

Mating-type Locus Characterization and Variation in

Pyrenophora semeniperda

Julie Leanna Henry

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

Master of Science

Craig E. Coleman, Chair
Susan E. Meyer
Brad D. Geary

Department of Plant & Wildlife Sciences

Brigham Young University

July 2015

Copyright © 2015 Julie Leanna Henry

All Rights Reserved

ABSTRACT

Mating-type Locus Characterization and Variation in *Pyrenophora semeniperda*

Julie Leanna Henry
Department of Plant and Wildlife Sciences, BYU
Master of Science

Pyrenophora semeniperda is a generalist fungal pathogen that occurs primarily on monocot seed hosts. It is in the phylum Ascomycota, which includes both self-compatible (homothallic) and self-incompatible (heterothallic) species. Homothallic fungal species contain complementary mating-type (*MAT*) idiomorphs in a single unikaryotic strain, while heterothallic strains contain a single *MAT* idiomorph requiring interaction between strains of complementary mating-types for sexual reproduction to occur. Because the majority of *P. semeniperda* strains contained either *MAT1* or *MAT2*, this species was provisionally categorized as heterothallic. However, many strains contain both *MAT* idiomorphs and appear to be homothallic. These results warranted a closer look at the *MAT* idiomorphs and the structure of the *P. semeniperda* genome in order to assure accurate characterization of the *MAT* locus. Additionally, an assessment of the geographic distribution of *MAT* idiomorphs provides us with insight into the genetic diversity of *P. semeniperda* and the reproductive strategies that it employs. In this study, we characterized the *P. semeniperda* *MAT* locus and assessed the idiomorph distribution of 514 isolates from 25 *P. semeniperda* populations collected from infected *Bromus tectorum* (cheatgrass) seeds. Additionally, we used simple sequence repeat (SSR) and *MAT* idiomorph length polymorphisms to demonstrate the existence of dikaryotic strains and pseudohomothallism in this fungus. We identified a unique variable number tandem repeat (VNTR) within each idiomorph of the *MAT* locus of *P. semeniperda*. Presence of the VNTR in all *MAT* loci analyzed from strains collected in the Intermountain West suggests ancient proliferation of this repeat. The persistence and effectiveness of *P. semeniperda* strains in the cheatgrass pathosystem depend not only on the density of the fungus in the soil, but also on the genetic heterogeneity of each population. Our study suggests that *P. semeniperda* genetic diversity is increased both through *MAT* locus-dependent sexual reproduction and asexually through anastomosis.

Keywords: *MAT* locus, *MAT* idiomorph, VNTR, SNPs, SSRs, life cycle, *Pyrenophora semeniperda*

ACKNOWLEDGEMENTS

I would first like to express my appreciation for all of the knowledge the genetics professors in BYU's Plant and Wildlife Department have imparted to me and for their support of my academic pursuits and research these last couple of years. I am likewise grateful for the friendships that I have made with fellow scholars in the genetics lab and for their humor, opinions, and brilliant insights. I particularly appreciate the assistance with data generation and analysis given by Sara Greenfield, Justin Page, and Dr. Joshua Udall. I would like to express my appreciation to my graduate advisor Dr. Coleman for giving me a chance to work in his lab as an inexperienced undergrad and later extending the opportunity to continue on as a graduate student. His expertise, encouragement, and knowledge have been invaluable to me as I have worked through my degrees. I would also like to thank my graduate committee members Dr. Susan Meyer and Dr. Bradley Geary for their expertise and guidance during the course of this research and through the challenging writing process. Thanks to Suzette Clement of the US Forest Service Shrub Sciences Laboratory in Provo, UT for providing *P. semeniperda* cultures and Steven Harrison for providing extracted DNA for this research. Special thanks to the University of Utah Huntsman Cancer Institute and the Sequencing Center at Brigham Young University for their invaluable assistance with sequencing and fragment analysis.

TABLE OF CONTENTS

Title Page	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Mating-type locus characterization and variation in <i>Pyrenophora semeniperda</i>	1
<i>Introduction</i>	1
<i>Materials and Methods</i>	4
DNA extractions	4
<i>MAT</i> locus primer development and genotyping	4
<i>P. semeniperda</i> <i>MAT</i> locus sequencing and haplotype phylogeny generation	5
<i>MAT</i> and SSR fragment length analysis	6
Illumina library preparations and contig analysis	8
<i>Results</i>	8
<i>MAT</i> idiomorph discovery and genotyping	8
Characterization of the <i>MAT</i> locus and construction of haplotype phylogenies	9
Calculation of selective pressures acting on the <i>MAT</i> idiomorphs	11
SSR and <i>MAT</i> locus polymorphisms and dikaryotic frequencies	11

<i>Discussion</i>	13
Homology of <i>P. semeniperda</i> <i>MAT</i> locus structure and sequence with other <i>Pyrenophora</i> species	13
Do <i>MAT</i> idiomorph frequencies, structures, and selective pressures infer functional <i>MAT</i> idiomorphs?	14
<i>MAT</i> locus sequence diversity	15
Implications of <i>MAT</i> idiomorph frequencies and ratios in pseudohomothallic <i>P. semeniperda</i> strains	17
The utility of <i>MAT</i> idiomorphs as genetic markers	18
<i>Conclusion</i>	18
<i>References</i>	20
<i>Tables</i>	24
<i>Figures</i>	26

LIST OF TABLES

Table 1	<i>MAT</i> idiomorph frequencies and identified haplotypes per collection site	24
Table 2	Heterothallic, pseudohomothallic, and overall <i>MAT</i> idiomorph frequencies	25

LIST OF FIGURES

Figure 1	<i>MAT</i> idiomorph distribution map	26
Figure 2	<i>MAT</i> locus characterization	27
Figure 3	Percentages of unikaryotic and dikaryotic strains with different <i>MAT</i> idiomorph combinations	28
Figure 4	<i>MAT1</i> haplotype phylogeny	29
Figure 5	<i>MAT2</i> haplotype phylogeny	30

Mating-type Locus Characterization and Variation in

Pyrenophora semeniperda

INTRODUCTION

Pyrenophora semeniperda is an ascomycete seed pathogen that primarily infects grass seeds. It has been isolated from *Bromus tectorum* (cheatgrass) soil seedbank samples collected from throughout the Intermountain West. Integral to understanding of *P. semeniperda* and its potential as a seed pathogen is knowledge of its introduction history, life cycle and reproductive strategies. Examination of the population genetics of the *P. semeniperda* – *B. tectorum* pathosystem revealed that the fungus was likely introduced into its current environment along with its host *B. tectorum* (Boose et al., 2011). Genetic diversity is high among *P. semeniperda* strains: 12 Internal Transcribed Spacer (ITS) haplotypes with an average genetic distance of 1.3% between the types have been identified (Boose et al., 2011) and mycelial growth rate and cytochalasin B production can vary by up to 4 times (Meyer et al., 2015). However, little is known about the possible role of sexual reproduction in the life cycle of *P. semeniperda*, as this fungus is most commonly observed in an asexual (anamorph) reproductive stage. Although seemingly uncommon, the sexual (teleomorph) state has been identified morphologically in both the field (Shoemaker, 1966) and in culture (Paul, 1969), and no correlation was found between strain virulence, ITS genotypes, and habitat of origin (Stewart, 2009). However, when a genome-wide single nucleotide polymorphism (SNP) analysis was run, low linkage disequilibrium was observed in only 18 of 36 Intermountain West populations (Soliai, 2011). The observed linkage disequilibrium was a function of clonal reproduction or inbreeding, so it was expected since the

fungus primarily reproduces asexually. These data suggest that a combination of sexual and asexual reproduction actively occurs in natural *P. semeniperda* populations (Milgroom, 1996; Xu, 2006).

Since no genetically-based studies specifically addressing the sexual life cycle of *P. semeniperda* have been published, we chose to explore the mating-type (*MAT*) locus, which is the primary locus underlying sexual reproduction in ascomycete fungi. The *MAT* locus has been the focus of numerous studies because it is occupied by idiomorphs of a gene known to regulate sexual reproduction in ascomycetes, *MAT1-1* and *MAT1-2* (Metzenberg and Glass, 1990; Coppin et al., 1997; Kronstad and Staben, 1997; Fraser and Heitman, 2004). *MAT1-1* and *MAT1-2* are idiomorphs rather than alleles because, with the possible exception of their respective DNA binding domains, they are unrelated by descent from a common sequence (Metzenberg and Glass, 1990; Turgeon and Yoder, 2000; Martin et al., 2010). The proposed nomenclature for mating-type 1 and mating-type 2 for our purposes is *MAT1-1-1* and *MAT1-2-1* (Turgeon and Yoder, 2000) respectively, but for the remainder of the paper they will be referred to as *MAT1* and *MAT2*. *MAT* idiomorphs produce transcription factors that initiate interaction between complementary mating types through the production of specific pheromones and pheromone receptors (Debuchy et al., 2010; Martin et al., 2010; Tsong et al., 2007) and later regulate internuclear recognition as the sexual cycle progresses (Casselton, 2002; Stanton and Hull, 2007; Turgeon and Debuchy, 2007). Individual idiomorphs are identified by the presence of one of the two DNA-binding regions: the α -box region (*MAT1*) or the high mobility group (HMG) domain (*MAT2*) (Galagan et al., 2005). In general, while the two *MAT* idiomorphs are highly dissimilar, it has been postulated that the α -box domain is likely derived from an ancestral HMG gene (Glass et al., 1990; Debuchy et al., 2010; Martin et al., 2010).

Because *P. semeniperda* is a filamentous ascomycete in the fungal subkingdom Dikarya, strains have the potential to house two genetically-distinct nuclei within their mycelium. Generally, filamentous ascomycetes can become dikaryotic through asexual vegetative hyphal fusion (anastomosis) or sexual structure fusion. Two different nuclear states produce self-fertile mycelium: a homothallic strain contains a single haploid nucleus possessing both *MAT1* and *MAT2*, whereas a pseudohomothallic strain contains two genetically-distinct nuclei, one with *MAT1* and the other with *MAT2*. Heterothallic strains are sexually self-incompatible because each mycelium possesses just one of the two *MAT* idiomorphs (Kronstad and Staben, 1997; Billiard et al., 2011; Coppin et al., 1997; Giraud et al., 2008). The *MAT* locus has been characterized in three other *Pyrenophora* species, all of which are economically important monocot leaf pathogens. *Pyrenophora teres* and *P. graminea* have been identified as heterothallic (Rau et al., 2007). In contrast, *P. tritici-repentis* is homothallic, possessing both *MAT1* and *MAT2* in tandem, separated by approximately 1 kb of DNA sequence, within the *MAT* locus (Lepoint et al., 2010).

Our study of the genetic factors underlying sexual reproduction in *P. semeniperda* is among the first for a seed pathogen. Our primary objectives were to locate and identify the *MAT* idiomorphs in *P. semeniperda*, to characterize these idiomorphs and to determine the distribution of the mating-types within and across Intermountain Western U.S. populations. Furthermore, we evaluated the possible evolutionary significance of *MAT* allelic variation and tested the hypothesis that *P. semeniperda* strains possessing both *MAT* idiomorphs have a dikaryotic pseudohomothallic nuclear composition rather than a true unikaryotic homothallic *MAT* locus structure. The data produced by our *MAT* locus study

provides evidence for active roles for both sexual reproduction and anastomosis in the *P. semeniperda* life cycle.

MATERIALS AND METHODS

DNA Extractions

Pyrenophora semeniperda samples were obtained from *B. tectorum* seeds collected from the seed bank at 25 different sites in CO, ID, NV, UT, and WA (Fig. 1). The process of procuring and growing pure *P. semeniperda* cultures is described in Boose et al. (2011). Much of the DNA used in this research had previously been extracted for that study. DNA for the remaining strains was obtained using the following protocol: the tissue was freeze-dried for 24 hours prior to pulverization in a 2mL screwtop container with 4-6 3mm bashing beads. Following these steps, the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) was used to produce DNA template for genetic analyses.

***MAT* locus primer development and genotyping**

To capture the *MAT* idiomorphs in *P. semeniperda*, species-specific primers were designed using Primer3 in the Geneious software package (Biomatters, San Francisco, CA) from scaffolds containing *MAT1* in isolate CCB6 (approx. 1,500 bp; Soliai et al. 2014) and *MAT2* in isolate STR15 (approx. 1,200 bp; unpublished data). Idiomorph-specific primers were used to amplify *MAT1* and *MAT2* in all 514 strains using GoTaq® HotStart Master Mix (Promega, Madison, WI) and amplification products were visualized by agarose gel electrophoresis. The thermocycler parameters were set according to the manufacturer's recommendations.

***P. semeniperda* MAT locus sequencing and haplotype phylogeny generation**

Following genotyping, *MAT* loci were sequenced from 76 different *P. semeniperda* strains using three sets of primers for each idiomorph designed using Primer3. Primers used for *MAT1* amplification were (forward / reverse): CCCACTCTCTTCTCACTTTC / GTAACGCTAGTAGGGCATAG; TTCGTGACCAACTCACCAAG / TGAATCCAGTTCCCGCATTG; ATGATCTCATTGCCGCTGTC / ATAGGCCTGCAGGTACAACCTC. Primers used for *MAT2* amplification were (forward / reverse): CTTCTGCTCAATTTCTCACTAC / CTATCGGTCGATGAGGAAAC; CTATTGAGGCTGCTCCAC / GGCAAGCATACCAACTGAAG; CAAGTATAGCCCTAGGAAGC / TTCATGGGACGTGGAGTTC. Gene fragments were amplified using GoTaq® HotStart Master Mix (Promega, Madison, WI) and reactions were cleaned up using a QIAquick PCR Purification kit (Qiagen, Valencia, CA) before submitting the amplification products to the Brigham Young University DNA Sequencing Center (Provo, UT). The identity of the mating-type idiomorphs was confirmed searching the GeneBank non-redundant database using BLAST.

The haplotype consensus sequences of each *MAT* idiomorph generated from the 76 *MAT* sequence contigs were then aligned using the Geneious software package (Geneious Consensus Align: Cost Matrix= 65% similarity, Gap open penalty= 12, Gap extension penalty= 3, Global alignment with free end gaps) in order to calculate the variation between them as a result of single nucleotide polymorphisms (SNPs), gaps, and the unique variable number repeats (VNTR) region present in both idiomorphs. Each 18 bp repeat within the VNTR region was calculated as a single polymorphism and all SNPs and gaps contributed equally to the calculations. One-thousand bootstraps for 10% of the data for each bootstrap were calculated and a neighbor-joining tree of each of the 1000 trees was

generated. A bootstrap consensus tree was then created by converging the individual bootstrap trees. The same approach was used to generate the genetic distance phylogenies (Figure 4).

MAT1 and *MAT2* haplotype phylogenies were generated using the PHYLIP package (Felsenstein, 1989).

Haplotype sequence alignments were also used to determine potential evolutionary forces that are acting on the *MAT1* and *MAT2* idiomorphs. For these calculations, the $\omega = K_a/K_s$ system was applied in which K_a is the number of nonsynonymous substitutions per nonsynonymous site and K_s is the number of synonymous substitutions per synonymous site (Hurst, 2002; Yang and Bielawski, 2000). Substitutions were calculated by comparing the different haplotypes against the most common haplotype of each *MAT* idiomorph (*MAT1* haplotype 1-2 and *MAT2* haplotype 2-6). SNPs were calculated from nucleotide sequence and K_a and K_s values were determined following translation of the mating-type nucleotide sequences into amino acids.

Following translation of the *MAT* haplotypes into amino acid sequence, the accuracy of the resulting putative peptides was confirmed by comparison with *P. teres* (accession AAY35015 *MAT1* and *MAT1* ADM21346.1), *P. graminea* (accession *MAT1* ABI37047.1 and *MAT2* ABI37066), and *P. tritici-repentis* (accession *MAT1* CAP08779.1 and *MAT2* CAP08744) full-length *MAT1* and *MAT2* proteins.

***MAT* and SSR fragment length analysis**

Seven loci were selected to detect the frequencies of dikaryotic and unikaryotic strains within 128 isolates from across 25 populations in CO, ID, NV, UT, and WA. Of the SSR loci that had been previously identified (Meyer et al., 2008), 5 highly polymorphic sites with trinucleotide repeats were selected and primers were designed to flank them. Additionally,

fluorescent *MAT1* and *MAT2*-specific primers were designed to bind to conserved sequence that flank the VNTR regions present in both mating-types that was identified by aligning 40+ sequences for each of the idiomorphs. The primers used for *MAT1* detection were (forward / reverse): GGTGACTTCGCTTTGAACGTTGTCGCT / GATTCACATCCGCGGATCTCACTCCCT; and the primers used for *MAT2* detection were (forward / reverse): AAGATGAGCCAGCAGTTCATGGGACGT / GACTCTGCCCTCCTTGAGTGCGTTCAT.

The names of the 5 polymorphic SSR loci that had been previously identified for *P. semeniperda* were: BFOFF9, BFOFF19, BFOFF25, BFOFF37, and BFOFF41 (Meyer et al., 2015). The process of SSR and *MAT*-typing included multiplex PCR reactions with fluorescent-labeled primers using the Type-it Microsatellite PCR Kit protocol (Qiagen Valencia, CA) followed by capillary fragment analysis and peak calling using the Geneious software package (Biomatters, San Francisco, CA). The primers used for SSR amplification were:

TATTGTTGAGGGATTGGTTGTG / TCCGCAGTTGGAATAAACGTCAAAAC (BFOFF9);
ACGTCTCTCCCACATGGCTCCCC / GGCTTAGAGAACCCTGTCCGCCATGG
(BFOFF19); CCGATCAGTCGTGCTCCTTCTAC /
GCAGACTCCTTGCATGAGAATCGTCCA (BFOFF25); CATGCCAAAGGGTGTCTTGGG
/ GCAACAAGCAAAAGAGCAGCGAGAAC (BFOFF37);
CCAGCCACCACGTCTGCCCTGA / GCCAGTCTGCCAGTCTCGGTCTTGG (BFOFF41).

The thermocycler settings were set at the following: initial heat activation 5 minutes at 95C, (22 cycles) denature for 30 seconds at 95C / anneal for 90 seconds at 60C / extend for 30 seconds at 72C, with a final extension for 30 seconds at 60C. Samples were diluted 1:50 and submitted to the Brigham Young University Sequencing Center for capillary fragment analysis.

Illumina library preparations and contig analysis

Single-read DNA libraries were prepared for *P. semeniperda* isolates CCB06 and STR15 using the instructions included in the Illumina TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA). Libraries were submitted to the Huntsman Cancer Institute (Salt Lake City, UT) for HiSeq sequencing. Reads generated from both isolates were used to address the question of homothallic vs. pseudohomothallic: STR15 because it consistently tested positive for both *MAT1* and *MAT2* and CCB06 because it tested positive exclusively for *MAT1*. Rau et al. (2007) reported that the 6.5 kb 5' region immediately upstream of the *MAT* locus is a highly conserved open reading frame (ORF) in *Pyrenophora*, so 30bp k-mers were generated from this region as well as from the *MAT* idiomorphs following Illumina sequencing. K-mers are subsequences of length k that are generated from contiguous nucleotide sequence (Compeau et al., 2011). For this application, k-mers were used as a way to quantify the number of reads generated from each region of interest as a broad way to determine the frequency with which each region is present within the sequenced strain. The k-mers were aligned and quantified against the STR15 and CCB6 Illumina contigs to compare k-mer frequencies of the conserved 5' region against the individual frequencies of both *MAT1* and *MAT2*.

RESULTS

***MAT* idiomorph discovery and genotyping**

Both *MAT1* and *MAT2* idiomorphs were identified in *P. semeniperda* by performing a BLAST search with *MAT* idiomorph sequence from related species *P. teres* (Rau et al. 2007; Lu et al. 2010) against strain CCB06 (Soliai et al. 2014) and strain STR15

(unpublished data) scaffolds. With this information we were able to design *P. semeniperda*-specific primers and successfully genotype 514 strains from 25 populations across the Intermountain West.

Based on data collected from PCR amplification of both idiomorphs, we observed the following mating-type frequencies: 34% contained only *MAT1*, 28% contained only *MAT2*, and 38% contained both mating-types (Fig. 1). The overall idiomorph frequency across all tested populations of *MAT1* was 53% and of *MAT2* was 47%. Distribution of the mating-type idiomorphs is relatively uniform across ID, UT, and WA: each genotype is typically represented in approximately 25-50% of the tested strains. However, in 3 out of 4 populations in NV (PVM, BFL, SWR) we found that the frequency of *MAT2*-only isolates is very low (0-8%) with the populations consisting of an approximate ratio of 2:1 of *MAT1* to *MAT2* idiomorphs.

Characterization of the *MAT* locus and construction of haplotype phylogenies

Once we had determined the *MAT* locus idiomorph distribution among and within populations, our next step was to take a closer look at the structure of the gene itself. Sequencing and constructing the *MAT* locus from multiple strains allows for a closer look at the genetic variation and evolution of the *P. semeniperda* mating-type idiomorphs, so we sequenced and assembled the *MAT* idiomorphs from 76 strains. The 76 strains that we chose represented 22 populations across the region and all three *MAT* locus genotypes (*MAT1*, *MAT2*, and *MAT1/MAT2*) identified from the genotyping data.

The stated base pair positions for the full-length *MAT1* and *MAT2* idiomorph sequences are relative to the start codon of each. We sequenced *MAT1* from 43 strains: the putative *MAT1* intron is 53 bp long and is located at

base position 250-302, interrupting the DNA sequence that codes for the α -box which is located from bases 63-639. We identified an 18 bp variable number tandem repeat (VNTR) unique to the *MAT1* idiomorph that occurs from 3 to 10 times without disrupting the reading frame. The genomic (gDNA) *MAT1* sequence length varies from 1306-1430 base pairs depending on the number of 18 bp repeats present (Fig. 2A). We also sequenced *MAT2* from 40 strains: the putative *MAT2* intron is 55 bp long and is located at base position 553-608, interrupting the sequence that codes for the HMG box which is located from bases 421-651. We identified an 18 bp VNTR unique to the *MAT2* idiomorph that occurs from 3 to 5 times without disrupting the reading frame. The gDNA *MAT2* sequence length varies from 1120 - 1156 bp depending on the number of 18 bp repeats present (Fig. 2B).

The most similar homologs to *P. semeniperda* *MAT1* and *MAT2* were found in *P. teres* with pairwise identity matches of 62% and 68%, respectively (Rau et al. 2007; Lu et al. 2010). Each *MAT* idiomorph was found to possess a different 18 bp repeat, which are unique to *P. semeniperda* based on BLASTn searches against the NCBI nucleotide database. We found that these 18 bp VNTRs are a defining source of variation within the mating-type idiomorphs. Nine *MAT1* and 8 *MAT2* haplotypes were identified based on VNTR differences and SNPs. The 3 most-represented *MAT1* haplotypes are 1-1 (8), 1-2 (19), and 1-3 (8) and the 3 most-represented *MAT2* haplotypes are 2-1 (7), 2-2 (7), and 2-6 (18). The among-population distribution of the different haplotypes for both *MAT1* and *MAT2* are found in Table 1.

We constructed haplotype-based phylogenies using haplotypes groups generated based on the putative CDS sequences for both *MAT1* and *MAT2*. We created these phylogenies in order to determine the bootstrap values underlying the consensus neighbor-joining phylogenies for each idiomorph.

Calculation of selective pressures acting on the *MAT* idiomorphs

The ratio of non-synonymous (K_a) and synonymous (K_s) amino acid changes ($\omega = K_a/K_s$) was then explored because we had found sufficient variation within 76 *P. semeniperda* strains to identify 9 *MAT1* haplotypes and 8 *MAT2* haplotypes. The most common haplotype of each idiomorph was used as the standard. Amino acid substitution ratios within each idiomorph indicate what kind of selective pressure is operating on them: diversifying selection ($\omega > 1$), stabilizing selection ($\omega < 1$), or neutral selection ($\omega = 1$). The VNTR region was excluded from the calculations because it would significantly inflate the ω values and there is uncertainty regarding any function it may have. The VNTR region within each idiomorph of the *MAT* locus of *P. semeniperda* was identified when the idiomorphs were sequenced. Selective pressures were anticipated to be variable between different sites (Zaffarano et al. 2011), so the combined ω value of each idiomorph was compared with the ω value of the DNA-binding domain found in each: the *MAT1* α -box and the *MAT2* HMG domain. Overall, $\omega = 1.03$ for *MAT1* translated exon sequence and, more specifically, $\omega = 0.31$ within the α -box and $\omega = 1.63$ outside of the α -box. Overall, $\omega = 1.24$ for *MAT2* translated exon sequence and, more specifically, $\omega = 1$ within the HMG domain and outside of the HMG domain $\omega = 1.29$.

SSR and *MAT* locus polymorphisms and dikaryotic frequencies

As a result of *MAT* genotyping, we discovered that both idiomorphs can coexist within a single strain. This lead us to address the question of whether individual nuclei are homothallic (possessing both mating-types) or heterothallic (possessing either *MAT1* or *MAT2*), in which case the strain containing both mating-types would be a pseudohomothallic dikaryote. Fragment analysis of amplified simple sequence repeat (SSR) loci and the VNTR regions found in each

mating-type were chosen to answer this question. This method enabled us to clearly distinguish between *MAT* idiomorphs that contain different numbers of VNTRs, making it possible to detect the presence of multiple *MAT* idiomorphs in strains that only possess one mating-type. Since these organisms are haploid, polymorphisms detected at any of the 5 SSR loci or possession of two *MAT1* or two *MAT2* idiomorphs within one strain, indicate that it is dikaryotic: polymorphisms were detected in 100% of the pseudohomothallic isolates, 19% of the *MAT1*-only isolates, and 35% of the *MAT2*-only isolates. Overall, 41% of the 128 strains tested were dikaryotic and the mating-type frequencies among these dikaryotes were 45% *MAT1* and 55% *MAT2* (Fig. 4). Our estimate of the frequency of dikaryotic strains is conservative due to the low number of SSR loci tested and the inability to distinguish between *MAT1* or *MAT2* haplotypes that possess identical VNTR lengths, but are unique from one another due to SNPs.

Based on the discovery that strains can possess either *MAT1*, *MAT2*, or both idiomorphs, we hypothesized that the idiomorphs were likely located on separate molecules and that *P. semeniperda* strains exist in a heterothallic (*MAT1* or *MAT2*) state unless anastomosis produces a dikaryotic organism that possesses multiple *MAT* loci, each located within separate haploid nuclei (pseudohomothallic). We used whole genome shotgun Illumina data generated from strains STR15 (contains both mating-types) and CCB6 (contains only *MAT1*) to test the hypothesis that strains containing both mating types are pseudohomothallic dikaryotes. Our approach involved production of Kmers, which in this case were 31 bp long and represent all the possible subsequences from a read of length K generated from 6.5 kb of highly conserved sequence immediately upstream of the *MAT* locus. We found that these k-mers are present at roughly double the frequency of either the *MAT1* or *MAT2* k-mers in STR15 and are present at the same frequency as *MAT1* in CCB6.

DISCUSSION

Our research is the first to map the distribution and frequency of the two *P. semeniperda* *MAT* idiomorphs and to characterize the *MAT* locus. Genotyping has shed light on the distribution and frequencies of the two mating-types and fragment analysis has revealed the propensity of the fungus to undergo anastomosis to form dikaryotic strains. Additionally, *MAT1* or *MAT2* were sequenced from 76 strains which enabled a sequence-level analysis of idiomorph variation and distribution across the Intermountain West. *MAT* locus sequencing also provided valuable insight into the potential selective pressures acting on the mating-type idiomorphs and their DNA-binding domains. Lastly, we explore the use of this locus as a possible genetic marker to better understand the potential for sexual reproduction in *P. semeniperda* populations based on previous success with other ascomycete genera (Turgeon et al. 1998).

Homology of *P. semeniperda* *MAT* locus structure and sequence with other *Pyrenophora* species

The structure of the *MAT* idiomorphs are highly conserved both in heterothallic *Pyrenophora* species *P. semeniperda*, *P. graminea*, and *P. teres*, as well as in homothallic *P. tritici-repentis*. The opposite mating-type genomes of the three heterothallic species (*P. semeniperda*, *P. graminea*, and *P. teres*) exclusively contain either *MAT1* or *MAT2* (Unpublished; Lu et al., 2010; Rau et al., 2007 respectively) and the homothallic species (*P. tritici-repentis*) contains both *MAT1* and *MAT2* oriented head-to-tail in tandem (Lepoint et al., 2010). Illumina k-mer frequency data generated from strain STR15, as well as the occurrence of SSR and *MAT* locus polymorphisms within single strains support our conclusion that *P.*

semeniperda is heterothallic. Genomic sequences of these organisms show that, with the exception of *MAT2* in *P. tritici-repentis* since it is in tandem with and immediately downstream of *MAT1*, a conserved ORF is located immediately upstream of both of the mating-type idiomorphs. The function of this 5' ORF has not yet been identified. All four of the aforementioned *Pyrenophora* species contain the conserved α -box region in *MAT1* and the HMG domain in *MAT2*, both of which are interrupted by an approximately 53 bp and 55 bp intron respectively. The VNTR region is unique to *P. semeniperda*, making it a distinguishing feature of this organism's *MAT* locus structure across all fungi.

Do *MAT* idiomorph frequencies, structures, and selective pressures infer functional *MAT* idiomorphs in *P. semeniperda*?

At this point, *MAT* locus gene expression data has not been generated for *P. semeniperda*, so we have no transcriptome-based molecular evidence for *MAT* locus functionality. Therefore, we used the data generated from *MAT* genotyping and sequencing to address the question of functionality. When idiomorph frequencies for all populations were combined the *MAT1*:*MAT2* ratio was approximately 1:1, which indicates that it is evolutionarily advantageous to equally preserve both mating-types (Pielou, 1977). This is because a 1:1 ratio of complementary mating-types facilitates sexual interaction and random mating between unikaryotic heterothallic strains (Taylor et al., 1999; Xu, 2006). It is possible that the three NV populations in which one mating-type dominates may be the result of the founder effect, but additional research would be required for a supported conclusion. Additionally, the very presence of unikaryotic heterothallic strains (59% of all isolates tested) indicates sexual reproduction since ascospores (sexual spores produced following karyogamy and re-segregation of genetically unique nuclei) are unikaryotic

(Coppin et al., 1997). If there were no active method of re-segregating nuclei, all strains would eventually become dikaryotic.

The hypothesis of functioning *MAT* idiomorphs is also supported by the absence of premature stop codons in all of the haplotypes of both idiomorphs because it implies that selection has only favored mutations that maintain the production a full-length protein. However, analysis of the selective pressures operating on the *MAT* idiomorphs yielded somewhat inconclusive results. We calculated the selection pressures because they can potentially provide insight into the preservation of a gene's function, particularly within the conserved α -box and HMG DNA binding domains, as loss of function in these domains would render the *MAT* transcription factor unable to bind target DNA sequence. We found that the *MAT1* α -box has been subjected to strong stabilizing selective pressure, while the *MAT2* HMG domain has experienced neutral selective pressure. Both stabilizing and neutral selective pressures can support the retention of gene function, particularly stabilizing selection. However, we determined that diversifying selection has acted on the non-DNA-binding exon regions of both idiomorphs. Since diversifying selection is the result of retained mutations, it is normally detrimental to the native function of the gene. All in all, we do not have sufficient evidence to support the functionality of the *MAT* locus in *P. semeniperda*.

***MAT* locus sequence diversity**

Sequence diversity within the *MAT* locus is found in both *MAT1* and *MAT2* as a result of the accumulation of SNPs and proliferation of VNTR sequence. As a result of SNPs and VNTR variation, 9 *MAT1* and 8 *MAT2* haplotypes were identified which is comparable to the high level of diversity found in *P. semeniperda* ITS sequence (Boose et al., 2011). The haplotype

phylogenies illustrate the diversity that is present within each *MAT* locus idiomorph due to the proliferation of the VNTR region and the presence of SNPs (Fig. 4). Upon comparison of the number of *MAT1* and *MAT2* haplotypes identified in *P. semeniperda* with a closely-related species (*P. teres*; Rau et al., 2007), it seems somewhat common for polymorphisms to remain in the *MAT* locus. Interestingly, polymorphisms are frequently retained in the first and second codon positions of non-DNA binding domains of the CDS which often results in a non-synonymous amino acid substitution. This allowance for amino acid substitutions has resulted in diversifying selection to act on the mating-type idiomorphs in *P. semeniperda* with as yet unknown implications for the function or sexual fitness of the strains possessing them.

A high level of sequence length variation, both nucleotide and amino acid, is introduced into *MAT1* and *MAT2* by the different VNTR sequences within both idiomorphs. These VNTR sequences have not been reported in the *MAT* idiomorphs of other ascomycete genera, making them unique to *P. semeniperda*. The 18 bp VNTR region within *MAT1* ranges from 54-162 nucleotides long, resulting in an additional 18-60 amino acids in the resulting protein. The length impact of the 18 bp VNTR region within *MAT2* is slightly less since it generally contains fewer repeats than *MAT1*: the additional 18-90 nucleotides add 18-30 amino acids to the protein. Since the VNTR sequences are unique to this fungus, we are unable to determine any possible functional impact the additional amino acids may have or potential role that they play in the *MAT* idiomorph proteins. Answers to these questions would become clearer with future *P. semeniperda* *MAT* locus functional studies involving gene expression analysis and a comparison of haplotypes possessing different numbers of VNTRs to identify any possible differences caused by the added amino acids.

Implications of *MAT* idiomorph frequencies and ratios in pseudohomothallic *P. semeniperda* strains

Of the 128 *P. semeniperda* isolates that underwent *MAT* locus and SSR fragment analysis, 41% were identified as dikaryotic. Within the dikaryotic strains, the frequency of each *MAT* combination was: 9% *MAT1/MAT1*, 12% *MAT2/MAT2*, and 20%, *MAT1/MAT2*. Since *P. semeniperda* is an ascomycete, a likely explanation for dikaryotic strains that contain two of the same idiomorph is asexual hyphal fusion, or anastomosis, between compatible strains.

Anastomosis normally only occurs within the mycelium of a single fungal organism in order to facilitate nutrient transfer to the exclusion of genetically dissimilar mycelium. The putative genes underlying anastomosis have not yet been identified in *Pyrenophora* species, but they may provide insight into the asexual mechanism that may increase genetic diversity in *P. semeniperda* populations. Of the dikaryotic strains, 49% contained one each of the *MAT* idiomorphs (pseudohomothallic) and the remaining 51% contained two *MAT* idiomorphs of the same mating-type (Figure 3). Dikaryotic strains possessing any of three possible idiomorph combinations (*MAT1/MAT1* : *MAT1/MAT2* : *MAT2-1/MAT2*) occur in an approximately 1:2:1 ratio, which is predicted from the frequencies of the two idiomorphs (Table 2) and is supported by a goodness of fit test using chi-square ($\chi^2 = 0.0007 < 3.841$; Hardy 1908). Since dikaryotic strains are in Hardy-Weinberg equilibrium, these data support that anastomosis operates independently of the *MAT* locus. Genetic diversity is likely increased in *P. semeniperda* populations through anastomosis between genetically unique strains under the direction of currently undetermined genes.

The utility of *MAT* idiomorphs as genetic markers

The mating-type idiomorphs can be valuable genetic markers despite the complications that dikaryotic mycelium can present, particularly when they contain two alleles of either *MAT1* or *MAT2*. This is due to the presence of a VNTR region in both *MAT* idiomorphs that, when primers are designed to flank the VNTR, can reveal the presence of two alleles of the same idiomorph. The fragment length diversity of both *MAT1* and *MAT2* that is introduced by the VNTR region provides insight into the genetic diversity and the prevalence of dikaryotic strains present in *P. semeniperda* populations. This contrasts with genotype data generated through amplification of full-length idiomorphs, which has reliable idiomorphic but not allelic resolution. However, both detection methods have utility for population-level analysis of mating-type distribution which has implications for the reproductive potentials of *P. semeniperda* populations. This is because *P. semeniperda* is self-incompatible and populations that disproportionately contain one mating-type over the other may exclusively reproduce asexually due to the absence of complementary strains in close proximity.

CONCLUSION

Our study coupled *MAT* locus identification and characterization with SSR fragment analysis in order to gain a more comprehensive understanding of the sexual life cycle and population structure of *P. semeniperda*. While evidence indicates that *P. semeniperda* reproduces sexually, *MAT* idiomorph expression data would provide valuable insight into the question of whether or not the idiomorphs are functional and would be a worthwhile focus of future research. Our data have indicated that asexually-generated dikaryotic mycelium are a common stage in the life cycle of this fungus. Its propensity to undergo anastomosis may provide a way for

researchers to produce unique strains with desirable traits asexually instead of inducing two complementary mating-type strains to reproduce sexually.

Although our data indicate that *P. semeniperda* has a functional *MAT* locus, additional research into the actual role that sexual reproduction plays in the life cycle of this fungus is necessary. It is important to have a working knowledge of the environmental conditions and compatible strain accessibility that cause the fungus to initiate sexual reproduction. Furthermore, when the sexual cycle concludes with the release of recombinant haploid spores, how do these unique recombinant strains contribute to the pathogenic and adaptive success of this organism? These additional *MAT* locus data would contribute to a better understanding of the vegetative and sexual life cycle of *P. semeniperda* and to future studies of the genetic diversity and evolutionary history of this organism.

REFERENCES

- Boose, D., Harrison, S., Clement, S., Meyer, S. 2011. Population genetic structure of the seed pathogen *Pyrenophora semeniperda* on *Bromus tectorum* in western North America. *Mycologia* 103, 85-93.
- Billiard, S., López -Villavicencio, M., Devier, B., Hood, M.E., Fairhead, C., Giraud, T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol. Rev.* 86, 1–6.
- Casselman, L.A. 2002. Mate recognition in fungi. *Heredity* 88, 142–147.
- Compeau, P., Peyzner, P., Teslar, G. 2011. How to apply De Bruijn graphs to genome assembly. *Nat. Biotech.* 29, 987-991.
- Coppin, E., Debuchy, R., Arnaise, S., Picard, M. 1997. Mating types and sexual development in filamentous *Ascomycetes*. *Microbiol. Mol. Biol. Rev.* 61, 411–428.
- Debuchy, R., Berteaux-Lecellier, V., Silar P. 2010. Mating systems and sexual morphogenesis in *Ascomycetes*. *Cellular Molec. Biol. Fila. Fungi W.A. D.C.: ASM Press*, 501-535.
- Felsenstein, J. 1989. PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164-166.
- Fraser, J. A. Heitman, J. 2004. Evolution of fungal sex chromosomes. *Mol. Microbiol.* 51, 299-306.
- Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., et al. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438, 1105-1115.

- Giraud, T., Yockteng, R., López -Villavicencio, M., Refrégier, G., Hood, M.E. 2008. Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot. Cell* 7, 765–775.
- Glass, N.L., Grotelueschen, J., Metzenberg, R.L. 1990. *Neurospora crassa* A mating-type region. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4912-4916.
- Hardy, G.H. 1908. Mendelian proportions in a mixed population. *Science* 28: 49-50.
- Hurst, L. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18, 486-489.
- Kronstad, J.W., Staben, C. 1997. Mating type in filamentous fungi. *Ann. Rev. Genet.* 31, 213-244.
- Lepoint, P., Renard, M.E., Legreve, A., Duveiller, E., Maraite, H. 2010. Genetic diversity of the mating type and toxin production genes in *Pyrenophora tritici-repentis*. *Phytopathology* 100, 474-483.
- Lu, S., Platz, G.J., Edwards, M.C., Friesen, T.L. 2010. Mating type locus-specific polymerase chain reaction markers for differentiation of *Pyrenophora teres f. teres* and *P. teres f. maculate*, the causal agents of barley net blotch. *Phytopathology* 100, 1298-1306.
- Martin, T., Lu, S.W., Tilbeurgh, H., Ripoll, D.R., Dixelius, C., Turgeon, B.G., Debuchy, R. 2010. Tracing the origin of the fungal $\alpha 1$ domain places its ancestor in the HMG-box superfamily: implication for fungal mating-type evolution. *PLoS ONE* 12, e15199.
- Metzenberg, R. L., Glass, N. L. 1990. Mating type and mating strategies in *Neurospora*. *Bioessays* 12, 53-59.

- Meyer, S.E., Beckstead, J., Allen, P.S., Smith, D.C. 2008. A seed bank pathogen causes seedborne disease: *Pyrenophora semeniperda* on undispersed grass seeds in western North America. *Can. J. Plant Pathol.* 30, 525-533.
- Meyer, S.E., Masi, M., Clement, S., Davis, T.L., Beckstead, J. 2015. Mycelial growth rate and toxin production in the seed pathogen *Pyrenophora semeniperda*: resource trade-offs and temporally varying selection. *Plant Path.* DOI: 10.1111/ppa. 12377.
- Milgroom, M.G. 1996. Recombination and the multilocus structure of fungal populations. *Annu. Rev. Phytopathol.* 34, 457-477.
- Paul, A.R. 1969. The production of *Pyrenophora semeniperda* in culture. *Trans. British Mycol. Soc.* 52, 373-379.
- Rau, D., Attene, G., Brown, A.H.D., Nanni, L., Maier, F.J., Baftnas, V., Saba, E., Schafer, W., Papa, R. 2007. Phylogeny and evolution of mating-type genes from *Pyrenophora teres*, the causal agent of barley “net blotch” disease. *Curr. Genet.* 51, 377-392.
- Shoemaker, R.A. 1966. A pleomorphic parasite of cereal seeds, *Pyrenophora semeniperda*. *Canadian Jour. Bot.* 44, 1451-1456.
- Soliai, M.M., Meyer, S.E., Udall, J.A., Elzinga, D.E., Hermansen, R.A., Bodily, P.M., Hart, A.A., Coleman, C.E. 2014. De novo genome assembly of the fungal plant pathogen *Pyrenophora semeniperda*. *PLoS ONE.* 9, e87045
- Stanton, B.C., Hull, C.M. 2007. Mating-type locus control of cell identity. In: Heitman, J., Kronstad, J.W., Taylor, J.W., Casselton, L.A. (Eds.), *Sex in Fungi. Mol. Determination Evol. Implications.* ASM Press, WA, DC, pp. 35-57.
- Stewart, T.E. 2009. The grass seed pathogen *Pyrenophora semeniperda* as a biocontrol agent for annual brome grasses. All Theses and Dissertations. Paper 2267.

- Taylor, J.W., Jacobson, D.J., Fisher, M.C. 1999. The evolution of asexual fungi: reproduction, speciation and classification. *Annu. Rev. Phytopathol.* 37, 197-246.
- Tsong, A.E., Brian, B.T., Johnson, A.D. 2007. Rewiring transcriptional circuitry: mating-type regulation in *Saccharomyces cerevisiae* and *Candida albicans* as a model for evolution. *Sex Fungi* W.A. D.C.: ASM Press, 75-89.
- Turgeon, B.G. 1998. Application of mating type gene technology to problems in fungal biology. *Annu. Rev. Phytopathol.* 36, 115-37.
- Turgeon, G.B., Debuchy, R. 2007. *Cochliobolus* and *Podospora*: mechanisms of sex determination and the evolution of reproductive lifestyle. In: Heitman, J., Kronstad, J.W., Taylor, J.W., Casselton, L.A. (Eds.), *Sex in Fungi. Mol. Determination Evol. Implications*. ASM Press, WA, DC, pp. 35-57.
- Turgeon, B. G., Yoder, O. C. 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet. Biol.* 31, 1-5
- Xu, J. 2006. Fundamentals of fungal molecular population genetic analyses. *Curr. Issues Mol. Biol.* 8, 75-89.
- Yang, Z., Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15, 496-503.
- Zaffarano, P.L., Valentin, Q., Duò, A., Grünig, C.R. 2011. Sex in the PAC: A hidden affair in dark septate endophytes. *BMC Evol. Biol.* 11, 282.

TABLES

Table 1 *MAT* idiomorph frequencies and identified haplotypes per collection site

Table 1

Name	Collection Site ¹	Number of strains genotyped	<i>MAT</i> idiomorph frequencies ²		<i>MAT</i> haplotypes identified ³	
			<i>MAT1-1-1</i>	<i>MAT1-2-1</i>	<i>MAT1-1-1</i>	<i>MAT1-2-1</i>
DIN	Dinosaur, CO	25	0.64	0.36	1-2 (2)	2-3, 2-6
CCB	Cindercone Butte, ID	25	0.48	0.52	1-2, 1-3	2-6
CCF	Cindercone Flat, ID	19	0.50	0.50	1-3, 1-8	2-6
BFL	Bedell Flat, NV	24	0.67	0.33	1-1, 1-2	2-1
LSC	Lower Smoke Cr., NV	22	0.50	0.50	1-6	2-6 (2)
PVM	Peavine Mtn., NV	18	0.70	0.30		
SWR	Stillwater, NV	17	0.61	0.39		
CKT	Cricket Mtns., UT	17	0.45	0.55	1-1	2-1
CNF	Confusion Range, UT	16	0.44	0.56		
DOG	Dog Valley, UT	20	0.58	0.42	1-2 (2)	2-2 (2)
DUJ	Dutch John, UT	17	0.44	0.56		
GUS	Gusher, UT	16	0.38	0.62	1-2	2-2
HRN	House Range, UT	19	0.26	0.74	1-1	2-1
MIL	Milk Ranch, UT	18	0.28	0.72		
SQC	Santaquin Canyon, UT	23	0.63	0.37	1-1, 1-3	2-6 (2)
STR	Strawberry, UT	19	0.53	0.47	1-2 (3)	2-3, 2-6
TMC	Ten Mile Cr., UT	22	0.66	0.34	1-2 (2), 1-4	2-2, 2-6 (2)
WHV	White's Valley, UT	22	0.53	0.47	1-2, 1-3, 1-4	2-3, 2-6
WRK	Whiterocks, UT	26	0.57	0.43	1-1, 1-2 (2)	2-1, 2-6
DRL	Dr. Lefcourt, WA	27	0.56	0.44		2-5, 2-6, 2-7
FSH	Fishtrap, WA	25	0.56	0.44	1-3	2-2, 2-6
KAH	Kahlotus, WA	15	0.43	0.57	1-2, 1-3, 1-7	2-1, 2-4, 2-6
MAR	Marcellus, WA	20	0.55	0.45	1-2, 1-8	2-3, 2-6
PKC	Packer Cr., WA	18	0.56	0.44	1-3, 1-5, 1-9	2-6, 2-8
SDM	Saddle Mtn., WA	15	0.30	0.70	1-2, 1-3	2-2, 2-6

¹ Collection sites and acronym designations are listed for the individual populations used for genotyping and sequencing.

² *MAT* idiomorph frequencies were calculated for a population-level analysis, combining unikaryotic and dikaryotic frequencies.

³ The number of strains within a population that contain the indicated haplotype is 1 unless a different value is shown in parentheses. *MAT1* Genbank haplotype accession numbers are KP718750-KP718758 and *MAT2* haplotype accession numbers are KP718741-KP718748.

Table 2 Heterothallic, pseudohomothallic, and overall *MAT* idiomorph frequencies

Table 2

Population ¹	Strain Genotype Frequencies ²			Idiomorph Frequencies	
	<i>MAT1</i>	<i>MAT2</i>	<i>MAT1</i> / <i>MAT2</i>	<i>MAT1</i>	<i>MAT2</i>
BFL	0.42	0.08	0.50	0.67	0.33
CCB	0.24	0.28	0.48	0.48	0.52
CCF	0.26	0.26	0.48	0.50	0.50
CNF	0.31	0.44	0.25	0.44	0.56
CKT	0.12	0.23	0.65	0.45	0.55
DIN	0.56	0.28	0.16	0.64	0.36
DOG	0.30	0.15	0.55	0.58	0.42
DRL	0.48	0.37	0.15	0.56	0.44
DUJ	0.35	0.47	0.18	0.44	0.56
FSH	0.28	0.16	0.56	0.56	0.44
GUS	0.31	0.56	0.13	0.38	0.62
HRN	0.10	0.58	0.32	0.26	0.74
INV	0.50	0.50	0.00	0.50	0.50
KAH	0.33	0.47	0.20	0.43	0.57
LSC	0.23	0.23	0.54	0.50	0.50
MAR	0.35	0.25	0.40	0.55	0.45
MIL	0.06	0.50	0.44	0.28	0.72
MLV	1.00	0.00	0.00	1.00	0.00
PKC	0.28	0.17	0.55	0.56	0.44
PVM	0.39	0.00	0.61	0.70	0.30
SDM	0.13	0.54	0.33	0.30	0.70
SQC	0.52	0.26	0.22	0.63	0.37
STR	0.26	0.21	0.53	0.53	0.47
SWR	0.41	0.06	0.53	0.61	0.39
TMC	0.50	0.18	0.32	0.66	0.34
WRK	0.38	0.24	0.38	0.57	0.43
WHV	0.41	0.36	0.23	0.53	0.47
Dikaryotic Isolate Frequencies					
	0.21	0.30	0.49	0.45	0.55
Combined Population Frequencies					
	0.34	0.28	0.38	0.53	0.47

¹ *MAT* idiomorph frequencies of over 500 strains were calculated for each listed population from a combination of genotyping and fragment analysis data.

² Dikaryotic isolate frequencies were determined from fragment analysis data alone because the genotyping approach wasn't sensitive enough to detect *MAT1*-only and *MAT2*-only dikaryotes.

FIGURES

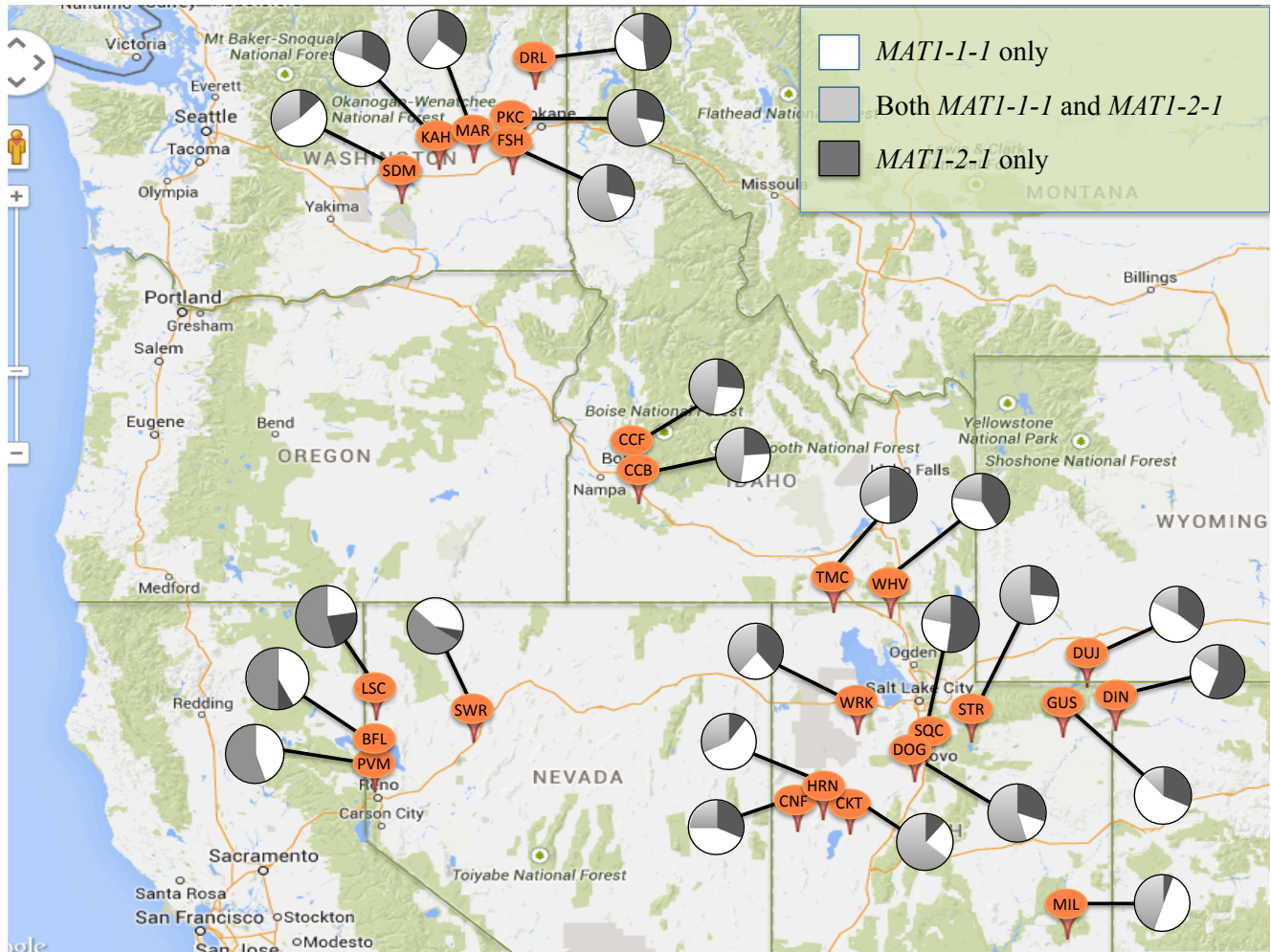
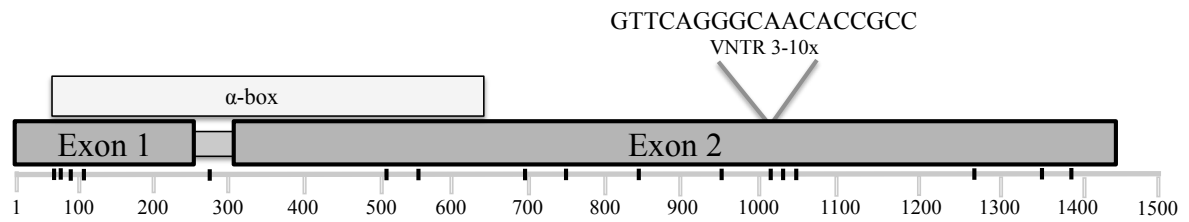


Figure 1 *MAT* idiomorph distribution map 514 *P. semeniperda* strains from 25 collection sites located in CO, ID, NV, UT, and WA were tested for both mating type idiomorphs. Circle graphs represent the percentage of strains one or both of the *MAT* idiomorphs: *MAT1*-only (white), *MAT2*-only (dark grey), and both *MAT1* and *MAT2* (light grey).

2a *MAT1*



2b *MAT2*

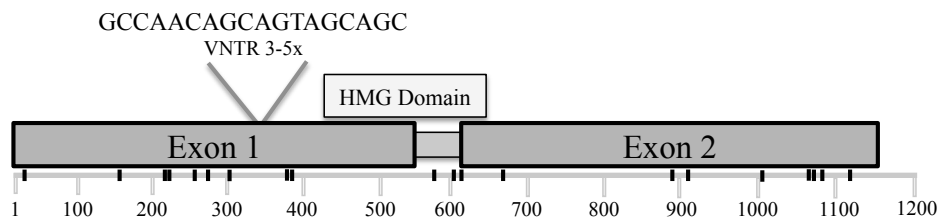


Figure 2 MAT locus characterization *MAT1* (2A) CDS spans 1,304-1,358 bp depending on the number of 18 bp VNTRs present in the gene. DNA sequence that codes for the α -box is located at nucleotide position 63-639 and is interrupted by an intron located at nucleotide position 250-302. *MAT2* (2B) CDS spans 1,120-1,156 bp depending on the number of 18 bp VNTRs present in the gene. DNA sequence that codes for the HMG domain is located at nucleotide position 421-651 and is interrupted by an intron located at nucleotide position 553-608. The vertical black lines represent the location of SNPs identified from nucleotide alignments of the 9 *MAT1* haplotypes (2a) and the 8 *MAT2* haplotypes (2b).

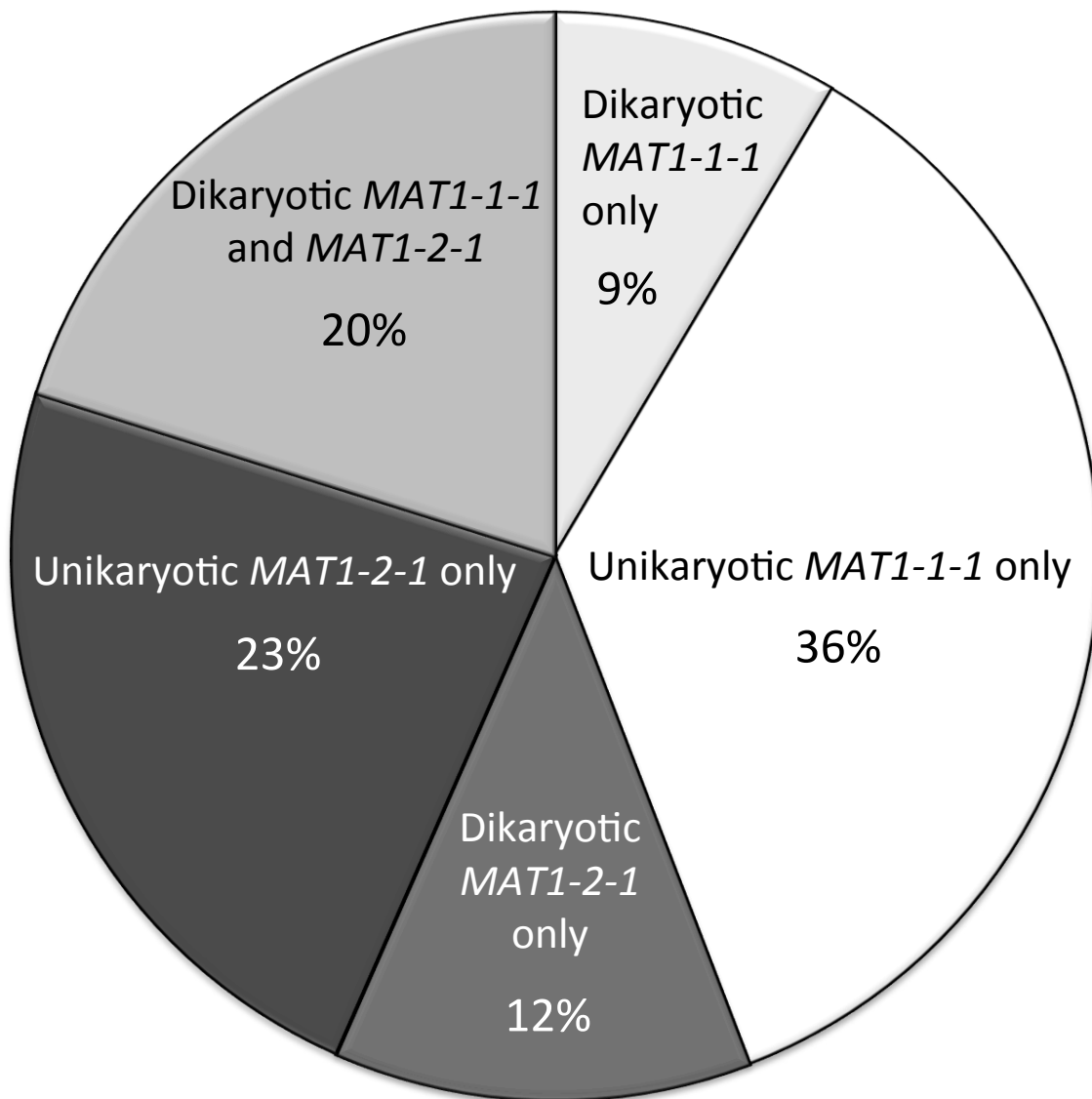
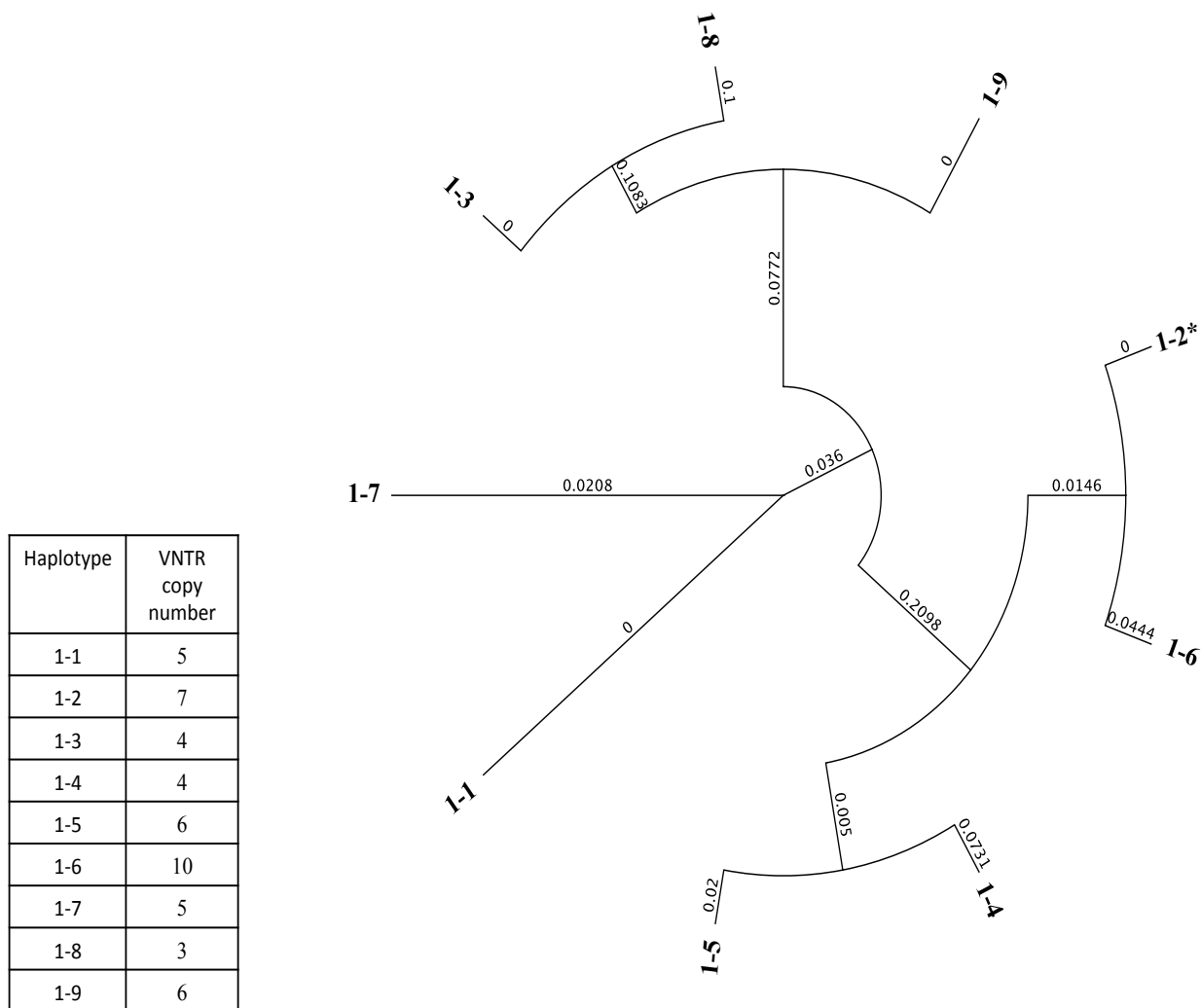


Figure 3 Percentages of unikaryotic and dikaryotic strains with different *MAT* idiomorph combinations

Percentages were calculated from the SSR and *MAT* idiomorph fragment analysis data generated from 128 isolates representing the 25 populations used for genotyping.



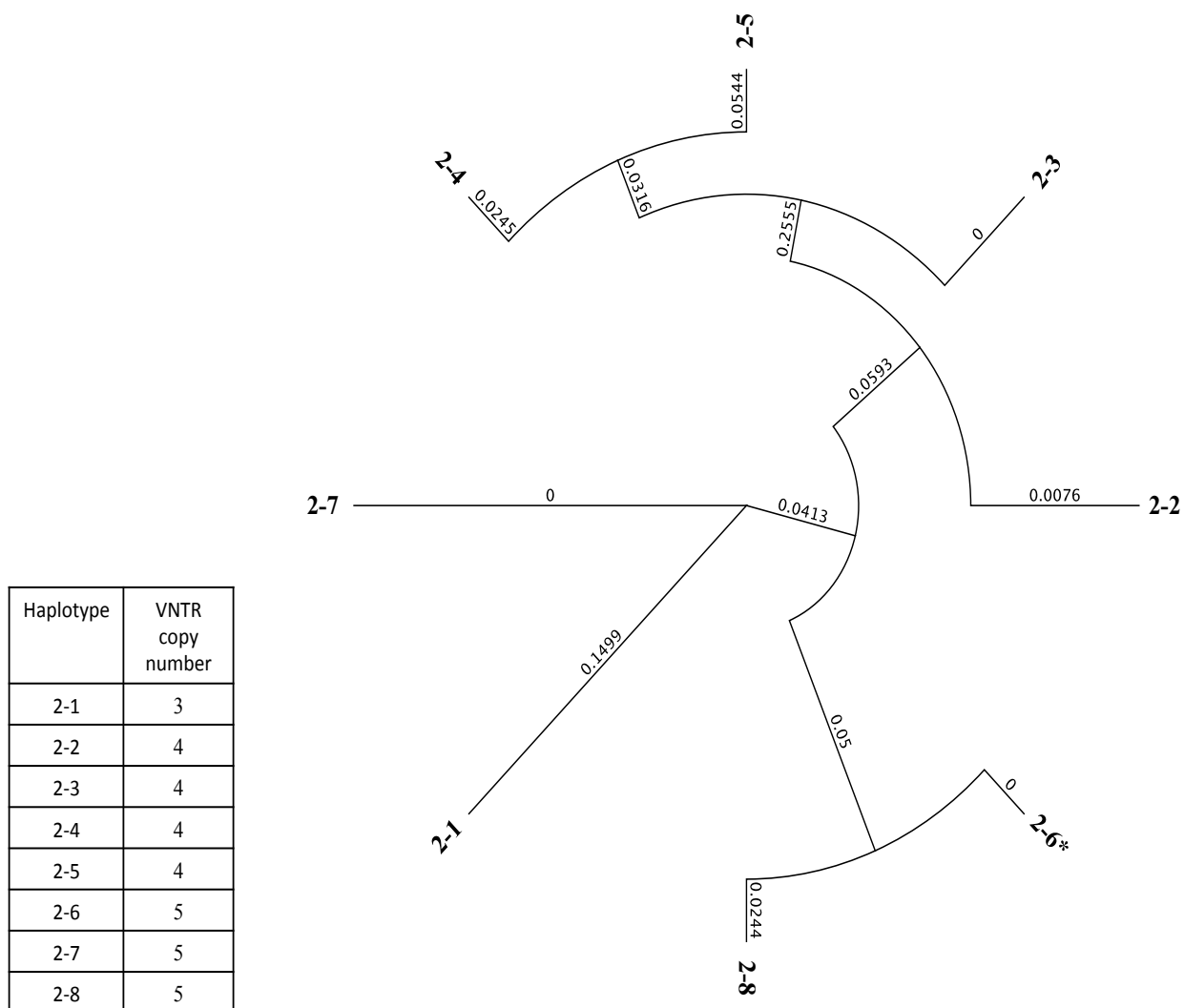


Figure 5 *MAT2* haplotype phylogeny Neighbor-joining phylogenetic tree of *MAT2* haplotypes with genetic distances generated by PHYLIP. Bootstrap values for each branch ranged from 30-100%. The neighbor-joining tree was determined in PHYLIP based on SNPs, gaps, and the VNTR region found within the CDS. Each 18 bp repeat within the VNTR region was calculated as a single polymorphism and all SNPs and gaps contributed equally to the calculations. Table lists the VNTR copy number per haplotype.