



## The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species

Eva H. Stukenbrock, Thomas Bataillon, Julien Y. Dutheil, et al.

*Genome Res.* 2011 21: 2157-2166 originally published online October 12, 2011  
Access the most recent version at doi:[10.1101/gr.118851.110](https://doi.org/10.1101/gr.118851.110)

---

**References** This article cites 41 articles, 4 of which can be accessed free at:  
<http://genome.cshlp.org/content/21/12/2157.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2011 by Cold Spring Harbor Laboratory Press

## Research

# The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species

Eva H. Stukenbrock,<sup>1,5,6</sup> Thomas Bataillon,<sup>1</sup> Julien Y. Dutheil,<sup>1</sup> Troels T. Hansen,<sup>1</sup> Ruiqiang Li,<sup>2</sup> Marcello Zala,<sup>3</sup> Bruce A. McDonald,<sup>3</sup> Jun Wang,<sup>2,4</sup> and Mikkel H. Schierup<sup>1</sup>

<sup>1</sup>Bioinformatics Research Center, Aarhus University, C.F. Moellers Alle, DK-8000 Aarhus C, Denmark; <sup>2</sup>BGI-Shenzhen, Shenzhen 518083, China; <sup>3</sup>ETH Zurich, Institute of Integrative Biology, 9082 Zurich, Switzerland; <sup>4</sup>Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark

The fungus *Mycosphaerella graminicola* emerged as a new pathogen of cultivated wheat during its domestication ~11,000 yr ago. We assembled 12 high-quality full genome sequences to investigate the genetic footprints of selection in this wheat pathogen and closely related sister species that infect wild grasses. We demonstrate a strong effect of natural selection in shaping the pathogen genomes with only ~3% of nonsynonymous mutations being effectively neutral. Forty percent of all fixed nonsynonymous substitutions, on the other hand, are driven by positive selection. Adaptive evolution has affected *M. graminicola* to the highest extent, consistent with recent host specialization. Positive selection has prominently altered genes encoding secreted proteins and putative pathogen effectors supporting the premise that molecular host–pathogen interaction is a strong driver of pathogen evolution. Recent divergence between pathogen sister species is attested by the high degree of incomplete lineage sorting (ILS) in their genomes. We exploit ILS to generate a genetic map of the species without any crossing data, document recent times of species divergence relative to genome divergence, and show that gene-rich regions or regions with low recombination experience stronger effects of natural selection on neutral diversity. Emergence of a new agricultural host selected a highly specialized and fast-evolving pathogen with unique evolutionary patterns compared with its wild relatives. The strong impact of natural selection, we document, is at odds with the small effective population sizes estimated and suggest that population sizes were historically large but likely unstable.

[Supplemental material is available for this article.]

The ascomycete fungus *Mycosphaerella graminicola* is a major pathogen of wheat with a global distribution. *M. graminicola* originated in the Middle East ~11,000 yr ago from wild species infecting other grasses (Stukenbrock et al. 2007). Speciation was associated with adaptation to domesticated wheat and its associated agro-ecosystem. In a previous study, we characterized the two progenitor pathogen species *Mycosphaerella* S1 and *Mycosphaerella* S2 (hereafter called S1 and S2) endemic to the Middle East. S1 and S2 were isolated from three distantly related wild grass species: *Elymus repens*, *Dactylis glomerata*, and *Lolium multiflorum*. *M. graminicola* is mainly found on cultivated wheat but is able to infect other grass species (Seifbarghi et al. 2009). The recent divergence of *M. graminicola*, S1, and S2 implies that only a limited number of evolutionary changes are likely to have occurred. Here we use a population genomics approach to investigate the extent of genome divergence between *M. graminicola*, S1, and S2. We identify footprints of natural selection and study patterns of genome evolution. We compare the evolution of a pathogen species on a domesticated host with evolution on “wild” host species. In plants and animals, domestication

is often associated with genome-wide increased rates of nonsynonymous mutations resulting from a relaxation of purifying selection (Wang et al. 1999; Dobney and Larson 2006; Lu et al. 2006; Cruz et al. 2008). In studies of domesticated fungal species the trend is less clear. A comparison of laboratory and wild strains of the yeast fungus *Saccharomyces cerevisiae* showed an increase in nonsynonymous substitution rate in the domesticated species, indicating a relaxation of purifying selection (Gu et al. 2005). But in the fungus *Aspergillus oryzae*, also domesticated for use in fermentation of food products, no visible effects of domestication on the proteome were found compared with its wild sister lineage *Aspergillus flavus* (Rokas 2009). Besides the environmental change associated with domestication itself, other parameters such as population size, clonal versus sexual reproduction, and demography may explain the different outcomes from evolutionary studies of domesticated species. The effect of “domestication” of a parasite, i.e., coevolution with its host during domestication, has so far never been studied. A parasite on a domesticated host is also expected to experience strong directional selection generated by changes in the host. The closely related *Mycosphaerella* species allowed us to investigate the evolutionary consequence of coevolution and “domestication” of *M. graminicola* by comparing evolutionary rates in the agricultural and the wild species.

We hypothesize that sustained selection on wild hosts in S1 and S2 and adaptation to a new domesticated host in *M. graminicola* have been the primary drivers of the speciation process in these pathogens. Genes encoding secreted proteins potentially interact

<sup>5</sup>Present address: Max Planck Institute Marburg, Karl von Frisch Str. 10, 35043 Marburg, Germany.

<sup>6</sup>Corresponding author.

E-mail [eva.stukenbrock@mpi-marburg.mpg.de](mailto:eva.stukenbrock@mpi-marburg.mpg.de).

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.118851.110>.

with host molecules and are therefore subjected to heterogeneous selection pressures imposed by hosts. These genes may therefore evolve more rapidly than other genes during host specialization. Positive selection in secreted proteins has been found in several plant pathogens, supporting the hypothesis that the evolution of these proteins is driven by coevolutionary arms races with host molecules (Rohmer et al. 2004; Stukenbrock and McDonald 2007; Win et al. 2007; Barrett et al. 2009; Stukenbrock et al. 2010). Our genome data set allows us to assess the extent of adaptive evolution in proteins targeted for secretion and to compare this with the extent of adaptive evolution in protein-coding genes throughout the genome. We furthermore aimed at identifying positively selected genes as candidates playing a major role in speciation or host specialization.

In a previous study we investigated the evolutionary history of the *Mycosphaerella* pathogens by applying an “Isolation with Migration” coalescence model (Nielsen and Wakeley 2001) to a small sequence data set from a population sample of the species (Stukenbrock et al. 2007). We estimated recent speciation times and ancient migration rates between the different species. The evolutionary genomic approach we apply here also includes a coalescence analysis that infers parameters of the ancestral species and the process of species divergence (Hobolth et al. 2007; Dutheil et al. 2009). It has previously been applied to the great apes to estimate ancestral population sizes and speciation times as well as to infer the amount of incomplete lineage sorting that results in incongruence between the genealogy of some regions of the genome and the species phylogeny (Dutheil et al. 2009).

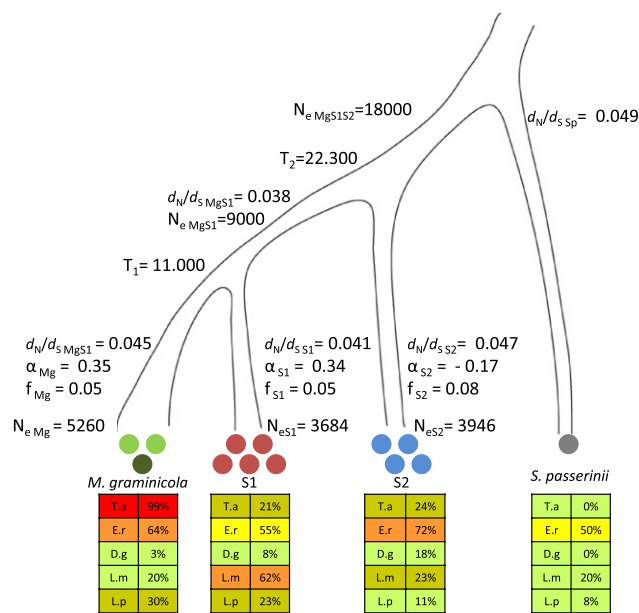
## Results and Discussion

### Host specialization among closely related pathogen species

A detached leaf assay was used to measure virulence on cultivated and wild hosts through the percentage of leaf area covered by fungal fruiting structures called pycnidia (Supplemental Fig. S1). Detached leaves of *Triticum aestivum* (cultivar Drifter), *E. repens*, *D. glomerata*, *L. perenne*, and *L. multiflorum*, corresponding to the original host genera of the three fungal species, were inoculated with a standardized spore suspension for each pathogen strain. We also included the more distantly related *Mycosphaerella* pathogen *Septoria passerinii* in the experiment. The four species are related as outlined in Figure 1. *M. graminicola* is more virulent on the wheat host ( $P = 0.001$ ) than the other species (Fig. 1, lower panel) but retained the ability to infect and propagate on other hosts (Table 1). S1 and S2 showed limited ability to infect wheat compared with *M. graminicola*. S1 is more virulent ( $P = 0.001$ ) on *L. multiflorum*, while S2 is more virulent on *D. glomerata* ( $P = 0.03$ ) (Fig. 1). We found a high variability in susceptibility among different plant individuals for each of the wild grasses, which likely is due to the fact that the seeds we used for testing originated from a genetically heterogeneous population. The variation for susceptibility among plants of the outcrossing wild grass hosts was considerably higher than among the wheat plants that all originated from one inbred cultivar. We conclude that these pathogens can show significant host adaptation even though they may coexist on the same host.

### Generation of a population genomic data set

To infer patterns of genome evolution in *M. graminicola*, S1, S2, and the barley pathogen *S. passerinii*, we sequenced the genomes of 12 fungal individuals (two *M. graminicola*, five S1, four S2, and one *S. passerinii*) (Fig. 1). The isolates of *M. graminicola*, S1, and S2, were



**Figure 1.** Species tree of *M. graminicola*, S1, S2, and *S. passerinii*. Circles indicate the number of genomes sequenced in each species. In our analyses we used the reference sequence of *M. graminicola* marked as a dark green circle. The lower panels of the figure summarize the ability of each species to infect the five grass species (T.a.) *Triticum aestivum*, (E.r.) *Elymus repens*, (D.g.) *Dactylis glomerata*, (L.m.) *Lolium multiflorum*, and (L.p.) *Lolium perenne*. The extent of host adaptation was estimated as percentage of leaf area covered by pycnidia among the different host species. Colors corresponding to the degree of infection are shown as 0%–20% (light green), 21%–40% (brownish green), 41%–60% (yellow), 61%–80% (orange), 81%–100% (red). We used a coalescence approach to estimate the split between *M. graminicola* and S1 ( $T_1$ ) and the more ancient split between S2 and the *M. graminicola*-S1 branch ( $T_2$ ). The population sizes of the ancestral species are indicated as  $N_{eMgS1}$  and  $N_{eMgS1S2}$  parameters. The evolutionary rate is estimated as the ratio of nonsynonymous to synonymous mutation rates  $d_N/d_S$  for each branch. Shown are also estimates of the parameter  $\alpha$  quantifying the extent of adaptive evolution and the parameter  $f$  as a measure of the strength of purifying selection in each of the branches of *M. graminicola*, S1, and S2 using *S. passerinii* as the outgroup species. Contemporary effective population sizes were estimated from polymorphism data for each species and are given as  $N_{eMg}$ ,  $N_{eS1}$ , and  $N_{eS2}$ .

all collected at the center of origin of the pathogen in the Middle East and were selected according to their host species, mating type locus, and previously determined genomic fingerprint (Supplemental Table S1; Stukenbrock et al. 2007). De novo assembly of paired-end Illumina reads resulted in  $>40\times$  coverage genome sequences (Supplemental Table S2). We compared contigs using the annotated reference genome of *M. graminicola* (Goodwin et al. 2011) as the backbone to construct a  $12 \times 28$  Mb alignment (Supplemental Table S3). The genome alignment was compared with electrophoretic karyotypes of four of the isolates to assess collinearity of chromosomes. The genomes show a high degree of synteny for the 13 essential chromosomes but a considerably lower degree of synteny among the smaller dispensable chromosomes (chromosomes 14–21 in the sequenced reference isolate) (Mehrabi et al. 2007; Wittenberg et al. 2009; Goodwin et al. 2011). Interestingly, we find that six and four of the small chromosomes are also dispensable in S1 and S2, respectively, as revealed by whole chromosome presence/absence polymorphisms (Supplemental Table S4). We conclude that presence/absence polymorphism of whole chromosomes is a trait shared among the three species,

**Table 1.** Percentage of leaf area covered by pycnidia on five different host species inoculated by isolates of *M. graminicola*, S1, S2, and *S. passerinii*

		<i>Triticum aestivum</i>	<i>Elymus repens</i>	<i>Dactylis glomerata</i>	<i>Lolium multiflorum</i>	<i>Lolium perenne</i>
<i>M. graminicola</i>	IPO 323	100 ± 0	100 ± 0	0 ± 0	28 ± 17	58 ± 51
	STIR01A26b	98 ± 5	25 ± 50	10 ± 20	13 ± 13	5 ± 10
S1	STIR01A48b	100 ± 0	68 ± 46	0 ± 0	20 ± 18	28 ± 25
	STIR04 3.11.1	45 ± 39	85 ± 30	20 ± 34	75 ± 13	8 ± 10
	STIR04 4.3.1	10 ± 14	20 ± 40	0 ± 0	63 ± 32	28 ± 22
	STIR04 2.2.1	10 ± 20	43 ± 42	3 ± 5	55 ± 40	30 ± 48
	STIR04 5.9.1	25 ± 26	30 ± 47	18 ± 29	48 ± 33	48 ± 40
S2	STIR04 5.3	15 ± 17	98 ± 5	0 ± 0	70 ± 36	0 ± 0
	STIR04 1.1.1	10 ± 8	63 ± 38	8 ± 5	18 ± 21	10 ± 8
	STIR04 1.1.2	13 ± 10	70 ± 35	3 ± 5	30 ± 14	8 ± 15
	STIR04 3.13.1	23 ± 15	90 ± 20	15 ± 13	10 ± 0	13 ± 13
	STIR04 3.3.2	50 ± 14	65 ± 44	45 ± 6	33 ± 39	15 ± 10
<i>S. passerinii</i>	<i>S. passerinii</i> P 63	0 ± 0	50 ± 38	0 ± 0	20 ± 23	8 ± 10

The experiment used a detached leaf assay. The percentages and their standard deviations (indicated with ±) are based on averages of four replicates for each isolate on each host.

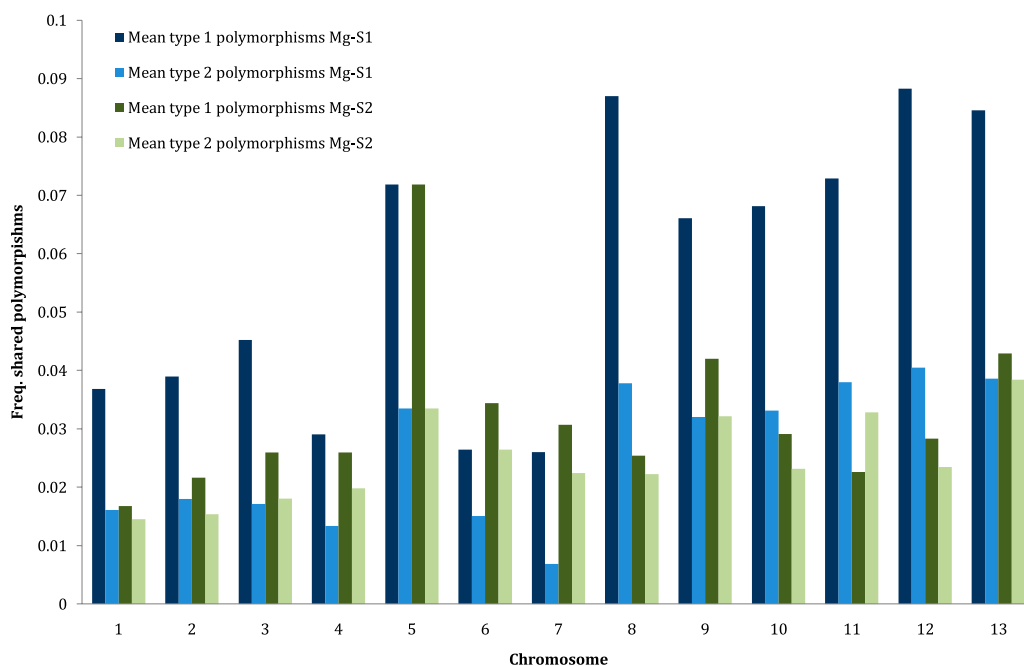
consistent with the hypothesis that dispensable chromosomes were already present in the ancestors of *M. graminicola*.

### Polymorphism and divergence in the *M. graminicola* species complex

The genomes of *M. graminicola* and S1 are on average 6% divergent at the nucleotide level, while *M. graminicola* and S2 are diverged by 10% (Supplemental Table S5). The amount of divergence (Dxy) and within-species nucleotide diversity ( $\pi$ ) increases with decreasing chromosome size (Supplemental Fig. S2). Average divergence between *M. graminicola* and S1 is only three to four times higher than the polymorphism within each species. Given this modest amount of divergence, we expect some polymorphisms to predate speciation

and thus to be shared between the two species. Between *M. graminicola* and S2 we do not expect polymorphism to be shared by descent to the same extent. Alternatively, shared polymorphisms may result from recurrent mutations, introgression, and/or balancing selection. We investigated shared polymorphisms by counting the proportion of shared polymorphic sites in 10-kb windows and distinguishing between sites segregating for the same polymorphism (type 1) and sites that actually contain different polymorphisms (type 2) (Supplemental Fig. S2; Supplemental Table S5).

We found higher amounts of shared polymorphisms between the two most closely related species (Supplemental Table S5), supporting the view that all shared polymorphisms cannot be derived by recurrent mutation. Furthermore, *M. graminicola* and S1 exhibit relatively more type 1 than type 2 shared sites, suggesting that a sizeable fraction of type 1 sites indeed are shared by descent (Fig. 2). The proportion of type 2 sites is similar between the *M. graminicola* and S1 and *M. graminicola* and S2 comparisons, suggesting that these most likely are caused by recurrent mutation. The increase in type 2 polymorphisms with decreasing chromosome size is again likely due to the higher amounts of polymorphisms in all species on smaller chromosomes. However, the excess of type 1 polymorphisms in *M. graminicola* and S1 shows a less clear correlation with chromosome sizes (Fig. 2). We hypothesize that changes in effective population size along the genomes can be important in determining how much polymorphism is shared between species (see also co-alescent analysis below).



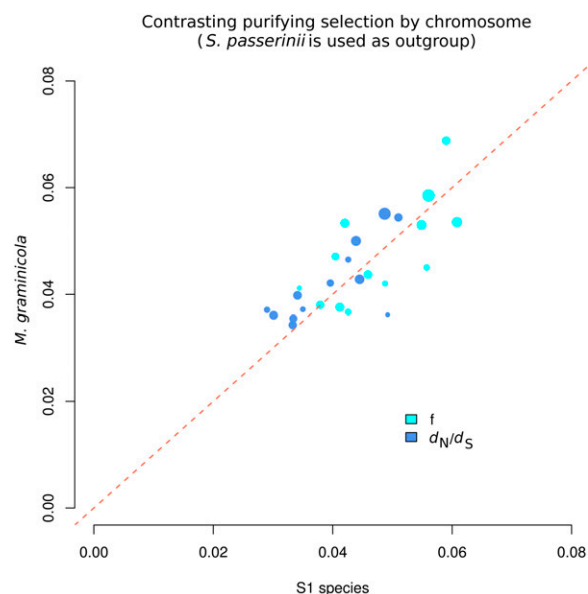
**Figure 2.** Proportion of polymorphic sites in *M. graminicola* that are also polymorphic in S1. Shared polymorphisms are distinguished as either type 1: same nucleotide variation between species; or type 2: different nucleotide polymorphisms between species.

Finally, some windows exhibit extremely high proportions of shared polymorphisms, which are best explained by balancing selection or recent introgression (Supplemental Fig. S3). Haplotype comparisons in these windows do not show increased similarity between species (data not shown) as expected if shared polymorphisms were originating from introgression. Balancing selection therefore remains as a likely hypothesis to be tested by functional studies of candidate regions.

### Footprints of natural selection during pathogen domestication

We investigated patterns of evolution and natural selection in the domesticated and wild species in two ways: (1) by a phylogenetic analysis of patterns of nonsynonymous and synonymous substitution rates between species (Yang 2007) from all aligned genes and (2) by contrasting amounts of synonymous and nonsynonymous substitutions between species with patterns of within-species polymorphism using a modified McDonald-Kreitman approach (Welch 2006). When referring to estimation of evolutionary rates obtained from branch-specific models according to Yang (Yang and Nielsen 2000) implemented in the CODEML program, we use the term  $d_N/d_S$ . For analyses based on the Nei and Gojobori method (Nei and Gojobori 1986; Nei and Kumar 2000) and the McDonald-Kreitman approach (McDonald and Kreitman 1991; Eyre-Walker 2006; Welch 2006), we use the terms  $K_a/K_s$  for the between-species ratio of nonsynonymous to synonymous substitution and  $P_N/P_S$  for the within-species ratio of nonsynonymous to synonymous mutations (for parameter definitions, see Supplemental Table S6).

We first estimated  $d_N/d_S$  on the four species phylogeny of *M. graminicola*, S1, S2, and *S. passerinii*. We find that purifying selection—as measured through the  $d_N/d_S$  ratio—is strong in the *Mycosphaerella* species complex with an average  $d_N/d_S$  of only 0.045. This value is considerably lower than values obtained from whole genome analyses of two other ascomycetes. In the domesticated *A. oryzae* species, Rokas estimated a  $d_N/d_S$  ratio of 0.45 and in the wild sister species *A. flavus* 0.51 (Rokas 2009). In laboratory and wild strains of *S. cerevisiae*,  $d_N/d_S$  ratios of 0.20 and 0.10, respectively, were reported (Gu et al. 2005). The branch-specific  $d_N/d_S$  is significantly larger on the terminal *M. graminicola* branch compared with its closest relative S1 ( $d_N/d_S = 0.041$ ; paired Wilcoxon test of 88,495 aligned codons on the 13 essential chromosomes,  $p = 3.7 \times 10^{-5}$ ) (Fig. 3, dark blue circles). Interestingly, when comparing the “wild” pathogen species S1 and S2 and using *S. passerinii* as outgroup, we find that S2 has a higher  $d_N/d_S$  ratio than both *M. graminicola* and S1 (Fig. 1). To interpret the meaning of these differences, we note that  $d_N/d_S$  can increase either due to relaxed purifying selection or increased positive selection as both processes inflate the rate of nonsynonymous mutations. The relative roles of positive and purifying selection can be disentangled by analyzing jointly the counts of polymorphisms and divergence in coding regions with an approach developed by Eyre-Walker (2006) and extended by Welch (2006). We pooled the McDonald-Kreitman tables (McDonald and Kreitman 1991) of aligned genes for each chromosome and found that purifying selection in *M. graminicola* and S1 was very strong: ~95% of nonsynonymous mutations were selected against, while in S2 the proportion was 92% (Fig. 1). *S. passerinii* was consistently used as outgroup in the analyses of adaptive and purifying selection to account for the high amount of incomplete lineage sorting between *M. graminicola*, S1, and S2 (see below). For individual chromosomes, there is no clear trend for differential amounts of purifying selection between *M. graminicola* and S1,



**Figure 3.** Contrasting evolutionary rates on the *M. graminicola* and S1 branches of the phylogeny. Plotted is  $d_N/d_S$  in *M. graminicola* versus S1 (blue dots) and the parameter “ $f$ ” (green dots) showing the fraction of nonsynonymous mutations experiencing strong purifying selection as a measure of purifying selection in the two species. Each dot represents the average measure of a chromosome, and the size of the dot corresponds to the relative sizes of the different chromosomes (on a log scale). While  $d_N/d_S$  is significantly higher in *M. graminicola*, the strength of purifying selection is highly similar in the two species. *S. passerinii* is used as the outgroup in both analyses.

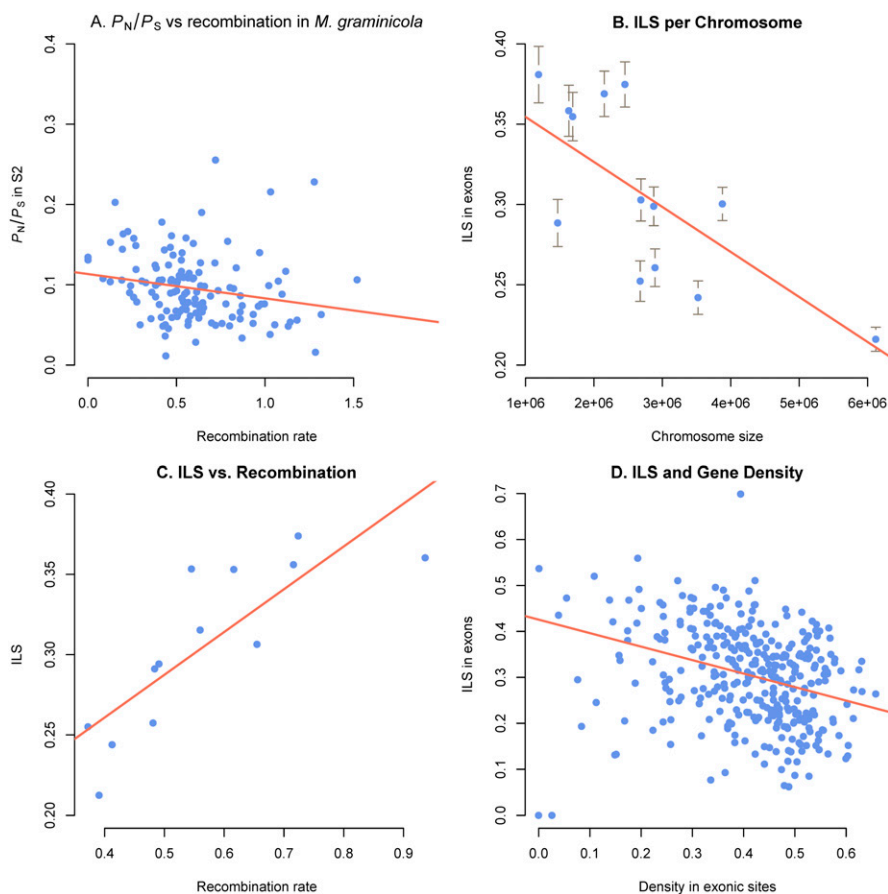
suggesting that deleterious mutations were purged with the same efficacy on the different chromosomes in these two species (Fig. 3, light blue circles).

The proportion, alpha ( $\alpha$ ), of the remaining amino acid changing mutations that were fixed by positive selection is overall marginally higher in *M. graminicola* compared with S1 and negative (although not significantly different from zero) in S2. In *M. graminicola* ~35% of the nonsynonymous mutations were fixed by positive selection, while signals of adaptive evolution were absent in S2. This also further argues that the increased  $d_N/d_S$  of the *M. graminicola* branch is driven at least partly by adaptive evolution.

Next, we correlated the ratio of amounts of nonsynonymous to synonymous polymorphism ( $P_N/P_S$ ) in *M. graminicola* to recombination rates across chromosomes. Recombination rate estimates were computed in adjacent 100-kb windows using a coalescence analysis described in detail below. We found a weak but significant negative correlation between  $P_N/P_S$  and recombination rate in *M. graminicola* (Kendall's rank correlation,  $\tau = -0.186$ ,  $P = 0.0012$ ) (Fig. 4A). The negative correlation may indicate a greater efficacy of selection in the *M. graminicola* genome compared with the other two species and again argues against a relaxation of purifying selection during domestication in *M. graminicola*. Taken together, the different analyses all support that the increase in  $d_N/d_S$  on the *M. graminicola* branch compared with its wild sister species results from higher amounts of adaptive evolution rather than mere accumulation of deleterious mutations.

### Positive selection in signal peptides

The genome alignment of *M. graminicola*–S1–S2 included more than 9000 protein-coding genes out of the 10,952 genes predicted



**Figure 4.** (A) Plot showing the negative correlation between  $P_N/P_S$  in *M. graminicola* and recombination rate. Both parameters were estimated as averages in windows of 100 kb (160 windows). One outlier window with an average value of  $P_N/P_S = 0.5$  was excluded from the plot and regression analyses. (B) The frequency of incomplete lineage sorting (ILS) in exons plotted against chromosome size (chromosome 1–13) measured in base pairs. The strong negative correlation demonstrates a higher extent of ILS on smaller chromosomes reflecting the higher recombination rate on the smaller chromosomes. (C) Plot illustrating the positive correlation between the frequency of ILS on each essential chromosome (1–13) and average recombination rates for the same chromosomes. (D) The frequency of ILS in exonic sites plotted against exon density in 100-kb windows. The negative correlation illustrates that regions with a higher density of coding sites show less ILS compared to regions with fewer coding sites.

in *M. graminicola*. We measured  $K_a$  and  $K_s$  to identify genes under positive selection. Only 27 genes show significantly higher  $K_a$  over  $K_s$  ( $P < 0.05$ , Z-test). None of these genes encode proteins with known function, but 11 of them were found to be positively selected in more than one species, supporting a putative role in divergence and/or host specialization (Fig. 5A).

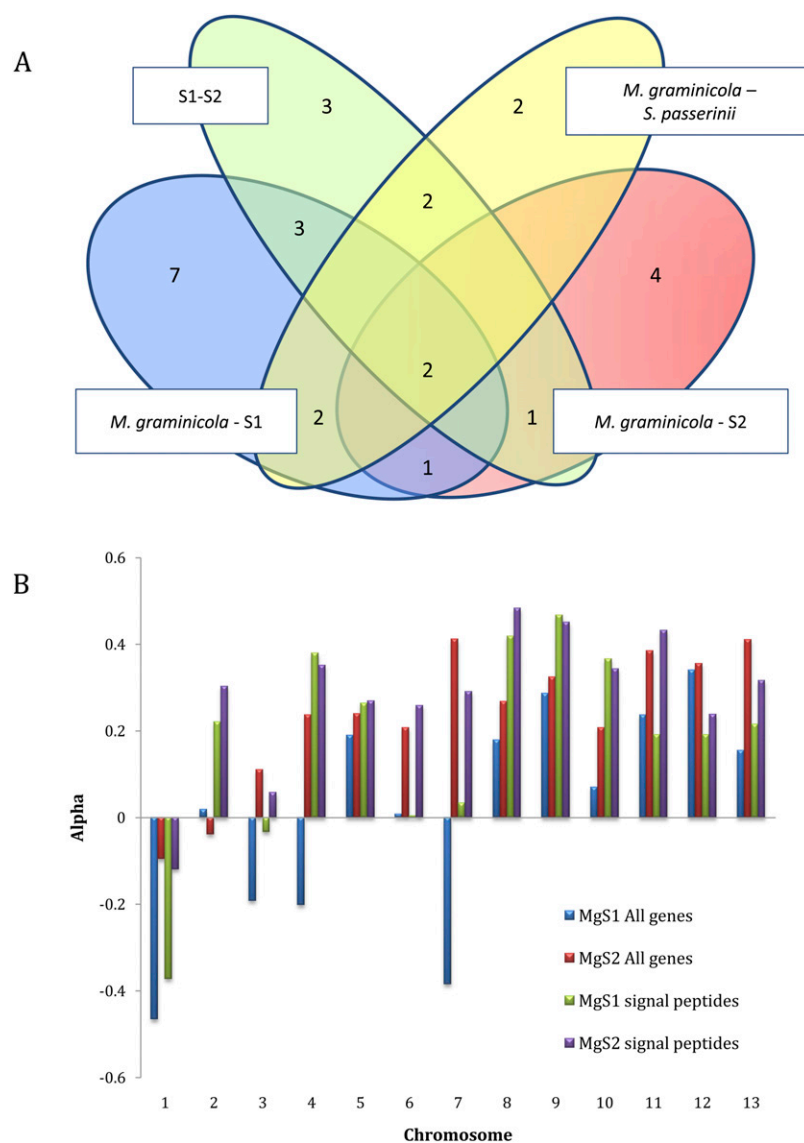
Pathogen effectors are secreted proteins that act to modify the plant immune system and thereby allow further access of the fungus inside the host (Kamoun 2007). We asked whether putative effectors exhibit more positively selected nonsynonymous mutations driven by adaptation to different host targets. Eight hundred two proteins targeted for secretion by the presence of a signal peptide were computationally identified (Kamoun 2007), and the proportion of nonsynonymous divergence attributable to positive selection,  $\alpha$ , was assessed (Welch 2006). Overall, genes encoding secreted proteins exhibited a significantly higher amount of adaptive evolution (Fig. 5B; Supplemental Fig. S4), supporting the hypothesis of a coevolutionary arms race between host and pathogen molecules. This difference remained significant after correcting for difference in base composition between the two

sets of genes (comparing *M. graminicola* and S1:  $P$ -value  $< 0.001$  and comparing *M. graminicola* and S2:  $P$ -value  $< 0.01$ ) (Supplemental Table S7). We additionally observed strong differences in the magnitude of  $\alpha$  between chromosomes that likely reflect the differences in recombination rates and effective population sizes of the chromosomes (Fig. 6).

We classified aligned genes using the Gene Ontology (GO) browser on the JGI *M. graminicola* genome portal. GO categories could be assigned to 1406 (15%) of the genes aligned between *M. graminicola*, S1, S2, and *S. passerinii* (Supplemental Table S8). We compared  $\alpha$  and  $f$  values for the 27 most abundant GO categories and found significant support for heterogeneity of  $f$  and  $\alpha$  among GO categories (permutation test; see Methods). The highest value of  $\alpha$  was found between *M. graminicola* and S1 in the GO categories “amino acid transport” ( $\alpha = 0.61$ , 65 genes assigned), “response to antibiotics” ( $\alpha = 0.22$ , 22 genes), and “carbohydrate transport” ( $\alpha = 0.20$ , 27 genes). Between *M. graminicola* and S2 the highest  $\alpha$  values were likewise the GO categories “amino acid transport” ( $\alpha = 0.59$ ) and “response to antibiotics” ( $\alpha = 0.32$ ). Another GO category significantly affected by selection in *M. graminicola* but not in S1 and S2 is “signal transduction” ( $\alpha = 0.16$  for *M. graminicola* and S1 and  $\alpha = -0.04$  for S1 and S2), suggesting that changes in regulatory pathways have driven the divergence and specialization of *M. graminicola*.

#### Reproductive isolation through significant alteration of a mating-type locus

Although the detached leaf assay indicated that coinfection of the same host species can occur, we previously found no evidence of contemporary gene exchange between species based on gene flow analyses of large-scale population samples of the pathogens (Stukenbrock et al. 2007). Experimental crosses of *M. graminicola* and S1 were unsuccessful (G Kema, pers. comm.), further supporting the assumption that these species are reproductively isolated even though they exist in sympatry. However, the genome data reveal islands of shared polymorphisms that between *M. graminicola* and S1 could result from introgression. Reproductive barriers in ascomycetes can evolve by divergent selection on genes involved in the recognition of self and non-self. Several genes are involved in mating, but we focused on the two major idiomorphs of *M. graminicola* *MAT1-1* and *MAT1-2* (Waalwijk et al. 2002). We assessed the level of polymorphisms and substitutions in *MAT1-1* and *MAT1-2* of *M. graminicola*, S1, and S2 (Supplemental Table S9). The two idiomorph genes are highly conserved within species and contained no polymorphism in a global population sample including more than 70 isolates (data not shown). However, between species, the *MAT1-2* locus has undergone several amino acid



**Figure 5.** (A) Venn diagram demonstrating the number of genes showing evidence of positive selection between four different combinations of species. As illustrated, several genes showed a significantly higher  $K_a$  value in more than one pairwise species combination. (B) The parameter  $\alpha$  estimated for all genes and for a subset of genes encoding signal peptides. The figure contrasts estimates of the two groups of genes on chromosomes 1–13 in the two combinations of species *M. graminicola* and S1 (MgS1) and *M. graminicola* and S2 (MgS2). Although the chromosomes show a high extent of heterogeneity, the signal peptide encoding genes overall have a significantly higher extent of adaptive evolution.

changes between *M. graminicola* and S1 and a change in start position between *M. graminicola* and S2. The *MAT1-1* idiomorph, on the other hand, has accumulated fewer (5 vs. 22 between *M. graminicola* and S1 and 30 vs. 50 between *M. graminicola* and S2) amino acid changes. Altogether this suggests that reproductive isolation between the three species could have been mediated in part through selection on one of the major genes involved in mating.

#### Recent species divergence and extensive incomplete lineage sorting in pathogen genomes

We applied a coalescent hidden Markov model (coalHMM) (Hobolth et al. 2007; Dutheil et al. 2009) estimating and exploiting

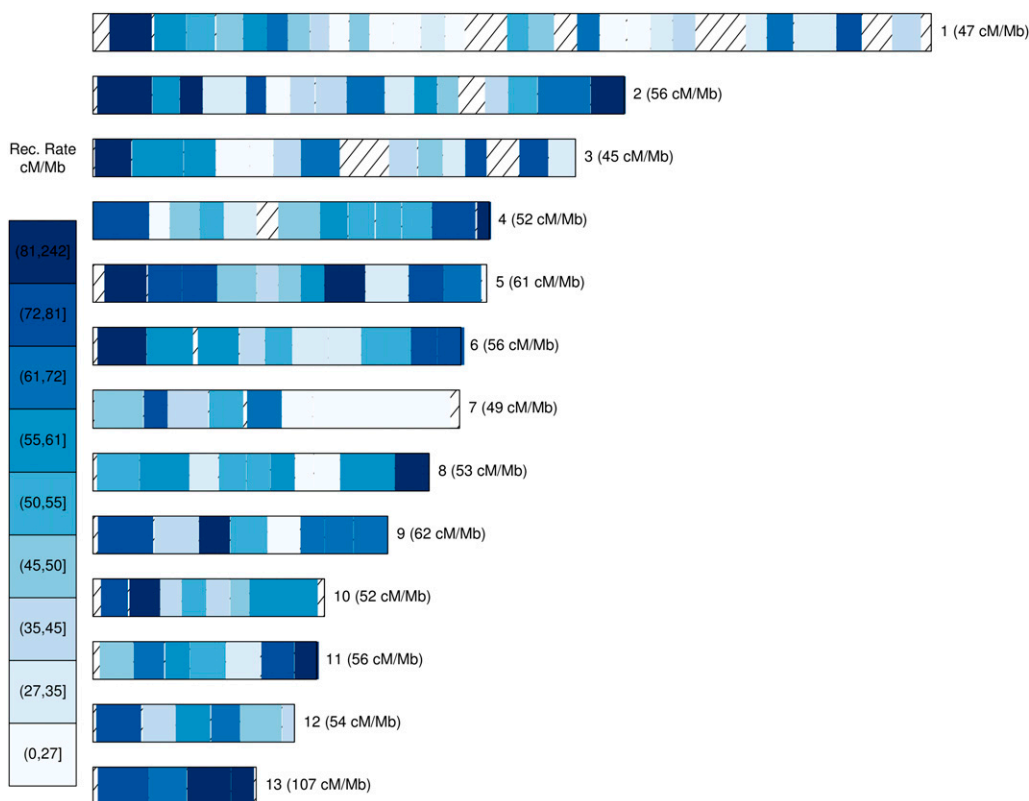
incomplete lineage sorting (ILS) between *M. graminicola*, S1, and S2 to assess speciation times, ancestral population sizes, and recombination rates. All parameters were estimated in units of substitutions, and in order to convert them to years and actual population sizes, we used the previously estimated speciation time of 11,000 yr as calibration (Stukenbrock et al. 2007). This assumption corresponds to a mutation rate per site per cell cycle of  $3.3 \times 10^{-8}$ . This assumed mutation rate matches well the direct estimate of mutation rate from yeast (Lynch et al. 2008). It is, however, important to acknowledge the possible variation in cell cycles per year of the grass pathogens existing in highly different environments, which, in addition, have been changing since species divergence.

The average recombination rate in the ancestral species of *M. graminicola* and S1 was estimated to be on average 46 cM/Mb (Fig. 6; Supplemental Table S10). Using 100-kb stretches of aligned sequence, we constructed a recombination map of the ancestral species that shows great heterogeneity in recombination rates (Fig. 6) but with higher recombination rates near the telomeres and higher recombination rates on smaller chromosomes, matching previous estimates from a genetic map of *M. graminicola* (Wittenberg 2007). We note that this is the first broad-scale recombination map reconstructed without any crossing data.

The coalHMM analysis provided a divergence time for the S2 species of ~23,000 yr. We estimate the effective population size of the S2 ancestor to be ~18,000, while the population size of the ancestor of *M. graminicola* and S1 was only 9000. Using the same mutation rate, we estimate from contemporary amounts of polymorphism (SNPs in the three genomes of *M. graminicola*, in the five of S1, and in the four S2 genomes) that the current effective population size is ~5000 for *M. graminicola* and ~3500–4000 for S1 and S2 if one sexual cycle per year occurs (Fig. 1).

If there are more sexual cycles per year, then these estimates should be multiplied by the number of sexual cycles per year. We place most confidence in the relative sizes and differences between the three species because the actual sizes as explained above rely on several assumptions for calibration. Our previous work also showed that *M. graminicola* has increased in population size while spreading around the world together with its wheat host (Banke and McDonald 2005; Stukenbrock et al. 2007). In *M. graminicola* Zhan and McDonald (2004) previously estimated an  $N_e$  of 25,000 using information from RFLP markers. This number is within the same range as the  $N_e$  values estimated here for the ancestral *Mycosphaerella* species (Fig. 1).

Interestingly, the small effective size estimates appear incompatible with the very strong selection observed, which predicts



**Figure 6.** Recombination map of chromosomes 1–13 derived from coalHMM estimates of the parameter  $\rho$  measured in units of substitutions.  $\rho$  was estimated in windows of 100 kb along aligned chromosomes of four isolates: the *M. graminicola* reference genome, one genome of S1 and S2, and the outgroup sequence of *S. passerinii*. We converted  $\rho$  to recombination rate measured in centimorgans per megabase (cM/Mb). Color differences indicate differences in recombination rate across chromosomes. The nine categories correspond to the nine quartiles of the observed distribution of recombination rate. Chromosome-wise values shown in parentheses are the median of all corresponding regions. Stripped regions show regions where  $\rho$  could not be estimated due to insufficient alignment data.

that the selection coefficient  $[N_e * s] \gg 10$  for 96%–97% of the genome. In *Drosophila melanogaster*, which exhibits a similar strength of purifying selection, effective sizes based on contemporary polymorphisms are estimated in the millions. Yet it has recently been proposed that even these estimates of  $N_e$  are smaller than the  $N$  relevant for natural selection due to fluctuations in population size (Karasov et al. 2010). We propose that fluctuations in population sizes must be much more dramatic in the *M. graminicola* species complex.

The pathogen genomes consist of a mosaic of sequences with different evolutionary histories. Thirty percent of the sites in the genome alignment are not concordant with the species tree and instead show either closer genomic relationship between *M. graminicola* and S2 or S1 and S2 as an effect of incomplete lineage sorting (ILS). The distribution of sites with ILS is nonrandom (Fig. 4B): We find a strong positive correlation between recombination rate and ILS (Kendall's rank correlation,  $\tau = 0.7$ ,  $P < 4 \times 10^5$ ) (Fig. 4C). We interpret this as evidence for less genetic hitchhiking and less background selection on the more highly recombining chromosomes. This also provides a partial explanation for the higher rate of adaptive evolution observed on the smaller chromosomes (Fig. 5B) since the efficacy of selection is expected to be larger where the recombination rate is higher. A more direct way to investigate the effect of selection on patterns of ILS is to correlate the density of genes (as proxies for intensity of selection) with ILS for each 100-kb segment. Figure 4D shows that the

amount of ILS observed is strongly negatively correlated with exon density (Kendall's  $\tau = -0.22$ ,  $P < 2.2 \times 10^{16}$ ). It is not possible from this observation to determine whether this is due to more background selection or more selective sweeps in more gene-rich regions, but the selection analysis suggests that it might well be a mixture of selection against 95% of amino acid changing mutations and selection for (sweeps of) 2% of new amino acid-changing mutations (the remainder being effectively neutral).

The ILS correlations all suggest that natural selection (positive and negative) affects linked variation in these species despite recombination rates about 50 times larger than for the human genome. We can roughly estimate that selective sweeps constantly occur throughout the pathogen genomes: We estimate fixation of one new amino acid substitution by positive selection happening approximately once a year since *M. graminicola* and S1 diverged in the following way. The number of amino acid differences between the species is 188,000, of which ~40% or 75,200 have been fixed by positive selection. The average genomic divergence time is 29,000 yr (11,000 yr of speciation plus coalescence in the common ancestor of expected  $2 \times 9000 [N_e]$  ancestor of *M. graminicola* and S1] = 18,000 yr). So 75,200 selective sweeps have occurred over 58,000 yr. This rate of adaptive evolution in the *Mycosphaerella* pathogens exceeds estimates from whole genome studies in *Drosophila* by a factor of 20 (Andolfatto 2005). However, a recent study also quantifying adaptive evolution by  $\alpha$  has found that adaptive

protein divergence on the X chromosome reaches 85%–90% in *Drosophila* species (Andolfatto et al. 2011).

## Conclusion

Using a population genomics approach, we demonstrate increased adaptive evolution in *M. graminicola* compared with S1 and S2. Differences in patterns of evolution between S1 and S2 demonstrate that natural populations of plant pathogens become specialized on different hosts by very different processes.

Positive selection has in particular affected genes encoding secreted proteins that likely interact with host molecules and potentially act as pathogen effectors. The pattern of overall strong positive selection on these genes is likely mediated by a coevolutionary arms race with host molecules.

The amount of incomplete lineage sorting in the genomes of *M. graminicola*, S1, and S2 provides evidence that genome divergence dates much further back than species divergence. Our analyses—based both on contemporary levels of polymorphism and patterns of ILS—are consistent with rather small present and ancestral effective population sizes of the species (Fig. 1). However, estimates of selective constraints point to fairly intense purifying selection and thus rather large (long-term effective) population sizes. Our study strongly supports the idea that the actual population size of *M. graminicola* must have been very high during the process of adaptation to a new domesticated host. Specialization on domesticated wheat did not entail any visible “domestication cost” that is otherwise often found in crop or livestock lineages (Lu et al. 2006; Haudry et al. 2007).

## Methods

### Host inoculation

Plants were obtained from seeds of the bread wheat variety Drifter that is highly susceptible to *M. graminicola* and from seeds of *E. repens* and *D. glomerata* provided by Otto Hauenstein Samen AG, Rafz Switzerland; *D. glomerata*, *L. perenne*, and *L. multiflorum* were provided by Fenaco Saemeraizentrum, Winterthur, Switzerland. To study host adaptation, we used a detached leaf assay (Browne and Cooke 2004) with some modifications. Host leaf material was produced in a greenhouse. Each fungal strain was inoculated onto detached leaves taken from individual plants using a spore concentration of  $10^7$  per milliliter. Cotton swabs saturated with the spore solution were wiped across 5.5-cm leaf segments. Inoculated leaves were air-dried for 10 min and placed into a plastic box containing agar amended with benzimidazole. Sealed boxes were placed in the dark for 24 h at 18°C and subsequently placed in growth chambers at 70% humidity under alternating light/dark cycles with 16 h under white fluorescent lights (15 kLux) at 18°C followed by 8 h of darkness at 15°C. The percentage of leaf area covered by pycnidia (asexual fruiting bodies produced by the *Mycosphaerella* pathogens on leaf surfaces) at 22 d after inoculation was used as a measure of pathogenicity. The assay was repeated four times for each fungal isolate on each of the plant hosts. The significance of differences among means was evaluated using *t*-tests.

### Electrophoretic karyotypes

Pulsed field gel electrophoresis was used to determine chromosome numbers and sizes in four of the resequenced isolates. DNA plugs were prepared by embedding intact spores in agarose and treating with EDTA, SDS, and protease XIV (Sigma-Aldrich) for 48 h at 60°C. PFGE was conducted with a contour-clamped homoge-

neous electric field (CHEF) Bio-Rad DR-II apparatus in 120 mL of  $1 \times$  TAE, 0.8% to 1% agarose (Invitrogen) with different running conditions for large, medium, and small chromosomes (Stukenbrock et al. 2010). The standard chromosome size markers used were from *S. cerevisiae* (small chromosomes), *Hansenula wingei* (medium chromosomes), and *Schizosaccharomyces pombe* (large chromosomes). Gels were stained with ethidium bromide for 20 min, and karyotypes were made visible using a UV transilluminator. Chromosome bands were identified with visual inspection, and gel pictures were further analyzed using the software QuantityOne (Bio-Rad).

### Genome assembly and alignment

Fungal genomes were sequenced by Illumina sequencing. Paired-end reads of 35–70 bp with insert size 150–500 nt were assembled using the SOAP de novo assembler (Li et al. 2008). Assembled contigs were aligned to the 21 annotated chromosomes of the JGI reference sequence of *M. graminicola* using LastZ included in the genome alignment software package BLASTZ (Schwartz et al. 2003). Alignment parameters were set as previously described in Stukenbrock et al. (2010). We merged the aligned chromosomes into a multiple genome alignment of 12 sequences using the reference sequence and coordinates as backbone. The 12 genome sequences can be accessed at the NCBI database (see “Data Access” below).

### Population genomic analyses

We developed a Java-based genome browser to extract counts of the chromosome alignments (available from <http://cs.au.dk/~tth/viewer/>). We recorded total counts of substitutions, polymorphisms, and shared polymorphisms in windows of 50 kb, 25 kb, and 10 kb according to the parameter of interest and the chromosome sizes. For inferences of nonsynonymous and synonymous substitutions, we used coordinate information from the JGI gene prediction in the reference sequence.

Evolutionary rates of the *M. graminicola*, S1, and S2 lineages were estimated by CODEML in the PAML software package using a branch model that allows for different  $d_N/d_S$  ratios (Yang and Nielsen 2000; Yang 2007). The parameters  $\alpha$  and  $f$  were estimated from counts of polymorphism and divergence in synonymous and nonsynonymous categories using the likelihood framework implemented in the MK-Test software (Welch 2006; Obbard et al. 2009). We built a series of models, ranging from models assuming strict neutrality, where  $f$  and  $\alpha$  are constrained to, respectively, 1 and 0, to models containing purifying or/and positive selection with varying intensities according to gene categories (or chromosomes). Given the sampling variance associated with these parameters (in particular  $\alpha$ ), we pooled MK tables by categories (gene categories or chromosomes). *S. passerinii* was used as outgroup for all analyses due to the high estimates of incomplete lineage sorting between *M. graminicola*, S1, and S2 obtained by the coalHMM approach. To compare the fit of these alternative models, we computed Akaike’s Information Criterion (AIC) of each model  $M_i$  as  $AIC(M_i) = -2 \ln L(M_i) + 2p(M_i)$ , where  $\ln L(M_i)$  is the (natural) log likelihood of Model  $M_i$  and  $p(M_i)$  the number of free parameters fitted. Previous work based on simulations suggests that AIC is a sensible choice for model selection in that context (Welch 2006). We evaluated 200 distinct models obtained by permutation of genes between categories. The resulting distribution of AIC allowed us to evaluate the significance of the partition of genes based on GO categories.

We estimated both parameters for pairs of species for the whole set of aligned genes and for a subset of genes encoding signal peptides. Signal peptides were identified using the software SignalP

(Emanuelsson et al. 2007). Only genes with a hidden Markov model score >90 were assigned as signal peptides. We compared  $\alpha$  and  $f$  estimates for signal peptides and nonsignal peptide genes. As genes with signal peptides appeared to have a significantly higher GC content, we performed an additional comparison by creating a subset of genes without signal peptides in which the GC content was within the range of “meanSP  $\pm$  2 stderr” (meanSP is the average GC and stderr the standard deviation of genes with signal peptide).

To identify single genes with evidence of positive selection, we used the Nei and Gojobori approach (Nei and Gojobori 1986) and assessed  $K_a$  and  $K_s$  rates for all genes with more than 100 aligned codons and used a Z test to identify genes with  $K_a > K_s$ .

### Mating-type sequences

Partial mating-type genes were amplified in a population sample of more than 70 isolates including *M. graminicola* collected from different geographical regions and S1 and S2 isolates collected in Iran. PCR conditions and sequencing were as previously described (Zhan et al. 2002; Banke et al. 2004). Primers are listed in Supplemental Table S11. The program DnaSP v.5 was used to estimate nucleotide differences between the sequences (Rozas et al. 2003).

### Coalescence analyses

The coalHMM approach used to infer coalescence parameters is described in Hobolth et al. (2007) and Dutheil et al. (2009) (Supplemental Fig. S5). We used four genome sequences for the analyses including the *M. graminicola* reference sequence, the S1 sequence of STIR04.3.11.1, the S2 sequence of STIR04.1.1.1, and the *S. passerinii* sequence of P63 (Supplemental Table S2). Analyses were initially performed on full chromosomes, on 200-kb aligned fragments, and on 100-kb fragments. All runs gave consistent parameter estimates.

We initially generated non-overlapping 1-kb windows along each chromosome. Windows in which at least one sequence had >40% gaps were discarded. The contiguous remaining windows were concatenated into alignment chunks and given as input to the coalHMM software. One set of parameters was estimated per alignment, but the Markov chain was reset after each alignment chunk. Next, we used the coalhmm program with the ILS model (2009 version) (Dutheil et al. 2009) to estimate the speciation times, ancestral population sizes, and recombination rate for each chromosome. Convergence was reached for all 13 essential chromosomes. We used a posterior decoding procedure to reconstruct the most likely genealogy for each position in the genome alignment. The amount of ILS for a given region was then computed as the proportion of sites in the region for which states Mg-S2 or S1-S2 (the states are illustrated in Supplemental Fig. S5) had the maximum posterior probability.

### Data access

The genome project has been deposited at DDBJ/EMBL GenBank under the following accession numbers: STIR04 A48b: AEYQ00000000, STIR04 A26b: AFIP00000000, STIR04 2.2.1: AFIQ00000000, STIR04 3.1.1: AFIO00000000, STIR04 4.3.1: AFIR00000000, STIR04 5.3: AFIS00000000, STIR04 5.9.1: AFIT00000000, STIR04 1.1.1: AFIU00000000, STIR04 1.1.2: AFIV00000000, STIR04 3.13.1: AFIW00000000, STIR04 3.3.2: AFIX00000000, *S. passerinii* P63: AFY00000000. The versions described in this paper are the first versions of the assembled genome data.

### Acknowledgments

We thank Freddy Bugge Christiansen and three anonymous reviewers for helpful comments to the manuscript, Gert Kema for testing the ability of *M. graminicola* and S1 to cross, and Tryggvi Stefansson for analyzing virulence data. The genome sequencing was financed by a research grant from the Faculty of Sciences, University of Copenhagen, Denmark and by a “L’Oreal Denmark for Women in Science Fellowship,” both to E.H.S. The research was additionally funded by the Danish Research Council through a postdoctoral grant to E.H.S. and an ERA-PG grant to M.H.S. The ETH Genetic Diversity Center was used to collect the *MAT1-1* and *MAT1-2* sequence data. The *M. graminicola* assembly used as reference in our study was produced by the U.S. Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community.

### References

- Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- Andolfatto P, Wong KM, Bachtrog D. 2011. Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species. *Genome Biol Evol* **3**: 114–128.
- Banke S, McDonald BA. 2005. Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. *Mol Ecol* **14**: 1881–1896.
- Banke S, Peschon A, McDonald BA. 2004. Phylogenetic analysis of globally distributed *Mycosphaerella graminicola* populations based on three DNA sequence loci. *Fungal Genet Biol* **41**: 226–238.
- Barrett LG, Thrall PH, Dodds PN, van der Merwe M, Linde CC, Lawrence GJ, Burdon JJ. 2009. Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. *Mol Biol Evol* **26**: 2499–2513.
- Browne RA, Cooke BM. 2004. Development and evaluation of an in vitro detached leaf assay for pre-screening resistance to *Fusarium* head blight in wheat. *Eur J Plant Pathol* **110**: 91–102.
- Cruz F, Vilá C, Webster MT. 2008. The legacy of domestication: Accumulation of deleterious mutations in the dog genome. *Mol Biol Evol* **25**: 2331–2336.
- Dobney K, Larson G. 2006. Genetics and animal domestication: new windows on an elusive process. *J Zool (Lond)* **269**: 261–271.
- Dutheil JY, Ganapathy G, Hobolth A, Mailund T, Uyenyama MK, Schierup MH. 2009. Ancestral population genomics: The coalescent hidden Markov model approach. *Genetics* **183**: 259–274.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953–971.
- Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol* **21**: 569–575.
- Goodwin SB, Ben M’Barek S, Dhillon B, Wittenberg A, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet* **7**: e1002070. doi: 10.1371/journal.pgen.1002070.
- Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM. 2005. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **102**: 1092–1097.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, et al. 2007. Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Mol Biol Evol* **24**: 1506–1517.
- Hobolth A, Christensen OF, Mailund T, Schierup MH. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* **7**. doi: 10.1371/journal.pgen.0030007.
- Kamoun S. 2007. Groovy times: filamentous pathogen effectors revealed. *Curr Opin Plant Biol* **10**: 358–365.
- Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet* **6**: e1000924. doi: 10.1371/journal.pgen.1000924.
- Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713–714.
- Lu J, Tang T, Tang H, Huang J, Shi S, Wu C-I. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet* **22**: 126–131.
- Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of

- the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci* **105**: 9272–9277.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**: 652–654.
- Mehrabi R, Taga M, Kema GHJ. 2007. Electrophoretic and cytological karyotyping of the foliar wheat pathogen *Mycosphaerella graminicola* reveals many chromosomes with a large size range. *Mycologia* **99**: 868–876.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.
- Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet* **5**: e1000698. doi: 10.1371/journal.pgen.1000698.
- Rohmer L, Guttman DS, Dangl JL. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. *Genetics* **167**: 1341–1360.
- Rokas A. 2009. The effect of domestication on the fungal proteome. *Trends Genet* **25**: 60–63.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human–mouse alignments with BLASTZ. *Genome Res* **13**: 103–107.
- Seifbarghi S, Razavi M, Aminian H, Zare REH. 2009. Studies on the host range of *Septoria* species on cereals and some wild grasses in Iran. *Phytopathol Mediterr* **48**: 422–429.
- Stukenbrock EH, McDonald BA. 2007. Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. *Mol Plant Pathol* **8**: 321–332.
- Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. 2007. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol Biol Evol* **24**: 398–411.
- Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. 2010. Whole genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. *PLoS Genet* **6**: e1001189. doi: 10.1371/journal.pgen.1001189.
- Waalwijk C, Mendes O, Verstappen ECP, de Waard MA, Kema GHJ. 2002. Isolation and characterization of the mating-type idiomorphs from the wheat *Septoria* leaf blotch fungus *Mycosphaerella graminicola*. *Fungal Genet Biol* **35**: 277–286.
- Wang R-L, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236–239.
- Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* **173**: 821–837.
- Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ, Kamoun S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**: 2349–2369.
- Wittenberg AHJ. 2007. “Genetic mapping using the Diversity Arrays Technology (DArT).” PhD thesis, Wageningen University, Wageningen, The Netherlands.
- Wittenberg AHJ, van der Lee TAJ, Ben M'Barek S, Ware SB, Goodwin SB, Kilian A, Visser RGF, Kema GHJ, Schouten HJ. 2009. Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. *PLoS ONE* **4**: e5863. doi: 10.1371/journal.pone.0005863.
- Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**: 32–43.
- Zhan J, McDonald BA. 2004. The interaction among evolutionary forces in the pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genet Biol* **41**: 590–599.
- Zhan J, Kema GHJ, Waalwijk C, McDonald BA. 2002. Distribution of mating type alleles in the wheat pathogen *Mycosphaerella graminicola* over spatial scales from lesions to continents. *Fungal Genet Biol* **36**: 128–136.

Received December 6, 2010; accepted in revised form August 10, 2011.