Origin and Domestication of the Fungal Wheat Pathogen *Mycosphaerella graminicola* via Sympatric Speciation

Eva H. Stukenbrock,* Søren Banke,* Mohammad Javan-Nikkah,† and Bruce A. McDonald*

*Plant Pathology, Institute of Integrative Biology, Eidgenössische Technische Hochschule, Zurich, Land und Forst Wissenschaft, Zurich, Switzerland; and †Department of Plant Protection, Faculty of Agriculture, University of Tehran Plant Protection Building, Karaj, Iran

The Fertile Crescent represents the center of origin and earliest known place of domestication for many cereal crops. During the transition from wild grasses to domesticated cereals, many host-specialized pathogen species are thought to have emerged. A sister population of the wheat-adapted pathogen *Mycosphaerella graminicola* was identified on wild grasses collected in northwest Iran. Isolates of this wild grass pathogen from 5 locations in Iran were compared with 123 *M. graminicola* isolates from the Middle East, Europe, and North America. DNA sequencing revealed a close phylogenetic relationship between the pathogen populations. To reconstruct the evolutionary history of *M. graminicola*, we sequenced 6 nuclear loci encompassing 464 polymorphic sites. Coalescence analyses indicated a relatively recent origin of *M. graminicola*, coinciding with the known domestication of wheat in the Fertile Crescent around 8,000–9,000 BC. The sympatric divergence of populations was accompanied by strong genetic differentiation. At the present time, no genetic exchange occurs between pathogen populations on wheat and wild grasses although we found evidence that gene flow may have occurred since genetic differentiation of the populations.

Introduction

Archaeological findings provide evidence of the important changes that took place in human culture when the first farming practices were initiated about 12,000 years ago in the Fertile Crescent (Childe 1953; Moore et al. 2000; Zohary and Hopf 2000; Gopher et al. 2002). Genetic data can support archaeology as genome-wide measures of genetic similarity have traced the origin of several domesticated cereals to wild populations of naturally occurring grasses that persist in the Middle East (see Salamini et al. 2002 for review). Intensive selection and cultivation of selected phenotypes altered populations of wild grasses into domesticated varieties of crops. For the cereal crops, the main morphological features that changed through this selection process were seed size, stiffness of the ear rachis, and the ease with which the seeds are released from the glumes (Sharma and Waines 1980; Elias et al. 1996; Taenzler et al. 2002). The process of domestication of certain grasses not only changed the genetic structure of the plant populations but also changed the physical and biotic environment for pathogens associated with the selected grass species. In wild host populations, characterized by patchy distributions, low plant density, high genetic diversity, and uneven distribution of nutrients, the pathogen is exposed to substantial environmental fluctuations, including factors such as humidity, UV radiation, and nutrient availability (Burdon 1993). Agricultural host populations are characterized by higher plant densities, more even nutrient distribution, and greater genetic uniformity, creating a more uniform environment that is less prone to environmental fluctuation and that is also more conducive to pathogen transmission. It is likely that the initiation of agricultural practices and the development of new crop species also led to the emergence of new pathogens or to significant changes through local adaptation in already existing pathogen populations.

One of the most serious pathogens on cultivated wheat today is the fungus *Mycosphaerella graminicola* that is found worldwide with its host, *Mycosphaerella graminicola* is known to be host specific and does not occur on other host species (Eyal et al. 1973, 1985; Saadaoui 1987; van Ginkel and Scharen 1987). Genetic analysis of virulence and resistance in the wheat—*M. graminicola* pathosystem suggests that pathogenicity is controlled by several loci and is likely inherited as a quantitative trait (Kema, Annone, et al. 1996; kema, Sayoud, et al. 1996; Zhan et al. 2005). Population genetic studies at both regional and continental scales demonstrated high levels of genetic diversity and gene flow between pathogen populations (Zhan et al. 2003; Banke and McDonald 2005). *Mycosphaerella graminicola* reproduces both clonally through the production of splash-dispersed pycnidiospores and sexually by the formation of ascospores. Ascospores are wind dispersed (Shaw and Royle 1989) and may provide an important means of long-distance dispersal together with the continuous human-mediated transport of infected plant material (Brown and Hommoller 2002). Phylogeographic studies indicated that the center of origin of *M. graminicola* is the Middle East and that the pathogen was dispersed throughout the world during the spread of wheat cultivation (McDonald et al. 1999; Banke et al. 2004). Ancient patterns of migration inferred from DNA sequence data supported this hypothesis by showing an asymmetrical migration of the pathogen from European and Israeli populations to “New World” countries (Banke and McDonald 2005). Estimates of recent migration patterns using microsatellite markers support an asymmetrical migration of the pathogen from the “Old World” but additionally include North America as a contemporary major donor of pathogen immigrants (Banke and McDonald 2005). This pattern of contemporary migration is consistent with the intensive cultivation of wheat on the North American continent today and global wheat trading patterns.

We hypothesized that *M. graminicola* emerged as a new pathogen during the process of wheat domestication. Sister populations of *M. graminicola* on other grass hosts have, up until now, not been available to test this hypothesis. The closest known relative of *M. graminicola* is the
coalescence parameters such as founding population sizes, ancestral population (Nielsen and Wakeley 2001; Hey and genes between populations after divergence from a common migration model which takes into account the exchange of To account for this possibility, we applied an isolation with scenario complicates the inference of coalescence patterns. A regular exchange of genetic information may result in shared polymorphisms that create the appearance of recent occurring on wild and agricultural host populations. This exchange of genes would have been possible during the pro-
tions occurring on wild grasses and populations adapting to with rice domestication.

5

CI

derived estimates suggested an early origin of the rice-
ficting pathogen Septoria passerinii; however, it is not known whether divergence of these 2 pathogen species coincided with the domestication of wheat and barley. Recently collected isolates of Mycosphaerella pathogens from noncultivated grasses in Iran provided us the opportunity to study the evolutionary history of M. graminicola. DNA sequence data can be combined with genealogical-based analyses to infer processes of population and species divergence and to identify evolutionary forces that have shaped the genetic structure of an organism (Avise 2000; Hey and Machado 2003). For fungi, such studies are limited. Fisher et al. (2001) used a genealogical approach to characterize the biogeographic expansion of the human pathogen Coccidioides immitis into South America. They concluded that the expansion coincided with the human colonization of South America but the estimated timeframe of the expansion had a very large confidence interval (CI) (2,700–228,000 years). In a recent study of the rice blast pathogen Magnaporthe oryzae, Couch et al. (2005) assessed the divergence time for a host shift event from millet to cultivated rice. Their Bayesian-derived estimates suggested an early origin of the rice-infecting pathogen approximately 9,000 years ago (95% CI = 2,500–28,000 years ago), which could be associated with rice domestication.

The emergence of a wheat-adapted Mycosphaerella population is likely to have occurred as a sympatric process where no geographical barriers separated pathogen populations occurring on wild grasses and populations adapting to the new domesticated crops. Because of the close proximity of wild and cultivated host populations and the long-distance dispersal of airborne ascospores, a continuous exchange of genes would have been possible during the process of divergence between Mycosphaerella populations occurring on wild and agricultural host populations. This scenario complicates the inference of coalescence patterns. A regular exchange of genetic information may result in shared polymorphisms that create the appearance of recent divergence even when the actual splitting occurred long ago. To account for this possibility, we applied an isolation with migration model which takes into account the exchange of genes between populations after divergence from a common ancestral population (Nielsen and Wakeley 2001; Hey and Nielsen 2004). The method allows the estimation of multiple coalescence parameters such as founding population sizes, changes in population size, the time of population formation, and gene flow. Previous studies used this coalescence approach to describe the divergence of Drosophila species, chimpanzee subpopulations, and ancient human populations (Hey and Nielsen 2004; Won and Hey 2005; Hey 2005).

The transition from hunter-gatherer cultures led to the emergence of new human infectious diseases that evolved from herd-domesticated animals and successfully spread in the associated human populations (Diamond 1999). We hypothesized a parallel emergence of new plant pathogens mediated by the domestication of crop plants and the development of agriculture. It is probable that the dense host populations and different environmental conditions existing in agroecosystems selected for new pathogen populations that were locally adapted to the agroecosystem, setting the stage for pathogen population differentiation and rapid speciation. To test these hypotheses, we here inferred the evolutionary history of M. graminicola by comparing gene genealogies of closely related populations on uncultivated grasses from the center of origin of both host and pathogen.

Materials and Methods

Fungal Isolates

Uncultivated grasses of 3 different genera, Agropyron repens, Dactylis glomerata, and Lolium multiflorum, infected by a fungal pathogen in the genus Mycosphaerella were collected in the northwestern Iranian province Ardabil (table 1). This region is one of the most important locations for wheat cultivation in Iran. Little is known about the origin and evolution of the 3 grass hosts; however, they are known to be native to Europe, North Africa, and temperate Asia (Hussey et al. 1997). Pathogen isolation and DNA extraction were performed as previously described (Linde et al. 2002). Using repetitive sequence–based polymerase chain reaction (Rep-PCR) fingerprints to differentiate clonal lineages, we selected 28 unique isolates for sequencing. One hundred twenty-three strains of M. graminicola isolated from cultivated wheat collected in Iran, Israel, Oregon, and Switzerland were used to infer the coalescence of Mycosphaerella populations on grasses and wheat. Mycosphaerella graminicola populations were described previously (McDonald et al. 1999; Linde et al. 2002; Zhan et al. 2003; Banke et al. 2004; Juergens et al. 2006). Additionally, we included 10 isolates of the closely related barley

<table>
<thead>
<tr>
<th>Pathogen Population Included in the Analysis</th>
<th>Pathogen</th>
<th>Number of Isolates</th>
<th>Population</th>
<th>Location</th>
<th>Host Plant</th>
<th>Year</th>
<th>Sampling Strategy</th>
<th>Collectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycosphaerella sp. (S1 and S2)</td>
<td>4</td>
<td>Iran</td>
<td>Ardabil, Khalkhal</td>
<td>Lolium multiflorum</td>
<td>2004</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td>Mycosphaerella graminicola</td>
<td>3</td>
<td>Iran</td>
<td>Ardabil, Khalkhal</td>
<td>Dactylis glomerata</td>
<td>2004</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>Iran</td>
<td>Ardabil, Khalkhal</td>
<td>Lolium multiflorum</td>
<td>2004</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Iran</td>
<td>Ardabil, Khalkhal</td>
<td>Agropyron repens</td>
<td>2004</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Iran</td>
<td>Ardabil, Khalkhal</td>
<td>Lolium multiflorum</td>
<td>2004</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td>Mycosphaerella graminicola</td>
<td>16</td>
<td>Iran</td>
<td>Mehran</td>
<td>Triticum aestivum</td>
<td>2001</td>
<td>Random</td>
<td>Javan-Nikkhah M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>Israel</td>
<td>Sa’ad</td>
<td>Triticum aestivum</td>
<td>1992</td>
<td>Hierarchical</td>
<td>Yarden O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>United States</td>
<td>Oregon</td>
<td>Triticum aestivum</td>
<td>1990–1992</td>
<td>Hierarchical</td>
<td>Mundt CC and McDonald BA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Switzerland</td>
<td>Zurich, Winterthur</td>
<td>Triticum aestivum</td>
<td>1999</td>
<td>Hierarchical</td>
<td>McDonald BA</td>
<td></td>
</tr>
<tr>
<td>Septoria passerinii</td>
<td>10</td>
<td>United States</td>
<td>North Dakota</td>
<td>Hordeum vulgare</td>
<td>1995</td>
<td>Random</td>
<td>Long D</td>
<td></td>
</tr>
</tbody>
</table>
pathogen *S. passerinii* collected in Walsh, North Dakota (kindly provided by Goodwin SB) (table 1).

Polymerase Chain Reaction and Sequencing

Internal transcribed spacer (ITS) amplification was performed as described by White et al. (1990) using the primer combination ITS4/ITS5. For population genetic and coalescence analyses, we sequenced 6 DNA loci previously described in *M. graminicola*: β-tubulin (321 bp, a transcribed sequence), Leu (111-bp promoter and 189-bp transcribed sequences of the 3-isopropylmalate dehydrogenase gene), *HPPD* (560 bp, a promoter sequence of the 4-hydroxyphenylpyruvate dioxygenase gene), *STS2* (402 bp of an anonymous RFLP locus), *STL10* (1,104 bp of an anonymous RFLP locus), and *STS43* (430 bp of an anonymous RFLP locus). Sequence data from the loci β-tubulin, Leu, *STS2*, and *STL10* were used in previous studies to determine phylogeny and patterns of migration in *M. graminicola* (Banke et al. 2004; Banke and McDonald 2005). *STS43* is a nuclear RFLP locus that was used in previous population genetic studies of *M. graminicola* (McDonald et al. 1999). The *HPPD* locus was originally sequenced and described by Keon and Hargreaves et al. (1998). Sequence data not published in previous studies can be accessed in National Center for Biotechnology Information GenBank under the accession numbers DQ535489–DQ535722. Amplification and sequencing of all loci were performed as described in Banke et al. (2004). Sequences were aligned and manually adjusted using the programs Sequencher 4.0 (Gene Code, Ann Arbor, MI) and Bioedit Sequence Alignment Editor (http://www.mbio.mahidol.ac.th/Downloads/Mol-Bio/Bioedit/Bioedit.htm).

Data Analysis

Gene Diversity

As most polymorphic sites are located in noncoding DNA, we did not differentiate between silent or nonsilent mutations and all polymorphic sites were analyzed. To initially characterize the phylogenetic relationship between pathogen species, we used the ITS alignment to construct a maximum parsimony tree by heuristic searches in PAUP* 4.0 (Swofford 1996). Trees were built by 100 iterations of random taxon addition with a different number seed for each iteration. Maximum parsimony analyses were also performed using sequence alignments of the 6 DNA loci. Within the *Mycosphaerella* populations from uncultivated grasses, we identified 2 clades, here named S1 (representing 27 isolates) and S2 (representing 34 isolates). Haplotype diversity was calculated for each of the 6 loci using the program Dnasp version 3.53 (Nei 1987; Rozas J and Rozas R 1999). To test for deviation from neutral evolution, we performed Fu and Li’s *D* and *F* (Fu and Li 1993) and Tajima’s *D* tests (Tajima 1989) for all sites in each gene using Dnasp (Rozas J and Rozas R 1999). Maximum parsimony, haplotype diversity, and deviation from neutrality were assessed using entire sequence alignments.

Compatible Alignments

Alignments were processed in the SNAP workbench Java program package, which implements and coordinates several programs to analyze gene genealogies and population parameters (Price and Carbone 2005). The program SNAP map was used to collapse sequences into haplotypes, removing indel mutations and excluding infinite-site violations (Aylor et al. 2006). Conflicting sites showing homoplasy were initially identified among variable sites using the compatibility methods SNAP clade and SNAP matrix in the SNAP workbench. To generate compatible sequence alignments for coalescence analyses, Carbone and Kohn (2001) suggested the removal of haplotypes. Here, instead, we chose to manually remove conflicting sites in each gene alignment to generate compatible nonrecombining alignments for the parsimony and the coalescence analysis. Table 2 shows the number of sites before and after manipulation of the alignments. The majority of polymorphic sites removed were found within the 4 *M. graminicola* populations, consistent with a high number of recombination events. Removal of conflicting sites may have introduced a bias into the data set. The majority of conflicting sites in the 6 genes were found between *M. graminicola* isolates, and we believe that the removal of conflicting sites did not bias the migration or coalescence estimates between the 2 groups.

Haplotype Networks

We analyzed the haplotype alignments of all loci using the program TCS 1.3 (Clement et al. 2000). This program applies a statistical parsimony method to infer unrooted cladograms based on Templeton’s 95% parsimony connection limit (Templeton 1992). All polymorphic sites in the compatible sequence alignments were used to generate the haplotype networks.

LAMARC Parameter Estimation

Population growth parameters, theta: θ values and recombination rates were inferred using the program LAMARC 2.0 (Kuhner et al. 2004). To compare parameters in pathogen populations on uncultivated grasses and wheat, we compared a pooled sample of the 4 *M. graminicola* populations and a pooled sample of S1 and S2. The pooling of populations was done in order to compare a sample of pathogens from cultivated wheat and from uncultivated grasses. Because geographical populations of *M. graminicola* showed little subdivision, these populations were analyzed as one population sample. Due to the smaller sample sizes of S1 and S2, we pooled these populations into one sample. The larger sample sizes improved the performance of the likelihood analysis. The LAMARC algorithm permits intragenic recombination and therefore allowed us to use the complete sequence alignments for each locus. All polymorphic sites were used to assess the population parameters. To estimate θ and a population growth parameter, we used 10 initial chains with 2,000 genealogies sampled and 2 final chains with 20,000 genealogies sampled. Analyses were run with fixed migration rates obtained from analyses performed in the IM program (see below) (Nielsen and Wakeley 2001). Because of large credibility intervals, we also performed LAMARC analyses without fixed migration estimates to determine the influence of these on the LAMARC analyses. Population parameters were inferred using both Bayesian and Maximum Likelihood
analyses. 95% credibility intervals were assessed only from the Bayesian search.

**Isolation with Migration Parameter Estimation**

Migration rates backward in time due to divergence of pathogen populations on uncultivated grasses and wheat were estimated using an isolation with migration model implemented in the Bayesian-based isolation with migration (IM) program (Nielsen and Wakeley 2001). The method estimates the likelihood function of the demographic parameters \( \theta \) (\( \theta \) population diversity), \( t \) (time), \( m \) (the number of migrants per generation between the 2 populations), \( m_t \) (the time of migration events into each population), and \( \mu \) (mutation rate) given 1) a demographic model, 2) the entire set of data (the DNA sequences) and the likelihood of their possible underlying genealogies calculated using a Markov chain Monte Carlo approach, and 3) a mutational model. In this case, we assumed the infinite-site model of Kimura (1969), which is intended for analyses of relatively recent cases of population splitting. Analyses were also performed using the finite-site model of Hasegawa et al. (1985), which allows for recurrent mutations, differences in nucleotide frequencies, and a transition/transversion bias. Highly similar results were obtained using the 2 models. Here only results obtained with the infinite-site model are shown. Nonrecombining sequence alignments were applied as in-files for the IM program.

Initially, pairwise population comparisons were carried out between the 4 *M. graminicola* populations. The IM program estimates mutation rates for each locus and subsequently scales all demographic parameters by the overall mutation rate. Bayesian posterior likelihood distributions for \( \theta \) values were set at 100, maximum migration rates \( (m) \) at 50, and maximum divergence time \( (t) \) at 10. Inheritance scalar was set at 1 for autosomal loci. We repeated every run 5 times with different random number seeds to ensure convergence of the estimates. Marginal histograms were compared between all independent runs for all parameters. Lower and upper bounds of the estimated 90% Highest Posterior Density intervals were calculated for each parameter.

Based on the low population subdivision found with the IM analyses, the 4 *M. graminicola* populations were pooled to form one panmictic population sample. To infer the coalescence of *M. graminicola* from the most closely related wild grass population S1, we conducted an analysis of these 2 population samples. The IM program estimated Bayesian posterior likelihood distributions of \( \theta \) of the ancestral founding population and of the 2 present-day populations; \( t \), the population splitting time; \( m \), gene flow rates between populations; and \( s \), the fraction of the ancestral population that founded the *M. graminicola* population.
Independent analyses were conducted using either mutation scalars for each locus (see below) or the geometric mean across all loci as inferred in IM. Analyses were conducted using 5,000,000 simulations and a burn in time of 100,000. Initially, scalar for maximum θ values was set at 200, maximum migration rates (m) at 50, and maximum divergence time (t) between M. graminicola and S1 at 50 and between M. graminicola and S2 at 100. The first analyses showed large CIs for the t estimates, and we therefore performed several analyses optimizing the maximum t value to 4 to have a smaller CI. We used Metropolis Coupling implemented with 5 chains and a 2-step increment model as described by Hey and Nielsen in the IM documentation (2006). Every run was repeated 5 times with different random number seeds.

**Calculation of Mutation Rates**

The IM analysis uses the estimate μ: mutation rate per locus per year of each locus to infer t (time since divergence of populations). To convert the estimates of t to real time, we needed an overall mutation rate per locus per year. To scale μ, we used an estimate of mutation rate in exons in the β-tubulin locus inferred by Kasuga et al. (2002). The estimate of Kasuga et al. (2002) is given as mutation rate per site per year. We converted this estimate to mutation rate per locus per year using the observed sequence divergence between M. graminicola and S. passerinii in the β-tubulin locus. A mutation rate for the β-tubulin locus was calculated and used to convert IM estimates of μ for all 6 loci. The mean mutation rate across all loci was used to calculate t.

**Conversion of IM Coalescence Parameters to Real Time**

To convert estimates of coalescence time to real time, the overall scaled mutation rate was used. The mean mutation rate across the 6 loci was estimated to be 2.0 \times 10^{-5}. We assumed a generation time of M. graminicola of 1 year (Kema, Verstappen, et al. 1996; Zhan et al. 2004) and the coalescence time t in years can then be calculated as: 

\[
t = r/\mu
\]

(where t is the IM time estimate and μ is the overall mutation rate).

**Genetree Population Divergence Time**

We also used the program Genetree (Griffiths and Tavaré 1994), incorporated in the SNAP program package to infer the sequence coalescence for each gene. Genetree allows the coalescence parameters to be estimated under a full multipopulation coalescence model. For this analysis, we therefore used a full data set of all populations including M. graminicola, S1, S2, and S. passerinii. The program estimates the ancestral history of each haplotype and shows the distribution of mutations on a coalescence scale. This allowed us to compare the divergence of haplotypes between and within each population for each gene.

Due to intergenic recombination, we were not able to combine sequence alignments from the 6 loci. Instead, coalescence parameters were estimated for each locus separately using the compatible sequence alignments of each gene. The coalescence analyses were performed assuming subdivision between the M. graminicola populations, the S1 and S2 populations, and S. passerinii. As we assume different population sizes among the wheat-adapted pathogens and the wild grass pathogens, we inferred the coalescence of populations with different population sizes. To do this, we applied information obtained from the IM estimates about the relative differences in population sizes. To determine the order of coalescence events for haplotypes backward in time, it is necessary to know the amount of migration between populations. Haplotypes from populations linked by migration will coalesce before haplotypes from unlinked populations. As suggested by Carbone and Kohn (2001), we constructed migration matrices for each locus, indicating the number of migrants exchanged between populations, using the program Migrate (Beerli and Felsenstein 2001). These migration matrices were then used as backward migration matrices for ancestral inference in Genetree. The genealogy with the highest root probability was determined by first performing 500,000 simulations of the coalescent with 10 different starting random number seeds. From these runs, the tree with the highest root probability was selected showing the relative divergence time for the pathogen populations and the distribution of mutations along the branches. To convert coalescence units to real time, we used the S1—M. graminicola population split as a calibration point. As this parameter was estimated across all 6 loci, it provided a robust measure of t. As the coalescence time scale of the gene tree is linear, it was possible to assess the time of the split of S. passerinii and S2, as well as the diversification of M. graminicola.

**Results**

Sixty-one Mycosphaerella isolates were obtained from wild grasses collected at 5 different locations in northwest Iran. We differentiated clonal lineages using Rep-PCR fingerprints and selected 28 unique isolates from the 5 Iranian populations for detailed sequence analysis.

**Gene Diversity and Tests of Neutrality**

Based on ITS sequence data, M. graminicola and S. passerinii form a monophyletic cluster within the Mycosphaerella clade (Goodwin et al. 2001). To evaluate the phylogenetic proximity of the wild grass Mycosphaerella isolates to M. graminicola, we initially performed a maximum parsimony analysis based on ITS sequences using S. passerinii as outgroup. Three ITS sequence types were identified among the wild grass populations. The ITS type 1 was identical to ITS sequenced from M. graminicola, the ITS type 2 differed by 1 nucleotide substitution and 4 indel mutations, and the ITS type 3 differed by 2 nucleotide substitutions and 5 indel mutations. In comparison, the number of nucleotide substitutions between M. graminicola and S. passerinii was 13.

We amplified and sequenced 6 polymorphic DNA sequence loci, a total of 3,080 bp, from the wild grass Mycosphaerella populations and from 4 populations of M. graminicola and 10 isolates of S. passerinii. A total of 464 polymorphic sites were identified among the 6 loci.
representing more than 10% of the total number of sequenced base pairs. The majority of the polymorphic sites were found only within the wheat-adapted *M. graminicola* population. Numbers of haplotypes and gene diversity for each locus for *M. graminicola*, S1, and S2 are summarized in table 2. Maximum parsimony analyses were performed for each of the 6 sequence loci to assess phylogenetic relationships (data not shown). Consistent tree topologies identified 2 clusters within the collection of *Mycosphaerella* spp. from wild grasses, here named S1 and S2 (S1 representing the ITS type 1 and S2 the ITS types 2 and 3). None of the haplotypes identified among the 6 loci from S1 and S2 were shared with *M. graminicola*.

Neutrality tests were performed for each locus in the S1 and S2 *Mycosphaerella* clades and for *M. graminicola*. Evidence of nonneutral evolution was found in the HPPD locus for the S1 cluster in Leu and STS2 and for *M. graminicola* in the STL10 locus (table 2). For these 3 loci, the significant statistical values were negative. A significant test result is consistent with either population growth or shrinkage or background selection (Fu 1997). As the majority of nucleotide positions were located in noncoding DNA, we hypothesized that the deviation from neutrality was due to change in effective population size. This hypothesis was supported by the population growth estimates inferred by the program LAMARC 2.0 as described below.

### Haplotype Networks

Haplotype networks were inferred from all 6 loci using compatible sequence alignments. No pattern of geographical association was revealed among the *M. graminicola* haplotypes, and the most frequent haplotype for the 6 loci was present in all 4 geographical populations. All 6 networks exhibited a star-like shape for the *M. graminicola* haplotypes. Haplotype networks of the STS2 and HPPD loci are shown in figures 1a and b. These 2 networks had the smallest number of missing steps. The STS2 network contained 26 *M. graminicola* haplotypes, 2 S1 haplotypes, 1 S2 haplotype, and 3 *S. passerinii* haplotypes. The total number of STS2 and HPPD haplotypes was higher (table 2); however, as conflicting sites were removed from the alignment prior to the analysis, the number of haplotypes was decreased. The number of mutational steps between *S. passerinii* and the S2 population was 72 and between S2 and *M. graminicola* was 18. Between S1 and *M. graminicola*, there were 13 mutations. Homoplasy was present in the STS2 network in the *M. graminicola* clades as shown by the loop between haplotype numbers 21, 3, 35, and 25.

In the HPPD network 29 haplotypes were recognized in the *M. graminicola* cluster, 2 in the S1, 2 in the S2, and 2 in the *S. passerinii* cluster. There were 76 differences between *S. passerinii* and *M. graminicola*. Between the S1 group and *M. graminicola* populations there were 8 differences, and between S2 and *M. graminicola* 9 differences. For both genes, the number of steps between *M. graminicola* haplotypes did not exceed 3 mutations, suggesting that almost all possible haplotypes were sampled for this group.

Population Expansion

The population parameters \( \theta \) for haploids \( \theta = 2N_e \mu \), where \( N_e \) is effective population size and \( \mu \) is mutation rate) and population growth rates were calculated using both Maximum Likelihood and Bayesian analyses in LAMARC. \( \theta \) values were used as a relative measure of effective population sizes. We found consistent values of \( \theta \) using the 2 analytical strategies; however, the growth rate estimates were less consistent and CIs for this parameter were large. The Bayesian estimates of \( \theta \) and growth rate are summarized in table 3. The overall pattern observed for the 6 genes suggests a larger effective population size of *M. graminicola* \( \theta = 0.022 \) when compared with the pooled S1 and S2 \( \theta = 0.0063 \). For individual loci, \( \theta \) ranged from 0.02 (\( \beta \)-tubulin) to 0.16 (HPPD) in *M. graminicola* and from 0.004 (STS2) to 0.017 (Leu) in the S1 and S2 population. The overall growth rate of the *M. graminicola* population (13.86) is consistent with a large effective population size and a population expansion. The negative growth rate values in the S1–S2 population (−139.91) indicate a decline in population size or a consistently small population size. For individual loci in *M. graminicola*, growth rates ranged from 1.51 (Leu) to 472.08 (STS2) and for the S1–S2 population from −462.28 (STS2) to 18.75 (Leu). These estimates may reflect a slower rate of evolution of the Leu locus as compared with the highly variable STS2 locus.

The analyses were also performed without fixed migration values. Estimates of \( \theta \) and population growth did not change significantly, neither was the range of the credibility intervals reduced. In these analyses, the overall \( \theta \) values were 0.022 (0.008–0.18) for the *M. graminicola* population and 0.007 (0.002–0.05) for the S1 population. Population growth estimates were 5.84 (−60.3–581.05) and −144.23 (−467.61–843.54) for *M. graminicola* and S1, respectively.

Recombination rates inferred by LAMARC varied among the different loci, ranging from 2.23 \times 10^{-5} (HPPD) to 0.16 (STS2) (table 3). The low levels of recombination found in the \( \beta \)-tubulin, HPPD, STL10, and STL43 suggest that mutation had a greater influence on the evolution of these loci than recombination.

### Coalescence and Migration

Estimates of directional gene flow between the 4 geographical *M. graminicola* populations were consistent with relatively high rates of migration among populations (table 4). Based on these estimates and previous results showing high genetic identity and low global \( F_{ST} \) values (Linde et al. 2002; Zhan et al. 2003; Juergens et al. 2006), we considered the 4 *M. graminicola* populations as one panmictic population. The pooled *M. graminicola* sample was compared with the S1 wild grass population. Estimates of directional gene flow backward in time since the divergence of *M. graminicola* and the S1 population were low, with an overall movement of genes from the wild grass S1 population to the wheat-adapted population (table 5 and fig. 2). From S1 to *M. graminicola*, the migration parameter \( m_{S1} \) was estimated to be 0.80 and from *M. graminicola* to S1 \( m_{S1} = 0.15 \). CIs are given in table 5.
Fig. 1.—Parsimony haplotype networks for the 2 loci STS2 (a) and HPPD (b). Origin of haplotypes from the *Septoria passerinii*, the Iranian S1 and S2 populations, and the 4 geographical populations of *Mycosphaerella graminicola* are indicated by different colors. Number for each haplotype is given. Numbers in parentheses refer to number of isolates with this haplotype if more than 1. Dots are hypothetical missing intermediate haplotypes.
The mean time of migration events was estimated over the 6 loci to assess when the majority of migration events took place. Converting the migration time estimates to real time, we find that the mean migration time corresponded to 8,250–12,750 years ago. The estimates suggest that the majority of migration events occurred at the time when the 2 populations were initially diverging and that recent migration events are uncommon.

Using the IM program, we inferred coalescence parameters for the split between *M. graminicola* and the S1 clade. To assess the efficiency of the Markov Chain mixing, we used the estimates of parameter autocorrelation $\rho_k$ and Effective Sample Size (ESS). $\rho_k$ is calculated from the parameter $k$, which is based on the number of steps between the pairs of values that are included in the IM parameter calculation. $\rho_k$ values close to zero indicate a high $k$ value, which is expected for independent samples. Similarly, the ESS is a measure of the number of independent points that have been sampled for each parameter and can be used as a guide to how well the Markov Chain is mixing. The parameters $\rho_k$ and ESS obtained from our IM analyses indicated that parameter estimations were not biased by strong autocorrelations in the Markov Chain mixing (table 5).

Estimates of coalescence parameters are summarized in table 5. $\theta$ of the ancestral population of S1 and *M. graminicola* was 10.97. Consistent with the LAMARC estimates, the IM analysis showed significantly higher $\theta$ values for *M. graminicola* ($\theta = 51.3$) than the wild grass S1 population ($\theta = 0.26$). However, the parameter $s$ (the fraction of the ancestral population which founded *M. graminicola*) suggests that the *M. graminicola* population was founded by only a very small fraction ($s = 0.02$) of the ancestral population (fig. 2). A large fraction ($1 - s = 0.98$) of the ancestral population founded the present-day S1 population. These estimates are consistent with a strong expansion of *M. graminicola* and a decline in the S1 population since the population split. Due to the large credibility intervals for the splitting factor $s$ (table 5), we also performed the IM analyses using a model that assumes constant population size. Estimates under this model are also shown in table 5. Estimates inferred by this model remained within the same range; however, credibility intervals for $\theta$ values of *M. graminicola* and S1 were narrower.

We calculated the overall mutation rate for the $\beta$-tubulin locus. This gave an overall mutation rate of $3.2 \times 10^{-6}$ mutations per locus per year. This estimate was used to calibrate the IM mutation rate parameter $u$ for the other 5 loci. Mutation rates are summarized in table 3. The highest mutation rate of $4.0 \times 10^{-5}$ substitutions per locus per year was found in the Leu locus. The lowest mutation rate was found in the $\beta$-tubulin locus. Assuming a generation time of 1 year for these *Mycosphaerella* pathogens, we calculated an overall mutation rate of $2.0 \times 10^{-5}$ per locus per year. Converting estimates of $t$ to real time, we find that the population split between the wheat-adapted *M. graminicola* population and the wild grass S1 population occurred approximately 10,500 years ago (95% CI: 2,000–19,500). The model parameters $\theta$ and $m$ were also converted to demographic parameters (table 5). These estimates suggest approximate values of effective population size and migration rates for *M. graminicola* and S1. The effective

---

**Table 3**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Population Growth, Rate per Locus</th>
<th>S1 and S2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>$\pi$</td>
<td>$\mu$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>$\beta$-tubulin</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$4.0 \times 10^{-5}$</td>
<td>$3.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>HPPD</td>
<td>$9.5 \times 10^{-6}$</td>
<td>$4.0 \times 10^{-5}$</td>
<td>$1.8 \times 10^{-5}$</td>
</tr>
<tr>
<td>STS43</td>
<td>$8.0 \times 10^{-5}$</td>
<td>$4.0 \times 10^{-5}$</td>
<td>$2.0 \times 10^{-5}$</td>
</tr>
<tr>
<td>Leu</td>
<td>$9.5 \times 10^{-5}$</td>
<td>$4.0 \times 10^{-5}$</td>
<td>$2.0 \times 10^{-5}$</td>
</tr>
<tr>
<td>Overall</td>
<td>$1.8 \times 10^{-5}$</td>
<td>$4.0 \times 10^{-5}$</td>
<td>$2.0 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

**NOTE.**—Estimates of $\theta$, $\mu$, and $\gamma$ were calculated using the IM Program (Minin and Suchard 2003; Minin et al. 2004) and the S1 population size.

---

The estimates suggest that migration events were uncommon.

Using the IM program, we inferred coalescence parameters for the split between *M. graminicola* and the S1 clade. To assess the efficiency of the Markov Chain mixing, we used the estimates of parameter autocorrelation $\rho_k$ and Effective Sample Size (ESS). $\rho_k$ is calculated from the parameter $k$, which is based on the number of steps between the pairs of values that are included in the IM parameter calculation. $\rho_k$ values close to zero indicate a high $k$ value, which is expected for independent samples. Similarly, the ESS is a measure of the number of independent points that have been sampled for each parameter and can be used as a guide to how well the Markov Chain is mixing. The parameters $\rho_k$ and ESS obtained from our IM analyses indicated that parameter estimations were not biased by strong autocorrelations in the Markov Chain mixing (table 5).

Estimates of coalescence parameters are summarized in table 5. $\theta$ of the ancestral population of S1 and *M. graminicola* was 10.97. Consistent with the LAMARC estimates, the IM analysis showed significantly higher $\theta$ values for *M. graminicola* ($\theta = 51.3$) than the wild grass S1 population ($\theta = 0.26$). However, the parameter $s$ (the fraction of the ancestral population which founded *M. graminicola*) suggests that the *M. graminicola* population was founded by only a very small fraction ($s = 0.02$) of the ancestral population (fig. 2). A large fraction ($1 - s = 0.98$) of the ancestral population founded the present-day S1 population. These estimates are consistent with a strong expansion of *M. graminicola* and a decline in the S1 population since the population split. Due to the large credibility intervals for the splitting factor $s$ (table 5), we also performed the IM analyses using a model that assumes constant population size. Estimates under this model are also shown in table 5. Estimates inferred by this model remained within the same range; however, credibility intervals for $\theta$ values of *M. graminicola* and S1 were narrower.

We calculated the overall mutation rate for the $\beta$-tubulin locus. This gave an overall mutation rate of $3.2 \times 10^{-6}$ mutations per locus per year. This estimate was used to calibrate the IM mutation rate parameter $u$ for the other 5 loci. Mutation rates are summarized in table 3. The highest mutation rate of $4.0 \times 10^{-5}$ substitutions per locus per year was found in the Leu locus. The lowest mutation rate was found in the $\beta$-tubulin locus. Assuming a generation time of 1 year for these *Mycosphaerella* pathogens, we calculated an overall mutation rate of $2.0 \times 10^{-5}$ per locus per year. Converting estimates of $t$ to real time, we find that the population split between the wheat-adapted *M. graminicola* population and the wild grass S1 population occurred approximately 10,500 years ago (95% CI: 2,000–19,500). The model parameters $\theta$ and $m$ were also converted to demographic parameters (table 5). These estimates suggest approximate values of effective population size and migration rates for *M. graminicola* and S1. The effective
population size of the pooled M. graminicola sample exceeded 1 million, whereas the S1 effective population size was 6,500. The migration rate per year for M. graminicola into S1 was $3 \times 10^{-6}$ and for S1 into M. graminicola $1.6 \times 10^{-5}$. However, these estimates may be imprecise in this case where single spores represent individuals and the yearly production of sexual and asexual fungal spores may exceed billions.

We also used the program Genetree (Griffiths and Tavaré 1994) to infer the gene history of the individual loci STS2, $\beta$-tubulin, HPPD, and STL10. We did not have information from the S2 population and S. passerinii for Leu or STS43 or from S. passerinii for the STL10 locus. Overall tree topologies and the relative divergence of M. graminicola from S1 and S2 were consistent for STS2, $\beta$-tubulin, HPPD, and STL10. The ancestral distributions of mutations and coalescence events for the loci STS2 and $\beta$-tubulin are illustrated in figures 3a and b. The S. passerinii cluster branched into the trees at the deepest point, suggesting an ancient divergence of S. passerinii from the other Mycosphaerella clades. The S1 and M. graminicola clades represented the youngest lineages of the trees. The STS2 RFLP locus showed a higher resolution of the different population groups. Less resolution was shown in the $\beta$-tubulin locus suggesting a more conserved evolution of some regions of this locus, which also showed the lowest mutation rate. In the $\beta$-tubulin and STL10 loci, we found evidence of introgression into M. graminicola as a few isolates were positioned in the same groups as the S1 population. This pattern most likely reflects gene flow from S1 into M. graminicola after divergence. The M. graminicola isolates clustering with the S1 population were from the Iranian (as observed in the $\beta$-tubulin locus) and the Israeli (as observed in the STL10 locus) populations. In both cases, the divergence of these isolates did not occur recently and may reflect ancient gene flow events. Scaling the coalescence time in the S2ST gene tree to real time according to the split between M. graminicola and the S1 population inferred in the IM analyses, we calculate that the divergence from S. passerinii occurred approximately 68,500 (95% CI: 13,700–126,700) years ago. In the $\beta$-tubulin gene tree, the divergence of S. passerinii and S2 is not well resolved, and both clusters are positioned as the roots of the tree. However, the number of mutations separating S. passerinii from M. graminicola and S1 is 15 and only 10 for the S2 population. This suggests that the actual divergence of S. passerinii is more ancient than the divergence of S2. The Genetree coalescence estimates from the STS2 locus suggested a divergence time of M. graminicola from the S2 clade approximately 18,500 (95% CI: 3,700–34,225) years ago. After divergence from the S1 clade, the M. graminicola populations experienced a strong expansion

### Table 4

<table>
<thead>
<tr>
<th>Source population</th>
<th>M. graminicola (migration, m)</th>
<th>Sink Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran (16)</td>
<td>0.16 (0.08–0.30)</td>
<td>21.25 (7.95–54.15)</td>
</tr>
<tr>
<td>Israel (40)</td>
<td>11.65 (6.35–18.75)</td>
<td>1.25 (0.05–4.05)</td>
</tr>
<tr>
<td>Oregon (32)</td>
<td>0.35 (0.05–1.55)</td>
<td>0.85 (0.25–2.15)</td>
</tr>
<tr>
<td>Switzerland (32)</td>
<td>13.15 (5.25–24.55)</td>
<td>7.30 (2.30–14.30)</td>
</tr>
</tbody>
</table>

Note.—ESS is given for each population in parentheses. Source populations are shown on the left side and sink populations are indicated along the top. 95% credibility intervals are given in parentheses.

### Table 5

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model Parameters, Constant Population Size</th>
<th>Model Parameters, Population Size Change</th>
<th>Demographic Parameters, Population Size Change</th>
<th>p* (k = 1,000,000)</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_3$</td>
<td>10.97 (5.24–23.36)</td>
<td>10.97 (5.24–21.45)</td>
<td>274,250 (131,000–536,250)</td>
<td>−0.256</td>
<td>8</td>
</tr>
<tr>
<td>$\theta_{5b}$</td>
<td>11.2035 (2.62–25.98)</td>
<td>5.25 (4.05–9.37)</td>
<td>1,281,250 (101,250–9,934,250)</td>
<td>0.078</td>
<td>22</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.09 (0.09–1.67)</td>
<td>0.26 (0.09–161.14)</td>
<td>6,500 (2250–4,028,500)</td>
<td>−0.108</td>
<td>6</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.16 (0.08–0.30)</td>
<td>0.21 (0.04–0.39)</td>
<td>10,500 (2,000–19,500)</td>
<td>−0.322</td>
<td>4</td>
</tr>
<tr>
<td>$m_{5b}$</td>
<td>0.03 (0.03–0.48)</td>
<td>0.15 (0.01–0.7)</td>
<td>$3 \times 10^{-5}$</td>
<td>−0.279</td>
<td>4</td>
</tr>
<tr>
<td>$m_b$</td>
<td>0.58 (0.03–1.93)</td>
<td>0.80 (0.40–1.40)</td>
<td>$1.6 \times 10^{-5}$</td>
<td>0.267</td>
<td>4</td>
</tr>
<tr>
<td>$m_{asg}$</td>
<td>0.092 (0.038–0.19)</td>
<td>0.255 (0.225–0.30)</td>
<td>$8.7 \times 10^{-5}$</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$m_s$</td>
<td>0.126 (0.063–0.257)</td>
<td>0.165 (0.12–0.21)</td>
<td>$8.25 \times 10^{-5}$</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$s$</td>
<td>NA</td>
<td>0.02 (0.004–0.22)</td>
<td>NA</td>
<td>0.013</td>
<td>6</td>
</tr>
</tbody>
</table>

Note.—$NA$, not applicable. Analyses were performed assuming either a constant population size or allowing population size change. Estimates are shown as model parameters and as converted demographic parameters for the analyses allowing population size change. As a measure of effective population size, the program estimates $\theta$ for haploid $\theta = 2N_e \mu$, where $N_e = $ effective population size and $\mu = $ mutation rate inferred over all loci. $\tau$ is given for each present-time population and for the common ancestral population. An estimate of divergence time of the 2 descendant populations is given as $t$. Migration from M. graminicola to S1 was estimated as $m_{sg}$ and from S1 to M. graminicola as $m_s$. $m_{asg}$ (gene flow rates per locus per generation). Mean migration times over all loci are shown as $mt$. $\theta$ is the fraction of the ancestral population that founded the M. graminicola population. 95% credibility intervals are shown in parentheses. $p*$ is the autocorrelation for the distance value $l$ indicating the number of steps between the pairs of values that were included in the parameter calculation. The measure ESS shows the number of independent points that have been sampled for each parameter.
as demonstrated by the diversification of *M. graminicola* haplotypes in the gene trees. The approximate time of this diversification was 5,000–6,000 years ago. Independent analyses of 4 loci showed consistent coalescence times between pathogen populations within the same CIs.

**Discussion**

The genetic data presented here provide evidence that the divergence of the wheat-adapted pathogen *M. graminicola* from an ancestral population infecting wild grasses in the Middle East occurred approximately 10,500 years ago. This coalescence event coincided with the beginning of agricultural-based societies in the Fertile Crescent and the domestication of wild grasses. Our findings demonstrate that the domestication of an agricultural crop was simultaneously accompanied by the domestication of a fungal pathogen.

We found 1–3 fixed nucleotide substitutions at the *ITS* locus among the *Mycosphaerella* populations. Phylogenetic relationships in the clade *Mycosphaerella* were described in a previous study using *ITS* sequences (Goodwin et al. 2001). It was found that the average number of within-species nucleotide substitutions in the *Mycosphaerella* genus ranged from 0 to 7. Based on *ITS* sequences, we could not resolve the S1 and S2 populations from *M. graminicola*, indicating a close phylogenetic relationship below the species level. The lack of nucleotide variation in the *ITS* locus suggests that this gene is unsuitable to resolve clades of pathogens that have been separated for less than 11,000 years. In contrast to the *ITS* locus, the 6 loci that were used for population and coalescence analyses showed high amounts of nucleotide variability within and between populations. Tests of nonneutral evolution could not be rejected for any of the 6 loci, but we believe that the observed pattern of gene diversity is most likely due to changes in population size.

The most frequent *M. graminicola* haplotypes were present in all 4 geographical populations, supporting previous results showing that the majority of gene diversity is distributed within individual populations and that populations are not geographically differentiated (McDonald et al. 1999; Linde et al. 2002; Zhan et al. 2003). The Iranian *M. graminicola* haplotypes clearly belonged to the wheat-adapted population though they originated from the same geographical region as S1 and S2. Isolates of the S1 and S2 populations were all collected very close to cultivated wheat fields but still differed from the wheat-adapted pathogens. We found low haplotype and gene diversity in the S1 and S2 populations even though these isolates originated from 5 different locations and had a broader host range (both S1 and S2 were collected from 3 different hosts). The small number of mutations separating haplotypes indicates that a large fraction of the total diversity was sampled.

The star-like shape of the *M. graminicola* populations in the haplotype networks is consistent with a significant population expansion (figs. 1a and b). We expected to find a higher haplotype and gene diversity in the pathogen populations isolated from uncultivated grasses as a result of diversifying selection imposed by different environments and host species (Burdon 1993). However, the low genetic diversity observed in the S1 and S2 populations is consistent with small effective population sizes of these populations and may reflect a different population biology than *M. graminicola*.

Results from the LAMARC and the IM analyses were both consistent with a significant expansion in *M. graminicola* populations and a decline in the S1 and S2 populations. A new version of the "isolation with migration" model allowed us to estimate the fraction of the ancestral pathogen population that gave rise to the *M. graminicola* population. Given the contemporary θ values, it is possible to extrapolate the relative increase or decrease in population sizes since the time of divergence. The IM estimates of demographic parameters suggest a recent split where only a small fraction of the ancestral pathogen population founded a new population during the domestication of the wheat host. A much larger fraction of the ancestral population founded the S1 group, which today persists on uncultivated grasses. The increase in *M. graminicola* effective population size was most likely caused by the corresponding increase in host population size following the spread of agriculture. We hypothesize that the decrease in size of the S1 and S2 populations reflects the significant decline in wild host populations that occurred during the conversion of wild grass habitats into agricultural fields. We are not able to determine whether the increase or decrease in population size of the respective pathogen populations occurred in a linear way through time or whether some historical events were particularly favorable or unfavorable for the respective pathogen populations. However, expansion of agriculture has favored pathogen propagation and dispersal by providing dense and uniform host populations that cover a large area. The spread of agriculture to other regions moved domesticated plants and animals across Europe and Asia (Clark 1965; Ammerman and Cavalli-Sforza 1971), and many centuries later wheat was introduced into New World countries by the European colonialism. Banke and McDonald (2005) showed evidence for past migration of
M. graminicola from the Middle East and Europe to New World countries, suggesting that the pathogen traveled with its host. The increase in wheat cultivation is also reflected in the burst of M. graminicola diversification and population expansion that the coalescence analyses dated to approximately 5,000–6,000 years ago.

We estimated the directional amount of gene flow occurring among M. graminicola and the S1 population since their divergence. Our results indicate that only a limited amount of gene flow has occurred between the 2 sympatrically diverging populations and that the small amount of genetic exchange detected has been mainly from populations on wild grasses to wheat-infecting populations. This points to a possible mechanism for the introgression of new avirulence genes or pathogenicity factors into M. graminicola from the S1 clade. However, estimates of migration time showed that the majority of gene flow events occurred at the time when populations were initially diverging approximately 10,000 years ago. Evidence of gene flow was also present in the distribution of haplotypes in the Genetree analyses. The Iranian M. graminicola population contained one isolate, which diverged from the S1 group later than the remaining M. graminicola haplotypes (fig. 3b). The position of this isolate in the β-tubulin gene tree suggests that gene flow and recombination also took place between the 2 diverged populations after genetic differentiation had already occurred. In summary, our data indicate that the pathogen host shift did not occur as one single event but probably happened through a series of introgressions of isolates from uncultivated grasses into the wheat-infecting population. Multiple host shifts also occurred in the evolution of the rice blast pathogen, which similarly emerged as a new pathogen by the domestication of rice (Couch et al. 2005).

Consistent with our results, other studies have reported extensive genetic differentiation between diverging fungal species (Bicknell and Douglas 1970; Horgen et al. 1984; Vilgalys and Johnson 1987; Dettman et al. 2003). The high amount of genetic diversity observed between relatively recently separated species suggests that speciation in Mycosphaerella was not only associated with host adaptation but was also accompanied by significant genetic differentiation.

FIG. 3.—Coalescence-based gene genealogies of the STS2 locus (a) and β-tubulin (b) showing the distribution of mutations in 4 geographical populations of Mycosphaerella graminicola (Mg), the 2 Iranian S1 and S2 populations, and S. passerineii (Sp USA). The split between M. graminicola and S1 was used to convert the coalescence scale to real time. The number of haplotypes represented in each branch is shown below the trees. The vertical lines enclose the position of haplotypes derived from the S1 population. Time scale is shown at the right side of each genealogy.
have been discussed and addressed in many studies (e.g., Burnett 1983; Hibbet et al. 1995; Natvig and May 1996; Brasier 1987; Petersen and Hughes 1999; Taylor et al. 2000; Harrington et al. 2002; Kohn 2005). Reinforcement has been suggested as an important mechanism for divergence between sympatrically diverging populations (Kohn 2005) and has been demonstrated in several studies (Anderson et al. 1980; Capretti et al. 1990; Stenlid and Karlsson 1991; Dettman et al. 2003). Reinforcement is the active mechanism that leads to an increase in prezygotic reproductive isolation and reduced hybrid fitness, thereby favoring intraspecific mating and over time resulting in genetic differentiation between species (Dobzhansky 1951). Reproductive success between M. graminicola, S1, and S2 was not tested in this study, but such tests could provide additional knowledge regarding the mechanisms underlying the sympatric speciation of Mycosphaerella populations.

The coalescence analyses performed in this study provide an approximate picture of the divergence of Mycosphaerella populations since the split from S. passerinii. The coalescence time was calculated using mutation rates from each of the 6 loci. The scaled mutation rates shown in this study are within the same range as mutations rates inferred in IM from a multilocus data set of human populations (Hey 2005) and as experimentally shown in yeast (Zeil and DeVisser 2001). As our estimates were calculated across 6 loci from 360 individuals using a Bayesian-based approach, we believe these mutations rates provide robust estimates. Based on the 6 inferred mutation rates, the Genetree coalescence analyses suggest that S. passerinii diverged from the 3 other Mycosphaerella groups approximately 68,500 years ago. This implies that speciation of M. graminicola and S. passerinii did not occur simultaneously with the domestication of wheat and barley but instead occurred much earlier. The divergence of these pathogen species may have been related to the specialization of S. passerinii to a Hordeum host.

Mycosphaerella graminicola and the S1 group represent the youngest phylogenetic clades that diverged from the S2 clade approximately 18,500 years ago. Mycosphaerella graminicola and the S1 group split 10,500 years ago, consistent with the genetic and archeological data obtained from crop plants in the Fertile Crescent showing speciation of the host plants (Salamin et al. 2002). Coalescence analyses of M. oryzae similarly showed that the rice blast pathogen emerged simultaneously with the domestication of rice approximately 9,000 years ago (Couch et al. 2005). The strong population expansion and global spread of M. graminicola demonstrate the accelerated pathogen evolution that was mediated by human practices and favored by the huge land areas covered by wheat. This cospeciation of M. graminicola and wheat took place within a very short evolutionary time frame, specifically, during the 10,000 to 12,000-year period in which historical changes in the host (i.e., the increase in wheat cultivation and the global dispersal) had a major impact on the parasite demography. A similar picture has emerged from coalescence analyses of the malaria parasite Plasmodium falciparum (Joy et al. 2003). Changes in human culture and the introduction of agricultural practices also led to an accelerated coevolution of the malaria mosquito Anopheles gambiae and P. falciparum, both experiencing strong population increases and migration to other regions. These examples illustrate how agricultural practices shaped the evolution of some of our most important modern pathogens.

**Acknowledgments**

We would like to thank M. Zala for technical support and Ignazio Carbone and Rasmus Nielsen for helpful suggestions to the coalescence analyses. Additionally, we would like to thank 3 anonymous reviewers for helpful comments on the manuscript. This work was funded by the Swiss Federal Institute of Technology (ETH), Zurich, and the Swiss National Science Foundation (grant 3100A0-104145).

**Literature Cited**


Shaw MW, Royle DJ. 1989. Airborne inoculum as a major source of Septoria tritici (Mycosphaerella graminicola) infections in winter wheat crops in the UK. Plant Pathol. 8:35–43.