Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species

Eva Holtgrewe Stukenbrock[a,b,1], Freddy Bugge Christiansen[b], Troels Toftebjerg Hansen[b], Julien Yann Dutheil[a,c], and Mikkel Heide Schierup[b]

[a]Fungal Biodiversity Group, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany; [b]Bioinformatics Research Center, Aarhus University, 8000 Aarhus C, Denmark; and [c]Institute for Evolutionary Sciences, Centre National de la Recherche Scientifique, Universite Montpellier, 34095 Montpellier, France

Edited by Michael L. Arnold, University of Georgia, Athens, GA, and accepted by the Editorial Board May 22, 2012 (received for review January 26, 2012)

In a genome alignment of five individuals of the ascomycete fungus *Zymoseptoria pseudotritici*, a close relative of the wheat pathogen *Z. tritici* (synonym *Mycosphaerella graminicola*), we observed peculiar diversity patterns. Long regions up to 100 kb without variation alternate with similarly long regions of high variability. The variable segments in the genome alignment are organized into two main haplotype groups that have diverged ∼3% from each other. The genome patterns in *Z. pseudotritici* are consistent with a hybrid speciation event resulting from a cross between two divergent haploid individuals. The resulting hybrids formed the new species without backcrossing to the parents. We observe no variation in 54% of the genome in the five individuals and estimate a complete loss of variation for at least 30% of the genome in the entire species. A strong population bottleneck following the hybridization event caused this loss of variation. Variable segments in the *Z. pseudotritici* genome exhibit the two haplotypes contributed by the parental individuals. From our previously estimated recombination map of *Z. tritici* and the size distribution of variable chromosome blocks untouched by recombination we estimate that the hybridization occurred ∼380 sexual generations ago. We show that the amount of lost variation is explained by genetic drift during the bottleneck and by natural selection, as evidenced by the correlation of presence/absence of variation with gene density and recombination rate. The successful spread of this unique reproductively isolated pathogen highlights the strong potential of hybridization in the emergence of pathogen species with sexual reproduction.

species dynamics | incipient species | population genomics | genome mosaic | genome scans

The emergence of new species can be driven by natural selection in sufficiently isolated populations. Differentiation can occur when the populations specialize to different ecological niches, or speciation can result from the fixation and accumulation of distinct mutations in isolated populations (1). New species may however arise abruptly through hybridization between distinct species generating fertile progenies with novel genetic and phenotypic traits (2). One of the best-described examples is the wild sunflower *Helianthus paradoxus*, which derives from a cross between *Helianthus annuus* and *Helianthus petiolaris* (3). *H. paradoxus* exemplifies a hybrid species that has diverged ecologically from its parents by the acquisition of new phenotypic qualities allowing it to colonize new habitats (4, 5). In unicellular and filamentous fungi, hybridization may occur by the fusion of haploid cells or hyphae from different species. In epichloid endophytes interspecific hybridization have occurred repeatedly leading to the formation of multiple heteroploid asexual lineages of endophytes (6). In pathogenic species, lineages with new virulence traits or host specificities have been documented to emerge through hybridization of genetically distinct species (7–9). The defining traits in these successful hybrids may stem from new complementary and favorably interacting gene combinations allowing the successful invasion of new hosts (2). So far little is known about the population genetics or evolution of hybrid pathogen species. The population genetic properties of a young hybrid population may stall the advance of beneficial gene combinations. A hybrid speciation can involve a strong population bottleneck if very few individuals found the species. In such cases the establishment of a viable hybrid species and the capacity for rapid adaptation is negatively affected by the unavoidable loss of variation (10). If and how natural selection operates in a hybrid population to cope with new gene combinations and genome-wide loss of variation is till now poorly understood.

We here present evidence for a recent hybrid speciation event in the haploid plant pathogenic fungus *Zymoseptoria pseudotritici* (described as S1 in ref. 11), a close relative of the prominent wheat pathogen *Zymoseptoria tritici* (synonym *Mycosphaerella graminicola* (12). This young hybrid provided us a unique model to study genome evolution and the effects of natural selection and random genetic drift in a hybrid population. In a genome alignment of five *Z. pseudotritici* individuals, we found highly unusual patterns of sequence diversity. Long segments of very high haplotype differentiation are interspersed between likewise long segments showing no variation. This genome-wide pattern of genetic variation within a species is very different from that found within other closely related *Zymoseptoria* species and in other fungal species where genome-wide variation has been described at the population level (13). *Z. pseudotritici* and *Z. tritici* diverged recently in the Fertile Crescent during the initiation of agriculture and the domestication of wheat (14). Whereas *Z. tritici* is specialized onto wheat and has spread globally with its host, *Z. pseudotritici* is found on a range of very different grass species at the center of origin of the *Zymoseptoria* species in the Middle East. *Z. pseudotritici* was sampled frequently across a 500-km transect in northern Iran, suggesting that it is a common pathogen of uncultivated grasses in the region. We use population genomics and evolutionary analyses to reconstruct the speciation event leading to *Z. pseudotritici* and assess the subsequent effects of natural selection during early genome evolution of this hybrid species. The population genomics analysis of *Z. pseudotritici* represents a unique, detailed investigation of genome evolution in a recently emerged hybrid species.

Author contributions: E.H.S., F.B.C., and M.H.S. designed research; E.H.S., F.B.C., and M.H.S. performed research; T.T.H., J.Y.D., and M.H.S. contributed new reagents/analytic tools; E.H.S., F.B.C., and M.H.S. analyzed data; and E.H.S., F.B.C., and M.H.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.L.A. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) database, www.ncbi.nlm.nih.gov/genome/browse/ [accession nos. AEYQ00000000 (STIR04 A48b), AFIP00000000 (STIR04 A26b), AFIQ00000000 (STIR04 2.2.1), AFI000000000 (STIR04 3.1.1), AFIR00000000 (STIR04 4.3.1), AFIS00000000 (STIR04 5.3), AFIT00000000 (STIR04 5.9.1), AFIU00000000 (STIR04 1.1.1), AFIV00000000 (STIR04 1.1.2), AFIW00000000 (STIR04 1.1.2), AFIX00000000 (STIR04 3.3.2)].

[1]To whom correspondence should be addressed. E-mail: eva.stukenbrock@mpi-marburg.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201403109/-/DCSupplemental.

Results

Genome of *Z. pseudotritici* Is a Mosaic of Nonvariable and Highly Diverged Segments. Isolates of *Z. pseudotritici* were collected in the northern province of Ardabil in Iran (Dataset S1).

The nuclear genomes of five individuals of *Z. pseudotritici* were sequenced and aligned against the 15 chromosomes of the *Z. tritici* reference genome (15). Of these chromosomes, 1–13 are described as essential in *Z. tritici*, whereas the two smallest chromosomes are dispensable (15). Synteny of the essential chromosomes is conserved between species and our analysis focuses mainly on the sequence composition of these 13 chromosomes (26-Mb filtered alignment). We scanned the *Z. pseudotritici* alignment in nonoverlapping windows of 1,000 bp. A window is termed nonvariable in case of zero or one segregating site. The frequency of windows with only one segregating site comprises only 4% of these nonvariable windows showing that windows with zero segregating sites constitute the vast majority of windows in this category (Fig. 1*A*). Across the 13 chromosomes, we categorized a total of 25,171 windows corresponding to a total alignment of 25 Mb. Among these windows, 54% are nonvariable. The long segments without variation alternate across the 13 chromosomes with segments of high levels of variation as illustrated in Fig. 1*B*. These variable segments are mainly comprised of windows showing only two haplotypes. In fact >90% of the variable 1-kb windows show only two haplotypes. Within each of the two haplotypes, the sequences of the different individuals are identical. We assess the distance between the two haplotypes by counting the number of segregating sites per window and find an average of 30 variable sites per 1 kb (3%) (Dataset S2 and Fig. 1*A*).

The length of nonvariable and variable segments in *Z. pseudotritici* is counted as the number of consecutive windows without or with variation, respectively. We find segments of sizes up to 100 kb without variation and variable segments of sizes up to 67 kb. The genome-wide average length of the nonvariable segments is 10.8 kb and that of the variable segments is 9.6 kb.

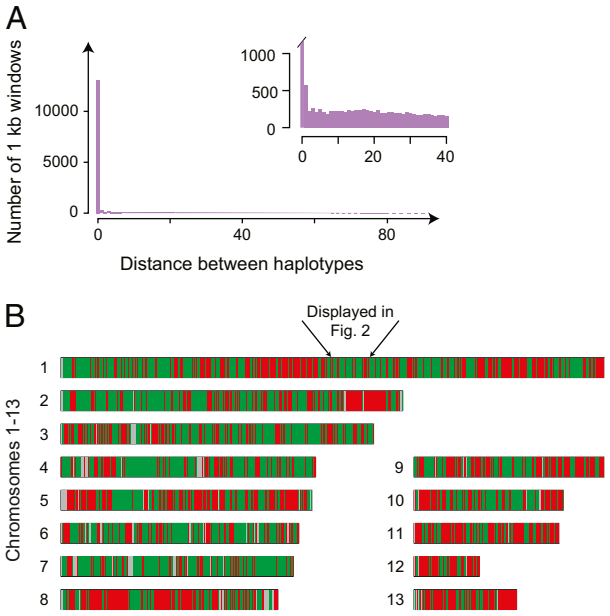

Fig. 1. (*A*) Frequencies of 1-kb windows with 0, 1, 2, 3. . .. segregating sites in five isolates of *Z. pseudotritici*. The plot illustrates that there is not a gradual decline in the number of polymorphic sites, but rather windows with 0 segregating sites is a distinct category. (*B*) Genome-wide distribution of variable (red) and nonvariable (green) segments across 13 aligned chromosomes of five isolates of *Z. pseudotritici*. Windows where the frequencies of gaps exceed 50% (gray) were excluded from the analysis.

The local variation across 400 kb in the *Z. pseudotritici* alignment is illustrated in Fig. 2*A*. Scanning across the variable segments we find that the alignments consist of blocks of different "topologies." We here refer to topologies as the pattern of association between the five aligned sequences and we depict these as trees (Fig. 2*B*). In total there are 15 possible topologies in the variable segments of the *Z. pseudotritici* alignment. An alignment of five sequences segregating two haplotypes can either show four identical sequences differing from the fifth, or three identical sequences differing from the remaining two (Fig. 2*B*). We number the five topologies of the first type 1–5 according to which of the five genomes harbor the singleton haplotype. The 10 topologies of the second type are numbered 6–15 according to the sequence combinations of the pair of identical sequences: 12,13,. . .,15,23,. . .,45. For each 1-kb window, we determined the underlying topology and find that variable segments often consist of mosaics of alternating sequence blocks of different topologies (Fig. 2*A*). The frequencies of the 15 topologies in the variable windows differ and show an excess of windows with the topologies 1–5 (Fig. 3 and Dataset S3) in agreement with sampling theory and the abundance of singletons in a population sample (16). Accounting for the different topologies in blocks within the variable segments, we find that the genome-wide average length of blocks of sequence with the same topology is 5.8 kb, i.e., about half the size of the variable and the nonvariable segments.

In summary, when describing the distribution of variation within and among the five genomes of *Z. pseudotritici*, we find two striking patterns. (*i*) Half of the genome has no variation, suggesting that the loss of variation occurred very recently. (*ii*) Locally in variable regions, only two haplotypes exist and the sequences within each haplotype are identical. These patterns suggest that the variation observed in *Z. pseudotritici* originate from only two haploid individuals.

Species-Wide Absence of Variation. Our population genomics dataset represents five *Z. pseudotritici* individuals and some of the nonvariable segments in these five genomes might show variation in a sample of more individuals. To describe and assess variation in a larger sample of *Z. pseudotritici* representing multiple individuals from the four local Iranian populations, we resequenced ∼600 bp in nine randomly chosen loci in segments without variation in 15 isolates of *Z. pseudotritici* (*SI Materials and Methods*). The homologous loci were also sequenced in several individuals of *Z. tritici*. In seven of these nine loci, we found no new alternative haplotypes. In each of the remaining two loci we identified a unique haplotype distinct from the other sequences. This discovery thereby classifies the loci as variable instead of nonvariable (Dataset S4). We also sequenced fragments of the Intergenic Transcribed Spacer (ITS) rDNA and of the mating type loci Mat1-1 and Mat1-2 in the 15 *Z. pseudotritici* isolates and found these completely nonvariable. We further tested whether the variable segments still only comprise two haplotypes when including additional sequences from more *Z. pseudotritici* isolates and we resequenced ∼600 bp of eight randomly chosen variable loci in 15 isolates. We could indeed show that even in alignments of 15 sequences, the *Z. pseudotritici* population only exhibits two haplotypes. In *Z. tritici*, the homologous loci (both nonvariable and variable) each contained polymorphic sites and segregated into several different haplotypes (Dataset S4). The resequencing of individual loci in multiple individuals of *Z. pseudotritici* thereby supports that the patterns observed in the five genomes are genuine and not caused by sampling, assembly, or alignment artifacts. Extrapolating from the resequencing data (excluding the mating type locus) and using the estimated percentages of the variable and nonvariable segments in the five genomes, we can roughly estimate that $1 - (46\% + 2/10 \times 54\%) = 43\%$ of the genome in the *Z. pseudotritici* species is without variation.

We further investigated the overall loss of variation in *Z. pseudotritici* by a population genetic analysis of the observed frequency of nonvariable windows and the frequencies of the

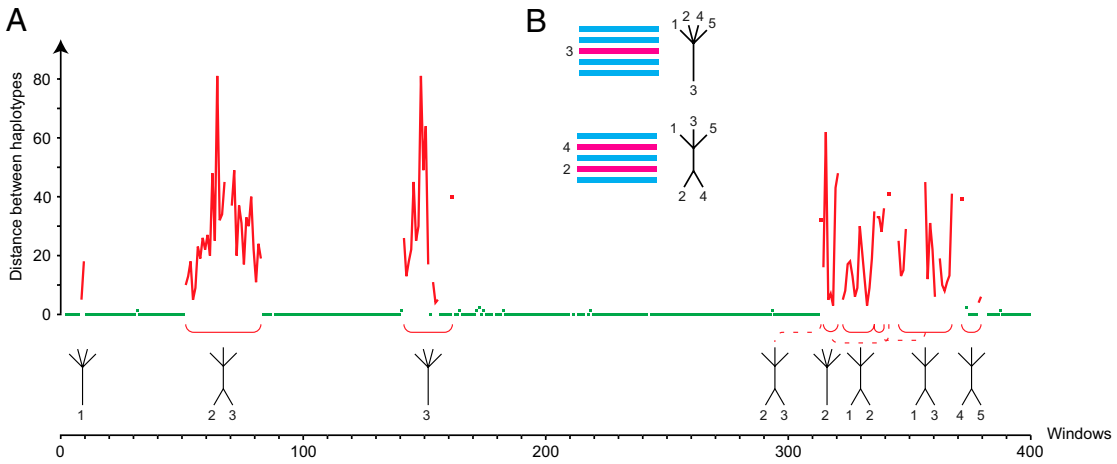

Fig. 2. (*A*) Close-up of 400 aligned 1-kb windows on chromosome 1 (Fig. 1*B*) in five *Z. pseudotritici* isolates illustrating the alternating patterns of variable (red) and nonvariable (green) segments. The vertical axis gives the distance in each window between the two haplotypes defined as H1 and H2 (see below). This distance between H1 and H2 reflects, according to our model, the distance between the parents of the hybrid species. (*B*) The pattern of association of sequences in the variable segments of *Z. pseudotritici* can be illustrated as topologies. With five individuals, there are 15 possible topologies. In *Z. pseudotritici* the five sequences always segregate into two haplotypes either where one single sequence differs from the remaining four or where two identical sequences differ from the remaining three. The two "ends" of the topology trees represent the haplotypes H1 and H2. For each aligned window in *Z. pseudotritici* the specific topology was assigned (*A*).

different topologies (*SI Results*). We use the observations that each variable 1-kb window segregates in two haplotypes and that these haplotypes are equally frequent over all variable windows. Furthermore, we assume that *Z. pseudotritici* emerged by a homoploid hybridization event and that the two haplotypes were equally frequent after the first crossing event. We can thereafter assess the distributions of the haplotype frequencies as described by Kimura (17). The best fit to the observed frequency of 54% nonvariable windows is obtained after 0.8 N_e generations of random genetic drift (0.8 ± 0.1), where N_e is the effective population size (*SI Results*). At this time, 34% of the windows are estimated to have lost variation in the species (0.34 ± 0.06).

The two independent approaches, the resequencing of multiple individuals and the population genetic modeling approach, both support that *Z. pseudotritici* has undergone a significant loss of variation, resulting in long segments of no variation for about one-third of the genome in the species.

***Z. pseudotritici* Emerged by a Recent Homoploid Hybridization Event.** The striking mosaic of variable and nonvariable segments cannot be explained by recurrent selective sweeps in a panmictic population. We do not observe a decrease in diversity in regions bordering the variable segments (Fig. 2*A*), which is otherwise expected under a scenario of selective sweeps (18). Furthermore, thousands of selective sweeps leaving large nonvariable regions would have resulted in a marked depletion of variation in the entire genome according to population genetics theory. Finally, a selective interpretation fails to explain the presence of only two haplotypes in variable windows.

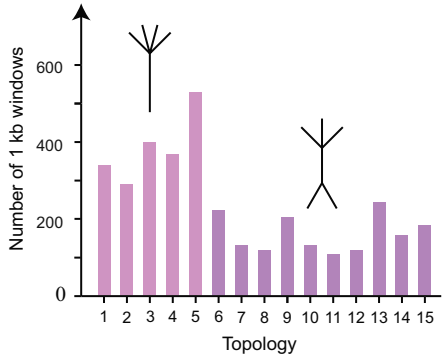

Fig. 3. Distribution of windows assigned to each of the 15 possible topologies. In total, 13,579 variable windows could be assigned to one of the topologies. For the topologies 1–5 one sequence differ from the other four and for the topologies 6–15 two sequences different from the other three (Fig. 2*B*).

Instead the genome-wide mosaic of segments of variation and no variation in *Z. pseudotritici* is consistent with a recent hybridization event (Fig. 4*A*). Under our proposed scenario, a diploid F_1 zygote formed by the fusion of two distinct haploid individuals underwent mitosis and meiosis to form a haploid F_1 hybrid swarm (Fig. 4*B*). In the *Z. pseudotritici* hybrid the two different haplotypes in variable windows reflect the preservation of sequences from both parental individuals. The 10% of the variable 1-kb windows exhibiting more than two haplotypes, show a pattern of variation consistent with an ancestral recombination event between the two parental haplotypes. The nonvariable segments in the *Z. pseudotritici* genome represent sequences maintained from only one of the two parental individuals. The absence of additional haplotypes in variable segments suggests that individuals of the initial hybrid swarm and the subsequent population have not backcrossed with either of the parental species. Individuals from the F_1 swarm must therefore have been able to perform sexual fusion and subsequent meiotic divisions. The finding of both mating type alleles homologous to the mating type idiomorphs described in *Z. tritici* (19) supports this. If frequent sexual recombination occurs in the hybrid species we expect even frequencies of the two mating type idiomorphs in a population sample of *Z. pseudotritici*. To test this hypothesis, we screened a collection of *Z. pseudotritici* isolates including 27 isolates collected from four local populations in Iran (Dataset S1). Isolates carrying the Mat1-1 and the Mat1-2 idiomorph are present in each of the four populations and the overall ratio of the Mat1-1 and Mat1-2 locus is ~3/5. These observations are consistent with the general effects of recombination seen in the *Z. pseudotritici* genome. The presence of both mating types in each of the local populations and their observed ratio strongly supports that sexual recombination can occur readily in the hybrid *Z. pseudotritici*.

According to our model, the two haplotypes in the variable segments of the *Z. pseudotritici* genome represent homologous sequences from the parental individuals. Thereby, the extent of

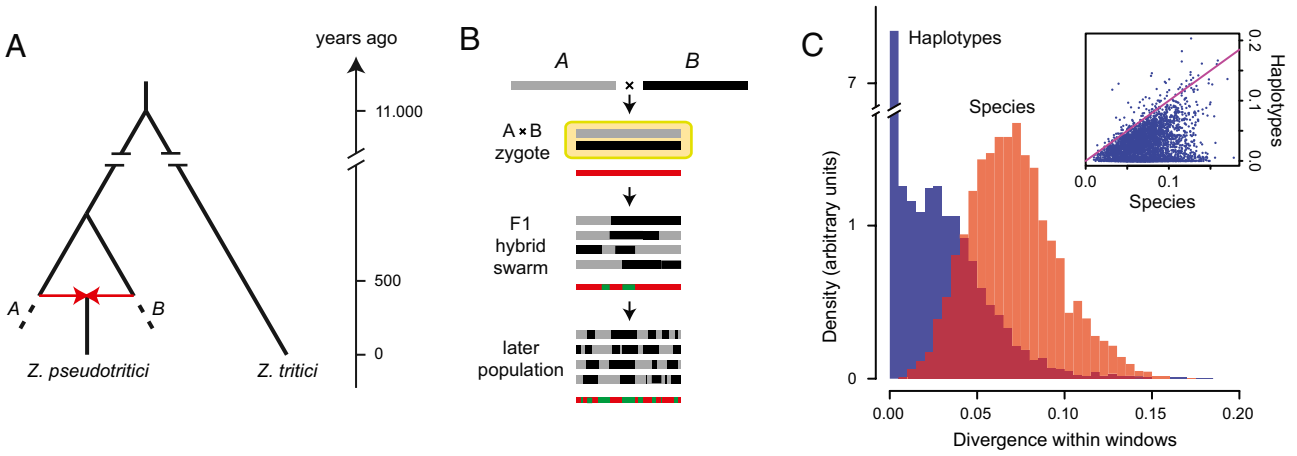

Fig. 4. (*A*) The genomic pattern in *Z. pseudotritici* is consistent with a speciation history where a common ancestor of the two hybrid parents diverged recently after the emergence of the *Z. tritici* lineages. Two haploid individuals from these diverged lineages formed viable hybrid offsprings through one sexual cross ~500 y ago. (*B*) Initial fusion between the two individuals gave rise to a diploid zygote. Following meiotic recombination the genomes of the haploid individuals in the F_1 hybrid swarm are mosaics of the parental chromosomes. Some segments of the genomes may be inherited and fixed from only one of the two parents. These segments will remain nonvariable in the subsequent generations of the hybrid swarm, and they exist alongside variable segments contributed by both parental species. Through time, further crossover and random genetic drift will lead to the genomic mosaic observed today. In the extant population of *Z. pseudotritici*, variable windows show only two haplotypes, suggesting that the hybrid swarm did not backcross to the parental species, but instead propagated as a saltatory biological species. (*C*) Parental haplotypes (H1 and H2) are more closely related to each other (blue bars) than to *Z. tritici* (red bars) as shown by the two histograms of divergence estimates for all aligned windows. In the subplot the divergence between the two *Z. pseudotritici* haplotypes H1 and H2 are plotted against the net divergence between *Z. tritici* and the two haplotypes for windows on chromosome 1. The majority of the 7,076 compared segments are found below the line of identity, showing the differentiation between the two *Z. pseudotritici* haplotypes H1 and H2 to be consistently lower than the divergence to *Z. tritici*.

differentiation between the two sequences reflects the extent of divergence between the parental lineages. To infer the evolutionary history of the hybrid and its parents, we compared the divergence between the two *Z. pseudotritici* haplotypes and their divergence from the sister species *Z. tritici*. For each window, we randomly designated two haplotypes H1 and H2. By comparing the distributions of nucleotide differentiation between H1 and H2 and the divergence between H1 and homologous windows in *Z. tritici*, it is evident that the parents of the *Z. pseudotritici* hybrid were more closely related to each other than any of them to *Z. tritici* (Fig. 4*C* and Fig. S1). Because the divergence of *Z. tritici* and *Z. pseudotritici* is estimated to have occurred ~11,000 y ago (14), the divergence of the two *Z. pseudotritici* parental species must have occurred more recently.

The average divergence of H1 and H2 measured across the 13 core chromosomes is 0.0336. This is notably larger than the typical within-species nucleotide diversity in the closely related species *Z. tritici* (0.0106) and *Zymoseptoria ardabiliae* (0.0088) as reported in ref. 11 (Dataset S2). However, the average divergence between *Z. pseudotritici* and the two sister species *Z. tritici* and *Z. ardabiliae* is 0.061 and 0.107, respectively, thus these two species are considerably more diverged from *Z. pseudotritici* than the two parental species were at the time of hybridization. Nevertheless, the threefold higher level of within-species variation in variable segments of *Z. pseudotritici* suggests that the parents of the hybrid were distinct lineages, although closely related. The variable and nonvariable segments show the same average extent of divergence to *Z. tritici* and *Z. ardabiliae* supporting that the two different components of the *Z. pseudotritici* genome have a very similar evolutionary history.

Only two dispensable chromosomes, 16 and 17, are shared between the five *Z. pseudotritici* individuals analyzed here (*SI Results*). Estimating the extent of divergence between *Z. pseudotritici* and *Z. tritici* on chromosome 17 we find a similar extent of divergence as observed on the essential chromosomes. Chromosome 16, however, shows a significantly higher level of differentiation from chromosome 16 in *Z. tritici* (Fig. S2). We speculate that chromosome 16 has undergone a different evolutionary history relative to the core chromosomes in one of the parents of *Z. pseudotritici*. Two additional dispensable chromosomes 20 and 21, showing presence/absence polymorphism in *Z. pseudotritici*, are present in four and absent in one of the five studied genomes. They show very little variation, indicating that they each originate from only one of the hybrid parents.

Timing of the Homoploid Hybridization Event. The mosaic of topology blocks and segments with and without variation represents a record of recurrent recombination in *Z. pseudotritici*. We use the length distribution of these blocks and segments to infer the number of sexual generations since the hybridization. Our previous whole genome coalescence analyses gave an estimated average recombination rate of 46 cM/Mb in the common ancestor of *Z. tritici*, *Z. pseudotritici*, and *Z. ardabiliae* (11). Only the fragmentation within variable segments, observed as blocks of different topologies (Fig. 2*B*), records the effects of recombination during the entire history of *Z. pseudotritici*. The fragmentation into variable and nonvariable segments on the other hand represents mainly recombination events that occurred in the early history of the species while population sizes were small. The average length of the blocks of distinct topologies is 5.8 kb and this value allows by the following argument a crude estimation of the elapsed time since speciation. A recombination rate of 46 cM/Mb is expected to produce a recombination event for every 2.2 Mb in the formation of the initial hybrid swarm. In the next sexual generation each of the resulting blocks have on average undergone one recombination event and the average size of conserved blocks has been reduced to 2.2/2 Mb. In the third sexual generation half of these blocks undergo yet another recombination event and the block size is on average reduced to 2.2/3 Mb, and so on. After 380 generations of recombination, the average length of the blocks of distinct topologies is expected to be 2.2/380 Mb = 5.8 kb, which is equal to the observed value. The calculation thereby proposes that the emergence of *Z. pseudotritici* occurred 380 sexual generations ago. A source of error in the calculation is additional fragmentations due to sequence or alignment errors, and this may inflate the estimated number of generations. Furthermore, a concentration of recombination events into hotspots would also affect

the average length of segments. Recombination hot spots would however concentrate recombination within particular 1-kb windows and we would not observe all of the recombination events that have actually occurred, causing an underestimation of the number of generations since the hybridization event. To date the hybrid speciation, we assume at least one sexual cycle per year and possibly two to three as in *Z. tritici*, giving an age of the order of 200 y.

An alternative estimation of the time since the hybridization event can be derived from the number of 1-kb windows with a single segregating site. If sequencing error is low as suggested by the long stretches without variation, most of these windows have acquired the segregating site as a unique mutation since the hybridization event. From the number of windows with two, three, and four segregating sites (Fig. 1*A*), we estimate that two-thirds of the mutations in the single variant windows are due to a unique mutation. These windows constitute 4% of the windows classified as nonvariable, with the remaining 96% with zero segregating sites. Assuming a star phylogeny of the five samples, the total branch length since variation was lost is $5Gu$, where G is the number of generations and u the mutation rate per generation. We can estimate $G = 0.04(2/3)/(5^{*}10^{-8}{*}1000) \sim 550$ generations since the hybridization event, assuming a mutation rate per sexual generation of 10^{-8} (20). Despite the very crude approximations, this estimate is in agreement with the above estimate obtained from the recombination pattern.

The size of blocks formed by recombination decrease through the history of the species. A sequence block that loses variation in the early history of the species is therefore expected to be longer than a sequence block that loses variation later in the history of the species. We use this assumption to address whether variation was lost early or late in evolution of the *Z. pseudotritici* genome. The variable and nonvariable segments are considerably longer than the blocks considered above (on average about 10 kb). The mere size of the nonvariable segments strongly suggests that the major part of the fixations and loss of variation occurred early in the history of *Z. pseudotritici*. This finding is consistent with the expected high rate of variation loss due to random genetic drift in the small initial population.

Effect of Natural Selection in Evolution of the Hybrid Genome. To assess the impact of natural selection in shaping the genome of *Z. pseudotritici*, we compare the distribution of segment lengths to various chromosomal parameters. Most notably, the frequencies and average lengths of variable and nonvariable segments are heterogeneous across the 13 core chromosomes (Fig. 1*B* and Dataset S5). On the 6-Mb–long chromosome 1, 60% of the aligned 1-kb windows are nonvariable (Fig. 5*A*) and the length of segments without variation are on average 13 kb (Fig. 5*B*). The variable segments on chromosome 13 are on average 13 kb long, and 35% of the windows are nonvariable. This heterogeneity in segment distributions is correlated with recombination rate. We find a negative correlation between the chromosomal average of the recombination rate and the length of nonvariable segments within the chromosome ($\tau = -0.641$, $P = 0.001616$, Kendall's rank correlation test) (Fig. 5 *B* and *C*).

The effect of natural selection on an allele depends on the size of the ancestral linkage block surrounding the specific allele. These flanking sequences will hitchhike with the allele, while they are eroded by recombination. A higher local recombination rate will diminish the lengths of the hitchhiking sequences more rapidly. The ensuing loss of variation in sequences flanking an allele fixed by selection is therefore expected to be less on chromosomes with a higher recombination rate than on those with a lower recombination rate. Thereby the loss of variation in flanking sequences of selected alleles is expected to be larger on chromosome 1 than on chromosome 13.

Gene density in the *Zymoseptoria* genomes is high, in *Z. tritici* it is on average ~280 genes per megabase. In *Z. pseudotritici* we found significantly more exonic sites in the nonvariable segments (6,898,606 bp and a mean density of 0.5411 exonic sites) relative to the polymorphic fraction (5,436,699 and a mean density of 0.5009 exonic sites) ($P = 0.01431$, Wilcoxon rank sum test). In total our alignment includes 9,490 predicted genes with at least 50 codons aligned between the five isolates. Of these 5,432 display no genetic variation, highlighting the implication of the genetic bottleneck in the species. We observe no correlation of segment distribution with rates of nonsynonymous or synonymous segregating sites (*SI Results*).

Discussion

The frequency and distribution of segments with null variation in the genome of the plant pathogenic fungus *Z. pseudotritici* suggests that the species was formed ~380 sexual generations ago by hybridization of closely related although distinct lineages of *Zymoseptoria* individuals. The parents of the hybrid were 3% divergent. This genetic distance still allowed the formation a viable and reproductively competent zygote. The observed level of genome-wide differentiation between the parental haplotypes, however, is two- to fourfold larger than the variation within other *Zymoseptoria* species, suggesting that the genetic differentiation has developed in populations that evolved separately over a long time period. We demonstrate that the common ancestor of the hybrid parents is younger than the split between *Z. tritici* and *Z. pseudotritici* 11,000 y ago (14). The recent origin of the hybrid parents and the recent formation of their hybrid descendant

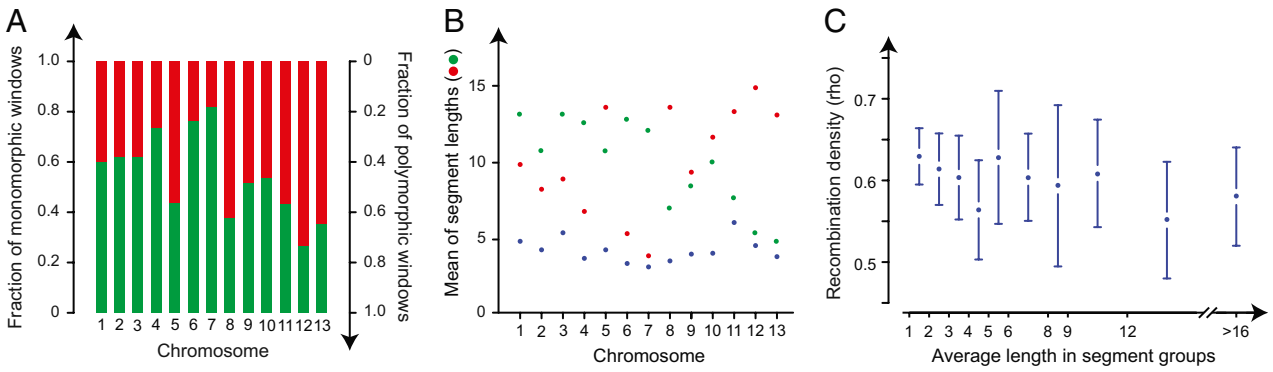

Fig. 5. (*A*) Frequency of variable (red) and nonvariable (green) windows differs across the 13 core chromosomes. (*B*) Average lengths of the variable segments (red dots) and and nonvariable (green dots) are shown for the 13 core chromosomes. Variable and nonvariable segments are within the same size range, whereas the average lengths of the topology blocks (blue) are considerably smaller. (*C*) Average length of segments is correlated with the local recombination rate measured as ρ. Longer segments are located in regions of lower recombination rates and shorter segments in regions of high recombination rate.

indicate that *Zymoseptoria* comprise a dynamic complex of closely related species, including lineages or incipient species that can evolve new pathogenic potentials by mutual hybridizations. The separation of lineages might relate to geographical isolation. However, the long-distance wind dispersal of the sexually produced ascospores of *Zymoseptoria* species argues against this suggestion. Instead, ecological divergence driven by local adaptation or host specialization, as observed for other plant pathogenic species (21, 22), seems a more likely explanation for the isolation and differentiation of lineages within *Zymoseptoria*.

We have not been able to identify the parental species of *Z. pseudotritici* in our collections of pathogens from the studied region in Iran. Still the parent species may exist as pathogens on grass hosts not included in our sampling. This possibility is indeed likely and interesting because virulence assays and comparative genomics could reveal the barriers to gene flow between the hybrid and its parents. We present strong evidence that backcrosses of the hybrid species to its parental species did not occur. This argues for distinct host specificities of *Z. pseudotritici* and both parental species. Experimental inoculation of both *Z. tritici* and *Z. pseudotritici* on different host species including wheat, *Elymus repens* and *Lolium perenne*, demonstrates that both species can infect and propagate asexually on multiple hosts. Different isolates, however, show a large variability in their degree of virulence, and in both species significant differences in virulence between isolates have been observed (11, 23). This suggests specialization to occur not only at the species level but also at the population level. Local variation in the degree of virulence on the various host species may form the intrinsic barriers to gene flow between populations coexisting in the same geographical location.

The stochastic effects in the recombination of parental genomes and random genetic drift appear as the major forces in the evolution of the hybrid genome in *Z. pseudotritici*. Variation was removed from about a third of the genome, including a similar fraction of the genes, and for the remaining genes only the two parental alleles exist. The loss of variation is not randomly scattered over the genome. We show a strong negative correlation between local average recombination rates and the frequency and length of segments without variation. This trend is likely to be caused by directional selection favoring one of the parental alleles of particular genes.

This study is a unique report on genome evolution in the early stages of a hybrid speciation event. The successful spread of *Z. pseudotritici* across a large geographical area demonstrates a mechanism and an astonishing rate by which a unique pathogenic species can emerge. Such rapid evolution and spread of a unique pathogenic species can be forwarded by agricultural trade and management practices allowing unique pathogens to emerge through hybridization between introduced and local species.

Materials and Methods

Population Genomic Analyses. Sequencing, assemblies, and alignments of fungal genomes including five isolates of *Z. pseudotritici*, two *Z. tritici*, and one *Z. ardabiliae* isolate is described in ref. 11. The genome alignment includes 27 Mb covering 15 chromosomes. We used an in-house developed Java-based genome browser to extract counts and parameter estimates from alignments of the 15 individual chromosomes.

For each window containing polymorphic sites, we determined the topology of the underlying sequence genealogy and assigned one of 15 possible topologies. For 9% (2,297 windows) of the windows topology assignment was not possible due to incompatibilities in the alignment. These windows were also included in the quantification of variable segments. To determine the amount of nucleotide differentiation between the hybrid parental species, we randomly isolated sequences from each end of the unrooted *Z. pseudotritici* topologies. We assessed the net divergence between haplotypes and between each of the two haplotype sequences to *Z. tritici* and *Z. ardabiliae*.

For each 1-kb window we computed the density of exons using the gene annotation of the *Z. tritici* reference sequence (http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html) (15). We further used the annotation information with the two-species alignment (*Z. tritici* and *Z. pseudotritici*) to estimate the rates of synonymous and nonsynonymous mutations (P_N and P_S) within species (24) and to categorize genes encoding signal peptides (25, 26). Estimates of recombination rates across chromosomes were taken from ref. 11. They are based on a whole genome coalescence hidden Markov model approach (27).

ACKNOWLEDGMENTS. We thank Deborah Charlesworth for helpful comments to the manuscript and Petra Happel for assistance in the laboratory. The genome sequencing of *Zymoseptoria* isolates was funded by the Danish Research Council. The *Z. tritici* assembly used as reference in our study was produced by the Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community.

1. Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323: 737–741.
2. Greig D, Louis EJ, Borts RH, Travisano M (2002) Hybrid speciation in experimental populations of yeast. *Science* 298:1773–1775.
3. Ungerer MC, Baird SJE, Pan J, Rieseberg LH (1998) Rapid hybrid speciation in wild sunflowers. *Proc Natl Acad Sci USA* 95:11757–11762.
4. Lexer C, Welch ME, Durphy JL, Rieseberg LH (2003) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of Helianthus paradoxus, a diploid hybrid species. *Mol Ecol* 12:1225–1235.
5. Rieseberg LH, et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216.
6. Schardl CL, et al. (1994) Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*. *Genetics* 136:1307–1317.
7. Farrer RA, et al. (2011) Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proc Natl Acad Sci USA* 108:18732–18736.
8. Brasier CM, Cooke DEL, Duncan JM (1999) Origin of a new Phytophthora pathogen through interspecific hybridization. *Proc Natl Acad Sci USA* 96:5878–5883.
9. Brasier CM, Kirk SA (2010) Rapid emergence of hybrids between the two subspecies of Ophiostoma novo-ulmi with a high level of pathogenic fitness. *Plant Pathol* 59: 186–199.
10. Kimura M (1995) Limitations of Darwinian selection in a finite population. *Proc Natl Acad Sci USA* 92:2343–2344.
11. Stukenbrock EH, et al. (2011) The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Res* 21:2157–2166.
12. Quaedvlieg W, et al. (2011) *Zymoseptoria* gen. nov.: A new genus to accommodate *Septoria*-like species occurring on graminicolous hosts. *Persoonia* 26:57–69.
13. Liti G, et al. (2009) Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.
14. 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.
15. Goodwin SBBMBS, et al. (2011) Finished genome of *Mycosphaerella graminicola* reveals stealth pathogenesis and extreme plasticity. *PLoS Genet* 7:e1002070.
16. Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3: 87–112.
17. Kimura M (1955) Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci USA* 41:144–150.
18. Barton NH (2000) Genetic hitchhiking. *Phil Trans Royal Soc London. Ser B Biol Sci* 355: 1553–1562.
19. 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.
20. Lynch M, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272–9277.
21. Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol* 25:387–395.
22. Giraud T, Villaréal LM, Austerlitz F, Le Gac M, Lavigne C (2006) Importance of the life cycle in sympatric host race formation and speciation of pathogens. *Phytopathology* 96:280–287.
23. Kema GHJ, van Silfhout CH (1997) Genetic variation for virulence and resistance in the wheat-Mycosphaerella graminicola pathosystem III. Comparative seedling and adult plant experiments. *Phytopathology* 87:266–272.
24. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
25. Stukenbrock EH, Dutheil JY (2012) Comparing fungal genomes: Insight into functional and evolutionary processes. *Methods Mol Biol* 835:531–548.
26. 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.
27. Dutheil JY, et al. (2009) Ancestral population genomics: The coalescent hidden Markov model approach. *Genetics* 183:259–274.