DISSERTATION

FROM FIELDS TO GENOMES: TOWARDS A COMPREHENSIVE UNDERSTANDING OF THE LIFESTYLE AND EVOLUTION OF *CLAVICEPS PURPUREA* THE ERGOT FUNGUS

Submitted by

Stephen Andrew Wyka

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: Vamsi Nalam
Co-Advisor: Kirk Broders

Amy Charkowski
Stephen Pearce
Courtney Jahn

Copyright by Stephen Andrew Wyka 2020

All Rights Reserved

ABSTRACT

FROM FIELDS TO GENOMES: TOWARDS A COMPREHENSIVE UNDERSTANDING OF THE LIFESTYLE AND EVOLUTION OF *CLAVICEPS PURPUREA* THE ERGOT FUNGUS

Claviceps purpurea (ergot), an ascomycete and member of the family Clavicipitacea, is considered a pathogen of all grass species (family Poaecea) including economically important cereal crops which infects ovaries resulting in the development of a fungal sclerotium rather than a plant seed. Ergot infections poses significant impacts to agriculture and livestock due to various toxic alkaloids present in the sclerotia. Severe ergot poisoning in humans and livestock, ergotism, can cause corrosion/loss of extremities from gangrene, internal bleeding, diarrhea, and reduced pregnancy and abortion. Due to these serious health concerns, strict restrictions are placed on the amount of ergot contaminated grain that can be accepted for food and livestock feed. However, these toxic alkaloids are also heavily researched in the field of pharmacology and have been shown to provide some beneficial aspects in human medicine. Despite the abundance of pharmacological and agricultural research on *C. purpurea* researchers have been unsuccessful in identifying crop or wild grass varieties that have resistance to ergot infection, leading to critical challenges in the control of ergot disease outbreaks. Recent studies have also suggested that *C. purpurea* is more of a conditional defensive mutualist as opposed to a plant pathogen. Taken together, these factors demonstrate that there are still gaps of knowledge surrounding the epidemiology, lifestyle, evolution, and adaptability of this species. We implemented a comprehensive analysis into the life history of *C. purpurea* through a combination of field surveys, greenhouse inoculations, and deep genomic data mining to help elucidate these gaps.

Field surveys were conducted to investigate the role wild grass populations surrounding cereal crop fields play in epidemiology of ergot outbreaks. Results revealed that unmanaged grasses along ditch banks, even in drought years, represent significant inoculum reservoirs of ergot, particularly when *Bromus* spp. are present, and should be a focal point in future research for better disease control. Greenhouse inoculations were conducted to elucidate the effects of *C. purpurea* infections on hosts through inoculations of a single isolate on two commercial cereal crops in a controlled setting. Our results show that the effect of *C. purpurea* infections can range from negative to positive, depending on infection rate, plant species, and plant tissue, but overall showed a general trend of neutral effects. However, we did observe a potential for increased root growth as infection rates increased, which could signify an interesting plant-microbe interaction that imparts a benefit, of infection, on highly rhizomatous grass hosts such as *Bromus* spp.. Lastly, through a collaborative effort we sequenced, assembled, and annotated 50 *Claviceps* genomes, representing 21 species, for a comprehensive comparison of genome architecture, plasticity, and evolution within the genera. We also conducted a detailed analysis of *C. purpurea* through construction of a pangenome and investigations of the recombination and positive selection landscape across the genome. Our genus-wide comparison revealed that despite having nearly identical life-strategies, these closely related species have substantially altered genomic architectures and plasticity that are likely driving genome adaptation. One key difference we observed was a shift from characteristic one-speed genomes in narrow host-range *Claviceps* species of sections *Citrinae* and *Paspalorum* to two-speed genomes in broader host-range lineages of sections *Pusillae* and *Claviceps*. *Claviceps purpurea* was observed to have a large accessory genome that is likely influenced by a large effective population size, high recombination rates, and transposable element (TE) mediated gene duplication. Due to a lack of

repeat-point induced (RIP) mutation, prolific TE expansion is likely controlled by high recombination rates, which subsequently may be influencing the overall trend of purifying selection observed within the species. However, secondary metabolites genes were found to have the highest rates of positive selection on codons within genes, indicating that these genes are a primary factor affecting the diversification of the species into new ecological niches and to potentially help maintain its global distribution and broad host range.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisors and committee members for their helpfulness and encouragement during my degree. They were always available to answer any and all of my questions and, if not, were able to lead me in the right direction. I am incredibly grateful to my advisor, turned co-advisor, Kirk Broders for his unending hard-work and dedication to constantly being able to read and provide feedback for my many drafts and manuscripts. It has been a wonderful six years being your graduate student, for my master's and now PhD, and I have learned so much under your mentorship.

I would also like to deeply thank my wife, Evgenia Wyka, for her love, compassion, and support during these stressful times. Especially, her devotion for moving with me to Colorado to purse my degree and for her assistance is helping raise and care for our two kids while I spent the time to write this dissertation, I could not have done it without you. In addition, I would like to express my appreciation to my computer science friend Rick Emery. About two years ago when I had zero computational knowledge and had difficulties writing the simplest conditional statements in Python, Rick was there to help guide me through my first script to re-organize a dataset and calculate some results in Python. Since then he was always available to help answer any scripting related questions, even my 2 am questions of "Why isn't Python doing what I am telling it to do", to which the general replies were "Python is working correctly, the problem is you". Despite this, Rick was always able to help me refine my scripting ability, which allowed me to develop my computational knowledge and ultimately analyze all the genomic data presented in this dissertation.

I would also like to thank Stephen Mondo for his advice into genomic analyses available and assistance in interpreting the results. In addition, I would like to thank Judith Jolly for all her

help in facilitating and conducting the ergot field surveys in the San Luis Valley, Colorado. For their gracious generosity in sharing data and biological and genomic resources, I would like to thank my collaborators Miao Liu and Miroslav Kholařík. Lastly, I would like to thank all of my family and friends for their support and encouragement throughout my education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Claviceps purpurea (Fr.) Tul., the ergot fungus, is an ascomycete from the family Clavicipitaceae. This fungal species has a long history as a plant pathogen, which is known to infect 400+ species of grasses, and is an important source of potential active secondary metabolites that have been studied for centuries for medicinal purposes (Lee 2009; Alderman *et al.* 2004). Infection of the plant occurs as spores enter unpollinated florets and begin to colonize on the host's stigma. These infections eventually lead to the complete colonization of the flower's ovary and the fungus produces a sclerotia (sexual fruiting body) in place of the seed, thereby reducing yield and sexual reproductive potential of the plant. Ergot was so commonly associated with rye that early botanical drawings from 1658 of rye also contained ergot sclerotia (Miedaner and Geiger 2015). The first documented use of ergot, as a prescription in medicine, is from 1582 (Ainsworth 1976), but one report could date the use as far back as 1474 (Píchová *et al.* 2018 *Appendix*). At low doses, ergot extracts were being used to accelerate childbirth or to induce abortions (Miedaner and Geiger 2015). Larger concentrations, such as ingestions of whole ergot sclerotia or bread baked with contaminated grain, often lead to a disease now called ergotism but has been previously known as St. Anthony's fire or "holy fire" (Lee 2009). Documented reports of ergotism epidemics date back to 857 AD and occurred throughout the Middle Ages, however, some reports suggest that this disease was recorded in a sacred text from around 350 BCE (Miedaner and Geiger 2015). It is even suggested that ergotism played a role in the witch trials of Salem, Massachusetts, USA in 1692 (Caporael 1976) and those in Finnmark, Norway in the 17th century (Alm 2003). These correlations are most likely due to the hallucinogenic effects of an active alkaloid produced from ergot, ergotamine, which was used in

1938 by Albert Hofmann to first synthesize lysergic acid diethylamide (LSD) (Lee 2009). Today ergot is still a serious concern for farmers and livestock producers as ingestion of ergot and its alkaloids causes harm to all mammals. Due to these serious health concerns, strict restrictions are placed on the amount of ergot contaminated grain that can be accepted for food and livestock feed.

However, despite its long research history the control of ergot disease proves challenging, as researchers have been unsuccessful in identifying germplasm with genetic resistance to ergot infections and studies have shown mixed results in the effectiveness of fungicide applications (Wood and Coley-Smith 1980; Evans *et al.* 2000; Schumann 2000; Gladders *et al.* 2001; Anon 2002; Bailey *et al.* 2003; Menzies 2004; Pageau and Lajeunesse 2006; Bayles *et al.* 2009; Oxley *et al.* 2009; Menzies and Turkington 2015). Current management strategies to control ergot outbreaks rely on whole-farm integrated approaches such as crop rotation with non-susceptible hosts, cleaning seed prior to sowing, deep plowing, and post-harvest field burnings to help reduce initial inoculum levels within a field for subsequent years, as sclerotia generally do not last longer than one year (Campbell and Freisin 1959; Cunfer and Seckinger 1977; Bretag 1981; Johnston *et al.* 1996; Alderman 2006; Bayles *et al.* 2009; Uppala *et al.* 2011; Menzies and Turkington 2015). While ergotism is no longer a threat in much of the world, due to grain restrictions and advancement of seed cleaning technology, recent reports from the San Luis Valley, Colorado (Judith Jolly *personal communication*) and across Canada (Miao Liu *personal communication*) indicate persistent threats of ergot outbreaks on cereal crops. These outbreaks are causing negative effects on the growers through the rejection on grain harvests, which is fueling continued research into *C. purpurea*. Recent studies have also suggested that *C. purpurea* is more of a conditional defensive mutualist to its host as opposed to a plant pathogen (Wäli *et*

al. 2013), which questions the true nature of *C. purpurea* and perhaps provides insight into the absence of known resistance genes in commercial and wild grasses. Taken together, these observations reflect gaps of knowledge surrounding the epidemiology, lifestyle, evolution, and adaptability of this species.

My overall goal was to better understand the evolutionary and lifestyle history of *Claviceps purpurea* by utilizing a multi-disciplinary approach to reexamine the *Claviceps*-grass pathosystem centered around four main objectives:

1. Utilize the barley production system in the San Luis Valley, CO to examine the role plant community compositions, along ditch banks, play in the epidemiology of ergot disease outbreaks.
2. Elucidate the effects *C. purpurea* has on its host through inoculations of a single isolate on two commercial cereal crops *Hordeum vulgare* (barley) and *Triticum aestivum* (wheat) in a controlled greenhouse setting.
3. Sequence and annotate 50 *Claviceps* genomes, representing 19 species, to characterizing the genomic plasticity and architecture of species to identify factors associated with the evolution of the genus.
4. Identify factors governing the evolution and adaptive potential of *C. purpurea* through construction of the pangenome and analysis of the recombination and positive selection landscape of the species.

Chapter 1: Field Surveys Demonstrate *Bromus* spp. are the Primary Inoculum Reservoirs for *Claviceps purpurea* in the San Luis Valley of Colorado

INTRODUCTION

The plant ovarian parasite *Claviceps purpruea* (Fr.) Tul. (ergot) poses significant impacts to agriculture and livestock due to various toxic alkaloids present in the sexual fruiting bodies (sclerotia) which can cause severe ergot poisoning in humans and livestock (ergotism) (Schumann 2000; Bailey *et al.* 2003). Due to these serious health concerns, the USDA has placed strict regulations on the amount of ergot contaminated grain that can be accepted for food and livestock feed with varying thresholds existing for select grain types. In general ergot levels of > 0.1% net weight are rejected for grain sale (USDA 2016). Ergot epidemics are often sporadic in nature as outbreaks are heavily dependent on favorable weather conditions for the dispersal of the disease throughout crop fields, therefore, once weather conditions are conducive outbreaks are likely to occur, even in regions where epidemics were not present in preceding years (Bayles *et al.* 2009). This suggests that ergot reservoirs might be ubiquitous around crop fields which provide sufficient inoculum to cause outbreaks upon favorable conditions (i.e. warmer temperatures and increased rainfall) (Campbell 1957, 1959; Mantle *et al.* 1977; Bayles *et al.* 2009).

Some C*laviceps* are limited to a single genus but others may infect an entire subfamily of plants, as is the case for *C. purpurea* which can infect 400+ species of Poaecea (Taber 1985; Alderman *et al.* 2004; Píchová et al. 2018). The initial source of inoculum comes from the overwintering sclerotia, which may be left on the soil during harvest or sown with seeds. In the spring (April – June), during the onset of precipitation and warmer temperatures, stalks

containing perithecia with ascospores emerge from sclerotia (Brown 1947; Rapilly 1968; Mantle *et al* 1977; Bayles *et al.* 2009). Upwards of 60 stalks can be formed from a single sclerotia over a period of several months (Sprague 1950; Wood and Coley-Smith 1982). Ascospores are forcibly ejected from perithecia into the air and dispersed by wind. Ascospores must land on unfertilized host stigma to germinate and cause infection within 24 hours. After a week, the embryo of infected florets is replaced with fungal mycelium and a mucoid mass conidia droplet (honeydew) forms on the surface of the grain flower. Conidia are then dispersed throughout the summer by physical contact, rain splash, or insects. Over time the infection of the floret continues, a sclerotium develops, and the sclerotium falls to the ground, is eaten by grazing animals, or is harvested with the grain (Bayles *et al.* 2009). The polycyclic nature of the fungus allows continuous infections to develop resulting in accumulation of inoculum reservoirs of sclerotia in the field and on nearby alternative hosts. These sclerotia do not usually survive for more than one year (Cunfer and Seckinger 1977; Taber 1985; Anon 2002).

Researchers have not yet identified germplasm with genetic resistance to ergot infections (Menzies 2004; Pageau and Lajeunesse 2006; Bayles *et al.* 2009; Oxley *et al.* 2009; Menzies and Turkington 2015). Similarly, fungicides do not consistently control ergot, suggesting that the use of fungicides for control of ergot is not economically practical (Wood and Coley-Smith 1980; Evans *et al.* 2000; Schumann 2000; Gladders *et al.* 2001; Anon 2002; Bailey *et al.* 2003; Pageau and Lajeunesse 2006). Current management strategies used to control ergot outbreaks consist of whole-farm integrated approaches, such as crop rotation with non-susceptible hosts, cleaning seed prior to sowing, deep plowing, and post-harvest field burnings to help reduce initial inoculum levels within a field for subsequent years (Campbell and Freisin 1959; Bretag 1981; Johnston *et al.* 1996; Alderman 2006; Bayles *et al.* 2009; Uppala *et al.* 2011; Menzies and

Turkington 2015). Additional strategies include reductions in the amount of overhead irrigation to help reduce the development and spread of ascospores and conidia throughout a field (Alderman 2006; Dave Dougherty *personal communication*). However, many of these practices are not ideal options for growers in the water-limited and wildfire prone western U.S. and all of these methods focus on control of ergot inoculum within a crop field. Despite the broad host range of ergot, control of grass hosts surrounding cereal crop fields to reduce additional inoculum reservoirs is often overlooked (Campbell 1957; Bayles *et al.* 2009). Although ergot has a broad host range, it may not form sclerotia well on all grass hosts. Targeted grass control approaches around field edges with grass species that pose the greatest threat to ergot outbreaks may provide additional ergot management. However, the grass species population abundances and exact species that should be targeted may vary among grain-growing regions.

Recent outbreaks of *C. purpurea* on barley (*Hordeum vulgare* L.) are now more prevalent in geographically isolated regions of the San Luis Valley, Colorado; Worland, Wyoming; and in Huntley and Shepherd, Montana (Dave Dougherty *personal communication*; Judith Jolly *personal communication*). Ergot outbreaks are primarily caused by climatic shifts, which alter the synchrony of susceptible crop growth stages with inoculum presence (Bayles *et al.* 2009). In addition, infected grass margins represent a significant inoculum reservoir resulting in increased disease severity, particularly in unmanaged or naturally regenerated margins that are colonized by highly susceptible hosts. For example, in Europe, black-grass (*Alopecurus myosuroides* Huds.) panicle numbers has a positive linear relationship with ergot numbers in harvested wheat (Mantle *et al.* 1977; Bayles *et al.* 2009). In the United States the invasive smooth brome (*Bromus inermis* Leyss.), a highly rhizomatous cold-season perennial, is frequently reported with ergot. This invasive species in now ubiquitous across North America and often out competes wild

grasses in naturally regenerated areas (Sprague 1950; Campbell 1957; Romo and Grilz 1990; Nagel *et al.* 1994; Murphy and Grant 2005; Otfinowski *et al.* 2007; Dillemuth *et al.* 2009; Fink and Wilson 2011). Smooth brome is a known host of *C. purpurea* in Colorado, Wyoming, and Montana, is very abundant in these regions (Sprague 1950), and *C. purpurea* isolates from smooth brome can infect barley Campbell (1957). In addition, Smooth brome has a high incidence of infection and spreads rapidly through rhizome proliferation (Romo and Grilz 1990; Otfinowski *et al.* 2007), so smooth brome and other unmanaged grass species alongside grain fields may serve as important inoculum reservoirs and their presence may be correlated with ergot outbreaks in cereal crops.

Devastating economic losses caused by ergot are rather infrequent, due to the dependency of climatic conditions for widespread infections (Bayles *et al.* 2009). However, recent data indicates future increases in total annual precipitation and extreme precipitation events throughout the U.S. and other parts of the world (Karl 1998; Milly *et al.* 2002; Rosenzweig *et al.* 2002; Alter *et al.* 2015). This may result in an increase in the frequency of ergot outbreaks as regions experience increased precipitation from year to year, particularly if field margins are composed of highly susceptible unmanaged grass species that harbor *C. purpurea* inoculum. Therefore, it is necessary to understand the ecology of *C. purpurea* on highly susceptible grass species and the impact of plant community compositions on the accumulation of inoculum reservoirs. This study utilized the barley production system in the San Luis Valley, CO to examine the role plant community compositions play in the epidemiology of ergot disease outbreaks.

MATERIALS AND METHODS

Region information

This survey was conducted in the San Luis Valley located in south central Colorado (Fig. 1.1), which is a high-altitude intermountain basin with an average elevation of 2,335 m. The valley floor is nearly flat and surrounded by two mountain ranges with peaks > 4,100 m, the San Juan Range to the west and the Sangre de Cristo Range to the east. The valley is split by the flow of the Rio Grande, which flows east from the San Juan Range. Much of the San Luis Valley is used for grazing, with farming being concentrated around the towns of Alamosa, Monte Vista, and Center. Primary crops include potatoes, barley, alfalfa, wheat, lettuce, quinoa, and more recently hemp, and crops are typically irrigated with center-pivot systems. The predominant crop rotation observed in the valley is a two-year rotation of potatoes and barley. In addition, there are also fields that continuously plant barley for several years (Judith Jolly *personal communication*). On average, since 2000, 20,272 ± 4,335 hectares of barely is planted each year in the San Luis Valley, CO (https://nass.usda.gov). Most of the barley in this region is grown as certified seed for Molson Coors and is thus under strict regulations regarding disease presence.

Precipitation data

Precipitation data was obtained from CoAgMET (https://coagmet.colostate.edu/) for sites Center (37.7067, -106.1440), Center #2 (37.8288, -106.03830), and Blanca (37.3905, -105.5570) from 2000 to 2019. Sites were averaged together to get mean cumulative precipitation and number of precipitation days for the spring (April - June) in the San Luis Valley, CO.

Site information and field surveys

During the first week of July, August, and September of 2017, 2018 (excluding August), and 2019 surveys were conducted at 12 sites, resulting in a final sample size of 96 survey sites.

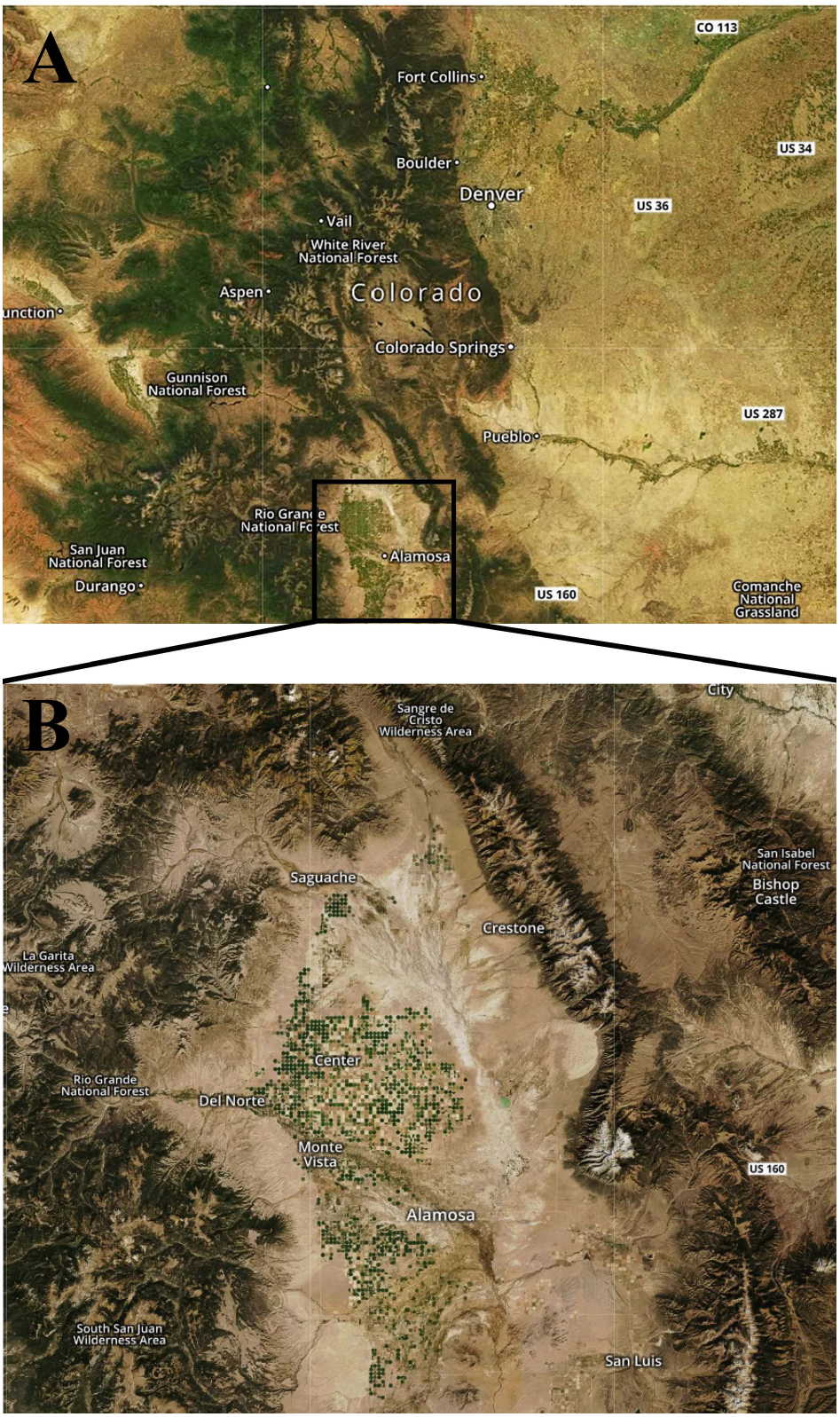


Figure 1.1: Satellite image of **A:** Colorado, USA and **B:** San Luis Valley, CO. Image obtained from: https://ngmdb.usgs.gov/.

Sites were ranked with ergot disease history based on input from Judith Jolly, the Molson Coors agronomist of the region, and the growers: **No**: site did not have any ergot history; **Low**: site had low ergot history (i.e. grain was neither rejected nor cleaned); **High**: site had severe ergot history (i.e. grain was either rejected or needed to be cleaned). Due to the use of crop rotation in the valley, in many of the cases, we were unable to perform repeated measures of the same site in consecutive years. Therefore, new sites were selected each year and exact locations were kept anonymous as requested by the growers.

To sample for ergot, 0.5 m^2 quadrats were placed every 3.05 m along two 30.5 m transects along ditch banks, adjacent to barley fields, at each site for a total of 20 quadrats (10 m^2) per site. To facilitate the speed of each survey, plants were divided into four groups: **Brome**: all *Bromus* spp. (dominantly smooth brome (*B. inermis* Lyess.), as well as meadow brome (*B. biebersteinii* Roem. & Schult.; syn: *B. erectus*, *B. riparius*) and mountain brome (*B. marginatus* Nees ex Steudel; syn: *B. carinatus*); **Wheatgrass**: western wheatgrass (*Pascopyrum smithii* (Rydb.) Á. Löve), intermediate *wheatgrass (Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey*)*, slender wheatgrass (*Elymus trachycaulus* (Link) Gould ex Shinners), quackgrass (*E. repens* (L.) Gould), and crested wheatgrass (*Agropyron cristatum* (L.) Gaertn.); **Other Grass**: all other grass species (*Poa* spp.*,* green foxtail (*Setaria viridis* (L.) P. Beauv.), oats (*Avena sativa* L.), barley (*H. vulgare*), sleepy grass (*Achnatherum robustum* (Vasey) Barkworth), timothy (*Phleum pratense* L.), blue gramma (*Bouteloua gracilis* (Willd. ex Kunth.) Lag. ex Griffiths), and foxtail barley (*H. jubatum* L.)); **Non-grass**: all non-grass plant species. Plant area coverage was estimated, within each quadrat, as the proportion of the quadrat covered by plants of a given group. Disease incidence was measured as the proportion of the total number of grass inflorescences per quadrat with at least 1 ergot sclerotia or honeydew droplet (Fisher *et al.* 2007).

Site level disease incidence was computed by averaging quadrats within each transect and then across transects. If infection occurred, the grass group (brome, wheatgrass, or other grass) was noted to have infection but the exact number of infected inflorescences per grass group was not measured. If no ergot was present along the two transects, the site was walked and visually checked to determine whether ergot was present at the site. During the August surveys, barley fields were surveyed in the same manner. Furthermore, during the September surveys we counted the number of sclerotia per inflorescence in each quadrat for up to five randomly sampled infected inflorescences from each grass group.

Statistical analyses and software

We quantified ditch bank plant composition as importance values (IV), which are calculated by adding relative coverage, relative density, and relative frequency together to measure the overall influence of each of the four plant groups per site (Curtis and McIntosh 1951). Since plant densities were not measured, importance values were only computed from relative plant area coverage and the relative frequency based on the number of quadrats per site plants were found in. Both disease incidence and plant group IV's were analyzed using analysis of variance (ANOVA) using JMP Pro v13, with student's t-test for pairwise mean comparison, to evaluate the differences between site disease history for each sampling month and plant group, respectively. Linear regressions were used to examine correlations between plant group IVs and disease incidence using R v3.6.1. We analyzed differences between sclerotia counts per inflorescence using Mann-Whitney U test with Benjamini-Hochberg false discovery rate (FDR) multi-test correction using the Python modules SciPy v1.3.1 and statsmodel v0.11.0. Figures were created using the Python module Matplotlib v3.1.1 and R v3.6.1.

Density and inoculum estimations

Inflorescence densities were estimated by multiplying inflorescence counts (from disease incidence calculations) and plant area cover (from IV calculations). During surveys we did note whether infection occurred on grass groups (brome, wheatgrass, or other grass) within each quadrat. From this data we estimated disease incidence per grass group. For example, if a quadrat had a total of 100 inflorescences with an area coverage of 50% brome, 25% wheatgrass, and 25% other grass our estimated inflorescence densities would be 50 brome, 25 wheatgrass, and 25 other grass inflorescences per quadrat (0.5 m^2). If we found 20 infected inflorescences and disease was only noted on brome then our estimated infection per plant group would be 100% brome, 0% wheatgrass, and 0% other grass resulting in an estimate of 20 infected brome inflorescences for the quadrat. However, if disease was noted on brome and wheatgrass, our estimated infection per plant group would be 50% brome, 50% wheatgrass, and 0% other grass resulting in an estimate of 10 infected brome and 10 infected wheat inflorescences for the quadrat. Based on these estimates we further estimated the amount of sclerotia coming from each grass group (primary inoculum for the subsequent year) by multiplying the number of infected inflorescences by the mean number of sclerotia per inflorescence observed for each grass group at each site. Statistics were not computed for estimates.

RESULTS

Precipitation in the San Luis Valley, CO

Between 2000 and 2019 the San Luis Valley saw a mean cumulative spring precipitation of 37.28 ± 22.59 mm with a mean of 13.87 ± 5.50 precipitation days (Fig. 1.2). The year with the highest number of precipitation days was 2015, which corresponded to the greatest incidence of ergot in barley in the valley since 2006 (Judith Jolly *personal communication*). The outbreak

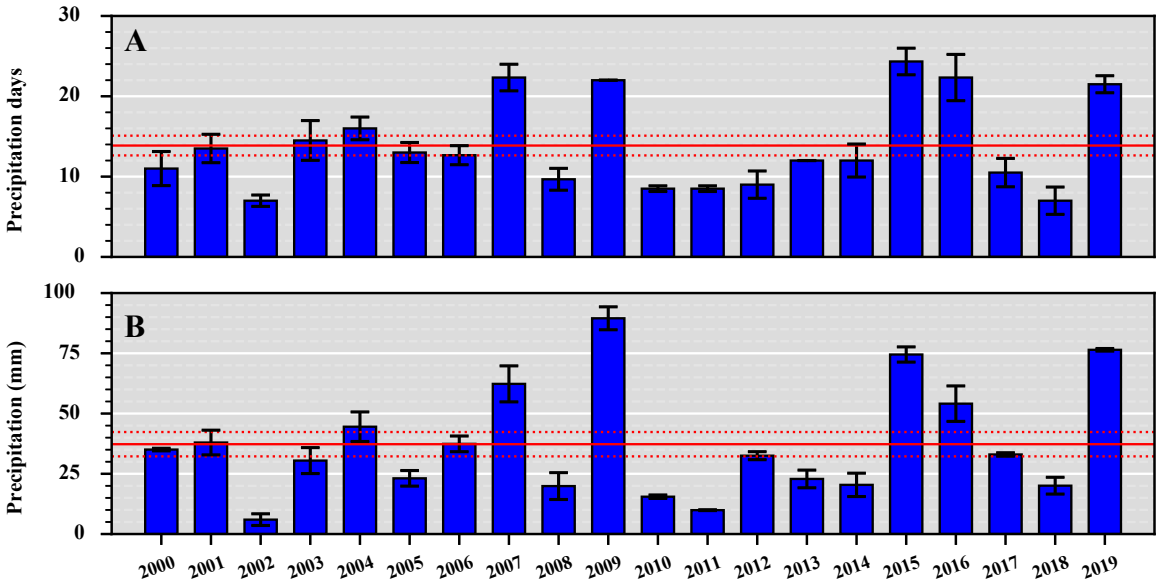


Figure 1.2: Spring precipitation (April – June) in the San Luis Valley, CO from 2000 to 2019. **A)** Number of precipitation days, **B)** Cumulative precipitation. Bars represent standard error. Red solid line represents overall means for the region with dotted lines representing standard error.

could have also been exacerbated due to shipments of ergot-contaminated wheat seed planted in the valley that year (Judith Jolly *personal communication*). However, it is likely that the increased precipitation in the region facilitated the spread of the disease, due to the dependence of ergot outbreaks on climatic factors (Bayles *et al.* 2009). Ergot disease pressure was also prevalent in 2016 (Judith Jolly *personal communication*; Dave Dougherty *personal communication*), which similarly experienced above average spring precipitation (Fig. 1.2). During the first two years of the survey (2017 and 2018), the San Luis Valley experienced below average spring precipitation and were considered drought years for the region (Fig. 1.2). The final year of the survey (2019) saw above average spring precipitation similar to 2015 and 2016 (Fig. 1.2).

Disease incidence

Infection rates in unmanaged grasses along ditch banks significantly increased throughout the summer ($F = 19.33$, $P < 0.0001$), irrespective of site disease history. The lowest levels (0.6% ± 2.5%) were observed in July and highest levels (14.8% ± 12.8%) of disease incidence were observed in September (Fig. 1.3). During July, if ergot was not initially present within transects, it was difficult to detect ergot through visual checks of the entire site as honeydew droplets were generally the dominant stage during this period. However, during the months of August and September ergot was present at all sites. Within each month, sites with a history of high ergot severity had the highest levels of disease incidence but were only significantly higher than both sites with no ($P < 0.001$) and low ($P = 0.006$) ergot history during the month of September (Fig. 1.4). Disease incidence during the month of August at sites with high ergot history were only significantly higher than sites with no ergot history ($P = 0.005$) (Fig. 1.4). In all months, sites

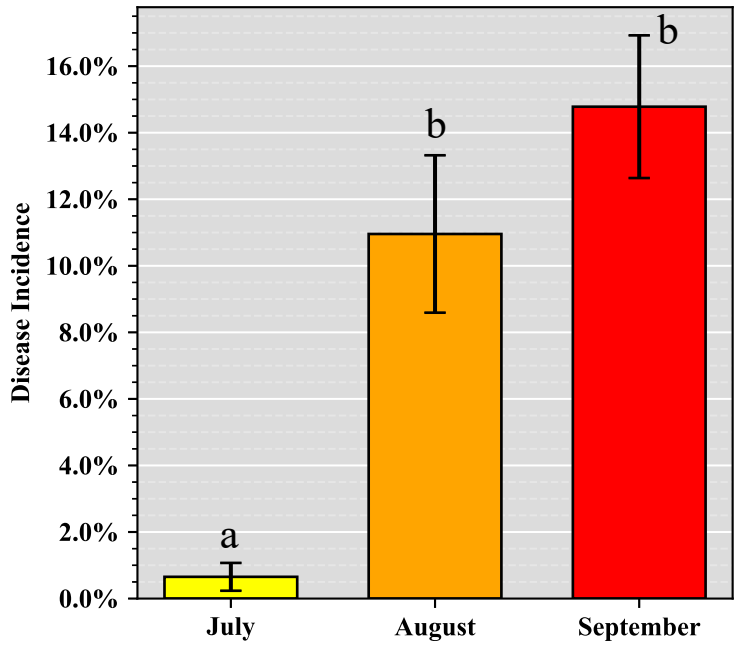


Figure 1.3: Monthly mean comparison of percent of infected inflorescences per site of wild grasses along ditch banks adjacent to barley fields in the San Luis Valley, CO. Student's t-test was used for mean separation after analysis of variance, letters not connected by the same letter are significantly different.

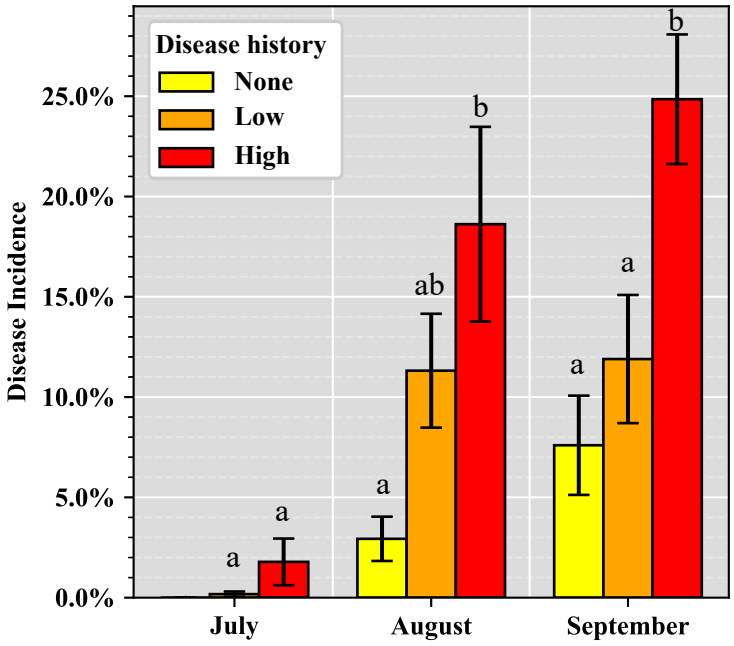


Figure 1.4: Site disease history rating mean comparison, of percent of infected inflorescences per site of wild grasses along ditch banks adjacent to barley fields in the San Luis Valley, CO. Student's t-test was used for mean separation after analysis of variance, letters not connected by the same letter are significantly different, within each month.

with a low history of ergot disease had higher levels of disease incidence than sites with no ergot history but were not significantly different ($P > 0.05$) (Fig. 1.4).

Infection rates within barley fields showed extremely low disease incidence with only 14 barley inflorescences with ergot out of greater than 10,000 inflorescences examined. Three of the barley infections occurred in 2017 and 11 occurred in 2019. The lower disease incidence in 2017 was associated with below average precipitation and number of precipitation days during the spring (Fig. 1.2), providing less favorable conditions for the spread of ergot sclerotia into barley fields. The increased precipitation observed in the spring of 2019 (Fig. 1.2), may have led to the increase in barley infections. However, due to the low incidence of barley infections, we were unable to confidently analyze the effects unmanaged grass compositions and disease incidence have on within field disease incidence.

Plant community composition and infections

Plant community composition was assessed using importance values (IV) of the four categorized plant groups (brome, wheatgrass, other grass, and non-grass). There was no significant differences in IVs between months for each of the plant groups ($F = 0.322$, $P = 0.726$; $F = 0.289$, $P = 0.75$; $F = 0.0162$, $P = 0.984$; $F = 0.22$, $P = 0.803$, respectively), therefore, months were pooled for further analysis. The ditch banks at sites with a history of high ergot severity were primarily composed of brome (116.19 ± 34.93 IV) followed by wheatgrass (61.41 ± 57.12 IV), non-grass (29.43 ± 26.28 IV), and other grass species (5.55 ± 9.12 IV) (Fig. 1.5). Sites with low ergot history showed a similar pattern of community composition. However, ditch banks were composed of significantly less brome species and significantly more non-grass species than sites with high ergot history ($P < 0.0001$, $P = 0.002$, respectively) (Fig. 1.5). We observed a shift in brome and non-grass compositions at sites with no ergot history where ditch banks were

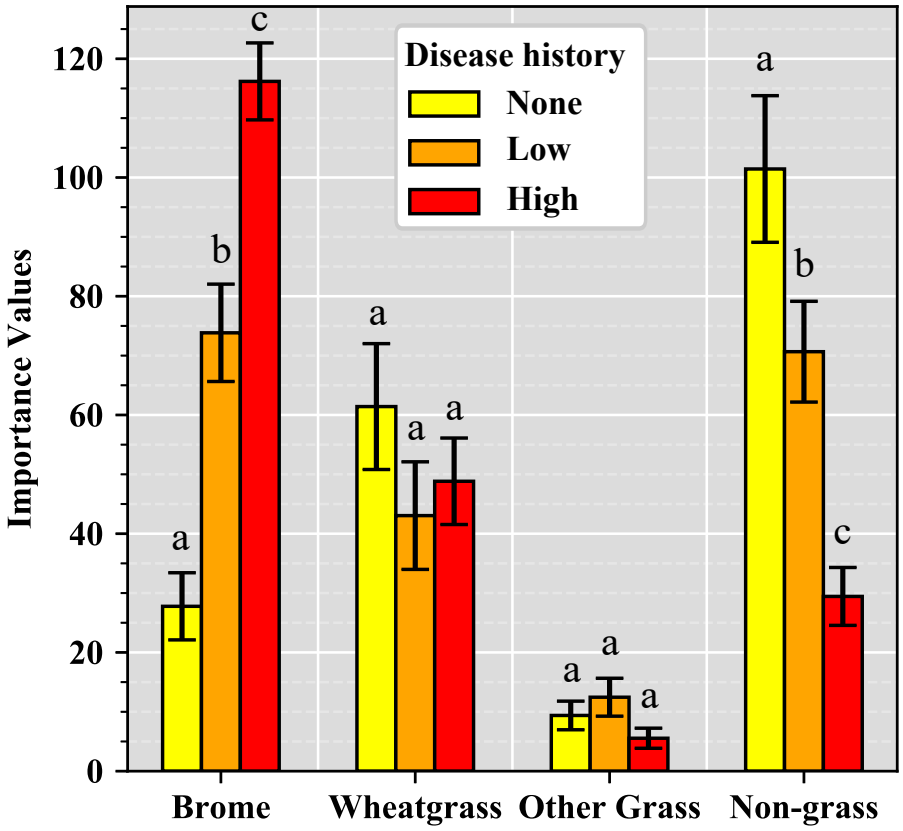


Figure 1.5: Between site disease history rating mean comparison of community composition (importance values) of plant groups per site within ditch banks adjacent to barley fields in the San Luis Valley, CO. Student's t-test was used for mean separation after analysis of variance, letters not connected by the same letter are significantly different, within each plant group.

primarily composed of non-grass species (101.44 ± 66.54 IV) and brome species (27.77 ± 6.99 IV) were the third most abundant plant group (Fig.1. 5). The compositions of brome and non-grass species at sites with no ergot history were significantly different from sites with a history of ergot disease ($P \leq 0.002$) (Fig. 1.5). The compositions of wheatgrass and other grass species did not significantly differ between disease history ratings ($F = 1.03$, $P = 0.3614$; $F = 1.84$, $P = 0.166$, respectively). Other grass species represented the lowest abundances across all disease history ratings (Fig. 1.5).

We further examined the effect of community composition (IV) on disease incidence. Our results indicate that during the month of July community composition did not correlate with disease incidence. This was expected as infections found in July were most likely the result of primary infection from sclerotia on the ground and not secondary infection via conidia. There was also no linear relationship between the composition of wheatgrass and other grass species with disease incidence ($P >> 0.05$) (Fig. 1.6). For sampling points in August and September, there was a positive correlation ($P = 0.007$, Adj. $R^2 = 0.254$; $P < 0.001$, Adj. $R^2 = 0.291$; respectively) between communities with a higher proportion of brome species and disease incidence, while there was a negative correlation ($P = 0.03$, Adj. $R^2 = 0.161$; $P = 0.002$, Adj. $R^2 = 0.224$; respectively) between the composition of non-grass species with disease incidence. We observed that the positive correlations of brome composition and the negative correlations of non-grass compositions became more correlated and significant from August to September (Fig. 1.6). While communities with a higher composition of non-grass species had a reduction in potential for ergot infection (i.e. through exclusion of potential hosts), these data suggest that brome species are more often infected than wheatgrass and other grass species (Fig. 1.6).

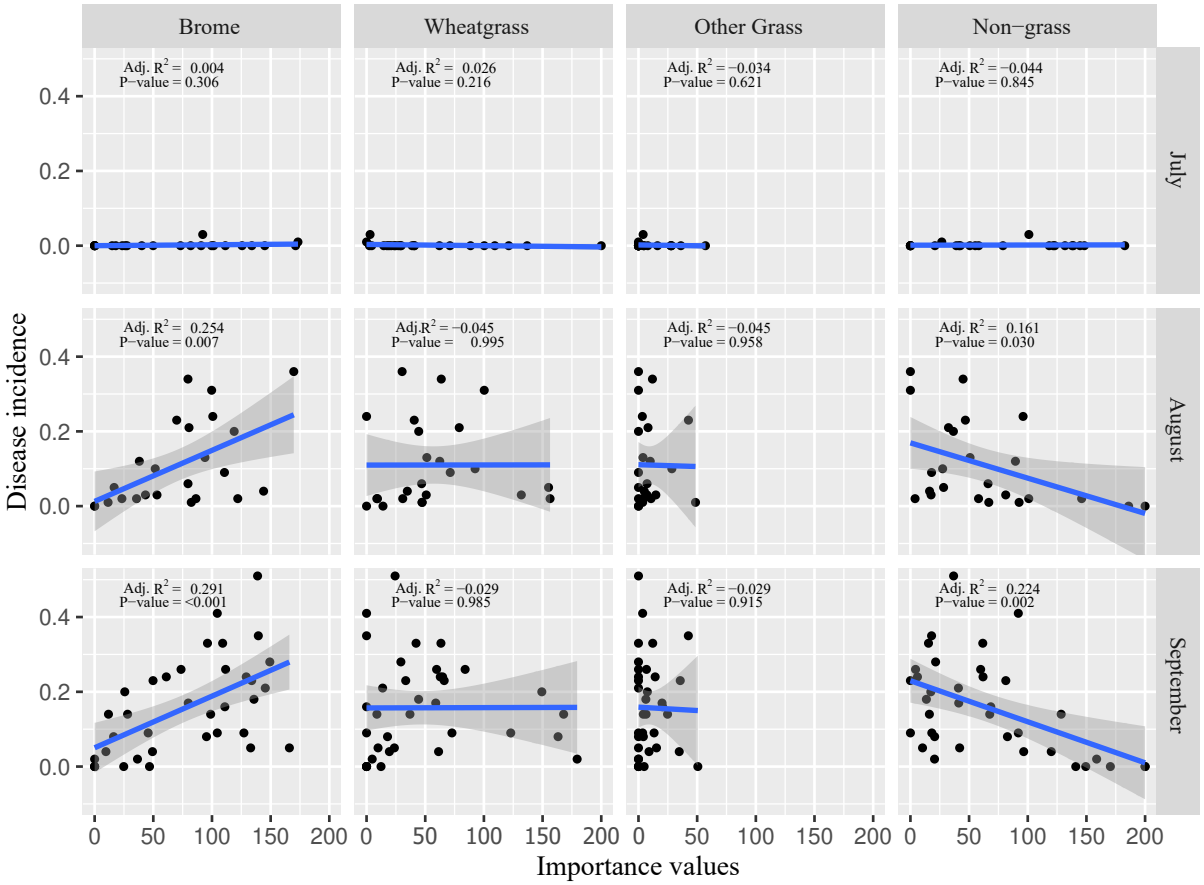


Figure 1.6: Regression for effects of each plant groups community composition (importance values) on the mean percent of infected inflorescences per site, for each month, along ditch banks adjacent to barley fields in the San Luis Valley, CO. Shadowed region depict 95% confidence intervals. Adjusted R^2 and *P*-values are imposed onto corresponding plots.

During September the number of ergot sclerotia per inflorescence on up to 5 randomly sampled infected inflorescences from each grass group were counted. This resulted in a sample size of 875 brome, 587 wheatgrass, and 21 other grass infected inflorescences. The increased sample size of brome species further suggests that brome species are more frequently infected than other grass species, but this could also be attributed to the overall increased abundance of brome species (Fig. 1.4). We found no significant differences in the number of sclerotia per inflorescence between sites based on history of ergot disease severity for both brome and other grass species ($P > 0.15$) (Fig. 1.7). Wheatgrass species had significantly more sclerotia per inflorescence at sites with a high history of ergot ($P \leq 0.001$) (Fig. 1.7). Overall, brome species showed the highest amount of sclerotia per inflorescence (median = 4, mean = 6.84 ± 9.16) with some inflorescences containing > 60 sclerotia. Other grass species showed the second highest amount of sclerotia per inflorescence (median = 3, mean = 3.19 ± 2.14) and wheatgrasses showed the lowest (median = 2, mean = 2.75 ± 2.66). Significant differences were only observed between brome and wheatgrass species ($P < 0.0001$) (Fig. 1.7), with a corrected $P = 0.081$ for the comparison between brome and other grass.

Inoculum estimates from grass groups

Inflorescence density, infected inflorescence density, and inoculum accumulation were estimated using inflorescence numbers, area coverage, and sclerotia per inflorescence values to gain insight into the magnitude of inoculum reservoirs associated with unmanaged grasses surrounding crop fields. Brome species had the highest estimated density of inflorescences at sites with a low and high history of ergot, while wheatgrass has the highest density at sites with no history of ergot (Fig. 1.8 A). The inflorescences of brome species were more often infected, except at sites with no ergot history (Fig. 1.8 B). Due to the high abundance (Fig. 1.5, 1.8 A) and

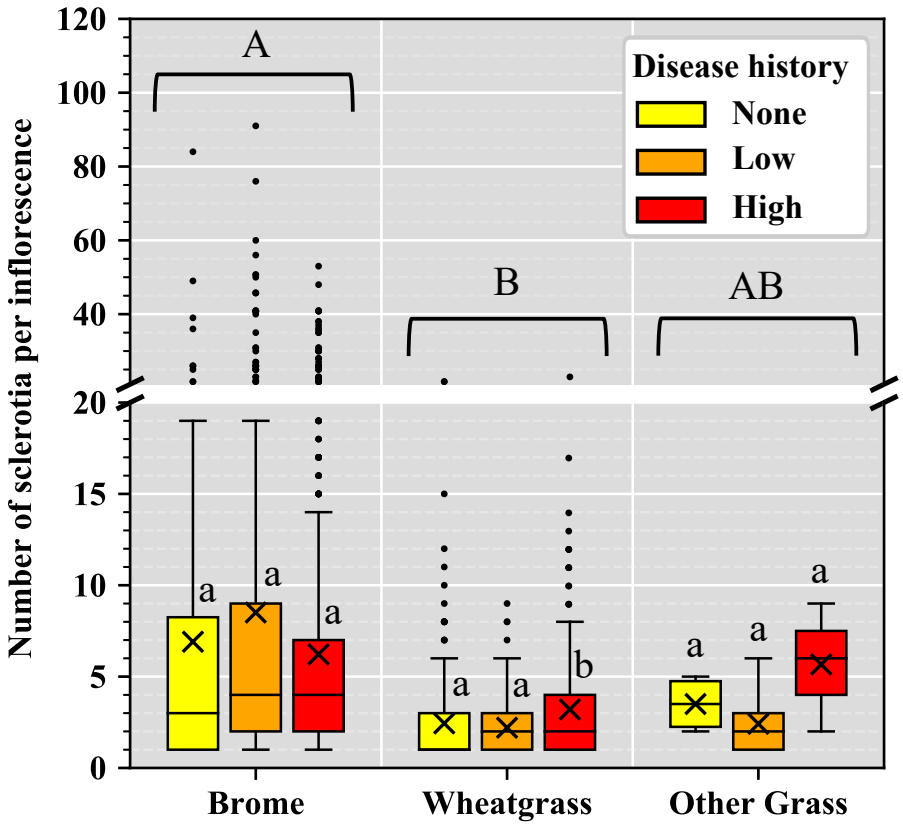


Figure 1.7: Boxplot distributions of the number of sclerotia per inflorescence from randomly sampled infected inflorescences from each quadrat during September surveys along ditch banks adjacent to barley fields in the San Luis Valley, CO. Horizontal lines within each boxplot represents median, while "X" represent mean. Mann-Whitney U test with Benjamini-Hochberg false discovery rate multi-test correction, was used for distribution separation. Letters not connected by the same letter are significantly different ($P \leq 0.05$); lower case letters correspond to differences between site ratings for each plant group, capital letters correspond to differences between plant groups. Disease history corresponds to the history of ergot disease severity in adjacent barley fields. Sample sizes: brome = 875, wheatgrass = 587, other grass = 21. It should be noted that the comparions between Brome and Other Grass was $P = 0.081$.

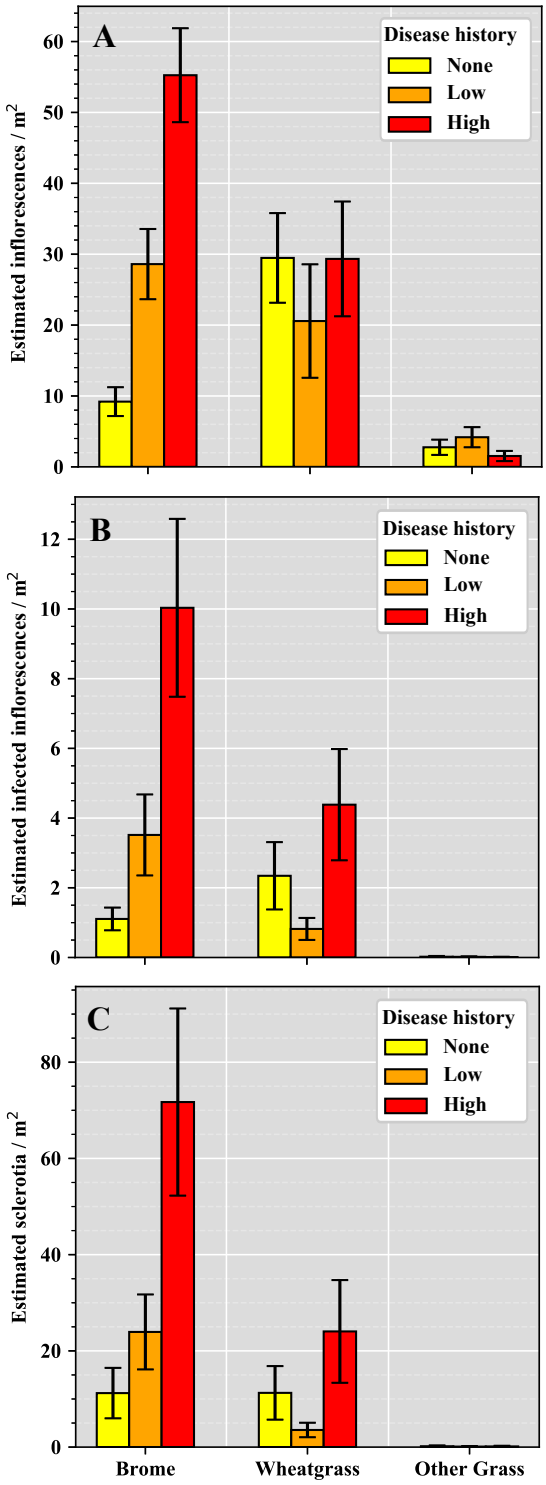


Figure 1.8: Mean estimates of each grass group for each site disease history rating for: **A**: inflorescence density; **B**: infected inflorescence density; **C**: sclerotia formation density per site within ditch banks adjacent to barley fields in the San Luis Valley, CO. Bars represents standard error.

high susceptibility (Fig. 1.7) of brome species we estimated high densities of sclerotia coming from brome, particularly at sites with high ergot history (71.7 ± 67.3 sclerotia/m^2) (Fig. 1.8 C). Summing sclerotia densities across all grass groups, at each disease history rating suggests that unmanaged grasses, along ditch banks, can produce upwards of 22.7, 27.6, 95.9 sclerotia/m^2 at sites with no, low, and high ergot history, respectively.

DISCUSSION

We identified factors that likely contribute to increased risk of ergot outbreaks by assessing unmanaged grass populations along ditch banks surrounding crop fields that varied in previous ergot disease incidence and severity. Sites that had a high history of ergot disease within the crop field also had ditch banks that were largely colonized with *Bromus* spp. (brome). In contrast, sites with no ergot history were surrounded by plant communities that were largely composed of non-grass species, which are not ergot hosts (Fig. 1.5). When looking only at grass species, our data suggests that brome species have a more influential impact than other grass species on ergot inoculum accumulation, and thus, ergot outbreak potential (Fig. 1.5-8).

Weedy and native plant species are known to interact with management of crop pathogens, such as the establishment of pathogen reservoirs in alternative hosts (Wisler and Norris 2005). Many of the known cases of suitable alternative hosts providing a significant impact to crop health regard bacterial or viral pathogens (Wisler and Norris 2005), however, it is well-known that fungal rust pathogens (i.e. *Puccina*, *Cronartium*, etc) require an alternative host to complete its lifecycle. Therefore, the control of these obligate alternative hosts of rusts are often integrated into rust management strategies (Wisler and Norris 2005). While *C. purpurea* does not need an alternative host to complete its lifecycle, alternative grass species may represent an example of a fungal pathogen reservoir that pose a significant threat to ergot management

strategies. In fact, our data revealed that comparison of community compositions (importance values) between grass species in ditch banks indicate that composition of brome within the plant community is correlated with the level of ergot disease incidence (Fig. 1.6). This was not the case for wheatgrass or other grass species, which had no correlation with disease incidence (Fig. 1.6) and showed similar levels of composition between sites of different ergot disease history (Fig. 1.5). Brome species also had the highest number of sclerotia per inflorescence regardless if the site had a history of ergot or not. This suggests that brome species have similar susceptibility patterns regardless of the history of ergot disease in adjacent crop fields. The lowest mean number of sclerotia per inflorescence in brome species was observed at sites with a history of high ergot incidence (Fig. 1.6) and the highest compositions of brome (Fig. 1.5), indicating that the increased composition of brome species might negatively impact infection potential. This may be the result of increased ground canopy cover which would influence the amount of ascospores (primary inoculum) escaping the canopy to reach inflorescences from sclerotia on the soil. Spread of conidia (secondary inoculum) should not be affected by increased plant densities since conidia are spread from inflorescence to inflorescence which are typically above the ground canopy; but could be affected if the plant communities are composed of grass species of varying heights. There were no significant differences in the number of sclerotia per brome inflorescence between sites with or without a history of ergot (Fig. 1.7). This implies that conidial spread is the primary factor in increased disease incidence leading to accumulation of inoculum reservoirs (Bayles *et al.* 2009). This is evident from the polycyclic nature of ergot which shows continual increases in disease incidence throughout the summer (Fig. 1.3, 1.4, 1.6). Overall, this suggests that the abundance of brome is a primary factor associated with the establishment of inoculum reservoirs. These brome reservoirs occurred more frequently around

fields with a history of high ergot disease incidence and severity, suggesting that the abundance of brome correlates to the potential of ergot outbreaks (Fig. 1.4, 1.5, 1.8).

Bayles *et al.* (2009) completed an extensive study on ergot epidemiology and the factors that grassy field margins have on ergot disease within wheat fields. In general, Bayles *et al.* (2009) suggested that ascospores released from sclerotia (primary inoculum) are unlikely to be significant sources of infection within wheat fields. Their attempts to capture ascospore production was generally unsuccessful even when continuously operating Burkard spore samplers were located next to buried sclerotia for three consecutive years. This suggests that ascospore dispersal is more widespread and randomly distributed around the sclerotia source (i.e. greater dispersal distance and not localized to the source). The few cases in which they were able to capture ascospores indicate that ascospores were released between April and June, prior to wheat anthesis. These cases correlated to observation of peak perithecia production in late May, which coincided with black-grass flowering (Bayles *et al.* 2009). Due to the synchrony of ascospore production, grasses flowering, and occurrence of rain that needs to be achieved for ascospore infection, it is believed that ascospore infection rates are rather low and are more likely to infect grasses surrounding crop fields which are composed of multiple species with varying flowering times (Brown 1947; Campbell and Freisin 1959; Rapilly 1968; Mantle *et al.* 1976; Alderman and Barker 2003; Bayles *et al.* 2009; Uppala *et al.* 2011). Additionally, grasses with more open florets (i.e. open-pollinated) would be more susceptible to ascospore infections than grasses with more closed florets (i.e. self-pollinating), as closed florets act as a physical barrier and have been shown to provide increased avoidance to infection (Bayles *et al.* 2009). These infected unmanaged grasses then represent local sources of secondary inoculum (conidia dispersal) which pose significant risk to cereal crops mainly situated at the edges of crop fields.

This is further exacerbated by increased incidence of late tiller development (Campbell 1957; Campbell and Freisin 1959; Mantle *et al.* 1977; Bayles *et al.* 2009).

Our data support the findings in Bayles *et al.* (2009) and provides new knowledge which expands our understanding of ergot epidemiology as we were able to correlate the history of ergot disease, within barley fields, to plant community compositions of adjacent ditch banks. While we did not examine ascospore production, our observations of July disease incidence suggest that ascospores were likely released during the month of June. This is evident as most of the infections noted in July were honeydew droplets, which typically form 1 - 2 weeks after infection, although some sclerotia were present in July suggesting that ascospores could have been released at the start of June. These patterns follow previous observations of ergot ascospore production (Brown 1947; Campbell and Freisin 1959; Rapilly 1968; Alderman 1993) and corresponds to anthesis of *Bromus inermis* (smooth brome) (Reynolds and Smith 1962; Looman 1983; Great Plains Flora Association 1986; Alex 1998). Smooth brome is generally self-sterile (Beddows 1931; Smith 1944; Cheng 1946; Adams 1953; Wilsie *et al.* 1952; McKone 1985), has open florets, produces pollen that loses viability within 24 hours (Domingo 1941), and has generally low seed set (~ 30%) (McKone 1985). These factors suggest smooth brome is highly susceptible to ergot infection (Sprague 1950; Campbell 1957) (Fig. 7). In addition, smooth brome continuously tillers throughout the summer and fall (Lamp 1952; Reynolds and Smith 1962; Eastin *et al.* 1964; Alex 1998) providing additional inflorescences susceptible to ergot infection. This is evident as we found that disease incidence was correlated with increases in brome composition with stronger, more significant, correlations occurring throughout the summer (Fig. 1.6). This contrasts with wheatgrasses, which were also observed to produce additional tillers throughout the summer but were not found to be correlated with ergot disease

incidence within unmanaged grass communities (Fig. 1.6). In addition, wheatgrasses often have more closed florets suggesting a lower potential for ergot infections (Bayles *et al.* 2009) (Fig. 1.7). These results indicate that smooth brome represents the primary source, from grass species examined, of local secondary inoculum which pose the greatest risk for the potential of ergot outbreaks within cereal crops, particularly to late emerging tillers at the edges of crop fields (Bayles *et al.* 2009). Eleven of the 14 infected barley tillers we observed were located within the first 9 m from the field edge. Even though disease is present in July, it appears that secondary spread of ergot from unmanaged grasses to the crop field is still largely dependent on favorable weather conditions or increased incidences in late tiller emergence, which were beyond the scope of this study. Our data reveals that these unmanaged grass communities experience yearly ergot infections, even in below average precipitation years (Fig. 1.2). While disease incidence in cereal crops fields remains sporadic largely due to the presence or absence of specific environmental variable such as precipitation, the amount of inoculum present in nearby unmanaged grass communities represent a consistent supply of primary and secondary inoculum. These reservoirs will result in ergot outbreaks in adjacent cereal crop fields in years when the weather is conducive. Therefore, management of wild and weedy grass communities should reduce the amount of primary inoculum available for the next season's disease cycle, which should also delay and reduce the incidence and severity of secondary inoculum produced throughout the summer. Further research is required to identify best practices for the management of wild and weedy field margins. In addition, research should continue to identify highly susceptible grass species that play significant roles in the accumulation of ergot inoculum surrounding cereal crop fields in different agricultural regions.

CONCLUSION

Overall, our results have shown that unmanaged grasses along ditch banks represent significant inoculum reservoirs of *Claviceps purpurea*, particularly if *Bromus* spp. are present. Even in years with limited precipitation, primary inoculum from sclerotia are able to infect susceptible unmanaged grasses surrounding crop fields, most notably open-pollinated grass species (Bayles *et al.* 2009). In addition, ergot inoculum accumulates within these unmanaged grass communities, throughout the summer (Fig. 1.3, 1.4) as continual tiller production and the spread of conidia (secondary inoculum) throughout these communities culminate in the establishment of large inoculum reservoirs. Estimates of inoculum production indicate these reservoirs could produce upwards of 90 sclerotia/m^2 (Fig. 1.8 C), representing the start of the disease cycle (primary inoculum) in the following spring. The spread of the primary inoculum into crop fields is generally infrequent due to dependence on favorable weather conditions and the synchrony of susceptible crop growth stages with inoculum presence. However, recent ergot outbreaks in 2015 and 2016 demonstrate that when these events do coordinate, there is ample inoculum from neighboring grass communities that provide the initial inoculum for significant disease spread and ergot production. Therefore, future research should focus on management strategies of these reservoirs to reduce the overall probability of ergot outbreaks within cereal crop fields which should be implemented alongside current whole-farm approaches for the control of ergot.

Chapter 2: Friend or Enemy? Greenhouse Inoculations reveal that *Claviceps purpurea* is a "Frenemy" with its Host

INTRODUCTION

Plant pathogens have traditionally been characterized as organisms that cause disease on plants, often leading to mortality or significant disruptions to plant health. These interactions are caused by the pathogen depleting host resources through different mechanisms depending on the pathogen's lifestyle (i.e. necrotrophic, biotrophic). Although, in the case of agricultural crops, plant pathogens include all parasitic organisms that negatively affect yield, quality, or profits. For instance, *Botrytis cinerea* (Pers.), the grey mold fungus, causes severe damage to the fruit during pre- and post-harvest, with occasional damage to other parts of the plant. While destruction of the fruit would provide a negative effect to the plant through reducing potential offspring, it could be argued that the most severe negative effects are felt as lost sellable products and profits by the grower. This results in the outsized importance of academic and industry research programs to develop novel resistant plants and fungicides to control this disease that rarely causes host mortality. Similarly, *Ustilago maydis* ((DC.) Corda, Icones), corn smut, is another important agricultural plant pathogen with primary damage to production being the infection of the corn ear, although, teliospore formation occurs on maize leaves which have been shown to negatively affect photosynthesis (Horst *et al.* 2008). These plant pathogens can disrupt the health of their host but generally have a more profound effect on the economically important portion of the plant (i.e. fruit). Interestingly, these fungal pathogens also claim a potential beneficial use. *Botrytis cinerea* is used in some aspects of wine production and the galls formed by *U. maydis* are edible and often farmed to produce huitlacoche (Tracy *et al.* 2007; Freeman

and Beattie 2008). Therefore, these pathogens can be seen as a friend or enemy depending on the positive or negative economic impact on the human and not the plant host.

Similarly, *Claviceps purpurea* (ergot), a grass fungal pathogen, causes significant impacts to agriculture and livestock production but also has human-driven beneficial aspects, such as the use of its secondary metabolites in pharmacology (Sorbe 1978; Tfelt-Hansen *et al*. 2000; Cincotta *et al.* 2005; Schiff 2006). While *C. purpurea* continues to have negative effects on agricultural and livestock production, recent studies question whether *C. purpurea* is more of a plant pathogen or symbiont in nature. These studies postulate that *C. purpurea* might be in a context-dependent mutualistic relationship with its host (Raybould *et al.* 1998; Fisher *et al.* 2007; Wäli *et al.* 2013). This suggests that the negative effects *C. purpurea* imposes on agricultural production is in context of its various toxic alkaloids that significantly disrupt human and animal health (ergotism) and not the negative effect it imposes on its host. Although, *Claviceps purpurea* does pose a negative effect on its grass host as the fungus infects unfertilized florets and develops a fungal mass (sclerotium) instead of a grass seed, however, this is generally seen as a weak effect. Ergot infections are fully restricted to individual ovaries and only access the plant's vascular system near the tip of the rachilla with no hyphae growth beyond that point (Tudzynski and Scheffer 2004). The fungus-host interaction also does not induce necrosis and actively manages to maintain host cell viability, to obtain nutrients from living tissue, likely through a complex cross-talk of fungal phytohormone production (Hinsch *et al.* 2015, 2016; Oeser *et al.* 2017; Kind *et al.* 2018a, 2018b). Furthermore, *C. purpurea* has an extensive range of hosts including the entire subfamily of Pooideae (Píchová *et al.* 2018). Despite this and the great interest in *Claviceps*, researchers had been generally unsuccessful in finding completely resistant varieties of crops or wild grasses (Wäli *et al.* 2013). It was only recently that 4 QTLs in durum

wheat were identified to reduce honeydew production and sclerotia weight (Gordon *et al.* 2020). While these QTLs can help reduce field level infections, they do not prevent *C. purpurea* from completing its lifecycle. These factors indicate that there does not appear to be strong selective pressure in the Pooideae family to gain resistance to *C. purpurea*, implying that the negative cost of ergot infection for the host is generally low (Wäli *et al.* 2013). While *C. purpurea* is often classified as a weak or mild plant pathogen, studies have indicated that *C. purpurea* infections contribute direct and indirect benefits to its host, revealing aspects of a mutualistic relationship (Raybould *et al.* 1998; Fisher *et al.* 2007; Wäli *et al.* 2013).

Wäli *et al.* (2013) demonstrated that sheep avoided forage containing *C. purpurea* and that the frequency of infected inflorescences was higher in pastures than surrounding ungrazed fields, suggesting that infected grasses have a protective effect against mammalian grazing. They further documented that the presence of ergot infection had no significant effect on successful seed production, as the proportion of ergot per inflorescence increased the proportion of successful seed decreased. This indicates that lower infection rates could benefit the host, while higher rates would be more detrimental. This is evident in other mutualistic relationships where the total fitness effects often vary between antagonistic to mutualistic depending on the conditions (Johnsson *et al.* 1997; Saikkonen *et al.* 1998, 2004; Leung and Poulin 2008; Rodriguez *et al.* 2012). Similarly, in wild salt marsh *Spartina* species infections of *C. spartina* result in negative to positive effects of seed set and weight, depending on the severity of ergot infections (Raybould *et al*. 1998; Fisher *et al.* 2007). Raybould *et al.* (1998) reported that inflorescences with a ≥ 10% infection were associated with reduction in seed set and lower seed weight, while inflorescences with < 10% infection showed an increase in seed set compared to uninfected. Similarly, Fisher *et al.* (2007) found that seed weight was higher in infected plants (0

- 10% infection severity) than uninfected plants in their 2002 survey, but not in their 2001 survey. Lastly, the populations of *Sporobolus anglicus* ((C.E. Hubb) P.M. Peterson & Saarela) (previously known as *Spartina anglica*) were not affected by persistent *C. spartina* infections; even between the years 1985 and 1995 where high infection levels (> 70% of inflorescences) were consistently reported (Raybould *et al*. 1998). This suggests that *Claviceps* infections do not significantly reduce the reproductive ability of *Spartina* populations, even during extreme outbreaks.

These studies concentrated their efforts on inflorescences and seed set where the perceived effects of fitness (gain or loss) would be most pronounced (Wäli *et al.* 2013). However, the plant species examined in these studies have sporadic seed set and are more often spread through vegetative growth of rhizomes, with no documentation of the effects *C. purpurea* could be having on other aspects of plant growth (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli *et al.* 2013). Furthermore, infected plant samples examined were taken from wild or cultivated populations of grasses across different habitats with infections likely occurring from multiple fungal strains, potentially confounding results. Our study aims to elucidate the effects *C. purpurea* has on its host through inoculations of a single isolate on two commercial cereal crops *Hordeum vulgare* (barley) and *Triticum aestivum* (wheat) in a controlled greenhouse setting. These results will provide a deeper understanding into the interactions occurring between *C. purpurea* and its grass host which will provide new insights into ergot epidemiology and direct future research studies of plant-microbe interactions.

MATERIALS AND METHODS

Greenhouse conditions

This study was conducted in the greenhouses at the Colorado State University Plant Growth Facility. Rooms were set to a 16 hr photoperiod, supplemented with 400-watt high pressure sodium lights, with temperatures of 18°C - 21°C, and no humidity control.

Inoculum preparation

Claviceps purpurea (isolate Clav04) was selected from pilot studies to be the most aggressive isolate available. Clav04 was originally isolated from *Bromus inermis* (smooth brome) from the San Luis Valley, Colorado, see Chapter 3 Material and Methods, pg. 67 for isolation method from sclerotia. Pure cultures were grown on Potato Dextrose Agar (PDA) and incubated at room temperature for two weeks. Plates were subsequently flooded with sterile DI H_2O and spores were dislodged using glass spreaders. The resulting suspensions were centrifuged at 4000 x *g* for 2 minutes. Supernatant was then discarded, spores were re-suspended in sterile DI H_2O, merged, and re-centrifuged to obtain a pellet. Pellets were suspended in a sterile 60% sucrose solution and stored at 4°C until used, with a maximum of one month (Menzies 2004). Prior to inoculations conidial suspensions were mixed with DI H_2O with 0.01% Tween 20 to make a final concentration of 10^6 conidia/mL.

Experimental design

This experiment was performed on barley (variety Moravian 69) from 2017 – 2019 and wheat (variety Oxen) from 2018 – 2020. Due to limited supply of barley, seeds were initially planted into plug trays. At three-leaf stage, barley seedlings were transplanted into 25.5 cm deep cone pots (650 mL) of fritted clay (Greens Grade) primed with 50 ppm of 24-18-6 (N/P/K) fertilizer (Miracle Grow) and ~350 mg (1/16 teaspoon) Mantra (Nufarm Americas Inc.). For

wheat, three wheat seeds were planted per primed cone pot and thinned to one seedling at the three-leaf stage. Each plant was fertilized weekly with 30 mL of 200 ppm 15-5-15 fertilizer (Cal-Mag Special; Plant Marvel Laboratories, Inc.) until the first round of inoculation and watered as needed throughout the experiment. For each variety three experimental trials, with 240 starting plants per trial, were performed and pooled together.

Despite controlled conditions, plants exhibited variations in growth rate which was more prominent in barley, most likely due to transplantation. Due to this and the infection requirements of *C. purpurea* (i.e. unpollinated florets and mature stigma) plants were blocked by their developmental stage and inoculated, in groups of 20, when applicable. In barley this corresponded to the boot stage, just prior to head emergence (Feekes 10) (Appendix 1 Fig. A1.1). In wheat this corresponded to just after head emergence but prior to anthesis (Feekes 10.1) (Appendix 1 Fig. A1.2). For inoculations, the spike of the primary (1°) tiller was carefully exposed (if applicable, i.e. barley). Florets were inoculated to have one of the following disease severities based on percent of spike infected; 5%, 15%, and 25%. The number of florets inoculated was rounded to the nearest whole number to best achieve the desired infection rate. Due to this, final infection rates varied around the desired rate, therefore, 5% represents an infection rates of 1 - 9%, 15% represents 10 - 19%, and 25% represents ≥ 20%. Uninfected plants (0%) acted as controls. Randomly chosen florets, excluding the top two florets, were syringe injected with a 10^6 conidia/mL suspension. In addition, the center floret in each wheat spikelet was never inoculated due to the delayed maturation rate of these florets. Tips of florets were cut to facilitate inoculation and reduce mechanical damage to the stigma and ovary. Inoculations were repeated the next day to help ensure infection. Plants were then placed in a generalized randomized block design (GRBD), 20 randomized treatment plants (5 replicates x 4

treatments) per block of inoculation day, with each block randomized on the bench. Plants that did not reach the required maturation stage 20 days after the first group was inoculated were not inoculated and were removed from the experiment. Plants from different inoculation days were harvested separately to keep the infection time period consistent.

Plant measurements

Inoculated plants (5 - 25% infection) that did not show any signs of infections (no sclerotia production) were not harvested for further analysis. In addition, plants with infections on secondary (2°) tillers, which could have been caused by dispersal of conidia from honeydew droplets of 1° tillers, were removed from the experiment. Glassine bags were not used to prevent spread from 1° tillers as preliminary tests show increased mortality and abortion of 1° tillers when used, particularly in barley where the heads removed from the boots were very weak. The amount of plants removed from conidial spread were low; more plants were removed due to inadequate infection and slow maturation rates. Harvested plants were measured for: sterile and fertile (filled) seeds per spike of 1° and 2° tillers; sclerotia and mean seed dry weight from 1° tiller; mean seed dry weight from 2° tillers; and dry plant and root biomass. Sterile seed counts of 1° tillers included sclerotia counts as ergot sclerotia, in nature, represent a non-fertile seed. Shoots were clipped at the soil interface and bagged in paper bags for drying. Roots were carefully extracted from fritted clay through an initial wash of a shower head sprayer and final wash in a water bath where all remaining shoot tissue were removed. All seeds, sclerotia, shoot, and root tissue were dried in an oven at 60°C for 4 days and weighed on an analytical scale. After weighing, eight randomly selected seeds per plant (4 from 1° tiller and 4 from 2° tillers) were sown into plug trays to examine germination rates, which were determined after 1 week of growth in general purpose potting soil.

Statistical analysis and software

Responses for plant tillers were broken down into tiller-level responses (1° and 2° tillers) and a plant-level response by grouping 1° and 2° tillers for each plant together. Each plant response parameter was analyzed using Mixed Model using the GRBD design in JMP Pro v13 (SAS Institute Inc.) using the following model (Model 1):

[1] Y = Inoculation day + Infection % + (Inoculation day * Infection %) + ε + Trial + Position[Trial]

Inoculation day, infection percent, and their interaction were treated as fixed effects and trial number and the position on the bench top nested within trial number as random effects. In addition, Tukey's honestly significant difference (HSD) was used to determine significance between infection rates using the least square means from the models.

In barley, the success of achieving desired infection rates was low (Fig. 2.1 A) with lack of success resulting from the inability of *C. purpurea* to colonize each inoculated floret. It is well known that inoculations of *C. purpurea* are difficult to achieve due to the small window in which host stigma are susceptible (Willingale and Mantle 1987), which is why most studies are performed on male sterile lines to increase the rate of infection. However, this approach would not be applicable to our experiment. Kind *et al.* (2018a) additionally discussed that emasculation of *Brachypodium distachyon* ((L.) P. Beauv.) prior to inoculation of *C. purpurea* often resulted in increases in necrotized ovaries due to mechanical wounding. Therefore, our approach sought to achieve sufficient statistical power by use of a large number of starting plants. Due to the lack of success in achieving infection rates we created two datasets of exact infection rates (Dataset 1; n = 302) where the final infection percent matched inoculation percent (Fig. 2.1 A) and a modified infection rate dataset (Dataset 2; n = 408) where the final infection percent replaced the inoculation percent (Fig. 2.1). For example, plants with a 15% inoculation percent but only had a 9% final infection rate were classified as 5% infection percent.

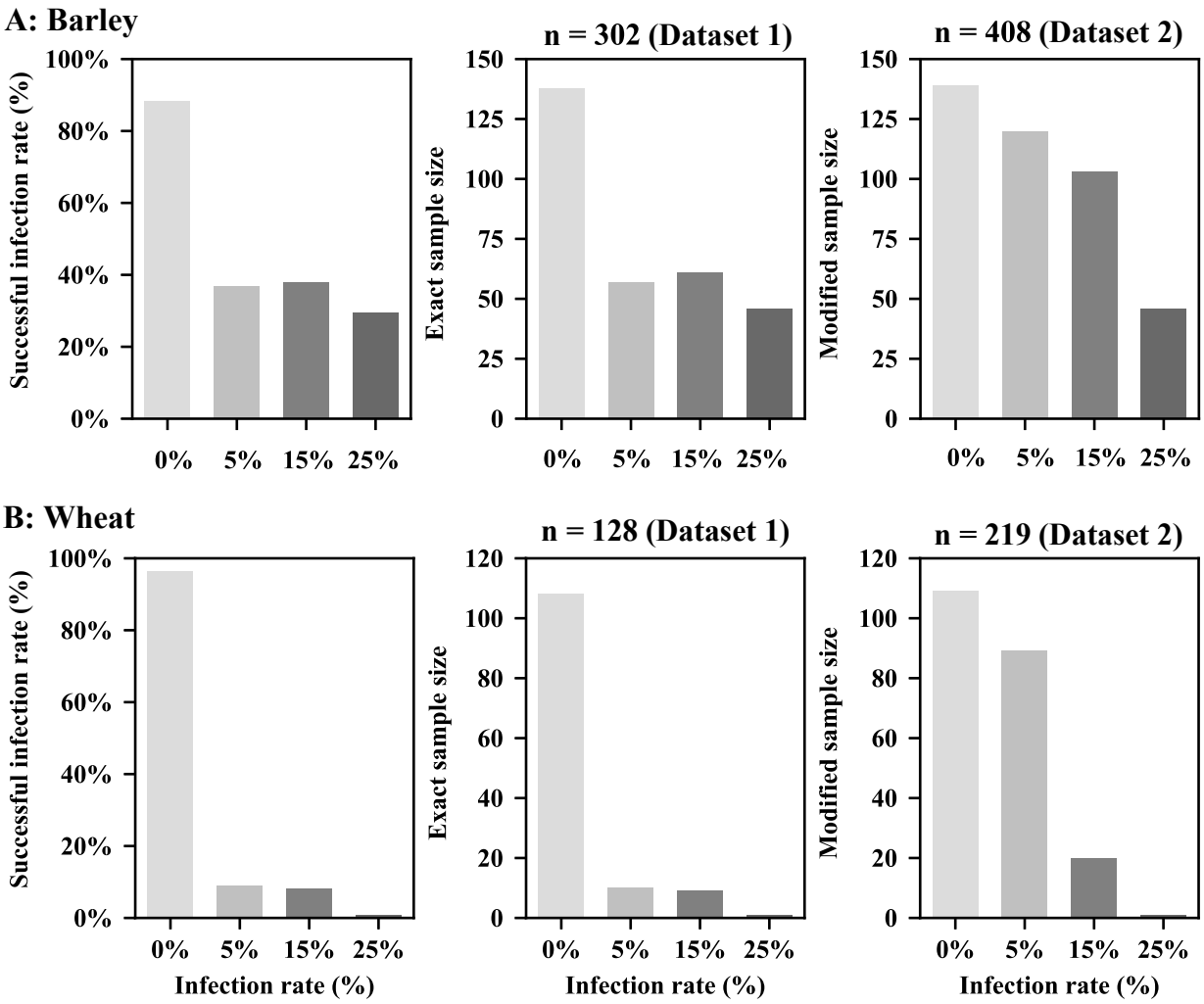


Figure 2.1: Bar graphs for each plant **A:** Barley and **B:** Wheat. Graphs show the percentage of plants with a final infection rate equal to the inoculation rate, the exact sample sizes for each successful infection rate (dataset 1), and the modified same size for each infection rate (dataset 2), respectively.

Any plant that had an initial inoculation but resulted in a final infection rate of 0% were not added to the controls, as they were not harvested.

In wheat, the success rate was more reduced (Fig. 2.1 B) with very few inoculated plants achieving the desired infection rate (Dataset 1; n = 128). By using the same modifying method, used in barley, we were able to increase the sample size of the 5% and 15% infection rates (Dataset 2; n = 219). Since there was only 1 plant with a final infection rate of 25%, this plant was removed from the dataset. We did not compute statistics for wheat dataset 1 due to the drastic differences in sample sizes between infected plants and controls.

Additional analyses and exploratory modeling were performed using the python modules scikit-learn v0.22.2, SciPy v1.3.1, R v3.6.1, and JMP Pro v13. Figures were created using the Python modules Matplotlib v3.1.1, seaborn v0.10.0, and R v3.6.1.

RESULTS

In general, most models showed improvements through decreases in Akaike information criterion with small sample correlation (AICc) and root mean squared error (RMSE), and increases in adjusted R^2 by using the modified sample sizes (Dataset 2) versus the exact sample sizes (Dataset 1). Due to this and the absence of models for Dataset 1 of wheat, we focused our discussion on the results of Dataset 2 for barley and wheat. However, all results from Dataset 1 of barley can be found in Appendix 1 Tables A1.1-6 and Figures A1.3, A1.4. We additionally utilized REML to observe the variance components of the random effects (trial and position[trial]) for each of the datasets and models. The percent of random effects varied greatly between dataset, model, plant variety, and plant measurements with trial generally having the most effect with a range of 0 – 58.5% and 0 – 39.1% of the proportion of total error in barley and wheat, respectively (Appendix Tables A1.5-10). The effects of position[trial] ranged between 0 –

17.2% and 0 – 24.9% of the proportion of total error in barley and wheat, respectively (Appendix Tables A1.5-10) The proportion of residual error ranged between 32.7 – 99.8% across all analyses (Appendix Tables A1.5-10). However, we were not concerned with the effects of inoculation treatments and the interactions between trial (i.e. year) and position on the bench. For this reason, and to reduce unnecessary and lengthy analysis, we did not analyses all combinations of models with the interactions of random effects as fixed effects and simply left these factors as random effects.

Model 1, initial model

Barley

There was variability in the fit of models for individual plant responses as adjusted R^2 values ranged from 0.056 for the seed weight of 2° tillers to 0.63 for 1° tiller fertility (Table 2.1, 2.2). Inoculation date, which corresponds to maturation rate, generally had more of an effect on plant growth with significant negative effects observed for total plant fertility, number of seeds per tiller, and 1° and 2° tiller seed weight (Table 2.1, 2.2). Demonstrating that late maturing plants have reduced growth compared to early maturing plants, which was expected. Infections of *C. purpurea* were only observed to have a significant effect on 1° tiller fertility where higher infection rates (≥ 15%) showed a significant decrease in fertility ($P < 0.001$; Table 2.2; Fig. 2.2), plants with a 5% infection rate were not significantly different than controls ($P > 0.05$). There was also a significant interactive effect of inoculation day and infection rate for 1° tiller fertility ($P = 0.014$) (Table 2.2). This interaction indicates that lower inoculation days (i.e. more vigorous plants) with lower infection rates showed a negative effect on 1° tiller fertility, while higher inoculation days with lower infection rates showed a positive effect on 1° tiller fertility. In some

Table 2.1: Mixed model results for Dataset 2 using Model 1 of plant-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.501	0.187	-85.9
Parameter Estimates							
Intercept	0.6360	0.1091	5.830	0.018			
Inoc. day	-0.0102	0.0040	-2.550	0.014			
Inf%[.05-0]	0.0288	0.0239	1.200	0.229			
Inf%[.15-.05]	-0.0504	0.0258	-1.950	0.052			
Inf%[.25-.15]	-0.0135	0.0345	-0.390	0.696			
(Inoc. day-7.46)*Inf%[.05-0]	0.0059	0.0040	1.450	0.147			
(Inoc. day-7.46)*Inf%[.15-.05]	-0.0019	0.0046	-0.410	0.684			
(Inoc. day-7.46)*Inf%[.25-.15]	0.0010	0.0062	0.170	0.866			
Fixed Effects							
Inoc. day			6.485	0.014			
Inf%			1.841	0.139			
Inoc. day*Inf%			0.820	0.484			
Plant Germination					0.282	0.302	270.1
Parameter Estimates							
Intercept	0.7295	0.1106	6.590	0.007			
Inoc. day	-0.0053	0.0058	-0.910	0.366			
Inf%[.05-0]	-0.0202	0.0396	-0.510	0.610			
Inf%[.15-.05]	-0.0027	0.0422	-0.070	0.948			
Inf%[.25-.15]	-0.0599	0.0578	-1.040	0.300			
(Inoc. day-7.23)*Inf%[.05-0]	0.0061	0.0068	0.890	0.372			
(Inoc. day-7.23)*Inf%[.15-.05]	0.0026	0.0075	0.350	0.725			
(Inoc. day-7.23)*Inf%[.25-.15]	-0.0052	0.0103	-0.500	0.617			
Fixed Effects							
Inoc. day			0.829	0.366			
Inf%			0.752	0.522			
Inoc. day*Inf%			0.543	0.653			
Root Biomass					0.150	0.097	-626.8
Parameter Estimates							
Intercept	0.1997	0.0164	12.160	<0.001			
Inoc. day	-0.0020	0.0017	-1.230	0.221			
Inf%[.05-0]	0.0081	0.0124	0.650	0.514			
Inf%[.15-.05]	-0.0009	0.0134	-0.070	0.948			
Inf%[.25-.15]	0.0001	0.0178	0.000	0.997			
(Inoc. day-7.46)*Inf%[.05-0]	0.0014	0.0021	0.690	0.494			
(Inoc. day-7.46)*Inf%[.15-.05]	0.0002	0.0024	0.070	0.947			
(Inoc. day-7.46)*Inf%[.25-.15]	0.0056	0.0032	1.760	0.080			
Fixed Effects							
Inoc. day			1.519	0.221			
Inf%			0.183	0.908			
Inoc. day*Inf%			2.046	0.107			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.103	0.006	-2715.1
Parameter Estimates							
Intercept	0.0374	0.0016	23.520	<0.001			
Inoc. day	-0.0002	0.0001	-1.610	0.110			
Inf%[.05-0]	0.0003	0.0008	0.350	0.723			
Inf%[.15-.05]	-0.0007	0.0008	-0.920	0.359			
Inf%[.25-.15]	-0.0007	0.0011	-0.640	0.519			
(Inoc. day-7.23)*Inf%[.05-0]	0.0001	0.0001	0.790	0.431			
(Inoc. day-7.23)*Inf%[.15-.05]	0.0000	0.0001	-0.010	0.994			
(Inoc. day-7.23)*Inf%[.25-.15]	0.0001	0.0002	0.640	0.526			
Fixed Effects							
Inoc. day			2.585	0.110			
Inf%			0.734	0.532			
Inoc. day*Inf%			0.578	0.630			
Avg. Seed/Tiller					0.475	4.246	2411.8
Parameter Estimates							
Intercept	13.0798	2.3479	5.570	0.019			
Inoc. day	-0.1915	0.0898	-2.130	0.038			
Inf%[.05-0]	0.7618	0.5430	1.400	0.162			
Inf%[.15-.05]	-1.0354	0.5867	-1.760	0.078			
Inf%[.25-.15]	-0.4526	0.7826	-0.580	0.563			
(Inoc. day-7.46)*Inf%[.05-0]	0.1695	0.0918	1.850	0.066			
(Inoc. day-7.46)*Inf%[.15-.05]	-0.0809	0.1056	-0.770	0.444			
(Inoc. day-7.46)*Inf%[.25-.15]	-0.0014	0.1406	-0.010	0.992			
Fixed Effects							
Inoc. day			4.550	0.038			
Inf%			1.735	0.159			
Inoc. day*Inf%			1.155	0.327			
Plant Biomass					0.321	0.619	859.7
Parameter Estimates							
Intercept	1.8904	0.2591	7.300	0.007			
Inoc. day	-0.0145	0.0118	-1.240	0.221			
Inf%[.05-0]	0.0762	0.0791	0.960	0.336			
Inf%[.15-.05]	0.0028	0.0855	0.030	0.974			
Inf%[.25-.15]	-0.1071	0.1139	-0.940	0.348			
(Inoc. day-7.46)*Inf%[.05-0]	0.0049	0.0133	0.370	0.715			
(Inoc. day-7.46)*Inf%[.15-.05]	-0.0021	0.0153	-0.140	0.891			
(Inoc. day-7.46)*Inf%[.25-.15]	0.0181	0.0205	0.880	0.377			
Fixed Effects							
Inoc. day			1.528	0.221			
Inf%			0.605	0.612			
Inoc. day*Inf%			0.428	0.733			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table 2.2: Mixed model results for Dataset 2 using Model 1 of tiller-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.630	0.198	-34.9
Parameter Estimates							
Intercept	0.5100	0.1553	3.28	0.069			
Inoc. day	-0.0071	0.0045	-1.58	0.120			
Inf%[.05-0]	-0.0277	0.0253	-1.10	0.274			
Inf%[.15-.05]	-0.0428	0.0274	-1.56	0.119			
Inf%[.25-.15]	-0.0582	0.0365	-1.60	0.112			
(Inoc. day-7.46)*Inf%[.05-0]	0.0127	0.0043	2.96	0.003			
(Inoc. day-7.46)*Inf%[.15-.05]	-0.0014	0.0049	-0.28	0.779			
(Inoc. day-7.46)*Inf%[.25-.15]	-0.0034	0.0066	-0.53	0.600			
Fixed Effects							
Inoc. day			2.508	0.120			
Inf%			5.708	<0.001			
Inoc. day*Inf%			3.572	0.014			
Avg. 1° Tiller Seed Weight					0.114	0.006	-2250.9
Parameter Estimates							
Intercept	0.0421	0.0020	20.840	<0.001			
Inoc. day	-0.0003	0.0001	-2.770	0.006			
Inf%[.05-0]	0.0003	0.0009	0.350	0.730			
Inf%[.15-.05]	-0.0004	0.0009	-0.450	0.655			
Inf%[.25-.15]	-0.0016	0.0014	-1.170	0.243			
(Inoc. day-6.89)*Inf%[.05-0]	0.0001	0.0001	0.600	0.551			
(Inoc. day-6.89)*Inf%[.15-.05]	0.0001	0.0002	0.590	0.559			
(Inoc. day-6.89)*Inf%[.25-.15]	0.0000	0.0002	0.160	0.871			
Fixed Effects							
Inoc. day			7.700	0.006			
Inf%			0.759	0.518			
Inoc. day*Inf%			0.612	0.608			
1° Tiller Germination					0.065	0.255	109.4
Parameter Estimates							
Intercept	0.9051	0.0547	16.540	<0.001			
Inoc. day	-0.0071	0.0041	-1.730	0.085			
Inf%[.05-0]	-0.0377	0.0352	-1.070	0.285			
Inf%[.15-.05]	0.0328	0.0378	0.870	0.385			
Inf%[.25-.15]	0.0002	0.0572	0.000	0.998			
(Inoc. day-6.89)*Inf%[.05-0]	0.0026	0.0058	0.440	0.659			
(Inoc. day-6.89)*Inf%[.15-.05]	0.0026	0.0066	0.390	0.694			
(Inoc. day-6.89)*Inf%[.25-.15]	0.0073	0.0094	0.780	0.439			
Fixed Effects							
Inoc. day			2.988	0.085			
Inf%			0.446	0.720			
Inoc. day*Inf%			0.749	0.524			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.367	0.195	-52.0
Parameter Estimates							
Intercept	0.7012	0.0982	7.140	0.012			
Inoc. day	-0.0006	0.0037	-0.170	0.862			
Inf%[.05-0]	0.0198	0.0274	0.720	0.471			
Inf%[.15-.05]	0.0116	0.0296	0.390	0.694			
Inf%[.25-.15]	0.0342	0.0403	0.850	0.398			
(Inoc. day-6.39)*Inf%[.05-0]	-0.0047	0.0053	-0.880	0.381			
(Inoc. day-6.39)*Inf%[.15-.05]	0.0047	0.0060	0.780	0.436			
(Inoc. day-6.39)*Inf%[.25-.15]	-0.0008	0.0074	-0.100	0.919			
Fixed Effects							
Inoc. day			0.030	0.862			
Inf%			1.112	0.344			
Inoc. day*Inf%			0.301	0.825			
Avg. 2° Tiller Seed Weight					0.056	0.007	-2226.6
Parameter Estimates							
Intercept	0.0347	0.0010	34.140	<0.001			
Inoc. day	-0.0004	0.0001	-3.300	0.001			
Inf%[.05-0]	-0.0001	0.0009	-0.060	0.955			
Inf%[.15-.05]	-0.0015	0.0010	-1.520	0.129			
Inf%[.25-.15]	0.0012	0.0013	0.930	0.351			
(Inoc. day-6.32)*Inf%[.05-0]	0.0000	0.0002	0.220	0.827			
(Inoc. day-6.32)*Inf%[.15-.05]	0.0000	0.0002	0.190	0.846			
(Inoc. day-6.32)*Inf%[.25-.15]	0.0002	0.0002	0.920	0.361			
Fixed Effects							
Inoc. day			10.874	0.001			
Inf%			1.075	0.360			
Inoc. day*Inf%			0.602	0.614			
2° Tiller Germination					0.202	0.414	435.3
Parameter Estimates							
Intercept	0.6441	0.1158	5.560	0.004			
Inoc. day	-0.0117	0.0088	-1.330	0.186			
Inf%[.05-0]	-0.0242	0.0597	-0.410	0.685			
Inf%[.15-.05]	0.0076	0.0642	0.120	0.905			
Inf%[.25-.15]	-0.0371	0.0866	-0.430	0.669			
(Inoc. day-6.32)*Inf%[.05-0]	0.0025	0.0117	0.220	0.830			
(Inoc. day-6.32)*Inf%[.15-.05]	0.0052	0.0129	0.400	0.691			
(Inoc. day-6.32)*Inf%[.25-.15]	-0.0052	0.0159	-0.320	0.746			
Fixed Effects							
Inoc. day			1.779	0.186			
Inf%			0.1572	0.925			
Inoc. day*Inf%			0.1433	0.934			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

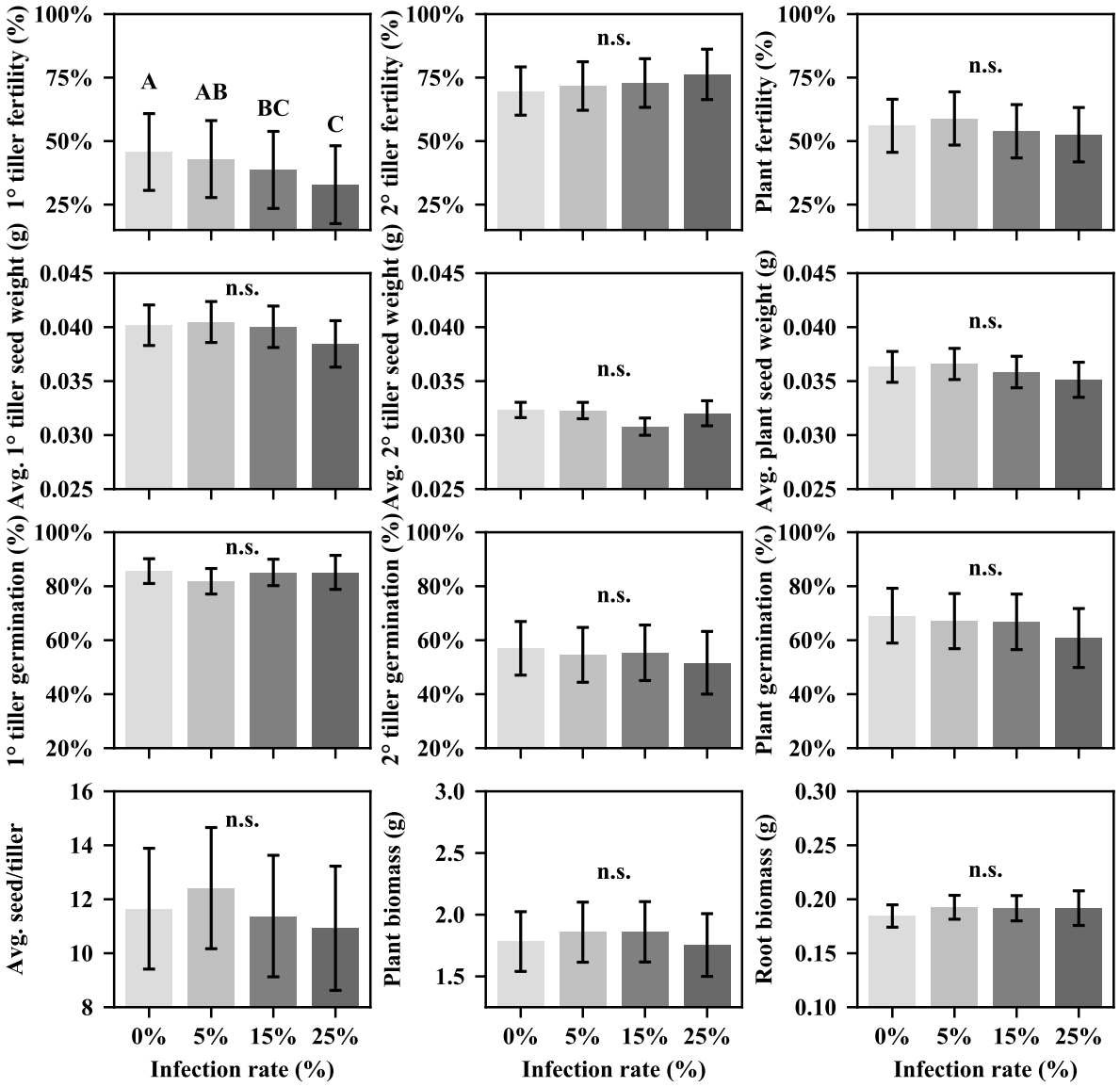


Figure 2.2: Least square means of plant responses for each infection rate in barley using Dataset 2 and Model 1. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

cases, plant responses such as root biomass, plant biomass, and germination rates showed no significant effects of any fixed effect (P >0.05; Table 2.1, 2.2).

Wheat

Model fit was also variable for plant responses as adjusted R^2 ranged from -0.017 for 1° tiller germination rates to 0.467 for plant biomass (Table 2.3, 2.4). Plant responses in wheat were generally comparable to those observed in barley, as many results showed that infection rate did not have any significant effect on plant growth and that inoculation day was often observed to have more of an effect on plant growth than infection rate (Tables 2.1-4). However in wheat, infection rate was shown to have more significant effects than in barley. As infection rates increased, significant effects were observed in the decrease of 1° tiller fertility, plant fertility, and the number of seeds per tiller ($P \leq 0.001$; Table 2.3, 2.4; Fig. 2.3). However, we found that 1° tiller seed weights were significantly higher in plants with a 5% infection rate compared to controls ($P = 0.0291$; Table 2.3, 2.4; Fig. 2.3). In addition, there was a significant interactive effect of inoculation day and infection rate on root biomass ($P = 0.037$) (Table 2.3). This interaction indicates that a lower inoculation day and lower infection rate had a negative effect, while a higher inoculation day and lower infection rate had a positive effect on root biomass.

Effects of sclerotia weight

Claviceps purpurea is an obligate biotrophic organism which completes its life cycle on its host and thus needs to obtain nutrients from its host to survive. This could suggest that if a plant had more ergot (through incidence or weight of sclerotia) there would be less nutrients utilized by the plant. Therefore, we investigated whether the final weight of sclerotia from each plant influenced plant responses. We first examined this relationship through linear regressions for barley (Fig. 2.4) and wheat (Fig. 2.5).

Table 2.3: Mixed model results for Dataset 2 using Model 1 of plant-level responses in wheat.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.139	0.125	-229.4
Parameter Estimates							
Intercept	0.6292	0.0454	13.870	0.025			
Inoc. day	-0.0020	0.0034	-0.590	0.558			
Inf%[.05-0]	-0.0676	0.0181	-3.740	<.001			
Inf%[.15-.05]	0.0072	0.0334	0.210	0.830			
(Inoc. day-3.66)*Inf%[.05-0]	-0.0089	0.0053	-1.660	0.098			
(Inoc. day-3.66)*Inf%[.15-.05]	0.0046	0.0117	0.400	0.691			
Random Effect							
Inoc. day			0.345	0.558			
Inf%			7.337	0.001			
Inoc. day*Inf%			1.393	0.251			
Plant Germination					0.038	0.104	-305.0
Parameter Estimates							
Intercept	0.9629	0.0220	43.860	0.000			
Inoc. day	0.0012	0.0030	0.390	0.698			
Inf%[.05-0]	0.0167	0.0152	1.100	0.272			
Inf%[.15-.05]	-0.0524	0.0281	-1.860	0.064			
(Inoc. day-3.66)*Inf%[.05-0]	-0.0024	0.0045	-0.530	0.595			
(Inoc. day-3.66)*Inf%[.15-.05]	0.0085	0.0098	0.870	0.383			
Random Effect							
Inoc. day			0.151	0.698			
Inf%			1.936	0.147			
Inoc. day*Inf%			0.420	0.658			
Root Biomass					0.311	0.083	-382.6
Parameter Estimates							
Intercept	0.2329	0.0169	13.750	<.001			
Inoc. day	-0.0109	0.0030	-3.690	0.001			
Inf%[.05-0]	0.0053	0.0121	0.440	0.664			
Inf%[.15-.05]	0.0245	0.0229	1.070	0.285			
(Inoc. day-3.66)*Inf%[.05-0]	0.0090	0.0036	2.500	0.013			
(Inoc. day-3.66)*Inf%[.15-.05]	-0.0010	0.0080	-0.130	0.899			
Random Effect							
Inoc. day			13.613	0.001			
Inf%			0.874	0.419			
Inoc. day*Inf%			3.354	0.037			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.183	0.003	-1242.8
Parameter Estimates							
Intercept	0.0289	0.0011	26.430	0.009			
Inoc. day	-0.0002	0.0001	-1.810	0.073			
Inf%[.05-0]	0.0008	0.0005	1.750	0.082			
Inf%[.15-.05]	-0.0006	0.0011	-0.510	0.609			
(Inoc. day-2.97)*Inf%[.05-0]	0.0003	0.0001	2.270	0.025			
(Inoc. day-2.97)*Inf%[.15-.05]	-0.0005	0.0007	-0.670	0.504			
Random Effect							
Inoc. day			3.293	0.073			
Inf%			1.553	0.215			
Inoc. day*Inf%			2.655	0.074			
Avg. Seed/Tiller					0.111	3.967	1073.5
Parameter Estimates							
Intercept	18.8315	0.6513	28.910	<.001			
Inoc. day	-0.0357	0.1543	-0.230	0.818			
Inf%[.05-0]	-2.1064	0.6148	-3.430	0.001			
Inf%[.15-.05]	-1.6618	1.4072	-1.180	0.239			
(Inoc. day-2.93)*Inf%[.05-0]	-0.3253	0.2157	-1.510	0.133			
(Inoc. day-2.93)*Inf%[.15-.05]	-0.7560	1.0061	-0.750	0.453			
Random Effect							
Inoc. day			0.053	0.818			
Inf%			7.824	0.001			
Inoc. day*Inf%			1.571	0.211			
Plant Biomass					0.467	0.462	358.7
Parameter Estimates							
Intercept	1.5261	0.1798	8.490	0.025			
Inoc. day	-0.0689	0.0248	-2.780	0.010			
Inf%[.05-0]	0.0192	0.0679	0.280	0.778			
Inf%[.15-.05]	-0.0056	0.1289	-0.040	0.965			
(Inoc. day-3.66)*Inf%[.05-0]	0.0349	0.0202	1.730	0.086			
(Inoc. day-3.66)*Inf%[.15-.05]	-0.0288	0.0448	-0.640	0.522			
Random Effect							
Inoc. day			7.713	0.010			
Inf%			0.040	0.960			
Inoc. day*Inf%			1.496	0.227			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table 2.4: Mixed model results for Dataset 2 using Model 1 of tiller-level responses in wheat.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.289	0.132	-199.9
Parameter Estimates							
Intercept	0.7078	0.0268	26.440	0.000			
Inoc. day	-0.0058	0.0050	-1.160	0.260			
Inf%[.05-0]	-0.1518	0.0194	-7.840	<.001			
Inf%[.15-.05]	0.0333	0.0368	0.910	0.367			
(Inoc. day-3.66)*Inf%[.05-0]	-0.0040	0.0057	-0.700	0.485			
(Inoc. day-3.66)*Inf%[.15-.05]	0.0098	0.0126	0.780	0.437			
Random Effect							
Inoc. day			1.347	0.260			
Inf%			31.341	<.001			
Inoc. day*Inf%			0.420	0.658			
Avg. 1° Tiller Seed Weight					0.093	0.003	-1810.5
Parameter Estimates							
Intercept	0.0302	0.0010	31.140	0.007			
Inoc. day	0.0000	0.0001	0.230	0.819			
Inf%[.05-0]	0.0011	0.0004	2.570	0.011			
Inf%[.15-.05]	-0.0003	0.0008	-0.400	0.688			
(Inoc. day-3.66)*Inf%[.05-0]	0.0003	0.0001	2.410	0.017			
(Inoc. day-3.66)*Inf%[.15-.05]	-0.0001	0.0003	-0.460	0.645			
Random Effect							
Inoc. day			0.053	0.819			
Inf%			3.354	0.037			
Inoc. day*Inf%			2.959	0.054			
1° Tiller Germination					-0.017	0.079	-425.2
Parameter Estimates							
Intercept	0.9868	0.0135	72.890	<.001			
Inoc. day	-0.0001	0.0026	-0.050	0.960			
Inf%[.05-0]	0.0031	0.0117	0.260	0.793			
Inf%[.15-.05]	0.0099	0.0218	0.460	0.648			
(Inoc. day-3.66)*Inf%[.05-0]	0.0025	0.0034	0.730	0.463			
(Inoc. day-3.66)*Inf%[.15-.05]	-0.0022	0.0074	-0.300	0.765			
Random Effect							
Inoc. day			0.003	0.960			
Inf%			0.175	0.839			
Inoc. day*Inf%			0.273	0.762			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.243	0.126	-133.4
Parameter Estimates							
Intercept	0.5677	0.0763	7.440	0.066			
Inoc. day	0.0015	0.0056	0.270	0.791			
Inf%[.05-0]	0.0205	0.0220	0.930	0.354			
Inf%[.15-.05]	-0.0359	0.0542	-0.660	0.509			
(Inoc. day-2.97)*Inf%[.05-0]	-0.0117	0.0071	-1.660	0.100			
(Inoc. day-2.97)*Inf%[.15-.05]	-0.0405	0.0337	-1.200	0.232			
Random Effect							
Inoc. day			0.071	0.791			
Inf%			0.551	0.577			
Inoc. day*Inf%			2.345	0.100			
Avg. 2° Tiller Seed Weight					0.237	0.003	-1183.8
Parameter Estimates							
Intercept	0.0277	0.0015	18.500	0.017			
Inoc. day	-0.0002	0.0001	-1.450	0.150			
Inf%[.05-0]	0.0005	0.0006	0.860	0.390			
Inf%[.15-.05]	-0.0009	0.0014	-0.640	0.525			
(Inoc. day-2.97)*Inf%[.05-0]	0.0002	0.0002	0.880	0.380			
(Inoc. day-2.97)*Inf%[.15-.05]	-0.0008	0.0008	-0.890	0.377			
Random Effect							
Inoc. day			2.111	0.150			
Inf%			0.485	0.617			
Inoc. day*Inf%			0.703	0.497			
2° Tiller Germination					0.132	0.226	32.4
Parameter Estimates							
Intercept	0.9195	0.0652	14.100	0.006			
Inoc. day	-0.0013	0.0108	-0.120	0.904			
Inf%[.05-0]	0.0587	0.0395	1.490	0.139			
Inf%[.15-.05]	-0.1808	0.0982	-1.840	0.068			
(Inoc. day-2.69)*Inf%[.05-0]	-0.0062	0.0127	-0.480	0.629			
(Inoc. day-2.69)*Inf%[.15-.05]	0.0287	0.0606	0.470	0.637			
Random Effect							
Inoc. day			0.015	0.904			
Inf%			2.365	0.098			
Inoc. day*Inf%			0.206	0.814			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

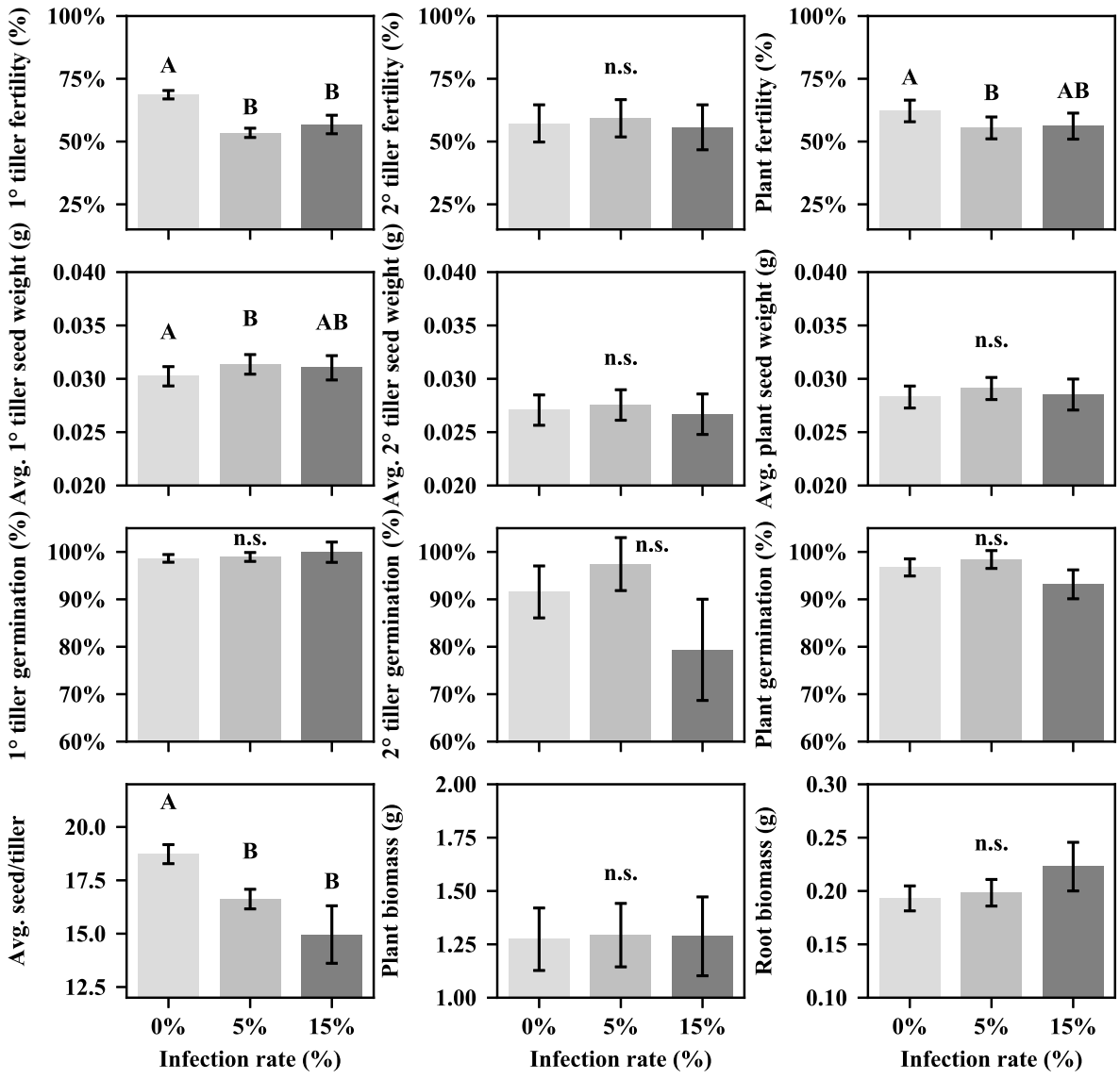


Figure 2.3: Least square means of plant responses for each infection rate in wheat using Dataset 2 and Model 1. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

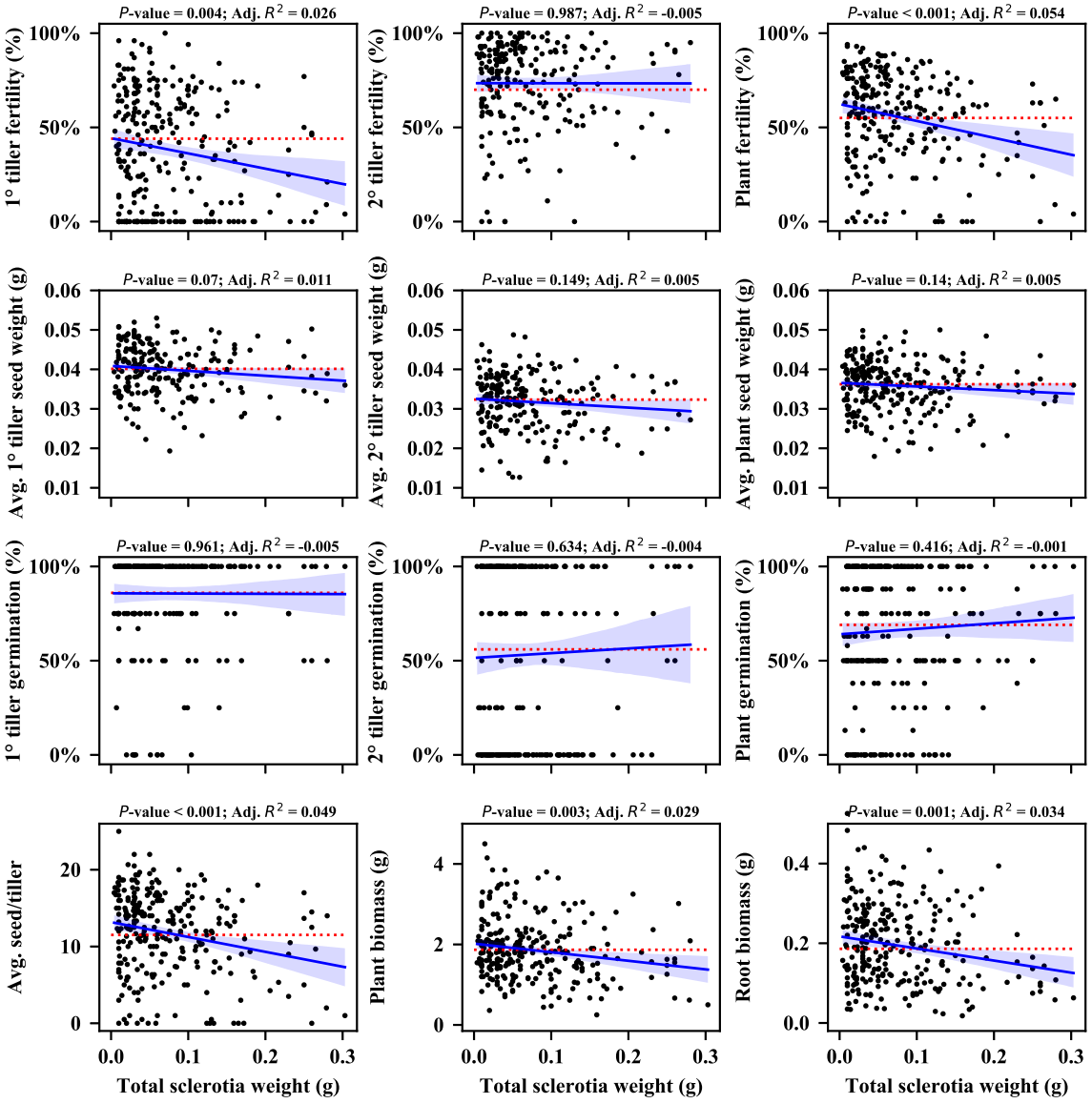


Figure 2.4: Linear regressions for the effect of total sclerotia weight on multiple plant responses in barley. Shadowed region signifies 95% confidence intervals. Dotted red line indicated arithmetic mean of controls. Adjusted R^2 and *P*-values are imposed on top of corresponding plots.

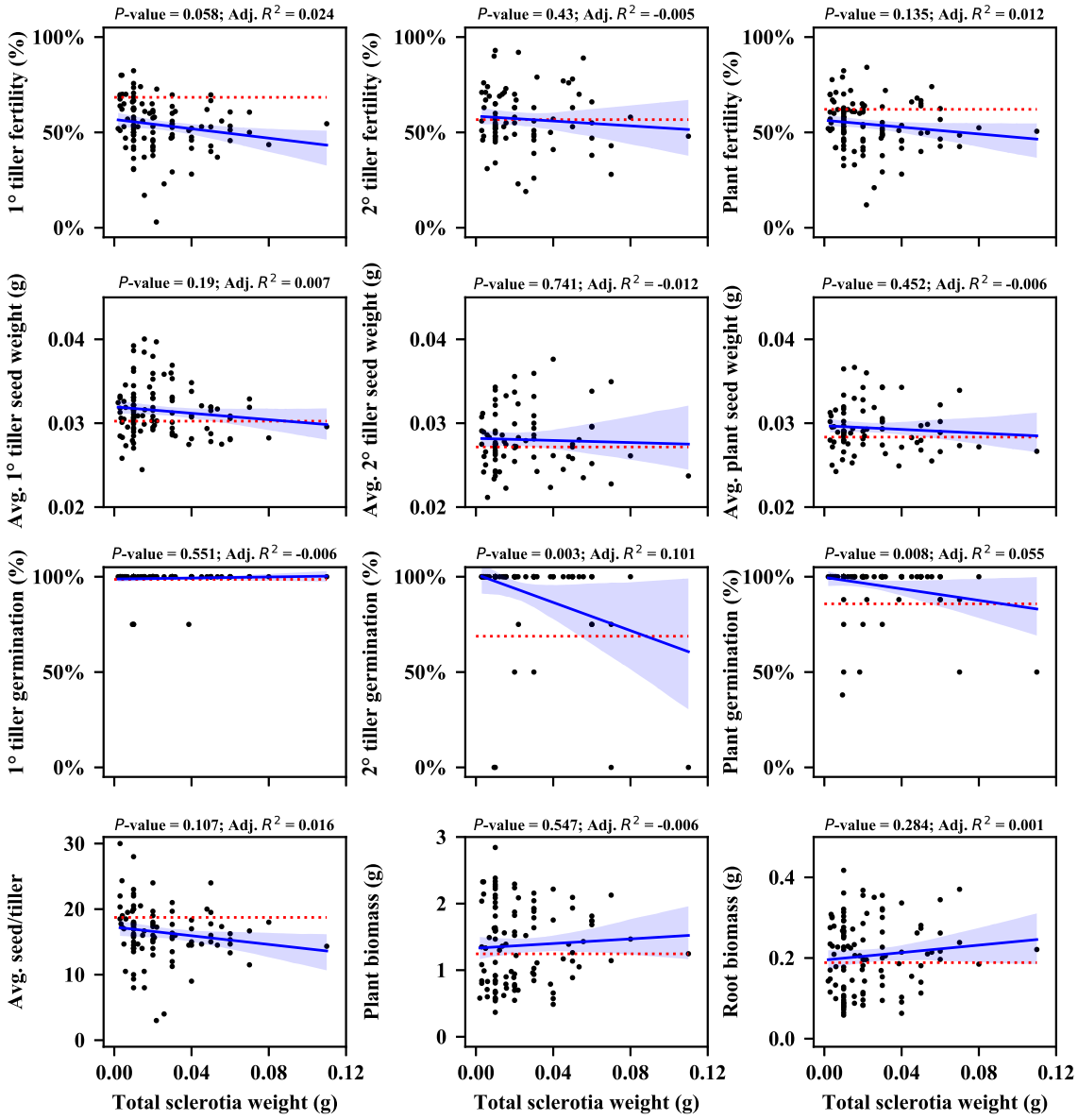


Figure 2.5: Linear regressions for the effect of total sclerotia weight on multiple plant responses in wheat. Shadowed region signifies 95% confidence intervals. Dotted red line indicated arithmetic mean of controls. Adjusted R^2 and *P*-values are imposed on top of corresponding plots.

In barley, total sclerotia weight was found to be significantly negatively correlated with 1° tiller fertility, plant fertility, number of seeds per tiller, plant biomass, and root biomass ($P < 0.01$; Fig. 2.4), with 1° tiller seed weight nearly significant ($P = 0.07$). In wheat, 2° tiller germination rates and plant germination rates had significant negative correlations ($P < 0.01$; Fig. 2.5), while 1° tiller fertility was near significant ($P = 0.058$). Although, adjusted R^2 values were low in both plant species ranging from -0.006 to 0.101 (Fig. 2.4, 2.5). This data indicates that heavier sclerotia produced on plants result in reductions of plant growth, potentially through the nutrient acquisition by the biotrophic organism.

Model 2, inclusion of sclerotia weight

Due to correlative evidence of sclerotia weight with plant responses and the biotrophic lifestyle of *C. purpurea* we re-analyzed our data through an exploratory model (Model 2) by adding total sclerotia weight as a fixed effect:

[2] Y = Inoculation day + Infection % + (Inoculation day * Infection %) + Total sclerotia weight + ε + Trial + Position[Trial]

Due to linear dependencies we were unable to examine the interaction between sclerotia weight and infection rate. The addition of sclerotia weight in Model 2 showed improvement in many plant responses compared to Model 1 through increased adjusted R^2 values and decreased RMSE and AICc values (Tables 2.1-8), but generally both models were comparable. However, in Model 2 sclerotia weight was often found to have a significant effect ($P < 0.05$) on plant responses and its addition to the model resulted in the increased observations of significant effects ($P < 0.05$) from infection rate (Tables 2.1-8), primarily in barley.

<u>Barley:</u>

Results from Model 2 were comparable to those found in Model 1 (Tables 2.1, 2.2, 2.5, 2.6), however, we observed some key differences in the results of Model 2. The most notable

Table 2.5: Mixed model results for Dataset 2 using Model 2 of plant-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.520	0.183	-98.1
Parameter Estimates							
Intercept	0.6363	0.1104	5.760	0.018			
Inoc. day	-0.0102	0.0039	-2.580	0.013			
Inf%[.05-0]	0.0654	0.0252	2.590	0.010			
Inf%[.15-.05]	-0.0014	0.0282	-0.050	0.961			
Inf%[.25-.15]	0.0102	0.0343	0.300	0.766			
(Inoc. day-7.46)*Inf%[.05-0]	0.0065	0.0040	1.640	0.101			
(Inoc. day-7.46)*Inf%[.15-.05]	0.0014	0.0046	0.300	0.761			
(Inoc. day-7.46)*Inf%[.25-.15]	-0.0023	0.0061	-0.380	0.705			
Total sclerotia weight	-0.8934	0.2250	-3.970	<0.001			
Fixed Effects							
Inoc. day			6.652	0.013			
Inf%			2.421	0.066			
Inoc. day*Inf%			1.475	0.221			
Total sclerotia weight			15.768	<0.001			
Plant Germination					0.286	0.301	269.5
Parameter Estimates							
Intercept	0.7292	0.1097	6.650	0.007			
Inoc. day	-0.0053	0.0058	-0.910	0.366			
Inf%[.05-0]	-0.0464	0.0424	-1.100	0.274			
Inf%[.15-.05]	-0.0368	0.0466	-0.790	0.431			
Inf%[.25-.15]	-0.0753	0.0583	-1.290	0.197			
(Inoc. day-7.23)*Inf%[.05-0]	0.0056	0.0068	0.830	0.408			
(Inoc. day-7.23)*Inf%[.15-.05]	0.0003	0.0076	0.040	0.967			
(Inoc. day-7.23)*Inf%[.25-.15]	-0.0026	0.0104	-0.250	0.800			
Total sclerotia weight	0.6398	0.3757	1.700	0.089			
Fixed Effects							
Inoc. day			0.828	0.366			
Inf%			1.693	0.168			
Inoc. day*Inf%			0.312	0.817			
Total sclerotia weight			2.901	0.089			
Root Biomass					0.176	0.096	-635.6
Parameter Estimates							
Intercept	0.1997	0.0162	12.350	<0.001			
Inoc. day	-0.0020	0.0016	-1.250	0.214			
Inf%[.05-0]	0.0256	0.0131	1.960	0.051			
Inf%[.15-.05]	0.0227	0.0147	1.550	0.122			
Inf%[.25-.15]	0.0116	0.0178	0.650	0.516			
(Inoc. day-7.46)*Inf%[.05-0]	0.0017	0.0021	0.840	0.402			
(Inoc. day-7.46)*Inf%[.15-.05]	0.0018	0.0024	0.740	0.461			
(Inoc. day-7.46)*Inf%[.25-.15]	0.0040	0.0032	1.260	0.210			
Total sclerotia weight	-0.4301	0.1169	-3.680	<0.001			
Fixed Effects							
Inoc. day			1.567	0.214			
Inf%			3.256	0.022			
Inoc. day*Inf%			2.557	0.055			
Total sclerotia weight			13.548	<0.001			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.103	0.006	-2706.7
Parameter Estimates							
Intercept	0.0374	0.0016	23.260	<0.001			
Inoc. day	-0.0002	0.0001	-1.620	0.107			
Inf%[.05-0]	0.0007	0.0008	0.810	0.419			
Inf%[.15-.05]	-0.0002	0.0009	-0.280	0.781			
Inf%[.25-.15]	-0.0005	0.0011	-0.420	0.674			
(Inoc. day-7.23)*Inf%[.05-0]	0.0001	0.0001	0.840	0.400			
(Inoc. day-7.23)*Inf%[.15-.05]	0.0000	0.0001	0.230	0.816			
(Inoc. day-7.23)*Inf%[.25-.15]	0.0001	0.0002	0.440	0.660			
Total sclerotia weight	-0.0096	0.0072	-1.330	0.185			
Fixed Effects							
Inoc. day			2.639	0.107			
Inf%			0.302	0.824			
Inoc. day*Inf%			0.688	0.560			
Total sclerotia weight			1.767	0.185			
Avg. Seed/Tiller					0.492	4.179	2395.3
Parameter Estimates							
Intercept	13.0855	2.3739	5.510	0.020			
Inoc. day	-0.1915	0.0884	-2.170	0.035			
Inf%[.05-0]	1.5387	0.5741	2.680	0.008			
Inf%[.15-.05]	0.0071	0.6423	0.010	0.991			
Inf%[.25-.15]	0.0549	0.7822	0.070	0.944			
(Inoc. day-7.46)*Inf%[.05-0]	0.1828	0.0904	2.020	0.044			
(Inoc. day-7.46)*Inf%[.15-.05]	-0.0104	0.1056	-0.100	0.922			
(Inoc. day-7.46)*Inf%[.25-.15]	-0.0722	0.1397	-0.520	0.606			
Total sclerotia weight	-18.999	5.1250	-3.710	<0.001			
Fixed Effects							
Inoc. day			4.690	0.035			
Inf%			2.574	0.054			
Inoc. day*Inf%			1.698	0.167			
Total sclerotia weight			13.743	<0.001			
Plant Biomass					0.351	0.606	845.8
Parameter Estimates							
Intercept	1.8905	0.2551	7.410	0.006			
Inoc. day	-0.0145	0.0119	-1.210	0.230			
Inf%[.05-0]	0.1956	0.0833	2.350	0.019			
Inf%[.15-.05]	0.1603	0.0931	1.720	0.086			
Inf%[.25-.15]	-0.0323	0.1134	-0.280	0.776			
(Inoc. day-7.46)*Inf%[.05-0]	0.0072	0.0131	0.550	0.584			
(Inoc. day-7.46)*Inf%[.15-.05]	0.0087	0.0153	0.570	0.568			
(Inoc. day-7.46)*Inf%[.25-.15]	0.0066	0.0202	0.330	0.743			
Total sclerotia weight	-2.8854	0.7420	-3.890	<0.001			
Fixed Effects							
Inoc. day			1.475	0.230			
Inf%			3.797	0.011			
Inoc. day*Inf%			0.730	0.535			
Total sclerotia weight			15.123	<0.001			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table 2.6: Mixed model results for Dataset 2 using Model 2 of tiller-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.636	0.196	-40.9
Parameter Estimates							
Intercept	0.5103	0.1564	3.260	0.071			
Inoc. day	-0.0071	0.0044	-1.620	0.112			
Inf%[.05-0]	0.0020	0.0269	0.070	0.940			
Inf%[.15-.05]	-0.0028	0.0301	-0.090	0.925			
Inf%[.25-.15]	-0.0386	0.0367	-1.050	0.293			
(Inoc. day-7.46)*Inf%[.05-0]	0.0131	0.0042	3.100	0.002			
(Inoc. day-7.46)*Inf%[.15-.05]	0.0013	0.0050	0.260	0.792			
(Inoc. day-7.46)*Inf%[.25-.15]	-0.0061	0.0066	-0.930	0.352			
Total sclerotia weight	-0.7290	0.2406	-3.030	0.003			
Fixed Effects							
Inoc. day			2.620	0.112			
Inf%			0.427	0.734			
Inoc. day*Inf%			4.556	0.004			
Total sclerotia weight			9.181	0.003			
Avg. 1° Tiller Seed Weight					0.125	0.006	-2245.7
Parameter Estimates							
Intercept	0.0421	0.0020	20.720	<0.001			
Inoc. day	-0.0003	0.0001	-2.830	0.005			
Inf%[.05-0]	0.0011	0.0009	1.150	0.252			
Inf%[.15-.05]	0.0005	0.0010	0.470	0.637			
Inf%[.25-.15]	-0.0010	0.0014	-0.740	0.458			
(Inoc. day-6.89)*Inf%[.05-0]	0.0001	0.0001	0.690	0.488			
(Inoc. day-6.89)*Inf%[.15-.05]	0.0002	0.0002	0.930	0.354			
(Inoc. day-6.89)*Inf%[.25-.15]	0.0000	0.0002	-0.080	0.935			
Total sclerotia weight	-0.0177	0.0081	-2.200	0.029			
Fixed Effects							
Inoc. day			7.987	0.005			
Inf%			0.783	0.504			
Inoc. day*Inf%			0.975	0.405			
Total sclerotia weight			4.828	0.029			
1° Tiller Germination					0.063	0.256	111.9
Parameter Estimates							
Intercept	0.9049	0.0549	16.490	<0.001			
Inoc. day	-0.0071	0.0041	-1.720	0.086			
Inf%[.05-0]	-0.0416	0.0381	-1.090	0.276			
Inf%[.15-.05]	0.0284	0.0413	0.690	0.493			
Inf%[.25-.15]	-0.0027	0.0583	-0.050	0.963			
(Inoc. day-6.89)*Inf%[.05-0]	0.0025	0.0058	0.430	0.668			
(Inoc. day-6.89)*Inf%[.15-.05]	0.0023	0.0067	0.350	0.729			
(Inoc. day-6.89)*Inf%[.25-.15]	0.0076	0.0095	0.800	0.425			
Total sclerotia weight	0.0891	0.3340	0.270	0.790			
Fixed Effects							
Inoc. day			2.964	0.086			
Inf%			0.461	0.710			
Inoc. day*Inf%			0.718	0.542			
Total sclerotia weight			0.071	0.790			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.366	0.195	-49.4
Parameter Estimates							
Intercept	0.7012	0.0986	7.110	0.012			
Inoc. day	-0.0006	0.0037	-0.170	0.863			
Inf%[.05-0]	0.0255	0.0295	0.860	0.388			
Inf%[.15-.05]	0.0185	0.0324	0.570	0.568			
Inf%[.25-.15]	0.0394	0.0416	0.950	0.344			
(Inoc. day-6.39)*Inf%[.05-0]	-0.0047	0.0053	-0.870	0.383			
(Inoc. day-6.39)*Inf%[.15-.05]	0.0053	0.0061	0.870	0.386			
(Inoc. day-6.39)*Inf%[.25-.15]	-0.0014	0.0076	-0.190	0.848			
Total sclerotia weight	-0.1468	0.2784	-0.530	0.598			
Fixed Effects							
Inoc. day			0.030	0.863			
Inf%			0.912	0.435			
Inoc. day*Inf%			0.330	0.804			
Total sclerotia weight			0.278	0.598			
Avg. 2° Tiller Seed Weight					0.056	0.007	-2217.1
Parameter Estimates							
Intercept	0.0347	0.0010	34.410	<0.001			
Inoc. day	-0.0004	0.0001	-3.280	0.001			
Inf%[.05-0]	0.0001	0.0010	0.150	0.883			
Inf%[.15-.05]	-0.0013	0.0011	-1.180	0.240			
Inf%[.25-.15]	0.0014	0.0014	1.040	0.299			
(Inoc. day-6.32)*Inf%[.05-0]	0.0000	0.0002	0.210	0.832			
(Inoc. day-6.32)*Inf%[.15-.05]	0.0001	0.0002	0.300	0.761			
(Inoc. day-6.32)*Inf%[.25-.15]	0.0002	0.0002	0.810	0.417			
Total sclerotia weight	-0.0051	0.0092	-0.550	0.581			
Fixed Effects							
Inoc. day			10.789	0.001			
Inf%			0.672	0.570			
Inoc. day*Inf%			0.618	0.604			
Total sclerotia weight			0.306	0.581			
2° Tiller Germination					0.199	0.414	436.2
Parameter Estimates							
Intercept	0.6441	0.1150	5.600	0.004			
Inoc. day	-0.0117	0.0088	-1.340	0.185			
Inf%[.05-0]	-0.0399	0.0642	-0.620	0.535			
Inf%[.15-.05]	-0.0105	0.0701	-0.150	0.881			
Inf%[.25-.15]	-0.0516	0.0894	-0.580	0.564			
(Inoc. day-6.32)*Inf%[.05-0]	0.0024	0.0117	0.210	0.836			
(Inoc. day-6.32)*Inf%[.15-.05]	0.0035	0.0132	0.260	0.793			
(Inoc. day-6.32)*Inf%[.25-.15]	-0.0033	0.0162	-0.210	0.837			
Total sclerotia weight	0.3938	0.6007	0.660	0.513			
Fixed Effects							
Inoc. day			1.789	0.185			
Inf%			0.302	0.824			
Inoc. day*Inf%			0.081	0.971			
Total sclerotia weight			0.430	0.513			

† t ratio for parameter estimates, F ration for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

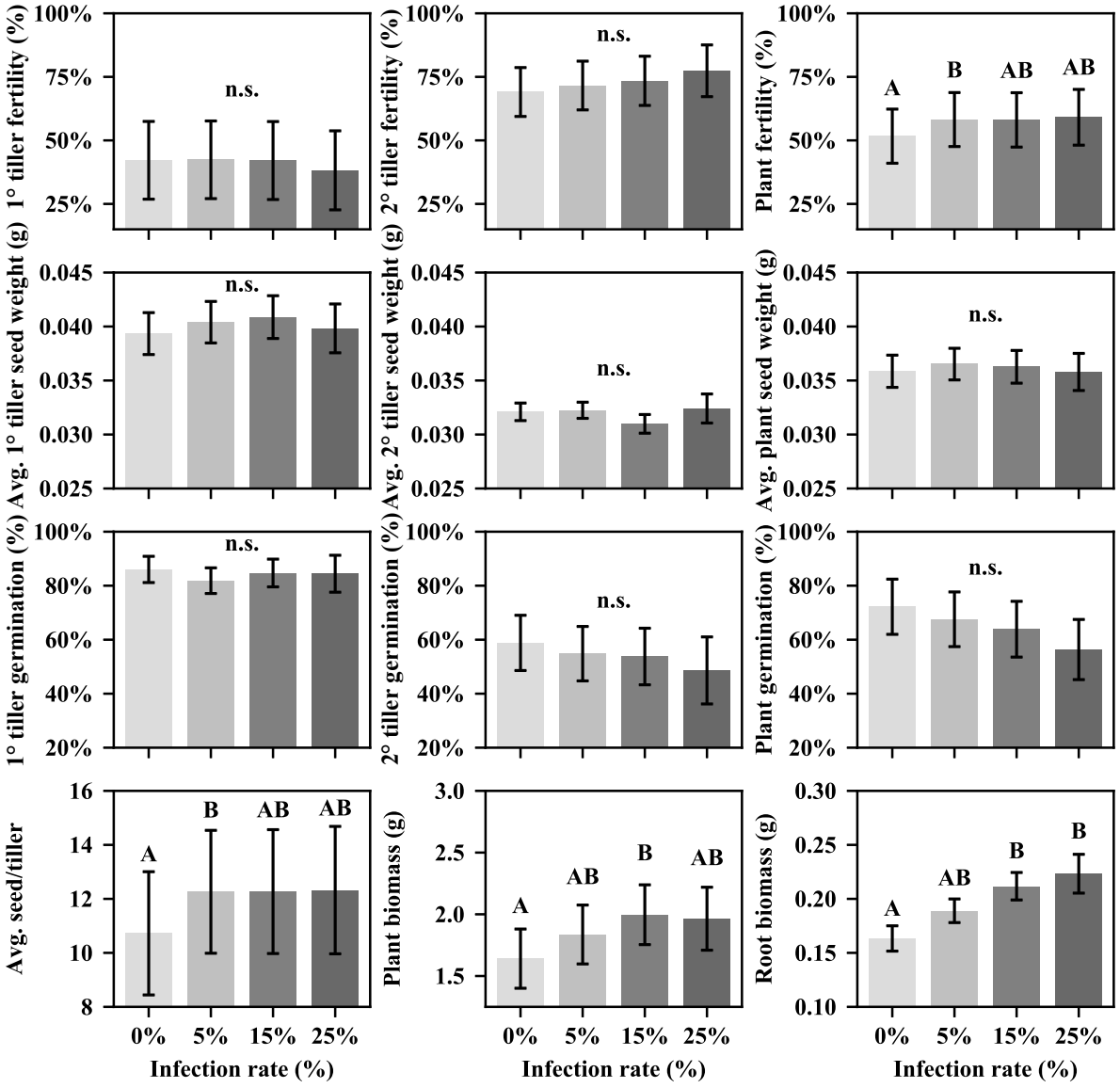


Figure 2.6: Least square means of plant responses for each infection rate in barley using Dataset 2 and Model 2. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

differences were the lack of significant effects of infection rate on 1° tiller fertility ($P > 0.05$; Table 2.6; Fig. 2.6) and the positive effects of increased infection rate on total plant fertility, number of seeds per tiller, and plant and root biomass (Table 2.5; Fig 2.6), which were often significantly higher than controls ($P < 0.05$). There was also a near significant effect ($P = 0.055$) of the interaction of inoculation day and infection rate (Table 2.5).. Similar results were found using Dataset 1 except that total plant fertility, number of seeds per tiller, and plant biomass were not found to be significantly higher in infected plants than controls ($P > 0.05$; Appendix 1 Tables A1.7, A1.8; Fig. A1.4).

Wheat:

Results from Model 2 of wheat were very comparable to the results from Model 1 (Table 2.3, 2.4, 2.7, 2.8; Fig. 2.3, 2.7). Similar to Model 1, 1° tiller fertility was found to be significantly lower in plants with a 5% and 15% infection rate compared to controls and plants with a 5% infection rate showed an increase in 1° tiller seed weight ($P < 0.05$; Fig. 2.7). However, in Model 2 both plant fertility and the number of seeds per tiller were not significantly lower in infected plants compared to controls ($P > 0.05$; Fig. 2.7).

Table 2.7: Mixed model results for Dataset 2 using Model 2 of plant-level responses in wheat.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.149	0.124	-229.4
Parameter Estimates							
Intercept	0.6297	0.0461	13.650	0.026			
Inoc. day	-0.0021	0.0034	-0.630	0.531			
Inf%[.05-0]	-0.0453	0.0220	-2.060	0.041			
Inf%[.15-.05]	0.0290	0.0366	0.790	0.429			
(Inoc. day-3.67)*Inf%[.05-0]	-0.0099	0.0054	-1.860	0.064			
(Inoc. day-3.67)*Inf%[.15-.05]	0.0036	0.0117	0.310	0.758			
Total_sclerotia_weight	-1.2127	0.6857	-1.770	0.078			
Random Effect							
Inoc. day			0.393	0.531			
Inf%			2.465	0.088			
Inoc. day*Inf%			1.765	0.174			
Total_sclerotia_weight			3.128	0.078			
Plant Germination					0.054	0.104	-305.6
Parameter Estimates							
Intercept	0.9636	0.0225	42.830	0.001			
Inoc. day	0.0010	0.0030	0.340	0.731			
Inf%[.05-0]	0.0375	0.0184	2.030	0.043			
Inf%[.15-.05]	-0.0331	0.0307	-1.080	0.282			
(Inoc. day-3.66)*Inf%[.05-0]	-0.0034	0.0045	-0.760	0.450			
(Inoc. day-3.66)*Inf%[.15-.05]	0.0077	0.0097	0.790	0.431			
Total sclerotia weight	-1.1276	0.5724	-1.970	0.050			
Random Effect							
Inoc. day			0.119	0.731			
Inf%			2.710	0.069			
Inoc. day*Inf%			0.460	0.632			
Total sclerotia weight			3.881	0.050			
Root Biomass					0.313	0.083	-378.7
Parameter Estimates							
Intercept	0.2329	0.0171	13.620	<.001			
Inoc. day	-0.0109	0.0030	-3.650	0.001			
Inf%[.05-0]	-0.0028	0.0148	-0.190	0.850			
Inf%[.15-.05]	0.0174	0.0251	0.690	0.490			
(Inoc. day-3.67)*Inf%[.05-0]	0.0094	0.0036	2.590	0.010			
(Inoc. day-3.67)*Inf%[.15-.05]	-0.0005	0.0080	-0.070	0.946			
Total sclerotia weight	0.4441	0.4650	0.950	0.341			
Random Effect							
Inoc. day			13.352	0.001			
Inf%			0.256	0.774			
Inoc. day*Inf%			3.625	0.028			
Total sclerotia weight			0.912	0.341			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.182	0.003	-1225.6
Parameter Estimates							
Intercept	0.0289	0.0011	26.820	0.009			
Inoc. day	-0.0002	0.0001	-1.850	0.069			
Inf%[.05-0]	0.0011	0.0006	2.040	0.044			
Inf%[.15-.05]	0.0001	0.0012	0.080	0.934			
(Inoc. day-2.98)*Inf%[.05-0]	0.0003	0.0001	2.140	0.034			
(Inoc. day-2.98)*Inf%[.15-.05]	-0.0004	0.0007	-0.530	0.600			
Total_sclerotia_weight	-0.0168	0.0160	-1.050	0.295			
Random Effect							
Inoc. day			3.411	0.069			
Inf%			2.081	0.129			
Inoc. day*Inf%			2.331	0.101			
Total_sclerotia_weight			1.105	0.295			
Avg. Seed/Tiller					0.129	3.941	1058.9
Parameter Estimates							
Intercept	18.8323	0.6488	29.020	<.001			
Inoc. day	-0.0359	0.1536	-0.230	0.816			
Inf%[.05-0]	-1.2574	0.7427	-1.690	0.092			
Inf%[.15-.05]	-0.7039	1.5156	-0.460	0.643			
(Inoc. day-2.93)*Inf%[.05-0]	-0.3644	0.2153	-1.690	0.092			
(Inoc. day-2.93)*Inf%[.15-.05]	-0.6867	1.0011	-0.690	0.494			
Total sclerotia weight	-44.472	22.158	-2.010	0.046			
Random Effect							
Inoc. day			0.055	0.816			
Inf%			1.518	0.222			
Inoc. day*Inf%			1.830	0.164			
Total sclerotia weight			4.028	0.046			
Plant Biomass					0.464	0.464	356.4
Parameter Estimates							
Intercept	1.5261	0.1808	8.440	0.026			
Inoc. day	-0.0688	0.0250	-2.760	0.011			
Inf%[.05-0]	0.0030	0.0832	0.040	0.971			
Inf%[.15-.05]	-0.0149	0.1419	-0.100	0.917			
(Inoc. day-3.67)*Inf%[.05-0]	0.0357	0.0204	1.750	0.082			
(Inoc. day-3.67)*Inf%[.15-.05]	-0.0282	0.0452	-0.620	0.533			
Total sclerotia weight	0.8804	2.6107	0.340	0.736			
Random Effect							
Inoc. day			7.603	0.011			
Inf%			0.006	0.994			
Inoc. day*Inf%			1.534	0.218			
Total sclerotia weight			0.114	0.736			

† t ratio for parameter estimates, F ratio for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table 2.8: Mixed model results for Dataset 2 using Model 2 of tiller-level responses in wheat.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.312	0.130	-205.1
Parameter Estimates							
Intercept	0.7094	0.0269	26.380	<.001			
Inoc. day	-0.0062	0.0048	-1.300	0.206			
Inf%[.05-0]	-0.1149	0.0233	-4.930	<.001			
Inf%[.15-.05]	0.0696	0.0397	1.750	0.081			
(Inoc. day-3.67)*Inf%[.05-0]	-0.0058	0.0057	-1.020	0.310			
(Inoc. day-3.67)*Inf%[.15-.05]	0.0077	0.0124	0.620	0.536			
Total sclerotia weight	-2.0044	0.7256	-2.760	0.006			
Random Effect							
Inoc. day			1.695	0.206			
Inf%			13.914	<.001			
Inoc. day*Inf%			0.572	0.565			
Total sclerotia weight			7.630	0.006			
Avg. 1° Tiller Seed Weight					0.096	0.003	-1794.5
Parameter Estimates							
Intercept	0.0302	0.0010	31.530	0.007			
Inoc. day	0.0000	0.0001	0.210	0.837			
Inf%[.05-0]	0.0015	0.0005	2.770	0.006			
Inf%[.15-.05]	0.0002	0.0009	0.230	0.822			
(Inoc. day-3.66)*Inf%[.05-0]	0.0003	0.0001	2.270	0.024			
(Inoc. day-3.66)*Inf%[.15-.05]	-0.0002	0.0003	-0.570	0.569			
Total sclerotia weight	-0.0189	0.0165	-1.150	0.252			
Random Effect							
Inoc. day			0.042	0.837			
Inf%			3.849	0.023			
Inoc. day*Inf%			2.581	0.078			
Total sclerotia weight			1.318	0.252			
1° Tiller Germination					-0.021	0.080	-420.1
Parameter Estimates							
Intercept	0.9868	0.0136	72.500	<.001			
Inoc. day	-0.0001	0.0026	-0.050	0.964			
Inf%[.05-0]	0.0013	0.0143	0.090	0.930			
Inf%[.15-.05]	0.0079	0.0241	0.330	0.744			
(Inoc. day-3.67)*Inf%[.05-0]	0.0026	0.0034	0.750	0.453			
(Inoc. day-3.67)*Inf%[.15-.05]	-0.0021	0.0075	-0.280	0.778			
Total sclerotia weight	0.0991	0.4403	0.230	0.822			
Random Effect							
Inoc. day			0.002	0.964			
Inf%			0.056	0.946			
Inoc. day*Inf%			0.284	0.753			
Total sclerotia weight			0.051	0.822			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.243	0.127	-130.9
Parameter Estimates							
Intercept	0.5680	0.0769	7.390	0.066			
Inoc. day	0.0014	0.0057	0.250	0.805			
Inf%[.05-0]	0.0329	0.0270	1.220	0.225			
Inf%[.15-.05]	-0.0203	0.0606	-0.340	0.738			
(Inoc. day-2.98)*Inf%[.05-0]	-0.0123	0.0071	-1.720	0.087			
(Inoc. day-2.98)*Inf%[.15-.05]	-0.0384	0.0340	-1.130	0.261			
Total sclerotia weight	-0.6254	0.7799	-0.800	0.424			
Random Effect							
Inoc. day			0.061	0.805			
Inf%			0.834	0.437			
Inoc. day*Inf%			2.391	0.095			
Total sclerotia weight			0.643	0.424			
Avg. 2° Tiller Seed Weight					0.232	0.003	-1166.9
Parameter Estimates							
Intercept	0.0277	0.0015	18.710	0.017			
Inoc. day	-0.0002	0.0001	-1.480	0.144			
Inf%[.05-0]	0.0008	0.0007	1.230	0.221			
Inf%[.15-.05]	-0.0002	0.0015	-0.130	0.894			
(Inoc. day-2.98)*Inf%[.05-0]	0.0001	0.0002	0.770	0.440			
(Inoc. day-2.98)*Inf%[.15-.05]	-0.0007	0.0008	-0.770	0.442			
Total sclerotia weight	-0.0182	0.0196	-0.930	0.356			
Random Effect							
Inoc. day			2.178	0.144			
Inf%			0.785	0.458			
Inoc. day*Inf%			0.534	0.588			
Total sclerotia weight			0.857	0.356			
2° Tiller Germination					0.138	0.225	29.5
Parameter Estimates							
Intercept	0.9222	0.0657	14.030	0.007			
Inoc. day	-0.0019	0.0105	-0.180	0.859			
Inf%[.05-0]	0.1058	0.0481	2.200	0.030			
Inf%[.15-.05]	-0.1314	0.1084	-1.210	0.227			
(Inoc. day-2.98)*Inf%[.05-0]	-0.0085	0.0127	-0.670	0.503			
(Inoc. day-2.98)*Inf%[.15-.05]	0.0350	0.0604	0.580	0.563			
Total sclerotia weight	-2.3659	1.3860	-1.710	0.090			
Random Effect							
Inoc. day			0.032	0.859			
Inf%			3.400	0.036			
Inoc. day*Inf%			0.352	0.704			
Total sclerotia weight			2.914	0.090			

† t ratio for parameter estimates, F ratio for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

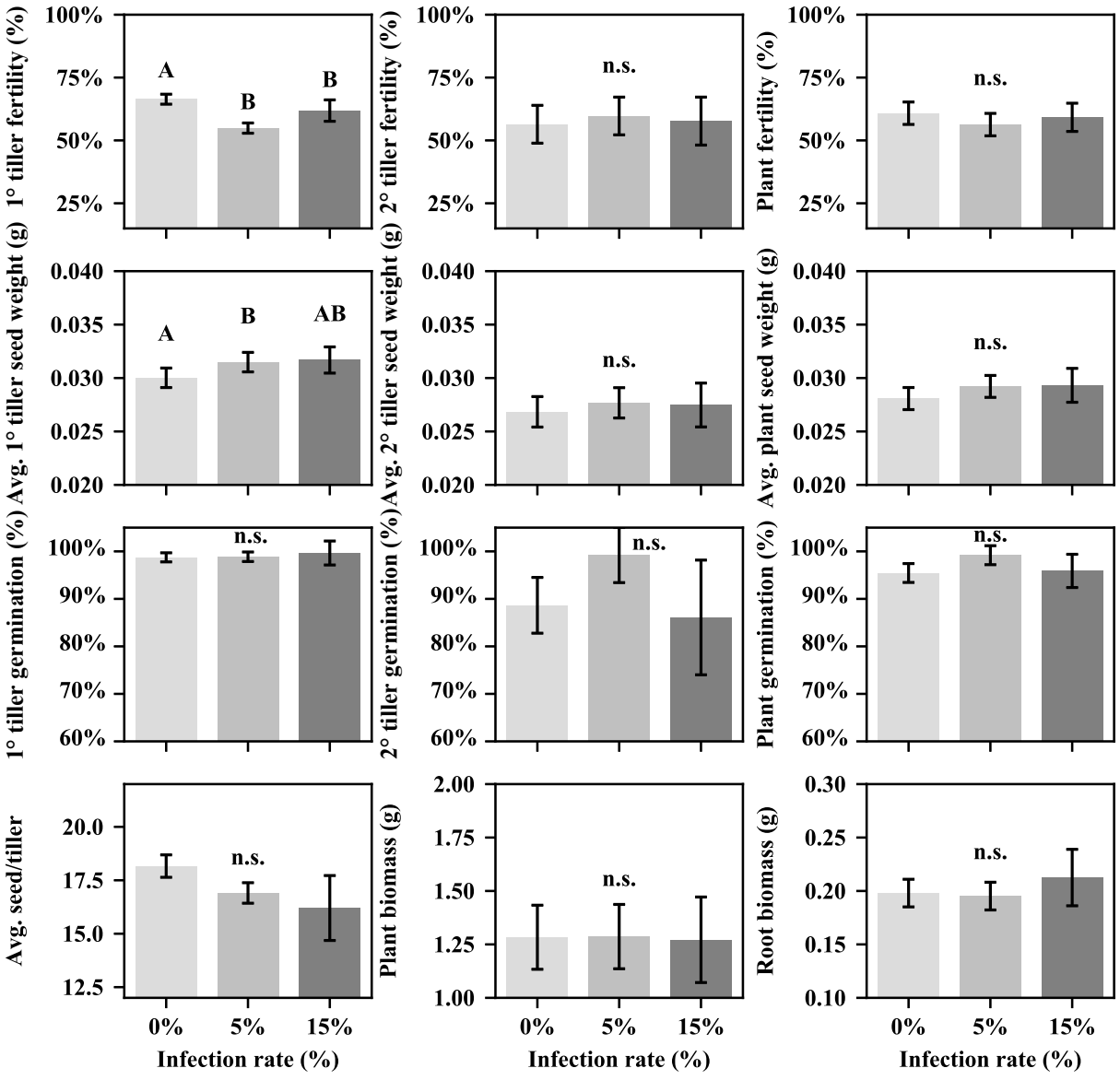


Figure 2.7: Least square means of plant responses for each infection rate in wheat using Dataset 2 and Model 2. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

DISCUSSION

Our results show that the effect of *Claviceps purpurea* infections can range from negative to positive, depending on infection rate, plant species, and plant tissue but generally result in neutral effects (Fig. 2.6, 2.7). Due to this variability, we cannot confidently identify *C. purpurea* as a friend or an enemy to its plant hosts but our results, from a controlled greenhouse setting, support the data from field collected samples of wild and cultivated hosts demonstrating the effects that infections of *Claviceps* can have on its host (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli *et al.* 2013). While further research is needed to better understand the precise relationship *C. purpurea* has with its host, our study provides new insight into their interaction and knowledge that will help guide future studies looking to examine this relationship further.

Our study provides evidence of the difficulties in *Claviceps* inoculation experiments due to the low success rate of obtaining desired infection rates (Fig. 2.1). These obstacles can be overcome if seed production is not a desired result. The utilization of sterile lines or self-incompatible grass species would facilitate this by allowing for greater success of infection and easier inoculation applications through sprays rather than syringe injections, which can be burdensome. It is common to use sterile lines for *C. purpurea* for inoculations, however, to our knowledge no study has looked at the effect of *C. purpurea* on plant biomass or roots. Efforts tend to focus on inflorescences and seed set where the perceived effects of fitness (gain or loss) would be most pronounced (Wäli *et al.* 2013), or when researchers are more interested in examining the infection process through microscopy or RNAseq (Hinsch *et al.* 2015, 2016; Oeser *et al.* 2017; Kind, *et al.* 2018a). Our study did not use sterile lines as we sought to investigate observations from field-based studies (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli

et al. 2013) in a controlled greenhouse setting and thus wanted to examine multiple aspects of plant growth, including seed production.

Furthermore, the effect of sclerotia weight on plant health can be debated and requires further research which we believe can be pursued through future studies of *Claviceps* inoculations. This continued research would greatly benefit our understandings of biotrophic-host interactions as direct measurements of the biotrophic organism could represent an indirect measurement of nutrients extracted from the host (Gray *et al.* 1990). Due to its restrictive growth, which colonizes plant ovaries and produces a single hypha that penetrates into the plant's vascular system near the tip of the rachilla with no hyphae growth beyond that point (Tudzynski and Scheffer 2004), *Claviceps* represents an excellent organism to study this interaction. In contrast, other biotrophic fungal organisms often produce extensive hyphal systems inter- and intra-cellularly throughout host tissue, making it difficult to get an accurate representation of the amount of fungal tissue present. While further research would elucidate the effect of sclerotia weight on plant health, our data suggests that this parameter often improved models and had significant effects on plant responses and thus should be incorporated into data analysis to provide a more accurate representation of the effects of *C. purpurea* infections. Therefore, our following discussion will focus on the results from Model 2.

Our results support previous reports of no significant effects of infection on seed germination and significant increases in seed weight on mildly infected tillers (< 10%) (Fig. 2.6, 2.7) (Raybould *et al*. 1998; Fisher *et al.* 2007). However, comparisons between our data and previous studies should only be done through examination of our 1° tiller data as previous studies only examined inflorescences that contained ergot. Since we did not infect our 2° tillers, our results allow us to observe the indirect effects that *Claviceps* infections can have on other

parts of the plant. For example, while infections on the 1° tiller resulted in trends of decreased fertility in barley (Fig. 2.6) and a significant decrease in wheat (Fig. 2.7), we observed trends of increased 2° tiller fertility in barley and wheat (Fig. 2.6, 2.7). When averaged over all tillers, plant fertility was seen to increase at higher infection rates compared to controls but was only significant at a 5% infection rate (Fig. 2.6). In general, positive effects of infection on plant responses were mostly significant at the 5% infection rate correlating to previous studies as higher infection rates have been shown to result in more adverse effects on plant health, through the decrease in successful seed production (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli *et al.* 2013). However, our data suggests that higher infection rates could have beneficial effects on other aspects of plant growth, particularly plant and root biomass (Fig. 2.6; Appendix 1 Fig. A1.4). Trends of increased root biomass were also observed in wheat (Fig. 2.7), however, this was not significant ($P > 0.05$) and we were not able to assess the effects of 25% infection rates. These positive effects of infection, on non-infected plant parts, such as increased 2° tiller fertility, plant biomass, and root biomass could be coping mechanisms of the host to tolerate infection through reallocation of resources. This reallocation could increase investments to flower heads and seed production on uninfected inflorescences (Raybould *et al*. 1998; Wäli *et al.* 2013), or to rhizomatous growth. While reallocation of resources might not fit the definition of a beneficial effect of infection, this result could be viewed as an indirect beneficial effect depending on the biology and the ecological lifestyle of the host.

Sporobolus anglicus is an invasive perennial marsh grass that originated in England but is now found globally (Ranwell 1967; Dethier and Hacker 2004). The seed production of *S. anglicus* is highly variably with seed set ranging from 2.3% - 77.1%, depending on the surrounding species compositions (Marks and Truscott 1985; Gray *et al.* 1990). Despite its

variable seed set, *S. anglicus* can produce large clonal populations through rapid asexual spread of rhizomes (Dethier and Hacker 2004). Raybould *et al*. (1991a, 1998) stated that populations of *S. anglicus*, in Poole Harbour, UK, were essentially a single genotype and thus most likely colonized the area through clonal expansion of rhizomes. These populations also have been reported to harbor epidemics of *Claviceps spartina* ((R. A. Duncan & J. F. White) Pazoutová *et* M. Kolařík) with > 70% disease incidence of inflorescences reported between 1985 and 1995. However, despite these epidemics *C. spartina* infections did not affect the population dynamics of *Spartina* in Poole Harbour (Raybould *et al*. 1998). Similar high infection levels (> 80%) of *C. spartina* have also been reported on populations of *Spartina* spp. in Mississippi (Eleuterius and Meyers 1974) and Europe (Gray *et al.* 1990). These results have postulated that severe epidemics of *C. spartina* will have little effect of plant fitness of *Spartina* spp., due to their low seed set and rapid proliferation of rhizomes (Jarosz and Davelos 1995). Although, no study has examined the effects of *Claviceps* infections on root production. As we observed that higher infection rates of *C. purpurea* result in greater root biomass (Fig. 2.6), it could be postulated that species with low natural seed set are open conduits for heavy ergot infections potentially resulting in increased clonal expansion through rhizomes.

Benefits of *Claviceps* infections have also been reported through an avoidance study which demonstrated that sheep avoided grazing grass containing *C. purpurea*, suggesting that infected grasses have a protective effect against mammalian grazing (Wäli *et al.* 2013). Interestingly, this study was conducted on *Festuca rubra* L. sl., a rhizomatous perennial grass species that often has low seed set and low self-fertility (Smith 1944; Harberd 1960; Ensign and Weiser 1975; Skálová *et al.* 1997; Pecháčková *et al.* 1999). While Wäli *et al.* (2013) did not examine below ground growth, they found that the frequency of *C. purpurea* infected

inflorescences were higher in pastures than surrounding ungrazed fields, suggesting a selective pressure in grazed pastures towards susceptible cultivars. As susceptible cultivars become infected they would avoid being grazed, allowing them to proliferate more. The success and proliferation of these susceptible cultivars could be the result of increases in *C. purpurea* infections followed by increased reallocation of resources to root/rhizome growth (Fig. 2.6). While grazing by sheep does not occur in all grass lands, grazing aversion by other mammals is likely as mammalian herbivores have been shown to detect and avoid endophyte-infected plants that possess similar alkaloid profiles to *C. purpurea* (Clay 1988; Parbery 1996; Panaccione *et al.* 2006; Uhlig *et al.* 2007; Krska and Crews 2008; Wäli *et al.* 2013). Therefore, the combination of grazing avoidance and increased rhizome growth, due to *C. purpurea* infections, could facilitate the spread and colonization of rhizomatous grass species. This may be further stimulated in grass species with low self-fertility and low natural seed set. Such an interaction could be an elaborate co-evolved symbiosis in which *C. purpurea* ensures its continued propagation by causing heavy infections on low seed set of rhizomatous grasses. These infections can further lower seed production (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli *et al.* 2013), which limits sexual reproduction and genetic variability. Due to this reduction of sexual reproduction plants might be forced to reallocate resources to asexual rhizomatous growth. This can result in reduced cultivar and species diversity in the surrounding grass community through facilitated expansion of highly susceptible genotypes. Repetition of this cycle could then produce large clonal populations, which both ensures the continued disease cycle *Claviceps* spp. and the colonization of these susceptible genotypes.

Similar relationships have been reported in the closely related (tribe Clavicipitaceae) grass endophyte *Atkinsonella hypoxylon* ((Peck) Diehl) which was shown to provide a

competitive advantage to its host *Danthonia* spp. against *Anthoxanthum* spp. in natural populations (Kelly and Clay 1987). Other closely related grass endophytes (i.e. *Epichloe*, *Balansia*) have been shown to provide mutualistic growth and reproductive benefits to their host and provide a mechanism for anti-herbivory (Diehl 1950; Bradshaw 1959; Clay 1984, 1986, 1987, 1988; Latch *et al.* 1985). While the production of toxic ergot alkaloids of these endophytes provides the host protection from herbivory (Clay 1988), the physiological and genetic basis for the observed growth and reproductive benefits have not been examined (Kuldau and Bacon 2008). Some research attributes enhanced growth effects to the production of synthetic growth hormones or phytohormones as some grass endophytes have been shown to produce auxins *in vitro* (Porter *et al.* 1985, De Battista *et al.* 1990). It has been recently discovered that *C. purpurea* produces cytokinins and auxins within the fungus and releases them into the host for establishment of the biotrophic interaction by manipulating the host's cytokinin levels (Hinsch *et al.* 2015, 2016; Oeser *et al.* 2017; Kind, *et al.* 2018a, 2018b). Interestingly, one of the most abundant cytokinins produced by *C. purpurea* are *cis*-zeatin (cZ)-derivatives (Hinsch *et al.* 2015), which are thought to be related to the switch from vegetative to reproductive growth after pollination in wheat and other cereals crops (Galuska *et al.* 2008).

Further research is needed to clarify the effect of *C. purpurea* infections on root and rhizome growth. Since our study utilized commercial cereal crops, for an overview of plant responses from seed to root production, we cannot provide evidence for *C. purpurea* infection on rhizome growth. However, our observations of significantly increased root biomass in infected barley (Fig. 2.6) and trends of increased root biomass in wheat (Fig. 2.7) suggest that rhizomes might be similarly affected. Future work looking into this interaction should focus on highly rhizomatous species that have low seed set. Using grass species that also have low self-fertility

or are male sterile would facilitate this endeavor as the success of inoculations would be greatly increased, which will provide greater statistical power with fewer complications. We believe that *Bromus inermis* could represent an excellent candidate for future studies due to its self-incompatibility and rhizomatous nature. In addition, *B. inermis* has been recently identified as a highly susceptible grass species that is responsible for large inoculum reservoirs of *C. purpurea* surrounding cereal crop fields in the San Luis Valley of Colorado (Chapter 1). *Bromus inermis* is also an introduced species that has become invasive and has outcompeted many wild grass species as it spread throughout North America (Sprague 1950; Campbell 1957; Romo and Grilz 1990; Nagel *et al.* 1994; Murphy and Grant 2005; Otfinowski *et al.* 2007; Dillemuth *et al.* 2009; Fink and Wilson 2011). It would be interesting to further investigate if *C. purpurea* facilitated the colonization and successful invasion of *B. inermis*.

Chapter 3: Whole genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus *Claviceps* are the result of varying mechanisms driving genome evolution and host range expansion[1]

INTRODUCTION

Fungi, particularly phytopathogenic species, are increasingly being utilized to gain insight into the evolution of eukaryotic organisms, due to their adaptive nature and unique genome structures (Gladieux *et al.* 2014; Dong *et al.* 2015). Adaptation and diversification of fungal species can be mediated by changes in genome architecture and plasticity such as genome size, transposable element (TE) content, localization of TEs to specific genes, genome compartmentalization, gene duplication rates, recombination rates, and presence/absence polymorphism of virulence factors (Dong *et al.* 2015; Möller and Stukenbrock 2017). The presence or absence of repeat-induced point (RIP) mutation, a fungi specific mechanism, is also an important mechanism for fungal genome evolution, as RIP works on a genome-wide scale to silence transposable elements and duplicated genes, which can also "leak" onto neighboring genes (Galagan *et al.* 2003, 2004; Raffaele and Kamoun 2012; Möller and Stukenbrock 2017; Urguhart *et al.* 2018). It is becoming increasingly evident that variations in these factors can be used to classify genomes as a one-speed (one-compartment), such as the powdery mildew fungi *Blumeria graminis* f.sp. *hordei* and f.sp *tritici*, two-speed (two-compartment), such as the late blight pathogen *Phytophthora infestans*, or multi-speed (multi-compartment) such as the multi-host pathogen *Fusarium oxysporum* (Dong *et al.* 2015; Frantzeskakis *et al.* 2019). These

[1] This chapter contain preprint work from: Wyka SA, Mondo SJ, Liu M, Dettman J, Nalam V, Broders KD. 2020. Whole genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus Claviceps are the result of varying mechanisms driving genome evolution and host range expansion. bioRxiv doi: 10.1101/2020.04.13.039230

different “speeds” are characterized by their potential adaptability such that one-speed genomes are often considered less-adaptable, while two-speed and multi-speed genomes are often considered more-adaptable (Dong *et al.* 2015; Möller and Stukenbrock 2017; Frantzeskakis *et al.* 2019).

The ergot fungi of the genus *Claviceps* (Ascomycota, Hypocreales) are biotrophic species that share a specialized ovarian-specific non-systemic parasitic lifestyle with their grass hosts (Píchová *et al.* 2018). Infections are fully restricted to individual unpollinated ovaries (Tudzynski and Scheffer 2004), and the fungus actively manages to maintain host cell viability to obtain nutrients from living tissue through a complex cross-talk of genes related to pathogenesis, such as secreted effectors, secondary metabolites, or cytokinin production (Hinsch *et al.* 2015, 2016; Oeser *et al.* 2017; Kind *et al.* 2018a, 2018b). Species of *Claviceps* are most notably known for their production of toxic alkaloids and secondary metabolites but are also known for their expansive host range and negative impact on global cereal crop production and livestock farming. These negative effects on human and livestock health are the primary reason *Claviceps* species are referred to as plant pathogens. However, under the light of co-evolution with their grass hosts some *Claviceps* species are considered conditional defensive mutualists with their hosts as they prevent herbivory and can improve host fitness (Raybould *et al.* 1998; Fisher *et al.* 2007; Wäli *et al.* 2013).

The genus *Claviceps* contains 59 species divided into four sections; sects. *Claviceps*, *Pusillae*, *Citrinae*, and *Paspalorum* (Píchová *et al.* 2018). It was postulated that sects. *Citrinae* and *Paspalorum* originated in South America, while sect. *Pusillae* experienced speciation throughout the Eocene, Oligocene, and Miocene as these species encountered newly emergent PACMAD warm-season grasses (subfamilies Panicoideae, Aristidoideae, Chloridoideae,

Micrairoideae, Arundinoideae, Danthonioideae) when an ancestral strain was transferred from South America to Africa (Píchová *et al.* 2018). In contrast, the crown node of sect. *Claviceps* is estimated at 20.4 Mya and was followed by a radiation of the section corresponding to a host jump from ancestral sedges (Cyperaceae) to the BOP clade (cool-season grasses; subfamilies Bambusoideae, Oryzoideae (syn: Ehrhartoideae) (Soreng *et al.* 2015), Pooideae) in North America (Bouchenak-Khelladi *et al.* 2010; Píchová *et al.* 2018). Section *Claviceps* has the largest host range with *C. purpurea sensu stricto* (s.s.) having been reported on up to 400 different species in clade BOP (Alderman *et al.* 2004, Píchová *et al.* 2018) across six tribes, and retains the ability to infect sedges (Cyperaceae) (Jungehülsing and Tudzynski 1997). In contrast, sect. *Pusillae* is specialized to the tribes Paniceae and Andropogoneae, and sects. *Citrinae* and *Paspalorum* only infect members of tribe Paspaleae and tribe Cynodonteae, respectively (Píchová *et al.* 2018). The shared specialized infection life cycle of the *Claviceps* genus, the drastic differences in host range potential of different species, and geographic distribution represent a unique system to study the evolution and host adaptation of eukaryotic organisms.

Despite their ecological and agriculture importance, little is known about the evolution and genomic architecture of these important fungal species in comparison to other cereal pathogens such as species in the genera *Puccinia* (Cantu *et al.* 2013; Kiran *et al.* 2016, 2017), *Zymoseptoria* (Grandaubert *et al.* 2015, 2019; Estep *et al.* 2015; Poppe *et al.* 2015; Testa *et al.* 2015b; Wu *et al.* 2017; Stukenbrock and Dutheil 2018a), or *Fusarium* (Kvas *et al.* 2009; Ma *et al.* 2010; Rep and Kistler 2010; Watanabe *et al.* 2011; Sperschneider *et al.* 2015). Unfortunately, the lack of genome data for the *Claviceps* genus has hampered our ability to complete comparative analyses to identify factors that are influencing the adaptation of *Claviceps* species across the four sections in the genus, and the mechanisms by which species of sect. *Claviceps*

have adapted to such a broad host range, in comparison to the other three sections. Here we present the sequences and annotations of 50 *Claviceps* genomes, representing 19 species, for a comprehensive comparison of the genus to understand evolution within the genus *Claviceps* by characterizing the genomic plasticity and architecture in relation to adaptive host potential. Our analysis reveals the trajectory from specialized one-speed genomes (sects. *Citrinae* and *Paspalorum*) towards adaptive two-speed genomes (sects. *Pusillae* and *Claviceps*) through co-localization of transposable elements around predicted effectors and a putative loss of RIP resulting in tandem gene duplication coinciding with increased host range potential.

MATERIALS AND METHODS

Sample acquisition

Field collected samples (Clav) were surfaced sterilized, allowed to grow as mycelia, and individual conidia transferred to make single spore cultures. Thirteen cultures were provided by Dr. Miroslav Kolařík from the Culture Collection of Clavicipitaceae (CCC) at Institute of Microbiology, Academy of Sciences of the Czech Republic. Raw Illumina reads for samples (LM28, LM582, LM78, LM81, LM458, LM218, LM454, LM576, and LM583) were downloaded from NCBI's SRA database. Raw Illumina reads from an additional 21 LM samples were generated by Dr. Liu's lab (AAFC), sequencing protocol of these 21 samples followed Wingfield *et al.* (2018). Summarized information can be found in Appendix 2 Table A2.1.

Preparation of genomic DNA

Cultures grown on cellophane PDA plates were used for genomic DNA extraction from lyophilized mycelium following a modified CTAB method (Doyle and Doyle 1987; Wingfield *et al.* 2018) without using the RNase Cocktail™ Enzyme Mix, only RNASE A was used. DNA contamination was checked by running samples on a 1% agarose gel and a NanoDrop One^{c}

(Thermo Fishcer Scientific). Twenty samples (7 Clav and 13 CCC) were sent to BGI-Hong Kong HGS Lab for 150-bp paired-end Illumina sequencing on a HiSeq^{TM} 4000.

Genome assembly

Preliminary data showed that raw reads of LM458 were contaminated with bacterial DNA but showed strong species similarity to Clav32 and Clav50. To filter out the bacterial DNA sequences, reads of LM458 were mapped against the assembled Clav32 and Clav50 genomes using BBSplit v38.41 (Bushnell 2014). All forward and reverse reads mapped to each of the genomes were concatenated and made non-redundant, respectively. Both sets were then interleaved to remove duplicates and used for further analysis. Reads for all 50 samples were checked for quality with FastQC v0.11.5 (Andrews 2010) and trimmed with Trimmomatic v0.36 (Bolger *et al.* 2014) using the commands (SLIDINGWINDOW:4:20 MINLEN:36 HEADCROP:10) to remove poor quality data, only paired end reads were used. To better standardize the comparative analysis all 50 sample were subject to *de novo* genome assembly with Shovill v0.9.0 (https://github.com/tseemann/shovill) using SPAdes v3.11.1 (Nurk *et al.* 2013) with a minimum contig length of 1000 bp.

The reference genomes of *C. purpurea* strain 20.1 (SAMEA2272775), *C. fusiformis* PRL 1980 (SAMN02981339), and *C. paspali* (F.Stevens & J.G. Hall) RRC 1481 (SAMN02981342) were downloaded from NCBI. Proteins for *C. fusiformis* and *C. paspali* were not available on NCBI so they were extracted from GFF3 files provided by Dr. Chris Schardl and Dr. Neil Moore, University of Kentucky, corresponding to the 2013 annotations (Schardl *et al.* 2013) available at http://www.endophyte.uky.edu. Reference genomes were standardized for comparative analysis with our 50 annotated genomes, by implementing a protein length cutoff of

50 aa and removal of alternatively spliced proteins in *C. fusiformis* and *C. paspali*, only the longest spliced protein for each locus remained.

Transposable elements

Transposable elements (TE) fragments were identified following procedures for establishment of *de novo* comprehensive repeat libraries set forth in Berriman *et al.* (2018) through a combined use of RepeatModeler v1.0.8 (Smit and Hubley 2015), TransposonPSI (Hass 2010), LTR_finder v1.07 (Xu and Wang 2007), LTR_harvest v1.5.10 (Ellinghaus *et al.* 2008), LTR_digest v1.5.10 (Steinbiss *et al.* 2009), Usearch v11.0.667 (Edgar 2010), and RepeatClassifier v1.0.8 (Smit and Hubley 2015) with the addition of all curated fungal TEs from RepBase (Bao *et al.* 2015). RepeatMasker v4.0.7 (Smit *et al.* 2015) was then used to soft mask the genomes and identify TE regions. TE content was represented as the proportion of the genome masked by TE regions determined by RepeatMasker, excluding simple and low complexity repeats. These steps were automated through construction of a custom script, TransposableELMT (https://github.com/PlantDr430/TransposableELMT).

Divergence landscapes for TEs in all 53 *Claviceps* genomes were generated using a custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/TE_divergence_landscape.py) and the RepeatMasker output results. The RepeatMasker results were also used with the respective GFF3 file from each genome to calculate the average distance (kb) of each gene to the closest TE fragment on the 5' and 3' flanking side. Values were calculated for predicted effectors, non-effector secreted genes, non-secreted metabolite genes, and all other genes using a custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/TE_closeness.py).

Genome annotation

AUGUSTUS v3.2.2 (Mario *et al.* 2008) was used to create pre-trained parameters files using the reference *C. purpurea* strain 20.1, available EST data from NCBI, and wild-type RNAseq data (SRR4428945) created in Oeser *et al.* (2017). RNA-seq data was subject to quality check and trimming as above. All three datasets were also used to train parameter files for the *ab initio* gene model prediction software's GeneID v1.4.4 (Blanco *et al.* 2007) and CodingQuarry v2.0 (Testa *et al.* 2015a). GeneID training followed protocols available at http://genome.crg.es/software/geneid/training.html. For CodingQuarry training, RNA transcripts were created *de novo* using Trinity v2.8.4 (Grabherr *et al.* 2011) on default settings and EST coordinates were found by mapping the EST data to the reference genome using Minimap2 v2.1 (Li 2018).

Gene models for the 50 genomes were then predicted with GeneID and CodingQuarry using the trained *C. purpruea* parameter files. CodingQuarry prediction was also supplemented with transcript evidence by mapping the available EST and RNA-seq *C. purpurea* data to each genome using Minimap2. BUSCO v3 (Waterhouse *et al.* 2018) was run on all 50 genomes using the AUGUSTUS *C. purpurea* pre-trained parameter files as the reference organism and the Sordariomyceta database. The resulting predicted proteins for each sample were used as training models for *ab initio* gene prediction using SNAP (Korf 2004) and GlimmerHMM v3.0.1 (Majoros *et al.* 2004). Lastly, GeMoMa v1.5.3 (Keilwagen *et al.* 2016) was used for *ab initio* gene prediction using the soft-masked genomes and the *C. purpruea* 20.1 reference files.

Funannotate v1.6.0 (Palmer and Stajich 2019) was then used as the primary software for genome annotation. Funannotate additionally uses AUGUSTUS and GeneMark-ES (Ter-Hovhannisyan *et al.* 2008) for *ab initio* gene model prediction, Exonerate for transcript and

protein evidence alignment, and EVidenceModeler (Hass *et al.* 2008) for a final weighted consensus. All *C. purpurea* EST and RNAseq data were used as transcript evidence and the Uniport Swiss-Prot database and proteins from several closely related species (*C. purpurea* strain 20.1, *C. fusiformis* PRL1980, *C. paspali* RRC1481, *Fusarium oxysporum f.sp. lycopersici* 4287 ((Sacc.) W.C. Snyder & H.N. Hansen), *Pochonia chlamydosporia* 170 ((Goddard) Zare & W. Gams, Nova Hedwigia 72: 334 (2001)), *Ustilago maydis* 521, and *Epichloe festucae* F1 ((Latch, M.J. Chr. & Samuels) C.W. Bacon & Schardl (2014))) were used as protein evidence. The AUGUSTUS pre-trained *C. purpurea* files were used as BUSCO seed species along with the Sordariomyceta database and all five *ab initio* predictions were passed through the --other_gff flag with weights of 1. The following flags were also used in Funannotate "predict": --repeats2evm, --optimize_augustus, --soft_mask 1000, --min_protlen 50. BUSCO was used to evaluate annotation completeness using the Dikarya and Sordariomyceta databases (odb9) with --prot on default settings.

Functional annotation

Functional analysis was performed using Funannotate "annotate". The following analyses were also performed on the three reference *Claviceps* genomes. Secondary metabolite clusters were predicted using antiSMASH v5 (Blin *et al.* 2019) with all features turned on. Functional domain annotations were conducted using eggNOG-mapper v5 (Huerta-Cepas *et al.* 2016, 2019) on default settings and InterProScan v5 (Jones *et al.* 2014) with the --goterms flag. Phobius v1.01 (Käll 2007) was used to assist in prediction of secreted proteins. In addition to these analyses Funannotate also performed domain annotations through an HMMer v3.2.1 (Wheeler and Eddy 2013) search against the Pfam-A v32.0 database and dbCAN v8.0 CAZYmes database,

a BLASTp search against the MEROPS v12.0 protease database, and secreted protein predictions with SignalP v4.1 (Nielsen 2017).

For downstream analysis, proteins were classified as secreted proteins if they had signal peptides detected by both Phobius and SignalP and did not possess a transmembrane domain as predicted by Phobius and an additional analysis of TMHMM v2.0 (Krogh *et al.* 2001). Effector proteins were identified by using EffectorP v2.0 (Sperschneider *et al.* 2018), with default settings, on the set of secreted proteins for each genome. Transmembrane proteins were identified if both Phobius and TMHMM detected transmembrane domains. Secondary metabolite proteins were identified if they resided within metabolite clusters predicted by antiSMASH. Proteins were classified as having conserved protein domains if they contained any Pfam or IPR domains.

Orthogroup identification and classification

OrthoFinder v2.3.3 (Emms and Kelly 2019) was run on default settings using Diamond v0.9.25.126 (Buchfunk *et al.* 2015) to infer groups of orthologous gene clusters (orthogroups) based on protein homology and MCL clustering. To more accurately place closely related genes into clusters an additional 78 fungal genomes (Appendix 2 Table A2.3) with emphasis on plant associated fungi of the order Hypocreales were added. To standardize, all 78 additional genomes were subject to a protein length cutoff of 50 amino acids and genomes downloaded from http://www.endophyte.uky.edu had alternatively spliced proteins removed. For downstream analysis, orthogroups pertaining to the 53 *Claviceps* genomes were classified as secreted, predicted effectors, transmembrane, metabolite, and conserved domain orthogroups if ≥ 50% of the *Claviceps* strains present in a given cluster had at least one protein classified as such.

Phylogenomics and genome fluidity

Phylogenetic relationship of all 53 *Claviceps* genomes, with *Fusarium graminearum* (Schwabe, Flora Anhaltina), *F. verticillioides* ((Sacc.) Nirenberg, Mitteilungen der Biologischen Bundesanstalt für Land- und Forstwirtschaft), *Epichloe festucae* and *E. typhina* ((Pers.: Fr.) Tulasne), as outgroups, was derived from 2,002 single-copy orthologs obtained from our OrthoFinder defined gene clusters (described above). This resulted in a dataset of 114,114 amino acids sequences which were concatenated to create a super-matrix and aligned using MAFFT v7.429 (Katoh and Standely 2013) on default settings. Uninformative sites were removed using Gblocks v0.91 (Castresana 2000) on default settings. Due to the large scale of the alignment maximum likelihood reconstruction was performed using FastTree v2.1.11 (Price *et al.* 2010) using the WAG model of amino acid substitution with the -gamma, -spr 4, -mlacc 2, -slownni, and -slow flag with 1000 bootstraps. MEGA X (Sudhir *et al.* 2018) was used for neighbor-joining reconstruction using the JTT model of amino acid substitution with gamma distribution and maximum parsimony reconstruction using the tree bisection reconstruction algorithm with 100 repeated searches. Nodal support for both NJ and MP reconstructions were assessed with 1000 bootstraps. In addition, an alignment and ML reconstruction was performed on each of the 2,002 protein sequences following the procedure as above (MAFFT, Gblocks, FastTree). A density consensus phylogeny was created from all gene trees using the program DensiTree v2.2.5 (Bouckaert and Heled 2014). PhyBin v0.3-1 (Newton and Newton 2013) was used to cluster trees from three datasets (1: *Claviceps* genus without outgroups, 2: sect. *Pusillae* species, 3: sect. *Claviceps* species) together to identify frequencies of concordant topologies using the --complete flag with --editdist=2. To reduce noise, from abundant incomplete lineage sorting in sect. *Claviceps*, we implemented a --minbranchlen=0.015 for our *Claviceps* genus dataset.

Following methodologies established in Kislyuk *et al.* (2011) genomic fluidity, which estimates the dissimilarity between genomes by using ratios of the number of unique gene clusters to the total number of gene clusters in pairs of genomes averaged over randomly chosen genome pairs from within a group on *N* genomes, was used to assess gene cluster dissimilarity within the *Claviceps* genus. For a more detailed description refer to Kislyuk *et al.* (2011). Datasets containing gene clusters from representative members of sect. *Pusillae*, sect. *Claviceps*, *Clavieps* genus, and all *C. purpurea* strains were extracted from our OrthoFinder defined gene clusters. Additional species- and genus-wide gene cluster datasets from the additional 78 fungal genomes were extracted for comparative purposes. All section- and genus-wide datasets contained one representative isolate from each species to reduce phylogenetic bias. Each extracted dataset was used to calculate the genomic fluidity using a custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/pangenome_fluidity.py). The result files for each dataset were then used for figure creation and two-sample two-sided z-test statistics (Kislyuk *et al.* 2011) using a custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/combine_fluidity.py).

Gene compartmentalization

A custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/genome_speed_hexbins.py) was used to calculate local gene density measured as 5' and 3' flanking distances between neighboring genes (intergenic regions). To statistically determine whether specific gene types had longer intergenic flanking regions than all other genes within the genome we randomly sampled 100 from each group of genes (specific gene vs. other genes) 1,000 times for both the 5' and 3' flanking distances. Mann-Whitney U test was used to test for significance on all 2,000 subsets

corrected with Benjamini-Hochberg. Corrected p-values were averaged per flanking side and then together to get a final p-value. Genes that appeared on a contig alone were excluded from analysis. For graphical representation, genes that were located at the start of each contig (5' end) were plotted along the x-axis, while genes located at the end of each contig (3' end) were plotted along the y-axis.

RIP analyses

For all 53 genomes a self BLASTp v2.9.0+ search was conducted to identify best hit orthologs within each genome with a cutoff e-value of 10^{-5} and removal of self-hits. This process was automated, using a custom script (https://github.com/PlantDr430/CSU_scripts/blob/master/RIP_blast_analysis.py). We further examined if gene pairs with a pairwise identity of ≥ 80% were located next to each other and/or separated by five or fewer genes. Fifty-six important *Claviceps* genes (Appendix 2 Table A2.5) including the *rid-1* homolog (Freitag *et al.* 2002) were used in a BLASTp analysis to identify the number of genes present that passed an e-value cutoff of 10^{-5}, 50% coverage, and 35% identity. Genes that appeared as best hits for multiple query genes were only recorded once for their overall best match. In addition, the web-based tool The RIPper (van Wyk *et al.* 2019) was used on default settings (1 kb windows in 500 bp increments) to scan whole-genomes for presence of RIP and large RIP affected regions (LRARs).

Statistical analysis and software

Statistics and figures were generated using Python3 modules SciPy v1.3.1, statsmodel v0.11.0, and Matplotlib v3.1.1. Heatmaps were generated using ComplexHeatmap v2.2.0 in R (Gu 2016).

RESULTS

Genome Assembly and Annotation

To provide a comprehensive view of variability across *Claviceps*, we sequenced and annotated 50 genomes (19 *Claviceps* spp.), including *C. citrina* the single species of sect. *Citrinae*, six species belonging to sect. *Pusillae*, and 44 genomes (12 species) belonging to sect. *Claviceps*, of which 23 genomes belong to *C. purpurea* s.s. (Table 3.1; Appendix 2 Table A2.1). The assemblies and annotations were of comparable quality to the reference strains (Table 3.1). A more detailed representation of the assembly and annotation statistics can be seen in Table 3.1 and Appendix: Fig. A2.1, Table A2.2.

Overall, species of sect. *Claviceps* had better assemblies and annotations than species of other sections regarding contig numbers, N50s, and BUSCO completeness scores (Table 3.1). Nearly all species of sect. *Claviceps* showed higher BUSCO scores than the references, while species of sects. *Pusillae* and *Citrinae* generally showed lower scores, likely due to their higher TE content (average 34.9% ± 11.0%; Table 3.1). Exceptions to the low BUSCO scores were *C. digitariae* (Hansf.) and *C. maximensis* (T. Theis) (sect. *Pusillae*), which had lower TE content 20.0% and 19.8%, respectively, than the rest of the species in sect. *Pusillae* (Table 3.1).

Phylogenomics and genome fluidity

Orthologous gene clusters (orthogroups), which contain orthologs and paralogs,were inferred from protein homology and MCL clustering using OrthoFinder. Across the 53 *Claviceps* isolates and outgroups species *Fusarium graminearum*, *F. verticillioides*, *Epichloe festucae*, and *E. typhina*, we identified 2,002 single copy orthologs. We utilized a super-matrix approach to infer a maximum likelihood (ML) species tree, based on these protein sequences. Results showed statistical support for four sections of *Claviceps* with a near concordant topology to the Bayesian

Table 3.1: Assembly and annotations statistics for the three reference *Claviceps* genomes and the 50 *Claviceps* genomes used in this study.

			Host of Origin							TE§		BUSCO Completeness	
Organism	Strain	Section	Family / Tribe	Genus / species	Read Coverage	Genome size (Mb)	Contig (#)	N50	Genomic GC (%)	content (%)	Gene count	Dikarya	Sordario-myceta
References:													
C. purpruea	20.1	Claviceps	Triticeae	*Secale cereale*	--	32.1	1,442†	46,498†	51.6%	10.9%	8,703	95.30%	94.70%
C. fusiformis	PRL1980	Pusillae	Paniceae	*Pennisetum typhoideum*	--	52.3	6,930	19,980	37.3%	47.5%	9,304	96.70%	94.90%
C. paspali	RRC1481	Paspalorum	Paspaleae	*Paspalum* sp.	--	28.9	2,304	26,898	47.7%	17.5%	8,400	94.30%	93.30%
This study:													
C. purpruea	Clav04	Claviceps	Bromeae	*Bromus inermis*	46x	31.8	3,288	21,051	51.7%	10.1%	8,824	95.50%	94.10%
C. purpruea	Clav26	Claviceps	Triticeae	*Hordeum vulgare*	59x	30.8	1,361	49,697	51.7%	9.1%	8,737	97.70%	96.50%
C. purpruea	Clav46	Claviceps	Triticeae	*Secale cereale*	58x	30.8	1,409	49,302	51.7%	9.7%	8,597	98.00%	96.60%
C. purpruea	Clav55	Claviceps	Poeae	*Lolium perenne*	59x	30.7	1,525	44,299	51.8%	9.8%	8,480	97.10%	95.90%
C. purpruea	LM4	Claviceps	Triticeae	*Tricosecale*	64x	30.6	1,296	47,441	51.8%	10.0%	8,470	97.00%	95.80%
C. purpruea	LM5	Claviceps	Triticeae	*Hordeum vulgare*	67x	30.5	1,258	51,505	51.8%	9.0%	8,508	96.90%	95.50%
C. purpruea	LM14	Claviceps	Triticeae	*Hordeum vulgare*	49x	30.6	1,297	49,955	51.8%	10.0%	8,422	97.40%	95.60%
C. purpruea	LM28	Claviceps	Triticeae	*Triticum aestivum*	49x	30.6	1,343	51,635	51.7%	9.6%	8,713	97.30%	96.10%
C. purpruea	LM30	Claviceps	Triticeae	*Secale cereale*	64x	30.6	1,224	51,374	51.8%	9.4%	8,526	97.00%	95.50%
C. purpruea	LM33	Claviceps	Triticeae	*Secale cereale*	45x	30.5	1,398	44,564	51.8%	9.2%	8,557	96.30%	95.50%
C. purpruea	LM39	Claviceps	Triticeae	*Triticum turgidum subsp. durum*	81x	30.5	1,282	48,443	51.8%	10.1%	8,591	97.10%	96.10%
C. purpruea	LM46	Claviceps	Triticeae	*Triticum turgidum subsp. durum*	79x	30.6	1,291	50,932	51.8%	9.6%	8,455	97.00%	95.80%
C. purpruea	LM60	Claviceps	Poeae	*Avena sativa*	81x	30.6	1,259	47,464	51.7%	9.3%	8,498	97.00%	95.80%
C. purpruea	LM71	Claviceps	Poeae	*Alopercurus myosuroides*	168x	30.5	1,400	45,114	51.8%	9.6%	8,472	97.10%	95.60%
C. purpruea	LM207	Claviceps	Triticeae	*Elymus repens*	53x	30.5	1,352	45,388	51.8%	9.2%	8,475	97.00%	95.70%
C. purpruea	LM223	Claviceps	Bromeae	*Bromus riparius*	74x	30.8	1,297	46,577	51.7%	10.5%	8,438	97.00%	95.70%
C. purpruea	LM232	Claviceps	Poeae	*Phalaris canariensis*	53x	30.7	1,348	49,571	51.7%	9.4%	8,512	96.60%	95.70%
C. purpruea	LM233	Claviceps	Poeae	*Phalaris canariensis*	49x	30.6	1,331	50,327	51.8%	9.9%	8,717	96.70%	95.90%
C. purpruea	LM461	Claviceps	Triticeae	*Elymus repens*	37x	30.5	1,440	44,216	51.8%	8.4%	8,656	96.60%	95.20%
C. purpruea	LM469	Claviceps	Triticeae	*Triticum aestivum*	75x	30.5	1,257	48,403	51.8%	10.0%	8,394	97.30%	96.00%
C. purpruea	LM470	Claviceps	Triticeae	*Elymus repens*	26x	30.5	1,797	32,579	51.8%	9.0%	8,591	96.50%	95.30%
C. purpruea	LM474	Claviceps	Triticeae	*Hordeum vulgare*	64x	30.6	1,354	47,245	51.8%	9.4%	8,500	96.80%	95.70%
C. purpruea	LM582	Claviceps	Triticeae	*Secale cereale*	89x	30.7	1,600	39,003	51.8%	9.6%	8,518	97.20%	95.40%
C. aff. purpruea	Clav52	Claviceps	Poeae	*Poa pratensis*	60x	29.6	1,334	48,893	51.8%	8.2%	8,316	96.80%	96.20%
C. quebecensis‡	Clav32	Claviceps	Triticeae	*Hordeum vulgare*	64x	28.7	1,068	58,118	51.6%	4.5%	8,232	98.00%	96.60%
C. quebecensis‡	Clav50	Claviceps	Triticeae	*Elymus sp.*	59x	28.8	1,075	66,795	51.6%	6.9%	8,046	97.50%	96.30%
C. quebecensis‡	LM458	Claviceps	Poeae	*Ammophila* (plant)	78x	28.4	1,166	45,693	51.6%	6.1%	8,055	97.10%	95.80%
C. occidentalis‡	LM77	Claviceps	Poeae	*Phleum pratense*	58x	28.7	1,728	29,222	51.4%	6.0%	8,162	96.10%	94.70%
C. occidentalis‡	LM78	Claviceps	Bromeae	*Bromus inermis*	64x	28.8	1,689	29,608	51.4%	6.0%	8,231	95.80%	94.70%
C. occidentalis‡	LM84	Claviceps	Bromeae	*Bromus inermis*	164x	28.9	1,404	36,685	51.4%	6.0%	8,221	97.00%	95.40%
C. ripicola‡	LM218	Claviceps	Poeae	*Phalaris arundinacea*	146x	31.1	1,072	60,464	51.4%	10.3%	8,327	96.70%	95.70%
C. ripicola‡	LM219	Claviceps	Poeae	*Phalaris arundinacea*	55x	30.8	1,239	55,312	51.4%	9.5%	8,381	96.80%	95.80%
C. ripicola‡	LM220	Claviceps	Poeae	*Phalaris arundinacea*	91x	30.9	1,223	54,100	51.4%	9.3%	8,449	97.10%	95.90%
C. ripicola‡	LM454	Claviceps	Poeae	*Ammophila breviligulata*	156x	31.2	1,508	40,844	51.4%	8.4%	8,562	97.10%	96.10%
C. spartinae	CCC535	Claviceps	Zoysieae	*Sporobolus anglicus*	60x	29.3	1,456	42,688	51.4%	7.1%	8,433	97.50%	95.90%
C. arundinis	LM583	Claviceps	Molinieae	*Phragmites australis*	69x	30.6	996	70,672	51.4%	9.8%	8,235	96.80%	95.70%
C. arundinis	CCC1102	Claviceps	Molinieae	*Phragmites australis*	61x	30.3	896	91,905	51.4%	8.3%	8,486	97.70%	96.50%

† The reference strain *C. purpurea* 20.1 was additionally assembled into 191 scaffolds with a scaffold N50 of 433,221.
‡ Newly identified species Liu *et al. Accepted*)
§ Transposable element (TE) content represented as percent of the genome masked by TEs

Table 3.1: Continued

			Host of Origin							TE[§]		BUSCO Completenes	
Organism	**Strain**	**Section**	**Family / Tribe**	**Genus / species**	**Read Coverage**	**Genome size (Mb)**	**Contig (#)**	**N50**	**Genomic GC (%)**	**content (%)**	**s Gene count**	**Dikarya**	**Sordario-myceta**
C. humidiphila	LM576	Claviceps	Poeae	*Dactylis sp.*	77x	31.2	1,236	55,717	51.5%	9.9%	8,440	97.00%	95.90%
C. perihumidiphila‡	LM81	Claviceps	Triticeae	*Elymus albicans*	140x	31.2	1,003	67,487	51.5%	11.0%	8,291	97.10%	95.90%
C. cyperi	CCC1219	Claviceps	Cyperaceae (family)	*Cyperus esculentus*	56x	26.6	1,921	27,113	51.7%	8.9%	7,673	97.70%	95.40%
C. capensis	CCC1504	Claviceps	Ehrharteae	*Ehrharta villosa*	66x	27.7	1,136	59,777	51.7%	6.2%	8,037	97.60%	95.70%
C. pazoutovae	CCC1485	Claviceps	Stipeae	*Stipa dregeana*	61x	27.6	1,304	42,785	51.7%	6.8%	7,941	97.50%	96.00%
C. monticola	CCC1483	Claviceps	Brachypodieae	*Brachypodium sp.*	58x	27.8	1,144	56,619	51.6%	7.0%	7,977	98.10%	96.50%
C. pusilla	CCC602	Pusillae	Andropogoneae	*Bothriochloa insculpta*	52x	45.9	5,068	15,010	40.4%	42.1%	8,735	90.90%	88.30%
C. lovelessii	CCC647	Pusillae	Eragostidinae	*Eragrostis sp.*	53x	41.1	5,300	12,480	42.1%	33.9%	8,862	91.60%	88.20%
C. digitariae	CCC659	Pusillae	Paniceae	*Digitaria eriantha*	57x	33.4	1,773	32,638	44.8%	20.0%	8,285	95.90%	94.70%
C. maximensis	CCC398	Pusillae	Paniceae	*Megathyrsus maximus*	58x	33.0	829	81,956	44.9%	19.8%	7,943	98.30%	96.50%
C. sorghi	CCC632	Pusillae	Andropogoneae	*Sorghum bicolor*	60x	35.6	3,660	16,225	44.4%	30.4%	8,208	89.90%	87.10%
C. africana	CCC489	Pusillae	Andropogoneae	*Sorghum bicolor*	56x	37.7	1,781	37,639	42.5%	34.0%	8,119	95.00%	91.50%
C. citrina	CCC265	Citrinae	Cynodonteae	*Distichlis spicata*	64x	43.5	4,772	16,294	41.5%	51.7%	7,821	92.20%	88.20%

† The reference strain *C. purpurea* 20.1 was additionally assembled into 191 scaffolds with a scaffold N50 of 433,221.

‡ Newly identified species Liu *et al. Accepted*)

§ Transposable element (TE) content represented as percent of the genome masked by TEs

five-gene phylogeny in Píchová *et al.* (2018). This topology was supported by neighbor-joining and maximum parsimony super-matrix analyses (Appendix: Fig. A2.2, A2.3). Notable exceptions were the placement of *C. paspali* (sect. *Paspalorum*) which grouped closer to *C. citrina* (Pazoutová, Fucík., Leyva-Mir & Flieger) (sect. *Citrinae*) instead of sect. *Claviceps,* and *C. pusilla* which grouped closer to *C. fusiformis* instead of *C. maximensis* (Fig. 3.1). We also found that sect. *Claviceps* diverged from a common ancestor with sect. *Pusillae* as opposed to sect. *Paspalorum*. Our results provide support for the deeply divergent lineages of sects. *Pusillae*, *Paspalorum*, and *Citrinae* with a long divergent branch resulting in sect. *Claviceps* (Fig. 3.1).

Each of the 2,002 protein orthogroups were also independently aligned and analyzed in the same manner as our super-matrix phylogeny from representative isolates of each species. A density consensus tree of all 2,002 topologies was concordant with our super matrix analysis but reveals evidence of incomplete lineage sorting (ILS), particularly within sect. *Claviceps* (Appendix 2 Fig. A2.4). Analysis of topology clustering revealed similar trends for divergence of sections, although, half of the genes examined represented 19 different topologies within the genus (Appendix 2 Fig. A2.5). A closer examination of the sections revealed that most of the topology variation was most likely due to sect. *Claviceps* with the top six topologies represented by 362 of the genes (18.1%) (Appendix 2 Fig. A2.6), suggesting an abundance of ILS. While grouping of species generally held true to Fig. 3.1, variation was more related to the order of branches, with *C. cyperi* (Loveless), *C. arundinis* (Pažotouvá *et* M. Kolařík), *C. humidiphila*, and *C. perihumidiphila* (M. Liu) showing the most variability. In contrast, the top six topologies of sect. *Pusillae* were represented by 1,666 genes (83.6%), although, we still observed differences in the placements of *C. digitarie* and *C. maximensis* (Appendix 2 Fig. SA2.7). These results

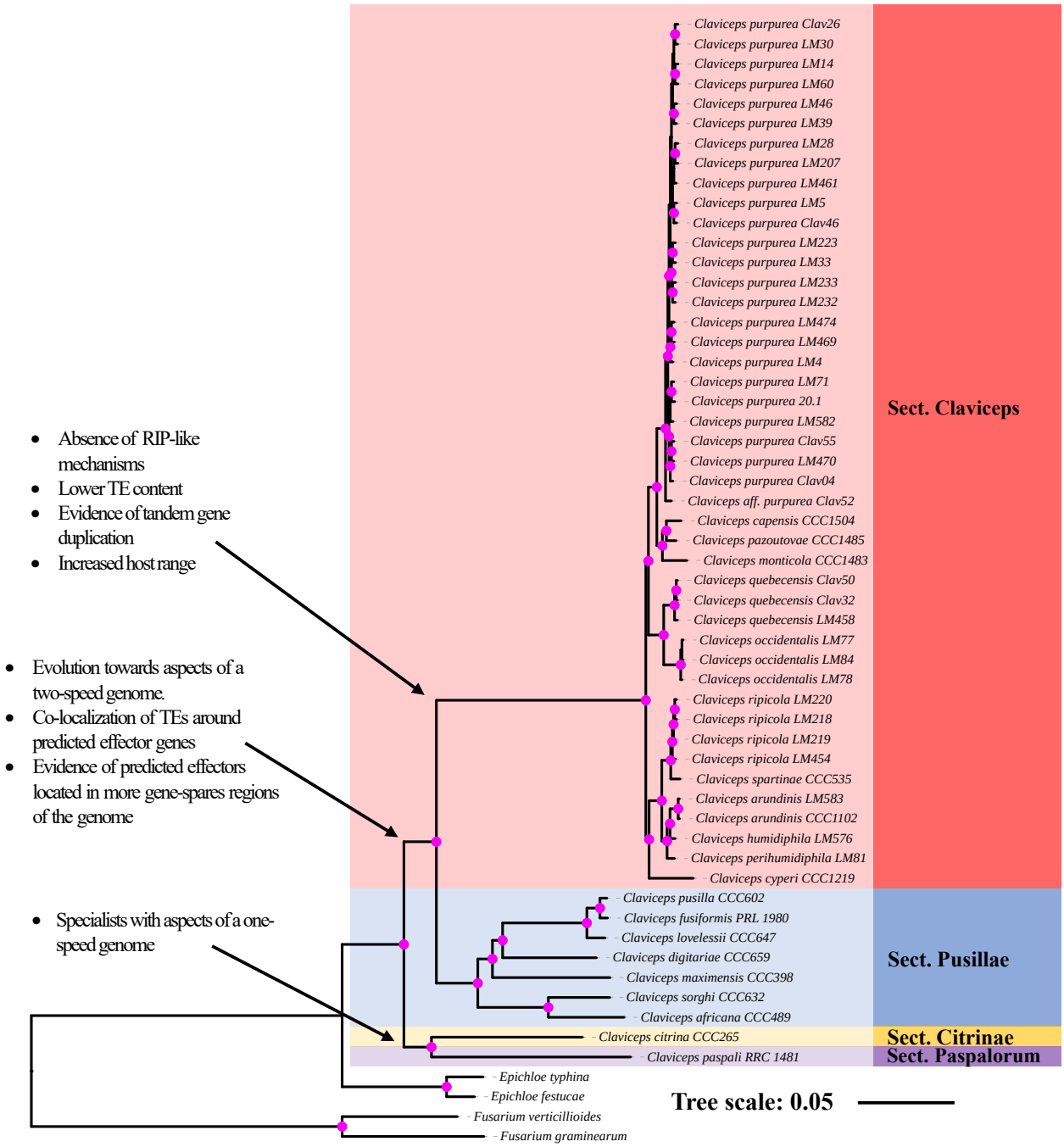
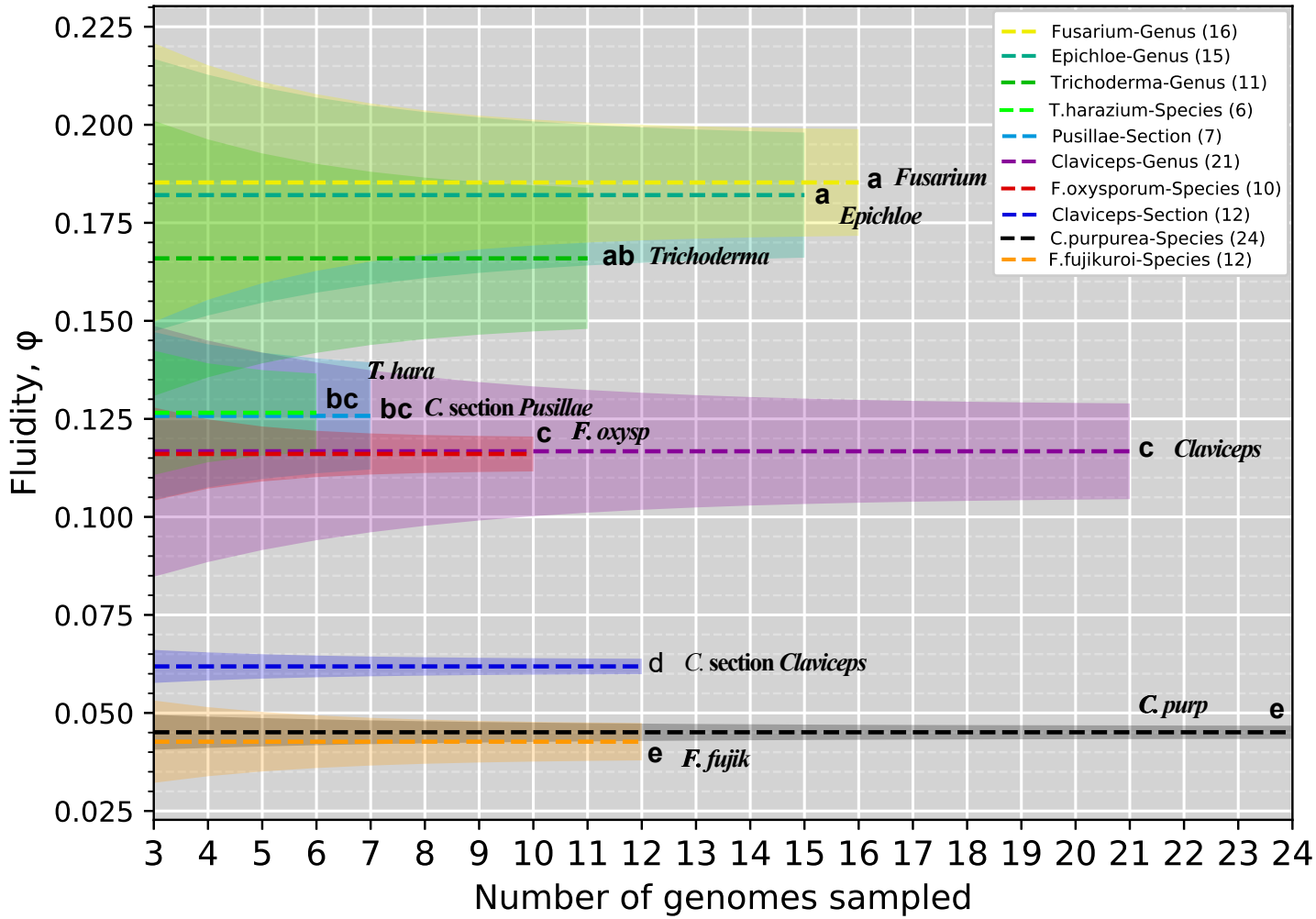


Figure 3.1: Maximum likelihood phylogenetic reconstruction of the *Claviceps* genus using amino acid sequences of 2,002 single copy orthologs with 1000 bootstrap replicates. Pink dots at branches represent bootstrap values ≥ 95. Arrows and descriptions indicate potential changes in genomic architecture between *Claviceps* sections identified in this study.

indicate the presence of ILS within sect. *Claviceps*, sect. *Pusillae*, and across the genus (Appendix 2 Fig. A2.5-7) but a consensus supporting our ML species tree (Fig. 3.1, Appendix 2 Fig. A2.4).

To further elucidate trends of divergence within the genus we examined genomic fluidity using all 82,267 orthogroups from our previous OrthoFinder analysis (Kislyuk *et al.* 2011). Genomic fluidity estimates the dissimilarity between genomes by using ratios of the number of unique orthogroups to the total number of orthogroups in pairs of genomes averaged over randomly chosen genome pairs from within a group on *N* genomes. For example, a fluidity value of 0.05 indicates that randomly chosen pairs of genomes in a group will on average have 5% unique orthogroups and share 95% of their orthogroups (Kislyuk *et al.* 2011). Section *Claviceps*, which is composed of 12 different species, showed a relatively small genomic fluidity (0.0619 ± 0.0019) with little variation, indicating that the pairwise orthogroup dissimilarity between randomly sampled genomes was quite low (Fig. 3.2). The amount of variation between 12 different *Claviceps* species was similar to the variation between 24 *C. purpurea* s.s. isolates, however, there was a significant difference between the fluidities (Fig. 3.2; Appendix 2: Table A2.4; $P < 0.0001$). In comparison, the fluidity of sect. *Pusillae* (0.126 ± 0.014; $P < 0.0001$) was two times greater than the fluidity of sect. *Claviceps* and exhibited greater variation, indicating greater dissimilarities in orthogroups between randomly sampled species of sect. *Pusillae* (Fig. 3.2).

Overall, our ML phylogeny (Fig. 3.1) and genome fluidity analysis (Fig. 3.2) indicate a large evolutionary divergence separating sect. *Claviceps*. Our subsequent analyses of the genomic architecture of all *Claviceps* species examine factors that could be associated with the evolutionary divergence of sect. *Claviceps* and those driving cryptic speciation.



Genomic fluidity (dashed lines) for specified groups within the order Hypocreales. Species level groups contain multiple isolates of a given species, while section and genus level groups contain one strain from representative species to remove phylogenetic bias. Shaded regions represent standard error and were determined from total variance, containing both the variance due to the limited number of samples genomes and the variance due to subsampling within the sample of genomes. Letters correspond to significant difference between fluidities determined through a two-sided two-sample z-test ($P < 0.05$; Appendix 2 Table A2.4). Legend is in descending order based on fluidity, and names are additionally appended to mean lines for clarity.

Transposable element divergences and locations

Transposable element (TE) divergence landscapes revealed an overrepresentation of LTR elements in sects. *Pusillae*, *Citrinae*, and *Paspalorum*. All three sections showed a similar large peak of LTRs with divergences between 5 – 10% (Fig. 3.3; Appendix 2 Fig. A2.8), indicating a relatively recent expansion of TEs. The landscapes of sects. *Pusillae*, *Citrinae*, and *Paspalorum* are in striking contrast to species of sect. *Claviceps* which showed more similar abundances of LTR, DNA, LINE, SINE, and RC (helitron) elements. Species of sect. *Claviceps* showed broader peaks of divergence between 5 – 30% but also showed an abundance of TEs with ~ 0% divergence suggesting very recent TE expansion (Fig. 3.3; Appendix 2 Fig. A2.8). The TE landscape of *C. cyperi* showed a more striking peak of divergence between 5 – 10% that more closely resembled the TE divergences of sects. *Pusillae*, *Paspalorum*, and *Citrinae*. However, the content of the TE peak in *C. cyperi* largely contained DNA, LINE, and unclassified TEs as opposed to LTR's (Appendix 2 Fig. A2.8).

To identify where genes were located in relation to TEs, we calculated the average distance (kb) of each gene to the closest TE fragment. This analysis was performed for predicted effectors, secreted (non-effector) genes, secondary metabolite (non-secreted) genes, and all other genes. Secreted genes and predicted effectors of sects. *Claviceps* and *Pusillae* species were found to be significantly closer to TEs compared to other genes within each respective section (Fig. 3.4; $P < 0.05$), suggesting that these genes could be located in more repeat-rich regions of the genome. It should be noted that we did observe a significant difference ($P < 0.001$, Welch's test) in TE content between sect. *Pusillae* (32.5% ± 9.59%) and sect. *Claviceps* (8.80% ± 1.52%). In both sects. *Claviceps* and *Pusillae* secondary metabolite genes were located farther away from TEs (Fig. 3.4; $P < 0.05$), i.e. repeat-poor regions of the genome.

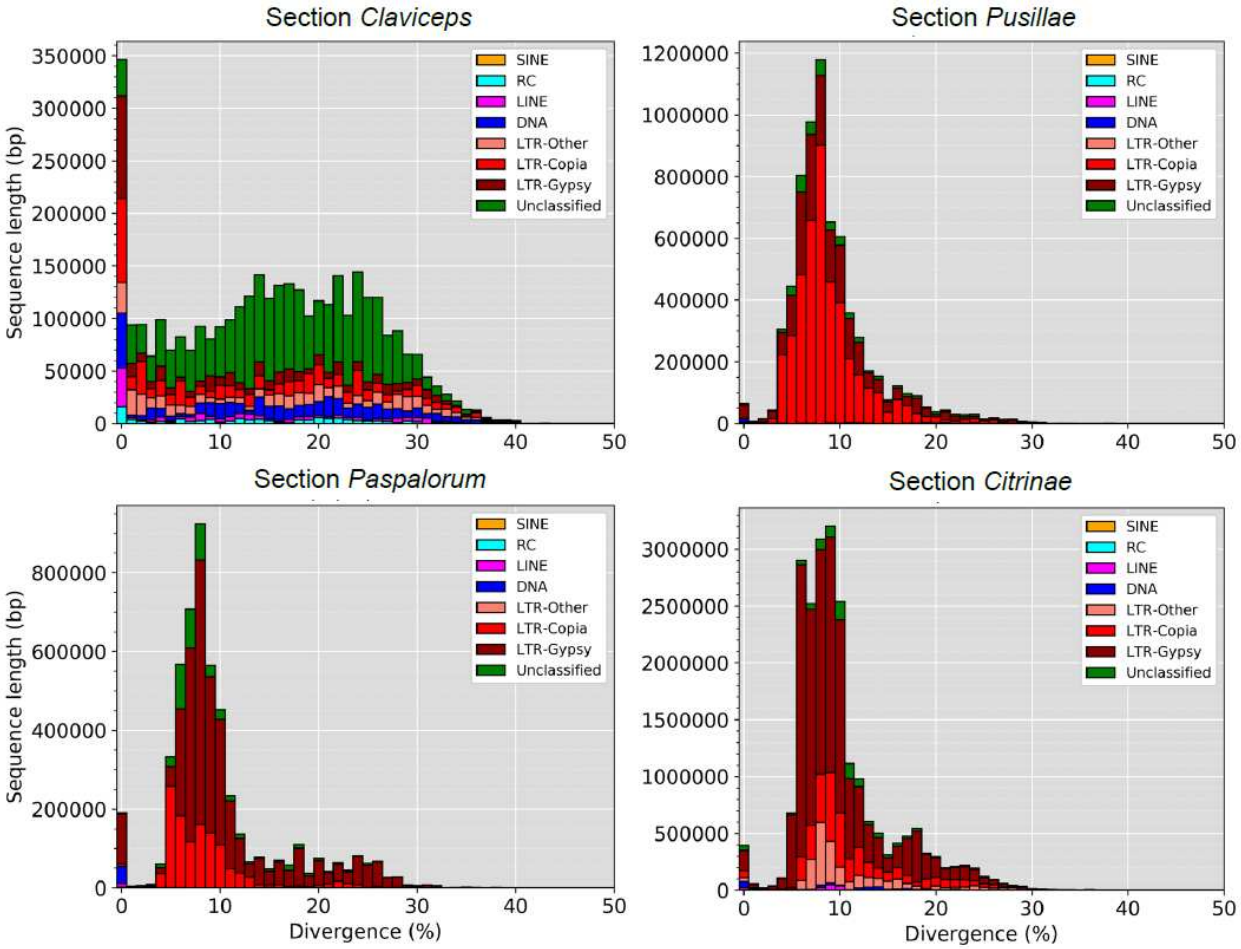


Figure 3.3: Transposable element (TE) fragment divergence landscapes for representative species of each *Claviceps* section; *C. purpurea* 20.1 (sect. *Claviceps*), *C. maximensis* CCC398 (sect. *Pusillae*), *C. paspali* RRC1481 (sect. *Paspalorum*), and *C. citrina* (sect. *Citrinae*). Stacked bar graphs show the non-normalized sequence length occupied in each genome (y-axis) for each TE type based on their percent divergence (x-axis) from their corresponding consensus sequence. Landscape for all remaining isolates can be seen in Appendix 2 Fig. S8.

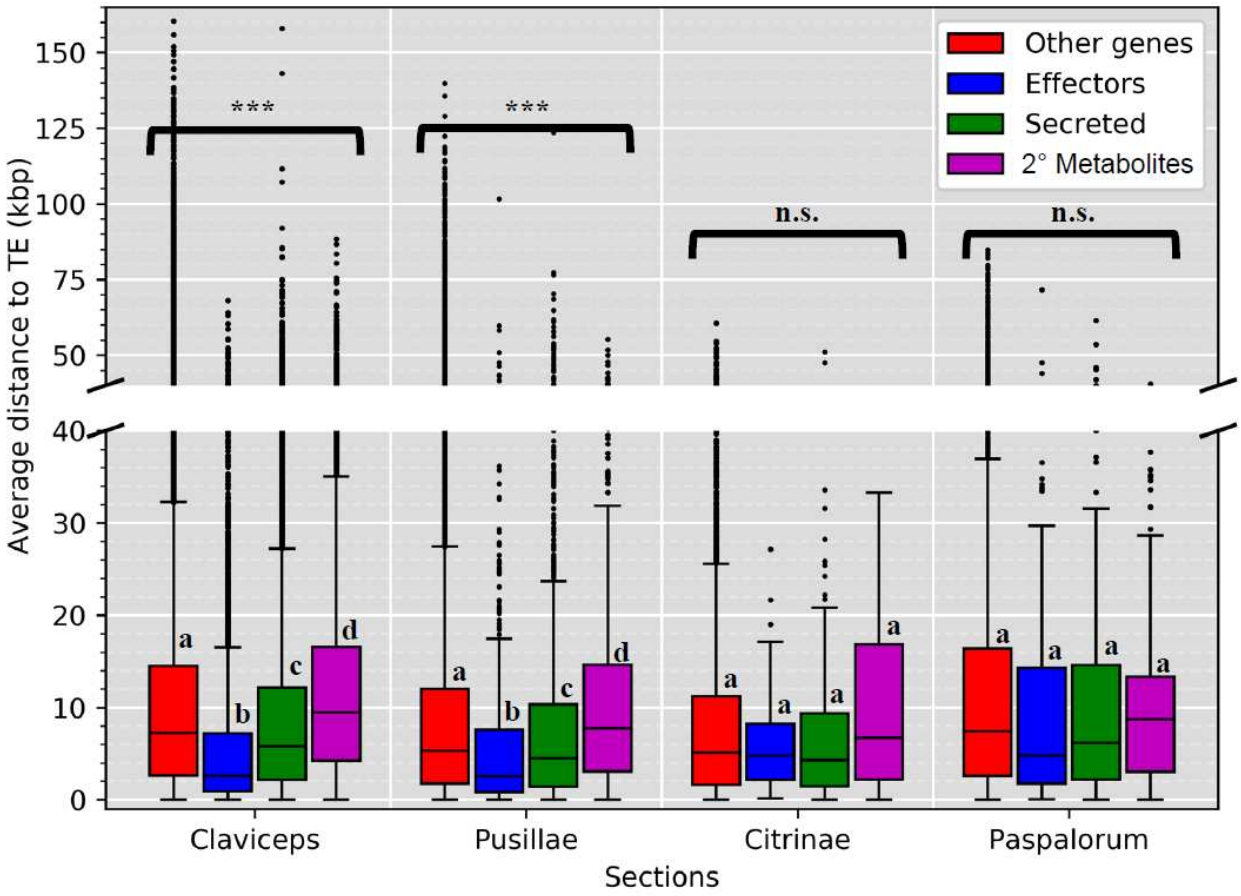


Figure 3.4: Boxplot distributions of predicted effectors, secreted (non-effectors), secondary metabolite (non-secreted) genes and other genes (i.e. genes that are not effectors, secreted, or secondary (2°) metabolite genes) in *Claviceps* sections showing the mean distance (kbp) of each gene to the closest transposable element fragment (5' and 3' flanking distances were averaged together). Kruskal Wallis (*P*-value; * < 0.05, ** < 0.01, *** < 0.001, n.s. = not significant). Pairwise comparison was performed with Mann-Whitney U-test with Benjamini-Hochberg multi-test correction. Letters correspond to significant differences between gene categories within sections ($P < 0.05$). Plots for all individual isolates can been seen in Appendix 2 Fig. A2.9.

These trends hold true for individual isolates, with a notable exception of *C. pusilla* (sect. *Pusillae*) showing no significant differences in the proximity of TEs to specific gene types (Appendix 2 Fig. A2.9; $P > 0.05$). Variation existed in whether particular isolates had significant differences between all other genes compared to secreted genes and secondary metabolite genes, but all species in sects. *Claviceps* and *Pusillae* (aside from *C. pusilla*) had predicted effector genes located significantly closer to TEs (Appendix 2 Fig. A2.9; $P < 0.05$). No significant differences in the proximity of TEs to specific gene types were observed in sects. *Citrinae* and *Paspalorum* (Fig. 3.4; $P > 0.05$), suggesting that TE's are more randomly distributed throughout these genomes.

Genome compartmentalization

To further examine genome architecture, we analyzed local gene density measured as flanking distances between neighboring genes (intergenic regions) to examine evidence of genome compartmentalization (i.e. clustering of genes with differences in intergenic lengths) within each genome. Results showed that all 53 *Claviceps* strains exhibited a one-compartment genome (lack of large-scale compartmentalization). Although, there was a tendency for more genes with larger intergenic regions in sects. *Claviceps* and *Pusillae* compared to sects. *Citrinae* and *Paspalorum* (Fig. 3.5; Appendix 2 Fig. A2.10).

To further clarify evolutionary tendencies, we evaluated whether gene types showed a difference in their flanking intergenic lengths compared to other genes within their genomes. Results showed that predicted effector genes in sect. *Claviceps* had significantly larger intergenic flanking regions compared to other genes, indicating they may reside in more gene-sparse regions of the genome ($P < 0.05$, Fig. 3.5, Appendix 2 Fig. A2.10). Only *C. digitariae* and *C. lovelessi* ((Pažoutová, M.Kolařík & Frederickson) M. Kolařík) (Appendix 2 Fig. A2.10; $P <$

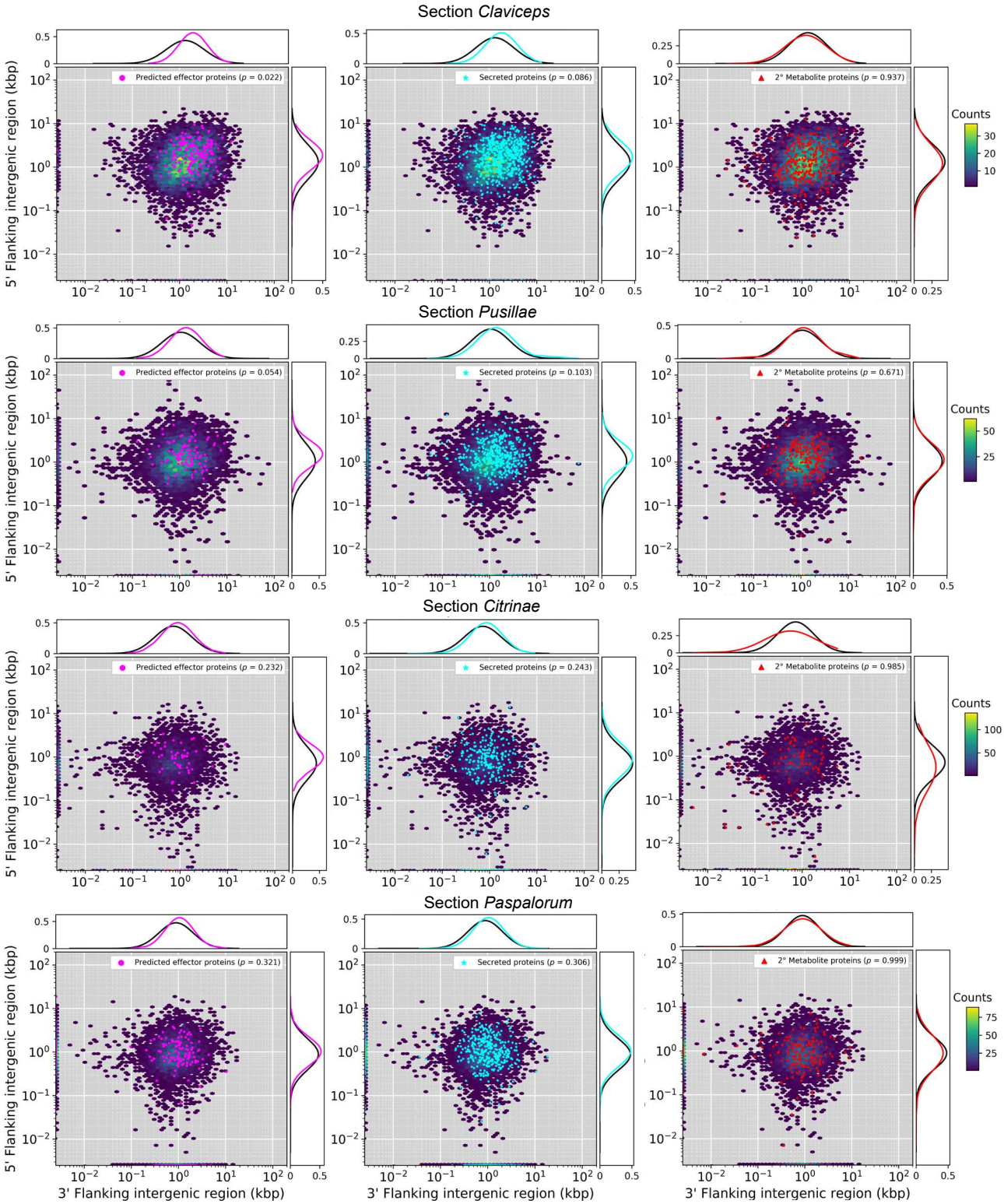


Figure 3.5: Gene density as a function of flanking 5' and 3' intergenic region size (y- and x-axis) of representative isolates of each of the four sections within the *Claviceps* genus; *C, purpurea* 20.1 (sect. *Claviceps*), *C. maximensis* CCC398 (sect. *Pusillae*), *C. paspali* RRC1481 (sect. *Paspalorum*), and *C. citrina* (sect. *Citrinae*). Colored hexbins indicate the intergenic lengths of all genes with color-code indicating the frequency distribution (gene count) according to the legend on the right. Overlaid markers indicate specific gene types corresponding to legends in the top right within each plot. Line graphs (top and right of each plot) depict the frequency distributions of specific gene types (corresponding legend color) and all other genes not of the specific type (black). For visualization purposes the first genes of contigs (5' end) are plotted along the x-axis and the last gene of each contig (3' end) are plotted along the y-axis. For information on statistical test see Chapter 3 Materials and Methods, pg. 74 and for plots of all remaining isolates see Appendix 2 Fig. A2.10.

0.01, $P = 0.024$, respectively) of sect. *Pusillae* had predicted effector genes with significantly larger intergenic regions than other genes, although, *C. fusiformis* (Loveless) and *C. pusilla* (Ces.) were near significant (Fig. 3.5, Appendix 2 Fig. A2.10; $P = 0.054$, $P = 0.056$, respectively). Flanking intergenic lengths of secreted genes also showed larger intergenic lengths and were often significantly larger than other genes (Fig. 3.5; Appendix 2 Fig. A2.10). In contrast, secondary metabolite genes exhibited a widespread distribution of intergenic lengths that were not significantly different than other genes in all 53 *Claviceps* strains ($P > 0.05$, Fig. 3.5; Appendix 2 Fig. A2.10).

RIP analysis

To test for effects of RIP, we assessed the bi-directional similarity of genes against the second closet BLASTp match within each isolate's own genome (Galagan *et al.* 2003; Urguhart *et al.* 2018), supported by a BLASTp analysis against the *rid-1* RIP gene of *Neurospora crassa* (Shear & B.O. Dodge), and calculations of RIP indexes in 1 kb windows (500 bp increments) using The RIPper (van Wyk *et al.* 2019). Results showed that sects. *Pusillae*, *Citrinae*, and *Paspalorum* had homologs of *rid-1*, fewer genes with close identity (≥ 80%), on average 27.4% ± 11.4% of their genomes affected by RIP, a mean RIP composite index of -0.03 ± 0.21, and 325 ± 138 large RIP affected regions (LRAR) covering 3,984 kb ± 2,144 kb of their genomes, indicating past or current activity of RIP (Fig. 3.6; Table 3.2, 3.3; Appendix 2 Table A2.5). This is further supported by an average GC content of 42.84% ± 3.03% (Table 3.1) in sects. *Pusillae*, *Citrinae*, and *Paspalorum*, which is on average 8.81% lower than in sect. *Claviceps* which shows an absence of RIP (reported below). The presence of RIP in sects. *Pusillae*, *Citrinae*, and *Paspalorum* was unexpected given the abundance of TEs within genomes of these sections (Table 3.1; Fig. 3.3; Appendix 2 Fig. A2.8) as RIP should be working to silence and inactive

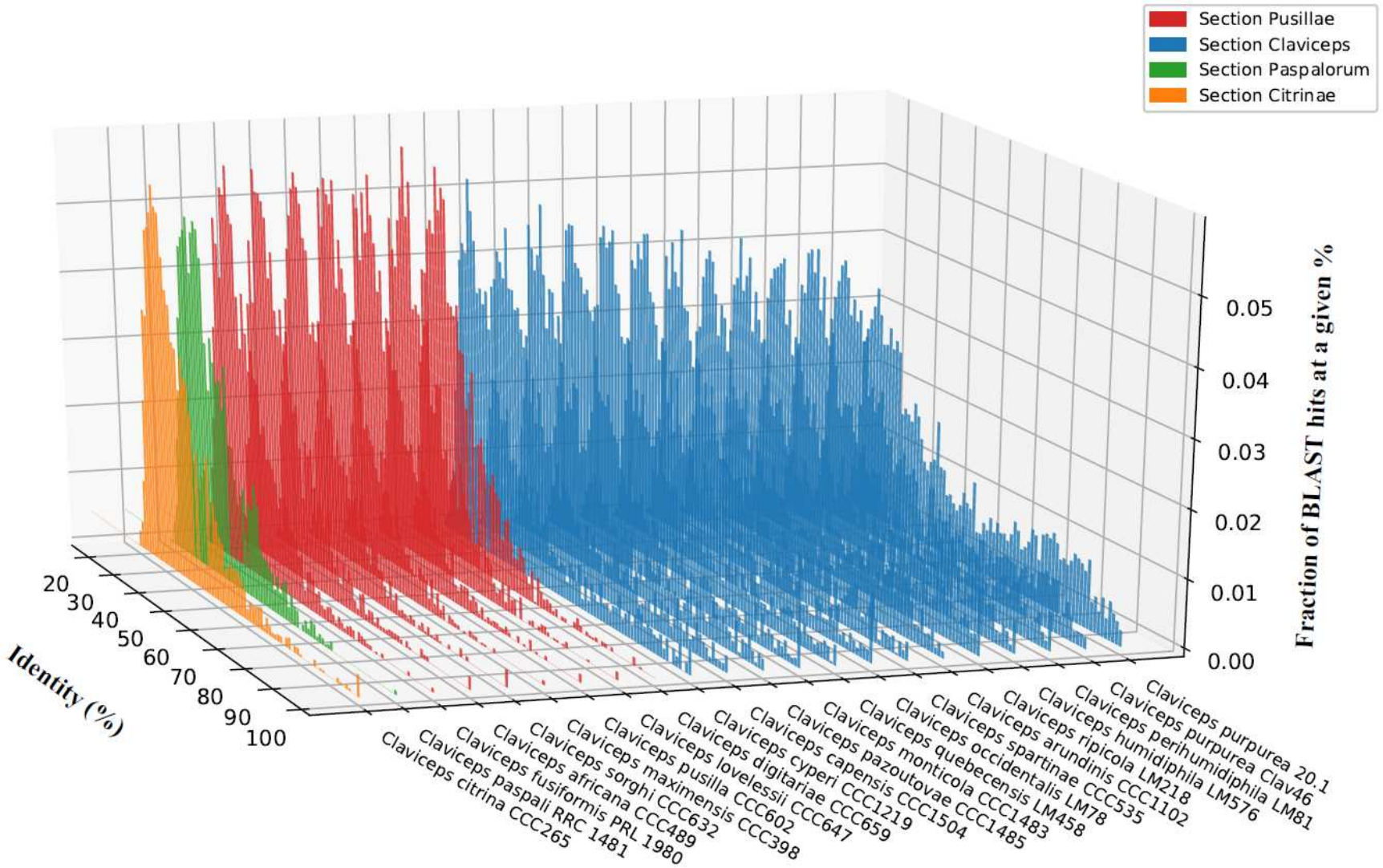


Figure 3.6: Representative isolates of each *Claviceps* species showing the fraction of BLAST hits at a given % identity (y-axis) within each isolate (z-axis) at a given percent identity (x-axis) from the second closet BLASTp match of proteins within each isolate's own genome. Two *C. purpruea* s.s. isolates are shown to compare a newly sequenced genome versus the reference.

Table 3.2: Number of duplicated genes and unique gene pairs with a pairwise identity ≥ 80% and the proportion of these gene pairs that are located next to each other (separated by 0 genes) and separated by five or fewer genes (≤ 5 genes) for all 53 *Claviceps* genomes.

				Separation	
Species	**Strain**	**Gene pairs† (#)**	**Duplicated genes (#)**	**0 genes**	**≤ 5 genes**
C. purpurea	20.1	997	846	11.74%	30.19%
C. purpurea	Clav04	578	710	8.65%	11.94%
C. purpurea	Clav26	429	587	17.48%	30.77%
C. purpurea	Clav46	415	553	18.55%	29.64%
C. purpurea	Clav55	373	523	16.09%	26.81%
C. purpurea	LM4	426	591	19.95%	34.51%
C. purpurea	LM5	412	536	15.78%	30.83%
C. purpurea	LM14	352	493	19.6%	32.95%
C. purpurea	LM28	404	542	14.36%	25.99%
C. purpurea	LM30	352	511	21.88%	38.35%
C. purpurea	LM33	395	528	18.23%	32.66%
C. purpurea	LM39	393	521	17.3%	28.24%
C. purpurea	LM46	383	550	15.4%	29.24%
C. purpurea	LM60	374	519	21.39%	34.22%
C. purpurea	LM71	332	484	17.77%	31.02%
C. purpurea	LM207	354	515	17.8%	27.68%
C. purpurea	LM223	348	487	21.84%	34.2%
C. purpurea	LM232	424	542	15.33%	26.42%
C. purpurea	LM233	673	616	11.59%	24.37%
C. purpurea	LM461	401	557	14.96%	27.93%
C. purpurea	LM469	361	489	20.5%	32.96%
C. purpurea	LM470	410	410	16.34%	27.8%
C. purpurea	LM474	319	496	18.81%	30.72%
C. purpurea	LM582	386	512	13.99%	24.09%
C. aff. purpurea	Clav52	235	355	20.0%	31.06%
C. capensis	CCC1504	144	247	13.89%	21.53%
C. pazoutovae	CCC1485	182	270	14.29%	20.33%
C. monticola	CCC1483	174	272	13.22%	21.84%
C. occidentalis	LM78	173	278	18.5%	26.59%
C. occidentalis	LM77	151	250	17.88%	28.48%
C. occidentalis	LM84	431	313	10.9%	18.79%
C. quebecensis	LM458	176	259	19.32%	26.14%
C. quebecensis	Clav32	189	284	14.29%	24.34%
C. quebecensis	Clav50	161	258	14.91%	26.09%
C. ripicola	LM218	386	523	16.84%	31.61%
C. ripicola	LM219	393	490	15.78%	28.5%
C. ripicola	LM220	412	412	16.02%	28.64%
C. ripicola	LM454	434	546	13.13%	21.43%
C. spartinae	CCC535	251	368	10.36%	16.33%
C. arundinis	CCC1102	431	518	11.6%	22.51%
C. arundinis	LM583	362	442	15.19%	26.24%
C. humidiphila	LM576	401	538	14.96%	23.44%
C. perihumidiphila	LM81	351	494	19.66%	33.62%
C. cyperi	CCC1219	193	244	5.7%	7.77%
C. pusilla	CCC602	9	17	0.0%	0.0%
C. fusiformis	PRL 1980	4	8	0.0%	0.0%
C. lovelessii	CCC647	7	14	0.0%	0.0%
C. digitariae	CCC659	10	18	0.0%	0.0%
C. maximensis	CCC398	3	6	0.0%	0.0%
C. sorghi	CCC632	12	23	0.0%	8.33%
C. africana	CCC489	8	16	0.0%	0.0%
C. citrina	CCC265	24	34	4.17%	4.17%
C. paspali	RRC 1481	1	2	0.0%	0.0%

† Unique pairs (i.e. pairs of gene A : gene B and gene B : gene A are not counted twice).

Table 3.3: Means, standard deviations, and additional statistics of repeat-induced point mutation (RIP) composite indexes and large RIP affected regions (LRARs) for all 53 *Clvaiceps* genomes computed using The RIPper on default settings.

Species	Strain	Section	RIP composite index†	RIP affected windows (#)	RIP genomic content (%)	LRARs (#)	LRARs length (kb)	LRARs genomic content (kb)	LRARs composite index†	LRARs GC content (%)
C. purpurea	20.1	Claviceps	-0.61±0.3	80	0.12%					
C. purpurea	Clav04	Claviceps	-0.59±0.3	136	0.21%					
C. purpurea	Clav26	Claviceps	-0.59±0.3	86	0.14%					
C. purpurea	Clav46	Claviceps	-0.59±0.3	88	0.14%					
C. purpurea	Clav55	Claviceps	-0.59±0.3	96	0.15%					
C. purpurea	LM4	Claviceps	-0.59±0.3	75	0.12%					
C. purpurea	LM5	Claviceps	-0.59±0.3	75	0.12%					
C. purpurea	LM14	Claviceps	-0.59±0.3	66	0.11%					
C. purpurea	LM28	Claviceps	-0.59±0.3	79	0.13%					
C. purpurea	LM30	Claviceps	-0.59±0.3	71	0.12%					
C. purpurea	LM33	Claviceps	-0.59±0.3	76	0.12%					
C. purpurea	LM39	Claviceps	-0.59±0.3	68	0.11%					
C. purpurea	LM46	Claviceps	-0.59±0.3	84	0.14%					
C. purpurea	LM60	Claviceps	-0.59±0.3	61	0.1%					
C. purpurea	LM71	Claviceps	-0.59±0.3	79	0.13%					
C. purpurea	LM207	Claviceps	-0.59±0.3	78	0.13%					
C. purpurea	LM223	Claviceps	-0.59±0.3	94	0.15%					
C. purpurea	LM232	Claviceps	-0.59±0.3	64	0.1%					
C. purpurea	LM233	Claviceps	-0.59±0.3	73	0.12%					
C. purpurea	LM461	Claviceps	-0.59±0.3	72	0.12%					
C. purpurea	LM469	Claviceps	-0.59±0.3	76	0.12%					
C. purpurea	LM470	Claviceps	-0.59±0.3	90	0.15%					
C. purpurea	LM474	Claviceps	-0.59±0.3	94	0.15%					
C. purpurea	LM582	Claviceps	-0.59±0.3	81	0.13%					
C. aff. purpurea	Clav52	Claviceps	-0.59±0.3	65	0.11%					
C. capensis	CCC1504	Claviceps	-0.61±0.3	41	0.07%					
C. pazoutovae	CCC1485	Claviceps	-0.61±0.3	39	0.07%					
C. monticola	CCC1483	Claviceps	-0.59±0.3	45	0.08%					
C. occidentalis	LM78	Claviceps	-0.55±0.3	119	0.2%					
C. occidentalis	LM77	Claviceps	-0.55±0.3	133	0.23%					
C. occidentalis	LM84	Claviceps	-0.55±0.3	111	0.19%					
C. quebecensis	LM458	Claviceps	-0.57±0.3	88	0.15%					
C. quebecensis	Clav32	Claviceps	-0.57±0.3	83	0.14%	2	4.5±0.01	9.0	0.7±0.1	54.1±0.4%
C. quebecensis	Clav50	Claviceps	-0.57±0.3	76	0.13%					
C. ripicola	LM218	Claviceps	-0.57±0.3	82	0.13%					
C. ripicola	LM219	Claviceps	-0.57±0.3	87	0.14%					
C. ripicola	LM220	Claviceps	-0.58±0.3	85	0.14%					
C. ripicola	LM454	Claviceps	-0.57±0.3	82	0.13%					
C. spartinae	CCC535	Claviceps	-0.58±0.3	73	0.12%					
C. arundinis	CCC1102	Claviceps	-0.58±0.3	55	0.09%					
C. arundinis	LM583	Claviceps	-0.58±0.3	60	0.1%					
C. humidiphila	LM576	Claviceps	-0.59±0.3	64	0.1%					
C. perihumidiphila	LM81	Claviceps	-0.58±0.3	80	0.13%					
C. cyperi	CCC1219	Claviceps	-0.60±0.3	90	0.17%	1	4.5±0	4.5	0.7±0.0	53.2±0.0%
C. pusilla	CCC602	Pusillae	0.15±1.0	36,205	38.36%	564	13.7±11.2	7,739	1.3±0.2	25.2±4.1%
C. fusiformis	PRL 1980	Pusillae	0.03±1.3	18,107	16.67%	274	5.9±1.6	1,610	1.9±0.5	4.8±3.2%
C. lovelessii	CCC647	Pusillae	-0.02±1.0	25,695	30.29%	399	10.8±6.6	4,320	1.3±0.2	23.6±3.8%
C. digitariae	CCC659	Pusillae	-0.25±0.9	13,661	20.17%	271	11.5±7.5	3,109	1.4±0.2	22.1±3.4%
C. maximensis	CCC398	Pusillae	-0.24±0.9	13,517	20.37%	148	14.6±13.5	2,156	1.4±0.1	21.4±2.5%
C. sorghi	CCC632	Pusillae	0.01±1.0	21,622	29.57%	348	13.8±13.4	4,804	1.4±0.2	23.7±2.8%
C. africana	CCC489	Pusillae	0.04±1.0	25,266	33.09%	289	22.0±19.7	6,362	1.3±0.1	20.1±3.9%
C. citrina	CCC265	Citrinae	0.36±1.1	43,520	48.69%	503	9.9±6.4	4,957	1.3±0.2	29.8±4.0%
C. paspali	RRC 1481	Paspalorum	-0.35±1.0	5,351	9.05%	131	6.1±1.6	799	1.9±0.4	4.1±2.8%

† *Composite Index Value* [(TpA/ ApT) – (CpA + TpG/ ApC + GpT)], positive values imply RIP

these TEs. While we did not directly test the activity of TEs within our genomes, due to lack of RNAseq data, the peaks of low TE nucleotide divergence (< 10%) in sects. *Pusillae*, *Citrinae*, and *Paspalorum* (Fig. 3.3, Appendix 2 Fig. A2.8) suggest recent activity of TEs (Frantzeskakis *et al.* 2018).

In comparison, species in sect. *Claviceps* lack *rid-1* homologs, showed larger amounts of gene similarity, and a lack of evidence of RIP with only 0.13% ± 0.03% of their genomes putatively affected by RIP, and a mean RIP composite index of -0.59 ± 0.01 suggesting that RIP is inactive (Fig. 3.6; Table 3.2, 3.3; Appendix 2 Table A2.5). Gene pairs sharing a ≥ 80% identity to each other were often located near each other. On average 27.02% ± 5.91% of the pairs were separated by five or fewer genes, and 15.95% ± 3.50% of the pairs were located next to each other, indicating signs of tandem gene duplication within the section (Table 3.2). *Claviceps cyperi* showed the smallest proportions of highly similar tandem genes (7.77% and 5.7%) compared to other species within sect. *Claviceps*. Additional variations in the proportions of highly similar tandem genes between other species of sect. *Claviceps* were not evident as these proportions appeared to vary more between isolate than species (Table 3.2).

Gene cluster expansion

The proteome of *Claviceps* genomes were used to infer orthologous gene clusters (orthogroups) through protein homology and MCL clustering using OrthoFinder. Our results revealed evidence of orthogroups expansion within sect. *Claviceps* as species contained more genes per orthogroup than species of the other three sections (Figure 3.7). To identify the types of gene clusters that were showing putative expansion we filtered our clusters by two criteria; 1) at least one isolates had two or more genes in the orthogroup, 2) there was a significant

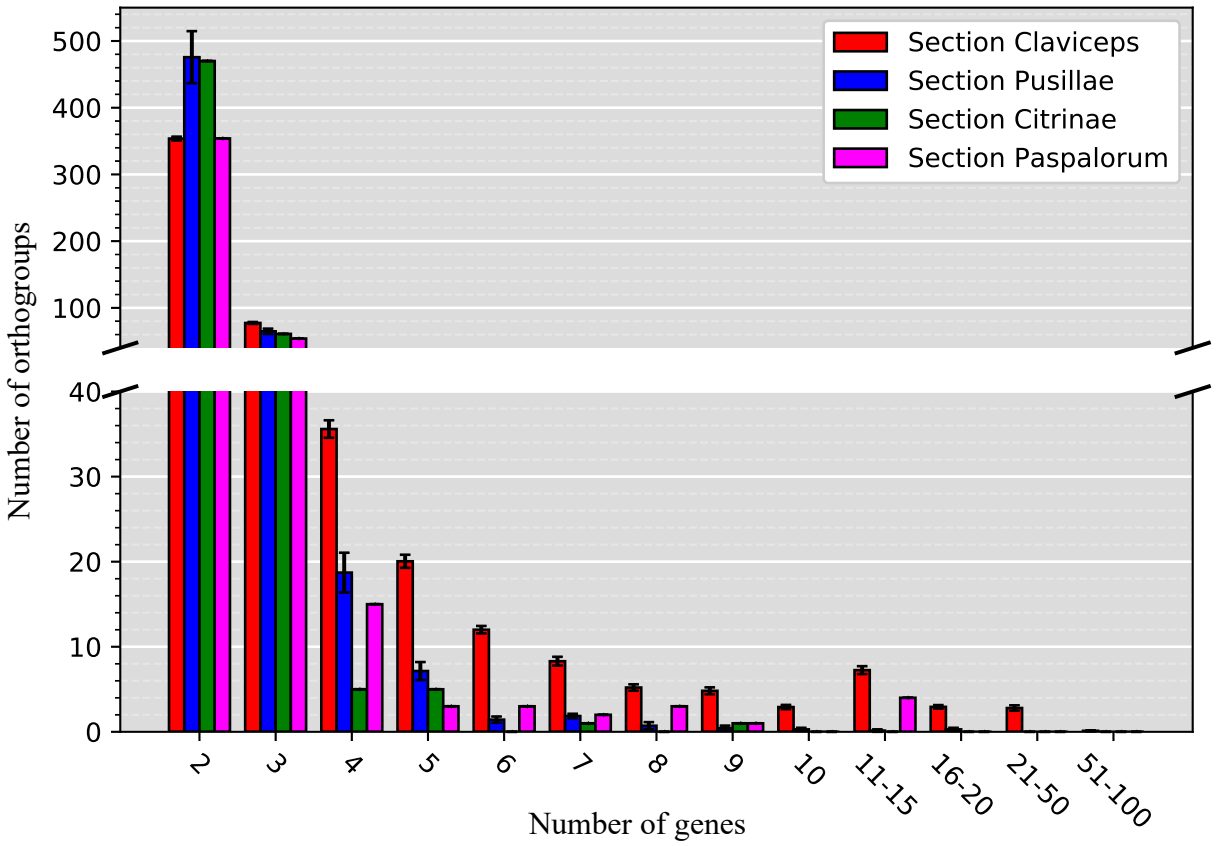


Figure 3.7: Mean number of orthogroups (y-axis) in each section of the genus *Claviceps* containing X number of genes (x-axis), not including single gene orthogroups for better visualization of paralogs. Bars represent standard error.

difference in the mean number of genes per orthogroup between all 44 isolates in sect. *Claviceps* and the 9 isolates from sects. *Pusillae*, *Citrinae*, and *Paspalorum* ($\alpha \leq 0.01$, Welch's test).

Overall, we identified 863 (4.7%) orthogroups showing putative expansion. We observed extensive expansion (orthogroups with observations of ≥ 10 genes per isolate) present in many unclassified, predicted effectors, secreted (non-effector) orthogroups, and orthogroups encoding genes with conserved domains (Fig. 3.8; Appendix 2 Fig. A2.11, A2.12). Transmembrane orthogroups also showed evidence of expansion with several isolates having 5 - 10 genes. Orthogroups with secondary metabolite genes showed the lowest amount of expansion (Fig. 3.8). Overall, sect. *Claviceps* showed expansion in a greater number of orthogroups than sect. *Pusillae*, *Citrinae*, and *Paspalorum* in all categories except transmembranes (Appendix 2 Fig. A2.13). Orthogroups with an average ≥ 5 genes per isolate, within sect. *Claviceps*, contained a variety of functional proteins, with generally more proteins encoding protein/serine/tyrosine kinase domains (Appendix 2 Table A2.6). Additional details can be obtained from Appendix 2 Table A2.7 (ordered orthogroups corresponding to heatmaps; Fig. 3.8 and Appendix 2 Fig. A2.11, A2.12) and Appendix 2 Table A2.8 (orthogroups identification and functional annotation of all proteins).

Within sect. *Claviceps* patterns of gene counts per orthogroup appeared to break down with variations in the number of genes per orthogroups with some presence/absences occurring between isolates and species. Notably, *C. cyperi* (CCC1219) showed the lowest amount of expansion, across all taxa, in comparison to other species of sect. *Claviceps*. In addition, *C. spartinae* (CCC535)*, C. capensis* (Van der Linde, K. Pešicová & M. Kolařík) (CCC1504), *C. monticola* (Van der Linde, K. Pešicová and M. Kolařík) (CCC1483), *C. pazoutovae* (Van der Linde, K. Pešicová and M. Kolařík) (CCC1485), *C. occidentalis* (M. Liu) (LM77, 78, 84),

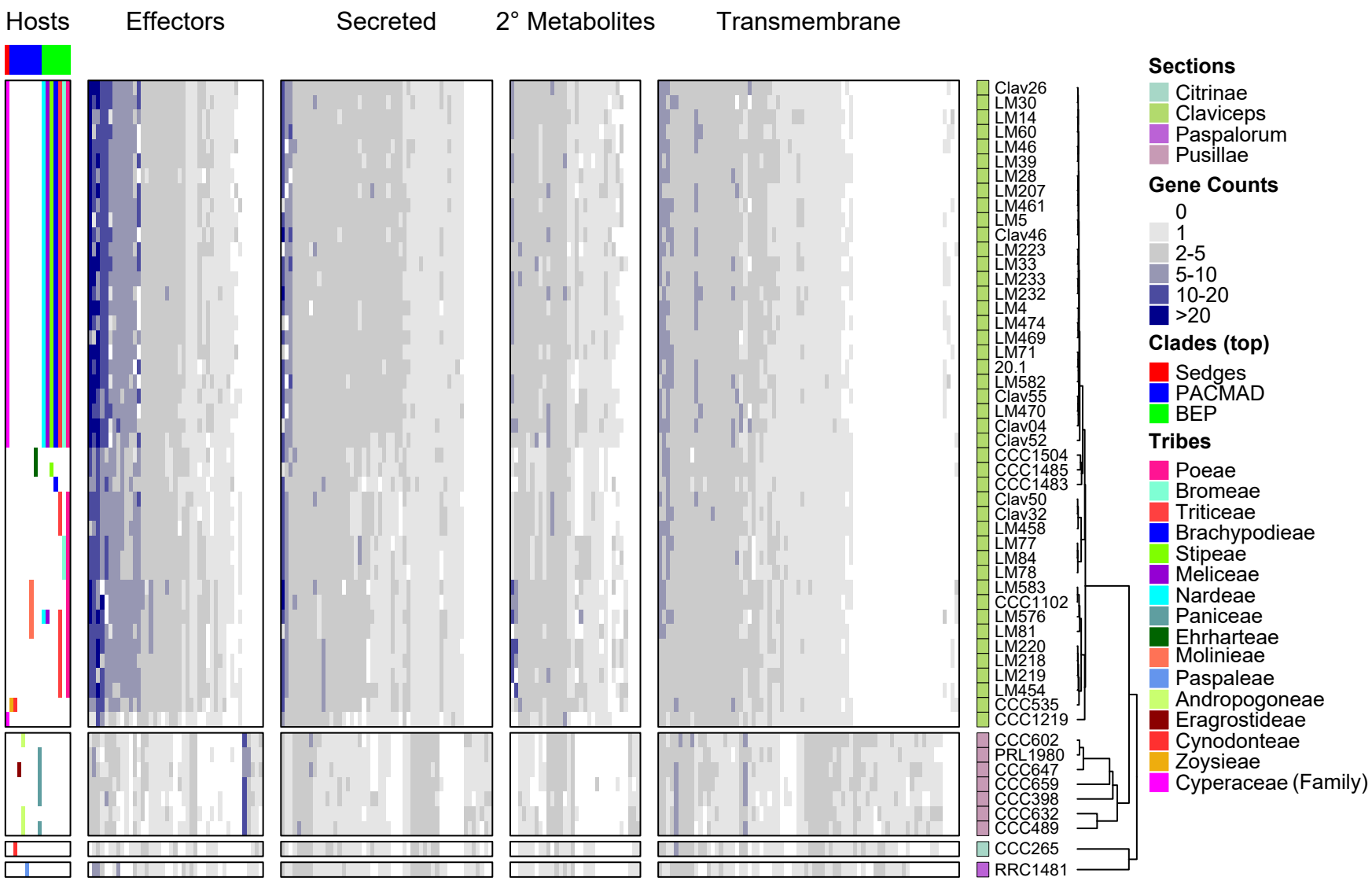


Figure 3.8: Heatmap of gene counts in orthogroups for all 53 *Claviceps* strains ordered based on ML tree in Fig. 3.1 and separated by sections. Orthogroups are separated based on their classification and are only represented once (i.e. secondary (2°) metabolite orthogroups shown are those that are not already classified into the effector or secreted orthogroups) and are ordered based on hierarchical clustering, see Appendix 2 Table A2.7 for list of orthogroups corresponding to the order shown in the heatmaps. The host spectrum (right) is generalized across species, as no literature has determined the existence of race specific isolates within species, is shown on the left side of the figure determined from literature review of field collected samples (Supplementary Material in Píchová *et al.* 2018) and previous inoculation tests Campbell (1957) and Liu *et al.* (*Accepted*). For heatmap of conserved domains see Appendix 2 Fig. A2.11 and for unclassified gene families see Appendix 2 Fig. A2.12.

and *C. quebecensis* (M. Liu and J. Cayouette) (LM458, Clav32, 50) also showed lower expansion (Fig. 3.8, Appendix 2 Fig. A2.11, A2.12). However, no patterns were observed linking the variation in expansions with the literature determined host range of different species within sect. *Claviceps*.

DISCUSSION

Our comparative study of 50 newly annotated genomes from four sections of *Claviceps* has provided us with enhanced understanding of evolution in the genus through knowledge of factors that could be contributing to its diversification. Our results have revealed that despite having nearly identical life-strategies, these closely related species have substantially altered genomic architecture and plasticity, which may drive genome adaptation. One key difference we observe is a shift from characteristic aspects of one-speed genomes (i.e. less adaptable) in narrow host-range *Claviceps* species (sects. *Citrinae* and *Paspalorum*) towards aspects of two-speed genomes (i.e. more adaptable) in broader host-range lineages of sects. *Pusillae* and *Claviceps* (Fig. 3.1).

The basal species of the genus, *C. citrina* (sect. *Citrinae*) and *C. paspali* (sect. *Paspalorum*), are characterized by a proliferation of TEs, particularly LTRs, which do not appear to be co-localized around particular gene types (Fig. 3.4). Coupled with a lack of large-scale genome compartmentalization (Fig. 3.5), these two species can be considered to fit the concept of a one-speed genome which are often considered to be less adaptable and potentially more prone to being purged from the biota (Dong *et al.* 2015; Frantzeskakis *et al.* 2019). This could help explain the lack of section lineages and restricted host range to one grass tribe, as similar patterns of large genome size, abundant TE content, and equal distribution of TEs has been observed in the specialized barley pathogen *Blumeria graminis* f.sp. *hordei* ((DC.) Speer,

Sydowia) (Frantzeskakis *et al.* 2018). Although, rapid adaptive evolution within *B. graminis* f.sp. *hordei*, has been suggested to occur through copy-number variation and/or heterozygosity of effector loci (Dong *et al.* 2015; Frantzeskakis *et al.* 2018, 2019). Our results show a lack of gene duplication occurring in sects. *Citrinae* and *Paspalorum*, likely due to the presence of RIP. However, even with the presence of RIP there was a high LTR content in these species (Fig. 3.3). This suggests that these LTR elements have found a way to avoid RIP or indicate that these species harbor a less active version of RIP as is found in several fungal species (Kachroo *et al.* 1994; Nakayashiki *et al.* 1999; Davière *et al.* 2001; Graïa *et al.* 2001; Ikeda *et al.* 2002; Chalyet *et al.* 2003; Kito *et al.* 2003). Nonetheless, due to the high abundance of TEs (Fig. 3.4) and presence of RIP (Table 3.3), we hypothesize that aspects of RIP "leakage" could be a likely mechanism for evolution in *C. citrina* and *C. paspali* (and similarly sect. *Pusillae*), as has been shown to occur in other fungi (Fudal *et al.*, 2009; Van de Wouw *et al.*, 2010; Hane *et al.*, 2015). It should be noted that since the estimated divergence of sect. *Citrinae* 60.5 Mya (Píchová *et al.* 2018) it has remained monotypic. It was only recently that unknown lineages of sect. *Paspalorum* were identified (Oberti *et al.* 2020), although, these lineages were found on the same genera of host as *C. paspali* (*Paspalum* spp.) supporting our hypothesis that species within sect. *Paspalorum* have restricted host ranges. These recent findings further suggest that lack of additional lineages within these sections could be due to limited records of *Claviceps* species in South America, where the genus is thought to have originated (Píchová *et al.* 2018). Further research into South American populations of *Claviceps* will provide significant insight into the evolution of these two sections.

Members of sect. *Pusillae* also exhibited a proliferation of TEs, however, as this section diverged from sects. *Citrinae* and *Paspalorum* the genomic architecture evolved such that TEs

co-localized around predicted effector genes (Fig. 3.4). This proximity of TEs to effectors persisted in sect. *Pusillae* species (except *C. pusilla*; Appendix 2 Fig. A2.9) and sect. *Claviceps* species potentially resulting in the large intergenic regions flanking predicted effector genes (Fig. 3.5, Appendix 2 Fig. A2.10). Together, these genomic alterations indicate aspects of a two-speed genome (Dong *et al.* 2015; Möller and Stukenbrock 2017). We hypothesize that these observed genomic changes influenced the divergence and adaptability of sects. *Pusillae* and *Claviceps* (Fig. 3.1) (Raffaele and Kamoun 2012; Stukenbrock 2013; Möller and Stukenbrock 2017).

Furthermore, our analyses suggest that the divergence of sect. *Claviceps*, from sect. *Pusillae*, is associated with a loss of RIP (Fig. 3.1, 3.6; Table 3.3). In the absence of RIP, the gene-sparse regions rich in TEs and effectors could be hot spots for duplication, deletion, and recombination (Galagan *et al.* 2003, 2004; Raffaele and Kamoun 2012; Dong *et al.* 2015; Faino *et al.* 2016; Möller and Stukenbrock 2017; Frantzeskakis *et al.* 2018, 2019). This would explain the observations of tandem gene duplication within the section (Fig. 3.6-8; Table 3.2; Appendix 2 Fig. A2.11-13), which may facilitate rapid speciation, as has been postulated in several smut fungi (Kämper *et al.* 2006; Schirawski *et al.* 2010; Dutheil *et al.* 2016). In fact, *C. cyperi*, a species of sect. *Claviceps* and thought to be ancestral from ancestral state reconstructions of host range (Píchová *et al.* 2018), showed the least amount of gene cluster expansion and tandem duplication (Fig. 3.8; Table 3.2; Appendix 2 Fig. A2.11, A2.12). Potentially indicating that gene duplication is contributing to the divergence of new species, as other species in sect. *Claviceps* have increased genome size, gene count, and number of closely related gene pairs (≥ 80% identity) (Table 3.1, 3.2). Within sect. *Claviceps* gene duplication is likely facilitated by recombination events during annual sexual reproduction (Esser and Tudzynski 1978). Future studies on recombination will be critical to our understanding of the mechanisms driving gene

duplication and elucidating factors associated with the observations of abundant incomplete lineage sorting (Pease and Hahn 2013) within the section.

Substantially altered genomic architecture and plasticity between *Claviceps* sections was observed in this study, yet it is unclear whether the evolution of these genomes were caused by contact with new hosts and different climates as ancestral lineages migrated out of South America (Píchová *et al.* 2018) or that the evolution towards a two-speed genome provided an advantage in adapting to new hosts or environments. Further research is needed to clarify this point. As sects. *Pusillae* and *Claviceps* have larger host ranges (5 tribes and 13 tribes, respectively) and increased levels of speciation (Píchová *et al.* 2018), they represent ideal systems to test this hypothesis. It is postulated that sect. *Pusillae* was transferred to Africa (ca 50.3 Mya), while sect. *Claviceps* originated in North America (ca. 20.7 Mya), and it is likely that the common ancestor shared between these sections (Fig. 3.1) had strains that were transferred to Africa, likely due to insect vectors via transatlantic long-distance dispersal (Píchová *et al.* 2018). The strains that remained, in South America, likely persisted but appeared to not speciate for roughly 30 Mya (Píchová *et al.* 2018), despite having aspects of a more adaptable two-speed genome (Fig. 3.4, 3.5). Limited sampling records could be a factor contributing to this lack of speciation during this 30 My period, but it could also be suggested that the ancestral species of sects. *Claviceps* did not diverge due to a lack of diversification of host species (Píchová *et al.* 2018). It is well known that *Claviceps* species share a rather unique relationship with their hosts (strict ovarian parasites). The evolution of *Claviceps* appears to be primarily driven by the evolution and diversification of the host species (Píchová *et al.* 2018). This can be inferred from divergence time estimates which show that the crown node of sect. *Pusillae* aligns with the crown node of PACMAD grasses (ca. 45 Mya) (Bouchenak-Khelladi *et al.* 2010, Píchová *et al.*

2018), suggesting that these two organisms radiated in tandem after ancestral strains of sect. *Pusillae* were transferred to Africa. Similarly, the estimated crown node of sect. *Claviceps* corresponds with the origin of the core Pooideae (Poeae, Triticeae, Bromeae, and Littledaleae), which occurred in North America (ca. 33-26 Mya) (Bouchenak-Khelladi *et al.* 2010; Sandve and Fjellheim 2010).

Such a large difference between the estimate divergence age (~ 30 My) and long divergence branch (Fig. 3.1) between sect. *Clavcieps* and the other three sections (Píchová *et al.* 2018) suggests that a sudden event sparked the adaptive radiation within this section (Fig. 3.1). Under an assumption that ancestral strains of sect. *Claviceps* were infecting sedges (Cyperaceae), as is seen in the basal species *C. cyperi*, a host jump to BOP grasses could have ignited the rapid speciation of sect. *Claviceps*, similar to the suggested tandem radiation of sect. *Pusillae* with the PACMAD grasses in Africa. However, unknown factors might be responsible for the drastic genomic changes (i.e. putative loss of RIP) observed in sect. *Claviceps*, as no such changes were observed in sect. *Pusillae*. The radiation of the core Pooideae occurred after a global super-cooling period (ca. 33 - 26 Mya) in North America. During this period, Pooideae experienced a stress response gene family expansion which enabled adaptation and diversification to cooler, more open, habitats (Kellogg 2001; Sandve and Fjellheim 2010). As gene cluster expansion was observed in sect. *Claviceps* (the only section to infect BOP grasses) it suggests that the same environmental factors that caused the radiation of Pooideae could have similarly affected sect. *Claviceps* (Kondrashov 2012) and might have resulted in the host jump to Pooideae, and potentially other BOP tribes. Interestingly, one of the orthogroups significantly expanded in sect. *Claviceps* (OG0000016) contains proteins associated with a cold-adapted (Alias *et al.* 2014) serine peptidase S8 subtilase (MER0047718; S08.139) (Appendix 2 Table A2.6). Although the

crown node of sect. *Claviceps* is estimated at approximately 5 - 10 Ma before the radiation of the core Pooideae, the 95% highest posterior density determined in Píchová *et al.* (2018) could indicate both radiation events occurred at similar times.

Further examination of *Claviceps* species in South and Central America needs to be conducted to better elucidate the evolution and dispersal of the genus (Píchová *et al.* 2018). Efforts should focus on the elusive *C. junci* (J.F. Adams), a pathogen of Juncaceae (rushes), which is thought to reside in sect. *Claviceps* based on morphological and geographic characteristics (Langdon 1952; Píchová *et al.* 2018). This species, and potentially others, will provide further insight into the early evolution of sect. *Claviceps* and could bridge the current gap between the environmental factors that sparked the radiation of the core Pooideae and sect. *Claviceps*.

Chapter 4: A large accessory genome, high recombination rates, and selection of secondary metabolite genes help maintain global distribution and broad host range of the fungal plant pathogen *Claviceps purpurea*[2]

INTRODUCTION

Pangenomes can provide useful insight into a species distribution and lifestyle through examination of gene functional diversity, abundance, and distribution into core and accessory genomes. These variations often provide fitness advantages and promote adaptive evolution of the organism (Araki *et al.* 2006; Hartmann *et al.* 2018; Brynildsrud *et al.* 2019). In prokaryotes the existence of more open pangenomes (large accessory) has been suggested to be the result of adaptive evolution that allows organisms, with large long-term effective population sizes, to migrate into new ecological niches (McInerney *et al.* 2017). Whereas closed pangenomes (larger core) are found to be associated with more obligate and specialized organisms (McInerney *et al.* 2017). Similar results have been identified in fungal species, where a range of saprotrophic to opportunistic yeasts were found to have accessory genomes representing ~ 9 – 19% of the genes (McCarthy and Fitzpatrick 2019), while *Zymoseptoria tritici* ((Desm.) Quaedvl. and Crous), a global wheat pathogen, had 40% of genes in the accessory genome (Badet *et al.* 2020). This increase in the *Z. tritici* accessory genome reflects the global distribution of this pathogen that must continuously adapt to overcome new host resistances and multiple cycles of annual fungicide applications (Sánchez-Vallet *et al.* 2018; Badet *et al.* 2020). While the identification of pangenome sizes provide valuable knowledge of polymorphic gene content, which can be used

[2] This chapter contain preprint work from: Wyka SA, Mondo SJ, Liu M, Nalam V, Broders KD. 2020. A large accessory genome, high recombination rates, and selection of secondary metabolite genes help maintain global distribution and broad host range of the fungal plant pathogen *Claviceps purpurea*. bioRxiv doi: 10.1101/2020.05.20.106880

to infer the lifestyle of the species, a combination of pangenomic and alternative genomic analyses provide a deeper understanding of the primary factors that are contributing to pangenome structure and the adaptive trajectory of the organism.

Claviceps purpurea is a biotrophic ascomycete plant pathogen that has a specialized ovarian-specific non-systemic lifestyle with its grass hosts (Píchová *et al.* 2018). Despite the specialized infection pattern, *C. purpurea* has a broad host range of ~ 400 grass species across eight grass tribes, including economically important cereal crops such as wheat, barley, and rye and has a global distribution (Píchová *et al.* 2018). However, the mechanisms that underlie the evolutionary success of this species is still understudied. Unlike other pathogens of cereal crops, researchers have been unsuccessful in identifying qualitative resistance (R) genes in crop or wild grass varieties (Menzies and Turkington 2015; Menzies *et al.* 2017; Gordon *et al.* 2020). Menzies *et al.* (2017) did note the potential for a complex virulence and host susceptibility relationship of *C. purpurea* on durum and hexaploid wheat varieties, however, virulence was determined if sclerotia weighed > 81 mg; indicating that *C. purpurea* is able to initiate its biotrophic interaction but might be arrested during the final stages of sclerotia development. During infection the fungus does not induce necrosis or hypersensitive response (host mediated cell death) in its host, instead it actively manages to maintain host cell viability to obtain nutrients from living tissue through a complex cross-talk of fungal cytokinin production (Hinsch *et al.* 2015, 2016; Oeser *et al.* 2017; Kind *et al.* 2018a, 2018b). Furthermore, Chapter 3 revealed evidence of tandem gene duplication occurring in genes often associated with pathogenicity or evasion of host defenses (effectors), which could implicate their role in the success of the species, however, the factors that were influencing these duplication events remain unclear.

Claviceps purpurea is also known for its diverse secondary metabolite profile of ergot alkaloids and pigments (Schardl *et al.* 2013; Tudzynski and Neubauer 2014; Neubauer *et al.* 2016; Flieger *et al.* 2019). Fungal secondary metabolites can play important roles in plant-host interactions as virulence factors but can also increase the fitness of the fungus through stress tolerance (Avalos and Carmen Limon, 2015; Píchová *et al.* 2018; Pusztahelyi *et al.* 2019). It was also recently postulated that the evolution of *C. purpurea* was associated with a host jump and subsequent adaptation and diversification to cooler, more open habitats (Píchová *et al.* 2018; Chapter 3). In addition, likely due to the toxicity of ergot alkaloids, grass grazing mammals showed avoidance in grazing grass infected with *C. purpurea*, suggesting a potential for beneficial effects for the host plant (Wäli *et al.* 2013). This along with other evidence of neutral to positive effects of infection to host plants (Raybould *et al.* 1998; Fisher *et al.* 2007) suggest that *C. purpurea* is a conditional defensive mutualist (Wäli *et al.* 2013).

In this study, we implement a comprehensive population genomic analysis to gain a deeper understating of factors governing the evolution and adaptive potential of *C. purpurea*. Using 24 isolates, from six countries and three continents, we construct the pangenome and subsequently use single-copy core orthologs to identity genes under positive selection. Full genome alignments were further utilized to estimate population recombination rates and predict recombination hotspots. We observed a large accessory genome likely maintained by a large effective population size and high recombination rates, which subsequently influence an overall trend of purifying selection and likely help defend against TE expansion. In addition, we observed that the *lpsA1* and *lpsA2* genes of the well-known ergoline biosynthetic cluster were likely the result of a recombination event.

MATERIALS AND METHODS

Genome data

Haploid genome data from a collection of 24 isolates was utilized in this study to provide a comprehensive analysis of *Claviceps purpurea*. The 32.1 Mb reference genomes of *C. purpurea* strain 20.1 was sequenced in 2013 using a combination of single and paired-end pyrosequencing (3 kb fragments) resulting in a final assembly of 191 scaffolds (Schardl *et al.* 2013; NCBI: SAMEA2272775). The remaining 23 isolates were recently sequenced, assembled, and annotated in Chapter 3 (NCBI BioProject: PRJNA528707), representing a collection of isolates from USA, Canada, Europe, and New Zealand (Table 4.1). The reference genome was subject to an amino acid cutoff of 50 aa to match the other 23 isolates. In this study, we report the pangenome of *C. purpurea*, analysis of the population genomic recombination, and the landscape of genes under positive selection.

Gene functional and transposable element (TE) annotations utilized were those reported in Chapter 3. In brief, secondary metabolite clusters were predicted using antiSMASH v5 (Blin *et al.* 2019), with all genes belonging to identified clusters classified as "secondary (2°) metabolites". Functional domain annotations were conducted using InterProScan v5 (Jones *et al.* 2014), HMMer v3.2.1 (Wheeler and Eddy 2013) search against the Pfam-A v32.0 and dbCAN v8.0 CAZYmes databases, and a BLASTp 2.9.0+ search against the MEROPs protease database v12.0 (Rawlings *et al.* 2018). Proteins were classified as secreted proteins if they had signal peptides detected by both Phobius v1.01 (Käll 2007) and SignalP v4.1 (Nielsen 2017) and did not possess a transmembrane domain as predicted by Phobius and TMHMM v2.0 (Krogh *et al.* 2001). Effector proteins were identified by using EffectorP v2.0 (Sperschneider *et al.* 2018) on the set of secreted proteins for each genome. Transmembrane proteins were identified if both

Table 4.1: Collection and annotation statistics for the 24 *Claviceps purpurea* genomes used in this study.

Strain ID†	Origin	Host	Genome size (Mb)	Genomic GC (%)	TE‡ content (%)	Gene count	BUSCO§ score (%)
20.1	Germany	*Secale cereale*	32.1	51.6%	10.87%	8,703	95.5%
Clav04	USA: Colorado	*Bromus inermis*	31.8	51.7%	10.05%	8,824	97.7%
Clav26	USA: Colorado	*Hordeum vulgare*	30.8	51.7%	9.07%	8,737	98.0%
Clav46	USA: Wyoming	*Secale cereale*	30.8	51.7%	9.68%	8,597	97.1%
Clav55	New Zealand	*Lolium perenne*	30.7	51.8%	9.80%	8,480	97.0%
LM4	Canada: Manitoba	*Tricosecale*	30.6	51.8%	10.04%	8,470	96.9%
LM5	Canada: Manitoba	*Hordeum vulgare*	30.5	51.8%	8.95%	8,508	97.4%
LM14	Canada: Saskatchewan	*Hordeum vulgare*	30.6	51.8%	9.96%	8,422	97.3%
LM28	Canada: Saskatchewan	*Triticum aestivum*	30.6	51.7%	9.58%	8,713	97.0%
LM30	Canada: Saskatchewan	*Hordeum vulgare*	30.6	51.8%	9.35%	8,526	96.3%
LM33	Canada: Manitoba	*Hordeum vulgare*	30.5	51.8%	9.20%	8,557	97.1%
LM39	Canada: Saskatchewan	*T. turgidum subsp. durum*	30.5	51.8%	10.11%	8,591	97.0%
LM46	Canada: Alberta	*T. turgidum subsp. durum*	30.6	51.8%	9.64%	8,455	97.0%
LM60	Canada: Manitoba	*Avena sativa*	30.6	51.7%	9.29%	8,498	97.1%
LM71	United Kingdom	*Alopercurus myosuroides*	30.5	51.8%	9.59%	8,472	97.0%
LM207	Canada: Manitoba	*Elymus repens*	30.5	51.8%	9.18%	8,475	97.0%
LM223	Canada: Manitoba	*Bromus riparius*	30.8	51.7%	10.53%	8,438	96.6%
LM232	Canada: Manitoba	*Phalaris canariensis*	30.7	51.7%	9.36%	8,512	96.7%
LM233	Canada: Manitoba	*Phalaris canariensis*	30.6	51.8%	9.89%	8,717	96.6%
LM461	Canada: Quebec	*Elymus repens*	30.5	51.8%	8.42%	8,656	97.3%
LM469	Canada: Ontario	*Triticum aestivum*	30.5	51.8%	10.01%	8,394	96.5%
LM470	Canada: Ontario	*Elymus repens*	30.5	51.8%	8.95%	8,591	96.8%
LM474	Canada: Ontario	*Hordeum vulgare*	30.6	51.8%	9.38%	8,500	97.2%
LM582	Czech Republic: Bezdedice	*Secale cereale*	30.7	51.8%	9.55%	8,518	95.5%

† NCBI BioProject: PRJNA528707

‡ Transposable element content presented in Chapter 3, as a proportion of genomic sequences

§ Benchmarking Universal Single-Copy Orthologs Dikarya database (odb9)

Phobius and TMHMM detected transmembrane domains. Transposable elements fragments were identified following procedures for establishment of *de novo* comprehensive repeat libraries set forth in Berriman *et al.* (2018) through a combined use of RepeaModeler v1.0.8 (Smit & Hubley 2015), TransposonPSI (Hass 2010), LTR_finder v1.07 (Xu & Wang 2007), LTR_harvest v1.5.10 (Ellinghaus *et al.* 2008), LTR_digest v1.5.10 (Steinbiss *et al.* 2009), Usearch v11.0.667 (Edgar 2010), and RepeatClassifier v1.0.8 (Smit & Hubley 2015) with the addition of all curated fungal TEs from RepBase (Bao *et al.* 2015). RepeatMasker v4.0.7 (Smit *et al.* 2015) was then used to identify TE regions and soft mask the genomes. These steps were automated through construction of a custom script, TransposableELMT (https://github.com/PlantDr430/TransposableELMT) (Chapter 3).

Pangenome analysis

The pangenome was constructed using OrthoFinder v2.3.3 (Emms *et al.* 2019), on all genes identified from the 24 genomes, to infer groups of orthologous gene clusters (orthogroups). OrthoFinder was run using BLASTp on default settings. For downstream analysis, gene clusters were classified as secreted, predicted effectors, transmembrane, secondary (2°) metabolites, carbohydrate-degrading enzymes (CAZys), proteases (MEROPs), and conserved domain (conserved) clusters if ≥ 50% of the strains present in a gene cluster had at least one protein classified as such. Gene clusters not grouped into any of the above categories were categorized as unclassified.

Core and pangenome size curves were extrapolated from resampling of 24 random possible combinations for each pangenome size of 1 - 24 genomes and modelled by fitting the power law regression formula: $y = Ax^B + C$ using the curve_fit function in the Python module

Scipy v1.4.1. These processes were automated through the creation of a custom python script (https://github.com/PlantDr430/FunFinder_Pangenome).

Positive selection

To investigate the positive selection landscape of genes we collected a total of 6,243 single-copy orthologs across all 24 genomes (See Table 4.2 for detailed report). For each ortholog cluster sequences were aligned using MUSCLE v3.8.1551 (Edgar 2004) on default settings and values of dN, dS, and dN/dS (omega, ω) were estimated using the YN00 (Yang and Nielsen 2000) method in PAML v4.8 using default parameters. Each ortholog was then individually examined for evidence of positive selection. Guide trees were generated for each ortholog cluster using FastTree version 2.1.10 SSE3 and positive selection was detected using the CodeML algorithm (Yang 2007) in PAML v4.8 with parameters: NSites = 0 1 2 3 7 8, CodonFreq = 2, seqtype = 1, kappa = 0.3, omega = 1.3, ncatG = 10. Due to high average nucleotide similarities in pairwise BLASTn searches within each ortholog (Appendix 3 Fig. A3.4) we utilized a stringent filtering method to enhance our confidence in the selection of genes with positive selection signatures. Orthologs were only identified as being under positive selection if they were significant at $\alpha \leq 0.01$ using a likelihood ratio test (df -2, $\chi 2$ critical value = 9.13) in both the M7 vs. M8 and M2 vs. M1 model comparisons. In addition, orthologs also needed to contain at least one specific amino acid residue significantly ($\alpha \leq 0.01$) identified as being under positive selection using the Bayes Empirical Bayes algorithm integrated into PAML (Yang 2007), in both the M8 and M2 models.

For statistical purposes, each gene cluster was only characterized by one functional category in the order displayed in Table 4.2 (i.e. secreted genes are those not already classified as effectors, etc). After filtering for positive selection, gene functional categories were examined for

Table 4.2: PAML and CodeML processing information and filtering of core orthogroups for calculation of dN/dS (ω) ratios and examination of positive selection signatures.

Total gene clusters (Pangenome)	10,540		
Single-copy gene clusters	6,244		
Number of clusters with N/A PAML results	43		
Cluster Classification (non-redundant) †:	Total Pangenome	Total Core‡	Single copy§
Effectors	257	100 (38.9%)	84 (84.0%)
Secreted	366	278 (75.9%)	253 (91.0%)
2° Metabolites	313	202 (64.5%)	181 (89.6%)
Transmembrane	1,210	998 (82.5%)	949 (95.1%)
MEROPs	167	149 (89.2%)	143 (96.0%)
CAZys	75	68 (90.7%)	66 (97.1%)
Conserved	4,754	3,985 (83.8%)	3,808 (95.6%)
Unclassified	3,398	778 (22.9%)	717 (92.2%)

† For statistical purposes classification is structured such that each cluster is only represented once (in the order provided), i.e. secreted clusters are those not already classified as effectors, etc.
‡ Percentage out of total pangenome
§ Percentage out of total core

enrichment of Pfam, Iprscan, MEROPs, CAZy, and smCOGs domains, as well as, gene ontology (GO) terms (See Chapter 4 Material and Methods, pg. 113).

Genome alignment, SNP calling, and recombination

Procedures followed Stukenbrock and Dutheil (2018a), for creation of a fine-scale recombination map of fungal organisms and identification of recombination hotspots. A brief description will be provided below, for a more detailed methodology and explanation of algorithms refer to Stukenbrock and Dutheil (2018a), Auton *et al.* (2014), and Wall and Stevison (2016).

LastZ and MultiZ from the TBA package (Blanchette *et al.* 2004) was used to create the population genome alignment projected against the reference genome, *C. purpurea* strain 20.1 (Schardl *et al.* 2013). Alignments in MAF format were filtered using MafFilter v.1.3.1 (Dutheil *et al.* 2014) following Stukenbrock and Dutheil (2018a). Final alignments were merged according to the reference genome and subsequently divided into nonoverlapping windows of 100 kb. MafFilter was additionally used to compute genome-wide estimates of nucleotide diversity (Watterson's θ) and Tajima's D in 10 kb windows. Single nucleotide polymorphisms (SNPs) were called by MafFilter from the final alignment. Principal Component Analysis (PCA) and a Maximum-Likelihood phylogeny were conducted with fully resolved biallelic SNPs (Table 4.3) using the R package SNPRelate v1.18.1 (Zheng *et al.* 2012) and RAxML v8.2.12 (Stamatakis 2014) using GTRGAMA and 1000 bootstrap replicates, respectively.

The following process was automated through the creation of a custom python script (https://github.com/PlantDr430/CSU_scripts/blob/master/Fungal_recombination.py). LDhat (Auton and McVean 2007) was used to estimate population recombination rates (ρ) from the filtered alignment using only fully resolved biallelic positions. A likelihood table was created for

Table 4.3: Summary statistics of whole-genome alignment filtering and SNP calls for *Claviceps purpurea*.

	C. purpurea strain 20.1	
Number of scaffolds	191	
Size of reference genome (bp)	32,091,443	
Number of exonic sites in reference genome (bp)	12,774,951 (39.8%)	
Number of haplotypes	24	
Summary Genome alignment:	Total Alignment Length (bp)	Number of alignment blocks
MultiZ alignment	27,523,755	16,330
Keep blocks with all strains	27,517,978	15,861
MAFFT in 10 kb windows	27,378,024	15,870
Filter 1	26,198,304	57,891
Filter 2	24,959,120	97,532
Merged per contigs (N's filled in)	31,389,412	154
Total number of SNPs	1,152,999	
Total number of analyzed SNPs (biallelic, no unresolved state) and percent of total SNPs	1,076,901 (93.4%)	
Total number of SNPs in exons and percent of total	370,045 (32.1%)	
Total number of analyzed SNPs in exons (biallelic, no unresolved state) and percent of total analyzed SNPs in exons	358,258 (96.8%)	
Diversity in 10 kb windows:	Median	
Watterson's Θ	0.01196	
Tajima's D	-0.82522	

the θ value 0.01, corresponding to the genome-wide Watterson's θ of *C. purpurea* (Table 4.3; Julien Dutheil *personal communication*), and LDhat was run with 10,000,000 iterations, sampled every 5000 iterations, with a burn-in of 100,000. The parameter ρ relates to the actual recombination rate in haploid organism through the equation $\rho = 2N_e \times r$, *where* N_e is the effective population size and *r* is the per site rate of recombination. However, without knowledge of N_e we cannot confidently infer *r* and thus sought to avoid the bias of incorrect assumptions. Therefore, we reported the population recombination rate (ρ).

Resulting recombination maps were filtered to remove pairs of SNPs for which the confidence interval of the recombination estimate was higher than two times the mean (Stukenbrock and Dutheil 2018a). Average recombination rates were calculated, in regions, by weighing the average recombination estimate between every pair of SNPs by the physical distance between the SNPs. Using the reference annotation file (Schardl *et al.* 2013), we calculated the average recombination rates for features in each gene: 1) exons, 2) introns, 3) 500 bp upstream, and 4) 500 bp downstream with a minimum of three filtered SNPs. Flanking upstream and downstream regions correspond to the 5′ and 3′ regions for forward stranded genes and the 3′ and 5′ regions for reverse stranded genes. We also calculated the average recombination rate for each intergenic region between the upstream and downstream regions of each gene. Introns were added to the GFF3 file using the GenomeTools package (Gremme *et al.* 2013). The original recombination maps produced from LDhat (Julien Dutheil *personal communication*) were converted from bp to kb format for use in LDhot (Auton *et al.* 2014) to detect recombination hotspots 1000 simulations and --windlist 10 to create 20 kb background windows (Wall and Stevison 2016). Only hotspots with a value of ρ between 5 and 100 and

width < 20 kb were selected for further analysis (Auton *et al.* 2014; Wall and Stevison 2016; Stukenbrock and Dutheil 2018a).

Statistical and enrichment analyses

Statistics and figures were generated using Python3 modules SciPy v1.3.1, statsmodel v0.11.0, Matplotlib v3.1.1, and seaborn v0.10.0. All multi-test corrections were performed with Benjamini-Hochberg false discovery rate procedure. Enrichment analyses were tested using Fischer's Exact test with a cutoff $\alpha = 0.05$. Uncorrected p-values were corrected using Benjamini-Hochberg and Bonferroni multi-test correction with a false discovery rate (FDR) cutoff of $\alpha = 0.05$. Corresponding p-values from correction tests were averaged together to get a final p-value. Enrichment was performed on protein domain names and GO terms. Orthogroups were only associated with a domain or GO term if ≥ 50% of the strains present in the gene cluster had one gene with the term. This process was automated through creation of a custom python script (https://github.com/PlantDr430/CSU_scripts/blob/master/Domain_enrichment.py).

RESULTS

Pangenome analysis

We constructed a pangenome of *Claviceps purpurea* from 24 isolates representing a collection from three continents and six countries (Table 4.1). Taking advantage of plentiful isolates available from Canada, we sampled more heavily from different provinces and on different host plants. The principal component and phylogenetic analysis revealed substantial genetic variation among the samples, such as LM470 (Canada) and Clav04 (USA) grouping closer to isolates from Europe and the isolate from New Zealand (Appendix 3 Fig. A3.1). In addition, across Canada and USA, isolates from similar regions rarely clustered together and were often intermixed (Appendix 3 Fig.A3.1 B). These results agree with the results from a

multi-locus genotyping of extended samples from Canada and midwestern USA (Liu *et al. unpublished data*). Previous reports (Chapter 3) showed that *C. purpurea* isolates had similar genome size (30.5 Mb – 32.1 Mb), genomic GC content (51.6% - 51.8%), TE content (8.42% - 10.87%), gene content (8,394 – 8,824), and BUSCO completeness score (95.5% - 98.0%) (Table 4.1). The pangenome consisted of 205,354 genes which were assigned to 10,540 orthogroups. We observed 6,558 (62.22%) orthogroups shared between all 24 isolates (core genome), of which 6,244 (59.2%) were single-copy gene clusters, while the remaining core orthogroups, 314 (3%), contained paralogs (2 – 8 paralogs per cluster). The accessory genome consisted of 3,982 (37.78%) orthogroups with 2,851 (27.05%) shared by at least two isolates (but not all) and 1,131 (10.73%) were lineage-specific (singletons) found in only one isolate (Fig. 4.1; Appendix 3 Table A3.1). Within the accessory genome (including lineage-specific orthogroups) we observed 592 (5.6%) orthogroups contained paralogs, with some isolates containing > 20 genes per cluster (Fig. 4.1 C; Appendix 3 Table A3.1).

We utilized multiple gene functional categories to get a deeper understanding of how gene of different functions were structured within the pangenome. As a proportion of orthogroups within each pangenome category (core, accessory, and singleton) we found that the core genome was significantly enriched in orthogroups that contained genes with conserved protein domains (conserved) (5,471; 84%), transmembrane domains (transmembrane) (1,038; 16%), peptidase and protease domains (MEROPs) (211, 3.2%), and orthogroups of carbohydrate-active enzymes (CAZys) (212, 3.2%) ($P < 0.01$, Fisher's exact test, Fig. 4.2 A, E-G). Effector proteins play major roles in plant-microbe interactions, often conveying infection potential of the pathogen. A total of 257 predicted effector orthogroups were identified; 100 (38.9%) were core, 143 (55.6%) were accessory, and 14 (5.4%) were singletons. Predicted

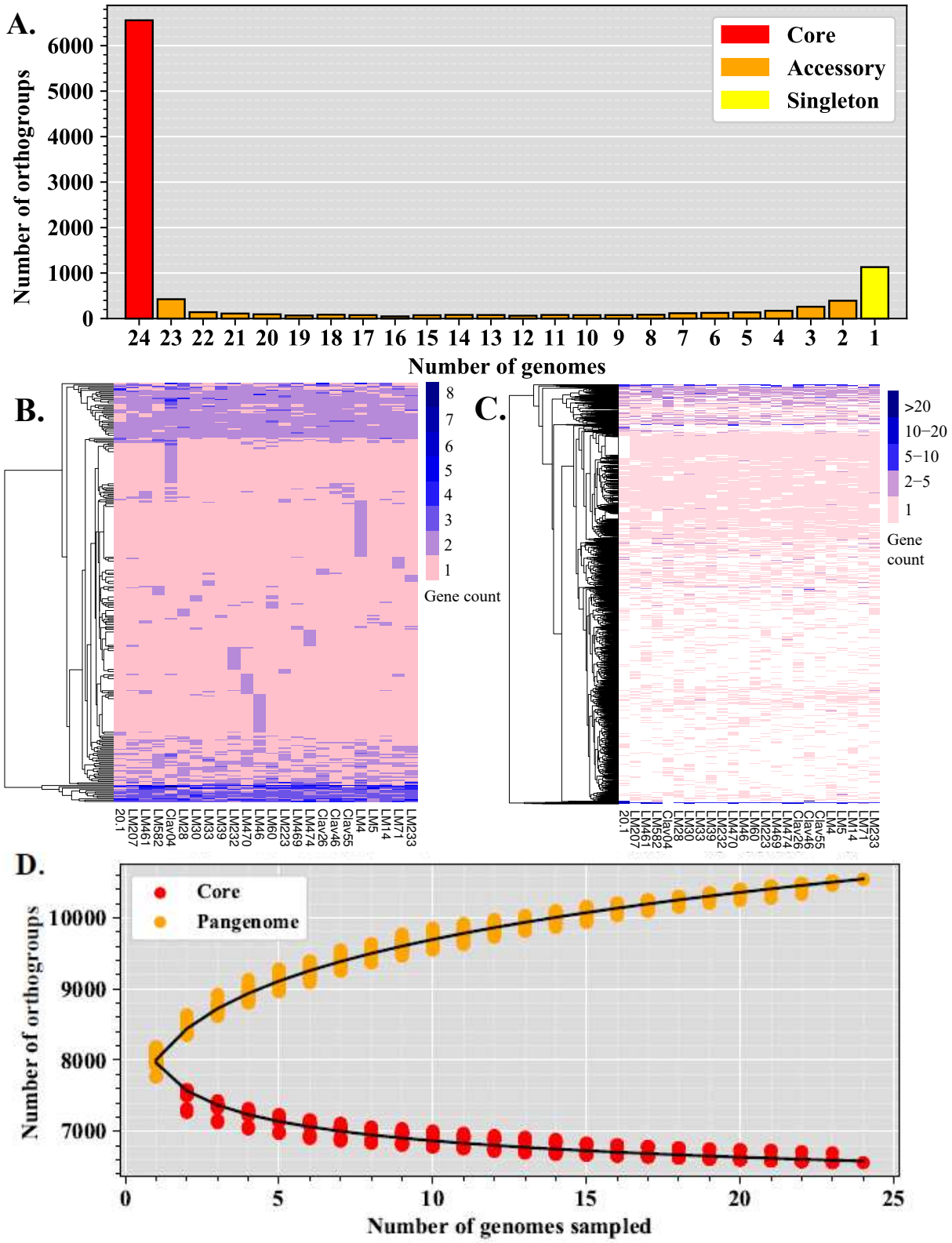


Figure 4.1: Analysis of predicted protein function across the pangenome. Graphs indicate the proportion of orthogroups within each pangenome category of categorized protein function determined if ≥ 50% of the isolates present in the orthogroups had at least one gene classified as such. **A**) Containing conserved protein domains, **B**) genes found in secondary (2°) metabolite clusters, **C**) possessing predicted secreted signals, **D**) predicted to be effectors, **E**) containing transmembrane domains, **F**) containing MEROPs domains for proteases and peptidases, **G**) contain CAZY enzymes, **H**) all unclassified orthogroups not falling into a previous category. Different letters (within each classification) represent significant differences determined by multi-test corrected Fisher exact test ($P < 0.01$).

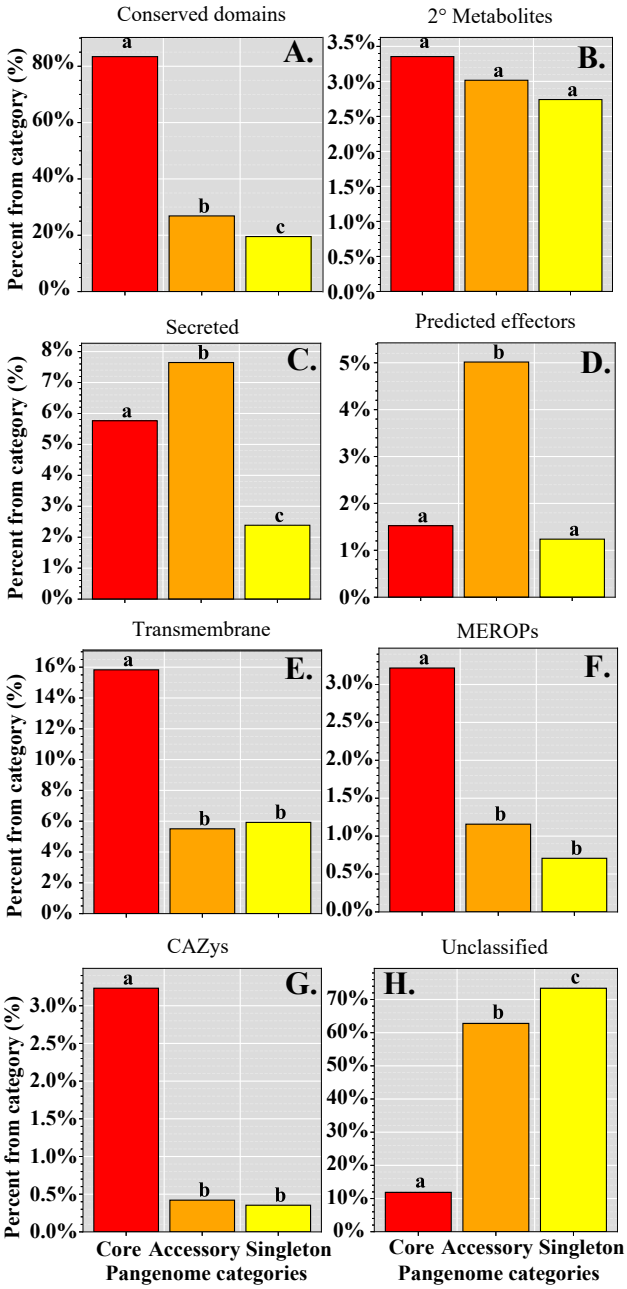


Figure 4.2: Analysis of predicted protein function across the pangenome. Graphs indicate the proportion of orthogroups within each pangenome category of categorized protein function determined if ≥ 50% of the isolates present in the orthogroups had at least one gene classified as such. **A**) Containing conserved protein domains, **B**) genes found in secondary (2°) metabolite clusters, **C**) possessing predicted secreted signals, **D**) predicted to be effectors, **E**) containing transmembrane domains, **F**) containing MEROPs domains for proteases and peptidases, **G**) contain CAZY enzymes, **H**) all unclassified orthogroups not falling into a previous category. Different letters (within each classification) represent significant differences determined by multi-test corrected Fisher exact test ($P < 0.01$).

effectors and orthogroups coding for secreted proteins, which also contribute to host-pathogen interactions, were significantly enriched in the accessory genome (143, 5%; 218, 7.6%; respectively) ($P < 0.01$, Fisher's exact test, Fig. 4.2 D, C). Although, the accessory and singleton genomes were largely composed of unclassified orthogroups (1791; 62.8%; 830, 73.4%; respectively) ($P < 0.01$, Fisher's exact test, Fig. 4.2 H). Lastly, we observed that orthogroups containing secondary (2°) metabolite genes were similarly represented within all pangenome categories ($P > 0.05$, Fisher's exact test, Fig. 4.2 B).

As expected, core orthogroups were found to be significantly enriched in general housekeeping and basic cellular functions and development such as protein and ATP binding, nucleus and membrane cellular components, and transmembrane transport, metabolic, and oxidation-reduction processes (Appendix 3 Table A3.2). Protein domains in core orthogroups were significantly enriched for several WD40-repeat domains, P-loop nucleoside triphosphate hydrolase (IPR027417), armadillo-type fold (IPR016024), and a major facilitator (PF07690) (Appendix 3 Table A3.2). When narrowing the focus to orthogroups with paralogs, core paralogous orthogroups were enriched in cytochrome P450 domains, and domains associated with trehalose activity (Appendix 3 Table A3.3). In contrast, the accessory genome was only found to be enriched in a fungal acid metalloendopeptidase domain (MER0001399) and the singleton genome had enrichment for a Tc5 transposase DNA-binding domain (PF03221) (Appendix 3 Table A3.2). Accessory paralogs were found to be enriched in several protein kinases, Myb-like domains, phosphotransferases, as well as DNA integration and a MULE transposase domain (Appendix 3 Table A3.3). Overall, our results reveal a large accessory pangenome enriched with genes associated with host-pathogen interactions and an abundance of

orthogroups containing paralogs (8.6%), indicating the presence of prolific gene duplication occurring within the species.

Positive selection landscape

To further understand the evolution of genes within the pangenome we investigated the positive selection landscape on protein coding genes using 6,244 single-copy orthologs to compute the ratio of non-synonymous substitutions to synonymous substitutions (dN/dS). Ratios of dN/dS (omega, ω) can provide information of evolutionary forces shaping an organism as genes with $\omega > 1$ may indicate positive or diversifying selection, $\omega = 1$ may indicate neutral evolution, and $\omega < 1$ may indicate negative or purifying selection (Jeffares *et al.* 2015).

Overall, we saw low dN and dS values across all functional categories (Appendix 3 Fig. A3.3), corresponding to low ω ratios (Fig. 4.3). This suggests a general trend of purifying selection within *C. purpurea*, although we did identify orthogroups with ω values > 1 (63, 1%), of which 25 (40%) were unclassified (Fig. 4.3, Appendix 3 Table A3.4). Notable BLASTp results showed that two conserved genes were related to transcription factors (OG0001193, $\omega = 1.13$, related to subunits Tfc3; OG0004135, $\omega = 1.21$, related to Cys6) and two were related to DNA repair (OG0001034, $\omega = 1.05$, related to mismatch repair PMS1; OG0004027, $\omega = 1.13$, related to XLF (XRCC4-like factor)) (Appendix 3 Table A3.5). The gene with the highest ω was a transmembrane gene related to a bacteriophage N adsorption protein (OG0001093, $\omega = 9.79$) (Appendix 3 Table A3.5). Overall, core unclassified genes showed the highest ω values but were not significantly different than predicted effector genes ($P >> 0.05$, multi-test corrected Mann-Whitney U Test, Fig. 4.3). In contrast, transmembrane, MEROPs, CAZys, and proteins with conserved domains showed the lowest ω values, indicating that these genes are more often pressured towards purifying selection.

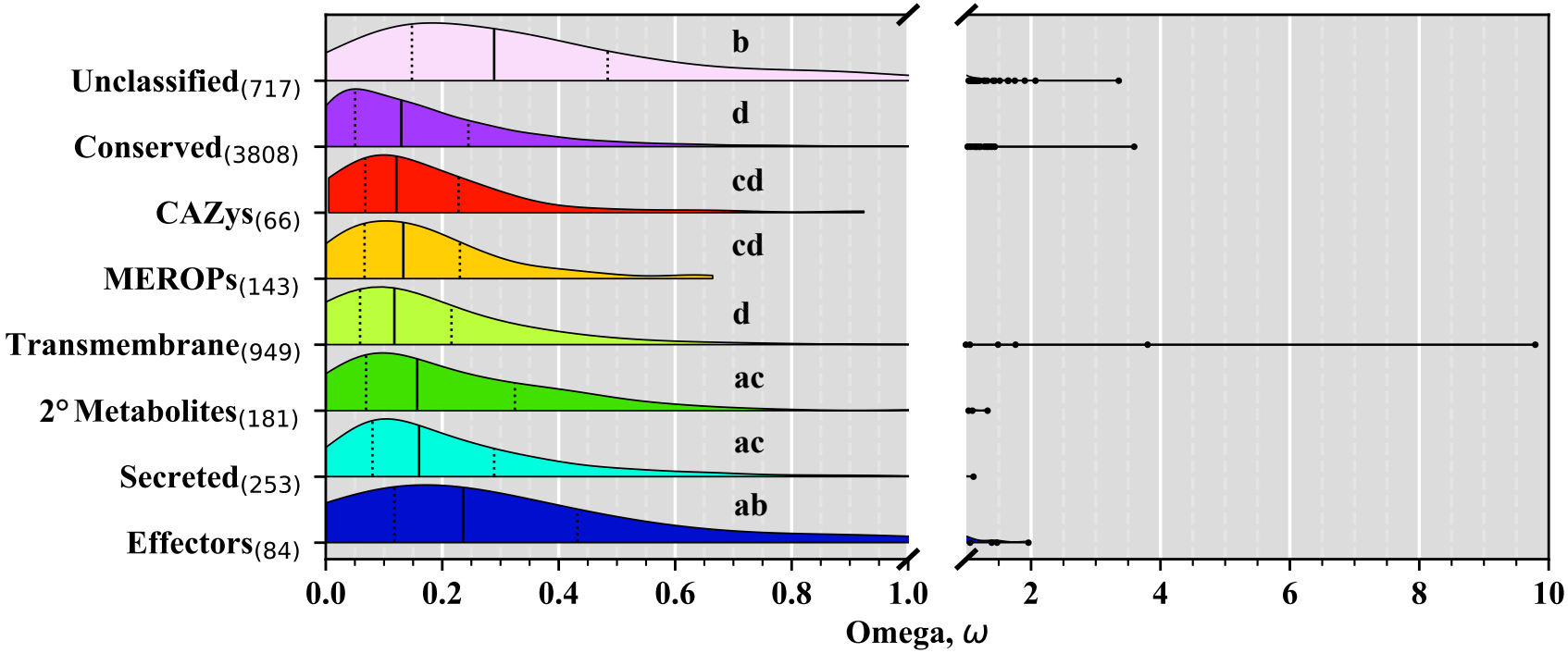


Figure 4.3: Distribution violin plots of omega (ω, dN/dS) ratios for core single-copy orthogroups protein functional categories. Solid vertical lines within each plot represent the median, while dotted lines represent the 25th and 75th quartile, respectively. Different letters represent significant differences determined by Kruskal-Wallis with *post hoc* multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$).

While ω values, calculated across the entire gene, can provide useful insight on the selective landscape of genes, positive selection and evolution occur at the codon triplet level and can occur in genes where ω, across the entire gene, is < 1 (Goldman and Yang 1994). For this reason, we utilized the CodeML algorithm (Yang 2007) was used to more accurately and confidently identify genes with positive selection signatures. Our results revealed a total of 986 positively selected genes (15.8%) that passed our stringent filtering (Fig. 4.4 A). The majority were genes encoding conserved domains (557, 56.5%) followed by unclassified genes (192, 19.5%). While conserved genes made up the largest portion of genes under putative positive selection, unclassified genes showed the highest proportion of genes with positive selection signatures (26.8%) followed by secondary (2°) metabolite genes (21.5%) (Fig. 4.4 B). We observed an enrichment of positively selected secondary metabolite genes that contained domains for polyketide synthase domains, and several phosphopantetheine domains, as well as, metabolic and catalytic GO terms (Table 4.2, Fisher's exact test, Fig. 4.4 B, $P \leq 0.05$, Appendix 3 Table A3.6). In addition, five genes in two known secondary metabolite clusters showed evidence of positive selective signatures: three genes (*easE* ω = 0.51, *lpsB* ω = 0.34, and *lps*C ω = 0.33) in the well-known ergoline biosynthetic cluster (ergot alkaloids) (Schardl *et al.* 2013) and two genes (*tcpC* ω = 0.37 and *tcp*P ω = 0.37) in the epipolythiodiketopiperazine biosynthetic cluster (Dopstadt *et al.* 2016). Additionally, one of the three genes responsible for the biosynthesis of fungal cytokinins, a pisatin demethylase cytochrome P450 (Hinsch *et al.* 2015, 2016), had signatures of positive selection (OG0000984, ω = 0.19, CCE30328.1, Appendix 3 Table A3.6). Transmembrane genes saw enrichment of three multicopper oxidase domains

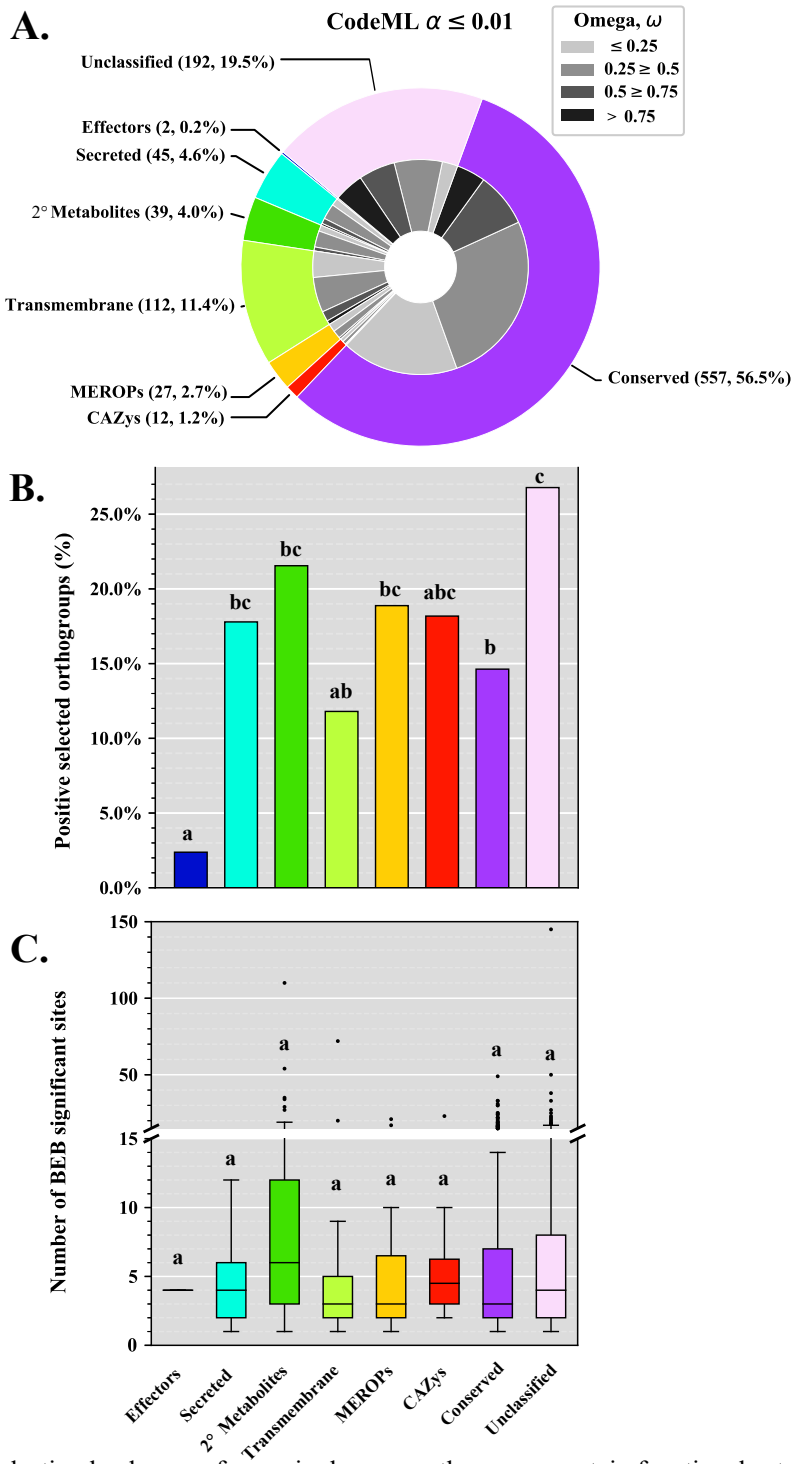


Figure 4.4: Positive selection landscape of core single-copy orthogroups protein functional categories as predicted by PAML with the CodeML algorithm. Genes with positive selection signatures were selected after a stringent filtering around an $\alpha \leq 0.01$. **A**) The total number of orthogroups in functional categories with signatures of positive selection. **B**) The proportion of orthogroups in each functional category based on the number of orthogroups examined in each category (outer circle). Omega (ω, dN/dS) ratios of orthogroups within each functional category (inner circle). **C**) The number of codons with selection signatures in the M8 model of CodeML, as determined by the Bayes Empirical Bayes (BEB) algorithm with an $\alpha \leq 0.01$. Different letters represent significant differences determined by Kruskal-Wallis with *post hoc* multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$). See Appendix 3 Figure A3.5 for results from a less stringent filtering of $\alpha \leq 0.05$.

($P < 0.05$, Fisher's exact test, Appendix 3 Table A3.6). Of which two transmembrane orthogroups, that contained genes with these domains, also encoded for the laccase CAZy enzymes AA1_1, AA1_2, and AA1_3 (OG0005604, ω = 0.38 and OG0002895, ω = 0.22) (Appendix 3 Table A3.1).

There was limited positive selection among predicted effector genes (Fig. 4.4B). Only two predicted effector genes (Fig. 4.4 A), corresponding to a proportion of 2.4% of the 84 predicted effector genes examined (Table 4.2, Fig. 4.4B), had evidence of positive selection. Suggesting that effectors might not be under pressure to evolve to overcome host defenses. These two predicted effector genes (OG0003219, ω = 0.76, EffectorP mean score = 0.90 ± 0.028; OG0006565, ω = 1.96, EffectorP mean score = 0.78 ± 0.051) did not have any associated protein domains (Appendix 3 Table A3.1, Appendix 3 Table A3.4). We also did not observe any evidence of positive selection in the 10 known virulence factors of *C. purpurea* (Mey *et al.* 2001, 2002; Oeser *et al.* 2002; Scheffer *et al.* 2005a, 2005b; Giesbert *et al.* 2008; Rolke *et al.* 2008; Bormann and Tudzynksi 2009) (Appendix 3 Table A3.1, Appendix 3 Table A3.4). In addition, we found no domain enrichment in positively selected secreted genes and CAZys. MEROPs genes only showed enrichment in an alpha/beta hydrolase fold domain (Appendix 3 Table A3.6).

Overall, our results reveal a significant lack of positive selection on predicted effector genes, but a larger proportion of core unclassified and secondary metabolite genes with signatures of positive selection (Fig. 4.4). It should be noted that secondary metabolite genes also showed the highest number of codons per gene with signatures of positive selection, as determined by the Bayes Empirical Bayes (BEB) algorithm integrated into PAML, however, we did not observe significant differences between gene classifications (Fig. 4.4 C).

Recombination landscape

Recombination is also an important potential driver of genome evolution and plays a central role in the adaptability of parasitic organisms to overcome host defenses (Morran *et al.* 2011). Our genome-alignments contained 154 of the original 191 scaffolds of the reference strain (20.1) (Table 4.3). These 37 missing scaffolds totaled 222,918 bp (average = 6,192 ± 5,676 bp) and corresponded to 59 genes. Thirty-one of the scaffolds contained genes that were only part of the accessory genome of which six scaffolds contained two or more genes (Appendix 3 Table A3.7), suggesting that these scaffolds represent blocks of genetic material that could be lost or gained from isolate to isolate. Most of the genes found on these scaffolds encoded conversed domains associated with either reverse transcriptase, integrases, or helicases (Appendix 3 Table A3.7), which suggest unplaced repetitive content. Although, one scaffold (scaffold 185) did possess a gene encoding a conserved domain for a centromere binding protein (Appendix 3 Table A3.7). Together these observations could indicate the potential for dispensable chromosomes, as dispensable and mini-chromosomes often contain higher repetitive content (Peng *et al.* 2019).

From our shared alignments we recovered 1,076,901 biallelic SNPs corresponding to a median nucleotide diversity (Watterson's θ) of 0.01196 and a Tajima's D of -0.82522 calculated from 10 kb non-overlapping windows (Table 4.3). The resulting SNPs were used to infer the population recombination rate (ρ) from the linkage disequilibrium between SNPs based on *a priori* specified population mutation rate θ, which was set to 0.01 based on our nucleotide diversity (Watterson's θ) (Table 4.3) (Stukenbrock and Dutheil 2018a). The *C. purpurea* genome recombination landscape was highly variable as some scaffolds showed highly heterogenous landscapes, other scaffolds showed intermixed large peaks of recombination, while others still

had more constantly sized peaks across the regions (Fig. 4.5, Appendix 3 Fig. A3.6). Overall, the mean genomic population recombination rate in *C. purpurea* was $\rho = 0.044$. We also examined recombination in specific sequence features and gene type through comparison of mean population recombination rates in exons, introns, 500-bp upstream and downstream of the coding DNA sequence, and intergenic regions based on the annotation of the reference genome (strain 20.1). The distribution of population recombination rates were comparable across different gene features and gene functional categories, although, some significant differences were observed (Fig. 4.6). In general, we found upstream regions to have the lowest recombination rates, while downstream regions have the highest recombination rates (Fig. 4.6). The decreased recombination in upstream regions might be the result of mechanisms trying to conserve promotor regions. This trend was observed across different functional gene categories, except in predicted effector genes where exons showed the highest recombination rates and downstream regions with the lowest, although these were not significantly different (Fig. 4.6 B). Across functional categories, secreted genes and transmembrane genes showed the highest recombination rates within each gene feature but were not always significantly different (Fig. 4.6 C).

Due to the observation of paralogs (Fig. 4.1) and evidence of tandem gene duplication in *C. purpurea* (Chapter 3) we investigated the extent recombination might have influenced these events. We found that duplicated genes had lower population recombination rates than all other genes within the genome (Fig. 4.6 D), suggesting that other factors are influencing gene duplication. Due to the absence of RIP (Chapter 3), transposable elements (TE) are likely a contributing factor. To investigate the association of duplicated genes with TEs we calculated the

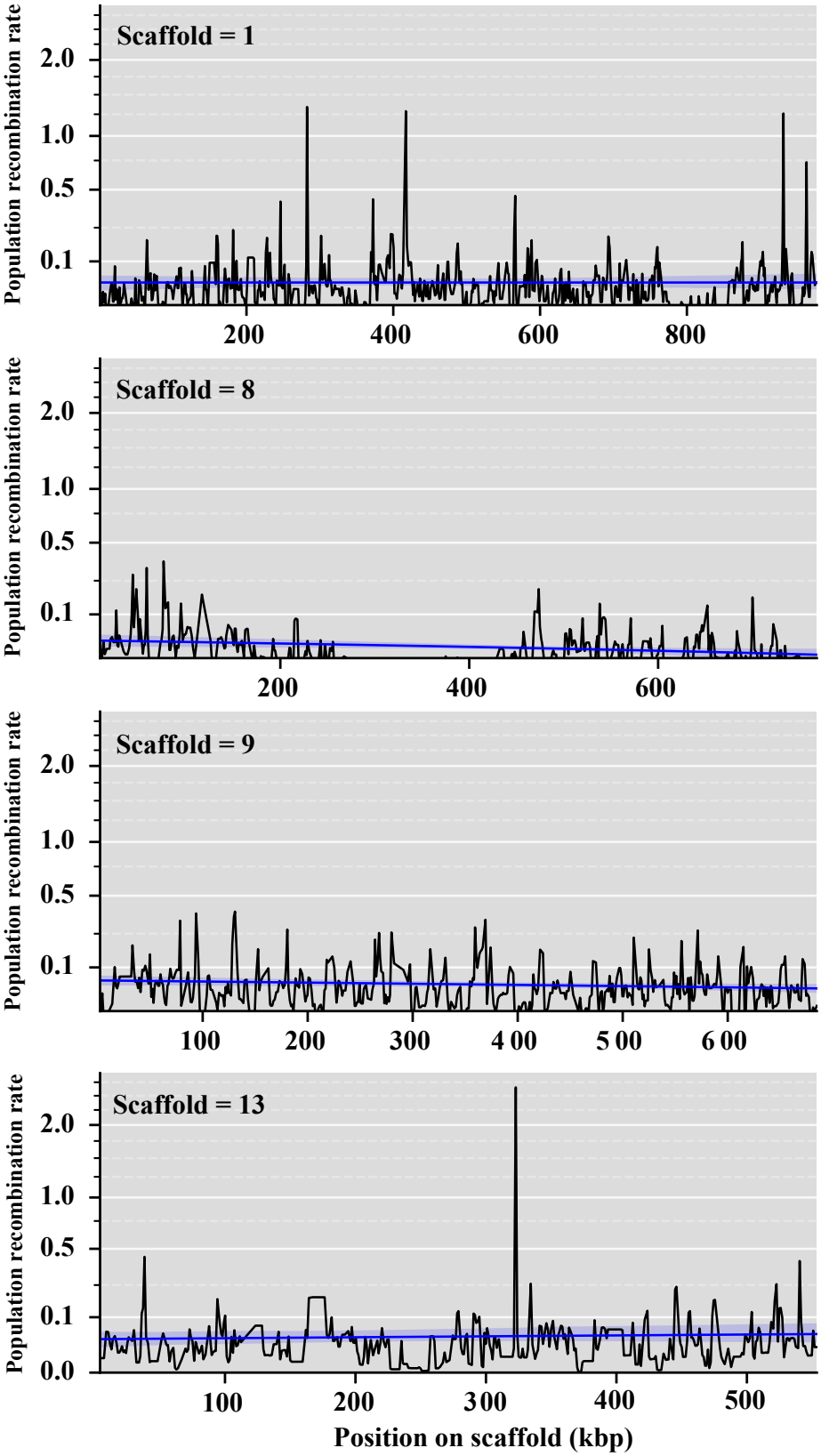


Figure 4.5: Estimates of population recombination rates (ρ), in non-overlapping 1 kb windows, across four representative scaffolds displaying the different variation observed across the *C. purpurea* genome. Smoothing curves were calculated from population recombination rates in 10 kb windows. See Appendix 3 Fig. A3.6 for remaining scaffolds.

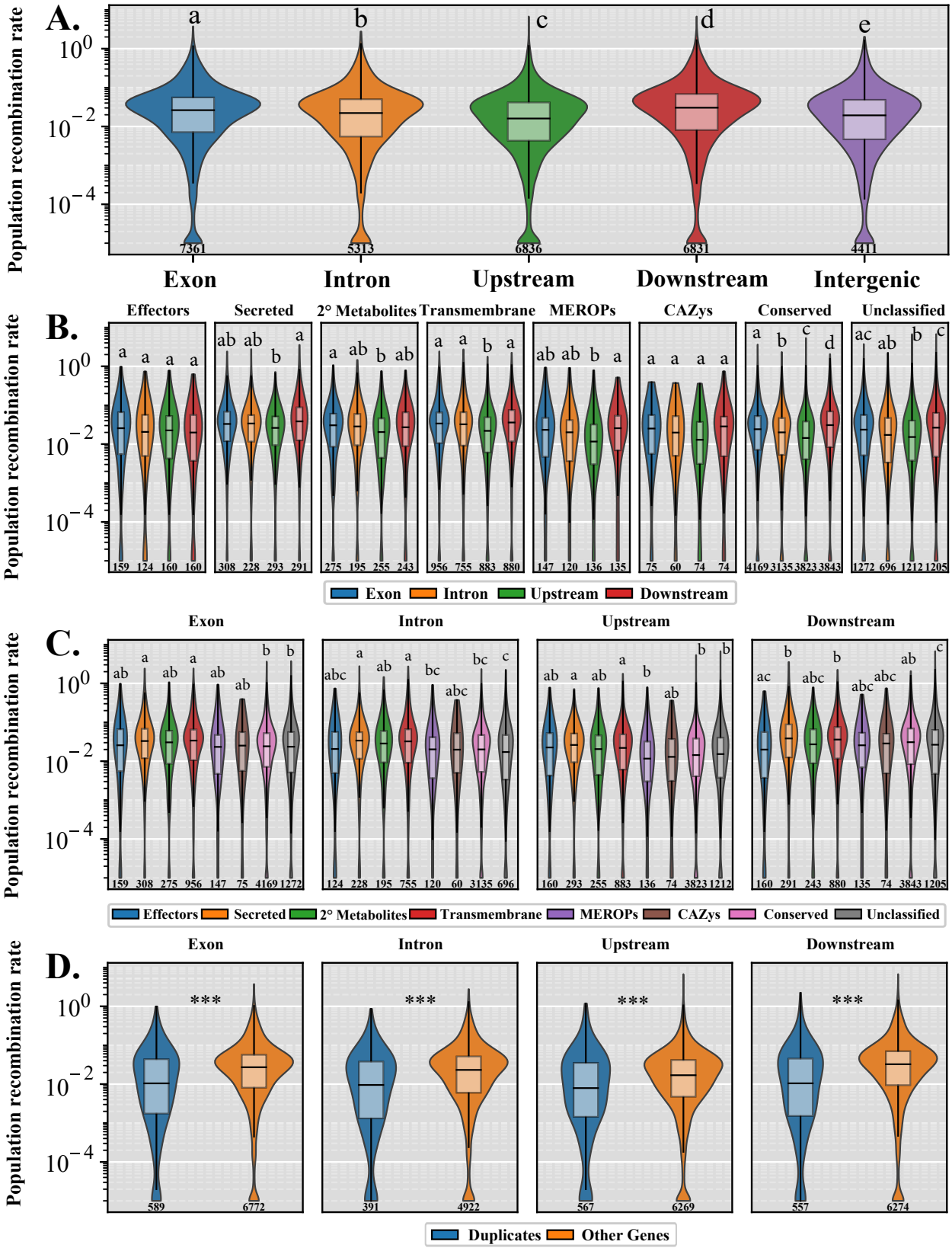


Figure 4.6: Fine-scale recombination patterns across the *C. purpurea* genome. Plots indicate the distribution of estimated population recombination rates (ρ) between different gene features (exons, introns, 500bp upstream and downstream) (**A)** and genes of different functional categories and classification (**B-D**). Different letters represent significant differences determined by Kruskal-Wallis with *post hoc* multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$) between data within each plotting window, *** $P < 0.0001$. Sample sizes are embedded below each plot.

average distance of genes to long terminal repeat (LTR) retrotransposons and the average number of flanking LTRs. Results showed duplicated genes were significantly closer to LTRs and had significantly more flanking LTRs than predicted effector and other genes ($P < 0.0001$, multi-test corrected Mann-Whitney U Test, Appendix 3 Fig. A3.7).

As we observed distinct peaks of recombination (Fig. 4.5, Appendix 3 Fig. A3.6), we further utilized LDhot to call statistically significant recombination hotspots by analysis of the intensity of recombination rates in 3 kb (1 kb increments) windows compared to background recombination rates in 20 kb windows (Auton *et al.* 2014; Wall and Stevison 2016; Stukenbrock and Dutheil 2018a). After implementing a cut-off of $\rho \geq 5$ and length of 20 kb (Wall and Stevison 2016) we retained only five recombination hotspots, ranging from 11 kb to 18.5 kb in length (Fig. 4.7). We observed a recombination hotspot located between the *lpsA1* and *lpsA2* genes of the ergoline biosynthetic cluster, suggesting that this gene duplication event was likely the result of recombination (Fig. 4.7 D). Association of gene functional category and TEs within hotspots varied between region. Some hotspots showed a greater association with duplicated genes and TEs (Fig. 4.7 B-D), while others showed a lower association (Fig. 4.7 A, E). In general, genes with conserved protein domains showed the highest presence within hotspots (Appendix 3 Fig. A3.8). It should be noted that some unclassified genes and genes with conserved protein domains associated with hotspots were also found to be overlapping regions identified as repeats (Fig. 4.7 A-C, E). Protein domains found within these genes were associated with ankyrin (IPR002110) and tetratricopeptide (IPR013026) repeats. Only 5 of the 846 duplicated genes (Chapter 3) found throughout the reference genome were located within predicted recombination hotspots (Fig. 4.7, Appendix 3 Fig. A3.8). While Chapter 3 showed that gene cluster expansion was prevalent among predicted effectors, we only found one

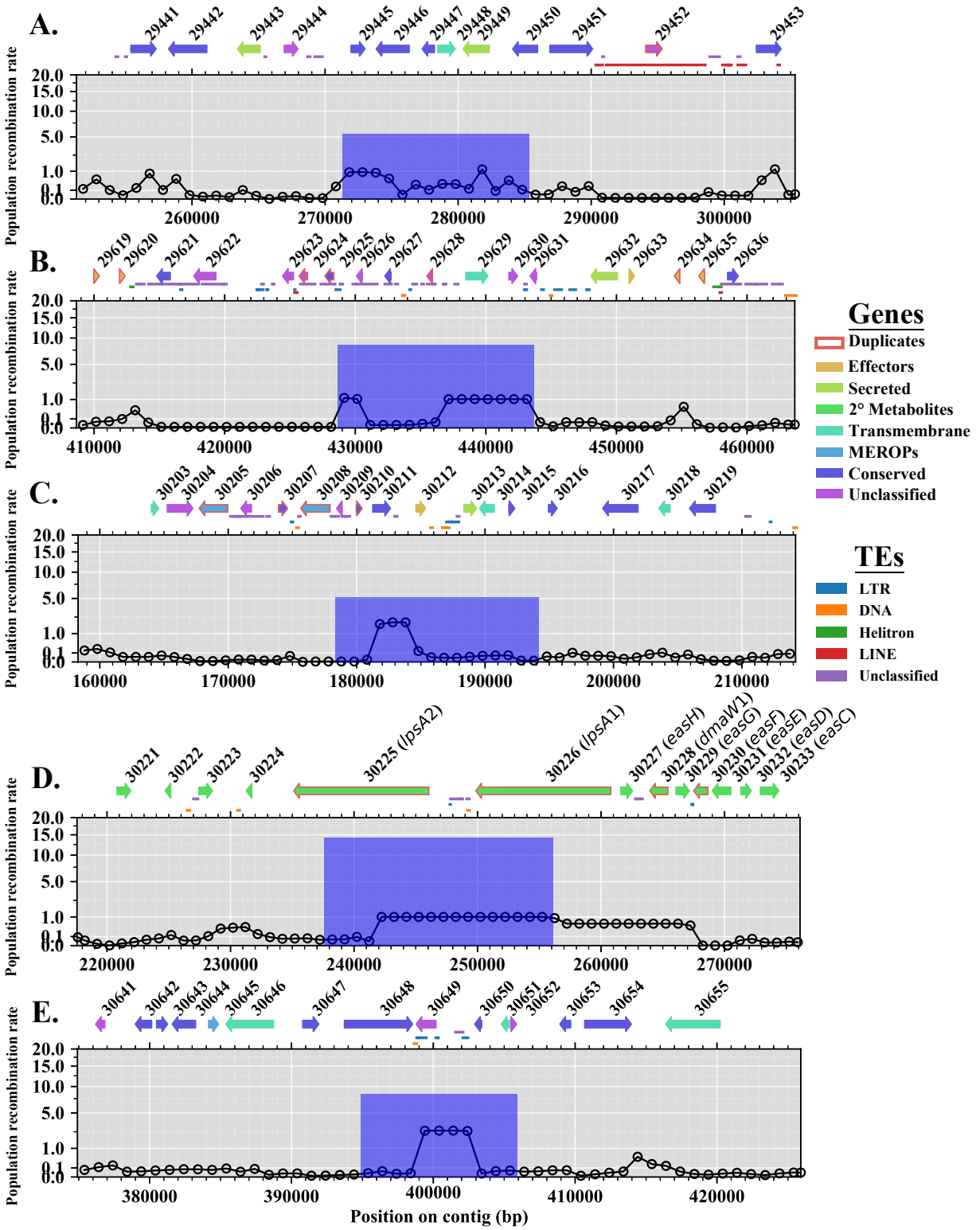


Figure 4.7: Distribution of recombination hotspots predicted in *C. purpurea* by LDhot with associated genes and transposable elements (TEs). Lines indicate background population recombination rates (ρ) estimated in non-overlapping 1 kb windows. Blue bars represent the position, intensity, and width of the predicted hotspots. Genes within the hotspot window and surrounding (± 20 kb) region are depicted by arrows with protein ID's of the reference (strain 20.1) from NCBI. Genes identified as duplicated (≥ 80% identity) from Chapter 3 are outlined in red. TEs are depicted by lines between genes and the corresponding hotspot graph. Colors of arrows and lines correspond to the legend on the right.

non-duplicated predicted effector (CCE30212.1) located within a recombination hotspot (Fig. 4.7 C). Together these results suggest that while recombination may result in important gene duplication; it is not the primary driver of gene duplication within *C. purpurea*.

DISCUSSION

Our establishment of a *Claviceps purpurea* pangenome from 24 isolates, as well as, the detection of core genes with signatures of positive selection and analysis of the recombination landscape have provided knowledge into how high recombination rates, gene duplication, and selection of secondary metabolite genes are driving the genomic evolution and adaptation of the species.

The pangenome of *C. purpurea* reveals a large accessory genome with 37.78% accessory orthogroups (27.05% accessory + 10.73% singleton) in comparison to four model fungal pangenomes (*Saccharomyces cerevisiae*, *Candida albicans*, *Cryptococcus neoformans*, and *Aspergillus fumigatus*), which found around 9 – 19% of their genes in the accessory genome (McCarthy and Fitzpatrick 2019). Our results are more comparable to the pangenome of the fungal pathogen *Zymoseptoria tritici* which had an accessory genome comprised of 40% (30% accessory + 10% singleton) of genes (Badet *et al.* 2020). Similar to *C. purpurea*, *Z. tritici* is a globally distributed biotrophic fungal pathogen of grasses, notably wheat, suggesting that fungal species with similar life strategies, hosts, and ecological environments could possess comparable pangenome structures as they are under similar evolutionary pressures. Similar factors of lifestyle, effective population size, and habitat have been reported to influence pangenome sizes in bacteria (McInerney *et al.* 2017). In fact, *C. purpurea* and *Z. tritici* both experienced enrichment of predicted effector orthogroups in the accessory genome and enrichment of carbohydrate-active enzymes (CAZys) orthogroups in the core genome (Fig. 4.2) (Badet *et al.*

2020), conveying a comparable similarity between gene functions within pangenome structure regarding the pathogenic lifestyle of these organisms. In addition, Badet *et al.* (2020) suggested that the large accessory genome of *Z. tritici* is maintained due to TE activity and a large effective population size as a result of observations of high SNP density, rapid decay in linkage disequilibrium, and high recombination rates (Croll *et al.* 2015; Hartmann *et al.* 2017; Stukenbrock and Dutheil 2018a). The same mechanisms could also explain the large accessory genome observed in *C. purpurea*.

We observed an abundance of orthogroups containing paralogs (8.6%), potentially due to a lack of RIP (Chapter 3). This presence of gene duplication and association with LTR retrotransposons (Appendix 3 Fig. A3.7) could be contributing to the large size of the accessory genome, potentially through pseudogenization or neofunctionalization. In fact, unclassified genes had the highest ω (dN/dS) ratios (Fig. 4.3) and the highest proportion of genes with signatures of positive selection (Fig. 4.4). While this analysis was only conducted on single-copy core genes, it suggests that some of the unclassified accessory genes (Fig. 4.2 H) may be undergoing similar evolutionary trends,. In addition, the abundance of duplication in accessory unclassified genes (Appendix 2 Fig. A2.12) and their small sizes (Appendix 3 Fig. A3.2)can further suggest the presence of pseudogenization and/or neofunctionalization. Badet *et al.* (2020) suggested that TEs were likely contributing to *Z. tritici* accessory genome due to their correlations of TE content with genome size and observations of transcribed TEs. We observed a similar correlation of TE content with genome size ($P = 0.004$, Adj. $R^2 = 0.28$), however, our genome sizes and TE content (30.5 Mb – 32.1 Mb, 8.42% - 10.87%, respectively) were not as variable as in *Z. tritici*, which also had a twofold higher TE content (Badet *et al.* 2020). This suggests that TEs play a more important role in *Z. tritici* genome expansion, however, only 0.2% of the orthogroups in *Z.*

tritici contained paralogs suggesting that gene duplication is not as common in *Z. tritici* as it is in *C. purpurea* (8.6% paralogs). The lack of gene duplication in *Z. tritici* is likely due to the presence of RIP (Testa *et al.* 2015b), which should also reduce TE expansion through silencing (Galagan *et al.* 2003, 2004; Urguhart *et al.* 2018). While we lack RNAseq data to observe TE transcription within *C. purpurea*, observations of TEs with 0% divergence in *C. purpurea* (Chapter 3) suggest recent TE activity. The observed lack of recombination associated with duplicated genes (Fig. 4.6 D) and association of duplicated genes with LTR transposons (Appendix 3 Figure A3.7) would suggest that gene duplication in *C. purpurea* is mediated by transposon activity.

Furthermore, we identified 37 missing scaffolds in our population genome alignment with 31 of these containing only genes present in the accessory genome, suggesting the potential for blocks of DNA that could be lost/gained between isolates. Of these accessory scaffolds 15 contained genes encoding conversed domains associated with either reverse transcriptase, integrases, or helicases and one scaffold possessed a gene encoding a conserved domain for a centromere binding protein (Appendix 3 Table A3.7). Together these could indicate the potential for dispensable mini-chromosomes, as dispensable and mini-chromosomes often contain higher repetitive content (Peng *et al.* 2019). However, even the combination of all 37 missing scaffolds (0.22 Mb) would represent the smallest mini-chromosome known in plant pathogens; 3-fold smaller than *Leptosphaeria maculans* (Balesdent *et al.* 2013), 2-fold smaller than *Nectria haematococca* (Mahmoud and Taga 2012), and 7-fold smaller than *Magnaporthe oryzae* (Peng *et al.* 2019). Many of these scaffolds contained repeated N's sequences from scaffolding (Schardl *et al.* 2013) and increased repeat content (Appendix 3 Table A3.7) suggesting that our Illumina based genomes might not have captured the true nature of these scaffolds. Therefore, we did not

process these elements further but believe that these are an important aspects of *C. purpurea* evolution and should be a focal point of future research with the advantage of long-read sequencing to more confidently understand their function. Due to these transcriptase rich unplaced scaffolds, the lack of RIP, association of duplicated genes with LTR transposons, and observation of TEs with 0% divergence (Chapter 3), we believe transposons and/or transcriptases are influencing gene duplication in *C. purpurea*.

Due to the potential for transposon mediated gene duplication, it was remarkable to find relatively low TE content (~8 - 10%) within *C. purpurea*, especially in the absence of RIP. Other genomic mechanism, such as recombination, may help to limit TE expansion and increases in genome size. Tiley and Burleigh (2015) found a strong negative correlation between global recombination rate, genome size, and LTR retrotransposon proportion across 29 plant species, indicating that higher recombination rates actively reduce genome size likely through the removal of LTR elements. A similar function may be affecting LTR content in *C. purpurea*, which would explain the observed differences in LTR content between *Clavicpes* section *Claviceps* (low LTR content, RIP absent) and *Claviceps* sections *Pusillae*, *Paspalorum*, and *Citrinae* (high LTR content, RIP present) (Chapter 3).

On average we observed a twofold higher mean population recombination rate (ρ = 0.044) in *C. purpurea* than *Z. tritici* (ρ = 0.0217) and tenfold higher than *Z. ardabiliae* (ρ = 0.0045) (Stukenbrock and Dutheil 2018a). As ρ is a function of effective population size and recombination rate per site ($\rho = 2N_e \times r$), these increases could be the result of the increment in recombination rate per site (r) and/or differences in the effective population size (N_e). Differences in ρ between the two *Zymoseptoria* species was postulated to be due to increased actual recombination rates as it was found that the nucleotide diversity (Watterson's $\theta = 2 N_e \times \mu$,

where μ is mutation rate) was 1.6 times higher in *Z. tritici* (0.0139) than *Z. ardabiliae* (0.00866). Under an assumption that both *Z. tritici* and *Z. ardabiliae* have comparable mutation rates, N_e of *Z. tritici* would only be 1.6 times higher than *Z. ardabiliae*, therefore, the 5 fold higher ρ would likely be caused by higher recombination rates per site (Stukenbrock and Dutheil 2018a).. Our observed Watterson's θ of 0.012 in *C. purpurea* (Table 4.2) is comparable to *Z. tritici*, suggesting that if mutation rates and effective population sizes are comparable than the twofold increase in ρ is likely influenced by higher recombination rates per site in *C. purpurea*. Although, *Z. tritici* is a heterothallic organism while *C. purpurea* is homothallic (Esser and Tudzynski 1978) but *C. purpurea* also frequently out-crosses in nature (Amici *et al.* 1967; Tudzynski 2006), suggesting that these factors may provide a difference in effective population sizes between these organisms. In addition, mutation rates might differ between *C. purpurea* and *Z. tritici* for several reasons. Selection pressure associated with agriculture control methods could be driving the mutation of *Z. tritici*, which is subjected to multiple annual fungicide treatments (Torriani *et al.* 2015) and multiple cultivars with various qualitative and quantitative resistance sources (Brown *et al.* 2015). In contrast, control of *C. purpurea* is focused on cultural practices as fungicides have proven inefficient and no resistance crop germplasm has been identified (Menzies and Turkington 2015). While fungicides and crop resistance affect the population structure of *Z. tritici* (Estep *et al.* 2015; Hayes *et al.* 2016; Welch *et al.* 2018), it is plausible to believe they might affect mutation rate or select for strains with a higher mutation or recombination rates. However, we are unaware of any study that has directly examined whether fungicides or crop resistance can have direct or indirect effects on mutation rates. An alternative, and more plausible, hypothesis to explain an increased mutation rate in *Z. tritici* would be associated with the function of RIP, which identifying repeat/duplicated sequences within a

genome and introduces C:G to T:A mutations to effectively silence these regions (Galagan *et al.* 2003, 2004; Urquhart *et al.* 2018). It has also been reported that RIP can "leak" into neighboring non-repetitive regions and introduce mutations, thus, accelerating the rate of mutations, particularly those in closer proximity to repeat regions (Fudal *et al.*, 2009; Van de Wouw *et al.* 2010; Hane *et al.* 2015). If the mutation rate is increased in *Z. tritici*, either due to RIP "leakage" or selective pressure from fungicides or host resistance the nucleotide diversity in *Z. tritici* could be the result of high mutation rates, whereas the nucleotide diversity in *C. purpurea* could be influenced by higher effective population size and/or recombination rates per site. Higher recombination rates were found to increase the efficacy of purifying selection in both plants (Tiley and Burleigh 2015) and *Z. tritici* (Grandaubert *et al.* 2019). Similarly, *C. purpurea* had an overall trend of purifying selection with skewness towards lower ω values (Fig. 4.3) and an observed correlation of higher population recombination rates around genes with lower ω ratios (Appendix 3 Fig A3.9), further suggesting the potential for higher recombination rates in *C. purpurea*.

Additional support, for higher recombination rates per site in *C. purpurea*, could be extrapolated from recombination hotspots, or lack thereof. While we observed evidence of a heterogenous recombination landscapes with several scaffolds showing large peaks in population recombination rates (Fig. 4.5, Appendix 3 Fig. A3.6), we only predicted five recombination hotspots (Fig. 4.7), which is in stark contrast to the ~ 1,200 hotspots identified in *Z. tritici* (Stukenbrock and Dutheil 2018b, *Updated dataset*). On average, we did observe higher population recombination rates across scaffolds compared to the rates observed across chromosomes of *Zymoseptoria* (Stukenbrock and Dutheil 2018a), suggesting that the background recombination rate in *C. purpurea* is higher and "flatter", potentially limiting the detection of

hotspots (Auton *et al.* 2014). Overall, this indicates that *C. purpurea* exhibits high actual recombination rates as a defense mechanism to combat TE expansion.

While these higher recombination rates are likely influencing the trend of strong purifying selection observed in *C. purpurea*, it might not be the sole factor responsible for the low number of predicted effector genes with signatures of positive selection (Fig. 4.4). Wäli *et al.* (2013) classified *C. purpurea* as a conditional defense mutualist with its plant host, as they found that sheep avoided grazing infected grasses and observed that infection rates were higher in grazed pastures compared to ungrazed fields. Other researchers have observed neutral to positive effects of seed set, seed weight, and plant growth on infected plants compared to uninfected plants (Raybould *et al*. 1998; Fisher *et al.* 2007; Wäli *et al.* 2013; Chapter 2). These factors, along with the broad host range of *C. purpurea* (400+ grass species) and lack of known crop resistance (R) genes, could suggest a lack of strong selection for resistance, to *C. purpurea*, in grass species (Wäli *et al.* 2013). This could help explain the lack of positive selection observed in predicted core effector genes, implying that effectors are not under strong selection pressure to compete in the evolutionary arms race against host defense. However, it should be noted that positive selection analyses are computed from single-copy core orthologs. Observations of significant enrichment of predicted effector genes in the accessory genome of *C. purpurea* and duplication of effector gene clusters (Chapter 3) could implicate their role in diversity of infection potential (Sánchez-Vallet *et al.* 2018), however, no host specific races of *C. purpurea* have been identified.

Claviceps purpurea, which is suggested to have an ancestral state of plant endophytism (Píchová *et al.* 2018) is also closely related to several mutualistic grass endophytes (i.e. *Epichloë*, *Balansia*, *Atkinsonella*) which have been known to provide beneficial aspects to their hosts

mostly through production of secondary metabolites and plant hormones (Clay 1988; Song *et al.* 2016; Xia *et al.* 2018). *Claviceps purpurea* is well-known for its secondary metabolite production and, as we observed, had the second highest proportion of genes with positive selection signatures, the highest number of codons under selection per gene (Fig. 4.4 B, C), and were enriched in polyketide synthase domains and phosphopantetheine domains. We also observed two orthogroups, with signatures of positive selection, containing domains for laccase CAZy enzymes - with some laccases facilitating the biosynthesis of melanin in fungi (Lee *et al.* 2019) - and selection signatures on the cytochrome P450 associated with fungal cytokinin biosynthesis (Hinsch *et al.* 2015). Secondary metabolites are known to increase stress tolerance in fungi (i.e. against UV radiation, oxidative stresses, or colder climates) as has been shown with several groups of pigments, such as melanin and carotenoids (Avalos and Carmen Limon 2015). Therefore, the evolution of secondary metabolites in *C. purpurea* (i.e. ergot alkaloids, ergochromes, or other pigments) can theoretically increase fitness by altering infection potential, stress tolerance, or antimicrobial resistance (Píchová *et al.* 2018; Pusztahelyi *et al.* 2019). The difference in the proportion of secondary metabolites genes under positive selection pressure, compared to predicted effectors, indicates that the evolution of secondary metabolite genes in *C. purpurea* is more important to the success of the species than adaptation of core effector proteins. This is in contrast to many fungal plant pathogens of cereal crops, such as *Z. tritici* and the rust fungi in the genus *Puccinia*, that rely on adaptation and diversification of effector proteins for success, particularly due to breeding of crop varieties with R genes (Sánchez-Vallet *et al.* 2018; Badet *et al.* 2020). The selective pressure on secondary metabolites in *C. purpurea* could help explain its evolutionary history as it was recently postulated that evolution of *Claviceps* section *Claviceps*, of which *C. purpurea* resides, occurred tandemly with the radiation

of the core Pooideae (Poeae, Triticeae, Bromeae, and Littledaleae) and was associated with adaptation and diversification to cooler, more open habitats (Kellogg 2001; Sandve and Fjellheim 2010; Píchová *et al.* 2018; Chapter 3). In addition, the speciation among *C. purpurea* and closely related species demonstrate varied levels of adaptation to ecological niches (Pažoutová *et al.* 2000, 2002, 2015; Douhan *et al.* 2008; Van der Linde *et al.* 2016; Negård *et al.* 2015; Shoukouhi *et al.* 2019; Liu *et al. Accepted*). Similar evolutionary trends towards positive selection of secondary metabolites could be influencing the divergence of these species as well. In fact, all members of *Claviceps* section *Claviceps*had genomes that lack RIP, exhibit gene duplication, and have comparable TE content (Chapter 3), suggesting that the genomic mechanisms identified in this study might be characteristic of section *Claviceps* as a whole.

CONCLUSION

Overall, we observed that the *Claviceps purpurea* pangenome is composed of a large accessory genome that is likely influenced by a large effective population size, high recombination rates, and TE mediated gene duplication. Pseudogenization and neofunctionalization might also be contributing due to the observed TE activity, observations of higher ω ratios, signatures of positive selection in core single-copy unclassified genes, and small size of many accessory unclassified genes. Due to a lack of RIP, prolific TE expansion is likely controlled by high recombination rates, which subsequently may be influencing the overall trend of purifying selection. However, secondary metabolites genes were found to have the highest rates of positive selection on codons within genes, indicating that these genes are a primary factor affecting the diversification of the species into new ecological niches and to potentially help maintain its global distribution and broad host range.

Conclusion and future direction

This study has provided a greater understanding of the epidemiology, lifestyle, evolution, and adaptability of *Claviceps purpurea* and a substantial increase in genomic resources to which fuel continued research into *C. purpurea*, the genus *Claviceps*, and basic evolutionary theories of adaptable fungal species. Our research has revealed that *Bromus* spp. represent a vast inoculum reservoir of *C. purpurea* in the San Luis Valley, CO, even in drought or low rainfall years. Due to the invasiveness and susceptibility of *Bromus* spp., it is likely that these grass species represent primary factors in increasing the potential of ergot outbreaks wherever *Bromus* spp. constitute a majority of the composition in unmanaged grasses surrounding cereal crop fields. In addition, due to its rapid spread through rhizomes and our observations of the potential for increased root growth with higher infection rates of *C. purpurea*, it raises questions as to whether there is an elaborate coevolved symbiosis between these two species and other highly rhizomatous grass species. Such an interaction could ensure the continued propagation of *C. purpurea* by causing heavy infections on low seed set rhizomatous grasses. These infections can further lower seed production, thus, limiting sexual reproduction and genetic variability of the host. Due to this reduction, plants may be forced to reallocate resources to asexual rhizomatous growth, thereby reducing cultivar and species diversity in the surrounding grass community through facilitated expansion of highly susceptible genotypes. If such a scenario exists it could represent the relative importance of native and invasive grasses in maintaining ergot inoculum in agricultural ecosystems, thereby adding additional significance to the impact of invasive grasses on our agriculture economy and food safety. Future research should continue to clarify this interaction as well as develop management strategies of these inoculum reservoirs to reduce the

overall probability of ergot outbreaks within cereal crop fields. These strategies should be implemented alongside current whole-farm approaches for the control of ergot as our genomic data suggests that the search for resistance crop varieties may continue to prove challenging; due to the adaptability of *C. purpurea* through duplication of effector genes and high recombination rates.

Our genomic analyses have revealed that *C. purpurea* is a highly adaptable species as a result of large accessory genome that is maintained through a large effective population size, high recombination rates, and transposable elements (TE) mediated gene duplication. This is complemented by the potential for pseudogenization and neofunctionalization due to the observed gene duplication, observations of higher ω ratios, signatures of positive selection in core unclassified genes, and small size of many accessory unclassified genes. These unknown and developing genes could represent novel genes promoting the success of the species, however, further research is needed to clarify these events. It is more plausible to postulate that the observed abundance of secondary metabolites genes under positive selection are primary factor affecting the diversification of the species and help maintain its global distribution and broad host range. These results support current pangenome theories that large accessory genomes promote the adaptive evolution of organisms which allows these organisms to migrate into new ecological niches. In fact, this theory directly aligns with the observed cryptic speciation that is occurring within the pre-molecular concept of *C. purpurea sensu lato* (within section C*laviceps*), with evidence pointing towards adaptation to ecological niches. This is supported through evidence that these cryptic species also exhibit gene duplication, similar TE content, genome size, and genomic architecture and plasticity, suggesting that the genomic mechanisms identified in *C. purpurea* might be characteristic of section *Claviceps* as a whole. This adaptability and

observed lack of core effectors under positive selection would help explain the difficulties in identifying resistance genes in two contrasting ways. The adaptability coupled with an enriched accessory genome of effectors could represent an actively evolving repertoire of virulence factors that allow *C. purpurea* to compete in the evolutionary arms race against host defense. On the other hand, the lack of core effectors undergoing positive selection imply that these core effectors are not under strong selection pressure to compete in the evolutionary arms race against host defense. The evidence of conditional mutualistic interactions of *C. purpurea* and its host might suggest that the latter is occurring, which could indicate that the search for resistance genes within hosts will continue to be unsuccessful. However, due to the global distribution of *C. purpurea* its broad host range of 400+ grass species it would not be unsurprising if both contrasting mechanisms are working in tandem for the success of the species.

References

Adams, MW (1953) Cross- and self-incompatibility in relation to seed-setting in *Bromus inermis*. *Botanical Gazette* **115,** 95-105.

Ainsworth GC (1976) Introduction to the History of Mycology. Cambridge University Press 367

Alderman SC (1993). Aerobiology of *Claviceps purpurea* in Kentucky bluegrass. *Plant disease* **77,** 1045-1049.

Alderman SC (2006) Ergot: biology and control. *Ergot DVD Transcript*. Available from: https://www.ars.usda.gov/ARSUserFiles/81/ErgotDVDtranscript.pdf

Alderman SC, Barker RE (2003) Evaluation of resistance to ergot, caused by *Claviceps purpurea*, in Kentucky bluegrass, based on incidence and severity estimates. *Plant Disease* **87,** 1043-1047.

Alderman SC, Halse RR, White JF (2004) A reevaluation of the host range and geographical distribution of *Claviceps* species in the United States. *Plant Disease* **88,** 63–81.

Alex JF (1998) Ontario weeds. Publication 505. Ontario Ministry of Agriculture, Food and Rural Affairs, Guelph, ON. pp. 304

Alias N, Ahmad Mazian M, Salleh AB, Basri M, Rahman RN (2014) Molecular cloning and optimization for high level expression of cold-adapted serine protease from Antarctic yeast *Glaciozyma antarctica* PI12. *Enzyme Research* **2014,** 197938.

Alm T. (2003) The witch trials of Finmark, Northern Norway, during the 17th century: Evidence for ergotism as a contributing factor. *Economic Botany* **57,** 403–416.

Alter RE, Fan Y, Lintner BR, Weaver CP (2015) Observational evidence that great plains irrigation has enhanced summer precipitation intensity and total in the Midwestern United States. *Journal of Hydrometeorology* **16,** 1717-1735.

Amici AM, Scotti T, Spalla C, Tognoli L (1967) Heterokaryosis and alkaloid production in *Claviceps purpurea*. *Applied microbiology* **15,** 611-615

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://wwwbioinformaticsbabrahamacuk/projects/fastqc.

Anon. 2002 Managing ergot in cereal crops. HGCA Topic Sheet No. 56. HGCA, London (http//www.hgca.com)

Araki H, Tian D, Goss EM, Jakob K, Halldorsdottir SS, Kreitman M, *et al.* (2006) Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. *Pnas* **103,** 5887-92.

Auton A, McVean G (2007) Recombination rate estimation in the presence of hotspots. *Genome Research* **17,** 1219-1227

Auton A, Myers S, McVean G (2014) Identifying recombination hotspots using population genetic data. a*rXiv*: 1403.4264

Avalos J, Carmen Limon M (2015) Biological roles of fungal carotenoids. *Current Genetics* **61,** 309–324

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D (2020) A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology* **18,** 12.

Bailey KL, Gossen BD, Gugel RK, Morrall RAA (2003) Diseases of Field Crops in Canada. 3rd ed. Saskatoon, SK: *The Canadian Phytopathological Society and University Extension Press*, University of Saskatchewan; p. 89–93.

Balesdent M-H, Fudal I, Ollivier B, *et al.* (2013) The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene conferring avirulence towards *Brassica rapa*. *New Phytologist* **198,** 887-98.

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6,** 11

Bayles R, Fletcher M, Gladders P, Hall R, Hollins W, Kenyon D, Thomas J (2009) Towards a sustainable whole-farm approach to the control of ergot. Project Report No. 456 to the Home Grown Cereals Authority, Stoneleigh Park, UK; p. 170.

Beddows AR (1931) Seed setting and flowering in various grasses. *Bull. Welsh Plant Breed* Stn, Aberystwyth, Ser. H, **12,** 5-99.

Berriman M, Coghlan A, Tsai IJ (2018) Creation of a comprehensive repeat library for newly sequenced parasitic worm genome. *Protocol Exchange* doi: 101038/protex2018054.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research* **14,** 708–715.

Blanco E, Parra G, Guigo R (2007) Using geneid to identify genes. *Current Protocols in Bioinformatics* **18**, 431-4328.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **47,** W81-87.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114-20.

Bormann J, Tudzynski P (2009) Deletion of Mid1, a putative stretch-activated calcium channel in *Claviceps purpurea*, affects vegetative growth, cell wall synthesis and virulence. *Microbiology* **155,** 3922-3933.

Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR (2010) Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society* **162,** 543–557.

Bouckaert R, Heled J (2014) DensiTree 2: Seeing Trees Through the Forest, doi: http://dx.doi.org/10.1101/012401.

Bradshaw AD (1959) Population differentiation in *Agrostis tenuis* Sibth. II. The incidence and significance of infection by *Epichloe typhina*. *New Phytologist* **58,** 310-892.

Bretag TW (1981) Effect of burial on survival of sclerotia and production of stromata by *Claviceps purpurea*. *Transactions of the British Mycological Society* **77,** 658-660.

Brown AM, (1947) Ergot of cereals and grasses. *Canadian Phytopathological Society* **15,** 15.

Brown JKM, Chartrain L, Lasserre-Zuber P, Saintenac C (2015) Genetics of resistance to *Zymoseptoria tritici* and applications to wheat breeding. *Fungal Genetics and Biology* **79,** 33-41.

Brynildsrud O, Gulla S, Feil EJ, Nørstebø SF, Rhodes LD (2016) Identifying copy number variation of the dominant virulence factors msa and p22 within genomes of the fish pathogen *Renibacterium salmoninarum*. *Microbial Genomics* **2,** e000055.

Buchfunk B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12,** 59-60.

Bushnell, B (2014) BBMap: A fast, accurate, splice-aware aligner. Available at: https://sourceforgenet/projects/bbmap/.

Campbell WP (1957) Studies on ergot infection in gramineous hosts. *Canadian Journal of Botany* **35,** 315–320.

Campbell WP, Freisin HA (1959) The control of ergot in cereal crops. *Plant Disease Reporter*, **43,** 1266-1267.

Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders DGO, Uauy C (2013) Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f sp *tritici*reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* **14,** 270.

Caporael LR (1976) Ergotism: The Satan loosed in Salem? *Science* **192,** 21–26.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17,** 540-552.

Chalvet F, Grimaldi C, Kaper F, Langin T, Dabousii MJ (2003) Hop, an active Mutator-like element in the genome of the fungus *Fusarium oxysporum. Molecular Biology and Evolution* **20,** 1362–1375.

Cheng CF (1946) Self-fertility studies in three species of commercial grasses. *Jounral of American Society of Agronomy* **38,** 873-881.

Cincotta AH, Meier AH, Jr MC (2005) Bromocriptine improves glycaemic control and serum lipid profile in obese Type 2 diabetic subjects: a new approach in the treatment of diabetes. *Expert Opinion on Investigational Drugs*. doi: 10.1517/ 13543784.8.10.1683.

Clay K (1984) The effect of the fungus *Atkinsonella hypoxylon* (Clavicipitaceae) on the reproductive system and demography of the grass *Danthonia spicata*. *New Phytologist* **98,** 165-175.

Clay K (1986) Induced vivipary in *Cyperus virens* and the transmission of the fungus *Balansia experi*. *Canadian Journal of Botany* **64,** 2984-2988.

Clay K (1987) Effects of fungal endophytes on the seed and seedling biology of *Lolium perenne* and *Festuca arundinacea*. *Oecologia* **73,** 358-362.

Clay K (1988) Fungal endophytes of grasses: a defensive mutualism between plants and fungi. *Ecology* **69,** 10-16.

Croll D, Lendenmann MH, Stewart E, BA MD (2015) The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics* **201,** 1213-28.

Cunfer BM, Seckinger A (1977) Survival of *Claviceps purpurea* and *C. paspali* sclerotia. *Mycologia* **69,** 1142-1148.

Curtis JT, McIntoch RP (1951) An upland forest continuum in the prairie-forest border region of Wisconsin. *Ecology* **32,** 476-496.

Davière JM, Langin T, Daboussi MJ (2001) Potential role of transposable elements in the rapid reorganization of the *Fusarium oxysporum* genome. *Fungal Genetics and Biology* **34,** 177–192.

De Battista JP, Bacon CW, Severson R, Plattner RD, Bouton JH (199) Indole acetic acid production by the fungal endophyte of tall fescue. *Agronomy Journal* **82,** 878-880.

Dethier MN, Hacker SD (2004). Improving management practices for invasive cordgrass in the Pacific Northwest: a case study of *Spartina anglica*. *Sea Grant Washington*.

Diehl WW (1950) *Balansia* and the Balansiae in America. *Agriculture Monograph 4*. United States Department of Agriculture. Washington, D.C., USA.

Dillemuth FP, Rietschier EA, Cronin JT (2009) Patch dynamics of a native grass in relation to the spread of invasive smooth brome (*Bromus inermis*). *Biological Invasions* **11,** 1381–1391.

Domingo WE (1941) Bulk Emasculation and Pollination of Smooth Bromegrass, *Bromus inermis*. *Agronomy Journal* **33,** 993-1002.

Dong S, Raffaele S, Kamoun S (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics and Development* **35,** 57–65.

Dopstadt J, Neubauer L, Tudzynski P, Humpf H-U (2016) The epipolythiodiketopiperazine gene cluster in *Claviceps purpurea*: dysfunctional cytochrome p450 enzyme prevents formation of the previously unknown clapurines. *PLoS ONE* **11,** e0158945.

Douhan GW, Smith ME, Huyrn KL, Westbrook A, Beerli P, Fisher AJ (2008) Multigene analysis suggests ecological speciation in the fungal pathogen *Claviceps purpurea*. *Molecular Ecology* **17,** 2276-2286.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19,** 11-15.

Dutheil JY, Gaillard S, Stukenbrock EH (2014) MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC genomics* **15,** 53.

Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Münsterkötter M, Güldener U, Schirawski J, Kahmann R (2016) A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. *Genome Biology and Evolution* **8,** 681–704.

Eastin JD, Teel MR, Langston R (1964) Growth and Development of Six Varieties of Smooth Bromegrass (*Bromus inermis* Leyss.) with Observations on Seasonal Variation of Fructosan and Growth Regulators. *Crop science* **4,** 555-559.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, **1792-1797**

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460-2461.

Eleuterius LN, Meyers SP (1974) *Claviceps purpurea* on *Spartina* in coastal marshes. *Mycologia* **66,** 978-986.

Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, a efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9,** 18.

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20,** 238.

Ensign RD, Weiser GC (1975) Root and Rhizome Development of Some Kentucky Bluegrass and Red Fescue Cultivars. *Agronomy Journal* **67,** 583-585.

Esser K, Tudzynski P (1978) Genetics of the ergot fungus *Claviceps purpurea*. I. Proof of a monoecious life-cycle and segregation patterns for mycelial morphology and alkaloid production. *Theoretical Applied Genetics* **53,** 145-149.

Estep LK, Torriani SFF, Zala M, Anderson NP, Flowers MD, McDonald BA, Mundt CC, Brunner PC (2015) Emergence and early evolution of fungicide resistance in North American populations of *Zymoseptoria tritici*. *Plant Pathology* **64,** 961-971.

Evans VJ, Jenkyn JF, Gladders P, Mantle PG (2000) Fungicides for the control of ergot in cereal crops. In: *Proceedings Brighton Crop Protection Conference-Pests and Diseases, 2000* pp. 511-514.

Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ (2016) Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research* **26,** 1091–1100.

Fink KA, Wilson SD (2011) *Bromus inermis* invasion of a native grassland: diversity and resource reduction. *Botany* **89,** 157-164.

Fisher AJ, DiTomaso JM, Gordon TR, Aegerter BJ, Ayres DR (2007) Salt marsh *Claviceps purpurea* in native and invaded Spartina marshes in Northern California. *Plant Disease* **91,** 380–386.

Flieger M, Stodůlková E, Wyka SA, *et al.* (2019) Ergochromes: heretofore neglected side of ergot toxicity. *Toxins* **11,** 439.

Frantzeskakis L, Kracher B, Kusch S, *et al.* (2018) Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19,** 381.

Frantzeskakis L, Kusch S, Panstruga R (2019) The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology* **20,** 3-7

Freeman BC, Beattie GA (2008). An overview of plant defenses against pathogens and herbivores. *The Plant Health Instructor*.

Freitag M, Williams RL, Kothe GO, Selker EU (2002) A cytosine methyltransferase homologue is essential for repeat_induced point mutation in *Neurospora crassa*. *PNAS* **99,** 8802-8807.

Fudal I, Ross S, Brun H, Besnard A-L, Ermel M, Kuhn M-L, Balesdent M-H, Rouxel T (2009) Repeat-induced point mutation (RIP) as an alternative mechanism of evolution towards virulence in *Leptosphaeria maculans*. *Molecular Plant-Microbe Interactions* **22,** 932-941.

Galagan JE, Calvo SE, Borkovich KA *et al*. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422,** 859-868.

Galagan JE Selker EU (2004) RIP: the evolutionary cost of genome defense. *TRENDS in Genetics* **20,** 417-423.

Galuska P, Spicahl L, Kopecny D, Tarkowski P, Frebortova J, Sebela M, Frebort I (2008) Metabolism of plant hormones cytokinins and their function in signaling, cell differentiation and plant development. *Studies in Natural Products Chemistry* **34,** 203-264.

Giesbert S, Schürg T, Scheele S, Tudzynski P (2008) The NADPH oxidase Cpnox1 is required for full pathogenicity of the ergot fungus *Claviceps purpurea*. *Molecular Plant Pathology* **9,** 317-327.

Gladders P, Evans VJ, Jenkyn JF, Lockley KD, Mantle PG (2001) Laboratory and field testing of fungicide for control of ergot in wheat and rye. Home Grown Cereals Authority (HGCA), London.

Gladieux P, Ropars J, Badouin H, *et al.* (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular ecology* **23,** 753-773.

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* **11,** 725-736.

Gordon A, McCartney C, Knox RE, *et al.* (2020) Genetic and transcriptional dissection of resistance to *Claviceps purpurea* in the durum wheat cultivar Greenshank. *Theoretical and Applied Genetics* doi: 10.1007/s00122-020-03561-9.

Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29,** 644-52.

Graïa F, Lespinet O, Rimbault B, Deguard-Chablat M, Coppin E, Picard M (2001) Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. *Molecular Microbiology* **40,** 586–595.

Grandaubert J, Bhattacharyya A, Stukenbrock EH (2015) RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. *Genes, Genomics, Genetics* **5,** 1323-1333.

Grandaubert J, Dutheil JY, Stukenbrock EH (2019) The genomic determinants of adaptive evolution in a fungal pathogen. *Evolution Letters* **3,** 299-312.

Gray AJ, Drury M, Raybould AF (1990) *Spartina* and the ergot fungus *Claviceps purpurea*-a singular contest? Blackwell Scientific Publications. Pp.63-79.

Great Plains Flora Association (1986) Flora of the Great Plains. University Press of Kansas, Lawrence, KS. pp 1402.

Gremme G, Steinbiss S, Kurtz S (2013) GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10,** 645–656.

Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32,** 2847-9.

Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP (2015) Repeat-induced point mutation: a fungal-specific, endogenous mutagenesis process. In: van den Berg MA

Maruthachalam K, editors. *Genetic transformation systems in fungi*. Vol. 2. Springer International Publishing, p. 55-68.

Harberd DJ (1960) Association-analysis in plant communities. *Nature* **185,** 53-54.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D (2017) A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal* **11,** 1189-204.

Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T (2018) Gene presence–absence polymorphism in castrating Anther-Smut fungi: recent gene gains and phylogeographic structure. *Genome Biology and Evolution* **10,** 1298-314.

Hass B (2010) TransposonPSI. Available online at: http://transposonpsisourceforgenet.

Hass BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology* **9,** R7.

Hayes LE, Sackett KE, Anderson NP, Flowers MD, Mundt CC (2016) Evidence of selection for fungicide resistance in *Zymoseptoria tritici* populations on wheat in western Oregon. *Plant disease* **100,** 483-489.

Hinsch J, Vrabka J, Oeser B, Novák O, Galuszka P, Tudzynski P (2015) *De novo* biosynthesis of cytokinins in the biotrophic fungus *Claviceps purpurea*. *Environmental Microbiology* **17,** 2935-2951.

Hinsch J, Galuszka P, Tudzynski P (2016) Functional characterization of the first filamentous fungal tRNA-isopentenyltransferase and its role in the virulence of *Claviceps purpurea*. *New Phytologist* **211,** 980-992.

Horst RJ, Engelsdorf T, Sonnewald U, Voll LM (2008) Infection of maize leaves with *Ustilago maydis* prevents establishment of C4 photosynthesis. *Journal of plant physiology* **165,** 19-28.

Huerta-Cepas J, Szklarczyk D, Heller D, *et al.* (2019) eggnog 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47,** D309-314

Huerta-Cepas J, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2016) Fast genome-wide functional annotation through orthology assignment by eggnog-mapper. *Molecular Biology and Evolution* **34,** 2115-2122.

Ikeda K, Nakayashiki H, Tosa Y, Mayama S (2002) Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Molecular Microbiology* **45,** 1355–1364.

Jarosz AM, Davelos AL (1995) Effects of disease in wild plant populations and the evolution of pathogen aggressiveness. *New Phytologist* **129,** 371-387.

Jeffares DC, Tomiczek B, Sojo V, dos Reis M (2015) A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. In: Peacock C. (eds) Parasite Genomics Protocols. Methods in Molecular Biology, vol 1201. Humana Press, New York, NY

Johnson NC, Graham JH, Smith FA (1997) Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *The New Phytologist* **135,** 575-585.

Johnston WJ, Golob CT, Sitton JW, Schultz TR (1996). Effect of temperature and postharvest field burning of Kentucky bluegrass on germination of sclerotia of *Claviceps purpurea*. *Plant disease* **80,** 766-768.

Jones P, Binns D, Chang H-Y, *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30,** 1236-1240.

Jungehülsing U, Tudzynski P (1997) Analysis of genetic diversity in *Claviceps purpurea* by RAPD markers. *Mycological Research* **101,** 1–6.

Kachroo P, Leong SA, Chatto BB (1994) Pot2, an inverted repeat transposon from the rice blast fungus *Magnaporthe grisea*. *Molecular and General Genetics* **245,** 339–348.

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

Kämper J, Kahmann R, Bölker M *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444,** 97–101.

Karl TR (1998) Secular trends of precipitation amount, frequency, and intensity in the USA. *Bulletin of the American Meteorological Society* **79,** 231–341.

Katoh K, Standley, DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30,** 772-780.

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F (2016) Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* **44,** e89.

Kelley SE, Clay K (1987) Interspecific competitive interactions and the maintenance of genotypic variation within the populations of two perennial grasses. *Evolution* **41,** 92-103.

Kellogg, EA (2001) Evolutionary history of the grasses. *Plant Physiology* **125,** 1198-1205.

Kind S, Schurack S, Hinsch J, Tudzynski P (2018a) Brachypodium distachyon as alternative model host system for the ergot fungus *Claviceps purpurea*. *Molecular Plant Pathology* **19,** 1005-1011.

Kind S, Hinsch J, Vrabka J, Hradilová M, Majeská-Čudejková M, Tudzynski P, Galuska P (2018b) Manipulation of cytokinin level in the ergot fungus *Claviceps purpurea* emphasizes its contribution to virulence. *Current Genetics* **64,** 1303-1319.

Kiran K, Rawal HC, Dubey H, Jaswal R, Bhardwaj SC, Prasad P, Pal D, Devanna BN, Sharma TR (2017) Dissection of genomic features and variations of three pathotypes of *Puccinia striiformis* through whole genome sequencing. *Scientific Reports* **7,** 42419.

Kislyuk AO, Haegeman B, Bergman NH, Weitz JS (2011) Genomic fluidity: an integrative view of gene diversity within microbial populations. *BMC Genomics* **12,** 32.

Kito H, Takahashi Y, Sato J, Fukiya S, Sone T, Tomita F (2003) Occan, a novel transposon in the Fot1 family, is ubiquitously found in several *Magnaporthe grisea* isolates. *Current Genetics* **42,** 322–331.

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B* **279,** 5048-5057.

Korf I (2004) Gene finding in novel Genomes. *BMC Bioinformatics* **5,** 59.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305,** 567-580.

Krska R, Crews C (2008) Significance, chemistry and determination of ergot alkaloids: A review. *Food Additives and Contaminants* **25,** 722-731.

Kuldau G, Bacon C (2008) Clavicipitaceous endophytes: Their ability to enhance resistance of grasses to multiple stresses. *Biological Control* **46,** 57-71.

Kvas M, Marasas WFO, Wingfield BD, Wingfield MJ, Steenkamp ET (2009) Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. *Fungal Diversity* **34,** 1-21.

Lamp HF (1952) Reproductive Activity in *Bromus inermis* in Relation to Phases of Tiller Development Contributions from the Hull Botanical Laboratory 633. *Botanical gazette* **113,** 413-438.

Langdon RFN (1952) Studies on ergot (PhD thesis). Queensland University, Brisbane, Australia

Latch GCM, Hunt WF, Musgrave DR (1985) Endophytic fungi affect growth of perennial ryegrass. *New Zealand Journal of Agricultural Research* **28,** 165-168.

Lee D, Jang E-H, Lee M, Kim S-W, Lee Y, Lee K-T, Bahn Y-S (2019) Unraveling melanin biosynthesis and signaling networks in *Cryptococcus neoformans*. *mBio* doi:10.1128/mBio.02267-19

Lee MR (2009) The history of ergot of rye (*Claviceps. purpurea*) I: From antiquity to 1900. *The Journal of the Royal College of Physicians of Edinburgh* **39,** 179–184.

Leung TLF, Poulin R (2008) Parasitism, commensalism, and mutualism: exploring the many shades of symbiosis. *Vie et Milieu – Life and Environment* **58,** 107-115.

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094-3100.

Liu M, Shoukouhi P, Hicks C, Bisson K, Spoule A, Wyka SA, Broders K, Popovic Z, Menzies JG. (2020) Four new ergot species based on morphology, alkaloid production, pathogenicity and DNA sequences analyses. *Mycologia Accepted*.

Looman J (1983) 111 range and forage plants of the Canadian prairies. Research Branch, Agriculture Canada. Ottawa, ON. pp. 255.

Ma L-J, van der Does HC, Borkovich KA, *et al*. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464,** 367-373.

Mahmoud A, Taga M (2012) Cytological karyotyping and characterization of a 410 kb minichromosome in *Nectria haematococca* MPI. *Mycologia* **104,** 845-56.

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* **20,** 2878-2879.

Mantle PG, Shaw S, Doling DA (1977) Role of weed grasses in the etiology of ergot disease in wheat. *Annals of Applied Biology* **86,** 339-351

Mario S, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24,** 637–644.

Marks TC, Truscott AJ (1985) Variation in seed production and germination of Spartina anglica within a zoned saltmarsh. *The Journal of Ecology* **73,** 695-705.

McCarthy CGP, Fitzpatrick DA (2019) Pan-genome analysis of model fungal species. *Microbial Genomics* **5,** e000243

McInerney JO, McNally A, O'Connell MJ (2017) Why prokaryotes have pangenomes. *Nature Microbiology* **2,** 17040

McKone MJ (1985) Reproductive biology of several bromegrasses (*Bromus*): breeding system, pattern of fruit maturation, and seed set. *American Journal of Botany* **72,** 1334–1339.

Menzies JG (2004) The reactions of Canadian spring wheat genotypes to inoculation with *Claviceps purpurea*, the causal agent of ergot. *Canadian Journal of Plant Science* **84,** 625-629.

Menzies JG, Turkington TK (2015) An overview of the ergot (*Claviceps purpurea*) issue in western Canada: challenges and solutions. *Canadian Journal of Plant Pathology* **37,** 40-51.

Menzies JG, Klein-Gebbinck HW, Gordon A, O'Sullivan DM (2017) Evaluation of *Claviceps purpurea* isolates on wheat reveals complex virulence and host susceptibility relationships. C*anadian Journal of Plant Pathology* **39,** 307-317.

Mey G, Oeser B, Lebrun MH, Tudzynski P (2001) The biotrophic, non-appressorium-forming grass pathogen *Claviceps purpurea* needs a *Fus3/Pmk1* homologous mitogen-activated protein kinase for colonization of rye ovarian tissue. *Molecular Plant-Microbe Interactions* **15,** 303-312.

Mey G, Held K, Scheffer J, Tenberge KB, Tudzynski P (2002) CPMK2, an SLT2-homologous mitogen-activated protein (MAP) kinase, is essential for pathogenesis of *Claviceps purpurea* on rye: evidence for a second conserved pathogenesis-related MAP kinase cascade in phytopathogenic fungi. *Molecular Microbiology* **46,** 305-318.

Miedaner T, Geiger HH (2015) Biology, genetics, and management of ergot (*Claviceps* spp.) in rye, sorghum, and pearl millet. *Toxins* **7,** 659-678.

Milly PCD, Wetherald RT, Dunne KA, Delworth TL (2002) Increasing risk of great floods in a changing climate. *Nature* **415,** 514–517.

Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* **15,** 756-771.

Morran LT, Schmidt OG, Gelarden IA, Parrish RC 2nd, Lively CM (2011) Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science* **333,** 216-8

Murphy RK, Grant TA (2005) Land management history and floristics in mixed-grass prairie, North Dakota, USA. *Natural Areas Journal* **25,** 351–358.

Nagel HG, Nicholson RA, Steuter AA (1994) Management effects on Willa Cather Prairie after 17 years. *Prairie Naturalist* **26,** 241–250.

Nakayashiki H, Nishimoto N, Ikeda K, Tosa Y, Mayama S (1999) Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Molecular and General Genetics* **261,** 958–966.

Negård M, Uhlig S, Kauserud H, Andersen T, Høiland K, Vrålstad T. (2015) Links between genetic groups, indole alkaloid profiles and ecology within the grass-parasitic *Claviceps purpurea* species complex. *Toxins* **7,** 1431–1456.

Neubauer L, Dopstadt J, Humpf H-U, Tudzynski P (2016) Identification and characterization of the ergochrome gene cluster in the plant pathogenic fungus *Claviceps purpurea*. *Fungal Biology and Biotechnology* **3,** 2.

Newton RR, Newton IL (2013) PhyBin: binning trees by topology. *Peer J.* 1:e187; doi: 10.7717/peerj.187.

Nielsen H (2017) Predicting secretory proteins with SignalP In: *Protein function prediction* (eds. Kihara D). *Methods in Molecular Biology* **1611** Humana Press, New York, NY.

Nurk S, Bankevich A, Antipov D, *et al.* (2013) Assembling genomes and mini-metagenomes from highly chimeric reads. *Research in Computational Molecular Biology RECOMB 2013 Lecture Notes in Computer Science*, vol 7821 Springer, Berlin, Heidelberg.

Oberti H, Dalla Rizza M., Reyno R, Murchio S, Altier N, Abreo E. (2020) Diversity of *Claviceps paspali* reveals unknown lineages and unique alkaloid genotypes. *Mycologia* **112,** 1-14.

Oeser B, Gessner-Ulrich K, Deing P, Tudzynski P (1993) pClK1 and pClT5 - two linear mitochondrial plasmids from unrelated *Claviceps purpurea* strains: a comparison. *Plasmid* **30,** 274-280.

Oeser B, Heidrich PM, Muller U, Tudzynski P, Tenberge KB (2002) Polygalacturonase is a pathogenicity factor in the *Claviceps purpurea*/rye interaction. *Fungal Genetics and Biology* **36,** 176-186.

Oeser B, Kind S, Schurack S, Schmutzer T, Tudzynski P, Hinsch J (2017) Cross-talk of the biotrophic pathogen *Claviceps purpurea* and its host *Secale cereale*. *BMC Genomics* **18,** 273.

Otfinowski R, Kenkel NC, Catling PM (2007) The biology of Canadian weeds. 134. *Bromus inermis* Leyss. *Canadian Journal of Plant Science* **87,** 183–198.

Oxley SJP, Havis ND, Hoad SP (2009) Understanding ergot risk in spring barley. Project Report No. 457 to the Home Grown Cereals Authority, Stoneleigh Park, UK. p. 37.

Pageau D, Lajeunesse J (2006) Évaluation de la résistance à l'ergot (*Claviceps purpurea*) chez l'orge et le blé. *Phytoprotection*. **87,** 63–68. doi:10.7202/013974ar.

Palmer J, Stajich J (2019) nextgenusfs/funannotate: funannotate v1.6.0 (Version 1.6.0) Zenodo doi: 105281/zenodo3354704.

Panaccione DG, Cipoletti JR, Sedlock AB, Blemings KP, Schardl CL, Machado C, Seidel GE (2006) Effects of ergot alkaloids on food preference and satiety in rabbits, as assessed with gene-knockout endophytes in perennial ryegrass (*Lolium perenne*). *Journal of agricultural and food chemistry* **54,** 4582-4587.

Parbery DG (1996) Trophism and the ecology of fungi associated with plants. *Biological Reviews* **71,** 473-527.

Pažoutová S, Olšovská J, Linka M, Kolínská R, Flieger M (2000) Chemoraces and habitat specialization of *Claviceps purpurea* populations. *Applied and Environmental Microbiology* **66,** 5419–5425.

Pažoutová S, Raybould AF, Honzátko A, Kolínská R (2002) Specialised population of *Claviceps purpurea* from salt marsh *Spartina* species. *Mycological Research* **106,** 210–214.

Pažoutová S, Pešicová K, Chudíčková M, Šrůtka P, Kolařík M (2015) Delimitation of cryptic species inside *Claviceps purpurea*. *Fungal Biology* **119,** 7–26.

Pease JB, Hahn MW (2013) More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* **67,** 2376-2384.

Pecháčková S, During HJ., Rydlová V, Herben T (1999) Species-specific spatial pattern of below-ground plant parts in a montane grassland community. *Journal of Ecology* **87,** 569-582.

Peng Z, Oliveria-Garcia E, Lin G, *et al.* (2019) Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genetics* **15,** e1008272.

Píchová K, Pažoutová S, Kostovčík M, Chudíčková M, Stodulůkvá E, Novák P, Flieger M, van der Linde E, Kolařík M (2018) Evolutionary history of ergot with a new infrageneric classification (Hypocreales: Clavicipitaceae: *Claviceps*). *Molecular Phylogenetics and Evolution* **123,** 73-87.

Poppe S, Dorcheimer L, Happel P, Stukenbrock EH (2015) Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLoS Pathogens* **11,** e1005055.

Porter JK, Bacon CW, Cutler HG, Arrendale RF, and Robbins JD (1985) In vitro auxin production by *Balansia epichloe*. *Phytochemistry* **24,** 1429-1431.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490.

Pusztahelyi T, Holb IJ, Pócsi I (2015) Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science* **6,** doi: 10.3389/fpls.2015.00573

Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10,** 417–430.

Ranwell DS (1967) World resources of *Spartina townsendii* (*sensu lato*) and economic use of *Spartina* marshland. *Journal of applied ecology* **4,** 239-256.

Rapilly F (1968) Studies on ergot of wheat-*Claviceps purpurea* (Fr.) Tul. In: *Annales des epiphyties*. Vol, 19, No. 2, p. 305.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn R (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research* **46,** 624-632.

Raybould AF., Gray AJ, Lawrence MJ, Marshall DF (1991) The evolution of *Spartina anglica* CE Hubbard (Gramineae): origin and genetic variability. *Biological Journal of the Linnean Society* **43,** 111-126.

Raybould AF, Gray AJ, Clarke RT (1998) The long-term epidemic of *Claviceps purpurea* on *Spartina anglica* in Poole Harbour: pattern of infection, effects on seed production and the role of *Fusarium heterosporum*. *New Phytologist* **138,** 497–505.

Rep M, Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species. *Current Opinion in Plant Biology* **13,** 420-426.

Reynolds JH, Smith D (1962) Trend of Carbohydrate Reserves in Alfalfa, Smooth Bromegrass, and Timothy Grown Under Various Cutting Schedules 1. *Crop science* **2,** 333-336.

Rodriguez Estrada QE, Jonkers W, Kistler HC, May G (2012) Interaction between Fusarium verticillioides, *Ustilago maydis*, and *Zea mays*: An endophyte, a pathogen, and their shared plant host. *Fungal Genetics* and *Biology* **49,** 578–587.

Rolke Y, Tudzynski P (2008) The small GTPase Rac and the p21-activated kinase Cla4 in *Claviceps purpurea*: interaction and impact on polarity, development, and pathogenicity. *Molecular Microbiology* **68,** 405-423.

Romo JT, Grilz PL, Driver EA (1990) Invasion of the Canadian prairies by an exotic perennial. *Blue Jay* **48,** 131-135.

Rosenzweig C, Tubiello FN, Goldberg R, Mills E, Bloomfield J (2002) Increased crop damage in the US from excess precipitation under climate change. *Global Environmental Change* **12,** 197-202.

Saikkonen K, Faeth SH, Helander M, Sullivan TJ (1998) Fungal endophytes: a continuum of interactions with host plants. *Annual review of Ecology and Systematics* **29,** 319-343.

Saikkonen K, Wäli P, Helander M., Faeth SH (2004) Evolution of endophyte–plant symbioses. *Trends in plant science* **9,** 275-280.

Sánchez-Vallet A, Fouché S, Fudal I, Hartmann F, Soyer JL, Tellier A, Croll D (2018) The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology* **56,** 21-40.

Sandve SR, Fjellheim S (2010) Did gene family expansions during the Eocene- Oligocene boundary climate cooling play a role in Pooideae adaptation to cool climates? *Molecular Ecology* **19,** 2075–2088.

Schardl CL, Young CA, Hesse U, *et al*. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics* **9,** e1003323.

Scheffer J, Chen C, Heidrich P, Dickman MB, Tudzynski P (2005a) A CDC42 homologue in *Claviceps purpurea* is involved in vegetative differentiation and is essential for pathogenicity. *Eukaryotic Cell* **4,** 1228-1238.

Scheffer J, Ziv C, Yarden O, Tudzynski P (2005b) The COT1 homologue CPCOT1 regulates polar growth and branching and is essential for pathogenicity in *Claviceps purpurea*. *Fungal Genetics and Biology* **42,** 107-118.

Schiff PL (2006) Ergot and its alkaloids. *American Journal of Pharmaceutical Education* 70.

Schirawski J, Mannhaupt G, Münch K *et al.* (2010) Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* **330,** 1546–1548.

Schumann GL (2000) Ergot. The Plant Health Instructor. doi:10.1094/PHII-2000-1016 01.

Shoukouhi P, Hicks C, Menzies JG, Popovic Z, Chen W, Seifert KA, Assabgui R, Liu M (2019) Phylogeny of Canadian ergot fungi and a detection assay by real-time polymerase chain reaction. *Mycologia* **111,** 1-13.

Skálová H, Pecháčková S, Suzuki J, Herben T, Hara T, Hadincová V, Krahulec F (1997) Within population genetic differentiation in traits affecting clonal growth: *Festuca rubra* in a mountain grassland. *Journal of Evolutionary Biology* **10,** 383-406.

Smit AFA, Hubley R (2015) RepeatModeler Open-10. Available online at: http://wwwrepeatmaskerorg.

Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-40. Available online at: http://wwwrepeatmaskerorg.

Smith DC (1944) Pollination and seed formation in grasses. *Journal of Agricultural Research* **68,** 79-95.

Song H, Nan Z, Song Q, Xia C, Li X, Yao X, Xu W, Kuang Y, Tian P, Zhang Q (2016) Advances in research on *Epichloë* endophytes in Chinese native grasses. F*rontiers in Microbiology* **7,** 1399

Sorbe B, (1978) Active pharmacologic management of the third stage of labor: a comparison of oxytocin and ergometrine. *Obstetrics & Gynecology* **52,** 694–697.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, Barberá P, Gillespie LJ, Zuloaga FO (2017) A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution* **55,** 259–290.

Sperchneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, and Taylor JM (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biology and Evolution* **7,** 1613-1627.

Sperschneider J, Dobbs PN, Gardiner DM, Singh KB, Taylor JM (2018) Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology* **19,** 2094-2110.

Sprague R (1950) Diseases of cereal grasses in North America. New York, USA: The Ronald Press Company.

Stamatakis A (2014) RAxML version 8: a tool for phylogeneitc analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312-1313

Steinbiss S, Willhoeft U, Gremme G, Kurtz S (2009) Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acids Research* **37,** 7002–7013.

Stukenbrock EH (2013) Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* **199,** 895–907.

Stukenbrock EH, Dutheil JY (2018a) Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscape and intragenic hotspots. *Genetics* **208,** 1209-1229.

Stukenbrock EH, Dutheil JY (2018b) Data and scripts for: Fine-Scale Recombination Maps of Fungal Plant Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots. Available from: https://gitlab.gwdg.de/molsysevol/ZtPopRec

Sudhir K, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35,** 1547-1549.

Taber WA (1985) Biology of *Claviceps*. In: *Biotechnology Series, Vol. 6. Biology of Industrial Microorganisms*. A. L. Demain and A. S. Nadine, eds. The Benjamin Cummings Publishing Co., Inc., New York. pp. 449-486.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18,** 1979-1990.

Testa A, Hane JK, Ellwood SR, Oliver RP (2015a) CodingQuarry: highly accurate hidden markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* **16,** 170.

Testa A, Oliver R, Hane J (2015b) Overview of genomic and bioinformatic resources for *Zymoseptoria tritici*. *Fungal Genetics and Biology* **79,** 13-16.

Tfelt-Hansen P, Saxena PR, Dahlöf C, *et al.* (2000) Ergotamine in the acute treatment of migraine. A review and European consensus. *Brain* **123,** 9–18.

Tiley GP, Burleigh JG (2015) The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evolutionary Biology* **15,** 194.

Torriani SFF, Melichar JPE, Mills C, Pain N, Sierotzki H, Courbot M (2015) *Zymoseptoria tritici*: a major threat to wheat production, integrated approaches to control. *Fungal Genetics and Biology* **79,** 8-12.

Tracy WF, Vargas C, Zepeda L, Pataky JK, Chandler MA (2007) Production and marketing of huitlacoche. *Issues in new crops and new uses* 233-236.

Tudzynski P, Scheffer JAN (2004) *Claviceps purpurea*: molecular aspects of a unique pathogenic lifestyle. *Molecular Plant Pathology* ***5*****,** 377-388.

Tudzynski P (2006) Genetics of *Claviceps purpurea*. In: Kren V, Cvak L (eds) Ergot: the genus *Claviceps*. Hardwood Academic Publishers (republished 2006 by Taylor and Francise e-Library), Amsterdam, The Netherlands, pp 25-56

Tudzynski P., Neubauer L. (2014) Ergot Alkaloids. In: Martín JF., García-Estrada C., Zeilinger S. (eds) Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites. Fungal Biology. Springer, New York, NY

Uhlig S, Vikoren T, Ivanova L, Handeland K, (2007) Ergot alkaloids in Norwegian wild grasses: a mass spectrometric approach. *Rapid Communications in Mass Spectrometry* **21,** 1651–1660.

Uppala S, Wu BM, Alderman S, Gilmore L (2011). Studies on epidemiology of ergot in Kentucky bluegrass. *Oregon State University Central Oregon Agricultural Research Center 2011 Annual Report.* Central Oregon Agricultural Research Center, Madras, OR, p. 13-22.

Urguhart AS, Mondo SJ, Makela MR *et al.* (2018) Genomic and genetic insights into a cosmopolitan fungus, *Paecilomyces variotii* (Eurotiales). *Frontiers in Microbiology* **9,** 3058.

USDA: official U.S. Washington D.C. (VA): United States Department of Agriculture: Grain Inspection, Packers & Stockyards Administration; c1988-2016 Available from: https://www.gipsa.usda.gov/fgis/usstandards.aspx.

Van de Wouw A, 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 Pathogens* **6,** e1001180

Van der Linde EJ, Pešicová K, Pažoutová S, Stodůlková E, Flieger M, Kolařík M (2016) Ergot species of the *Claviceps purpurea* group from South Africa. *Fungal Biology* **120,** 917-930.

Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp, ET. (2019) The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point RIP mutations. *PeerJ* **7,** e7447.

Wäli PP, Wäli PR, Saikkonen K, Tuomi J (2013) Is the pathogenic ergot fungus a conditional defensive mutualist for its host grass? *PLoS ONE* **8,** e69249.

Wall JD, Stevison LS (2016) Detecting recombination hotspots from patterns of linkage disequilibrium. *G3* **6,** 2265-2271.

Watanabe M, Yonezawa T, Lee K, Kumagai S, Sugita-Konishi Y, Goto K, Hara-Kudo Y (2011) Molecular phylogeny of the higher and lower taxonomy of the *Fusarium* genus and differences in the evolutionary histories of multiple genes. *BMC Evolutionary Biology* **11,** 322.

Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35,** 543-548.

Welch T, Feechan A, Kildea S (2018) Effect of host resistance on genetic structure of core and accessory chromosomes in Irish *Zymoseptoria tritici* populations. *European journal of plant pathology* **150,** 139-148.

Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29,** 2487-2489

Willingale J, Mantle PG (1987) Stigmatic constriction in pearl millet flowering infection by *Claviceps fusiformis. Physiological and Molecular Plant Pathology* **30,** 247-257.

Wilsie CP, Ching CB, Hawk VB (1952) Self-fertility and progeny performance in *Bromus inermis*. *Agronomy Journal* **44,** 605-609.

Wingfield BD, Liu M, Nguyen HDT, *et al.* (2018) Nine draft genome sequences of *Claviceps purpurea* s.lat., including *C. arundinis*, *C. humidiphila*, and *C.* cf. *spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti*, *Grosmannia galeiformis*, *Quambalaria eucalypti*, and *Teratosphaeria destructans*. *IMA Fungus* **9,** 401-418.

Wisler GC, Norris, RF (2005) Interactions between weeds and cultivated plants as related to management of plant pathogens. *Weed Science* **53,** 914-917

Wood G, Coley-Smith JR (1980) Observations on the prevalence and incidence of ergot disease in Great Britain with special reference to open-flowering male-sterile cereals. *Annals of Applied Biology* **95,** 41-46.

Wood G, Coley-Smith JR (1982) Epidemiology of ergot disease (*Claviceps purpurea*) in open-flowering male-sterile cereals. *Annals of Applied Biology* **100**, 73-82.

Wu B, Macielog AI, Hao W (2017) Origin and spread of spliceosomal introns: insights from the fungal clade *Zymoseptoria*. *Genome Biology and Evolution* **9,** 2658-2667.

Xia C, Christensen MJ, Zhang X, Nan Z (2018) Effect of *Epichloë gansuensis* endophyte transgenerational effects on the water use efficiency, nutrient and biomass accumulation of *Achnatherum inebrians* under soil water deficit. *Plant Soil* **424,** 555-571.

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35,** W265–268.

Yang Z (2007) PAML4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24,** 1586-1591

Yang Z, Neilsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17,** 32-34.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28,** 3326-3328.

Appendix 1: Friend or Enemy? Greenhouse Inoculations reveal that *Claviceps purpurea* is a "Frenemy" with its Host

Supplemental Figures and Tables for Chapter 2

Figure A1.1: Example of barley stage in which inoculations of *Claviceps purpurea* were performed.

Figure A1.2: Example of wheat stage in which inoculations of *Claviceps purpurea* were performed.

Table A1.1: Mixed model results for Dataset 1 using Model 1 of plant-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.502	0.196	-19.1
Parameter Estimates							
Intercept	0.6310	0.1173	5.380	0.023			
Inoc. day	-0.0093	0.0038	-2.440	0.019			
Inf%[.05-0]	0.0230	0.0318	0.720	0.470			
Inf%[.15-.05]	-0.0511	0.0369	-1.390	0.167			
Inf%[.25-.15]	-0.0019	0.0395	-0.050	0.963			
(Inoc. day-7.53)*Inf%[.05-0]	0.0043	0.0061	0.720	0.475			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0021	0.0071	0.300	0.765			
(Inoc. day-7.53)*Inf%[.25-.15]	-0.0010	0.0071	-0.140	0.893			
Fixed Effects							
Inoc. day			5.973	0.019			
Inf%			0.913	0.435			
Inoc. day*Inf%			0.657	0.579			
Plant Germination					0.334	0.296	212.7
Parameter Estimates							
Intercept	0.7336	0.1105	6.640	0.006			
Inoc. day	-0.0062	0.0063	-0.980	0.330			
Inf%[.05-0]	-0.0309	0.0496	-0.620	0.534			
Inf%[.15-.05]	0.0270	0.0570	0.470	0.636			
Inf%[.25-.15]	-0.0724	0.0618	-1.170	0.243			
(Inoc. day-7.23)*Inf%[.05-0]	0.0177	0.0094	1.880	0.062			
(Inoc. day-7.23)*Inf%[.15-.05]	-0.0010	0.0109	-0.090	0.929			
(Inoc. day-7.23)*Inf%[.25-.15]	-0.0136	0.0111	-1.220	0.222			
Fixed Effects							
Inoc. day			0.966	0.330			
Inf%			0.724	0.539			
Inoc. day*Inf%			2.014	0.112			
Root Biomass					0.119	0.104	-405.6
Parameter Estimates							
Intercept	0.2011	0.0167	12.080	<.001			
Inoc. day	-0.0021	0.0017	-1.240	0.221			
Inf%[.05-0]	0.0154	0.0169	0.910	0.361			
Inf%[.15-.05]	-0.0121	0.0196	-0.620	0.537			
Inf%[.25-.15]	0.0042	0.0209	0.200	0.841			
(Inoc. day-7.53)*Inf%[.05-0]	0.0006	0.0032	0.180	0.860			
(Inoc. day-7.53)*Inf%[.15-.05]	-0.0005	0.0038	-0.140	0.888			
(Inoc. day-7.53)*Inf%[.25-.15]	0.0077	0.0038	2.040	0.042			
Fixed Effects							
Inoc. day			1.528	0.221			
Inf%			0.294	0.830			
Inoc. day*Inf%			2.153	0.094			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.124	0.006	-1953.7
Parameter Estimates							
Intercept	0.0373	0.0015	24.670	<.001			
Inoc. day	-0.0001	0.0001	-1.380	0.173			
Inf%[.05-0]	0.0001	0.0010	0.130	0.895			
Inf%[.15-.05]	-0.0003	0.0011	-0.300	0.761			
Inf%[.25-.15]	-0.0010	0.0012	-0.790	0.428			
(Inoc. day-7.23)*Inf%[.05-0]	0.00004	0.0002	0.200	0.838			
(Inoc. day-7.23)*Inf%[.15-.05]	0.00003	0.0002	0.150	0.881			
(Inoc. day-7.23)*Inf%[.25-.15]	0.0002	0.0002	0.730	0.466			
Fixed Effects							
Inoc. Day			1.891	0.173			
Inf%			0.465	0.707			
Inoc. Day*Inf%			0.504	0.680			
Avg. Seed/Tiller					0.472	4.440	1813.8
Parameter Estimates							
Intercept	12.9546	2.4928	5.200	0.024			
Inoc. day	-0.1697	0.0852	-1.990	0.053			
Inf%[.05-0]	0.3792	0.7197	0.530	0.599			
Inf%[.15-.05]	-0.9844	0.8343	-1.180	0.239			
Inf%[.25-.15]	-0.0286	0.8936	-0.030	0.975			
(Inoc. day-7.53)*Inf%[.05-0]	0.1554	0.1369	1.130	0.257			
(Inoc. day-7.53)*Inf%[.15-.05]	-0.0362	0.1613	-0.220	0.823			
(Inoc. day-7.53)*Inf%[.25-.15]	-0.0260	0.1615	-0.160	0.872			
Fixed Effects							
Inoc. day			3.967	0.053			
Inf%			0.694	0.556			
Inoc. day*Inf%			0.639	0.591			
Plant Biomass					0.267	0.782	788.1
Parameter Estimates							
Intercept	1.8102	0.2613	6.930	0.004			
Inoc. day	0.0040	0.0152	0.260	0.795			
Inf%[.05-0]	0.0138	0.1277	0.110	0.914			
Inf%[.15-.05]	-0.00865	0.1476	-0.060	0.953			
Inf%[.25-.15]	-0.0784	0.1574	-0.500	0.619			
(Inoc. day-7.53)*Inf%[.05-0]	-0.0199	0.0241	-0.820	0.411			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0038	0.0284	0.130	0.893			
(Inoc. day-7.53)*Inf%[.25-.15]	0.0199	0.0285	0.700	0.485			
Fixed Effects							
Inoc. day			0.068	0.795			
Inf%			0.125	0.945			
Inoc. day*Inf%			0.391	0.759			

† t ratio for parameter estimates, F ratio for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table A1.2: Mixed model results for Dataset 1 using Model 1 of tiller-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.644	0.203	9.7
Parameter Estimates							
Intercept	0.5106	0.1609	3.170	0.075			
Inoc. day	-0.0071	0.0044	-1.620	0.113			
Inf%[.05-0]	-0.0107	0.0330	-0.330	0.745			
Inf%[.15-.05]	-0.0622	0.0382	-1.630	0.104			
Inf%[.25-.15]	-0.0556	0.0409	-1.360	0.176			
(Inoc. day-7.53)*Inf%[.05-0]	0.0072	0.0063	1.150	0.253			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0050	0.0074	0.680	0.499			
(Inoc. day-7.53)*Inf%[.25-.15]	-0.0043	0.0074	-0.580	0.564			
Fixed Effects							
Inoc. day			2.626	0.113			
Inf%			5.356	0.001			
Inoc. day*Inf%			1.850	0.138			
Avg. 1° Tiller Seed Weight					0.110	0.006	-1589.5
Parameter Estimates							
Intercept	0.0419	0.0020	21.460	0.000			
Inoc. day	-0.0003	0.0001	-2.540	0.012			
Inf%[.05-0]	-0.00003	0.0011	-0.030	0.976			
Inf%[.15-.05]	-0.0005	0.0013	-0.350	0.727			
Inf%[.25-.15]	-0.0012	0.0015	-0.780	0.436			
(Inoc. day-6.85)*Inf%[.05-0]	0.00005	0.0002	0.250	0.802			
(Inoc. day-6.85)*Inf%[.15-.05]	0.0001	0.0002	0.600	0.550			
(Inoc. day-6.85)*Inf%[.25-.15]	0.00002	0.0003	0.060	0.952			
Fixed Effects							
Inoc. day			6.426	0.012			
Inf%			0.546	0.651			
Inoc. day*Inf%			0.550	0.649			
1° Tiller Germination					0.032	0.261	105.6
Parameter Estimates							
Intercept	0.9166	0.0491	18.660	<.0001			
Inoc. day	-0.0083	0.0043	-1.930	0.056			
Inf%[.05-0]	-0.0264	0.0467	-0.570	0.572			
Inf%[.15-.05]	0.0174	0.0539	0.320	0.747			
Inf%[.25-.15]	0.0027	0.0628	0.040	0.966			
(Inoc. day-6.85)*Inf%[.05-0]	0.0090	0.0084	1.070	0.286			
(Inoc. day-6.85)*Inf%[.15-.05]	-0.0032	0.0102	-0.310	0.756			
(Inoc. day-6.85)*Inf%[.25-.15]	0.0061	0.0108	0.560	0.573			
Fixed Effects							
Inoc. day			3.708	0.056			
Inf%			0.107	0.956			
Inoc. day*Inf%			0.826	0.481			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.399	0.199	-6.9
Parameter Estimates							
Intercept	0.6970	0.1078	6.410	0.016			
Inoc. day	0.0005	0.0039	0.130	0.897			
Inf%[.05-0]	-0.0019	0.0350	-0.050	0.958			
Inf%[.15-.05]	0.0319	0.0409	0.780	0.436			
Inf%[.25-.15]	0.0409	0.0451	0.910	0.365			
(Inoc. day-6.58)*Inf%[.05-0]	-0.0001	0.0076	-0.010	0.988			
(Inoc. day-6.58)*Inf%[.15-.05]	0.0044	0.0088	0.500	0.614			
(Inoc. day-6.58)*Inf%[.25-.15]	-0.0048	0.0084	-0.580	0.565			
Fixed Effects							
Inoc. day			0.017	0.897			
Inf%			1.308	0.272			
Inoc. day*Inf%			0.169	0.917			
Avg. 2° Tiller Seed Weight					0.085	0.006	-1625.7
Parameter Estimates							
Intercept	0.0347	0.0010	33.620	<.0001			
Inoc. day	-0.0004	0.0001	-3.230	0.002			
Inf%[.05-0]	-0.0005	0.0011	-0.460	0.649			
Inf%[.15-.05]	-0.0004	0.0013	-0.340	0.738			
Inf%[.25-.15]	0.0006	0.0014	0.450	0.655			
(Inoc. day-6.48)*Inf%[.05-0]	-0.0002	0.0002	-0.620	0.536			
(Inoc. day-6.48)*Inf%[.15-.05]	0.0002	0.0003	0.650	0.519			
(Inoc. day-6.48)*Inf%[.25-.15]	0.0003	0.0003	1.040	0.301			
Fixed Effects							
Inoc. day			10.441	0.002			
Inf%			0.266	0.850			
Inoc. day*Inf%			0.918	0.433			
2° Tiller Germination					0.326	0.386	319.4
Parameter Estimates							
Intercept	0.6398	0.1319	4.850	0.009			
Inoc. day	-0.0116	0.0098	-1.180	0.242			
Inf%[.05-0]	-0.0400	0.0707	-0.570	0.572			
Inf%[.15-.05]	0.0372	0.0823	0.450	0.652			
Inf%[.25-.15]	-0.0197	0.0891	-0.220	0.825			
(Inoc. day-6.48)*Inf%[.05-0]	0.0251	0.0154	1.630	0.104			
(Inoc. day-6.48)*Inf%[.15-.05]	-0.0031	0.0176	-0.180	0.861			
(Inoc. day-6.48)*Inf%[.25-.15]	-0.0204	0.0165	-1.240	0.217			
Fixed Effects							
Inoc. day			1.399	0.242			
Inf%			0.124	0.946			
Inoc. day*Inf%			1.557	0.201			

† t ratio for parameter estimates, F ratio for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

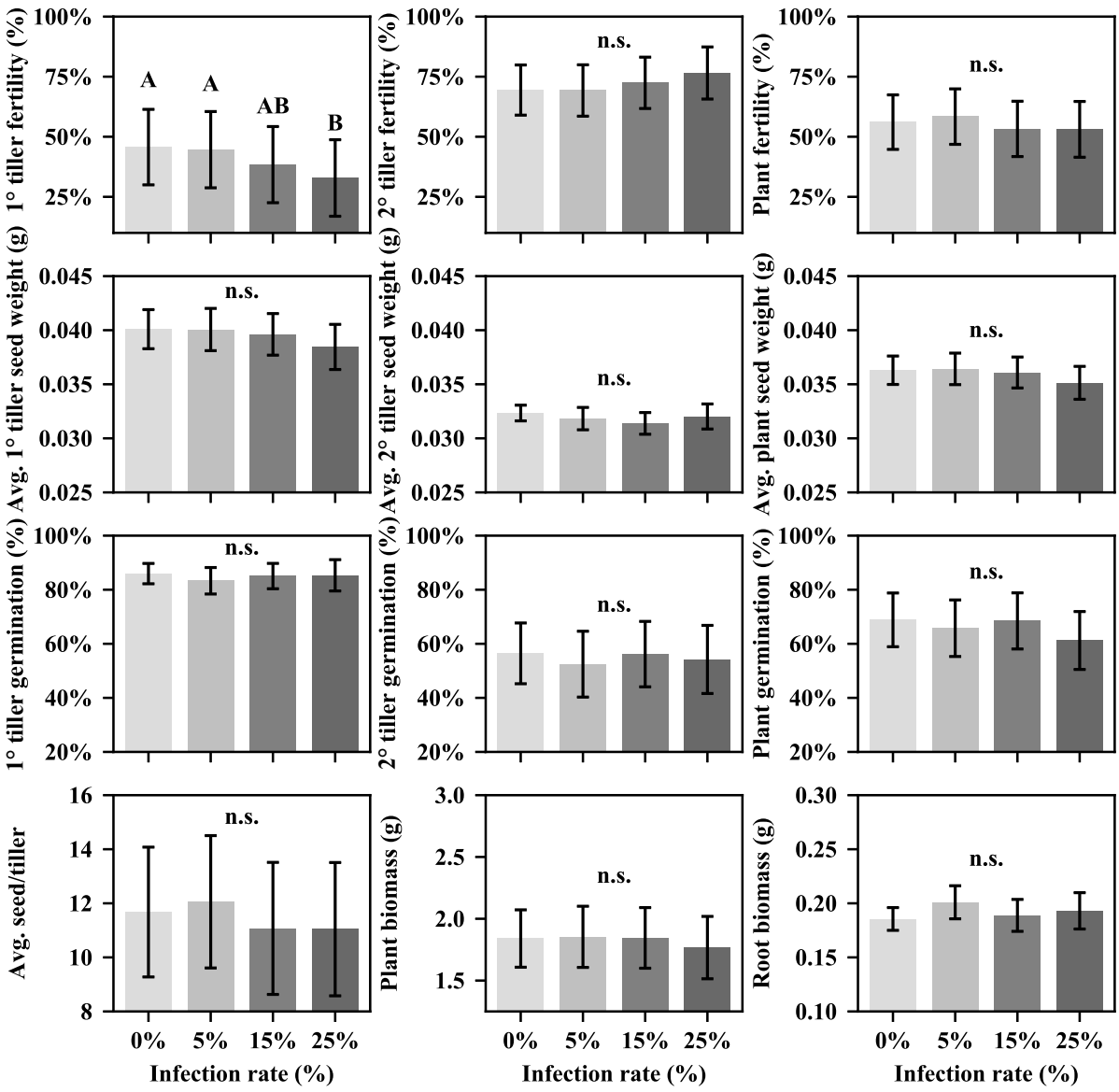


Figure A1.3: Least square means of plant responses for each infection rate in barley using Dataset 1 and Model 1. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

Table A1.3: Mixed model results for Dataset 1 using Model 2 of plant-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Plant Fertility					0.511	0.195	-21.0
Parameter Estimates							
Intercept	0.6316	0.1170	5.400	0.023			
Inoc. day	-0.0094	0.0038	-2.460	0.018			
Inf%[.05-0]	0.0448	0.0331	1.350	0.177			
Inf%[.15-.05]	-0.0156	0.0401	-0.390	0.698			
Inf%[.25-.15]	0.0178	0.0402	0.440	0.659			
(Inoc. day-7.53)*Inf%[.05-0]	0.0052	0.0060	0.860	0.392			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0033	0.0071	0.470	0.639			
(Inoc. day-7.53)*Inf%[.25-.15]	-0.0027	0.0071	-0.380	0.706			
Total Sclerotia Weight	-0.6375	0.2922	-2.180	0.030			
Fixed Effects							
Inoc. day			6.031	0.018			
Inf%			0.682	0.564			
Inoc. day*Inf%			0.985	0.400			
Total Sclerotia Weight			4.760	0.030			
Plant Germination					0.342	0.294	210.3
Parameter Estimates							
Intercept	0.7330	0.1109	6.610	0.006			
Inoc. day	-0.0062	0.0063	-0.990	0.329			
Inf%[.05-0]	-0.0628	0.0516	-1.220	0.225			
Inf%[.15-.05]	-0.0259	0.0620	-0.420	0.677			
Inf%[.25-.15]	-0.0980	0.0626	-1.570	0.119			
(Inoc. day-7.23)*Inf%[.05-0]	0.0166	0.0094	1.770	0.078			
(Inoc. day-7.23)*Inf%[.15-.05]	-0.0028	0.0109	-0.260	0.797			
(Inoc. day-7.23)*Inf%[.25-.15]	-0.0108	0.0111	-0.980	0.331			
Total Sclerotia Weight	0.9492	0.4539	2.090	0.038			
Fixed Effects							
Inoc. day			0.971	0.329			
Inf%			2.059	0.106			
Inoc. day*Inf%			1.545	0.203			
Total Sclerotia Weight			4.372	0.038			
Root Biomass					0.163	0.102	-413.6
Parameter Estimates							
Intercept	0.2008	0.0167	12.000	<.001			
Inoc. day	-0.0021	0.0017	-1.220	0.228			
Inf%[.05-0]	0.0337	0.0173	1.950	0.052			
Inf%[.15-.05]	0.0176	0.0209	0.840	0.402			
Inf%[.25-.15]	0.0210	0.0210	1.000	0.317			
(Inoc. day-7.53)*Inf%[.05-0]	0.0012	0.0031	0.390	0.696			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0006	0.0037	0.150	0.880			
(Inoc. day-7.53)*Inf%[.25-.15]	0.0063	0.0037	1.690	0.092			
Total Sclerotia Weight	-0.5344	0.1522	-3.510	0.001			
Fixed Effects							
Inoc. day			1.482	0.228			
Inf%			3.089	0.028			
Inoc. day*Inf%			2.326	0.075			
Total Sclerotia Weight			12.330	0.001			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
Avg. Plant Seed Weight					0.124	0.006	-1945.9
Parameter Estimates							
Intercept	0.0373	0.0015	24.740	<.001			
Inoc. day	-0.0001	0.0001	-1.400	0.167			
Inf%[.05-0]	0.0005	0.0010	0.520	0.602			
Inf%[.15-.05]	0.0003	0.0012	0.280	0.778			
Inf%[.25-.15]	-0.0006	0.0012	-0.500	0.619			
(Inoc. day-7.23)*Inf%[.05-0]	0.00005	0.0002	0.280	0.777			
(Inoc. day-7.23)*Inf%[.15-.05]	0.00006	0.0002	0.270	0.788			
(Inoc. day-7.23)*Inf%[.25-.15]	0.0001	0.0002	0.550	0.581			
Total Sclerotia Weight	-0.0122	0.0088	-1.390	0.166			
Fixed Effects							
Inoc. day			1.948	0.167			
Inf%			0.238	0.870			
Inoc. day*Inf%			0.553	0.647			
Total Sclerotia Weight			1.932	0.166			
Avg. Seed/Tiller					0.482	4.401	1805.1
Parameter Estimates							
Intercept	12.9704	2.4830	5.220	0.024			
Inoc. day	-0.1720	0.0855	-2.010	0.051			
Inf%[.05-0]	0.8923	0.7480	1.190	0.234			
Inf%[.15-.05]	-0.1424	0.9063	-0.160	0.875			
Inf%[.25-.15]	0.4365	0.9093	0.480	0.632			
(Inoc. day-7.53)*Inf%[.05-0]	0.1745	0.1360	1.280	0.201			
(Inoc. day-7.53)*Inf%[.15-.05]	-0.0073	0.1605	-0.050	0.964			
(Inoc. day-7.53)*Inf%[.25-.15]	-0.0667	0.1611	-0.410	0.679			
Total Sclerotia Weight	-15.0637	6.6014	-2.280	0.023			
Fixed Effects							
Inoc. day			4.043	0.051			
Inf%			0.615	0.606			
Inoc. day*Inf%			0.953	0.415			
Total Sclerotia Weight			5.207	0.023			
Plant Biomass					0.304	0.764	779.2
Parameter Estimates							
Intercept	1.8119	0.2659	6.810	0.004			
Inoc. day	0.0038	0.0160	0.240	0.814			
Inf%[.05-0]	0.1369	0.1311	1.040	0.298			
Inf%[.15-.05]	0.1839	0.1582	1.160	0.246			
Inf%[.25-.15]	0.0284	0.1582	0.180	0.858			
(Inoc. day-7.53)*Inf%[.05-0]	-0.0151	0.0237	-0.640	0.524			
(Inoc. day-7.53)*Inf%[.15-.05]	0.0095	0.0279	0.340	0.735			
(Inoc. day-7.53)*Inf%[.25-.15]	0.0104	0.0280	0.370	0.711			
Total Sclerotia Weight	-3.5241	1.1504	-3.060	0.002			
Fixed Effects							
Inoc. day			0.056	0.814			
Inf%			1.535	0.206			
Inoc. day*Inf%			0.184	0.908			
Total Sclerotia Weight			9.385	0.002			

† t ratio for parameter estimates, F ratio for Fixed effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

Table A1.4: Mixed model results for Dataset 1 using Model 2 of tiller-level responses in barley.

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
1° Tiller Fertility					0.645	0.203	9.9
Parameter Estimates							
Intercept	0.5111	0.1607	3.180	0.075			
Inoc. day	-0.0072	0.0044	-1.650	0.107			
Inf%[.05-0]	0.0053	0.0345	0.150	0.879			
Inf%[.15-.05]	-0.0357	0.0418	-0.850	0.394			
Inf%[.25-.15]	-0.0409	0.0419	-0.970	0.331			
(Inoc. day-7.530)*Inf%[.05-0]	0.0077	0.0063	1.230	0.221			
(Inoc. day-7.530)*Inf%[.15-.05]	0.0059	0.0074	0.800	0.422			
(Inoc. day-7.530)*Inf%[.25-.15]	-0.0055	0.0074	-0.730	0.464			
Total Sclerotia Weight	-0.4740	0.3054	-1.550	0.122			
Fixed Effects							
Inoc. day			2.716	0.107			
Inf%			0.843	0.471			
Inoc. day*Inf%			2.194	0.089			
Total Sclerotia Weight			2.410	0.122			
Avg. 1° Tiller Seed Weight					0.109	0.006	-1580.5
Parameter Estimates							
Intercept	0.0419	0.0019	21.510	0.000			
Inoc. day	-0.0003	0.0001	-2.550	0.011			
Inf%[.05-0]	0.0002	0.0012	0.210	0.833			
Inf%[.15-.05]	0.00000	0.0014	0.000	0.997			
Inf%[.25-.15]	-0.0009	0.0015	-0.560	0.573			
(Inoc. day-6.851)*Inf%[.05-0]	0.00006	0.0002	0.300	0.767			
(Inoc. day-6.851)*Inf%[.15-.05]	0.0002	0.0002	0.640	0.521			
(Inoc. day-6.851)*Inf%[.25-.15]	0.00000	0.0003	0.000	0.999			
Total Sclerotia Weight	-0.0081	0.0098	-0.830	0.407			
Fixed Effects							
Inoc. day			6.515	0.011			
Inf%			0.127	0.944			
Inoc. day*Inf%			0.615	0.606			
Total Sclerotia Weight			0.689	0.407			
1° Tiller Germination					0.029	0.262	107.6
Parameter Estimates							
Intercept	0.9159	0.0496	18.480	<.0001			
Inoc. day	-0.0083	0.0043	-1.910	0.058			
Inf%[.05-0]	-0.0335	0.0488	-0.690	0.493			
Inf%[.15-.05]	0.0062	0.0585	0.110	0.916			
Inf%[.25-.15]	-0.0047	0.0647	-0.070	0.942			
(Inoc. day-6.851)*Inf%[.05-0]	0.0087	0.0084	1.040	0.299			
(Inoc. day-6.851)*Inf%[.15-.05]	-0.0034	0.0102	-0.340	0.737			
(Inoc. day-6.851)*Inf%[.25-.15]	0.0064	0.0108	0.600	0.552			
Total Sclerotia Weight	0.2061	0.4118	0.500	0.617			
Fixed Effects							
Inoc. day			3.632	0.058			
Inf%			0.176	0.913			
Inoc. day*Inf%			0.785	0.503			
Total Sclerotia Weight			0.250	0.617			

	Estimate	Std Error	t/F† Ratio	*P* - value	Adj. R^2	RMSE‡	AICc§
2° Tiller Fertility					0.397	0.199	-4.9
Parameter Estimates							
Intercept	0.6910	0.1074	6.430	0.016			
Inoc. day	0.0005	0.0039	0.120	0.902			
Inf%[.05-0]	-0.0092	0.0369	-0.250	0.804			
Inf%[.15-.05]	0.0204	0.0448	0.460	0.648			
Inf%[.25-.15]	0.0326	0.0469	0.690	0.489			
(Inoc. day-6.578)*Inf%[.05-0]	-0.0003	0.0076	-0.040	0.969			
(Inoc. day-6.578)*Inf%[.15-.05]	0.0037	0.0089	0.410	0.680			
(Inoc. day-6.578)*Inf%[.25-.15]	-0.0038	0.0085	-0.450	0.654			
Total Sclerotia Weight	0.2236	0.3498	0.640	0.523			
Fixed Effects							
Inoc. day			0.015	0.902			
Inf%			0.318	0.812			
Inoc. day*Inf%			0.101	0.960			
Total Sclerotia Weight			0.409	0.523			
Avg. 2° Tiller Seed Weight					0.097	0.006	-1618.4
Parameter Estimates							
Intercept	0.0348	0.0010	34.700	<.0001			
Inoc. day	-0.0004	0.0001	-3.230	0.002			
Inf%[.05-0]	0.00001	0.0012	0.010	0.996			
Inf%[.15-.05]	0.0004	0.0014	0.260	0.793			
Inf%[.25-.15]	0.0012	0.0015	0.820	0.411			
(Inoc. day-6.477)*Inf%[.05-0]	-0.0001	0.0002	-0.590	0.556			
(Inoc. day-6.477)*Inf%[.15-.05]	0.0002	0.0003	0.850	0.397			
(Inoc. day-6.477)*Inf%[.25-.15]	0.0002	0.0003	0.770	0.442			
Total Sclerotia Weight	-0.0159	0.0110	-1.450	0.149			
Fixed Effects							
Inoc. day			10.452	0.002			
Inf%			0.326	0.807			
Inoc. day*Inf%			0.928	0.428			
Total Sclerotia Weight			2.094	0.149			
2° Tiller Germination					0.322	0.387	320.3
Parameter Estimates							
Intercept	0.6398	0.1313	4.870	0.009			
Inoc. day	-0.0117	0.0098	-1.190	0.240			
Inf%[.05-0]	-0.0509	0.0747	-0.680	0.496			
Inf%[.15-.05]	0.0209	0.0899	0.230	0.816			
Inf%[.25-.15]	-0.0317	0.0928	-0.340	0.733			
(Inoc. day-6.477)*Inf%[.05-0]	0.0247	0.0154	1.600	0.110			
(Inoc. day-6.477)*Inf%[.15-.05]	-0.0040	0.0177	-0.230	0.820			
(Inoc. day-6.477)*Inf%[.25-.15]	-0.0190	0.0168	-1.130	0.259			
Total Sclerotia Weight	0.3216	0.7022	0.460	0.647			
Fixed Effects							
Inoc. day			1.415	0.240			
Inf%			0.189	0.904			
Inoc. day*Inf%			1.387	0.248			
Total Sclerotia Weight			0.210	0.647			

† t ratio for parameter estimates, F ratio for Fixed Effects

‡ RMSE = Root mean squared error

§ AICc = Akaike information criterion with small sample correlation

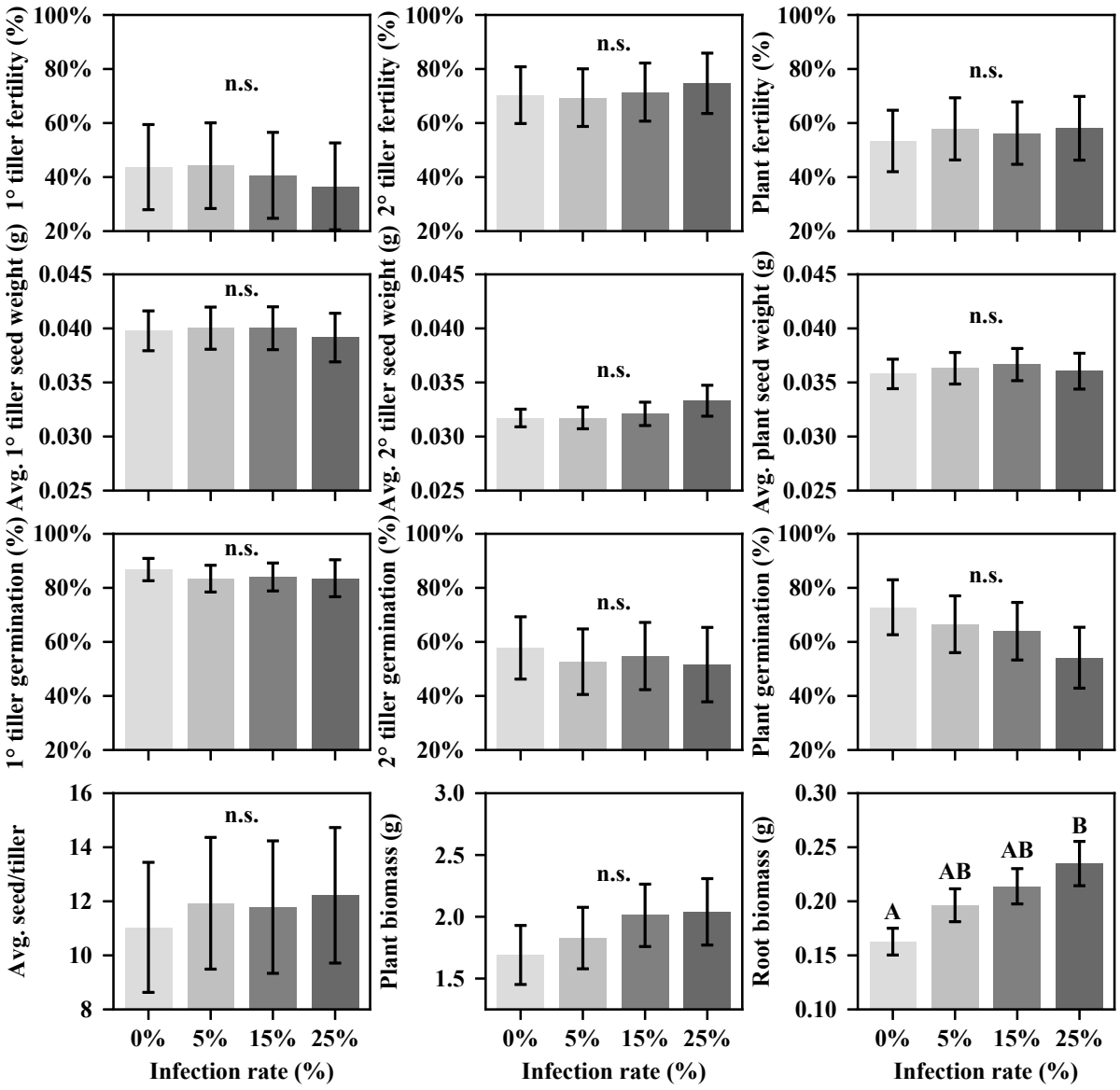


Figure A1.4: Least square means of plant responses for each infection rate in barley using Dataset 1 and Model 2. Tukey's HSD was used to test for significance ($P \leq 0.05$) of mean pairwise comparisons; letters not connected by the same letter are significant and n.s. indicates no significant difference for any pairwise comparison. Bars represent standard error.

Table A1.5: REML variance component results for Dataset 1 using Model 1 and 2 of plant-level responses in barley.

Model 1

	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total
Plant Fertility						
Random Effect						
Trial	0.96394	0.03711	0.03814	0.00982	1.75315	45.37
Position[Trial]	0.16062	0.00618	0.00306	0.00284	0.02236	7.56
Residual		0.03850	0.00337	0.03265	0.04607	47.070
Total		0.08179	0.03834	0.03882	0.27026	100
Avg. Plant Seed Weight						
Random Effect						
Trial	0.12556	4.23E-06	4.79E-06	1.03E-06	0.00041	10.76
Position[Trial]	0.04175	1.41E-06	1.49E-06	3.61E-07	0.00008	3.57
Residual		0.00003	3.05E-06	0.00003	0.00004	85.67
Total		0.00004	5.65E-06	0.00003	0.00005	100
Plant Germination						
Random Effect						
Trial	0.29502	0.02588	0.02854	0.00645	2.02842	19.93
Position[Trial]	0.18507	0.01624	0.00762	0.00770	0.05372	12.51
Residual		0.08774	0.00798	0.07398	0.10575	67.56
Total		0.12986	0.03005	0.08646	0.21675	100
Avg. Seed/Tiller						
Random Effect						
Trial	0.84014	16.55992	17.06848	4.37346	796.74053	42.20
Position[Trial]	0.15098	2.97586	1.50263	1.34934	11.10917	7.58
Residual		19.71085	1.72601	16.72153	23.58472	50.22
Total		39.24663	17.17756	19.40223	117.24290	100
Root Biomass						
Random Effect						
Trial	0.22697	0.13873	0.15491	0.03416	12.01832	16.43
Position[Trial]	0.15478	0.09461	0.04482	0.04463	0.31772	11.20
Residual		0.61124	0.05333	0.51885	0.73087	72.37
Total		0.84459	0.16673	0.59377	1.29697	100
Plant Biomass						
Random Effect						
Trial	0	0	0	0	0	0
Position[Trial]	0.08857	0.00097	0.00060	0.00038	0.00553	8.14
Residual		0.01092	0.00095	0.00927	0.01306	91.86
Total		0.01189	0.00102	0.01012	0.01417	100

Model 2

	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total
Plant Fertility						
Random Effect						
Trial	0.97225	0.03682	0.03788	0.00974	1.74942	45.36
Position[Trial]	0.17093	0.00647	0.00311	0.00303	0.02228	7.98
Residual		0.03787	0.00332	0.03212	0.04533	46.6
Total		0.08116	0.03808	0.03851	0.26854	100
Avg. Plant Seed Weight						
Random Effect						
Trial	0.126047	4.24E-06	4.78E-06	1.03E-06	0.00040	10.82
Position[Trial]	0.038155	1.28E-06	1.46E-06	3.10E-07	0.00013	3.28
Residual		0.00003	3.06E-06	0.00003	0.00004	85.89
Total		0.00004	5.65E-06	0.00003	0.00005	100
Plant Germination						
Random Effect						
Trial	0.30315	0.02627	0.02889	0.00656	2.01708	20.37
Position[Trial]	0.18458	0.01599	0.00751	0.00759	0.05293	12.41
Residual		0.08665	0.00790	0.07304	0.10448	67.21
Total		0.12892	0.03034	0.08530	0.21734	100
Avg. Seed/Tiller						
Random Effect						
Trial	0.84660	16.39535	16.91356	4.32673	793.46243	42.19
Position[Trial]	0.16026	3.10368	1.52360	1.43236	11.04289	7.98
Residual		19.36607	1.69818	16.42543	23.17823	49.83
Total		38.86510	17.02332	19.20528	116.22435	100
Root Biomass						
Random Effect						
Trial	0.23800	0.13908	0.15769	0.03379	13.63786	16.44
Position[Trial]	0.20908	0.12218	0.05259	0.06098	0.35674	14.45
Residual		0.58436	0.05119	0.49571	0.69925	69.10
Total		0.84562	0.17051	0.59043	1.31188	100
Plant Biomass						
Random Effect						
Trial	0	0	0	0	0	0
Position[Trial]	0.10662	0.00111	0.00063	0.00047	0.00510	9.63
Residual		0.01042	0.00091	0.00884	0.01247	90.37
Total		0.01153	0.00100	0.00979	0.01379	100

Table A1.6: REML variance component results for Dataset 1 using Model 1 and 2 of tiller-level responses in barley.

Model 1

	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total
1° Tiller Fertility						
Random Effect						
Trial	1.75264	0.07213	0.07363	0.01921	3.25449	58.35
Position[Trial]	0.25067	0.01032	0.00416	0.00534	0.02771	8.35
Residual		0.04116	0.00359	0.03494	0.04921	33.29
Total		0.12361	0.07376	0.05018	0.64541	100
2° Tiller Fertility						
Random Effect						
Trial	0.79252	0.03137	0.03219	0.00832	1.46524	42.65
Position[Trial]	0.06579	0.00260	0.00212	0.00084	0.03584	3.54
Residual		0.03958	0.00382	0.03304	0.04829	53.81
Total		0.07356	0.03242	0.03622	0.22188	100
Avg. 1° Tiller Seed Weight						
Random Effect						
Trial	0.22758	8.75E-06	9.34E-06	2.24E-06	0.00054	18.54
Position[Trial]	0	0	0	0	0	0
Residual		0.00004	3.62E-06	0.00003	0.00005	81.46
Total		0.00005	0.00001	0.00003	0.00008	100
Avg. 2° Tiller Seed Weight						
Random Effect						
Trial	0.01063	4.28E-07	1.19E-06	4.77E-08	2.15E+05	1.04
Position[Trial]	0.00904	3.64E-07	1.58E-06	3.23E-08	5.95E+22	0.89
Residual		0.00004	3.95E-06	0.00003	0.00005	98.07
Total		0.00004	3.88E-06	0.00003	0.00005	100
1° Tiller Germination						
Random Effect						
Trial	0.03496	0.00238	0.00347	0.00046	3.64204	3.38
Position[Trial]	0	0	0	0	0	0
Residual		0.06821	0.00643	0.05717	0.08281	96.62
Total		0.07060	0.00722	0.05834	0.08719	100
2° Tiller Germination						
Random Effect						
Trial	0.19817	0.02959	0.03596	0.00676	5.38932	13.69
Position[Trial]	0.24866	0.03713	0.01735	0.01766	0.12212	17.18
Residual		0.14930	0.01489	0.12393	0.18339	69.12
Total		0.21602	0.04056	0.15428	0.32413	100

Model 2

	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total
1° Tiller Fertility						
Random Effect						
Trial	1.75445	0.07200	0.07347	0.01918	3.23944	58.53
Position[Trial]	0.24321	0.00998	0.00407	0.00513	0.02720	8.11
Residual		0.04104	0.00359	0.03483	0.04909	33.35
Total		0.12302	0.07359	0.04985	0.64600	100
2° Tiller Fertility						
Random Effect						
Trial	0.78420	0.03112	0.03195	0.00825	1.45674	42.40
Position[Trial]	0.06551	0.00260	0.00213	0.00083	0.03654	3.54
Residual		0.03969	0.00384	0.03311	0.04844	54.06
Total		0.07341	0.03218	0.03626	0.21976	100
Avg. 1° Tiller Seed Weight						
Random Effect						
Trial	0.22609	8.70E-06	9.3E-06	2.23E-06	0.00054	18.44
Position[Trial]	0	0	0	0	0	0
Residual		0.00004	3.64E-06	0.00003	0.00005	81.56
Total		0.00005	9.97E-06	0.00003	0.00007	100
Avg. 2° Tiller Seed Weight						
Random Effect						
Trial	0.00548	2.18E-07	1.03E-06	1.91E-08	5.36E+27	0.54
Position[Trial]	0.01681	6.70E-07	1.69E-06	8.06E-08	2398.9028	1.64
Residual		0.00004	3.93E-06	0.00003	0.00005	97.82
Total		0.00004	3.83E-06	0.00003	0.00005	100
1° Tiller Germination						
Random Effect						
Trial	0.03652	0.00250	0.00360	0.00049	3.25818	3.52
Position[Trial]	0	0	0	0	0	0
Residual		0.06842	0.00647	0.05733	0.08309	96.48
Total		0.07092	0.00731	0.05852	0.08774	100
2° Tiller Germination						
Random Effect						
Trial	0.19480	0.02923	0.03556	0.00667	5.38151	13.54
Position[Trial]	0.24378	0.03658	0.01734	0.01725	0.12295	16.95
Residual		0.15005	0.01501	0.12448	0.18443	69.51
Total		0.21586	0.04023	0.15451	0.32283	100

Table A1.7: REML variance component results for Dataset 2 using Model 1 and 2 of plant-level responses in barley.

Model 1							**Model 2**						
	Var Ratio	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**		**Var Ratio**	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**
Plant Fertility							**Plant Fertility**						
Random Effect							Random Effect						
Trial	0.89140	0.03114	0.03228	-0.03213	0.09441	41.78	Trial	0.95433	0.03209	0.03321	0.00844	1.58839	43.32
Position[Trial]	0.24226	0.00846	0.00322	0.00216	0.01477	11.35	Position[Trial]	0.24886	0.00837	0.00317	0.00448	0.02085	11.29
Residual		0.03493	0.00258	0.03038	0.04059	46.87	Residual		0.03363	0.00249	0.02924	0.03908	45.39
Total		0.07453	0.03248	0.03693	0.22136	100	Total		0.07408	0.03340	0.03599	0.23086	100
Avg. Plant Seed Weight							**Avg. Plant Seed Weight**						
Random Effect							Random Effect						
Trial	0.14951	5.23E-06	5.63E-06	-5.80E-06	0.00002	12.81	Trial	0.15524	5.42E-06	5.82E-06	1.38E-06	0.00034	13.26
Position[Trial]	0.01785	6.24E-07	9.69E-07	-1.27E-06	2.52E-06	1.53	Position[Trial]	0.01581	5.53E-07	9.51E-07	9.20E-08	0.01460	1.35
Residual		0.00003	2.65E-06	0.00003	0.00004	85.66	Residual		0.00003	2.65E-06	0.00003	0.00004	85.39
Total		0.00004	6.19E-06	0.00003	0.00006	100	Total		0.00004	6.36E-06	0.00003	0.00006	100
Plant Germination							**Plant Germination**						
Random Effect							Random Effect						
Trial	0.30339	0.02761	0.02953	0.00706	1.72251	21.30	Trial	0.29907	0.02709	0.02900	0.00692	1.69878	21.08
Position[Trial]	0.12094	0.01101	0.00579	0.00486	0.04429	8.49	Position[Trial]	0.11988	0.01086	0.00573	0.00479	0.04386	8.45
Residual		0.09100	0.00697	0.07875	0.10637	70.21	Residual		0.09058	0.00695	0.07837	0.10590	70.48
Total		0.12961	0.03050	0.08577	0.21847	100	Total		0.12853	0.02997	0.08533	0.21550	100
Avg. Seed/Tiller							**Avg. Seed/Tiller**						
Random Effect							Random Effect						
Trial	0.78828	14.21466	14.78828	-14.76980	43.19916	38.97	Trial	0.83888	14.65105	15.20911	3.84545	739.87060	40.45
Position[Trial]	0.23435	4.22584	1.60038	1.089141	7.36253	11.59	Position[Trial]	0.23526	4.10880	1.55949	2.19752	10.26762	11.34
Residual		18.03253	1.32917	15.68669	20.94992	49.44	Residual		17.46493	1.28931	15.18979	20.29534	48.21
Total		36.47303	14.90024	18.74780	99.47152	100	Total		36.22477	15.31124	18.26918	103.21710	100
Root Biomass							**Root Biomass**						
Random Effect							Random Effect						
Trial	0	0	0	0	0	0	Trial	0	0	0	0	0	0
Position[Trial]	0.12907	0.00122	0.00053	0.00068	0.00297	11.43	Position[Trial]	0.12887	0.00118	0.00051	0.00059	0.00346	11.42
Residual		0.00948	0.00070	0.00843	0.01075	88.57	Residual		0.00919	0.00068	0.00800	0.01068	88.58
Total		0.01070	0.00083	0.00947	0.01221	100	Total		0.01038	0.00080	0.00897	0.01214	100
Plant Biomass							**Plant Biomass**						
Random Effect							Random Effect						
Trial	0.42172	0.16170	0.17068	-0.17283	0.49622	26.84	Trial	0.41948	0.15427	0.16380	0.03970	9.12275	26.31
Position[Trial]	0.14967	0.05739	0.02438	0.00960	0.10517	9.53	Position[Trial]	0.17448	0.06417	0.02604	0.03310	0.17358	10.95
Residual		0.38343	0.02825	0.33357	0.44543	63.64	Residual		0.36776	0.02714	0.31986	0.42735	62.74
Total		0.60251	0.17399	0.36730	1.16633	100	Total		0.58620	0.16728	0.35919	1.12451	100

Table A1.8: REML variance component results for Dataset 2 using Model 1 and 2 of tiller-level responses in barley.

Model 1							**Model 2**						
	Var Ratio	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**		**Var Ratio**	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**
1° Tiller Fertility							**1° Tiller Fertility**						
Random Effect							Random Effect						
Trial	1.70335	0.06659	0.06811	-0.06690	0.20007	56.70	Trial	1.76770	0.06786	0.06930	0.01807	3.07081	57.84
Position[Trial]	0.30085	0.01176	0.00409	0.00374	0.01978	10.01	Position[Trial]	0.28837	0.01107	0.00390	0.00616	0.02552	9.44
Residual		0.03909	0.00288	0.03402	0.04540	33.29	Residual		0.03839	0.00283	0.03340	0.04460	32.72
Total		0.11744	0.06822	0.04857	0.57712	100	Total		0.11732	0.06940	0.04791	0.60061	100
2° Tiller Fertility							**Tiller Fertility**						
Random Effect							Random Effect						
Trial	0.68233	0.02586	0.02646	-0.02599	0.07772	39.40	Trial	0.68699	0.02607	0.02668	0.00693	1.19411	39.53
Position[Trial]	0.04930	0.00187	0.00155	-0.00118	0.00491	2.85	Position[Trial]	0.05088	0.00193	0.00158	0.00062	0.02713	2.93
Residual		0.03790	0.00311	0.03249	0.04481	57.75	Residual		0.03795	0.00312	0.03252	0.04488	57.54
Total		0.06564	0.02664	0.03386	0.17758	100	Total		0.06596	0.02686	0.03396	0.17920	100
Avg. 1° Tiller Seed Weight							**Avg. 1° Tiller Seed Weight**						
Random Effect							Random Effect						
Trial	0.24908	9.64E-06	0.00001	3.11E-06	0.000241	19.94	Trial	0.25737	9.84E-06	0.00001	2.56E-06	0.00050	20.50
Position[Trial]	0	0	0	0	0.	0	Position[Trial]	0	0	0	0	0	0
Residual		0.00004	3.08E-06	0.00003	0.00004	80.06	Residual		0.00004	3.05E-06	0.00003	0.00005	79.53
Total		0.00005	0.00001	0.00004	0.00007	100	Total		0.00005	0.00001	0.00003	0.00008	100
Avg. 2° Tiller Seed Weight							**Avg. 2° Tiller Seed Weight**						
Random Effect							Random Effect						
Trial	0.00833	3.55E-07	8.68E-07	-1.35E-06	2.06E-06	0.83	Trial	0.00683	2.91E-07	8.17E-07	3.23E-08	0.00002	0.68
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		0.00004	3.52E-06	0.00005	0.00005	99.17	Residual		0.00004	3.40E-06	0.00004	0.00005	99.32
Total		0.00004	3.61E-06	0.00005	0.00005	100	Total		0.00004	3.45E-06	0.00004	0.00005	100
1° Tiller Germination							**1° Tiller Germination**						
Random Effect							Random Effect						
Trial	0.07024	0.00458	0.00538	0.00108	0.60364	6.56	Trial	0.07061	0.00462	0.00542	0.00109	0.60624	6.60
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		0.06523	0.00519	0.05615	0.07673	93.44	Residual		0.06542	0.00521	0.05630	0.07697	93.40
Total		0.06981	0.00743	0.05727	0.08701	100	Total		0.07004	0.00748	0.05742	0.08735	100
2° Tiller Germination							**2° Tiller Germination**						
Random Effect							Random Effect						
Trial	0.13445	0.02300	0.02685	0.00544	2.89654	10.90	Trial	0.13106	0.02249	0.02631	0.00531	2.88104	10.68
Position[Trial]	0.09947	0.01701	0.01037	0.00681	0.09339	8.06	Position[Trial]	0.09639	0.01654	0.01027	0.00653	0.09504	7.85
Residual		0.17105	0.01432	0.14613	0.20298	81.04	Residual		0.17161	0.01439	0.14656	0.20370	81.47
Total		0.21106	0.03070	0.16178	0.28699	100	Total		0.21064	0.03025	0.16197	0.28520	100

Table A1.9: REML variance component results for Dataset 2 using Model 1 and 2 of plant-level responses in wheat.

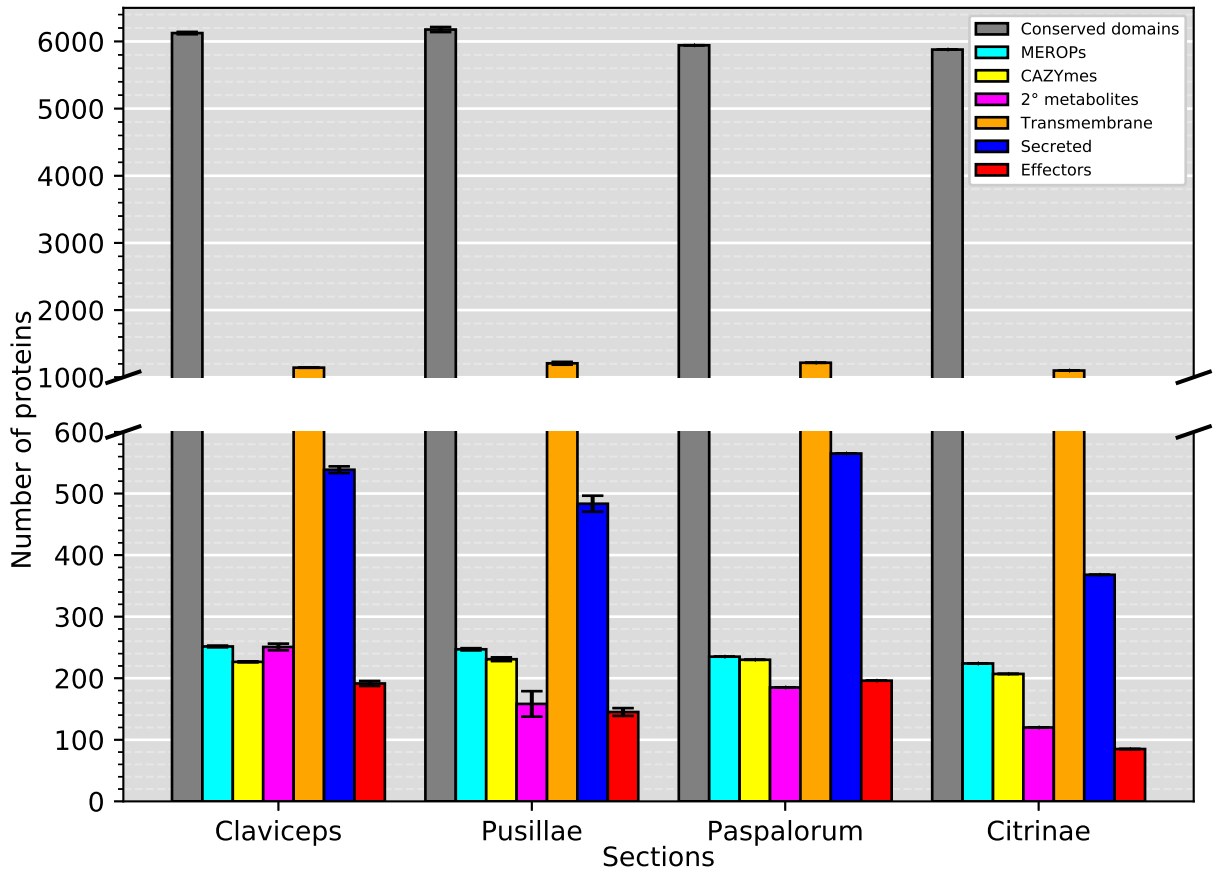
Model 1	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total	Model 2	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	% of Total
Plant Fertility							**Plant Fertility**						
Random Effect							Random Effect						
Trial	0.21981	0.00342	0.00515	0.00064	8.63353	18.02	Trial	0.23099	0.00357	0.00536	0.00067	8.62317	18.76
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		0.01556	0.00151	0.01298	0.01901	81.98	Residual		0.01545	0.00151	0.01287	0.01889	81.24
Total		0.01898	0.00537	0.01168	0.03615	100	Total		0.01902	0.00556	0.01153	0.03719	100
Avg. Plant Seed Weight							**Avg. Plant Seed Weight**						
Random Effect							Random Effect						
Trial	0.27547	1.91E-06	2.91E-06	3.54E-07	0.00589	21.38	Trial	0.26462	1.84E-06	2.82E-06	3.42E-07	0.00586	20.76
Position[Trial]	0.01284	8.89E-08	3.29E-07	8.35E-09	8.47E+13	1.00	Position[Trial]	0.00996	6.94E-08	3.29E-07	6.07E-09	4.69E+27	0.78
Residual		6.93E-06	8.64E-07	5.50E-06	8.99E-06	77.62	Residual		6.97E-06	8.76E-07	5.53E-06	9.06E-06	78.46
Total		8.92E-06	3.03E-06	5.05E-06	0.00002	100	Total		8.88E-06	2.94E-06	5.09E-06	0.00002	100
Plant Germination							**Plant Germination**						
Random Effect							Random Effect						
Trial	0.04063	0.00044	0.00085	0.00007	141.04322	3.90	Trial	0.04665	0.00050	0.00093	0.00008	68.86580	4.46
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		0.01090	0.00106	0.00908	0.01332	96.10	Residual		0.01075	0.00105	0.00896	0.01315	95.54
Total		0.01134	0.00135	0.00910	0.01452	100	Total		0.01125	0.00139	0.00895	0.01457	100
Avg. Seed/Tiller							**Avg. Seed/Tiller**						
Random Effect							Random Effect						
Trial	0	0	0	0	0	0	Trial	0	0	0	0	0	0
Position[Trial]	0.01980	0.31163	0.62508	0.04527	3.21E+05	1.94	Position[Trial]	0.02128	0.33052	0.63269	0.05003	9.77E+04	2.08
Residual		15.73812	1.71762	12.84793	19.73211	98.06	Residual		15.52940	1.70615	12.66127	19.50152	97.92
Total		16.04975	1.68298	13.20173	19.93554	100	Total		15.85992	1.67266	13.03163	19.72576	100
Root Biomass							**Root Biomass**						
Random Effect							Random Effect						
Trial	0	0	0	0	0	0	Trial	0	0	0	0	0	0
Position[Trial]	0.23720	0.00163	0.00073	0.00079	0.00508	19.17	Position[Trial]	0.24477	0.00168	0.00075	0.00082	0.00520	19.66
Residual		0.00685	0.00070	0.00567	0.00846	80.83	Residual		0.00687	0.00070	0.00567	0.00849	80.34
Total		0.00848	0.00096	0.00687	0.01074	100	Total		0.00855	0.00098	0.00691	0.01085	100
Plant Biomass							**Plant Biomass**						
Random Effect							Random Effect						
Trial	0.14663	0.03136	0.06342	0.00452	4.06E+04	9.60	Trial	0.14655	0.03159	0.06400	0.00455	4.31E+04	9.58
Position[Trial]	0.38053	0.08139	0.03367	0.04154	0.22555	24.92	Position[Trial]	0.38303	0.08256	0.03419	0.04212	0.22908	25.04
Residual		0.21389	0.02183	0.17681	0.26405	65.48	Residual		0.21555	0.02211	0.17802	0.26642	65.38
Total		0.32664	0.07189	0.22141	0.53013	100	Total		0.32971	0.07264	0.22341	0.53540	100

Table A1.10: REML variance component results for Dataset 2 using Model 1 and 2 of tiller-level responses in wheat.

Model 1							**Model 2**						
	Var Ratio	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**		**Var Ratio**	**Var Component**	**Std Error**	**95% Lower**	**95% Upper**	**% of Total**
1° Tiller Fertility							**1° Tiller Fertility**						
Random Effect							Random Effect						
Trial	0.00814	0.00014	0.00074	0.00001	9.68E+38	0.77	Trial	0.01499	0.00025	0.00088	0.00002	7.48E+14	1.40
Position[Trial]	0.05635	0.00098	0.00106	0.00025	0.06647	5.29	Position[Trial]	0.05517	0.00093	0.00104	0.00023	0.08375	5.16
Residual		0.01734	0.00180	0.01429	0.02148	93.94	Residual		0.01683	0.00176	0.01385	0.02088	93.44
Total		0.01846	0.00192	0.01521	0.02288	100	Total		0.01801	0.00193	0.01475	0.02249	100
2° Tiller Fertility							**2° Tiller Fertility**						
Random Effect							Random Effect						
Trial	0.65749	0.01046	0.01534	0.00201	17.61461	38.98	Trial	0.66219	0.01062	0.01557	0.00204	17.94419	39.10
Position[Trial]	0.02912	0.00046	0.00098	0.00006	2.63E+03	1.73	Position[Trial]	0.03124	0.00050	0.00102	0.00007	709.53845	1.85
Residual		0.01591	0.00202	0.01259	0.02076	59.29	Residual		0.01603	0.00205	0.01266	0.02096	59.05
Total		0.02684	0.01546	0.01116	0.12953	100	Total		0.02715	0.01569	0.01126	0.13198	100
Avg. 1° Tiller Seed Weight							**Avg. 1° Tiller Seed Weight**						
Random Effect							Random Effect						
Trial	0.16506	0.00000	0.00000	0.00000	0.00514	14.17	Trial	0.16021	1.43E-06	2.20E-06	2.63E-07	0.00520	13.81
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		8.93E-06	8.69E-07	7.44E-06	0.00001	85.83	Residual		8.92E-06	8.72E-07	7.43E-06	0.00001	86.19
Total		0.00001	2.42E-06	6.91E-06	0.00002	100	Total		0.00001	2.37E-06	6.92E-06	0.00002	100
Avg. 2° Tiller Seed Weight							**Avg. 2° Tiller Seed Weight**						
Random Effect							Random Effect						
Trial	0.36055	3.70E-06	5.60E-06	6.91E-07	0.01006	25.79	Trial	0.34811	3.61E-06	5.46E-06	6.74E-07	0.00993	25.20
Position[Trial]	0.03772	3.87E-07	5.33E-07	7.88E-08	0.00028	2.69	Position[Trial]	0.03317	3.44E-07	5.28E-07	6.34E-08	0.00117	2.40
Residual		0.00001	1.27E-06	8.15E-06	0.00001	71.52	Residual		1.04E-05	1.30E-06	8.22E-06	0.00001	72.40
Total		0.00001	5.73E-06	7.45E-06	0.00004	100	Total		1.43E-05	5.61E-06	7.52E-06	0.00004	100
1° Tiller Germination							**1° Tiller Germination**						
Random Effect							Random Effect						
Trial	0.00214	0.00001	0.00015	4.04E-06	8.98E+185	0.21	Trial	0.00233	0.00001	0.00015	3.36E-06	1.25E+160	0.23
Position[Trial]	0	0	0	0	0	0	Position[Trial]	0	0	0	0	0	0
Residual		0.00628	0.00061	0.00524	0.00767	99.79	Residual		0.00634	0.00062	0.00528	0.00775	99.77
Total		0.00630	0.00062	0.00524	0.00771	100	Total		0.00635	0.00063	0.00528	0.00779	100
2° Tiller Germination							**2° Tiller Germination**						
Random Effect							Random Effect						
Trial	0.08400	0.00428	0.00794	0.00067	580.43654	7.47	Trial	0.09185	0.00465	0.00833	0.00075	2.85E+02	8.19
Position[Trial]	0.04101	0.00209	0.00325	0.00038	8.51284	3.64	Position[Trial]	0.02925	0.00148	0.00316	0.00020	9.89E+03	2.61
Residual		0.05098	0.00647	0.04034	0.06647	88.89	Residual		0.05061	0.00650	0.03994	0.06621	89.20
Total		0.05735	0.01012	0.04173	0.08377	100	Total		0.05673	0.01036	0.04086	0.08410	100

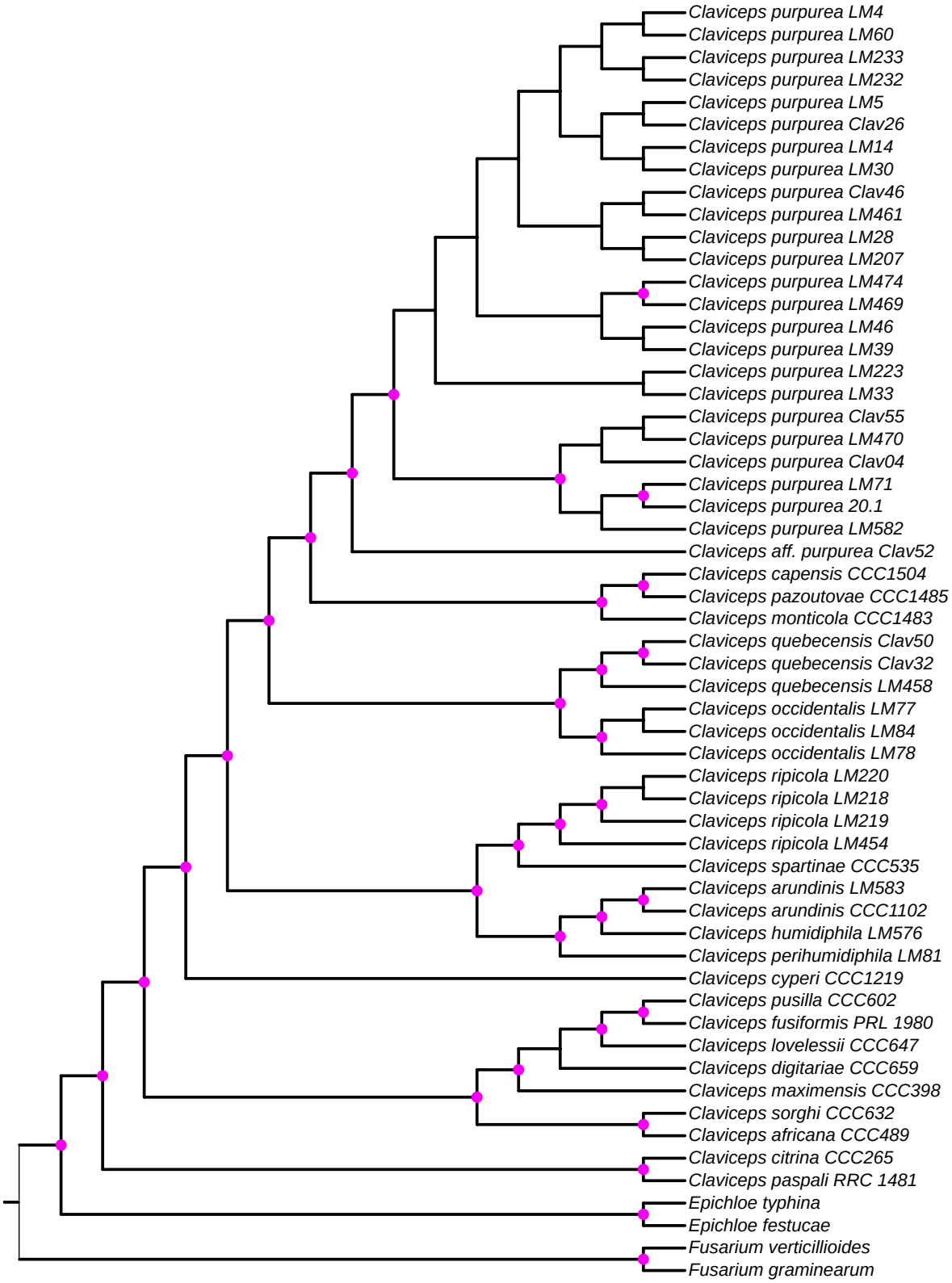
Appendix 2: Whole genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus *Claviceps* are the result of varying mechanisms driving genome evolution and host range expansion

Supplemental Figures and Tables for Chapter 3

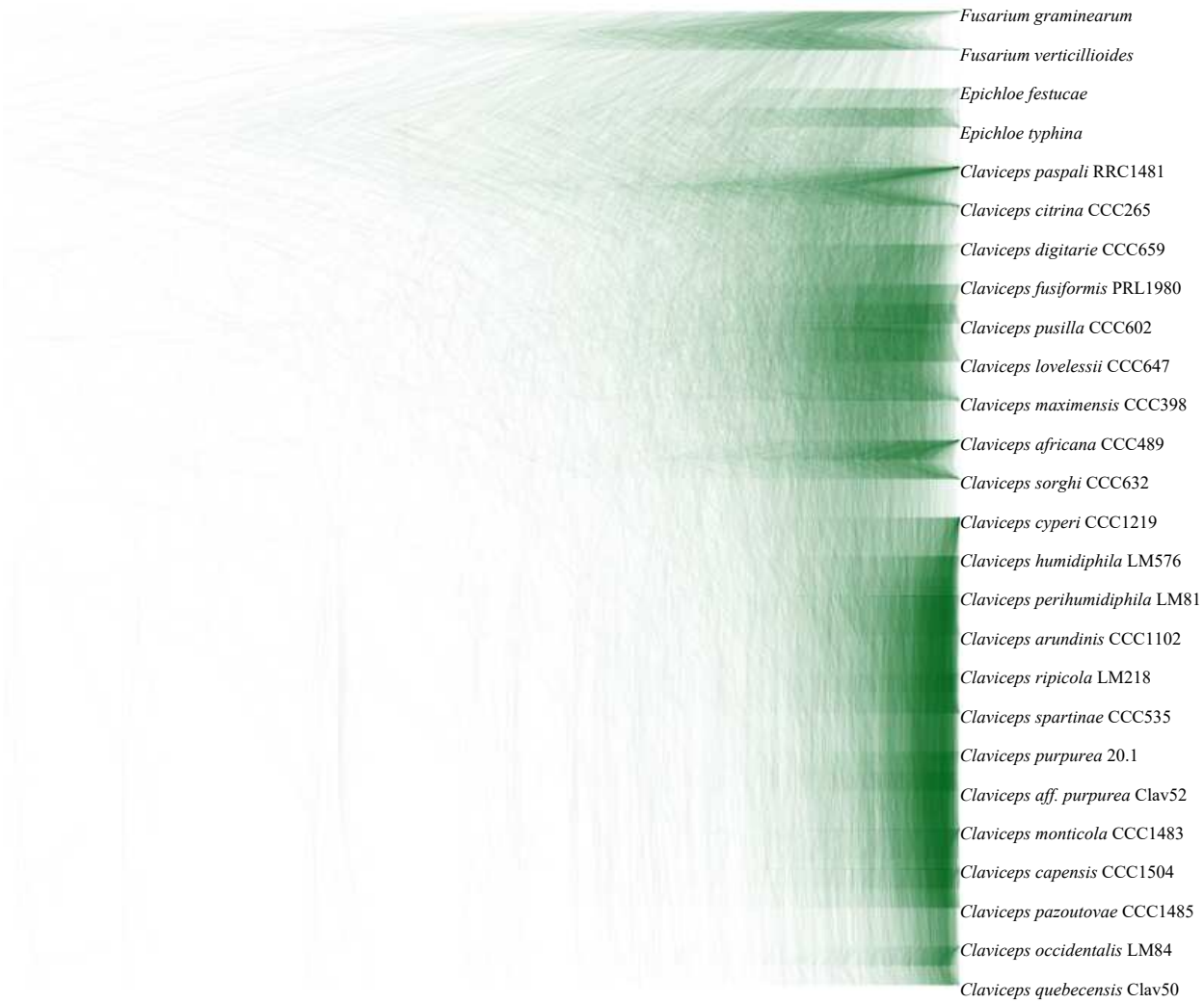


Appendix 2 Figure A2.1: Mean number of proteins in each section of the genus *Claviceps*. Bars represent standard error

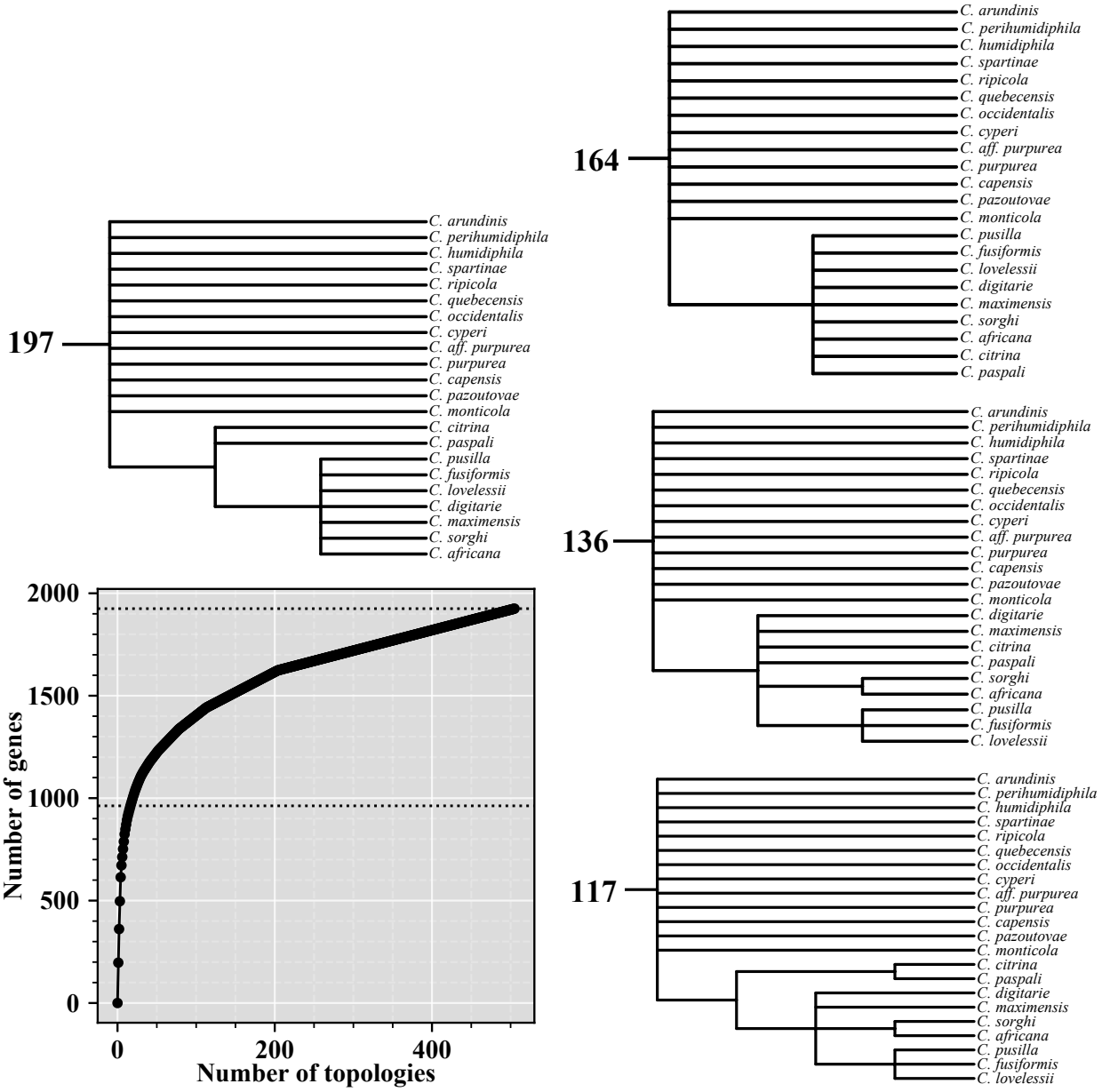
Appendix 2 Figure A2.2: Neighbor-joining phylogenetic reconstruction of the *Claviceps* genus using amino acid sequences of 2,002 single copy orthologs with 1000 bootstrap replicates. Pink dots at branches represent bootstrap values ≥ 95.



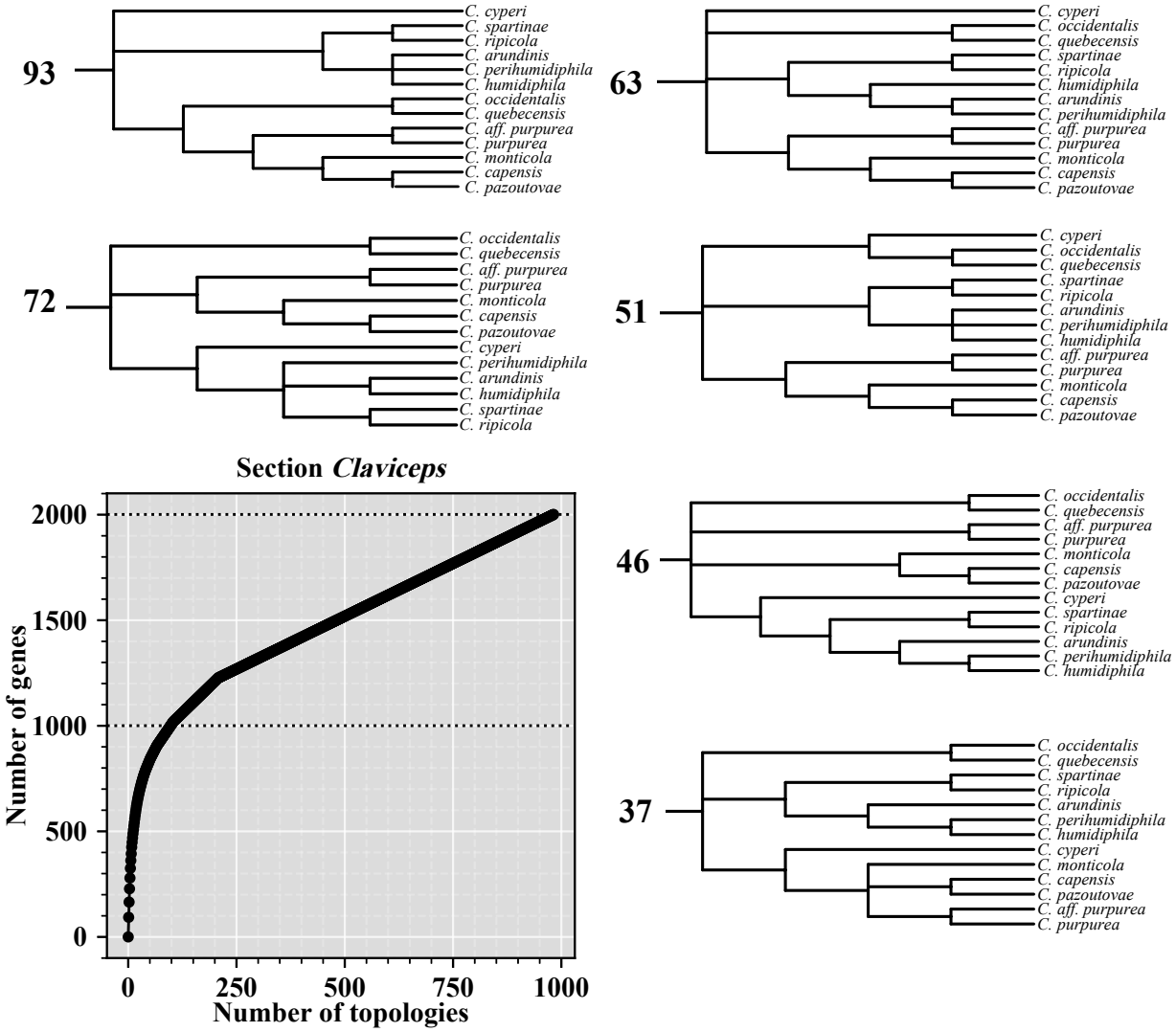
Appendix 2 Figure A2.3: Maximum parsimony phylogenetic reconstruction of the *Claviceps* genus using amino acid sequences of 2,002 single copy orthologs with 1000 bootstrap replicates. Pink dots at branches represent bootstrap values ≥ 95.



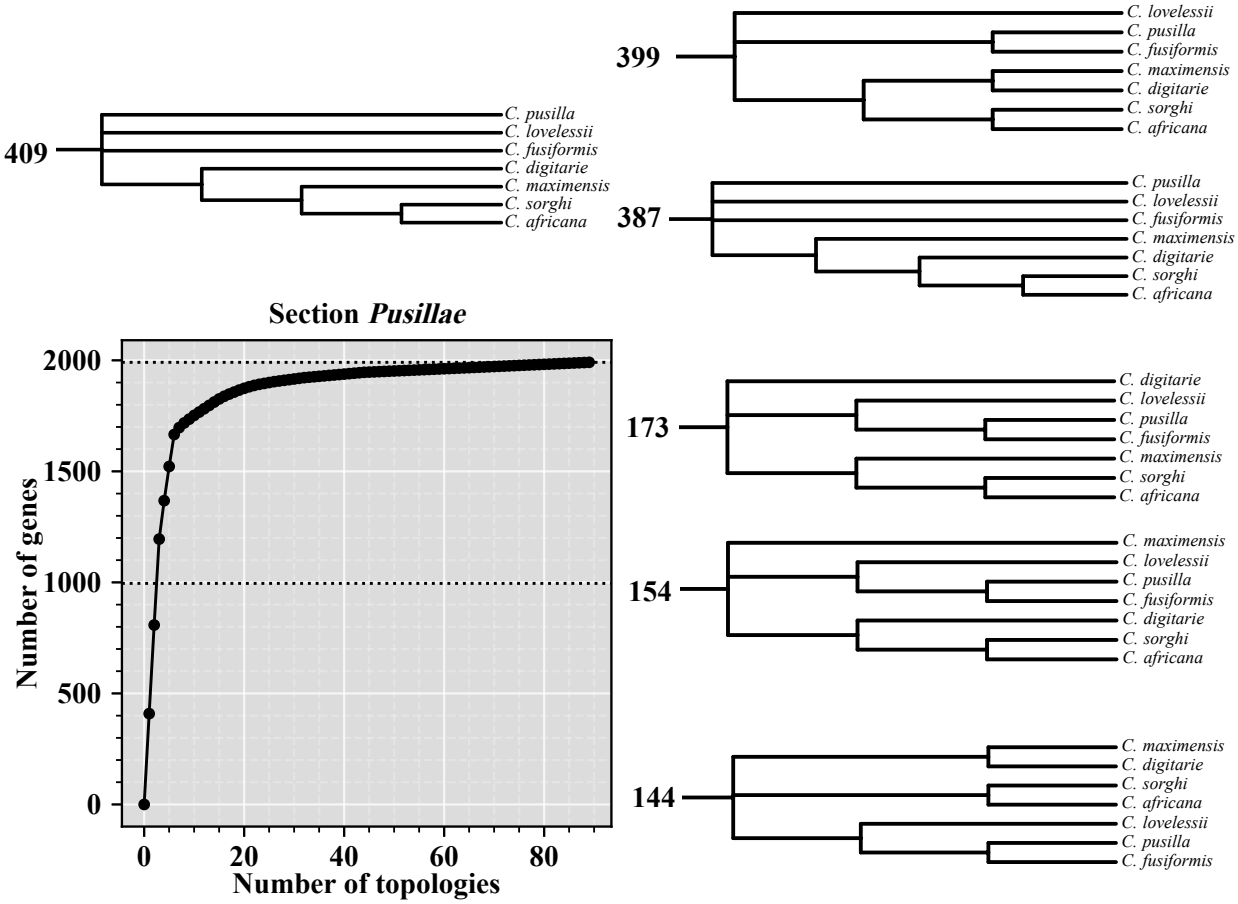
Appendix 2 Figure A2.4: Density consensus tree of 2,002 maximum likelihood phylogenetic reconstructions of the *Claviceps* genus using amino acid sequences of the single-copy orthologs with 1000 bootstrap replicates. Representative isolates from each species were used in this analysis for clarity. Thicker overlapping regions is an indicator of branch support. Tree order was determined by the most frequently occurring tree order.



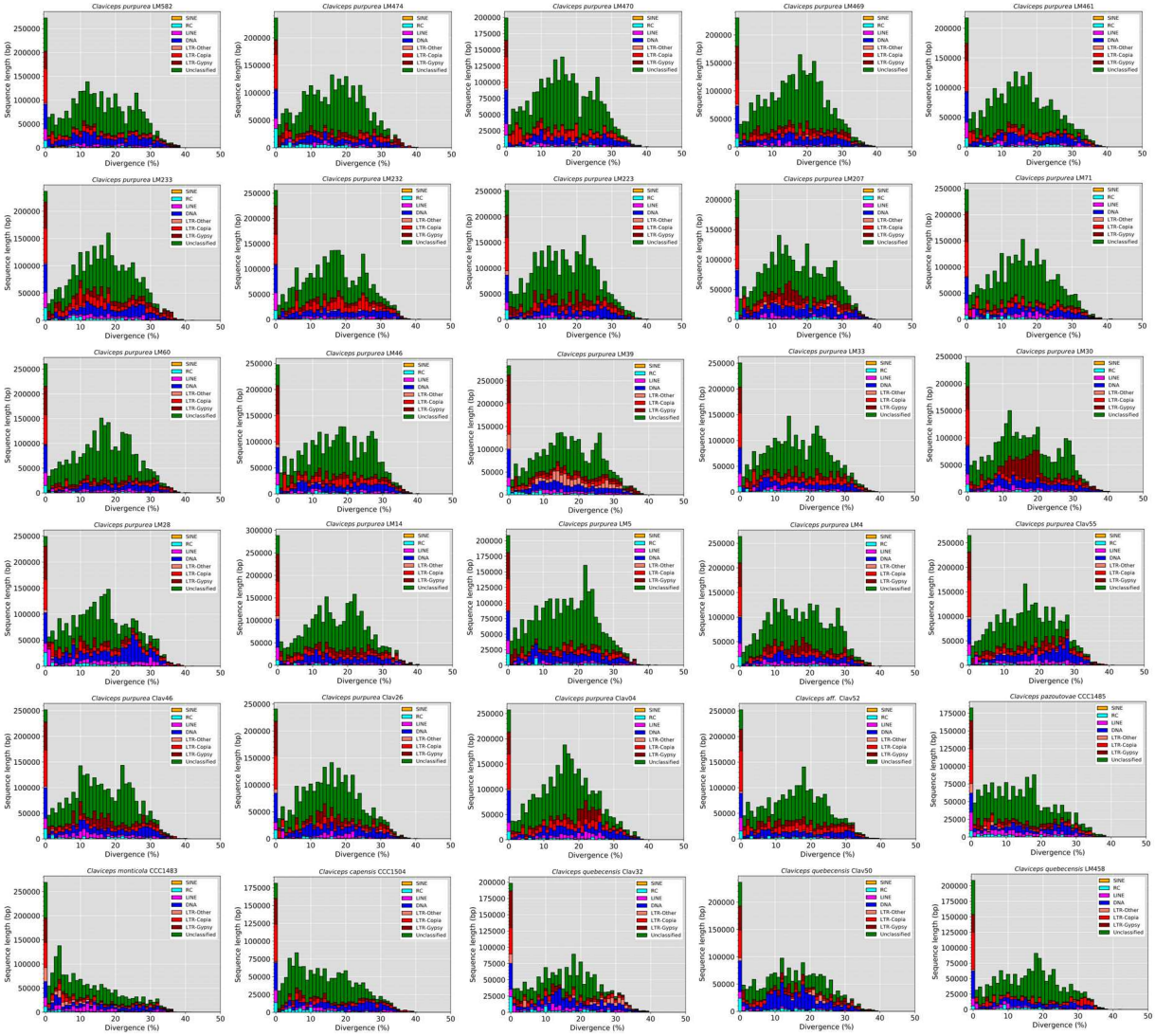
Appendix 2 Figure A2.5: Phylogenetic reconstructions and genealogy variation of gene trees for the *Claviceps* genus (excluding outgroups). (Line chart) Cumulative distribution of the number of genes per topology. Horizontal dotted lines indicate the half of the genes examined and total genes examined. (Trees) Four most frequent topologies with their corresponding frequencies.



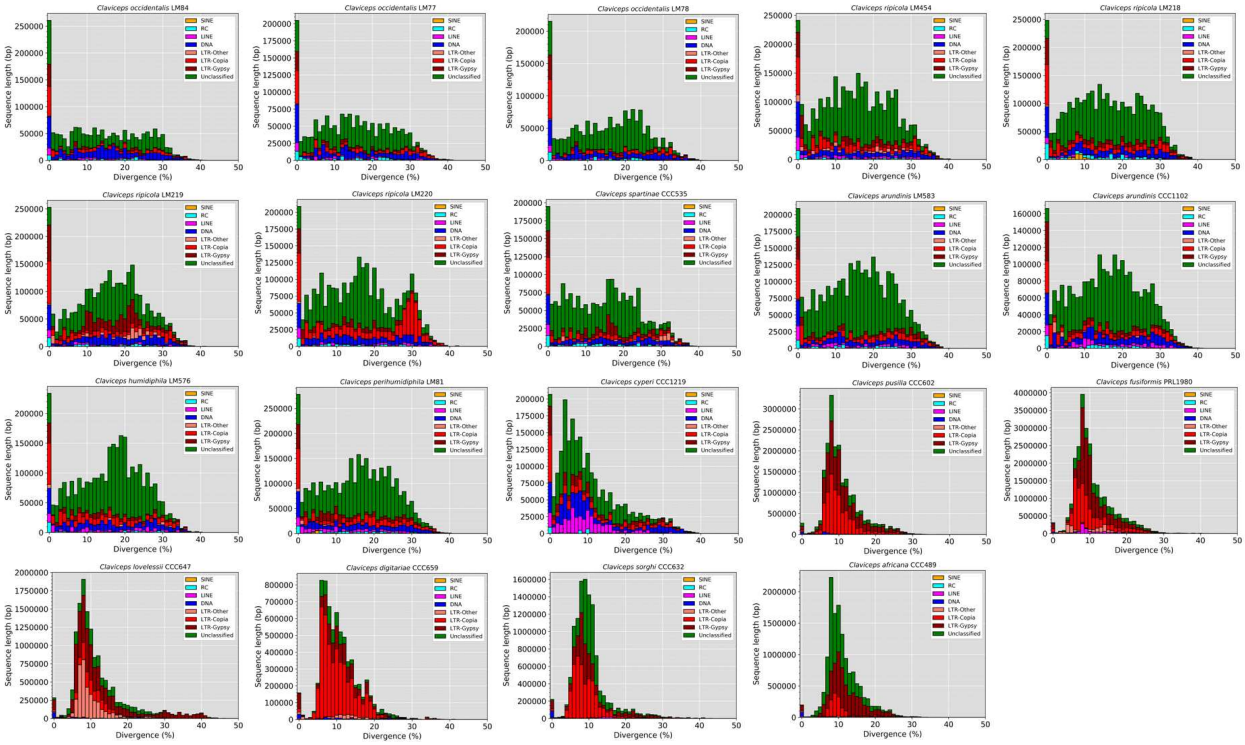
Appendix 2 Figure A2.6: Phylogenetic reconstructions and genealogy variation of gene trees for *Claviceps* section *Claviceps*. (Line chart) Cumulative distribution of the number of genes per topology. Horizontal dotted lines indicate the half of the genes examined and total genes examined. (Trees) Six most frequent topologies with their corresponding frequencies.



Appendix 2 Figure A2.7: Phylogenetic reconstructions and genealogy variation of gene trees for *Claviceps* section *Pusillae*. (Line chart) Cumulative distribution of the number of genes per topology. Horizontal dotted lines indicate the half of the genes examined and total genes examined. (Trees) Six most frequent topologies with their corresponding frequencies.

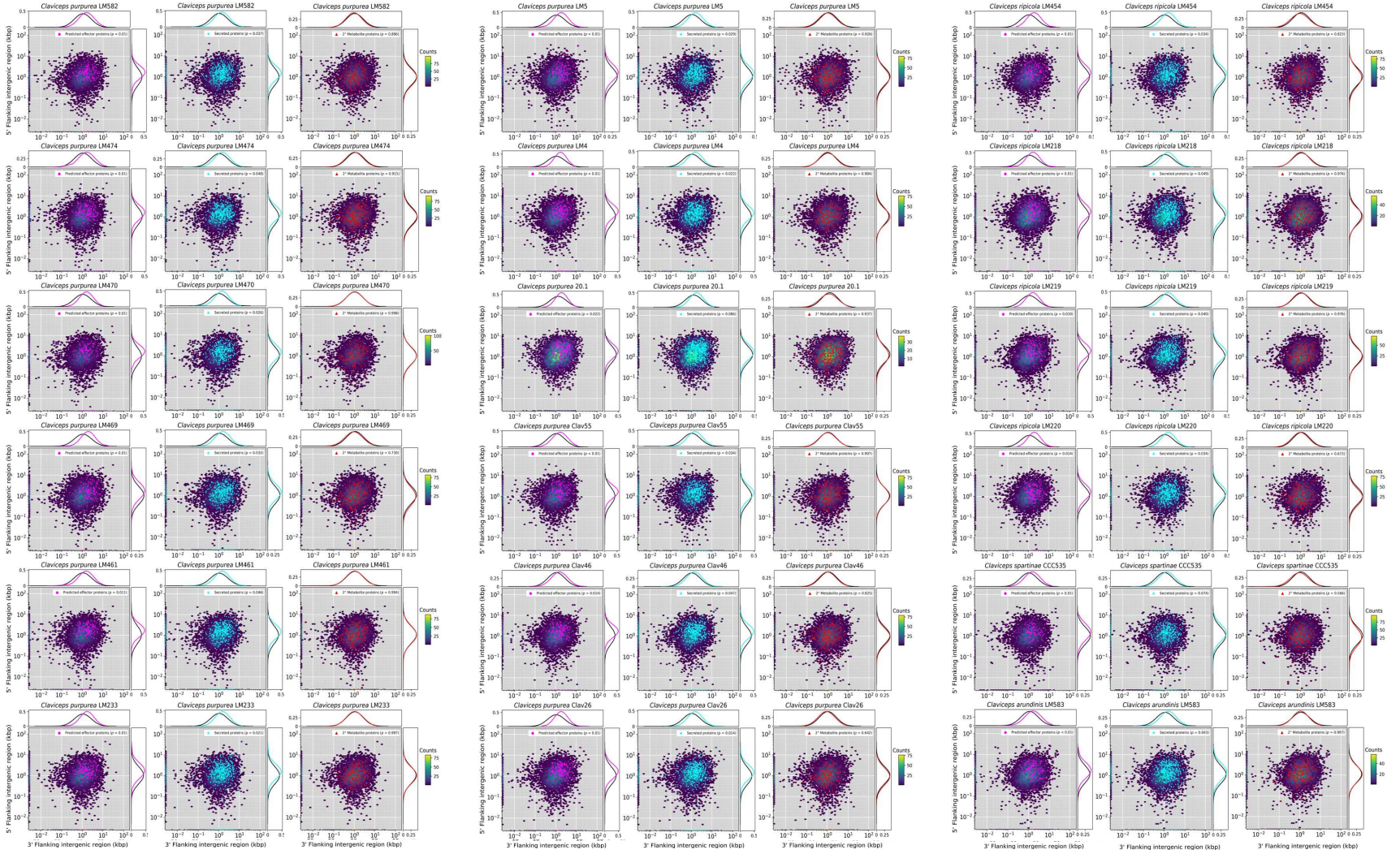


Appendix 2 Figure A2.8: Transposable element (TE) fragment divergence landscapes for *Claviceps* species. Stacked bar graphs show the non-normalized sequence length occupied in each genome (y-axis) for each TE type based on their percent divergence (x-axis) from their corresponding consensus sequence.

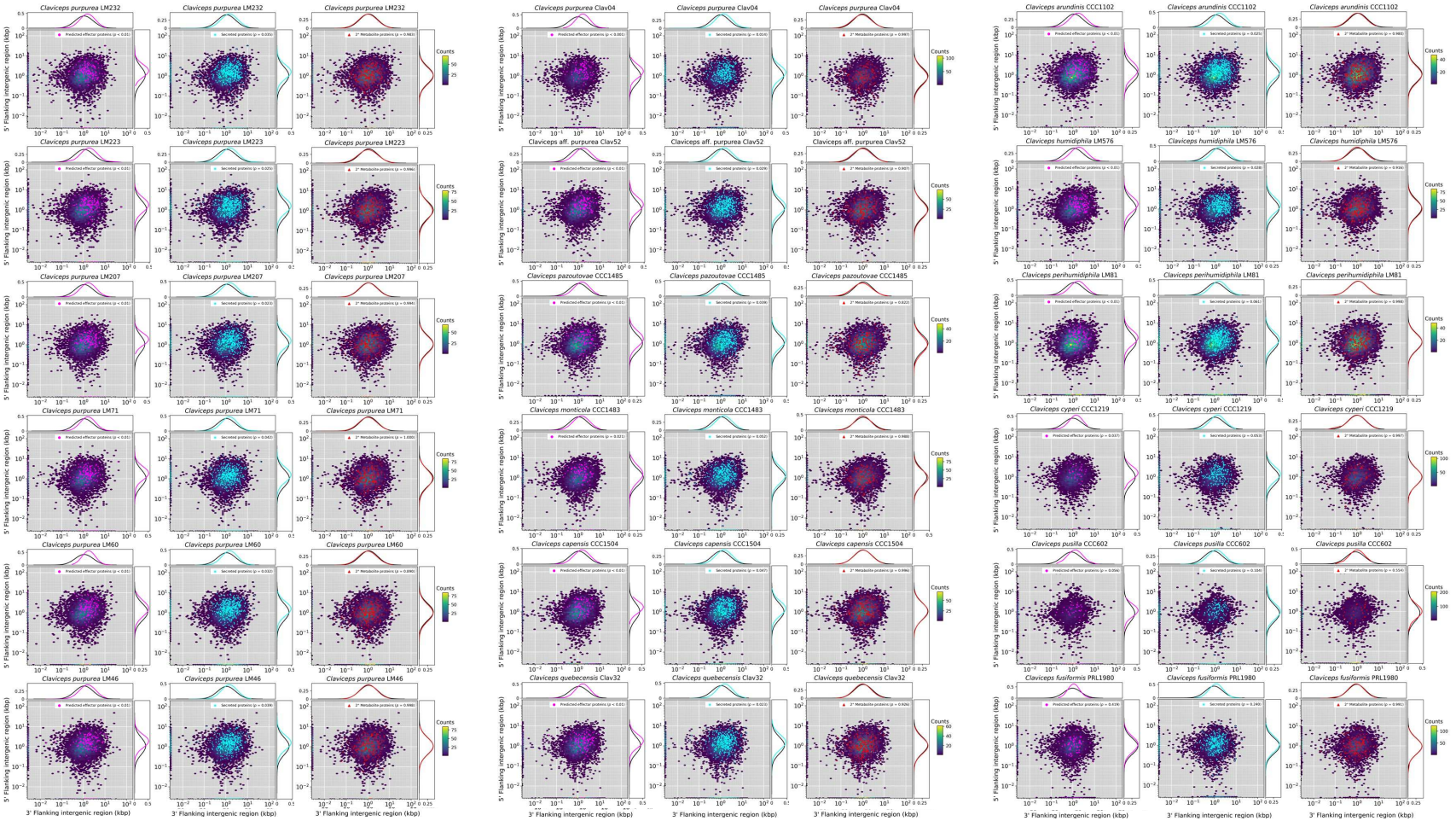


Appendix 2 Figure A2.8: Continued

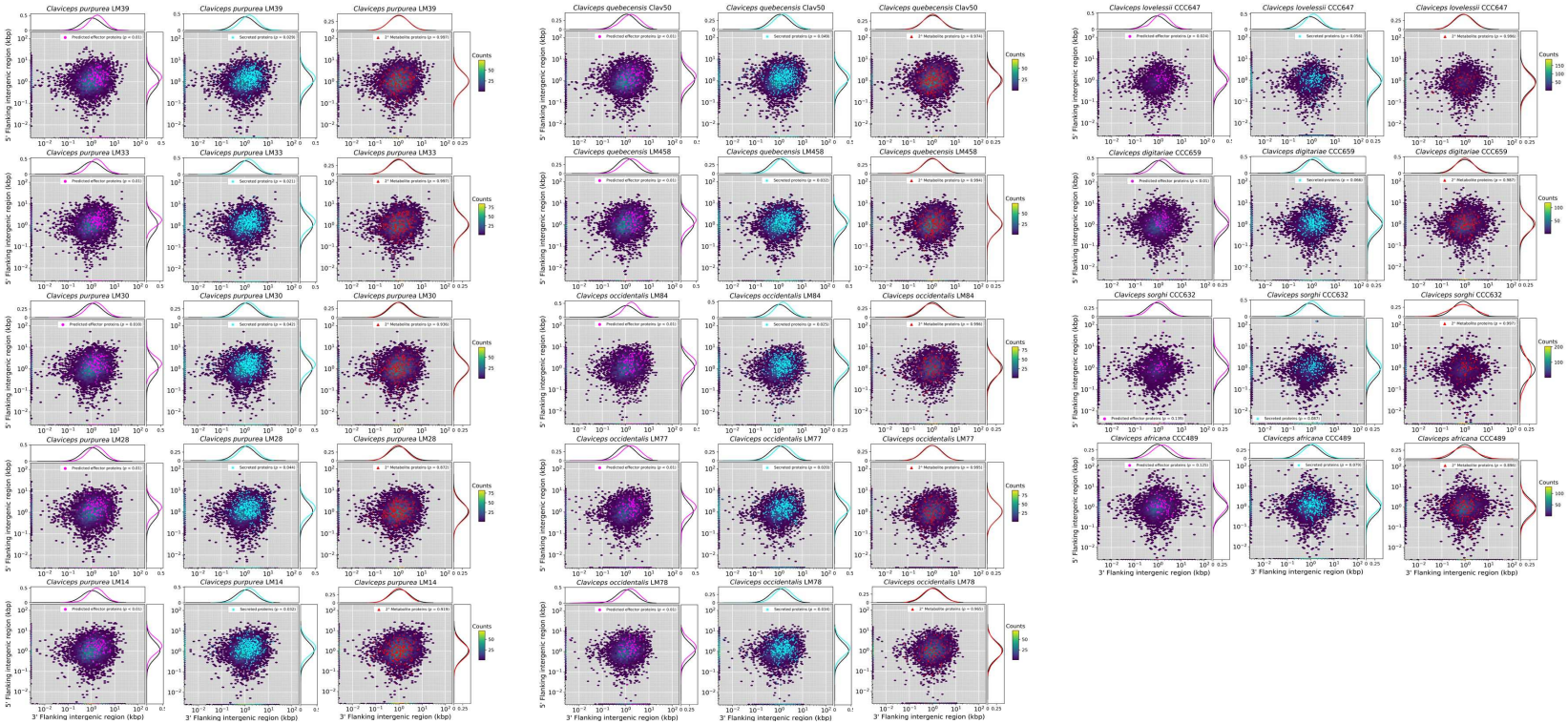
Appendix 2 Figure A2.9: Boxplot distributions of predicted effectors, secreted, secondary metabolite genes and other genes (i.e. genes that are not effectors, secreted, or secondary metabolite genes) in *Claviceps* species showing the mean distance (kbp) of each gene to the closest transposable element fragment (5' and 3' flanking distances were averaged together). Kruskal Wallis (P-value; * < 0.05, ** < 0.01, *** < 0.001, n.s. = not significant). Pairwise comparison was performed with two-sided Mann-Whitney U-test with Benjamini-Hochberg multitest correction. Letters correspond to significant differences between gene categories within sections ($\alpha = 0.05$).



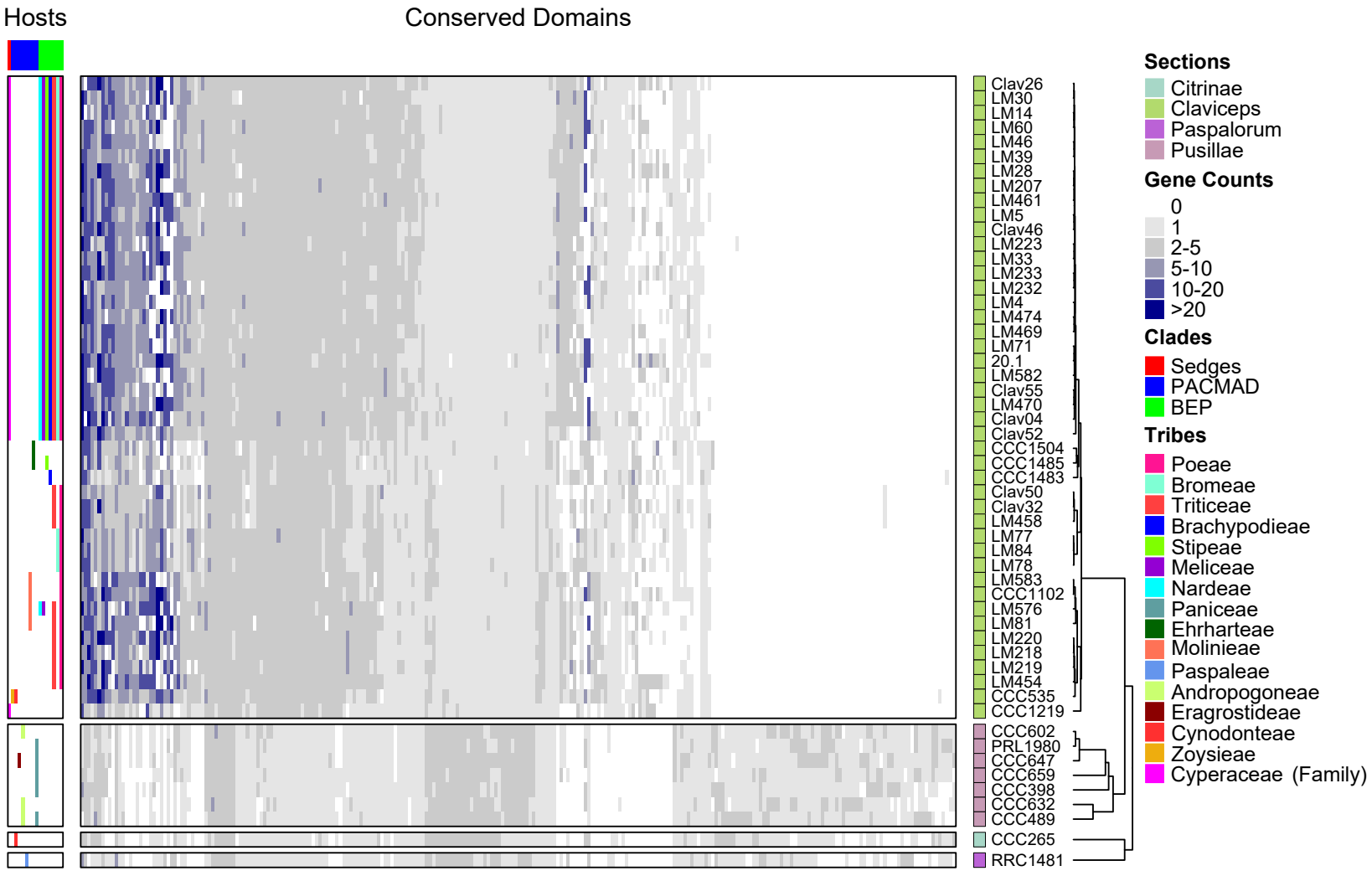
Appendix 2 Figure A2.10: 5' and 3' intergenic region size (y and x-axis) of *Claviceps* species. First genes of contigs are plotted along the bottom of the x-axis and genes at the end of each contig are plotted along the y-axis. Colored hexbins indicate the intergenic lengths of all genes with color-code indicating the frequency distribution (gene counts) according to the legend on the right. Overlaid markers indicate specific gene types corresponding to legends in the top right within each plot. Frequency distributions of specific gene types (corresponding legend color) and all other genes not of the specific type (black) are plotted along the x- and y-axis. For information on statistical test (See Chapter 3 Materials and Methods, pg 76).



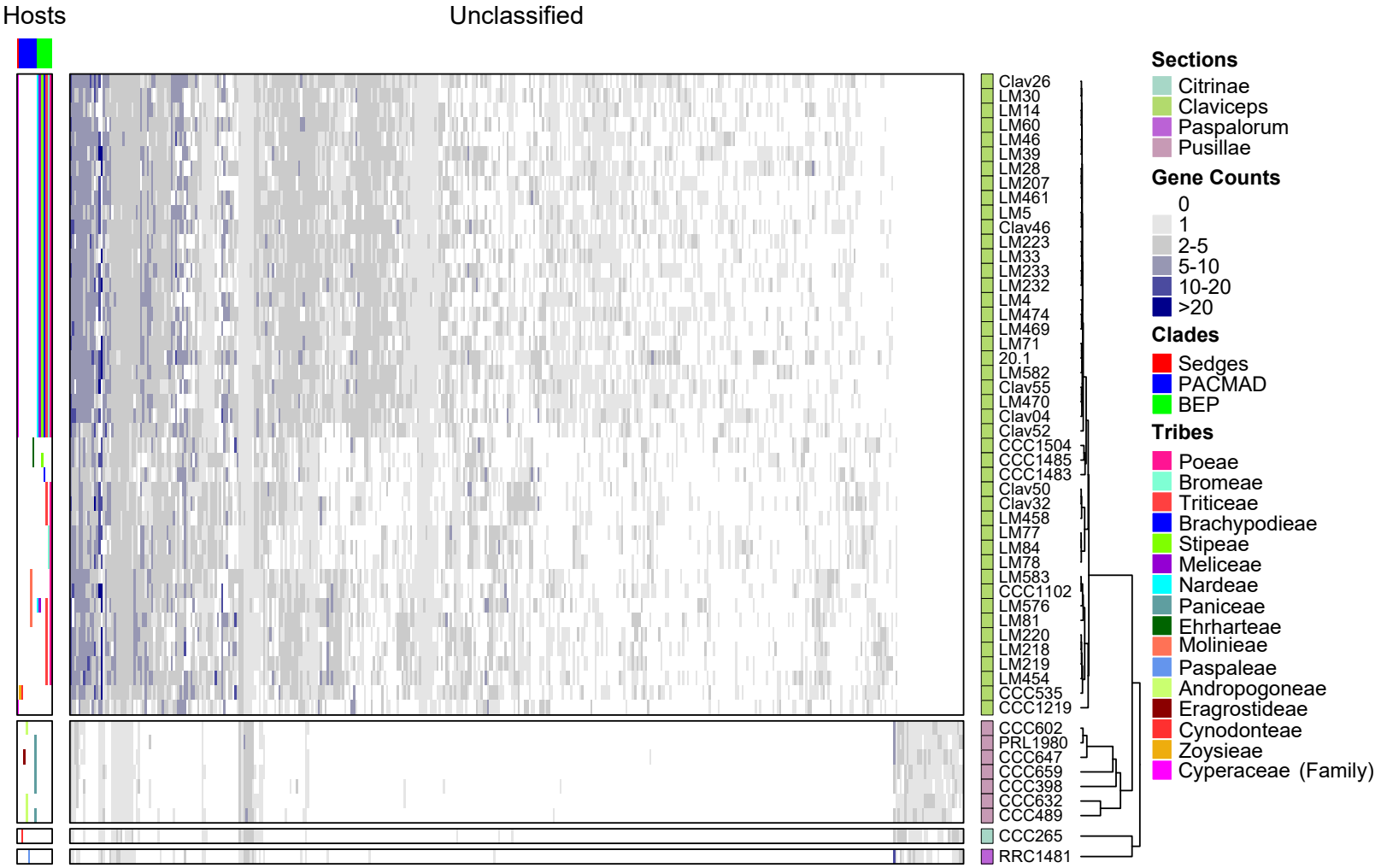
Appendix 2 Figure A2.10:
Continued



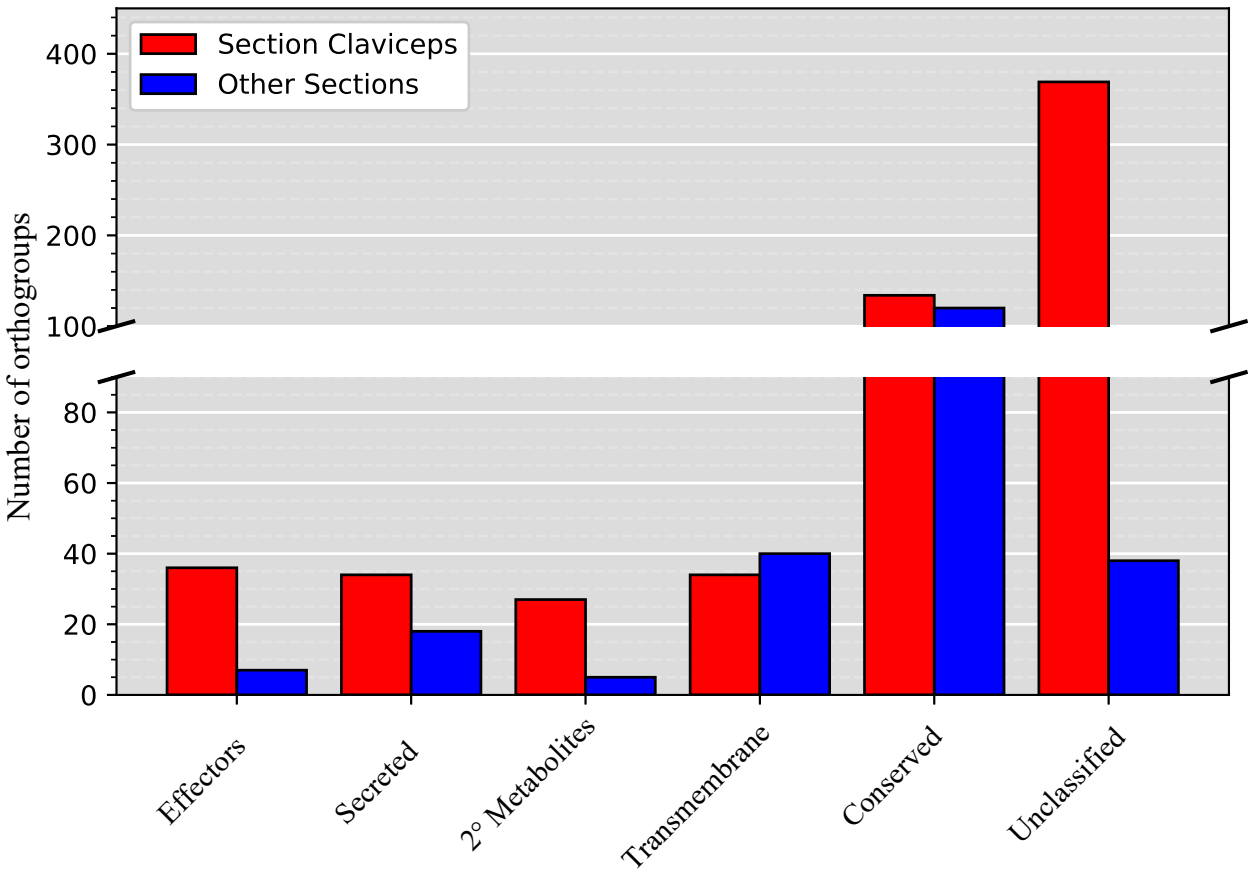
Appendix 2 Figure A2.10:
Continued



Appendix 2 Figure A2.11: Heatmap of gene counts in the remaining orthogroups containing genes encoding conserved protein domains for all 53 *Claviceps* strains ordered based on ML tree in Fig. 3.1 and separated by sections. Orthogroups are ordered based on hierarchical clustering. The host spectrum (left) is generalized across species, as no literature has determined the existence of race specific isolates within species, is shown on the left side of the figure determined from literature review of field collected samples (Supplementary Material in Píchová *et al.* 2018) and previous inoculation tests Campbell (1957) and Liu *et al.* (*Accepted*).



Appendix 2 Figure A2.12: Heatmap of gene counts in remaining orthogroups containing unclassified genes for all 53 *Claviceps* strains ordered based on ML tree in Fig. 3.1 and separated by sections. Orthogroups are ordered based on hierarchical clustering. The host spectrum (left) is generalized across species, as no literature has determined the existence of race specific isolates within species, is shown on the left side of the figure determined from literature review of field collected samples (Supplementary Material in Píchová *et al.* 2018) and previous inoculation tests Campbell (1957) and Liu *et al.* (*Accepted*).



Appendix 2 Figure A2.13: Number of orthogroups showing significantly ($P \leq 0.01$) greater expansion in respective *Claviceps* sections. Other sections include the combination of sects. *Pusillae*, *Citrinae*, and *Paspalorum*.

Appendix 2 Table A2.1: Collection information for *Claviceps* strains used in this study.

Organism	Strain	NCBI Accession	SRA Accession	Culture Collection	Location	Host	Collection Date
References:							
C. purpruea	20.1	SAMEA2272775	--	--	Germany	*Secale cereale*	~1988
C. fusiformis	PRL 1980	SAMN02981339	--	--	Africa: Cote d'Ivoire	*Pennisetum typhoideum*	~1958
C. paspali	RRC 1481	SAMN02981342	--	--	USA: Georgia, Mansfield	*Paspalum* sp.	~2001
This study:							
C. purpruea	Clav04	SAMN11159846	SRR8785178	‡	USA: Colorado, San Luis Valley	*Bromus inermis*	2016
C. purpruea	Clav26	SAMN11159847	SRR8785181	‡	USA: Colorado, San Luis Valley	*Hordeum vulgare*	2016
C. purpruea	Clav46	SAMN11159848	SRR8785180	‡	USA: Wyoming, Worland	*Secale cereale*	2016
C. purpruea	Clav55	SAMN11159850	SRR8785174	‡	New Zealand	*Lolium perenne*	2017
C. purpruea	LM4	SAMN11159851	SRR8785145	DAOMC:250624	Canada: Manitoba	*Tricosecale*	1996
C. purpruea	LM5	SAMN11159852	SRR8785146	DAOMC:250625	Canada: Manitoba	*Hordeum vulgare*	1996
C. purpruea	LM14	SAMN11159853	SRR8785147	DAOMC:250634	Canada: Saskatchewan	*Hordeum vulgare*	1996
C. purpruea	LM28	SAMN11159854	SRR6985966†	DAOMC:250647	Canada: Saskatchewan	*Triticum aestivum*	2000
C. purpruea	LM30	SAMN11159855	SRR8785151	DAOMC:250649	Canada: Saskatchewan	*Secale cereale*	2000
C. purpruea	LM33	SAMN11159856	SRR8785141	DAOMC:250652	Canada: Manitoba	*Secale cereale*	2015
C. purpruea	LM39	SAMN11159857	SRR8785142	DAOMC:250658	Canada: Saskatchewan	*T. turgidum subsp. durum*	2000
C. purpruea	LM46	SAMN11159858	SRR8785143	DAOMC:250663	Canada: Alberta	*T. turgidum subsp. durum*	2000
C. purpruea	LM60	SAMN11159859	SRR8785144	DAOMC:250680	Canada: Manitoba	*Avena sativa*	2005
C. purpruea	LM71	SAMN11159860	SRR8785148	DAOMC:250720	United Kingdom	*Alopercurus myosuroides*	2004
C. purpruea	LM207	SAMN11159861	SRR8785149	‡	Canada: Manitoba	*Elymus repens*	2014
C. purpruea	LM223	SAMN11159862	SRR8785164	DAOMC:250814	Canada: Manitoba	*Bromus riparius*	2014
C. purpruea	LM232	SAMN11159863	SRR8785161	DAOMC:250822	Canada: Manitoba	*Phalaris canariensis*	2014
C. purpruea	LM233	SAMN11159864	SRR8785162	‡	Canada: Manitoba	*Phalaris canariensis*	2014
C. purpruea	LM461	SAMN11159865	SRR8785163	DAOMC:251847	Canada: Quebec	*Elymus repens*	2016
C. purpruea	LM469	SAMN11159866	SRR8785165	‡	Canada: Ontario	*Triticum aestivum*	2016
C. purpruea	LM470	SAMN11159867	SRR8785166	‡	Canada: Ontario	*Elymus repens*	2016
C. purpruea	LM474	SAMN11159868	SRR8785167	‡	Canada: Ontario	*Hordeum vulgare*	2016
C. purpruea	LM582	SAMN11159869	SRR6985962†	DAOMC:251723	Czech Republic: Bezdedice	*Secale cereale*	2003
C. aff. purpruea	Clav52	SAMN11159849	SRR8785175	‡	USA: Washington	*Poa pratensis*	2017
C. quebecensis	Clav32	SAMN11159882	SRR8785176	‡	USA: Montana, Shephard	*Hordeum vulgare*	2016
C. quebecensis	Clav50	SAMN11159881	SRR8785177	‡	USA: Oklahoma, Hoop house Ardmore	*Elymus sp.*	2017
C. quebecensis	LM458	SAMN11159883	SRR6985957†	DAOMC:251898	Canada:Quebec, Cote Nord	*Ammophila* (plant)	2015
C. occidentalis	LM77	SAMN11159879	SRR8785179	DAOMC:250577	Canada: Alberta	*Phleum pratense*	2016
C. occidentalis	LM78	SAMN11159878	SRR6985960†	DAOMC:250578	Canada: Alberta, North Star	*Bromus inermis*	1956
C. occidentalis	LM84	SAMN11159876	SRR8785170	DAOMC:250590	Canada: British Columbia	*Bromus inermis*	2016
C. ripicola	LM218	SAMN11159875	SRR6985964†	DAOMC:251843	Canada: Manitoba, Grants Field Snowflake	*Phalaris arundinacea*	2014
C. ripicola	LM219	SAMN11159874	SRR8785169	DAOMC:250811	Canada: Manitoba	*Phalaris arundinacea*	2014
C. ripicola	LM220	SAMN11159873	SRR8785168	DAOMC:250812	Canada: Manitoba	*Phalaris arundinacea*	2014
C. ripicola	LM454	SAMN11159872	SRR6985963†	DAOMC:251845	Canada: Quebec, MRC Maria-Chapdelaine	*Ammophila breviligulata*	2014
C. spartinae	CCC535	SAMN11159888	SRR8785160	CCC:535	United Kingdom: Marchwood	*Sporobolus anglicus*	1999

† SRA data first published in Nguyen *et al.* 2018

‡ Cultures available at the lab of Dr. Vamsi Nalam, Colorado State University, Fort Collins, CO or Dr. Miao Liu Ottawa, Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada

Appendix 2 Table A2.1: Continued.

Organism	Strain	NCBI Accession	SRA Accession	Culture Collection	Location	Host	Collection Date
C. arundinis	LM583	SAMN11159894	SRR6985961†	DAOMC:251724/CCC:933	Czech Republic: Haklovy Dvory	*Phragmites australis*	2008
C. arundinis	CCC1102	SAMN11159893	SRR8785153	CCC:1102	France: D973 Rte de Beaune	*Phragmites australis*	2009
C. humidiphila	LM576	SAMN11159871	SRR6985959†	DAOMC:251717/CCC:434	Germany: Bavaria	*Dactylis sp.*	1998
C. perihumidiphila	LM81	SAMN11159877	SRR6985958†	DAOMC:250581	Canada: Alberta, Metiskow	*Elymus albicans*	1956
C. cyperi	CCC1219	SAMN11159895	SRR8785154	CCC:1219	South Africa: Kempton Park	*Cyperus esculentus*	2012
C. capensis	CCC1504	SAMN11159898	SRR8785171	CCC:1504T	South Africa: Cape Town, Western Cape	*Ehrharta villosa*	2014
C. pazoutovae	CCC1485	SAMN11159897	SRR8785152	CCC:1485T	South Africa: Hogsback, Eastern Cape	*Stipa dregeana*	2014
C. monticola	CCC1483	SAMN11159896	SRR8785150	CCC:1483T	South Africa: Hogsback, Eastern Cape	*Brachypodium sp.*	2014
C. pusilla	CCC602	SAMN11159889	SRR8785157	CCC:602	Zimbabwe: Matopos	*Bothriochloa insculpta*	2000
C. lovelessii	CCC647	SAMN11159891	SRR8785155	CCC:647T	Zimbabwe: Matopos	*Eragrostis sp.*	2001
C. digitariae	CCC659	SAMN11159892	SRR8785156	CCC:659	Africa: Botswana	*Digitaria eriantha*	--
C. maximensis	CCC398	SAMN11159886	SRR8785172	CCC:398	Paraguay: Chaco	*Megathyrsus maximus*	1997
C. sorghi	CCC632	SAMN11159890	SRR8785158	CCC:632	India: Karnataka, Jewargi, Gulbarga	*Sorghum bicolor*	2000
C. africana	CCC489	SAMN11159887	SRR8785159	CCC:489	Mexico: Celaya, Guanajuato	*Sorghum bicolor*	1998
C. citrina	CCC265	SAMN11159885	SRR8785173	CCC:265	Mexico: Texcoco (semillero), Mexico City	*Distichlis spicata*	1996

† SRA data first published in Nguyen *et al.* 2018

‡ Cultures available at the lab of Dr. Vamsi Nalam, Colorado State University, Fort Collins, CO or Dr. Miao Liu Ottawa, Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada

Appendix 2 Table A2.2: Number of genes with functional protein classifications for all 53 *Claviceps* genomes in this study.

			Protein function						
Organism	**Strain**	**Section**	**Conserved domains**	**MEROP domains**	**CAZY-mes**	**2° meta-bolites**	**Trans-membrane**	**Secreted**	**Predicted effectors**
References:									
C. purpruea	20.1	Claviceps	6560	255	243	321	1114	547	199
C. fusiformis	PRL1980	Pusillae	6178	249	244	227	1315	521	152
C. paspali	RRC1481	Paspalorum	5943	235	230	185	1216	565	196
This study:									
C. purpruea	Clav04	Claviceps	6233	246	231	199	1145	557	221
C. purpruea	Clav26	Claviceps	6221	263	229	255	1158	592	226
C. purpruea	Clav46	Claviceps	6181	266	229	281	1151	560	198
C. purpruea	Clav55	Claviceps	6113	260	231	253	1166	553	195
C. purpruea	LM4	Claviceps	6132	249	233	230	1149	584	231
C. purpruea	LM5	Claviceps	6179	257	231	287	1154	547	193
C. purpruea	LM14	Claviceps	6132	259	227	273	1164	529	172
C. purpruea	LM28	Claviceps	6260	254	233	279	1156	540	188
C. purpruea	LM30	Claviceps	6167	262	229	268	1156	588	220
C. purpruea	LM33	Claviceps	6193	263	226	264	1143	597	237
C. purpruea	LM39	Claviceps	6156	265	232	256	1150	584	228
C. purpruea	LM46	Claviceps	6125	261	225	250	1158	554	201
C. purpruea	LM60	Claviceps	6126	256	231	290	1147	553	195
C. purpruea	LM71	Claviceps	6095	240	221	248	1152	550	219
C. purpruea	LM207	Claviceps	6120	259	220	251	1140	571	224
C. purpruea	LM223	Claviceps	6076	234	230	250	1146	564	214
C. purpruea	LM232	Claviceps	6169	260	232	284	1156	555	199
C. purpruea	LM233	Claviceps	6294	264	230	266	1162	570	207
C. purpruea	LM461	Claviceps	6223	237	231	246	1154	550	210
C. purpruea	LM469	Claviceps	6106	255	229	273	1148	542	185
C. purpruea	LM470	Claviceps	6195	256	223	210	1144	533	184
C. purpruea	LM474	Claviceps	6118	242	222	290	1162	522	191
C. purpruea	LM582	Claviceps	6132	257	226	229	1141	545	198
C. aff. purpruea	Clav52	Claviceps	6078	252	225	250	1153	523	177
C. quebecensis	Clav32	Claviceps	6057	244	226	266	1157	522	174
C. quebecensis	Clav50	Claviceps	5986	248	228	260	1139	524	174
C. quebecensis	LM458	Claviceps	6007	243	226	235	1135	508	154
C. occidentalis	LM77	Claviceps	6020	243	222	186	1132	517	182
C. occidentalis	LM78	Claviceps	6052	246	223	163	1133	513	174
C. occidentalis	LM84	Claviceps	6088	244	223	189	1129	517	180
C. ripicola	LM218	Claviceps	6090	249	228	255	1136	545	203
C. ripicola	LM219	Claviceps	6133	255	226	270	1130	538	188
C. ripicola	LM220	Claviceps	6168	249	225	255	1122	564	211
C. ripicola	LM454	Claviceps	6163	261	224	261	1134	554	210
C. spartinae	CCC535	Claviceps	6146	260	227	249	1153	504	163
C. arundinis	LM583	Claviceps	6102	253	225	287	1140	526	186
C. arundinis	CCC1102	Claviceps	6213	259	226	288	1130	548	195
C. humidiphila	LM576	Claviceps	6147	253	227	282	1152	564	211
C. perihumidiphila	LM81	Claviceps	6106	238	225	270	1124	522	176
C. cyperi	CCC1219	Claviceps	5774	228	202	135	1072	392	97
C. capensis	CCC1504	Claviceps	5989	245	225	250	1145	475	137
C. pazoutovae	CCC1485	Claviceps	5954	232	223	225	1137	478	150
C. monticola	CCC1483	Claviceps	5908	248	220	207	1118	481	148
C. pusilla	CCC602	Pusillae	6276	252	227	76	1235	461	138
C. lovelessii	CCC647	Pusillae	6351	248	228	110	1230	525	174
C. digitariae	CCC659	Pusillae	6195	254	239	187	1213	513	158
C. maximensis	CCC398	Pusillae	6100	244	235	226	1172	468	126
C. sorghi	CCC632	Pusillae	6085	241	222	117	1139	425	123
C. africana	CCC489	Pusillae	6057	241	221	165	1151	471	145
C. citrina	CCC265	Citrinae	5879	224	207	120	1100	368	85

Appendix 2 Table A2.3: Additional annotated genomes used in OrthoFinder analysis for finding orthologous gene families.

Organism	Strain	Accession
Acremonium chrysogenum	ATCC 11550	SAMN02799700
Atkinsonella hypoxylon	B4728	http://www.endophyte.uky.edu/
Atkinsonella texensis	B6155	http://www.endophyte.uky.edu/
Balansia obtecta	B249	http://www.endophyte.uky.edu/
Clonostachys rosea	CBS125111	JGI:1032557
Epichloe amarillians	ATCC 200744	http://www.endophyte.uky.edu/
Epichloe aotearoea	ATCC MYA-1229	http://www.endophyte.uky.edu/
Epichloe baconii	ATCC 200745	http://www.endophyte.uky.edu/
Epichloe brachyelytri	E4804	http://www.endophyte.uky.edu/
Epichloe bromicola	AL0426/2	http://www.endophyte.uky.edu/
Epichloe coenophiala	e4163	http://www.endophyte.uky.edu/
Epichloe elymi	ATCC 201551	http://www.endophyte.uky.edu/
Epichloe festucae	Fl1	http://www.endophyte.uky.edu/
Epichloe gansuensis	CDM-2007b	http://www.endophyte.uky.edu/
Epichloe glyceriae	ATCC 200747	http://www.endophyte.uky.edu/
Epichloe inebrians	ATCC MYA-1228	http://www.endophyte.uky.edu/
Epichloe mollis	AL9924	http://www.endophyte.uky.edu/
Epichloe sylvatica	GR 10156	http://www.endophyte.uky.edu/
Epichloe typhina	ATCC 200736	http://www.endophyte.uky.edu/
Epichloe uncinata	CBS 102646	http://www.endophyte.uky.edu/
Fusarium ambrosium	NRRL 20438	SAMN07200640
Fusarium avenaceum	Fave_LH27	SAMN02850900
Fusarium fujikuroi		SAMEA4440726
Fusarium fujikuroi	B14	SAMEA4436914
Fusarium fujikuroi	C1995	SAMEA4440729
Fusarium fujikuroi	E282	SAMEA4440730
Fusarium fujikuroi	FGSC8932	SAMN03075939
Fusarium fujikuroi	FSU48	SAMEA4440731
Fusarium fujikuroi	IMI58289	SAMEA3724789
Fusarium fujikuroi	KSU3368	SAMN03075941
Fusarium fujikuroi	KSUX10626	SAMN03075940
Fusarium fujikuroi	m657	SAMEA4440732
Fusarium fujikuroi	MRC2276	SAMEA4440733
Fusarium fujikuroi	NCIM1100	SAMEA4440734
Fusarium graminearum	PH-1/NRRL 31084	SAMN02953593
Fusarium kuroshium	AF-12	SAMN07200645
Fusarium langsethiae	Fl201059	SAMN03274931
Fusarium longipes	NRRL 20695	SAMN08631279
Fusarium mangiferae	MRC7560	SAMEA3862491
Fusarium oxysporum f. sp. cepae	FoCFus2	SAMN05529097
Fusarium oxysporum f. sp. conglutinans	54008	SAMN02981380
Fusarium oxysporum f. sp. cubense	54006	SAMN02981379
Fusarium oxysporum f. sp. lycopersici	4287	SAMN02953675
Fusarium oxysporum f. sp. melonis	26406	SAMN02981378
Fusarium oxysporum f. sp. narcissi	N139	SAMN05526391
Fusarium oxysporum f. sp. pisi	HDV247	SAMN02981366
Fusarium oxysporum f. sp. radicis-cucumerinum	Forc016	SAMN04348764
Fusarium oxysporum f. sp. raphani	54005	SAMN02981381
Fusarium oxysporum f. sp. vasinfectum	25433	SAMN02981377

Appendix 2 Table A2.3: Continued.

Organism	Strain	Accession
Fusarium poae	2516	SAMN05178635
Fusarium proliferatum	ET1	SAMEA3862493
Fusarium pseudograminearum	CS3096	SAMN02981337
Fusarium solani (Nectria haematococca)	77-13-4	SAMN02746079
Fusarium sporotrichioides	NRRL 3299	SAMN08631227
Fusarium venenatum	A3/5	SAMEA2827224
Fusarium verticillioides	7600	SAMN02953630
Periglandula ipomoeae	IasaF13	http://www.endophyte.uky.edu/
Purpureocillium lilacinum	PLFJ-1	SAMN04404347
Saccharomyces cerevisiae	S288C	PRJNA43747
Stachybotrys chlorohalonata	IBT 40285	SAMN01819006
Trichoderma arundinaceum	IBT 40837	SAMN06320351
Trichoderma asperellum	CBS 433.97	SAMN00769595
Trichoderma atroviride	IMI 206040	SAMN02744066
Trichoderma citrinoviride	TUCIM 6016	SAMN05369575
Trichoderma gamsii	T6085	SAMN02849381
Trichoderma guizhouense	NJAU 4742	SAMN04535176
Trichoderma harzianum	CBS22695	SAMN00761861
Trichoderma harzianum	T6776	SAMN02851310
Trichoderma harzianum	Tr1	SAMN06219536
Trichoderma harzianum	TR274	SAMN07456232
Trichoderma harzianum	M10 v1	JGI:1185309
Trichoderma harzianum	T22 v1	JGI:1185313
Trichoderma longibrachiatum	ATCC 18648	SAMN00767620
Trichoderma parareesei	CBS 125925	SAMN03784587
Trichoderma reesei	QM6a	SAMN02746107
Trichoderma virens	Gv29-8	SAMN02744059
Ustilaginoidea virens	UV-8b	SAMN02693461
Ustilago maydis	521	SAMN02900459

Appendix 2 Table A2.4: *P*-values for genomic fluidity differences from two-sample two-sided z-test, bold numbers indicate significance.

	Fusarium-Genus	Epichloe-Genus	Trichoderma-Genus	T.harazium-Species	Pusillae-Section	Claviceps-Genus	F.oxysporum-Species	Claviceps-Section	C.purpurea-Species	F.fujikuroi-Species
Fusarium-Genus	-	8.74E-01	3.78E-01	**3.30E-04**	**1.54E-03**	**7.98E-05**	**4.01E-07**	**4.35E-21**	**6.44E-27**	**3.65E-25**
Epichloe-Genus	-	-	4.94E-01	**2.68E-03**	**6.47E-03**	**7.59E-04**	**4.39E-05**	**1.81E-14**	**2.04E-18**	**9.37E-18**
Trichoderma-Genus	-	-	-	5.29E-02	7.23E-02	**2.01E-02**	**6.26E-03**	**5.33E-09**	**1.12E-11**	**1.73E-11**
T.harazium-Species	-	-	-	-	9.63E-01	5.21E-01	3.35E-01	**2.04E-10**	**8.00E-16**	**2.65E-14**
Pusillae-Section	-	-	-	-	-	6.13E-01	4.96E-01	**3.34E-06**	**3.91E-09**	**7.56E-09**
Claviceps-Genus	-	-	-	-	-	-	9.56E-01	**2.98E-06**	**8.88E-10**	**2.95E-09**
F.oxysporum-Species	-	-	-	-	-	-	-	**3.42E-30**	**3.32E-53**	**7.78E-31**
Claviceps-Section	-	-	-	-	-	-	-	-	**2.05E-11**	**1.35E-04**
C.purpurea-Species	-	-	-	-	-	-	-	-	-	6.25E-01
F.fujikuroi-Species	-	-	-	-	-	-	-	-	-	-

Appendix Table A2.5: BLASTp results showing the number of non-duplicated hits passing an e-value cutoff of 10E-5 with 50% coverage an 35% identity of biologicaly important Claviceps genes for all 53 genomes used in this study.

Gene name	Gene Code	20.1	Clav 04	Clav 26	Clav 46	Clav 55	LM 4	LM 5	LM 14	LM 28	LM 30	LM 33	LM 39	LM 46	LM 60	LM 71	LM 207	LM 223	LM 232	LM 233	LM 461	LM 469	LM 470	LM 474	LM 582	Clav 52	LM 77	LM 78	LM 84	Clav 32	Clav 50	LM 458	LM 218	LM 219	LM 220	LM 454	LM 81	LM 576	LM 583	CCC 1102	CCC 535	CCC 1504	CCC 1483	CCC 1485	CCC 1219	CCC 489	CCC 659	CCC 647	CCC 398	CCC 602	CCC 632	PRL 1980	RRC 1481	CCC 265	Reference
Virulence factors																																																						
mpk2	CCE28226.1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	4	5	3	3	5	3	3	3	4	3	3	3	4	5	4	4	4	3	Mey *et al.* 2002
mpk1	CCE29906.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	4	Mey *et al.* 2001
mid1	CCE30023.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Bormann and Tudzynski 2009
cppg1	CCE33055.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Oeser *et al.* 2002
cppg2	CCE33054.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Oeser *et al.* 2002
cpnox1	CCE35216.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Giesbert *et al.* 2008
chm1-Cla4	CCE30913.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	3	3	2	2	2	2	2	3	3	Rolke *et al.* 2008
cdc42	CCE32231.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Scheffer *et al.* 2005a
rac1	CCE32733.1	6	5	5	6	6	6	5	6	6	6	6	5	6	6	6	5	6	6	6	6	6	5	6	5	6	6	6	6	6	6	5	6	6	6	6	6	6	6	5	6	5	6	6	6	6	5	6	4	5	7	6	5	Rolke *et al.* 2008
cot-1	CCE35250.1	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	3	3	2	3	3	3	3	3	3	3	3	3	2	3	2	3	3	2	2	3	2	2	3	Scheffer *et al.* 2005b
Phytohormones																																																							
cptRNA-ipt	CCE29200.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Hinsch *et al.* 2015
cpp450	CCE30328.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	1	2	1	Hinsch *et al.* 2015
cpipt-log	CCE30329.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Hinsch *et al.* 2015
cpiaah	CCE30767.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Mlynarčiková, 2015
RIP																																																							
rid-1	NCU02034																																													1	1	1	1	1	†	1	1	1	Freitag *et al.* 2002
ETP Gene Cluster																																																						
tcpZ	CCE28980.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpN	CCE28981.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpJ	CCE28982.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1		1	1	1	Dopstadt *et al.* 2016
tcpK	CCE28983.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpA	CCE28984.1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	5	3	3	4	2	4	3	4	Dopstadt *et al.* 2016
tcpG	CCE28985.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpC	CCE28986.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpD	CCE28987.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1		1		1			Dopstadt *et al.* 2016
tcpI	CCE28988.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016
tcpP	CCE28989.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1								Dopstadt *et al.* 2016
tcpT	CCE28990.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											Dopstadt *et al.* 2016

†BLAST hit with 29% coverage and 45% identity

Appendix Table A2.5: Continued

Gene name	Gene Code	20.1	Clav 04	Clav 26	Clav 46	Clav 55	LM4	LM5	LM14	LM28	LM30	LM33	LM39	LM46	LM60	LM71	LM207	LM223	LM232	LM233	LM461	LM469	LM470	LM474	LM582	Clav 52	LM77	LM78	LM84	Clav 32	Clav 50	LM458	LM218	LM219	LM220	LM454	LM81	LM576	LM583	CCC 1102	CCC 535	CCC 1504	CCC 1483	CCC 1485	CCC 1219	CCC 489	CCC 659	CCC 647	CCC 398	CCC 602	CCC 632	PRL 1980	RRC 1481	CCC 265	Reference
Ergochrome gene cluster																																																						
	CCE31570.1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1		1	1	1	Neubauer *et al.* 2016
	CCE31571.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1		1	1		Neubauer *et al.* 2016
	CCE31572.1	2	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1		2	2	2	2		2	2	2	Neubauer *et al.* 2016
	CCE31573.1	1	2													1		1					1							1	1	2		1			1	1	1	1					1											Neubauer *et al.* 2016
	CCE31574.1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1		1	1		Neubauer *et al.* 2016
	CCE31575.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	Neubauer *et al.* 2016
	CCE31576.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		2	1	1	1	1	1	1		Neubauer *et al.* 2016
	CCE31577.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	Neubauer *et al.* 2016
	CCE31578.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1									1	1	Neubauer *et al.* 2016
	CCE31579.1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Neubauer *et al.* 2016
	CCE31580.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Neubauer *et al.* 2016
	CCE31581.1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1		1	1		1		1	Neubauer *et al.* 2016
	CCE31582.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1		1	1		Neubauer *et al.* 2016
	CCE31583.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	Neubauer *et al.* 2016
	CCE31584.1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1			1	1	1	Neubauer *et al.* 2016
EAS gene cluster																																																						
easP	AET79200																																																				1		Schardl *et al.* 2013
lpsC	CCE30237.1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1						1				1						1	1	1		Schardl *et al.* 2013
easA	CCE30236.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	Schardl *et al.* 2013
lpsB	CCE30235.1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	2	1	1	1	2	2	2	2	1	2	1	1	1	1	2	2	1	1		1		1				1	Schardl *et al.* 2013
cloA	CCE30234.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1					1	2		Schardl *et al.* 2013
easC	CCE30233.1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	2	3	2	3	3	2	3	2		Schardl *et al.* 2013
easD	CCE30232.1	2	2	2	2	3	2	1	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	1	2	1	1	1	1	1	2	1	1	1	1	2	2	1	1	1		1	1	1	1	1	1		Schardl *et al.* 2013
easE	CCE30231.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	Schardl *et al.* 2013
easF	CCE30230.1	2	2	2	2	1	1	3	3	3	2	3	3	3	3	2	2	3	3	2	4	1	2	2	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1		1		1	1	1	1	1	Schardl *et al.* 2013
easG	CCE30229.1	1	1	1	1	1	1	1	1	1		1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1		1		1	1	1	Schardl *et al.* 2013
dmaW1	CCE30228.1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Schardl *et al.* 2013
easO	AET79193																																																			1		Schardl *et al.* 2013
easH	CCE30227.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1		1		1			1			Schardl *et al.* 2013
lpsA1	CCE30226.1	3	3	1	1	2	1	2	1	1	2	2	1	2	1	3	1	1	1	1	3	1	1	1	1	3	1	1	1	3	1	1	2	1	1	2	1	1	1	3	1	1	1	1	1		1		1		1	1	1	1	Schardl *et al.* 2013
lpsA2	CCE30225.1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	2	1	1	1							1		1	1	1	1	1	1				1	1	1	1	1			Schardl *et al.* 2013
dmaW2	CCE30257.1	2	1		1			1	2	1	1			1	1	1	1		1	1			1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				1									Schardl *et al.* 2013

†BLAST hit with 29% coverage and 45% identity

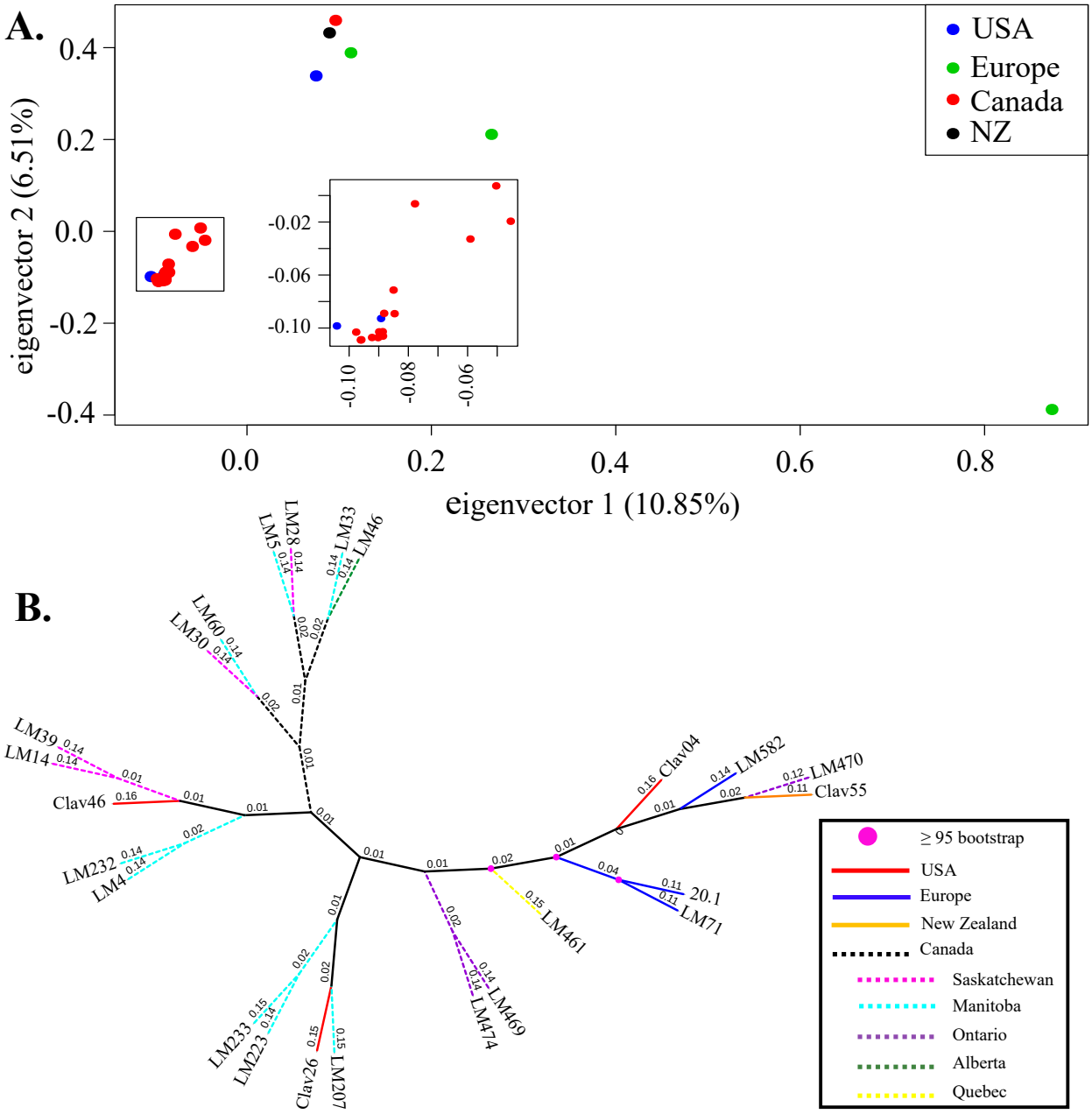
Appendix 2 Table A2.6: Please see Supplemental File 1 TableA2.6

Appendix 2 Table A2.7: Please see Supplemental File 1 TableA2.7

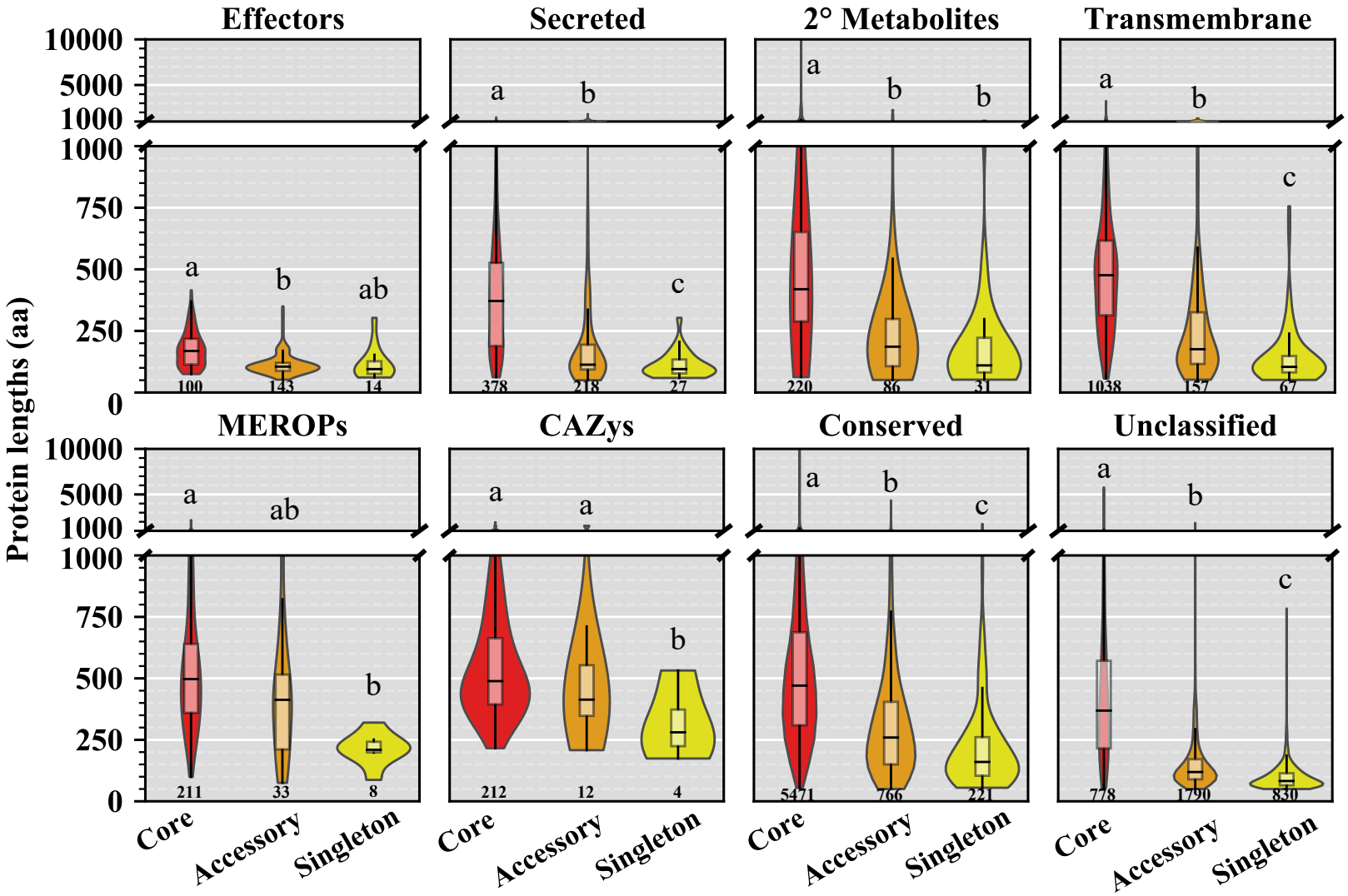
Appendix 2 Table A2.8: Please see Supplemental File 2 TableA2.8

Appendix 3: A large accessory genome, high recombination rates, and selection of secondary metabolite genes help maintain global distribution and broad host range of the fungal plant pathogen *Claviceps purpurea*

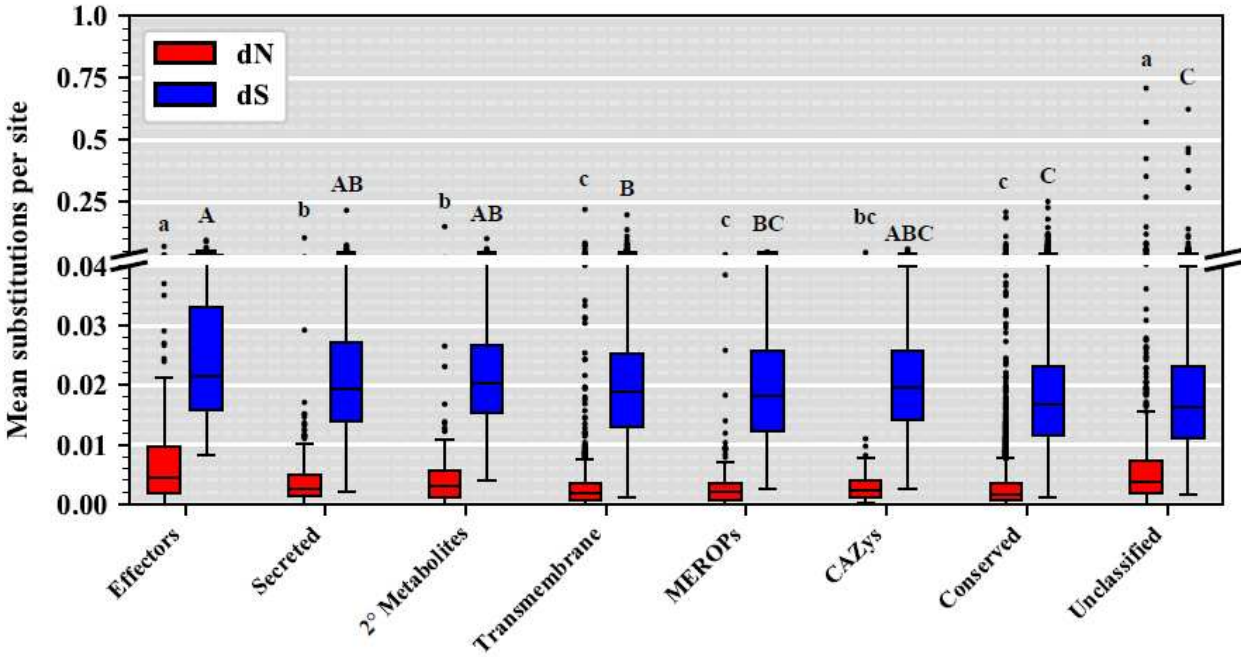
Supplemental Figures and Tables for Chapter 4



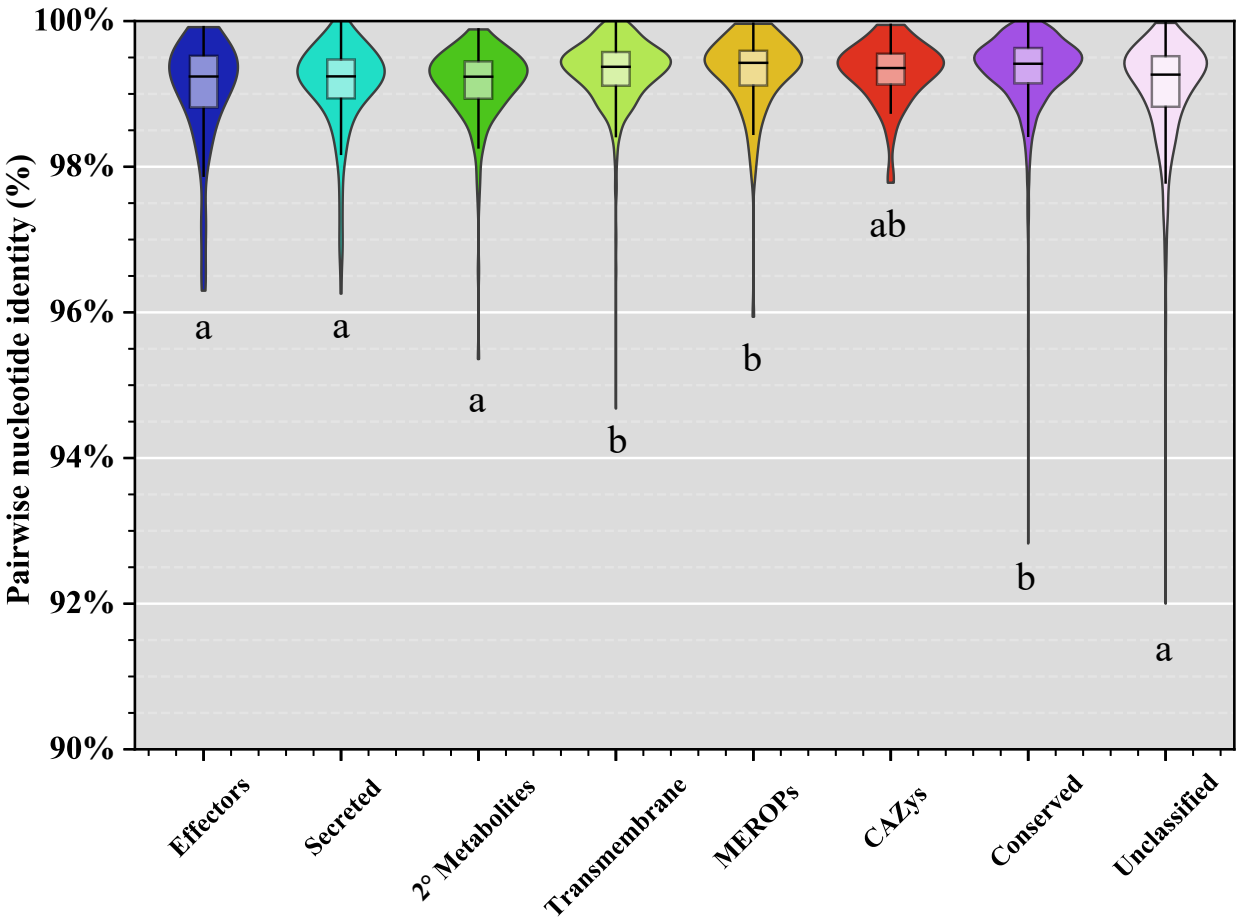
Appendix 3 Figure A3.1: Comparison of 24 *Claviceps purpurea* isolates used in this study, based on 1,076,901 bi-allelic SNPs generated from whole genome alignments (Table 2). A) Principal component analysis of SNP matrix with variation explained by each of the axes shown in parentheses. B) Maximum likelihood phylogeny of all isolates used in this study. Colors of branches depict geographical location and pink dots at nodes represent ≥ 95. Branches were reduced for visual purposes, branch lengths were instead written above each line.



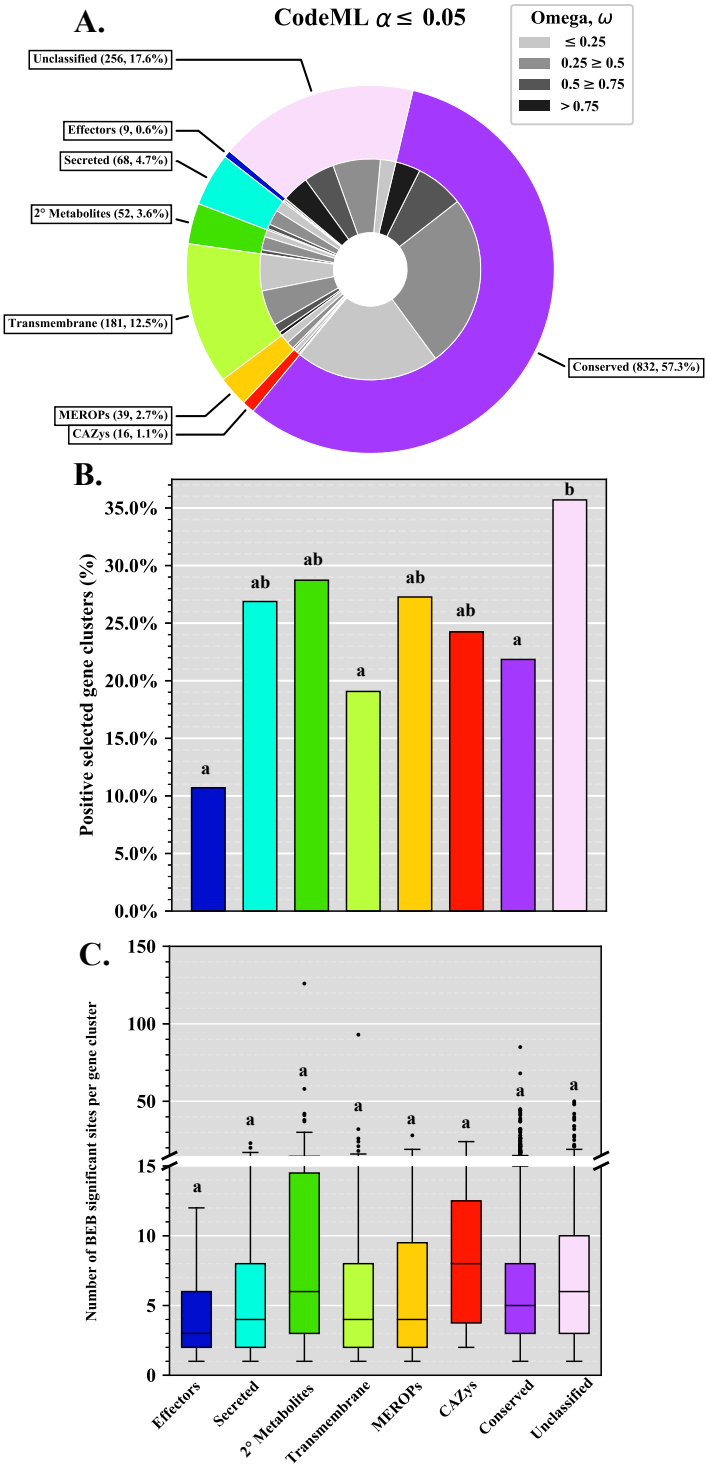
Appendix 3 Figure A3.2: Average protein lengths (aa) of all orthogroups identified in each of the *Claviceps purpurea* pangenome categories; core (shared between all isolate), accessory (shared between ≥ 2 isolates, but not all), and singletons (found in only one isolate). Orthogroups are categorized by predicted protein function if ≥50% of the isolates present in the orthogroups had one gene identified as such. Different letters represent significant differences determined by Kruskal-Wallis with post hoc multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$), with each functional category.



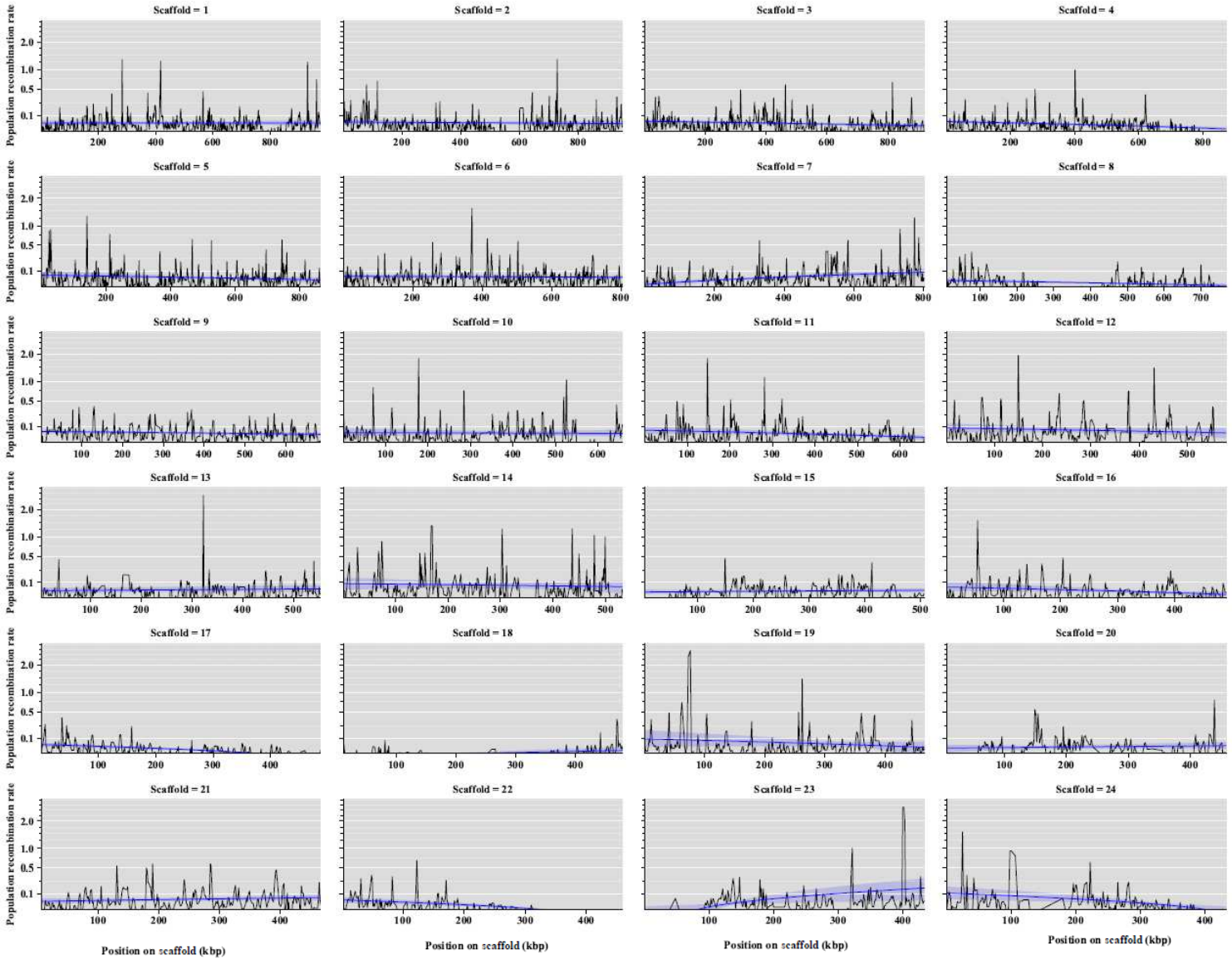
Appendix 3 Figure A3.3: Boxplot distributions of mean non-synonymous (dN) and synonymous (dS) substitution rate of core single-copy orthogroups in *Claviceps purpurea*, for each predicted functional category. Orthogroups are categorized by predicted protein function if ≥50% of the isolates present in the orthogroups had one gene identified as such. Different letters represent significant differences determined by Kruskal-Wallis with post hoc multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$), across each substitution category (lower case = dN comparison, upper case = dS comparison).



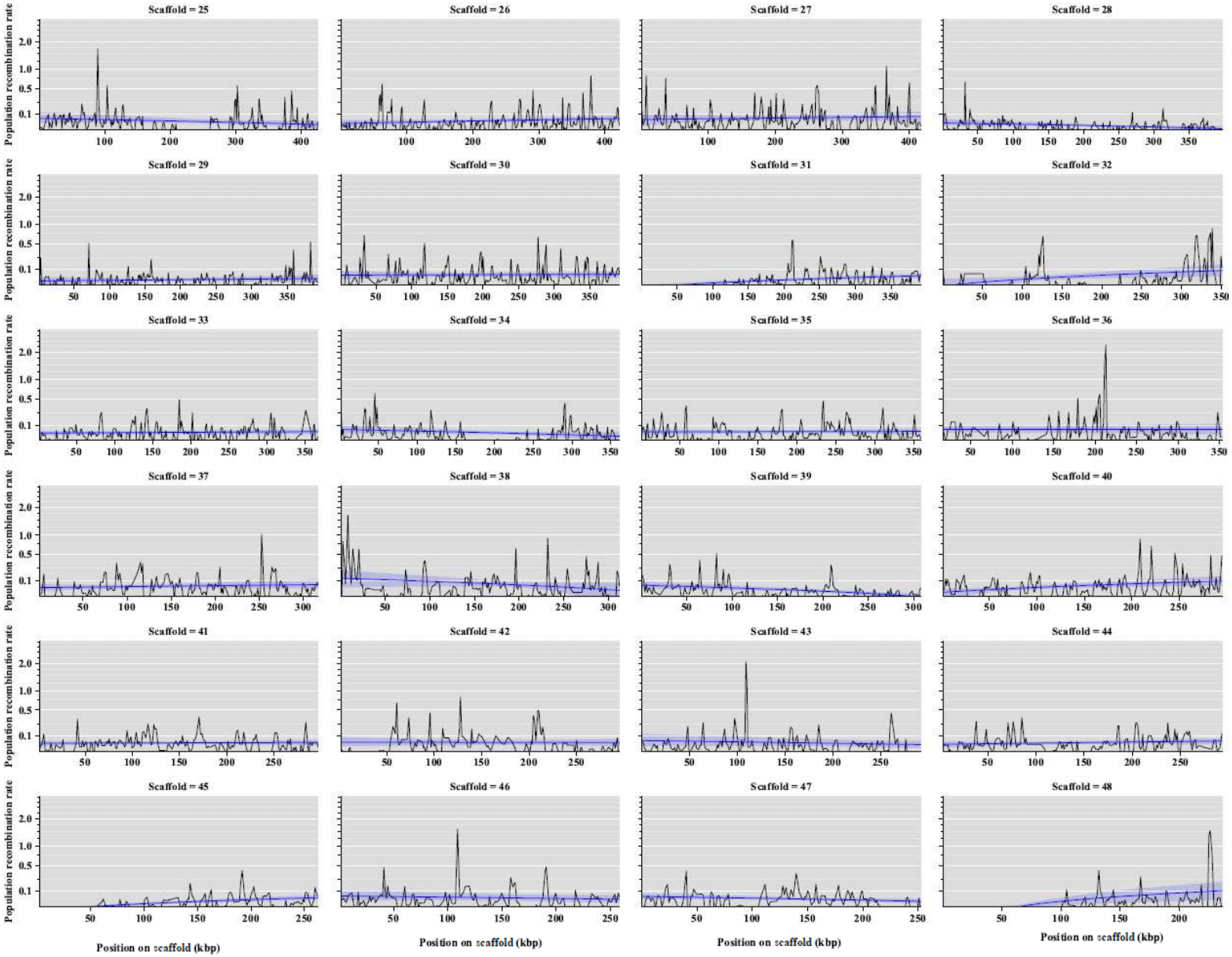
Appendix 3 Figure A3.4: Violin plot distributions of mean nucleotide identity (%) of core single-copy orthogroups in *Claviceps purpurea*, for each predicted functional category. Orthogroups are categorized by predicted protein function if ≥50% of the isolates present in the orthogroups had one gene identified as such. Different letters represent significant differences determined by Kruskal-Wallis with post hoc multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$).



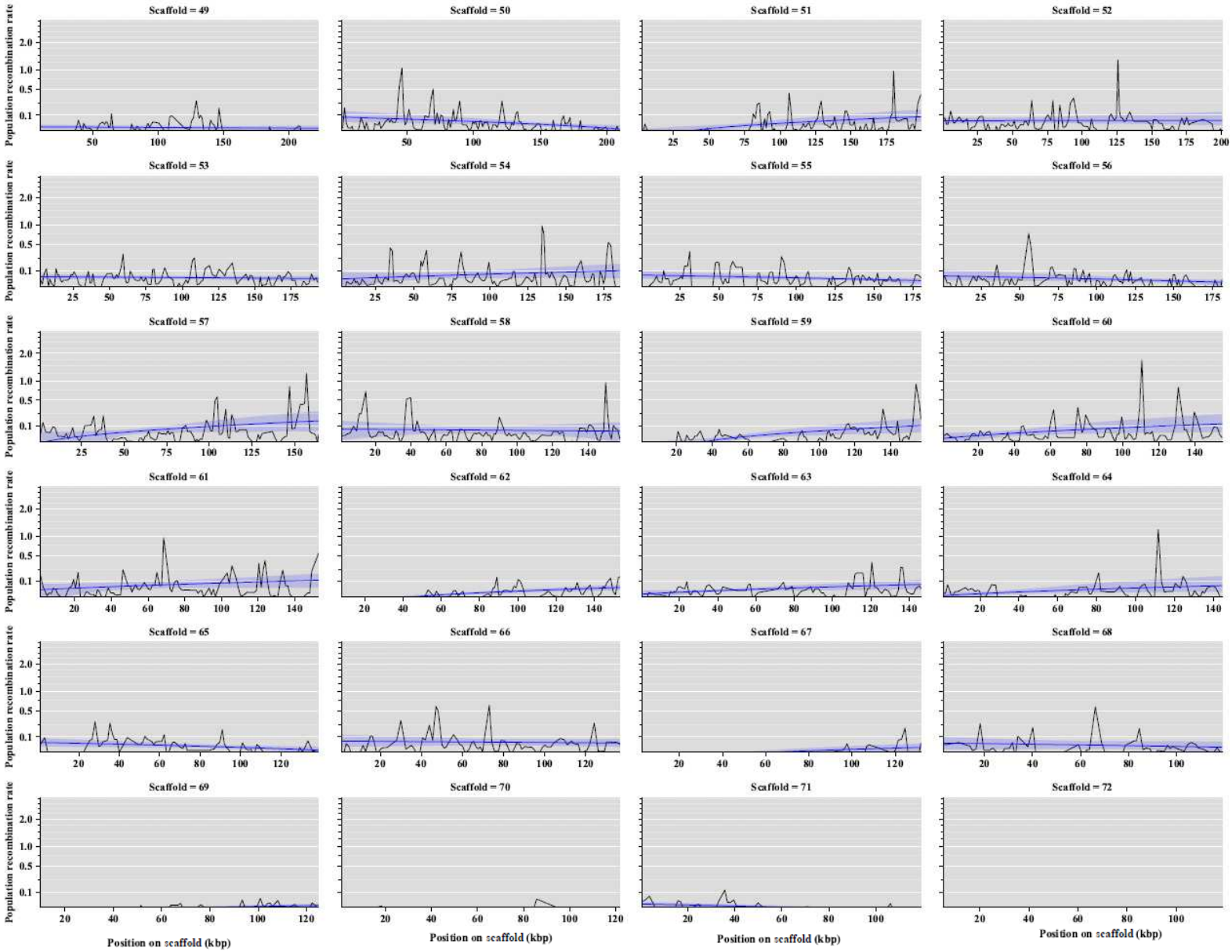
Appendix 3 Figure A3.5: Positive selection landscape of core single-copy orthogroups protein functional categories as predicted by PAML with the CodeML algorithm. Genes with positive selection signatures were selected after a stringent filtering around and $\alpha \leq 0.05$. A) The total number of orthogroups in functional categories with signatures of positive selection. B) The proportion of orthogroups in each functional category based on the number of orthogroups examined in each category (outer circle). Omega (ω, dN/dS) ratios of orthogroups within each functional category(inner circle). C) The number of codons with selection signatures in the M8 model of CodeML, as determined by the Bayes Empirical Bayes (BEB) algorithm with an $\alpha \leq 0.01$. Different letters represent significant differences determined by Kruskal-Wallis with post hoc multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$).



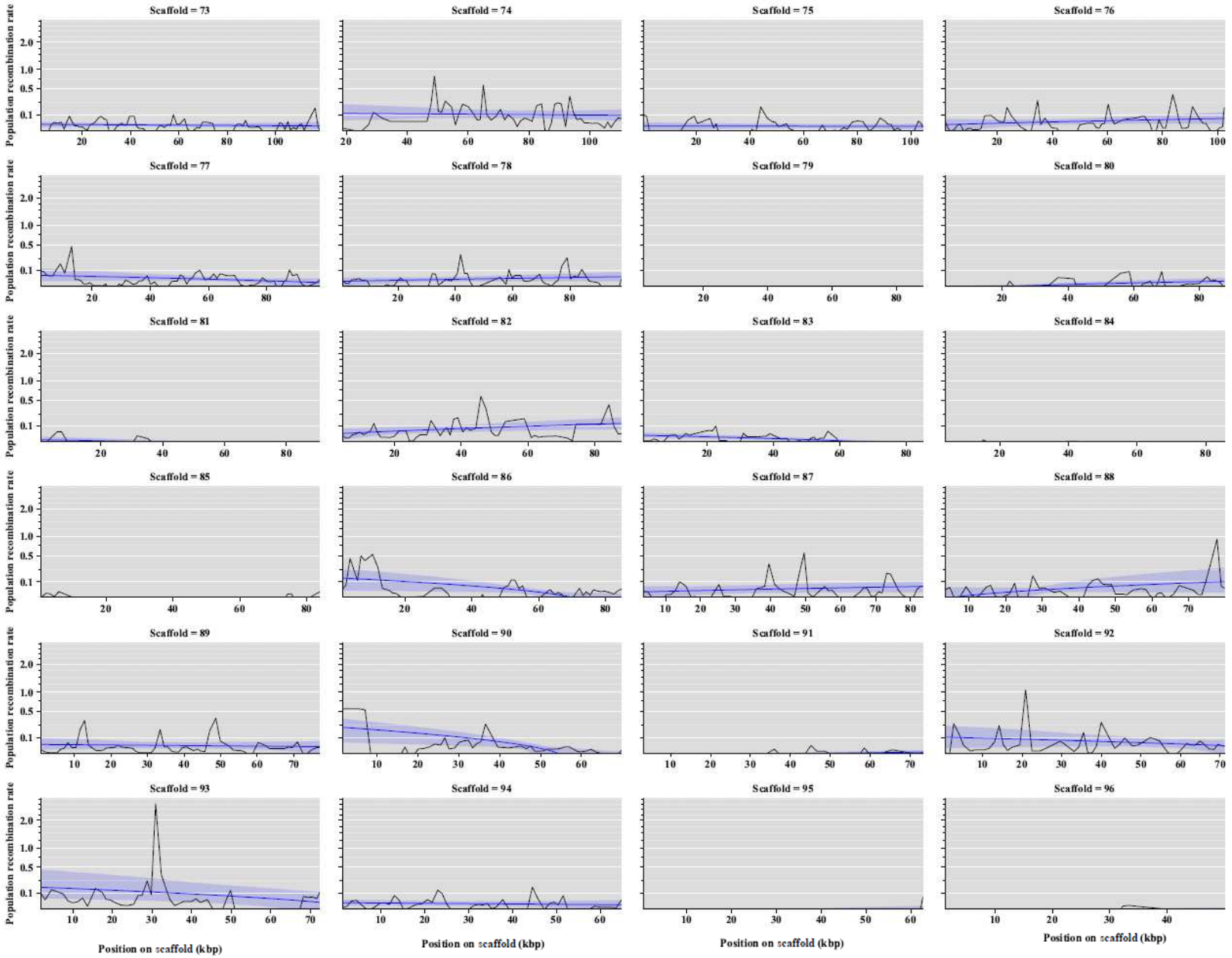
Appendix 3 Figure A3.6: Estimates of population recombination rates (ρ), in non-overlapping 1 kb windows, across four representative scaffolds displaying the different variation observed across the *Claviceps purpurea* genome. Smoothing curves were calculated from population recombination rates in 10 kb windows. Scaffolds with apparent lack of lines showed mean estimated population recombination rates of near 0 across the region.



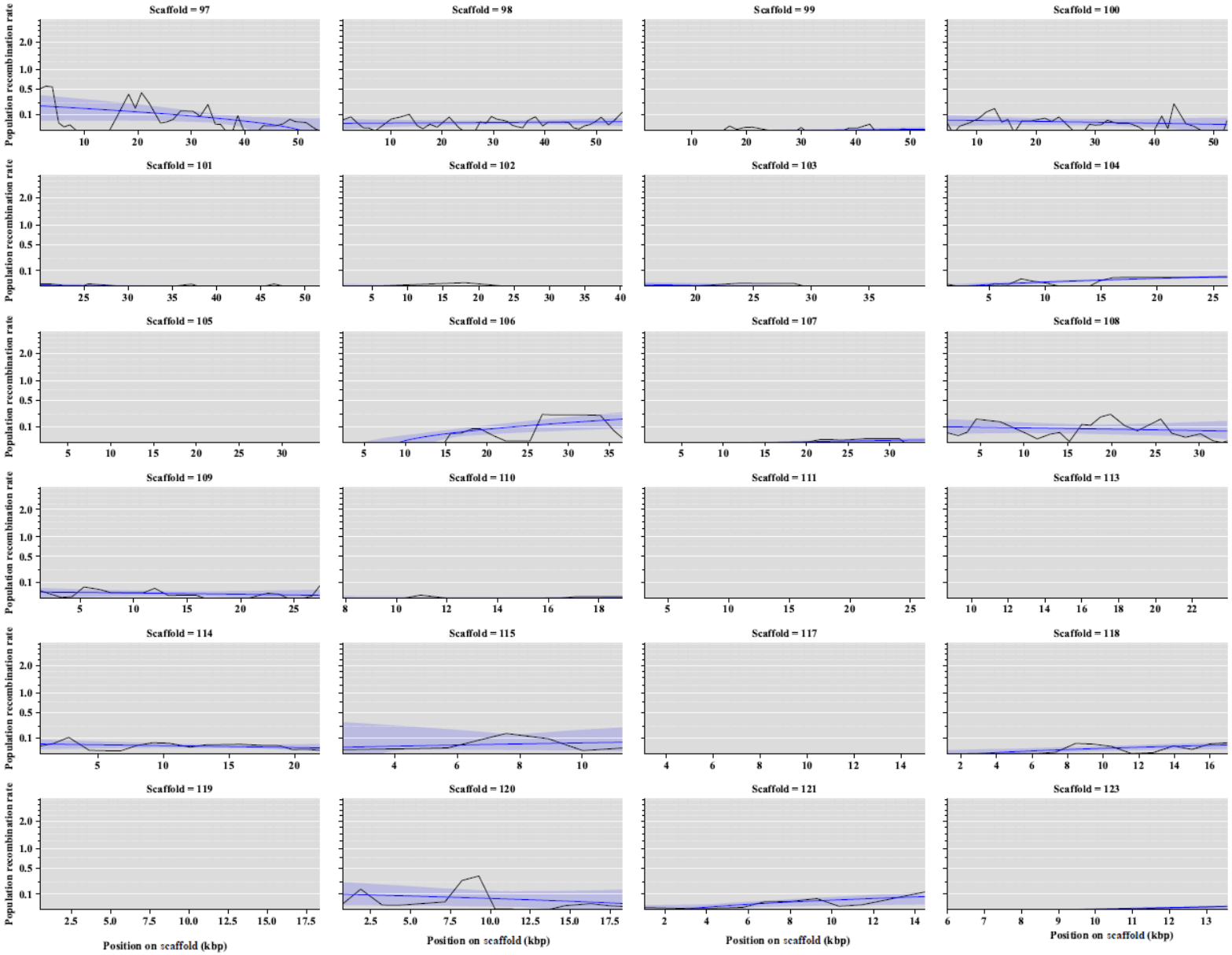
Appendix 3 Figure A3.6: Continued, scaffolds 25-48



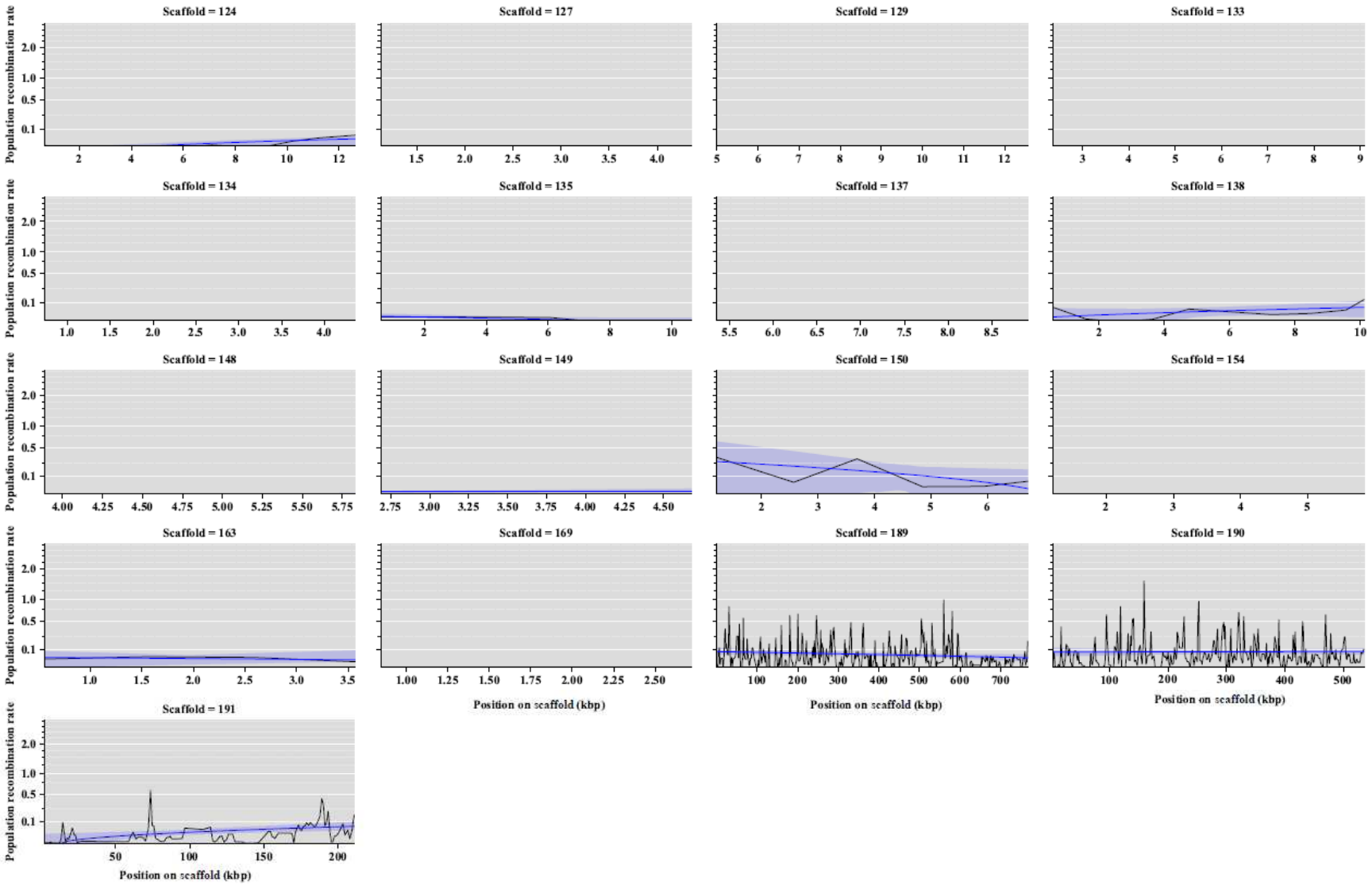
Appendix 3 Figure A3.6: Continued, scaffolds 49-72



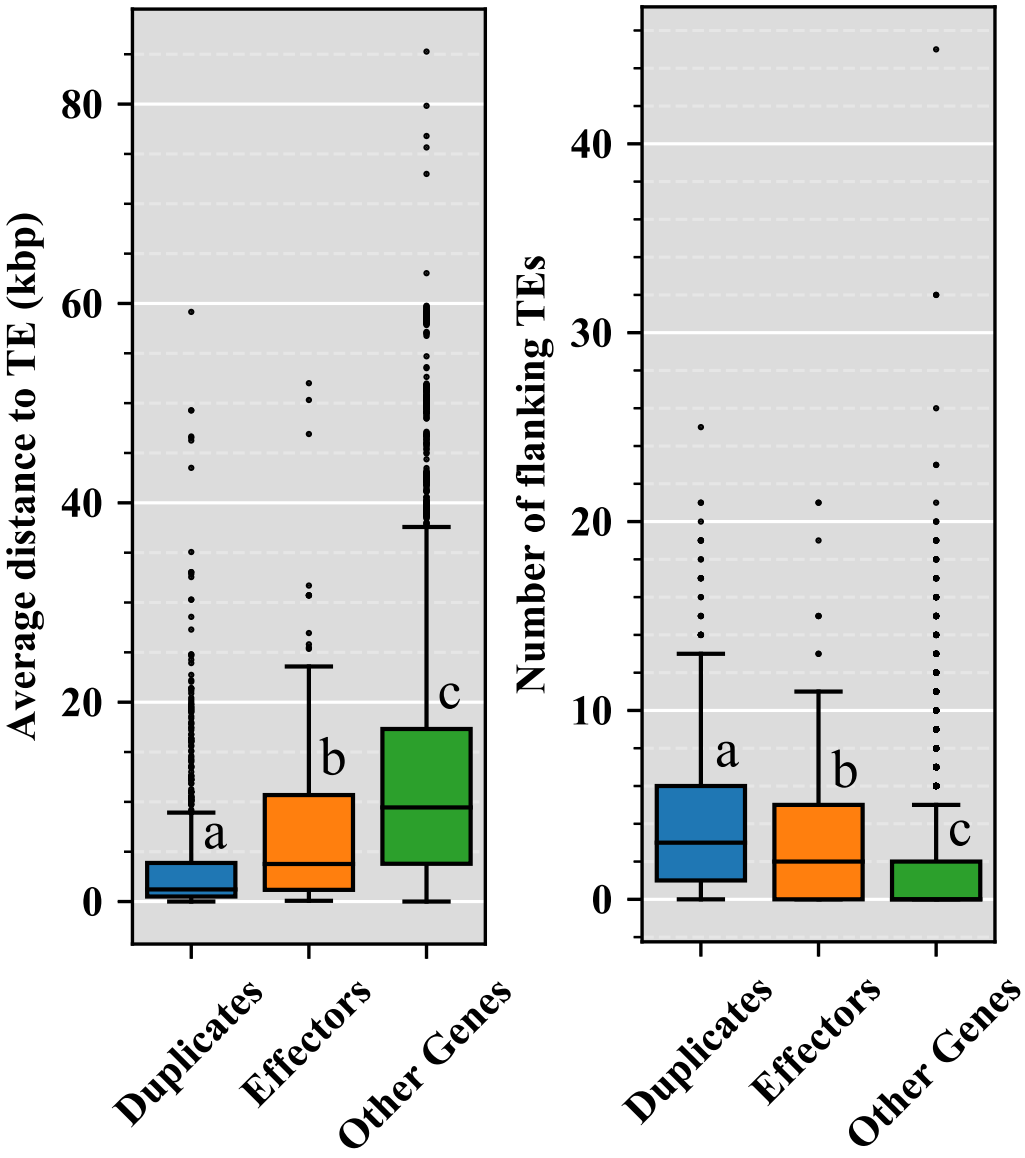
Appendix 3 Figure A3.6: Continued, scaffolds 73-96



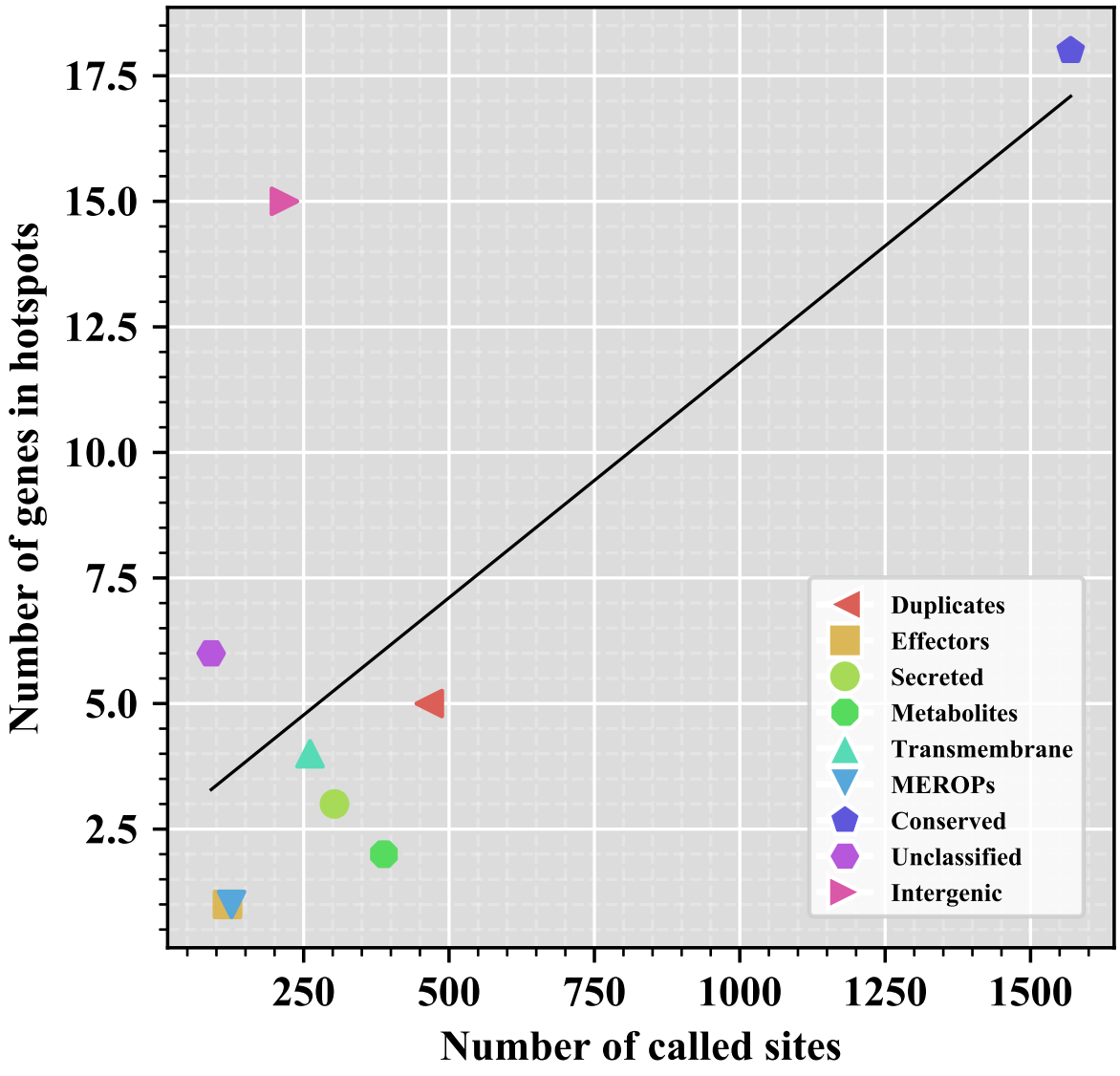
Appendix 3 Figure A3.6: Continued, scaffolds 97-123



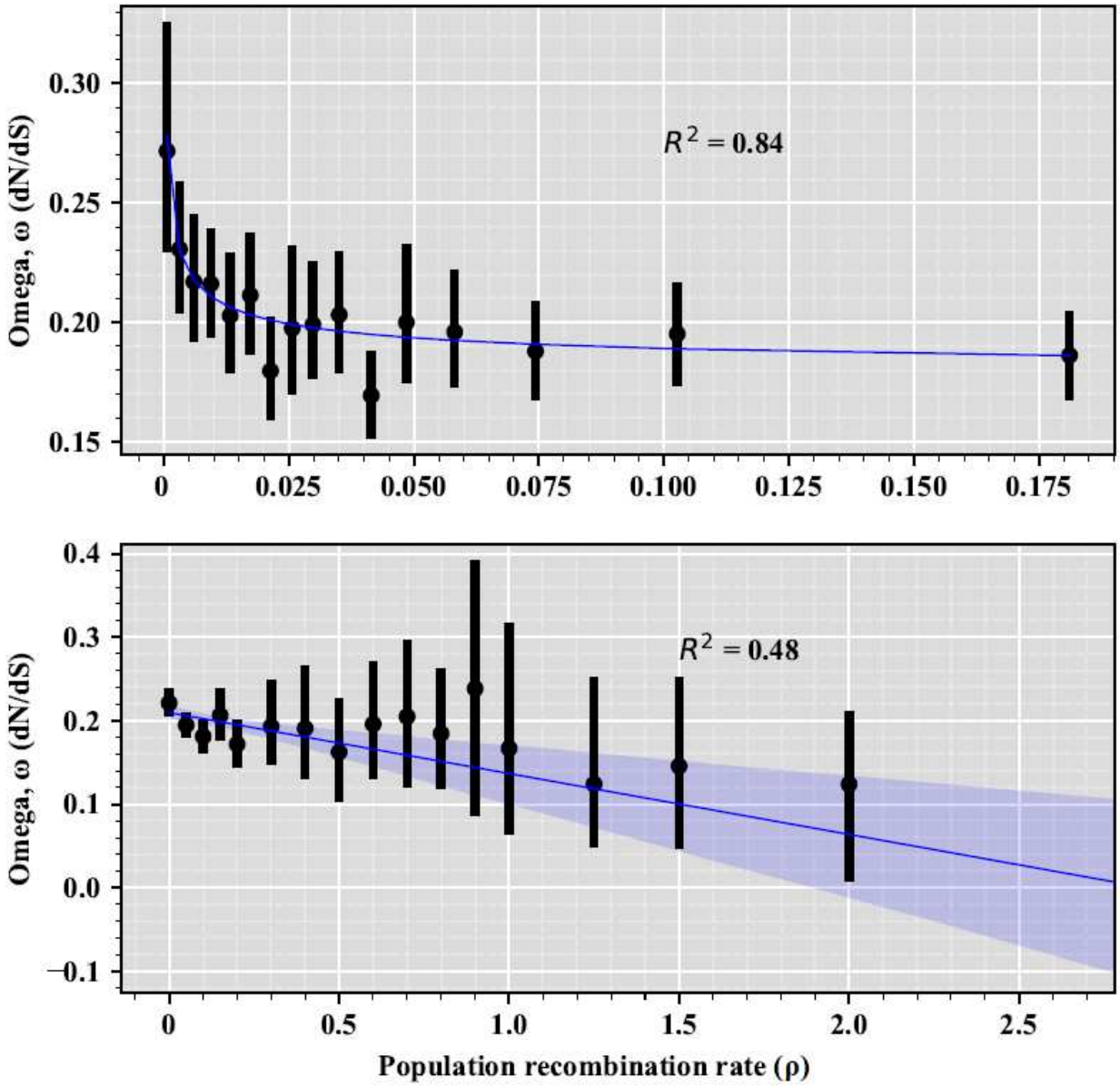
Appendix 3 Figure A3.6: Continued, scaffolds 124-191



Appendix 3 Figure A3.7: Boxplot distributions of putative duplicated genes (≥ 80% identity), predicted effectors, and all other genes in *Claviceps purpurea* showing the mean distance (kbp) of each gene to the closest transposable element (TE) fragment (5' and 3' flanking distances were averaged together) and the mean number of flanking TE fragments. Different letters represent significant differences determined by Kruskal-Wallis with post hoc multi-test corrected Mann-Whitney U Test ($\alpha \leq 0.01$).



Appendix 3 Figure A3.8: Number of genes and intergenic regions within the boundaries of the five predicted recombination hotspots in *Claviceps purpurea* as a function of the number of called sites. Line corresponds to ordinary least square regression.



Appendix 3 Figure A3.9: Correlation of estimated population recombination rate to omega, (ω, dN/dS) ratios of all core single-copy orthologs. Points represent median values and error bars indicate the first and third quartiles of each distribution. The x-axis was binned two different ways for clarity of visualization. A) Bins with equal point densities. Medians of bins were fit to a power law regression y = Ax-B + C. B) Bins with unequal densities centered around population recombination rate values of 0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.25, 1.5, 2. Original (non-binned) data was fit to a linear regression, shaded regions depicts 95% confidence interval.

Appendix 3 Table A3.1: Please see Supplemental File 3 TableA3.1

Appendix 3 Table S3.2: Domains that are significantly enriched (Fischer's Exact test with Benjamini-Hochberg and Bonferroni FDR cutoff α = 0.05) of all pangenome orthogroups.

Pangenome Category	Pfams	Association	*P*-value	Iprscan	Association	*P*-value	GO	Association	*P*-value	MEROPs	Association	*P*-value
Core	PF00400	WD domain, G-beta repeat	0.016	IPR027417	P-loop containing nucleoside triphosphate hydrolase	0.0001	GO:0005515	protein binding (M)	2.86E-06			
	PF07690	Major Facilitator Superfamily	0.024	IPR016024	Armadillo-type fold	0.025	GO:0016021	integral component of membrane (C)	0.0001			
				IPR036322	WD40-repeat-containing domain superfamily	0.029	GO:0055114	oxidation-reduction process (B)	0.0001			
				IPR015943	WD40/YVTN repeat-like-containing domain superfamily	0.037	GO:0005524	ATP binding (M)	0.0002			
				IPR017986	WD40-repeat-containing domain	0.045	GO:0003824	catalytic activity (M)	0.0002			
				IPR001680	WD40 repeat	0.046	GO:0055085	transmembrane transport (B)	0.0007			
							GO:0005634	nucleus (C)	0.0019			
							GO:0016491	oxidoreductase activity (M)	0.0023			
							GO:0008270	zinc ion binding (M)	0.0078			
							GO:0016020	membrane (C)	0.0084			
							GO:0008152	metabolic process (B)	0.0132			
							GO:0005488	binding (M)	0.0172			
							GO:0003676	nucleic acid binding (M)	0.0381			
Accessory										MER0001399	penicillolysin / fungal acid metalloendopeptidase	0.016
Singleton	PF03221	Tc5 transposase DNA-binding domain	0.019									

Appendix 3 Table A3.3: Domains that are significantly enriched (Fischer's Exact test with Benjamini-Hochberg and Bonferroni FDR cutoff α = 0.05) of pangenome orthogroups that contain paralogs.

Pangenome Category	Pfams	Association	*P*-value	Iprscan	Association	*P*-value	GO	Association	*P*-value
Core	PF00067	Cytochrome P450	0.0380	IPR017972	Cytochrome P450	0.028	GO:0004555	alpha,alpha-trehalase activity (M)	0.035
							GO:0005991	trehalose metabolic process (B)	0.035
							GO:0016705	oxidoreductase activity (M)	0.042
Accessory	PF01636	Phosphotransferase enzyme family	9.44E-06	IPR011009	Protein kinase-like domain superfamily	6.82E-09	GO:0006468	protein phosphorylation (B)	0.002
	PF00249	Myb-like DNA-binding domain	0.0010	IPR002575	Aminoglycoside phosphotransferase	1.01E-05	GO:0004672	protein kinase activity (M)	0.002
	PF10551	MULE transposase domain	0.0216	IPR008266	Tyrosine-protein kinase, active site	4.08E-04	GO:0015074	DNA integration (B)	0.039
	PF13921	Myb-like DNA-binding domain	0.0498	IPR009057	Homeobox-like domain superfamily	0.002			
				IPR000719	Protein kinase domain	0.005			
				IPR001005	SANT/Myb domain	0.013			
				IPR017877	Myb-like domain	0.034			

Appendix 3 Table A3.3: Continued

Pangenome Category	MEROPs	Association	*P*-value	smCOGS	Association	*P*-value	CAZys	Association	*P*-value
Core				SMCOG1127	condensation domain-containing protein	0.037	GH37	α,α-trehalase	0.0049
				SMCOG1034	cytochrome P450	0.049			
Accessory	MER0093133	acid prolyl endopeptidase	0.0081						
	MER0001399	penicillolysin / fungal acid metalloendopeptidase	0.03099						

Appendix 3 Table A3.4: Please see Supplemental File 3 TableA3.4

Appendix 3 Table A3.5: BLASTp results for classified single-copy core orthologs with an ω (dN/dS) ≥ 1, signifying the potential for positive selection across the gene. Results were filtered for hits >50% query cover, >35% percent identity, and E-evalue < 10E-5.

Orthogroup ID	Classification	Omega	Reference protein ID	Protein relation	Query Cover (%)	E-value	Percent Identity (%)	Accession
OG0000853	Conserved	1.17	CCE31653.1	related to DNA-directed RNA polymerase II chain RPB9	98%	6.0E-75	65.45%	XP_018146545.1
OG0000872	Conserved	1.28	CCE33395.1	related to Ubiquitin-conjugating enzyme E2C-binding protein	96%	1.0E-142	53.61%	XP_014546302.1
OG0000928	Conserved	1.33	CCE28346.1	probable VID24-required for vacuolar import and degradation of Fbp1p	98%	6.0E-133	59.66%	XP_007815430.1
OG0000930	Conserved	3.59	CCE28348.1					
OG0001034	Conserved	1.06	CCE30521.1	related to DNA mismatch repair protein PMS1	50%	6.0E-128	64.53%	KIE01082.1
OG0001091	Conserved	1.38	CCE34001.1	related to RNA polymerase II, large subunit	98%	3.0E-180	68.65%	XP_018145151.1
OG0001093	Transmembrane	9.79	CCE33999.1	related to bacteriophage N adsorption protein A c-term domain	100%	6.0E-41	88.57%	XP_018183192.1
OG0001193	Conserved	1.13	CCE29986.1	related to TFIIIC transcription initiation factor complex subunits Tfc3	98%	0.0E+00	44.23%	XP_007808740.1
OG0001195	Conserved	1.33	CCE29989.1	related to pantetheine-phosphate adenylyltransferase family protein	98%	1.0E-145	57.88%	XP_007808702.1
OG0001306	Effectors	1.39	CCE29309.1					
OG0001455	Conserved	1.16	CCE31996.1	related to BTB/POZ protein	74%	3.0E-38	40.53%	KAB8071045.1
OG0001456	Transmembrane	1.49	CCE34002.1	related to the member of the syntaxin family of t-SNAREs TLG2	98%	4.0E-152	71.67%	POR33458.1
OG0001471	Effectors	1.47	CCE35277.1					
OG0001489	Conserved	1.30	CCE32298.1	probable regulator of conidiation rca-1	40%	4.0E-47	59.35%	CCE35394.1
OG0001650	Transmembrane	3.80	CCE34357.1	related to peroxisomal ATP carrier	99%	0.0E+00	78.51%	TWU74751.1
OG0001912	Conserved	1.10	CCE29445.1	related to key lime pathogenicity protein	62%	3.0E-26	36.82%	XP_007808105.1
OG0002007	Conserved	1.20	CCE31134.1	related to pre-mRNA-splicing factor 38B	100%	3.0E-158	49.34%	OAA38572.1
OG0002052	Conserved	1.21	CCE28974.1	related to telomere capping protein	82%	1.0E-42	57.14%	EXU96111.1
OG0002180	Conserved	1.03	CCE32158.1	related to C2H2 type zinc finger domain	90%	1.0E-33	53.60%	XP_014548912.1
OG0002663	Conserved	1.15	CCE27977.1	related to single-stranded DNA-binding protein	93%	7.0E-68	72.59%	KND94862.1
OG0002947	Conserved	1.44	CCE34422.1	related to Homeodomain-like protein	92%	2.0E-63	37.39%	KID93050.1
OG0003047	Effectors	1.06	CCE29577.1					
OG0003117	Conserved	1.15	CCE26831.1	related to Acyl-CoA N-acyltransferase	93%	1.0E-119	63.64%	XP_014545930.1
OG0003335	Metabolites	1.03	CCE31863.1	related to DNA polymerase III subunits gamma and tau-like protein	100%	5.0E-74	44.38%	EXV05995.1
OG0003631	Metabolites	1.33	CCE30401.1					
OG0003871	Transmembrane	1.06	CCE32166.1					
OG0004027	Conserved	1.22	CCE31220.1	related to XLF (XRCC4-like factor) family protein	100%	0.0E+00	56.78%	XP_018140513.1
OG0004135	Conserved	1.21	CCE30681.1	related to transcription factor Cys6	51%	2.0E-118	48.41%	XP_014543556.1
OG0004212	Conserved	1.35	CCE32238.1					
OG0004282	Conserved	1.07	CCE30322.1	related to GTP-binding domain, HSR1-related protein	96%	2.0E-124	50.11%	KID93173.1
OG0004586	Conserved	1.01	CCE33815.1	related to ribonuclease h2 subunit c	72%	1.0E-13	37.50%	RFN42344.1
OG0005242	Conserved	1.43	CCE33985.1	related to zinc finger protein	97%	0.0E+00	73.91%	XP_018145163.1
OG0005619	Conserved	1.02	CCE29609.1	probable Vacuolar ATP synthase subunit G	47%	3.0E-57	84.68%	XP_013942611.1
OG0005679	Transmembrane	1.76	CCE27604.1					
OG0006470	Secreted	1.11	CCE27039.1	related to Zinc finger, CCHC-type	98%	8.0E-56	44.54%	OAA48434.1
OG0006472	Metabolites	1.09	CCE32432.1	related to alpha-1,3-mannosidase family protein	96%	2.0E-169	50.00%	KDB15318.1
OG0006565	Effectors	1.96	CCE31639.1					
OG0006715	Conserved	1.40	CCE29668.1					

Appendix 3 Table A3.6: Domains that are significantly enriched (Fischer's Exact test with Benjamini-Hochberg and Bonferroni FDR cutoff α = 0.05) of examined core orthogroups that show significant ($P < 0.01$) signatures of positive selection from PAML and CodeML.

Cluster Classification	Pfams	Association	*P*-value	Iprscan	Association	*P*-value	GO	Association	*P*-value	smCOGs	Association	*P*-value
Metabolites	PF00550	Phosphopantetheine attachment site	0.033	IPR006162	Phosphopantetheine attachment site	0.01	GO:0031177	phosphopantetheine binding (M)	0.013	SMCOG1002	AMP-dependent synthetase and ligase	0.033
				IPR036736	Polyketide synthase, phosphopantetheine-binding domain	0.047	GO:0008152	metabolic process (B)	0.022			
				IPR009081	Phosphopantetheine binding ACP domain	0.047	GO:0003824	catalytic activity (M)	0.037			
Transmembrane	PF00394	Cu-oxidase	0.049									
	PF07731	Cu-oxidase type 2	0.049									
	PF07732	Cu-oxidase type 3	0.049									
Conserved	PF07992	Pyridine nucleotide-disulphide oxidoreductase	0.049									
MEROPS	PF07859	alpha/beta hydrolase fold	0.038									

Appendix 3 Table A3.7: Scaffolds in the reference strain (*Claviceps purpurea* strain 20.1) absent in the whole genome alignment after processing and filtering.

Scaffold	Length (bp)	GC content (%)	Repeat† content (%)	Gene ID	Pangenome	Association	Pfam domains	IPRscan domains
128	12,799	49.38%	9.38%	CCE34836.1	Accessory			
				CCE34837.1	Accessory			
				CCE34838.1	Accessory	Reverse transcriptase	PF07727	IPR013103
130	12,395	51.45%	6.15%	CCE34842.1	Accessory	Reverse transcriptase, Ribonuclease H	PF05699	IPR008906, IPR012337
				CCE34843.1	Accessory	Reverse transcriptase		IPR012337
				CCE34844.1	Accessory	Chromatin, Reverse transcriptase, Integrase	PF00385, PF17921, PF17917, PF17919, PF00665, PF09337	IPR000953, IPR015416, IPR036397, IPR023780, IPR001584, IPR016197, IPR012337
				CCE34845.1	Accessory	Reverse transcriptase	PF00078	IPR000477
131	11,987	50.27%	12.15%	CCE34846.1	Accessory			
				CCE34847.1	Accessory	Reverse transcriptase, Ribonuclease H	PF05699	IPR008906, IPR012337
116	21,920	50.87%	9.16%	CCE34790.1	Accessory			
				CCE34791.1	Accessory			
				CCE34792.1	Accessory			
122	14,441	49.17%	18.15%	CCE34814.1	Accessory			
				CCE34815.1	Accessory			
				CCE34816.1	Accessory			
142	9,483	50.88%	11.50%	CCE34875.1	Accessory			
				CCE34876.1	Accessory	SKP1/BTB/POZ domain		IPR011333
155	5,238	52.08%	5.84%	CCE34906.1	Accessory			
157	5,176	56.39%	1.39%	CCE34909.1	Accessory	AAA proteins, Helitron, DNA Helicase	PF13245, PF13604, PF14214	IPR010285, IPR025476, IPR027417
158	5,154	51.49%	11.49%	CCE34910.1	Accessory			
159	4,760	50.29%	7.63%	CCE34911.1	Accessory	Reverse transcriptase		IPR013103
161	4,319	57.68%	6.71%	CCE34913.1	Accessory	Reverse transcriptase, Integrase	PF17921, PF17917, PF17919, PF00665, PF00078	IPR001584, IPR036397, IPR000477, IPR000953, IPR012337, IPR016197
162	4,248	47.93%	1.67%	CCE34914.1	Accessory	Chromatin, Reverse transcriptase, Integrase	PF00385, PF17921, PF17917, PF17919, PF00665, PF00078, PF09337	IPR000953, IPR000477, IPR001584, IPR023780, IPR036397, IPR015416, IPR012337, IPR016197
164	4,041	55.85%	9.45%	CCE34916.1	Accessory	Helitron		IPR025476
166	3,708	49.54%	3.88%	CCE34918.1	Accessory	Reverse transcriptase, Ribonuclease H	PF05699	IPR008906, IPR012337
167	3,608	48.39%	3.52%	CCE34919.1	Accessory			
168	3,435	54.91%	4.72%	CCE34920.1	Accessory	Zinc finger, Ribonuclease H		IPR001878, IPR036397
170	3,331	49.02%	5.67%	CCE34922.1	Accessory			
171	2,957	51.13%	22.12%	CCE34923.1	Accessory			
172	2,904	48.21%	4.34%	CCE34924.1	Accessory			
173	2,654	49.36%	4.07%	CCE34925.1	Accessory	Reverse transcriptase	PF07727	IPR013103

† Includes simple and low complexity repeats

Appendix 3 Table A3.7: Scaffolds in the reference strain (*Claviceps purpurea* strain 20.1) absent in the whole genome alignment after processing and filtering.

174	2,632	62.08%	1.60%	CCE34926.1	Accessory	Endonuclease, Reverse transcriptase, Zinc finger	PF14529	IPR005135, IPR000477, IPR036691, IPR001878
175	2,564	55.34%	8.35%	CCE34927.1	Accessory			
178	2,421	55.31%	8.84%	CCE34930.1	Accessory	Reverse transcriptase, Aspartic peptidase	PF17919	IPR021109
179	2,327	49.46%	4.94%	CCE34931.1	Accessory			
180	2,320	49.05%	3.32%	CCE34932.1	Accessory			
183	2,227	51.68%	17.42%	CCE34935.1	Accessory			
184	2,211	47.90%	9.81%	CCE34936.1	Accessory			
185	2,134	49.44%	4.59%	CCE34937.1	Accessory	Centromere, transcription activity	PF12550, PF16787	IPR038279, IPR022210, IPR031872
186	2,124	50.89%	4.19%	CCE34938.1	Accessory	Pkinase, Tyrosine-protein kinase	PF17667	IPR000719, IPR008266, IPR011009
187	2,039	53.95%	24.91%	CCE34939.1	Accessory	DNA helicase, P-loop triphosphate hydrolase		IPR003840, IPR027417
188	2,010	48.21%	14.63%	CCE34940.1	Accessory			
176	2,538	50.87%	16.12%	CCE34928.1	Core			
112	24,908	49.52%	12.43%	CCE34766.1	Accessory			
				CCE34767.1	Accessory			
				CCE34768.1	Singleton			
				CCE34769.1	Accessory	P-loop triphosphate hydrolase		IPR027417
				CCE34770.1	Accessory			
				CCE34771.1	Core			
126	13,693	48.78%	10.87%	CCE34829.1	Accessory	Reverse transcriptase, Ribonuclease H, Integrase	PF07727	IPR013103, IPR036397, IPR001584, IPR012337
				CCE34830.1	Accessory			
				CCE34831.1	Accessory			
				CCE34832.1	Singleton			
141	9,523	50.53%	4.90%	CCE34871.1	Singleton			
				CCE34872.1	Accessory	Chromatin, Reverse transcriptase, Integrase	PF00385, PF17921, PF17917, PF17919, PF00665, PF00078	IPR000953, IPR000477, IPR023780, IPR036397, IPR001584, IPR016197, IPR012337
				CCE34873.1	Accessory			
144	8,689	49.53%	18.44%	CCE34880.1	Core	Nucleosome assembly	PF00956	IPR002164, IPR037231
				CCE34881.1	Accessory			
				CCE34882.1	Core	Nucleosome assembly	PF00956	IPR002164, IPR037231

† Includes simple and low complexity repeats