet CM, Perez AA, Halmai J, Le VM, MacKay JP, Farnham PJ, and Segal DJ. (2017). dCas9-Based Epigenome Editing Suggests Acquisition of Histone Methylation Is Not Sufficient for Target Gene Repression. *Nucleic Acids Res* 45(17):9901-16 https://doi.org/10.1093/nar/gkx578

Ozaki A, Konishi R, Otomo C, Kishida M, Takayama S, Matsumoto T, Tanaka T, Kondo A. (2017).
Metabolic engineering of *Schizosaccharomyces pombe* via CRISPR-Cas9 gneome editing for
lactic acid production from glucose and cellobiose. *Metab Eng Commun*
https://doi.org/10.1016/j.meteno.2017.08.002

Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y, Lee J, Edraki A, Megha S, Sontheimer EJ,
Maxwell KL, Davidson AR. (2016). Naturally Occurring Off- Switches for CRISPR-Cas9. *Cell*
167(7): 1829-38 https://doi.org/10.1016/j.cell.2016.11.017

Perlin MH, Beckerson WC, Gopinath A, Cobbs GA. (2020). *Molecular and Cellular Genetics: laboratory
Studies 2nd Edition*. San Diego, CA: Cognella Academic Publishing.

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): 1-24

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 Mol Sci.* 158 (5): 568-574

Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from
transmembrane regions. *Nature Methods*. 8: 785-786

Pevsner J. (2015). *Bioinformatics and Functional Genomics*. Hoboken, NJ: John Wiley & Sons.

Pfeifer B, Wittelsbürger U, Ramos-Onsins S E, Lercher M J. (2014). POPGENOME: an efficient Swiss
army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31, 1929–
1936 Pierleoni A, Martelli PL, Casadio R. 2008. PredGPI: a GPI-anchor predictor. *BMC
Bioinformatics*. 23(9): 392

Plissonneau C, Stürchler A, Croll D. (2016). The evolution of orphan regions in genomes of a fungal
pathogen of wheat. *MBio*. 7(5). pii: e01231-16

Plissonneau C, Hartmann FE, Croll D. (2018). Pangenome analyses of the wheat pathogen Zymoseptoria
tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol*. 16(1): 5

Port F, Strein C, Stricker M, Rauscher B, Heigwer F, Zhou J, Beyersdörffer C, Frei J, Hess A, Kern

    K, Malamud R, Pavlovic B, Rädecke K, Schmitt L, Voos L, Valentini E, Boutros M. (2019). A

    large-scale resource for tissue-specific CRISPR mutagenesis in *Drosphila*. *BioRxiv*

    https://doi.org/10.1101/636076

Presnell JS, Browne WE. (2019). Krüppel-like factor gene function in the ctenophore *mnemiopsis*

    suggest an ancient role in promoting cell proliferation in metazoan stem cell niches. *BioRxiv*

    https://dx.doi.org/10.1101/527002

Price MA, Cruz R, Baxter S, Escalettes F, Rosser SJ. (2019). CRISPR-Cas9 *In Situ* engineering of

    subtilisin E in *Bacillus subtilis*. *PLOS One* https://doi.org/10.1371/journal.pone.0210121

Quinlan AR, Hall IM. (2010). BEDTools: a flexible suite of utilities for comparing genomic features.

    Bioinformatics. 26(6):841-2 https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for

    Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rabe, F., Ajami-Rashidi, Z., Doehlemann, G., Kahmann, R., and Djamei, A. (2013). Degradation of the

    plant defence hormone salicylic acid by the biotrophic fungus *Ustilago maydis*. Mol Microbiol 89:

    179-188.

Rasys AM, Park S, Ball RE, Alcala AJ, Lauderdale JD, Menke DB. (2019). CRISPR-Cas9 gene

    editing in lizards through microinjection of unfertilized oocytes. *BioRxiv*

    https://dx.doi.org/10.1101/591446

Refrégier G, Gac ML, Jabbour F, Widmer A, Shykoff J, Yockteng R, Hood ME, Giraud T. (2008).

    Cophylogeny of the anther smut fungi and their Caryophyllaceous hosts: Prevalence of host shifts

    and importance of delimiting parasite species for inferring cospeciation. *BMC Evolutionary*

    Biology. 8(100) https://doi.org/10.1186/1471-2148-8-100

Reich, C., Wise, J. A. (1990). Evolutionary origin of the U6 small nuclear RNA intron. *Mol Cell Biol*

    10(10):5548-52 https://doi.org/10.1128/MCB.10.10.5548

Reis H, Pfiffi S, Hahn M. (2005). Molecular and functional characterization of a secreted lipase from

    Botrytis cinerea. *Mol Plant Pathol*. 6(3): 257-67

    https://doi.org/10.1111/j.1364-3703.2005.00280.x

Rep M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiology Letters*. 253(1): 19-27 https://doi.org/10.1016/j.femsle.2005.09.014

Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *methanococcus mariplaludis*. *Nucleic Acids Res* 40(19): 9887-9896 https://doi.org/10.1093/nar/gks737

Richardson CD, Kazane KR, Feng SJ, Zelin E, Bray NL, Schäfer AJ, Floor SN, Corn JE. (2010). CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat Gen* 50: 1132-1139. https://doi.org/10.1038/s41588-018-0174-0

Robson RL, Jones R, Robson RM, Schwartz A, Richardson TH. (2015). *Azotobacter* genomes: the genome of *Azotobacter chroococcum* NCIMB 8003 (ATCC 4412). *PLOS ONE* https://doi.org/10.1371/journal.pone.0127997

Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyet JL, Tellier A, Croll D. (2018). The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol.* 56: 21-40

Sanchez L, Courteaux B, Hubert J, Kauffmann S, Renault JH, Clement C, Baillieul F, Dorey S. (2012). Rhamnolipids elicit defense responses and induce disease resistance against biotrophic, hemibiotrophic, and necrotrophic pathogens that require different signaling pathways in Arabidopsis and highlight a central role for salicylic acid. *Plant Physiol.* 160: 1630-1641 https://doi.org/10.1104/pp.112.201913

Sanders SM, Ma Z, Hughes JM, Riscoe BM, Givson GA, Watson AM, Flici H, Frank U, Schnitzler CE, Baxevanis AD, Nicotra ML. (2018). CRISPR/Cas9-mediated gene knockin in the hydroid *Hydractinia symbiolongicarpus*. *BMC Genomics* 19: 649 https://doi.org/10.1186/s12864-018-5032-z

Savchenko T, Walley JW, Chehab EW, Xiao Y, Kaspi R, Pye MF, Mohamed ME, Lazarus CM, Bostock RM, Dehesh K. (2010). Arachidonic acid: an evolutionarily conserved signaling molecule modulates plant stress signaling networks. *Plant Cell*. 22(10): 3193-3205 https://doi.org/10.1105/tpc.110.073858

Schäfer AM, Kemler M, Bauer R, Begerow D. (2010). The illustrated life cycle of *Microbotryum* on the

host plant *Silene latifolia*. *Botany*. 88: 875-885.

Schirawski J. *et.al.* 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*. 330(6010): 1546-1548

Scholz I, Lange SJ, Hein S, Hess WR, Backofen R. (2013). CRISPR-Cas systems in the Cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and Cmr2 protein. *PLOS ONE* https://doi.org/10.1371/journal.pone.0056470

Schuster M, Kahmann R. (2019). CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genet Biol.* 130: 43-53 https://doi.org/10.1016/j.fgb.2019.04.016

Schuster M, Schweizer G, Kahmann R. (2018). Comparative analyses of secreted proteins in plant pathogenic smut fungi and related basidiomycetes. *Fungal Genet Biol*. 112: 21-30

Schuster M, Schweizer G, and Kahmann R. (2016). Genome editing in *Ustilago maydis* using the CRISPR-Cas system. *Fungal Genet and Biol* 89: 3-9. https://doi.org/10.1016/j.fgb.2015.09.001

Sekine R, Kawata T, and Muramoto T. (2018). CRISPR/Cas9 Mediated Targeting of Multiple Genes in Dictyostelium. *Sci Rep-UK* 8(1): 8471 https://doi.org/10.1038/s41598-018-26756-z

Sidik SM, Huet D, Lourido S. (2018). CRISPR-Cas9-based genome-wide screening of *Toxoplasma gondii*. *Nat Protoc* 13(1): 307-323 https://doi.org/10.1038/nprot.2017.131

Sorlien EL, Witucki MA, Ogas J. (2018). Efficient production and identification of CRISPR/Cas9-generated gene knockouts in the model system *Danio rerio*. *J Vis Exp* (138), e56969. https://doi.org/10.3791/56989

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli l, Singh KB, Manners JM, Taylor JM. (2015). EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*. https://doi.org/10.1111/nph.13794

Square T, Romášek M, Jandzik D, Cattell MV, Klymokowsky M, Medeiros DM. (2015). CRISPR/Cas9-mediated mutagenesis in the sea lamprey *Petromyzon marinus*: a powerful tool for understanding ancestral gene functions in vertebrates. *Development* 142: 4180-4187. https://doi.org/10.1242/dev.125609

Stachler A-E, Marchfelder A. (2016). Gene repression in *Haloarchaea* using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B system. *J Biol Chem* 291(29): 15226-

15242 https://doi.org/10.1074/jbc.M116.724062

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. (2007). Advanced models for

detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.*

35: W506–11.

Stolfi A, Gandhi S, Salek F, Christiaen L. (2014). Tissue-specific genome editing in *Ciona* embryos by

CRISPR/Cas9. *Development* 141(21): 4115-20 https://doi.org/10.1242/dev.114488

Strobel G. (2018). The Emergence of Endophytic Microbes and their Biological Promise. *Journal of

Fungi*. 4(2): 57

Sugano SS, Nishihama R, Shirakawa M, Takagi J, Matsuda Y, Ishida S, Shimada T, hara-Nishimura

I, Osakabe K, Kohchi T. (2018). Efficient CRISPR/Cas9-based genome editing and its application

to conditional genetic analysis in *Marchantia polymorpha*. *PLOS ONE*

https://doi.org/10.1371/journal.pone.0205117

Sun D, Guo Z, Liu Y, Zhang Y. (2017). Progress and prospects of CRISPR/Cas9 systems in insects and

other arthropods. *Front Physiol* https://doi.org/10.3389/fphys.2017.00608

Tan SZ, Reisch CR, Prather KLJ. (2018). A robust CRISPR interference gene repression system in

Pseudomonas. *J Bacteriol* 200(7):1-12 https://doi.org/10.1128/JB.00575-17

Taning CNT, Van Eynde B, Yu N, Ma S, Smagghe G. (2017). CRISPR/Cas9 in insects: applications,

best practices and biosafety concerns. *J Insect Physiol* 98: 245-257.

https://doi.org/10.1016/j.jinphys.2017.01.007

Tao W, Yang A, Deng Z, Sun Y. (2018). CRISPR/Cas9-based editing of *Streptomyces* for discovery,

characterization, and production of natural products. *Front Microbiol* 9: 1660.

https://doi.org/10.3389/fmicb.2018.01660

Testa AC, Oliver AP, Hane JK. (2016). OcculterCut: A Comprehensive Survey of AT-Rich Regions in

Fungal Genomes. *Genome Biol Evol.* 8(6): 2044-64 https://doi.org/10.1093/gbe/evw121

Thrall PH. (2016). Epidemiological and evolutionary outcomes in gene-for-gene and matching allele

models. *Frontiers in Plant Science*. 6: 1084 https://doi.org/10.3389/fpls.2015.01084

Toh SS, Chen Z, Rouchka EC, Schultz DJ, Cuomo CA, Perlin MH. (2018). *Pas de deux*: An intricate dance

of anther smut and its host. G3: Genes, Genomes, Genetics| 8:2 505-518;

https://doi.org/10.1534/g3.117.300318

Toh SS, Chen Z, Schultz DJ, Cuomo CA, Perlin MH. (2017). Transcriptional analysis of mating and pre-infection stages of the anther smut, *Microbotryum lychnidis-dioicae*. *Microbiology.* 163: 410-420.

Toh SS, Treves DS, Barati MT, and Perlin MH. (2016). Reliable Transformation System for Microbotryum Lychnidis-Dioicae Informed by Genome and Transcriptome Project. *Arch Microbiol* 198(8): 813-25 https://doi.org/10.1007/s00203-016-1244-2

Toh SS, Perlin MH. (2015). Size does matter: staging of *Silene latifolia* floral buds for transcriptome studies. *Int J Mol Sci* 16: 22027-22045.

Tsuda K, Katagiri F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinions in Plant Biology*. 13(4): 459-65.

https://doi.org/10.1016/j.pbi.2010.04.006

Tu Q, Yin J, Fu J, Herrmann J, Li Y, Yin Y, Stewart AF, Müller R. (2016). Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. *Sci Rep-UK* 6; 246-48. https://doi.org/10.1038/srep24648

Urry LA, Cain ML, Wasserman SA, Minorsky PV, Reece JB. (2017). *Campbell Biology AP Edition, 11th Edition*. NY: Pearson.

USDA. (2017). Agricultural Research Service: Cereal Disease Lab: St. Paul, MN

https://www.ars.usda.gov/midwest-area/stpaul/cereal-disease-lab/docs/cereal-rusts/wheat-stem-rust/

Van de Wouw AP, Cozijnsen AJ, Hane JK, Brunner PC, McDonald BA, Oliver RP, Howlett BJ. (2010). Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathog*. 4;6(11): e1001180 https://doi.org/10.1371/journal.ppat.1001180

Van Valen L. (1973). A new evolutionary law. *Evolution Theory*. 1: 1-30

Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytol*. 198: 347-385

Vincent D, Rafiqui M, Job D. (2020). The multiple facets of plant-fungal interactions reveled through plant and fungal secretomics. *Frontiers in Plant Science*. 10(1626): 1-19

Vonk PJ, Escobar N, Wösten HAB, Lugones LG, Ohm RA. (2019). High-throughput targeted gene deletion in the model mushroom *Schizophyllum commune* using pre-assembled Cas9 ribonucleoptroteins. *Sci Rep-UK* 9:7632 https://doi.org/10.1038/s41598-019-44133-2

Wang F, Sun X, Shi X, Zhai H, Tian C, Kong F, Liu B, Yuan X. (2016). A global analysis of the polygalacturonase gene family in soybean (*Glycine max*). *PLoS One*. 11(9):e0163012 https://doi.org/10.1371/journal.pone.0163012

Wang L, Deng A, Zhang Y, Liu S, Liang Y, Bai H, Cui D, Oiu Q, Shang X, Yang Z, He X, Wen T. (2018). Efficient CRISPR-Cas9 mediated multiplex genome editing in yeast. *Biotechnol Biofuels* 11: 277 https://doi.org/10.1186/s13068-018-1271-0

Wang L, Wang L, Tan Q, Fan Q, Zhu H, Hong Z, Zhang Z, Duanmu D. (2016). Efficient inactivation of symbiotic nitrogen fixation related genes in *Lotus japonicus* using CRISPR-Cas9. *Front Plant Sci* 7:1333 https://doi.org/10.3389/fpls.2016.01333

Watters KE, Fellmann C, Bai HB, Ren SM, and Doudna JA. (2018). Systematic Discovery of Natural CRISPR-Cas12a Inhibitors. *Science* 362(6411): 236-9 https://doi.org/10.1126/science.aau5138

Wei Y, Qiu Y, Chen Y, Liu G, Zhang Y, Xu L, and Ding Q. (2017). CRISPR/Cas9 with Single Guide RNA Expression Driven by Small TRNA Promoters Showed Reduced Editing Efficiency Compared to a U6 Promoter. *RNA* 23(1)1-5 https://doi.org/10.1261/rna.057596.116

Wei W, Xin H, Roy B, Dai J, Miao Y, Gao G. (2014). Heritable genome editing with CRISPR/Cas9 in the silkworm, *Bombyx mori*. *PLOS ONE* https://doi.org/10.1371/journal.pone.0101210

Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, Armstrong MR, Grouffaud S, van West P, Chapman S, Hein I, Toth IK, Pritchard L, Birch PR. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*. 450: 115-118

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(3): 809–816.

Willmann R, Lajunen HM, Erbs G, Newman MA, Kolb D, Tsuda K, Katagiri F, Fliegmann J, Bono JJ,

Cullimore JV, Jehle AK, Götz F, Kulik A, Molinaro A, Lipka V, Gust AA, Nürnberger T. (2011). Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* 108: 19824–19829. https://doi.org/10.1073/pnas.1112862108

Witte H, Moreno E, Rödelsperger C, Kim J, Kim J-S, Streit A, Sommer RJ. (2015). Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificusi. Dev Genes Evol* 225(1): 55-62 https://doi.org/10.1007/s00427-014-0486-8

Wu R, Lucke M, Jang Y, Zhu W, Symeonidi E, Wang C, Fitz J, Xi W, Schwab R, Weigel D. (2018). An efficient CRISPR vector toolbox for engineering large deletions in *Arabidopsis thaliana*. *Plant Methods* 14: 65 https://doi.org/10.1186/s13007-018-0330-7

Wu X, Kriz AJ, and Sharp PA. (2014). Target Specificity of the CRISPR-Cas9 System. *Quantitative Biol* 2(2): 59-70 https://doi.org/10.1007/s40484-014-0030-x

Westbrook AW, Moo-Young M, Chou CP. (2018). Development of a CRISPR-Cas9 tool kit for comprehensive engineering of *Bacillus subtilis*. *Appl Environ Microb* 82(16): 4876-4895 https://doi.org/10.1128/AEM.01159-16

Xiao Y, Wang S, Rommelfanger S, Balassy A, Barba-Ostria C, Gu P, Galazka JM, Zhang F. (2018). Developing a Cas9-based tool to engineer native plasmids in *Synechocystis* sp. PCC 6803. *Biotechnol Bioeng* 115(9): 2305-2314 https://doi.org/10.1002/bit.26747

Xie C, Chen YL, Wang DF, Wang YL, Zhang TP, Li H, Liang F, Zhao Y, and Zhang GY. (2017). SgRNA Expression of CRIPSR-Cas9 System Based on MiRNA Polycistrons as a Versatile Tool to Manipulate Multiple and Tissue-Specific Genome Editing. *Sci Rep-UK* 7(1): 5795 https://doi.org/10.1038/s41598-017-06216-w

Xin LUO, Min LI, Bing SU. (2016). Application of the genome editing tool CRISPR/Cas9 in non-human primates. *Zool Research* 37(4): 241-219. doi: 10.13918/j.issn.2095-8137.2016.4.214

Xu J, Lian W, Jia Y, Li L, and Huang Z. (2017). Optimized guide RNA structure for genome editing via Cas9. *Oncotarget* 8(55): 94166-94171 https://doi.org/10.18632/oncotarget.21607

Yang L, Mali P, Kim-Kiselak C, Church G. (2014). CRISPR-Cas-mediated targeted genome editing in human cells. *Methods Mol. Biol.* 1114:245-267 https://doi.org/10.1007/978-1-62703-761-7_16

Yang Z, Nielsen R, Goldman N, Perdersen AM. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155(1): 431-449

Zeaiter Z, Mapelli F, Crotti E, Borin S. (2018). Methods for the genetic manipulation of marina bacteria *Electron J Biotechn* 33: 17-28. https://doi.org/10.1016/j.ejbt.2018.03.003

Zebec Z, Manica A, Zhang J, White MF, Schleper C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res* 42(8): 5280-4288. https://doi.org/10.1093/nar/gku161

Zhang T, Gao Y, Wang R, and Zhao Y. (2017). Production of Guide RNAs in Vitro and in Vivo for CRISPR Using Ribozymes and RNA Polymerase II Promoters. *BIO-PROTOCOL* 7(4)e2148 https://doi.org/10.21769/BioProtoc.2148

Zhang Z, Zhang Y, Gao F, Han S, Cheah KS, Tse H-F, Lian Q. (2017). CRISPR/Cas9 genome-editing system in human stem cells: current status and future prospects. *Mol Ther-Nucl Acids*. 9:230-241. https://doi.org/10.1016/j.omtn.2017.09.009

Zheng K, Wang Y, Li N, Jiang F-F, Wu C-X, Liu F, Chen, H-C, Liu, Z-F. (2018). Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Commun. Biol.* 1: 32

Zuo Q, Wang Y, Cheng S, Lian C, Tang B, Wang F, Lu Z, Ji Y, Zhao R, Zhang W, Jin K, Song J, Zhang Y, Li B. (2016). Site-directed genome knockout in chicken cell line and embryos can use CRISPR/Cas gene editing technology. *G3 (Bethesda)* 6(6): 1787-1792. https://doi.org/10.1534/g3.116.028803

APPENDIX

Supplemental Material:

<u>Supplemental Table C3-1</u>

Microbotryum	Nucleotide Sequence	# bp Sub.	Amino Acid Sequence	# aa Sub.
MVLG	ATGACCTCACAAGTGCGAA TGCAAGTCGAGAGTCGTG CCCAACGACGCGCAGGGG CCTACGCGTCCATGAGGTT GTTGCTCGCTCTGGTCTTC GCCCTCTGCACCTTAGCG CACCTGCCGACAACCAGT GCCGCACCGCTGGCTTCG GAGCAAATCTCGTCCGGT CTCGTCTTTCGACAAGAAC CACCCAGATGGTTACAATT CTCTCGGCCTCATGAGAAA GTGTCGCATCAAGGCAAA GATCATCTGGACTGGAAAA ACACGTCGCCTTCGCCGTT CACTTCCAGCGAACCATCG AGGCGTGTGAAACGTGAC GAGATGTGGGAGCAGTAC ATCGAGGGGGATGAGATC GACGGGGAGAAGAGCGAG GATGTTCGTGCAGGAGAT CCGGATGTTGCCGGGGAT GAAGTCCTGACAGACACC GAGATCGCGGGCGGAGCG GACGAAGCGGGCGAGGG GTCCACAGGGGAAAAGTG GTGGCAAGCCAGGAGACG ATTGCGTGAGAGGCGATC GGCCACCACAAGGGTTGT TCCGTAACGGCTCTTTGTT C		MTSQVRMQVESRAQRRA GAYASMRLLLALVFALCTL AHLPTTSAAPLASEQISSG LVFRQEPPRWLQFSRPHE KVSHQGKDHLDWKNTSP SPFTSSEPSRRVKRDEM WEQYIEGDEIDGEKSEDV RAGDPDVAGDEVLTDTEI AGGADEAGEGSTGEKW WQARRRLRERRSATTRV VP*	
MvSl	ATGACCTCACAAGTGCGAA TGCAAGTCGAGAGTCGTG CCCAACGACGCGCAGGGG CCTACGCGTCCATGAGGTT GTTGCTCGCTCTGGTCTTC GCCCTCTGCACCTTAGCG CACCTGCCGACAACCAGT GCCGCACCGCTGGCTTCG	0	MTSQVRMQVESRAQRRA GAYASMRLLLALVFALCTL AHLPTTSAAPLASEQISSG LVFRQEPPRWLQFSRPHE KVSHQGKDHLDWKNTSP SPFTSSEPSRRVKRDEM WEQYIEGDEIDGEKSEDV RAGDPDVAGDEVLTDTEI	0

	GAGCAAATCTCGTCCGGT CTCGTCTTTCGACAAGAAC CACCCAGATGGTTACAATT CTCTCGGCCTCATGAGAAA GTGTCGCATCAAGGCAAA GATCATCTGGACTGGAAAA ACACGTCGCCTTCGCCGTT CACTTCCAGCGAACCATCG AGGCGTGTGAAACGTGAC GAGATGTGGGAGCAGTAC ATCGAGGGGGATGAGATC GACGGGGAGAAGAGCGAG GATGTTCGTGCAGGAGAT CCGGATGTTGCCGGGGAT GAAGTCCTGACAGACACC GAGATCGCGGGCGGAGCG GACGAAGCGGGCGAGGG GTCCACAGGGGAAAAGTG GTGGCAAGCCAGGAGACG ATTGCGTGAGAGGCGATC GGCCACCACAAGGGTTGT TCCGTAACGGCTCTTTGTT C		AGGGADEAGEGSTGEKW WQARRRLRERRSATTRV VP*	
MvSd	ATGACCTCACAAGTGCGAA TGCAAGTCGAGAGTCGTG CCCAACGACGCGCAGGGG CCTACGCGTCCATGAGGTT GTTGCTCGCTCTGGTCTTC GCCCTCTGCACCCTAGCG CACCTGCCGACAACCAGT GCCGCACCGCTGGCTTCG GAGCAAATCTCGTCCGGT CTCGTCTTTCGACAAGAAC CACCCAGATGGTTACAATT CTCTCGGCCTCATGAGAAA GTGTCGCATCAAGACAAAG ATCATCTGGACTGGAAAAA CACGTCGCCTTCGCCGTT CACTTCCAGCGAACCATCG AGGCGTGTGAAACGTGAC GAGATGTGGGAGCAGTAC ATCGAGGGGGATGAGATC GACGGGGAGAAGAGCGAG GAAGTTCGTGCAGGAGAT CCGGATGTTGCCAGGGAT GAAGTTCTGACAGACACC GAGATCGCGGGCGAAGCG GACGAAGCGGGCGAGGG	6	MTSQVRMQVESRAQRRA GAYASMRLLLALVFALCTL AHLPTTSAAPLASEQISSG LVFRQEPPRWLQFSRPHE KVSHQDKDHLDWKNTSP SPFTSSEPSRRVKRDEM WEQYIEGDEIDGEKSEEV RAGDPDVARDEVLTDTEI AGEADEAGEGSTGEKWW QARRRLRERRSATTRVVP *	4

	GTCCACAGGGGAAAAGTG GTGGCAAGCCAGGAGACG ATTGCGTGAGAGGCGATC GGCCACCACAAGGGTTGT TCCGTAACGGCTCTTTGTT C			
MvSp	ATGAACTCACAAGCGCGAA TGCAAGTCGAGAGTCGTG CCCAACGGCGCGCGGGG GCCTACACGTCCCAGAGG TTGTTACTCGCTCTGGTCT TCGCCCTCTGCACCTTAGC ACACCTATCGACAACCAGA GCCGCACCGCTGGCTTCA GAGCAAATCTCGTCCCGTC TCGTCTTTCGACAAGAAAC ACCGCGATGGTTACGATTC TCTCGCCCTCATGAGAAG GCCTCGCATCAAGGCAAA CATCATCTGGACCGGAAAA ACACGTTGCCTTCGCCGTT CACTTCCAGCGAACCATCG AGGCGGGTGAAACGTGAC GAGATGTGGGAGCAGTAC ATTGAGGGGGATGAAATC GACGGGGAGAACAGCGAG AAACTTGGTCCAGGAGATC CGGATGTTGCCGAGGATG AAATTCTGACAAACATCGA GATCGCGGGCGAAGCTGA GGAAGCGGGCGAGTGGTC CATAGGAGAAAAGTGGTG GCAAGCCAGGAGACGATT GCGTGAGCGGCGTTTGGC CACCGCAAAGGTTGTTCC GTAACGGCGCTTTGTTC	51	MNSQARMQVESRAQRRA GAYTSQRLLLALVFALCTL AHLSTTRAAPLASEQISSR LVFRQETPRWLRFSRPHE KASHQGKHHLDRKNTLPS PFTSSEPSRRVKRDEMW EQYIEGDEIDGENSEKLG PGDPDVAEDEILTNIEIAG EAEEAGEWSIGEKWWQA RRRLRERRLATAKVVP*	29

Supplemental Table C3-1

Program	MvSl	MvSd	MvSp
SignalP 4.1	D Score = 0.695	D Score = 0.695	D Score = 0.412
EffectorP 1.0	Prob = 0.865	Prob = 0.686	Prob = 0.645
Pfam 32	Hits = 0	Hits = 0	Hits = 0
HMMER 3.1b1	Hits = 0	Hits = 0	Hits = 0

Supplemental File C5-1.

>Cas9
```
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
```

>Cas9n (D10A)
```
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
```

>dCas9 (D10A & H840A)
```
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
```

>Cas9-HF (N497A, R661A, Q695A, & Q926A)
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

>xCas9(3.7)-BE3 (A262T, R324L, S409I, E480K, E543D, M694I, E1219V)
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK**L**YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSG**D**QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF**I**QLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

Supplemental File C5-2.

>Trp_Promoter_Bacteria_Ecoli
ACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGC

>PTS_Promoter_Archaea_Hmediterranei
ACTGATGTTTGTTAACGACTAAAACGAGCAGAAATATTCGTATTCGAAAGGATTTTTGAG

>ADH1_Promoter_Fungi_Scerevisiae
Atccttttgttgtttccgggtgtacaatatggacttcctctttctggcaaccaaacccatacatcgggattcctataataccttcgtt
ggtctccctaacatgtaggtggcggaggggagatatacaatagaacagataccagacaagacataatgggctaaacaagactacaccaa
ttacactgcctcattgatggtggtacataacgaactaatactgtagccctagacttgatagccatcatcatatcgaagtttcactaccc
tttttccatttgccatctattgaagtaataataggcgcatgcaacttcttttctttttttttcttttctctctccccccgttgttgtctc
accatatccgcaatgacaaaaaaatgatggaagacactaaaggaaaaaattaacgacaaagacagcaccaacagatgtcgttgttccag
agctgatgaggggtatctcgaagcacacgaaactttttccttccttcattcacgcacactactctctaatgagcaacggtatacggcct
tccttccagttacttgaatttgaaataaaaaaaagtttgctgtcttgctatcaagtataaatagacctgcaattattaatctttttgttt
cctcgtcattgttctcgttcccttttcttccttgtttctttttctgcacaatatttcaagctataccaagcatacaatcaact

>Ubi_Promoter_Plant_Zmays
ctgcagtgcagcgtgacccggtcgtgcccctctctagagataatgagcattgcatgtctaagttataaaaaattaccacatattttttt
tgtcacacttgtttgaagtgcagtttatctatctttatacatatatttaaactttactctacgaataatataatctatagtactacaat
aatatcagtgtttagagaatcatatataatgaacagttagacatggtctaaaggacaattgagtattttgacaacaggactctacagtt
ttatctttttagtgtgcatgtgttctccttttttttttgcaaatagcttcacctatataatacttcatccattttattagtacatccatt
tagggtttagggttaatggtttttatagactaatttttttagtacatctattttattctattttagcctctaaattaagaaaactaaaa
ctctatttagtttttttatttaataatttagatataaaatagaataaaatttaaagtgactaaaaattaaacaaatacccttttaagaaat
taaaaaaactaaggaaacatttttcttgtttcgagtagataatgccagcctgttaaacgccgtcgacgagtctaacggacaccaaccag
cgaaccagcagcgtcgcgtcgggccaagcgaagcagacggcacggcatctctgtcgctgcctctggacccctctcgagagttccgctcc
accgttggacttgctccgctgtcggcatccagaaattgcgtggcggagcggcagacgtgagccggcacggcaggcggcctcctcctcct
ctcacggcacggcagctacgggggattcctttcccaccgctccttcgctttcccttcctcgcccgccgtaataaatagacaccccctcc
acaccctctttccccaacctcgtgttgttcggagcgcacacacacacaaccagatctcccccaaatccacccgtcggcacctccgcttc
aaggtacgccgctcgtcctcccccccccccctctctaccttctctagatcggcgttccggtccatggttagggcccggtagttctact
tctgttcatgtttgtgttagatccgtgtttgttagatccgtgctgctagcgttcgtacacggatgcgacctgtacgtcagacacgtt
ctgattgctaacttgccagtgtttctctttgggggaatcctgggatggctctagccgttccgcagacgggatcgatttcatgatttttttt
tgtttcgttcatagggtttggtttgccctttttcctttatttcaatatatgccgtgcacttgtttgtcgggtcatctttttcatgcttttt
ttttgtcttggttgtgatgatgtggtctggttgggcggtcgttctagatcggagtagaattctgtttcaaactacctggtggatttatt
aattttggatctgtatgtgtgtgccatacatattcatagttacgaattgaagatgatggatggaaatatcgatctaggataggtataca
tgttgatgcgggttttactgatgcatatacagagatgctttttgttcgcttggttgtgatgatgtggtgtggttgggcggtcgttcatt
cgttctagatcggagtagaatactgtttcaaactacctggtgtatttattaattttggaactgtatgtgtgtgtcatacatcttcatag

```
ttacgagtttaagatggatggaaatatcgatctaggataggtatacatgttgatgtgggttttactgatgcatatacatgatggcatat
gcagcatctattcatatgctctaaccttgagtacctatctattataataaacaagtatgttttataattattttgatcttgatatactt
ggatgatggcatatgcagcagctatatgtggatttttttagccctgccttcatacgctatttatttgcttggtactgtttctttttgtcg
atgctcaccctgttgtttggtgttacttctgcag
```

```
>CMV_Promoter_Animal_Mmusculus
cgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatag
taacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatg
ccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttg
gcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcac
ggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaac
tccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagct
```

Supplemental File C5-3.
```
>J23119_Promoter_Bacteria_Ecoli
TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT
```

```
>PTS_Promoter_Archaea_Hmediterranei
ACTGATGTTTGTTAACGACTAAAACGAGCAGAAATATTCGTATTCGAAAGGATTTTTGAG
```

```
>H1_Promoter_Eukaryotes_Hsapiens
GAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAA
GATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTT
TGGATTTGGGAATCGTATAAGAACTGTATGAGACCAC
```

```
>U3_Promoter_Eukaryotes_Osativa
AAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAACTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCA
GATGTGCAGTCAGGGACCATAGCACAAGACAGGCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTCG
GAAACCACGTGATGTGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATTGCAG
TATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCACATAGATCAAAGCTGATTTAAA
AGAGTTGTGCAGATGATCCGTGGCA
```

```
>U6_Promoter_Eukaryotes_Hsapiens
AAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAA
TTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTT
TTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
```

160

Primers:

Name	Sequence5'-> 3'	TM	Length	Date Ordered	Order Ref
MVLG_02245FW	GCAGAATTCATGACCTCACAAGTGCGAATGCAAGTCGAG	72.1	39	31-Oct-16	1163909
MVLG_02245RV	GCGGCCGCGGCACTGGTTGTCGGCAGGTGCGCTAAG	80.6	36	31-Oct-16	1163909
MvSd-IT02g04324FW	GCAGAATTCATGAAGCTGTCCACCTTGATCCTCACCCTT	72.1	39	31-Oct-16	1163909
MvSd-IT02g04324RV	GCGGCCGCCGCCACGGCAATGCTGCTGCCGACAAGA	80.6	36	31-Oct-16	1163909
MvSp880g00831FW	GCAGAATTCATGGTGTCCAAGCTGCTGGGCGCACTGGAC	76.3	39	31-Oct-16	1163909
MvSp880g00831RV	GCGGCCGCTGCGAGGGCTCTTGACAAAGGAAAGAAC	76.1	36	31-Oct-16	1163909
MvSp880g16237FW	GCAGAATTCATGTTCGTCTTGCTGCTCACCACCCCAGAT	73.1	39	31-Oct-16	1163909
MvSp880g16237FW	GCGGCCGCTGCCAGGAGACAACCAACGCGCAATCT	78.3	35	31-Oct-16	1163909
MVLG_05398F	GCAGAATTCATGCGCACCCCTTCCCTCGCGTTCGTCTTGCTT	76.4	42	12-Dec-16	1181690
MVLG_05398R	GCGGCCGCAGCGCTGACTGCAGACGTTAGACCAAGCAAG	64.1	39	12-Dec-16	1181690
MvSd01662FW	GCAGAATTCATGAGATTGCTCTTTGCTATCACCTTT	67.0	36	23-Jan-17	1196597
MvSd01662RV	GCGGCCGCCGCATGGATCATACACACCGCGAGGCTA	78.3	36	23-Jan-17	1196597
MvSd02874FW	GCAGAATTCATGAAGCTCCTCGCGATTGCCGTCGCC	74.9	36	23-Jan-17	1196597
MvSd02874RV	GCGGCCGCCAGCTTGACTTGCGGCGACCCTCATGGC	80.6	36	23-Jan-17	1196597
MvSd07159FW	GCAGAATTCATGCTTTTCCCCATCGTGTGCTTCACG	71.5	36	23-Jan-17	1196597
MvSd07159RV	GCGGCCGCCCCAGGTCAGGAATGTTTCAAGCAAGAG	76.1	36	23-Jan-17	1196597
MvSp01648FW	GCAGAATTCATGCGCTTCTCCATGCTCATCCCCGTT	72.6	36	27-Feb-17	1212819
MvSp01648RV	GCGGCCGCCGCCGATGACGGTAGCGATGAGCGAGGC	81.8	36	27-Feb-17	1212819
MvSp10116FW	GCAGAATTCATGCTCCCCAGGTTATCTCGCCATCTG	72.6	36	27-Feb-17	1212819
MvSp10116RV	GCGGCCGCCAGCAAGGCTAGGTCCTGGTGCAGAACA	78.3	36	27-Feb-17	1212819
MVLG_02245RVNew	GCGGCCGCCGGCACTGGTTGTCGGCAGGTGCGCTAA	80.6	36	27-Feb-17	1212819
MvSdG10114RVNew	GCGGCCGCCCGCCACGGCAATGCTGCTGCCGACAAG	81.8	36	27-Feb-17	1212819
MvSp-G10007RVNew	GCGGCCGCCTGCGAGGGCTCTTGACAAAGGAAAGAA	76.1	36	27-Feb-17	1212819
MvSp-G15012RVNew	GCGGCCGCCTGCCAGCGAGACAACCAACGCGCAATC	79.5	36	27-Feb-17	1212819
tRNATryGTAFw	GCAAAGCTTCCAATTCCTGTTCCCCCTGATTGG	70.9	33	27-Feb-17	1212819
tRNATryGTARv	TCCCGCCACGCGGGATCGAACCGCGGACCCTCA	79.6	33	27-Feb-17	1212819
MVLG5582TarFw	TTCGATCCCGCGTGGCGGGAGTGATTTGACGAGCCACGAGTTTTAGAGCTAGAAATAGC	78.0	59	27-Feb-17	1212819
MVLG5582TarRv	TTCGATCCCGCGTGGCGGGAGTGATTTGACGAGCCACGAGTTTTAGAGCTAGAAATAGC	78.0	59	27-Feb-17	1212819
gRNAScaffoldF	GCTATTTCTAGCTCTAAAACTCGTGGCTCGTCAAATCACTCCCGCCACGCGGGATCGAA	59.2	59	27-Feb-17	1212819

gRNAScaffoldR	AGCACCGACTCGGTGCCACTTTTTCAAGTT	68.7	30	27-Feb-17	1212819
PolIIITermF	AGTGGCACCGAGTCGGTGCTAAAAACTAGACCCAGCTTTCTTGTAGGTACC	76.0	51	27-Feb-17	1212819
PolIIITermR	GGTACCTACAAGAAAGCTGGGTCTAGTTTTTAGCACCGACTCGGTGCCACT	76.0	51	27-Feb-17	1212819
MVLG2523PromF	GCAGGTACCGTGCAACTGTGAGCCCCGACTCCG	77.1	33	27-Feb-17	1212819
MVLG2523PromR	GCAGGATCCCTACGTCATGCACTGACGATGGGA	73.4	33	27-Feb-17	1212819
CRISPRScrnF	GCTCTACCGATGCCTTACCA	62.4	20	24-Apr-17	1240208
CRISPRScrnR	GCTGCGAAACTCCTTCACTC	62.4	20	24-Apr-17	1240208
Cas9ScrF	CTCCACCATTCTCTCGGTCT	62.4	20	12-May-17	1249612
Cas9ScrR	ATTGGGCGATGAGGTTTTC	58.0	19	12-May-17	1249612
TarScrF	CAGCACCTAGCGTCGTTTTC	62.4	20	12-May-17	1249612
TarScrR	CCTTCGTGGCTCTCGTATTT	60.4	20	12-May-17	1249612
RnosScrF	CCGAGGACAACGAACAAAG	60.4	20	12-May-17	1249612
TnosScrR	GTAAAACGACGGCCAGTGAA	60.4	20	12-May-17	1249612
5585Seq50bpUpE1F	GCAACACTCCGCCGCGGCGGCACTGCAGCTGCT	79.6	33	21-Jul-17	1282410
5585Seq950inR	TCGTACGGGCACGAGATCTCAGCATCGGCAACC	74.6	33	21-Jul-17	1282410
5585Seq950inF	GGTTGCCGATGCTGAGATCTCGTGCCCGTACGA	74.6	33	21-Jul-17	1282410
5585Seq1.95KbinR	TAGGAACCAAAATCGGGGACGTTGATGCGTTTG	69.6	33	21-Jul-17	1282410
Cas9Seq50bpUpF	ACGACAAAGTGATGAAGCAGCTGAAGCGTAGGC	70.9	33	21-Jul-17	1282410
Cas9Seq50bpUpR	TCTCTAACTTTATAGAACTGAAAATCTTTGCGG	63.4	33	21-Jul-17	1282410
MV5585E2F	TTCGATCCCGCGTGGCGGGACACCGACACTACGAACTGTGTTTTAGAGCTAGAAATAGC	78.0	59	26-Jul-17	1286090
MV5585E2R	GCTATTTCTAGCTCTAAAAACAGTTCGTAGTGTCGGTGTCCCGCCACGCGGGATCGAA	78.0	59	26-Jul-17	1286090
MV5585E4F	TTCGATCCCGCGTGGCGGGAGCTTCCCGCTACTTCAAGAGTTTTAGAGCTAGAAATAGC	78.0	59	10-Aug-17	1294116
MV5585E4R	GCTATTTCTAGCTCTAAAACTCTTGAAGTAGCGGGAAGCTCCCGCCACGCGGGATCGAA	78.0	59	10-Aug-17	1294116
xcmlFW	GCACCACATCTGTCTTGGGGCGATGGCGATCAGTAT	73.8	36	10-Aug-17	1294116
BsrGIRv	GCATGTACACTATTCCTTTGCCCTCGGACGAGTGCT	72.6	36	10-Aug-17	1294116
MvSd04324R	GGATCCTTAGAATCCCACGGGAAAAATAGTGGT	68.4	33	27-Sep-17	1316106
MvSd07159F	GAATTCGAGAAGGTCCCGTCGCTCGATCGATGT	72.1	33	27-Sep-17	1316106
MvSd07159R	GGATCCTTACGGACTTCTGCGTCGTCCTCGGGC	75.8	33	27-Sep-17	1316106
MvSp00831F	GAATTCGACCCACCTGATTTGCGTCGCTTGACA	70.9	33	27-Sep-17	1316106
MvSp00831R	GGATCCTCATAGGATTACTGCACCTAAAGCCAA	68.4	33	27-Sep-17	1316106
MvSp01648F	GAATTCGGGGTCGTTCCCAAGGAGGCTCCTGTT	73.4	33	27-Sep-17	1316106
MvSp01648R	GGATCCTCAGAGCTTGACGTTGTCG	68.4	25	27-Sep-17	1316106
MvSp10116FW	GAATTCCGGCGGTCCGAGTCACATGCACTTGTC	73.4	33	27-Sep-17	1316106
MvSp10116R	GGATCCTCACCTCGAAGCTCCCGTCCACAGCAT	74.6	33	27-Sep-17	1316106
MvSp16237F	GAATTCTCCGAGCCTCGCACCGAAGGACACTGT	73.4	33	27-Sep-17	1316106
MvSp16237R	GGATCCACATTGGGATCACGAGCATGGTCGTGC	73.4	33	27-Sep-17	1316106
BsrGIFw	GCATGTACAGGCGATGGCGATCAGTATCACCTTGCT	72.6	36	27-Sep-17	1316106
MVLG_02245F	GAATTCGCACCGCTGGCTTCGGAGCAAATCTCG	73.4	33	15-Sep-17	1310443
MVLG_02245R	GGATCCGGGTGATGCTGTGGTAGTATGATAGCG	72.1	33	15-Sep-17	1310443

MVLG_05398F	GAATTCATTGCCGTTGGTGAACCCA	72.1	25	15-Sep-17	1310443
MVLG_05398R	GGATCCGGCGCTGGGATCCGGCGTSAGTGGGAAT	77.1	34	15-Sep-17	1310443
MvSd01662F	GAATTCCCTTCCCGCCGAAGACCACAGCGACAC	74.6	33	15-Sep-17	1310443
MvSd01662R	GGATCCTCATACATGGCCGTGTCCCTGGTAGGG	74.6	33	15-Sep-17	1310443
MvSd02874F	GAATTCACACCAGACCCTATGTCTGGTAGCAGC	70.9	33	15-Sep-17	1310443
MvSd02874R	GGATCCTCAGCAATGAAACTTGGAATGACAACC	68.4	33	15-Sep-17	1310443
MvSd04324F	GAATTCGCACCGGCACCCCTAATCGACGGCGGC	77.1	33	15-Sep-17	1310443
MvSd04324R	GGATCCTTAGAATCCCACGGGAAAAATAGTGGT	68.4	33	15-Sep-17	1310443
MvSd662FW	GCAGTTAATTAAATGAGATTGCTCTTTGCTATCACC	65.8	36	10-Oct-17	1321828
MvSd662RV	GCAGATATCGATCTCAGGGTCCTGCCCTATCTCAAT	71.5	36	10-Oct-17	1321828
MvSd874FW	GCAGTTAATTAAATGAAGCTCCTCGCGATTGCCGTC	70.4	36	10-Oct-17	1321828
MvSd874RV	GCAGATATCGATCAAGACGTGATCACGTGCGACACT	71.5	36	10-Oct-17	1321828
MvSd324FW	GCAGTTAATTAAATGAAGCTGTCCACCTTGATCCTC	68.1	36	10-Oct-17	1321828
MvSd324RV	GCAGATATCGCAAAAACGGTAATGCCGAATTAGACT	68.1	36	10-Oct-17	1321828
MvSd159FW	GCAGTTAATTAAATGCTTTTCCCCATCGTGTGCTTC	68.1	36	10-Oct-17	1321828
MvSd159RV	GCAGATATCGCTTTCCACTGCCTCGCATCTCCACCT	73.8	36	10-Oct-17	1321828
MvSp16237YSTFW	GCAGAATTCATGTTCGTCTTGCTGCTCACCACCCCAGATTGCGCGTTGGT TGTCTCGCTG	79.1	60	17-Oct-17	1325248
MvSp16237YSTRV	GCAGCGGCCGCTGCCAGCGAGACAACCAACGCGCAATCTGGGGTGGTGAG CAGCAAGACG	83.9	60	17-Oct-17	1325248
MvSd01662F	TGCATGACGTAGATGAGATTGCTCTTTGCTATCACC	69.2	36	01-Nov-17	1332464
MvSd02874F	TGCATGACGTAGATGAAGCTCCTCGCGATTGCCGTC	73.8	36	01-Nov-17	1332464
MvSd043245F	TGCATGACGTAGATGAAGCTGTCCACCTTGATCCTC	71.5	36	01-Nov-17	1332464
MvSd07159F	TGCATGACGTAGATGCTTTTCCCCAATCGTGTGCTC	71.5	36	01-Nov-17	1332464
MvSp00831F	TGCATGACGTAGATGGTGTCCAAGCTGCTGGGCGCA	74.9	36	01-Nov-17	1332464
MvSp01648F	TGCATGACGTAGATGCGCTTCTCCATGCTCATCCCC	73.8	36	01-Nov-17	1332464
MvSp10116F	TGCATGACGTAGATGCTCCCCAGGTTATCTCGCCAT	72.6	36	01-Nov-17	1332464
MVLG2523PromF	GCATTAATTAAGTGCAACTGTGAGCCCCGACTCCGT	71.5	36	01-Nov-17	1332464
MVLG2523PromR	CTACGTCATGCACTGACGATGGGATCGAGA	70.1	30	01-Nov-17	1332464
MvSp831R	GCAGATATCACGCGGTCGAACCCGAACCCGAAGCTA	58.3	36	11-Dec-17	1349588
MvSp1648R	GCAGATATCGCGGCACCCACTACCGGGTCTCTCTCT	61.1	36	11-Dec-17	1349588
MvSp10116R	GCAGATATCGCAGGTGGGGGACGGAGAAATTGCAGC	58.3	36	11-Dec-17	1349588
MVLG_05703Lf	GACGGGAGACAAACAAGCAT	60.4	20	15-Jan-18	1361290
MVLG_05703Rt	CTTCCTCGGGTGAATAGACG	62.4	20	15-Jan-18	1361290
MVLG_00961Lf	TGGTTTGACTTTGGCTGTTG	58.4	20	15-Jan-18	1361290
MVLG_00961Rt	CATTTGGTGGAAGATGAGCA	58.4	20	15-Jan-18	1361290
MVLG_06117Lf	ATGCCAAGGTTGTGCTCTTT	58.4	20	15-Jan-18	1361290
MVLG_06117Rt	ACTCATTCCAGGCCATGTTG	60.4	20	15-Jan-18	1361290
MVLG_03216Lf	GGCGAAACTCCTTTTGATGA	58.4	20	15-Jan-18	1361290
MVLG_03216Rt	TCGTGGTGGAAGTTGTGGTA	60.4	20	15-Jan-18	1361290
MVLG_05589Lf	ATCCCTGCTCTCTTCTTCAA	58.4	20	15-Jan-18	1361290
MVLG_05589Rt	GTCGACGTATTCGAGGTGGT	62.4	20	15-Jan-18	1361290

163

Cas9/gRNAFw	TTAATTAACCAATTCCTGTTCCCCCTGATTGGCCTA	68.1	36	11-Feb-18	1375008
Cas9/gRNARv	GATACTTCATGTTTGACAGCTTATCATCGGATCTAG	67.0	36	11-Feb-18	1375008
sgRNAFw	GCATTAATTAACCAATTCCTGTTCCCCCTGATTGGC	69.2	36	11-Feb-18	1375008
sgRNARv	TACTTTGTCACCTGCAGGTACAAGAAAGCTGGGTCT	70.4	36	11-Feb-18	1375008
5703-PFw	TGTACCTGCAGGTGACAAAGTAGTAMAGTGGAAGCGT	70.4	37	11-Feb-18	1375008
5703-PRv	GCATTGTACAACCGAGGCGCGAGCTTGCTTCCGAAA	73.8	36	11-Feb-18	1375008
HF-Cas9Fw	GCCTCGGTTGTACAATGCCGCCTAAGAAGAAACGCA	72.6	36	11-Feb-18	1375008
HF-Cas9Rv	GCAGATATCTCATGTTTGACAGCTTATCATCGGATC	68.1	36	11-Feb-18	1375008
5703-P_ALTFw	TGTACCTGCAGGTCCTGCCAGTGCTTCCTC	72.8	30	23-Feb-18	1381641
MVLG_p1A1-PRFw	TTCGATCCCGCGTGGCGGGATCCGTTTCAGCAATTTACGGTTTTTAGAGCTAGAAATAGC	49.2	59	01-Mar-18	1384908
MVLG_p1A1-PRRv	GCTATTTCTAGCTCTAAAACCGTAAATTGCTGAAACGGATCCCGCCACGCGGGATCGAA	49.2	59	01-Mar-18	1384908
MvSp831Fw	GGAGAATTCGACCCACCTGATTTGCGTCGCTTGACA	52.8	36	01-Mar-18	1384908
MvSd4324Fw	GGAGAATTCGCACCGGCACCCCTAATCGACGGCGGC	66.7	36	01-Mar-18	1384908
MVLG_00933F	GCAGAATCCATGTTATCTCGTCGAAAGTGGTCC	69.6	33	13-Mar-18	1391073
MVLG_00933R	GCAGCGGCCGCCGGCGCGAACGGGAGGCAGCGC	84.5	33	13-Mar-18	1391073
MVLG_00934F	GCAGAATCCATGTACCAAGACGCAGCAGCGCCG	74.6	33	13-Mar-18	1391073
MVLG_00934R	GCAGCGGCCGCCCGGCTGCACCGAATGTGTGTT	78.3	33	13-Mar-18	1391073
MvSd1141Gg04324R	GCAGGATCCGGAAGGATGGGGGAGGAT	72.2	27	28-Mar-18	1398840
Cas9TrunkFw	GATCAAGAAAGGTATCCTCCAGACCGTCAA	67.4	30	04-Apr-18	1402535
Cas9TrunkRv	TTGACGGTCTGGAGGATACCTTTCTTGATC	67.4	30	04-Apr-18	1402535
GibsonOriAmpFw	GCTGTCAAACATGAGATATCTCGCAGCCACCCACAGTA	72.2	38	23-Apr-18	1412131
GibsonOriAmpRv	ACAGGAATTGGTTAATTAAAGGGGATAACGCAGGAAAGA	67.8	39	23-Apr-18	1412131
pMvCC9CutFw	GCAGCAGCAGCATTAATTAACCAATTCCTGTTCCCCCTGATTGGC	73.8	45	17-May-18	1425299
pMvCC9CutRv	GCAGCAGCAGCAGATATCTCATGTTTGACAGCTTATCATCGGATC	72.9	45	17-May-18	1425299
sgRNAFw	GCAGATATCCCAATTCCTGTTCCCCCTGATTGG	70.9	33	25-May-18	1430012
sgRNARv	GCATCTAGATACAAGAAAGCTGGGTCTAGTTTT	65.9	33	25-May-18	1430012
Cas9Fw	GCATTAATTAAGTGCAACTGTGAGCCCCGACTC	69.6	33	25-May-18	1430012
Cas9RvBam	GATGTCGAGTCCGATGCTGT	62.4	20	06-Jul-18	1451819
AmpORI Fw	TTAATTAATGCTGCTGCTGATGAGTATTCAACATTTCCG	66.8	39	09-Oct-18	1500901
5703GibFw	AGCTTTCTTGTACCTGCAGGTCCTGCCAGTGCTTCCTCGAAAAAACGC	52.1	48	05-Sep-18	1483894
Pol III Gib Fw	TTGAAAAAGTGGCACCGAGTCGGTGCTAAAAACTAGACCCAGCTTTCTTGTACCTGCAGG	48.3	60	05-Sep-18	1483894
Pol III Gib Rv	CCTGCAGGTACAAGAAAGCTGGGTCCCTGCAGGTACAAGAAAGCTGGGTCACTTTTTCAA	48.3	60	05-Sep-18	1483894
Msd13216F	TTAATTAATTCCGGCGATTCGGTCCGGGCGAAAAACGT	71.2	38	05-Dec-18	1529639
Msd13216R	GATATCGGATGGAAAGCGTGGAGAGGAGGAGAGTGG	73.8	36	05-Dec-18	1529639
Msd12336F	TTAATTAAATGAACCATCGCTGTCGCACCATGGTCTGT	69.0	38	05-Dec-18	1529639
Msd12336R	TCTAGACTGAGCCCGAACCTGAGCCCGAACCTGAAC	74.9	36	05-Dec-18	1529639
Msd10910F	TTAATTAAGGCAATGTGCATCAGTCTCCACGGGGCGGA	72.2	38	05-Dec-18	1529639
Msd10910R	TCTAGACTTGTCTTTTCGTGCGTACALTGCCGTCGGC	72.6	37	05-Dec-18	1529639
Msd09295F	TTAATTAAGTTCGTCTCTTGGACGTAGACGGCCGAATG	70.1	38	05-Dec-18	1529639
Msd09295R	TCTAGACTAGCGTGTCCACTTGAGTTTGGTCCCAGA	71.5	36	05-Dec-18	1529639

Overlap-HYG-EcoRV	CGGGTACCGAGCTCGATATCTCATGTTTGACAGCTTATCATCGGATCT	74.1	48	16-Jan-19	1545605
5585_up_flFw	TTACCTAAAATTTGCAACTATGAT	54.3	24	20-Mar-19	1577478
5585_up_flRv	ACAGTTGCACTCGGTGGGTGGGGAGCTTTT	71.5	30	20-Mar-19	1577478
MVLG_2523_Fw	CACCCACCGAGTGCAACTGTGAGCCCCGAC	75.6	30	20-Mar-19	1577478
MVLG_Rv	TGCTCAGCATCTACGTCATGCACTGACGAT	68.7	30	20-Mar-19	1577478
CFP_Fw	CATGACGTAGATGCTGAGCAAGGGCGAGGA	71.5	30	20-Mar-19	1577478
CFP_Rv	GCGGCAACCATTACTTGTACAGCTCGTCCA	70.1	30	20-Mar-19	1577478
dwn_Fw	GTACAAGTAATGGTTGCCGCAGCAGAAAGT	67.4	30	20-Mar-19	1577478
dwn_Rv	CCTTGTAGTTGGGGTTGCATTTGA	62.9	24	20-Mar-19	1577478
5730P-Cas9 Fw	CACCCACCGATCCTGCCAGTGCTTCCTCGA	74.2	30	20-Mar-19	1577478
5730P-Cas9 Rv	GCGGCAACCACTTCTTCTTCTTGGCCTGTC	71.5	30	20-Mar-19	1577478
Tel5703PCas9Fw	CAGTGCACGAGCAATCGACCGTGCACACTAGACCCAGCTTTCTTGTAGGTA	76.8	51	20-Mar-19	1577478
Tel5703PCas9Rv	ACACTAGACCCAGCTTTCTTGTAGGACCCAGGGTITTCCCAGTCACG	79.0	47	20-Mar-19	1577478
Alt_dwn_Rv_1	TCTTCATCTGCACCAAGTCG	60.4	20	16-Apr-19	1591255
Alt_dwn_Rv_2	CCTTGTAGTTGGGGTTGCAT	60.4	20	16-Apr-19	1591255
tRNAgRNAcasHF1F	GGTACCCCAATTCCTGTTCCCCCTGATTGG	56.7	30	26-Apr-19	1596432
tRNAgRNAcasHF_1F	TCTAGAAAAAAAGCACCGACTCGGTGCCACT	48.4	31	26-Apr-19	1596432
80mer Fw	GTACAAGTAACGATGGTTGCCGCAGCAGAAAGTTCCGCACACCAACGGAC GGCCCTCGTGATTGGCGCCGGCGTCGGAGG	62.5	80	06-May-19	1600822
80mer Rv	CCTCCGACGCCGGCGCCAATCACGAGGGCCGTCCGTTGGTGTGCGGAACT TTCTGCTGCGGCAACCATCGTTACTTGTAC	62.5	80	06-May-19	1600822
Dwn Flank Rv	CCTCCGACGCCGGCGCCAAT	75.0	20	06-May-19	1600822
U6_ovlp Fw	GATTACGAATTCTTAATTAACGACCAGAGAGAGAGGCAGA	68.8	40	23-Aug-19	1644769
U6_ovlp Rv	TCGTGGCTCGTCAAATCACGTTGTAGAATGGAATTTTG	69.0	38	23-Aug-19	1644769
Tar/Scaf Fw	CAAAATTCCATTCTACAACGTGATTTGACGAGCCACGA	69.0	38	23-Aug-19	1644769
Tar/Scaf Rv	CGGGTACCGAGCTCGATATCAAAAAAGCACCGACTCGGTG	75.0	40	23-Aug-19	1644769
tRNA/sgRNA Fw	GGCCGTCGTTTTACTCTAGACCAATTCCTGTTCCCCCTGA	73.9	40	30-Aug-19	1648403
tRNA/sgRNA Rv	GGCGCGCCCCTAGGAGGCCTAAAAAAGCACCGACTCGGTG	79.0	40	30-Aug-19	1648403
U6/sgRNA	GGCCGTCGTTTTACTCTAGATAATACGTTCGTTCCGATGT	70.8	40	30-Aug-19	1648403
pGA2pGB Fw	TGGCCATGGAGGCCGAATTCTACCCATACGACGTACCAGAT	74.7	41	09-Oct-19	1665031
pGA2pGB Rv	CGCTGCAGGTCGACGGATCCTGCACGATGCACAGTTGAAG	77.0	40	09-Oct-19	1665031
5703-P Fw	GATTACGAATTCTTAATTAAAAAAAAAACGCAGCAACACATT	63.7	40	09-Oct-19	1665031
5703-P Rv	TTCGCACTTGTGAGGTCATCGGCATATGTGTCGTGAAGG	73.1	39	09-Oct-19	1665031
2245 Fw	CCTTCACGACACATATGCCGATGACCTCACAAGTGCGAA	73.1	39	09-Oct-19	1665031
2245 Rv	ACCATAGAGCCGCCAGAGCCGCCCGGAACAACCCTTGTGGTGG	80.0	43	09-Oct-19	1665031
EGFP Fw	TTCCGGGCGGCTCTGGCGGCTCTATGGTGAGCAAGGGCGAGGA	80.9	43	09-Oct-19	1665031
EGFP Rv	CTCTGCAGGTCGACTCTAGAGATCTAGTAACATAGATGAC	70.8	40	09-Oct-19	1665031
Cas9F	AGCGAGTGATCTTGGCTGAT	60.4	20	30-Oct-19	1673975
Cas9R	CGATGCTTGTTGTAGGCAGA	60.4	20	30-Oct-19	1673975
tRNAsgRNAaltF	CGAGCCACGAGTTTTAGAGC	62.4	20	30-Oct-19	1673975

tRNAsgRNAaltR	CGGACTAGCCTTATTTTAACTT	57.1	22	30-Oct-19	1673975
MvSl1693MvSdF	CGATTTCACGAAGCAGAACA	58.4	20	30-Oct-19	1673975
MvSl1693MvSdR	GTCGAGGGGTGACGAAAGTC	64.5	20	30-Oct-19	1673975
MvSd9295MvSlaltF	CCAGCATACTGTTGTGTCTGC	62.6	21	30-Oct-19	1673975
MvSd9295MvSlaltR	CCAGCATACTGTTGTGTCTGC	58.4	21	30-Oct-19	1673975
MVLG2245GFPF	TATATCATGGCCGACAAGCA	58.4	20	30-Oct-19	1673975
MVLG2245GFPR	GTTGTGGCGGATCTTGAAGT	60.4	20	30-Oct-19	1673975
MVLG5585SeqFw	CTCCGATAGCGACTCACCTC	64.5	20	08-Nov-19	1678214
MVLG5585SeqRv	CTCCGATAGCGACTCACCTC	64.5	20	08-Nov-19	1678214
MVLG2245SeqFw	GGCACTGGTAGGTATCTTCGAT	62.7	22	08-Nov-19	1678214
MVLG2245SeqRv	AGAGCGAGCAACAACCTCAT	60.4	20	08-Nov-19	1678214
MvSl1693SeqFw	AGAGCGAGCAACAACCTCAT	60.8	20	08-Nov-19	1678214
MvSl1693SeqRv	GAGATGTTCTGCTTCGTGAAATC	61.0	23	08-Nov-19	1678214
NewTargetRNAPFw	AGCGTCGTTITCCCGCCATA	62.4	20	08-Nov-19	1678214
NewTargetRNAPRv	AGCGTCGTTITCCCGCCATA	70.6	20	08-Nov-19	1678214
NewTargetScafFw	AGCGTCGTTITCCCGCCATA	54.2	20	08-Nov-19	1678214
NewTargetScafRv	GCATGCGGCGCGCCCCTAGG	72.7	20	08-Nov-19	1678214
MVLG2245TarOligoF	GTTCGATCCCGCGTGGCGGGACTTGCATTCGCACTTGTGGTTTTAGAGCTAGAAATAGC	78.0	59	08-Nov-19	1678217
MVLG2245TarOligoR	GCTATTTCTAGCTCTAAAACCACAAGTGCGAATGCAAGTCCCGCCACGCG	78.0	73	08-Nov-19	1678217
MvSl1693TarOligoF	TGCGAATGCAAGTCCCGCCACGC GTTCGATCCCGCGTGGCGGGCTGCGTTCGATCCCGCGTGGCGGGCTGCTA GAAATAGC	78.7	58	08-Nov-19	1678217
MvSl1693TarOligoR	GCTATTTCTAGCTCTAAAACAATGCGACCACAAGAGCAGCCCGCCACGCG GGATCGAAC	78.7	59	08-Nov-19	1678217
MVLG2245ScafF	ACTTGCATTCGCACTTGTGGTTTTAGAGCTAGAAATAGC	68.9	39	08-Nov-19	1678217
MVLG2245tRNAPR	CACAAGTGCGAATGCAAGTCCCGCCACGCGGGATCGAAC	77.3	39	08-Nov-19	1678217
U6_ovlp Fw	GGCCGTCGTTTTACTCTAGACGACCAGAGAGAGAGGCAGA	75.0	40	20-Nov-19	1682701
U6_ovlp Rv	CTCGTGGCTCGTCAAATCACGTTGTAGAATGGAATTTTGA	69.8	40	20-Nov-19	1682701
Tar/Scaf Fw	TCAAAATTCCATTCTACAACGTGATTTGACGAGCCACGAG	69.8	40	20-Nov-19	1682701
Tar/Scaf Rv	GGCGCGCCCCTAGGAGGCCTAAAAAAGCACCGACTCGGTG	79.0	40	20-Nov-19	1682701
MVLG5703ExpressF	CTCAATCACCCCTCGCTTC	62.3	19	05-Dec-19	1688287
MVLG5703ExpressR	CTGTTGCCTCATTTTGTCGT	58.4	20	05-Dec-19	1688287
t-g Fw	GCTTTGGAAGAGCATCAGAC	60.4	20	09-Dec-19	1689480
t-g Rv	CTAAAACTCGTGGCTCGTCA	60.4	20	09-Dec-19	1689480
g Fw	GTGATTTGACGAGCCACGAG	62.4	20	09-Dec-19	1689480
g Rv	CGGACTAGCCTTATTTTAACTTGC	61.2	24	09-Dec-19	1689480
1693 qrt Fw	AACGGTGAAGAAGGAGCAAA	58.4	20	09-Dec-19	1689480
1693 qrt Rv	TCGTGAGACGCCGAAGTAAA	60.4	20	09-Dec-19	1689480
MVLG_05589-P Fw	GATTACGAATTCTTAATTAACGCCAGTGTGCTGGAATTCG	68.8	40	12-Dec-19	1691181
MVLG_05589-P Rv	TTGCGTTTCTTCTTAGGCATTGGGCGAAGAGAGGATGCGA	72.9	40	12-Dec-19	1691181
Cas9-Frag Fw	TCGCATCCTCTCTTCGCCCAATGCCTAAGAAGAAACGCAA	72.9	40	12-Dec-19	1691181
Cas9-Frag Rv	CTGCGTGTATAACGTCGACGTGCCGTGCGTTTGAGGCGTG	77.0	40	12-Dec-19	1691181

MVLG5589PRv2245	ATTCGCACTTGTGAGGTCATTGGGCGAAGAGAGGATGCGA	73.9	40	12-Dec-19	1691181
MVLG_02245 Fw	TCGCATCCTCTCTTCGCCCAATGACCTCACAAGTGCGAAT	73.9	40	12-Dec-19	1691181
Short-U6 Fw	tctagaCGACCAGAGAGAGAGGCAGA	67.7	26	10-Jan-20	1698963
Short-U6 Rv	GTTGTAGAATGGAATTTTGA	52.2	20	10-Jan-20	1698963
Short-U6Tar Fw	GTGATTTGACGAGCCACGAG	62.4	20	10-Jan-20	1698963
Short-U6Tar Rv	aggcctAAAAAAGCACCGACTCGGTG	67.7	26	10-Jan-20	1698963
Cas9-1 Fw	CCTCTCTCGGCACCTATCAC	64.5	20	10-Jan-20	1698963
Cas9-1 Rv	TCCTCGTTCTCTTCGTTGTC	60.4	20	10-Jan-20	1698963
Cas9-2 Fw	CCTCTCTCGGCACCTATCAC	64.5	20	10-Jan-20	1698963
Cas9-2 Rv	ATCCTCGTTCTCTTCGTTGTC	60.6	21	10-Jan-20	1698963
Cas9-3 Fw	TGCTCACCCTCACTTTGTTC	60.4	20	10-Jan-20	1698963
Cas9-3 Rv	TGCTTCATCACTTTGTCGTC	58.4	20	10-Jan-20	1698963
4324-P Fw	gattacgaattcttaattaaCACGGTGCAGCATGACGAAG	68.8	40	10-Jan-20	1698963
4324-P Rv	GCGACCACAAGAGCAGCATTTTCGCCCACACATCACTCAA	73.9	40	10-Jan-20	1698963
MvSl-1064_01693Fw	TTGAGTGATGTGTGGGCGAAAATGCTGCTCTTGTGGTCGC	73.9	40	10-Jan-20	1698963
MvSl-1064_01693Rv	ctctgcaggtcgactctagaTCAATTGTTCTTGATGGTAA	69.8	40	10-Jan-20	1698963
Whole U6 Fw	GGCCGTCGTTTTACTCTAGATAATACGTTCGTTCCGATGT	70.8	40	20-Jan-20	1702482
Whole U6 Rv	CTCGTGGCTCGTCAAATCACGTTGTAGAATGGAATTTTGA	69.8	40	20-Jan-20	1702482
WU6-sgRNA Fw	TCAAAATTCCATTCTACAACGTGATTTGACGAGCCACGAG	69.8	40	20-Jan-20	1702482
WU6-sgRNA Rv	TGCGGCGCGCCCCTAGGAGGCCTAAAAAAGCACCGACTCG	79	40	20-Jan-20	1702482
β-Car-Hyg Fw	TGATTTGACGAGCCACGATGGTTGCCGCAGCGCCAGTGTGCTGGAATTCGCCCTTGCTGT	80.5	60	28-Jan-20	1706129
β-Car-Hyg Rv	CCTAATCAAGAAGTCAACTATGCGCTCCCTCTATTCCTTTGCCCTCGGACGAGTGCTGGG	78.4	60	28-Jan-20	1706129
MvSl_01693Fw60	GCCTCGTACCATGCTGCTCTTGTGGTCGCACGCCAGTGTGCTGGAATTCGCCCTTGCTGT	81.2	60	28-Jan-20	1706129
MvSl_01693Rv60	CGCTCCTCTAATTTCAATTGTTCTTGATGGCTATTCCTTTGCCCTCGGACGAGTGCTGGG	77.1	60	28-Jan-20	1706129
β-CarFw	CCGCTCGGGTCCTCAGGCTCGGCCT	76.1	25	28-Jan-20	1706129
β-CarRv	TGTACAAAGGTCTGTATGCTTCTAA	59.7	25	28-Jan-20	1706129
MvSl_01693Fw25	CGTTACGCCTCTTGTCCTTTTGCTT	64.6	25	28-Jan-20	1706129
MvSl_01693Rv25	CTTGTTGTGATGTATAAAAGATGTC	58	25	28-Jan-20	1706129
MvSl_01693 1kbFw	CGCCAACTTGGTTTCACTATGCGCC	67.9	25	28-Jan-20	1706129
MvSl_01693 1kbRv	GAAGGGGCGGGCAGGACAATCAAGA	69.5	25	28-Jan-20	1706129
1693 1kb_OvlFw	ACCATGATTACGAATTCTTACGCCAACTTGGTTTCACTATGCGCC	72	45	28-Jan-20	1706129
1693 1kb_OvlRv	CCTCTGCAGGTCGACTCTAGGAAGGGGCGGGCAGGACAATCAAGA	78.4	45	28-Jan-20	1706129
FixMVLG05589PFw	GATTACGAATTCTTAATTAAGTTTAAACTGAAGGCGGGAA	65.7	40	28-Jan-20	1706129
2245-GFPnoSP FW	TCGCATCCTCTCTTCGCCCAGCACCGCTGGCTTCGGAGCA	79	40	17-Feb-20	1715090
2245-GFPnoSP RV	CTCTGCAGGTCGACTCTAGAGATCTAGTAACATAGATGAC	70.8	40	17-Feb-20	1715090
New2245-GFPnoSP FW	TCGCATCCTCTCTTCGCCCAATGGCACCGCTGGCTTCGGA	78	40	25-Feb-20	1718964
New2245-GFPnoSP RV	TCCGAAGCCAGCGGTGCCATTGGGCGAAGAGAGGATGCGA	78	40	25-Feb-20	1718964

Freezer Stocks:

Freezer Box #1 Listed Index

Grid	Plasmid	Cell Type
A1	PMV_HYG	In DH5α
A2	BP 6175 3-1	In DH5α
A3	BP 859 C9	In DH5α
A4	MvSp_01648 In pYST0	In Yeast
A5	MVLG_5378 Signal Pep. In pYST0	In Yeast
A6	MvSd_04324-P In TOPO	In Yeast
A7	MvSd_02874-P In TOPO	In Yeast
A8	MvSd_07159-P In TOPO	In Yeast
A9	MvSd_01662 In pYST0	In Yeast
B1	pMv_Hyg	In DH5α
B2	Lu53	Lu53
B3	DH5α	DH5α
B4	MvSp_01648 In pYST0	In Yeast
B5	MVLG_02245 In pYST0 In Frame	In SEY
B6	MvSp A2 Stock	MvSp A2
B7	MvSp A2 Stock	MvSp A2
B8	MvSp A2 Stock	MvSp A2
B9	MvSp A2 Stock	MvSp A2
C1	EHA105	EHA105
C2	BsrGI-Hyg-BsrGI In TOPO	In DH5α
C3	*Bacilis subtilis*	*B. subtilis*
C4	-	-
C5	MvSp_16237 nIn pYST0 In Frame	In SEY
C6	MvSp A1 Stock	MvSp A1
C7	MvSp A1 Stock	MvSp A1
C8	MvSp A1 Stock	MvSp A1
C9	MvSp A1 Stock	MvSp A1
D1	pMS8 2-1	In DH5α
D2	pMS8 2-2	In DH5α
D3	Proto3	In DH5α
D4	MvSd_07159 In pYST0 In Frame	In SEY
D5	MvSd_01662 In pYST0 In Frame	In SEY
D6	MvSd A2 Stock	MvSd A2
D7	MvSd A2 Stock	MvSd A2
D8	MvSd A2 Stock	MvSd A2
D9	MvSd A2 Stock	MvSd A2
E1	pMS10 2-1	In DH5α
E2	pMS10 2-2	In DH5α
E3	Proto2 E1 target+Mv Promoter for Cas9	In DH5α

E4	MvSp_01648 In pYST0	In DH5α
E5	MvSp_01648 In pYST0	In DH5α
E6	MvSd A1 Stock	MvSd A1
E7	MvSd A1 Stock	MvSd A1
E8	MvSd A1 Stock	MvSd A1
E9	MvSd A1 Stock	MvSd A1
F1	pMs 73 A	In DH5α
F2	pMs 73 B	In DH5α
F3	Proto 1 pMS8+Frag	In DH5α
F4	MVLG_02245 In pYST0 In Frame	In SEY
F5	MVLG_05398 In pYST0 In Frame	In SEY
F6	MvSl p1A2 Stock	MvSl p1A2
F7	MvSl p1A2 Stock	MvSl p1A2
F8	MvSl p1A2 Stock	MvSl p1A2
F9	MvSl p1A2 Stock	MvSl p1A2
G1	Cas9 HF	In DH5α
G2	CRISPR Frag In TOPO	In DH5α
G3	MVLG_02523 Promoter In TOPO	In DH5α
G4	MvSp_00831 In pYST0 In Frame	In SEY
G5	MvSd_02874 In pYST0 In Frame	In SEY
G6	MvSl p1A1 Stock	MvSl p1A1
G7	MvSl p1A1 Stock	MvSl p1A1
G8	MvSl p1A1 Stock	MvSl p1A1
G9	MvSl p1A1 Stock	MvSl p1A1
H1	A2 STC	
H2	A2 STC	
H3	A2 STC	
H4	A2 STC	
H5	A2 STC	
H6	A2 STC	
H7	A2 STC	
H8	A2 STC	
H9	A2 STC	
I1	A2 STC	
I2	A2 STC	
I3	A2 STC	
I4	A2 STC	
I5	A2 STC	
I6	A2 STC	
I7	A2 STC	
I8	A2 STC	
I9	A2 STC	

Freezer Box #2 Listed Index

Grid	Plasmid	Cell Type
A1	Y2H 2245 In TOPO	In DH5α
A2	MvSd_09295 In TOPO	In DH5α
A3	MvSd_10910 In TOPO	In DH5α
A4	MvSd_12336 In TOPO	In DH5α
A5	MvSd_13216 In TOPO	In DH5α
A6	Y2H Control pGADT7 T7	In Y187
A7	Y2H Control pGBKT7 53	in Y187
A8	Nelson's Signal Peptide 5122	In Yeast
A9	Nelson's pYST0 Vector	In Yeast
B1	pGADT7 MVLG_02245	In DH5α
B2	MvSd_09295 In pMvHyg	In DH5α
B3	pGBKT7_Q1	In DH5α
B4	pGBKT7_Q3	In DH5α
B5	pGBKT7_Q5	In DH5α
B6	pGBKT7_Q6	In DH5α
B7	pGADT7_Q1	In Y187
B8	pGADT7_Q3	In Y187
B9	pGBKT7_2245	In AH109
C1	pGBKT7 MVLG_02245	In DH5α
C2	MvSd_09295 In pMvHyg	In DH5α
C3	pGB_Q1	in AH109
C4	pGB_Q3	in AH109
C5	pGB_Q5	in AH109
C6	pGB_Q6	in AH109
C7	pGADT7_Q5	In Y187
C8	pGADT7_Q6	In Y187
C9	pGB_2245 X pGA_Q6	Yeast Diploids
D1	-	-
D2	MvSd_09295 In pMvHyg	In DH5α
D3	pGB_Q1 X pGA_2245	Yeast Diploids
D4	pGB_Q3 X pGA_2245	Yeast Diploids
D5	pGB_Q5 X pGA_2245	Yeast Diploids
D6	pGB_Q6 X pGA_2245	Yeast Diploids
D7	pGB_2245 X pGA_Q1	Yeast Diploids
D8	pGB_2245 X pGA_Q3	Yeast Diploids
D9	pGB_2245 X pGA_Q5	Yeast Diploids
E1	Infected *S. latifolia* Prey Library	In Y187
E2	Infected *S. latifolia* Prey Library	In Y187
E3	Infected *S. latifolia* Prey Library	In Y187
E4	pMvHyg_tRNA_CRISPR	In DH5α
E5	T#140 CRISPR-Trans	MvSl p1A1
E6	pMvHyg_MvSl_01693	MvSd A1
E7	pMvHyg_MvSl_01693	MvSd A2
E8	pMvHyg_No-SP-MVLG_02245-GFP	MvSl p1A1
E9	pMvHyg_No-SP-MVLG_02245-GFP	MvSl p1A2
F1	pMvHyg_Cas9	In DH5α
F2	Infected *S. latifolia* Prey Library	In Y187
F3	pMvHyg_CRISPR	In DH5α
F4	pMvHyg_sgRNA T	In DH5α
F5	pMvHyg_sgRNA	In EHA105
F6	pBG_53 Control	In AH109
F7	-	-
F8	-	-
F9	pMvHyg_CRISPR Transformant #2	in p1A1

G1	pCFP	In DH5α
G2	*Burkholderia Cenocpacia* J2315	*B. Cenocpacia*
G3	pMvHyg_6175	In DH5α
G4	pMvHyg_sgRNA B	In DH5α
G5	Rosetta *E. coli*	Rosetta *E. coli*
G6	pGA_T7 Control	In Y187
G7	-	-
G8	-	-
G9	pMvHyg_CRISPR Transformant #3	in p1A1
H1	pET_Cas9 in ADD GENE Bacteria	ADD GENE Bacteria
H2	Star One Shot Cells	BL21 (DE3)
H3	pET_Cas9	BL21 (DE3)
H4	pMvHyg_sgRNA In TOPO	In DH5α
H5	pET_Cas9	In BL21 (DE3)
H6	pGBKT7	In DH5α
H7	-	-
H8	-	-
H9	pMvHyg_CRISPR Transformant #4	in p1A1
I1	MVLG_SAD F1 Gen Strain #1	MvSl p1A1
I2	MVLG_SAD F1 Gen Strain #2	MvSl p1A2
I3	MVLG_SAD F1 Gen Strain #1	MvSl p1A2
I4	MVLG_SAD F1 Gen Strain #2	MvSl p1A2
I5	pMvHyg_SAD	In DH5α
I6	pGADT7	In DH5α
I7	pMvHyg_MvSl-01693	In DH5α
I8	pMvHyg_MvSl-01693	In EHA105
I9	pMvHyg_CRISPR Transformant #5	in p1A1

Protocols:

ATMT of *Microbotryum*

Mating the Cells

- Prepare 1 LB + Kan + Spec plate, 1 YPD – 10% + Amp plate, 1 IM plate, 5 mL of IM broth , 5 mL YPD – 10% broth, and 10 YPD – 10% + Hyg + Cef plates per transformation.

  [See next page for media prep]

- Streak out *Agrobacterium* (EHA105) with your construct on LB + Kan + Spec and grow at 28°C for 2 days.
  - *(EHA105 uses a binary plasmid system and must be plated on Kan & Spec)
- Streak out *Microbotryum* on YPD – 10% + Amp and grow at 28°C for 2 days.
- Scrape a loopful of cells into 1 mL of IM broth (less is more).
- Vortex until completely suspended.
- Dilute 10X in IM broth and check OD.
- Dilute original stocks to 1E8 cells each
  - D *Microbotryum* OD_{600} 1 = 3.4E7 cells/mL
  - *Agrobacterium* OD_{600} 1 = 8E8 cells/mL
- Mix the following volumes of 1E8 cells/mL to achieve the listed ratios of total cells:

Resulting Ratio of Total Cells	Volume of *Agrobacterium* cells (µL)	Volume of *Microbotryum* cells (µL)	Volume of IM (µL)
1E7 *M*: 1E7*Agro*	100	100	0
1E6 *M*: 1E7*Agro*	100	10	90

- Spot 200 µL of each mixture onto a different sides of an IM plate
  - *(Make sure surface is level so that the mixtures don't run and merge)
- Let plates sit at room temperature, lid up, for 2-3 days.

Selecting for Transformants
- Scrape each mass of cells into 2 mL of YPD – 10% broth resuspend.
- Spread 200 µL of suspension onto a YPD 10% + Hyg + Cef plate.
- Allow 2-3 weeks of growth for transformed *Microbotryum* colonies to appear

Media Prep

<u>LB + Kan + Spec Plates</u>
0.5%   Yeast Extract
1.0%   Tryptone
1.0%   Sodium Chloride
2.0%   Agar
  >Add diH$_2$O to the appropriate volume
  >Autoclave on liquid cycle @ 121°C for 15min
  >Let cool and add 50 μg/mL Kan and 150 μg/mL Spec.

<u>YPD – 10% + Amp Plates</u>
1.0%   Yeast Extract   10% Dextrose
2.0%   Peptone   >Fill to 500mL with diH$_2$O
2.0%   Agar
  >Fill to 500mL with diH$_2$O
              >Autoclave on liquid cycle @ 121°C for 15min
              >Pour the dextrose into the YE/P/A flask without making bubbles
              >Let cool and add 50 μg/mL Ampicillin.
*(For broth, do not add agar)
*(For Hyg + Cef plates, add 150 μL/mL Hygromycin and 300 μL/mL Cefoxitin instead of Amp)

<u>IM Plates (250 mL)</u>
100 mL 2.5X MM Salts
0.225 g Dextrose
1.25 mLGlycerol
5.0 g   Agar
  >Add 135 mL diH$_2$O
  >Autoclave on liquid cycle @ 121°C for 15min
Dissolve 1.925 g MES in 12.5 mL diH$_2$O
Dissolve 4.75 mg Acetosyringone in 62.5 μL DMSO
  >Mix MES and Acetosyringone
  >Let cool and add MES/Aceto mixture to media

<u>IM Broth (50 mL)</u>
20 mL   2.5X MM Salts
45 mg   Dextrose
250 μL Glycerol
  >Add 27 mL diH$_2$O
  >Autoclave on liquid cycle @ 121°C for 15min
Dissolve 0.385 g MES in 2.5 mL diH$_2$O
Dissolve 0.95 mg Acetosyringone in 12.5 μL DMSO
          *(It is better to make this along with the mixture of the IM Plates and just 12.6 mL of the mixture)
  >Mix MES and Acetosyringone
  >Let cool and add MES/Aceto mixture to media

<u>2.5X MM Salts (1 L)</u>
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   6.2 mg
(NH$_4$)$_2$SO$_4$ 1.250 g

CRISPR Kit Transformation of Microbotryum Protoplasts

Preparing Cas9 Duplex

- Resuspend the crRNA and tracrRNA to 100 µM stock concentrations in the provided Nuclease-Free Duplex Buffer.

Normalized Amount Delivered (nmol)	Duplex Buffer Resuspension Volume (µL)
2	20
5	50
10	100
20	200
100	1000

- Store resuspended RNA oligos at -20°C
- For complete gene deletion, use two gRNAs, one that targets a PAM site at the 5' end and one that targets a PAM site at the 3' end.
- For each gRNA, prepare a 33 µM RNA duplex solution by mixing the crRNA and corresponding tracrRNA in equimolar concentration.

Component	Amount (µL)
100 µM Alt-R CRISPR-Cas9 crRNA	5
100 µM Alt-R CRISPR-Cas9 tracrRNA	5
Nuclease-Free Duplex Buffer	<u>5</u>
Total Volume	15

- Heat at 95°C for 5 minutes and then allow to cool to room temperature
    Note: The crNRA:tracrRNA guide complex can be used for 3 months with no loss in activity when stored at -20°C.
- Before use, thoroughly mix the stock Alt-R S.p. Cas9 Nuclease 3NLS by inverting the tube. Quick spin. Dilute Cas9 10X to a final concentration of 1 µg/µL (recommend using 1 µL of Cas9 and 9 µL of Cas9 working buffer, this will allow for 6 reactions).
- Combine the following:

Component		Amount (µL)
33 µM RNA duplex solution 1 (gRNA 1)		1.5
33 µM RNA duplex solution 2 (gRNA 2)		1.5
Cas9 nuclease (1 µg/µL)	1.5	
Cas9 Working Buffer		<u>22</u>
Total Volume		26.5

- Incubate at room temperature for 5 min to allow the assembly of the RNP complexes.

<u>Cas9 Working Buffer</u>:   800 µL
    20 mM HEPES (pH 7.5)   20 µL 1M HEPES (pH 7.5)
    150 mM KCl   150 µL 1M KCl
                                  > 630 µL Nuclease-Free Water

Transformation

- Retrieve protoplasts form -80°C storage and thaw on ice.
- Pipette 100 μL of STC solution into a round bottom 15 mL snap cap tube and place on ice.
- Add 7 μg of purified repair template (plasmid, cosmid, or linear construct). Do not add more than 20 μL of DNA.
- Add 26.5 μL of the RNP complex
- Add 200 μL of protoplasts using wide-bore tips, carefully and slowly pipetting the mixture to ensure even suspension without destroying the cells.
- Add 50 μL of 30% PEG solution. Mix carefully by gentle swirling or slowly pipetting up and down with wide-bore tips and incubate on ice for 50 min.
- Add 2 mL of 30% PEG solution, mixing carefully via gentle inversion. Incubate at room temp. for 20 minutes. Cells should clump together during this incubation, which can be verified microscopically. (During this step, add top agar to your regeneration plates)
- Add 2 mL of STC and mix thoroughly via gentle inversion of the tube.
- Pipette 500 μL of the transformation mixture to the top agar of the selection plates. Tilt plates to help spread the mixture evenly, but do not spread. (This will make up to 8 plates)
- Incubate at ~25°C until colonies form.
- Re-streak colonies onto YPD-10% 150 μg/ mL Hyg plates.

Materials

STC Solution: 10 mL
      1.2 M Sorbitol   2.18 g
      50 mM CaCl$_2$.2H$_2$O   500 μL 1 M solution
      10 mM Tris-HCL (pH 8.0)   100 μL 1 M solution

30% PEG Solution: 10 mL
      30% PEG 8000   3 g
      50 mM CaCl$_2$ 500 μL 1 M solution
      10 mM Tris-HCl (pH 8.0)   100 μL 1 M solution

Top Agar Selection Plates:
>Make 100 mL of bottom and 100 mL of top agar for 10 plates:
      YPD-10% Plates:
      1 g Yeast Extract
      2 g Peptone
      20 g Dextrose (Separate Flask)
      2 g Agar
       >Autoclave
       >Add 300 μg/mL Hyg-B to bottom agar

>Make bottom agar the same day you plan to do the transformations (storage lowers efficiency).
>Plate 10 mL of bottom agar, let solidify. During 20 minute incubation period, add 10 mL top agar.

Designing Primers for CRISPR Kit

Selecting Target

Targets for the Cas9 endonuclease should be 20 bp in length preceding an **NGG** PAM site. These targets should be unique in the genome, at least with regards to the last 10 bp before the PAM site. GGG PAM sites should also be avoided if possible.

Ex:

	Target		PAM
5'	GTGTGCGGAACTTTCTGCTG	3'	CGG

*Note: it does not matter in which direction the target is located as Cas9 cuts across both strands.

To increase likelihood of knockout, two targets should be designed (1 at the 5' end of the target gene and the other at the 3' end of the target gene). By using two targets, the target is excised completely from the genome.

When using the CRISPR Kit, a sgRNA is generated using hydrogen binding of two components, the crRNAs (the targets you design) and the tracrRNA (the scaffold that holds the crRNA to Cas9). These are generated automatically when you input your 20 bp target sequence into the program provided by the company to order your crRNAs. Because the tracrRNA and Cas9 are the same regardless of target, these are supplied separately and do not require adjustment.

Repair Template

In addition to your dual target sequences, a selectable marker with 60 bp of overlapping sequence to the genome immediately upstream for the 5' target and downstream of the 3' target should be generated. This can be accomplished by designing a 60 bp oligonucleotide primer for your insert with 20 bp for priming the sequence and 40 bp of overlap. While 40 bp of overlap may suffice for insertion in some systems, 60 bp is optimal. Furthermore, amplifying inserts from genomic DNA with such a large primer can be difficult. To increase your length of overlap, and to isolate inserts from genomic DNA, nested PCR is recommended.

To perform a nested PCR resulting in an amplified insert with the 60 bp corresponding overlap sequence, a series of 3 primers can be used (2 sets if the aforementioned 60 bp oligonucleotide successfully amplifies your target.

1) An initial 20 bp Forward and Reverse primer without overhangs can be used to amplify the target from genomic DNA. The resulting fragment should be run on a gel and gel purified for use as the template in step 2.
2) A second set of 60 bp oligonucleotides with 20 bp of primer to the fragment amplified in step 1 can be used to amplify the fragment, adding 40 bp of complementary sequence corresponding to the upstream of the 5' target and downstream of the 3' target.
3) A third set of 40 bp oligonucleotides with 20 bp of primer to the fragment amplified in step 2 can be used to amplify the fragment, adding an additional 20 bp of complementary sequence. This will yield an insert with 60 bp of complementary sequence to the boundaries of the excised target for the Cas9 kit.

Abdallah Et al., 2017
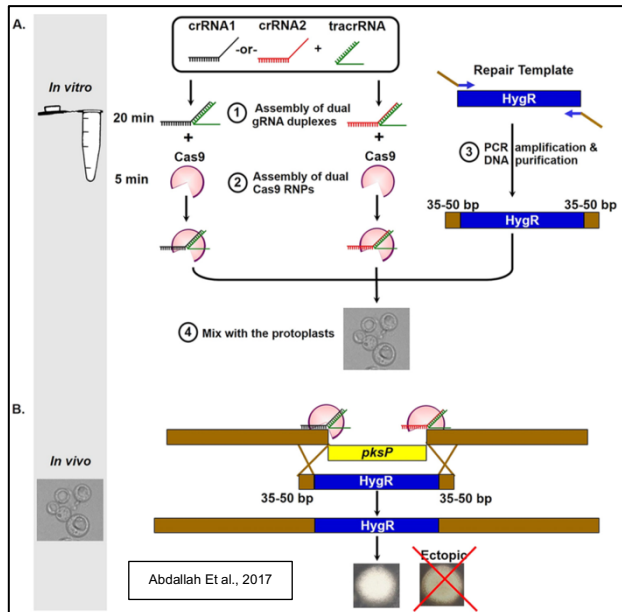
For the pilot study for use of the CRISPR Kit in *Microbotryum lychnidis-dioicae*, the 5589-Hyg resistance cassette from pMvHyg was amplified as the selectable insert for knockout of MVLG_05585.

*Note: because the flanking regions are different depending on your targets, the insert with complementary sequence must be generated from new primers every time a new target is used.

How to Order the Materials

Once you have designed your target sequences, the crRNAs can be ordered from the following site:
https://www.idtdna.com/site/order/oligoentry/index/crispr

In addition to your custom crRNAs, you will also need to order tracrRNA and the Cas9 endonuclease
*Note: the company sells multiple variants of the Cas9 endonuclease, including the single mutant nickase (Cas9n) and the double mutant Cas9 for CRISPRi (dCas9).

Primers for amplification of the selectable insert should be ordered through Eurofins.

Infecting Silene

Inoculating Seeds
- Grow A1 and A2 *Microbotryum* mating types separate on YPD-10% for 2 days.
- Inoculate 1 mL of water with a loop of cells separately from each mating type.
- Check optical density of mixtures and adjust to 1.0 A_{600nm}.
- Plate 40-50 seeds equal distance from each other on large 1.8% water agar plate.
- Mix equal volumes of each mating type mixture and spot ~50 μL onto each seed.
- Place plates face up in 12C° chamber (4C° will work if needed) for 48h.
- Move plates to growth chambers and spot the seeds with more water every 2-3 days when the water from the previous spotting is gone. Make sure plates are wrapped to prevent drying out.

Transplanting Sprouts
- Once the seeds begin to sprout and form the cotyledon, cut away the agar around the root and transplant into egg-shell crate using propagation mix as the soil. It is helpful to lightly pack the soil and "prime" it by adding a little water to the top to help it draw up water from the tray. Toothpicks can also be used to help prop sprouts up until they become established.
- Cover the tray with a lid to help keep humidity up.
- Water seeds from the bottom up using a tray tap water.

Transplanting Young Plants
- Once the sprouts are at the 4-8 leaves stage, transplant them into small tray pots using Sta-Green potting mix. When transplanting from an eggshell crate, first fill the small pots with soil and lightly pack. Prime the soil with water and let sit until there is no standing water (5-10 minutes). Poking holes can help this process along as packing the soil can make it slower to soak up the water. Once the soil is primed, create a plug for the entire eggshell crate soil plug and transplant the young plants.
- Continue to use the lid on the trays until the plants are too tall for it. It may then be useful to prop up taller plants with wooden rods.
- If plants have not bolted after 2 months of growth, begin watering them by alternating between Bloom-Booster fertilizer and tap water.

Making Competent DH5α Cells

- Make 50 mL LB liquid Media in a 200 mL flask and 5mL LB liquid Media in a 50 mL flask.

  <u>LB Liquid Media</u>
  0.5%   Yeast Extract
  1.0%   Tryptone
  1.0%   Sodium Chloride
          >Add diH$_2$O to the appropriate volume
          >Autoclave on liquid cycle @ 121°C for 15min

- T-streak fresh DH5α cells onto an LB plate and incubate at 37°C for 24h.
- Pick a single colony of DH5α and inoculate the flask of 5 mL LB.
- Shake the inoculum at 37°C overnight (~12-16h).
- Use a 1000 μL pipette to transfer 500 μL of the overnight culture to the 50 mL LB flask.
- Shake the new inoculum at 37°C, taking absorbance readings every half hour, starting after 2h of shaking, until the OD$_{600nm}$ is between 0.2-0.5 (about 3-4h in total).
  - While cells are shaking, label 50 X 2 mL microcentrifuge tubes "DH5α" and store at -20°C in -80°C storage box.
- Once the OD is between 0.2-0.5, split the culture into 2 aliquots of ~25 mL in 50 mL centrifuge tubes, weighing the tube with the inoculum on a gram scale and adjusting to equal weights.
- Store tubes on ice for 10min.
- Centrifuge tubes for 10min at 3000rpm at 4°C.
- Remove supernatant by pouring **carefully** so as not to dislodge the pellet, and then immediately place tubes back on ice.
- Use a 1000 μL pipette to remove any remaining supernatant, avoiding the pellet.
- Gently re-suspend the pellet in TSS buffer equal to 10% previous volume
  (i.e., use 2.5 mL TSS to re-suspend the cells in each tube).

  <u>TSS Buffer (50 mL)</u>
  5 g   PEG 8000
  1.5 mL   1M MgCl$_2$
  2.5 mL DMSO
          >Add **LB liquid media** up to 50 mL

- Very quickly, aliquot 100 μL of cell suspension to each of the 50 X 2 mL microcentrifuge tubes that you prepared earlier, trying to keep the box of tubes as cold as possible.
- **Immediately** store the box of microcentrifuge tubes at -80°C.

Plasmid Mini-Preps from Bacteria

- Shake bacteria in 4 mL of Circlegrow® broth with appropriate antibiotic for plasmid selection overnight at 37°C
- Spin down cells for 1 min at 14,000 rpm in a 2 mL microcentrifuge tube 2 mL at a time, pouring off the supernatant after each spin.
- After the second spin, bang the tubes dry on a paper towel and add 250 μL of resuspension buffer.
- Vortex the tubes to resuspend.
- Once the cells are resuspended, add 250 μL Lysis Solution. Close the caps and gently invert 5 times.
- Lay tube on side at room temperature for 5 min.
- After 5 min, add 250 μL Neutralization solution and invert several times to mix.
- Centrifuge at 14,000 rpm for 10 min at room temperature.
- After spin, remove supernatant to new 2 mL microcentrifuge tube, being careful to avoid any pellet or floating material.
- Add 750 μL of isopropanol to the supernatant.
- Centrifuge again at 14,000 rpm for 7 min.
- Aspirate (or pour off) the supernatant, being careful to avoid the DNA pellet.
- Dry the tubes until all residual isopropanol has evaporate (15 min in a spin-vac, longer in a heating block)
- Resuspend the DNA pellet in 50 μL TE (use filter tips)

Reagents:

Cell Resuspension Solution   <*We just order this premade
       50 mM Tris-HCl (pH 7.5)
       10 mM EDTA
       100 μg/mL RNase A

Lysis Solution (NaOH/SDS)
       100 μL 10N NaOH
       500 μL 10% SDS
       4.4 mL diH_2O

Neutralization Solution (1.32M Potassium Acetate)
       12.95 g of Potassium Acetate
       100 mL of diH_2O
           >pH the solution to 4.8 using Glacial Acetic Acid

TE Buffer
       10 mM Tris-HCl (pH 7.5)
       1 mM EDTA

Protoplasting Microbotryum

Making Enzyme Solution
- o Lysing Enzymes from Trichoderma harzianum   0.5 g
- o Driselase   0.5 g
- o 1M MgSO4   25 mL
- Mix and store overnight at 4°C
- Spin down solution at 11,000 rpm for 10 min.
- Filter supernatant into new sterile container (50 mL Falcon Tube)
- Store at 4°C

Protoplasting *Microbotryum* (Mv) Cells
- Grow Mv cells on YPD-10% for 2 days at 27°C
- Scrape a generous loop of cells into 5 mL of Enzyme Solution and shake on table top at medium-low speed overnight
- CAREFULLY add 5 mL of 1.2M mannitol <u>TOPWISE</u> (DO NOT MIX!)
- Centrifuge at 2,000 rpm for 20 min
- Using a P1000, carefully extract the middle layer solution without taking any of the bottom layer. (if you mix the layers you will NOT be able to re-separate them)
  - o 20 mL of solution makes about 6 mL of cells

Short-term Storage
- Aliquot 100 μL into 600 μL tubes and store in fridge until ready to use

Long-term Storage
- Aliquot 100 μL into 600 μL tubes and spin down cells at 14,000 rpm for 1 min
- Discard supernatant and re-suspend cells in 100 μL of STC via finger vortex
- Store in -80°C until ready to use (make sure to thaw on ice before use)

*Method by Naoko Fujita

Topo Cloning

Set up the following:

PCR product	0.5-4μL
Salt Solution	1μL
PCR-TOPO	1μL
Total	6μL

- Mix gently and incubate for exactly 5min at room temp
- Briefly centrifuge and place on ice. Proceed immediately to transformation
- Stir solution gently with pipette tip
- Add 4μL of the TOPO-Cloning reaction into a vial of One Shot cells and mix again
- Incubate on ice for 30 min
- Heat shock cells for 30 sec at 42°C (Gel room heat bath)
- Immediately transfer tubes to ice and incubate for 2 min
- Cap the tube tightly and shake the tube horizontally at 37°C for 30 min – 1hour
- Spread 50-100μL from each transformation on a plate (warm at room temp for 20min) and incubate overnight
- at 37°C
- Pick white colonies and 1 blue for comparison
- Plate colonies on LB with 50μL/mL Kanamycin or 200μL/mL Ampicillin
- Miniprep

Materials

LB Plates
1.0% Tryptone
0.5% Yeast Extract
1.0% NaCl
15g/L Agar
pH 7.0
> *For 1 liter, dissolve 10g tryptone, 5g yeast extract, 10g NaCl and 15g agar in 950 mL deionized water
> *Adjust the pH of the solution to 7.0 with NaOH and bring the volume up to 1 liter
> *Autoclave on liquid cycle for 20 min at 15 psi. Allow solution to cool to 55°C. Add antibiotic
> *Pour into plates

X-Gal Stock
40mg/mL stock solution
> *Dissolve 400mg X-Gal in 10mL dimethylformamide
> *Protect from light by storing in brown bottle at -20°C
> *Add 40μL to plates

IPTG Stock
100mM stock
> *Dissolve 238mg of IPTG in 10mL deionized H_2O
> *Filter-sterilize and store in 1 mL aliquots at -20°C
> *Add 40μL to plates

Transforming Competent *E. coli*

- Thaw competent cells on ice for 5 minutes.
- Add 5 μL of plasmid and finger vortex. Quick spin.
- Incubate on ice for 30 minutes
- Quickly, move tubes to 40°C water bath for 30 sec, and then return to ice for 2 min.
- Add 250 μL of Circlegrow© and shake tubes on their sides at 37°C for 1 hour.
- Spread 150 μL onto selective LB media.
- Incubate plates at 37°C overnight (~18h).

Yeast Miniprep

- Grow Cells in 3ml Dropout media to an $OD_{600} > 1$
  - -Overnight: use all 3mL
  - -Two days: use 1.5mL
- Pellet cells in a 1.5mL microfuge using a 5 min spin. Pour off supernatant. Repeat until all of the media is
- pelleted. Dry on paper towel.
- Resuspend by vortexing in 200μL of SCE/Zymolyase/2ME.
- Incubate at 37° for 30-60 min.
- Add 400μL 0.2N NaOH/1% SDS (lysis solution). Invert to mix.
- Incubate on ice 5min.
- Add 300μL cold 3M K/5M OAc. Invert to mix.
- Incubate on ice 5 min.
- Spin 2 minutes at top speed in a microcentrifuge.
- Pipette supernatant into a fresh tube. Repeat spin.
- Transfer entire volume to a fresh tube.
- Add 400μL isopropanol. Vortex and let stand for 5 min at room temp.
- Spin at top speed for 5 min at room temp.
- Pour off supernatant and wash pellet with 0.5mL 70% ethanol.
- Repeat spin. Pour off supernatant.
- Dry pellet and resuspend in 25μL TE.
- Transform into E. coli. Use 1μL for electrocompetent cells or 10μL for chemically competent cells.

Materials

SCE Solution (100mL)
18.2g   1M sorbitol (in H_2O) [74.8mL for 18.2g]
2.94g   0.1M sodium citrate pH 7.6 (dehydrate trisodium salt mwt. 294.10)
2.23g   0.06M EDTA

SCE/Zymolyase/2ME Solution
5mL   SCE
60μL   10mg/ml Zymolyase (in 1M Sorbitol)
10μL   2-mercaptoethanol **{add in hood}**

NaOH/SDS Solution (Lysis solution)
100μL   10N NaOH
500μL   10% SDS
4.4mL dH_2O

3M K/5M OAc
60mL   5M potassium acetate
11.5mL   glacial acetic acid
28.5ml dH_2O

CURRICULUM VITAE

William C. Beckerson

🏠 1110 Central Ave Louisville, KY 40208   📞 859.325.3133   ✉ Wcbeck01@louisville.edu

🐦 @WilliaMycete   🐦 @WBeckerson

EDUCATION

University of Louisville - Louisville, KY	2015-2020
Doctor of Philosophy in Biology	
University of Louisville - Louisville, KY	2015-2017
Master of Science in Biology	
Georgetown College - Georgetown, KY	2009-2013
Bachelor of Science in Biology / Minor: Business Management	

RESEARCH EXPERIENCE

Dissertation Research:

University of Louisville, US (PI: Michael Perlin, PhD)	2015-2020
Molecular analysis of secreted proteins in the *Microbotryum* genus	
Ruhr-Universität Bochum, DE (Dominik Begerow, PhD)	2018-2019
Introduction of CRISPR Cas9 transformation systems in *Microbotryum*	
Université Paris-Sud, FR (Tatiana Giraud, PhD)	2016-2018
Comparative genomics of species-specific effectors in *Microbotryum*	

Collaborations:

University of Louisville, US (Deborah Yoder-Himes, PhD)	2017-2020
Analyzing the impact of active learning on different student social personalities	

FUNDING

Extramural Funding ($9,605):

GSA Fungal Genetics Conference Travel Award	($250)	2019
DAAD Short Term Research Grant	($4,075)	2018
Chateaubriand STEM Fellowship	($5,280)	2016

Intramural Funding, University of Louisville ($3,236):

Biology 1970's Cohort Fund Grant	($200)	2019
Graduate Student Council Travel Grant	($350)	2019
Graduate Network of Arts and Science Travel Grant	($250)	2019
Graduate Student Council Travel Grant	($350)	2018
Arts & Science Research & Creative Activities Grant	($500)	2018
Biology Graduate Student Association Travel Grant	($175)	2018
Joint Arts & Science Research & Creative Activities Grant	($1,311)	2016
Co-written with Venkata S. Kuppireddy		
Graduate Network of Arts and Sciences Research Fund	($100)	2016

Significant Contributions to Other Grants ($296,889):
NSF Track I International Research Experience for Students (IRES)　　　($296,889)　　2018
　Co-written with: (PI) Dr. Michael H. Perlin　　Award number: 1824851

PUBLICATIONS

Publications in Progress:

Beckerson WC, Klenner S, Leiter R, Khanal S, Gold S, Dominik B, Perlin MH. (2020). The First
　　　Cut is the Deepest: Implementing CRISPR-Cas9 for site specific gene disruptions in the
　　　fungal pathogen complex *Microbotryum vioalaceum*. *xxxxxxx*. - In Progress

Beckerson WC, Klenner S, Sullivan PM, Rollnik T, Rodriguez de la Vega R, Giraud R, Begerow
　　　D, Perlin MH. (2020). To Each Their Own: Analyzing species-specific small secreted
　　　proteins in the *Microbotryum violaceum* species complex. *xxxxxxx*. - In Progress

Beckerson WC, Long G, Rodriguez de la Vega R, Giraud R, Perlin MH. (2020). Breaker of
　　　Chains: Functional characterization of the conserved *Microbotryum* effector
　　　MVLG_0225. *xxxxxxx*. - In Progress

Beckerson WC, Perlin MH. (2020). Director's Cut: How to design a CRISPR Cas9 construct for
　　　use in a new system. *xxxxxxx*. - In Progress

Peer Reviewed Articles:

Beckerson WC, Anderson JO, Perpich JD, Yoder-Himes D. (2020). An Introvert's Perspective:
　　　Analyzing the impact of active learning on social personalities in an upper-level biology
　　　course. *Journal of College Science Teaching*. 49:3, 47-57
　　　https://www.nsta.org/store/product_detail.aspx?id=10.2505/4/jcst20_049_03_47

Beckerson WC, de la Vega RCR, Hartmann FE, Duhamel M, Giraud T, Perlin MH. (2019).
　　　Cause and Effectors: Whole genome comparisons reveal shared but rapidly evolving
　　　effector sets among host-specific plant-castrating fungi. *mBio*. mBio 10:e02391-19
　　　https://doi.org/10.1128/mBio.02391-19

Kuppireddy VS, Uversky VN, Toh SS, Tsai M-C, Beckerson WC, Cahill CC, Carman B, Perlin
　　　MH. (2017). Identification and initial characterization of effectors of an anther smut
　　　fungus and the potential host target proteins. *International Journal of Molecular Science*.
　　　18, 2489
　　　https://doi.org/10.3390/ijms18112489

Textbooks:

Perlin MH, Beckerson WC, Gopinath A, Cobbs G. (2020). *Molecular and Cellular Genetics:
　　　Laboratory Studies*. San Diego, CA: Cognella Academic Publishing. 2nd Edition

Perlin MH, Beckerson WC, Gopinath A, Cobbs G. (2018). *Molecular and Cellular Genetics:
　　　Laboratory Studies*. San Diego, CA: Cognella Academic Publishing. 1st Edition

HONORS AND AWARDS

Graduate Student Publication Award, UofL	2020
Graduate School of Arts and Sciences Student Spotlight, UofL	2020
https://louisville.edu/graduate/student-spotlight/student-spotlight-february-2020	
Introductory Biology Lab Development Award, UofL	2019
Graduate Student Research Presentation Award, UofL	2019
Biology Department Service Award, UofL	2019
College of Arts and Science Student Profile, UofL	2016
https://louisville.edu/artsandsciences/academics/graduate-education/student-profiles/beckerson	

CONFERENCE PRESENTATIONS

Oral Presentations:

Ruhr-Universität Bochum *Microbotryum* Symposium, DE	2019

An Unorthodox CRISPR Approach for and Unorthodox Fungus

Asilomar Fungal Genetics Conference: Smut Convergence, US	2019

Cause and Effectors: Secretome comparison of members from the anther-smut pathogen species complex, *Microbotryum violaceum*

Gordon Research Seminar on Cellular and Molecular Fungal Biology, US	2018

The First Cut is the Deepest: Implementing CRISPR Cas9 as a transformation system for site specific gene disruptions in the fungal pathogen species complex *Microbotryum violaceum*

Kentucky Academy of Science Conference, US	2016

Identifying unique small secreted proteins in divergent species of the fungal pathogen complex *Microbotryum violaceum*

Ruhr-Universität Bochum *Microbotryum* Symposium, DE	2016

Analyzing the role of protein-protein interactions in host/pathogen co-evolution

Poster Presentations:

National Association of Biology Teachers: Professional Development Conference, US	2019

An Introvert's Perspective: Analyzing the impact of active learning on multiple levels of class social personalities in an upper-level biology course

Asilomar Fungal Genetics Conference, US	2019

Cause and Effectors: Secretome comparison of members from the anther-smut pathogen species complex, *Microbotryum violaceum*

Gordon Research Conference on Cellular and Molecular Fungal Biology, US	2018

The First Cut is the Deepest: Implementing CRISPR Cas9 as a transformation system for site specific gene disruptions in the fungal pathogen species complex *Microbotryum violaceum*

PROFESSIONAL PRESENTATIONS

Invited Talks:

Georgetown College Invited Speaker Seminar, US	2019

Cause and Effectors: How rapidly evolving effectors lead to host-specificity between *Microbotryum* and Caryophyllaceae

Belknap Academic Building Anniversary Event, US	2019

An Introvert's Perspective: Analyzing the impact of active learning on multiple levels of class social personalities in an upper-level biology course

Ruhr-Universität Bochum Invited Speaker, DE	2019

The History and Future of CRISPR Cas9

Ruhr-Universität Bochum Invited Speaker, DE	2018

The First Cut is the Deepest: CRISPR Cas9 and how to get started

Georgetown College Invited Speaker Seminar, US	2016

Here and Back Again: A GCPALS tale

Chateaubriand Fellow Research Plan, FR	2016

Identification of Small-Secreted Proteins in the *Microbotryum* genus

Departmental Talks:

University of Louisville Awards Day, US	2019

An Introvert's Perspective: The effect of social personality on active learning

University of Louisvilla GRADtalks, US	2019

Cause and Effectors: Comparing the secretomes of anther-smuts

University of Louisville GNAS Invited Speaker, US	2019

Cause and Effectors: Secretome comparison of members from the anther-smut pathogen species complex, *Microbotryum violaceum*

University of Louisville Awards Day, US	2018

Searching for "Nuclear" Arms: Identifying species-specific small secreted proteins from the fungal pathogen species complex *Microbotryum violaceum*

University of Louisville Awards Day, US	2017

The First Cut is the Deepest: How to get started using CRISPR Cas9 in YOUR lab

University of Louisville Awards Day, US   2016
   Identification of protein-protein interactions of host/pathogen co-evolution in *Microbotryum violaceum*

TEACHING EXPERIENCE

Adjunct Faculty of Record, Georgetown College   2019
BIO 111: Introductory Biology for Majors

1 section	75 min/class	24 students	Twice/week	Fall 2019

BIOL 111: Introductory Biology Lab

1 section	110 min/class	24 students	Once/week	Fall 2019

Invited Group Lecturer for Biotechnology Methods, University of Louisville   2018
BIOL 416: Biotechnology Methods   (Yeast-Two-Hybrid Systems)

2 sections	240 min/class	4 students	Twice/week	Fall 2018

Microbiology Teaching Innovation Learning Lab, University of Louisville   2017-2019
BIO 357: General Microbiology

1 section	75 min/class	64 students	Eight/Semester	Fall 2019

BIO 357: General Microbiology

1 section	75 min/class	49 students	Four/Semester	Fall 2018

BIO 357: General Microbiology

1 section	75 min/class	43 students	Four/Semester	Spring 2018

BIO 357: General Microbiology

1 section	75 min/class	65 students	Four/Semester	Fall 2017

Graduate Teaching Assistant, University of Louisville   2015-2020
BIOL 331: Genetics and Molecular Biology

2 sections	110 min/class	20 & 20 students	Twice/week	Spring 2020

BIOL 331: Genetics and Molecular Biology

2 sections	110 min/class	20 & 22 students	Twice/week	Fall 2019

BIOL 331: Genetics and Molecular Biology

2 sections	110 min/class	20 & 22 students	Twice/week	Spring 2019

BIOL 331: Genetics and Molecular Biology

1 section	110 min/class	19 students	Twice/week	Fall 2018

BIOL 104: introduction to Biological Systems

2 sections	110 min/class	14 students	Three/week	Summer 2018

BIOL 331: Genetics and Molecular Biology

2 sections	110 min/class	20 & 21 students	Twice/week	Spring 2018

BIOL 331: Genetics and Molecular Biology

1 section	110 min/class	8 students	Twice/week	Fall 2017

BIOL 258: Microbiology

2 sections	90 min/class	15 & 6 students	Four Days/week	Summer 2017

BIOL 331: Genetics and Molecular Biology

2 sections	110 min/class	17 & 21 students	Twice/week	Spring 2017

BIOL 331: Genetics and Molecular Biology

1 section	110 min/class	16 students	Twice/week	Fall 2016

BIOL 244: Principles of Biology

2 sections	110 min/class	27 & 28 students	Twice/week	Spring 2016

BIOL 104: introduction to Biological Systems

3 sections	110 min/class	33, 33, & 33 students	Once/week	Fall 2015

STUDENTS MENTORED

Ms. Rebecca Turney	Perlin Lab, University of Louisville	2020
Mr. Lucas Engelhardt	Begerow Lab, Universität Bochum	2019
Mr. Phillip Sullivan	Perlin Lab, University of Louisville	2018-2019
Ms. Grace Long	Perlin Lab, University of Louisville	2018-2019
Mr. Lloyd Bartley	Perlin Lab, University of Louisville	2017-2018
Mr. Adney Rakotoniaina	Perlin Lab, University of Louisville	2017
Ms. Brittany Carman	Perlin Lab, University of Louisville	2016
Ms. Catarina Cahill	Perlin Lab, University of Louisville	2015

UNIVERSITY SERVICES

Biology Undergraduate Student Association: Graduate Student Panel		2019
Biology Faculty Search Committee – Graduate Student Representative		2019
German Club		2018-2019
2018-2019	Member	
Student Grievance & Discipline Committee		2016-2017
2016-2017	Natural Science Division Representative	
Graduate Network of Arts & Sciences		2016-2018
2017-2018	Vice President	
2017	Natural Science Rep. for Grant Review Committee	
2016-2018	Department of Biology Representative	
Biology Graduate Student Association		2015-2020
2019-2020	President	
2018-2019	Graduate Student Rep.	
2016-2017	Social Chair	
2016 & 2020	Webmaster	
2016-2020	Member	

COMMUNITY INVOLVEMENT / OUTREACH

Beer with a Scientist – Monnik Beer Company, Louisville US	2020
Our Friends the Fungi: The many types of fungi and the history of how we've used them	
Orlando Science Center: Spooky Science Week	2019
Zombie Hunt: Using iNatrualist to find zombie ants	
Citizen Science Initiative: the Zombie Fungus Foray	2019
Website: https://wcbeck01.wixsite.com/thezombiefungusforay	
iNaturalist: https://www.inaturalist.org/projects/the-zombie-fungus-foray	
Skype a Scientist	2019-2020
2020 *Creekside Middle School: Sixth Grade Class – Bentonville, AR*	
2019 *Marie Curie Institute: Fourth and Fifth Grade Class – Amsterdam, NY*	
Corry Area High School: Ninth Grade Class – Corry, PA	
Newark Central: Second Grade Class – Newark, NY	
E.K. Powe Elementary School: First Grade Class (AKA the Sea Crew) – Durham, NC	
Annunciation Catholic School: Seventh Grade Class – Denver, CO	
Guest Speaker at University of Louisville: Meet the Professor	2019
Science Information Literacy & Oral Communication	
DuPont Manual Regional Science Fair Judge, Louisville KY	2018
Louisville Regional Science & Engineering Fair Judge, KY	2018
Guest Speaker at Lexington Christian Academy High School	2018
The history of genetic modification of our food	
Guest Speaker at Lexington Christian Academy High School	2016
What is a GMO?	
ExBEERiment – Socialize with Science at the Louisville Science Center	2016

PROFESSIONAL DEVELOPMENT / SERVICES

Training/Workshops:

University of Central Florida NIH One Day Virtual Conference	2020
Moving classes to a remote option for COVID-19	2020

       Training with Panopto, Blackboard Collaborate, and Remote Assessment Tools

Research Academy RUHR: Open Access Science Workshop, DE	2019
Faculty Search Committee: Diversity Training, US	2019

Professional Societies/Organizations:

National Association of Biology Teachers	2019-current

       2019-2020   Community Science Committee

Genetics Society of America	2018-current
Kentucky Academy of Science	2014-current

Peer Review

|	
Society for Molecular Plant-Microbe Interactions	2020

DIVERSITY STATEMENT

As a first-generation college graduate, I am deeply committed to providing opportunities for individuals of all socioeconomic, educational, religious, gender, age, sexuality, nationality, disability, and racial backgrounds. Science benefits from the flow of different ideas and life experiences, and I take steps to be consciously inclusive towards all groups. I also conduct myself under the fundamental premise that quality education should be available to every individual. This includes prioritizing publication of research and data strictly in open access journals and platforms, and actively reaching out to, and providing financial support for, individuals underrepresented in STEM. Financial compensation in addition to the lab experience gained by high-school and undergraduate students is essential to alleviate barriers of entry that disproportionally affect minoritized groups and is vital to improve retention of these students. Financial support includes compensation for the student's work, as well as funds to attend conferences and present their research. As a graduate student, I mentored a diverse group of undergraduate students from a wide range of backgrounds, 6/8 of whom are from traditionally underrepresented groups in Biology. I am committed to continuing this practice of inclusivity and have established a Community Science project oriented towards increasing the participation of Latin(x) individuals, both in academia and in the local community, for future research endeavors.

RESEARCH TRAVEL

Ruhr-Universität Bochum, DE	2019 May-Aug

Graduate student leader for the IRES grant: Infection assays and electron microscopy of transgenic *Microbotryum*

Ruhr-Universität Bochum, DE	2018 June-July

Collaborative research project transforming *Microbotryum* with CRISPR Cas9.

Université Paris-Sud, FR	2016 May-Aug

Collaborative research learning horticulture and infection techniques and implementing agrobacterium-mediated transformation methods for *Microbotryum*.

RWTH University Aachen, DE	2016 June

Group outing with collaborators to discuss future research directions within the *Microbotryum* community.

Ruhr-Universität Bochum, DE	2016 June

Microbotryum symposium with collaborators from Germany and France.

USDA Georgia, US	2015 Dec

Learning agrobacterium-mediated transformation techniques in fungi to apply to *Microbotryum* with Dr. Scott Gold
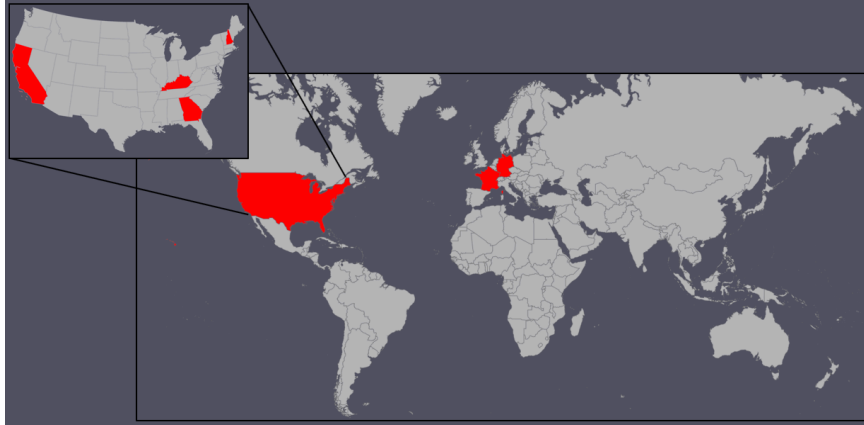
LANGUAGE PROFICIENCIES

English
IRL level 5 – Native Proficiency

German
IRL level 2 – Limited Working Proficiency

French
IRL level 2 – Limited Working Proficiency

Spanish
IRL level 1 – Elementary Proficiency

REFERENCES

Dr. Michael Perlin	University of Louisville, US/ PI	*Michael.perlin@louisville.edu*
Dr. Tatiana Giraud	Université Paris-Sud, FR/ Collaborator	*Tatiana.giraud@u-psud.fr*
Dr. Dominik Begerow	Ruhr-Universität Bochum, DE/ Collaborator	*Dominik.begrow@rub.de*
Dr. Scott Gold	USDA Georgia, US/ Committee Member	*Scott.gold@ARS.USDA.Gov*