2011-03-17

De novo Genome Assembly and SNP Marker Development of *Pyrenophora semeniperda*

Marcus Makina Soliai
Brigham Young University - Provo

De novo Genome Assembly and SNP Marker Development of

Pyrenophora semeniperda

Marcus Makina Soliai

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Craig E. Coleman, Chair
Susan E. Meyer
Joshua A. Udall

Department of Plant & Wildlife Sciences

Brigham Young University

April 2011

ABSTRACT

De novo Genome Assembly and SNP Marker Development of

Pyrenophora semeniperda

Marcus Makina Soliai
Department of Plant and Wildlife Sciences, BYU
Master of Science

Pyrenophora semeniperda (anamorph *Drechslera campulata*) is a necrotrophic fungal seed pathogen of a variety of grass genra and species, including *Bromus tectorum*, an exotic grass that has invaded many natural ecosystems of the U.S. Intermountain West. As a natural seed pathogen of *B. tectorum*, *P. semeniperda* has potential as a biocontrol agent due to its effectiveness at killing dormant *B. tectorum* seeds; however, few genetic resources exist for this fungus. Here, the genome assembly of a *P. semeniperda* isolate using 454 GS-FLX genomic and paired-end pyrosequencing techniques is presented. The total assembly is 32.5 Mb and contains 11,453 gene models greater than 24 amino acids. The assembly contains a variety of predicted genes that are involved in pathogenic pathways typically found in necrotrophic fungi. In addition, 454 sequence reads were used to identify single nucleotide polymorphisms between two isolates of *P. semeniperda*. In total, 20 SNP markers were developed for the purposes of recombination assesment of 600 individual *P. semeniperda* isolates representing 36 populations from throughout the U.S. Intermountain West. Although 17 of the fungal populations were fixed at all SNP loci, linkage disequilibrium was determined in the remaining 18 populations. This research demonstrates the effectiveness of the 454 GS-FLX sequencing technology, for *de novo* assembly and marker development of filamentous fungal genomes. Many features of the assembly match those of other *Pyrenophora* genomes including *P. tritici-repentis* and *P. teres* f. *teres*,

which lend validity to our assembly. These findings present a significant resource for examining and furthering our understanding of *P. semeniperda* biology.

Keywords: 454 sequencing, genome assembly, SNPs, linkage disequilibrium

ACKNOWLEDGEMENTS

I would first like to express my deepest appreciation to my wonderful parents who have been such an inspiration and support throughout my life. I am grateful for the many friendships that I have made over the years in the genetics lab and for the humor, encouragement, and insight they have shown me. Special thanks goes out to my friend David Elzinga for his assistance, advice, and time that he spent helping me progress through my project and for teaching me all I know about the mysteries of the computer terminal and the greatness of Ubuntu. I would like to thank Russell Hermansen, Paul Bodily, Aaron Hart and Daniel Standage for their help with the genome annotation. Thanks to Suzette Clement at the Shrub Sciences Laboratory for supplying me with all of BFOD tissue needed for this project. Special thanks to my graduate advisor, Dr. Coleman, for his patience, advice, encouragement, counsel and all that he has taught me during my time here at BYU. I would also like to thank my graduate committee members, Dr. Meyer and Dr. Udall for their expertise and guidance that they have shown me during the challenging completion of this work. Finally, I would also like to give thanks to all of the genetics professors in the Plant and Wildlife Department for teaching and mentoring me while attending BYU.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 1: Literature Review

Introduction

The fungal kingdom is home to eukaryotic heterotrophs comprised of molds, yeasts, mushrooms, puffballs, truffles, rusts, and smuts. Conservative estimates place 1.5 million species in this kingdom; these are found in many diverse ecosystems throughout the world (Hawksworth, 2001). Their diverse and unique life histories have had many positive effects in industry, furthered our progress in medicine, and added to our knowledge in the biological sciences. In contrast to some of these benefits, fungi can have many devastating and detrimental consequences for many organisms. Many species of parasitic fungi can cause local infections on the skin of humans and animals while others pose life-threatening risks to immuno-compromised patients causing infection in the upper respiratory system, digestive tract, and the central nervous system. Phytopathogenic fungi infect plants and cause severe strain on ecosystems and economic hardship to farmers and others who depend on their crops for sustenance.

Ascomycete fungal species of the genus *Pyrenophora* are well known for their ability to infect and kill major grass species such as corn, barley, and wheat. In particular, *Pyrenophora semeniperda* is a generalist pathogen of many grass species, attacking host seeds and in some cases causing leaf spotting. Recently, there has been interest in using *P. semeniperda* as a biocontrol agent against cheatgrass (*Bromus tectorum*), an invasive grass species that threatens ecosystems throughout the Intermountain West (IMW) of the United States. *Pyrenophora semeniperda* has been suggested as an effective biocontrol for the prevention and control of cheatgrass as it has been shown to successfully infect their seed banks and prevent or severely retard germination (Medd & Campbell, 2005; Meyer, 2008). Selection of *P. semeniperda* fungal isolates capable of this task is paramount to the success of such a project, unfortunately very little is understood about the genetics of this fungus as genetic resources are currently limited.

Over the years, fungi have been important eukaryotic model organisms owing to their simple life cycles and the simplicity of culturing in the laboratory. Many fungi, including *P. semeniperda*, are haploids with genome sizes ranging anywhere between 10-90 Mb, which are relatively small by eukaryotic standards. Sequencing small genomes such as those of fungi remains a relatively simple task due to readily available sequencing technologies. Specifically, 454 sequencing technology provides many advantages when sequencing full genomes in terms of speed, accuracy, read length, and cost.

The *Pyrenophora* Genus

The fungal genus *Pyrenophora* (anamorph *Drechslera*) is host to many graminicolous ascomycete species including *P. teres*, *P. graminea*, *P. tritici-repentis*, and *P. semeniperda*. The geographical ranges of the these species are widespread and have been reported around the world including North and South America, Europe, Africa, and Australia. Species belonging to this group are pathogens of grasses and cereals and although most commonly found in barley, *Pyrenophora* species have also been observed in oats, wheat, and other non-crop grasses such as turf. A wide range of diseases are caused by *Pyrenophora* species, most notably net blotch in barley, barley stripe, and tan spot of wheat and have led to extensive yield loss in cereal crops year after year. Though the main factors of epidemiology of plant pathogens of *Pyrenophora* include leaf necrosis in their hosts, *P. semeniperda* is the only member of this genus that is capable of seed infection (Pfender & Wootke, 1988; Schilder & Bergstrom, 1994).

Pyrenophora is the teleomorphic (perfect or sexual) state of this fungus and is the easiest to differentiate from other closely related fungi however, the anamorph (imperfect or asexual) state of *Pyrenophora*, *Drechslera*, is most commonly observed. *Drechslera* was originally placed in the genus *Helminthosporium* with other graminicolous fungi such as *Cochliobolus* and *Setosphaeria* due to

similar morphology of their conidia (Shoemaker, 1959). *Drechslera* produce asexually; haploid spores called conidia are its primary mode of reproduction. These nonmotile multicellular spores are mostly straight and cylindrical in shape and are born from structures called conidiophores located at the tip of the fungal hypha or on specialized structures called stromata. Conidia are spread over short distances by wind currents, rain, cultivating equipment, and animals (Piening, 1968; Martin, 1984). Rain and air currents have proved to play an important role in the epidemiology of this fungus as they serve as modes of transportation for inter-field dispersion of conidia and an increased rate of infection of host plants (Franel, 1997). Meyer et al. (2008) have also demonstrated the potential for *P. semeniperda* conidia to be dispersed along with seeds of its host, effectively moving beyond the host canopy boundary layer.

In general, *Drechslera* survive the winter as mycelium, threadlike filaments that constitute the vegetative body of a fungus, or as conidia on seeds or plant debris; *Pyrenophora*/*Drechslera* species do not survive well as saprophytes in the soil due to competition from other saprophytic microorganisms (Pfender, 1988). Most species are favored by a humid environment and moderate to warm temperatures (20-30 °C) and are retarded by dry hot weather. In favorable conditions, infection can occur quickly, for example, *P. teres* conidia are capable of germinating as soon as 30 minutes after landing on a host (Hargreaves & Keon, 1983). Modes of infection after the initial germination event slightly differ from species to species and on the type of host tissue being infected.

Pyrenophora tritici-repentis

The homothallic ascomycete, *P. tritici-repentis* (anamorph *Drechslera tritici-repentis*) is an economically important pathogen in leaf disease of wheat causing losses grain yield throughout the world (Watkins et al., 1978; Rees & Platz, 1983). Infection by this fungus causes tan spot and chlorosis

in its hosts and has been reported to be responsible for grain losses ranging anywhere from 3 to 50% in North America (Rees & Platz, 1983; Hosford et al., 1987; Faris et al., 1997).

Toxin production has been studied extensively in *P. tritici-repentis*. Tomás and Bockus (1987) found toxic activity in *P. tritici-repentis* after collecting culture filtrate and exposing it to wheat cultivars. They suggested a relationship between susceptibility to the phytopathogen and sensitivity to the toxic filtrates. Various isolates of *P. tritici-repentis* have been found to vary in their virulence. Krupinsky (1987) inoculated 27 isolates of *P. tritici-repentis* from smooth bromegrass onto wheat and found that these isolates varied in their ability to cause disease. It was later discovered that an isolate of *P. tritici-repentis* (nec+) releases a host specific toxin, Ptr necrosis toxin, found to be associated with the induction of tan necrosis and chlorosis in the host (Tomás & Bockus, 1987; Balance et al., 1989; Lamari & Bernier, 1989; Lamari & Bernier, 1991). Ciuffetti and Tuori (1999) suggest that various isolates of *P. tritici-repentis* produce different combinations of toxin which offers an explanation as to why isolates of this species vary in infection severity.

Recently, researchers at the Broad Institute have been able to sequence the genome of *P. tritici-repentis* as it is an economically and scientifically important phytopathogen (Broad Institute, 2009). All 11 chromosomes have been sequenced and annotated with over 12,000 genes that have been identified and mapped to the *P. tritici-repentis* genome. The genome sequence of this species serves as a reference for the sequencing of *P. semeniperda* and other related fungi, as well as facilitating the work of comparative and functional genomics studies of many pathogenic fungi. Most importantly, the *P. tritici-repentis* genome provides valuable insight in understanding structural and functional aspects of the *P. semeniperda* genome.

Pyrenophora graminea

Pyrenophora graminea (anamorph *Drechslera graminea*) is distributed throughout the world and primarily infects barley, causing leaf stripe on seedlings and adult plants. This fungus is exclusively seed transmitted (Aragona et al., 2000) and can survive on the outer layers of the infected seed before obtaining another host. *Pyrenophora graminea* is particular about its environmental needs as it requires cool, humid conditions in order to thrive and properly infect its hosts. *Pyrenophora graminea* systemically infects its hosts (Aragona et al., 2000) while releasing toxins and other enzymes required for digestion of the infected plant. This in turn causes striping that is observed in affected barley leaves. Infected plants are stunted in their growth and produce sterile spikes; rarely will infected spikes produce seed while late forming and uninfected spikes are usually successful at seed production (Tekauz & Chiko, 1980). *Pyrenophora graminea* was responsible for considerable yield losses in the past with over a 68% loss reported in parts of Russia. Presently, this fungus is easily controlled through various planting methods and seed treatments enforced by barley farmers (Tekauz et al., 1985).

The anamorphs and teleomorphs of *P. graminea* are easily identified by conidia-releasing structures. Asexual forms of this fungi produce chamber-like structures called pycnidia which are pear-shaped fruiting bodies lined with conidiophores. The reproductive spores produced by pycnidia are refered to as pycnidiospores and are able to escape through holes (ostioles). The sexual form of *P. graminea* produce spherical or pear-shaped conidia-releasing structures called perithecia. The distinguishing feature of this structure is a single ostiole which releases conidia one at a time (Mathre, 1997). These structures are also found among other *Pyrenophora* fungi including *P. teres*.

Pyrenophora teres

Pyrenophora teres (anamorph *Drechslera teres*) is the causative agent of net blotch in barley. Net blotch is a common foliar disease in barley, causing between 10 to 40% in annual yield loss,

although, susceptible barley cultivars can experience complete yield loss if left untreated with the proper fungicides (Mathre, 1997). This fungus overwinters on barley seed or crop residue and spreads its conidia via air currents or rain. Like *P. graminea*, *P. teres* is most active in cool, humid weather between 10 – 15 °C when infection and spore production are significantly higher than in colder or warmer temperatures.

Reproductive methods and structures are similar to those in *P. graminea*. In fact, stable and fertile interspecific hybrids were created in a laboratory between *P. teres* and *P. graminea*, indicating a close genetic relationship (Smedegärd-Petersen, 1983). However, there is insufficient data to indicate that this hybridization occurs naturally (Bakonyi & Justesen, 2007). *P. teres* exists in two morphologically indistinguishable but genetically differentiated forms: *P. teres* f. *teres* (net form of net blotch, NFNB) and *P. teres* f. *maculata* (spot form of net blotch, SFNB) (Smedegärd-Petersen, 1971; Rau et al., 2007; Ellwood et al., 2010). In addition, the *P. teres* f. *teres* genome was recently sequenced using Illumina‘s Solexa sequencing platform. Sequencing data were used to develop genetic markers to construct a linkage map of this fungus and identify secreted proteins thought to be involved in the infection strategy of *P. teres* (Ellwood et al., 2010).

Pyrenophora semeniperda

Pyrenophora semeniperda (anamorph *Drechslera campulata*) is a generalist phytopathogen of a wide range of grass generas, infecting seeds and causing leaf spotting in its hosts, most often seen in the field as its *Drechslera* anamorph. Classification of *P. semeniperda* has been particularlly difficult since its discovery over a century and a half ago. As reviewed by Medd et al., *D. campulata* was originally described by Léveillé in 1841 on *Bromus sterelis* seeds and named *Angiopoma campanulatum* Lév. (Léveillé, 1841). Subsequent binary names for this anamorph include *Podosporiella verticillata* O‘Gara (O‘Gara, 1915), *Helminthosporium cyclops* Drchslera (Drechsler

1923), *Bioplaris cyclops* (Drechsler) Sprague (Sprague 1962), *Drechslera verticillata* (O'Gara) Shoem. (Shoemaker 1966), and finally *Drechslera campanulata* which is the widely accepted name of this fungus, based on Léveillé's original description (Medd et al., 2003). The teleomorph, *Pyrenophora semeniperda*, is not as common as its anamorph. It was originally named *Pleosphaeria semeniperda* Brittlebank & Adam (Brittlebank & Adam 1924) and *Pyrenophora horrida* H. Sydow (Sydow 1924) until Shoemaker (1966) reduced the names to *Pyrenophora semeniperda* (Medd et al., 2003).

This fungus is mainly found in temperate grasslands and winter cereal areas throughout the world (IMI, 1995). Léveillé (1841) originally described the fungus in France on infected seeds of *B. sterelis* and has recently been reported in Greece (Stewart et al., 2009). *P. semeniperda* is mainly found in temperate grasslands in South Africa, New Zealand, and North and South America. Barreto and Fortugno (1994) isolated the fungus in Argentina from naturally infected wheat and stated that the fungus occurs sporadically. Records indicate that this fungus is often found on durum (*Triticum durum*) and bread wheat (*Triticum aestivum*), *Bromus pictus*, and *Poa ligularis* in Argentina which suggests that *P. semeniperda* is well established in this country (Medd, 2003).

Black stromata protruding out of its host's seed is characteristic of infection by this fungus. Wallace (1959) reported successful infection in paleae, glumes, and lemmas of oats and wheat after inoculation with *P. semeniperda* conidia. Infection has no evident effect on seed maturation but has been observed to seriously reduce germination and emergence rates (O'Gara, 1915; Brittlebank & Adam, 1924; Kreitlow & Bleak, 1964). Campbell and Medd (2003) describe the infection process of *P. semeniperda* in wheat, of which little was known previously.

Pyrenophora semeniperda has been investigated as a possible biocontrol agent against cheatgrass. *Pyrenophora semeniperda* attacks ungerminated cheatgrass seeds in the seed bank, effectively causing mortality in dormant seeds (Beckstead, 2004; Meyer et al., 2008). The ability of *P.*

semeniperda to effectively cause cheatgrass mortality is directly related to the rate of germination as described by Beckstead et al. (2004); nondormant cheatgrass seeds in the transient seed bank are capable of germinating quickly in autumn thereby avoiding mortality from *P. semeniperda* infection whereas seeds in secondary dormancy (dormant seeds) can suffer high mortality rates from fungal infection (Meyer et al., 2008). Beckstead et al. (2004) believe that the infection strategy and success of *P. semeniperda* is highly dependent on whether or not it is capable of killing the seed before it germinates.

Cytochalasins

Many toxins produced by *Pyrenophora* species as well as those unique to *P. semeniperda* have been isolated and identified (Evidente et al., 2002). However, a group of fungal metabolites called cytochalasins are of particular interest due to their ability to bind to actin filaments and prevent elongation and polymerization. Cytochalasisns are comprised of a group of structurally diverse metabolites, well known for their distinctive biological characteristics. Their ability to bind to actin filaments is an important aspect of cytochalasin biochemistry that has proven useful in undertanding various celluar functions including actin polymerization, cell motility and division, and even apoptosis (Haidle & Myers, 2004). This group of fungal metabolites is also used extensively in cancer research, having been shown to have anti-tumoric effects in mouse models (Bogyo, 1991).

There are many structural varieties of cytochalasins; a tricyclic ring made of an isoindolone moiety is a structural feature common to all members of this group of fungal metabolites. Experiments reveal that cytochalasin biosynthesis involves the formation of acetate and methionine-derived octa- or nonaketide chain and the attachment of an amino acid; depending on the type of cytochalasin metabolite that is synthesized, phenylalanine, tryptophan, or leucine is incorporated.

Evidente et al. (2002) identified and described the structure of Cytochalasins Z1, Z2, and Z3, three previously unknown molecules produced by *P. semeniperda*. Cytochalasin F, T, deoxaphomin and cytochalasin B are also produced by this fungal species, the latter being made in relatively high amounts. Cytochalasins produced by *Pyrenophora* species are considered to be potent mycotoxins. For example, when incorporated into animal diets, *D. campulata* was shown to be highly toxic to sheep, goats and inbred rats (Schneider et al., 1985; Collette et al., 1988).

Cytochalasin B is a cell-permeable mycotoxin, capable of inhibiting cell division by blocking the formation of contractile microfilaments and inhibiting cellular movement (Carter, 1967). Although mitosis continues normally, cytokinesis is inhibited, causing multi-nucleated or denucleated cells to arise. Estensen & Plagemann (1972) found other mycotoxic effects of cytochalasain B, observing its ability to inhibit the transport of glucose, deoxyglucose, and glucosamine by Novikoff hepatoma cells. Although transfer of these molecules was adversely affected, intracellular phosphorylation and metabolism continued as in normal cells unexposed to cytochalasin B. Similar effects are observed in plant cells exposed to cytochalasin B such as disruption of the actin cytoskeleton, inhibition of glucose inhibitors and bi-nucleate cells (Huang et al., 2000; Horemans et al., 2008; Capio et al., 2004). Evidente et al. (2002) suggested the use of these metabolites as a natural herbicide instead of using a living agent. Experiments conducted by Campbell et al. (2003) showed that extracted metabolites of *P. semeniperda* from solid media containing various cytochalasin molecules inhibit root growth of tomato and wheat seedlings.

It was long supposed that the tricyclic ring system of cytochalasans is generated by a Diels-Alder-type reaction (Oikawa et al., 1992; Oikawa et al., 2004), however the genes and enzymes involved in the synthesis of these molecules been unknown. Investigating the molecular basis of cytochalasan biosynthesis, Schümann and Hertweck (2007) successfully cloned and sequenced the first

gene cluster thought to be involved in the early stages of cytochalasan biosynthesis, identified as a group of seven genes of what are now called the chaetoglobosin (*Che*) gene cluster in *Penicillium expansum*. RNA silencing methods were used to show that *Che*A (PKS-NRPS hybrid protein) plays an essential role in cytochalasan biosynthesis and provided the first and only insight into any molecular explanation of cytochalasan biosynthesis.

Bromus tectroum **Biocontrol**

Cheatgrass is an exotic winter annual native to Southern Europe and Southwestern Asia. It is an invasive grass species of the (IMW) and provides the fuel for disastrous wildfires that occur on sagebrush rangelands (Young et al., 1987), significantly increasing the fire frequency in these areas with a return time decreased from 50 to 200 years to 3 to 5 years (Wisenant, 1990; Meyer et al., 2001). Continuously disturbed areas caused by cultivation, overgrazing, fires, construction etc. encourage invasion, creating ecological advantages for cheatgrass. It was suggested that cheatgrass was accidentally introduced to the IMW as a seed contaminant at the end of the 19th century during the time of large-scale grazing (Knapp, 1996). Sheep, cattle, and feral horse grazing also helped to establish this grass in these areas, and in turn increased the frequency of wild fires and displaced native plant and animal species (Knapp, 1996). Landscapes once home to native plants and animals are now replaced with monocultures of cheatgrass and costing the US government millions of dollars annually to control wildfires caused by this weed (Knapp, 1996).

Conventional methods of cheatgrass control include the use of various herbicides, mechanical removal, and early season burning, all done to prevent seed dispersal. These traditional methods are sometimes expensive, dangerous, hard to control, and cause damage to native plant species (Meyer et al., 2008). The possibility of using fungal pathogens as bio-control agents against cheatgrass was investigated (Meyer et al., 2008), including the use of *P. semeniperda* to attack the cheatgrass seed

bank. Results indicate a 100 percent mortality rate of dormant cheatgrass seeds inoculated with *P. semeniperda* conidia while non-dormant seeds showed a reduced mortality rate between 8–13%. All seeds inoculated with conidia show growth of black stromatal bodies, indicative of infection, however, only seeds that exhibit these symptoms before germination were actually killed by the pathogen (Beckstead et al., 2007; Meyer et al., 2008).

It is not uncommon for fungi to be used as biological controls of plant or animal populations. For example, the fungus *Entomorphaga* was used to kill gypsy moths that were destroying trees in the eastern United States and is being investiged for use in controlling other insect populations, thereby replacing chemical insecticides. Another fungus, *Candida oleophila*, is grown on the surface of harvested fruits and vegetables before other spoilage fungi have a chance to infect, preventing high yield losses. In addition to beneficial effects, fungi have also had detrimental consequences on entire populations, for example, the introduced ascomycete fungus, *Cryphonectria parasitica*, virtually annihilated all but small isolated populations of the chestnut tree in the United States.

Although potential exists for *P. semeniperda* to be used against cheatgrass, concerns have been raised in regard to problems that might arise if *P. semeniperda* were to be used as a biocontrol agent in the field. One worry is that *P. semeniperda* is only effective at eliminating the carryover seed bank. Unless other forms of control are used to attack the transient (current-year) seed bank then *P. semeniperda* may only have a low impact as a biocontrol agent against cheatgrass (Meyer et al., 2008). As mentioned previously, many races or isolates of this fungus exist in nature, each of which produces varying levels of toxins, suggesting differences in virulence and abilities to successfully infect host seeds, therefore, artificial selection of fungal isolates for increased virulence may provide an advantage in eradicating the transient seed bank (Krupinsky, 1987; Ciuffetti & Tuori, 1999; Campbell et al., 2003; Capio et al., 2004).

Phytopathogenic Fungal Evolution and Genetics

Although the fungal kingdom has proven to be a producer of many plant pathogens, most fungi are not pathogenic towards plants. In fact, phylogenetic studies show that phytopathogenic fungi appear disproportionately distributed, suggesting that these fungi have arisen multiple times throughout evolutionary history (Berbee, 2001). Many if not most of the plant pathogenic fungi tend to be host specific, *P. semeniperda* being an exception. Host specificity can be attributed to species-specific genetic factors, host, environment, and quantitative traits. Identification of these factors and an understanding of evolutionary mechanisms surrounding these fungi become imperative in understanding host specificity and infection strategies.

Host specificity and fungal evolution is affected by many factors which allow them to exploit their natural environments. High mutation rates are observed across species of fungal phytopathogens which may be a result of an ―arms race" between pathogen and host (Maor & Shirasu, 2005). As disease pressure increases on a host species, changes in host proteins will occur, favoring neutralization of pathogen virulence/disease factors. At the same time, the pathogen will be pressured to update these factors in order to adapt to a changing host targets, resulting in accelerated evolution of pathogen virulence factors and host recognition elements. Does and Rep (2007) suggest two consequences of this type of arms race: 1) loss of potential recognition targets (virulence gene loss), such as compounds with an obvious contribution to fitness, and 2) high mutation rates of secreted proteins involved in virulence, as a way to keep engaging their targets while escaping resistance genes in the plant host population.

These consequences may offer a reason for the staggered phylogenetic distribution of pathogenicity among fungi. However, gene loss and high mutation rates may be countered through genetic strategies not often observed in other eukaryotic organisms, an exceptional example being the gain of genetic material through horizontal gene transfer (HGT), the incorporation of genetic

information from one organism to another without being its descendant or offspring. HGT may be a relevant issue for the evolution of fungal pathogenicity because the acquirement of a new virulence gene or gene cluster could turn the recipient into a novel pathogen or allow it to infect multiple hosts. Interestingly, a case of HGT between *P. tritici-repentis* and *S. nodorum* in which an 11 kb fragment of DNA containing *ToxA* genes were shared between both species illustrates HGT and its correspondence to the emergence of a new disease (van der Does & Rep, 2007). In addition to HGT, chromosomal rearrangements, addition of conditionally dispensable chromosomes containing virulence factors, and virulence gene clusters contribute to the emergence of novel fungal pathogens.

Filamentous fungi are an important group of organisms that have contributed to man in many ways and have been instrumental in solving many fundamental problems in biology. There is no arguing their relative importance to the health of the general population and their effects on the global agricultural economy. Continued study of this incredibly diverse group of organisms will further contribute to our understanding of their genetics, pathogenicity strategies, and effects on ecosystems.

LITERATURE CITED

Aragona, M., Montigiani, M. & Porta-Puglia, A. (2000). Electrophoretic karyotypes of the phytopathogenic Pyrenophora graminea and P. teres. *Mycological Research* **104**, 853-857.

Bakonyi, J. & Justesen, A. F. (2007). Genetic Relationship of Pyrenophora graminea, P. teres f. maculata and P. teres f. teres Assessed by RAPD Analysis. *Journal of Phytopathology* **155**, 76-83.

Ballance, G. M., Lamari, L. & Bernier, C. C. (1989). Purification and characterization of a host-selective necrosis toxin from Pyrenophora tritici-repentis. *Physiological and Molecular Plant Pathology* **35**, 203-213.

Barreto, D., Fotugno, C. (1994). *Drechslera verticillata*: ceral pathogen. ITNA, Argentina **25**, 71-79.

Beckstead, J., Meyer, S. E., Molder, C. J. & Smith, C. (2007). A Race for Survival: Can Bromus tectorum Seeds Escape Pyrenophora semeniperda-caused Mortality by Germinating Quickly? *Annals of Botany* **99**, 907-914.

Berbee, M.L., Taylor, J.W. 2001. Fungal Molecular Evolution: Gene Trees and Geologic Time. Pp. (229-246) In: The Mycota Vol.VIIB, Systematics and Evolution. D.J. McLaughlin, E. McLaughlin and P.A. Lemke, eds. Springer, Berlin.

Bogyo, D., Fondy, S. R. E., Finster, L., Fondy, C., Patil, S. & Fondy, T. P. (1991). Cytochalasin-B-induced immunosuppression of murine allogeneic anti-tumor response and the effect of recombinant human interleukin-w. *Cancer Immunology Immunotherapy* **32**, 400-405.

Brittlebank, C. & Adam, D. (1924). A new disease of the Gramineae: *Pleospheria semeniperda*. *Nov. sp. Trans. Br. Mycol. Soc.* **10**, 123-127.

Broad Institute. "http://www.broadinstitute.org/annotation/genome/pyrenophora_tritici_repentis/." Pyrenophora tritici-repentis Database. August 12, 2009. Broad Institute. August 13, 2009.

Campbell, M. A., Medd, R. W. & Brown, J. F. (2003). Phytotoxicity of metabolites produced by Pyrenophora semeniperda in liquid culture. *Australian Journal of Experimental Agriculture* **43**, 1237-1244.

Capio, E. R., Tate, M. E. & Wallwork, H. (2004). Phytotoxic metabolites from Drechslera wirreganensis and D-campanulata. *Australasian Plant Pathology* **33**, 23-28.

Carter, S. B. (1967). Effects of Cytochalasins on Mammalian Cells. *Nature* **213**, 261-264.
Ciuffetti, L. M. & Tuori, R. P. (1999). Advances in the Characterization of the Pyrenophora tritici-repentis-Wheat Interaction. *Phytopathology* **89**, 444-9.

Collett, M. G., Fincham, J. E., Tustin, R. C., Joubert, G. & Marasas, W. F. (1988). The pathology of chronic Drechslera campanulata toxicosis in inbred rats. *Onderstepoort J Vet Res* **55**, 51-7.

Drechsler, C. (1923). Some graminicolous species of *Helminthosporium*: 1. *Journal of Agricultural Research* **24**, 641-740.

Ellwood, S., Liu, Z., Syme, R., Lai, Z., Hane, J., Keiper, F., Moffat, C., Oliver, R. & Friesen, T. (2010). A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. *Genome Biology* **11**, R109.

Estensen, R. D. & Plagemann, P. G. W. (1972). Cytochalasin B: Inhibition of Glucose and Glucosamine Transport. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 1430-1434.

Evidente, A., Andolfi, A., Vurro, M., Zonno, M. C. & Motta, A. (2002). Cytochalasins Z1, Z2, and Z3, Three 24-Oxa[14]cytochalasans Produced by Pyrenophora semeniperda. *ChemInform* **33**, 232-232.

Faris, J., Anderson, J., Francl, L. & Jordahl, J. (1997). RFLP mapping of resistance to chlorosis induction by <i>Pyrenophora tritici-repentis</i> in wheat. *TAG Theoretical and Applied Genetics* **94**, 98-103.

Franel, L. J. (1997). Local and mesodistance dispersal of *Pyrenophora tritici-repentis* conidia. *Canadian Journal of Plant Pathology* **19**, 247-255.

Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641-79.

Haidle, A. M. & Myers, A. G. (2004). An enantioselective, modular, and general route to the cytochalasins: synthesis of L-696,474 and cytochalasin B. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12048-53.

Hargreaves, J. A. & Keon, J. P. R. (1983). The binding of isolated mesophyll cells from barley leaves to hyphae of <i>Pyrenophora teres</i&gt. *Plant Cell Reports* **2**, 240-243.

Hawksworth, D. L. (2001). The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research* **105**, 1422-1432.

Horemans, N., Szarka, A., De Bock, M., Raeymaekers, T., Potters, G., Levine, M., Banhégyi, G. & Guisez, Y. (2008). Dehydroascorbate and glucose are taken up into Arabidopsis thaliana cell cultures by two distinct mechanisms. *FEBS Letters* **582**, 2714-2718.

Hosford, R., Larez, C. & Hammond, J. (1987). Interaction of wet period and temperature on *Pyrenophora tritici-repentis* infection and development in wheats of differing resistance. *Phytopathology* **77**, 1021-1027.

Huang, Z., Han, Y. & Howell, S. H. (2000). Formation of Surface Tubules and Fluorescent Foci in Arabidopsis thaliana Protoplasts Expressing a Fusion between the Green Fluorescent Protein and the Cauliflower Mosaic Virus Movement Protein. *Virology* **271**, 58-64.

Knapp, P. A. (1996). Cheatgrass (Bromus tectorum L) dominance in the Great Basin Desert : History, persistence, and influences to human activities. *Global Environmental Change* **6**, 37-52.

Kreitlow, K. & Bleak, A. (1964). *Podosporiella verticillata*, a soil-borne pathogen of some western Gramineae. *Phytopathology* **54**, 353-357.

Krupinsky, J. (1987). Pathogenicity on wheat of *Pyrenophora tritici-repentis* isolated from *Bromus inermis*. *Phytopathology* **77**, 760-765.

Lamari, L. & Bernier, C. (1989). Toxin of *Pyrenophora tritici-repentis*: Host-Specificity, Significance in Disease, and Inheritance of Host Reaction. *Phytopathology* **79**, 740-744.

Lamari, L. & Bernier, C. (1991). Genetics of Tan Necrosis and Extensive Chlorosis in Tan Spot of Wheat Caused by *Pyrenophora tritici-repentis*. *Phytopathology* **81**, 1092-1095.

Léveillé, J. (1841). Description de quelques espèces nouvelles de Champignons. *Ann. Sci. Nat.* **16**, 235-242.

Maor, R. & Shirasu, K. (2005). The arms race continues: battle strategies between plants and fungal pathogens. *Current Opinion in Microbiology* **8**, 399-404.

Martin, R., Clough, K. & Johnston, H. (1984). Vertical and horizontal distribution of conidia of Pyrenophora teres in barley fields. *Canadian Journal of Plant Pathology* **6**, 280-282.

Mathre, D. E. (1997). Compendium of Barley Diseases. *American Phytopathological Society*.
Medd, R. W., Murray, G. M. & Pickering, D. I. (2003). Review of the epidemiology and economic importance of Pyrenophora semeniperda. *Australasian Plant Pathology* **32**, 539-550.

Medd RW, C. M. (2005). Grass seed infection following inundation with Pyrenophora semeniperda. *Biocontrol Science and Technology* **15**, 21-36.

Meyer, Susan E.; Nelson, David L.; Clement, Suzette; Beckstead, Julie 2008. Cheatgrass (*Bromus tectorum*) biocontrol using indigenous fungal pathogens. In: Kitchen, Stanley G.; Pendleton, Rosemary L.; Monaco, Thomas A.; Vernon, Jason, comps. 2008. Proceedings-Shrublands under fire: disturbance and recovery in a changing world; 2006 June 6-8; Cedar City, UT. Proc. RMRS-P-52. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 61-67.

Meyer, S. E., S. C. Garvin, and J. Beckstead. 2001. Factors mediating cheatgrass invasion of intact salt desert shrubland. In: D. E. McArthur and D. J. Fairbanks [coMPS]. Shrubland ecosystem genetics and biodiversity: Proceedings. RMRS- P-21. Provo, UT: USDA Forest Service, Rocky Mountain Research Station. p 224-232.

Meyer, S. E., Beckstead, J., Allen, P. S. & Smith, D. C. (2008). A seed bank pathogen causes seedborne disease: *Pyrenophora semeniperda* on undispersed grass seeds in western North America. *Canadian Journal of Plant Pathology* **30**, 525-533.

O'Gara, P. (1915). A *Podosporiella* disease of germinating wheat. *Phytopathology* **5**, 323-325.

Oikawa, H., Murakami, Y. & Ichihara, A. (1992). Biosynthetic study of chaetoglobosin A: origins of the oxygen and hydrogen atoms, and indirect evidence for biological Diels-Alder reaction. *Journal of the Chemical Society, Perkin Transactions 1*, 2955-2959.

Oikawa, H. & Tokiwano, T. (2004). Enzymatic Catalysis of the Diels—Alder Reaction in the Biosynthesis of Natural Products. *ChemInform* **35**, no-no.

Pfender, W. (1988). Suppression of Ascocarp Formation in Pyrenophora tritici-repentis by *Limonomyces roseipellis*, a Basidiomycete from Reduced-Tillage Wheat Straw. *Phytopathology* **78**, 1254-1258.

Pfender, W. & Wootke, S. (1988). Microbial Communities of *Pyrenophora*-Infested Wheat Straw as Examined by Multivariate Analysis. *Microbial Ecology* **15**, 95-113.

Piening, L. (1968). Development of barley net blotch from infested straw and seed. *Canadian Journal of Plant Pathology* **48**, 623-625.

Rau, D., Attene, G., Brown, A., Nanni, L., Maier, F., Balmas, V., Saba, E., Schäfer, W. & Papa, R. (2007). Phylogeny and evolution of mating-type genes from <i>Pyrenophora teres</i>, the causal agent of barley ~~"~~net blotch" disease. *Current Genetics* **51**, 377-392.

Rees, R. & Platz, G. (1983). Effect of yellow spot on wheat: Comparison of epidemics at different stages of crop development. *Australian Journal of Agricultural Research* **34**, 39-46.

Schilder, A. & Bergstrom, G. (1994). Infection of wheat seed by *Pyrenophora tritici-repentis*. *Canadian Journal of Botany* **72**, 510-519.

Schneider, D., Marasas, W., Collett, M. & van der Weshuizen, G. (1985). An experimental mycotoxicosis in sheep and goats caused by *Drrechslera campanulata*, a fungal pathogen of green oats. *Onderstepoort Journal of Veterinary Research* **52**, 93-100.

Schümann, J. & Hertweck, C. (2007). Molecular Basis of Cytochalasan Biosynthesis in Fungi: Gene Cluster Analysis and Evidence for the Involvement of a PKS-NRPS Hybrid Synthase by RNA Silencing. *Journal of the American Chemical Society* **129**, 9564-9565.

Shoemaker, R. (1959). Nomenclature of *Drechslera* and *Bipolaris*, grass parasites segregated from *'Helminthosporium'*. *Canadian Journal of Botany* **37**, 879-887.

Shoemaker, R. (1966). A pleomorphic parasite of cereal seeds, *Pyrenophora semeniperda*. *Canadian Journal of Botany* **44**, 1451-1456.

Smedegärd-Petersen, V. (1971). *Pyrenophora teres* f. *maculata* f. nov and *Pyrenophora teres*, the causal agent of barley in Denmark. *K. Vet. Landbohøjsk. Årsskr.* , 124-144.

Smedegärd-Petersen, V. (1983). Cross fertility and genetic relationship between *Pyrenophora teres* and *P. graminea*. *Seed Science Technology* **11**, 673-680.

Stewart, T. E., Allen, P. S. & Meyer, S. E. (2009). First report of *Pyrenophora semeniperda* in Turkey and Greece. *Plant Disease* **93**, 1351.

Steyn, P. S. (1980). In The Biosynthesis of Mycotoxins, a Study in Secondary Metabolism. *Academic Press: New York*, 260-299.

Tekauz, A. & Chiko, A. W. (1980). Leaf stripe of barley caused by *Pyrenophora graminea*: Occurrence in Canada and comparisons with barley stripe mosaic. *Candian Journal of Plant Pathology* **2**, 152-158.

Tekauz, A., Harper, F. R. & Davidson, J. G. N. (1985). Effect of date of seeding and seed treatment fungicides on infection of barley by *Pyrenophora graminea*. *Canadian Journal of Plant Pathology* **7**, 408-416.

Tomás, A. & Bockus, W. (1987). Cultivar-Specific toxicity of culture filtrates of *Pyrenophora tritici-repentis*. *Phytopathology* **77**, 1337-1340.

van der Does, H. C. & Rep, M. (2007). Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Molecular Plant-Microbe Interactions* **20**, 1175-82.

Wallace, H. A. H. (1959). A rare seed-borne disease of wheat caused by *Podosporiella verticillata*. *Canadian Journal of Botany* **37**, 509-515.

Whisenant SG 1990: Changing fire frequencies on Idaho‘s Snake River Plains: ecological and management implications. In: McArthur, E. Durant; Romney, EvanM.; Smith, StanleyD.; Tueller, PaulT., comps. Proceedings-symposium on cheatgrass invasion, shrub die-off, and other aspects of shrub biology and management, Las Vegas, NV. Gen. Tech. Rep. INT-276. Ogdent UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station: 4-10.

Young, J. A., Evans, R. A., Eckert, R. E. & Kay, B. L. (1987). Cheatgrass. *Rangelands* **9**, 266-270.

Zhang, G. & Berbee, M. L. (2001). Pyrenophora Phylogenetics Inferred from ITS and Glyceradehyde-3-Phosphate Dehydrogenase Gene Sequences. *Mycologia* **93**, 1048-1063.

Chapter 2: *De novo* Genome Assembly of *Pyrenophora semeniperda*

Introduction

The fungal genus *Pyrenophora* (anamorph *Drechslera*) is comprised of many graminicolous ascomycete species including *P. teres*, *P. graminea*, *P. tritici-repentis*, and *P. semeniperda*. A wide range of plant diseases are caused by this genus and can be found on many hosts, some of which are fairly common and lead to extensive yield loss in cereal species each year, most notably barley net blotch, barley stripe, and wheat tan spot. *Pyrenophora s*pecies are generalist, necrotrophic pathogens of grasses and although most commonly found in barley, *Pyrenophora* species also infect oats, wheat, and other non-crop grasses such as turf. The geographical ranges of *Pyrenophora* species are widespread and distributed throughout the world. One of characteristics of the epidemiology of this plant pathogen is leaf infection in which the fungus causes tissue necrosis in its hosts (Pfender & Wootke, 1988; Schilder & Bergstrom, 1994). The anamorph (asexual) state of *Pyrenophora* is most commonly observed although the teleomorphic (sexual) state has also been described (Zhang & Berbee, 2001; Medd et al, 2003). The asexual form produces haploid spores called conidia as its primary mode of reproduction.

One of the better-known *Pyrenophora* species and an economically important pathogen in wheat is *Pyrenophora tritici-repentis* (Lamari et al., 1991). Successful infection by this fungus causes tan spot and chlorosis in its hosts and is responsible for grain losses ranging from 3 to 50% in North America (Rees & Platz, 1983; Hosford et al., 1987). The *P. tritici-repentis* genome was recently sequenced due to its economic and scientific importance as a major plant pathogen (Broad Institute, 2009). It is the first species in the genus to be sequenced and serves as a model for the sequencing of other ascomycete fungi and fungal grass pathogens. *P. tritici-repentis* is haploid with a nuclear genome of eleven chromosomes and an estimated genomes size of 37 Mb.

P. semeniperda (anamorph *Drechslera campulata*) is a generalist phytopathogen of a wide range of grass generas, infecting host seeds and in some cases causing leaf spotting in infected plants. Black stromata protruding out of dead seeds are characteristic of infection by this fungus. The host range of *P. semeniperda* was first described by Wallace (1959) who reported successful infection of paleae, glumes, and lemmas of oats and wheat. Currently it is believed to infect over 36 genera of annual and perennial grasses (Yonow et al., 2004). Infection does not have any effect on seed maturation but does reduce germination and emergence (O'Gara, 1915; Brittlebank & Adam, 1924; Kreitlow & Bleak, 1964).

Recently, interest was expressed in using *P. semeniperda* as a biocontrol agent against cheatgrass, an extremely invasive weed in the Intermountain West (IMW) of the United States (Medd & Campbell, 2005). Cheatgrass is a threat to many ecosystems, invading habitats of sensitive native plants and animals. As a natural pathogen of cheatgrass, *P. semeniperda* is effective at killing dormant seeds after conidial inoculation (Meyer et al., 2008) and its use as a biocontrol agent against cheatgrass may offer a better alternative to other expensive and dangerous conventional methods of biocontrol such as the use of herbicides, or early season burning. Nevertheless, there are very few genetic and genomic resources available to facilitate studies of *P. semeniperda* biology.

The advent of next-generation sequencing technology has accelerated the field of genetics and genomics. In particular, pyrosequencing by 454 Life Sciences provides low-cost, high-throughput sequencing capabilities enabling a variety of applications including metagenomics and whole genome and transcriptome sequencing which further our understanding of genomic structure, gene expression, as well as population and evolutionary genetics. 454 pyrosequencing technology has dramatically increased the amount of genetic and genomic resources available for many fungal species including *Sordaria macrospora* (Nowrousian et al., 2010), *Grosmannia clavigera* (DiGuistini et al., 2009), and

Pichia pastoris (De Schutter et al., 2009). The application of this sequencing technology to *P. semeniperda* research will expand current genetic and genomic resources for this fungus and provide insights and clues into the unique biology of fungal pathogenesis.

Here, the *de novo* assembly and annotation of the 32.5 Mb *P. semeniperda* genome are presented. This is the third *Pyrenophora* genome to be sequenced and, to our knowledge, the first member of the family Pleosporaceae to be solely sequenced via 454 pyrosequencing using whole genome and paired-end sequencing techniques. The small genome size, haploid state, and modest level of repetitive elements within many fungal genomes make the job of *de novo* assembly relatively simple compared to other large and complex eukaryotic genomes (Galagan et al., 2005). The *P. semeniperda* sequencing project has four main objectives: 1) Obtain a high-quality draft of the *P. semeniperda* genome using next-generation sequencing technology, 2) annotate the *P. semeniperda* genome using *P. tritici-repentis* and other fungal sequencing data as references 3) identify genes involved in pathogenicity, and 4) identify genomic structural variations between *P. semeniperda* and *P. tritici-repentis*. These objectives will help to elucidate factors involved in virulence and other molecular mechanisms that may be used to exploit biological functions to control expansion of the cheatgrass invasion. Moreover, the work presented here should add to the general knowledge of fungal biology and contribute to the discovery of novel mechanisms of pathogenicity and infection by other fungi.

Materials & Methods

DNA and RNA Isolation. Fungal cultures and tissue were prepared as described by Boose et al. (2011). DNA was isolated from mycelium of the CCB06 isolate of *P. semeniperda*, collected from a site at Cinder Cone Butte, ID, using the ZR Fungal/Bacterial DNA MiniPrep™ from Zymo Research

(Zymo Research Corporation, Orange, CA) following the manufacturer's protocol. DNA was quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE).

RNA was isolated from two *P. semeniperda* isolates using the ZR Fungal/Bacterial RNA MiniPrep™ from Zymo Research (Zymo Research Corporation, Orange, CA) and stored at -80° C; RNA was collected from multiple tissue types including mycelium, fingers, and conidia from *P. semeniperda* isolates including CCB06 and the WRK0 isolate, collected from a site at Skull Valley, UT. RNA quality and integrity was assessed for each extraction using the RNA Nano 6000 kit from Agilent Technologies and the 2100 Bioanalyzer Expert software (Agilent Technologies, Santa Clara, CA). RNA quantity was measured with the TBS-380 Mini-Fluorometer (Turner Biosystems, Sunnyvale, CA) in combination with the RiboGreen RNA quantitation reagent (Molecular Probes, Eugene, OR).

cDNA Library Construction and Normalization. RNA samples meeting sufficient quality and quantity criteria were pooled together for 1st strand synthesis and cDNA optimization. cDNA was synthesized from pooled RNA using the SMARTer™ PCR cDNA Synthesis Kit (ClonTech, Mountain View, CA) followed by PCR cycling of cDNA with the Advantage HF 2 PCR kit (ClonTech, Mountain View, CA). Modified oligo primers were used to allow for *MmeI* digestion for 5' and 3' adaptor excision: 5' Smart Oligo [5'-AAG CAG TGG TAA CAA CGC ATC CGA CGC rGrGrG-3']; 3' Oligo dT SmartIIA [5'-AAG CAG TGG TAA CAA CGC ATC CGA CTT TTT TTT TTT TTT TTT TTV N-3']; New SmartIIA [5'-AAG CAG TGG TAA CAA CGC ATC CGA C-3'] (Sandra Clifton, Personal Communication, Washington University).

Normalization of cDNA was accomplished with the Axxora Trimmer cDNA Normalization kit (AxxOra, San Diego, CA). *MmeI* (New England BioLabs, Ipswitch, MA) was used for 5' and 3' modified SMART adaptor excision followed by removal of excised 5' and 3' adaptor ends using

AMPure beads (Agencourt Bioscience Corporation, Beverly, MA) using manufactuer's recommended protocols.

454 Sequencing. In total, a half plate of a whole genome library, a full plate of a 3-kb paired end library, and a half plate of normalized cDNA were sequenced using the 454 Life Sciences DNA Sequencer FLX using GS FLX Titanium series reagents (454 Life Sciences, Bradford, CT). Titanium emPCR, library preparation, and sequencing were completed at the Brigham Young University DNA Sequencing Center (Provo, UT).

454 Reads Assembly and Genome Annotation. *De novo* genome assembly was accomplished using Newbler (454 Life Sciences, Bradford, CT). The genome assembly included whole genome and paired-end reads. Default settings were chosen for the assembly in Newbler. cDNA reads were assembled separately from genomic reads and default settings were chosen for transcript assembly.

The genome annotation pipeline MAKER was used to predict gene models within the *de novo* assembly of *P. semeniperda*. Expressed sequence tags (ESTs), derived from the cDNA library, were used to provide evidence for predicted genes within the *P. semeniperda* genome for the annotation pipeline. An in-house Perl script was created to expedite the naming process, as an automated naming scheme did not exist within the MAKER pipeline. Assembled scaffolds of the *P. semeniperda* genome along with naming from the automated script are viewable through GBrowse, a widely used online genome browser made publicly available by the Genetic Model Organization Database project (GMOD).

P. semeniperda gene models were imported into the Blast2GO suite (Conesa et al., 2005; Götz et al., 2008) for functional annotation analysis. GO annotations were made in accordance with the

recommended protocol in the Blast2Go tutorial. Default settings were chosen along with an e-value threshold set at ≤ e-06 for each step of the GO annotation process.

RepeatMasker. Repeated and low complexity sequences within the *P. semeniperda* genome were identified using RepeatMasker (v3.2.7) (Smit et al., 2010) with a fungal repeat library. A slow search was performed for increased accuracy, increasing the sensitivity of the search between 0-5%.

Genome Assembly Validation. To assess the accuracy of the genome assembly, PCR primers were designed to amplify low-complexity regions throughout the genome including regions of densely clustered repetitive elements of the genome. Primers were designed to flank N domains representing nucleotides between contigs that were oriented together by the paired-end sequence assembly, thereby allowing us to evaluate the authenticity of the assembly; primers were also designed in areas within scaffolds flanking the N domains thereby allowing us to determine the accuracy of the assembly of the 454 sequencing reads. Genomic regions were PCR amplified with the Advantage ® Genomic LA Polymerase Mix (Clontech, Mountain View, CA). PCR fragments were directly sequenced using Sanger sequencing technology (BYU Sequencing Center, Provo, UT). Many PCR fragments were too large to obtain sequence from a single run, therefore, the left and right ends of these fragments were sequenced independently and the distance between the ends was determined from the size of the PCR products and their position on the genome assembly. Sanger sequences were mapped to the *P. semeniperda* genome assembly using the gsMapper software (454 Life Sciences, Bradford, CT). Sanger sequences were then aligned to the genome assembly using Mauve (Darling et al., 2004) and visualized within Geneious (Drummond et al., 2010).

SyMap. SyMap v3.3 (Synteny Mapping and Analysis Program) (Soderlund et al., 2006) was used to detect and display syntenic relationships between *P. semeniperda* and *P. tritici-repentis*. SyMap, by default uses NUCmer (Delcher et al., 2002) for multiple genome alignments via a modified Smith-

Waterman algorithm (Smith & Waterman, 1981). GFFs (General Feature Files), files containing gene descriptive information and other features associated with the genome, of *P. semeniperda* were imported into SyMap after the alignment was completed.

Results & Discussion

Sequencing and assembly of the P. semeniperda genome. For sequencing, the CCB06 isolate of *P. semeniperda* from the Cinder Cone Butte, ID, USA population was selected. *Pyrenophora semeniperda* is haploid with an unknown chromosome number, although, electrophoretic and cytological karyotyping in related species reveals 9 chromosomes in *P. teres* (Ellwood et al., 2010) and anywhere between 8-11 chromosomes in *P. tritici-repentis* depending on the isolate (Aboukhaddour et al., 2009). A shotgun sequencing strategy was used with the 454 Life Sciences DNA Sequencer FLX to sequence a whole-genome sequencing library and a paired-end sequencing library with an insert size of 3 kb. The whole-genome library produced approximately 257 Mb averaging 371 bp in read length on a half-plate run. The full-plate 3 kb paired-end library produced over 469.9 Mb of sequence with an average read length of 362.06 bp. In total, 726.9 Mb of sequence was produced from 2,759,755 reads with an average read length of 366.7 bp between both of the sequencing libraries (genomic and paired); 28.11% of the total number of reads are paired-end (775,958 reads) (Table 1).

The 454 reads were assembled using the Newbler software package developed by 454 Life Sciences for *de novo* DNA sequence assembly, specifically for 454's medium-length reads. First, the 257-Mb from the genomic library was assembled into 7,890 contigs; 98.38% of the total bases were successfully aligned (252 Mb). A 6-fold coverage of the genome was achieved a genome size and N50 contig length of approximately 31 Mb and 6,499 bp, respectively.

Next, the 469.9 Mb from the 3-kb insert library was combined to the whole genome assembly, decreasing the number of contigs to 1,001; 98.62% (672 Mb) of the total 681 Mb and 95.00% (2,621,753 reads) of the raw reads were used in the assembly. The incorporation of the paired reads dramatically improved the assembly, arranging the 1,001 contigs into 54 scaffolds and closing most internal gaps between the contigs. The average distance between the paired reads is 2.6 kb with a standard deviation of 665.5 bp.

The *de novo* assembly of *P. semeniperda* resulted in a 17-fold coverage of the genome with an N50 scaffold length of 1.3 Mb. The *P. semeniperda* genome assembly currently contains 32.5 Mb, although the Newbler software estimates the genome to be nearly 40.1 Mb. The combination of reads from the 454 genomic and paired-end sequence libraries resulted in an increase of the N50 contig size from 6,499 bp to 104,587 bp and, as expected, the inclusion of paired reads drastically reduced the number of gaps as compared to the assembly of the raw read data from the genomic sequencing library alone (Table 2). The sequenced genome size of 32.5 Mb of *P. semeniperda* is similar to the reported size of 37.8 Mb for *P. tritici-repentis* (Broad Institute, 2009) and 41.9 Mb for *P. teres* f. *teres* (Ellwood et al., 2010).

Genome Assembly Validation. To assess the accuracy of the genome assembly, primers were designed to amplify areas within 8 different scaffolds of the assembly containing low-complexity and repetitive sequences. Amplified regions ranged in size from 600–5,600 bp (total 47.2 kb; average length 1.77 kb); approximately 18.7 kb of the PCR fragments represented amplified regions in which N domains were present within the genome assembly, thereby allowing us to evaluate the placement of contigs within the scaffolds in addition to determining the validity of the contigs themselves. Obvious misassemblies were not detected as Sanger sequences mapped to expected regions of the assembly. An example of the

validation assembly can be seen in Figure 1. GC content of the validation sites are expectedly low (average < 50%) as low genic regions were chosen for the validation.

Gene identification and annotation. The MAKER annotation pipeline (Cantarel et al., 2007) was used to annotate the *P. semeniperda* genome and to create a publicly accessible genome database. The pipeline was used to make gene predictions, align ESTs to the genome, and integrate these into protein-coding gene annotations. To provide evidence of gene identity and ease of detection, a normalized *P. semeniperda* cDNA library was prepared for 454 sequencing, producing over 110.8 Mb of raw sequence data, averaging approximately 331 bp in read length. The Newbler assembly of the cDNA sequence library generated 7,963 isogroups and over 7 Mb of total sequence length. In addition to EST evidence from *P. semeniperda*, 12,171 *P. tritici-repentis* gene models (Broad Institute, 2009) were used as a reference to reinforce confidence in *ab initio* gene predictions.

Each gene model was used as a query in a BLAST search against all protein sequences in the GenBank database, using an e-value cutoff threshold of ≤ e-20. All putative genes were given systematic names based on their location within the assembly including scaffold and start and end nucleotide positions. In addition, species name and gene function from the BLAST search were also included in the naming scheme of *P. semeniperda* genes if they received a positive hit from the search.

The MAKER pipeline predicted a total of 11,453 genes, of which 9,064 were given a functional name and of these 8,070 matched genes from *P. tritici-repentis*. There were 2,269 gene models that did not receive a positive hit to any of the *P. semeniperda* ESTs or *P. tritici-repentis* gene models above the BLAST cutoff value. These gene models were predicted using the *ab initio* gene prediction software SNAP, used as an external program in the MAKER pipeline. Because these genes were predicted *ab initio* and did not receive any positive hits from the ESTs or gene models, little evidence was available to support their presence within the *P. semeniperda* genome. However, upon further

investigation, BLASTP searches (e-value ≤ e-20) within all the known reference protein databases revealed that 1,942 of the 2,269 *ab* initio-predicted genes (85.5%) share homology with sequences with known proteins from public databases. There were 327 gene models from *P. semeniperda* that lack significant matches to the known proteins from any of the public databases as a result of the BLASTP search, representing 2.86% of all *P. semeniperda* gene models. This number is in accordance with expected values of less than 5% or 500 orphan genes (genes without an apparent homolog in any of the known sequence databases) per ascomycete genome (Nowrousian et al., 2010). The average coding sequence (CDS) length is 1,312 bp, the smallest and largest CDS having nucleotide lengths of 72 bp and 25,382 bp respectively. The genome viewer, GBrowse, is used to visualize the *P. semeniperda* genome annotation and predicted genes on an interactive web page, publicly available at http://psoda2.cs.byu.edu/cgi-bin/gbrowse/bfod00001/.

Most of the annotation success was due in large part to the similarity between the genomes of *P. tritici-repentis* and *P. semeniperda*. Furthermore, a positive hit to any of the *P. tritici-repentis* gene models during the annotation process was automatically named, even if the protein name was labeled "Hypothetical Protein". Thus, the number of predicted genes only represents those that were simply assigned a name and does not necessarily express the number of genes that actually have any attached biological or functional information.

General genomic features. The overall GC content of the *P. semeniperda* genome is 49.98% (32.029 Mb excluding N-runs) and slightly increasing to 52.53% in CDSs, representing 46.47% of the genome. Additionally, t-RNAscan-SE (Lowe & Eddy, 1997) detected 91 tRNA genes, located on 22 scaffolds (Table 3). These putative tRNA genes do not tend to group together as seen in other fungi such as *Saccharomyces pombe* (Wanchanthuek et al., 2005) but are relatively spread throughout the genome.

Over 3,979 orthologous groups were identified between *P. semeniperda* and *P. tritici-repentis* with the Inparanoid v4.0 program (Remm et al., 2001), describing genes derived from a common ancestor of these fungal species and therefore are likely to share similar molecular functions (Berglund et al., 2008). Also, 4,184 genes in *P. semeniperda* have been identified as in-paralogs (a result of gene duplication after a speciation event).

Gene ontology. Putative *P. semeniperda* genes (11,264 genes) were run through the Blast2GO (Conesa et al., 2005; Götz et al., 2008) suite to clarify and identify gene function within the genome. 6,180 (54.8%) of the query set of genes were successfully annotated totaling 24,357 GO terms. Annotated genes were placed into three categories: biological process (BP), molecular function (MF), and cellular components (CC). The most common GO terms, in terms of the number of query sequences, in the BP category (9,555 total terms), were metabolic process (36%), cellular process (31.3%), and localization (9.5%) (Figure 2). Abundant GO terms in the MF category include catalytic activity (44.4%), binding (40.8%), and transporter activity (5.7%) (Figure 3). Finally, the bulk of the cellular component GO terms are categorized in the cell (58.9%) followed by the organelle (27.3%) and macromolecular complex (10.7%) (Figure 4). These terms only describe putative genes within the *P. semeniperda* genome and in no way informs us about gene expression.

Repeat sequences and transposons. Transposable elements and repeated sequences are some of the most abundant sequences in eukaryotic genomes; for example, over 44% of the human genome (Mills et al., 2007) and more than 75% of the maize genome (Wolfgruber et al., 2009) are comprised of transposable and other repetitive elements. Fungal genomes, however, contain relatively small amounts of these elements when compared to other eukaryotes, rarely exceeding 5% of the genome (Wöstemeyer & Kreibic, 2002). Such low levels of transposable and repetitive elements may be due to defense mechanisms known as repeat-induced point mutations (RIP) (Hood et al., 2005) that protect

fungal genomes against highly repeated sequences. The *P. semeniperda* genome was analyzed for repetitive sequences and retro-elements with RepeatMasker 3.2.7 (Smit et al., 2010), which screens DNA sequences for interspersed repeats and low complexity elements. Interspersed repeats (retroelements and DNA transposons) were the most abundant elements identified by RepeatMasker, totaling 610.7 kb or 1.89% of the genome. Many of these components were identified as class I retroelements (447 elements), mainly Gypsy/DIRS1 LTRs (388 elements). 296 class II DNA transposable elements were also found, 293 of these being Tc1-IS630-Pogo DNA transposons. In total, 859,266 bp or 2.66% of the genome was identified as containing interspersed repeats or low complexity elements (Table 4) and is quantitatively consistent with repeat elements observed in other ascomycete fungi (Wöstemeyer & Kreibic, 2002).

Genome Rearrangements. Questions have been raised concerning the impact transposable and repetitive elements have on the genomic architecture and evolution of fungi (Wöstemeyer & Kreibich, 2002). The insertion of repetitive DNA into a genome can impact the regulation of neighboring genes and may provide sites for homologous and ectopic recombination (Thronburg et al., 2006; Maksakova & Mager, 2005; Mieczkowski et al., 2006). These recombination sites may play an important role in observed local or wide-scale chromosomal rearrangements in fungi as well as in other organisms (Kupiec & Petes', 1988; Lim & Simmons, 1994; Ladevèze et al., 1998; Lahn & Page, 1999; Daboussi, 2003). These observations may help improve the understanding of the organization of the *P. semeniperda* genome. To investigate the role transposable and repeat elements may have played in the genomic architecture of *P. semeniperda*, its genome assembly was aligned with that of *P. tritici-repentis*. The *P. tritici-repentis* genome was ideal for a whole genome alignment as many of its sequences are arranged into full chromosomes, thereby allowing easier identification of genomic rearrangements.

SyMap (Synteny Mapping and Analysis Program) (Soderlund et al., 2006) was used to align and visualize genomic synteny between *P. semeniperda* and *P. tritici-repentis*. All scaffolds above 0.5 Mb (19 scaffolds; Figure 5), representing 86.4% of the sequenced genome of *P. semeniperda*, were aligned all the 11 *P. tritici-repentis* chromosomes. The alignment produced 88% and 80% of syntenic coverage in *P. semeniperda* and *P. tritici-repentis* respectively, and 6,376 gene hits on *P. semeniperda* scaffolds. Although over 8,070 genes within the *P. semeniperda* genome were identified from *P. tritici-repentis*, only 19 scaffolds were included in the genome alignment. A circular view of the genome alignment is presented in Figure 6.

The dot-plot of the genome alignment reveals regions of synteny and colinearity (Figure 7). A total of 101 syntenic blocks with an identity range of 95% have been determined and are highlighted in blue. Inversions, deletions, and transpositions are clearly illustrated. Many of the observed rearrangements within the *P. semeniperda* scaffolds are localized within the same corresponding *P. tritici-repentis* chromosome (intrachromosomal rearrangements) (Figure 8). Interestingly, these types of large-scale rearrangements have also been observed when comparing the genomes of *P. anserina* and *N. crassa*, most of which were intrachromosomal (Pain & Fowler, 2008; Galagan et al., 2003; Espagne et al., 2008). Intrachromosomal rearrangements remain constant throughout the alignment with the exception of scaffold 28 which shows patterns of interchromosomal rearrangements, transposing to four different *P. tritici-repentis* chromosomes (Figure 9).

Large-scale genomic rearrangements such as these in *P. semeniperda* have been extensively studied in *Saccharomyces cerevisiae*, many of which are attributed to recombination events between retrotransposons and other repetitive elements (Mieczkowski et al., 2006). Further investigation of the chromosomal rearrangements in the *P. semeniperda* genome reveals large, concentrated areas of transposable and repetitive elements flanking syntenic blocks. BLASTX searches of these areas in *P.*

semeniperda scaffolds reveal the presence of transposases, *Gag* genes, and transcriptases. The location of these elements in such highly concentrated areas of the genome and between chromosomal _breakpoints' is highly indicative of regions of ectopic recombination and/or mechanisms of genome evolution; Ty elements in *S. cerevisiae* have been shown to be sources of chromosomal crossovers which cause deletions, duplications, inversions, and translocations, though by what mechanisms and under what conditions this occurs under is unknown (Mieczkowski et al., 2006). More molecular evidence is needed to make conclusions concerning the roles retrotransposons have in the genomic architecture and evolution of *P. semeniperda*.

Pathogenicity and infection-related genes. There are a number of strategies that fungi use to infect their hosts. To help better understand these strategies in *P. semeniperda*, a pathogenicity gene database containing 74 cloned genes that are necessary for disease development was constructed. This particular database contains amino acid sequences of genes controlling such factors as infection structures, cuticle and cell wall degradation enzymes, fungal toxins, signal cascade components, and other _novel' and unclassified pathogenicity factors as described by Idnurm & Howlett (2001). A BLASTP search was conducted between this database and *P. semeniperda* amino acid sequences at an e-value threshold of ≤ e-20. A total of 224 genes from *P. semeniperda* were identified as pathogenicity factor genes sharing homology with those in the pathogenicty database (Table 5). Over 23 of these genes represent proteins that respond to the host environment (Table 6); many of which provide resistance to plant chemical defenses (phytoalexins), preventing degradation, altered physiology, or other uncharacterized mechanisms (Idnurm & Howlett, 2001). Eight of these genes share homology with the pisatin demethylase gene in *Nectria haematococca*, an enzyme that degrades the phytoalexin, pisatin. Loss of the pisatin gene in *N. haematoccoca* results in a significant reduction in pathogenicity (Wasmann & VanEtten, 1995). Interestingly, 13 genes categorized as fungal toxins share homology with AM-toxin synthetase from *Alternaria alternata*, a generalist filamentous plant pathogen that causes leaf spots, rot

and blight in its hosts (Taba et al., 2009; Bashan et al., 1991). Disruption of the AM-toxin synthetase gene prevented *A. alternata* from causing disease symptoms in host plants (Johnson et al., 2000). In addition, 3 genes were identified that share homology with AK-toxin genes (Akt1, AktR-2, and Akt3-2) from *A. alternata*, which are host-specific toxins produced by *A. alternata* that cause severe necrosis on the fruit and leaves of susceptible hosts (Uemura et al., 2002).

To obtain a broader perspective of infection-related genes in *P. semeniperda*, a database of amino acid sequences encoding proteins that have been demonstrated to be toxic to plants was created, following the suggestion of Brett et al., (2006). These sequences code for such gene products as hydrolases, protease inhibitors, protein toxins, secondary metabolite biosynthesis, ABC transporters, and effector genes, all factors related to virulence in necrotrophic plant pathogens. In total, 919 genes from *P. semeniperda* share homology with those in the database (Table 7). Most of these genes (72%) were identified as having some sort of hydrolase activity, targeting various polysaccharides. For example, pectinases, cutinases, and chitinases were among some of the putative gene products identified in *P. semeniperda*, characteristic of necrotrophic phytopathogens (Table 8); these enzymes are involved in the breakdown of polysaccharide substrates found in the cell walls of plants.

Secreted proteins. The expansion of secreted protein gene families has been observed in the comparison of the phytopathogen ascomycete *S. nodorum* and *M. grisea* with the saprophyte *Neurospora crassa* (Dean et al., 2005; Hane et al., 2007; Ellwood et al., 2010), congruous with their roles as plant pathogenic fungi. *P. semeniperda* contains a substantial amount of putative secreted gene products (996), as predicted by WolfP-SORT (Horton et al., 2006), ranging in length from 180 - 5,845 bp. A significant portion of the *P. semeniperda* secretome (81%) is homologous to genes in *P. tritici-repentis*, as might be expected given their close phylogenetic relationship. This conforms to data of the

P. teres f. *teres* secretome, showing 85% of the predicted genes to share homology with genes from *P. tritici-repentis*.

Approximately 54% (546 sequences) of the secretion genes were annotated with GO terms using Blast2GO (Conesa et al., 2005; Gotz et al., 2008). Although there are some drawbacks and limitations with the existing annotations databases due to their incompleteness (Khatri & Draghici 2005), these GO terms provide a short synopsis of the types of secretion proteins that might be expected in *P. semeniperda*. Consistent with its role as a necrotrophic plant pathogen, many of these secretion genes are predicted to code for proteins and enzymes that target various polysaccharides (Table 9). As observed in our previous assessment of pathogenic-related sequences, putative secreted gene products containing hydrolase activity are homologous to other sequences containing cellulose binding domains, carboxypeptidase, as well as cell wall glucanase and glycosyl hydrolase activity. Many of these sequences were also annotated with GO terms for oxidation reduction and oxidoreductase activity, suggesting that these gene products have key roles in the process of cellulose and lignin degradation (Raíces et al., 2002). Furthermore, there is some intermittent clustering between these secretion genes, grouping together in pairs of 2 -3 spread throughout the genome.

Chaetoglobosin genes. Cytochalasans are comprised of a diverse group of fungal metabolites, well-known for their ability to bind to actin filaments and block polymerization and elongation, thus preventing cytokinesis while mitosis remains unaffected. Due to their ability to prevent the normal function of the cytoskeleton, many members of this molecular family have been identified as antibiotic, antiviral, anti-inflammatory, and as antitumoral agents (Schümann & Hertweck, 2007). Various forms of cytochalasan molecules have been described in many phytopathogenic fungi, including three previously unknown cytochalasans (Z1, Z2, and Z3) identified in *P. semeniperda*. The exact role of cytochalasans in fungal virulence pathways currently eludes researchers; although, Beckstead et al.

(2007) suggest that *P. semeniperda* may use these phytotoxins in the process of infection to effectively attack dormant or slowly germinating cheatgrass seeds which are most vulnerable to attack from this fungus.

It is understood that the tricyclic ring system of cytochalasans is generated by a Diels-Alder-type reaction (Oikawa et al., 1992; Oikawa et al., 2004). Recently, the genes encoding the enzymes responsible for the early stages of cytochalasan biosynthesis in *Penicillium expansum* were cloned by Schümann & Hertweck (2007). They identified 7 genes grouped together in what is now called the *chaetoglobosin* (*Che*) gene cluster. RNA silencing methods identified *Che*A (PKS-NRPS hybrid protein) as essential to cytochalasan biosynthesis (Schümann & Hertweck, 2007).

Homologs of the genes in the *Che* cluster from *P. expansum* were found in *P. semeniperda*. A BLASTP search was conducted between *Che* amino acid sequences from *P. expansum* and *P. semeniperda*. In total, 21 genes from *P. semeniperda* share homology with the *P. expansum Che* gene cluster having p-values of ≤ e-37. Protein domains of *Che* genes were compared between *P. expansum* and *P. semeniperda* to further validate the BLASTP results using a Pfam domain search. Putative *P. semeniperda Che* genes were not found in clusters as they were in *P. expansum* but are interspersed throughout the genome, covering 10 scaffolds of the current *P. semeniperda* assembly. We found 5 genes in *P. semeniperda* sharing homology with *Che*A, the gene shown to be necessary for cytochalasan formation in *P. expansum*. PKS-NRPS hybrid domains were also identified in these 5 genes as found in the *Che*A from *P. expansum* (Schümann & Hertweck, 2007). The protein domains of the *P. expansum Che* genes were also found among those of *P. semeniperda* including monooxygenase, transcription factors, and enoyl reductase domains (Table 10).

Conclusion and future work. The genome sequence, assembly and annotation of the CCB06 isolate of *P. semeniperda* are reported here. The assembly produced over 32 Mb in sequence length and an

estimated genome size of approximately 37 Mb, in range of other related fungi. Genome comparisons between *P. semeniperda* and *P. tritici-repentis* allow visualization of large-scale rearrangements between these related species and provide clues to evolutionary mechanisms used by this fungus. The *P. semeniperda* genome contains a rich diversity of putative genes, common to other plant pathogens, notably hydrolases, ABC transporters, cytochrome P450 and secreted gene products attributable to other necrotrophs. In addition, the genome sequence can provide information for the development of molecular markers which may be implemented in population or evolutionary studies of this organism. This assembly also provides researchers with genomic and genetic resources to advance *P. semeniperda* research and the means to further our understanding of other phytopathogenic fungi.

Future genomics work includes the sequencing and annotation of additional *P. semeniperda* isolates to provide deeper insight into genome structure and evolution in this species. Genome comparisons of *P. semeniperda* isolates using next-generation sequencing technology will be the first attempt, to our knowledge, that a filamentous fungi belonging to the same species will be analyzed on this scale.

LITERATURE CITED

Aboukhaddour, R., Cloutier, S., Ballance, G. M. & Lamari, L. (2009). Genome characterization of Pyrenophora tritici-repentis isolates reveals high plasticity and independent chromosomal location of ToxA and ToxB. *Molecular Plant Pathology* **10**, 201-212.

Adachi, K. & Hamer, J. E. (1998). Divergent cAMP Signaling Pathways Regulate Growth and Pathogenesis in the Rice Blast Fungus Magnaporthe grisea. *The Plant Cell Online* **10**, 1361-1374.

Ahn, J.-H. & Walton, J. D. (1997). A Fatty Acid Synthase Gene in Cochliobolus carbonum Required for Production of HC-Toxin, Cyclo(d-Prolyl-l-Alanyl- d-Alanyl- l-2-Amino-9,10-Epoxi-8-Oxodecanoyl). *Molecular Plant-Microbe Interactions* **10**, 207-214.

Bailey, A., Mueller, E. & Bowyer, P. (2000). Ornithine Decarboxylase of Stagonospora (Septoria) nodorum Is Required for Virulence toward Wheat. *Journal of Biological Chemistry* **275**, 14242-14247.

Barrett, K., Gold, S. & Kronstad, J. W. (1993). Identification and complementation of a mutation to constitutive filamentous growth in *Ustilago maydis*. *Molecular Plant-Microbe Interacteraction* **6**, 274-283.

Bashan, Y., Levanony, H. & Or, R. (1991). Wind Dispersal of Alternaria alternata, a Cause of Leaf Blight of Cotton. *Journal of Phytopathology* **133**, 225-238.

Beckstead, J., Meyer, S. E., Molder, C. J. & Smith, C. (2007). A Race for Survival: Can Bromus tectorum Seeds Escape Pyrenophora semeniperda-caused Mortality by Germinating Quickly? *Annals of Botany* **99**, 907-914.

Berglund, A.-C., Sjölund, E., Östlund, G. & Sonnhammer, E. L. L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Research* **36**, D263-D266.

Brittlebank, C. & Adam, D. (1924). A new disease of the Gramineae: *Pleospheria semeniperda*. *Nov. sp. Trans. Br. Mycol. Soc.* **10**, 123-127.

Broad Institute. "http://www.broadinstitute.org/annotation/genome/pyrenophora_tritici_repentis/." Pyrenophora tritici-repentis Database. August 12, 2009. Broad Institute. August 13, 2009.

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A. & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188-196.

Cheng, Y.-Q., Ahn, J.-H. & Walton, J. D. (1999). A putative branched-chain-amino-acid transaminase gene required for HC-toxin biosynthesis and pathogenicity in Cochliobolus carbonum. *Microbiology* **145**, 3539-3546.

Choi, W. & Dean, R. A. (1997). The Adenylate Cyclase Gene MAC1 of Magnaporthe grisea Controls Appressorium Formation and Other Aspects of Growth and Development. *The Plant Cell* **9**, 1973-1983.

Chumley, F. G. & Valent, B. (1990). Genetic analysis of melanin-deficient, nonpathogenic mutants of *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions* **3**, 135-143.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. & Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.

Daboussi, M.-J. & Capy, P. (2003). TRANSPOSABLE ELEMENTS IN FILAMENTOUS FUNGI. *Annual Review of Microbiology* **57**, 275-299.

Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* **14**, 1394-1403.

De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., Van de Peer, Y. & Callewaert, N. (2009). Genome sequence of the recombinant protein production host Pichia pastoris. *Nat Biotech* **27**, 561-566.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N. D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., Soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J. E. & Birren, B. W. (2005). The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature* **434**, 980-986.

Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research* **30**, 2478-2483.

Dickman, M. B., Podila, G. K. & Kolattukudy, P. E. (1989). Insertion of cutinase gene into a wound pathogen enables it to infect intact host. *Nature* **342**, 446-448.

DiGuistini, S., Liao, N., Platt, D., Robertson, G., Seidel, M., Chan, S., Docking, T. R., Birol, I., Holt, R., Hirst, M., Mardis, E., Marra, M., Hamelin, R., Bohlmann, J., Breuil, C. & Jones, S. (2009). De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology* **10**, R94.

Drummond, A.J, Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A. (2010) Geneious v5.3, Available from http://www.geneious.com.

Dufresne, M., Bailey, J. A., Dron, M. & Langin, T. (1998). clk1, a serine/threonin protein kinase-encoding gene, is involved in pathogenicity of *Colletotrichum lindemuthianum* on common bean. *Molecular Plant-Microbe Interactions* **11**, 99-108.

Dufresne, M., Perfect, S., Pellier, A.-L., Bailey, J. A. & Langin, T. (2000). A GAL4-like Protein Is Involved in the Switch between Biotrophic and Necrotrophic Phases of the Infection Process of Colletotrichum lindemuthianum on Common Bean. *The Plant Cell Online* **12**, 1579-1590.

Dürrenberger, F. & Kronstad, J. (1999). The <i>ukc1</i> gene encodes a protein kinase involved in morphogenesis, pathogenicity and pigment formation in <i>Ustilago maydis</i&gt. *Molecular and General Genetics MGG* **261**, 281-289.

Ellwood, S., Liu, Z., Syme, R., Lai, Z., Hane, J., Keiper, F., Moffat, C., Oliver, R. & Friesen, T. (2010). A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. *Genome Biology* **11**, R109.

Enkerli, J., Bhatt, G. & Covert, S. F. (1998). Maackiain Detoxification Contributes to the Virulence of Nectria haematococca MP VI on Chickpea. *Molecular Plant-Microbe Interactions* **11**, 317-326.
Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B., Couloux, A., Aury, J.-M., Segurens, B., Poulain, J., Anthouard, V., Grossetete, S., Khalili, H., Coppin, E., Dequard-Chablat, M., Picard, M., Contamine, V., Arnaise, S., Bourdais, A., Berteaux-Lecellier, V., Gautheret, D., de Vries, R., Battaglia, E., Coutinho, P., Danchin, E., Henrissat, B., Khoury, R., Sainsard-Chanet, A., Boivin, A., Pinan-Lucarre, B., Sellem, C., Debuchy, R., Wincker, P., Weissenbach, J. & Silar, P. (2008). The genome sequence of the model ascomycete fungus Podospora anserina. *Genome Biology* **9**, R77.

Evidente, A., Andolfi, A., Vurro, M., Zonno, M. C. & Motta, A. (2002). Cytochalasins Z1, Z2, and Z3, Three 24-Oxa[14]cytochalasans Produced by Pyrenophora semeniperda. *ChemInform* **33**, 232-232.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P. C., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C. & Birren, B. (2003). The genome sequence of the filamentous fungus Neurospora crassa. *Nature* **422**, 859-868.

Galagan, J. E., Henn, M. R., Ma, L.-J., Cuomo, C. A. & Birren, B. (2005). Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Research* **15**, 1620-1631.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420-3435.

Hall, C., Brachat, S. & Dietrich, F. S. (2005). Contribution of Horizontal Gene Transfer to the Evolution of Saccharomyces cerevisiae. *Eukaryotic Cell* **4**, 1102-1115.

Hane, J. K., Lowe, R. G. T., Solomon, P. S., Tan, K.-C., Schoch, C. L., Spatafora, J. W., Crous, P. W., Kodira, C., Birren, B. W., Galagan, J. E., Torriani, S. F. F., McDonald, B. A. & Oliver, R. P. (2007).

Dothideomycete–Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat Pathogen Stagonospora nodorum. *The Plant Cell Online* **19**, 3347-3368.

Have, A. t., Mulder, W., Visser, J. & van Kan, J. A. L. (1998). The Endopolygalacturonase Gene Bcpg1 Is Required for Full Virulence of Botrytis cinerea. *Molecular Plant-Microbe Interactions* **11**, 1009-1016.

Hood, M. E., Katawczik, M. & Giraud, T. (2005). Repeat-Induced Point Mutation and the Population Structure of Transposable Elements in Microbotryum violaceum. *Genetics* **170**, 1081-1089.

Hosford, R., Larez, C. & Hammond, J. (1987). Interaction of wet period and temperature on *Pyrenophora tritici-repentis* infection and development in wheats of differing resistance. *Phytopathology* **77**, 1021-1027.

Hotopp, J. C. D., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Torres, M. C. M., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H. & Werren, J. H. (2007). Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. *Science* **317**, 1753-1756.

Idnurm, A. & Howlett, B. J. (2001). Pathogenicity genes of phytopathogenic fungi. *Molecular Plant Pathology* **2**, 241-255.

Isshiki, A., Akimitsu, K., Yamamoto, M. & Yamamoto, H. (2001). Endopolygalacturonase Is Essential for Citrus Black Rot Caused by Alternaria citri but Not Brown Spot Caused by Alternaria alternata. *Molecular Plant-Microbe Interactions* **14**, 749-757.

Johnson, K. B., Stockwell, V. O., Sawyer, T. L. & Sugar, D. (2000). Assessment of environmental factors influencing growth and spread of *Pantoea agglomerans* on and among blossoms of pear and apple. *Phytopathology* **90**, 1285-1294.

Johnson, R. D., Johnson, L., Itoh, Y., Kodama, M., Otani, H. & Kohmoto, K. (2000). Cloning and characterization of a cyclic peptide synthetase gene from *Alternaria alternata* apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity. *Molecular Plant-Microbe Interacteraction* **13**, 742-753.

Khatri, P. & Drăghici, S. Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* **21**, 3587-3595.

Kim, Y.-K., Kawano, T., Li, D. & Kolattukudy, P. E. (2000). A Mitogen-Activated Protein Kinase Kinase Required for Induction of Cytokinesis and Appressorium Formation by Host Signals in the Conidia of Colletotrichum gloeosporioides. *The Plant Cell Online* **12**, 1331-1344.

Kreitlow, K. & Bleak, A. (1964). *Podosporiella verticillata*, a soil-borne pathogen of some western Gramineae. *Phytopathology* **54**, 353-357.

Kubo, Y., Takano, Y., Endo, N., Yasuda, N., Tajima, S. & Furusawa, I. (1996). Cloning and structural analysis of the melanin biosynthesis gene SCD1 encoding scytalone dehydratase in Colletotrichum lagenarium. *Appl. Environ. Microbiol.* **62**, 4340-4344.

Kupiec, M. & Petes', T. (1988). Allelic and Ectopic Recombination between Ty Elements in Yeast. *Genetics Society of America* **119**, 549-559.

Ladevèze, V., Aulard, S., Aulard, N., Périquet, G. & Lemeunier, F. (1998). Hobo transposons causing chromosomal breakpoints. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 1157-1159.

Lahn, B. T. & Page, D. C. (1999). Four Evolutionary Strata on the Human X Chromosome. *Science* **286**, 964-967.

Lamari, L. & Bernier, C. (1991). Genetics of Tan Necrosis and Extensive Chlorosis in Tan Spot of Wheat Caused by *Pyrenophora tritici-repentis*. *Phytopathology* **81**, 1092-1095.

Lim, J. K. & Simmons, M. J. (1994). Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. *BioEssays* **16**, 269-275.

Lowe, T. M. & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955-964.

Maksakova, I. A. & Mager, D. L. (2005). Transcriptional Regulation of Early Transposon Elements, an Active Family of Mouse Long Terminal Repeat Retrotransposons. *J. Virol.* **79**, 13865-13874.

Mayorga, M. E. & Gold, S. E. (1999). A MAP kinase encoded by the ubc3 gene of Ustilago maydis is required for filamentous growth and full virulence. *Molecular Microbiology* **34**, 485-497.

Medd, R. W. & Campbell, M. A. (2005). Grass seed infection following inundation with Pyrenophora semeniperda. *Biocontrol Science and Technology* **15**, 21-36.

Medd, R. W., Murray, G. M. & Pickering, D. I. (2003). Review of the epidemiology and economic importance of Pyrenophora semeniperda. *Australasian Plant Pathology* **32**, 539-550.

Meyer, Susan E.; Nelson, David L.; Clement, Suzette; Beckstead, Julie 2008. Cheatgrass (*Bromus tectorum*) biocontrol using indigenous fungal pathogens. In: Kitchen, Stanley G.; Pendleton, Rosemary L.; Monaco, Thomas A.; Vernon, Jason, comps. 2008. Proceedings-Shrublands under fire: disturbance and recovery in a changing world; 2006 June 6-8; Cedar City, UT. Proc. RMRS-P-52. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 61-67.

Mieczkowski, P. A., Lemoine, F. J. & Petes, T. D. (2006). Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. *DNA Repair* **5**, 1010-1020.

Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends in Genetics* **23**, 183-191.

Mitchell, T. K. & Dean, R. A. (1995). The cAMP-Dependent Protein Kinase Catalytic Subunit Is Required for Appressorium Formation and Pathogenesis by the Rice Blast Pathogen Magnaporthe grisea. *The Plant Cell* **7**, 1869-1878.

Namiki, F., Matsunaga, M., Okuda, M., Inoue, I., Nishi, K., Fujita, Y. & Tsuge, T. (2001). Mutation of an Arginine Biosynthesis Gene Causes Reduced Pathogenicity in Fusarium oxysporum f. sp. melonis. *Molecular Plant-Microbe Interactions* **14**, 580-584.

Nowrousian, M., Stajich, J. E., Chu, M., Engh, I., Espagne, E., Halliday, K., Kamerewerd, J., Kempken, F., Knab, B., Kuo, H.-C., Osiewacz, H. D., Pöggeler, S., Read, N. D., Seiler, S., Smith, K. M., Zickler, D., Kück, U. & Freitag, M. (2010). <italic>De novo</italic> Assembly of a 40 Mb Eukaryotic Genome from Short Sequence Reads: <italic>Sordaria macrospora</italic>, a Model Organism for Fungal Morphogenesis. *PLoS Genet* **6**, e1000891.

O'Gara, P. (1915). A *Podosporiella* disease of germinating wheat. *Phytopathology* **5**, 323-325.
Oikawa, H. & Tokiwano, T. (2004). Enzymatic Catalysis of the Diels—Alder Reaction in the Biosynthesis of Natural Products. *ChemInform* **35**, no-no.

Pain, A. & Hertz-Fowler, C. (2008). Genomic adaptation: a fungal perspective. *Nat Rev Micro* **6**, 572-573.

Panaccione, D. G., Scott-Craig, J. S., Pocard, J.-A. & Walton, J. D. (1992). A Cyclic Peptide Synthetase Gene Required for Pathogenicity of the Fungus Cochliobolus carbonum on Maize. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6590-6594.

Pfender, W. & Wootke, S. (1988). Microbial Communities of *Pyrenophora*-Infested Wheat Straw as Examined by Multivariate Analysis. *Microbial Ecology* **15**, 95-113.

Raíces, M., Montesino, R., Cremata, J., García, B., Perdomo, W., Szabó, I., Henriksson, G., Hallberg, B. M., Pettersson, G. & Johansson, G. (2002). Cellobiose quinone oxidoreductase from the white rot fungus Phanerochaete chrysosporium is produced by intracellular proteolysis of cellobiose dehydrogenase. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1576**, 15-22.

Rees, R. & Platz, G. (1983). Effect of yellow spot on wheat: Comparison of epidemics at different stages of crop development. *Australian Journal of Agricultural Research* **34**, 39-46.

Remm, M., Storm, C. E. V. & Sonnhammer, E. L. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology* **314**, 1041-1052.

Rogers, L. M., Flaishman, M. A. & Kolattukudy, P. E. (1994). Cutinase Gene Disruption in Fusarium solani f sp pisi Decreases Its Virulence on Pea. *The Plant Cell Online* **6**, 935-945.

Ruiz-Roldán, M. C., Maier, F. J. & Schäfer, W. (2001). PTK1, a Mitogen-Activated-Protein Kinase Gene, Is Required for Conidiation, Appressorium Formation, and Pathogenicity of Pyrenophora teres on Barley. *Molecular Plant-Microbe Interactions* **14**, 116-125.

Sanders, I. R. (2006). Rapid disease emergence through horizontal gene transfer between eukaryotes. *Trends in Ecology & Evolution* **21**, 656-658.

Schilder, A. & Bergstrom, G. (1994). Infection of wheat seed by *Pyrenophora tritici-repentis*. *Canadian Journal of Botany* **72**, 510-519.

Schoonbeek, H., Del Sorbo, G. & De Waard, M. A. (2001). The ABC Transporter BcatrB Affects the Sensitivity of Botrytis cinerea to the Phytoalexin Resveratrol and the Fungicide Fenpiclonil. *Molecular Plant-Microbe Interactions* **14**, 562-571.

Schümann, J. & Hertweck, C. (2007). Molecular Basis of Cytochalasan Biosynthesis in Fungi: Gene Cluster Analysis and Evidence for the Involvement of a PKS-NRPS Hybrid Synthase by RNA Silencing. *Journal of the American Chemical Society* **129**, 9564-9565.

Shieh, M. T., Brown, R. L., Whitehead, M. P., Cary, J. W., Cotty, P. J., Cleveland, T. E. & Dean, R. A. (1997). Molecular genetic evidence for the involvement of a specific polygalacturonase, Pc2, in the invasion and spread of Aspergillis flavus in cotton bolls. *Appl. Environ. Microbiol.* **63**, 3548-3552.

Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*. 1996-2010 <http://www.repeatmasker.org>.

Smith, T. F. & Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol* **147**, 195-197.

Soderlund, C., Nelson, W., Shoemaker, A. & Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Research* **16**, 1159-1168.

Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G. & Valent, B. (1998). Magnaporthe grisea Pathogenicity Genes Obtained Through Insertional Mutagenesis. *Molecular Plant-Microbe Interactions* **11**, 404-412.

Taba, S., Takara, A., Nasu, K., Miyahira, N., Takushi, T. & Moromizato, Z.-i. (2009). Alternaria leaf spot of basil caused by <i>Alternaria alternata</i> in Japan. *Journal of General Plant Pathology* **75**, 160-162.

Tanaka, A., Shiotani, H., Yamamoto, M. & Tsuge, T. (1999). Insertional Mutagenesis and Cloning of the Genes Required for Biosynthesis of the Host-Specific AK-Toxin in the Japanese Pear Pathotype of Alternaria alternata. *Molecular Plant-Microbe Interactions* **12**, 691-702.

Tanaka, A. & Tsuge, T. (2000). Structural and Functional Complexity of the Genomic Region Controlling AK-Toxin Biosynthesis and Pathogenicity in the Japanese Pear Pathotype of Alternaria alternata. *Molecular Plant-Microbe Interactions* **13**, 975-986.

Thornburg, B. G., Gotea, V. & Makalowski, W. (2006). Transposable elements as a significant source of transcription regulating signals. *Gene* **365**, 104-110.

Tonukari, N. J., Scott-Craig, J. S. & Walton, J. D. (2000). The Cochliobolus carbonum SNF1 Gene Is Required for Cell Wall–Degrading Enzyme Expression and Virulence on Maize. *The Plant Cell Online* **12**, 237-248.

Truesdell, G. M., Yang, Z. & Dickman, M. B. (2000). A Gα subunit gene from the phytopathogenic fungus *Colletotrichum trifolii* is required for conidial germination. *Physiological and Molecular Plant Pathology* **56**, 131-140.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M.-K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S. & Boore, J. L. (2006). Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science* **313**, 1261-1266.

Uemura, I., Miyagawa, H. & Ueno, T. (2002). Asymmetric total synthesis of AK-toxins. *Tetrahedron* **58**, 2351-2358.

Urban, M., Bhargava, T. & Hamer, J. E. (1999). An ATP-driven efflux pump is a novel pathogenicity factor in rice blast disease. *EMBO J* **18**, 512-521.

van der Does, H. C. & Rep, M. (2007). Virulence Genes and the Evolution of Host Specificity in Plant-Pathogenic Fungi. *Molecular Plant-Microbe Interactions* **20**, 1175-1182.

Wallace, H. A. H. (1959). A rare seed-borne disease of wheat caused by *Podosporiella verticillata*. *Canadian Journal of Botany* **37**, 509-515.

Wanchanthuek, P., Hallin, P. F., Gouveia-Oliveira, R. & Ussery, D. (2006). Comparative Genomics Using Fungi as Models. *Springer*.

Wasmann, C. C. & VanEtten, H. D. (1995). Transformation-mediated chromosome loss and disruption of a gene for pisatin demethylase decrease the virulence of *Nectria haematococca* on pea. *Molecular Plant-Microbe Interacteraction* **9**, 793-803.

Watkins, J. E., Odvody, G. N., Boosalis, M. G. & Partridge, J. E. (1978). An apidemic of tan spot of wheat in Nebraska. *Plant Dis. Rep.* **62**, 132-134.

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D.-H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., Allison, J., Birchler, J. A., Jiang, J., Dawe, R. K. & Presting, G. G. (2009). Maize Centromere Structure and Evolution: Sequence Analysis of Centromeres 2 and 5 Reveals Dynamic Loci Shaped Primarily by Retrotransposons. *PLoS Genet* **5**, e1000743.

Wöstemeyer, J. & Kreibich, A. (2002). Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Current Genetics* **41**, 189-198.

Xu, J. R., Urban, M., Sweigard, J. A. & Hamer, J. E. M. P.-M. I., 10:187-194. (1997). The CPKA gene of *Magnaporthe grisea* is essential for appressorial penetration. *Molecular Plant-Microbe Interacteraction* **10**, 187-194.

Xu, J.-R., Staiger, C. J. & Hamer, J. E. (1998). Inactivation of the Mitogen-Activated Protein Kinase Mps1 from the Rice Blast Fungus Prevents Penetration of Host Cells but Allows Activation of Plant Defense Responses. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 12713-12718.

Yang, Z. & Dickman, M. B. (1999). Colletotrichum trifolii Mutants Disrupted in the Catalytic Subunit of cAMP-Dependent Protein Kinase Are Nonpathogenic. *Molecular Plant-Microbe Interactions* **12**, 430-439.

Yonow, T., Kriticos, D. J. & Medd, R. W. (2004). The Potential Geographic Range of Pyrenophora semeniperda. *Phytopathology* **94**, 805-812.

Zhang, G. & Berbee, M. L. (2001). Pyrenophora Phylogenetics Inferred from ITS and Glyceradehyde-3-Phosphate Dehydrogenase Gene Sequences. *Mycologia* **93**, 1048-1063.

Chapter 3: SNP Marker Development in *Pyrenophora semeniperda*

Introduction

The frequency of sexual reproduction in *Pyrenophora semeniperda* is currently unknown (Paul, 1969; Boose et al., 2011). Due to limited genetic resources, genetic markers have not been readily available to assess genetic recombination in populations of this fungus. SNP marker development in this fungal species shows promise for detection of recombination levels. Some of the main advantages of using SNP markers over other conventional marker systems include the ability to fully automate the genotyping process, lower costs, and to take advantage of a higher density of DNA polymorphisms (Khlestkina & Salina, 2006). In the past, identification of SNPs was somewhat expensive and time consuming, but with the current advances in genomics and biotechnology, the task of identifying markers has become much easier.

The development of genetic markers, specifically single nucleotide polymophisms (SNPs), in many phytopathogenic fungi have allowed researchers to determine fungicide resistance (Bäumler et al., 2003), identify species introduction events (Delmotte et al., 2008), and understand the history, ecology and evolution of fungal populations (Xu et al., 2007). With the advent of next-generation sequencing, biological research has accelerated at unprecedented rates, allowing the sequencing of whole genomes at a fraction of the cost and time than in the past. These advances have not only led to an increase in the amount of fungal genomes that have been sequened, including several important plant pathogenic fungi (Xu et al., 2006), but has also allows researchers to adapt the technology for genetic marker discovery (Bai et al., 2010; Ellwood et al., 2010; Perry & Rowe, 2010).

Population studies of *P. semeniperda* have been hindered due to the lack of genetic resources. Genetic studies of the genus *Pyrenophora* in general and specifically of *P. semeniperda* have been studied using ITS and the glyceraldehyde-3-phosphate dehydrogenase (*gdp*) gene (Zhang et al., 2001; Boose et al., 2011). Recent genetic diversity tests based on ITS sequences suggest that *P. semeniperda*

populations were established from a common source (Boose et al., 2011). However, the role and frequency of recombination in this species is currently unknown due to the lack of sensitive molecular markers available for this species and because the sexual state is difficult to observe and grow under laboratory conditions (Paul, 1969; Boose et al., 2011).

Here, a method of SNP discovery using pyrosequencing technology by 454 Life Sciences in *P. semeniperda* is presented. 20 polymorphic SNP markers have been identified and used to genotype 600 *P. semeniperda* isolates, representing 36 ecotypes, with the KASPar genotyping system. Current work on the genome assembly of *P. semeniperda* enables us to place these markers in the genome. Also, linkage disequilibrium tests have been run on the SNP genotyping data to assess recombination occurrence among the various *P. semeniperda* populations throughout the IMW.

Materials & Methods

Tissue growth and DNA extraction. Fungal cultures were prepared as described by Boose et al. (2011). DNA was extracted from dried mycellium with the ZR Fungal/Bacterial DNA KitTM from Zymo Research (Zymo Research Coporation, Orange, CA) following the manufactuer's protocol and quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmingtond, DE).

454 Sequencing. Genomic libraries were made from DNA extracted from the CCB06 and STR15 isolates of *P. semeniperda*. These isolates were collected from sites located in Cinder Cone Butte, ID and Strawberry, UT respectively and were chosen based on strong differences in ITS sequence data between these isolates (Boose et al., 2011). A half-plate of genomic library from each isolate was sequenced using a whole-genome shotgun sequencing strategy with the 454 Life Sciences DNA Sequencer FLX and GS FLX Titanium series reagents (454 Life Sciences, Bradford, CT). Titanium

emPCR, library preparation, and sequencing were completed at the Brigham Young University DNA Sequencing Center (Provo, UT).

Assembly and SNP Detection. The 454 sequencing reads were assembled with the CLC Genomics Workbench (CLC bio, Cambridge, MA) using the *de novo* assembly option and default parameters. An ace file of the assembled reads was created in the CLC Genomics Workbench and imported into an in-house, custom-made Perl script called SNP Finder, used to detect species-specific SNPs without the need of a reference genome; SNP Finder parses the ace file and is able to distinguish between CCB06 and STR15 sequence reads. The minimum coverage threshold was set to a value of 15. Residues were only counted as SNPs if they were contained within at least 30% of the overlapping sequences, meaning that the minor allele had a minimum frequency of 30%. SNPs were also counted as ~~"t~~rue" if each allele was 90% homozygous in each isolate. This threshold was used to help distinguish true SNPs from sequencing errors.

Primer design. Sequences containing SNPs were chosen with a minimum assembly coverage of 20. BLASTX was used to check against sequences in the NCBI database to determine if sequences containing SNPs of interest were contained within coding regions. SNP distribution was checked using the assembled genome of the *P. semeniperda* isolate CCB06. Primer sequences were designed for individual true SNPs using Primer Picker (KBiosciences, Beaverly, MA; Table 11).

SNP validation and genotyping. The KASPar Reagent kit (KBiosciences, Beverly, MA) was used for validation and genotyping following recommended protocol from KBiosciences. All DNA template was diluted to a concentration of approximately 4 ng/ul, and a 1 ul (4 ng) sample was taken and dried-down for subsequent PCR. A touchdown PCR program was used with the following thermocycling parameters: 95 ºC for 15 min, 95 ºC for 20 sec, 61-55ºC for 60 sec (10 cycles; the annealing

temperature is incrementally decreased with each cycle: 61, 60.3, 59.7, 59, 58.3, 57.7, 57, 56.3, 55.7, 55 °C), 95 °C for 20 sec, 55 °C for 60 sec (20 cycles).

PCR products were analyzed with the PHERAstar *Plus* HTS Microplate Reader (BMG Labtech, Cary, NC). Genotypes were assigned using the KlusterCaller software package (KBiosciences, Beverly, MA). Primer sets were validated using a panel of 15 isolates. A minor-allele frequency of 20% (3 of 15) between isolates used in the validation was chosen to classify a SNP as polymorphic. Because DNA was extracted from haploid tissue, heterozygotes are not present in the data.

Linkage disequilibrium. Linkage disequilibrium (LD) was tested using the TASSEL software (Trait Analysis by aSSociation, Evolution and Linkage) (Bradbury et al., 2007). The SNP data were first converted to create multiple alignments by gamete using the genotype converter within the TASSEL software. The k-nearest-neighbor algorithm (Cover & Hart, 1967) was used to impute missing data; a Manhatten distance was used in the imputation with 3 K neighbors, and a minimum frequency of row data of 0.80, calculated with an unweighted average. Unbiased p-values were determined by running 100,000 permutations on the converted data to calculate the proportion of permuted haplotype distributions that were less probable than the observed haplotype distribution under the null hypothesis of independence (Weir, 1996).

Results & Discussion

Sequencing and assembly of fungal isolates. DNA sequences of the *P. semeniperda* isolates CCB06 and STR15 were produced using Roche's 454 Life Sciences sequencing platform. In total, the 454 sequencing runs yielded approximately 257 Mb and 278 Mb from the genomic libraries of CCB06 and

STR15 respectively for a total of 535 Mb. 1,375,580 total reads were produced from both genomic libraries averaging 389 bp in length. Over 24,575 contigs with an average length of 1,607 bp were assembled using the CLC Genomics Workbench software package. The total assembly size is approximately 39.49 Mb in length with an average coverage of approximately 9.44. A summary of the assembly is presented in Table 12.

SNP marker development and genotyping. Species-specific SNPs were identified from the assembled reads of CCB06 and STR15 with SNP Finder. 11,566 true species SNPs were identified, making up 0.0293 percent of the genome assembly (between CCB06 and STR15). Summary statistics from SNP Finder are presented in Table 13. The average SNP coverage was approximately 19.29 with a standard deviation of 16.49. A majority of the reported SNPs were transitions (76.11%) either between A/G or C/T residues.

Twenty polymorphic SNP markers were chosen from a validation panel of 15 *P. semeniperda* isolates. The marker loci are from randomly selected anonymous genomic regions and, to the best of our knowledge, are located in non-coding sites. Randomly selected markers were determined to be distributed throughout the genome, averaging 0.91 Mb away from each other when found on the same scaffold (ranging from 0.42 - 1.11 Mb). The markers were kept as far away from each other as possible to help clarify LD and recombination events. A total of 600 *P. semeniperda* isolates from 36 different geographical locations were genotyped using the KASP SNP genotyping system from KBiosciences.

Haplotype groups. In total, 44 haplotypes were identified (Table 14). Haplotype I is the most common haplotype, observed for at least half of all 600 haplotypes. 18 of the populations were fixed at all SNP loci, though, a low sample size (N) may have contributed to this effect in 10 of the reported populations where $N \leq 14$; all fixed populations shared the same haplotype (haplotype I). The remaining 17 populations were analyzed further for nucleotide diversity and LD.

Estimates of nucleotide diversity. Weighted averages of nucleotide diversity (π), nucleotide polymorphism (θ [2Nμ]), and Tajima's D (Tajima, 1989), were calculated for the global population (all isolates) as well as for individual populations using the TASSEL software. A description of the nucleotide diversity of SNP loci is presented in Table 15. Populations were subdivided into three groups, groups I, II, and II, according to SNP diversity measurements and Tajima's D values; group I contains the most diverse populations followed by less diverse populations in group II and finally group III which are fixed at all loci. Levels of SNP diversity is relatively high in the global population ($\pi = 0.164$). Group I populations in Table 15 have high SNP diversity with levels as high as $\pi = 0.442$ and $\theta = 0.276$. Segregation site levels (S) remain high in this group, contributing to the high SNP diversity levels. Tajima's D is positive for most of the group I populations indicating a decreasing population size and/or a balancing selection (Tajima, 1989).

Group II populations show diversity levels 4 – 5 times lower than those in Group I, with values ranging from $0.009 \leq \pi \geq 0.184$ and $0.014 \leq \theta \geq 0.257$. Segregation sites are significantly lower, contributing to the drop in diversity levels, with exception of the CCF and SQC populations. The Tajima's D values for all Group II populations are negative, suggestive of a population size expansion, due to a bottleneck or selective sweep, and/or a purifying selection (Kousathanas et al., 2011). The genotypes for these populations indicate that they may be moving toward fixation at all SNP loci (haplotype I).

Linkage disequilibrium. LD was calculated for the purpose of identifying recombination within these fungal populations. Calculations were made using the TASSEL software and LD estimators, r^2 and D', were used to identify LD between SNP loci. Statistical data of the LD results ($p \leq 0.05$) are reported in Table 16. To gain a perspective on the historical LD in *P. semeniperda*, LD was first calculated for all 600 isolates. The LD levels have a wide range of values within the combined data set for both

estimators (Figure 10) which may indicate a breakage of ancestral LD between SNP loci. LD levels within populations tended to be higher than for those measured for all of the isolates together, with r^2 and D‘ levels ranging from 0.191 – 1 and 0.212 – 1 respectively. Considering that *P. semeniperda* is primarily asexual, it is not surprising that such high levels of LD are observed.

Despite high levels of r^2 and D‘, values < 1 may indicate a breakage of ancestral LD between SNP loci. D‘ values < 1 are observed in many of the populations and indicate that the complete ancestral LD has been disrupted. It would be difficult to conclude that such breakage could be a result of recombination, recurrent mutation, or gene conversion as intermediate levels of D‘ are biologically difficult to interpret and should not be used to measure the strength of LD (Ardlie et al., 2002). It should be noted that estimates of D‘ are strongly inflated in small samples, even for SNPs with common alleles, but especially with rare SNP alleles. D‘ values have a tendency to overestimate the magnitude of LD; such values can be obtained even when markers are in fact in linkage equilibrium. r^2 values < 1 are also observed in these populations, though high, they do indicate a disruption in LD as indicated by its counterpart, D‘. Populations with positive Tajima‘s D values support the idea of LD disruption as these populations also have D‘ values < 1.

Conclusion and future work. Next-generation sequencing technology can be applied to SNP marker development and used for population studies in *P. semeniperda* as has been shown in this study. Nucleotide diversity and linkage disequilibrium calculations may indicate the occurrence of recombination in *P. semeniperda* populations or at least the disruption of ancestral linkage disequilibrium whether by recombination, recurrent mutation or gene conversion, though more evidence is needed to make any conclusions. Such ambiguity regarding sexual reproduction in *Pyrenophora* species is not uncommon. Rau et al., (2003) reported variable levels of linkage disequilibrium across *P. teres* populations, displaying non-significant to highly significant levels of

linkage disequilibrium. This led them to conclude that relative frequencies of sexual and asexual reproduction vary across different environments within *P. teres* populations and which may also be the case with *P. semeniperda*.

ITS sequences have also been analyzed in *P. semeniperda* (Boose et al., 2011) which show that these fungal populations were established from a common source; Boose et al. (2011) hypothesize that this fungus probably arrived with cheatgrass from its native Eurasian range. SSR markers have also been developed and used to determine recombination in this fungal species. These data along with the SNP data of this study will be used to assess genetic recombination in *P. semeniperda* populations and to clarify their modes of reproduction.

Furthermore, future work also includes identifying the idiomorphic alleles at the *MAT* locus, *MAT-1* and *MAT-2*, in *P. semeniperda* populations. This locus is found in heterothallic ascomycetes such as *P. semeniperda* and is referred to as the mating type locus; the sexual cycle is initiated only when two fungal strains carrying different *MAT* alleles interact (Kronstad & Staben, 1997; Turgeon, 1998). The identification of these alleles will help to determine if recombination is even possible within *P. semeniperda* populations and will assist with interpreting linkage disequilibrium results and future 'breeding' programs of this fungus.

LITERATURE CITED

Ardlie, K. G., Kruglyak, L. & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3**, 299-309.

Bai, X., Zhang, W., Orantes, L., Jun, T.-H., Mittapalli, O., Mian, M. A. R. & Michel, A. P. (2010). Combining Next-Generation Sequencing Strategies for Rapid Molecular Resource Development from an Invasive Aphid Species, *Aphis glycines*. *PLoS ONE* **5**, e11370.

Bäumler, S., Sierotzki, H., Gisi, U., Mohler, V., Felsenstein, F. G. & Schwarz, G. (2003). Evaluation of Erysiphe graminis f sp tritici field isolates for resistance to strobilurin fungicides with different SNP detection systems. *Pest Management Science* **59**, 310-314.

Boose, D., Harrison, S., Clement, S. & Meyer, S. (2011). Population genetic structure of the seed pathogen Pyrenophora semeniperda on Bromus tectorum in western North America. *Mycologia* **103**, 85-93.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Cheung, F., Win, J., Lang, J., Hamilton, J., Vuong, H., Leach, J., Kamoun, S., Andre Levesque, C., Tisserat, N. & Buell, C. R. (2008). Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. *BMC Genomics* **9**, 542.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N. D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., Soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J. E. & Birren, B. W. (2005). The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature* **434**, 980-986.

Delmotte, F., Giresse, X., Richard-Cervera, S., M'Baya, J., Vear, F., Tourvieille, J., Walser, P. & Labrouhe, D. T. d. (2008). Single nucleotide polymorphisms reveal multiple introductions into France of Plasmopara halstedii, the plant pathogen causing sunflower downy mildew. *Infection, Genetics and Evolution* **8**, 534-540.

DiGuistini, S., Liao, N., Platt, D., Robertson, G., Seidel, M., Chan, S., Docking, T. R., Birol, I., Holt, R., Hirst, M., Mardis, E., Marra, M., Hamelin, R., Bohlmann, J., Breuil, C. & Jones, S. (2009). De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology* **10**, R94.

Ellwood, S., Liu, Z., Syme, R., Lai, Z., Hane, J., Keiper, F., Moffat, C., Oliver, R. & Friesen, T. (2010). A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. *Genome Biology* **11**, R109.

Evidente, A., Andolfi, A., Vurro, M., Zonno, M. C. & Motta, A. (2002). Cytochalasins Z1, Z2, and Z3, Three 24-Oxa[14]cytochalasans Produced by Pyrenophora semeniperda. *ChemInform* **33**, 232-232.

Khlestkina, E. & Salina, E. (2006). SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. *Russian Journal of Genetics* **42**, 585-594.

Kousathanas, A., Oliver, F., Halligan, D. L. & Keightley, P. D. (2011). Positive and Negative Selection on Noncoding DNA Close to Protein-Coding Genes in Wild House Mice. *Molecular Biology and Evolution* **28**, 1183-1191.

Kronstad, J. W. & Staben, C. (1997). Mating type in filamentous fungi. *Annual Review of Genetics* **31**, 213-244.

Meyer, S. E., Quinney, D., Nelson, D. L. & Weaver, J. (2007). Impact of the pathogen Pyrenophora semeniperda on Bromus tectorum seedbank dynamics in North American cold deserts. *Weed Research* **47**, 54-62.

Paul, A. R. (1969). The production of Pyrenophora semeniperda in culture. *Transactions of the British Mycological Society* **52**, 373-379.

Perry, J. C. & Rowe, L. (2011). Rapid Microsatellite Development for Water Striders by Next-Generation Sequencing. *Journal of Heredity* **102**, 125-129.

Rau, D., Brown, A. H. D., Brubaker, C. L., Attene, G., Balmas, V., Saba, E. & Papa, R. (2003). Population genetic structure of *Pyrenophora teres* Drechs. the causal agent of net blotch in Sardinian landraces of barley (*Hordeum vulgare L.*). *TAG Theoretical and Applied Genetics* **106**, 947-959.

Soanes, D. M., Skinner, W., Keon, J., Hargreaves, J. & Talbot, N. J. (2002). Genomics of Phytopathogenic Fungi and the Development of Bioinformatic Resources. *Molecular Plant-Microbe Interactions* **15**, 421-427.

Stewart, T. E., Allen, P. S. & Meyer, S. E. (2009). First report of *Pyrenophora semeniperda* in Turkey and Greece. *Plant Disease* **93**, 1351.

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**, 585-595.

Turgeon, B. G. (1998). Application of mating type gene technology to problems in fungal biology. *Phytopathology* **36**, 115-137.

Weir, B. G. (1996). Genetic data analysis II. Sunderland, MA: Sinauer.

Xu, J., Guo, H. & Yang, Z.-L. (2007). Single nucleotide polymorphisms in the ectomycorrhizal mushroom Tricholoma matsutake. *Microbiology* **153**, 2002-2012.

Xu, J. Y., Xu, G. B. & Chen, S. L. (2009). A new method for SNP discovery. *BioTechniques* **46**, 201-208.

Yonow, T., Kriticos, D. J. & Medd, R. W. (2004). The Potential Geographic Range of Pyrenophora semeniperda. *Phytopathology* **94**, 805-812.

Zhang, G. & Berbee, M. L. (2001). Pyrenophora Phylogenetics Inferred from ITS and Glyceradehyde-3-Phosphate Dehydrogenase Gene Sequences. *Mycologia* **93**, 1048-1063.

TABLES

Table 1. Primary alignment metrics of the genome assembly.

	Genomic Library	Paired Library	cDNA Library
no. of reads	691,955	2,067,802	333,014
no. of bases	233,183,002	469,930,696	110,891,692
avg. read length	371.47	362.06	358.2

Table 2. Progression of the *P. semeniperda* Newbler assembly.

	Shotgun	Paired-end
Peak Depth	6	17
Est. Genome Size	42.6	40.1
Read Status:		
Num. Aligned Reads	678259	2621753
Num Aligned Bases	252629934	672385374
Scaffold Metrics:		
Number of Scaffolds		54
Number of Bases		32539172
Avg Scaffold Size		602577
N50* Scaffold Size		1473690
Largest Scaffold Size		2679305
Large Contig Metrics:		
Number of Contigs	7325	7777
Number of Bases	31201036	32316274
Avg Contig Size	4259	41591
N50 Contig Size	6499	104587
Largest Contig Size	42958	871975
All Contig Metrics:		
Number of Contigs	7890	1001
Number of Bases	31367229	32370938

*The N50 length describes 50% of the bases in the assembly that are in a contig/scaffold of at least that length.

Table 3. General features of the *P. semeniperda* genome

GC percentage in coding region	52.53
GC percentage in non-coding regions	47.73
tRNA genes	91
protein coding genes (CDSs)	11,453
percent coding	46.47
avg. CDS size (min/max)	1,312 bp (72 bp/25,382bp)

Table 4. Identified interspersed and simple repeats in *P. semeniperda*.

	number of elements	length occupied (bp)	% of genome sequence
Retroelements	447	362220	1.12%
LINEs:	8	2295	0.01%
R1/LOA/Jockey	1	79	0%
LTR elements:	439	359925	1.11%
Ty1/Copia	51	38056	0.12%
Gypsy/DIRS1	388	321869	1%
DNA transposons	296	248288	0.77%
hobo-Activator	1	40	0%
Tc1-IS630-Pogo	293	248200	0.77%
MuDR-IS905	1	41	0%
Unclassified:	2	192	0%
Total interspersed repeats:		610700	1.89%
Small RNA:	47	12191	0.04%
Simple repeats:	2988	138005	0.43%
Low complexity:	1521	98370	0.30%

Table 5. Pathogenicity genes cloned from phytopathogenic fungi

Gene	No. of genes in *P. semeniperda*	Fungus	Name and putative function	References
Novel' and nonclassified pathogenicity genes				
SNF1	7	*Cochliobolus carbonum*	Regulation of catabolite-repressed genes; glucose starvation	Tonukari et al., 2000
CLTA1	1	*Colletotrichum lindemuthianum*	Regulator of biotrophic/necrotrophic switch	Dufresne et al., 2000
Cuticle and cell wall degradation				
CutA	1	*Fusarium solani* f.sp. *pisi*	Cutinase	Dickman et al., 1989 Rogers et al., 1994
BcPGA1	1	*Botryotinia fuckeliana*	Endopolygalacturonase	ten Have et al., 1998
Acpg1	1	*Alternaria citri*	Endopolygalacturonase	Isshiki et al., 2001
pecA	1	*Aspergillus flavus*	Endopolygalacturonase	Shieh et al., 1997
Fungal Toxins				
Tox2 locus		*C. carbonum*	HC-toxin biosynthesis	
HTS1	2		Peptide synthetase	Panaccione et al., 1992
TOXC	2		β subunit fatty acid synthase	Ahn & Walton, 1997
TOXF	1		Amino acid synthesis	Cheng et al., 1999
AK-toxin		*Alternaria*		
Akt1	1	*alternata* Japanese	Carboxyl-activating protein	Tanaka et al., 1999
AktR-2	1	pear pathotype	Regulatory gene	Tanaka & Tsuge, 2000
Akt3-2	1		Hydratase/isomerase	Tanaka & Tsuge, 2000

Table 5. *continued.*

Gene	No. of genes in *P. semeniperda*	Fungus	Name and putative function	References
AMT	8	*A. alternata* apple pathotype	Non-ribosomal peptide synthase; AM-toxin biosynthesis	Johnson et al., 2000
Infection Structure				
BUF	4	*Magnaporthe grisea*	Polyhydroxynaphthalene reductase; melanin biosynthesis	Chumley & valent, 1990
SD1	1	*C. lagenarium*	Scytalone dehydrogenase; melanin biosynthesis	Kubo et al., 1996
Responding to host environment				
atrB	6	*B. fuckeliana*	ABC transporter; plant toxin efflux?	Schoonbeek et al., 2001
ABC1	2	*M. grisea*	ABC transporter; plant toxin efflux?	Urban et al., 1999
ARG1	1	*Fusarium oxysporum* f.sp. *Melonis*	Argininosuccinate lyase; arginine biosynthesis	Namiki et al., 2001
PTH2	2	*M. grisea*	Carnitine acyl transferase	Sweigard et al., 1998
PTH3	1	*M. grisea*	Imidazole glycerol phosphate dehydrogenase; histidine biosynthesis	Sweigard et al., 1998
MAK1	1	*N. haematococca*	Maakiain detoxification	Enkerli et al., 1998
PTH9	1	*M. grisea*	Neutral trehalase; stress protection	Sweigard et al., 1998
odc1	1	*Stragonospora nodorum*	Ornithine decarboxylase; polyamine biosynthesis	Bailey et al., 2000
PDA1	8	*Nectria haematococca*	Pisatin demethylase; pisatin detoxification	Wasmann & VanEtten, 1995

Table 5. *continued*

Gene	No. of genes in *P. semeniperda*	Fungus	Name and putative function	References
Signal cascade components				
cpkA/pth4	3	*M. grisea*	Catalytic subunit cAMP-dependent protein kinase	Mitchell & Dean, 1995
pkaC	1	*C. trifolii*	Catalytic subunit cAMP-dependent protein kinase	Xu et al., 1997 Yang & Dickman, 1999
ptk1	1	*Pyrenophora teres*	MAP kinase	Ruiz-Roldán et al., 2001
kpp2	1	*U. maydis*	MAP kinase	Mayorga and Gold, 1999
MPS1	1	*M. grisea*	MAP kinase	Xu et al., 1998
EMK1	1	*C. gloeosporioides*	Protein kinase kinase	Kim et al., 2000
SUM1	1	*M. grisea*	Regulatory subunit, cAMP-dependent protein kinase	Adachi & Hammer, 1998 Barrett et al., 1993
ctg1	2	*Colletotrichum trifolii*	α subunit of G protein	Truesdell et al., 2000

Table 6. Total pathogenicity genes cloned from Phytopathogenic fungi also found in *P. semeniperda*

Novel' and nonclassified pathogenicity genes	8
Cuticle and cell wall degradation	4
Fungal Toxins	16
Infection Structure	5
Responding to host environment	23
Signal Cascade	11
Total	67

Table 7. Putative infection-related genes in *P. semeniperda*

Gene product	No. of putative genes
Hydrolases	
Protease	13
Serine protease	5
Metalloprotease	66
Cysteine proteases	12
Glycosyl hydrolases	132
Pectinases	
Pectinesterase	98
Pectate lyase	12
Cutinase	12
Chitinase	5
Lipase	94
Phospholipases	143
Phospholipase C	71
Phospholipase D	4
Protease inhibitors, all	3
Cystatin	0
Kazal	3
Protein Toxins	
Crn family	0
NPP family	1
PcF family	0
Secondary Metabolite Biosynthesis	
NR peptide synthase	16
PKT synthase	25
Cytochrome P450	73
ABC transporters, all	40
MRP	16
PDR	52
MDR	14
Effectors	
Elicitin	9

Table 8. Total putative infection-related genes in *P. semeniperda*

Hydrolases	667
Protease inhibitors	6
Protein Toxins	1
Secondary Metabolite Biosynthesis	114
ABC transporters	122
Effectors	9
Total	919

Table 9. GO terms assoc. with secreted gene products

GO identifier	Description	No. of genes
Molecular Function		
GO:0016787	Hydrolase activity	210
GO:0016491	Oxidoreductase activity	114
GO:0043167	Ion binding	97
GO:0016740	Transferase activity	56
GO:0000166	Nucleotide binding	54
GO:0048037	Cofactor binding	39
GO:0030246	Carbohydrate binding	28
GO:0005515	Protein binding	19
GO:0046906	Tetrapyrrole binding	17
GO:0001871	Pattern binding	17
GO:0016874	Ligase activity	11
GO:0016853	Isomerase activity	10
GO:0016829	Lyase activity	9
Biological process		
GO:0055114	Oxidation reduction	62
GO:0006508	Proteolysis	32
GO:0006629	Lipid metabolic process	27
GO:0043581	Mycelium development	26
GO:0050794	Regulation of cellular process	16
GO:0044248	Cellular catabolic process	13
GO:0006464	Protein modification process	13
GO:0006032	Chitin catabolic process	9
GO:0051591	Response to cAMP	8
GO:0045493	Xylan catabolic process	8

Molecular function GO terms are limited to level 3; GO term associated secretion gene products; adapted from Ellwood et al., 2010.

Table 10. Putative *Che* gene in P. semeniperda

***Che* Gene (*P. expansum*)**	**No. of genes (*P. semeniperda)***	**Putative Gene Function**	**Pfam Domain**
A	5	PKS-NRPS protein	AMP-Binding Domain
B	2	enoyl reductase	-
C	1	transcription factor	-
D	1	monooxygenase	p450
E	5	monooxygenase	-
F	3	transcription factor	-
G	4	monooxygenase	p450

Table 11. SNP primer sequences

SNP Loci		Primer Sequences	SNP Loci		Primer Sequences
1	ALC	GAAGGTGACCAAGTTCATGCTTGGCACGAACGGTGCTAACTC	**23**	ALA	GAAGGTGACCAAGTTCATGCTGCCGCTTCGCAGTCATGTATGAT
	ALT	GAAGGTCGGAGTCAACGGATTTGGCACGAACGGTGCTAACTT		ALG	GAAGGTCGGAGTCAACGGATTCCGCTTCGCAGTCATGTATGAC
	C	CTGGTGTTGCGAGTGCAATTGACAA		C	GCAACTTCGTGCAGCTGATGATGAA
2	ALA	GAAGGTGACCAAGTTCATGCTTGTAGTAAAGGAGGGTTCCACATT	**25**	ALC	GAAGGTGACCAAGTTCATGCTCCAAGTTACGAGACGTAAGTCCG
	ALG	GAAGGTCGGAGTCAACGGATTGTAGTAAAGGAGGGTTCCACATC		ALT	GAAGGTCGGAGTCAACGGATTCCAAGTTACGAGACGTAAGTCCA
	C	GAAGTGTAGCTACGTCTTCCCTTCTA		C	ATGATCTCATGCAGAGCAATATTACCCTT
3	ALA	GAAGGTGACCAAGTTCATGCTCTTCCCACCGTCCGTAAATGA	**27**	ALA	GAAGGTGACCAAGTTCATGCTGCACGAGAGCTTGTAAGTGGCAA
	ALG	GAAGGTCGGAGTCAACGGATTCTTCCCACCGTCCGTAAATGG		ALG	GAAGGTCGGAGTCAACGGATTCACGAGAGCTTGTAAGTGGCAG
	C	GCGCTATTCCCATTCCTCCATGATA		C	ATTGCCCGTTACTATAACGCACCGA
5	ALG	GAAGGTGACCAAGTTCATGCTAAGACTGGTGGTGTCATCGAGC	**29**	ALC	GAAGGTGACCAAGTTCATGCTGATAAAGGTGCGCTTCACCGATG
	ALT	GAAGGTCGGAGTCAACGGATTAAAGACTGGTGGTGTCATCGAGA		ALT	GAAGGTCGGAGTCAACGGATTAGATAAAGGTGCGCTTCACCGATA
	C	GCACCACGCACCACTTAGCGTT		C	CCATGGCTAGAATTAGGGCGCTAA
6	ALC	GAAGGTGACCAAGTTCATGCTGAACACATCACTGCGCCAGTC	**30**	ALC	GAAGGTGACCAAGTTCATGCTGGTCGCCTTGCTGTGGAGTAC
	ALT	GAAGGTCGGAGTCAACGGATTGGAACACATCACTGCGCCAGTT		ALT	GAAGGTCGGAGTCAACGGATTGGTCGCCTTGCTGTGGAGTAT
	C	CCATGCCGCAATGAGCAGACTGA		C	CCTCGACCCCTCTGGCCAAAA
8	ALC	GAAGGTGACCAAGTTCATGCTCAACATGGTCGTCACATCTCGC	**31**	ALA	GAAGGTGACCAAGTTCATGCTCCGTCATCTTCTCAAACACTACTT
	ALG	GAAGGTCGGAGTCAACGGATTCAACATGGTCGTCACATCTCGG		ALG	GAAGGTCGGAGTCAACGGATTCCGTCATCTTCTCAAACACTACTC
	C	GGCAGCAGACGCTTAATTGCACAT		C	GGCGTGCAATTGTTGCTGTTCAGAT
10	ALA	GAAGGTGACCAAGTTCATGCTCAATGGTTGAGGCAATGCCGCAA	**32**	ALA	GAAGGTGACCAAGTTCATGCTCGAGAAATCGGACGGCGCTGA
	ALG	GAAGGTCGGAGTCAACGGATTAATGGTTGAGGCAATGCCGCAG		ALG	GAAGGTCGGAGTCAACGGATTGAGAAATCGGACGGCGCTGG
	C	CAGCTCTGCTGCAAATTATTTCGTTGTAA		C	GTCTCTCCCTAAGACAGTCCATGTT

Table 11. *continued*

SNP Loci		Primer Sequences	SNP Loci		Primer Sequences
12	ALA	GAAGGTGACCAAGTTCATGCTGGAAGGAGATGAAGAAACGAGTGTA	**33**	ALA	GAAGGTGACCAAGTTCATGCTCAGACCAAGTCAAGATATACAAATAAATAATT
	ALT	GAAGGTCGGAGTCAACGGATTGGAAGGAGATGAAGAAACGAGTGTT		ALG	GAAGGTCGGAGTCAACGGATTCAGACCAAGTCAAGATATACAAATAAATAATC
	C	GCGCTAATCACTTGGTCCGACCA		C	CTCAATACGCACCTTGTCCACTGTA
14	ALA	GAAGGTGACCAAGTTCATGCTGCGATGAAGCTATTCGGAGTACA	**34**	ALA	GAAGGTGACCAAGTTCATGCTTGAAAGACGTTGTTGGTGGCGT
	ALG	GAAGGTCGGAGTCAACGGATTGCGATGAAGCTATTCGGAGTACG		ALG	GAAGGTCGGAGTCAACGGATTGAAAGACGTTGTTGGTGGCGC
	C	GGAGGATCTGGCCCAGCTTGTA		C	GCGTCGTCTCTCCTCGGAGAA
15	ALC	GAAGGTGACCAAGTTCATGCTAAAGAGTAAATTAGAAAAAGCAAAGGAGTG	**36**	ALA	GAAGGTGACCAAGTTCATGCTGGCACAGGCGCGGTGGA
	ALT	GAAGGTCGGAGTCAACGGATTGAAAGAGTAAATTAGAAAAAGCAAAGGAGTA		ALG	GAAGGTCGGAGTCAACGGATTGGCACAGGCGCGGTGGG
	C	CTCAAATCACGCTCCTTCTTCTGCAT		C	TGATGCCAGTGACCTCGCTTACTA

Al' - denotes the SNP-specific primer; 'C' - denotes the common primer

Table 12. Summary of CCB06 & STR15 genome assembly

	Count	Average Length	Total Bases
Reads	1,375,380	388.97	535,056,975
Contigs	24,577	1,607	39,493,729
N50 Length	3,957		
Coverage	9.44		

Table 13. SNP Finder summary statistics

True SNP Results		
Total SNPs found	11566	
Total contigs containing SNPs	4890	
Total SNps / assembly length	0.000293	
True SNP Distribution		
A/C SNP count	674	(5.83%)
A/G SNP count	4408	(38.11%)
A/T SNP count	547	(4.73%)
C/G SNP count	839	(7.25%)
C/T SNP count	4395	(38.00%)
G/T SNP count	703	(6.08%)
True SNP Coverage Statistics		
SNP average coverage	19.29	
SNP coverage standard deviation	16.49	

Table 14. Haplotypes identified in *P. semeniperda* populations

Haplotype	**1**	**2**	**3**	**5**	**6**	**8**	**10**	**12**	**14**	**15**	**23**	**25**	**27**	**29**	**30**	**31**	**32**	**33**	**34**	**36**
I	**T**	**G**	**G**	**G**	**T**	**C**	**G**	**T**	**G**	**C**	**A**	**A**	**G**	**C**	**A**	**C**	**G**	**C**	**T**	**A**
II	C	A	A	T	.	G	A	A	.	.	G	.	A	.	G	G	.	.	C	G
III	C	A	.	.	.	G	.	.	A	.	G	.	.	T	.	G	.	A	.	G
IV	C	.	.	T	C	G	.	A	.	.	G	.	A	T	G	G	.	A	C	G
V	C	.	.	.	.	G	.	A	.	.	G	.	.	T	.	G	.	A	C	G
VI	C	.	.	T	.	G	A	A	.	.	G	.	A	T	.	G	.	A	C	G
VII	C	.	.	T	.	G	A	A	.	.	G	.	A	T	.	G	.	A	C	G
VIII	C	.	.	T	.	G	A	A	.	.	G	.	A	T	.	G	.	A	C	G
IX	C	A	A	.	C	G	.	A	A	T	G	G	A	T	G	G	.	A	C	G
X	C	A	A	T	C	G	A	A	A	T	G	G	A	T	G	G	A	A	C	G
XI	C	.	A	.	C	G	A	A	.	T	G	G	A	T	G	G	.	A	C	G
XII	C	A	A	T	.	G	A	A	A	T	G	G	A	T	.	G	.	A	C	G
XIII	C	.	A	T	.	G	A	A	A	T	G	G	A	T	G	G	.	A	C	G
XIV	C	.	A	.	.	G	A	A	.	.	G	.	A	T	G	G	A	A	C	G
XV	C	A	A	T	.	G	A	A	A	.	G	.	.	T	G	G	.	A	C	G
XVI	C	A	A	.	.	G	.	A	.	T	G	.	A	T	G	G	.	A	C	G
XVII	C	.	.	.	.	G	A	A	A	.	G	.	A	T	G	G	.	A	C	G
XVIII	C	.	.	.	.	G	A	A	A	T	G	.	A	T	.	G	.	A	C	G
XIX	C	.	.	.	.	G	A	A	A	T	G	.	A	T	G	G	.	A	C	G
XX	C	.	.	.	.	G	.	A	A	T	G	.	A	T	G	G	.	A	C	G
XXI	C	A	.	.	.	G	.	A	A	T	G	.	A	T	G	G	.	A	C	G
XXII	C	A	A	.	.	G	A	A	A	T	G	.	A	T	G	G	.	A	C	G

Table 14. *continued*

Haplotype	1	2	3	5	6	8	10	12	14	15	23	25	27	29	30	31	32	33	34	36
XXIII	C	A	.	.	.	G	A	A	A	T	G	.	A	T	G	G	.	A	C	G
XXIV	C	.	.	.	.	G	A	A	.	T	G	.	A	T	G	G	.	A	C	G
XXV	C	A	.	.	.	G	A	A	.	T	G	.	A	T	.	G	.	A	C	G
XXVI	C	A	.	.	.	G	A	A	.	T	G	.	A	T	G	G	.	A	C	G
XXVII	.	.	.	.	-	.	-	.	.	.	-	-	.	.	.	.	A	-	.	.
XXVIII	.	.	.	.	.	G	A	A	A	.	G	.	.	.	.	.	A	.	C	.
XXIX	.	.	.	.	.	G	A	A	A	.	G	.	.	.	.	.	.	.	C	G
XXX	.	.	.	.	.	G	A	A	A	.	G	.	.	T	.	.	.	.	C	.
XXXI	.	.	.	.	.	G	A	A	A	.	.	.	.	.	.	.	.	.	C	.
XXXII	.	.	.	.	.	G	A	A	A	.	G	.	.	.	.	.	.	.	C	.
XXXIII	.	A	.	.	.	C	.	.	A	.	.	.	.	.	.	G	.	.	.	G
XXXIV	.	A	.	.	.	G	.	A	A	.	.	.	.	.	.	.	.	.	.	.
XXXV	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
XXXVI	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
XXXVII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
XXXVIII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
XXXIX	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
XL	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
XLI	.	A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
XLII	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
XLIII	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
XLIV	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.

Dots indicate identity with the reference sequence (haplotype I); dashes indicate an unknown allele.

Table 15. Nucleotide diversity within *P. semeniperda* populations

	Population	Site Count*	AvgSiteCount	S**	π per SNP§	θ per SNP	Tajima's D	Haplotypes
	All populations	20	20	20	0.164	0.143	0.361	
	BFL	20	20	16	0.311	0.214	1.621	I, III, V, VI, VII, XVII, XX
	CCB	20	20	20	0.269	0.259	0.142	I, X, XVI, XXV, XXXII, XXXVIII
	DRL	20	20	19	0.29	0.241	0.699	I, II, XII, XV, XXIII, XXXII, XXXVII, XLII, XLIV
	FSH	20	17.688	19	0.313	0.26	0.744	I, IV, XIV, XVIII, XXII, XXIX, XXX, XXXII, XL, XLII, XLIII
	HRN	20	20	1	0.014	0.014	0.021	I, XLII
Group I	KAH	20	19.2	18	0.323	0.276	0.677	I, XI, XIX, XXIII, XL, XLII, XLIII
	MAR	20	19.428	19	0.442	0.264	2.526	I, IX, XVIII, XXI, XXII, XXIV, XXXVII
	MLV	20	20	1	0.017	0.017	0.014	I, XLII
	PKC	20	18.758	17	0.258	0.256	0.04	I, XXIII, XXVI, XXVII, XXVIII, XLII, XLIV
	PVM	20	20	14	0.336	0.189	2.744	I, VIII, XVII, XXXII
	SDM	20	18.477	6	0.092	0.09	0.093	I, XXXI, XXXIV, XLI, XLII, XLIII
	BRU	20	20	1	0.025	0.027	-0.612	I, XXXVII
	CCF	20	20	18	0.184	0.257	-1.075	I, XIII, XIV, XXXII, XXXVII, XLIII
	CF	20	20	1	0.025	0.027	-0.612	I, XLII
Group II	CKT	20	20	2	0.022	0.029	-0.634	I, XLII, XLIII
	CNF	20	20	1	0.007	0.015	-1.155	I, XLIII
	DIN	20	20	3	0.011	0.038	-1.733	I, XXXVI, XLII, XLIII
	DUJ	20	20	1	0.009	0.014	-0.591	I, XLII
	GUS	20	20	2	0.013	0.027	-1.158	I, XXXVII, XL

Table 15. *continued*

	Population	Site Count*	AvgSiteCount	S**	π per SNP§	θ per SNP	Tajima's D	Haplotypes
	MIL	20	20	1	0.009	0.014	-0.591	I, XLII
	MOM	20	20	1	0.016	0.021	-0.933	I, XLIII
	PAK	20	20	3	0.022	0.044	-1.377	I, XXXVIII, XLII, XLIII
	SQC	20	19.565	7	0.037	0.094	-1.897	I, XXXVII, XLII
Group II	STR	20	19.473	4	0.027	0.057	-1.504	I, XXXV, XXXVII, XLIII
	SWR	20	20	1	0.005	0.014	-1.164	I, XLII
	VERN	20	20	1	0.01	0.017	-1.111	I, XXXIX
	WHV	20	20	7	0.035	0.096	-2.002	I, XXXII, XXXV, XLIII
	WRK	20	20	2	0.013	0.027	-1.175	I, XXXVII, XLII
	BRD	20	20	0	0	0	-	I
	DFF	20	20	0	0	0	-	I
	DOG	20	20	0	0	0	-	I
Group III	INV	20	20	0	0	0	-	I
	JIM	20	20	0	0	0	-	I
	LSC	20	20	0	0	0	-	I
	TMC	20	20	0	0	0	-	I
	WRS	20	20	1	0.033	0.033	-	I

* Site Count indicates the number of SNPs sites per isolate within the population; **S = segregation sites, or polymorphic sites;
§ π = diversity, average pair-wise divergence

Table 16. Comparison of LD, measured using r^2 and D'

	r^2					D'					N
	Mean	Median	Upper quartile	Lower quartile	Minimum	Mean	Median	Upper quartile	Lower quartile	Minimum	Mean
All populations	0.339	0.271	0.548	0.123	0.01	0.739	0.81	0.911	0.635	0.212	600
BFL	0.737	0.799	1	0.555	0.308	0.981	1	1	1	0.555	24
CCB	0.616	0.496	0.765	0.46	0.28	0.979	1	1	1	0.678	27
CCF	0.808	0.627	1	0.627	0.627	1	1	1	1	1	19
DRL	0.49	0.395	0.718	0.303	0.191	0.946	1	1	1	0.613	29
FSH	0.643	0.674	0.787	0.45	0.246	0.975	1	1	1	0.74	21
KAH	0.706	0.74	1	0.375	0.375	0.983	1	1	1	0.688	14
MAR	0.744	0.814	0.869	0.654	0.234	0.986	1	1	1	0.79	21
PKC	0.862	1	1	1	0.422	0.965	1	1	1	0.65	15
PVM	0.953	1	1	0.835	0.835	1	1	1	1	1	23
SDM	0.578	0.531	0.673	0.428	0.428	0.966	1	1	1	0.666	16
SQC	1	1	1	1	1	1	1	1	1	1	23
WHV	1	1	1	1	1	1	1	1	1	1	22

N is the average sampling size per population

FIGURES

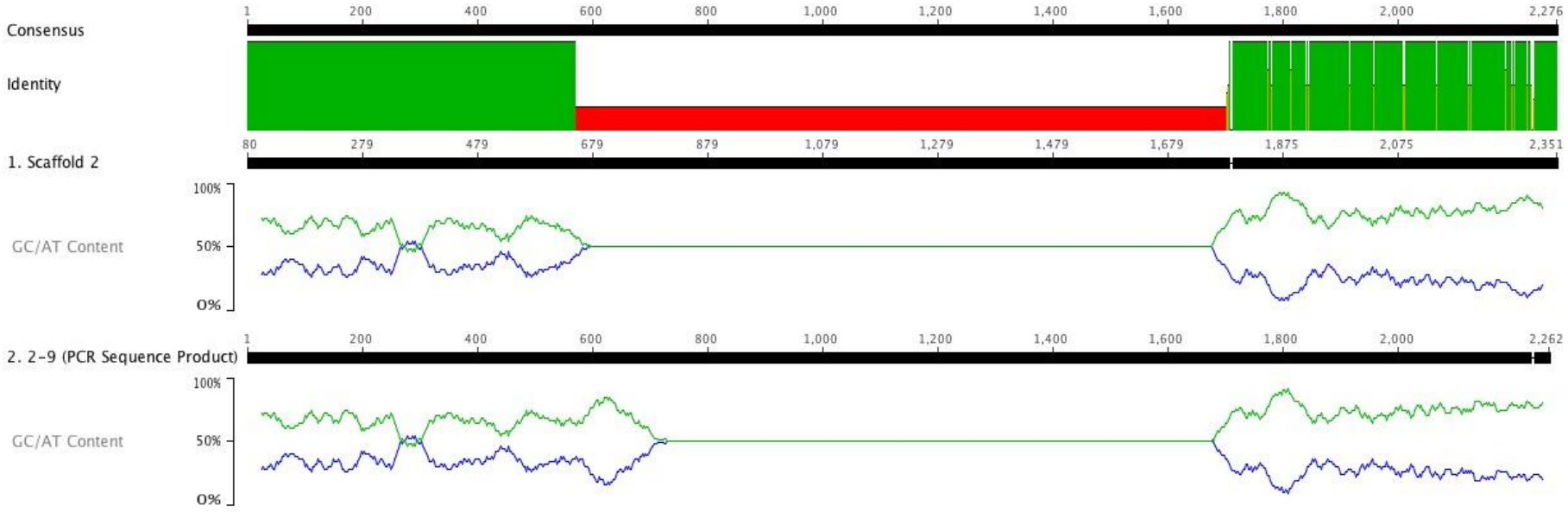

Figure 1 Comparison of the *P. semeniperda* 454 sequence scaffold compared to Sanger-sequenced PCR fragments. The ends of the PCR fragment (2-9) were sequenced and aligned to scaffold 2 of the *P. semeniperda* genome assembly. Positive alignments are presented in green while the N domain is presented in red. GC content (blue line) remains relatively low throughout the sequence. There is slight overlap of the N domain in scaffold 2 and the left sequence of 2-9.

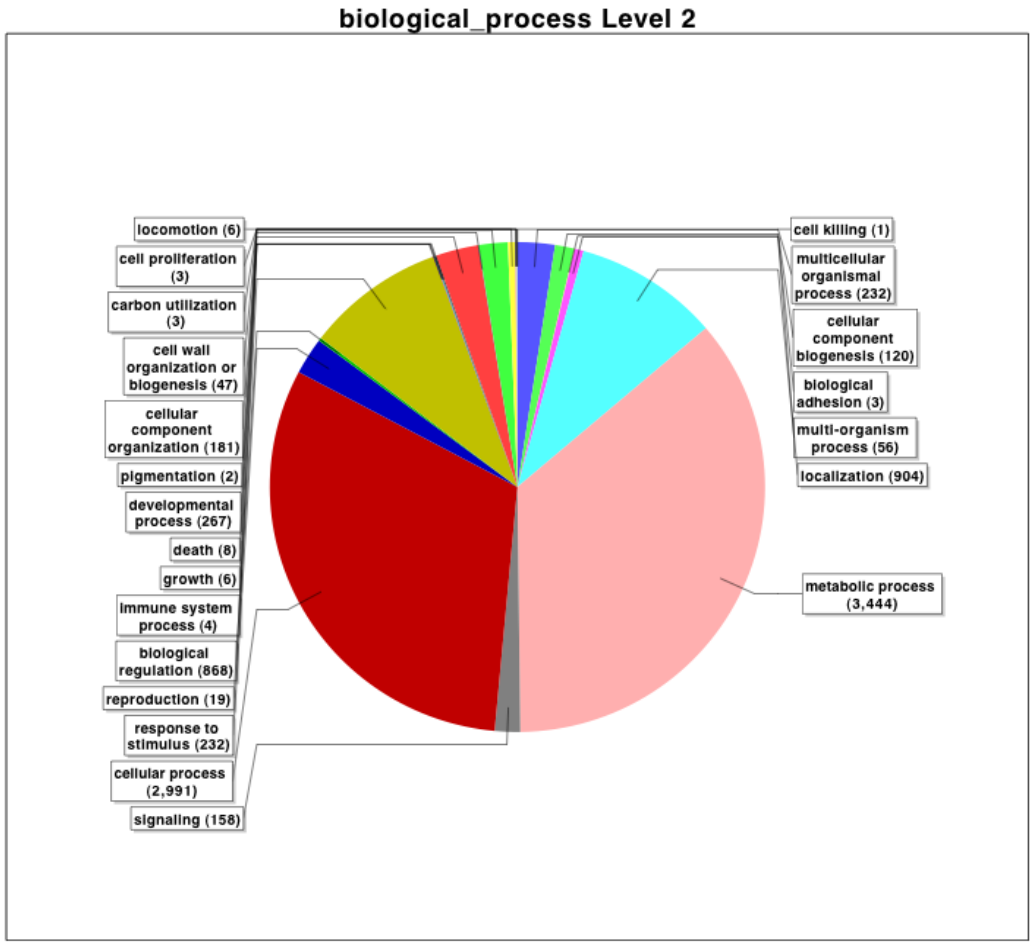

Figure 2 Blast2GO functional annotation of putative genes from *P. semeniperda* for Biological Process. Level 2 GO terms are presented in the figure.

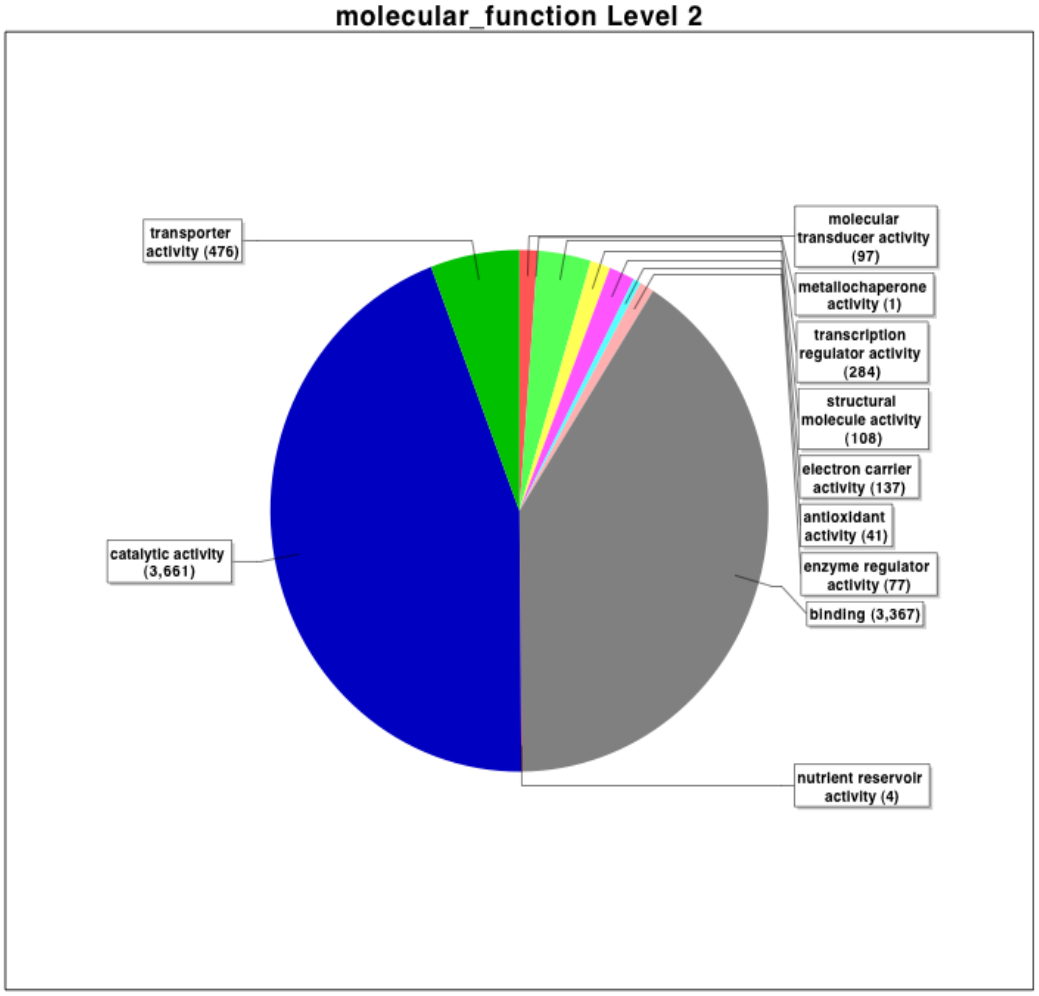

Figure 3 Blast2GO functional annotation of putative genes from *P. semeniperda* for Molecular Function. Level 2 GO terms are presented in the figure.

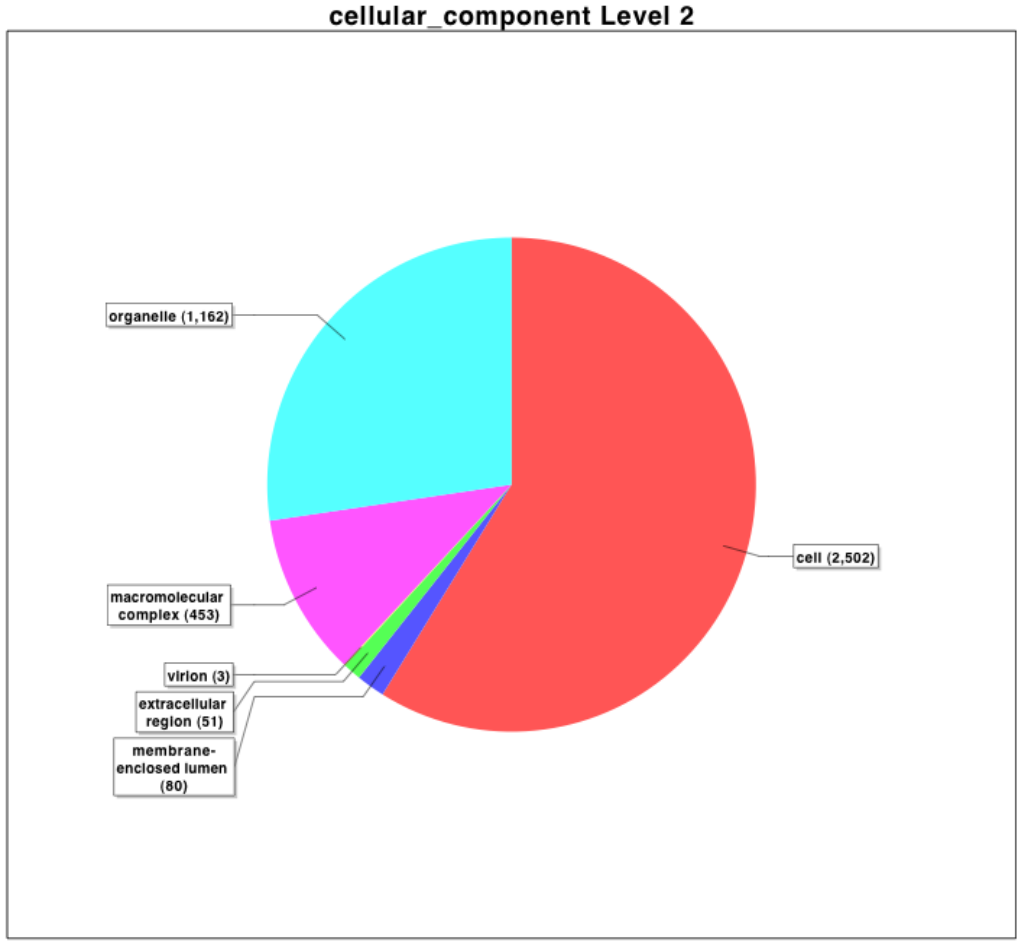

Figure 4 Blast2GO functional annotation of putative genes from *P. semeniperda* for Cellular Components. Level 2 GO terms are presented in the figure.

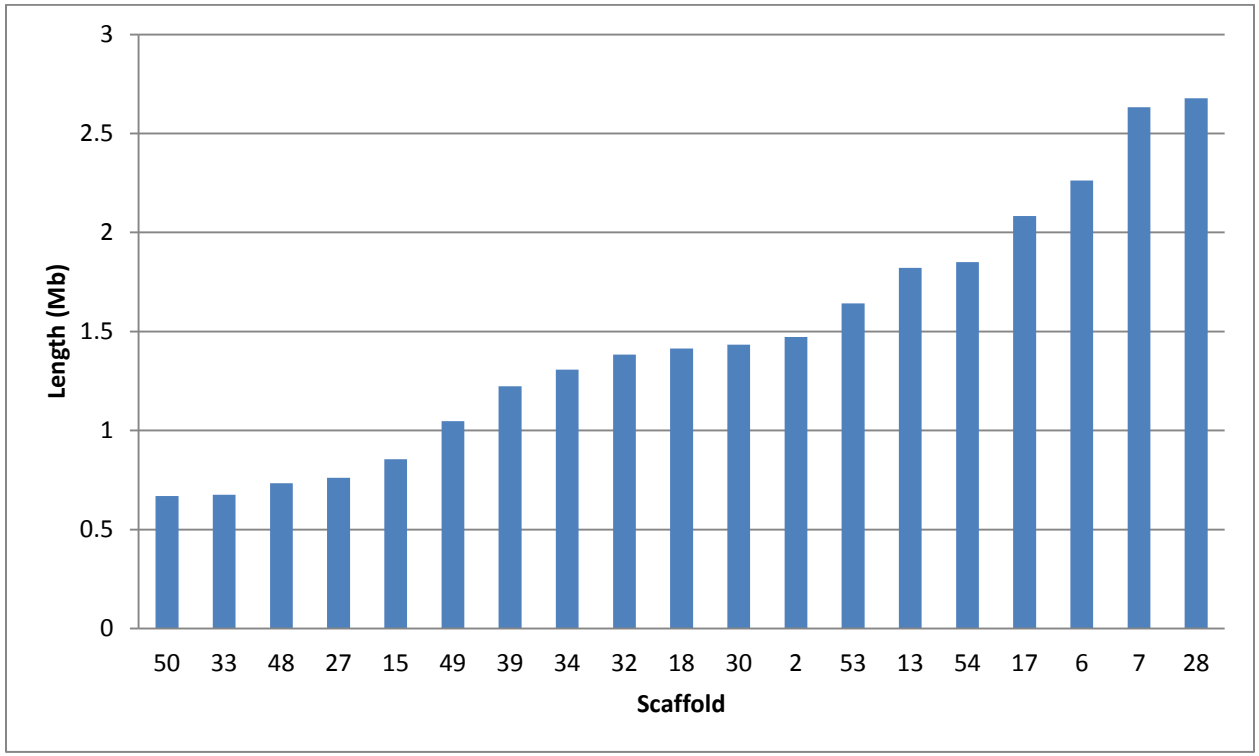

Figure 5 Length of scaffolds used in the genome alignment between *P. tritici-repentis* and *P. semeniperda*. All scaffolds used in the alignment were above 0.5 Mb.

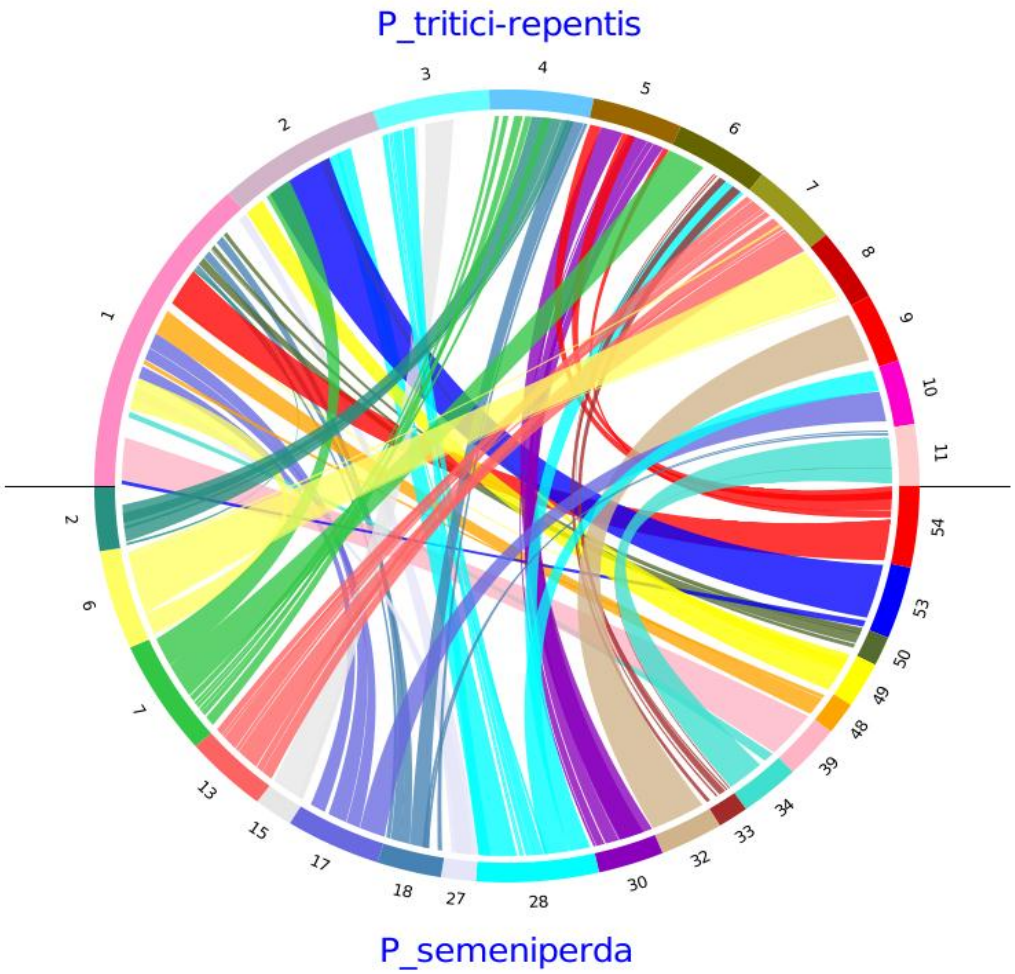

Figure 6 Circular view of the genome alignment. Homology is shown between 19 scaffolds of *P. semeniperda* (bottom half of circle) to the 11 chromosomes of *P. tritici-repentis* (top half of circle).

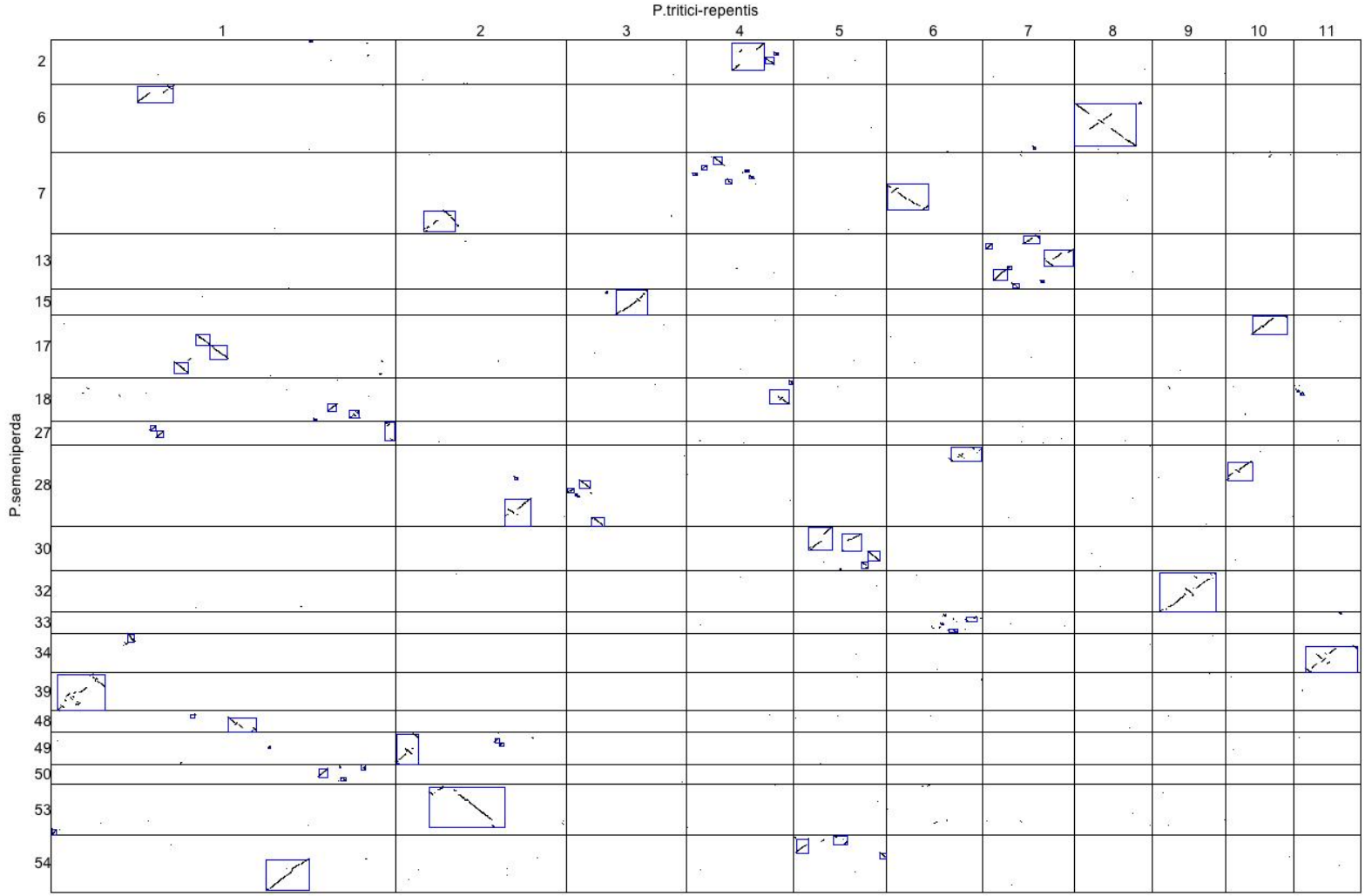

Figure 7 SyMap, dot plot of genome alignment between *P. semeniperda* (Y-axis) and *P. tritici-repentis* (X-axis).

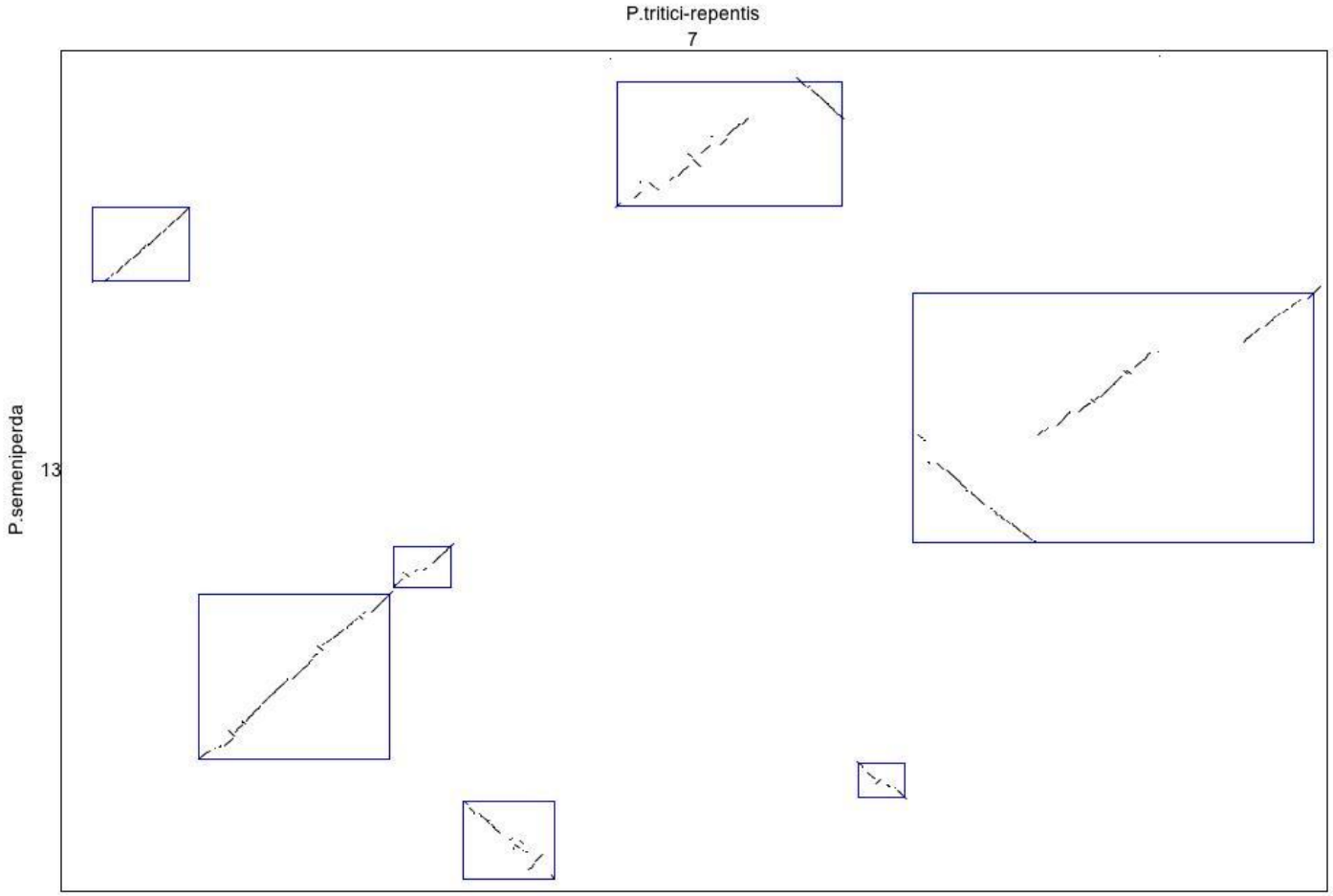

Figure 8 SyMap, dot plot (95% identity) of alignment between scaffold 13 (*P. semeniperda*) and chromosome 7 (*P. tritici-repentis*) illustrating intra-chromosomal rearrangements.

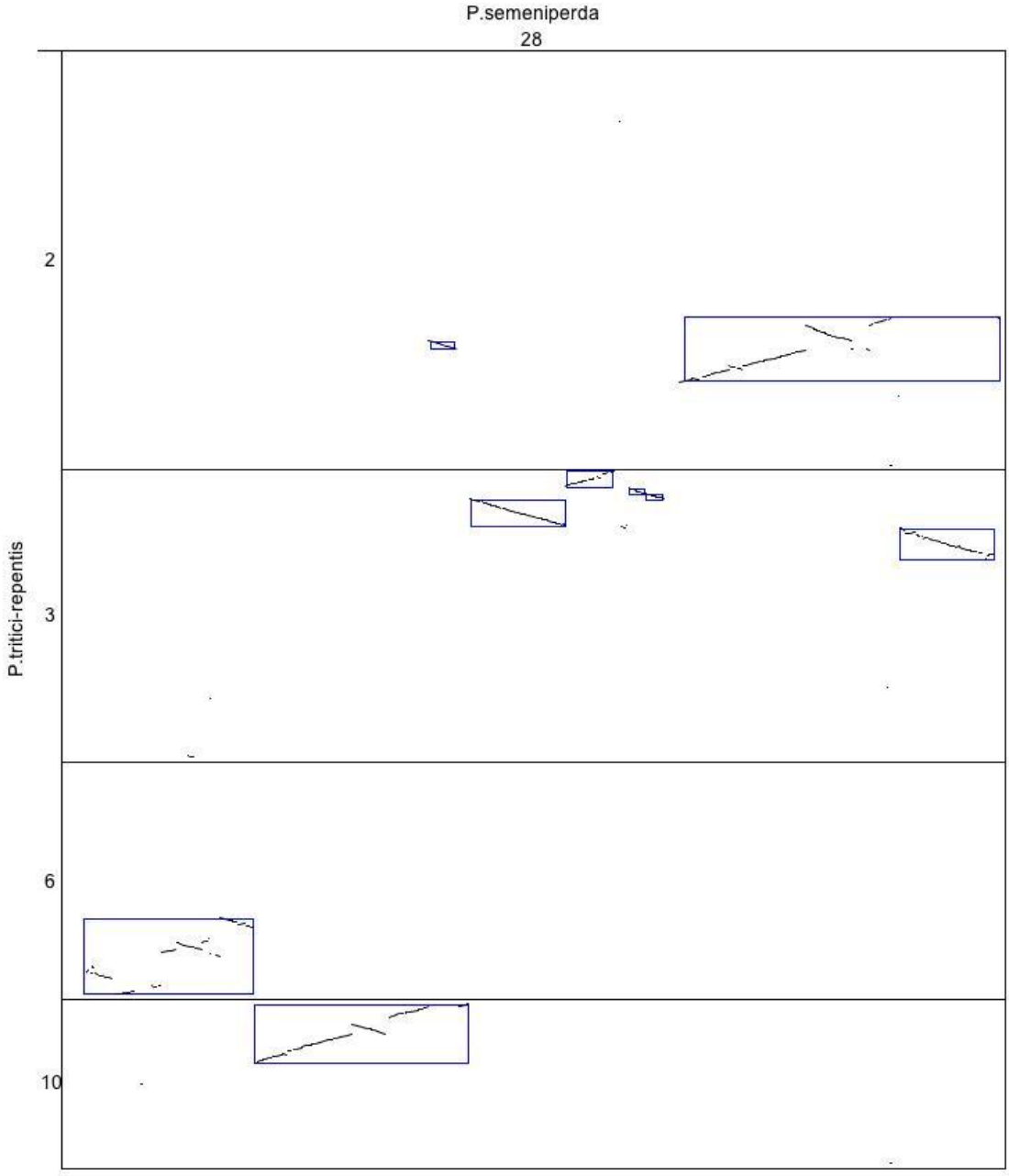

Figure 9 SyMap, dot plot (95% identity) of alignment between scaffold 28 from *P. semeniperda* (X-axis) and chromosomes 2, 3, 6 and 10 of *P. tritici-repentis* (Y-axis).

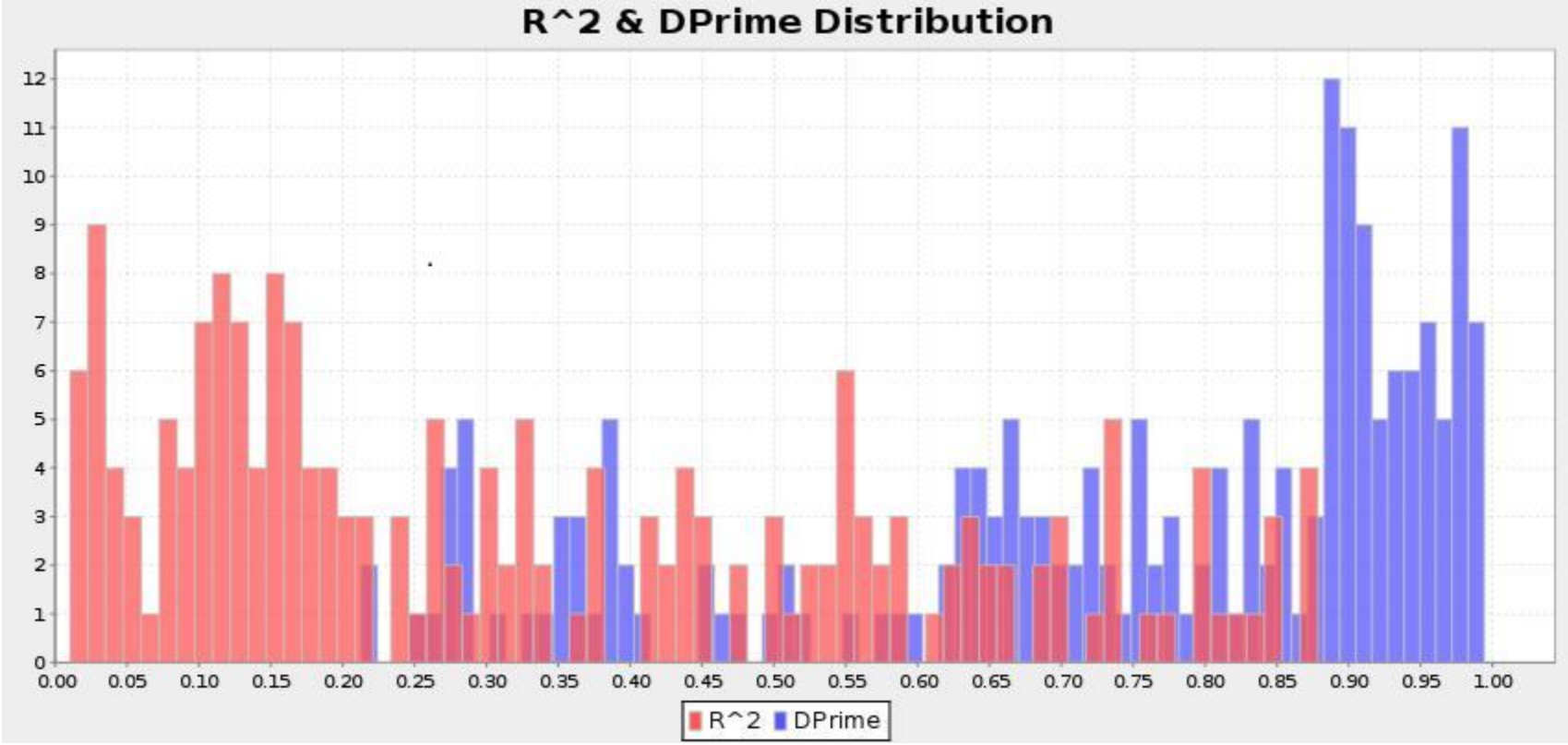

Figure 10 Distribution of LD estimator values (r^2 and D‘) per sample across all populations.