Evolution and acquisition of necrotrophic effectors by the fungal wheat pathogen Phaeosphaeria nodorum and close relatives

Author(s):
McDonald, Megan C.

Publication Date:
2012

Permalink:
https://doi.org/10.3929/ethz-a-007558763

Rights / License:
In Copyright - Non-Commercial Use Permitted
Evolution and acquisition of necrotrophic effectors by the fungal wheat pathogen *Phaeosphaeria nodorum* and close relatives

A dissertation submitted to the

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

MEGAN CAMILLA MCDONALD

BSc, University of Arizona
Born August 17th, 1985
Citizen of The United States of America

Accepted on recommendation of

Prof. Bruce A. McDonald, examiner
Prof. Beat Keller, co-examiner

2012
This work is dedicated to my family: Mom, Dad, Morgie, Mady, Grandma Fran, Grandma Margie and Granddad
Thanks for your unfaltering love and support.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SUMMARY</strong></td>
</tr>
<tr>
<td><strong>ZUSAMMENFASSUNG</strong></td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong></td>
</tr>
<tr>
<td>General Introduction</td>
</tr>
<tr>
<td><strong>CHAPTER 2</strong></td>
</tr>
<tr>
<td>Phylogenetic and population genetic analyses of <em>Phaeosphaeria nodorum</em> and its close relatives indicate cryptic species and an origin in the Fertile Crescent</td>
</tr>
<tr>
<td><strong>CHAPTER 3</strong></td>
</tr>
<tr>
<td>Horizontal acquisition of three necrotrophic effectors in the fungal wheat pathogen <em>Phaeosphaeria nodorum</em></td>
</tr>
<tr>
<td><strong>CHAPTER 4</strong></td>
</tr>
<tr>
<td><em>SnTox3</em> Acts in Effector Triggered Susceptibility to Induce Disease on Wheat Carrying the <em>Snn3</em> Gene</td>
</tr>
<tr>
<td><strong>CHAPTER 5</strong></td>
</tr>
<tr>
<td>The Cysteine Rich Necrotrophic Effector <em>SnTox1</em> Produced by <em>Stagonospora nodorum</em> Triggers Susceptibility of Wheat Lines Harboring <em>Snn1</em></td>
</tr>
<tr>
<td><strong>CHAPTER 6</strong></td>
</tr>
<tr>
<td>General Conclusions and Outlook</td>
</tr>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
</tr>
<tr>
<td><strong>PUBLICATIONS</strong></td>
</tr>
<tr>
<td><strong>CURRICULUM VITAE</strong></td>
</tr>
<tr>
<td><strong>APPENDICES</strong></td>
</tr>
</tbody>
</table>
SUMMARY

This thesis examines the evolutionary history of the fungal wheat pathogen *Phaeosphaeria nodorum* and three of its necrotrophic effector genes (NEs) *SnToxA*, *SnTox3* and *SnTox1*.

Over 355 *Phaeosphaeria* isolates from a global collection were sequenced at four neutral loci. Inferred phylogenetic and coalescent analyses clustered the samples in nine distinct sister species. Two species, *P. nodorum* and *P. avenaria* f. sp. *tritici* 1 (*Pat1*), comprise about 85% of the sampled isolates. Evidence of hybridization was found between *P. nodorum* and *Pat1* in 4% of the isolates examined. Measures of private allelic richness at both simple sequence repeats (SSRs) and sequenced loci reveal higher levels of diversity in Iranian *P. nodorum* populations than in other parts of the world. Additionally four of the nine sister species were found exclusively in Iran. Higher genotypic diversity and species richness strengthen the existing hypothesis that *P. nodorum* originated from the ancient Fertile Crescent.

Over 1000 isolates were screened for the presence/absence of each NE. Only two of the nine identified species, *P. nodorum* and *Pat1*, were found to carry NEs. In *P. nodorum*, contingency chi-square tests showed that NE frequency was dependent upon the population from which it was sampled. The observed number of multi-locus genotypes did not deviate from expectations under random mating. NE sequence diversity was assessed by sequencing about 200 individuals from a global collection. The population harboring the highest sequence diversity was different for each effector locus, and never coincided with the highest diversity found in neutral markers. The coalescent tree built from neutral markers revealed that NEs in *P. nodorum* were likely acquired horizontally. Shared toxin sequence alleles between *P. nodorum* and *Pat1* suggest that *Pat1* acquired NEs from *P. nodorum* via interspecific hybridization.

*SnTox3* and *SnTox1* were cloned and experimentally demonstrated to act as NEs. Both proteins induce a necrotic response *in planta* similar to the natural defense response known as the hypersensitive response (HR). The NEs of *P. nodorum* interact in a gene-for-gene manner with un-cloned wheat susceptibility genes. Preliminary evidence suggests that classic plant resistance NB-LRR genes encode the host susceptibility genes.
The *Phaeosphaeria* species complex provides a unique opportunity to compare the panmictic pathogenic species’ *P. nodorum* and *Pat1* with sister species found within a restricted spatial distribution. This study highlights the impact that horizontal gene transfer may have on pathogen emergence processes in the agro-ecosystem. Horizontal acquisition of NEs by *P. nodorum* enabled this fungus to become a global pest. Moreover, horizontal acquisition of NEs from *P. nodorum* has led to the emergence of at least two wheat pests, *Pyrenophora tritici-repentis* and *Pat1*. 
Diese Arbeit untersucht die Evolutionsgeschichte des weizenpathogenen Pilzes Phaeosphaeria nodorum und drei seiner nekrotrophen Effektorgene (NEs) SnToxA, SnTox3 und SnTox1.


SnTox3 und SnTox1 wurden kloniert und es wurde experimentell bestätigt, dass sie als NEs agieren. Beide Proteine induzieren eine nekrotische Reaktion in planta ähnlich der

Der *Phaeosphaeria* Artenkomplex bietet eine einzigartige Möglichkeit, die panmiktischen Pathogenarten *P. nodorum* und *Pat 1* mit Schwesterarten aus einer begrenzten räumlichen Verteilung zu vergleichen. Diese Studie zeigt Auswirkungen auf, die horizontaler Gentransfer auf den Prozess der Pathogenentstehung in einem agronomischen Ökosystem haben könnte. Die horizontale Aufnahme von NEs von *P. nodorum* hat zur Entstehung von mindestens zwei Weizenschädlingen geführt, *Pyrenophora tritici repensis* und *Pat 1*. 
Chapter 1:

General Introduction
The Pathogen

*Phaeosphaeria nodorum* is a haploid fungal pathogen of wheat, within the class Dothideomycetes, order Pleosporales (Solomon et al., 2006; Zhang et al., 2009). The primary hosts are domesticated wheats (*Triticum aestivum* and *Triticum durum*) and Triticale (Solomon et al., 2006). The pathogen is known world-wide as the causal agent of Stagonospora nodorum blotch (SNB) or glume blotch. The fungal symptoms are necrotic lesions on leaves preceded by a yellow halo (Cunfer and Ueng, 1999). The yellow halo is indicative of the production of necrotrophic effectors, also known as host-selective toxins, that are secreted preceding fungal growth in order to kill host cells (Solomon et al., 2006). There are three necrotrophic effectors (NEs) that have been described and characterized, *SnTox1, SnTox3* and *SnToxA* (Friesen et al., 2008; Liu et al., 2009; 2012). These genes and their evolutionary history within *P. nodorum* are the primary focus of this dissertation.

The pathogen *P. nodorum* is globally disseminated and causes major yield losses, particularly, in Australia and parts of the United States (Cowger and Silva-Rojas, 2006; Solomon et al., 2006). Historically, the disease was often mistaken for *Zymoseptoria tritici*, formerly *Mycosphaerella graminicola* (anamorph: *Septoria tritici*) (Quaedvlieg et al., 2011), as both pathogens form brown necrotic lesions on the leaves. Both pathogens also have similar looking asexual and sexual fruiting structures. The asexual stage appears as black or brown pycnidia in the leaf surface. Sexual ascospores are produced in pseudothecia and are believed to be the main over-wintering structure (Halama et al., 1999). *P. nodorum*, unlike *Z. tritici*, is able to infect and form fruiting structures on the glume and is classified as a necrotroph, feeding exclusively from dead plant tissue.

Agricultural plant pathogens pose a serious threat to global food security (Strange and Scott, 2005), yet much remains unknown about the origin of these diseases. Furthermore, fungal pathogens in the agricultural ecosystem (agro-ecosystem) have been shown to respond extremely rapidly to changes in control strategies (i.e. through the rapid defeat of resistant cultivars or fungicides)(McDonald and Linde, 2002; Torriani et al., 2009). Evolutionary studies, that use large collections of isolates, spanning different geographic areas, are able to describe the
potential an organism has to overcome deployed control strategies (Stukenbrock and McDonald, 2009). This chapter provides a brief introduction to the methods and concepts used to infer the evolutionary potential of *P. nodorum*, as well as summarizes current knowledge of its evolutionary history.

### Population genetics and known evolutionary history of *P. nodorum*

Co-evolution between domesticated plants and their corresponding pathogens has occurred mainly in the last 10-15 thousand years of human history (Balter, 2007). While the genetic variation in cultivated crops has gradually narrowed over time, pathogen populations remain extremely genetically diverse (Stukenbrock and McDonald, 2008). Understanding the breadth of pest diversity, reproductive mode and potential for long-distance dispersal are all key to properly implementing and maintaining the efficacy of modern control strategies (McDonald and Linde, 2002). The field of population genetics provides ample tools to study the diversity of an organism. In combination with molecular markers, populations collected over broad spacial areas allow geneticists to observe the movement of genetic material both within and among populations (Stukenbrock and McDonald, 2008). This information can then be used to infer the evolutionary history of the organism in question, as well as identify the potential a pest has to evade control strategies or invade new ecological niches.

### Basics of population genetics

Population genetics is defined as the study of genetic variation within a locus or loci among individuals of a population or populations (Hedrick, 2011). Genetic markers are used in population studies to describe the “genetic structure” of populations. The genetic structure of an organism refers to the amount and distribution of genetic diversity within or between populations (McDonald and McDermott, 1993). High pathogen diversity is a serious concern, as it provides a broad range of different pathogen genotypes upon which natural selection may act (McDonald and Linde, 2002). A basic measure of diversity involves counting the number of different alleles (variants) at a given locus, termed allelic richness. Similarly, the number of private alleles within
a population is an indication of high genetic diversity. A secondary measure of genetic diversity is the effective population size ($N_e$). The larger $N_e$ the more genetically diverse the underlying population (McDonald and Linde, 2002). Larger populations contain more rare alleles or mutations than smaller populations.

There are five major evolutionary processes that can shape the genetic structure of populations; natural selection, recombination, mutation, genetic drift and gene flow (McDonald and Linde, 2002). These five processes are briefly defined below followed by a summary of prior population genetic studies of *P. nodorum*.

**Intuitively,** natural selection is the process that “selects” for the best suited combination of traits in a given environment. Individuals whose overall phenotype best fits their environment are expected to be the most successful, leaving the most offspring for the next generation.

**Recombination** is the process of shuffling genetic information into new untested combinations. Recombination is considered one of the main mechanisms of generating new combinations of genes (genotypes) upon which selection may act.

**Mutation** is the source of novel genetic changes in DNA sequence. The result of mutation is the creation of new alleles or variants at a given locus. The vast majority of mutations are deleterious, however beneficial mutations are expected to increase in frequency in populations if they are favored by selection.

**Genetic drift** is the stochastic change in allelic frequencies. This process is seen mostly in small populations and often leads to the fixation of alleles or genotypes within a population. One common scenario where drift is observed is in founder events. Founder events occur when a new population is started from a small number of individuals. In agriculture, this is seen when infected plant material is transported to novel environments, i.e. the bringing of wheat to colonies in South Africa and Australia.

**Gene-flow** is the movement of genes or genotypes between populations that would otherwise be genetically distinct from each other. Gene-flow can be uni- or bi-directional and is usually quantified by measuring the number of migrants per generation between populations. Populations that mainly donate genetic material are referred to as sources, whereas populations that mainly receive information are called sinks.
Global genetic structure of *P. nodorum*

*P. nodorum* has been the subject of several population genetic studies using two different sets of neutral markers; restriction fragment length polymorphisms (RFLPs) and simple sequence repeats (SSRs) (Keller et al., 1997; Sommerhalder et al., 2006; Stukenbrock et al., 2006). These studies describe the basic genetic structure and reproductive mode of *P. nodorum*. Stuckenbrock et al. (2006) measured the allelic richness in populations of *P. nodorum* on five different continents. The highest allelic richness was found in Europe, followed closely by China and North America. The estimated $N_e$ sizes were also found to be highest in these regions but not significantly different from each other. Smaller effective population sizes were found in isolated, wheat producing regions, such as South Africa and Australia (Stukenbrock et al., 2006).

Fungal pathogens that undergo regular cycles of recombination are considered more dangerous than strictly clonal pathogens, due to the ability to quickly re-arrange virulent alleles into new combinations (McDonald and Linde, 2002). *P. nodorum* is a heterothallic fungus that requires two opposite mating types to undergo sexual recombination (Bennett et al., 2003). In field populations that undergo regular cycles of sexual reproduction mating-type alleles are expected to occur in equal frequency (Milgroom, 1996). Population collections of *P. nodorum* showed that both mating-types occur in equal frequencies (Sommerhalder et al., 2006), consistent with a sexually recombining population. The contribution of ascospores to primary infection was measured in a two year mark-release-re-capture experiment (Sommerhalder et al., 2010). This study concluded that sexual recombinants of the released isolates contribute significantly to primary infection at the beginning of the growing season. Moreover an increase in the number of recombinant genotypes sampled within the same growing season indicated that *P. nodorum* may produce sexual ascospores during the growing season, undergoing more than one sexual cycle per year (Sommerhalder et al., 2010).

**Centers of origin**

One of the major challenges of evolutionary biologists is to understand and describe the origins of disease causing organisms. For agricultural pests this search usually starts at the center of origin (CofO) of the host crop. The CofO of a species is defined as the geographical place
from which an organism originates (Stukenbrock and McDonald, 2008). For crop plants the CofO was first investigated by Vavilov in the 1940’s and 50’s (Vavilov, 1992). Vavilov performed one of the first extensive samplings of crops and their wild relatives, identifying seven major centers of crop diversity. In a review, Smith (1969) carefully summarized many of the main ideas proposed by Vavilov to define CofO (Smith, 1969). Some of his most important points were: (1) centers of origin have higher levels of neutral genetic diversity in comparison to other regions; (2) centers of origin have higher numbers of wild relatives of the cultivated crop with less interspecific hybridization; (3) centers can be supported by archeological or other forms of historical data. Though over 50 years has passed since the first publication of this work, these criteria are still used for defining CofO today (Burger et al., 2008).

In his chapter on the origins of wheats Vavilov rejected an earlier proposed origin in Central Asia. Instead, he proposed the ancient Fertile Crescent as the CofO of wheat. Vavilov also observes that wheat species carrying 21 chromosomes (now known to be hexaploid) are highly susceptible to rusts and other fungal diseases. He also notes slightly higher disease resistance in tetraploid wheat species and highest levels of resistance in the diploid wheat species *Triticum monococcum* (Vavilov, 1992). The recognition of variation in crop susceptibility to disease was a major motivation for seeking out new sources of genetic variation that could improve crop productivity. CofO were considered important areas where this genetic variation could be found (Vavilov, 1992).

Domestication of plants has been shown to be accompanied by the domestication and specialization of their corresponding fungal pests (Couch et al., 2005; Stukenbrock et al., 2007; Lê Van et al., 2012). Unlike plants (Balter, 2007), fungal pathogens do not form easily identifiable fossils or remains that could be used to verify their presence at a particular time (Berbee and Taylor, 2010). This necessitates the use of genetic data from contemporary populations to infer the evolutionary history of the organism. As described above, genetic diversity and the detection of wild-species remain important criteria for identifying pathogen CofO. Modern evolutionary models have made it possible to estimate gene-flow using neutral markers, such as sequence loci or SSRs. High diversity, coupled with high migration rates (gene-flow) away from the region are strong evidences for positive identification of the CofO. This
method has been successfully applied to several filamentous plant pathogens including *Phytophthora infestans* on potato (Gomez-Alpizar et al., 2007), *Zymoseptoria tritici* (*Mycosphaerella graminicola*) on wheat (Stukenbrock et al., 2007), *Rhyncosporium secalis* on barley (Zaffarano et al., 2008), and *Magnaporthe oryzae* on rice (Couch et al., 2005).

**Hypothesized CofO of *P. nodorum***

Populations of *P. nodorum* were shown to have high levels of genetic diversity in North America, Europe and China. Migration rates (gene-flow) between continents was found to be relatively high. Consistent with high levels of gene flow, Stukenbrock et al. (2006) also found relatively little population subdivision. Europe, China and North America were all identified as source populations, donating high numbers of migrants to other continental regions. Colonial populations, such as South Africa and Australia, were identified as sinks. These results are consistent with the hypothesis of human mediated transport of the pathogen from Europe or Asia to other parts of the world (Stukenbrock et al., 2006). Based on these findings, the authors proposed that the pathogen CofO coincided with that of its host, in the ancient Fertile Crescent (Balter, 2007; Burger et al., 2008). A lack of populations from these regions, however, prevented the authors from further testing this hypothesis.

Collections made on wheat in Iran in 2005 and 2010 enabled further testing of the Middle East CofO hypothesis. Chapter two of this thesis uses new sequence markers with SSR loci to asses the level of genetic diversity within *P. nodorum* Iranian populations and compares it to the previously published values from other populations.

**Species recognition concepts and species diversity associated with *P. nodorum***

**Defining a species**

A prerequisite for any population genetic study is the unambiguous identification of distinct taxa. Fungal taxonomical classification was traditionally done based on differences in morphology, often grouping very distantly related organisms in the same species due to the lack of distinct characters (Taylor et al., 2000). However, as DNA sequencing became more
affordable and widely applied, many genetically distinct groups were identified within traditionally classified species (Taylor et al., 2000; Kohn, 2005; Giraud et al., 2008).

Genetic tools, in particular gene sequencing, led to the development of phylogenetic species recognition (PSR). PSR uses gene trees to identify monophyly, or patterns of ancestry and decent, in identifiable clusters of individuals (Nixon and Wheeler, 1990). This method was further refined to include trees constructed from multiple, independent loci. The limit of the species was then defined as the boundary between concordance and conflict of the combined multiple gene trees. Concordant trees share the same topology due to fixation of formerly polymorphic loci, driven by genetic isolation between individuals. (Taylor et al., 2000). This method, known as genealogical concordance phylogenetic species recognition (GCPSR), is currently the most widely used species recognition method for fungi (Giraud et al., 2008).

Traditional phylogenetic trees are often used to define species under the GCPSR concept, however the concept was originally developed with coalescent gene genealogies (Baum and Shaw, 1995). Coalescent gene trees are similar in appearance to phylogenetic trees, however branch lengths are determined by estimation of the time from when two individuals last shared a common ancestor (for gene trees this is the last shared nucleotide polymorphism) (Griffiths and Tavare, 1999), not the genetic distance between two individuals (i.e. the number of nucleotide differences). The most recent common ancestors (MRCAs) of a gene or species are represented in the tree at the nodes connecting two branches. In addition to constructing gene-trees, coalescent genealogy samplers are also capable of estimating population sizes ($N_e$), population expansion, gene flow and time to divergence (Kuhner, 2009). The ability to estimate these parameters makes coalescent models extremely powerful tools for elucidating the evolutionary histories of populations within a species.

*Species complex associated with P. nodorum*

*P. nodorum* was one of many fungal pathogens subject to mis-classification based on morphology. When originally described, the teleomorphic stage of *Stagonospora nodorum* was assigned to the genus *Leptosphaeria*, which was reclassified as *Phaeosphaeria* in the late 1960’s (Cunfer and Ueng, 1999). In agriculture, fungal plant pathogens often carry the name of the host
from which they were collected (i.e. *Magnaporthe oryzae* on rice). *Phaeosphaeria*-like species have also been described on several other graminaceous hosts, mainly barley and oats (Ueng and Chen, 1994; Ueng et al., 1995). The oat infecting species, *Phaeosphaeria avenaria* (*Stagonospora avenae*) was originally classified based on host preference and elongated spores in comparison to *P. nodorum* (Richardson and Noble, 1970). Similarly, *Phaeosphaeria avenaria* f. sp. *tritici* (*Pat*) was found on wheat or barley but had spore lengths more similar to *P. avenaria* than *P. nodorum* (Shaw, 1957; Richardson and Noble, 1970). This species complex was the subject of several single gene phylogeny studies (Malkus et al., 2005; Reszka et al., 2005; Arkadiusz et al., 2006).

In addition to investigating the CofO of *P. nodorum*, Chapter two elaborates on the relationships among *Phaeosphaeria* sp. using a multi-locus sequencing approach from four neutral loci and several hundred individuals.

**Non-vertical genetic exchanges between fungi**

One of the unique features of filamentous fungi is the ability to undergo regular hyphal fusion (anastomosis), that results in the formation of heterokaryons (cells with multiple genetically distinct nuclei). Anastomosis provides fungi occupying the same ecological niches with a unique opportunity to exchange genetic information outside of species boundaries. Horizontal exchanges of genetic material and hybridization have both resulted in the emergence of several novel fungal diseases (Olson and Stenlid, 2002; Mehrabi et al., 2011). The mechanisms attributed to hybridization and horizontal gene transfer (HGT) along with several examples of disease emergence are discussed below.

**Mechanisms associated with hybridization and HGT**

There are three main processes that allow regular heterokaryon formation; vegetative compatibility, conidial anastomosis tubes (CATs) and mating-type fusion. Vegetative compatibility allows the fusion of vegetative hyphae under independent genetic control from genes that control the sexual cycle (Glass et al., 2000). This process is believed to be important
for maintaining homeostasis and intra-hyphal communication within a fungal body (Glass et al., 2000). Fusion of hyphal tips can also occur between two genetically distinct individuals that encounter each other in the environment. Successful fusions between isolates of the same species are controlled by heterokaryon (het) incompatibility loci (Glass et al., 2000). Fusions between non-compatible individuals results in the rapid lysis of interacting cells (Glass and Kaneko, 2003).

Another form of asexual anastomosis observed in filamentous fungi are CATs. Much less is know about the genetic control of CATs, however both intra- and interspecific fusions between CATs has been observed (Gabriela Roca et al., 2005). CATs are thin, short tubes that connect germinating conidia and are distinct from germ tubes. CAT formation is not genetically connected to germ tube formation but has been linked to the genes that control vegetative compatibility (het genes)(Gabriela Roca et al., 2005).

Anastomosis is also required for the fusion of two individuals of opposite mating-types in heterothallic mating systems (Glass et al., 2000). Fusion between mating types often occurs between specialized fungal structures and in ascomycetes, like P. nodorum, is under the genetic control of mating-type loci (Kronstad and Staben, 1997; Bennett et al., 2003).Vegetative incompatibility between mating-types of different species is referred to as a pre-zygotic reproductive barrier. Interestingly these barriers appear to be stronger between species that occur in the same ecological niche. Species brought together from geographically separated regions were shown to have a higher proportion of isolates undergoing interspecific hyphal fusions (Giraud et al., 2008). Alternatively some distinct species are able to undergo anastomosis but are unable to successfully complete the sexual cycle (Kohn, 2005).

Pathogen emergence through hybridization

Inter-specific hybrids have been reported to be unstable and less fit than offspring generated from normal crosses (Schardl and Craven, 2003). Sterility is also commonly reported property of hybrid offspring (Schardl and Craven, 2003). Despite the instability, or low-fitness associated with interspecific hybrids, there are several examples of novel fungal plant diseases that arose through hybridization (Brasier, 2000). Newcombe et al. described the emergence of the
Melampsora medusae x Melampsora occidentalis hybrid, which infects a poplar clone generated from the cross of each fungi’s host species’ (Newcombe et al., 2000). The resulting hybrid was found to have an expanded host range, causing disease on the new poplar cross and on both original host trees (Newcombe et al., 2000). In the fungal-like oomycetes, hybridization between Phytophthora cambivora and P. fragariae was reported after an outbreak on alder trees (Brasier et al., 1999). P. cambivora is a common pathogen against hardwood trees, while P. fragariae attacks strawberry and raspberry. The new disease on alder represented the hybrid’s ability to invade a totally new ecological niche outside of either parent species’ original host range (Newcombe et al., 2000). Recent evidence from the model organisms Arabidopsis and Drosophila, showed that gene expression of parental genes was likely to be much higher or lower expressed in hybrid offspring (Landry et al., 2007). This was proposed as the genetic mechanism through which hybrids explore different fitness landscapes, enabling the persistence of hybrid species in new ecological niches (Giraud et al., 2008).

Pathogen emergence through horizontal gene (chromosome) transfer

Horizontal gene transfer (HGT) refers to the stable integration of genetic material after transfer between two individuals (Mehrabi et al., 2011). This process is not necessarily mutually exclusive from hybridization and the mechanisms named as the main drivers of HGT are the same as those described for hybridization (Mehrabi et al., 2011). HGT has resulted in the emergence of several novel plant diseases. Friesen et al. (2006) described the HGT of SnToxA from P. nodorum to Pyrenophora tritici-repentis, resulting in the emergence of tan-spot on wheat. The disease, first reported in 1941, led to the discovery of the necrotrophic effector (known at that time as a host-selective toxin) ToxA in Pyrenophora tritici-repentis. The gene was later found in the genome sequence of P. nodorum and subsequent population genetic analysis confirmed that the direction of transfer was from P. nodorum into P. tritici-repentis (Friesen et al., 2006). Another disease emergence linked with HGT was discovered after the outbreak of Southern corn leaf blight in 1970, caused by Cochliobolus heterostrophus. The disease, which only affected male-sterile cytoplasm corn varieties, was determined to be caused by a 1.2Mb piece of DNA that contained genes necessary for production of a long-chain polyketide toxin.
Genes encoded in the 1.2 Mb piece, were not found in other *Cochliobolus* spp. and codon usage was found to be different from other closely related species. Though not as clearly shown as the example from *P. nodorum*, HGT was deemed the most-likely source of the toxin producing gene cluster (Yang et al., 1996). These and several other examples show the potential risk associated with novel pathogen emergence associated with HGT.

Horizontal chromosome transfer (HCT), as the name implies, is the stable transfer of a chromosome or chromosomes between individuals. Horizontally transferred chromosomes are often found to be un-necessary for normal growth and reproduction (Mehrabi et al., 2011). The most well described example of this phenomenon is the lineage specific (LS) chromosomes of *Fusarium* spp. Co-incubation of pathogenic strains of *F. oxysporum* f.sp. *lycopersici* with non-pathogenic strains of *F. oxysporum* resulted in the horizontal transfer of an entire LS chromosome. The chromosome contained genes known to be associated with virulence against tomato (Stergiopoulos and de Wit, 2009). Non-virulent isolates that gained the LS chromosome during co-incubation, were shown to be pathogenic against tomato after acquisition (Ma et al., 2010).

Chapter 2 discusses the potential role of hybridization between sister species of *Phaeosphaeria* that resulted in the horizontal transfer of NEs. Comparison of the genetic diversity of the NE loci with neutral loci diversity in *P. nodorum* in Chapter 3, provides further evidence to support this hypothesis.

**Fungal Effectors**

*General properties*

Fungal effectors have been recognized since the early 1940’s with the discovery of gene-for-gene interactions in the flax-rust pathosystem (Agrios, 2005). Currently, effectors are defined as any molecule or protein that modulates the host physiological state, either to induce or suppress natural immune responses. These proteins contribute to pathogen virulence by altering the immune response of the host in a way that benefits pathogen growth or survival. The pioneering work for fungal effectors was performed with the tomato pathogen *Cladosporium*
fulvum. Using isogenic lines of tomato, plant pathologists identified a small cysteine rich peptide associated with induction of HR-like necrosis on compatible tomato lines (van Kan et al., 1991). Similar biochemical approaches were used to identify effectors in other fungal plant systems, including Flor’s flax rust system, the rice blast pathogen Magnaporthe oryzae and many others (Koeck et al., 2011). While the classic approach to identify effector candidates was highly successful, this process is tedious and requires extensive phenotyping on host plants. A recent breakthrough in effector identification was due to advances in fungal genome sequencing.

The standout example of the use of genomics to identify candidate effectors comes from the Irish potato famine pathogen, Phytophthora infestans. This fungus-like organism has a very large genome (~ 240 mega bases), 75% of which is repetitive DNA (Haas et al., 2009). Despite the challenges associated with this complex genome organization, a comparative genomic study between several closely related species identified the conserved RxLR translocation motif (Bhattacharjee et al., 2006). Further comparative genomic analyses unveiled over 200 RxLR proteins in each of four closely related Phytophthora species (Raffaele et al., 2010). Wet lab validation of candidate effector proteins showed induction of HR in compatible interactions. Several genome and secretome projects in true fungi have identified fungal effectors, however unlike in oomycetes, no conserved motifs or translocation signals were observed (Oliva et al., 2010).

**Effectors and the boom and bust cycle in agriculture**

Plant genomes are enriched in resistance genes, which encode proteins that recognize a wide variety of parasites (Jones and Dangl, 2006). In agriculture these genes are bred into new cultivars to confer immunity, known as major-gene resistance, to specific pathogens. These resistant cultivars are distributed as commercial seed, resulting in a “boom” of resistance gene use. Unfortunately, major-gene resistance has been found to be very unstable in the field, often becoming ineffective only a few years after widespread use (McDonald and Linde, 2002). The rapid loss in R-gene efficacy and subsequent disuse of the gene is referred to as the “bust” phase of the cycle. The majority of plant R-genes share a conserved structure, composed of a nucleotide-binding site domain and a highly variable recognition domain, composed of leucine
rich repeats (abbreviated NBS-LRRs). This class of genes has been shown to be responsible for induction of rapid and localized cell death, known as the hypersensitive response (HR). Typical morphological and biochemical characteristics of HR induced cell-death are cell-shrinkage, condensation of chromatin, nuclear breakdown, DNA laddering and membrane blebbing (Wolpert et al., 2002). For biotrophic pathogens, induction of HR in resistant cultivars results in localized cell death near the pathogen point of entry, which prevents pathogen establishment in the host (Ellis et al., 2000). Recognition by NBS-LRRs has been shown to be highly specific and often only recognizing a single allelic variant of the target pathogen ligand (Dodds et al., 2006).

**Necrotrophic effectors and resistance/susceptibility genes**

Necrotrophic effectors (NEs) encompass a wide variety of fungal metabolites ranging from linear polyketols, cyclic tetra-peptides to small molecular weight proteins (Wolpert et al., 2002). Initially termed host selective/specific toxins, NEs are typically active against a single plant host. Absence of the gene (or genes) responsible for NE production or absence of the host sensitivity gene, results in non-compatible host parasite interactions and no disease development (Wolpert et al., 2002). This interaction is known as the inverse gene-for-gene model (Stukenbrock and McDonald, 2009). The first described NE was discovered after the emergence of Victoria blight of oats, caused by *Cochliobolus victoriae*. The disease was found on oat varieties carrying the Victoria-type major resistance gene, *Pc2*, which conferred resistance to the crown rust pathogen *Puccinia coronata* (Wolpert et al., 2002). Pathogenicity on Victoria oats was determined to be caused by production of a toxin by *C. victoriae*. The toxin was subsequently named victorin. Attempts to separate the R-gene locus and sensitivity locus targeted by victorin were unsuccessful, leading to the hypothesis that the two genes were either tightly linked or the same gene (Wolpert et al., 2002). It was noted, that necrosis induced by victorin highly resembled that of resistance gene induced HR (Wolpert et al., 2002). This led to the hypothesis that NEs targeted R-genes in plants to induce defense associated cell death. Much later, molecular cloning of the gene conferring victorin sensitivity in *Arabidopsis thaliana* revealed the *LOV1* gene, a coiled-coil NBS-LRR gene; a classic R-gene (Lorang et al., 2007).
Chapters four and five of this thesis characterize the interaction of \textit{SnTox3} and \textit{SnTox1} with sensitive wheat varieties. Evidence that NE function by inducing HR is presented. This work further supports the hypothesis that NEs manipulate host defense responses in order to kill host cells from which the fungus derives its energy.

\textbf{The present work}

This thesis is divided into two first author chapters (chapters two and three), followed by two published, collaborative works (chapters four and five). Chapter two focuses on the population genetics and evolutionary biology of neutral loci in \textit{Phaeosphaeria nodorum} and several related species. Chapter three builds upon chapter two, by comparing the evolutionary history at neutral loci with that of multiple unlinked NE loci. Finally, chapters three and four describe the molecular cloning and characterization of two NE loci \textit{SnTox3} and \textit{SnTox1}.
REFERENCES


Chapter 2:
Phylogenetic and population genetic analyses of *Phaeosphaeria nodorum* and its close relatives indicate cryptic species and an origin in the Fertile Crescent

Megan C. McDonald, Mohammad Razavi, Timothy L. Friesen, Patrick C. Brunner, Bruce A. McDonald

Under review in Fungal Genetics and Biology: re-submitted April 27th, 2012
ABSTRACT

The origin of the fungal wheat pathogen Phaeosphaeria nodorum remains unclear despite earlier intensive global population genetic and phylogeographical studies. We sequenced 1,683 bp distributed across three loci in 355 globally distributed Phaeosphaeria isolates, including 74 collected in Iran near the center of origin of wheat. We identified nine phylogenetically distinct clades, including two previously unknown species tentatively named P1 and P2 collected in Iran. Coalescent analysis indicates that P1 and P2 are sister species of P. nodorum and the other Phaeosphaeria species identified in our analysis. Two species, P. nodorum and P. avenaria f. sp. tritici 1 (Pat1), comprised ~85% of the sampled isolates, making them the dominant wheat-infesting pathogens within the species complex. We designed a PCR-RFLP assay to distinguish P. nodorum from Pat1. Approximately 4% of P. nodorum and Pat1 isolates showed evidence of hybridization. Measures of private allelic richness at both SSR and sequence loci suggest that the center of origin of P. nodorum coincides with its host in the Fertile Crescent. We hypothesize that the origin of this species complex is also in the Fertile Crescent, with four species out of nine found exclusively in the Iranian collections.
INTRODUCTION

It is hypothesized that domestication of plants was accompanied by the domestication and specialization of their corresponding fungal pathogens in a pathogen emergence process known as host-tracking (Stukenbrock and McDonald, 2009). Signatures of host-tracking include elevated levels of genetic diversity and the presence of closely related ancestral pathogen species on undomesticated hosts at the domesticated host’s center of origin. Specialization on the tracked host is believed to come at the cost reduced ability to successfully colonize a broader range of plant species (Barrett, et al., 2009). This specialization may lead to the emergence of new host-specialized “domesticated” pathogen species from the ancestral “wild” source populations. Two compelling examples of this process have been identified in agro-ecosystems. Stukenbrock et al. (2007) found evidence to support host tracking in Zymoseptoria tritici (formerly Mycosphaerella graminicola, Quaedvlieg et al., 2011), a fungal pathogen of wheat. In Iran, near the ancient Fertile Crescent where wheat originated (Balter, 2007), Stukenbrock et al. (2007) found elevated levels of genetic diversity and two newly described grass-infecting species that were shown to be the closest living relatives of the globally distributed wheat pathogen. Phylogeographic studies on the rice blast pathogen Magnaporthe oryzae demonstrated a single shift of the modern pathogen onto domesticated rice (Couch et al., 2005). Both Couch et al. (2005) and Stukenbrock et al. (2011) demonstrated elevated levels of pathogenicity on the tracked host and reduced virulence on closely related hosts, indicative of host-specialization.

While host-tracking in the agro-ecosystem appears to lead to host-specialized pathogens, little is known about the ecological role of the “wild” species before their emergence as global agricultural pests. A recent model proposed by Schulz and Boyle (2005), suggests that endophytes act antagonistically towards their host to facilitate colonization. This antagonism may be balanced by an equally strong defense response from the host. Pathogens could emerge from endophytes when this balance is disrupted and the immune responses of the plant are no longer sufficient to contain the endophyte (Schulz and Boyle, 2005). Experimental endophyte studies have been limited mainly to the Clavicipitaceae family. However recent reviews of endophyte diversity using both cultivable and molecular methods suggest that endophyte
diversity is grossly under-estimated (Arnold, 2007, Saunders and Kohn, 2010). Studies of agricultural endophyte communities suggest that host genotypes can have a strong effect on the structure of the endophyte community within a plant (Saunders and Kohn, 2009). Viewed within the context of the host-tracking hypothesis, it is plausible that alteration of the host environment through domestication could result in a disruption of the “balanced antagonism” and facilitate the emergence of a pathogen from an endophyte (Schulz and Boyle, 2005).

Host specialization in plant pathogenic fungi has often been used to define species groups, but this criterion can mistakenly separate fungi that are still exchanging genes (Cai et al., 2011). The most common current concept to delimit fungal species is based on Genealogical Concordance Phylogenetic Species Recognition (GCPSR) (Taylor et al., 2000). Briefly, this concept establishes species boundaries by examining concordance between multi-gene phylogenies. This concept has proven particularly useful in identifying groups of cryptic fungal species (Giraud et al. 2008). Divergence at multiple independent loci requires the interruption or gradual reduction of gene flow between two groups or populations. In allopatry, the reduction of gene flow occurs because of a physical or external barrier prohibiting genetic exchange among individuals (i.e. allopatric hosts or barriers to dispersal) (Giraud et al., 2008). In sympatry, the probability of mating should depend only on the genotype of the individuals (Kondrashov, 1986). Under this definition two species specializing on different grass hosts in the same area would be in sympatry. Specialization on one or the other host, a product of the host genotype, would reduce the chances of mating, and therefore result in sympatric speciation (Giraud et al., 2008).

*Phaeosphaeria nodorum* (E. Müller) (anamorph: *Stagonospora nodorum* (Berk.) Castellani and E.G. Germano) is a globally distributed wheat pathogen that causes significant damage in Australia and parts of North America (Solomon et al., 2006). Previous population genetic analyses, using both RFLP and microsatellite markers in nearly 1000 global isolates, concluded that *P. nodorum* populations exhibit high levels of gene flow and high effective population sizes in Europe, North America and China (Keller et al., 1997; McDonald et al., 1999; Stukenbrock et al., 2006). Attempts to identify the center of origin of *P. nodorum* based on these populations were inconclusive, which was attributed to a lack of population samples from the Fertile Crescent (Stukenbrock et al., 2006). A complex of *Phaeosphaeria* cereal pathogens has
appeared sporadically in the taxonomic literature since the 1950s. These morphology-based studies classified \textit{P. nodorum} with different fungal relatives and resulted in several name changes; earlier names included \textit{Septoria nodorum} and \textit{Leptosphaeria nodorum}. Reviews by Cunfer et al. (1999) and Solomon et al. (2006) provide a more complete history of \textit{P. nodorum} naming conventions. More recent phylogenetic studies placed \textit{P. nodorum} within the order Pleosporales, sub-clade Phaeosphaeriaceae (Zhang et al., 2009).

Three main \textit{Phaeosphaeria}-like species infecting cereals were identified in the 1950s based on spore morphology, spore production, formation of sexual structures and host specialization (Shaw, 1957a; b). Isolates that were most pathogenic on wheat and showed heterothallic mating type behavior formed the group now called \textit{P. nodorum}. Isolates collected from oats or other hosts were initially named \textit{Leptosphaeria avenaria}, with the species name reflecting the host preference. These isolates also exhibited heterothallic mating behavior (Shaw, 1957b). Subsequent molecular studies confirmed that this group also fell into the Phaeosphaeriaceae clade and it was renamed \textit{Phaeosphaeria avenaria} (anamorph: \textit{Stagonospora avenae}). A third group of isolates was non-pathogenic on oats, weakly pathogenic on wheat and other cereals and homothallic. These isolates were morphologically similar to \textit{P. avenaria} but pathogenic on wheat and were therefore named \textit{Phaeosphaeria avenaria} f. sp. \textit{tritici} (\textit{Pat}) (Shaw, 1957a; Ueng et al., 1995).

Field observations of \textit{P. nodorum}-like symptoms on barley and other \textit{Hordeum} sp. led to studies that divided \textit{P. nodorum} into wheat or barley-infecting biotypes (Martin and Cooke, 1979). These studies confirmed field observations of stable host specialized biotypes (Osbourn et al., 1986). Restriction fragment length polymorphisms (RFLPs) (Ueng and Chen, 1994) and ITS sequences (Ueng et al., 1998) confirmed genetic differences between the two biotypes and led to the proposal to split \textit{Pat} into three groups, \textit{Pat1}, \textit{Pat2} and \textit{Pat3}. Since 1998, the species complex was analyzed using six additional single gene phylogenies including mating-type loci (Bennett et al., 2003; Ueng et al., 2003), \(\beta\)-tubulin (\textit{tubA}) (Malkus et al., 2005), \(\beta\)-glucosidase (\textit{bgl1}) (Reszka et al., 2005), RNA polymerase II (Arkadiusz et al., 2006) and histidine synthase (\textit{his}) (Wang et al., 2007). Until now, the phylogenetic relationships among these \textit{Phaeosphaeria} sp. were
assessed using single loci and a small number of isolates collected from different hosts. Results have not been consistent across all studies.

In this study we included over 300 *Phaeosphaeria* isolates collected from farmers’ fields on five continents over 25 years in addition to over 50 isolates collected from wild grasses on different continents. While many of these isolates were used previously in population genetic studies of *P. nodorum*, we included several new collections from Iran, located near wheat’s center of origin. We used a three-gene phylogeny to determine the relationships among 355 isolates within the *Phaeosphaeria* sp. complex, using both traditional phylogenetic and newer coalescent methods. In addition, sequences of mating type loci were analyzed for many isolates in each species. Finally we assessed allelic richness and other measures of genetic diversity within *P. nodorum* collections from Iran to test the hypothesis that this pathogen originated in the Fertile Crescent.

**MATERIALS AND METHODS**

**Fungal collections and DNA extraction**

Isolates used in this study are described in Table 1. Single oozing cirri from lesions were isolated with a flame-sterilized needle. Isolates were grown on Petri dishes containing Yeast Sucrose Agar (YSA, 10g/L Yeast Extract, 10g/L sucrose, 1.2% agar) amended with 50 µg of kanamycin. Single colonies were transferred to 50 mL Yeast Sucrose Broth (YSB, 10g/L Yeast Extract, 10g/L sucrose) and grown on a rotary shaker for 3 days at 120 rpm at 18°C. Fungal biomass was lyophilized, ground into a powder and total DNA was extracted using the DNeasy Plant Mini DNA extraction kit (Qiagen GmbH, Hilden, Germany) following the manufacturer’s instructions. Extracted DNA was diluted 1:500 with sterile water for PCR amplification.

**PCR amplification, restriction enzyme digestion, sequencing and alignment**

PCR amplification was performed in 20 µl reactions containing 0.05 µM of each primer (Microsynth, Balgach Switzerland), 1 × Dream Taq Buffer (MBI Fermentas), 0.4 µM dNTPs (MBI Fermentas) and 0.5 units of Dream Taq DNA polymerase (MBI Fermentas). The PCR
cycle parameters were: 2 min initial denaturation at 96°C followed by 35 cycles of 96°C for 30 s, anneal for 45 s, and extend at 72°C for 1 min. A final 7 min extension was made at 72°C. A portion of the β-tubulin gene (annotated as SNOG06791.2 in the *P. nodorum* genome sequence available at http://genome.jgi-psf.org/Stano2/Stano2.home.html), the internal transcribed spacer (ITS) of the ribosomal gene cluster, and a portion of a β-xylosidase gene (SNOG06123.2) were amplified and sequenced in all isolates (Hane et al., 2007). The mating type idiomorphs *MAT1-1* (SNOG00375.2) and *MAT1-2* were sequenced for a subset of isolates. The conserved mating type (*MAT*) primers for *P. nodorum* (Bennett et al., 2003) were used to amplify and sequence the *MAT* loci from all *Phaeosphaeria* sp. Sample size permitting, we assessed the ratio of *MAT1-1* to *MAT1-2* alleles. Primer sequences and annealing temperatures specific for each pair are listed in Appendix A.1.

Sequencing reactions were conducted in 10 µl volume using the BigDye® Terminator v3.1 Sequencing Standard Kit (Life Technologies, Applied Biosystems) with both the forward and reverse primers. The cycling parameters were 96°C for 2 min followed by 55 or 99 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. PCR products were cleaned with the illustra™ Sephadex™ G-50 fine DNA Grade column (GE Healthcare) according to the manufacturer’s recommendations and sequenced with a 3730xl Genetic Analyzer (Life Technologies, Applied Biosystems). Alignment of forward and reverse sequences was performed in SeqScape software V2.5 (Life Technologies, Applied Biosystems). Quality screening and ambiguous base calls were edited by hand using SeqScape. Final alignments were exported and re-aligned using ClustalW, implemented online using the Max Planck Institute of Bioinformatics Toolkit (http://toolkit.tuebingen.mpg.de). Sequences of each isolate were concatenated using the utility function implemented in SNAP Workbench (Price and Carbone, 2005). Each locus was tested for neutrality using the standard measurements implemented in DnaSP v5 (Rozas and Rozas, 1995; Librado and Rozas, 2009).

**Phylogenetic relationships and coalescence dating**

Bayesian and maximum-likelihood methods were used to determine the phylogenetic relationship among isolates. Orthologous sequences from *Pyrenophora tritici-repentis* were used
to root the phylogenies. The program jModelTest v0.1 (Guindon and Gascuel, 2003; Posada, 2008) was used to select the most appropriate nucleotide substitution model for each locus. Top models were chosen based on the corrected Akaike Information Criterion (AICc). For β-tubulin the top AICc model was TrN+I+G. For ITS the top model was GTR+G and for β-xylosidase the top model was TIM2+G. The top model of the concatenated dataset was TIM2+I+G. The most commonly implemented model which was also ranked by AICc within the top 12 nucleotide models for all three loci was GTR+G.

The program RaxML v7.2.1 (Stamatakis, 2006; Stamatakis et al., 2008) was used to calculate the tree with the highest likelihood and MrBayes v4.1.2 (Ronquist and Huelsenbeck, 2003; Altekar et al., 2004) was used for Bayesian posterior sampling. To reduce computation time, the data set was condensed into distinct sequence haplotypes, excluding indels and infinite site violations, using the program MAP implemented in SNAP workbench (Aylor et al., 2006). The MAT idiomorphs were analyzed separately using RaxML v7.2.1.

For RaxML runs we used the “–f a” algorithm which performs a rapid bootstrap analysis while simultaneously searching for the best maximum likelihood tree. Bootstrap scores were plotted onto the highest scoring maximum likelihood tree with 10,000 bootstrap replicates. The author recommended nucleotide model GTR-CAT was selected over the GTR+G model due to computational advantages within RAxML (Stamatkis, 2006). For Bayesian analysis the nucleotide substitution model GTR was used with a gamma distribution to account for variation in mutation rates at different sites. The default 4 gamma categories were used. The chain length was set to 15 million generations and posterior samples were recorded every 1,000 generations. The first 25% of trees (3,750) were discarded as burn-in. To ensure convergence, replicate independent runs were conducted to ensure that both runs reached similar Ln-likelihood scores.

Coalescent and species tree construction were conducted using the Bayesian Markov Chain Monte Carlo (MCMC) methods implemented in the program BEAST v1.6.1 (Drummond and Rambaut, 2007). *BEAST was used for calculation of the species tree (Heled and Drummond, 2010). *BEAST allows uncoupling of the gene trees to estimate the species tree, thereby taking into account incongruence of gene trees among loci in the posterior sample of the estimated species tree. The data set was partitioned by locus in the nexus input file so that
parameters for each gene could be estimated separately. The entire nucleotide dataset was used for this analysis. Simulations were conducted with all loci under the HKY or GTR model. Due to computational time limits, rate variation was left out of the nucleotide model. All 355 *Phaeosphaeria* isolates were assigned to species groups a priori based on the results from the maximum-likelihood and Bayesian phylogenetic analyses. Runs were conducted using a strict molecular clock under the Yule’s speciation model specified in BEAUti. Posterior sampled trees were visualized after burn-in of 25% of the sampled trees. The posterior sampling was visualized in the form of consensus species trees using DensiTree (Bouckaert, 2010).

Relative time of divergence was also estimated in *BEAST*. For estimation of divergence times for the individual species, all individuals were assigned a priori to groups based on the results from the previous MrBayes and RaxML analyses. Additional taxonomic groups were added to estimate time to the most recent common ancestor (TMRCA) of groups of species at nodes deeper in the tree (i.e. TMRCA of the ancestor of *P. nodorum* and the other *Phaeosphaeria* sp.). Estimation of all parameters was conducted under two different clock models; strict and relaxed exponential. The complementary program BEAUti (Drummond and Rambaut, 2007) was used to prepare the XML file needed for input into *BEAST*. Runs were assessed for convergence using the Estimated Sample Size (ESS) values for each parameter calculated in the program Tracer v1.5 (Rambaut and Drummond, http://beast.bio.ed.ac.uk/Tracer). As suggested by the authors all runs were conducted until ESS values exceeded 100 for all estimated parameters. Means and 95% highest posterior density (HPD) intervals for all runs and combined runs were also calculated and plotted using Tracer.

**Allelic diversity in *P. nodorum* and *Pat1***

Stukenbrock et al. (2006) previously described the allelic diversity and migration patterns among global collections of *P. nodorum*. This dataset was restricted to the nine largest single field collections made in China, Australia, Texas, North Dakota, Oregon, New York, South Africa and Switzerland. We added our largest Iranian field population, collected in 2005, to this dataset to compare allelic richness at eight microsatellite (SSR) loci. In a second analysis we included another Iranian field population sampled from the same region in 2010 in order to
increase the sample size from Iran. Amplification of SSR loci was performed using labeled primers as described previously (Stukenbrock et al., 2005). Amplicons were separated in a 3730xl Genetic Analyzer (Life Technologies, Applied Biosystems). The software Genemapper (Life Technologies, Applied Biosystems) was used for genotyping.

The program ADZE v1.0 was used to visualize average allelic richness and private allelic richness across all 8 loci (Szpiech et al., 2008). The program plots both values as a function of sample size (g), which allows comparison across populations with unequal sample sizes. For P. nodorum and Pat1, sample sizes were adequate to measure sequence diversity at the regional level. The number of alleles per locus and pairwise nucleotide diversity were both calculated in R using the pegas library (Paradis, 2010).

The minimum number of recombination events (Rm) and number of pairs of sites with four gametes was also assessed using DnaSP v5 (Librado and Rozas, 2009). The program TCS v1.2 was used to visualize the multi-locus haplotype network of the Pat1 species (Clement et al., 2000). This program generates a haplotype network within which possible recombination events are allowed in the network construction. Pie charts showing the frequency of multi-locus haplotypes in each population were generated using the pegas package in R (Paradis, 2010).

We developed a PCR-restriction fragment length polymorphism (PCR-RFLP) assay using fixed species polymorphisms to distinguish between P. nodorum and Pat1. The 962 bp amplicon from β-xylosidase (SNOG06123.2) was used as the template DNA. NEB Cutter v2.0 was used to identify species-specific restriction enzyme recognition sites. PCR amplicons were digested with 2 units of the restriction enzyme ScaI (NE Biolabs,) at 37°C for ninety minutes, followed by 15 minutes of heat inactivation at 65°C. Digested products were separated in 1.8% agarose gels and visualized under UV light using ethidium bromide staining.

**Southern Blotting and Hybridization**

Genomic DNA was blotted onto nylon membranes using a Bio-Dot microfiltration apparatus (BIO-RAD) following the instructions in the user manual. Radioactive probes were synthesized using purified PCR product (purification same as above). Probe labeling with 32P was performed with 25 to 50 ng/µL of template DNA using the NEBlot Kit (NE Biolabs)
following the manufacturer’s recommendations. Unincorporated nucleotides were filtered out using illustra NICK™ columns (GE Healthcare). DNA hybridization, membrane washing and image acquisition were performed as described previously (McDonald and Martinez, 1990).

RESULTS

Phylogenetic analyses

The complete ITS region (502 bp) and partial nucleotide sequences of the β-tubulin (490 bp) and β-xylosidase genes (689 bp) were concatenated and used for phylogenetic analysis of 355 Phaeosphaeria isolates (Figure 1). This multi-locus dataset used the closest known pathogenic relatives of P. nodorum, Leptosphaeria maculans and Pyrenophora tritici-repentis, as outgroups. Based on the genes used in this study, P. tritici-repentis is more closely related to the Phaeosphaeria sp. complex than L. maculans. Therefore, all further analyses requiring an outgroup were performed with P. tritici-repentis. All three loci displayed neutral evolution. The dataset collapsed into 149 multi-locus haplotypes distributed among nine distinct phylogenetic groups (Figure 1). The complete dataset is available for download through GenBank: accession numbers JQ757169 - JQ758513.

P. nodorum formed the largest clade, found on wheat leaves, ears and seeds, barley, triticale, crested wheatgrass (Agropyron desertorum) and other wild grasses. Pat1 was the second largest clade, but found only on wheat leaves, ears and seeds. Six Pat3 isolates were isolated from triticale and wheat leaves in Denmark. In addition to these described species, several new clades were discovered. To be consistent with earlier publications the new clades were named using the P. avenaria f. sp. tritici (Pat) notation adopted by Malkus et al. (2005). All Pat4 isolates were isolated from Elymus tauri in Iran. Pat6 isolates were found on both E. tauri and Agropyron tauri in Iran. Pat5 was isolated from Elymus angustus and Bromus inermis in North Dakota between 1982 and 1984 (Krupinsky, 1982). These Pat5 isolates have been characterized as non-pathogenic on wheat (T. Friesen, unpublished data).

We found 29 P. nodorum isolates, 19 Pat1 isolates and two un-described species, provisionally named P1 (Phaeosphaeria 1, seven isolates) and P2 (Phaeosphaeria 2, two isolates) in Iranian wheat fields. Unrestricted BLAST searches in GenBank using the P1 and P2
sequences found significant similarity to both P. nodorum and P. avenaria for the β-tubulin and β-xylosidase loci. Unrestricted searches with the ITS region of P1 found >98% sequence identity to several fungi described as Leptosphaeria sp., Pleosporales sp. or unknown fungal sp. The GenBank accession numbers, location of sampling, sampling method, and type of study for hits with 100% query coverage and >98% identity are as follows, GB accessions: AB693774.1 (unknown), GU985210.1 (China, cultured, endophyte), GU985209.1 (China, cultured, endophyte), GU985203.1 (China, cultured, endophyte), HM537061.1 (China, unknown, endophyte), GU078648.1 (China, uncultured, rice paddy soil), FJ537121.1 (unknown), AJ608969.1 (Bulgaria, cultured, plastic), FN5481557.1 (Germany, cultured, endophyte), EF505560.1 (Ohio, uncultured, endophyte). Most of these studies described fungal leaf endophytes on a wide variety of hosts. The ITS BLAST search for P2 found 99% identity to two uncultured fungal clones found in air samples in Germany (GB accessions: GQ999266.1, FJ820759.1). Other matches with lower identity (96%) included the soil fungal community of grasslands in New Mexico (GB: EU479755.1, EU479754.1), a fungal community of ants in Texas (GB: HQ608028.1) and endophytes of cupressaceous trees (GB: EF420004.1). The ITS sequence similarity of P1 and P2 to P. nodorum averaged 89% for both groups. This is moderately higher than the ITS similarity between P. nodorum and P. tritici-repentis (85%). A complete list of the top 250 BLAST hits for both P1 and P2 is available in Appendix C and D.

Both tree construction methods separated all nine phylogenetic groups into distinct clades with both bootstrap and posterior probability support estimates (Figure 1). Maximum-likelihood trees constructed using each locus separately gave similar results (Appendix B.1, 2 and 3). All clade-defining branches were highly supported using both Bayesian and maximum-likelihood methods, but some intermediate branches had low bootstrap values. The number of isolates within each clade is given in Table 2. Fixed nucleotide differences between species were used to design a PCR-RFLP assay to distinguish P. nodorum from Pat1 (Figure 2). This method was tested on 96 P. nodorum isolates from eight different global populations and 68 Pat1 isolates from 4 different global populations. The method identified the species correctly for every isolate tested. The number of fixed differences and the average number of nucleotide substitutions between each species pair is summarized in Appendix A.2.
The mating-type primers successfully amplified the corresponding loci from *P. avenaria* and all *Pat* clades, but did not amplify the more distant P1 and P2 clades. Southern hybridization of P1 and P2 DNA with *P. nodorum* MAT1-1 and MAT1-2 probes was also negative. A summary of mating type assignments for the entire dataset is presented in Table 2. For the MAT1-1 idiomorph, 80 *P. nodorum* isolates were sequenced and collapsed into seven nucleotide haplotypes (5 with synonymous mutations, 2 with non-synonymous mutations). For the MAT1-2 idiomorph, five haplotypes (4 synonymous, 1 non-synonymous) were found among 117 *P. nodorum* isolates. A summary of the non-synonymous and synonymous substitutions found between mating type idiomorphs among all species is given in Appendix A.3.

All *Pat1* isolates carried only the MAT1-1 allele. Absence of the MAT1-2 locus was confirmed with Southern hybridization. MAT1-1 sequences in 38 *Pat1* isolates collapsed into a single haplotype. Both MAT1-1 and MAT1-2 alleles were found among the *Pat3* and *Pat5* isolates. *P. avenaria*, *Pat4*, and *Pat6* isolates were all MAT1-2, but small sample sizes may explain the lack of MAT1-1. Figures 1B and 1C show the maximum likelihood tree with 1000 bootstrap replicates for MAT1-1 and MAT1-2, respectively.

**Evidence for genetic exchange between *P. nodorum* and *Pat1***

We observed three multi-locus haplotypes (six individuals) that possessed a β-tubulin haplotype identical to a *Pat1* haplotype, with all other sequences classifying these isolates as *P. nodorum*. These six isolates were sampled at the same time from crested wheatgrass (*A. cristatum*) plants growing in close proximity to each other. Two of the three multi-locus haplotypes contained isolates with opposite *P. nodorum* mating-types, indicating that they were not the same clone. The mating-type of the third multi-locus haplotype could not be determined.

Five multi-locus haplotypes (eight isolates) carried a *P. nodorum* β-tubulin allele but *Pat1* alleles for the other genes. Three of these haplotypes had β-tubulin sequences that were identical to three different *P. nodorum* haplotypes. Two of the β-tubulin haplotypes were unique and appear to have arisen through intragenic recombination of existing *P. nodorum* point mutations (Figure 3). These isolates originated from a seed collection made in Canada that encompassed at
least several hundred square kilometers. One of the eight isolates had a *P. nodorum* Mat1-1 allele and a second isolate carried a *P. nodorum* Mat1-2 allele.

A comparison of the minimum number of recombination events within each locus for *P. nodorum*, *Pat1* and the hybrid isolates is shown in Table 3. In isolates that were strictly classified as *Pat1*, there was no evidence for intra-locus recombination for any locus. In *P. nodorum* there was evidence for several recombination events within loci. In the hybrid *P. nodorum* isolates carrying a *Pat1* β-tubulin sequence there was no evidence of recombination. The *Pat1* isolates carrying a *P. nodorum* β-tubulin, however, showed evidence of intragenic recombination at the β-tubulin locus. At the multi-locus level there was evidence for at least ten recombination events in *P. nodorum*. *Pat1* isolates showed a minimum of one recombination event between loci. *Pat1* hybrid isolates showed some evidence of recombination at the multi-locus level.

**Estimates of species divergence using *BEAST***

A posterior sample of species trees was generated using *BEAST*. A sample containing ~96% of all posterior trees collapsed into nine different consensus trees, shown in Figure 4A. All remaining consensus trees represented less than one percent of the posterior sample. This analysis established a clear separation between the time to most recent common ancestor (TMRCA) of P1 and P2 and the remaining *Phaeosphaeria* species. P1 and P2 split earlier from the other *Phaeosphaeria* sp., as shown by the high branching points in Figure 4A. Crossing over of branches between P1 and P2 indicate that the analysis was not able to resolve the precise relationships among these groups. Given this lack of consistency in the species trees in the posterior sample, it is unclear whether P1 or P2 shares the most recent common ancestor (MRCA) with the *Phaeosphaeria* sp. clade. Estimates for the TMRCA for each individual species or groups of species are shown in Figure 4B. The TMRCA estimate for the entire *Phaeosphaeria* sp. complex is marked by a red asterisk in both Figure 4A and 4B. The flattened posterior probability density peak in Figure 4B indicates a wide confidence interval for this estimate, which is illustrated in the varying height of the final node in the consensus trees. Additional informative sequence loci will be needed to provide more resolution for the older splitting events.
Enclosed in the red box in Figure 4A are the remaining species, termed “young *Phaeosphaeria*”. To make the tree more legible, all isolates belonging to the *P. avenaria* (*tritici*) complex (marked by a red bracket in Figure 1) were condensed into a single group. This simplification did not change the topology compared to an analysis where all groups were classified separately. The TMRCA of the young *Phaeosphaeria* is labeled with a black asterisk in both Figures 4A and B. In Figure 4B the narrowed peak for the TMRCA indicates a smaller confidence interval for this parameter. The smaller confidence interval is also seen in the illustration of the consensus trees, where there is a high amount of overlap for the height of the nodes defining the young *Phaeosphaeria* clade. There were several crossing branches among consensus trees, making it difficult to conclude which of the species shared a common ancestor before the ancestor marked by the black asterisk. *Pat1* is estimated to be the youngest group, followed by *Pat6, P. nodorum*, and the *P. avenaria* (*tritici*) complex (Fig. 4B).

**Allelic diversity at the hypothesized center of origin**

Allelic richness at eight micro-satellite loci in the largest Iranian *P. nodorum* population was compared with previously analyzed single field populations (Stukenbrock et al., 2006). As sample size increased, the Iranian population showed allelic richness similar to China, Switzerland and New York. The Iranian field population, however, contained a larger number of private alleles (Fig. 5A). The larger Iranian population, created by merging an Iranian field population from 2005 with a second field population sampled from the same region in 2010, showed a similar trend (Fig. 5B). The larger sample size allowed ADZE to extrapolate to a higher final sample size, where the rarefaction curves appear to level off, with Iran having the highest level of private allelic richness in comparison to the other populations.

Sequence diversity for *P. nodorum* and *Pat1* was also compared, but differences in sampling strategy and sample size between populations limited our ability to interpret these comparisons. Table 4 summarizes the average nucleotide diversity by locus and by regions of comparable sample size. Overall, the distribution of genetic diversity shown in the sequence data is strikingly similar to the main findings obtained for the SSR markers (Fig. 5A and B). Average nucleotide diversity was comparable between Iran, North America and Europe, but more private
alleles were found in Iran. Average nucleotide diversity was generally lower in the South African population (Table 4). Table 5 summarizes the number of multi-locus haplotypes and number of private multi-locus haplotypes found within regions. The P. nodorum collections from Iran, North America and Central Asia had considerably more private haplotypes than other regions (Table 5).

Tables 4 and 5 also show the nucleotide diversity and number and proportion of private haplotypes for the Pat1 collections. The Pat1 sequence diversity was generally lower than the average diversity found for P. nodorum (Table 4). The percentage of shared Pat1 haplotypes was relatively equal among regions (Table 5). Overall, P. nodorum had a higher number of total and private haplotypes when compared to Pat1 collected in the same region. To better illustrate the distribution of diversity in Pat1, a multi-locus haplotype network was constructed, excluding all hybrid isolates (Figure 6). The number of multi-locus haplotypes was highest in Canada. The network shows several loops among haplotypes, indicating that recombination played a role in generating new multi-locus haplotypes. This is consistent with the minimum number of recombination events shown in Table 3. These results show all possible recombination routes to newly generated haplotypes, not the minimum number of recombination events.

**DISCUSSION**

This study provides evidence from two independent neutral marker data sets that the origin of the wheat pathogen P. nodorum coincides with its host in the Fertile Crescent. We also clarify the relationships among wheat-infecting Phaeosphaeria sp. and several newly described relatives. Two of the nine species, P. nodorum and Pat1, are widely distributed in wheat fields globally, accounting for ~85% of sampled isolates. Pat1 was found almost exclusively in seed or ear collections. We hypothesize that this pattern indicates specialization of Pat1 to infect ears. Coalescent analysis identified a shared MRCA between P. nodorum and six additional species. This analysis indicates a rapid radiation of species from the shared ancestor, with three out of seven species described as wheat or oat pathogens. Coalescent analysis also revealed two new species closely related to the Phaeosphaeria sp. complex. Sequences from either P1 or P2 were
divergent enough to root the phylogeny of the remaining members of the Phaeosphaeria sp. complex.

Differentiating Phaeosphaeria sp.

Our analysis separated all clades with high posterior probability support, including non-synonymous differences between mating-type alleles (Appendix A.3). Non-synonymous changes between these idiomorphs are consistent with reproductive isolation between species. With the exception of potential hybridization between *P. nodorum* and *Pat1*, discussed below, the multi-locus dataset from this study conforms to the GCPSR definition of distinct species (Taylor et al., 2000). The host range for each species is unknown, but *P. nodorum* was found on at least five different grass species in our collections. We postulate that the relatives of *P. nodorum* also retain the ability to colonize several grass hosts.

*P. avenaria* f. sp. *tritici* was originally described using 44 isolates collected from barley, rye, wheat, oats and other grasses (Shaw, 1957a). These isolates were found to be non-pathogenic on oats and therefore given the formae specialis name *tritici*. Pathogenicity tests showed that about half of the isolates were weakly pathogenic on wheat, while the other half produced only traces of infection (Shaw, 1957a). Tests are currently underway to determine whether the recently discovered *Pat* sp. 3-6 are also pathogenic on wheat.

These results also significantly expand on earlier findings of a Phaeosphaeria species complex occurring on several plant hosts (Ueng et al., 1995; 1998; Malkus et al., 2005; Arkadiusz et al., 2006; Wang et al., 2007). We add three new *P. avenaria* f. sp. *tritici* (*Pat*) groups, isolated primarily from wild grasses, to the previously described *Pat1* and *Pat3* (Figure 1). The *Pat2* group described previously (Ueng et al., 1998) was not found in our collections. Our collection included eight isolates sampled from barley in North Dakota, which did not form a genetically distinct clade. These results indicate that *P. nodorum* barley biotypes are encompassed within the diversity observed within the *P. nodorum* species (Ueng et al., 1995). The two new, more distant species *P1* and *P2* were isolated from wheat ears in Iran.

*P1* and *P2* represent two un-described relatives of *P. nodorum*. We hypothesize that these species are relatives of *P. nodorum* that fall within the Phaeosphaeriaceae family described by
Zhang et al. (2009). The closest known pathogenic relative outside of this family remains *P. tritici-repentis*, which belongs to the sister family Pleosporaceae. It is unclear whether P1 and P2 are widely distributed outside of the Fertile Crescent. BLAST search results of ITS sequences indicate that it may be possible to find these species in the environment in Europe and China. Our P1 and P2 isolates were collected from wheat ears in Golestan Province in Iran. Additional collections made in the same province in 2010, in an attempt to re-sample these species, recovered only *Pat1* and *P. nodorum* from pycnidia on wheat ears.

The ITS region of P1 and P2 share at least 98% sequence identity with several GenBank accessions compared to an average 89% identity with *P. nodorum*. Several additional GenBank hits had higher sequence similarity to P1 and P2 than *P. nodorum*. Many of these fungi were collected during endophyte diversity screenings of soil, agricultural crops and trees. Without further sequences from these isolates it is not possible to correctly place them in the wheat-associated *Phaeosphaeria sp.* phylogeny presented in this manuscript. A more extensive sampling of grass-associated endophytes could reveal a wealth of currently un-named species that may be more closely related to *P. nodorum* than P1 and P2. The possible endophytic origins of *P. nodorum* are discussed below.

A third collection was made in 2011 from wild grass hosts, *A. tauri* and *E. tauri*, in the same region of Iran. This collection resulted in the discovery of *Pat4* and *Pat6*. Despite differences in host of origin, the species from wild grasses were more closely related to *P. nodorum* than P1 and P2, with average ITS identity for *Pat4* and *Pat6* of 94% and 96% respectively. *Pat4* was close enough to *P. avenaria* and the other *Pat* lineages to be included in the *P. avenaria* f. sp. *tritici* complex. *Pat6* split the phylogeny and was equidistant from the *P. nodorum* and *Pat* lineages, with an average nucleotide divergence of 103 nucleotides from *P. nodorum* and 94 nucleotides from *Pat3* (Appendix A.2). The discovery of *Pat4*, *Pat5* and *Pat6* enriched the species tree, filling a large gap between *P. nodorum* and *P. avenaria*. This new phylogeny provides an excellent framework for placing additional species or enlarging the sample size of smaller groups.
Evidence for specialization in Pat1 and hybridization with *P. nodorum*

In contrast to the small number of isolates sampled for P1, P2 and Pat3-6, Pat1 was found in populations on different continents and accounted for ~20% of the isolates in the total data set. The prevalence of Pat1 in collections from Canada and Iran suggested that it would be useful to have a simple method to distinguish between *P. nodorum* and Pat1. We used our sequence data to develop a PCR-RFLP assay based on fixed species polymorphisms (Figure 2) and validated it using more than 160 isolates. This test will be useful to determine if a collection contains mainly *P. nodorum* or Pat1 isolates, but it would not be able to detect the presence of hybrid isolates in the sample. Pat1 was isolated mainly from infected seeds or ears. We hypothesize that this correlation reflects specialization by Pat1 to infect ears, suggesting that Stagonospora leaf blotch is caused mainly by *P. nodorum* while Stagonospora ear blotch is caused by a combination of *P. nodorum* and Pat1. Further experiments will be needed to test this hypothesis.

It remains unclear whether Pat1 is the homothallic species described more than 50 years ago (Shaw, 1957a). Malkus et al. (2005) described their Pat1 isolates as homothallic citing a previous publication by Ueng et al. (1998). While the 1998 publication by Ueng et al. does not explicitly deal with mating behavior, it cites two unpublished observations of perithecial formation in Pat1 cultures started from single spores (G.M. Hoffmann and E. Arseniuk, unpublished data). Retention of sexual recombination after the loss of the MAT1-2 idiomorph has been observed in several *Neurospora* sp. (Lin and Heitman, 2007). The loops associated with the Pat1 multi-locus haplotype network (Figure 6) indicate that recombination has occurred between loci. Tests for minimum number of recombination events also suggested that Pat1 has undergone at least one inter-genic recombination event. Combining these earlier observations with our findings, we consider it likely that the Pat1 species described here is the same homothallic species first described by Shaw et al. (1957).

Our sample included 14 isolates that show evidence of hybridization between *P. nodorum* and Pat1. The finding of three different *P. nodorum* β-tubulin alleles in otherwise Pat1 isolates suggests that hybridization occurred independently at least three times. In fungi it remains unclear for how long interspecies fertility may persist after speciation (Giraud et al., 2008).
Recent studies on the sister species of *Microbotryum violaceum* indicated that the two sister species diverged several hundred thousand years ago, but showed evidence for very recent gene flow through hybridization (Gladieux et al. 2010). Hybridization at the β-tubulin locus between sexual and asexual endophytic Clavicipitaceae species has also been reported (Moon et al., 2004, Schardl et al. 1994, Tsai et al. 1994).

We hypothesize that the hybrid strains result from sexual recombination between *Pat1* and *P. nodorum*. Two *Pat1* hybrid isolates possessed either a *P. nodorum Mat1-1* allele or a *Mat1-2* allele. Though two individuals is a small sample to support our hypothesis, earlier studies recognized hybrid isolates based on the presence of multiple gene copies within single isolates (Tsai et al. 1994, Schardel et al., 1994). We did not find multiple copies for any gene in any of the 14 hybrid isolates. Given that these 14 isolates represent >4% of the 314 *Pat1, P. nodorum* and hybrid isolates included in the analysis, it appears that hybridization between these species may be relatively common. It is unclear whether these hybrid strains are aneuploids that contain duplicated chromosomes from each species, or are hybrids that have subsequently backcrossed with a parent species (introgressed). Genome-scale sequencing of these isolates will likely resolve many questions related to gene flow among *Pat1* and *P. nodorum*.

**Distribution of diversity for *P. nodorum* and other *Phaeosphaeria* sp.**

The collection from a wheat field in Iran provided us with an opportunity to compare the genetic diversity of *P. nodorum* at its hypothetical center of origin with other field populations. The Iranian population had a higher number of private alleles at the SSR loci, but average allelic richness was similar to other populations. Sequence data reflected the patterns observed for SSRs, with the Iranian population having the largest number of private alleles. Together these data support the hypothesis that the center of origin of *P. nodorum* is in the Fertile Crescent. Analyses of gene flow can provide additional support for a hypothesized center of origin by detecting ancient migration away from the center of diversity. But sufficient sample sizes are required for these analyses to consistently estimate levels and directions of gene flow among regions. Our attempts to analyze ancient migration using our small field collection of Iranian *P. nodorum* isolates were inconclusive.
Species richness offers additional support to the hypothesis of an origin in the Fertile Crescent. Four related species, P1, P2, Pat1 and P. nodorum were collected from the same wheat field in Iran in 2005. Although our BLAST search results indicate that close relatives of P1 and P2 may be present outside of the Fertile Crescent, the isolation of all four species from a single wheat field indicates higher species richness in Iran compared to other wheat fields around the world. A new sample from the same region in 2010 recovered both Pat1 and P. nodorum from infected ears. Pat4 and Pat6 were collected from wild grasses in the same region in 2011.

We consider two possible scenarios to explain the observed pattern of global diversity. First, recent political isolation of Iran may have restricted movement of the fungus on infected seed and grain, restricting gene flow and allowing divergence of the Iranian population through genetic drift. We consider this scenario highly unlikely because Iran experienced 1000s of years of international trade as a major hub on the Silk Road prior to its recent isolation. Our earlier work indicated that P. nodorum populations around the world recombine regularly (McDonald et al., 1999; Sommerhalder et al., 2006; Stukenbrock et al., 2006) and our field experiments showed that airborne ascospores contribute significantly to epidemic development in Swiss wheat fields (Sommerhalder et al., 2010). Thus we believe that airborne ascospores are regularly carried over long distances, enabling significant gene flow over spatial scales of 100s-1000s of km. To explain the observed global pattern of diversity, we hypothesize that the Iranian population represents an ancient pathogen population near the center of origin of the pathogen in the Fertile Crescent. We further hypothesize that recent international trade of infected wheat seed has enabled global dissemination of large, diverse populations of this fungus among the major wheat producing regions around the world, masking historical gene flow and lowering genetic differentiation between populations as we reported earlier (Stukenbrock et al., 2006). Larger population collections from the Fertile Crescent will be needed to conduct additional gene flow analyses that may identify a more precise origin of P. nodorum.

Endophytic origins of the Phaeosphaeria sp. complex

While our data suggests that the pathogen P. nodorum originated in the Fertile Crescent, we believe the Phaeosphaeria species complex likely originated in Eurasia long before the
development of agriculture. The significant divergence found at highly conserved loci, including ITS and β-tubulin, indicate that this species complex is much older than *Zymoseptoria tritici*, another wheat pathogen thought to have emerged in the Fertile Crescent around 10,000 years ago (Stukenbrock et al., 2007). *P. nodorum* belongs to the Pleosporales, the largest order within the *Dothideomycetes* (Zhang et al., 2009). The Pleosporales are represented in the fossil record in the Eocene, giving a minimum age of 55-35 Ma (Mindell et al., 2007), but the order is thought to have emerged 280-175 Ma (Berbee and Taylor, 2010). Pollen records indicate that cool season grasses (including ancestors of wheat and oats) were widespread and abundant throughout Eurasia by the late Miocene (15 Ma - 7 Ma). These grasses may have provided the first niches to be colonized by grass-associated saprophytic or endophytic fungi.

In a large phylogenetic study of the Pleosporales, *P. nodorum* and *P. avenaria* were placed in a clade associated mainly with endophytes and saprobes of terrestrial and water associated grasses (Zhang et al., 2009). The prevalence of highly similar BLAST hits for P1 and P2 connected to endophyte studies is striking. In 250 BLAST hits of 90% similarity or higher, 61 (25%) included the word endophyte in the GenBank accession file. Similarly 62 BLAST hits for P1 included the word endophyte and many of these accession numbers were the same between the two BLAST searches. The spatial distribution of BLAST hits with >98% identity to P1 was limited to studies conducted in Eurasia, with the exception an isolate collected from a maize plant in Ohio. For P2 two exact matches were found in air samples collected in Germany. The spatial distribution of BLAST hits became global at an identity of 97% or less for both species.

*Phaeosphaeria* sp. have long been recognized as common endophytes of grasses (Newell, 1996; Bergbauer and Newell, 1992). Several *Phaeosphaeria* sp. were described as beneficial, seed transmitted endophytes of the reed grass *Phragmites australis* (Ernst et al., 2003). Seed transmission is a common attribute of endophytic fungi (Clay and Schardl, 2002; Saikkonen et al., 2004) and is a noted feature of *P. nodorum* (Solomon et al., 2006). The Pat5 species described in this study is known to be avirulent on a collection of susceptible wheat cultivars (T. Friesen, unpublished data.). Pat5 was collected exclusively from wild grasses, including *Agropyron cristatum* (crested wheatgrass), *Agropyron intermedium* and *Bromus inermis* (smooth brome). All of these grass species are native to northeastern Europe and were imported into the
United States for use as cold tolerant pasture grasses (www.plants.USDA.gov). \( P. avenaria \) and other unknown \( Phaeosphaeria \) sp. were recently isolated from asymptomatic leaf tissue in a diversity study examining fungal endophytes of \( D. glomerata \) in Spain (Sánchez Márquez et al., 2007). Combining these findings with our collections from Iran, it appears that all of these \( Phaeosphaeria \) sp. have roots in Eurasian grasslands, leading us to hypothesize that this is the original source of the species complex.

We hypothesize that ancestral \( Phaeosphaeria \) sp. were present in the ancient Fertile Crescent living as endophytes or pathogens of wild grasses. The species we now call \( P. nodorum \) and \( Pat1 \) are domesticated pathogenic species that outcompeted their close relatives to colonize domesticated wheat. Because the closest relatives of \( P. nodorum \) and \( Pat1 \) are not extinct, this collection of cryptic species provides us with an excellent opportunity to elucidate the molecular mechanisms that enabled \( P. nodorum \) and \( Pat1 \) to emerge as important pathogens on wheat.

**SUPPORTING MATERIAL**

Appendix A. Supplementary Tables  
Supplementary tables for this chapter.

Appendix B. Supplementary Figures  
Supplementary figures for this chapter.

Appendix C. BLAST results for fungal species P1  
A summary of the top 250 BLAST hits for P1. Not printed, in excel format available upon request.

Appendix D. BLAST results for fungal species P2  
A summary of the top 250 BLAST hits for P2. Not printed, in excel format available upon request.

**ACKNOWLEDGEMENTS**

The authors would like to thank all collectors who provided isolates or infected leaves. We also acknowledge the Genetic Diversity Center (GDC) in Zurich for providing excellent sequencing facilities. This project was funded by the ETH Zurich.
Table 1. List of regional collections, sample sizes and host source for all isolates.

<table>
<thead>
<tr>
<th>Populations</th>
<th>Year</th>
<th>Collector</th>
<th>Location</th>
<th>N</th>
<th>Host Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>2005</td>
<td>R. Sommerhalder</td>
<td>Golestan Province</td>
<td>42</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>M. Razavi</td>
<td>Golestan Province</td>
<td>15</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>M. Razavi</td>
<td>Golestan Province</td>
<td>17</td>
<td><em>Elymus tauri; Agropyron tauri</em></td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>2005</td>
<td>R. Sommerhalder</td>
<td></td>
<td>12</td>
<td>Wheat</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>2003-04</td>
<td>H. Maraite, E. Duveiller</td>
<td></td>
<td>23</td>
<td>Wheat</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>2004</td>
<td>H. Maraite, E. Duveiller</td>
<td></td>
<td>5</td>
<td>Wheat</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>2003</td>
<td>H. Maraite, E. Duveiller</td>
<td></td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>Russia</td>
<td>2003</td>
<td>H. Maraite, E. Duveiller</td>
<td></td>
<td>4</td>
<td>Wheat/Durum Wheat</td>
</tr>
<tr>
<td>China</td>
<td>2001</td>
<td>R. Wu</td>
<td>Fujian Province</td>
<td>8</td>
<td>Wheat</td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>2005</td>
<td>T.W. Hollins</td>
<td>Cambridge</td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>Denmark</td>
<td>2005</td>
<td>E. Stukenbrock</td>
<td>Jylland</td>
<td>37</td>
<td>Triticale/ Wheat</td>
</tr>
<tr>
<td>Sweden</td>
<td>2005</td>
<td>E. Blixt</td>
<td></td>
<td>15</td>
<td>Wheat</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1994</td>
<td>S. Keller</td>
<td>Zurich</td>
<td>7</td>
<td>Wheat</td>
</tr>
<tr>
<td>North America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>1991</td>
<td>R. Clear</td>
<td>Alberta</td>
<td>18</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Manitoba</td>
<td>7</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Saskatchewan</td>
<td>23</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td>United States</td>
<td>1986-92</td>
<td>J. Krupinsky</td>
<td>North Dakota (CWG)</td>
<td>33</td>
<td><em>A. cristatum</em></td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>M. Schmidt</td>
<td>Oregon</td>
<td>10</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>G. Milus, M. Gray</td>
<td>Arkansas</td>
<td>8</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>J. Krupinsky</td>
<td>North Dakota (AVR)</td>
<td>2</td>
<td><em>Agropyron sp., A. intermedium</em></td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>J. Krupinsky</td>
<td>North Dakota (AVR)</td>
<td>4</td>
<td>Altai wild rye, Smoothe Brome</td>
</tr>
<tr>
<td></td>
<td>1984</td>
<td>J. Krupinsky</td>
<td>North Dakota (AVR)</td>
<td>3</td>
<td>Smoothe Brome</td>
</tr>
<tr>
<td></td>
<td>1985</td>
<td>J. Krupinsky</td>
<td>North Dakota (VIR)</td>
<td>1</td>
<td>Wild Rye</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>L. Francl</td>
<td>North Dakota</td>
<td>3</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>J. Krupinsky</td>
<td>North Dakota (VIR)</td>
<td>8</td>
<td>Barley</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>P. Lips</td>
<td>OH03 Sn1501(VIR)</td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>T. Friesen</td>
<td>North Dakota</td>
<td>12</td>
<td>Wheat</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>Z. Pretorius</td>
<td>Western Cape</td>
<td>32</td>
<td>Wheat</td>
</tr>
<tr>
<td>SN15</td>
<td>1979</td>
<td>J. Krupinsky</td>
<td>North Dakota (AVR)</td>
<td>1</td>
<td>Grass</td>
</tr>
<tr>
<td>SN79</td>
<td></td>
<td>T. Friesen</td>
<td>North Dakota (VIR)</td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>SN4</td>
<td></td>
<td>K. Clarke</td>
<td>Australia (Mt. Barker)</td>
<td>1</td>
<td>Oat</td>
</tr>
<tr>
<td><em>P. avenaria</em></td>
<td>2009</td>
<td>Aus. Dept. Agriculture</td>
<td>Australia</td>
<td>1</td>
<td>Wheat</td>
</tr>
</tbody>
</table>

**Total**: 355

---

*a* Isolates for which genome sequence is available

*b* Isolates collected from North Dakota on Crested Wheat Grass (CWG)

*c* Populations which have known phenotypes avirulent (AVR) or virulent (VIR)

*d* Collections used for the first time in this publication
Table 2. Number of isolates and mating type assignment by region and species

<table>
<thead>
<tr>
<th>Region</th>
<th>Species</th>
<th>N Isolates</th>
<th>N MAT1-1</th>
<th>N MAT1-2</th>
<th>N NDa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td><em>P. nodorum</em></td>
<td>29</td>
<td>10</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat1</td>
<td>19</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat4</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat6</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>7</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Central Asia</td>
<td><em>P. nodorum</em></td>
<td>40</td>
<td>18</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Pat1</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>China</td>
<td><em>P. nodorum</em></td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td><em>P. nodorum</em></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td><em>P. avenaria</em></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td><em>P. nodorum</em></td>
<td>31</td>
<td>14</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pat3</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Sweden</td>
<td><em>P. nodorum</em></td>
<td>15</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Switzerland</td>
<td><em>P. nodorum</em></td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>North America</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td><em>P. nodorum</em></td>
<td>10</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td><em>P. nodorum</em></td>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat1</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat1 hybrid</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>North Dakota AVR</td>
<td><em>P. nodorum</em></td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>North Dakota AVR</td>
<td><em>Pat5</em></td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>North Dakota VIR</td>
<td><em>P. nodorum</em></td>
<td>11</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>North Dakota CWG</td>
<td><em>P. nodorum</em></td>
<td>27</td>
<td>11</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><em>P. nod. hybrid</em></td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Arkansas</td>
<td><em>P. nodorum</em></td>
<td>8</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>Pat1</td>
<td>41</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pat1 hybrid</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>South Africa</strong></td>
<td><em>P. nodorum</em></td>
<td>32</td>
<td>17</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>P. nodorum</em></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>P. avenaria</em></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>355</td>
<td>177</td>
<td>151</td>
<td>26</td>
</tr>
</tbody>
</table>

a ND- not determined because amplification did not work
Table 3. *P. nodorum* and *Pat1* minimum number of intragenic recombination events by locus

<table>
<thead>
<tr>
<th>Locus</th>
<th>Multi-locus Species Classification</th>
<th>P. nodorum</th>
<th>Pat1</th>
<th>P. nod hybrid</th>
<th>Pat1 hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Isolates</td>
<td></td>
<td>227</td>
<td>73</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td><strong>β-tubulin</strong></td>
<td>N Haplotypes</td>
<td>32</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>N sites(^a)</td>
<td>479</td>
<td>479</td>
<td>479</td>
<td>479</td>
</tr>
<tr>
<td></td>
<td>N polymorphic(^b)</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Pairs 4-gamete</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Min Rec(^c)</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><strong>β-xylosidase</strong></td>
<td>N Haplotypes</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>N sites(^a)</td>
<td>687</td>
<td>687</td>
<td>687</td>
<td>687</td>
</tr>
<tr>
<td></td>
<td>N polymorphic(^b)</td>
<td>68</td>
<td>10</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Pairs 4-gamete</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Min Rec(^c)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>ITS region</strong></td>
<td>N Haplotypes</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>N sites(^a)</td>
<td>468</td>
<td>475</td>
<td>475</td>
<td>475</td>
</tr>
<tr>
<td></td>
<td>N polymorphic(^b)</td>
<td>18</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Pairs 4-gamete</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Min Rec(^c)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Multi-locus</strong></td>
<td>N Haplotypes</td>
<td>92</td>
<td>17</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>N sites(^a)</td>
<td>1633</td>
<td>1641</td>
<td>1641</td>
<td>1641</td>
</tr>
<tr>
<td></td>
<td>N polymorphic(^b)</td>
<td>44</td>
<td>17</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Pairs 4-gamete</td>
<td>327</td>
<td>12</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Min Rec(^c)</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

\(^a\) Number of nucleotide sites for within species alignment  
\(^b\) Number of polymorphic nucleotide sites  
\(^c\) The minimum number of recombination events
Table 4. Sequence diversity by locus in *P. nodorum* and *Pat1* populations of comparable sample size.

### *P. nodorum*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Population</th>
<th>N</th>
<th>N haplotypes</th>
<th>Nuc Diversity&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-tubulin</td>
<td>Iran</td>
<td>29</td>
<td>5</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Denmark</td>
<td>31</td>
<td>5</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td>29</td>
<td>9</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>32</td>
<td>5</td>
<td>0.008</td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>Iran</td>
<td>29</td>
<td>5</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Denmark</td>
<td>31</td>
<td>6</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td>29</td>
<td>6</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>32</td>
<td>4</td>
<td>0.001</td>
</tr>
<tr>
<td>ITS</td>
<td>Iran</td>
<td>29</td>
<td>6</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Denmark</td>
<td>31</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td>29</td>
<td>3</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>32</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Multi-locus</td>
<td>Iran</td>
<td>29</td>
<td>16</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Denmark</td>
<td>31</td>
<td>14</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td>29</td>
<td>20</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>32</td>
<td>10</td>
<td>0.003</td>
</tr>
</tbody>
</table>

### *Pat1*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Population</th>
<th>N</th>
<th>N haplotypes</th>
<th>Nuc Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-tubulin</td>
<td>Iran</td>
<td>19</td>
<td>1</td>
<td>N\A</td>
</tr>
<tr>
<td></td>
<td>Azerbaijan</td>
<td>5</td>
<td>1</td>
<td>N\A</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>41</td>
<td>2</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>North Dakota</td>
<td>8</td>
<td>1</td>
<td>N\A</td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>Iran</td>
<td>19</td>
<td>3</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>Azerbaijan</td>
<td>5</td>
<td>2</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>41</td>
<td>5</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>North Dakota</td>
<td>8</td>
<td>4</td>
<td>0.002</td>
</tr>
<tr>
<td>ITS</td>
<td>Iran</td>
<td>19</td>
<td>3</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>Azerbaijan</td>
<td>5</td>
<td>2</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>41</td>
<td>4</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>North Dakota</td>
<td>8</td>
<td>4</td>
<td>0.0023</td>
</tr>
<tr>
<td>Multi-locus</td>
<td>Iran</td>
<td>19</td>
<td>5</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>Azerbaijan</td>
<td>5</td>
<td>2</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>41</td>
<td>11</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>North Dakota</td>
<td>8</td>
<td>5</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

<sup>a</sup>Pairwise nucleotide diversity (π)
Table 5. Multi-locus haplotype diversity for *P. nodorum* and *Pat1* by population

### *P. nodorum* multi-locus haplotype diversity by region

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>h&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Private h</th>
<th>% Shared h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>South Africa</td>
<td>32</td>
<td>10</td>
<td>3</td>
<td>70%</td>
</tr>
<tr>
<td>Europe</td>
<td>54</td>
<td>20</td>
<td>11</td>
<td>45%</td>
</tr>
<tr>
<td>North America</td>
<td>63</td>
<td>35</td>
<td>27</td>
<td>23%</td>
</tr>
<tr>
<td>Central Asia</td>
<td>40</td>
<td>32</td>
<td>25</td>
<td>22%</td>
</tr>
<tr>
<td>Middle East</td>
<td>29</td>
<td>16</td>
<td>13</td>
<td>19%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>226</td>
<td>93</td>
<td>77</td>
<td>17%</td>
</tr>
</tbody>
</table>

### *Pat1* multi-locus haplotype diversity by region

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>h&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Private h</th>
<th>% Shared h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>41</td>
<td>11</td>
<td>7</td>
<td>36%</td>
</tr>
<tr>
<td>Central Asia</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>50%</td>
</tr>
<tr>
<td>Middle East</td>
<td>19</td>
<td>5</td>
<td>3</td>
<td>40%</td>
</tr>
<tr>
<td>North America</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>60%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>73</td>
<td>18</td>
<td>13</td>
<td>28%</td>
</tr>
</tbody>
</table>

<sup>a</sup> h is the number of nucleotide haplotypes
Fig. 1.
Multi-gene haplotype phylogenetic reconstruction of nine *Phaeosphaeria* sp. (A) On the left is the Bayesian maximum clade credibility tree with posterior probabilities. On the right is the maximum-likelihood tree with bootstrap values. Isolate names are listed under the multi-locus haplotype number preceded by the letter “H”. Collection locations, year of collection and host are listed in parentheses underneath or next to each clade name. Below is shown the un-rooted maximum-likelihood trees for (B) the *MAT1-1* and (C) *MAT1-2* idiomorphs. Conserved *P. nodorum* MAT primers successfully amplified the corresponding loci for six out of eight *Phaeosphaeria* sp. Branches are labeled with bootstrap support.
Fig. 2.
PCR-RFLP assay designed to distinguish between *P. nodorum* and *Patl*. The 962 base pair amplicon corresponding to SNOG_06123.2 (β-xylosidase) is amplified in both species. *Patl* isolates possess one specific restriction site for *ScaI*, resulting in two fragments visible at 695 bp and a second at 267 bp, respectively. Faint bands of incomplete digestion are present in the *Patl* samples. *P. nodorum* isolates lack a *ScaI* restriction site, leaving the uncut 962 bp amplicon.
Fig. 3.
Hypothetical parents of unique β-tubulin haplotypes found in hybrid Pat1 isolates. All point mutations were found in *P. nodorum* haplotypes. Single recombination events can explain the two unique β-tubulin haplotypes found in *Pat1* hybrids.
Fig. 4A.
The *BEAST top nine consensus species trees, taken from a posterior sample of 7,500 trees. These species trees were constructed after the modeling of each gene tree separately. The top four consensus trees are shown with solid lines and adjusted in opacity based on their frequency within the posterior sample; this frequency is shown in the legend. The remaining trees (<6% of the posterior each) are shown in dashed lines with equal opacities, their relative frequencies in the posterior are shown in the legend. The height of the branches corresponds to relative time. For the trees shown, time is scaled with a strict molecular clock fixed to one based on the β-tubulin locus. Coalescent events, at nodes, show most recent common ancestors (MRCA) for each species. The tree is divided into two parts by a red box. This box denotes the consistent separation of the young *Phaeosphaeria* from P1 and P2 in all nine consensus trees. The *P. avenaria (tritici)* complex (shown in Fig. 1A) is collapsed into a single group to simplify the figure.
**Fig. 4B.**
Posterior probability estimates of the time to most recent common ancestor (TMRCA) for each species population or group of species shown in part A. The posterior probability densities for each TMRCA are plotted against relative time (x-axis). The time axis is scaled with the same molecular clock used in part A. Most recent speciation events are close to zero, with time running into the past as you move from left to right along the axis. These estimates were calculated assuming a Yule’s speciation model. The black and red asterisks mark the time estimates that correspond with the nodes on the consensus trees depicted in part (A).
Fig. 5.
ADZE rarefaction results from 8 SSR loci for single field populations of *P. nodorum* (A) and a larger Iranian field population formed by merging two field populations collected in the same region in different years (B). The rarefied mean allelic richness (left) and private allelic richness (right) are shown in relation to increasing sample size (g). Maximum g samples sizes are limited by the smallest population sample size.
Fig. 6.
Parsimony-based multi-locus haplotype network of Pat1 generated in TCS. The size of the circles corresponds to the number of isolates belonging to each haplotype. The haplotypes are divided into colors based on the population from which the isolates were collected. Missing haplotypes are denoted with white open circles. Each white open circle indicates a single point mutation. Loops are indicative of recombination. The network shows all possible recombination events, not the minimum number of events.
REFERENCES


Microbe Interact. 16, 580–587.


Sommerhalder, R.J., McDonald, B.A., Zhan, J., 2006. The frequencies and spatial distribution of mating types in *Stagonospora nodorum* are consistent with recurring sexual reproduction. Phytopathology 96, 234–239.


Tsai, H.F., Liu, J.S., Staben, C., Christensen, M.J., Latch, G.C.M., Siegel, M.R. Schardl, C.L., 1994. Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloë* species. PNAS 91, 2542-2546.


differentiation from *S. nodorum*. Mycol. Res. 102, 607–614.


Chapter 3:

Horizontal acquisition of three necrotrophic effectors in the fungal wheat pathogen Phaeosphaeria nodorum

Megan C. McDonald, Richard P. Oliver, Timothy L. Friesen, Patrick C. Brunner and Bruce A. McDonald

Submitted to PLoS Pathogens June 2012
ABSTRACT

*Phaeosphaeria nodorum* (anamorph *Stagonospora nodorum*) secretes three characterized necrotrophic effectors (NEs) *SnTox1*, *SnTox3* and *SnToxA*. Over 1000 isolates from 16 global populations were assayed for the presence or absence of each NE in order to generate a multi-effector genotype for each isolate. Only one of 16 populations showed a significant deviation from the frequency of multi-effector genotypes expected under random mating, indicating that selection has not favored isolates carrying more or fewer NEs. Earlier population genetic studies using neutral markers showed that *P. nodorum* exhibits limited population differentiation over large geographical scales as a consequence of high levels of gene flow within and among continents. At NE loci, however, statistically significant differences in NE frequency were found even across relatively small spatial scales, consistent with the hypothesis that differences in frequencies of NE sensitivity genes in associated wheat populations leads to population differentiation for corresponding pathogen effector genes. Diversity at each NE locus was assessed by sequencing ~200 individuals from around the world. The population harboring the highest sequence diversity was different for each effector locus, but never coincided with the highest diversity for neutral markers found in Fertile Crescent populations where the pathogen likely originated. Screening for the three NEs in *P. nodorum* and eight sister species revealed that only one sister species, *Phaeosphaeria avenaria tritici 1* (*Pat1*), also carried these NEs. Coalescent analysis of the species complex indicates that the three NEs in *P. nodorum* were most likely acquired horizontally. *Pat1* has many fewer NE alleles, all of which were shared with *P. nodorum*. Coupled with the fact that a small number of hybrids between *P. nodorum* and *Pat1* were observed in a previous phylogenetic study, it appears that the NEs in *Pat1* were acquired from *P. nodorum* via interspecific hybridization. We conclude that the horizontal acquisition of necrotrophic effectors was the key innovation that enabled the emergence of *P. nodorum* and *Pat1* as globally important wheat pathogens.
INTRODUCTION

Necrotrophic fungal pathogens secrete a variety of necrotrophic effectors (NE) (syn host-specific/selective toxins) that interact in a gene-for-gene manner with host susceptibility genes [1]. Typical plant defense responses involve the induction of pathogenicity related genes that lead to the production of anti-microbial compounds, the accumulation of reactive oxygen species and localized cell death. This defense response is known as the hypersensitive response (HR) [2]. A growing body of evidence supports the hypothesis that necrotrophic pathogens have taken advantage of the HR, using small secreted proteins to activate HR preceding fungal growth [3,4]. These small, secreted proteins are referred to as necrotrophic effectors (NEs).

Effectors are a class of pathogen proteins or metabolites whose function is to alter or suppress the host’s normal immune response. Three NEs have been described for the fungal wheat pathogen *Phaeosphaeria nodorum* [3-5]. Each of these NEs is a small, secreted peptide that displays a presence/absence polymorphism in natural field populations. *SnTox1* is a cysteine-rich peptide that was shown to exhibit significant diversifying selection [3]. *SnTox3* has no known homology to any proteins available in public databases [4]. *SnToxA* has limited homology to a prokaryotic gene and also exhibits significant diversifying selection [6]. Transformation with any of the three NEs into a non-pathogenic fungal isolate was sufficient to induce necrosis on susceptible wheat cultivars [3-5]. CHEF gel analysis indicates that each of the NEs are located on different chromosomes [3]. This is unusual as other fungal effectors have been shown to occur in clusters (i.e. the RxLR effectors of *Phytophthora* sp. [7]) or exclusively on pathogenicity associated chromosomes (i.e. mobile pathogenicity chromosomes in *Fusarium* sp. [8]).

Very little is known about the evolutionary origins of effector genes. While there is a growing list of shared properties associated with effector genes (for reviews see: Kamoun et al., 2007 and Stergiopoulos et al., 2009 [9,10]), the lack of homology with other proteins, even within the genome, makes it difficult to trace the evolutionary history of these genes. Population genetic studies of effector loci have provided important insights into the evolutionary processes that affect NE loci. A population genetic analysis of the *NIP1* gene in *Rhynchosporium commune* showed that alteration of the effector protein sequence or deletion of the effector allele could lead to virulence [11]. Studies on the obligate biotroph of flax, *Melampsora lini*, also revealed high
levels of non-synonymous substitutions and a presence/absence polymorphism [12,13]. Virulence alleles in *L. maculans* were attributed to both deletion of the *AvrLm6* locus and introduction of early stop codons by repeat-induced polymorphism (RIP) [14].

For necrotrophic effectors, deletion of the effector gene leads to loss of the virulent phenotype. Up until now, NE studies have focused on individual genes or small groups of effector genes in a small number of individuals (eg, [3,4,13,15,16]) though it is clear that fungal populations are large and capable of harboring high levels of NE diversity. As more effector genes are discovered and characterized, a key question has become, how did this class of genes originate within fungal pathogens? Understanding the evolutionary origins of these genes could provide significant insights into the mechanisms involved in pathogen emergence.

The horizontal transfer of *SnToxA* from *P. nodorum* to *Pyrenophora tritici-repentis* is thought to have led to the emergence of *P. tritici-repentis* as the tan spot pathogen on wheat [5]. Detection of this horizontal gene transfer (HGT) event was made possible by the high sequence similarity between the *SnToxA* allele found in *P. nodorum* and the *PtrToxA* allele found in *P. tritici-repentis*. It remains unknown if the NEs present in *P. nodorum* are the result of a long co-evolutionary process between the pathogen and its hosts, or alternatively if the NEs were acquired horizontally from an unknown donor.

This study focuses on the population genetics and evolutionary history of *SnTox1*, *SnTox3* and *SnToxA* in *Phaeosphaeria nodorum* and its closest known relatives. We assessed the global distribution and geographical diversity for all three effectors. We determined the presence or absence of each gene in over 1000 global isolates using both PCR and Southern hybridization. We calculated the frequency of each NE over spatial scales ranging from fields to continents and generated multi-effector genotypes to determine if selection was operating on the combination of NEs. To gain insight into the ancestral origin of these NEs in *P. nodorum*, we assessed the presence or absence of each NE in eight recently described sister species. Finally, we sequenced each NE in several hundred global strains to compare NE sequence diversity with previously published population genetic studies based on neutral markers.

**MATERIALS AND METHODS**

**Isolate Collection**

Isolates used in this study are described in Table 1. Fungal hyphae were transferred to 50
mL Yeast Sucrose Broth (YSB, 10g/L Yeast Extract, 10g/L sucrose) and grown on a rotary shaker for 3 days at 120 rpm at 18°C. Fungal biomass was lyophilized, ground into a powder and total DNA was extracted using the DNeasy Plant Mini DNA extraction kit (Qiagen GmbH, Hilden, Germany) following the manufacturer’s instructions. Extracted DNA was diluted 1:500 with sterile water for PCR amplification.

**Data Collection**

PCR amplification was performed in 20 µl reactions containing 0.05 µM of each primer (Microsynth, Balgach Switzerland), 1 × Dream Taq Buffer (MBI Fermentas), 0.4 µM dNTPs (MBI Fermentas) and 0.5 units of Dream Taq DNA polymerase (MBI Fermentas). The PCR cycle parameters were: 2 min initial denaturation at 96°C followed by 35 cycles of 96°C for 30 s, anneal for 45 s at 72°C for 1 min. A final 7 min extension was made at 72°C. Annealing temperature and primers used have been published, see Friesen et al. [5] for *SnToxA*, Liu et al. [4] for *SnTox3* and Liu et al. [3] for *SnTox1*.

Sequencing reactions were conducted in 10 µl volume using the BigDye® Terminator v3.1 Sequencing Standard Kit (Life Technologies, Applied Biosystems) with both the forward and reverse primers. The cycling parameters were 96°C for 2 min followed by 55 or 99 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. PCR products were cleaned with the illustra™ Sephadex™ G-50 fine DNA Grade column (GE Healthcare) according to the manufacturer’s recommendations and sequenced with a 3730xl Genetic Analyzer (Life Technologies, Applied Biosystems). Alignment of forward and reverse sequences was performed in SeqScape software V2.5 (Life Technologies, Applied Biosystems). Quality screening and ambiguous base calls were edited by hand using SeqScape. Final alignments were exported and re-aligned using ClustalW, implemented online using the Max Planck Institute of Bioinformatics Toolkit (http://toolkit.tuebingen.mpg.de). Sequences that have been published previously are *SnToxA* accession numbers DQ423483; EF108451–EF108463; *SnTox3* accession number FJ823644; *SnTox1* accession numbers JN791682-JN791693. New sequence haplotypes described in this manuscript are now deposited in GenBank under accession numbers XXX-XXX.

Genomic DNA was blotted onto nylon membranes using a Bio-Dot microfiltration
apparatus (BIO-RAD) following the instructions in the user manual. Radioactive probes were synthesized using purified PCR product (purification same as above). Probe labeling with $^{32}$P was performed with 25 to 50 ng/µL of template DNA using the NEBlot Kit (NE Biolabs) following the manufacturer’s recommendations. Unincorporated nucleotides were filtered out using illustra NICK™ columns (GE Healthcare). DNA hybridization, membrane washing and image acquisition were performed as described previously[35].

**Data Analysis**

*Contingency $\chi^2$ tests and Fisher’s Exact Tests*

Effector presence or absence was summarized as allele frequencies within each population. Contingency $\chi^2$ tests were used to test for independence of effector frequencies between populations. The program CHIFISH was used to calculate both Fisher’s exact test and Pearson’s $\chi^2$ test for each locus separately. P-values for independence based on the combined effector frequencies were calculated by summing the $\chi^2$ test statistics and degrees of freedom, or by combining the P values from Fisher’s exact test using Fisher’s method [36,37]. A Bonferroni correction was applied to correct P-values for multiple testing. After confirming the independence of each population the expected number of multi-locus genotypes were calculated. The frequency of each effector within a population was used as the expected probability of sampling an individual carrying the corresponding allele. For three bi-allelic loci there are eight possible multilocus genotypes ($ToxA+Tox3+Tox1+$; $ToxA-Tox3-Tox1-$ etc.). The expected number of multilocus genotypes was compared with the observed number of multilocus genotypes using Fisher’s exact test in each population (Fisher, 1970). This test was deemed to be most appropriate for large multinomial datasets with small expected values (n<5) [38]. Fisher’s exact tests were performed using the statistics package implemented in R. For contingency tables larger than 2x2, an estimation of the P-value was made using Monte Carlo simulations. Our test was conducted with MonteCarlo=TRUE with 10,000 replicates (Patefield, 1981). Each population was tested for independence separately, so no p-value corrections were applied.

*Gene Diversity and Haplotype Networks*

Effector sequences were collapsed into haplotypes using Map, implemented in the Java-based
program package SNAP workbench [1,39,40]. This program implements several individual programs into one platform in order to facilitate analysis of population parameters [2,39]. The haplotype alignment was used in TCS 1.3 to generate the most parsimonious haplotype network [3,4,41]. This program utilizes statistical parsimony methods to infer unrooted cladograms based on Templeton’s 95% parsimony connection limit. Mutational steps resulting in non-synonymous changes were identified manually using Sequencher v4.8 (Gene Codes Corp., Ann Arbor, MI, USA).

**Rarefaction**

Rarefaction was used to identify regions of highest sequence diversity among populations with unequal sampling sizes. Rarefaction analysis was conducted using the method described by Simberhof [3-5,42] and implemented in an online calculator (http://www.biology.ualberta.ca/jbrzusto/rarefact.php#Inputs). Rarified sample sizes from 10 to 20 were used to estimate the number of sequence alleles by global region. The output given is the rarefied allele count and standard deviation at each of the given sample sizes. Standard deviations were used to calculate the standard error (SE). The number of sequence alleles, plus the 95% confidence interval (1.96*SE), were plotted for each sample size and population using R.

**RESULTS**

**Frequencies of necrotrophic effectors differ among populations but are randomly assorted within populations**

Effector presence or absence was determined with gene-specific PCR primers for over 1000 isolates across seven major regions, including North America, Europe, Iran, Central Asia, China, South Africa and Australia (Table 1). Results from PCR assays were compared to Southern hybridization assays using 193 isolates for *SnToxA*, 284 isolates for *SnTox3* and 242 isolates for *SnTox1*. The average disagreement between the PCR and hybridization assays across all three loci was 4%. Sixteen out of the 30 discrepancies involved the highly polymorphic *Tox1* locus and 6 of these were in the Central Asian population. When the two assays were not in agreement, the result from the hybridization assay was included in further analyses. Effector
frequencies were calculated for each region and for the 16 sub-populations; variances were estimated using 100 bootstrapped replicates with replacement by region (Figure 1). *SnToxA* was present at a lower global frequency than the other two effectors. *SnTox1* was found at a frequency >60% in all continental regions. All three effectors were present at high frequencies in South Africa and Australia.

Effector frequencies differed significantly among field populations for all three loci according to contingency chi-squared tests and Fisher’s exact tests (Table S1). In many cases, effector frequencies among field populations within a region (e.g. Arkansas and Texas within North America or Sweden and Denmark in Europe) also differed significantly.

The complete multi-locus effector presence/absence genotype was determined for 945 strains. Given three effector loci there are eight possible multi-locus presence/absence genotypes. Based on the observed effector frequencies in each population, the expected number for each multi-locus genotype under random mating was estimated and compared to the observed number of multi-locus genotypes (Figure 2). Fisher’s exact test was used to determine if the number of observed genotypes differed from the number of expected genotypes. The only population with a significant deviation from expected proportions was Oregon (Table 2). Values for European and North American populations were summed to generate the graph shown in Figure 2. Sample sizes for each NE locus are given along with the number of isolates for which complete genotypes were scored (Figure 2).

**Effector sequence alleles are diverse and globally distributed**

We sequenced *SnTox1*, *SnTox3* and *SnToxA* from 295, 203 and 178 individuals, respectively. The 295 *SnTox1* sequences collapsed into 18 nucleotide haplotypes, which encoded 14 different proteins (Figure 3A). The *SnTox1* haplotype network contains three loops, indicating that some of the alleles originated through intragenic recombination. *SnTox3* collapsed into 11 nucleotide haplotypes encoding six different proteins (Figure 3B). The 178 sequences from *SnToxA* collapsed into 17 unique haplotypes. Two *SnToxA* alleles with nonsense mutations were detected, one reported earlier in South Africa [6] and a second one found in New York. Excluding these two nonsense haplotypes, there were nine different SnToxA proteins (Figure 3C). Evidence for intragenic recombination was also found at the SnToxA locus.
**Effector sequence diversity does not correspond with the center of diversity for neutral markers in *P. nodorum***

The region with the highest diversity for neutral markers in *P. nodorum* corresponds with the acknowledged [17] origin of its wheat host in the ancient Fertile Crescent [18,19]. Allele diversity for each effector locus was compared among populations using rarefaction analysis across sample sizes ranging from 10-20 individuals. The rarefied number of alleles and 95% confidence intervals for each sample size are shown in Figure 4. For *SnToxA* the South African population clearly contained the highest number of alleles. For *SnToxI* both Europe and North America had the highest number of alleles. The average number of alleles found within each population was much lower for *SnTox3*, with North America and Australia showing the highest number of alleles. While the highest diversity based on neutral sequence loci and microsatellite loci was found in Iran [19], the highest diversity for each effector locus was always outside of Iran (Figure 4). The number of private alleles (alleles found in only one region) is summarized in Figure 4D. Private sequence alleles comprised 56% (10 out of 18) of the *SnToxI* haplotypes, 64% (7 out of 11) of *SnTox3* haplotypes and 76% (13 out of 17) of *SnToxA* haplotypes.

**Evidence for effector acquisition by horizontal transmission**

*P. nodorum* is one of nine members of a species complex infecting wheat and other grasses [19]. The presence or absence of the three NEs was determined in all nine species using both PCR and Southern hybridization. The three effectors were found only in *P. nodorum* and *Phaeosphaeria avenaria* f. sp. *tritici* 1 (*Pat1*). Figure 5 shows the coalescent multi-locus species trees (adapted from [19]) for the nine *Phaeosphaeria sp.* as well as the frequency of each effector within *P. nodorum* and *Pat1*. Only one *Pat1* isolate carried *SnTox3*. Based on the positions of the effectors in the maximum clade credibility tree (Figure 5) the most parsimonious explanation for occurrence of the effectors only in *P. nodorum* and *Pat1* is acquisition through HGT.

The number of sequence alleles for each effector in both species is shown in Figure 5. The 37 *SnTox1* sequences from *Pat1* collapsed into a single haplotype that was shared with both North American and Iranian *P. nodorum* isolates. The single *SnTox3* sequence found in *Pat1* was also the most frequent *P. nodorum* haplotype. Three *SnToxA* alleles were found among the 14 *Pat1* isolates sequenced. One *SnToxA* haplotype sequence was unique to *Pat1* but had the same
amino acid sequence as the most frequent effector allele in the SnToxA network. The other two Pat1 alleles were found in the two most common P. nodorum haplotypes. The overall diversity of Pat1 sequence alleles was far lower than found in P. nodorum, with all but one of the alleles identical in sequence to a common P. nodorum allele. This pattern is consistent with the hypothesis that Pat1 acquired all three NEs recently from P. nodorum.

DISCUSSION

Phylogenetic analysis of the Phaeosphaeria sp. complex revealed that only two out of nine closely related species carry the NEs SnTox1, SnTox3 and SnToxA. All three NEs are diverse at the amino acid level with a high proportion of population-specific sequence alleles. Rarefaction analysis indicated that the center of diversity for each NE did not correspond with previous population genetic studies that identified the highest levels of neutral diversity in Iran [19]. This population genetic study is only the second to compare neutral marker diversity with diversity at effector loci to infer the evolutionary history of effector genes. The earlier population study on the barley scald pathogen Rhynchosporium commune found that the highest diversity for neutral marker loci, including DNA sequences, RFLPs and SSRs corresponded with the highest diversity for the NIP1 effector gene in Scandinavia [11,20], indicating that NIP1 shared the same evolutionary history, and likely the same common ancestor, as the other R. commune genes. In contrast, the phylogenetic placement of the NEs in the Phaeosphaeria species complex and the lack of geographical correspondence between toxin diversity and neutral marker diversity provide strong evidence that the NEs of P. nodorum do not share the same evolutionary history as the neutral loci. Evidence for hybridization between P. nodorum and Pat1 was reported recently [19]. This finding is important because it provides a plausible explanation for the reduced NE sequence diversity found in Pat1. We propose that Pat1 acquired SnTox1, SnTox3 and SnToxA from P. nodorum through the formation of rare interspecific hybrids. Later we will discuss how this study and a growing number of other studies support the importance of horizontal gene transfer as a source of effectors in fungal pathogens.

Our analyses showed that the likelihood of an individual carrying one of the three NEs is population dependent. P. nodorum is a sexual pathogen with a large effective population size that
exhibits high levels of gene flow over continental scales (i.e. between Oregon, Texas and New York within North America) [18,21]. Based on gene flow estimates from neutral microsatellite loci, the expected frequencies of NEs in pathogen populations are not expected to differ among populations within a continent. Instead we found significant differences in NE frequencies among many field populations that did not differ for neutral markers (Supplementary Table S1). We believe that the differences in NE frequency among populations reflect differences in the frequencies of the corresponding host NE sensitivity genes among regions. As already shown for ToxA and Tsn1 [3,4,5] we hypothesize that the activity of Tox1 and Tox3 depends upon an interaction with a corresponding host sensitivity protein which is present in some wheat cultivars but absent in others. In the absence of a host sensitivity protein, pathogen strains encoding the corresponding NE have no fitness advantage and the NE gene is expected to be essentially neutral and subject to genetic drift. But in fields planted to wheat carrying a NE sensitivity gene, pathogen strains carrying the corresponding NE have a selective advantage [22] and the frequency of the NE will increase accordingly in these pathogen populations. This highlights one of the main challenges associated with controlling globally disseminated pathogens with a high capacity for gene flow. Susceptible cultivars planted within the dissemination range of pathogen populations carrying a NE could rapidly select for pathogen populations carrying the NE. Thus breeding efforts should be coordinated across large geographical regions to eliminate known susceptibility genes and reduce the frequencies of NEs at the continental scale. This type of effort is now underway in Australia to eliminate the Tsn1 gene that encodes susceptibility to ToxA [1,23].

A similar study that illustrated the dynamics of multiple effector loci in large natural populations was recently completed using the flax rust pathogen Melampsora lini. Thrall et al. [24] found dramatic fluctuations in the frequency of M. lini avirulence alleles across multiple loci and they were able to correlate these fluctuations with the susceptibility of the corresponding host populations. Their analyses showed how rapidly the genotype frequencies of host and pathogen can change in a gene-for-gene system experiencing antagonistic co-evolution [24]. While we did not measure the sensitivity of the host in each field population, the large differences in local NE frequency despite high levels of neutral gene flow suggest that there is very strong selection operating on NEs at the field level.

Despite significant differences in NE frequency between field populations, the distribution
of multi-effector genotypes within all but one of the 16 field populations did not differ from the expectation under random mating of neutral markers (Table 2). This indicates that selection has not favored particular combinations of effectors at the population level and that there is not a selective advantage associated with carrying an excess of NEs. This finding is also consistent with the hypothesis that host cultivar was the main determinant of NE frequency in these pathogen populations.

The historical classification of avirulence or toxin genes into fungal races is analogous to the multi-effector genotyping conducted in this study [16]. Traditionally, pathogens carrying particular combinations of avirulence or effector genes were classified into races. The “cost of carrying” was believed to drive the loss or alteration of the avirulence gene. Our finding of random associations among effector alleles within a population suggests that there is little fitness cost associated with carrying particular combinations of necrotrophic effectors. We hypothesize that the “carrying cost” of these effector genes is low in the absence of the corresponding host sensitivity allele. In the absence of the susceptibility allele we predict that these NE loci will be effectively neutral in the pathogen genome, and their rate of loss will be proportional to the strength of genetic drift.

The haplotype networks presented in Figure 3 show a prevalence of non-synonymous mutations. It was reported previously that SnToxA and SnTox1 are under significant positive selection [3,6], while SnTox3 did not show evidence of positive selection. The detection of diversifying selection operating on protein-coding genes with unknown function has become a powerful tool for identifying effectors in fungal genomes [25-28]. It is often assumed that pathogens undergo rapid diversification of their effector genes to avoid recognition by the plant immune system. Dodds et al. performed experiments that supported this assumption in the gene-for-gene interactions between M. lini and flax [13] and this assumption makes sense in the context of biotrophic pathogens that rely on avoidance of the plant immune system for survival. For necrotrophic pathogens it was hypothesized that the inverse process was operating, whereby diversification of the NE proteins resulted from the pathogen tracking changes in the host susceptibility alleles [6]. An alternative hypothesis to explain the diversity seen in NEs is that selection has favored mutant NE alleles that increase pathogen fitness through a quantitative increase in virulence. This hypothesis is supported by recent experimental studies showing that the most frequent ToxA protein variant is significantly more active against identical wheat Tsn1
(sensitivity) alleles with a corresponding increase in pathogen fitness [22]. The hypothesis of selection for increased toxin activity is very attractive because it provides an explanation for the quantitative variation in virulence often observed in gene-for-gene interactions. The haplotype networks of both SnTox1 and SnTox3 exhibit two or more frequent and widely distributed protein variants that differ at two or more amino acid positions. Under the hypothesis of selection for increased NE activity, the most common protein variants in these networks would be expected to induce significantly more necrosis than the less common protein variants in the network. Experimental testing of this hypothesis is now underway.

There has been a rapid expansion of literature describing fungal effectors as small, secreted proteins that interact with the host to suppress or alter the immune response (reviewed in [29], [9,30]. As more effectors have been identified and characterized an important question has become; what are the evolutionary origins of these effectors and how did pathogens acquire them? For some filamentous plant pathogens, acquisition of effectors appears to be through the horizontal transfer of conditionally dispensable chromosomes (CDC). This was experimentally demonstrated in the asexual species Fusarium oxysporum, where co-incubation of two unrelated strains resulted in the horizontal transfer of an entire chromosome containing a cluster of effector genes. Similarly, a host specific toxin conferring virulence on tomato is located on a CDC in Alternaria alternata that appears to be horizontally transmitted [31,32]. Asexual lineages of the rice blast pathogen Magnaporthe oryzae have experienced multiple translocations of the AVR-Pita gene between chromosomes [15]. These studies indicated that asexual pathogens can transmit effectors without undergoing meiotic recombination.

The rapid increase in genome sequences has led to the detection of HGT events involving single genes among more distantly related organisms. For example, genome sequencing of P. nodorum coupled with population genetic analyses revealed the horizontal transfer of SnToxA from P. nodorum to Pyrenophora triticirepentis [5]. Richards et al. used comparative genomics to demonstrate extensive cross-kingdom HGT from fungi to plant pathogenic oomycetes [33]. A recent study examining host specificity in Fusarium oxysporum f. sp. lini revealed the horizontal acquisition of pea pathogenicity genes (PEP) from Nectria haematococca [34]. In each of these cases HGT was detected by finding highly similar or identical protein sequences between divergent fungal lineages.

Our approach differs from these examples in that we used a phylogenetic analysis to reveal
the presence or absence of NE genes in a complex of closely related and sympatric pathogen species. We coupled the phylogenetic analysis with population genetic analyses that measured NE allelic diversity in a global sample of isolates and made a comparison to the allelic diversity observed at neutral loci. Taken together, our results indicate that the necrotrophic effectors of *P. nodorum* and *Pat1* were not acquired vertically through descent from a recent common ancestor.

In Figure 5 the most recent common ancestor of *Phaeosphaeria nodorum* is shared with six additional species. Within these seven species, necrotrophic effectors were found only in *P. nodorum* and *Pat1*. McDonald et al. provided evidence of hybridization between *P. nodorum* and *Pat1* [19] and *Pat1* was found to share all of its effector sequence alleles with *P. nodorum*. For *SnTox3* and *SnToxA* these sequence alleles were the most common ones found in *P. nodorum*. Based on the evidence of hybridization and the lower NE sequence diversity found in *Pat1*, we postulate that all three effectors in *Pat1* were acquired horizontally from *P. nodorum*.

The geographical region that harbored the highest sequence diversity was different for each effector gene. The highest diversity for *SnToxA* was found in South Africa, for *SnTox3* in North America and Australia, and for *SnTox1* in Europe. The finding of higher effector diversity in “New World” populations where wheat cultivation began only during the last few hundred years following the arrival of European colonists (South Africa ~350 years ago, Australia ~200 years ago and North America ~500 years ago) suggests that each of these effectors may have been acquired relatively recently and in different locations. None of the populations harboring the highest effector diversity overlapped with the hypothesized center of origin of *P. nodorum* in the ancient Fertile Crescent. The original source of the three effector genes remains unknown, but the ToxA protein has a domain consistent with a prokaryotic origin [5]. McDonald et al. [19] present evidence that *P. nodorum* existed as a species before the domestication of wheat, likely as a fungal endophyte on grasses. We propose that it was the acquisition of these necrotrophic effector genes, likely outside of the Fertile Crescent, which enabled *P. nodorum* to emerge as a specialized and damaging pathogen on wheat. Hybridization with its sister species *Pat1* enabled the horizontal transfer of these genes to a new fungal species, resulting in the emergence of another damaging, though closely related, pathogen. This study highlights the impact that horizontal gene transfer can have on the process of pathogen emergence in an agro-ecosystem.
ACKNOWLEDGEMENTS

The authors acknowledge the Genetic Diversity Center at ETH Zurich for providing sequencing facilities and the ETH Zurich for funding this research. We also thank the Australian Grains Research and Development Corporation and the USDA-NIFA AFRI Microbial Biology Program-Competitive grant # 2009-04265 for additional funding.
### Table 1. List of regional collections, sample sizes and host source for all isolates by species

<table>
<thead>
<tr>
<th>Regions</th>
<th>Year</th>
<th>Collector</th>
<th>Location</th>
<th>N</th>
<th>Host Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. nodorum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>2005</td>
<td>R. Sommerhalder</td>
<td>Golestan Province</td>
<td>27</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>M. Razavi</td>
<td>Golestan Province</td>
<td>13</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td>Central Asia</td>
<td>2005</td>
<td>R. Sommerhalder</td>
<td>Azerbaijan</td>
<td>7</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2003-04</td>
<td>H. Maraite, E. Duveiller</td>
<td>Kazakhstan</td>
<td>32</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>H. Maraite, E. Duveiller</td>
<td>Tajikistan</td>
<td>7</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>H. Maraite, E. Duveiller</td>
<td>Kyrgyzstan</td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>H. Maraite, E. Duveiller</td>
<td>Russia</td>
<td>9</td>
<td>Wheat / Durum Wheat</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>Syria</td>
<td>2</td>
<td>Wheat</td>
</tr>
<tr>
<td>China</td>
<td>2001</td>
<td>R. Wu</td>
<td>Fujian Province</td>
<td>115</td>
<td>Wheat</td>
</tr>
<tr>
<td>Europe</td>
<td>1992</td>
<td></td>
<td>England</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>E. Stukenbrock</td>
<td>Denmark</td>
<td>54</td>
<td>Triticale / Wheat</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>E. Blixt</td>
<td>Sweden</td>
<td>55</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>S. Keller</td>
<td>Switzerland</td>
<td>113</td>
<td>Wheat</td>
</tr>
<tr>
<td>United States</td>
<td>1991</td>
<td>G. Milus, M. Gray</td>
<td>Arkansas</td>
<td>94</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>G. Shaner</td>
<td>Indiana</td>
<td>8</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>L. Francl</td>
<td>North Dakota</td>
<td>12</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>J. Krupinsky</td>
<td>North Dakota</td>
<td>2</td>
<td>Agropyron sp.</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>J. Krupinsky</td>
<td>North Dakota</td>
<td>9</td>
<td>Barley</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>P. Lips</td>
<td>OH03 Sn1501</td>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>T. Friesen</td>
<td>North Dakota</td>
<td>6</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>G. Bergstrom</td>
<td>New York</td>
<td>46</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>M. Schmidt</td>
<td>Oregon</td>
<td>92</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>B. McDonald</td>
<td>Texas</td>
<td>92</td>
<td>Wheat</td>
</tr>
<tr>
<td>South Africa</td>
<td>1994</td>
<td>P. Crous</td>
<td>Southwestern Cape</td>
<td>74</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>Z. Pretorius</td>
<td>Western Cape</td>
<td>112</td>
<td>Wheat</td>
</tr>
<tr>
<td>Australia</td>
<td>2001</td>
<td>B. McDonald</td>
<td>Narrogin</td>
<td>73</td>
<td>Wheat</td>
</tr>
<tr>
<td><strong>Total P. nodorum</strong></td>
<td></td>
<td></td>
<td></td>
<td>1065</td>
<td></td>
</tr>
<tr>
<td><strong>Pat1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>1991</td>
<td>R. Clear</td>
<td>Alberta</td>
<td>65</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>R. Clear</td>
<td>Manitoba</td>
<td>7</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>R. Clear</td>
<td>Saskatchewan</td>
<td>36</td>
<td>Wheat (seed)</td>
</tr>
<tr>
<td>United States</td>
<td>1993</td>
<td>L. Francl</td>
<td>North Dakota</td>
<td>12</td>
<td>Wheat</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>T. Friesen</td>
<td>North Dakota</td>
<td>8</td>
<td>Wheat</td>
</tr>
<tr>
<td>Iran</td>
<td>2005</td>
<td>M. Razavi</td>
<td>Golestan Province</td>
<td>6</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>M. Razavi</td>
<td>Golestan Province</td>
<td>18</td>
<td>Wheat (ears)</td>
</tr>
<tr>
<td><strong>Total Pat1</strong></td>
<td></td>
<td></td>
<td></td>
<td>152</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Fisher’s exact tests for observed vs. expected genotypes by population

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Central Asia</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>England</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Denmark</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Sweden</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Switzerland</td>
<td>0.98</td>
</tr>
<tr>
<td>United States</td>
<td>Arkansas</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Indiana</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>North Dakota</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Oregon</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>0.26</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td>0.43</td>
</tr>
</tbody>
</table>

* Significant deviation from the number of observed vs. expected genotypes
Figure 1. Frequency of NEs is different between regions.
Bar plots showing the frequency of each NE in seven global regions. Bootstrap re-sampling with replacement was performed on each dataset to generate variance for the mean frequency of each necrotrophic effector. 95% confidence intervals are plotted for the bootstrap replicates.
Figure 2. The observed number of multi-effector genotypes matches the expected under an assumption of random mating.

The total number of isolates scored for the presence/absence of each locus is shown under N and alongside is the percent presence of each locus in the sample. The bar plots show the observed multi-effector genotypes compared to the expected number of genotypes calculated from the effector presence/absence data. The total number of individuals for which all three NE were scored is listed as N= in the top right corner of the bar plot.
Figure 3. Haplotype networks for each NE and sample sizes for sequence data by region.
A. The *SnTox1* haplotype network for the 18 sequence alleles. Circle sizes reflect the frequency of each haplotype and colors correspond to the global region where it was found. Haplotype numbers, denoted “H#” correspond to previously published sequences. Newly described haplotypes are marked with §. Non-synonymous mutations are marked with asterisks and synonymous mutations or mutations in the intron are labeled with “S” and “I” respectively. Open circles represent missing haplotypes in the network. Loops in the network indicate potential intra-locus recombination events. Unique features of the network are labeled with red lines or boxes and a short description. Black brackets indicate mutations that occurred within the same codon. B. The *SnTox3* haplotype network. C. The *SnToxA* haplotype network. The two nonsense haplotypes are labeled with a red hexagon.
Figure 4. The rarefied sequence allele count for each NE and number of private sequence alleles within each global region. A, B, C.
Chapter 3

The rarefied mean number of alleles for sample sizes ranging from 10 to 20 randomly chosen individuals from each population. Vertical bars represent the 95% confidence interval for the mean number of alleles. D. Pie charts showing the number and geographic location of private sequence alleles by NE locus.
Figure 5. Two out of nine *Phaeosphaeria sp.* carry NEs.

The multilocus maximum-clade credibility tree adapted from M.C. McDonald et al. [19] is shown. The tree is a coalescent tree with relative time on the x-axis and posterior probabilities for the branches shown. The two species within which NEs were found were *P. nodorum* and *Pat1*. The number of NE sequence alleles (N alleles) is shown above. The percent NE presence within the species is shown below.

Chapter 3
REFERENCES


19. McDonald MC, Razavi M, Friesen TL, Brunner PC, McDonald BA (In review)
Phylogenetic and population genetic analyses of *Phaeosphaeria nodorum* and its close relatives indicate cryptic species and an origin in the Fertile Crescent. Fungal Genet Biol.


Chapter 4:

*SnTox3 Acts in Effector Triggered Susceptibility to Induce Disease on Wheat Carrying the *Snn3* Gene*

Zhaohui Liu, Justin D. Faris, Richard P. Oliver, Kar-Chun Tan, Peter S. Solomon, Megan C. McDonald, Bruce A. McDonald, Alberto Nunez, Shunwen Lu, Jack B. Rasmussen, Timothy L. Friesen


* M. C. McDonald contributed the section on genetic diversity of *SnTox3*, Figure 10 and sequence information for Figure 11.
ABSTRACT

The necrotrophic fungus *Stagonospora nodorum* produces multiple proteinaceous host-selective toxins (HSTs) which act in effector triggered susceptibility. Here, we report the molecular cloning and functional characterization of the SnTox3-encoding gene, designated *SnTox3*, as well as the initial characterization of the SnTox3 protein. *SnTox3* is a 693 bp intron-free gene with little obvious homology to other known genes. The predicted immature SnTox3 protein is 25.8 kDa in size. A 20 amino acid signal sequence as well as a possible pro sequence are predicted. Six cysteine residues are predicted to form disulfide bonds and are shown to be important for SnTox3 activity. Using heterologous expression in *Pichia pastoris* and transformation into an avirulent *S. nodorum* isolate, we show that *SnTox3* encodes the SnTox3 protein and that *SnTox3* interacts with the wheat susceptibility gene *Snn3*. In addition, the avirulent *S. nodorum* isolate transformed with *SnTox3* was virulent on host lines expressing the *Snn3* gene. *SnTox3*-disrupted mutants were deficient in the production of SnTox3 and avirulent on the *Snn3* differential wheat line BG220. An analysis of genetic diversity revealed that *SnTox3* is present in 60.1% of a worldwide collection of 923 isolates and occurs as eleven nucleotide haplotypes resulting in four amino acid haplotypes. The cloning of *SnTox3* provides a fundamental tool for the investigation of the *S. nodorum*-wheat interaction, as well as vital information for the general characterization of necrotroph-plant interactions.
INTRODUCTION

Diseases caused by necrotrophic pathogens are believed to be distinctly different from those caused by biotrophs where the pathogen requires a living host to grow and sporulate. The interaction between biotrophic pathogens and their hosts is controlled via the interaction of pathogen effector molecules which, if recognized by the corresponding resistance gene product, results in localized programmed cell death (PCD) and activation of plant defense which leads to a resistant or incompatible interaction. This interaction has been referred to as effector triggered immunity (ETI) [1]. In contrast, necrotrophic pathogens induce host cell death during infection. Host-selective toxins (HSTs) are molecules produced by some necrotrophic fungi that induce a necrotic reaction and promote a susceptible/compatible interaction in the host [2]. Most HSTs are small secondary metabolites and their production is under the control of complex genetic and enzymatic pathways [3]. But several proteinaceous HSTs have been identified more recently [4-8].

*Stagonospora nodorum*, the causal agent of Stagonospora nodorum blotch (SNB), is a necrotrophic fungus that causes major yield losses worldwide by infecting the leaves and glumes of wheat [9-11]. SNB resistance is quantitatively inherited with little obvious evidence of an isolate-specific relationship between the pathogen and the host [12]. Early research on SNB indicated that resistance was mostly controlled by multiple genes located on chromosomes throughout the genome, with each showing minor association with disease [13, 14].

We recently identified four proteinaceous HSTs from *S. nodorum* and showed they play an important role in disease development by interacting with corresponding host sensitivity/susceptibility gene products [14]. One of these HSTs, ToxA, is a small proteinaceous HST produced by approximately 40% of *S. nodorum* isolates worldwide and is present in multiple forms in *S. nodorum* [15]. ToxA interacts with the product of the wheat gene *Tsn1* [16] and fungal strains carrying ToxA are significantly more virulent on wheat lines carrying *Tsn1*. ToxA is also expressed by the tan spot pathogen *Pyrenophora tritici-repentis* [4] and induces necrosis in a process which involves uptake into the host cytoplasm, translocation to the chloroplast and disruption of photosynthesis [17].
SnTox1 and SnTox2 also appear to be small proteins, though their encoding genes have not yet been cloned [18, 19]. Preparations of SnTox1 and SnTox2 induce necrosis on wheat lines carrying the toxin sensitivity genes *Snn1* and *Snn2*, and fungal isolates producing these toxins are more virulent on wheat lines carrying *Snn1* and *Snn2*, respectively. The presence of these toxins also varies between isolates and the strain Sn79-1087 appears to produce no toxins at all and is avirulent on all tested wheat lines [15].

We previously characterized the reaction of a necrosis-inducing activity that defines a fourth host locus called *Snn3*. Culture filtrate fractions of *S. nodorum* strain Sn1501 were used to define a QTL on chromosome 5BS in a cross between wheat lines BR34 and Grandin. SnTox3 is proteinaceous and was reported to be 10-30 kDa in size [20]. QTL analysis showed that the SnTox3-*Snn3* interaction contributed significantly to disease development [20].

Here we report the molecular cloning of the SnTox3-encoding gene *SnTox3*, and show that SnTox3 plays an important role in disease by interacting directly or indirectly with the product of the wheat sensitivity gene *Snn3*. The cloning of *SnTox3* provides a critical tool for the molecular and biochemical characterization of the host-pathogen interaction in the *S. nodorum*-wheat pathosystem.

**MATERIALS AND METHODS**

**Plant materials**

A population of recombinant inbred (RI) lines derived from a cross between the Brazilian hard red spring wheat (HRSW) breeding line BR34 and the North Dakota HRSW variety ‘Grandin’ was developed and provided by Dr. James A. Anderson, University of Minnesota. This population, consisting of 118 F$_{7:9}$ lines, segregates for both toxin and disease reaction to *S. nodorum*. Parental line BR34 is toxin insensitive and disease resistant and Grandin is toxin sensitive and disease susceptible. Several BG RI lines have been selected as toxin differential lines including BG261 (SnToxA sensitive only, [15]), BG223 (SnTox2 sensitive only, [19]) and BG220 (SnTox3 sensitive only, [20]). The entire population, with parental lines and selected toxin differential lines were used in this study for testing toxin production and/or virulence of different genetically modified *P. pastoris* and *S. nodorum* fungal strains along with wild type
strains or isolates. All plants were grown in plastic cones containing SB100 professional grow mix (Sungrow Horticulture, Dellevue, WA) in the greenhouse at an average temperature of 21°C with a 14-h photoperiod.

**Toxin bioassay and fungal inoculation**

Toxin bioassay and fungal inoculation were conducted at the two- to three-leaf stage following the procedures previously described [15]. For toxin bioassays, the necrotic reaction was recorded as either sensitive or insensitive based on the presence or absence of necrosis in the infiltrated area. For fungal inoculation, disease ratings followed the 0-5 disease scale developed by Liu et al. [21] with 0 being highly resistant and 5 highly susceptible. Two replicates of at least three plants of each line were evaluated for toxin sensitivity and three replicates of three plants of each line were evaluated for fungal inoculation. An overall average from the three replicates was calculated and used in statistical analysis.

**Purification of SnTox3**

SnTox3 was partially purified and characterized as described in Friesen et al. [20] using *S. nodorum* isolate Sn1501. Recently the highly virulent North Dakota *S. nodorum* isolate Sn4, was identified as a better toxin producer, therefore this isolate was used for production of SnTox3. Culture filtrate production and partial purification was done as described by Friesen et al. [20] with some modifications. Briefly, initial purification steps were conducted using the ÄKTA prime plus (GE Healthcare, Piscataway, NJ) liquid chromatography system. In order to obtain enough protein for mass spectrometry, ~300 ml of 3 week old fungal culture filtrate was dialyzed overnight against water using a 3.5 kDa molecular weight cutoff dialysis tubing (Fisher Scientific, Pittsburgh, PA). The dialyzed culture filtrate was loaded onto a HiPrep SPXL 16/10 cation exchange column (GE Healthcare Piscataway, NJ) after pre-equilibration with the same buffer used in dialysis. After loading the sample, the column was washed with 50 ml of 20 mM sodium acetate buffer pH 5.0 followed by a gradient elution of 0-300 mM sodium chloride plus 20 mM sodium acetate pH 5.0 at a flow rate of 5.0 ml/min over 20 min. The 5 ml fractions were collected and tested on the SnTox3 differential line BG220. The most active fraction was
used for size exclusion chromatography using a HiLoad 16/60 Superdex 30 prep-grade gel filtration column. The sample was injected using a 5 ml loop and 20 mM sodium acetate, 50 mM NaCl, pH 5.0 running buffer was used with a flow rate of 1 ml/min and a fraction size of 5 ml. The size-based fraction containing SnTox3 was further separated using SDS-PAGE gel electrophoresis. The sample was loaded into a precast 16.5 % tris-tricine polyacrylamide gel (Bio-Rad, Hercules, CA) and subjected to electrophoresis in a Bio-Rad Mini PROTEAN 3 system (Bio-Rad, Hercules, CA) using the buffer system of Shägger and von Jagow [42]. The gel was stained with Coomassie Blue solution (0.2% Coomassie Blue R250, 7.5% acetic acid and 50% ethanol) to visualize the protein bands. Based on a pre-stained protein standard (Bio-Rad, Hercules, CA), the bands within the expected size range (10-30 kDa) were excised individually from the gel and subjected to mass spectrometric analysis.

**Mass spectrometry and identification of the SnTox3-encoding gene**

Matrix-assisted laser desorption/ionization mass spectrometry with automated tandem time of flight fragmentation of selected ions (MALDI-TOF/TOF) of trypsin digested proteins were acquired with a 4700 Proteomics Analyzer mass spectrometer (Applied Biosystems, Framingham, MA) in the positive reflectron mode. Spectra were obtained by averaging 1000 and 2500 acquired spectra in the MS and MS/MS mode, respectively. Post source decay fragmentation MS/MS spectra of selective peptides were obtained with 1 keV acceleration voltage. Conversion of TOF to mass (Da) for the monoisotopic ions, [M + H]⁺, was based on calibration of the instrument with a peptide standard calibration kit (Applied Biosystems). The MS and MS/MS spectra were combined and searched against the sequence of the *S. nodorum* protein database (http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum) using the Mascot (Matrix Science, Inc. Boston, MA) search engine through GPS Explorer Software (Applied Biosystems) with a 50 ppm and 0.1 Da error tolerance for MS and MS/MS spectra, respectively, one missed trypsin cleavage allowance, oxidation of methionine, and carbamidomethyl derivatization of reduced cysteine as a variable modification. The signal to noise ratio for peak filtering was set to 10 for MS and 20 for MS/MS. The threshold for proteins from database searches (MS + MS/MS and MS/MS) was set within a ≥95% confidence interval.
Protein bands likely containing SnTox3 were subjected to digestion using Trypsin Gold, mass spectrometry grade (Promega Co., Madison, WI), following manufacture procedures and ZipTips protocols for sample cleaning and spotting in a MALDI plate. Alpha-cyano-4-hydroxycinnamic acid was used as a matrix for mass spectrometry analysis. Using the S. nodorum genome sequence (http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum), each identified protein was used to identify the genomic DNA sequence for the encoding gene, along with its 5’ and 3’ flanking region. Primers for identified genes (SNOG_08981 and SNOG_16063): 8981g1F with 8981g1R, and 16063g1F with 16063g1R (Supporting Table 1) were designed using the web-based program Primer3 [43] (http://frodo.wi.mit.edu/). PCR was performed with an annealing temperature of 60 °C to verify the presence or absence of each gene in both virulent and avirulent isolates. Candidate proteins were subjected to further validation if they were found to be absent in the avirulent isolate Sn79-1087 and present in both SN15 and Sn4.

Fungal isolates and Southern analysis of the occurrence of the SnTox3-encoding gene

Nine S. nodorum isolates, collected from different locations, were used to investigate the occurrence of the SnTox3-producing gene. The origin and toxin production of those isolates are listed in Table 1. Three related fungal species isolates were also included in this study including: P. teres Pt-15A (obtained from B. Steffenson Univ. of MN St. Paul, MN USA), P. tritici-repentis Ptr86-124, (obtained from L. Lamari Univ. of Manitoba, Winnipeg, Canada), and Mycosphaerella graminicola Str975, (obtained from C. Hollingsworth, Univ. of MN Crookston, MN USA). Fungal DNA extractions were carried out as previously described [44, 45]. A total of 5µg of genomic DNA of each isolate was completely digested with the restriction enzyme EcoRI at 37 °C. DNA blotting, Southern probe preparation, hybridization, and signal detection were performed as described by Faris et al [46]. The full length SnTox3 gene, a 693bp fragment, was PCR amplified from SN15 with primer pair 8981cF and 8981cR (see Supporting Table 1 for sequence) and used as a probe.
**SnTox3 full length cDNA amplification and 5’ and 3’ RACE**

Total RNA was prepared from Sn4 using TRIzol Reagents (Invitrogen, Carlsbad, CA) according to the manufacturer’s protocol and treated with RNase-free DNase I (Promega Madison, WI). The one step RT PCR kit (Invitrogen Carlsbad, CA) containing reverse-transcriptase and a DNA Taq polymerase mix was used to amplify a full length cDNA of SnTox3 using primers 8981cF and 8981cR (Supporting Table 1). To make sure there was no DNA contamination, one control PCR reaction was set up with only Taq polymerase (New England Biolabs Ipswich, MA). The amplified fragment was excised from the agarose gel, gel-purified and cloned into the TopoTA vector (Invitrogen, Carlsbad CA). To obtain the 5’ and 3’ end of the cDNA, 5’ and 3’ RACE were performed using the Smart RACE cDNA amplification kit (Stratagene, La Jolla, CA) according to the instructions in the user’s manual. The amplified 5’ and 3’ RACE fragments were gel-purified and cloned into the TopoTA vector. All of the plasmids were prepared with the plasmid DNA miniprep kit (Qiagen, Valencia, CA) and sequenced from both directions with M13 forward and reverse primers. The sequences were then used to assemble the full length cDNA, including the 5’ and 3’ untranslated region (UTR) based on the genomic sequence.

The search of DNA sequence with similarity to *SNOG_08981* cDNA was performed using BLASTN [47] and BLASTX [48] against the public NCBI non-redundant (NR) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the annotated *S. nodorum* database at the Broad institute website. (http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum.2/Blast.html).

**RNA isolation and quantitative real time Q-PCR**

To investigate the transcriptional expression pattern of *SNOG_08981* during infection, SN15 conidia were inoculated onto the susceptible wheat variety ‘Amery’ and total RNA was isolated from inoculated leaves at 3, 5, 7 and 10 days after inoculation. RNA isolation, cDNA synthesis and gene transcript abundance analysis were performed as previously described [49]. Intron-spanning primers ActinF and ActinR (Supporting Table 1) designed to amplify actin (*Act1*; Genbank accession number EAT90788) were used to check all cDNA samples [50] and
were shown to be free of genomic DNA (data not shown). \textit{SnTox3} expression was examined using Tox3qPCRF and Tox3qPCRR primers (Supporting Table 1). \textit{Act1} was used as a constitutively expressed control for the normalisation of \textit{SnTox3} expression using ActinqPCRf and ActinqPCRR primers (Supporting Table 1). All reactions were performed in technical duplicates from pooled biological triplicates.

**Heterologous expression of \textit{SnTox3} in \textit{Pichia pastoris}**

The yeast strain \textit{Pichia pastoris} X33 and vector pGAPZA for gene expression were provided in the commercial kit developed for constitutive expression and purification of recombinant proteins (Invitrogen, Carlsbad CA). The pGAPZA vector is \(~2.9\) kb and contains the Zeocin resistance gene (\textit{Sh ble}) as a selectable marker for use in both \textit{E. coli} and yeast. The full length of \textit{SnTox3} cDNA in TopoTA was re-amplified with primer pairs 8981cF\textunderscore EcoRI and 8981cR\textunderscore ApaI (see Supporting Table 1) which contain EcoRI and ApaI restriction sites, respectively. The re-amplified \textit{SnTox3} cDNA with restriction sites was cloned into the TopoTA vector. After sequencing confirmation for sequence identity, the full length \textit{SnTox3} cDNA was released from the TopoTA vector by digestion with EcoRI and ApaI and directionally cloned into the pGAPZA vector. The expression constructs were sequenced with the primer pGAPF and \textit{3’Aox1} (see Supporting Table 1) for checking the gene identity and the translation reading frame. At least \(5\ \mu\text{g}\) of the construct DNA linearized with \textit{Bsr}DI was used for yeast transformation. The preparation and transformation of competent \textit{P. pastoris} cells were done using the \textit{Pichia} EasyComp kit (Invitrogen, Carlsbad, CA) following the steps described in the user manual.

Four different positive \textit{Pichia} clones were selected from the transformation plate to test toxin expression. Each clone was picked and cultured in \(1\) ml YPD liquid medium (\(1\%\) yeast extract, \(2\%\) peptone and \(2\%\) dextrose) in \(15\) ml tubes. Since SnTox3 are expected to be secreted, \textit{P. pastoris} cells were centrifuged at \(6000\ \times\text{g}\) for \(5\) min and the supernatant was used for testing toxin activity.
Transformation and disruption of the $SnTox3$ gene

$SnTox3$ expression vector construction: The previously published vector, pDAN [15], containing a cpc-1::hygromycin-resistance gene cassette was used to carry the $SnTox3$ gene for transformation into the avirulent isolate Sn79-1087. The primer pair 8981g1F_XbaI and 8981g1R_XbaI (Supporting Table 1) containing an XbaI restriction site was used to amplify a ~1.2 kb genomic region (Figure 2B) of $SnTox3$ which contained a putative $SnTox3$ promoter region and terminator. This ~1.2 kb fragment was first cloned into the TopoTA vector to form the plasmid pTopoSnTox3, followed by sequence confirmation of the XbaI restriction site and $SnTox3$ sequence using M13 forward and reverse primers. The remaining steps to put $SnTox3$ into the pDAN vector followed the same procedure as that used for $SnToxA$[15]. The resulting plasmid was screened and verified using the primer pairs 8981g1F_XbaI and 8981g1R_XbaI (Supporting Table 1).

$SnTox3$ disruption vector construction: $SnTox3$ has a unique SalI restriction site located near the middle of the ORF that provides a site for the insertion of the hygromycin-resistance gene to form the gene disruption vector. The plasmid pTopoSnTox3 was cut open via SalI digestion and re-ligated with the hygromycin resistance gene ($HYG$) cassette which had been released from the pLP605KO vector [51] using the restriction enzyme XhoI. The recombinant DNA was transformed into DH5α $E. coli$ competent cells (Invitrogen, Carlsbad, CA) and screened with PCR using the primer pair 8981g1F_XbaI and HY (Supporting Table 1). In the recombinant vector, the $SnTox3$ fragment was therefore separated by the $HYG$ gene cassette as ~700 bp of 5' region and ~500 bp of 3' region (Figure 6B) for homologous recombination. The whole construct was released from the TopoTA backbone as a ~3.8 kb fragment using XbaI before transformation into Sn1501.

Fungal protoplasting and transformation: The fungal protoplasting and PEG-mediated transformation methods described by Solomon et al. [45] were used to introduce the expression vector into the avirulent isolate Sn79-1087 and the disruption vector into the virulent isolate Sn1501. The regenerated colonies were screened for gene integration using the corresponding primer pair (8981g1F and 8981g1R for $SnTox3$ transformation in Sn79-1087, 8981g0F and HY for $SnTox3$ disruption in Sn1501). RT-PCR with primers 8981cF and 8981cR and Southern
analysis with full length SnTox3 was performed on the selected transformants to validate SnTox3 transformation and disruption. The S. nodorum actin gene primer [50] was used to amplify the actin gene as an internal control in RT-PCR.

**Molecular mapping and QTL analysis**

A wheat genetic map containing 787 DNA markers has been developed in the BG population [52, 53]. The marker density of the maps makes this population ideal for mapping genes and doing QTL analysis of the S. nodorum-wheat interaction. In this population, sensitivity to SnToxA, SnTox2 and SnTox3 have been mapped to wheat chromosome arms 5BL [15, 22], 2DS [19] and 5BS [20], respectively. To verify that SnTox3 has been produced by genetically modified yeast strains and S. nodorum isolates the culture filtrate or partially purified toxin prep from the transformed isolates were infiltrated on the 118 recombinant inbred lines to map the sensitivity loci and verify that they correlate with the previously mapped toxin sensitivity loci. One Sn79-1087 SnTox3 transformant and two Sn1501 SnTox3 knockout transformants were also inoculated onto this population using wild type and ectopic transformants as controls. The molecular mapping and QTL analysis in the BG population was performed as previously described [20].

**Prediction of signal peptides and disulfide bonds in SnTox3 protein**

Two web-based programs SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) and WolfPsort (http://wolfpsort.org/) were used to predict the cellular location of SnTox3 and determine the presence and length of a signal peptide. The protein sequence was submitted to the website PredictProtein to run DISULFIND [54] to identify the disulfide bond prediction. To investigate if the predicted disulfide bonds were important for toxin activity, the P. pastoris culture containing SnTox3 was treated with Dithiothreitol (DTT) (Fisher Scientific, Pittsburgh, PA) at two different concentrations including 5 mM and 10 mM, with water added as a negative control. The treated culture filtrates were assayed on line BG220 along with an untreated negative control.
Genetic diversity analysis

PCR amplification was used to screen 923 isolates for the presence or absence of \textit{SnTox3}. PCR primer pairs, 8981cF-R and 8981g1F-R were both used to confirm \textit{SnTox3} presence. The annealing temperature of all PCRs was 55°C. Sequencing of the PCR products was performed with the same primers using an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA). Alignment of forward and reverse sequences for each isolate was performed in SeqScape software V2.5 (Applied Biosystems, Foster City, CA). The same software was used for translation and identification of protein haplotypes. Gene diversity (Pi) was measured using the nucleotide alignment software DNAsp [55].

\textit{S. nodorum} isolates were obtained from eight major geographical regions; North America (358), Central America (41), Europe (192), Iran (47), Australia (57), Central Asia (49), East Asia (107), and South Africa (74). The location, year of collection and frequency of \textit{SnTox3} deletions is summarized for each population in Supporting Table 2. Each of these populations was characterized previously for \textit{SnToxA} [39]. Isolation procedures and DNA extractions were performed according to the procedures described previously [56]. A total of 250ng of genomic DNA was spotted as per manufacturer instructions using the BioDot Microfiltration Apparatus (BioRAD, Hercules, CA). Hybridization and signal detection was performed as previously described [46].

The other three protein variants of SnTox3 were amplified from a representative isolate and transformed into \textit{P. pastoris} following the same procedure described earlier for SnTox3 in isolate Sn4. In order to test the effectiveness of each toxin variant, culture filtrates were produced and assayed on the SnTox3 differential line BG220.

\section*{RESULTS}

\subsection*{Identification of the SnTox3-encoding gene}

Progeny line BG220 from the BR34 x Grandin recombinant inbred population was previously identified as a SnTox3 differential line [20] and was used in identifying and characterizing this toxin. Ion exchange and size exclusion chromatography was used for the initial purification of BG220-reactive material as previously described [20]. A large peak and
three smaller peaks were observed in size exclusion chromatography (Supporting Figure 1); the large peak was associated with SnTox3 activity and was analyzed by SDS-PAGE (Figure 1). The protein gel was visualized as containing a light band and a dark band within the size range of 6.5 to 32.0 kDa. The light band was estimated at 18.0 kDa and the darker band was estimated at 13.0 kDa. Mass spectra obtained from the two bands was used to search the *S. nodorum* protein database [11] leading to the identification of two predicted proteins, SNOG_08981 and SNOG_16063, corresponding to the light and dark protein bands in the gel, respectively.

PCR amplification revealed that *SNOG_16063* was present in all strains tested whereas *SNOG_08981* was absent in the avirulent isolate Sn79-1087 (Supporting Figure 2). Since Sn79-1087 did not produce BG220-reactive material, this indicated that *SNOG_08981* could correspond to the SnTox3 encoding gene. Further analysis of the MALDI-TOF/TOF MS and MS/MS spectra of the trypsinized peptides identified a total of 44% of the predicted immature SNOG_08981 protein sequence (see Supporting Figure 3 for mascot search results of SNOG_08981). This included the majority of the C-terminal region but did not include the first 20 amino acids predicted to be the signal sequence nor did it account for amino acids 21-72 of the N-terminus (Figure 2).

**Genomic context of SNOG_08981**

The predicted *SNOG_08981* gene was located on the end of supercontig14 [11] which is ~1.2 Mb in length and is currently estimated to harbor 455 genes (Figure 2A, [11], Hane and Oliver, unpublished). Between *SNOG_08981* and the end of the contig is a ~10 kb region of AT-rich sequence. No additional genes are predicted in this AT-rich region except a polyprotein and a putative transposase (Figure 2A). Some of the AT-rich sequence is annotated as part of the Elsa family of degraded retrotransposons [11]. Of the four genes upstream of *SNOG_08981*, two of them are functionally conserved, and the other two are hypothetical genes based on the genome sequence annotation [11].
**SNOG_08981 contains no introns and has no identified homologs in other fungi**

The full-length cDNA of the *SNOG_08981* gene was identified by reverse-transcription and rapid amplification of cDNA ends (RACE). The gene comprises an 80 bp 5’ UTR, a 146 bp 3’ UTR and a single exon of 693-nucleotides encoding a protein with 230 amino acids (Figure 2A and B). A putative TATA box was located 95 bp upstream of the start codon whereas no obvious CAAT box was identified (Figure 2B).

The *SNOG_8981* cDNA sequence identified was different from that found in the genome database (http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum). The genome sequence archives at this region were retrieved to perform re-assembly and the difference was shown to be due to an error in the sequence assembly. The reassembled sequence fully matched our cDNA sequence. The corrected *SNOG_08981* sequence was submitted to NCBI as Genbank accession number FJ823644.

Using the *SNOG_08981* ORF sequence in BLASTN or BLASTX searches of the NCBI nr database, no significant similarity to sequences from other organisms was detected. The *S. nodorum* hypothetical protein, SNOG_10812, was the only hit in a BLASTX search with a score at 67.0 bits and an E-value at 1e-09. Other hits were obtained but were identified at E-values greater than 0.1.

**Southern analysis of SNOG_08981 in S. nodorum strains**

Using the full-length gene as a probe, Southern analysis of genomic DNA from a small sub-set of *S. nodorum* isolates indicated that the *SNOG_08981* gene was present as a single copy gene in six of the eight *S. nodorum* virulent isolates, but absent in the other two virulent isolates, as well as the avirulent isolate (Figure 3). All *SNOG_08981*-containing isolates showed the same size hybridizing band as SN15 except for the Danish isolate SnCP2052 where a larger band, suggests there is sequence variation flanking *SNOG_08981* in this isolate (Figure 3).

The detectable toxin activity from different *S. nodorum* isolates was summarized in Table 1. The six *S. nodorum* isolates found containing *SNOG_08981* were also known to produce SnTox3 in culture (Table 1). The *SNOG_08981* probe did not hybridize to the virulent isolates Sn2000 [18, 21], or Sn50; accordingly, culture filtrates from these strains were also unable to
induce necrosis on BG220 indicating Sn2000 and Sn50 do not produce SnTox3 (Table 1). Sn2000 has been used in QTL analysis in the BG population and no QTL associated with disease susceptibility was identified at the Sn3 locus [21], [22]. However, these two isolates produce SnToxA, SnTox1 and/or other toxins, which make them virulent on wheat lines carrying either Tsn1 or Snn1 or both [15, 18, 21] (unpublished data).

**Gene expression of SNOG_08981 is maximized at the early stages of infection**

SNOG_08981 expression was examined at three, five, seven, and ten days post-infection on the SnTox3 sensitive wheat c.v. ‘Amery’ using quantitative real time RT-PCR. The transcription levels of SNOG_08981 were maximal at three days post infection coinciding with the hyphal proliferation and the onset of lesion development on the leaf (Figure 4). Transcription levels were significantly reduced from five days post-infection which coincided with a transition from hyphal proliferation to asexual sporulation. This observation was essentially confirmed with microarray analysis of the S. nodorum transcriptome at the same post-infection time points (data not shown).

**Heterologous expression of SNOG_08981 in Pichia pastoris**

A yeast expression construct containing the full-length cDNA of SNOG_08981 was transformed into P. pastoris. The culture filtrate from positive yeast clones were harvested and used to infiltrate the toxin differential lines BG261 (SnToxA-sensitive; Tsn1), BG223 (SnTox2-sensitive; Snn2), and BG220 (SnTox3-sensitive; Snn3) as well as parental lines BR34 and Grandin [15, 19, 20]. The culture filtrates produced a necrotic reaction on Grandin (Snn3) and BG220 (Snn3), but not on BR34 (snn3), BG261 (snn3) and BG223 (snn3) (Figure 5). The culture filtrate from the yeast transformed with an empty vector (negative control) did not induce necrosis on any of the differential lines (Supporting Figure 4). All BG lines sensitive to partially purified SnTox3 [20] were also sensitive to SNOG_08981 transformed yeast culture filtrates, and thereby the sensitivity was mapped to the Snn3 locus on wheat chromosome arm 5BS in the BG population as previously described [20]. This strongly indicates that SNOG_08981 is the SnTox3-encoding gene and therefore we designated it as SnTox3.
**SnTox3 renders an avirulent *S. nodorum* strain virulent in host lines containing Snn3**

Culture filtrate of *S. nodorum* isolate Sn79-1087, which was isolated from a wild grass, produces no necrosis on any wheat line we have studied. Furthermore it is completely avirulent to all wheat lines that have been tested. Expression of ToxA in Sn79-1087 rendered it pathogenic on BG261 and made it capable of producing active ToxA [15]. A fungal transformation construct containing the genomic region of the *SnTox3* gene and the hygromycin B resistance gene was transformed into Sn79-1087. Two putative transformants, Sn79+SnTox3A and Sn79+SnTox3B were selected and analyzed for *SnTox3* gene integration, toxin production, and virulence change compared to the wild type. Southern analysis using the *SnTox3* full-length cDNA as a probe showed that the Sn79+SnTox3A and Sn79+SnTox3B transformants harbored at least two and one copies of *SnTox3*, respectively. Both transformants contained the same size restriction fragment (~ 7.0 kb) suggesting that one of the *SnTox3* integrations was present in the same place in the Sn79-1087 genome (Figure 6A). RT-PCR also confirmed the transcription activity of *SnTox3* in the two transformants (Figure 6D).

Culture filtrates of Sn79-1087, Sn79+SnTox3A and Sn79+SnTox3B were infiltrated into BG220. Necrosis was not induced using the wild type Sn79-1087 strain, whereas the two transformants caused a strong necrotic reaction (Figure 7A). Culture filtrates from the two transformants were also infiltrated into all 118 RI lines in the BG mapping population to verify that sensitivity mapped to the Snn3 locus. As expected, sensitivity to culture filtrates of both transformants in the BG population was conferred by a single gene and mapped to chromosome arm 5BS at the Snn3 locus [20]. This provides additional evidence that *SNOG_08981* is responsible for the production of SnTox3.

To test if the addition of *SnTox3* changes the specificity of Sn79-1087, the conidia from strain Sn79+Tox3A were harvested and used to inoculate the BG differential lines as well as BR34 and Grandin. This isolate caused typical *S. nodorum* lesions on Grandin and BG220 which contain Snn3 (Figure 7B). No visible lesions were observed on the snn3 lines BR34, BG261 and BG223. This isolate was further inoculated onto the entire BG population and QTL analysis showed that the Snn3 locus explained the majority of the variation in susceptibility (59%) to the fungus and no other QTL was identified (Figure 8). From these data, we can conclude that the
addition of the *SnTox3* gene is sufficient to change an avirulent *S. nodorum* strain into a virulent strain by inducing effector triggered susceptibility on host lines containing *Snn3*.

**SnTox3**-disrupted virulent isolates are reduced in virulence specific to *Snn3* wheat lines.

*SnTox3* was disrupted in Sn1501, a strain which is deficient in the *SnToxA* gene but contains *SnTox3* (Figure 3) and produces SnTox3 (Table 1). As *SnTox3* is closely linked to a large region of retrotransposon-related elements (Figure 2A), a gene replacement knockout strategy was unavailable. Instead we inserted the hygromycin B resistance gene (*HYG*) cassette into the *SnTox3* coding region to construct a gene disruption vector (see Material and Methods for vector construction). In this vector, the HYG cassette was flanked by ~700 bp of *SnTox3* 5’ region at one end and ~500 bp of *SnTox3* 3’ region at the other (Figure 6B). Two sets of primers: 8981g0F and HY, 8981g1F and 8981g1R were used to screen and verify transformants (see Figure 6B for primer locations and Supporting Table 1 for sequence). Three out of 68 resulting strains contained the disrupted *SnTox3* cassette, whereas the remaining 65 transformants harbored ectopic integrations. Two of the *SnTox3*-disrupted strains, designated Sn1501ΔTox3A and Sn1501ΔTox3B, and one ectopic strain, designated Sn1501Ect, were selected for further analysis. PCR, Southern blotting and RT-PCR confirmed the gene-disruption (Figure 6A, 6C and 6D).

The culture filtrates of four strains were tested on the SnTox3 differential line BG220 and the results showed that the wild type (Sn1501) and ectopic transformants (Sn1501Ect) were able to produce SnTox3. The two strains harboring the mutated *SnTox3* did not produce active SnTox3 *in vitro*, suggesting that SnTox3 was non-functional in these two mutants (Figure 7C). Sn1501 wild type, Sn1501ΔSnTox3A, Sn1501ΔSnTox3B, and Sn1501Ect were inoculated onto the SnTox3 differential line BG220. Sn1501 and Sn1501Ect induced typical lesions on BG220, whereas only small white flecks were present without any visible lesions on the leaves inoculated with Sn1501ΔSnTox3A and Sn1501ΔSnTox3B, indicating the mutants were avirulent on BG220 (Figure 7D). Compared to the wild type, the two *SnTox3*-disrupted strains also showed decreased virulence toward the parental line Grandin which also contains *Snn3*; however, *SnTox3* disruption
did not change the reaction towards parental line BR34 or the SnToxA and SnTox2 differential lines BG261 and BG223 respectively, which do not harbor Snn3 (Supporting Figure 5).

The four strains Sn1501, Sn1501ΔSnTox3A, Sn1501ΔSnTox3B, and Sn1501Ect were also inoculated onto the entire BG population in order to quantify the effect of the SnTox3-Snn3 interaction using QTL analysis. For wild type Sn1501, as previously reported [20], significant QTLs were detected on the distal end of chromosome 2DS and 5BS which corresponded to Snn2 (SnTox2 sensitivity) and Snn3, respectively. The Snn2 and Snn3 loci explained 46% and 10% of the variation in disease, respectively. The same two QTLs with similar effects (46% and 13%) were found to be associated with susceptibility to the Sn1501Ect strain. For the two SnTox3-disrupted strains Sn1501ΔSnTox3A and Sn1501ΔSnTox3B, QTL analysis showed that the Snn2 locus explained 60 and 64% of the variation, respectively, and the effects of the Snn3 locus were not significant (Figure 8) showing that SnTox3 specifically interacts with Snn3 and the loss of SnTox3 does not negatively affect the SnTox2-Snn2 interaction. Together, these results demonstrate that SnTox3 codes for the SnTox3 protein which plays a significant role in disease on lines carrying the wheat HST-sensitivity gene Snn3.

**Preliminary characterization of the SnTox3 protein**

The predicted SnTox3 pre-protein contains 230 amino acids with a calculated mass of 25.85 kDa. SnTox3 activity was detected in culture filtrates and both SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) and WolfPsort (http://wolfsort.org/) signal peptide prediction software predicted that the first 20 amino acids are a signal peptide (Figure 2) necessary for the secretion of the protein.

SnTox3 contains six cysteine residues. Using DISULFINDF at the website of PredictProtein (http://www.predictprotein.org), six cysteine residues in the SnTox3 protein were predicted to form three disulfide bonds with a confidence level of 9 (0-9 scale) (Figure 2B). The best connectivity pattern based on the prediction is C89-C209, C154-C203, and C166-C218. These disulfide bridges may play an important role in the stability of the protein. Dithiothreitol (DTT) treatment of *P. pastoris* SnTox3 culture filtrates eliminated necrotic activity (Figure 9) suggesting the importance of at least one of the disulfide bonds in SnTox3 activity.
Diversity and function of SnTox3 in global S. nodorum populations

A total of 923 samples were collected from eight major geographical regions and were screened for presence/absence of SnTox3 using PCR. The SnTox3 deletion frequencies ranged from 3.51% in Australia to 61.7% in the Middle East, with a global deletion frequency of 39.9%. (Supporting Table 2). Dot Blot Hybridization, with a subset of isolates, confirmed that isolates with no PCR amplicon did not contain a copy of SnTox3.

Sequence diversity of SnTox3 was assessed by sequencing the gene for 245 isolates. Eleven nucleotide haplotypes were identified which encoded four amino acid sequences (Supporting Table 2). Nucleotide diversity was estimated using the pair-wise difference measurement Pi [23] and shown in Figure 10. There were 9 synonymous and 8 non-synonymous substitutions across 654 nucleotide sites. Isolates Sn4, SnCP2052, ARKW40, and KZ3.4.10 were used to amplify and clone SnTox3 alleles representing protein variants 1, 2, 3, and 4, respectively. The amino acid differences for the four protein variants are shown in Figure 11. Protein variant 1 composed 66% of the sequenced isolates followed by protein variant 2 which composed 31% of the sequences. Protein variant 1 was used in the P. pastoris expression experiments described earlier, but the other three protein variants were also heterologously expressed in P. pastoris and were all found to induce necrosis on the SnTox3 differential line BG220 (Figure 11, Supporting Table 2).

DISCUSSION

Until recently, the S. nodorum-wheat pathosystem was thought to be based on the interaction of a suite of non-specific toxins and cell wall degrading enzymes [24]. No specific variation in virulence on the host was recognized. Resistance was quantitative and defined as a plethora of mostly weak and environment-specific QTLs [14]. As a result, breeding for disease resistance was based on simple phenotypic assessment; no molecular markers were in use.

The assembly of critical tools including functional genomics, host mapping populations, and the genome sequence has allowed the role of pathogen-produced effectors (HSTs) to be revealed. We previously characterized the SnToxA-Tsn1 interaction in which some isolates
produced different variants of ToxA that interact directly or indirectly with the product of the wheat gene *Tsn1* to produce a necrotic reaction that benefits the pathogen [15]. Three additional toxin/receptor interactions have been partially characterized [18-20]. Together, these findings have opened up a rational route to the genetic control of this disease [14, 25] where pathogen produced effectors (HSTs) interact with dominant host gene products. But in contrast to effector triggered immunity (ETI) found in many biotrophic systems [26], the end result in the *S. nodorum*-wheat interaction is effector triggered susceptibility (ETS).

Proteomic analysis of purified active fractions identified the candidate gene *SNOG_08981*. A range of tests presented here provided convincing and comprehensive evidence that the candidate gene *SNOG_08981* corresponds to *SnTox3* and that addition or ablation of SnTox3 changes the specificity of the corresponding fungal strains on the differential line harboring the host gene *Snn3*.

*SnTox3* is an intron-free 693 nucleotide gene encoding a predicted 230 amino acid immature protein. MS and MS/MS spectra of the digested protein identified 44% of the total sequence, including amino acids from position 73 to the C-terminus. The first 20 amino acids make up a predicted signal sequence that likely serves as a signal peptide for secretion. This is consistent with the detection of SnTox3 activity in the culture filtrate of fungal isolates that harbor SnTox3. As amino acids 21 through 72 are not accounted for by MS and also SnTox3 was observed as an ~18 kDa protein on SDS-PAGE, it is likely that SnTox3 is a pre-pro protein similar to ToxA and that the mature protein requires the cleavage of an additional N-terminal region including part or all of amino acids 21 to 72. It is interesting to note that immediately before the first peptide identified by mass spectrometry, there are four residues (LSKR) similar to an LRKR sequence in the effector protein Six1 which is predicted as a Kex2-like protease recognition site [27]. The calculated molecular weight for the remainder of the protein is 17.88 kDa similar to what we observed for SnTox3 using SDS-PAGE. Additional work is ongoing to verify the N-terminal region of the mature protein.

SnTox3 is a unique protein and presently has no obvious homology to other proteins present in current protein databases other than *S. nodorum*. Many pathogen proteins that are secreted into host environments containing many plant proteases are rich in cysteine residues.
These cysteine residues often form disulfide bonds that play a critical role in folding or stability of the protein [27, 28]. SnTox3 contains six cysteine residues which are predicted to form three disulfide bonds and the loss of SnTox3 activity after reduction of the protein by DTT treatment indicates that the formation of at least one disulfide bond is critical for its biological function.

SnTox3 is expressed at high levels especially early in infection as well as in culture. The in planta transcription profile is similar to that of ToxA [15]. It is consistent with a model whereby the role of SnTox3 is to induce necrosis in susceptible host cells ahead of tissue colonization. Once the host tissue becomes necrotic, it is appropriate that SnTox3 expression is reduced as the fungus prepares to disseminate via sporulation.

In the assembled S. nodorum genome, SnTox3 is located on the end of supercontig 14 (~1.2 Mb), however it is not clear whether this supercontig constitutes an entire chromosome. Although SnToxA lies in the middle of a very short supercontig (supercontig 55, ~32 Kb), CHEF gel analysis showed it to be in a 2.35 Mb chromosome [15]. Using Southern analysis of the same CHEF blot, an SnTox3 probe hybridized to a significantly smaller band (data not shown), indicating SnToxA and SnTox3 are located in different chromosomes. It is striking that both SnTox3 and SnToxA are flanked by an AT-rich sequence that contains long terminal repeat (LTR) retrotransposons [11]. We note that many other fungal avirulence and HST genes are surrounded by repetitive DNA. AvrLm1 and AvrLm6 conferring avirulence in Leptosphaeria maculans were also found residing in a genomic region rich in LTR transposons [29, 30]. Similarly, avirulence genes in Fusarium oxysporum f.sp. lycopersici, were surrounded by repetitive elements [31, 32], and Magnaporthe grisea avirulence genes are also found near repetitive elements and telomeres [33]. The significance of the presence of these genes in repetitive regions has yet to be determined, but it does suggest another means to identify candidate genes from sequenced fungal genomes.

Recent research in the area of pathogen effectors and host recognition has led to the theory that two layers of plant resistance are present. In the initial layer of plant defense, pathogen or microbe associated molecular patterns (PAMPs or MAMPs) are recognized by pattern recognition receptors (PRRs) at the plant cell surface, leading to PAMP triggered immunity (PTI) [26]. Pathogens can produce effectors to elude or inhibit PTI and these effectors
can then be recognized in a second layer of host defense known as effector triggered immunity [26]. In this second layer of defense, the plant receptors are typically nucleotide binding-leucine rich repeat (NB-LRR) proteins which trigger resistance after direct or indirect recognition of the corresponding pathogen produced effector. Interestingly, two recent discoveries of host genes involved in HST susceptibility have been shown to be in the NB-LRR family [34, 35].

The *S. nodorum*-wheat interaction is a necrotrophic pathogen system where, like many necrotrophic systems, the pathogen induces cell death via HSTs and then thrives on dying tissue, leading to sporulation of the pathogen. Based on what is known about the ToxA-*Tsn1* model, Ptr ToxA and presumably SnToxA is recognized by the host within a few hours of infiltration [36] and internalized into the cell [17] followed by a cascade of events involving host transcription and translation [37, 38] leading to a susceptible response. Evidence suggests that *Tsn1*, the wheat gene associated with ToxA sensitivity in the *S. nodorum* and *P. tritici-repentis* systems, is also a member of the NB-LRR family (Faris et al. unpublished data) implicating another resistance-like gene in effector triggered susceptibility by a necrotrophic fungus.

In the *S. nodorum*-wheat interaction, evidence has been shown for a complex of at least four pathogen-produced effector proteins each with a distinct corresponding dominant host gene that confers susceptibility [14], one of which is likely a member of the NB-LRR gene family. The *S. nodorum*-wheat interaction is the inverse of the biotrophic host pathogen model of effector triggered immunity, in that cell death is favorable to pathogen virulence rather than to host resistance. The current work adds strength to the hypothesis that disease in the SNB system is a result of multiple effector proteins which, when produced by the pathogen, lead to effector triggered susceptibility (ETS) rather than effector triggered immunity (ETI), but possibly via pathways involving resistance-like genes. This is likely the case in other necrotrophic systems involving both proteinaceous and non proteinaceous HSTs.

A screen of a global sample of *S. nodorum* field populations revealed that 39.9% of tested strains lacked *SnTox3* and there was considerable variation in the observed frequency of *SnTox3* among geographical regions. Very similar findings emerged from an analysis of *SnToxA* using the same isolates, though *SnToxA* was deleted in almost two thirds of the tested isolates [39]. Data for both toxins were obtained from 849 of the field isolates; 164 isolates tested positive for
both toxins, while 293 isolates had neither toxin. The latter finding suggests that these two toxins are not required for pathogenicity and that additional toxins or other factors are likely to be active in modern field populations of \textit{S. nodorum}. The remaining isolates carried only one out of the two toxins discussed above. For \textit{SnToxA} we hypothesized that the frequency of the gene in \textit{S. nodorum} populations reflected the frequency of the \textit{Tsn1} sensitivity gene in the corresponding wheat populations [15, 39]. Some evidence to support this hypothesis came from recent work in Australia where all of the tested \textit{S. nodorum} isolates carried \textit{ToxA}. Oliver et al. [40] showed that more than 90\% of modern Australian wheat cultivars carry the corresponding \textit{Tsn1} sensitivity allele. We propose the same mechanism to explain the observed differences in \textit{SnTox3} frequencies among regional populations, with frequencies of \textit{SnTox3} reflected by the frequencies of \textit{Snn3} in the corresponding wheat populations. Sequence analysis of \textit{SnTox3} loci revealed diversity at both the nucleotide and amino acid level that was similar to that observed for \textit{SnToxA} [39]. \textit{SnTox3} was sequenced from 245 isolates. Two of the protein variants comprised approximately 97\% of the isolates, with the other two variants being rare. All four proteins exhibited full biological activity, and thus virulence, on wheat varieties carrying \textit{Snn3}.

The availability of \textit{SnTox3} will allow wheat geneticists and breeders to use \textit{SnTox3} to identify and eliminate \textit{Snn3} in germplasm. \textit{SnTox3} is only the second gene cloned from the SNB system that encodes a proteinaceous necrosis inducing toxin and therefore can be used in conjunction with \textit{SnToxA} to continue to evaluate the mechanism by which \textit{S. nodorum} infects wheat. This susceptibility inducing effector protein is one of only a handful of proteinaceous HSTs identified and therefore is a valuable addition to the expanding body of knowledge specific to necrotrophic plant pathogen interactions, especially in the important Dothideomycete class. On a broader level, the cloning of \textit{SnTox3} provides an additional tool to investigate how necrotrophic fungi benefit from susceptibility-inducing effector proteins.

\textit{S. nodorum} is a member of the Dothideomycete class of fungi which is a newly classified and large fungal taxon including many important necrotrophic pathogens [41]. In addition to \textit{S. nodorum} and \textit{P. tritici-repentis}, proteinaceous HSTs have recently been reported from other Dothideomycete species [6-8]. It is possible that multiple interactions similar to the \textit{S. nodorum}-wheat interaction are also present in other Dothideomycete disease systems. As the first
Dothideomycete species to be sequenced, *S. nodorum* may serve as a model to study the necrotrophic pathogen lifestyle including the interaction with its host.

**SUPPORTING MATERIAL**

All supplementary files for this chapter are available online

**Supporting Table S1**
Primers used in this study.
Found at: doi:10.1371/journal.ppat.1000581.s001 (0.04 MB DOC)

**Supporting Table S2**
*P. nodorum* populations used in the investigation of genetic doversity of *SnTox3*.
Found at: doi:10.1371/journal.ppat.1000581.s002 (0.56 MB DOC)

**Figure S1**
Partial purification of *SnTox3* by ion exchange (A) and size exclusion (B) chromatography.
Found at: doi:10.1371/journal.ppat.1000581.s003 (0.02 MB TIF)

**Figure S2**
PCR testing of SNOG_08981 and SNOG_16063 identified from mass spectrometry for presence in Sn79-1087.
Found at: doi:10.1371/journal.ppat.1000581.s004 (2.55 MB TIF)

**Figure S3**
Protein view of SNOG_08981 from the mascot search report.
Found at: doi:10.1371/journal.ppat.1000581.s005 (0.89 MB TIF)

**Figure S4**
Toxin bioassay of *P. pastoris* transformed with an empty express vector.
Found at: doi:10.1371/journal.ppat.1000581.s006 (3.82 MB TIF)

**Figure S5**
Virulence analysis of Sn1501 and its *SnTox3* disrupted strains as well as an ectopic transformant on BR34, Grandin and the SnToxA and SnTox2 toxin differential lines.
Found at: doi:10.1371/journal.ppat.1000581.s007 (9.53 MB TIF)

**ACKNOWLEDGMENTS**

The authors would like to thank Danielle Holmes, Leonard Cook and Meghan Rolfsrud for technical assistance. This work was supported by USDA-ARS CRIS projects 5442-22000-043-00D and 5442-22000-030-00D.
Table 1. Fungal isolates or strains used in the yeast expression, SnTox3 transformation, SnTox3 disruption, protein haplotypes analysis, and Southern hybridization.

<table>
<thead>
<tr>
<th>Isolate or strain§</th>
<th>Origin/Resource</th>
<th>SnToxA</th>
<th>SnTox1</th>
<th>SnTox2</th>
<th>SnTox3</th>
<th>SnTox3‡</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sn4</td>
<td>ND, USA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Friesen, et al. unpublished</td>
</tr>
<tr>
<td>Sn15</td>
<td>Australia</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>[15]</td>
</tr>
<tr>
<td>Sn1501</td>
<td>OH, USA</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>[20]</td>
</tr>
<tr>
<td>Sn50</td>
<td>ND, USA</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>Friesen, et al. unpublished</td>
</tr>
<tr>
<td>Sn2000</td>
<td>ND, USA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>[18], [21]</td>
</tr>
<tr>
<td>SnCP2052</td>
<td>Denmark</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Friesen, et al. unpublished</td>
</tr>
<tr>
<td>Sn6</td>
<td>ND, USA</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>[19]</td>
</tr>
<tr>
<td>SnOH69-1</td>
<td>OH, USA</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Friesen, et al. unpublished</td>
</tr>
<tr>
<td>Sn79-1087</td>
<td>ND,USA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>[15]</td>
</tr>
<tr>
<td>Sn79+SnTox3A</td>
<td>This study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>Sn79+SnTox3B</td>
<td>This study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>Sn1501ΔSnTox3A</td>
<td>This study</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>This study</td>
</tr>
<tr>
<td>Sn1501ΔSnTox3B</td>
<td>This study</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>This study</td>
</tr>
<tr>
<td>Sn1501Ect</td>
<td>This study</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>This study</td>
</tr>
</tbody>
</table>

§Sn1501 and SnOH69-1 were provided by Dr. Pat Lipps, Ohio State University; the avirulent isolate Sn79-1087 was provided by Dr. Joseph Krupinsky, retired, Northern Great Plains Research Laboratory USDA-ARS, Mandan, ND, USA.

†Toxin activity is detectable (+), not detectable (-) or not tested (?).

‡Identifies the presence (+) or absence (-) of the SnTox3 (SNOG_08981) gene in isolates or strains.
Figure 1. SDS-PAGE image of a size exclusion chromatography fraction containing SnTox3 activity. Lane 1. Genescript Smart His-tagged protein standard; Lane 2 and 3, duplicate protein samples of the size exclusion chromatography fraction 17-12 (Figure S1B); Lane 4. Bio-Rad pre-stained protein standard. The molecular weight (kDa) of each protein is labeled and the arrows indicate the two protein bands that were excised from the gel. The larger band is estimated at 18 kDa and contains SnTox3.
Figure 2. Genomic location, gene structure and nucleotide and amino acid sequence of SNOG_08981. A. The genomic location and gene structure of SNOG_08981. Top: Supercontig14 was schematically drawn to show one end containing SNOG_08981 (Gene 1 -
light blue-color filled arrow) as well as the four other SNOG genes (2, 3, 4 and 5, black color filled arrows) present in the flanking region. A large amount of AT rich sequence closely flanking SNOG_08981 was labeled that contains a polyprotein gene and a putative transposase gene. Below: A close up of a 1.6 kb genomic region (dash boxed) indicates a single exon, the 59- and 39- UTR, and the putative promoter region. B. Nucleotide sequence of a 1.6 kb genomic region of SNOG_08981 (SnTox3) and the deduced amino acid sequence of the SNOG_08981 protein (SnTox3). The underlined peptide sequence was detected in mass spectrometry. The amino acids highlighted in blue indicate the predicted signal peptide. The six cysteine residues highlighted in green are predicted to form 3 disulfide bonds. The bold DNA sequence indicates the start codon (ATG), stop codon (TAG) and the SalI restriction site. DNA sequence highlighted in yellow is UTR and the red highlighted region is a putative TATA box. Bold italicized sequence indicates the primer sites used to amplify the genomic region for transformation.
Figure 3. Southern analysis of SNOG_08981 in S. nodorum and related fungi. Southern hybridization was performed using a probe from the full length SNOG_08981 ORF as well as lDNA. Size ladders include Hyperladder I and lDNA HindIII fragments. The isolate designations starting with Sn or SN are S. nodorum isolates, and Ptr86-124, St975, and Pt-15A, are isolates of Pyrenophora tritici-repentis, Mycosphaerella graminicola and P. teres, respectively.
Figure 4. Expression pattern of SnTox3 in planta. The expression level of SnTox3 was examined and compared to that of the Act1 gene at 3, 5, 7 and 10 days after inoculation. The x axis shows the number of days post-infection. The y axis represents relative gene expression levels normalized to Act1. Standard error bars are shown.
Figure 5. Toxin bioassay of the Pichia pastoris X33 strain transformed with SNOG_08981. Leaves of toxin differential and parental lines, including BR34, Grandin, BG261, BG220 and BG223 were infiltrated with culture filtrate from P. pastoris X33 transformed with SNOG_08981. Wheat lines containing Snn3 (Grandin and BG220) were sensitive whereas lines not containing Snn3 (e.g. containing snn3) (BR34, BG261, and BG223) were insensitive.
Figure 6. Molecular characterization of SnTox3 transformation in Sn79-1087 and disruption in Sn1501. A. Verification of the SnTox3 transformation into Sn79-1087 and disruption in Sn1501 using Southern analysis. Genomic DNA from the avirulent isolate Sn79-1087 and its two SnTox3 transformed strains Sn79+SnTox3A and Sn79+SnTox3B were digested with XhoI. The virulent
SnTox3 containing isolate Sn1501 and its two SnTox3-disrupted strains Sn1501DSnTox3A and Sn1501DSnTox3B as well as a strain with an ectopic integration, Sn1501Ect, were digested with EcoRI. The blot was probed with a full length SnTox3 cDNA. The two red dots indicate the bands of 7.5 kb and 3.0 kb specifically present in the SnTox3-disrupted strains (see Figure 7B). A 7.9 kb band is present in the wild type and ectopic strain. More ectopic insertions were detected by southern in Sn1501DSnTox3A and Sn1501Ect, compared to Sn1501DSnTox3B. B. SnTox3 disruption strategy in Sn1501. The SnTox3 gene in Sn1501 was disrupted by insertion of the hygromycin B resistance gene (orange bar) into a SalI restriction site located in the center of the SnTox3 gene. A 1.2 kb genomic region of SnTox3 (black bar) amplified with primers 8981g1F_XbaI and 8981g1R_XbaI was cloned and then linearized with SalI. The linearized vector was re-ligated with a cpc-1:HYG :tryptophan cassette that was released from the pLP605KO vector [51] using the restriction enzyme XhoI. The resulting vector was linearized via XbaI before transformation into Sn1501 protoplasts. The disruption of SnTox3 can be identified using primers 8981g0F with HY and verified by primers 8981g1F and 8981g1R. C. PCR screening and verification of SnTox3-disrupted mutants in Sn1501. The primer 8981g0F (g0F) located outside of the 1.2 kb region along with primer HY amplify a 2.9 kb fragment in strains which have SnTox3 disrupted by the insertion. No amplification is observed in wild type and ectopics. The primers 8981g1F1 (g1F) and 8981g1R (g1R) amplify a 1.2 kb fragment in the wild type and ectopic strains indicating the SnTox3 gene remains intact, while they amplify a 3.8 kb fragment (1.2 kb of SnTox3 region plus 2.6 kb of cpc-1:HYG :tryptophan cassette) in the SnTox3-disrupted strains. D. RT-PCR verification of different mutated strains. The primers 8981cF and 8981cR which amplify the SnTox3 ORF region were used to test the presence of transcripts of SnTox3 in different isolates and genetically modified strains including Sn79-1087 and its two transformed SnTox3+ strains (Sn79+SnTox3A and Sn79+SnTox3B) and Sn1501 and its two SnTox3-disrupted strains Sn1501DSnTox3A and Sn1501DSnTox3B along with the ectopic transformant Sn1501Ect. The S. nodorum actin gene [49] was used as an internal control and water was used as a PCR negative control.
Figure 7. Toxin bioassay and virulence analysis of the SnTox3 genetically modified strains. A. Toxin bioassay of Sn79-1087 with its two transformed SnTox3+ strains Sn79+SnTox3A and Sn79+SnTox3B. Leaves of BG220 (SnTox3 differential line) were infiltrated with culture filtrates of Sn79-1087 wild type, Sn79+SnTox3A and Sn79+SnTox3B and photographed 3 days after infiltration showing culture filtrates from the wild type isolate did not induce necrosis while the strains transformed with SnTox3 (Sn79+SnTox3A and Sn79+SnTox3B) induced necrosis. B. Virulence of the SnTox3 transformed strain Sn79+SnTox3A. Conidia from Sn79+SnTox3A were
inoculated onto BR34, Grandin, BG220, BG261 and BG223 showing this strain was virulent on Grandin and BG220 which harbors Snn3, while it remained avirulent on BR34, BG261, and BG223 which harbors snn3. C. Toxin bioassay of SnTox3 disrupted strains. Leaves of BG220 (Snn3) were infiltrated with culture filtrate of Sn1501 wild type, Sn1501DSnTox3A, Sn1501DSnTox3B, and the ectopic strain Sn1501Ect and photographed 3 days after infiltration. Culture filtrates from the wild type and ectopic strains were able to induce necrosis while those from the two SnTox3-disrupted strains were unable to induce necrosis. D. Virulence comparison of Sn1501 wild type and the SnTox3 disrupted strains on BG220 (Snn3). Inoculation of conidia from Sn1501 wild type, Sn1501DSnTox3A, Sn1501DSnTox3B and Sn1501Ect were inoculated onto BG220 (Snn3) showing the loss of virulence of SnTox3 disrupted strains on the SnTox3 differential line BG220.
Figure 8. Interval map of chromosome 2D (Top) and 5B (bottom). The map shows QTL of susceptibility associated with Snn2 and Snn3 loci in the BG population after being inoculated with genetically modified fungal strains. Strains are depicted by different colors as indicated. A centiMorgan scale is on the left of the map and markers are shown in their relative position along the right. An LOD scale is shown along the x axis, and the critical LOD threshold of 3.0 is indicated by the dotted line.
Figure 9. SnTox3 activity is sensitive to Dithiothreitol (DTT). Leaves of BG220 were infiltrated with culture filtrates of P. pastoris expressing SnTox3 that had been treated for 2 h with 0 mM, 5 mM and 10 mM DTT.
Figure 10. Calculation of nucleotide diversity (Pi) at the SnTox3 coding region. The calculations were done using a sliding window of size 10 with a step size of 5. Pi values of 0 indicate conserved regions of the gene for all 245 sequenced isolates.
Figure 11. The four protein variants of SnTox3 and their biological activity. The different haplotypes of SnTox3 were amplified from the corresponding representative isolates and expressed in P. pastoris. The unfilled bars represent the complete SnTox3 protein and the number under the bar shows the position of each sequence variation indicated by pairs of amino acids (the first is the protein variant 1 and the substitution is in parenthesis). The biological activity of yeast clones expressing each protein haplotype is shown to the right.
REFERENCES


Chapter 4


Chapter 5:
The Cysteine Rich Necrotrophic Effector \textit{SnTox1} Produced by \textit{Stagonospora nodorum} Triggers Susceptibility of Wheat Lines Harboring \textit{Snn1}

Zhaohui Liu, Zengcui Zhang, Justin D. Faris, Richard P. Oliver, Robert Syme, Megan C. McDonald, Bruce A. McDonald, Peter S. Solomon, Shunwen Lu, Weilin L. Shelver, Steven Xu, Timothy L. Friesen

Published in PLoS Pathogens 2012 Vol 8 (1), doi:10.1371/journal.ppat.1002467

* M. C. McDonald contributed the section on diversity and population genetics of \textit{SnTox1}, tests for positive selection and Table 1.
ABSTRACT

The wheat pathogen *Stagonospora nodorum* produces multiple necrotrophic effectors (also called host-selective toxins) that promote disease by interacting with corresponding host sensitivity gene products. SnTox1 was the first necrotrophic effector identified in *S. nodorum*, and was shown to induce necrosis on wheat lines carrying *Snn1*. Here, we report the molecular cloning and validation of SnTox1 as well as the preliminary characterization of the mechanism underlying the SnTox1-*Snn1* interaction which leads to susceptibility. SnTox1 was identified using bioinformatics tools and verified by heterologous expression in *Pichia pastoris*. SnTox1 encodes a 117 amino acid protein with the first 17 amino acids predicted as a signal peptide, and strikingly, the mature protein contains 16 cysteine residues, a common feature for some avirulence effectors. The transformation of SnTox1 into an avirulent *S. nodorum* isolate was sufficient to make the strain pathogenic. Additionally, the deletion of SnTox1 in virulent isolates rendered the SnTox1 mutated strains avirulent on the *Snn1* differential wheat line. SnTox1 was present in 85% of a global collection of *S. nodorum* isolates. We identified a total of 11 protein isoforms and found evidence for strong diversifying selection operating on SnTox1. The SnTox1-*Snn1* interaction results in an oxidative burst, DNA laddering, and pathogenesis related (PR) gene expression, all hallmarks of a defense response. In the absence of light, the development of SnTox1-induced necrosis and disease symptoms were completely blocked. By comparing the infection processes of a GFP-tagged avirulent isolate and the same isolate transformed with SnTox1, we conclude that SnTox1 may play a critical role during fungal penetration. This research further demonstrates that necrotrophic fungal pathogens utilize small effector proteins to exploit plant resistance pathways for their colonization, which provides important insights into the molecular basis of the wheat-*S. nodorum* interaction, an emerging model for necrotrophic pathosystems.
INTRODUCTION

Like other parasites, fungal pathogens secrete a battery of molecules known as effectors during the infection process. These effectors can alter plant biological processes resulting in successful colonization [1, 2]. Conversely, recognition of effectors by the plant innate immune system can initiate a defense response resulting in effector-triggered immunity (ETI) [3, 4]. ETI is characterized by the accumulation of reactive oxygen species (ROS), transcriptional induction of pathogenesis-related (PR) genes and production of antimicrobial chemical compounds, eventually leading to rapid and localized plant cell death, known as the hypersensitive response (HR) [5]. In ETI, the perception of the fungal effector is mediated by the corresponding plant resistance gene (R) which acts in a gene-for-gene manner [6, 7]. Currently, it is believed that this localized suicide of plant cells induced by ETI halts further growth of the biotrophic fungal pathogen, which requires living plant tissue for survival.

Necrotrophic fungal pathogens are known to produce host selective toxins (HSTs), including low molecular weight metabolites and small secreted proteins that function as essential determinants of pathogenicity or virulence [8, 9]. HSTs can therefore be viewed as effectors of necrotrophic pathogenicity and hence we prefer the term necrotrophic effector (NE) [10, 11]. These effectors play significant roles in determining the outcomes of plant-pathogen interactions by specifically interacting (directly or indirectly) with the products of corresponding host genes [12, 13]. However, in contrast to ETI in the classical gene-for-gene model, the necrosis induced by effectors from necrotrophic fungal pathogens results in disease susceptibility; thus, it can be described as effector-triggered susceptibility (ETS) [14, 15], a term which was originally used in reference to biotrophic systems [4].

The molecular basis of necrosis-induced ETS involving necrotrophic fungi is still largely unknown, but has in several cases exhibited the hallmarks of programmed cell death (PCD); DNA laddering, heterochromatin condensation, cell shrinkage, callose deposition and an oxidative burst [9, 16, 17]. ToxA, a necrotrophic effector found in both Pyrenophora tritici-repentis and Stagonospora nodorum, causes the loss of plasma membrane integrity and the accumulation of hydrogen peroxide [18, 19]. Microarray analysis revealed that several wheat
genes involved in defense response and signaling pathways were strongly regulated by the ToxA-
Tsn1 interaction [20, 21].

Interestingly, three plant genes involved in susceptibility to necrotrophic effectors (Pc, the sorghum sensitivity gene corresponding to PC toxin; LOV1, the Arabidopsis sensitivity gene corresponding to victorin; and Tsn1, the wheat sensitivity gene corresponding to ToxA) have been cloned and shown to be resistance-like genes containing both nucleotide binding (NB) and leucine-rich repeat (LRR) domains [15, 22, 23]. This has led to speculation that necrotrophic fungal pathogens may utilize plant resistance signaling pathways to subvert PCD and enable pathogen growth [15, 24].

Stagonospora nodorum, an ascomycete fungus (teleomorph: Phaeosphaeria nodorum), is the causal agent of wheat Stagonospora nodorum blotch (SNB), a globally distributed and economically important disease [25]. *S. nodorum* is a typical necrotrophic fungal pathogen [10, 26]. In recent years, it has been shown that this pathosystem is based largely on interactions involving proteinaceous necrotrophic effectors and corresponding host sensitivity genes that, when occurring together, result in ETS. To date, six interactions have been reported including SnTox1-Snn1 [27], SnToxA-Tsn1 [28, 29], SnTox2-Snn2 [12], SnTox3-Snn3-B1 [30], SnTox4-Snn4 [31], and SnTox3-Snn3-D1 [32]. In addition, several other effector-host gene interactions have been identified (Friesen and Faris, Oliver and Tan, unpublished data). Therefore, the wheat-*S. nodorum* system is emerging as a model to investigate the molecular mechanisms of necrotrophic pathogenesis [13]. One of our research goals has been to clone necrotrophic effector genes and decipher their molecular and biochemical functions.

Of the *S. nodorum* effector genes, SnToxA and SnTox3 have been cloned with the aid of the *S. nodorum* genome sequence information [14, 29, 33]. The SnToxA gene is essentially identical to the ToxA gene isolated from the wheat tan spot pathogen *P. tritici-repentis*. Mature ToxA consists of a 13.2 kDa protein containing two cysteine residues as well as an RGD-containing vitronectin-like motif that is present in a solvent-exposed loop in the active protein [34-38]. The RGD motif has been shown to be essential for internalization and internalization has been shown to be critical for the induction of necrosis [37, 39, 40] SnTox3 encodes an
approximately 17.5 kDa mature protein with six cysteine residues and has no homology to genes in the public databases [14].

Here, we report the molecular cloning and characterization of the SnTox1 gene which encodes the SnTox1 protein, and we show that SnTox1 is specifically recognized by the corresponding wheat sensitivity/susceptibility gene Sn1. The characterization of the SnTox1-Sn1 interaction provides strong evidence that necrotrophic fungal pathogens use necrotrophic effectors to subvert the host resistance mechanism to cause disease.

MATERIALS AND METHODS

Bioinformatics for prioritizing candidate genes

A series of experimental and bioinformatic criteria associated with effectors were evaluated to produce a candidate gene ranking of the predicted genes in the S. nodorum genome. These criteria were based on the known and predicted properties of effectors. Genes matching different criteria were given scores from 1 to 6. The sum of scores for each gene was ranked and the top 100 genes were considered. The criteria used data from mass-spectrometry analyses of culture filtrates, a genome sequence scan of the strains Sn4 and Sn79-1087, an in planta microarray experiment and various bioinformatics analyses. The criteria were as follows: predicted to be less than 30 kDa (1 point), cysteine rich (> 1 standard deviation more cys residues than expected of a protein of that size) (2 points), detected by MS in culture filtrates (6 points), located within 5 kb of repetitive sequences (2 points), absence of homologues in the NCBI nr database (2 points), presence of RXLR or RGD motifs (2 points), predicted to be secreted (3 points), presence of a modified version of the gene in Sn4 (3 points), absence of the gene in Sn79-1087 (4 points), and a gene expression profile similar to ToxA and Tox3 (3 points).

Yeast expression of SNOG_20078 for verification that it was SnTox1

The total RNA of 7 day old mycelium of SN15 grown in Fries media [27] was prepared using the RNeasy plant mini kit (Qiagen) and treated with RNase-free DNase I (Promega). First-strand cDNA was synthesized from 2 µg of total RNA using TaqMan Reverse Transcription Reagents (Applied Biosystems). The coding region of SNOG_20078 was amplified from the
above cDNA sample using primers 20078CF_EcoRI and 20078CR_ApaI containing the indicated restriction site (Table S4). The cloning of SNOG_20078 into the corresponding sequencing and expression vectors, yeast transformation, and preparation of culture filtrates from yeast cultures all followed the procedure described by Liu et al. [14]. The pGAPZ A vector containing the SNOG_20078 gene was linearized with AvrII before transformation. Culture filtrates of the yeast culture transformed with the SNOG_20078 coding region were infiltrated into wheat lines including BR34 (snn1), Grandin (snn1), BG220 (snn1), BG223 (snn1), BG261 (snn1), W-7984 (Snn1), Chinese Spring (Snn1), Opata85 (snn1), and ND495 (snn1). Because the culture filtrates caused necrosis on W-7984 and CS, which both possess Snn1 [27], it was infiltrated onto CS 1BS-18, CS ems237, and the ITMI population [27] for verification of SnTox1 based on its interaction with Snn1. CS 1BS-18 carries a deletion in the distal end of chromosome 1B that harbors the Snn1 locus [27]. CS ems237 is an SnTox1 insensitive mutant derived from CS by EMS (ethane methyl sulfonate) mutagenesis (Faris et al., unpublished data).

**Generation of an SnTox1 antibody and western blot analysis of SnTox1**

A 14 amino acid long peptide (sequence: CKNGKQAAHEAQKQ), designated SnTox1:50-63, was synthesized by GenScript (Piscataway, NJ). The peptide SnTox1:50-63 (4.7 mg, 0.003 mmole) was first conjugated to bovine serum albumin (BSA, 20 mg, 0.0003 mmole, Sigma-Aldrich, St Louis, MO) in the presence of 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC, 20 mg, Pierce Biotechnology, Rockford, IL) in 2 mL of 100 mM 2-(N-morpholino) ethanesulfonic acid buffer, pH 6 overnight at 4 °C. The protein was separated from EDC through a size-based column (D-Salt Excellulose, Thermo Scientific, Rockford, IL) and concentrations were determined by the method of Bradford (Bio-Rad Laboratories, Inc. Hercules, CA) using BSA as the calibration standard. Success of the conjugation reaction was assessed on a 13% SDS-PAGE gel. One hundred milligrams of the immunogen were immunized into New Zealand White Rabbits at 3 week intervals for a total of six immunization cycles. The final sera were collected eight days after immunization and were used for western blot analysis.
To prepare the SnTox1 protein sample for western blot analysis, 5 mL of culture filtrate from an SnTox1 yeast culture and control yeast culture (yeast strain transformed with an empty vector) was precipitated by adding 20 mL of methanol and incubating in a -20 freezer overnight. After centrifuging for 10 min at 13,000 rpm on a HERMLE Z 323K centrifuge with a 220.80 V02 rotor (Labnet), the pellet was retained, air dried and re-suspended in 500 µL of a 1× sample loading buffer. Protein gels were loaded with 50 µL of the resulting sample solution. SDS-PAGE, protein transferring, and color development followed a routine protocol described in Meinhardt et al. [36]. To ensure the quality of protein sample preps, the same amount of sample solution was also run on a gel and visualized by coomassie blue staining.

**Amplification of full-length transcript of SnTox1**

The same RNA extracted from SN15 was used to amplify the 5’ and 3’ ends of the cDNA of SnTox1. The 5’ and 3’ RACE were performed using the Smart RACE cDNA amplification kit (Stratagene, LaJolla, CA) according to the instructions in the user manual with gene-specific primers 20078CF and 20078CR (Table S4). The procedure described by Liu et al. [14] was followed for the cloning and sequencing of the amplified 5’ and 3’ RACE fragments. The obtained sequences from 5’ and 3’ RACE fragments were used to assemble the full length cDNA and determine the 5’ and 3’ UTRs based on the SN15 genome sequence.

**Searches for SnTox1 homologs and protein alignments**

SnTox1 and Avr4 homologs were identified from the NCBI non-redundant (nr) protein database (http://www.ncbi.nlm.nih.gov/BLAST/) using BLAST searches. The chitin-binding domains of Avr4 and its homologues were identified using Reverse Position-Specific (RPS)-BLAST searches (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Amino acid alignments were performed using the MegAlign programs from Lasergene 8.1 software (DNASTAR Inc. Madison, WI). Three-dimensional (3D) structure-based sequence alignment of the putative chitin-binding motifs identified in SnTox1 with those of ChtBD1 and ChBD2 proteins were performed following the previously published data on related structures [43, 44, 75].
Analyses of *SnTox1* diversity

*SnTox1* presence and absence was screened in 777 *S. nodorum* isolates from seven geographical regions: Australia, Central Asia, East Asia, Europe, Middle East, North America, South America and South Africa (Table S1) using PCR with primer pair Tox1F_Coding and Tox1R_Coding (Table S4). A secondary PCR screen using the conserved primer pair Tox1_XF and Tox1_XR (Table S4) was conducted to confirm questionable PCR amplicons. PCR amplification was performed in 20 µl reactions containing 0.05 µM of each primer (supplied by Microsynth), 1X Dream Taq Buffer (Fermentas), 0.4 µM dNTPs (Fermentas) and 0.5 units of Dream Taq™ DNA polymerase (Fermentas). The PCR cycle parameters were: 2 min initial denaturation at 96°C followed by 35 cycles of 96°C for 30 s, 58°C for 45 s and 72°C for 1 min. A final 5 min extension was made at 72°C. To demonstrate the wide distribution of *SnTox1*, a subset of the global collection (79 isolates), along with 10 avirulent isolates and several related fungal species including Pti2 (*P. tritici-repentis*), ND89-19 (*P. teres f. teres*), Sm15A (*P. bromi*) and S. tr 9715 (*M. graminicola*) (Table S2) were evaluated in a dot blot analysis. For dot blot analysis, the DNA of fungal samples was isolated using a BioSprint 15 instrument (QIAGEN) with the corresponding kit (QIAGEN). The DNA samples were blotted onto a nylon membrane using a Bio-Dot microfiltration apparatus (BIO-RAD) following the instructions in the user manual. The entire *SnTox1* coding region was PCR amplified from the genomic DNA of SN15 and used as a probe for Southern blot analysis. Probe preparation, DNA hybridization, membrane washing and image acquisition followed the protocol described by Faris et al. [76]. The membrane was stripped and hybridized to the *S. nodorum* actin gene probe to ensure the quality for all the DNA samples.

Sequences for the entire coding region were obtained using the primer pair Tox1UTR_F and Tox1UTR_R and the primer pair Tox1Fout and Tox1Rout (Table S4). In cases of poor amplification primer pair Tox1F_Coding and a new conserved reverse primer, Tox1R_Conserved (Table S4), were used to confirm observed sequence variation. Sequencing reactions were conducted in 10 µl volume using the BigDye Terminator v3.1 Sequencing Standard Kit (Applied Biosystems) with both the forward and the reverse primer. The cycling parameters were 96°C for 2 min followed by 55 or 99 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. The
products were cleaned with the illustra Sephadex G-50 fine DNA Grade column (GE healthcare) according to the manufacturer’s recommendations and then sequenced with a 3730x/ Genetic Analyzer (Applied Biosystems). Alignment of forward and reverse sequences for each isolate was performed in SeqScape software V2.5 (Applied Biosystems, Foster City, CA). Translation and identification of protein haplotypes was also performed using this software.

**Tests for Positive Selection in SnTox1**

Codeml implemented in the software PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) was used to test for positive diversifying selection [77]. The program uses four different codon substitution models implemented in a maximum-likelihood framework to test which model, neutral or selection, best fits the data. Each model assumes a different range of values for the estimated value \( \omega \) (the ratio of non-synonymous to synonymous nucleotide substitutions). Under purifying selection, non-synonymous substitutions are expected to be rare, thus \( \omega \) will remain below 1. If non-synonymous mutations offer a selective advantage, they will be fixed at a higher rate than synonymous mutations and \( \omega \) will be greater than one. We compared the null model M1a (neutral), which assumes two site classes, purifying (\( 0 < \omega_0 < 1 \)) or neutral (\( \omega_1 = 1 \)) to the alternative model M2a (selection), which adds another class of diversifying sites (\( \omega_2 > 1 \)). We also compared the more complex null model M7 (neutral) that assumes a beta distribution for \( 0 < \omega < 1 \), with the alternative model M8 (selection) which also assumes a beta distribution and adds an additional site class with \( \omega_2 > 1 \). A likelihood ratio test was used to compare the likelihood estimate scores. The model simultaneously calculates the posterior probability for each codon that belongs to a particular site class (e.i. \( \omega > 1 \)). If the posterior probability for a codon is high and it belongs to the site class with \( \omega > 1 \), positive selection can be inferred for that codon, known as Bayes Empirical Bayes [78].

**Investigation of the genomic region of SnTox1 in different S. nodorum isolates**

Based on the annotated SN15 genome sequence, four genes SNOG_07153-SNOG_07156 were predicted within a ~7.6 kb region containing SnTox1. Primers were designed (Table S4) to amplify the gene region from start to stop codon for the four genes identified in the avirulent
isolate Sn79-1087. Since all the genes except SNOG_07154 were present in Sn79-1087, one new primer designed in SNOG_07153 (20078g3R, Table S4) was used with the SNOG07155 forward primer to amplify the whole region in several virulent isolates as well as Sn79-1087. The amplified fragments with different sizes were cloned into the pCR-4 TOPO cloning vector (Invitrogen) for sequencing. The sequences from different isolates were analyzed manually to identify the variations in the SnTox1 genomic region with the aid of the genome sequence (http://genome.jgi-psf.org/Stano1/Stano1.home.html).

SnTox1 gene transformation into the avirulent isolate Sn79-1087

A ≈1.1 kb sequence of the SnTox1 genomic region including a putative promoter and terminator was amplified from the Sn2000 isolate using primers 20078gF_XbaI and 20078gR_XbaI, each containing an XbaI restriction site sequence (Table S4). The amplified fragment was cloned into the pCR-4 TOPO vector (Invitrogen) for sequencing to verify the identity and XbaI restriction sites. The SnTox1 gene fragment was then released from pCR-4 TOPO plasmid and cloned into the pDAN vector that carries the cpc-1:hygR (hygromycin-resistance gene) cassette. The resulting plasmid, designated pDAN-SnTox1 (Figure S4) containing the 1.1 kb genomic region containing SnTox1 and hygR was used to transform Sn79-1087 protoplasts. Plasmid DNA was prepared through the regular alkaline lysis method as described by Sambrook and Russell [79] followed by the purification of the plasmid DNA using precipitation with PEG 8000 [79]. The plasmid DNA was linearized with EcoRV and concentrated to 1 µg/µl for transformation. The fungal protoplasting and PEG-mediated transformation followed the procedure described by Liu et al. [14]. The regenerated clones were screened by PCR with primers 20078gF_XbaI and 20078gR_XbaI (Table S4) and verified by Southern analysis [76]. The culture filtrate production, and infiltration and fungal inoculation with Sn79-1087 and SnTox1 transformed strains followed the protocol described previously [27, 45].
**SnTox1 gene replacement in the virulent isolate Sn2000**

The knock out of *SnTox1* was performed using a split marker strategy which employed two rounds of PCR to generate replacement fragments as described by Catlett et al. [80] (Figure S4). In the first round of PCR, the 800 bp of 5’ flanking region and 825 bp of 3’ flanking region of *SnTox1* were amplified from Sn2000 using two pairs of primers 20078KOF1 with 20078KOF2 and 20078KOF3 with 20078KOF4 (Figure S4, Table S4). Simultaneously, overlapping marker fragments HY and YG of the hygromycin phosphotransferase cassette (HYG) were amplified from pDAN with two pairs of primers, M13F with HY and M13R with YG (Table S4, Figure S4). All amplified fragments were gel purified and then used in a second round of PCR. Two reactions were set up for the second round of PCR with one to fuse and amplify the *SnTox1* 5’ flanking region with the HY fragment and the other to fuse and amplify the *SnTox1* 3’ flanking region with the YG fragment by adding the corresponding first round templates and primers. At least 100 µl of PCR reaction was set up for each reaction in the second round. Standard PCR conditions and Taq polymerase (NEB BioLabs) were used for both rounds of amplification except that round 2 used a longer extension time due to the longer template. A small amount of product from the second round of PCR was evaluated on a 1.0% agarose gel to ensure a successful fusion and amplification for each fragment. The remaining product was combined and concentrated by routine ethanol precipitation [79]. The pellet was finally re-suspended in 20 µl of TE (10mM Tris and 1mM EDTA) for transformation of Sn2000 protoplasts.

Fungal protoplasting and transformation followed the procedure described by Liu et al. [14]. The regenerated clones were screened using the PCR primers 20078KOF and 20078KOR (Table S4) which amplifies the partial coding region of *SnTox1* that was replaced by the hygromycin-resistance gene cassette. The ectopic transformant and two knock out transformants were verified by Southern blot analysis using the *SnTox1* coding region as a probe. Spores of the knock out and ectopic strains as well as wild type Sn2000 were inoculated onto wheat lines W-7984, and CS for testing the effect of the *SnTox1* knock out.
QTL analysis of SnTox1 knock out strains

The International Triticeae Mapping Initiative (ITMI) mapping population was originally used to map the Snn1 gene, which confers sensitivity to SnTox1, and quantitative trait loci conferring resistance/susceptibility to Sn2000 [27,45]. The same 106 recombinant inbred (RI) lines of this population were used to evaluate the genetically modified fungal strains including two Sn2000 SnTox1 knock out transformants (Sn2000ΔSnTox1-9 and Sn2000ΔSnTox1-15), one Sn2000 ectopic transformant (Sn2000ΔSnTox1-ECT), and the wild type Sn2000 as a control. All strains were evaluated with three biological replications by inoculating their conidia onto the ITMI population as previously described [45]. The disease rating was conducted 7 days post inoculation using a 0-5 rating scale as described by Liu et al. [45]. Composite interval mapping with the average of three disease readings was performed as previously described [30].

Disulfide bond prediction

The web-based program DISULFIND (http://disulfind.dsi.unifi.it/) was used to predict if a particular cysteine residue was involved in the formation of a disulfide bond (DB_state) as well as the confidence level of the prediction. The state of each cysteine residue was predicted as either involved (1) or not involved (0) in a DB. The scale of confidence of disulfide bonding state prediction ranges from 0 (low) to 9 (high) [46]. The web-based program DiANNA 1.1 (http://clavius.bc.edu/~clotelab/DiANNA/) [47] was used to determine the best connectivity prediction of cysteine residues in SnTox1.

qPCR analysis of SnTox1 transcription during infection

The secondary leaves of CS (~2 week old plants) were inoculated with a fungal strain modified from the avirulent isolate Sn79-1087 by addition of the SnTox1 gene. The leaf tissues were collected from the inoculated leaves at 1h, 3h, 6h, 12h, 24h, 2d (day), 3d, 4d, 5d, 6d, and 7d post inoculation. The RNA was extracted from leaf samples using the RNeasy Plant Mini Kit (QIAGEN) and treated with RNase-free DNase I (Promega). RNA sample quantification, cDNA synthesis, and gene transcript abundance analysis were performed as previously described [15]. The gene specific primers SnTox1qPCRF and SnTox1qPCRR (Table S4) designed within the
two exons were used for the SnTox1 gene in qPCR. The previously reported primers ActinF and ActinR [14] were used for the S. nodorum actin gene as internal control.

**Dithiothreitol (DTT) and heat treatment of the SnTox1 protein**

Because all cysteine residues were predicted to form disulfide bonds, the protein stability of SnTox1 was investigated by DTT and heat treatment. For DTT treatment, the SnTox1 P. pastoris culture filtrate were treated with DTT (Fisher Scientific, Pittsburgh, PA) at final concentrations of 0, 20, or 40 mM and incubated at room temperature for 2 h or 4 h. For heat treatment, the P. pastoris culture filtrate was sealed in a 2 ml centrifuge tube and heated for 30 min or 1 h on a hot plate setting at 100 °C. All treated culture filtrates were then infiltrated into CS leaves.

**DAB staining of hydrogen peroxide**

The fully expanded secondary leaves of CS and CS ems237 were infiltrated with the culture filtrates from yeast transformed with SnTox1 or culture filtrates from yeast transformed with an empty vector (as control). At 24, 48 and 72 hours post infiltration, leaf samples were collected and leaf segments with an infiltrated area were cut and stained in a freshly made 1 mg/ml 3′-3′ diaminobenzidine (DAB) (Sigma) solution. The preparation of a DAB staining solution and the staining process followed a procedure described by Thordal-Christensen et al. [48]. The stained leaf tissue was cleared for chlorophyll by placing them on a paper pre-soaked with ethanol/acetic acid solution (3:1, V/V) in a petri dish and incubating overnight. The cleared leaves were rinsed and stored in a lactic acid/glycerol/H₂O solution (1:1:1, v/v/v).

**DNA laddering analysis**

The fully expanded secondary leaves of CS and CS ems237 were infiltrated with SnTox1 yeast culture filtrates and control culture filtrates. Leaf samples were taken at 1, 2, 4, 8, 10, 24, 36, 48, 60, and 72 hour post infiltration. The DNA was extracted from the collected leaf samples using the CTAB method [17]. The 5 µl of DNA from each sample were separated on a 2%
agarose gel. The gel was stained in ethidium bromide solution for 1 hour, destained in water for 1h and photographed using a Gel LOGIC 100 image system (Kodak).

**Wheat defense gene expression using RT-PCR and qPCR**

The fully expanded secondary leaves of CS and CS ems237 were infiltrated with SnTox1 yeast culture filtrates and control culture filtrates. Five centimeter segments of infiltrated leaf tissue was collected at 1, 2, 4, 8, 10, 24, 36, 48, 60, and 72 hour post infiltration. Three leaves from different plants were collected as three replications for each time point. Total RNA was extracted from all leaf samples using the RNeasy plant kit (QIAGEN) and treated with RNase-free DNaseI (Promega). The RNA quantification and first strand cDNA synthesis were conducted as previously described [15]. Using the cDNA samples, we examined the expression of a total of 28 wheat genes that have been reported or predicted to be involved in the defense response [Lu et al. unpublished, 21] (Table S3). The RT-PCR and agarose gel electrophorsis were performed using a standard procedure. The same cDNA samples from the three replications were used to conduct the qPCR analysis for three genes: PR-1-A1, chitinase (PR-3), and thaumatin-like protein (PR-5) following the description by Faris et al. [15].

**Generation of GFP tagged strains and microscopy**

The gGFP vector [81] was used to transform the green fluorescence protein gene into two fungal strains that were only different in the production of SnTox1. One was the avirulent Sn79-1087 that did not produce SnTox1 nor did it cause disease, and the other was an Sn79-1087 SnTox1 transformant (Sn79+SnTox1A1) that expressed SnTox1 and caused disease on Snn1 lines. Since Sn79+SnTox1A1 already carried the hygromycin resistance resulting from the SnTox1 transformation, the plasmid pII99 [82] containing geneticin resistance, was used with gGFP for co-transformation of this fungus. The plasmid DNA preparation, fungal protoplasting, and fungal transformation followed the same methods described above. For all transformations, at least 20 µg of each plasmid DNA linearized with the corresponding restriction enzyme (gGFP with BglII and pII99 with EcoRV) was used. The transformants with the strongest GFP signal
were selected for both strains under the Nikon Eclipse TE-2000U microscope equipped with a GFP filter and UV light (Nikon, Japan).

The two GFP-tagged fungal strains were inoculated onto both genotypes of Snn1 (CS) and snn1 (CS ems237) as described in Liu et al. [45]. The inoculated leaves were collected at 1h, 3h, 6h, 12h, 24h, 2d, 4d, and 7d post inoculation. The leaves were cut into 5 cm long segments and directly mounted onto glass slides. The specimens were examined immediately using a Zeiss Axioplan 2 Imaging Research Microscope with ApoTome confocal component (Carl Zeiss Light Microscopy, Germany) equipped with filter blocks with spectral properties matching those of GFP.

**RESULTS**

*SNOG_20078 was identified as the SnTox1-encoding gene*

Whole genome reference sequences have proven to be powerful for the identification of fungal and oomycete effector genes [1, 41]. The annotated *S. nodorum* genome sequence supports a minimum of 10,762 nuclear genes with 1,782 predicted to encode extracellular proteins [33]. A specific set of criteria was used to prioritize the genes and generate a list of candidates. The criteria (size less than 30 kDa, predicted to be secreted, expressed in planta, etc, see Materials and Methods) were based on the characteristics of the previously cloned *SnToxA* and *SnTox3* genes. We focused on the top 100 genes and as expected, *SnTox3* and *SnToxA* were identified among them. PCR analysis was conducted to confirm the absence of genes in the *S. nodorum* avirulent isolate Sn79-1087 (data not shown). Genes meeting these criteria were expressed in the *Pichia pastoris* heterologous expression system [14]. This process and the subsequent screening of a set of differential lines (see Materials and Methods) led us to identify *SNOG_20078* as the SnTox1-encoding gene.

Culture filtrates of *P. pastoris* strain X33 transformed with the coding region of *SNOG_20078* cDNA were infiltrated into the leaves of W-7984, Chinese Spring (CS), CS 1BS-18 and CS ems237. W-7984 and CS carry the dominant Snn1 allele that confers sensitivity to SnTox1 [27]. CS 1BS-18 and CS ems237 are nearly identical to CS, but harbor mutations at
the Snn1 locus, resulting in insensitivity to SnTox1. Necrosis developed in the SnTox1-sensitive lines W-7984 and CS, but not in CS 1BS-18 and CS ems237 (Figure 1) suggesting that SNOG_20078 was the SnTox1-encoding gene. To map the gene conferring sensitivity, the same culture filtrates were subsequently infiltrated into the entire ITMI mapping population, which segregates for Snn1/snn1. All lines sensitive to the partially purified native SnTox1 [27] were also sensitive to the culture filtrates of the SNOG_20078 transformed yeast strain. This strongly indicated that SNOG_20078 was the SnTox1-encoding gene and therefore we designated it SnTox1.

**SnTox1 gene structure and genomic location**

SnTox1 is located in supercontig 10 of the assembled SN15 genome sequence ([33], Figure 2A). Within a ≈7.6 kb region, there are three genes upstream (SNOG_07154-SNOG_7156) and one downstream (SNOG_07153) of SnTox1 (Figure 2A). Similarly, there is a short truncated molly-type retrotransposable element (183 bp) sequence following SnTox1 (http://genome.jgi-psf.org/cgi-bin/browserLoad/?db=Stano1&position=scaffold_10). The sequencing of the 5’ and 3’ RACE fragments revealed three exons as well as 5’ and 3’ untranslated regions (UTRs) in the full-length transcript of SnTox1 (Figure 2B). Putative TATA and CAAT boxes were identified 114 bp and 570 bp upstream of the ATG start site, respectively (Figure S1).

The SnTox1 protein consists of 117 amino acids with the first 17 amino acids predicted as a signal peptide. Interestingly, 16 of the remaining 100 amino acids are cysteine residues (Figure 2C). No prosequence was predicted using the web-based program ProP 1.0 (http://www.cbs.dtu.dk/services/ProP/) and after the cleavage of the predicted signal sequence the mature protein was estimated to be 10.33 kDa. To demonstrate that SnTox1 was produced in yeast culture and to verify the size of SnTox1, we applied western blot analysis to the protein samples prepared from SnTox1 yeast culture filtrates. The antibody for SnTox1 was generated from rabbit immunized with a BSA-conjugated 14 amino acid long SnTox1 peptide (see Material and Methods). A single band was only observed in protein samples prepared from SnTox1 yeast culture filtrates, but not from the control culture filtrates (yeast strain transformed with an empty
vector). Furthermore, the western band was visualized between the size standard of 10 and 15 kDa, but much closer to 10 kDa (Figure S2). The estimated size of SnTox1 obtained from the western blot agreed with the predicted molecular weight of 10.33 kDa for the mature protein.

A BlastP search of the NCBI non-redundant database with the SnTox1 protein sequence as a query led to the identification of three putative proteins with unknown functions, one from *S. nodorum* (SNOG_06487) and two from *P. triticum-repentis* (PTRT_04748 and PTRT_03544) with similarities of 38%, 56%, and 43%, respectively. The conserved amino acids between SnTox1 and these proteins were mostly distributed in the predicted signal sequence and the N-terminal region of the mature protein (Figure S3).

**SnTox1 contains a C-terminal chitin-binding-like motif**

Amino acid alignment with manual adjustment indicated that SnTox1 contained local similarity with cysteine-rich *Cladosporium fulvum* Avr4-like fungal effectors (Figure 3A) from *Cercospora beticola*, *Mycosphaerella fijiensis* [42] and two ascomycete human pathogens, *Microsporum gypseum* and *Geomyces pannorum* (this study). These conserved motifs were identified within the chitin-binding domain (ChtBD) including the C-terminal conserved chitin-binding (CB) motif (Figure 3A). Three-dimensional (3D) structure-based sequence alignment suggested that the putative CB motif in SnTox1 was more similar to those of plant-specific ChtBDs (ChtBD1, or CBM18 superfamily, pfam00187) than to Avr4 proteins, which are related to invertebrate ChtBDs (ChBD2, or the CBM14 superfamily, pfam01607) [43] (Figure 3B). SnTox1 contained all secondary-structure-related residues including the strictly conserved β-strand-forming “CCS” motif found only in plant-specific ChtBD1 proteins [44] (Figure 3B). In contrast, all Avr4-like proteins lacked the “CCS” motif and had a loosely conserved “QWN” motif at the same positions as that found in the antimicrobial protein tachycitin, a representative ChtBD2 [44]. There were several insertions found between conserved regions in SnTox1 which also lacked the C-terminal extension after the conserved CB motif, suggesting a significant sequence divergence between SnTox1 and Avr4-like proteins.
SnTox1 is present in most virulent isolates and absent in avirulent isolates

The distribution of SnTox1 in different S. nodorum isolates and related fungal species (Table S1 and S2) was investigated using PCR assays and DNA dot blots. Among the 777 isolates that were sampled from wheat fields around the world, 85% (661) possess the SnTox1 gene (Table S1). Dot blot analysis of a subset of a global collection (Table S2) showed that SnTox1 was absent in all S. nodorum isolates collected from wild grasses which are avirulent on wheat (Figure 4A). Additionally, SnTox1 was absent in related fungal species including P. tritici-repentis, P. teres, P. bromi and M. graminicola.

To investigate sequence variation in SnTox1, the gene was PCR-amplified and sequenced from 159 global S. nodorum isolates. We found 12 different nucleotide haplotypes, 11 of which encode different protein isoforms, consistent with strong diversifying selection. The 11 protein isoforms involve amino acid changes at eight positions within SnTox1; however, all cysteine residues remain unchanged across all isoforms (Figure 2D). The nucleotide sequences of all 12 haplotypes have been submitted to GenBank and the accession number for each haplotype is provided at the end of the text. Four codons exhibit significant positive selection using PAML (Table 1). These findings provide strong evidence that positive diversifying selection, consistent with a co-evolutionary process, has been operating on SnTox1.

To investigate sequence variation of the SnTox1 genomic region in virulent and avirulent isolates, we used PCR to amplify the four genes flanking SnTox1 (SNOG_07153, SNOG_07154, SNOG_07155, and SNOG_07156, see Figure 2A for their locations). Only SNOG_07154 located directly upstream of SnTox1 could not be amplified from the avirulent isolate Sn79-1087 (data not shown), which suggested that a region containing all or part of SNOG_07154 as well as the entire SnTox1 sequence may be missing in Sn79-1087. PCR primers were designed within the two genes SNOG_07153 and SNOG_07155 and used to amplify DNA from different virulent isolates as well as Sn79-1087. The amplified fragment in SN15 was ~4.1 kb as expected but 4.5kb in Sn1501 and 2.3kb in Sn79-1087 (Figure 4B). Sequencing revealed that a 3.1 kb region including SnTox1 and the last 85 bp of the 3’ end of SNOG_07154 coding region was replaced by a 1.3 kb sequence in Sn79-1087 (Figure 4C). The 1.3 kb insertion in Sn79-1087 does not share homology with any other known sequence in the NCBI database.
revealed that two indels occur in the SnTox1 genomic region with one indel of 400 bp in the upstream, and the other indel of 167 bp at the end of the 3’UTR region (Figure 4C).

Addition of the SnTox1 gene to an avirulent isolate is sufficient to change the host range

The avirulent isolate Sn79-1087 does not produce any known S. nodorum necrotrophic effectors, nor does it induce a susceptible response on any of the wheat lines that we have tested. Therefore, a 1.1 kb SnTox1 genomic region (Figure S1) containing the native promoter, open reading frame, and the native terminator was cloned into the pDAN vector (Figure S4A) and transformed into Sn79-1087. Southern blot analysis indicated all but one transformant possessed the 1.1 kb SnTox1 fragment (Figure S4B). Transformants A1 and A3, designated Sn79+SnTox1A1 and Sn79+SnTox1A3, were selected for further analysis. We confirmed that culture filtrates of Sn79-1087 did not cause necrosis nor did spore inoculations cause disease on CS, which contains Snn1 (Figure 5A). However, infiltration of culture filtrates from Sn79+SnTox1A1 and Sn79+SnTox1A3 produced necrosis on the leaves of CS (Figure 5A) and inoculation of CS with conidia of Sn79+SnTox1A1 and Sn79+SnTox1A3 produced lesions on the leaves of CS (Figure 5B). The two transformants did not cause disease on CS 1BS-18 or CS ems237, which lack a functional Snn1 gene (Figure 5B).

Deletion of SnTox1 in virulent isolates renders them nonpathogenic on Snn1 differential lines

The virulent isolate Sn2000 was used in the original identification of SnTox1 and Snn1 [27]. Therefore, this isolate was used to conduct gene knock outs of SnTox1. We exploited a PCR-based split marker method to replace the majority of the SnTox1 gene with the hygromycin resistance gene (hygR). The transformants were verified using Southern blot analysis with a probe amplified from the SnTox1 region that was replaced by hygR (Figure S4C). In two transformants designated Sn2000ΔSnTox1-9 and Sn2000ΔSnTox1-15, the SnTox1 gene was successfully replaced, and one transformant designated Sn2000ΔSnTox1-ECT was identified as an ectopic insertion due to it being hygromycin resistant but still having an intact and functional SnTox1 gene (Figure S4D).
Spores of the three transformed fungal strains along with wild type Sn2000 were inoculated onto the Snn1 differential wheat line W-7984 [27]. The ectopic strain Sn2000ΔSnTox1-ECT induced similar reaction as the wild type including defined tan necrotic lesions with widespread small white flecking, whereas the two knockout strains induced almost no reaction on the leaves (Figure 6A) indicating SnTox1 is an important virulence factor for Sn2000. Sn2000ΔSnTox1-9 and the Sn2000 wild type were also inoculated onto CS. Compared to the wild type, the virulence of Sn2000ΔSnTox1-9 on CS was substantially reduced, but not completely eliminated (Figure 6B), which is due to CS having at least one additional necrotrophic effector sensitivity gene that likely interacts with another effector produced by Sn2000 (Faris and Friesen, unpublished).

The wheat ITMI population was used to originally map the QTL associated with disease susceptibility caused by Sn2000, in which two significant QTL were identified, one on chromosome 1BS corresponding to the Snn1 locus and the other on chromosome 4BL, explaining 48% and 9% of the disease, respectively [45]. We inoculated the three fungal strains: Sn2000ΔSnTox1-9, Sn2000ΔSnTox1-15 and Sn2000ΔSnTox1-ECT along with wild type Sn2000 onto the ITMI population. For Sn2000, as expected, we detected two significant QTL with one being at the Snn1 locus and the other being on chromosome 4BL accounting for 50 and 17% of the disease variation, respectively. A very similar result was obtained for Sn2000ΔSnTox1-ECT where the Snn1 QTL and the QTL on chromosome 4BL were detected explaining 50 and 15% of the variation, respectively (Figure 6C). However, in the inoculation with the two SnTox1 knockout strains, the QTL conferred by Snn1 on chromosome 1BS became undetectable showing no association with disease, but the QTL on chromosome 4B was retained and became more significant overall accounting for 40-50% of the disease variation (Figure 6C). The QTL analysis in the ITMI population clearly demonstrated that SnTox1 codes for the SnTox1 protein which plays a significant role in disease by interacting with the host gene Snn1.
The transcription of *SnTox1* peaks at 3 days post inoculation correlating with the onset of necrotic lesion development

*SnTox1* had a very similar expression pattern as *SnToxA* and *SnTox3* during infection in a microarray analysis that examined the expression of all fungal genes at 3, 5, 7, and 10 days post inoculation (DPI) in the wheat cultivar ‘Amery’ inoculated with SN15 (Ip-Cho and Oliver unpublished data). The analysis showed that the expression of all three genes was highest at 3 DPI (Figure S5). In this work, *SnTox1* expression was examined after inoculation of CS with *Sn79+SnTox1A1*, in which no other toxin-sensitivity gene interactions were involved. In the current study, relative expression of *SnTox1* to the fungal actin gene was examined at 10 time points ranging from 1 h to 7 d post inoculation using relative-quantitative PCR. Our analysis confirmed that *SnTox1* expression was maximized at 3 DPI (Figure 7A). The expression of *SnTox1* showed a slow increase between 6 and 12 HPI, increasing to about the same level as the actin gene at 24 HPI and increasing dramatically to 2.5 times higher than the actin gene expression at 48 HPI (Figure 7A). Once gene expression peaked at 3DPI, the *SnTox1* transcription levels started to drop significantly from 3 to 4 DPI and returned to similar levels as the actin gene between 5 and 6 DPI. The accelerated increase of *SnTox1* expression from 24 HPI to 3 DPI indicates that *SnTox1* plays an important role in the early stage of infection.

The symptom development was examined macroscopically on CS inoculated with *Sn79+SnTox1A1* (Figure 7B). Disease symptoms were first visible on leaves at 2 DPI as white flecks and progressed into larger necrotic and chlorotic lesions. Interestingly, tan necrotic lesions start to develop at 3 DPI within the chlorotic areas, which correlates with the maximum expression of *SnTox1* (Figure 7A). By 5 DPI, necrotic lesions became evident and the chlorotic areas enlarged (Figure 7B). The overall phenotype of the lesions changed very little from 5 to 7 DPI with only a slight change in size of individual lesions (Figure 7B).

**The SnTox1 protein contains 16 cysteine residues that likely form multiple disulfide bonds and are important for SnTox1 stability**

The SnTox1 protein contains 16 cysteine residues all of which are predicted to be involved in the formation of disulfide bonds with confidence levels greater than 7 (0 to 9 scale,
Figure 8A). The prediction software DiANNA [47] was used to identify the most likely connectivity of the cysteine residues as following: 1-11, 2-5, 3-6, 4-13, 7-9, 8-16, 10-12, and 14-15 (Figure 8B). The stability of SnTox1 was tested by incubation of an SnTox1-containing yeast culture filtrate with different concentrations of dithiothreitol (DTT) and different incubation periods. The complete elimination of SnTox1 activity required 4 h in 40 mM DTT (Figure 8C). Additionally, the stability of SnTox1 was tested by directly heating the SnTox1 yeast culture filtrates on a hot plate. Strikingly, the culture filtrates maintained necrotic activity even after boiling for 30 min and did not completely lose activity until after 1 h (Figure 8D). Together, these results show that SnTox1 is a highly stable protein with the ability to resist physical and chemical degradation.

**SnTox1 triggers an oxidative burst**

The oxidative burst is one of the best-known biochemical responses of plant cells during a resistance response. The oxidative burst can be visualized by 3′-3′ diaminobenzidine (DAB) staining for H$_2$O$_2$ production [48]. Chinese Spring (CS, Snn1) wheat leaves were infiltrated with SnTox1 yeast culture filtrate or a control yeast culture filtrate and collected at 48 h post-infiltration. The CS ems237 line (snn1) was included for infiltration and DAB staining as a comparison. Leaves were stained with 1 mg/ml DAB solution followed by clearing of chlorophyll. Dense brown DAB staining was observed on the leaves of CS (Snn1) infiltrated with SnTox1, but DAB staining did not appear on leaves of CS infiltrated with the control culture filtrates deficient in SnTox1, nor did DAB staining appear when SnTox1 was infiltrated into leaves of CS ems237 lines (snn1) (Figure 9A), clearly showing that the production of H$_2$O$_2$ is induced only during the SnTox1-Snn1 interaction. A control without DAB staining was also conducted on CS leaves infiltrated with SnTox1 yeast culture. After clearing the leaf, no browning was observed indicating that, in the absence of DAB, the SnTox1 reaction itself was not able to cause brown staining on the leaf (Figure 9A).

The production of H$_2$O$_2$ was also detected during the fungal infection. The CS leaves inoculated with Sn79+SnTox1A1 were collected daily from 1 to 7 days post inoculation and stained with DAB followed by the same procedure for leaf clearing. The accumulation of brown
staining on the leaf was readily visible from 2 DPI (Figure 9B). The generation of reactive oxygen species (ROS) associated with a hypersensitive response \textit{in planta} often occurs in the chloroplast [49]. Using DAB stained CS leaves from the SnTox1 infiltration, we observed that chloroplasts had the highest intensity of brown color (Figure 9C).

\textbf{SnTox1 triggers stronger up-regulation of PR-genes}

Up-regulation of plant defense or signaling pathway genes including pathogenesis-related (PR) genes are hallmarks of a resistance response. Using RT-PCR, we examined the transcription level of 28 wheat genes (Table S3) in CS (\textit{Snn1}) and CS ems237 (\textit{snn1}) leaves that were collected at different time points from 1 h to 72 h after being infiltrated with SnTox1 culture filtrates as well as control culture filtrates. Three genes including PR-1-A1, a thaumatin-like protein gene, and a chitinase were found to be significantly up-regulated in CS leaves infiltrated with SnTox1 compared to the control leaf samples infiltrated with culture filtrates deficient in SnTox1 (Figure 10A). In the CS ems237 line which has a mutated \textit{snn1} gene, a transcript was undetectable for the \textit{PR-1-A1} gene and was at a significantly lower level for the \textit{thaumatin} and \textit{chitinase} genes as detected by RT-PCR (Figure 10A). Quantitative PCR (qPCR) analysis confirmed the higher expression of the three genes in SnTox1 infiltrated CS leaves compared to the control infiltrated CS leaves. Not only did all three genes show maximum expression at 36 HPI, but each had at least two-fold higher expression in SnTox1-infiltrated samples than the control (Figure 10B). qPCR also showed much lower expression of the three genes in the CS ems237 line infiltrated with either SnTox1 or the control yeast culture filtrates in comparison to CS infiltrated with control culture filtrates (Figure 10B). The reason for this is not clear, but it could be explained by the idea that \textit{Snn1} may play a role in sensing other environmental stimuli that can trigger PR gene expression.

\textbf{The SnTox1-\textit{Snn1} interaction triggers DNA laddering}

Programmed cell death (PCD) triggered by biotrophic effectors is often evidenced by DNA laddering in plants [16, 50]. To determine if the necrosis induced by SnTox1 on \textit{Snn1} lines was a result of PCD, we isolated DNA from infiltrated CS leaf samples and checked for evidence
of DNA laddering. For negative comparisons where no necrosis developed, DNA fragmentation was also examined in CS leaves infiltrated with control culture filtrates (no SnTox1) and CS ems237 (mutated \textit{snn1}) infiltrated with SnTox1 or control culture filtrates. In the CS leaf samples infiltrated with SnTox1, DNA laddering was detected as early as 10 h after infiltration and was most evident at 36 h after infiltration (Figure 11); however, in the leaf samples from the other three treatments, no DNA laddering was observed at any time point (Figure 11), indicating that SnTox1-induced necrosis on lines harboring \textit{Snn1} is a result of host-controlled PCD.

**Light is required for the development of necrosis and disease induced by SnTox1**

Light has been found to be important in the development of necrosis induced by necrotrophic effectors from \textit{P. tritici-repentis} and \textit{S. nodorum} [12, 37]. Therefore, we investigated whether the development of necrosis induced by SnTox1 as well as the disease development caused by the SnTox1-\textit{Snn1} interaction was light dependent. After infiltration with SnTox1 yeast culture filtrates, CS plants were incubated in a growth chamber but covered for 2 days. The plants in the dark did not exhibit a necrotic reaction in the infiltrated area on the leaves, while the plants grown in the same growth chamber without covering showed necrosis (Figure 12) indicating the development of necrosis induced by SnTox1 is light dependent. Interestingly, necrosis did develop on the dark treated plants once they were treated with a 12 h light-dark cycle for 2 additional days.

A very similar situation was observed in the inoculation experiment. CS leaves showed no disease symptoms at 3 days post inoculation when plants were kept in the dark and similar to the infiltration experiment, the lesions developed once the dark-treated plants were moved to the light again (Figure 12).

**SnTox1 has an important function during the penetration period**

To investigate the role of SnTox1 in disease development, we tagged both the avirulent isolate Sn79-1087 and the pathogenic strain \textit{Sn79+SnTox1A1} with GFP and compared their infection processes by fluorescence microscopy in wheat lines CS (\textit{Snn1}) and the \textit{Snn1} mutant, CS ems237 (\textit{snn1}). The inoculation of CS with the SnTox1-producing strain Sn79+SnTox1A1
resulted in an infection (susceptible interaction); however, the other three combinations (CS inoculated with Sn79-1087, CS ems237 inoculated with Sn79-1087, and CS ems237 inoculated with Sn79+SnTox1A1) gave no disease (resistant interaction) (Figure 13). Within 24 HPI, there was little difference observed between resistant and susceptible interactions. During this period, conidia germinated, grew short hyphae and began the penetration process. The pathogen was able to initiate penetration in both types of reactions visualized by the formation of the indistinct penetration structure called a hyphopodia [26]; Figure 13 A, B) and by autofluorescence of the damaged epidermal cell walls (Figure 13 A, B). We observed mainly direct penetration of the leaf surface over both periclinal and anticlinal epidermal cell walls.

A strong green autofluorescence was observed beneath the epidermis by 2 DPI in the susceptible interaction, suggesting that the pathogen had successfully penetrated through the epidermal cell layer and started the infection of mesophyll cells (Figure 13C). However, in the resistant interaction, the pathogen grew extensively on the leaf surface and no green autofluorescence was visible (Figure 13D). At 4 DPI, the infection area had enlarged in the susceptible interaction as shown by more mesophyll cells producing a fluorescent signal (Figure 13E). On the leaves of the resistant interactions, most of the fungal mycelium was dead, likely due to scarcity of nutrients, and only a few hyphae continued to grow over the leaf surface with repeated unsuccessful attempts to penetrate (Figure 13 F). The susceptible interaction had induced widespread lesion formation on the leaves by 7 DPI, however, no symptoms were found on the leaves of the resistant interaction (Figure 13 G, H). Examination under the fluorescent microscope of the necrotic lesion formed from the susceptible reaction revealed the extensive growth of fungal mycelium within the lesion (Figure S6).

**Light is required for penetration during infection**

The fungal infection process was also compared microscopically on Snm1-containing plants that were either grown under a normal light/dark cycle or in complete darkness after inoculation. The pathogen was able to germinate and generate hyphopodia within 24 HPI in both conditions (data not shown). However at 48 HPI, only the plants grown in a normal light/dark cycle showed successful penetration through the epidermal cell layer and the initiation of the
infection of mesophyll cells, evidenced by the autofluorescence of the mesophyll cells (Figure 14A). In the plants that were kept in the dark, no autofluorescence was observed in the mesophyll cells and the pathogen still remained on the leaf surface without having successfully penetrated the epidermis (Figure 14B).

DISCUSSION

Molecular cloning of SnTox1

The necrotrophic fungal pathogen *S. nodorum* produces multiple necrotrophic effectors (host-selective toxins) that function as virulence factors during the infection process. The cloning of these necrotrophic effector genes is an essential step in the characterization and elucidation of the molecular and biochemical mechanism of fungal pathogenesis in the wheat-*S. nodorum* pathosystem. Besides the traditional biochemical and genetic tools, new genomic strategies have been recently applied for the identification and cloning of effector genes in a number of fungi and oomycetes as more genome and other sequence data becomes available. A typical procedure would include a process of data mining to identify candidate genes that meet a set of specific criteria followed by gene validation through functional analysis. High throughput functional genomics [1] as well as comparative genomics and association genetics [41] have been successfully used for the identification of pathogen effector genes in fungi and oomycetes. In the current study, we used a set of criteria to mine the *S. nodorum* genome sequence dataset for the identification of necrotrophic effector genes. This strategy led to the successful identification of SnTox1 from *S. nodorum*. Through heterologous expression, gene transformation, and gene disruption, we have provided convincing evidence that the candidate gene SNOG20078 (Gene ID: 5974395) is the SnTox1-encoding gene. This research further highlights the value of genome sequence data along with efficient bioinformatics tools in identifying effector genes. We are continuing to use this strategy to identify additional *S. nodorum* necrotrophic effector genes.
The genomic location of *SnTox1*

*SnTox1* was identified using a set of criteria based on the cloned *S. nodorum* effector genes *SnToxA* and *SnTox3*; however, the *SnTox1* gene does have some unique features. Unlike many previously identified effector genes including those from *Leptosphaeria maculans* [51-53], *Magnaporthe grisea* [41, 54], *Fusarium oxysporum* f. sp. *lycopersici* [55, 56], *Blumeria graminis* f. sp. *hordei* [57], and those from several *Phytophthora* species [58], *SnTox1* lies in a gene-rich region and was flanked closely by other genes. Except for a short (~300bp) sequence predicted to be a truncated molly-type RE, no other obvious RE or AT-rich sequences were identified within the 300 kb genome region surrounding *SnTox1* (http://genome.jgi-psf.org/Stano1/Stano1.info.html) showing that not all effector genes are associated with an abundance of repetitive or transposable elements.

The occurrence of effector genes in close proximity to one another has also been reported for several fungal and oomycete pathogens [53, 59-62]. This does not appear to be the case for *S. nodorum*. The three *S. nodorum* effector genes (*SnToxA*, *SnTox1*, and *SnTox3*) were located on different supercontigs and have been shown by pulse field gel electrophoresis and Southern analysis to reside on 2.35, 1.88 and 1.66 Mb chromosomes, respectively, in SN15 (data not shown) indicating that these genes are not clustered.

The population genetics of *SnTox1*

Using a worldwide collection of 777 *S. nodorum* isolates, *SnTox1* was found to be present in ~85% of isolates, a markedly higher frequency than found for *SnToxA* (~36%) and *SnTox3* (~60%) [14, 63]. Like the other NEs, *SnTox1* was shown to have a presence/absence polymorphism within individual wheat fields. This type of polymorphism has been reported in other fungal pathosystems, as reviewed in Stergiopoulos and de Wit [64]. The frequency of *SnTox1* varied significantly across regional populations. We hypothesize that regional differences in the frequency of *SnTox1* reflect regional differences in the frequency of *Snn1*. However this correlation was not apparent when tested on a small worldwide collection of wheat. We found that *Snn1* is most prevalent in durum wheat lines and much more rare among hexaploid bread wheat lines throughout the world (data not shown). This could indicate that the maintenance of
Sn1 in durum wheat is associated with another important trait. Widespread deployment of wheat cultivars lacking Sn1 could cause the frequency of SnTox1 to decrease if there is a fitness cost associated with producing the effector. But the large effective population sizes of S. nodorum [65] make the complete loss of the effector through genetic drift unlikely.

Observed diversity at the SnTox1 locus was found to fit a model of diversifying selection significantly better than a neutral model. Positive selection was found for 4 of the SnTox1 codons, consistent with the growing list of prokaryotic and eukaryotic effector candidates that exhibit positive selection [66]. None of the non-synonymous substitutions were found in the signal peptide, the putative chitin-binding domain, the putative Avr4-like domain or any of the cysteine codons. This suggests that the effector’s functional domains were preserved, while more flexible amino acid sites were subject to diversifying selection. Possible differences in activity between different protein variants of SnTox1 are currently being tested.

**SnTox1 is a small secreted and cysteine rich necrotrophic effector**

Similar to SnToxA and SnTox3, SnTox1 was shown to play a significant role in disease development. Results presented here on the SnTox1-Snn1 interaction provide further evidence that the necrotrophic wheat-S. nodorum system is largely based on specific host-effector interactions that act in ETS [14, 15], which essentially has the opposite outcome of ETI that operates in many biotrophic systems [3, 4].

One of the most striking features of the SnTox1 protein as a necrotrophic effector is the high cysteine residue content. This feature is often associated with fungal avirulence gene products such as the Avr and ECP effectors from Cladosporium fulvum [64], SIX (secreted in xylem) effectors from Fusarium oxysporum f.sp. lycopersici [55], and Nip1 from Rhynchosporium secalis [67]. The predicted mature SnTox1 protein has 100 amino acids, 16 of which are cysteine residues, the richest of all effectors that have been identified. The high content of cysteine residues and high stability suggest that SnTox1 may function in the plant apoplastic space which is abundant in plant defense components. We are currently investigating the subcellular location of SnTox1. Most small cysteine-rich secreted effectors from the tomato fungal pathogen C. fulvum such as Avr2, Avr4, Avr9, and ECP2 are thought to function
exclusively in the apoplast to inhibit and protect against plant hydrolytic enzymes [64]. ECP6, another *C. fulvum* effector containing LysM chitin binding domains was recently found functioning apoplastically as a scavenger of fungal chitin to prevent it from eliciting PAMP-triggered immunity *in planta* [68]. Interestingly, we observed that SnTox1 has some similarity to *C. fulvum* Avr4 within the chitin-binding domain and in the positions of six of the cysteine residues at the C-terminus. However, further tests are needed to determine the binding activity and functional roles of the putative CB domain in SnTox1. The presence of a potential chitin binding domain provides a point of investigation for an added function for SnTox1 in addition to its interaction with *Snn1*.

**The SnTox1-*Snn1* interaction enables direct penetration**

Successful penetration is a prerequisite for a pathogen to establish itself and fulfill its colonization *in planta*. For *S. nodorum*, previous studies have observed direct penetration through the junction of epidermal cells [69] or penetration through stomata [70] or both [26]. Based on our observation using GFP-tagging and confocal fluorescent microscopy, the fungus predominantly used direct penetration through the junction of epidermal cells, and stomatal entry was not evident. We have observed that the fungal mycelium grew over guard cells and anchored the penetration point between the junction of the guard cell and the adjacent epidermal cell instead of going through the stomata (data not shown). Although the avirulent isolate belongs to *S. nodorum*, the preference for direct penetration, which is different from that reported by Solomon et al. [26], may be due to its adaptation to wild grasses from which it was originally isolated.

It was our observation that the fungus could initiate direct penetration by producing a hyphopodia in both the resistant and the susceptible interaction with little difference, which agrees with previous reports [69] indicating that SnTox1 is not required for hyphopodia formation and the initial degradation of the cuticle layer and the cell wall between the junctions of the epidermal cells. Hydrolytic enzymes or other mechanisms may be employed by the fungus to breach the initial physical barrier. Several cell wall-degrading enzymes such as amylase, pectin methyl esterase, polygalacturonases, xylanases, and cellulase have been found to be
produced in vitro and during the infection of wheat leaves by S. nodorum [71]. As infection progressed, the pathogen was unable to penetrate through the epidermal cell layer and therefore could not reach the mesophyll cells to establish a successful infection without the SnTox1-Snn1 interaction. This suggests that SnTox1 is significant in the initial penetration process across the epidermal cell layer. Our hypothesis is that SnTox1 interacts with Snn1 to induce cell death in epidermal cells, providing the fungus with nutrients for further invasive growth. In Cochliobolus victoria on oat and Arabidopsis systems, it was also observed that fungal penetration ceases following appressorium development and hyphae remain on the leaf surface in the absence of a compatible interaction, which requires both victorin and its corresponding sensitivity gene [22].

Our speculation was further supported by the fact that the inoculation of an Snn1 line (CS) with SnTox1 transformed avirulent isolates induced widespread necrosis (presumably programmed cell death) on leaves. Furthermore, inoculation with the SnTox1-knock out virulent strain lost the ability to cause this necrotic reaction. Additionally, qPCR revealed that SnTox1 expression was induced in planta starting as early as 12 HPI and increased at an accelerated rate from 12 to 24 HPI when the fungus was observed to penetrate. Collectively, this suggests that S. nodorum may use SnTox1 to induce cell death in the epidermal cells, providing a portal to enter the plant and subsequently feeding from dead cells to gain nutrients for further invasive growth.

**SnTox1 induces a light dependent reaction**

It is well known that plant defenses against pests and pathogens are commonly influenced by environmental conditions, including light. Many studies have demonstrated the requirement of illumination for the interaction of plants with a diversity of bacterial and fungal pathogens as well as the isolated pathogenic elicitors [72, 73]. The effect of light on the disease development of SNB was first noticed by Baker and Smith [69] who observed that the necrotic reaction and lesion coalescence tended to be suppressed in the absence of light. The necrotrophic effector ToxA, was also shown to induce light-dependent necrosis on Tsn1 lines [37]. Among the S. nodorum necrotrophic effectors published to date, all effectors except SnTox3 have been shown to be light dependent [13, 32]. Using heterologously expressed SnTox1 and the avirulent isolate carrying the SnTox1 gene, we showed clearly that the necrosis and disease susceptibility induced
by SnTox1 on *Snn1* lines were completely dependent on light. The requirement of light for resistance to biotrophic disease as well as susceptibility to necrotrophic disease suggests a common host mechanism shared by reactions to the two classes of disease interactions.

The molecular mechanism underlying the light dependency of plant pathogen interactions is still poorly understood; however, research on the ToxA-*Tsn1* interaction has shown that ToxA is internalized in the plant cell followed by localization to the chloroplast and induction of photosystem alterations (reviewed in [40]), providing a hint for the influence of light on this interaction. Recently, it was demonstrated that *Tsn1* is regulated by light and its expression is significantly suppressed in the dark [15], providing a possible explanation for the light dependency of the ToxA-*Tsn1* interaction. SnTox1 is cysteine rich and therefore possibly acts in the apoplastic space. If SnTox1 remains in the apoplastic space, different mechanisms would likely be involved even though both are dependent on light. In Arabidopsis, plants kept in the dark do not accumulate H$_2$O$_2$ in the chloroplasts and show significantly delayed HR cell death after a resistance signaling pathway is activated [49]. This indicates that light is required for H$_2$O$_2$ production in chloroplasts and that this H$_2$O$_2$ production is critical to programmed cell death. The DAB staining in CS (*Snn1*) leaves infiltrated with SnTox1 was found to be associated with the chloroplast and the CS plant infiltrated with SnTox1 showed no DAB staining if kept in the dark (data not shown), suggesting a similar mechanism underlying SnTox1-induced cell death. Very interestingly, we found that plants kept in the dark developed necrosis and disease symptoms once transitioned to a normal photoperiod. Therefore signal transduction appears to pause rather than stop in the absence of light. This may indicate that the SnTox1 signal is progressing to the chloroplast but this process is interrupted in the absence of light.

**SnTox1 functions like a biotrophic effector to induce PCD in *Snn1* host lines**

Biotrophic effectors often function as elicitors of programmed cell death (PCD) thereby activating the resistance response in host plants containing the corresponding resistance genes. The host resistance reaction begins with the direct or indirect recognition of the pathogen-produced effector by the resistance gene product, followed by a complicated signaling pathway and a series of biochemical and physiological responses in host plant cells [74]. The host
response often includes an oxidative burst, cell wall restructuring, PR-gene expression and antimicrobial compound production culminating in a localized cell death at the infection site. This PCD is known as a hypersensitive response and is typically aimed at halting further colonization by the pathogen [5]. A set of biochemical tests has shown that SnTox1 is able to induce resistance-like host responses and PCD evidenced by the \( \text{H}_2\text{O}_2 \) production, stronger expression of PR-genes, and DNA laddering in lines carrying \( Snn1 \). It is important to note that SnTox1 physiologically evoked a widespread necrotic flecking on the \( Snn1 \) line, which is symptomatically similar to the hypersensitive response in biotrophic disease systems. However, this necrosis spreads into larger lesions resulting in susceptibility (sporulation) rather than resistance (prevention or inhibition of sporulation). Together, this indicates that SnTox1 is likely functioning biochemically and physiologically similar to a biotrophic effector (avirulence factor) in the presence of \( Snn1 \) but with a different end result.

A number of other necrotrophic effectors have also been shown to invoke a host resistance response [9, 17, 40]. It has generally been thought that necrotrophic plant pathogenic fungi possess simplistic infection mechanisms that rely on lytic and degrading enzymes [11]. In contrast, biotrophic fungal pathogen interactions have been considered more sophisticated due to the formation of special penetration and feeding structures, secretion of effectors to overcome plant PAMP triggered immunity and a constantly changing effector complement to avoid recognition by the plant innate immune system. However, three genes conferring sensitivity to necrotrophic effectors as well as susceptibility to the corresponding necrotrophic fungal pathogens have been cloned, and all possess resistance gene-like features [15, 22, 23]. Therefore, it seems that necrotrophic fungal pathogens may subvert plant resistance mechanisms for their own good. Here, we clearly showed that SnTox1 is an important virulence factor for \( S. \ nodorum \) in the presence of \( Snn1 \) and that the host response to SnTox1 shows several similarities to a classical resistance response induced by many biotrophic effectors, however, the outcome of the host recognition was susceptibility rather than resistance.

\( SnTox1 \) is the third effector gene that we have cloned and characterized from \( S. \ nodorum \), which further strengthens the hypothesis that the wheat-\( S. \ nodorum \) pathosystem is based largely on host-effector interactions. The three effector genes cloned have provided molecular tools to
study the mechanisms underlying disease in this system, an emerging model for necrotrophic fungal diseases.

SUPPORTING MATERIAL

All supplementary files for this chapter are available online

Supporting Table S1
SnTox1 distribution and haplotypes in a global collection of S. nodorum isolates.
Found at: doi:10.1371/journal.ppat.1002467

Supporting Table S2
S. nodorum isolates and its related fungal species used in dot blot analysis of SnTox1 presence.
Found at: doi:10.1371/journal.ppat.1002467

Supporting Table S3
Defense response genes investigated in SnTox1-Snn1 interaction.
Found at: doi:10.1371/journal.ppat.1002467

Supporting Table S4
Primers used in this study.
Found at: doi:10.1371/journal.ppat.1002467

Figure S1
SnTox1 nucleotide and deduced amino acid sequences.
Found at: doi:10.1371/journal.ppat.1002467

Figure S2
Analysis of SnTox1 protein expressed in Pichia pastoris. A.
Found at: doi:10.1371/journal.ppat.1002467

Figure S3
Amino acid sequence alignment of SnTox1 with its homologs obtained from BlastP searches.
Found at: doi:10.1371/journal.ppat.1002467

Figure S4
Molecular manipulation and characterization in SnTox1 gene transformation and disruption.
Found at: doi:10.1371/journal.ppat.1002467

Figure S5
Gene expression patterns for SnTox4, SnTox3 and SnTox1 during infection revealed by microarray analysis.
Found at: doi:10.1371/journal.ppat.1002467

Figure S6
A close-up of a necrotic lesion induced by a GFP-tagged SnTox1 transformed fungal strain.
Found at: doi:10.1371/journal.ppat.1002467
The *S. nodorum* gene *SNOG_20078* has been deposited in Genbank with identity numbers of 5974395 for gene ID and XP_001797505.1 for protein ID. The nucleotide sequence of 12 different haplotypes of *SnTox1*, designated Tox1_H1-H13, was submitted to GenBank with accession numbers from JN791682 to JN791693. The other genes and proteins referred to in this paper included *Cladosporium fulvum* Avr4 protein (CAA69643.1), *Mycosphaerella fijiensis* Avr4-like protein (Protein ID: Mycfi:87167), *Cercospora beticola* Avr4-like protein (GU574324), *Microsporum gypseum* Avr4-like protein (GeneID:10030079) and *Geomycetes pannorum* Avr4-like protein (DY991214).
### Table 1. Summary of a likelihood ratio test using PAML for positive selection in the *SnTox1* gene.

<table>
<thead>
<tr>
<th>Model pairs</th>
<th>$d_s/d_S$</th>
<th>Log likelihood ($l^a$)</th>
<th>$P^b$</th>
<th>Positively selected codons $^c$</th>
<th>Posterior prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1a (nearlyneutral)</td>
<td>0.52</td>
<td>-553.98</td>
<td></td>
<td>48</td>
<td>0.975</td>
</tr>
<tr>
<td>M2a (selection)</td>
<td>4.49</td>
<td>-546.04**</td>
<td></td>
<td>68</td>
<td>0.978</td>
</tr>
<tr>
<td>M1a/M2a LRT</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td>72</td>
<td>0.997</td>
</tr>
<tr>
<td>M7 (beta)</td>
<td>0.50</td>
<td>-554.07</td>
<td></td>
<td>106</td>
<td>0.974</td>
</tr>
<tr>
<td>M8 (beta &amp; $\omega$)</td>
<td>4.49</td>
<td>-546.127**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M7/M8 LRT</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Asterisks indicate which model resulted in a statistically higher likelihood score for the *SnTox1* codon alignment.

$^b$ P-value for the likelihood ratio test (LRT) between the log-likelihoods of comparable models. In both cases the null hypothesis is rejected in favor of the selection model.

$^c$ Specific codons within the amino acid sequence were tested for positive selection. Positively selected codons with a Bayes Empirical Bayes posterior probability higher than 0.95 are listed on the right half of the table [78].
Figure 1. *SNOG_20078* encodes SnTox1.
The reaction of wheat lines W-7984 (*Snn1*), Chinese Spring (CS) (*Snn1*), CS 1BS-18 (*snn1*) and CS ems237 (*snn1*) to culture filtrates of a *Pichia pastoris* strain transformed with *SNOG_20078*. 
Figure 2. Genomic location, structure and deduced amino acid sequence of SnTox1.
A. SnTox1 genomic region. SnTox1 is located in supercontig 10 of the assembled SN15 genome sequence and is surrounded by four other predicted genes (boxed arrows, SNOG7153 to SNOG7156). A short truncated molly-type retrotransposon sequence (gray rectangle) closely follows SnTox1. B. SnTox1 gene structure. The full length transcript of SnTox1 and contains three exons (black rectangles) and both 5’ and 3’ untranslated regions (white rectangles). C. SnTox1 amino acid sequence. SnTox1 protein contains 117 amino acids with the first 17 (in bold) being a predicted signal sequence. The 16 cysteine residues are underlined. D. The alignment of 11 different SnTox1 protein isoforms. Two regions of SnTox1, from 36 to 87 and 97 to 117 are shown to indicate the variable amino acid positions (arrows). The remaining amino acids, (except for the cysteine residues) are shown as stars, indicating they are conserved.
Figure 3. Sequence alignment of SnTox1 with other proteins harboring chitin binding domains.

A. Sequence alignment of SnTox1 with Avr4-like proteins. Three motif-like sequences (TxC at C1, SCT-x-QC at C2 and C3, and CxxG at C4) and six conserved cysteine residues are marked by arrows. CfAvr4 = *Cladosporium fulvum* Avr4 protein (CAA69643.1), MfAvr4 = *Mycosphaerella fijiensis* Avr4-like protein (Protein ID: Mycf1:87167), CbAvr4 = *Cercospora beticola* Avr4-like protein (GU574324), MGYG_02307 = Avr4-like protein from *Microsporum gypseum* and GpDY991214 = Avr4-like protein identified from *Geomyses pannorum*. B. Local alignment of chitin binding (CB) domains from different proteins. Positions of the two antiparallel β-sheets (β4 and β5, arrows), helical turn (α1), disulfide bond (dashed line), and the active sites (asterisks) conserved in CB domains (based on Suetake et al. [44]) are indicated at the top and/or the bottom, respectively. SnTox1 contained all secondary-structure-related residues including the strictly conserved β-strand-forming “CCS” motif found only in plant-specific ChtBD1 proteins.
Figure 4. Distribution of SnTox1 in S. nodorum isolates and variation in its genomic region.
A. DNA dot blot analysis of SnTox1 gene distribution. Among 93 fungal isolates or species (Table S1), SnTox1 is absent in related fungal species (dash-lined box), S. nodorum avirulent isolates (box), and 6 virulent isolates not containing SnTox1(circles). The remaining virulent isolates (72 out of 79) all contained SnTox1. B. PCR amplification of the ~4.1 kb genomic region of SnTox1 in different S. nodorum isolates. Two primers located within SNOG_7153 and SNOG_7155 (see Fig 4C arrows for primer locations) were used to amplify the SnTox1 genomic region in four S. nodorum isolates including the avirulent isolate (Sn79-1087). A difference in fragment size was observed among isolates. C. A diagram of the SnTox1 genomic region in different S. nodorum isolates. Five predicted genes (arrow blocks) and a truncated molly type repeat (gray rectangle) are schematically drawn. The 4.1 Kb region was amplified with two primers in SNOG_7153 and SNOG_7155 (small black arrows) and was investigated by cloning and sequencing. In Sn79-1087, a portion of SNOG_7154 and the entire SnTox1 gene was missing and was replaced by a 1.3 kb region (rectangle). Additionally, two indels (dash-lined rectangle) were identified in the upstream region and the 3’UTR of SnTox1.
Figure 5. *SnTox1* makes an avirulent isolate pathogenic.

**A.** Reaction of Chinese Spring to infiltration with culture filtrates from the avirulent isolate Sn79-1087 and its *SnTox1* transformants. The culture filtrates from Sn79-1087 did not produce necrosis on CS, but the two *SnTox1* transformants Sn79+SnTox1A1 and Sn79+SnTox1A3 did. **B.** Disease reaction of CS (*Snn1*) and CS 1BS-18 (*Snn1* deleted) or CS ems237 (*Snn1* mutated) to the inoculation with the avirulent isolate Sn79-1087 and its *SnTox1* transformants. Sn79-1087 did not cause disease on CS, but the two Sn79+SnTox1A1 and Sn79+SnTox1A3 did produce tan necrotic lesions and widespread flecking on CS (*Snn1*). However, the two transformants were unable to cause disease on CS 1BS-18 and CS ems237 which carry the recessive allele (*snn1*).
Figure 6. *SnTox1* disruption affects virulence only on *Snn1* differential lines. 

A. Reaction of W-7984 to inoculation with Sn2000 and its *SnTox1* disrupted (Sn2000ΔSnTox1-9 and Sn2000ΔSnTox1-15) and ectopic (Sn2000ΔSnTox1-ECT) strains. Compared to the wild type and ectopic strain, the two *SnTox1* disrupted strains completely lost virulence on W-7984 which only contains the SnTox1 sensitivity gene (*Snn1*). 

B. Reaction of CS to inoculation with Sn2000 and its *SnTox1* disrupted strain (Sn2000ΔSnTox1-9). Compared to wild type, the *SnTox1* disrupted strain (Sn2000ΔSnTox1-9) showed significantly reduced virulence on CS which is not only sensitive to SnTox1 but to another necrotrophic effector produced by Sn2000. 

C. Interval
map of chromosome 1B (left) and 4B (right) from QTL mapping in the ITMI population inoculated with Sn2000, and its SnTox1 disrupted and ectopic strains. Strains are depicted by different colors as indicated. A centiMorgan scale is on the left of the map and markers are shown in their relative position along the right. The Snn1 locus on the tip of chromosome 1B is shown in red. An LOD scale is shown along the x axis, and the critical LOD threshold of 3.0 is represented by the dotted lines.
Figure 7. *SnTox1* expression is induced *in planta* during disease development.

A. Expression pattern of *SnTox1 in planta* during disease development. The x axis shows the time points post-inoculation when leaf samples were taken for qPCR. The y axis represents the relative gene expression levels normalized to *Act1*. Standard error bars from three replications are shown. B. Macroscopic examination of disease development on CS inoculated with Sn79+SnTox1A1. The samples of leaves of CS inoculated with Sn79+SnTox1A1 were collected and photographed at 24 h intervals post-inoculation.
Figure 8. SnTox1 protein is cysteine-rich and highly stable.
A. Disulfide bond prediction. The prediction was conducted using the web-based program DISULFIND (http://disulfind.dsi.unifi.it/). AA: amino acid sequence, DB_state: predicted disulfide bonding state (1=disulfide bonded, 0=not disulfide bonded); DB_conf: confidence of disulfide bonding state prediction (0=low to 9=high).
B. The best predicted connectivity of disulfide bonding. The best connectivity of eight disulfide bonds was determined using the web-based program DiANNA 1.1 (http://clavius.bc.edu/~clotelab/DiANNA/).
C. Dithiothreitol (DTT) treatment of SnTox1. Reaction of CS to the SnTox1 yeast culture filtrates that were treated with DTT at the indicated concentration for 2h or 4h at room temperature.
D. Heat treatment of SnTox1. Reaction of CS to the SnTox1 yeast culture filtrates that were heated to boiling on a hot plate for 30 min or 1h.
Figure 9. SnTox1 triggers H$_2$O$_2$ production.
A. H$_2$O$_2$ production in infiltrated leaves. CS (Snnl) or CS ems237(snnl) leaves were infiltrated with SnTox1 (+SnTox1) or control culture filtrates (-SnTox1) and stained with 1 mg/ml DAB (+DAB) followed by clearing of chlorophyll. Leaves infiltrated with SnTox1 were cleared without staining as a control (-DAB) to show that the toxin reaction itself does not cause the brown color. B. H$_2$O$_2$ production in the inoculated leaves. CS leaves were inoculated with Sn79+SnTox1A1 and collected at each day post inoculation for DAB staining followed by clearing of chlorophyll. Photograph was taken before and after DAB staining and clearing. 2 DPI and 7 DPI are shown. C. Location of H$_2$O$_2$ production in plant cells. The DAB-stained CS leaves were examined under the light microscope for the cellular location of DAB staining. The strong DAB staining was localized to the chloroplasts (white arrows) only in CS infiltrated with SnTox1 (400x magnification).
Figure 10. SnTox1-Snn1 interaction induces increased defense gene expression.
A. Expression of three pathogenesis-related (PR) genes using RT-PCR. CS (Snn1) and CS ems237-1 (snn1) leaf samples were collected at 1, 2, 4, 8, 12, 24, 36, 48, 60, and 72 h after infiltration with SnTox1 or control yeast culture filtrates. RT-PCR was conducted to compare the expression of three PR-genes (PR-1-A1, chitinase, and thaumatin) among four different interactions including CS infiltrated with SnTox1 culture filtrates (CS/+SnTox1), CS infiltrated with control culture filtrates (CS/-SnTox1), CS ems237 infiltrated with SnTox1 (ems/+SnTox1), and CS ems237 infiltrated with control culture filtrates (ems/-SnTox1). The wheat 18S gene was used as an RNA quantity control. B. Expression of three PR genes using qPCR. Comparisons were made among the four different interactions described above. The same RNA samples from RT-PCR were used in qPCR analysis. The relative expression level for each time point was normalized to the wheat 18S gene.
Figure 11. The SnTox1-Snn1 interaction induces programmed cell death.
Leaf samples of Chinese Spring (CS) and CS ems237 were collected at 1, 2, 4, 10, 24, 36, 48, 60 and 72 h post infiltration with SnTox1 yeast culture filtrates (+SnTox1) and control culture filtrates (-SnTox1). Genomic DNA was extracted from the collected leaf samples, run on a 2% agarose gel and stained with ethidium bromide. DNA fragmentation was detected only in leaves of CS infiltrated with SnTox1 culture filtrates, but not in leaves of CS infiltrated with control culture filtrates, CS ems237 infiltrated with SnTox1, or control culture filtrates.
Figure 12. The SnTox1-Snn1 interaction is light dependent.
The light dependency of the SnTox1-Snn1 interaction was tested in both infiltration (left panel) and inoculation (right panel) experiments. Chinese Spring plants (Snn1) were grown under a normal light-dark cycle (12 h photoperiod) or under complete dark conditions for 48 h after being infiltrated with SnTox1 culture filtrates or inoculated with conidia from Sn79+SnTox1A1. No necrosis or lesions developed in the plants under dark treatment in either experiment, however, once the dark treated plants were moved back to a normal light cycle, necrosis and lesions developed subsequently on the leaves after an additional 48 h.
Figure 13. Comparison of the fungal infection process with or without an SnTox1-Snn1 interaction.

The fungal infection process was examined and compared at 1, 2, 4 and 7 days post inoculation (DPI) between the susceptible interaction (A, C, E, G) and resistant interaction (B, D, F and H) using GFP-tagged fungal strains and confocal fluorescent microscopy. A, B. Infection at 1 DPI. The fungus was able to form penetration structures (hyphopodium, white arrow) in both interactions. C, D, Infection at 2 DPI. The fungus breached the epidermal layer and started infecting the mesophyll layer in the susceptible interaction shown by the autofluorescence (C). No autofluorescence was observed in the resistant interaction (D). E, F, Infection at 4 DPI. The SnTox1+ strain continued infecting mesophyll cells shown by the larger area having autofluorescence in the susceptible interaction (E), but the SnTox1- strain remained on the leaf surface and continued attempting to penetrate in the resistant interaction (F). G, H, infection at 7DPI. The disease lesions have fully developed in the susceptible interaction (G), but no disease symptom was observed in the resistant interaction (H). Scale bar =20 µm.
Figure 14. Light is required for penetration of the epidermis.
The fungal infection process was examined and compared between plants grown under a normal light cycle and those under complete darkness post inoculation. At 48 HPI, the pathogen successfully penetrated through the epidermal cell layer and started infection of mesophyll cells in the CS leaves that were grown under a normal light cycle (A), but not in the leaves that were grown in the dark (B). Scale bar=20 µm.
REFERENCES


Chapter 6:

General Conclusions and Outlook
In the agro-ecosystem rapid changes in host phenotypes, driven by human intervention, impose strong selective pressure on pathogens to respond. This creates an environment in which host-parasite interactions are likely to evolve faster than in natural ecosystems (Stukenbrock and McDonald, 2008). Phaeosphaeria nodorum and the associated species complex described in this thesis provide an excellent system with which to further examine the traits that drive closely related species toward different life strategies (i.e. pathogens from endophytes or saprobes) (Chapter 2). We propose that recent horizontal acquisition of NEs outside of the Center of Origin of P. nodorum, enabled this fungus to become a globally distributed pathogen of wheat (Chapter 3).

Global population genetics of P. nodorum and related species, reveal a highly diverse pathogen complex distributed across all continents where wheat is grown. In this study, Phaeosphaeria nodorum was by far the most commonly sampled pathogen on wheat, however, a significant proportion of isolates were P. avenaria tritici 1. The remaining species collections were relatively small in comparison, however, Pat3, P1 and P2 were all collected in wheat fields alongside P. nodorum. In total we described nine distinct species, four collected exclusively on wheat or triticale, one on oats and three from wild grasses. P. nodorum was collected from wheat, triticale and wild grasses. We consider it likely that further sampling of grasses in Eurasia would reveal additional unknown sister species to further enrich the species complex described in this thesis.

BLAST analysis suggested that P1 and P2 may be closely related to grass-associated endophytic fungi. P1 and P2, however, were collected from wheat heads in a field that also contained P. nodorum and Pat1. It remains unknown whether these species are opportunistic saprotrophs, endophytes or pathogens. Controlled pathogenicity tests with all nine species, on both domesticated and non-domesticated hosts, are needed in order to determine the trophic strategy of each species in the complex. Knowing the trophic strategies of sister species is important as it can dramatically shape the formulation of future experiments with this species complex. Sister species that are saprophytic or endophytic would prompt questions such as: Was it solely the acquisition of NEs that drove P. nodorum’s emergence as a pathogen? Is there potential for the creation of new pathogens through HGT between P. nodorum and the remaining
Phaeosphaeria sp.? Alternatively, if the Phaeosphaeria sister species are determined to be pathogenic, questions pertaining to the genetic determinants of pathogenicity, in the absence of NEs, would be extremely interesting to explore.

To investigate the underlying genetic differences between sister species we propose a genomic sequencing approach with multiple unique strains from each species. Costs of this project would be comparatively low, in comparison to the amount of information acquired, and this method has been successfully applied to assess the evolution of both inter- and intraspecific fungal populations (i.e. with the species complex associated with Zymoseptoria tritici and differentiated populations of Neurospora crassa (Ellison et al., 2011; Stukenbrock et al., 2011; Torriani et al., 2011). Furthermore, genome sequences could be used to resolve the ancestral relationships between species presented in Chapter 2. In particular, clarification of the ancestral relationships between P. nodorum and Pat1 would confirm or refute the current tree that shows P. nodorum and Pat1 do not share an immediate common ancestor. Higher confidence in this tree topology would strengthen the hypothesis of horizontal acquisition of NE by Pat1 from P. nodorum. Newer coalescent methods, recently published by Bryant et al. (2012) enable the estimation of coalescent species trees using unlinked bi-allelic markers, such as SNPs in genome assemblies. This model can also estimate the effective population sizes of each species in the analysis, including ancestors (Bryant et al., 2012). The use of a large number of unlinked SNPs could also provide enough resolution to define the relationships among the Phaeosphaeria avenaria tritici complex (Chapter 2), which were too closely related at conserved sequence loci to accurately estimate their ancestral relationships with each other.

Hybridization between sexual and asexual grass endophytes is commonly reported in the literature (Moon et al., 2004). Similarly, this thesis describes the hybridization between the heterothallic, sexually recombining species, P. nodorum, and the single mating-type species Pat1. The finding of P. nodorum mating-type alleles in Pat1 hybrids suggests that these strains may have arisen through interspecific sexual crosses. Hybridization between species can result in homoploids (individuals that contain the same number of chromosomes as their parent species) or allopolyploids (individuals that contain the number of chromosomes equal to the sum from their parent species) (Giraud et al., 2008). Allopolyploids often lose chromosomes from either
parent but have been shown to exhibit wider ranges of phenotypes than either of their parent species (Giraud et al., 2008). Allopolyploids are believed to result in immediate genetic isolation due to an inability to properly pair chromosomes in meiosis with a parent. Homoploids however, may undergo subsequent backcrosses with either parent, resulting in competition between hybrid individuals and subsequently backcrossed offspring (Giraud et al., 2008). Pulse field gel electrophoresis of each of the Pat1 x P. nodorum hybrids and several P. nodorum and Pat1 isolates from the same field collections would determine if the hybrids shown in this study are homo- or allopolyploids. Homoploid hybrids would be more likely to backcross with their parent species. This could be used as further evidence for the horizontal transfer of NE from P. nodorum to Pat1 via hybridization. An intriguing final experiment would be to attempt experimental sexual crosses between P. nodorum and Pat1. P. nodorum, however, is not easily crossed under laboratory conditions and formation of additional hybrids strains in outdoor conditions would pose serious conflicts of interest with bio-safety concerns.

Our work provides further evidence that the CofO of P. nodorum coincides with that of its host in the ancient Fertile Crescent. These results, however, were not irrefutably conclusive and additional populations from Iran or surrounding regions, such as Jordan, Eastern Turkey, Iraq or Afghanistan, would be needed to more precisely locate the CofO of the species. Similarly, the origin and reproductive mode of Pat1 remains an open question. Larger collections from Pat1 could determine if the species has a center of genetic diversity that coincides with P. nodorum, or if the species originates outside of the Fertile Crescent. Determining the reproductive mode of Pat1 is slightly more difficult. Assuming Pat1 to be a homothallic fungus that undergoes high rates of inbreeding, it might be possible to detect recent recombination events between unlinked loci if the effective population sizes of Pat1 are large enough. However if the populations are small, or there is a high proportion of clonal reproduction or selfing, it would be difficult to distinguish between ancient recombination in a now asexual species, or a sexual population that is highly inbred. Definitive proof of homothallic recombination would require detection of sexually produced ascospores in single isolate cultures. This was reported by Shaw in her original description of the fungus in the 1950’s (1957).

Beyond the evolutionary questions associated with species complexes and pathogen
emergence, this work also provides insight into the practical implementation of control strategies for *P. nodorum*. As discussed in Chapters four and five, prior to the discovery of NEs, the interaction between necrotrophic pathogens and their hosts was believed to be based on the production of non-specific toxins and lytic enzymes (Chapter 4 and 5). This hypothesis however, provided no clear explanation for the host-specificity of necrotrophic pathogens. It is now clear that virulence in *P. nodorum* is driven by the production and secretion of NEs that interact specifically with host-sensitivity genes. After infection or exposure to purified toxin the physiological host response is consistent with R-gene mediated induction of programmed cell death. This hypothesis was recently verified with the cloning of the *Tsn1* sensitivity locus, which revealed an NBS-LRR-like resistance gene (Faris et al., 2010). To date there are three cloned sensitivity genes, one each from arabidopsis (Lorang et al., 2007), wheat (Faris et al., 2010) and sorghum (Nagy and Bennetzen, 2008). All three sensitivity genes in each host were found to be NBS-LRR-like genes.

The sensitivity loci are dominant sensitive to NEs, providing breeders with an easy, fast method to to screen new cultivars for toxin sensitivity (Oliver and Solomon, 2010). In Australia, scientists have already screened 60 commercial cultivars for sensitivity to SnTox3 using simple infiltration techniques. This work showed that the vast majority of lines, 87%, were sensitive to SnTox3. (Waters et al., 2011). They propose that simple greenhouse screening with the purified toxins would enable breeders to quickly eliminate sensitive lines from further breeding, allowing breeders to focus their efforts on other, more difficult, crop traits (Waters et al., 2011).

Significant differences between toxin frequency in different regions suggests that presence of the sensitivity genes in local cultivars has a strong effect on the prevalence of NE in the corresponding fungal population (Chapter 3). This suggests that breeding programs that eliminate sensitive germplasm from their commercial cultivars would be successful in driving down the frequency of NE in the local populations of *P. nodorum*. However, high levels of gene flow between regions (Stukenbrock et al., 2006) highlights the risk that fungal migrants carrying NEs from a neighboring region could quickly re-invade populations with low NE prevalence. Maintenance of cultivars that lack sensitivity genes is critical for the continued success of breeding for disease “resistance” to *P. nodorum* and other NE producing fungi.
REFERENCES


203

ACKNOWLEDGEMENTS

Many heartfelt thanks go to ...

Prof. Bruce McDonald for his mentorship and for the opportunity to spend five wonderful years here in Switzerland.

Prof. Beat Keller for agreeing to be my co-examiner. Thank you for your time.

To everybody in the Plant Pathology group for just being awesome! I couldn’t have picked a better group to be a part of:

Special thanks:
To Patrick for tolerating my random question invasions of his office.
To Daniel, Ethan and Pascal for being wonderful lunch and work companions.
To Beat, for your inspiring passion for fungi (truffles) and other yummy things. Poke!
To Stefano!!!! For hugs and coffees and long discussions (but mostly for the hugs).
To Tryggvi for being pretty smart even though you’re from Iceland.
To Fabi for being an emblem of strength.
To Paola for wonderful laughter and girly giggles.
To Mark and Christine for help in the lab with my project, Danke!

To the running crew! Maxi, Johannes, Ste and Ste! Here’s to Berlin, Oslo, Istanbul and Vienna!

To my big brothers Roberto (the best guitar player that I know), Johannes and Stefano (second best guitar player that I know) for always keeping an eye on me.

To my dearest wonderful Maxi, for translation of my Summary, for being a wonderful roomie and all-around awesome person. Thanks for picking me up when I didn’t have the strength.

To the Geezers and the whole touch rugby crew for introducing me to the best sport ever.

For the laughs, coffees, hugs, lake bbqs and other wonderful things, thanks to all of my friends!

To Joce and Beck my extra sisters.

To Mom, Dad, Mady, Morgie, Justin, Grandma Fran, Fred, Joce and Peggy for coming across and ocean to share this day with me.

To my Grandparents for teaching my parents the value of education. Talk about a trickle down effect! :)

And last but not least, to Skype for many, many hours of shared moments with my loved ones far away.
Published Works


Unpublished Works

McDonald, M.C., Ravazi, M., Brunner, P.C., McDonald, B.A. Phylogenetic and population genetic analyses of Phaeosphaeria nodorum and its close relatives indicate cryptic species and an origin in the Fertile Crescent (In Review). Fungal Genetics and Biology

CURRICULUM VITAE

Megan Camilla McDonald

Born 17th August, 1985. Citizen of The United States of America

University Education

Bachelor of Science  
University of Arizona  
Molecular and Cellular Biology G.P.A – 3.87  
Magna Cum Laude  
August 2003-May 2007

PhD Doctor of Science  
ETH Zurich  
Plant Pathology Group: Institute of Integrative Biology  
Population Genetics of Phaeosphaeria nodorum  
September 2008- July 2012
### APPENDICIES

#### Appendix A.1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence 5’-3’</th>
<th>Anneal Temp</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>ITS4</td>
<td>TCC TCC GCT TAT TGA TAT GC</td>
<td>58</td>
<td>White et. al. 1990</td>
</tr>
<tr>
<td></td>
<td>ITS5</td>
<td>GGA AGT AAA AGT CGT AAC AAG G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-tubulin</td>
<td>Btub_542F</td>
<td>GTG CAT CTG TAG CTG CCT GTT G</td>
<td>57</td>
<td>This paper</td>
</tr>
<tr>
<td></td>
<td>Btub_1259R</td>
<td>CAG TTG TTA CCA GCA CCA GAC TG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-tubulin</td>
<td>Btub_1273F</td>
<td>CAG TCT GGT GCT GGT AAC AAC TG</td>
<td>55</td>
<td>This paper</td>
</tr>
<tr>
<td></td>
<td>Btub_1986R</td>
<td>CTC AAC GAA GTA GGA CGA GTT C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-xylosidase</td>
<td>Bxylo_9Fcod</td>
<td>CAA AGA ACC CAT TGT CAC ACA C</td>
<td>56</td>
<td>This Paper</td>
</tr>
<tr>
<td></td>
<td>Bxylo_970Rco</td>
<td>GCT GTT CTT CAG CCA ACT TGA T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mat1-1</td>
<td>MT1.1</td>
<td>CTT CAC GAC CGG CCA GAT AGT</td>
<td></td>
<td>Bennett et. al. 2003</td>
</tr>
<tr>
<td></td>
<td>MT1.2</td>
<td>CAG AGG CTT GTC GGG TTC AT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mat1-2</td>
<td>MT2.1</td>
<td>ACC CCG CCC CAT GAA CAA GTG</td>
<td></td>
<td>Bennett et. al. 2003</td>
</tr>
<tr>
<td></td>
<td>MT2.2</td>
<td>CTA GAC CGG CCC GAT CAA GAC CAA AGA AG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix A.2

Supplementary Table A.2: Summary of average number of nucleotide differences and number of fixed differences between ea. species

### B-tubulin

<table>
<thead>
<tr>
<th>Pairwise</th>
<th>N Isolates</th>
<th>N Nucleotides</th>
<th>N Haplotypes</th>
<th>N Var Sites</th>
<th>Length InDel</th>
<th>N mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. nodorum</strong></td>
<td>405</td>
<td>490</td>
<td>70</td>
<td>124</td>
<td>13</td>
<td>165</td>
</tr>
<tr>
<td>* Pat1</td>
<td>6.29</td>
<td>22.05</td>
<td>19.9</td>
<td>20.58</td>
<td>18.98</td>
<td>16.43</td>
</tr>
<tr>
<td>* Pat3</td>
<td>18</td>
<td>12</td>
<td>17.17</td>
<td>20.83</td>
<td>21.1</td>
<td>21.5</td>
</tr>
<tr>
<td>* Pat4</td>
<td>11</td>
<td>14</td>
<td>6.14</td>
<td>14.3</td>
<td>14.6</td>
<td>20.44</td>
</tr>
<tr>
<td>* Pat5</td>
<td>13</td>
<td>18</td>
<td>9.18</td>
<td>2.67</td>
<td>13.8</td>
<td>20.66</td>
</tr>
<tr>
<td>* Pat6</td>
<td>9</td>
<td>17</td>
<td>7</td>
<td>8</td>
<td>8.3</td>
<td>21.3</td>
</tr>
<tr>
<td>* P. averaria</td>
<td>0</td>
<td>12</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>12.67</td>
</tr>
<tr>
<td>* Pat1</td>
<td>9</td>
<td>13</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>* Pat2</td>
<td>31</td>
<td>37</td>
<td>37</td>
<td>3</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>44</td>
<td>45</td>
<td>47</td>
<td>44</td>
<td>46</td>
<td>43</td>
</tr>
</tbody>
</table>

### B-xyllosidase

<table>
<thead>
<tr>
<th>Pairwise</th>
<th>N Isolates</th>
<th>N Nucleotides</th>
<th>N Haplotypes</th>
<th>N Var Sites</th>
<th>Length InDel</th>
<th>N Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. nodorum</strong></td>
<td>370</td>
<td>690</td>
<td>51</td>
<td>261</td>
<td>3</td>
<td>357</td>
</tr>
<tr>
<td>* Pat1</td>
<td>3.48</td>
<td>51.64</td>
<td>49.11</td>
<td>47.25</td>
<td>94.5</td>
<td>63.89</td>
</tr>
<tr>
<td>* Pat3</td>
<td>45</td>
<td>3.49</td>
<td>20.25</td>
<td>26.25</td>
<td>29.79</td>
<td>58.79</td>
</tr>
<tr>
<td>* Pat4</td>
<td>41</td>
<td>12</td>
<td>6.33</td>
<td>16.44</td>
<td>23.5</td>
<td>52.03</td>
</tr>
<tr>
<td>* Pat5</td>
<td>43</td>
<td>23</td>
<td>10</td>
<td>2</td>
<td>34.5</td>
<td>51.08</td>
</tr>
<tr>
<td>* Pat6</td>
<td>45</td>
<td>22</td>
<td>11</td>
<td>16</td>
<td>9.53</td>
<td>61</td>
</tr>
<tr>
<td>* P. averaria</td>
<td>50</td>
<td>55</td>
<td>47</td>
<td>49</td>
<td>51</td>
<td>2</td>
</tr>
<tr>
<td>* Pat1</td>
<td>44</td>
<td>16</td>
<td>4</td>
<td>13</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>* Pat2</td>
<td>104</td>
<td>111</td>
<td>112</td>
<td>115</td>
<td>118</td>
<td>78</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>52</td>
<td>72</td>
<td>71</td>
<td>69</td>
<td>70</td>
<td>78</td>
</tr>
</tbody>
</table>

### ITS

<table>
<thead>
<tr>
<th>Pairwise</th>
<th>N Isolates</th>
<th>N Nucleotides</th>
<th>N Haplotypes</th>
<th>N Var Sites</th>
<th>Length InDel</th>
<th>N mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. nodorum</strong></td>
<td>387</td>
<td>503</td>
<td>35</td>
<td>112</td>
<td>104</td>
<td>145</td>
</tr>
<tr>
<td>* Pat1</td>
<td>3.02</td>
<td>9.71</td>
<td>12.09</td>
<td>29.09</td>
<td>9.96</td>
<td>18.64</td>
</tr>
<tr>
<td>* Pat3</td>
<td>6</td>
<td>2</td>
<td>7.81</td>
<td>26.07</td>
<td>7.64</td>
<td>18.29</td>
</tr>
<tr>
<td>* Pat4</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>24.83</td>
<td>8.5</td>
<td>20</td>
</tr>
<tr>
<td>* Pat5</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>16.5</td>
</tr>
<tr>
<td>* Pat6</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>* P. averaria</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>82</td>
<td>53</td>
<td>50</td>
<td>29</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>55</td>
<td>54</td>
<td>51</td>
<td>41</td>
<td>57</td>
<td>62</td>
</tr>
</tbody>
</table>

### Multi-locus (hybrids removed)

<table>
<thead>
<tr>
<th>Pairwise</th>
<th>N Isolates</th>
<th>N Nucleotides</th>
<th>N Haplotypes</th>
<th>N Var Sites</th>
<th>Length InDel</th>
<th>N mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. nodorum</strong></td>
<td>356</td>
<td>1653</td>
<td>149</td>
<td>465</td>
<td>120</td>
<td>650</td>
</tr>
<tr>
<td>* Pat1</td>
<td>9.25</td>
<td>83.04</td>
<td>80.1</td>
<td>86.751</td>
<td>82.47</td>
<td>103.36</td>
</tr>
<tr>
<td>* Pat3</td>
<td>66</td>
<td>4.36</td>
<td>44.58</td>
<td>62.76</td>
<td>7.78</td>
<td>59.99</td>
</tr>
<tr>
<td>* Pat4</td>
<td>56</td>
<td>29</td>
<td>15.5</td>
<td>45.1</td>
<td>46.5</td>
<td>94.17</td>
</tr>
<tr>
<td>* Pat5</td>
<td>62</td>
<td>44</td>
<td>20</td>
<td>21.4</td>
<td>55.77</td>
<td>97.6</td>
</tr>
<tr>
<td>* Pat6</td>
<td>63</td>
<td>47</td>
<td>24</td>
<td>31</td>
<td>12.33</td>
<td>101</td>
</tr>
<tr>
<td>* P. averaria</td>
<td>88</td>
<td>92</td>
<td>90</td>
<td>78</td>
<td>87</td>
<td>6</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>63</td>
<td>37</td>
<td>12</td>
<td>27</td>
<td>26</td>
<td>85</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>195</td>
<td>207</td>
<td>205</td>
<td>104</td>
<td>213</td>
<td>213</td>
</tr>
<tr>
<td>* P. *P. *P.</td>
<td>191</td>
<td>172</td>
<td>169</td>
<td>154</td>
<td>178</td>
<td>189</td>
</tr>
</tbody>
</table>
Appendix A.3

Supplementary Table A.3: Summary of Mating Type polymorphism between and within each species

| Maximum pairwise number of mutations between species: Non-synonymous |
| Maximum pairwise number of mutations between species: Synonymous |
| Total number of mutations within Species |

### MAT1-1 Polymorphism

#### Between-Species Pairwise

<table>
<thead>
<tr>
<th></th>
<th>P. nodorum</th>
<th>Pat1</th>
<th>Pat3</th>
<th>Pat5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. nodorum</td>
<td>7</td>
<td>12</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Pat1</td>
<td>22</td>
<td>0</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Pat3</td>
<td>12</td>
<td>16</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Pat5</td>
<td>16</td>
<td>19</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

#### Within Species

<table>
<thead>
<tr>
<th></th>
<th>P. nodorum</th>
<th>Pat1</th>
<th>Pat3</th>
<th>Pat5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Haplotypes</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Synonymous</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Nsynon</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Intron</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

### MAT1-2 Polymorphism

#### Between Species Pairwise

<table>
<thead>
<tr>
<th></th>
<th>P. nodorum</th>
<th>Pat3</th>
<th>Pat4</th>
<th>Pat5</th>
<th>Pat6</th>
<th>P. avenaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. nodorum</td>
<td>6</td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Pat3</td>
<td>22</td>
<td>0</td>
<td>5</td>
<td>11</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>Pat4</td>
<td>23</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>38</td>
<td>9</td>
</tr>
<tr>
<td>Pat5</td>
<td>22</td>
<td>7</td>
<td>12</td>
<td>0</td>
<td>43</td>
<td>15</td>
</tr>
<tr>
<td>Pat6</td>
<td>27</td>
<td>29</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>P. avenaria</td>
<td>22</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>31</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Within Species

<table>
<thead>
<tr>
<th></th>
<th>P. nodorum</th>
<th>Pat3</th>
<th>Pat4</th>
<th>Pat5</th>
<th>Pat6</th>
<th>P. avenaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Haplotypes</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Synonymous</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nsynon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intron</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix B.1

B.1 Maximum likelihood tree with β-tubulin haplotypes

Maximum likelihood tree constructed using β-tubulin haplotypes from 355 Phaeosphaeria sp. Names shown at the tip of each triangle are the original names of one member of the haplotype. Bootstrap values are shown on the branches. Support values are not shown for tips where all isolates collapsed into a single haplotype. If included, estimates on these branches would be 100.
Appendix B.2

B.2 Maximum likelihood tree with β-xylosidase haplotypes

Maximum likelihood tree constructed using β-xylosidase haplotypes from 355 Phaeosphaeria sp. Names shown at the tip of each triangle are the original names of one member of the haplotype. Bootstrap values are shown on the branches. Support values are not shown for tips where all isolates collapsed into a single haplotype. If included, estimates on these branches would be 100.
Appendix B.3

B.3 Maximum likelihood tree with ITS haplotypes

Maximum likelihood tree constructed using ITS haplotypes from 355 *Phaeosphaeria* sp. Names shown at the tip of each triangle are the original names of one member of the haplotype. Bootstrap values are shown on the branches. Support values are not shown for tips where all isolates collapsed into a single haplotype. If included, estimates on these branches would be 100.