Local adaptation drives the diversification of effectors in the fungal wheat pathogen *Parastagonospora nodorum* in the United States

Jonathan K. Richards[1], Eva H. Stukenbrock[2,3], Jessica Carpenter[4], Zhaohui Liu[4], Christina Cowger[5], Justin D. Faris[6], Timothy L. Friesen[4,6]*

1 Department of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, Louisiana, United States of America, **2** Department of Environmental Genomics, Christian-Albrechts University of Kiel, Kiel, Germany, **3** Max Planck Institute for Evolutionary Biology, Plön, Germany, **4** Department of Plant Pathology, North Dakota State University, Fargo, North Dakota, United States of America, **5** Plant Science Research Unit, USDA-ARS, Raleigh, North Carolina, United States of America, **6** Cereal Crops Research Unit, Edward T. Schaefer Agricultural Research Center, USDA-ARS, Fargo, North Dakota, United States of America

* timothy.friesen@usda.gov

Abstract

Filamentous fungi rapidly evolve in response to environmental selection pressures in part due to their genomic plasticity. *Parastagonospora nodorum*, a fungal pathogen of wheat and causal agent of septoria nodorum blotch, responds to selection pressure exerted by its host, influencing the gain, loss, or functional diversification of virulence determinants, known as effector genes. Whole genome resequencing of 197 *P. nodorum* isolates collected from spring, durum, and winter wheat production regions of the United States enabled the examination of effector diversity and genomic regions under selection specific to geographically discrete populations. 1,026,859 SNPs/InDels were used to identify novel loci, as well as *SnToxA* and *SnTox3* as factors in disease. Genes displaying presence/absence variation, predicted effector genes, and genes localized on an accessory chromosome had significantly higher pN/pS ratios, indicating a higher rate of sequence evolution. Population structure analyses indicated two *P. nodorum* populations corresponding to the Upper Midwest (Population 1) and Southern/Eastern United States (Population 2). Prevalence of *SnToxA* varied greatly between the two populations which correlated with presence of the host sensitivity gene *Tsn1* in the most prevalent cultivars in the corresponding regions. Additionally, 12 and 5 candidate effector genes were observed to be under diversifying selection among isolates from Population 1 and 2, respectively, but under purifying selection or neutrally evolving in the opposite population. Selective sweep analysis revealed 10 and 19 regions that had recently undergone positive selection in Population 1 and 2, respectively, involving 92 genes in total. When comparing genes with and without presence/absence variation, those genes exhibiting this variation were significantly closer to transposable elements. Taken together, these results indicate that *P. nodorum* is rapidly adapting to distinct selection pressures unique to spring and winter wheat production regions by rapid adaptive evolution and

various routes of genomic diversification, potentially facilitated through transposable element activity.

Author summary

Parastagonospora nodorum is an economically important pathogen of wheat, employing proteinaceous effectors to cause disease. Recognition of effectors by host susceptibility genes often leads to the elicitation of programmed cell death. However, little is known on the correlation between effector diversity and the spatial distribution of host resistance/ susceptibility or the genomic mechanisms of diversification. This research presents the genome resequencing of 197 *P. nodorum* isolates collected from spring, winter, and durum wheat production regions of the United States, enabling the investigation of genome dynamics and evolution. Results illustrate local adaptation to host resistance or susceptibility, as evidenced by population-specific evolution of predicted effector genes and positively selected selective sweeps. Predicted effector genes, genes exhibiting presence/absence variation, and genes residing on an accessory chromosome, were found to be diversifying more rapidly. Additionally, transposable elements were predicted to play a role in the maintenance or elimination of genes. A GWAS approach identified the previously reported *SnToxA* and *SnTox3* as well as novel virulence candidates as major elicitors of disease. These results highlight the flexibility of the *P. nodorum* genome in response to population-specific selection pressures and illustrate the utility of whole genome resequencing for the identification of putative virulence mechanisms.

Introduction

Plant pathogenic microorganisms, which are continually in evolutionary conflict with their respective hosts, have developed mechanisms by which they adapt and proliferate. The dynamic nature of fungal genomes, resulting in small and large-scale changes, provides diversification while maintaining essential functions. The prevalence of mobile elements often drives this flexibility, resulting in the creation, abolition, or translocation of genes [1,2]. Additionally, this activity of mobile elements significantly contributes to the diversification of asexually reproducing fungal pathogens through the rearrangement of chromosomes into lineage specific segments [3,4]. This phenomenon has recently been described as a compartmentalized 'two-speed' genome, consisting of regions of high gene density and equilibrated GC content, as well as a gene-sparse compartment [5,6,7,8]. The regions of low gene density and high repetitive content appear to be hotbeds of rapid evolution, harboring genes encoding virulence determinants known as effectors that manipulate host cellular processes to facilitate infection [7]. However, not all plant pathogenic fungi possess rigidly defined regions of low or high GC content; rather, they may have rapidly evolving genes and transposable elements dispersed evenly throughout the genome [9,10].

Fungal effector molecules act to modulate the host defense response in various ways, typically dependent on the lifestyle of the pathogen. Pathogen associated molecular patterns (PAMPs), such as the major fungal cell wall constituent chitin, may be recognized by the host and trigger an early defense response known as PAMP triggered immunity (PTI) [11,12,13]. Effectors can successfully subvert this response to permit infection. Ecp6, a fungal effector from *Cladosporium fulvum*, binds chitin and prevents the initiation of PTI [14]. In the context

of a biotrophic pathogen, host plants subsequently evolved the means to recognize virulence effectors, triggering a hypersensitive response and sequestration of the pathogen. Recognition is often mediated by host resistance genes in the nucleotide binding leucine-rich repeat (NB-LRR) gene family [15,16]. Within many biotrophic pathosystems, this interaction between a dominant host resistance gene and a biotrophic effector follows the gene-for-gene model [17]. However, necrotrophic pathogens use effectors to exploit the same host defense cellular machinery. Occurring in an inverse gene-for-gene manner, effectors are recognized by dominant host susceptibility genes, resulting in necrotrophic effector triggered susceptibility [18]. The lack of homology or conserved domains between effector proteins hinders efforts at novel effector discovery. This lack of similarity can be attributed to the rapid evolution in response to local selection pressure exerted by host resistance or susceptibility. However, the problem can be remedied via thorough genomic and genetic analyses.

Rapid evolution and positive selection of pathogen effector genes, as well as the host resistance/susceptibility genes, drive antagonistic co-evolutionary processes in plant-pathogen systems. The detection of positive selection permits the identification of genes potentially involved in local adaptation to host genotypes or environmental pressures [19]. Examination and comparison of nonsynonymous and synonymous substitution rates within genes has been used to detect evidence of purifying, neutral, or positive selection of putative pathogen virulence genes [20,21]. More recently, population whole-genome sequencing data has been leveraged to identify and compare selective sweep regions between pathogen populations, as well as closely related species, resulting from strong selection of a particular allele. The signature of such strong selection in population data is a genomic region devoid of genetic variation, and thus termed a selective sweep region. Badouin et al. [22] found a greater prevalence of selective sweeps in *Microbotryum lychnidis-dioicae* compared to its sister species *Microbotryum silenes-dioicae*. Additionally, candidate genes potentially involved in pathogenicity and host adaptation, as well as genes upregulated *in planta*, were identified within selective sweep regions. Analysis of a global collection of the barley pathogen *Rhynchosporium commune* identified three major genetic clusters that exhibited unique and generally non-overlapping selective sweep regions hypothesized to stem from ecological variations [23]. Interestingly, genes potentially implicated in response to abiotic stress and not host adaptation were found to be enriched within selective sweep regions. Similarly, evidence for selective sweeps was detected in a worldwide collection of *Zymoseptoria tritici*, with the majority of the identified regions being unique to one of the four postulated populations [24]. However, in this case, genes typically associated with pathogenicity or virulence, such as secreted proteins or cell wall degrading enzymes, were found to underlie selective sweeps. These studies revealed how selection pressures differing between local populations or species lead to the positive selection of genomic regions, providing evidence for local adaptation. The identification of sweep regions can lead to the identification of genes co-localizing with the sweeps, aiding the formation of hypotheses on niche adaptation.

Globally, wheat ranks as one of the most widely grown crops and supplies humans with an important source of calories. In 2017, approximately 218.5 million hectares were harvested globally [25]. Common wheat (*Triticum aestivum*), an allohexaploid (AABBDD), is classified by several characteristics including growth habit (spring or winter), bran color (red or white), and hardness of endosperm (soft or hard). Hard red spring wheat and hard red winter wheat are utilized primarily in the bread making industry, whereas flour from soft red winter wheat is often used for pastries and crackers [26]. Durum wheat (*Triticum turgidum* var. *durum*), an allotetraploid (AABB), is primarily used in the pasta industry [26]. Hard red spring wheat and durum wheat are typically grown in the Upper Midwest, hard red winter wheat is typically grown in the Great Plains, and soft red winter wheat cultivation predominates in the eastern

United States [27]. As all classes of wheat are threatened by disease epidemics on an annual basis, investigation of the evolutionary mechanisms of adaptation in important wheat pathogens is of upmost importance to protect the production of this important global commodity.

Parastagonospora nodorum, a haploid, necrotrophic fungal pathogen of both common wheat and durum wheat, causes significant yield losses and is an annual threat to global wheat production [28]. This pathogen has also emerged as a model organism for the study of plant-necrotrophic specialist interactions. Genomic resources, including the development of near complete reference genomes of diverse isolates bolstered by deep RNA sequencing, have greatly aided the investigation of pathogen virulence. Initially, the *P. nodorum* Australian isolate SN15 was sequenced with Sanger sequencing and subsequently improved via re-sequencing with short-read Illumina technology, RNA-seq data, and protein datasets [29,30,31]. Recently, we developed reference quality genome assemblies of *P. nodorum* isolates LDN03-Sn4 (hereafter referred to as Sn4), Sn2000, and Sn79-1087 using long-read sequencing technology [32]. These telomere to telomere sequences of nearly every *P. nodorum* chromosome, in addition to the annotation of 13,379 genes using RNA-seq, greatly improved the framework for effector discovery. A total of nine effector-host susceptibility factor interactions have been previously identified [18,33,34,35,36,37,38,39,40,41,42,43] with three of the pathogen effectors (SnToxA, SnTox1, and SnTox3) having been cloned and validated [35,36,42]. Novel interactions with susceptibility genes in winter wheat germplasm have also been identified in response to effectors produced by *P. nodorum* isolates collected in the southeastern United States, which differ from sensitivities and effectors identified in the hard red spring wheat production region of the Upper Midwest [44].

The functionally validated *P. nodorum* effector molecules display typical properties of pathogen effectors, including being small, secreted, cysteine-rich proteins embedded in or adjacent to repetitive regions of the genome. Additionally, these genes exhibit presence/absence variation (PAV), as they are completely missing in avirulent isolates [35,36,42]. This gain or elimination of entire genes, as well as intragenic polymorphisms, is rampant within *P. nodorum* populations [30], although little correlation between this diversity to the spatial distribution of host susceptibility has been made. Despite the evidence of frequent sexual reproduction in the field [45,46], attempts to develop bi-parental sexual populations of *P. nodorum* have been unsuccessful, and the validation of candidate effectors has been accomplished through computational, comparative, and reverse-genetics approaches [35,36,42].

Use of genome wide association studies (GWAS) overcomes the inability to develop bi-parental fungal populations in *P. nodorum* and allows candidate effectors to be mapped at high-resolution. Due to the increased accessibility of genome sequencing, GWAS can be relatively easily applied to fungi for effector identification [47]. Gao et al. [48] conducted GWAS in *P. nodorum* using restriction-associated DNA genotyping-by-sequencing (RAD-GBS) data from 191 isolates. Significant marker-trait associations (MTAs) were identified corresponding to effector genes *SnToxA* and *SnTox3*, illustrating the utility of this approach. Additionally, it was determined that linkage disequilibrium (LD) decayed rapidly in *P. nodorum*, highlighting the necessity of higher marker density for the successful identification of causal or linked single nucleotide polymorphisms (SNPs) [48]. GWAS was also used in another pathogen of wheat, *Zymoseptoria tritici*. Sequencing of 103 isolates and subsequent identification of 584,171 single nucleotide polymorphisms (SNPs) enabled the successful identification of *AvrStb6*, the gene conferring avirulence on wheat lines carrying a functional *Stb6* resistance gene [49,50].

As plant pathogens rapidly respond to the selection pressure placed upon them by regionally deployed host resistance/susceptibility genes [51], we hypothesized that this would result in local adaption of effector repertoires which would be visible as an accumulation of population-specific non-synonymous or loss-of-function mutations and genome-wide signatures of

selection. This research presents the whole-genome resequencing of 197 *P. nodorum* isolates collected from various wheat growing regions of the United States, enabling the investigation of population structure and regionally specific gene diversity, including effector diversification. Selective sweep and genic diversity analyses detected unique selection pressures on local *P. nodorum* populations, resulting in the regionally specific evolution of genes with a predicted virulence function, including putative effector genes that are hypothesized to have specific host susceptibility targets. Genes associated with presence/absence variation were found to be near repetitive elements, indicating a potential role of transposons in the maintenance or elimination of genic diversity. Additionally, a high level of genotypic diversity enabled robust genome-wide association analyses, which identified *SnToxA* and *SnTox3* as being significantly associated with disease severity on both spring and winter wheat lines. GWAS also detected a cell wall degrading enzyme and an entire gene cluster as novel candidates for virulence on winter wheat, laying the foundation for further dissection of this molecular pathosystem.

Results

Whole genome sequencing and variant identification

To compare patterns of genetic variation in populations of *P. nodorum*, we sequenced full genomes of 197 isolates (S1 Table) collected from spring, winter, and durum wheat production regions of the United States. Total sequence per isolate ranged from approximately 126 Mb to 3.01 Gb with a median value of 1.06 Gb. This corresponds to an approximate genome coverage (Sn4 genome size of 37.7 Mb) ranging from 3.34× to 79.8× with a median coverage of 28.08× (S2 Table).

Following filtering for genotype quality and read depth, 1,026,859 SNPs and insertions/deletions (InDels) were identified corresponding to a nucleotide diversity of 0.0062. Analysis of the functional effects of SNPs/InDels revealed a total of 226,803 synonymous and 160,159 non-synonymous polymorphisms within 13,238 genes, with 141 genes lacking any polymorphism (Table 1). Additionally, a total of 5,110 loss of function (LOF) mutations were detected, abolishing the function of 2,848 genes. These variants include 1,464 frameshift mutations, 2,815 premature stop codons, 457 losses of a stop codon, and 374 losses of a start codon (Table 1). The total SNP data set (including intergenic SNPs/InDels) was further filtered for a minor allele frequency of 5% and maximum missing data per marker of 30%, resulting in the identification of 322,613 SNPs/InDels to be used in association mapping analyses.

P. nodorum isolates missing genotype calls in greater than 50% of SNP/InDel sites and/or having average gene coverage of less than 95% were discarded from all coverage-based analyses, resulting in a final dataset of 175 *P. nodorum* isolates. Coverage analysis across the 13,379 annotated *P. nodorum* isolate Sn4 gene set identified 882 genes that were deleted in at least one isolate. Individual isolates were annotated as harboring between 5 and 343 gene losses which are distributed throughout the genome. Among genes exhibiting PAV, 70 genes encoded proteins with predicted secretion signals, including the previously characterized *SnToxA*, *SnTox1*, and *SnTox3* [35,36,42]. Overall, no significant differences were observed in the frequency of PAV between predicted effectors, secreted non-effectors, or non-secreted proteins, with PAV frequency rates of 7.8%, 5.2%, and 6.7%, respectively (Pairwise comparison of proportions, FDR adjusted p > 0.21 for all comparisons).

P. nodorum exhibits strong population structure in North America

We next set out to assess if the population structure of *P. nodorum* reflects adaptation to different host populations or geographical distribution. Using a genotypic dataset consisting of

Table 1. Functional variants identified in a population of 197 *P. nodorum* isolates.

Variant Category	Total	Population 1	Population 2
Synonymous SNP	226,803	194,313	151,965
Non-synonymous SNP	160,159	127,006	97,361
Frameshift	1,464	992	798
Gain of Stop Codon	2,815	2030	1406
Loss of Stop Codon	457	389	282
Loss of Start Codon	374	298	216

https://doi.org/10.1371/journal.pgen.1008223.t001

approximately one SNP per kb across the entire genome to mitigate potential marker pairs in linkage disequilibrium, STRUCTURE analysis revealed an optimal number of two sub-populations (S1 Fig). All isolates collected from North Dakota, Minnesota, and South Dakota formed a cluster, here termed Population 1 (S1 Fig). Isolates collected from Arkansas, Georgia, Maryland, New York, North Carolina, Ohio, Oregon, South Carolina, Tennessee, Texas, and Virginia formed a second cluster, here termed Population 2 (S1 Fig). Based on the 0.85 threshold of membership probability used to assign isolates to subpopulations, all 17 Oklahoma isolates were not clearly assigned to a cluster. To test the hypothesis that the Oklahoma isolates were admixed between the two major populations, a three-population test was conducted [52]. The f_3 statistic, z-value, and standard error were 16.50, 4.68, and 3.53, respectively. Both the f_3 statistic and z-score values being non-negative indicated a lack of evidence for admixture. This structure analysis separates isolates based on geographical location rather than wheat cultivar. However, the isolates collected from Oklahoma were not clearly placed in either population. Results from a three-population test indicated that isolates collected from Oklahoma were likely not admixed, but rather stem from a lineage of the major populations.

Principal components analysis (PCA) revealed similar results to those obtained by Bayesian clustering. The first two principal components, accounting for approximately 12.9% of the cumulative variation, separate Population 1 and Population 2. However, this analysis also indicated a separate cluster corresponding to eight isolates collected from Oregon (lower right of the plot), which likely represent a small subpopulation (Fig 1). However, due to a comparatively lower population size (n = 8), these isolates were removed from any subsequent comparisons between populations due to lack of proper representation (Fig 1).

High genetic variation in *P. nodorum* populations

We used our genome-wide inference of genetic variation to compare patterns of genetic diversity and polymorphism distribution in the *P. nodorum* population. Due to the larger number of SNB susceptibility targets that have been validated in spring wheat relative to winter wheat [18], we hypothesized that *P. nodorum* populations specific to spring wheat regions would contain a higher level of diversity and be under balancing selection to maintain a larger virulence gene repertoire. The F_{st} statistic was 0.181 when comparing Population 1 to Population 2, indicating moderate differentiation has occurred between isolates collected from each region. Overall, Population 1 had a higher level of nucleotide diversity (0.006) compared to Population 2 (0.004), as well as a greater proportion of segregating sites, as evidenced by Θ_W values of 0.005 and 0.004, respectively. Both groups had positive Tajima's D values caused by an excess of alleles of intermediate frequencies and absence of rare alleles, with Population 1 having the highest genome-wide Tajima's D value of 0.864 and Population 2 having a lower, yet still positive, Tajima's D value of 0.566 (Table 2; S2 Fig).
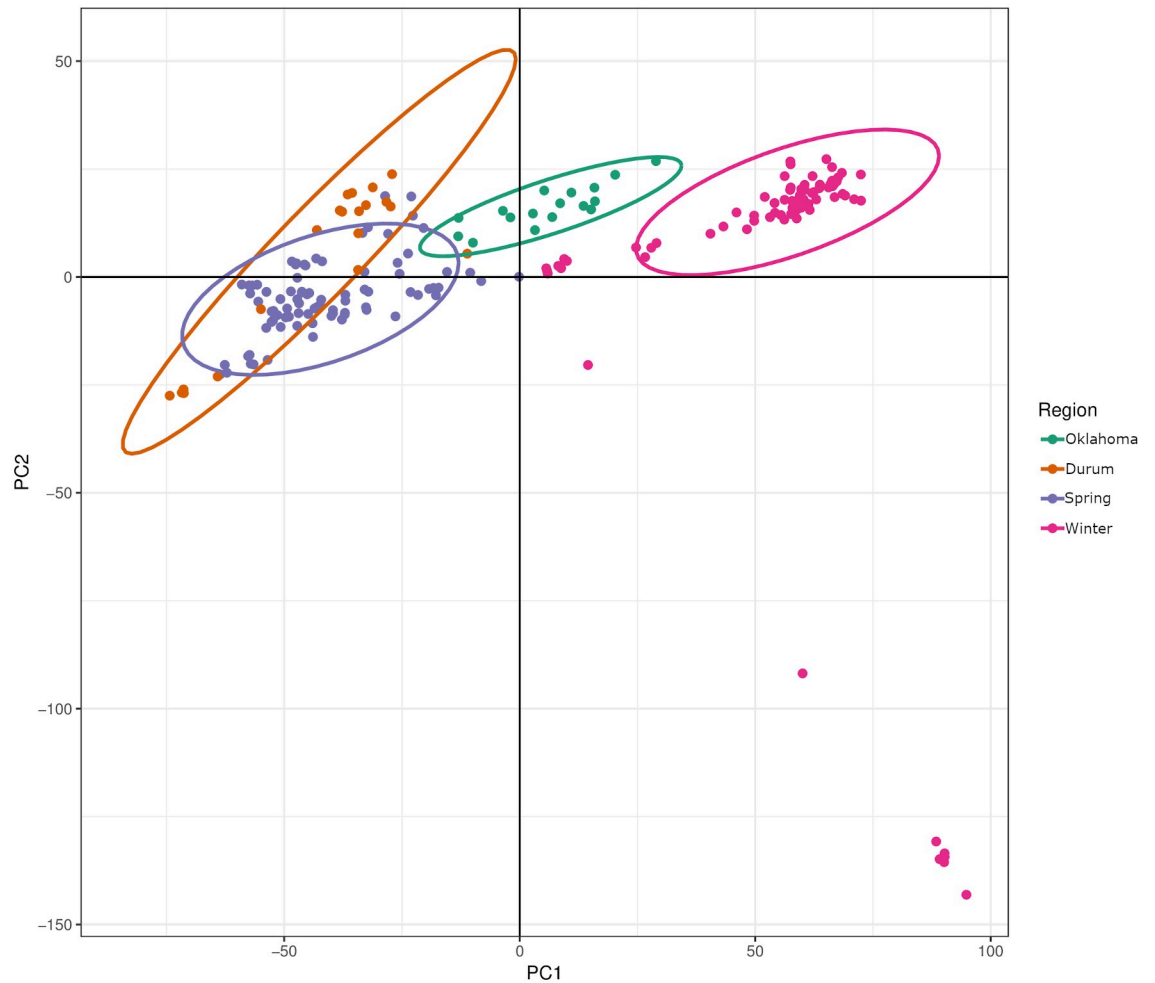
Fig 1. PCA using genotypic data of 50,000 randomly selected markers. PC1 is represented on the x-axis and PC2 is represented on the y-axis. Colored dots correspond to individual isolates and the color signifies the predominant wheat class (or state) of the region from which the isolate was collected. The isolates collected in Oregon are depicted in the lower right of the figure. The color legend is displayed on the right side of the figure.

Signatures of selective sweeps indicate recently acquired advantageous mutations

Examination of genomic regions having undergone selective sweeps sheds light on potentially beneficial genes being selected and facilitates the comparison of selective forces acting on

Table 2. Population genomics parameters of 172 *P. nodorum* isolates by population.

Population	π^1	$\Theta_W{}^2$	Tajima's D[3]	pN/pS[4]
Population 1	0.006	0.005	0.864	0.278
Population 2	0.004	0.004	0.566	0.294

[1]Nucleotide diversity calculated across entire genome

[2]Watterson's estimator calculated across entire genome

[3]Genome-wide Tajima's D value

[4]Average pN/pS

Table 3. Detection of selective sweeps in two *P. nodorum* populations.

Population	Sweep Regions	Median (kb)	Maximum (kb)	Genes	Genome Coverage (%)
Population 1	10	4.5	25.0	16	0.17
Population 2	19	9.0	92.0	76	1.02

different populations. We hypothesize that due to regional differences in wheat genotypes grown, as well as differing environmental cues, positive selection is acting on different regions of the *P. nodorum* genome within each population. Selective sweep analysis using SweeD revealed 42 and 46 regions having undergone selective sweeps in Population 1 and Population 2, respectively. To add further evidence of selective sweeps, predicted sweep regions were compared to population-specific Tajima's D values calculated in intervals across the genome. This analysis identified a total of 10 and 19 regions in Population 1 and Population 2, respectively (Table 3; Fig 2A), which were predicted as selective sweeps by SweeD and were located in a genomic region with a negative value of Tajima's D. Interestingly, no genes co-localizing with selective sweep regions were common to both populations. This indicates that different selection pressures, likely from regional wheat genotypes or the environment, are being exerted on isolates from each population for the maintenance of beneficial genes.

Sweep regions detected in Population 1 isolates had a median length of 4.5 kb with the largest region being 25.0 kb. Underlying these regions are 16 genes, of which, 7 are predicted hypothetical proteins with no known functional domains. A total of four genes encode predicted secreted proteins, including one predicted effector. Additionally, the presence/absence of five genes was polymorphic, which is significantly greater than expected (one gene expected; Fisher's Exact Test p = 0.003). Gene ontology enrichment analysis did not identify any significantly overrepresented gene functions, likely due to the low number of genes underlying predicted selective sweep in Population 1 (S1 File). A significant selective sweep region was detected on chromosome 8, flanking the *SnToxA* locus, but was not detected in Population 2 (Fig 2B). This reinforces the hypothesis that *SnToxA* was selected for in the Midwestern population due to the prevalence of *Tsn1* but was lost in Population 2 due to the lack of the *Tsn1* gene in the popular local winter wheat cultivars [54].

Analysis of isolates from Population 2 identified 19 selective sweep regions having a median length of 9.0 kb, maximum length of 92.0 kb, and a cumulative length of 385.4 kb, covering 1.02% of the genome (Table 3; Fig 2). The quantity and cumulative size of the predicted sweep regions are larger than those identified in Population 1, indicating that these may be more recent selection events. A total of 76 genes underlie the sweeps detected in Population 2, including 11 genes encoding predicted secreted proteins, one of which is a predicted effector protein. A total of 7 genes (9.2% of genes identified in regions of selective sweeps) exhibited PAV, which is not significantly greater than expected (five genes expected; Fisher's Exact test p > 0.05). Gene ontology enrichment analysis did not identify any overrepresented genes underlying selective sweeps specific to Population 2 (S1 File).

P. nodorum populations harbor different alleles and prevalence of *SnToxA*, *SnTox1*, and *SnTox3*

We then wanted to determine if host sensitivity conferred by a previously characterized effector had influenced *P. nodorum* population structure within the United States. Sensitivity to effectors SnToxA, SnTox1, and SnTox3 conferred by host genes *Tsn1*, *Snn1*, and *Snn3*, respectively, offers a distinct advantage to the pathogen through the strong induction of programmed cell death. The predominant presence of a host sensitivity gene, including novel genes or
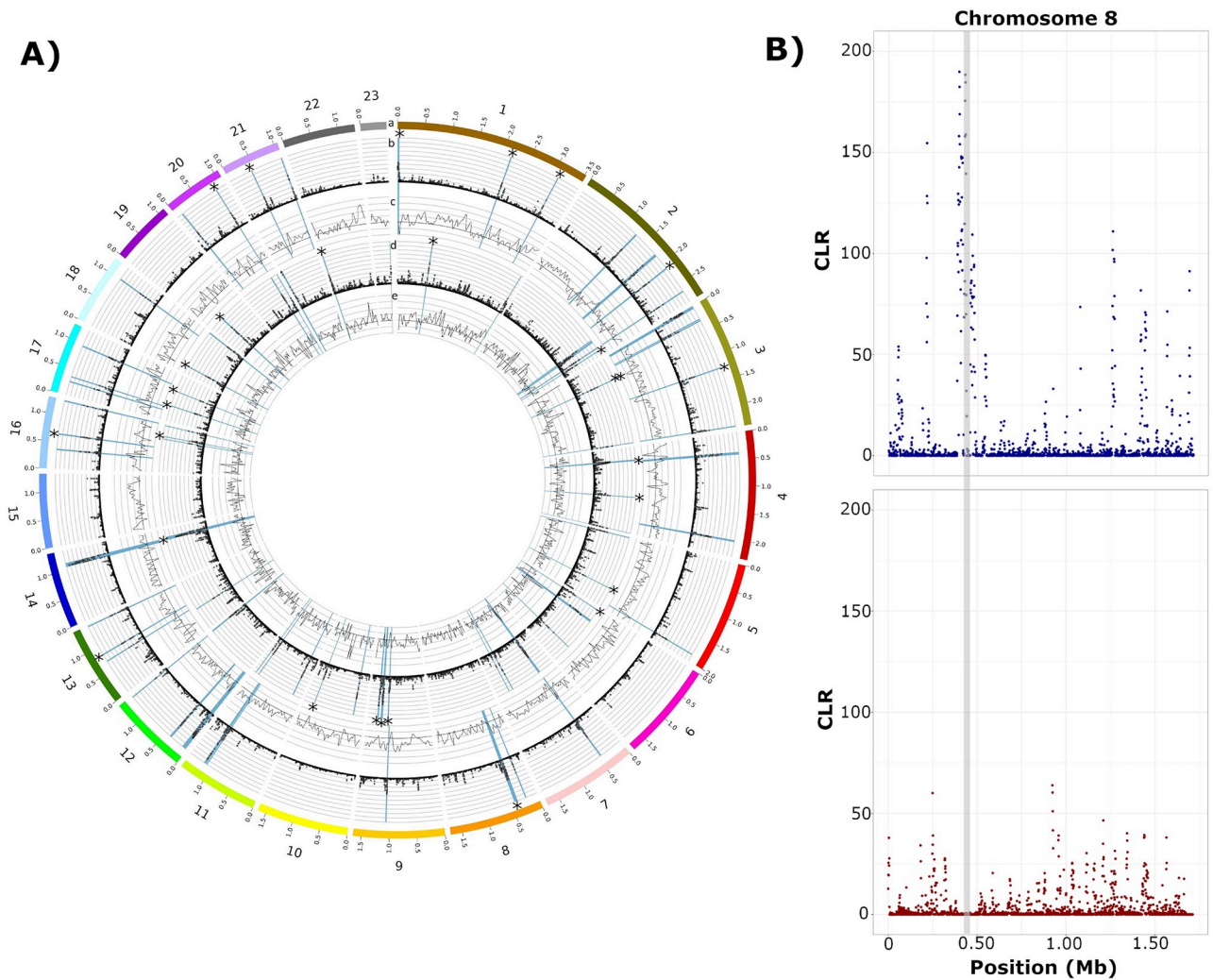
Fig 2. A) Plot illustrating the detected selective sweep loci from both *P. nodorum* subpopulations and sliding window Tajima's D analysis. '*' signifies a selective sweep region that was detected using SweeD and has a negative average value of Tajima's D. a) Chromosomes with sizes listed in Mb in 0.5 Mb increments b) Selective sweeps detected in Population 1. Genomic position is displayed on the x-axis. Likelihood values are displayed on the y-axis. Y-axis scale ranges from 0 to 400. Individual dots represent the likelihood value of a single 1 kb window. Highlighted blue regions are the 99th percentile of likelihood values. c) Tajima's D values of isolates in Population 1 in 50 kb windows in 25 kb steps. Genomic positions are displayed on the x-axis. Tajima's D values are displayed on the y-axis. The y-axis ranges from -2 to 4. The bold horizontal axis line is 0. d) Selective sweeps detected in Population 2. Genomic position is displayed on the x-axis. Likelihood values are displayed on the y-axis. Y-axis scale ranges from 0 to 400. Individual dots represent the likelihood value of a single 1 kb window. Highlighted blue regions are the 99th percentile of likelihood values. e) Tajima's D values of isolates in Population 2 in 50 kb windows in 25 kb steps. Genomic positions are displayed on the x-axis. Tajima's D values are displayed on the y-axis. The y-axis ranges from -2 to 4. The bold horizontal axis line is 0. B) Selective sweep analysis of *P. nodorum* chromosome 8 for Population 1 (top panel) and Population 2 (bottom panel). Position along the chromosome in megabases (Mb) is displayed on the x-axis. The composite likelihood ratio (CLR) is shown on the y-axis. The genomic region harboring *SnToxA* is highlighted in light grey.

https://doi.org/10.1371/journal.pgen.1008223.g002

interactions still undiscovered, may provide a strong selective force favoring the maintenance of a given effector, and therefore may influence population structure. Coverage analysis was used to determine the presence or absence of the three previously characterized *P. nodorum* effectors *SnToxA*, *SnTox1*, and *SnTox3* in isolates with sufficient coverage (n = 175). Overall, *SnToxA*, *SnTox1* and *SnTox3* were absent from 36.6%, 4.6%, and 41.1% of isolates, respectively (Table 4). Prevalence of effector genes was also examined by subpopulation. The clear majority of isolates retained a functional *SnTox1* gene, as 0% and 12.5% of isolates from Population 1

Table 4. Functional variants identified in *SnToxA*, *SnTox1*, and *SnTox3*.

Variant Category	*SnToxA*	*SnTox1*	*SnTox3*
Coding Sequence (bp)	534	351	693
Synonymous SNP	8	0	3
Non-synonymous SNP	3	9	5
Gain of Stop Codon	1	0	0
Loss of Start Codon	0	0	1
Deletion (%) Population 1	4.3	0	38.3
Deletion (%) Population 2	93.8	12.5	48.4
Overall Deletion (%)	36.6	4.6	41.1

and Population 2, respectively, harbored *SnTox1* gene deletions. *SnTox3* was observed to be absent from 38.3% and 48.4% of isolates from Population 1 and Population 2, respectively. A stark difference was observed when comparing the presence of *SnToxA* between the two populations. Within Population 1, containing isolates from the Upper Midwest, only 4.3% of isolates lacked *SnToxA*. However, among Population 2, isolates from the Southern, Eastern, and Pacific Northwest, 93.8% lacked *SnToxA* (Table 4). Additionally, *SnToxA* was present in 100% (n = 17) of the isolates collected from Oklahoma, which were not assigned to a major population. These results indicate that strong selection pressure has favored maintaining *SnTox1* in natural US populations, likely due to its dual function [42]. Additionally, *SnToxA* has been selected for in Population 1, but in contrast selected against in Population 2, likely due to regional differences in deployment of the host sensitivity gene *Tsn1*.

In addition to the PAV exhibited by all three effectors, functional diversity was also detected within the coding regions of each gene. Throughout the entire population, *SnToxA* harbored 12 mutations, including eight synonymous SNPs, three nonsynonymous SNPs, and one SNP introducing a premature stop codon (Table 4). A total of four unique nucleotide haplotypes and protein isoforms were detected. The SNP inducing a premature stop codon was found in a single isolate collected on winter wheat from Ohio. Within the *SnTox1* coding region, no synonymous changes were detected; however, nine nonsynonymous SNPs were identified (Table 4). The nine non-synonymous changes form nine unique nucleotide haplotypes and protein isoforms. *SnTox3* was found to contain a total of nine SNPs, including three synonymous changes, five nonsynonymous SNPs, and one SNP causing the loss of the start codon (Table 4). The detected *SnTox3* variants collapsed into five nucleotide haplotypes and four protein isoforms. The start codon mutation was only observed in one isolate, collected on durum wheat in western North Dakota.

Disease phenotyping

Previous research has identified the presence of novel necrotrophic effectors in *P. nodorum* isolates collected from the southeastern United States and cognate host sensitivities specific to winter wheat germplasm [44]. In order to better characterize disease reactions of winter wheat to a diverse pathogen collection and potentially identify genes contributing to virulence via GWAS, wheat lines Alsen (*Tsn1* control), Jerry (hard red winter), TAM105 (hard red winter), ITMI38 (*Snn3* control), Massey (soft red winter), and F/G95195 (soft red winter) were inoculated with 197 *P. nodorum* isolates collected from spring, winter, and durum wheat production regions of the United States. Average disease reactions on spring wheat line Alsen ranged from 0.75 to 4.50 with an average of 3.13 (Fig 3). Disease reaction on hard red winter wheat line Jerry ranged from 0.25 to 4.25 with an average of 2.93 (Fig 3). Disease reaction on hard red
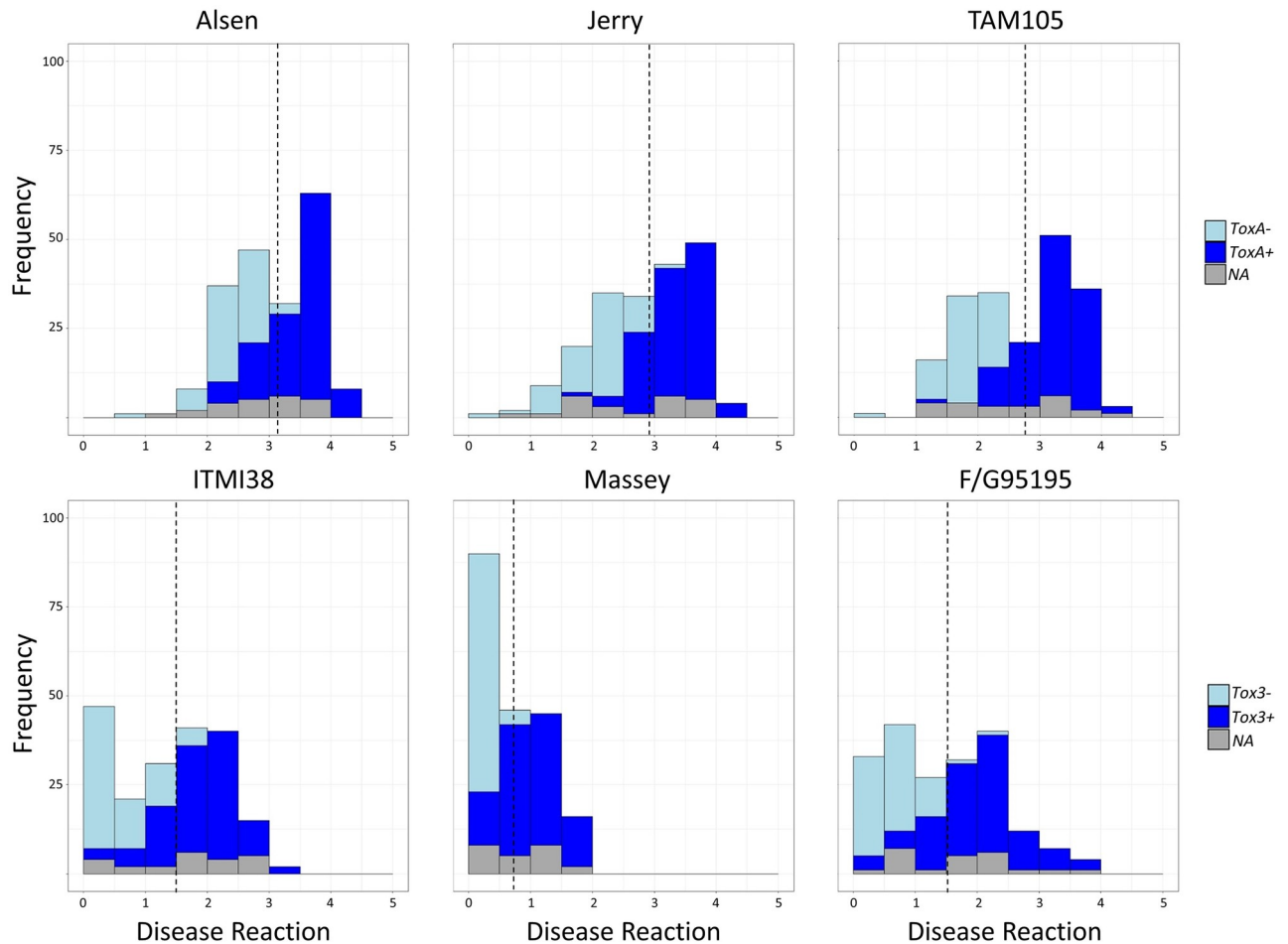
Fig 3. Histograms depicting the distribution of disease reactions to 197 *P. nodorum* isolates on wheat lines Alsen (*Tsn1*), Jerry, TAM105, ITMI38 (*Snn3*), Massey, and F/G95195. Disease reactions are displayed in bins along the x-axis and frequency is illustrated on the y-axis. Bars are shaded to illustrate the proportion of presence/absence of the corresponding effector (*SnToxA* or *SnTox3*) or missing data (NA) per bin. Light blue corresponds to the absence of the effector, dark blue corresponds to the presence of the effector, and grey corresponds to missing data.

https://doi.org/10.1371/journal.pgen.1008223.g003

winter wheat line TAM105 ranged from 0.13 to 4.13 with an average of 2.74 (Fig 3). An increased disease reaction correlated with the presence of a functional *SnToxA* gene, indicating that the SnToxA-Tsn1 interaction is largely responsible for disease on these wheat lines. Disease reaction on the recombinant inbred wheat line ITMI38 ranged from 0 to 3.50 with an average of 1.48 (Fig 3). Disease reaction on soft red winter wheat line Massey ranged from 0.00 to 2.00 with an average of 0.72 (Fig 3). Disease reaction on soft red winter wheat line F/G95195 ranged from 0.00 to 3.88 with an average of 1.50 (Fig 3). Disease reaction correlated with the presence of *SnTox3* which indicated that the SnTox3-Snn3 interaction is the main facilitator of disease on these wheat lines. Interestingly, the disease reactions on Massey were comparatively lower, indicating that although SnTox3 is an effective virulence factor, an underlying host resistance not present in the other wheat lines may exist.

GWAS provides new virulence enhancement candidates and insight into local levels of linkage disequilibrium

Previous studies have identified effectors that interact with host susceptibility genes derived from spring wheat germplasm. Novel effector-susceptibility gene interactions have been
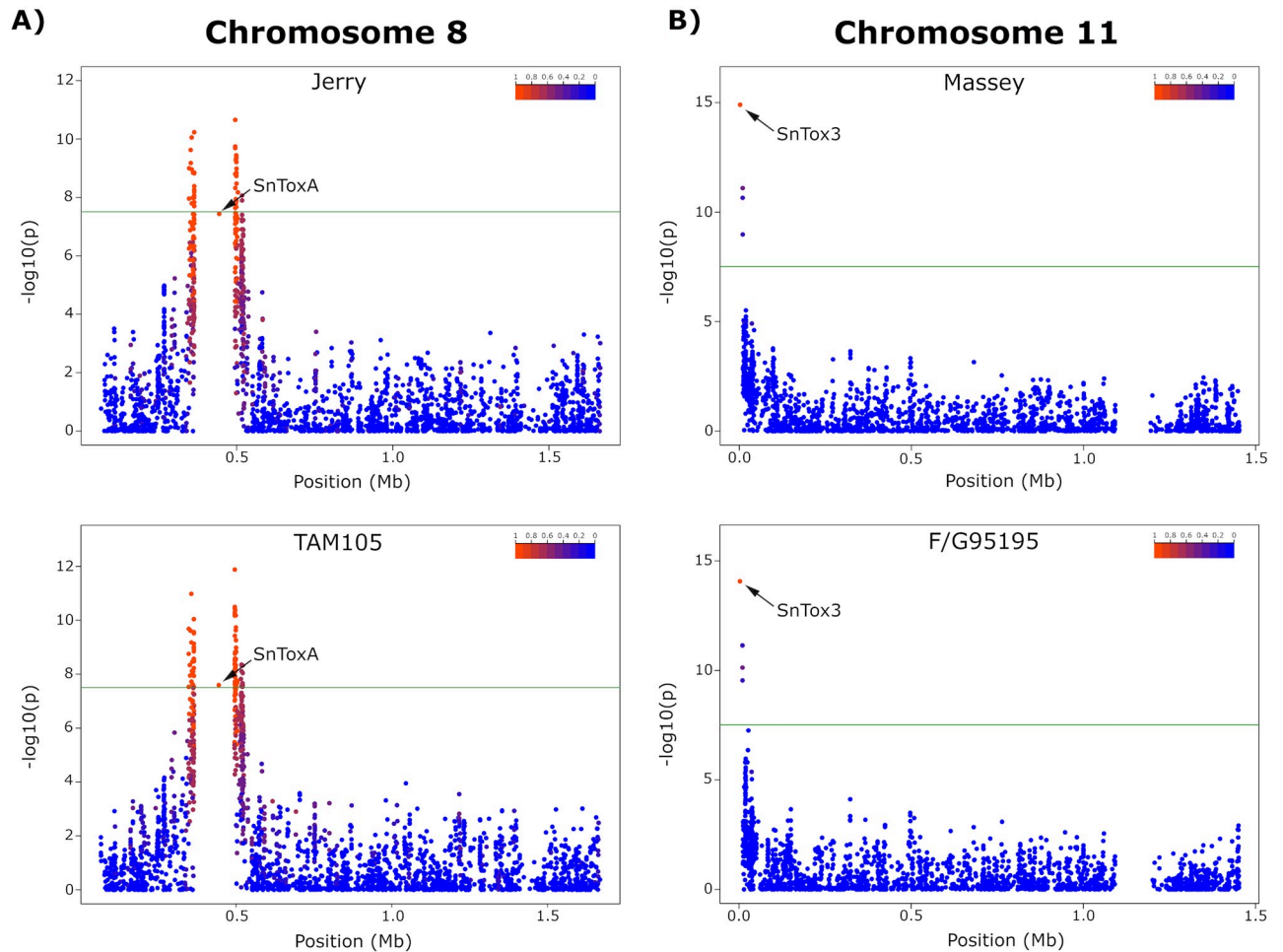
Fig 4. Genome-wide association study (GWAS) detecting significant associations with virulence on wheat lines Jerry, TAM105, Massey, and F/G95195. (A) Manhattan plot of *P. nodorum* chromosome 8 illustrating a highly significant association with the *SnToxA* locus for isolates inoculated onto hard red winter wheat lines Jerry and TAM105. Markers are represented by dots which are colored corresponding to the level of linkage disequilibrium (R^2) with the most significant marker and are order by position on the x-axis. The significance of each marker expressed as–log10(p) is displayed on the y-axis. (B) Manhattan plot of *P. nodorum* chromosome 11 illustrating a highly significant association with the *SnTox3* locus for isolates inoculated onto soft red winter wheat lines Jerry and TAM105. Markers are represented by dots which are colored corresponding to the level of linkage disequilibrium (R^2) with the most significant marker and are ordered by position on the x-axis. The significance of each marker expressed as–log10(p) is displayed on the y-axis.

https://doi.org/10.1371/journal.pgen.1008223.g004

identified that are specific to the winter wheat gene pool [44]. To identify candidate virulence genes underlying these novel interactions on winter wheat lines, as well as determine if previously identified effectors play a role in disease development, phenotypic and genotypic data were utilized to conduct a GWAS. Further filtering the quality SNPs/InDels identified from the *P. nodorum* natural population (n = 197) for a minor allele frequency of 5%, a total of 322,613 markers were used to identify associations with virulent phenotypes. Using a mixed linear model incorporating a kinship matrix, a total of 174 and 277 markers were identified as significant in *P. nodorum* for virulence on winter wheat lines Jerry and TAM105, respectively (Fig 4A). The same marker approximately 51.2 kb upstream of *SnToxA* on chromosome 8 was detected with the highest significance on both wheat lines and the PAV of *SnToxA* was also highly significant on each line (Fig 4A). *SnToxA* resides in an approximately 112.1 kb isochore region of chromosome 8 characterized by low GC content. Due to the repetitive nature of this

region, SNPs could not be reliably called, leaving the PAV of *SnToxA* as the only marker within this region. Among the 174 markers significantly associated with virulence on winter wheat line Jerry, 54 were in a 62.6 kb genomic region downstream and 111 were in an 87.0 kb region upstream of *SnToxA*. A total of eight markers were identified outside of the *SnToxA* locus, including four markers on chromosome 4, one marker on chromosome 10, two markers on chromosome 11, and one marker on chromosome 14 (S3 Fig). Out of the 277 significant markers identified for virulence on TAM105, 80 were in a 66.2 kb genomic region downstream and 173 were in an 88.1 kb region upstream of *SnToxA*. A total of 23 markers were detected outside of the *SnToxA* locus, including four markers on chromosome 1, one marker on chromosome 2, one marker on chromosome 3, three markers on chromosome 4, eight markers on chromosome 5, three markers on chromosome 9, two markers on chromosome 10, and one marker on chromosome 12 (S3 Fig).

Candidate genes involved in virulence on TAM105 were also identified at novel loci. The significant SNP on chromosome 1 at position 2,989,164 bp was within a glycosyl hydrolase family 11 gene, implicated in the degradation of plant cell walls. A high level of polymorphism was detected within this gene, as evidenced by 18 SNPs, including seven non-synonymous SNPs, all within the predicted functional domain. The most significant SNP was a non-synonymous change from valine to isoleucine at amino acid position 116. Directly flanking a significant SNP at position 1,484,536 on chromosome 5 by 1364 bp was an entire gene cluster exhibiting PAV. This cluster was absent from approximately 53% of the isolates and was comprised of a NAD(P)-binding monoxygenase, aldehyde dehydrogenase, aromatic ring opening dioxygenase, acyl esterase/dipeptidyl peptidase, and a transcription factor. In isolates harboring the cluster, the transcription factor was observed to harbor a high level of polymorphism, including 29 non-synonymous and 11 synonymous SNPs.

Interestingly, the PAV of *SnToxA* was not the most significant marker associated with virulence on winter wheat lines Jerry and Massey but still exhibited a high LD with the most significant marker ($R^2 = 0.97$). Additionally, LD extended approximately 48.6 kb downstream and 47.6 kb upstream of *SnToxA* before decaying to R^2 levels below 0.20, explaining the large number of SNPs identified in association with virulence. These results, confirmed by sensitive reactions of Jerry and TAM105 to infiltrations with SnToxA (Fig 5A), indicated that SnToxA was the major effector facilitating infection on these two winter wheat lines. However, other candidate genes contributing to disease development in a quantitative manner were identified at significant genomic loci and warrant further investigation.

Association mapping analyses using a mixed linear model incorporating three principal components (15.9% cumulative variation) as fixed effects and a kinship matrix as a random effect revealed four markers significantly associated with virulence on winter wheat line Massey. The most significant association corresponds to the PAV of *SnTox3* on chromosome 11, with the three remaining significant markers being SNPs located in an approximately 6.0 kb genomic region upstream of the *SnTox3* gene (Fig 4B). Using the same mixed linear model, a total of ten significant markers were detected in association with virulence on winter wheat line F/G 95195. Similar to virulence on Massey, *SnTox3* PAV was the most significant locus identified, with nine additional significant SNPs located in an ~23.1 kb upstream region (Fig 4B). LD decayed to levels below 0.20 approximately 6.5 kb upstream of *SnTox3*. As this gene is located near the telomere, no distal markers were identified; therefore, no LD estimations could be made downstream of *SnTox3*. As no markers outside of the *SnTox3* genomic region were significant, these results indicate that SnTox3 is the lone major segregating effector governing virulence on the winter wheat lines Massey and F/G 95195, which is confirmed by the sensitivity to infiltrations of SnTox3 (Fig 5B).
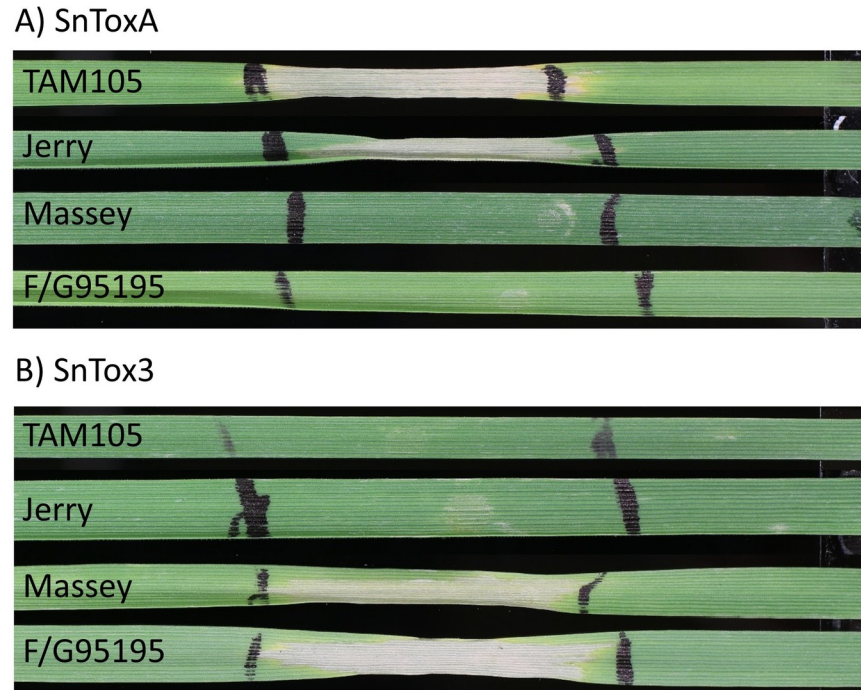
A) SnToxA

B) SnTox3

Fig 5. Sensitivity of winter wheat lines TAM105 (hard red), Jerry (hard red), Massey (soft red), and F/G95195 (soft red) to infiltrations with SnToxA (A) and SnTox3 (B). TAM105 and Jerry are sensitive to SnToxA but insensitive to SnTox3, indicating that they possess a functional *Tsn1* and lack *Snn3*. Massey and F/G95195 are insensitive to SnToxA but sensitive to SnTox3, indicating that they lack a functional *Tsn1* but harbor *Snn3*.

https://doi.org/10.1371/journal.pgen.1008223.g005

Genes residing on an accessory chromosome harbor an excess of non-synonymous SNPs

Examination of gene locations and corresponding pN/pS ratios revealed a significantly higher pN/pS ratio of genes residing on the accessory chromosome (Chromosome 23) compared to the core chromosomes (Pairwise Wilcoxon Rank Sum Test, $p < 2 \times 10^{-16}$, w = 333,700). Genes on the accessory chromosome had a median pN/pS value of 0.50 compared to the genome-wide median of 0.20 (Fig 6). This was also compared to a random subsample of equal size (n = 126 genes). The pN/pS medians of the five random subsamples ranged from 0.19–0.22, which is significantly different than the 126 genes residing on the accessory chromosome (Pairwise Wilcoxon Rank Sum Test, $p < 2 \times 10^{-16}$). This indicates that genes residing on the accessory chromosome are evolving faster, and this genomic compartment may serve as a hot-bed for the evolution of new traits.

Predicted effectors evolve faster than non-effectors

As effectors play an essential role in the development of disease and are directly affected by the selection pressure exerted by host resistance, we next wanted to determine if these genes are preferentially accumulating nonsynonymous changes, as well as if these changes are population-specific. Additionally, extreme forms of diversification including the abolition of gene function via loss-of-function mutations or the direct elimination of an effector gene were analyzed. Analysis of the *P. nodorum* isolate Sn4 annotated gene set revealed a total of 1,020 proteins containing a predicted secretion signal and lacking a predicted transmembrane domain. Further analysis using EffectorP revealed 219 of these proteins to be predicted effectors. This
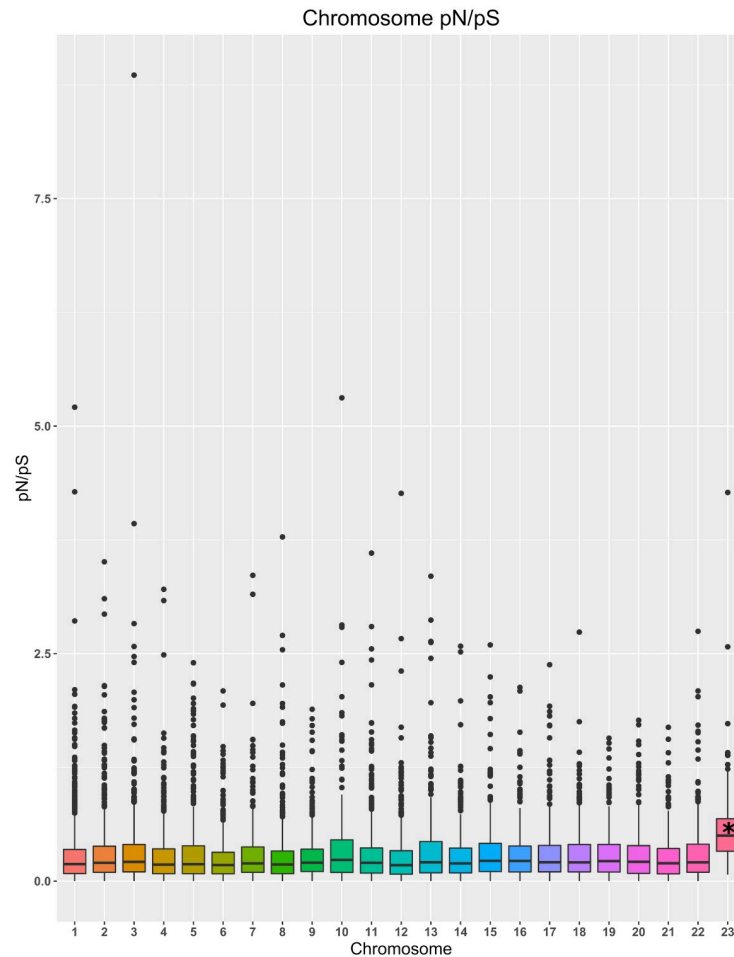
Fig 6. Distribution of pN/pS ratios of genes across all *P. nodorum* chromosomes. Specific chromosomes are displayed on the x-axis and pN/pS ratios are displayed on the y-axis. '*' indicates significantly different than each pairwise comparison at $p < 2 \times 10^{-16}$ (Pairwise Wilcoxon Rank Sum test, $p < 0.01$). Outliers with pN/pS values greater than 10 were omitted.

candidate effector list was used for further comparative analyses between the derived populations, excluding the isolates from Oklahoma and Oregon. Predicted effector proteins were observed to accumulate a greater number of nonsynonymous SNPs compared to secreted non-effectors or non-secreted proteins (Pairwise Wilcoxon Rank Test, $p < 0.01$) (Fig 7A). Genes encoding predicted effectors had average pN/pS ratios of 0.36 (median of 0.26), whereas secreted non-effectors and secreted proteins had average pN/pS ratios of 0.22 (median of 0.16) and 0.30 (median of 0.21), respectively. Also, the clear majority of predicted effector proteins did not have a predicted function, with only 27.6% having predicted functional domains, compared to 64% of secreted non-effector proteins. A set of five genes encoding predicted effector proteins, including previously characterized effector *SnTox3*, were found to have pN/pS ratios equal to or less than 1 in Population 2 but appeared to be diversifying in Population 1 (S3 Table). All five proteins lacked predicted functional domains. Also, seven predicted effectors lacking functional domains were found to have only non-synonymous SNPs (and pN/pS could not be calculated) in Population 1 but lacked non-synonymous changes in Population 2 (S3 Table). Conversely, four genes lacked an accumulation of nonsynonymous changes in Population 1 but had pN/pS ratios greater than 1 in Population 2. None of these proteins had
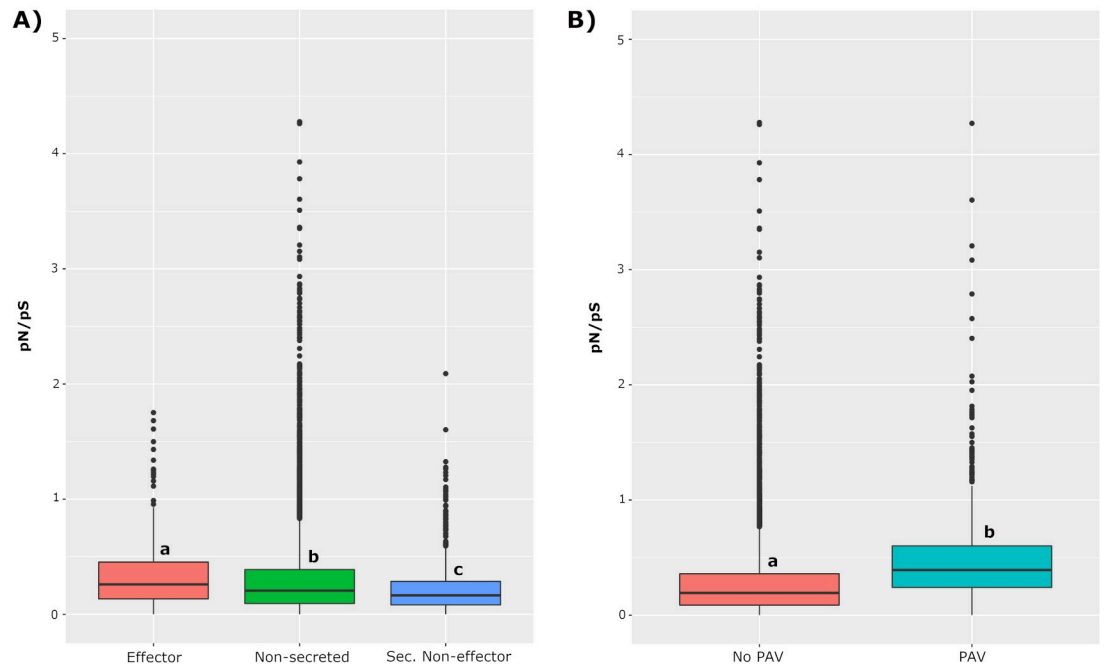
Fig 7. Distribution of pN/pS ratios among genes encoding (A) predicted effectors, secreted non-effectors, non-secreted proteins, and (B) genes exhibiting PAV. Categories are displayed on the x-axes and the log transformed pN/pS ratio is illustrated on the y-axes. Color legends are displayed to the right of each figure. Letter codes correspond to statistically different groups (Pairwise Wilcoxon Rank Test, p < 0.01). Outliers of pN/pS values greater than 10 are not displayed.

https://doi.org/10.1371/journal.pgen.1008223.g007

predicted functional domains. Additionally, a gene encoding a predicted effector with a chitin binding domain lacked nonsynonymous polymorphism in Population 1 but harbored nonsynonymous SNPs in Population 2 (S3 Table). These results indicate that effectors are preferentially accumulating nonsynonymous SNPs compared to other gene classes and that distinct suites of effectors are diversifying in specific geographic populations.

Differences were also observed with genes encoding predicted effector proteins harboring LOF mutations or PAV between populations. In total, LOF mutations affected 45 predicted effector genes, with 12 mutations being specific to Population 1 and 14 mutations being found only in Population 2. Overall, isolates from Population 1 had a significantly higher number of genes affected by LOF mutations, with 2,193 genes having frameshifts, loss of start codons, loss of stop codons, or gain of stop codons, compared to 1,849 detected in isolates from Population 2 (Kruskal-Wallis test, p-value < 0.001). The PAV for two predicted effectors, including *SnTox1*, was polymorphic in Population 2, but remained fixed in Population 1. Conversely, the PAV for predicted effectors was segregating in Population 1 but remained fixed in Population 2. Additionally, all genes (effector, secreted non-effector, or non-secreted) exhibiting PAV were observed to have significantly higher pN/pS ratios compared to those genes found in all isolates (Kruskal-Wallis test, $p < 2 \times 10^{-16}$), indicating that not only are they being strongly selected via elimination or gain of the entire gene, but the nucleotide sequences are evolving more rapidly (Fig 7B).

Repetitive elements are associated with PAV loci

As repetitive elements have been observed to be a large contributing factor in the dynamic nature of the fungal genome [1,2], we next wanted to examine the repeat content of
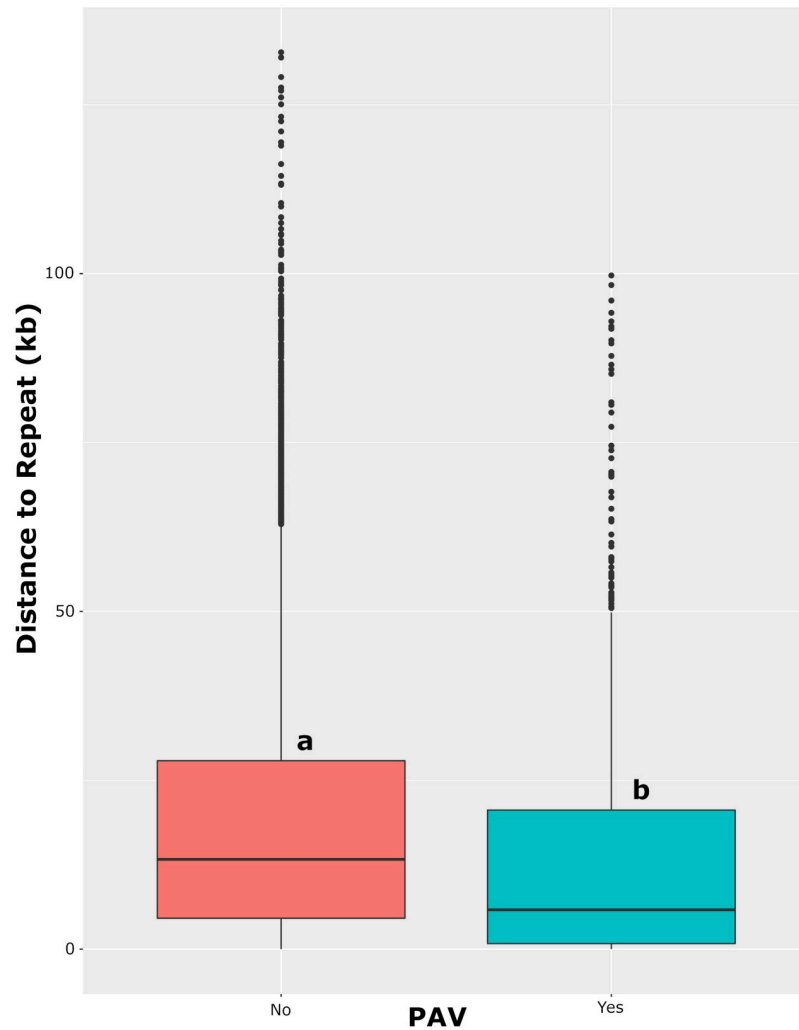
Fig 8. Distances of genes to the nearest repetitive element. Categories are listed on the x-axis and distance to nearest repeat on the y-axis. Different letter codes (a and b) correspond to statistically different groups (Kruskal-Wallis rank sum test p < 0.01).

P. nodorum and its relation to the diversity observed within the natural population. Repeat annotation classified 2,223,841 bp of the Sn4 genome as repetitive content, consisting of LTR/Copia (54.37%), LTR/Gypsy (22.33%), DNA/TcMar-Fot1 (11.64%), LINE/Penelope (2.60%), Satellites (0.29%), and unknown (8.77%) elements, all of which represented approximately 5.90% of the entire genome. Proximity of genes to the nearest repeat element was calculated to determine if transposable or repetitive elements influenced gene diversification. When comparing proximity to repetitive elements of genes exhibiting PAV, a large discrepancy was observed. Genes present in all isolates were a median distance of 13.3 kb away from the nearest repetitive element, while genes exhibiting PAV were significantly closer, being only a median distance of 5.8 kb away (Pairwise Wilcoxon Rank Sum Test, p < 2.0 x 10^{-16}; Fig 8). These results indicate that transposable or repetitive element activity may be influencing the gain or loss of genes and is a driving factor in the constantly evolving fungal genome.

Discussion

The dynamic nature of fungal genomes has been revealed following the increase in abundance of whole-genome sequences that facilitate the investigation of the way plant pathogenic fungi undergo diversification. Until recently, effector biology and genome research within *P. nodorum* has relied on relatively few genomic resources. Complementing the recently refined SN15 genome [31] and the establishment of nearly complete telomere-to-telomere reference genomes of three additional *P. nodorum* isolates [32], this research enhances our knowledge of genomic diversity within *P. nodorum* by revealing locale-specific effector diversification and evidence of host susceptibility genes influencing population structure. Due to the previous identification of the SnToxA effector, as well as characterization of the host susceptibility gene *Tsn1*, we were able to demonstrate the direct impact of the effector-host interaction on pathogen population allele frequencies and signatures of positive selection. This investigation also gives insight into the classes of genes undergoing diversifying selection through the accumulation of nonsynonymous SNPs, LOF mutations, and PAV, as well as identifying repetitive elements as a contributing factor to this diversification. Additionally, the results illustrate the utility of these data for association mapping strategies using linkage disequilibrium surrounding effector loci.

Population structure and genetic diversity

Population structure analyses clearly separated the natural population of *P. nodorum* isolates into two populations corresponding to geographic region. Due to the expansive region in which isolates belonging to Population 2 were collected compared to the three-state region giving rise to isolates from Population 1, we originally postulated that Population 2 would have higher levels of genetic diversity. However, a greater number of genic SNPs, both synonymous and nonsynonymous, were observed in Population 1 compared to Population 2. This disparity is also evidenced in the estimated π values for each population. One possible explanation is the age of the populations. If the United States *P. nodorum* population originated in the Upper Midwest and subsequently spread to the other wheat producing regions, more genome variants might have accumulated over time. Population size may also play a role in the observed differences and account for the increased level of nucleotide diversity observed in Population 1. Isolates in Population 2 may have recently experienced a bottleneck, resulting in a lower genome-wide Tajima's D compared to Population 1, as well as the identification of a larger number of potential selective sweep regions. Alternatively, differences may exist in the amount of selection pressure being placed on genes within each population. In the Upper Midwest, hard red spring, hard red winter, and durum wheat are cultivated, whereas predominantly just a single market class, soft red winter wheat, is grown in the Southern/Eastern United States. Additionally, it has been shown that differences exist in the presence of *P. nodorum* sensitivity genes between wheat germplasm in the Midwest and the South and Eastern United States [44,52]. If the genetic basis of host sensitivity of wheat grown in the Southern/Eastern United States is more uniform, with less variability in susceptibility genes present in the winter wheat cultivars grown, selection pressure may have a reduced effect on the genic variation of isolates collected from winter wheat regions. Conversely, with three wheat classes being grown in the Upper Midwest, the *P. nodorum* population may have been exposed to a diverse range of host resistance/susceptibility genes, therefore driving stronger genic variation.

Distribution and diversity of *SnToxA*, *SnTox1*, and *SnTox3*

A large difference was observed in the distribution of *SnToxA* among *P. nodorum* isolates. Only 4.3% of isolates from Population 1 had lost *SnToxA*, compared to the 93.8% of isolates

from Population 2 that lacked the gene (Table 2). Additionally, a selective sweep was detected in the genomic region flanking *SnToxA* specifically in Population 1, but not in Population 2 (Fig 2B). This dramatic difference in the prevalence of *SnToxA* is likely correlated with the absence of *Tsn1*, the dominant susceptibility gene that indirectly recognizes SnToxA, from the winter wheat cultivars planted in the soft red winter wheat region of the United States where most of Population 2 was collected. Previous research investigated the sensitivity of winter wheat breeding lines planted in the regions where isolates from the natural population were collected, particularly Georgia, Maryland, and North Carolina. Except for a single line from the breeding program at the USDA-ARS, Raleigh, North Carolina, all breeding lines from the regions were found to be insensitive to infiltrations of SnToxA, and therefore, lacked a functional *Tsn1* [54]. Conversely, previous research identified SnToxA sensitivity present in popular North Dakota spring wheat cultivars Glenn and Steele ND [55]. Additionally, *Tsn1* is present in the vast majority of wheat cultivars planted in Oklahoma [53]. The strong selective advantage given by SnToxA in the presence of a functional *Tsn1* explains the abundance of isolates harboring *SnToxA* in Oklahoma and the Upper Midwest. As *Tsn1* is fairly rare in the eastern winter wheat producing regions of the United States, the *SnToxA* gene was lost. A study by McDonald et al. [56] found *SnToxA* to be present in only 25% of a North American *P. nodorum* population, compared to an overall presence of 63.4% in the present study population. The large difference between these two studies is probably due to differences in the number of isolates used from the Upper Midwest and the rest of the United States. Previously, approximately 8% of the isolates used were from North Dakota, with the remaining isolates originating in winter wheat producing regions [55], compared to the population used in coverage analysis in this study, consisting of 59.5% of the isolates collected in the Upper Midwest where *Tsn1* is prevalent. The results from this study illustrate the powerful selection pressure placed on pathogen populations by the deployment of a single host susceptibility gene for the maintenance of the corresponding effector. Additionally, these results suggest that in the absence of the host susceptibility gene *Tsn1* in regional wheat cultivars, the presence of *SnToxA* may incur a fitness cost for the pathogen.

SnTox1 was the most prevalent *P. nodorum* effector found within this natural population, being present in 95.4% of the isolates (Table 4). This frequency was higher than that observed in a North American population by McDonald et al. [56], where it was found to be present in 70% of isolates examined, as well as in 84% of a global population. However, like the study by McDonald et al. [56], *SnTox1* was found to be the most widespread effector of the three characterized genes. Additionally, we detected nine unique haplotypes for *SnTox1* compared to the two private haplotypes previously detected in a North American population [56]. The wide-range distribution of *SnTox1* is likely due to its dual function in chitin binding and protection from wheat chitinases [57]. The ability of SnTox1 to trigger programmed cell death via recognition by Snn1 is a strong selective force [33,42] but is dependent on the presence/absence of *Snn1* in locally grown wheat cultivars. The more broad-range effector function of chitinase protection provided by SnTox1 likely explains its relatively high prevalence compared to the other necrotrophic effectors.

Effector diversification and selective sweeps are population dependent

Plant-pathogenic fungi are constantly adapting the way they infect their host, often using a suite of effectors to facilitate disease. These effectors are typically small, secreted proteins that lack known functional domains. Additionally, they may be cysteine-rich and exhibit signatures of diversifying selection [58,59]. Overall, the identified putative effectors in the *P. nodorum* genome exemplify these hallmarks, as evidenced by the accumulation of significantly more

non-synonymous changes compared to secreted non-effectors or non-secreted proteins. Additionally, 72.4% of predicted effector proteins lacked apparent functional domains. Interestingly, not only are effectors diversifying within the entire population, but effector diversification was also observed to occur within specific subpopulations. Previous studies have identified allelic differences between populations in functionally validated effector loci. Significant population structure was observed at the *AvrP123* and *AvrP4* loci between regions and local populations in *Melampspora lini* [60]. These differences were hypothesized to result from distinct interactions with local host populations. Brunner and McDonald [61] identified 30 isoforms of the effector AvrStb6 from *Zymoseptoria tritici*, with 27 isoforms being specific to a population, presumably due to local adaptation. Without relying on previously identified genes and using the genome-wide approach presented here, novel effectors that were observed to be diversifying in a population-specific manner may be prioritized for further investigation to expedite research on unexplored host-pathogen interactions. Selective sweep analysis identified 92 genes co-localizing with regions that have undergone a selective sweep. Interestingly, we identified multiple sweep regions specific to one of the two populations and no genes located in sweep regions were common to both populations. Similar results were observed in the wheat pathogen *Zymoseptoria tritici*, where the investigation of selective sweeps in four globally distinct populations revealed that genomic regions under selection were largely population specific [24]. This highlights the ability of plant-pathogenic fungi to rapidly fix advantageous mutations and subsequently shape a population.

Selective sweep analyses have also been used in a 'reverse ecology' approach to identify genomic regions associated with changes in regional climates. Ellison et al. [62] identified selective sweep regions in the model fungus *Neurospora crassa* with haplotypes specific to a population derived from Louisiana or the Caribbean. Following gene characterization within the sweep regions, it was hypothesized and later functionally validated that one sweep region corresponded to local adaptation to cold tolerance. In a similar manner, due to the large variation in climate between the Upper Midwest and the Southern/Eastern United States, it is likely that some of the selective sweeps identified in this study arose in response to regional environmental differences. Additionally, plant pathogenic fungi have developed effector proteins that have functions other than host colonization, such as competition with local microbial communities [63]. *P. nodorum* likely produces such effectors during its saprotrophic stage, and differences in the composition of local microbial populations may be contributing to the diversification or fixation of effector proteins within specific pathogen subpopulations. Combined, these results indicate that selection pressure exerted by specific host genes, regional climate conditions, or other biotic or abiotic environmental factors is driving effector/gene diversification as well as fixation and can be restricted to specific geographical regions.

It has been documented that multiple mutations often occur in adjacent sites, resulting in a multi-nucleotide mutation (MNM) that causes a nonsynonymous change [64]. Since MNMs are the result of a single mutation event and do not occur independently in a stepwise manner, they may lead to the overestimation of genes under positive selection [65] As the presence of MNMs was not accounted for in our analyses, the potential exists for false positives; however, we would not expect this phenomenon to differentially affect the gene classes under comparison here (effectors vs. secreted non-effectors vs. non-secreted proteins). Additionally, further investigation is warranted to determine the prevalence of MNMs in *P. nodorum*, as well as other plant pathogenic fungi, in order to more accurately measure positive selection.

Analysis of repeat content of the genome facilitated the comparison of the genomic location of genes and levels of diversity. The clear majority (90.9%) of identified repetitive DNA was classified within families of transposable elements. Genes exhibiting PAV were significantly closer to repetitive elements when compared to genes present in all isolates, suggesting that

transposable element activity may contribute to gene gain or loss. This phenomenon has also been observed in *Magnaporthe oryzae* [2] and provides a glimpse into one of the mechanisms giving fungal genomes their plasticity. Additionally, genes residing on the *P. nodorum* accessory chromosome were observed to be evolving faster than genes distributed elsewhere in the genome ([Fig 6]). Similar results have been observed in *Zymoseptoria tritici*, where up to eight accessory chromosomes have been observed in a single isolate [66]. The genes localized on these accessory chromosomes were found to have significantly higher pN/pS ratios when compared to genes residing on the core chromosomes, as well as a lower rate of recombination [67]. These results further support the hypothesis that transposable element activity mobilizes beneficial genes and, upon exposure to significant selection pressure in a given locale, the genes may become fixed in a pathogen subpopulation. Additionally, they indicate that the accessory chromosome of *P. nodorum* undergoes evolution at a different rate than the core chromosomes, possibly to accelerate adaptation to various selection pressures.

GWAS and LD decay

Using 322,613 SNP/InDel markers, as well as the PAV of *SnToxA* and *SnTox3* obtained through coverage analysis of whole-genome sequences, a robust framework for association mapping was created. This marker set translates to approximately one marker every 114 bp, overwhelmingly meeting our previous estimate of needing one marker every 7 kb to overcome LD decay [48]. Whole-genome sequencing provided the unique opportunity to examine LD surrounding validated effector loci. In the case of *SnTox3*, the PAV was the most significant marker detected, however, due to the sufficient marker density and LD extending to 6.5 kb, non-causal SNPs were also detected as significant. At the *SnToxA* locus, the PAV of the effector gene was the lone marker identified within the AT-rich isochore. These isochores are found interspersed throughout the genome and their repetitive nature directly hinders the identification of high-confidence nucleotide variants. In this case, the *SnToxA* PAV marker was not the most significant variant but was still in high LD with the most significant SNP residing in the region flanking the AT-rich isochore. This is likely due to missing data associated with individual isolates being removed from the PAV dataset due to low coverage across the entire gene set, subsequently reducing the significance of the marker. Interestingly, LD extended much further into the region flanking *SnToxA* compared to that of *SnTox3*. This resulted in the detection of two strong associations in the regions flanking the *SnToxA*-containing isochore. As *SnTox3* is in the subtelomeric region, its flanking region was likely more prone to recombination and the breakdown of LD, as subtelomeric regions are known to be recombination hotspots in fungi [68]. The opposite is likely true at the *SnToxA* locus. Due to the presence/absence nature of the region, recombination may be suppressed and therefore preserve LD.

GWAS also revealed a novel candidate virulence gene encoding a predicted hydrolase. Necrotrophic pathogens have long been thought to use CWD enzymes for the initiation of infection through the degradation of plant cells [66]. Although necessary for pathogenicity, cell wall degrading enzymes are not typically associated with direct or indirect interactions with host R genes, however, other classes of CWD enzymes have been observed to be under diversifying selection [69,70]. It is hypothesized that these proteins may trigger the plant basal immune response and are therefore diversifying to avoid this detection [71]. It is also possible that *P. nodorum* hydrolases are evolving rapidly to develop more efficient lysis of plant cells to enhance virulence. The identification of novel loci associated with virulence using association mapping methodology may facilitate the investigation into pathogen infection strategies, including pathways not necessarily involved in major gene interactions.

Conclusion

Prior to this study, focus had been placed on the investigation of diversity within previously identified effector genes. Whole-genome sequencing of 197 *P. nodorum* isolates collected from spring, winter, and durum wheat producing regions of the United States revealed the accumulation of non-synonymous changes in suites of effectors specific to geographical regions. Additionally, *SnToxA* was found to be present in nearly all isolates collected from regions where local wheat lines harbor *Tsn1* and absent from isolates collected from regions where *Tsn1* had been removed from locally grown cultivars. The *SnToxA* locus was detected near a selective sweep region, reinforcing this hypothesis. Selective sweep analysis also revealed unique genomic regions specific to *P. nodorum* subpopulations having undergone positive selection, indicating that strong and distinct selective forces are acting on each sub-population. Taken together, this illustrates the selective power that host susceptibility genes place on the pathogen populations and identifies candidate effector genes specific to the spring and winter wheat production regions of the United States. Also, the identified variants were successfully used in a GWAS to identify strong associations of *SnToxA* and *SnTox3* with disease reaction on winter wheat lines, as well as to detect novel virulence candidates. Vastly different patterns of LD were observed surrounding these loci, highlighting the importance of marker density for the successful identification of effector loci in association mapping. Further investigation of *P. nodorum* genomics and functional characterization of effector candidates is ongoing and will provide further information on how this destructive pathogen is interacting with its host.

Materials and methods

DNA extraction and whole genome sequencing

Dried agarose plugs of each isolate were placed in liquid Fries media [33] and cultured for approximately 48 hours. Fungal tissue was then collected, lyophilized, and homogenized using Lysing Matrix A (MP Biomedicals) by vortexing for 3 minutes on maximum speed. Genomic DNA was extracted using the Biosprint using the manufacturer's protocol. DNA was enzymatically fragmented using dsDNA fragmentase (New England Biolabs) and whole genome sequencing libraries were prepared using the NEBNext Ultra II Library kit (New England Biolabs) according the recommended protocol. NEBNext Multiplex Oligos for Illumina were used to uniquely index libraries and were subsequently sequenced at the Beijing Genome Institute (BGI) on an Illumina HiSeq 4000. Raw sequencing reads of each isolate were uploaded to the NCBI short read archive under BioProject PRJNA398070.

Variant identification

Quality of sequencing reads were analyzed using FastQC [72] and subsequently trimmed using trimmomatic [73]. Trimmed reads were mapped to the *P. nodorum* isolate Sn4 [32] reference genome (NCBI BioProject PRJNA398070) using BWA-MEM [74]. Sequencing coverage of the genome was calculated per isolate using GATK 'Depth of Coverage' using default settings. SNPs/InDels were identified using SAMtools 'mpileup' [75]. Variants were then filtered for individual genotype quality equal to or greater than 40 with at least three reads supporting the variant and all heterozygous calls were coded as missing data. For GWAS analysis, all SNPs/INDELs with missing data greater than 30% or a minor allele frequency less than 5% were removed from the dataset.

Population structure

SNP data were read into the R statistical environment [76] and converted to a genlight object using the package 'vcfR' [77]. PCA was conducted within the R package 'adegenet' [78] using a randomly selected subset of 50,000 markers.

A marker subset minimizing potential linkage disequilibrium between marker pairs consisting of approximately one SNP/10 kb was created in TASSEL v5 [79] and used to infer population structure using the software STRUCTURE [80]. A burn-in of 10,000 followed by 25,000 MCMC replications using the admixture model was completed for each k value from 1–8 with three iterations. Optimal sub-population level was determined using the method developed by Evanno et al. [81] and implemented in StructureHarvester [82]. STRUCTURE analysis was re-run using the optimal k value with a burn-in period of 10,000 and 100,000 MCMC. An individual isolate was assigned to a specific subpopulation if membership probability was greater than 0.85. Subsets corresponding to SNP/InDel calls for individuals within specific populations were created using VCFtools [83]. ADMIXTOOLS [52] was used to conduct a 3-population test to determine if evidence existed for the hypothesis that isolates collected from Oklahoma resulted from admixture between Population 1 (Upper Midwest) and Population 2 (South/East United States). Genotypic data was converted to PLINK format in TASSEL 5. Genotypic data was further converted to EIGENSTRAT format using 'convertf' provided in ADMIXTOOLS. A 3-population test was conducted using the two major populations identified using STRUCTURE as the sources and the isolates collected from Oklahoma as the target population. As recommended in the software documentation, the 'inbreed' option was selected due to the haploid nature of the organism. Additionally, as the f_3 statistic uses a measure of heterozygosity for normalization, as recommended in the documentation, the parameter 'OUTGROUP = YES' was used. Isolates collected from Oklahoma (n = 17), not clearly belonging to either major population, were removed from the datasets when comparing pN/pS ratios to obtain a more representative estimate of diversifying or purifying selection for each population.

Population genomics

Using genotypic data for 197 *P. nodorum* isolates and population assignments derived from STRUCTURE analysis, population genomics analyses were conducted in the R package 'PopGenome' specifying a ploidy level of one [84]. F_{st}, nucleotide diversity (π), Watterson's Θ, and Tajima's D were calculated across individual chromosomes and averaged to obtain a genome-wide value. Tajima's D was also calculated in 50 kb windows in 25 kb steps across each chromosome to examine regions that substantially lack allelic diversity and may provide evidence for selective sweeps.

A likelihood-based method implemented in SweeD [85] was used to detect regions of the genome that have undergone a selective sweep. Chromosomes were divided into approximate 1 kb grids and each population was analyzed separately. The option '-folded' was selected to consider the site frequency spectrum as folded, as to not distinguish between ancestral and derived states. Grids within the 99th percentile of likelihood values were extracted and examined for values of Tajima's D. If the value of Tajima's D fell below zero within an interval, it was considered as a selective sweep region. Selective sweep grids within 10 kb were considered a single region. Genes underlying selective sweep regions were extracted using BEDtools 'intersect' [86] and gene ontology enrichment analysis was conducted as described in a subsequent section.

Disease phenotyping and effector infiltration

The natural *P. nodorum* population (n = 197) used in this study consists of 51 isolates collected from spring wheat in North Dakota and Minnesota, 45 isolates collected from durum wheat in

North Dakota, nine isolates collected from winter wheat in South Dakota, and 92 isolates collected from winter wheat in the Eastern, Southern, and Pacific Northwest regions of the United States (S1 Table). These isolates were chosen for their diversity in geographical origin of isolation, as well as the differences in wheat class (spring, winter, and durum wheat). *P. nodorum* isolates were cultured as described by Friesen and Faris [87]. Briefly, dried agarose plugs stored at -20 ˚C of each isolate were brought to room temperature and rehydrated by placing on V8-potato dextrose agar (150 mL V8 juice, 3 g $CaCO_3$, 10 g Difco PDA, 10 g agar, 850 mL H_2O). Plugs were then spread across the plate to distribute the spores and the plates were then incubated at room temperature under a constant fluorescent light regimen for seven days. Pycnidial spores were harvested by flooding the plates with sterile H_2O and agitation with a sterile inoculation loop. Spores were counted using a hemocytometer, concentration was adjusted to 1×10^6 spores per mL, and two drops of Tween20 were added per 100 mL of spore suspension.

Disease reaction to the 197 *P. nodorum* isolates was assessed on six wheat cultivars including spring (Alsen) and winter types (Jerry, TAM105, Massey, and F/G95195), as well as a recombinant from a synthetic wheat population (ITMI38). These lines were chosen based on preliminary data that these lines may harbor different susceptibilities with ITMI38 being a differential line for the *Snn3*-SnTox3 interaction. Three seeds of each wheat genotype were sown into cones, comprising a single replicate as described by Friesen and Faris [87]. A border of wheat cultivar Alsen was planted to reduce the edge effect. Plants were grown in the greenhouse for approximately two weeks, until the two- to three-leaf stage. Inoculations were conducted as described by Friesen and Faris [87]. Briefly, plants were inoculated using a paint sprayer until leaves were fully covered, until runoff. Inoculated plants were placed in a mist chamber at 100% humidity for 24 hours and subsequently moved to a climate-controlled growth chamber at 21 ˚C with a 12-hour photoperiod for six days. Disease ratings were taken seven days post-inoculation using the 0–5 scale as described by Liu et al. [88]. For each wheat line, at least three replicates were conducted per *P. nodorum* isolate and the averages of the replicates were used as the phenotypic data for downstream analyses.

To produce effector proteins used in bio-assays to validate associations detected with known necrotrophic effectors, cells from glycerol stocks of *P. pastoris* expressing previously characterized SnToxA and SnTox3 were plated on YPD media amended with Zeocin (Invitrogen) at 25 μg/mL and incubated at 30 ˚C for 2–4 days until colony formation [35,36]. Colonies were picked and grown in 2 mL of liquid YPD media and grown for 24 hours at 30 ˚C and shaking at 250 rpm. A 500 μL aliquot of each starter culture was transferred to 50 mL of YPD media and incubated at 30 ˚C and shaking at 250 rpm for 48 hours. Cultures were transferred to 50 mL conical tubes and centrifuged at 3,166 rcf for 10 minutes. Supernatants were decanted and filtered using a 0.45 μm Durapore Membrane Filter (Merck Millipore Ltd.). Filtered supernatant was then lyophilized overnight and stored at -20 ˚C. Seeds of winter wheat lines Jerry, TAM105, Massey, and F/G95195 were sown in cones and grown under greenhouse conditions for approximately 14 days until the secondary leaf was fully emerged. Freeze-dried protein samples of SnToxA and SnTox3 were resuspended in ddH_2O and infiltrated into secondary wheat leaves using a needleless syringe. Plants were placed in a growth chamber at 21 ˚C with a 12-hour photoperiod and evaluated for necrotrophic effector sensitivity after three days.

Genome-wide association analysis

Association mapping was conducted using TASSEL v5 [79] and GAPIT [89,90]. A naïve model, as well as a model including the first three components derived from a PCA as fixed effects were used in association analyses in TASSEL v5. A kinship matrix (K) was calculated

using the EMMA algorithm. Models incorporating K as a random effect, as well as a model using a combination of PCA and K were analyzed in GAPIT. The most robust model was chosen for each trait by visualization of Q-Q plots produced by GAPIT, illustrating the observed vs expected unadjusted p-values. Marker p-values were adjusted using a false discovery rate (FDR) in the R Statistical Environment. Markers with a FDR adjusted p-value of 0.05 or less were deemed significant.

Linkage disequilibrium

Linkage disequilibrium (LD) was calculated between the most significant marker detected from association mapping analyses and each intrachromosomal marker in TASSEL v5 [79]. LD decay was determined by averaging R^2 values across sliding windows consisting of five markers and observing when average R^2 values fell below 0.20 for three consecutive windows. Position of the extent of LD was taken as the position of the marker at the center of the sliding window before decay.

Putative effector diversification analyses

The presence of predicted secretion signals and transmembrane domains were identified using DeepSig [91]. Effector candidates were predicted via EffectorP using protein sequences of secreted proteins [92]. For comparative analyses, proteins were categorized as effectors (identified by EffectorP), secreted non-effectors (predicted secreted proteins without transmembrane domain or effector predictions), or non-secreted. Genomic positions of all genes were obtained and used to calculate genome coverage with BEDtools 'coverage' [86]. A gene was classified as absent if sequencing coverage across the entire length of the gene was less than 40%.

A consensus genome sequence was obtained for each isolate using bcftools consensus, specifying "—sample" [93]. The perl script 'gff2fasta.pl' was used to extract the coding regions for each isolate separately (https://github.com/minillinim/gff2fasta/blob/master/gff2fasta.pl). Coding regions of each gene were aligned using Clustal Omega [94] and the subsequent alignments were concatenated. The average number of nonsynonymous (NSite) and synonymous sites (SSite) were calculated using egglib [95].

Synonymous (S) and non-synonymous (N) SNPs, as well as LOF mutations were identified using SNPeff [96] and SNPsift [97] with the *P. nodorum* isolate Sn4 reference genome annotation. LOF mutations include small InDels causing frameshifts, as well as SNPs/InDels resulting in the loss of a start codon, loss of a stop codon, or gain of a premature stop codon. Isolates from specific subpopulations were grouped and fixed SNPs/LOF mutations (minor allele frequency of 0%) were removed from each dataset. pN/pS ratios were obtained for each gene by the following formula: (N/NSites)/(S/SSites). Genes with pN/pS ratios greater than one were considered to be under diversifying selection, while genes with pN/pS ratios equal to or less than one were considered to be under neutral or purifying selection. Raw counts of LOF mutations were used in comparisons between subpopulations. Protein sequences of the *P. nodorum* isolate Sn4 annotation were input into InterProScan [98] to identify conserved domains and gene ontology. Gene ontology enrichment analysis was conducted in the R Statistical Environment using the package 'topGO' [99]. Analysis was conducted on both molecular function and biological process associated ontology terms using the classic algorithm and significance of overrepresented terms were determined using Fisher's exact test. GO terms with FDR adjusted p-values of less than 0.10 were declared significantly enriched.

Repetitive element annotation and analysis

De novo identification of repetitive element families in the Sn4 reference genome was conducted using RepeatModeler [100]. Identified repetitive element families were used as input to RepeatMasker [101] to annotate the repetitive regions of the Sn4 genome. Distance to the nearest repetitive element from each annotated gene was calculated using bedtools 'closest' [86].

Supporting information

S1 Fig. STRUCTURE analysis of 197 *P. nodorum* isolates. A) Evanno method indicated the optimal k value (number of subpopulations) to be two. B) Using a k-value of two, the 197 *P. nodorum* natural population is divided into two populations. Population 1 (red) consists of isolates from the Midwestern United States. Population 2 (blue) consists of isolates from the Southern/Eastern region of the United States, as well as Oregon.
(TIF)

S2 Fig. Boxplots illustrating the distribution of (A) nucleotide diversity, (B) Tajima's D, and (C) Watterson's Theta calculated in 50 kb windows within Population 1 and Population 2.
(TIF)

S3 Fig. Whole genome Manhattan plots illustrating significant associations with virulence on wheat lines TAM105, Jerry, Massey, and F/G95195. Dots represent individual SNPs/InDels. Markers are ordered by position and chromosomes are displayed on the x-axis. The–log10(p) value is displayed on the y-axis. The horizontal line represents the significance threshold at and FDR adjusted p-value of 0.05.
(TIF)

S1 File. Results of the gene ontology enrichment analysis conducted using topGO.
(XLSX)

S2 File. Phenotypic data.
(XLSX)

S1 Table. Collection information of the fungal isolates used in this study.
(XLSX)

S2 Table. Genome sequencing coverage statistics for each isolate.
(XLSX)

S3 Table. Genes under population-specific diversifying selection.
(XLSX)

S4 Table. STRUCTURE likelihoods.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Zhaohui Liu, Timothy L. Friesen.

Data curation: Jonathan K. Richards.

Funding acquisition: Timothy L. Friesen.

Investigation: Jonathan K. Richards, Jessica Carpenter.

Methodology: Eva H. Stukenbrock, Jessica Carpenter.

Project administration: Timothy L. Friesen.

Resources: Christina Cowger, Justin D. Faris, Timothy L. Friesen.

Supervision: Zhaohui Liu, Timothy L. Friesen.

Writing – original draft: Jonathan K. Richards, Timothy L. Friesen.

Writing – review & editing: Eva H. Stukenbrock, Christina Cowger, Justin D. Faris, Timothy L. Friesen.

References

1. Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, Dean RA. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. Genome Biology. 2006; 7:R16 https://doi.org/10.1186/gb-2006-7-2-r16 PMID: 16507177

2. Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, Kamoun S, Terauchi R. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 2016; 17:370 https://doi.org/10.1186/s12864-016-2690-6 PMID: 27194050

3. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 2013; 23: 1271–1282 https://doi.org/10.1101/gr.152660.112 PMID: 23685541

4. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2016; 26: 1091–1100 https://doi.org/10.1101/gr.204974.116 PMID: 27325116

5. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, Meyers BC, Nusbaum C, Kamoun S. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science. 2010; 330: 1540–1543 https://doi.org/10.1126/science.1193070 PMID: 21148391

6. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. PLoS Pathog. 2012; 8(4): e1002608 https://doi.org/10.1371/journal.ppat.1002608 PMID: 22570606

7. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Devel. 2015; 35: 57–65

8. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012; 10: 417–430 https://doi.org/10.1038/nrmicro2790 PMID: 22565130

9. Frantzeskakis L, Kracher B, Kush S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics. 2018; 19:381 https://doi.org/10.1186/s12864-018-4750-6 PMID: 29788921

10. Frantzeskakis L, Kusch S, Panstruga R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. Mol. Plant Pathol. 2019; 20(1): 3–7 https://doi.org/10.1111/mpp.12738 PMID: 30557450

11. Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell. 2006; 124(4): 803–814 https://doi.org/10.1016/j.cell.2006.02.008 PMID: 16497589

12. Zipfel C. Early molecular events in PAMP-triggered immunity. Curr Opin Plant Biol. 2009; 12(4): 414–420 https://doi.org/10.1016/j.pbi.2009.06.003 PMID: 19608450

13. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. Nature Rev Micro. 2013; 11: 800–814

14. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MHA, Thomma BPHJ. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science. 2010; 329: 953–955 https://doi.org/10.1126/science.1190859 PMID: 20724636

15. Jones JDG, Dangl JL. The plant immune system. Nature. 2006; 444: 323–329 https://doi.org/10.1038/nature05286 PMID: 17108957

16. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. Nature Rev Genet. 2010; 11: 539–548 https://doi.org/10.1038/nrg2812 PMID: 20585331

17. Flor HH. The complementary genic systems in flax and flax rust. Advances in genetics. 1956; 8: 29–54

18. Friesen TL, Meinhardt SW, Faris JD. The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. Plant J. 2007; 51: 681–692 https://doi.org/10.1111/j.1365-313X.2007.03166.x PMID: 17573802

19. Aguileta G, Refregier G, Tockteng R, Fournier E, Giraud T. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses, and protists. Infect. Genet. Evol. 2009; p(4): 656–670 https://doi.org/10.1016/j.meegid.2009.03.010 PMID: 19442589

20. Stukenbrock EH, McDonald BA. Geographical variation and positive diversifying selection in the host-specific toxin *SnToxA*. Mol. Plant Pathol. 2007; 8(3): 321–332 https://doi.org/10.1111/j.1364-3703.2007.00396.x PMID: 20507502

21. Stukenbrock EH, Jorgensen FG, Zala M, Hansen TT, McDonald BA, Shierup MH. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. PLoS Genet. 2010; 6(12): e1001189 https://doi.org/10.1371/journal.pgen.1001189 PMID: 21203495

22. Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, Snirc A, Le Prieur S, Jeziorski C, Branca A, Giraud T. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. Mol. Ecol. 2017; 26(7): 2041–2062 https://doi.org/10.1111/mec.13976 PMID: 28012227

23. Mohd-Assaad N, McDonald BA, Croll D. Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. Genome Biol. Evol. 2018; 10(5): 1315–1332 https://doi.org/10.1093/gbe/evy087 PMID: 29722810

24. Hartmann FE, McDonald BA, Croll D. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. Mol. Ecol. 2018; 27(12): 2725–2741 https://doi.org/10.1111/mec.14711 PMID: 29729657

25. FAOSTAT. Crop Production Data. http://www.fao.org/faostat/en/#data/QC

26. Cf Morris, Rose SP. Wheat. In Cereal grain Quality. 1996 (pp. 3–54) Springer, Dordrecht

27. USDA-ERS. Wheat Sector at a Glance. December 2018. https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/

28. Oliver RP, Friesen TL, Faris JD, Solomon PS. *Stagonospora nodorum*: From Pathology to Genomics and Host Resistance. Ann Rev Phytopathol. 2012; 50: 23–43

29. Hane JK, Lowe RGT, Solomon PS, Tan KC, Schoch CL, Spatafora JW, Crous PW, Kodira C, Birren BW, Galagan JE, Torriani SFF, McDonald BA, Oliver RP. Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell. 2007; 19: 3347–3368 https://doi.org/10.1105/tpc.107.052829 PMID: 18024570

30. Syme RA, Hane JK, Friesen TL, Oliver RP. Resequencing and comparative genomics of *Stagonospora nodorum*: Sectional gene absence and effector discovery. G3: Genes, Genomes, Genetics. 2013; 3(6): 959–969

31. Syme RA, Tan KC, Hane JK, Dodhia K, Stoll T, Hastie M, Furuki E, Ellwood SR, Williams AH, Tan YF, Testa AC, Gorman JJ, Oliver RP. Comprehensive annotation of the *Parastagonospora nodorum* reference genome using next-generation genomics, transciptomics and proteogenomics. PLoS One. 2016; 11(2): e0147221 https://doi.org/10.1371/journal.pone.0147221 PMID: 26840125

32. Richards JK, Wyatt NA, Liu Z, Faris JD, and Friesen TL. Reference quality genome assemblies of three *Parastagonospora nodorum* isolates differing in virulence on wheat. G3: Genes, Genomes, Genetics. 2018; 8(2): 393–399

33. Liu ZH, Faris JD, Meinhardt SW, Ali S, Rasmussen JB, Friesen TL. Genetic and physical mapping of a gene conditioning sensitivity in wheat to a partially purified host-selective toxin produced by *Stagonospora nodorum*. Phytopathology 2004; 94: 1056–1060 https://doi.org/10.1094/PHYTO.2004.94.10.1056 PMID: 18943793

34. Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, Fellers JP, Meinhardt SW, Rasmussen JB, Xu SS, Oliver RP, Simons KJ, Friesen TL. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci. 2010; 1007: 13544–13549

35. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet. 2006; 38: 953–956 https://doi.org/10.1038/ng1839 PMID: 16832356

36. Liu Z, Faris JD, Oliver RP, Tan KC, Solomon PS, McDonald MC, McDonald BA, Nunez A, Lu S, Rasmussen JB, Friesen TL. SnTox3 acts in effector triggered susceptibility to induce disease on wheat carrying the *Snn3* gene. PLoS Pathog. 2009; 5: e1000581 https://doi.org/10.1371/journal.ppat.1000581 PMID: 19806176

37. Zhang Z, Friesen TL, Xu SS, Shi G, Liu Z, Rasmussen JB, Faris JD. Two putatively homoeologous wheat genes mediate recognition of SnTox3 to confer effector-triggered susceptibility to *Stagonospora nodorum*. Plant J. 2011; 65: 27–38 https://doi.org/10.1111/j.1365-313X.2010.04407.x PMID: 21175887

38. Abeysekara NS, Friesen TL, Keller B, Faris JD. Identification and characterization of a novel host-toxin interaction in the wheat-*Stagonospora nodorum* pathosystem. Theor Appl Genet. 2009; 120: 117–126 https://doi.org/10.1007/s00122-009-1163-6 PMID: 19816671

39. Friesen TL, Chu C, Xu SS, Faris JD. SnTox5-*Snn5*: a novel *Stagonospora nodorum* effector-wheat gene interaction and its relationship with the SnToxA-*Tsn1* and SnTox3-*Snn3-B1* interactions. Mol Plant Pathol. 2012; 13: 1101–1109 https://doi.org/10.1111/j.1364-3703.2012.00819.x PMID: 22830423

40. Shi G, Friesen TL, Saini J, Xu SS, Rasmussen JB, Faris JD. The wheat *Snn7* gene confers susceptibility on recognition of the *Parastagonospora nodorum* necrotrophic effector SnTox7. The Plant Genome. 2015; 8(2)

41. Gao Y, Faris JD, Li Z, Kim YM, Syme RA, Oliver RP, Xu SS, Friesen TL. Identification and Characterization of the SnTox6-*Snn6* Interaction in the *Parastagonospora nodorum*-Wheat Pathosystem. Mol Plant Microbe Interact. 2015; 28(5): 615–625 https://doi.org/10.1094/MPMI-12-14-0396-R PMID: 25608181

42. Liu Z, Zhang Z, Faris JD, Oliver RP, Syme R, McDonald MC, McDonald BA, Solomon PS, Lu S, Shelver WL, Xu S, Friesen TL. The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring *Snn1*. PLoS Pathog. 2012; 8: e1002467 https://doi.org/10.1371/journal.ppat.1002467 PMID: 22241993

43. Shi G, Zhang Z, Friesen TL, Raats D, Fahima T, Brueggeman RS, Lu S, Trick HN, Liu Z, Chao W, Frenkel Z, Xu SS, Rasmussen JB, Faris JD. The hijacking of a receptor kinase-driven pathway by a wheat fungal pathogen leads to disease. Science Advances. 2016; 2(10): e1600822 https://doi.org/10.1126/sciadv.1600822 PMID: 27819043

44. Crook AD, Friesen TL, Liu ZH, Ojiambo PS, Cowger C. Novel necrotrophic effectors from *Stagonospora nodorum* and corresponding host sensitivities in winter wheat germplasm in the southeastern United States. Phytopathology. 2012; 203(5): 498–505

45. Sommerhalder RJ, McDonald BA, and Zhan J. The frequencies and spatial distribution of mating types in *Stagonospora nodorum* are consistent with recurring sexual reproduction. Phytopathology 2006; 96: 234–239 https://doi.org/10.1094/PHYTO-96-0234 PMID: 18944437

46. Cowger C, and Silva-Rojas HV. 2006. Frequency of *Phaeosphaeria nodorum*, the sexual stage of *Stagonospora nodorum*, on winter wheat in North Carolina. Phytopathology. 2006; 96:860–866. https://doi.org/10.1094/PHYTO-96-0860 PMID: 18943751

47. Sánchez-Vallet A, Hartmann FE, Marcel TC, Croll D. Nature's genetic screens: using genome-wide association studies for effector discovery. Mol Plant Pathol. 2018; 19(1): 3–6 https://doi.org/10.1111/mpp.12592 PMID: 29226559

48. Gao Y, Liu A, Faris JD, Richards J, Brueggeman RS, Li X, Oliver RP, McDonald BA, Friesen TL. Validation of genome-wide association studies as a tool to identify virulence factors in *Parastagonospora nodorum*. Phytopathology 2016; 106(10): 1177–1185 https://doi.org/10.1094/PHYTO-02-16-0113-FI PMID: 27442533

49. Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, Zala M, Ducasse A, Confais J, Compain J, Lapalu N, Amselem J, McDonald BA, Croll D, Palma-Guerrero J. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. New Phytologist. 2017; 214(2): 619–631 https://doi.org/10.1111/nph.14434 PMID: 28164301

50. Kema GHJ, Gohari AM, Aouini L, Gibriel HAY, Ware SB, van den Bosch F, Manning-Smith R, Alonso-Chavez V, Helps J, M'Barek SB, Mehrabi R, Diaz-Trujillo CD, Zamana E, Schouten HJ, van der Lee TAJ, Waalwijk C, de Waard MA, de Wit PJGM, Verstappen ECP, Thomma BPHJ, Meijer HJG, Seidl MF. Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin resistance. Nature genet. 2018; 50: 375–380 https://doi.org/10.1038/s41588-018-0052-9 PMID: 29434356

51. Rouxel T, Penaud A, Pinochet X, Brun H, Gout L, Delourme R, Schmit J, Balesdent M. A 10-year survey of populations of *Leptosphaeria maculans* in France indicates a rapid adaptation towards the *Rlm1* resistance gene of oilseed rape. Eur. J. Plant Pathol. 2003; 109(8): 871–881

52. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012; 192(3): 1065–1093 https://doi.org/10.1534/genetics.112.145037 PMID: 22960212

53. Friesen TL, Holmes DJ, Bowden RL, Faris JD. *ToxA* is present in the U.S. *Bipolaris sorokiniana* population and is a significant virulence factor on wheat harboring *Tsn1*. Plant Disease. 2018; 102(2): 2446–2452

54. Bertucci M, Brown-Guedira G, Murphy JP, Cowger C. Genes conferring sensitivity to *Stagonospora nodorum* necrotrophic effectors in Stagonospora nodorum blotch-susceptible U.S. wheat cultivars. Plant Disease. 2014; 98(6): 749–753

55. Zhang Z, Friesen TL, Simons KJ, Xu SS, Faris JD. Development, identification, and validation of markers for marker-assisted selection against the *Stagonospora nodorum* toxin sensitivity genes *Tsn1* and *Snn2* in wheat. Mol Breeding. 2009; 23: 35–49

56. McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA. Global diversity and distribution of three necrotrophic effectors in *Phaesosphaeria nodorum* and related species. New Phytol. 2013; 199 (1): 241–251 https://doi.org/10.1111/nph.12257 PMID: 23550706

57. Liu Z, Gao Y, Kim YM, Faris JD, Shelver WL, de Wit PJGM, Xu SS, Friesen TL. SnTox1, a *Parastagonospora nodorum* necrotrophic effector, is a dual-function protein that facilitates infection while protecting from wheat-produced chitinases. New Phytol. 2016; 211(3): 1052–1064 https://doi.org/10.1111/nph.13959 PMID: 27041151

58. Oliver RP, Solomon PS. New developments in pathogenicity and virulence of necrotrophs. Curr Opin Plant Biol. 2010; 13(4): 415–419 PMID: 20684067

59. Oliver RP, Friesen TL, Faris JD, Solomon PS. *Stagonospora nodorum*: From Pathology to Genomics and Host Resistance. Ann Rev Phytopathol. 2012; 50: 23–43

60. Barrett LG, Thrall PH, Dodds PN, van der Merwe M, Linde CC, Lawrence GJ, Burdon JJ. Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. Mol. Biol. Evol. 2009; 26(11): 2499–2513 https://doi.org/10.1093/molbev/msp166 PMID: 19633228

61. Brunner PC, McDonald BA. Evolutionary analyses of the avirulence effector AvrStb6 in global populations of *Zymoseptoria tritici* identify candidate amino acids involved in recognition. Mol. Plant Pathol. 2018; 19(8): 1836–1846

62. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc. Natl. Acad. Sci. 2011; 108(7): 2831–2836 https://doi.org/10.1073/pnas.1014971108 PMID: 21282627

63. Rovenich H, Boshoven JC, Thomma BPHJ. Filamentous pathogen effector functions: of pathogens, hosts, and microbiomes. Curr Opin Plant Biol. 2014; 20: 96–103 https://doi.org/10.1016/j.pbi.2014.05.001 PMID: 24879450

64. Schrider DR, Hourmozdi JN, Hahn MW. Pervasive multi-nucleotide mutational events in eukaryotes. Curr. Biol. 2011; 21(12): 1051–1054 https://doi.org/10.1016/j.cub.2011.05.013 PMID: 21636278

65. Venkat A, Hahn MW, Thornton JW. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nat. Ecol. Evol. 2018; 2: 1280–1288 https://doi.org/10.1038/s41559-018-0584-5 PMID: 29967485

66. Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, et al. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 2011; 7(6); e1002070 https://doi.org/10.1371/journal.pgen.1002070 PMID: 21695235

67. Grandaubert J, Dutheil JY, Stukenbrock EH. The genomic determinants of adaptive evolution in a fungal pathogen. Evol. Letters. 2019; 3(3): 299–312

68. Croll D, Lendenmann MH, Steward E, McDonald BA. The Impact of Recombination Hotspots on Genome Evolution of a Fungal Plant Pathogen. Genetics. 2015; 201: 1213–1228 https://doi.org/10.1534/genetics.115.180968 PMID: 26392286

69. Hématy K, Cherk C, Somerville S. Host-pathogen warfare at the plant cell wall. Curr Opin Plant Biol. 2009; 12(4): 406–413 https://doi.org/10.1016/j.pbi.2009.06.007 PMID: 19616468

70. Rowe HC, Kliebenstein DJ. Elevated genetic variation within virulence-associated *Botrytis cinera* polygalacturonase loci. Mol Plant Microbe Inter. 2007; 20(9): 1126–1137

71. Stukenbrock EH, McDonald BA. Population genetics of fungal and Oomycete effectors involved in gene-for-gene interactions. Mol Plant Microbe Inter. 2009; 22(4): 371–380

72. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010;175–176.

73. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15): 2114–2120 https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

74. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;1303.3997

75. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16): 2078–2079 https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

76. R Core Team. 2013. R: A language and environment for statistical computing.

77. Knaus BJ, Grunwald NJ. VCFR: a package to manipulate and visualize variant call format data in R. Mol Ecol Res. 2017; 17(1): 44–53

78. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008; 24(11): 1403–1405 https://doi.org/10.1093/bioinformatics/btn129 PMID: 18397895

79. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23(19): 2633–2635 https://doi.org/10.1093/bioinformatics/btm308 PMID: 17586829

80. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2): 945–959 PMID: 10835412

81. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005; 14(8): 2611–2620 https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

82. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Res. 2012; 4(2): 359–361

83. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G. The variant call format and VCFtools. Bioinformatics. 2011; 27(15): 2156–2158 https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

84. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. PopGenome: An efficient Swiss Army knife for population genomic analyses in R. Mol Biol Evol. 2014; 31(7): 1929–1936 https://doi.org/10.1093/molbev/msu136 PMID: 24739305

85. Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Molecular biology and evolution. 2013; 30(9):2224–34. https://doi.org/10.1093/molbev/mst112 PMID: 23777627

86. Quinlan AR, Hall IA. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6): 841–842 https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

87. Friesen TL, Faris JD. Characterization of plant-fungal interactions involving necrotrophic effector-producing plant pathogens. In Plant Fungal Pathogens: Methods and Protocols, Methods in Molecular Biology. Bolton MD and Thomma BPHJ (eds.). 2012; 835:191–207

88. Liu ZH, Friesen TL, Rasmussen JB, Ali S, Meinhardt SW, Faris JD. Quantitative trait loci analysis and mapping of seedling resistance to *Stagonospora nodorum* leaf blotch in wheat. Phytopathology. 2004; 94(10): 1061–1067 https://doi.org/10.1094/PHYTO.2004.94.10.1061 PMID: 18943794

89. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012; 28(18):2397–9. https://doi.org/10.1093/bioinformatics/bts444 PMID: 22796960

90. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE, Buckler ES. GAPIT version 2: an enhanced integrated tool for genomic association and prediction. The plant genome. 2016; 9(2).

91. Savojardo C, Martelli PL, Fariselli P, Casadio R. DeepSig: deep learning improves signal peptide detection in proteins. Bioinformatics. 2018; 2018: 1–7

92. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 2016; 210(2): 743–761 https://doi.org/10.1111/nph.13794 PMID: 26680733

93. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27(21): 2987–2993 https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

94. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Systems biol. 2011; 7(1): 539

95. De Mita S, Siol M. EggLib: processing, analysis, and simulation tools for population genetics and genomics. BMC genets. 2012; 13(1): 27

96. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome

of *Drosophila melanogaster* strain w1118. Fly. 2012; 6(2): 80–92 https://doi.org/10.4161/fly.19695 PMID: 22728672

97. Ruden DM, Cingolani P, Patel VM, Coon M, Ngueyn T, Land SJ, Lu X. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. Frontiers Genet. 2012; 3: 35

98. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001; 17(9):847–8 https://doi.org/10.1093/bioinformatics/17.9.847 PMID: 11590104

99. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version. 2010; 2(0)

100. Smit AFA and Hubley R. 2008. RepeatModeler Open-1.0. http://www.repeatmasker.org.

101. Smit AFA, Hubley R, and Green P. 2017. RepeatMasker Open-3.0. http://www.repeatmasker.org