Fine-Scale Recombination Maps of Fungal Plant Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots

Eva H. Stukenbrock*,[†,1] and Julien Y. Dutheil[‡,§,1]

*Environmental Genomics, and ‡Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany,
†Environmental Genomics, Christian-Albrechts University of Kiel, 24118, Germany, and §Institut des Sciences de L'Évolution de
Montpellier, Centre National de la Recherche Scientifique, Université Montpellier 2, 34095, France

ORCID IDs: 0000-0001-8590-3345 (E.H.S.); 0000-0001-7753-4121 (J.Y.D.)

ABSTRACT Meiotic recombination is an important driver of evolution. Variability in the intensity of recombination across chromosomes can affect sequence composition, nucleotide variation, and rates of adaptation. In many organisms, recombination events are concentrated within short segments termed recombination hotspots. The variation in recombination rate and positions of recombination hotspot can be studied using population genomics data and statistical methods. In this study, we conducted population genomics analyses to address the evolution of recombination in two closely related fungal plant pathogens: the prominent wheat pathogen *Zymoseptoria tritici* and a sister species infecting wild grasses *Z. ardabiliae*. We specifically addressed whether recombination landscapes, including hotspot positions, are conserved in the two recently diverged species and if recombination contributes to rapid evolution of pathogenicity traits. We conducted a detailed simulation analysis to assess the performance of methods of recombination rate estimation based on patterns of linkage disequilibrium, in particular in the context of high nucleotide diversity. Our analyses reveal overall high recombination rates, a lack of suppressed recombination in centromeres, and significantly lower recombination rates on chromosomes that are known to be accessory. The comparison of the recombination landscapes of the two species reveals a strong correlation of recombination rate at the megabase scale, but little correlation at smaller scales. The recombination landscapes in both pathogen species are dominated by frequent recombination hotspots across the genome including coding regions, suggesting a strong impact of recombination on gene evolution. A significant but small fraction of these hotspots colocalize between the two species, suggesting that hotspot dynamics contribute to the overall pattern of fast evolving recombination in these species.

KEYWORDS genome evolution; recombination analyses; recombination hotspots; fungal plant pathogens; effectors; *Zymoseptoria*

MEIOTIC recombination is a fundamental process that, in many eukaryotes, shapes genetic variation in populations and drives evolutionary changes. Studies based on experimental and empirical data have demonstrated that

recombination in sexual organisms plays a crucial role in defining genome-wide neutral and nonneutral nucleotide variation patterns (Begun and Aquadro 1992; Spencer *et al.* 2006), rates of protein evolution (Betancourt *et al.* 2009), transposable element (TE) distribution (Rizzon *et al.* 2002), GC content (Meunier and Duret 2004), and codon-usage bias (Marais *et al.* 2003). Despite the ubiquitous occurrence of recombination, however, the mechanisms that determine the genome-wide and temporal distribution of crossover events are still poorly understood in most species.

Accurate genome-wide recombination maps are essential for studying the genomics and genetics of recombination. Recombination rates have been recorded in many species by direct observations of meiotic events using genetic crosses or pedigrees (for example Broman *et al.* 1998; Jeffreys *et al.* 1998; McMullen *et al.* 2009). Pedigree studies, however, rely

on large numbers of individuals and produce only low-resolution rate estimates because of the relatively low number of meiotic events that can practically be observed (Stumpf and McVean 2003). Furthermore, many microbial eukaryotic species, including important pathogens, are difficult or even impossible to cross under laboratory conditions (Taylor *et al.* 2015). While experimental measures of recombination rate can be challenging in many species, advances in statistical analyses provide powerful tools to generate fine-scale recombination maps using population genomic data (*e.g.*, Myers *et al.* 2005; Chan *et al.* 2012; Wang and Rannala 2014). These methods are based on genome-wide patterns of linkage disequilibrium (LD) among single nucleotide polymorphisms (SNPs) and have the potential to capture the history of recombination events in a population sample. Thus, recombination studies based on population genomic data have provided detailed insights into the genomics of recombination in a range of species (Winckler *et al.* 2005; Horton *et al.* 2012; Singhal *et al.* 2015; Hunter *et al.* 2016). In many organisms, but not all, the majority of recombination events tend to concentrate in short segments termed recombination hotspots (Petes 2001; Chan *et al.* 2012). In the human genome, >25,000 recombination hotspots have been identified, with a number of them showing a >100-fold increase in recombination rates and exhibiting a strong impact on the overall recombination landscape and genome evolution in general (Myers *et al.* 2005; Winckler *et al.* 2005; Jeffreys and Neumann 2009).

Comparative analyses of recombination maps between closely related species have shed light on the dynamics of recombination landscapes in different taxa. A comparative analysis of recombination landscapes of chimpanzees and humans found a strong correlation of recombination rates at broad scales (whole-chromosome and megabase scale), whereas fine-scale recombination rates were considerably less conserved because of nonoverlapping recombination hotspots (Auton *et al.* 2012). The localization of recombination hotspots in primates and mice is in large part determined by PRDM9, a histone methyltransferase with an array of DNA-binding, Zn-finger domains (Myers *et al.* 2010). In some species—including species without PRDM9 such as yeast, plants, birds, and some mammals—recombination hotspots associate with particular functional features such as transcription start and stop sites as well as CpG islands (Horton *et al.* 2012; Choi *et al.* 2013; Lam and Keeney 2015; Singhal *et al.* 2015; Smeds *et al.* 2016). A model developed to explain the association of recombination hotspots and functional elements proposes that a depletion of nucleosome occupancy at these sites increases the accessibility of the recombination machinery (Kaplan *et al.* 2009; de Castro *et al.* 2012). Indeed, in the fission yeast *Schizosaccharomyces pombe* and the Brassicaceae plant *Arabidopsis thaliana*, meiotic recombination hotspots were shown to colocalize with nucleosome-depleted regions, supporting a link between chromatin structure and recombination in these species (de Castro *et al.* 2012; Wijnker *et al.* 2013).

Although many pathogens and parasites are sexual, the impact of recombination on the evolution of their genomes has been rarely addressed (Awadalla 2003). Genome studies have revealed exceptionally high rates of sequence evolution in some filamentous pathogens, including oomycetes and fungi (Raffaele and Kamoun 2012; Möller and Stukenbrock 2017). TEs, in particular, have been shown to play an important role in shaping the architecture and size of these pathogen genomes. TEs have often been found to be enriched in specific genomic compartments such as accessory chromosomes and repeat-rich regions that further encode virulence-related genes (reviewed in Möller and Stukenbrock 2017). Increased mutation rates in TE-rich regions have been shown to contribute to the rapid evolution of new virulence specificities in pathogens and can contribute to the rapid generation of new genetic variation in pathogen genomes in the absence of sexual recombination (Daverdin *et al.* 2012; de Jonge *et al.* 2013; Dutheil *et al.* 2016). While TEs may contribute to the rapid evolution of specific genome compartments there are few population genetic studies of genome evolution, genome evolution in eukaryotic pathogens. As recombination can be an important driver of overall genome evolution in pathogen species, we set out to investigate patterns of recombination in plant pathogenic fungi. We focused on the economically important wheat pathogen *Zymoseptoria tritici*, which causes septoria leaf blotch on wheat. *Z. tritici* originated in the Middle East during the Neolithic revolution and has coevolved and dispersed with its host since early wheat domestication (Stukenbrock *et al.* 2007). A close relative of *Z. tritici*, *Z. ardabiliae*, has been isolated from wild grass species in the Middle East (Stukenbrock *et al.* 2012). The two pathogen species diverged recently but have nonoverlapping host ranges and show some differences in morphology and host infection patterns (Stukenbrock *et al.* 2011, 2012). Both species undergo frequent sexual recombination, which results in the formation of ascospores that serve as a means of long distance wind dispersal and primary infection of new hosts (Stukenbrock *et al.* 2011). The colinear genomes of *Z. tritici* and *Z. ardabiliae* share 90% nucleotide similarity on average, thus providing an excellent resource for comparative analyses of genome evolution (Stukenbrock *et al.* 2011). The 40-Mb haploid genome of the reference *Z. tritici* isolate comprises 21 chromosomes, of which 8 are accessory chromosomes (Goodwin *et al.* 2011). These highly variable chromosomes are characterized by presence/absence variation, structural variation, high repeat content, and low gene densities (Goodwin *et al.* 2011; Grandaubert *et al.* 2015). Interestingly, the accessory chromosomes are partly conserved among several species in the genus *Zymoseptoria*, suggesting that these small chromosomes have been maintained over long evolutionary times, predating the divergence of species (Stukenbrock *et al.* 2011).

In a previous study, we applied a whole-genome coalescence approach to generate a map of incomplete lineage

sorting of the ancestral species of *Z. tritici* and another closely related species, *Z. pseudotritici* (Stukenbrock *et al.* 2011). We found evidence of a high recombination rate in the ancestral species (genome average 46 cM/Mb) and showed a significantly higher proportion of sites showing incomplete lineage sorting in regions with high recombination rate. The existence of high recombination rates in the genus *Zymoseptoria* was recently supported by experimental data. Croll *et al.* (2015) generated a linkage map of *Z. tritici* from two independent crosses of Swiss field isolates. This map, based on actual crossover events along the 40-Mb genome, confirms the high recombination rates (genome average 66 cM/Mb, measured in windows of 20 kb) in the present-day pathogen species. Interestingly, the study also reported large differences between the two independent crosses of *Z. tritici*, suggesting that the recombination landscape is highly dynamic in this pathogen (Croll *et al.* 2015).

In this study, we addressed the evolution of recombination rate in fungal pathogens. We applied a population genomics approach to generate a fine-scale recombination map of the two recently diverged species *Z. tritici* and *Z. ardabiliae*. This allowed us to infer and compare fine-scale, genome-wide patterns of recombination rates in the two species and investigate the evolution of recombination landscapes. We first confirm the exceptionally high recombination rates, as also observed in a previous coalescence-based genome analysis and as shown by experimental crosses (Stukenbrock *et al.* 2011; Croll *et al.* 2015). Furthermore, we identify 2578 and 862 recombination hotspots in *Z. tritici* and *Z. ardabiliae*, respectively. Intriguingly, detailed analyses of the recombination hotspots show not only a comparatively higher hotspot frequency in the wheat pathogen but also the occurrence of stronger hotspots in *Z. tritici*. Our findings confirm that recombination rate landscapes are highly dynamic across time in the two fungal pathogens. Furthermore, the prominence of dynamic recombination hotspots in genes suggests a high impact on gene evolution; a finding that is unprecedented in other species.

Materials and Methods

Genome data

The life cycle of *Z. tritici* is predominantly haploid and the genome analyses conducted here thus rely on haploid genome data. The 40-Mb reference genome of the *Z. tritici* isolate IPO323 was sequenced at the Joint Genome Institute using Sanger sequencing (Goodwin *et al.* 2011). Two Iranian *Z. tritici* isolates and four Iranian *Z. ardabiliae* isolates were sequenced in a previous study using Illumina sequencing (Supplemental Material, Table S1) (Stukenbrock *et al.* 2011). We used genome data from an additional 10 isolates of *Z. tritici* that originate from wheat fields in Denmark, France, and Germany (Grandaubert *et al.* 2017). In this study, we report

the genome sequences of 13 isolates of *Z. ardabiliae* that originate from wild grasses collected in the province of Ardabil in Iran (Table S1). DNA extraction was performed as previously described (Stukenbrock *et al.* 2011). Library preparation and paired-end sequencing using an Illumina HiSeq2000 platform were conducted at Aros, Skejby, Denmark.

The 13 resequenced *Z. ardabiliae* genomes were assembled from 100-bp, paired-end reads using the *de novo* assembly algorithm of the CLC Genomics Workbench version 5.5 (QIAGEN, Aarhus, Denmark). The assemblies were created using standard settings for paired-end reads. We used a previously published RNA-sequencing-based annotation to distinguish the parameter estimates for coding and noncoding sequences (Grandaubert *et al.* 2015). To predict the genes that encode effectors, we used the software EffectorP (Sperschneider *et al.* 2016), with default settings, on genes predicted by SignalP (Petersen *et al.* 2011) to encode a secreted protein.

Genome alignment and SNP calling

Genome alignments were separately created for each population using the MultiZ program from the TBA package (Blanchette *et al.* 2004). Default parameters were used, although LastZ was used instead of BlastZ for pairwise alignments. Genome alignments were projected against the two reference genomes of each species: IPO323 for *Z. tritici* and STO4IR-1.1.1 for *Z. ardabiliae* (Goodwin *et al.* 2011; Stukenbrock *et al.* 2011). The projected alignments in MAF format were filtered using the MafFilter program (Dutheil *et al.* 2014) with the following filters: (1) each syntenic block was realigned using MAFFT (Katoh *et al.* 2009), and blocks with >10 kb were split for computational efficiency; (2) only blocks where all individuals were present were retained (13 *Z. tritici* and 17 *Z. ardabiliae*); (3) a window of 10 bp was slid by 1 bp, and windows containing at least two indel events were discarded and the containing blocks were split; (4) a window of 10 bp was slid by 1 bp, and windows with a total of >100 gap characters were discarded and the containing blocks were split; and (5) all blocks were merged according to the reference genome with empty positions filled by "N," which resulted in one masked alignment per chromosome for *Z. tritici* and one masked alignment per contig for *Z. ardabiliae*. The chromosome and contig alignments were further divided into nonoverlapping windows of 1 Mb (data set 1) or 100 kb (data set 2). The MafFilter program was further used to estimate statistics on the alignments at each filtering step, and to compute the nucleotide diversity (Watterson's θ) from the final filtered genome alignments.

Estimating recombination

Filtered alignments (1-Mb windows, data set 1) were exported as fasta files for the LDhat and LDhelmet packages. The program "convert" from the LDhat package was used to convert fasta files into input loci files for the program "interval" (Auton

Table 1 Summary of genome alignment processing and whole-genome SNP analyses for *Z. tritici* and *Z. ardabiliae*

	Z. tritici		*Z. ardabiliae*	
Size of sequenced reference genome (bp)	39,686,251		31,546,591	
Number of exonic sites in reference genome (bp)	17,296,247 (43.6%)		15,570,421 (49.4%)	
Number of haplotypes	13		17	
Summary genome alignment	Total alignment length (Mb)	Number of alignment blocks	Total alignment length (Mb)	Number of alignment blocks
MultiZ alignment	40.8	21,500	32.4	22,296
Splitting in maximum 10 kb	40.8	21,904	32.4	23,001
MAFFT realignment	40.5	21,904	32.2	23,001
Keep blocks with all strains	27.7	6,445	28.2	7,117
Filter 1	27.5	15,703	28.0	18,402
Filter 2	27.3	18,785	27.7	26,074
Percentage of repeated sequences in initial alignment (%)	19.74		3.36	
Percentage of repeated sequences in final alignment (%)	0.93		1.38	
Total number of SNPs	1,483,950		1,069,014	
Total number of analyzed SNPs (biallelic, no unresolved state) and percent of total SNPs	1,438,385 (96.9%)		1,035,158 (96.8%)	
Total number of SNPs in exons and percent of total SNPs	713,733 (48.1%)		403,895 (37.8%)	
Total number of analyzed SNPs in exons (biallelic, no unresolved state), and percent of total analyzed SNPs in exons	690,096 (96.7%)		396,247 (98.1%)	
Summary SNP analyses	1-Mb windows	100-kb windows	1-Mb windows	100-kb windows
Minimum number of SNPs	143	0	0	0
Median number of SNPs	43,680	3,556	1,598	634
Maximum number of SNPs	1,02,400	15,170	33,680	20,110
Diversity (median of Watterson's θ in windows of 10 kb)	0.0139		0.008663	

and McVean 2007). Only fully resolved biallelic positions were exported (see Table 1 for the details of SNP numbers). Likelihood tables were generated for θ values of 0.0005, 0.005, and 0.05. The interval program was run with 10,000,000 iterations and sampled every 5000 iterations with a burn-in of 100,000 iterations. LDhelmet was run with the parameters suggested in the user manual (Chan *et al.* 2012; https://sourceforge.net/projects/ldhelmet/). Comparison of recombination maps on the same set of SNPs was performed using standard principal component analysis, as implemented in the R package ade4 (Dray and Dufour 2007). A table was computed with one column per method (LDhat and LDhelmet, each with θ set to 0.0005, 0.005, or 0.05) and one line per analyzed SNP pair, and the two first principal components were kept to plot a correlation circle (Figure 1A).

To assess the robustness of the recombination maps, alternative maps for *Z. tritici* were constructed using the same protocol (1) after discarding all singletons, and (2) after removing five individuals to ensure absence of population structure. All maps were compared to the previously published genetic map of Croll *et al.* (2015) in windows of 20 kb. Correlations were assessed using Kendall's rank correlation test, and confidence intervals were obtained by bootstrapping windows.

We calculated average recombination rates in windows and regions by taking the average of recombination estimates

between every pair of SNPs, weighted by the physical distance between the SNPs. Pairs of SNPs for which the confidence interval of the recombination estimate was higher than two times the mean were discarded and therefore not used in the average computation. Using the gene annotations available for the two reference species (Grandaubert *et al.* 2015), we calculated the following information for each gene: (1) the average recombination rate in exons, (2) the average recombination rate in introns, (3) the average recombination rate in the 500 bp flanking the 5′ region, and (4) in the 500 bp flanking the 3′ region. We also calculated the average recombination rate for each intergenic region (500 bp from/to genes). GFF3 files from Grandaubert *et al.* (2015) were retrieved and processed using the GenomeTools package to add intron annotations (Gremme *et al.* 2013). The resulting gene annotations were analyzed in R together with recombination maps (R Core Team 2013).

Assessment of LD-based recombination estimates by simulation

We used the SCRM coalescent simulator (Staab *et al.* 2015) to simulate polymorphism data with a constant mutation rate but variable recombination rate. Recombination rates were drawn randomly from an exponential distribution with a mean of 0.02. Segments with a piecewise constant
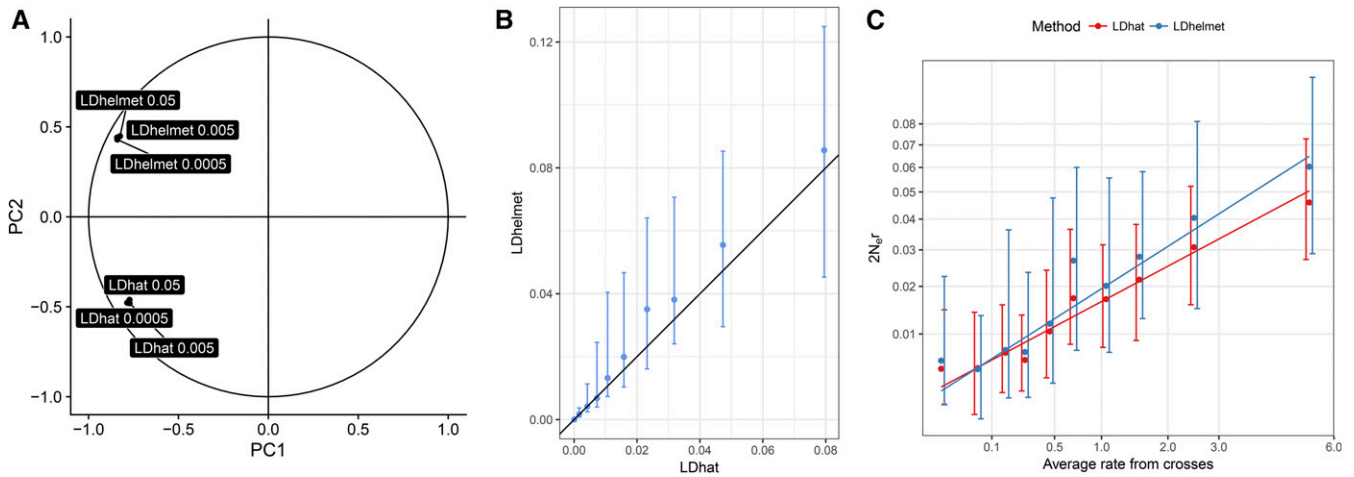
Figure 1 Correlations among recombination maps in *Z. tritici* show highly correlated estimates from two composite likelihood methods. (A) Correlation circle of the six population genomic recombination maps based on the two first factors of a principal component (PC) analysis. The programs LDhat interval (Auton and McVean 2007) and LDhelmet (Chan *et al.* 2012) were used with three distinct input-scaled effective population sizes (θ) of 0.0005, 0.005, and 0.05. (B) Correlation of the LDhat and LDhelmet maps with θ = 0.005. The LDhat map was discretized into 10 categories with equal numbers of points. The points and error bars represent the median and first and third quartile of the distribution for each category. (C) To assess the quality of the inferred recombination maps, genome-wide estimates of recombination were correlated with a genetic map obtained by experimental crossing of *Z. tritici* isolates. *y*-Axis: Population genomic maps were obtained by LDhat and LDhelmet with a scaled population size of 0.005. *x*-Axis: average recombination map from two independent crosses (Croll *et al.* 2015). Points and error bars represent the median and first and third quartile of the distribution for each category, obtained as in (B).

recombination rate were taken randomly from an exponential distribution with a mean of 100 kb. Sample sizes of 10, 30, and 100 individuals were tested for comparison with a population mutation rate equal to 0.05, 0.005, 0.0005, and 0.00005. We generated a locus of 10 Mb for simulations with θ equal to 0.005, 0.0005, and 0.00005; but only 1 Mb for simulations with θ equal to 0.05, as the resulting output file from LDhat would otherwise become excessively large due to the high number of SNPs. The true recombination rate used at each position of the alignment was recorded for later comparison. The output of SCRM was converted to LDhat input format using Python scripts. Recombination rates were estimated using the interval program from the LDhat package (Auton and McVean 2007). For simulations with θ = 0.05 and 0.005, a likelihood lookup table with θ = 0.01 was used; whereas a lookup table with θ = 0.001 was used for simulations with θ = 0.0005 and 0.00005. The inferred recombination rate at each position was then compared to the true rate. A variant of this simulation procedure was used to assess the impact of population structure on the inference of recombination rate. The SCRM coalescent simulator was used with a five-islands population model, with sample sizes of 2, 3, 4, 5, and 6 per deme, resulting in a total of 20 individuals. Migration rates were assumed to be all identical between demes, and values of $M = 4N_e \times m = 1$, 10, and 100 were tested. Regions of 1 Mb were simulated with θ = 0.005 for each migration rate.

Reference species alignment and comparison

The two reference strains IPO323 (*Z. tritici*) and ST11IR-11.4.1 (*Z. ardabiliae*) were aligned using LastZ (Blanchette

et al. 2004). The resulting genome alignment was used to map the coordinates of *Z. ardabiliae* SNPs to the *Z. tritici* genome, using the MafFilters LiftOver filter (Dutheil *et al.* 2014). A total of 893,171 (86%) positions could be mapped from *Z. ardabiliae* to *Z. tritici* and were used for further analyses. Nonoverlapping windows containing at least 100 analyzed SNPs in each species were generated for the comparison of recombination rates between the two species.

Multi-scale correlations

We calculated the average recombination rates in windows of varying sizes and retained only windows that contained at least 1% of the polymorphic positions. To enforce a similar statistical power among different window sizes, a number of windows were chosen randomly. The same number of randomly chosen windows was used for the distinct comparisons. To assess the sampling variance, 1000 independent samplings (with replacement) were performed for each window size. Window sizes of 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024 kb were tested, with 27 windows sampled in each case. We measured correlation coefficients using the Spearman, Kendall, and Pearson's correlation coefficients. Spearman and Kendall's coefficients are ranked based; therefore they do not assume binormality as Pearson's coefficient does. Because recombination rates are typically exponentially distributed, Pearson's coefficient was measured for the log rates instead of the raw ρ rates. Spearman's coefficient assumes that the variables are continuously distributed; therefore it does not resolve ties. Thus "jittering" was used to randomly resolve ties in the input variables (R function jitter, with

default parameters). Conversely, Kendall's coefficient assumes ordinal input variables. Therefore, using the three correlation measures allows assessing the robustness of the correlation signal. A graphical representation was performed using the ggplot2 package for R, which performed local polynomial regression fitting for the curves (Wickham 2016).

Mapping of hotspots

Hotspots were detected using the LDhot program (Auton *et al.* 2014). For computational efficiency, LDhot was run on the 100-kb alignments (data set 2). A background recombination map was first estimated for each alignment using the interval program of LDhat with a θ value of 0.005 (Auton and McVean 2007). The resulting maps were highly correlated with the maps based on 1-Mb alignments, and showed little effect of the discretization scheme. The background recombination map was used as input to LDhot with default parameter values and 1000 simulations.

Significant hotspots were filtered for further analysis. First, only the hotspots with a value of ρ between 5 and 100 across the hotspot coordinates were selected, because higher values are most likely artifacts and the performance of LDhot is low for weak hotspots (Auton *et al.* 2014). A few hotspots with extremely large sizes (>2 kb) were further discarded. This process identified 9133 hotspots in *Z. tritici* and 1287 hotspots in *Z. ardabiliae*. We calculated the mean background rate in each detected hotspot and in the two 20-kb flanking regions. We further selected hotspots for which the within-hotspot rate was at least 10 times higher than the flanking regions. Thus 2578 and 862 hotspots were identified in *Z. tritici* and *Z. ardabiliae*, respectively. The *Z. ardabiliae* hotspots were mapped onto the *Z. tritici* genome using MafFilter's LiftOver function (Dutheil *et al.* 2014). We considered a hotspot in *Z. tritici* as colocalizing with a hotspot in *Z. ardabiliae* if the distance between them was <1 kb, and if no other hotspot was found between the two. We compared statistics on the distribution of hotspots by randomizing the hotspot positions while keeping their original size, for each chromosome independently. To do so, we used the following procedure:

1. Compute the total "interhotspots" distance, L, as the sum of all distances between consecutive hotspots.
2. Draw random distinct positions uniformly in $[1, L]$. These positions are the starting positions of each randomized interval.
3. Order and then expand each interval to match its original size and compute the corresponding end positions. Correct the coordinates to account for previous intervals.

To account for variable coverage along the genome, we also simulated intervals corresponding to chromosome regions that were not included in our analysis, using the same procedure as for hotspot randomization. Each randomized set of hotspots therefore contains the same amount of callable sites as the actual analysis. We assessed the

significance of the number of colocalizing hotspots using 10,000 permutations.

Models of GC-content evolution

The two reference strains IPO323 (*Z. tritici*) and ST11IR-11.4.1 (*Z. ardabiliae*) were aligned using LastZ (Blanchette *et al.* 2004). Several filtering steps were further applied to the alignment. First, each synteny block was realigned using the MAFFT aligner (Katoh *et al.* 2009) after splitting blocks >10 kb for computational efficiency, which resulted in an alignment of 27,918,318 bp that included both species. Second, a window of 30 bp was slid by 1 bp along the alignment. Windows with >29 gaps in total between the two species were further discarded, which resulted in 27,237,601 filtered positions. To minimize the effect of selection on GC patterns, we further discarded regions in the alignment that were annotated as protein-coding genes in one or both species. This resulted in a total alignment of 9,143,114 bp. The alignment was further divided into windows ranging from 1 to 4 kb and only data from the essential chromosomes (*Z. tritici* chromosomes 1–13) were retained. The final alignment contained 2052 cleaned windows containing sequences for both species with no synteny break, and it encompassed 3,179,581 bp. A model of sequence evolution was independently fitted on each window using maximum likelihood (Dutheil and Boussau 2008). The HKY85 model was used as a basis allowing three frequency parameters $[(G + C)/(A + C + G + T), A/(A + T),$ and $G/(G + C)]$ in addition to the transition over transversion ratio (Hasegawa *et al.* 1985). We fitted a nonhomogeneous, nonstationary model of substitution, allowing us to estimate three distinct GC contents for *Z. tritici*, *Z. ardabiliae* and their common ancestor. Other parameters were considered constant between species and their ancestor. A molecular clock was assumed (so that the two branches leading to *Z. tritici* and *Z. ardabiliae* were equal in length) and a four-class gamma distribution of rates with a shape parameter fixed to 0.5 was used. We further calculated the observed GC content in each species for each window. The average recombination rate was calculated for each window containing at least 1% polymorphic positions (leaving 1642 windows).

A similar analysis was conducted using recombination rate estimated from Croll *et al.* (2015), which was calculated in 20 kb windows. The corresponding pairwise alignment regions were extracted and filtered, and coding regions from both species were discarded; resulting in 1948 windows of at least 1 kb where a nonhomogeneous, nonstationary model of substitution could be fitted.

Data availability

Sequence data has been deposited under the National Center for Biotechnology Information (NCBI) Illumina reads for *Z. ardabiliae* are available from NCBI under the Biosample accession numbers SAMN05818736–SAMN05818752. Illumina reads for *Z. tritici* are available from NCBI under the

BioProject accession number PRJNA312067. All scripts and data sets necessary to reproduce the analyses and figures in this manuscript may be accessed on FigShare under https://doi.org/10.6084/m9.figshare.3806244.v1.

Results and Discussion

Genome alignments and SNP calling

A total of 30 whole haploid genome sequences was used to infer the recombination landscapes of the two species *Z. tritici* and *Z. ardabiliae*. First, we generated *de novo* genome assemblies of 10 *Z. tritici* and 13 *Z. ardabiliae* isolates previously not studied (Table S1). The haploid genomes, including an additional three *Z. tritici* and four *Z. ardabiliae* genomes already published (Stukenbrock *et al.* 2011), were aligned for each species; resulting in multiple genome alignments of 40.8 Mb for *Z. tritici* and 32.4 Mb for *Z. ardabiliae*. (Table 1)

Recombination analyses rely on SNP data. However, erroneously called SNPs or alignment errors can greatly bias LD inference. To generate high-quality SNP data sets, we therefore filtered the genome alignments (see *Materials and Methods*) to retain only the alignment blocks in which all isolates were represented. This filtering yielded genome alignments of 27.7 and 28.2 Mb for *Z. tritici* and *Z. ardabiliae*, respectively (Table 1). We further filtered the alignments to mask ambiguously aligned positions, leading to a final alignment size of 27.3 Mb for *Z. tritici* and 27.7 Mb for *Z. ardabiliae*. Less than 2% of the final alignment contained repetitive sequences, including TEs. In the case of *Z. tritici*, repeat regions have been filtered out during the alignment quality checking; while in the case of *Z. ardabiliae*, for which no telomere-to-telomere sequencing is available, most repeats were poorly assembled and therefore virtually absent from the original alignment (Table 1). After filtering we identified 1.48 million SNPs in *Z. tritici* and 1.07 million SNPs in *Z. ardabiliae*, which corresponds to the nucleotide diversities, measured as Watterson's θ, of 0.0139 in *Z. tritici* and 0.0087 in *Z. ardabiliae* (Table 1). Thus, despite the larger sample size, *Z. ardabiliae* shows a much lower SNP density and sequence diversity than the wheat pathogen *Z. tritici*.

Inference of fine-scale recombination maps

We estimated and compared the local recombination rates in *Z. tritici* and *Z. ardabiliae* using two methods implemented in the LDhat (Auton and McVean 2007) and LDhelmet (Chan *et al.* 2012) packages. Both methods estimate the local population recombination rates based on the LD between SNPs in a given genome data set using a composite likelihood method. The methods infer the population-scaled recombination rate ρ across the genome, based on an *a priori* specified population mutation rate θ. The parameter ρ relates to the actual recombination frequency by the equation $\rho = 2N_e \times r$ for haploid individuals, where N_e is the effective population size and r is the per site rate of recombination

per generation across the region. Inferring r from ρ therefore requires knowledge of N_e. Furthermore, in *Zymoseptoria* species, sexual reproduction is not obligatory and may vary from year to year with environmental conditions and the availability of compatible hosts and mating partners, rendering the estimation of r very difficult without any additional knowledge of the amount of clonal reproduction. To avoid the bias of incorrect assumptions, we therefore further analyzed and compared the recombination maps of *Z. tritici* and *Z. ardabiliae* based on the parameter ρ.

As θ substantially varies along genomes and between species, we generated recombination maps using three scaled effective population size values as inputs ($\theta = 0.05, 0.005,$ and 0.0005). For both LDhat and LDhelmet, we find that the three different input θ values only have a marginal influence on the recombination rate estimates obtained (Figure 1A). We therefore proceeded with the recombination map estimated using a θ of 0.005, similar to the median of θ values estimated in 10-kb windows in *Z. tritici* ($\theta = 0.0139$) and in *Z. ardabiliae* ($\theta = 0.0087$) (Table 1).

To assess the performance of the two methods and input parameters for the fungal data sets, we first compared the inferred recombination maps of *Z. tritici* with data from previously published genetic maps (Croll *et al.* 2015). We compared both the LDhat and LDhelmet recombination maps with the genetic maps created from two sexual crosses of Swiss *Z. tritici* isolates, 3D7×3D1 and SW5×SW39 (Croll *et al.* 2015). The recombination maps estimated by LDhat and LDhelmet from SNP data both correlate with the genetic maps, confirming that the composite likelihood methods allow us to assess the recombination landscapes in the fungal pathogens (Figure 1B and Table 2). We find a significant correlation between the LDhat map and the two genetic maps (3D7×3D1: Kendall's rank correlation test, $\tau = 0.27$, and $P < 2.2e^{-16}$; SW5×SW39: Kendall's rank correlation test, $\tau = 0.23$, and $P < 2.2e^{-16}$). Using an average recombination rate of the 3D7×3D1 and SW5×SW39 crosses, the correlation further increases (Kendall's rank correlation test, $\tau = 0.29$, $P < 2.2e^{-16}$) (Figure 1B and Table 2). While correlated, the new recombination maps of *Z. tritici* encompasses >1 million SNPs and thereby provides a considerably finer resolution of the recombination landscape in *Z. tritici* than previously obtained from experimental crosses (based on ~23,000 SNPs) (Croll *et al.* 2015). The same correlation analyses using the LDhelmet map show consistent results with slightly lower correlations (Kendall's rank correlation test, $\tau = 0.24$ for the cross 3D7×3D1, 0.20 for the cross SW5×SW39, and 0.25 using the average of the two crosses; all $P < 2.2e^{-16}$) (Table 2). These correlations, although highly significant, have relatively small size effects. However, it is also noteworthy that the correlation between the two Swiss crosses 3D7×3D1 and SW5×SW39 is only 0.43 (Kendall's rank correlation test, $P < 2.2e^{-16}$), supporting a high variability in recombination even between individual

Table 2 Robustness of the population recombination map and correlation with crossover maps

Data	Map	3D7×3D1			SW5×SW39			Average		
		Correlation	*P*-value	C.I.	Correlation	*P*-value	C.I.	Correlation	*P*-value	C.I.
Unfiltered	LDhat	0.27	$2.22e^{-60}$	[0.24, 0.30]	0.23	$5.11e^{-44}$	[0.20, 0.26]	0.29	$4.61e^{-69}$	[0.26, 0.32]
	LDhelmet	0.24	$1.68e^{-47}$	[0.21, 0.27]	0.20	$8.18e^{-33}$	[0.17, 0.23]	0.25	$1.11e^{-52}$	[0.22, 0.28]
	Average	0.26	$2.59e^{-56}$	[0.23, 0.29]	0.22	$9.81e^{-40}$	[0.19, 0.25]	0.27	$4.71e^{-63}$	[0.24, 0.30]
	LDhat intergenic	0.20	$9.54e^{-32}$	[0.17, 0.24]	0.18	$9.54e^{-25}$	[0.15, 0.21]	0.22	$2.65e^{-37}$	[0.19, 0.25]
	LDhelmet intergenic	0.22	$9.73e^{-37}$	[0.19, 0.25]	0.19	$1.89e^{-28}$	[0.16, 0.23]	0.23	$1.79e^{-42}$	[0.20, 0.27]
	LDhat no singleton	0.21	$1.18e^{-36}$	[0.18, 0.24]	0.17	$1.81e^{-24}$	[0.14, 0.20]	0.21	$6.88e^{-39}$	[0.18, 0.24]
	LDhat no structure	0.25	$7.77e^{-54}$	[0.22, 0.29]	0.23	$7.90e^{-43}$	[0.19, 0.26]	0.26	$1.48e^{-59}$	[0.23, 0.29]
Filtered	LDhat	0.31	$3.36e^{-76}$	[0.27, 0.33]	0.28	$1.94e^{-64}$	[0.25, 0.31]	0.34	$2.10e^{-96}$	[0.31, 0.37]
	LDhelmet	0.26	$4.45e^{-57}$	[0.23, 0.29]	0.22	$3.35e^{-40}$	[0.19, 0.25]	0.28	$1.34e^{-64}$	[0.25, 0.31]
	Average	0.29	$9.54e^{-69}$	[0.26, 0.32]	0.25	$2.62e^{-51}$	[0.22, 0.28]	0.31	$5.65e^{-80}$	[0.28, 0.34]
	LDhat intergenic	0.20	$5.39e^{-30}$	[0.17, 0.23]	0.18	$3.70e^{-24}$	[0.15, 0.21]	0.22	$1.90e^{-36}$	[0.19, 0.25]
	LDhelmet intergenic	0.23	$1.81e^{-37}$	[0.19, 0.26]	0.19	$5.06e^{-26}$	[0.15, 0.22]	0.24	$6.40e^{-42}$	[0.20, 0.27]
	LDhat no singleton	0.29	$6.92e^{-66}$	[0.25, 0.32]	0.25	$6.22e^{-52}$	[0.22, 0.28]	0.31	$8.55e^{-79}$	[0.28, 0.34]
	LDhat no structure	0.29	$1.24e^{-71}$	[0.26, 0.32]	0.29	$2.58e^{-67}$	[0.26, 0.32]	0.32	$3.66e^{-88}$	[0.29, 0.35]

Correlation values are Kendall's τ. SW5×SW39 and 3D7×3D1 correspond to crosses in Croll *et al.* (2015). C.I., 95% confidence interval obtained by 10,000 bootstraps.

crosses of *Z. tritici* (Table 2). Based on the comparison of the outputs of LDhat and LDhelmet, we decided to use the LDhat map as our reference population map for the remainder of this study. We next investigated the impact of possible confounding factors on the recombination rate estimates, including SNP densities, possible sequencing errors, population structure, and natural selection.

SNP density and filtering based on confidence intervals: LDhat and LDhelmet have been developed for recombination analyses in animals (Auton and McVean 2007; Auton *et al.* 2012; Chan *et al.* 2012), and their performance on data from haploid eukaryotes with high recombination rates has not been tested. Therefore, we next assessed the robustness of the composite likelihood approach using simulations with distinct sample sizes and SNP densities. We report that the interval program infers recombination rate with the highest reliability for intermediate diversity levels (θ = 0.0005 or 0.005). Furthermore, while larger sample sizes decrease the variance in the estimate, we show that LDhat reliably infers recombination when as few as 10 haploid genomes are used (Figure 2). We observe that ρ generally tends to be underestimated, and its estimation variance is larger for small sample sizes. However, better estimates can be obtained by discarding all estimates where the width of the 95% confidence interval is larger than or equal to two times the mean. Interestingly, this filtering has the strongest effect for highly diverse regions (θ = 0.05), where the raw estimates of LDhat appear to be highly underestimated even for large sample sizes (*n* = 100). Discarding estimates with large confidence intervals efficiently suppresses this bias (Figure 2). We also note that the inference bias is stronger for low recombination rates, and that this effect is independent of the sample size (Figure 2). Based on these simulation results, we similarly filtered our recombination estimates based on the 95% confidence interval reported by LDhat.

This filtering discards 49 and 20% of all SNP pairs for *Z. tritici* and *Z. ardabiliae*, respectively. The large difference between the two data sets is imputable to the much higher nucleotide diversity of *Z. tritici*. When compared with the genetic map (Croll *et al.* 2015), the filtered map of *Z. tritici* shows a correlation of 0.34 (Kendall's rank correlation test, $P < 2.2e^{-16}$; Table 2). Interestingly, correlations between the genetic map and the LD map inferred with LDhat increase with increased window size: using 500-kb windows, the correlation becomes 0.43 (Kendall's τ, $P = 0.000206$).

Putative sequencing errors: Sequencing errors can affect LD estimates as they appear as unlinked singletons in population genomic data sets. Such potential effects are partially accounted for at two levels in our analyses. First, our data sets are based on *de novo* assembly of each individual genome, which already corrects for putative sequencing errors in the sequencing read outs. Second, many of the SNPs discarded for having a large confidence interval in the estimation of recombination rates by LDhat are singletons. To further assess the potential impact of putative sequencing errors, we ran LDhat on the *Z. tritici* data set after discarding all singletons and filtering for confidence interval as described above. The resulting recombination map appeared to be highly correlated to the map including all singletons (Figure S1 and Table 2), and the correlation of the filtered map with the crossover map was significant, but weaker than when including them (Kendall's τ = 0.3083, $P < 2.2e^{-16}$) (Table 2). We therefore conclude that putative sequencing errors have no significant impact on our inferred recombination maps.

Effect of population structure: Previous studies reported that *Z. tritici* strains are sampled from a globally panmictic population (Linde *et al.* 2002; Zhan *et al.* 2003). However, in a recent study based on whole-genome data, we report evidence for slight population structure, notably between
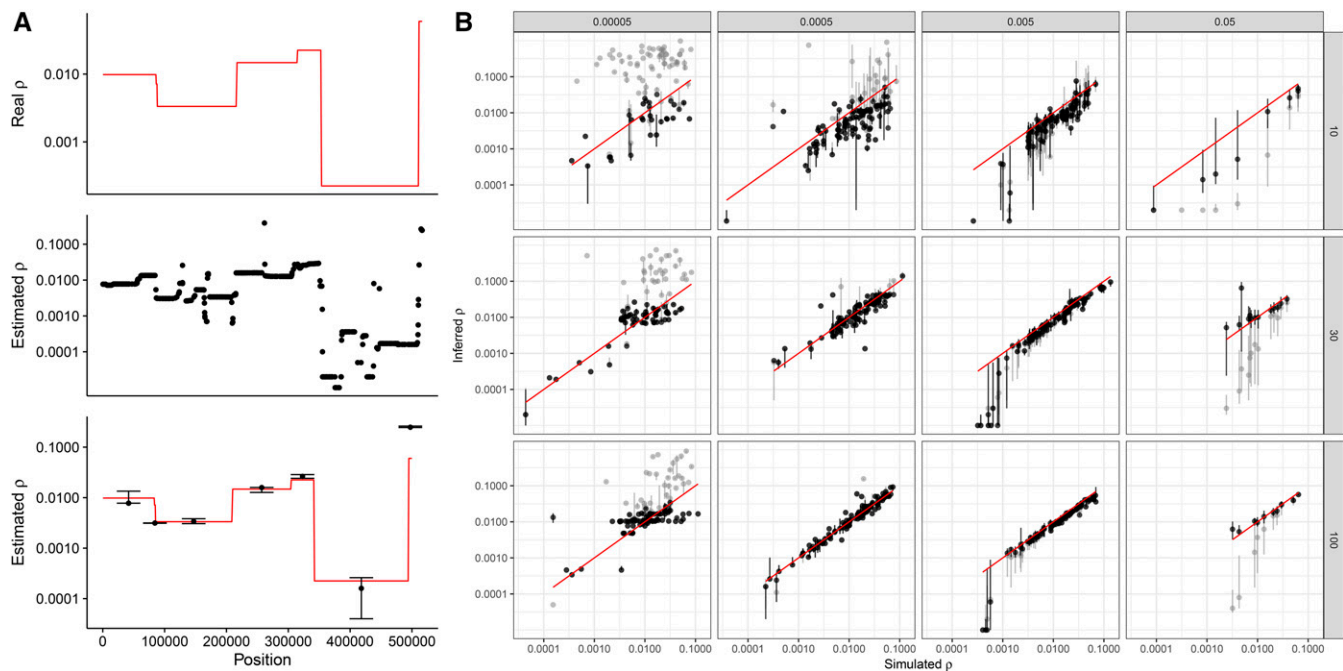
Figure 2 Effect of sample size and diversity on the estimation of recombination rate by LDhat. Regions of 10 Mb (1 Mb for regions with θ = 0.05) were simulated using a coalescent model with variable recombination rate: random segments were generated by sampling lengths from an exponential distribution and rates from the observed distribution of recombination rates. (A) Example of a 500-kb region, with variable recombination rate (red line), LDhat estimates between pairs of SNPs (middle panel), and median (with first and third quartiles as error bars) for each segment of uniform recombination (bottom panel). (B) Inferred *vs.* true recombination rate for different nucleotide diversity values (θ = 2N_e r) and sample sizes. Each ● corresponds to a region with constant recombination rate in the simulated alignment, as shown in (A). Bars indicate the first and third quartiles of LDhat estimates for the region. Gray points are raw estimates; black points are computed from filtered estimates (see *Materials and Methods*). The red diagonal line shows the 1:1 ratio. Columns indicate distinct population mutation rates and rows distinct sample sizes (number of haploid genomes).

Iranian isolates *vs.* European isolates (Grandaubert *et al.* 2017). To assess whether this structure could bias our recombination estimates, we generated a new recombination map using LDhat on a reduced sample of eight strains of *Z. tritici*. We excluded the two Iranian isolates in our data set as well as three German isolates forming a separate, yet nonsignificant, cluster. We report that the resulting map is highly correlated to our recombination map (Figure S1) as well as to the previously published genetic map (Kendall's τ = 0.3237, $P < 2.2e^{-16}$) (Table 2), suggesting that population structure has little effect on our inference of recombination rate. Because the resulting correlation with the crossover map was slightly lower than when using the complete data set, we use the complete data set for the further analyses.

As little is known about the population structure of *Z. ardabiliae*, we conducted additional simulations to assess the putative impact of structure on the inference of recombination rate. We used a five-islands structure model, with sample sizes equal to 2, 3, 4, 5, and 6 in each deme, with a total sample size of 20; which is comparable to the 17 genomes of the *Z. ardabiliae* data set analyzed here. Migration rates between demes were symmetrical and all equal, and we tested several rates. We find that while ρ is systematically overestimated in the presence of population structure,

it is remarkably proportional to the true value, in particular after filtering on the confidence intervals (Figure S2). Population structure, if any, is therefore not expected to bias our comparison of recombination rates along the genome. In addition, these results suggest that the true recombination rate in *Z. ardabiliae* is potentially even lower than the value reported here.

Coding sequences: Recombination inference based on patterns of LD is affected by various patterns of selection. The genomes of *Z. tritici* and *Z. ardabiliae* are gene dense and protein-coding genes occupy nearly 50% of the sequences. We therefore considered the impact of selection on our recombination inference in the two species, assuming lower selection in noncoding regions. To this end, we compared the previously published crossover map with estimates of ρ exclusively in the intergenic regions, excluding coding sequences and 500 bp up- and downstream of the annotated genes (Figure S1 and Table 2). These analyses based on noncoding sequences and filtering of SNPs based on the confidence interval of recombination rate estimates resulted in correlations of 0.22 for the LDhat map and the average of the two genetic crosses (Kendall's rank correlation test, $P < 2.2e^{-16}$) and 0.24 for the LDhelmet map (Kendall's rank correlation test, $P < 2.2e^{-16}$). Thus, the
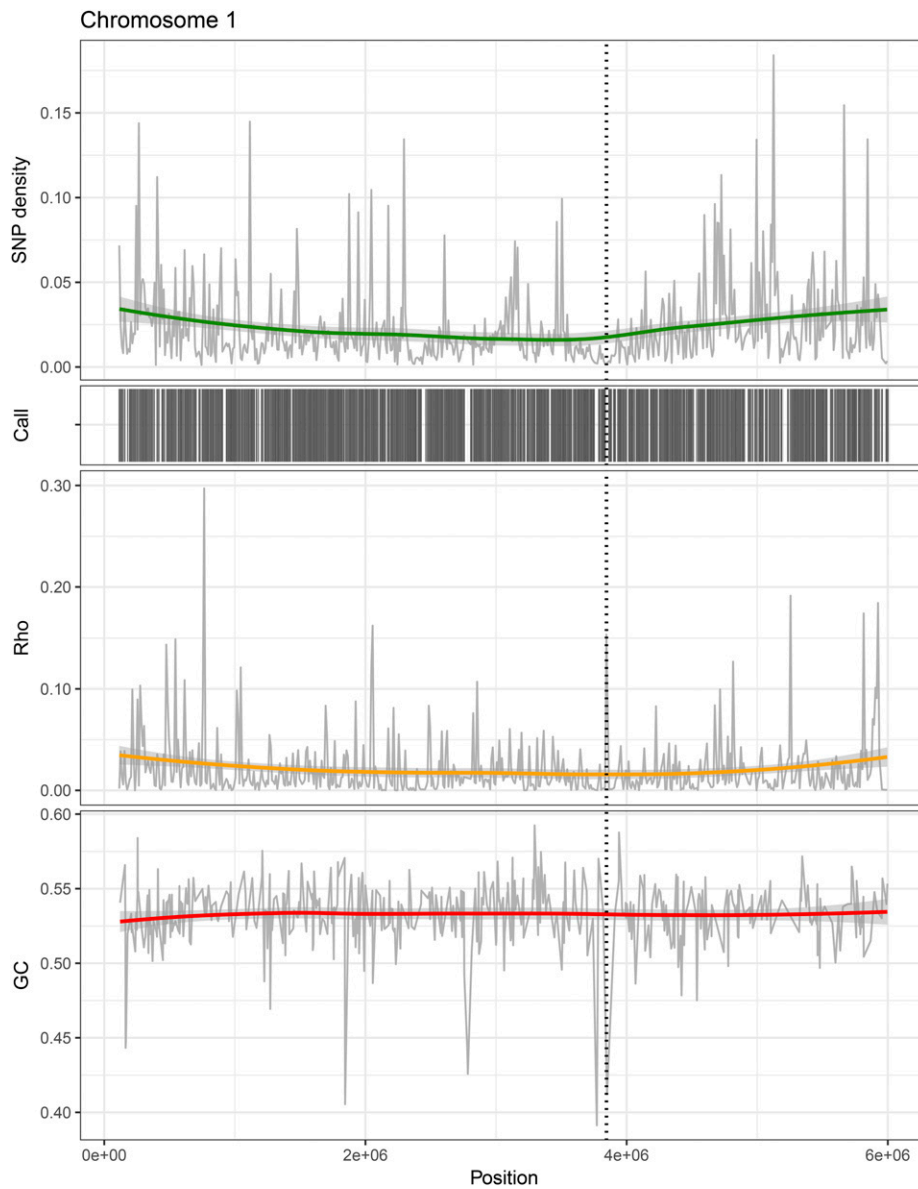
Figure 3 Variation in recombination rate across chromosomes. Based on the population genomics data of *Z. tritici* and *Z. ardabiliae*, genome-wide patterns of recombination are estimated. Patterns of variation across chromosome 1 of *Z. tritici* is shown as an example. (Upper panel) SNP density in 10-kb windows with corresponding smoothing curve. (Middle panel) Distribution of called sites along the chromosome in black, corresponding to the regions that were included in the analyses. (Lower panel) Estimates of the population recombination rate ρ show a highly heterogeneous, small-scale recombination landscape across the chromosomes. (D) Observed GC content. The position of the centromere is marked over the chromosome plots as a vertical stippled line.

best correlations of LD based on the recombination maps and genetic crosses are obtained when coding regions are included (Table 2). The finding suggests that the composite likelihood method provides robust estimates of recombination, even in regions likely to deviate from purely neutral evolution. Based on these simulation results, we chose to use the LDhat-inferred recombination rates on the full genome, with an input θ of 0.005 and filtered according to confidence intervals for both *Z. tritici* and *Z. ardabiliae*.

A fivefold higher population-scaled recombination rate in Z. tritici

The inference of ρ across the genomes of *Z. tritici* and *Z. ardabiliae* reveals highly heterogeneous recombination landscapes in both species (Figure 3 and File S1). We find a fivefold higher recombination rate in *Z. tritici* than in *Z. ardabiliae*: the mean values of ρ are 0.0217 and 0.0045 for *Z. tritici* and *Z. ardabiliae*, respectively. As $\rho = 2N_e \times r$, where r is the actual recombination rate

per generation per nucleotide and N_e is the effective population size, this fivefold difference might reflect differences in r or global differences in N_e. Furthermore, the inferred parameter ρ reflects the historical rates of recombination in the two species, which may have varied according to different demographic events since their divergence. Nonetheless, the nucleotide diversity estimated by Watterson's θ is 1.6 times higher in *Z. tritici* than in *Z. ardabiliae*, indicating that different population sizes alone cannot explain the observed difference in recombination rates, assuming that the two species have comparable mutation rates. The higher value of ρ estimated in *Z. tritici* thus likely reflects a higher actual recombination rate (in the past or presently) in the wheat pathogen compared to *Z. ardabiliae*.

Recombination on small arms of acrocentric chromosomes

Physical factors, such as chromosome length, centromere position, or distance to the centromere, have been reported
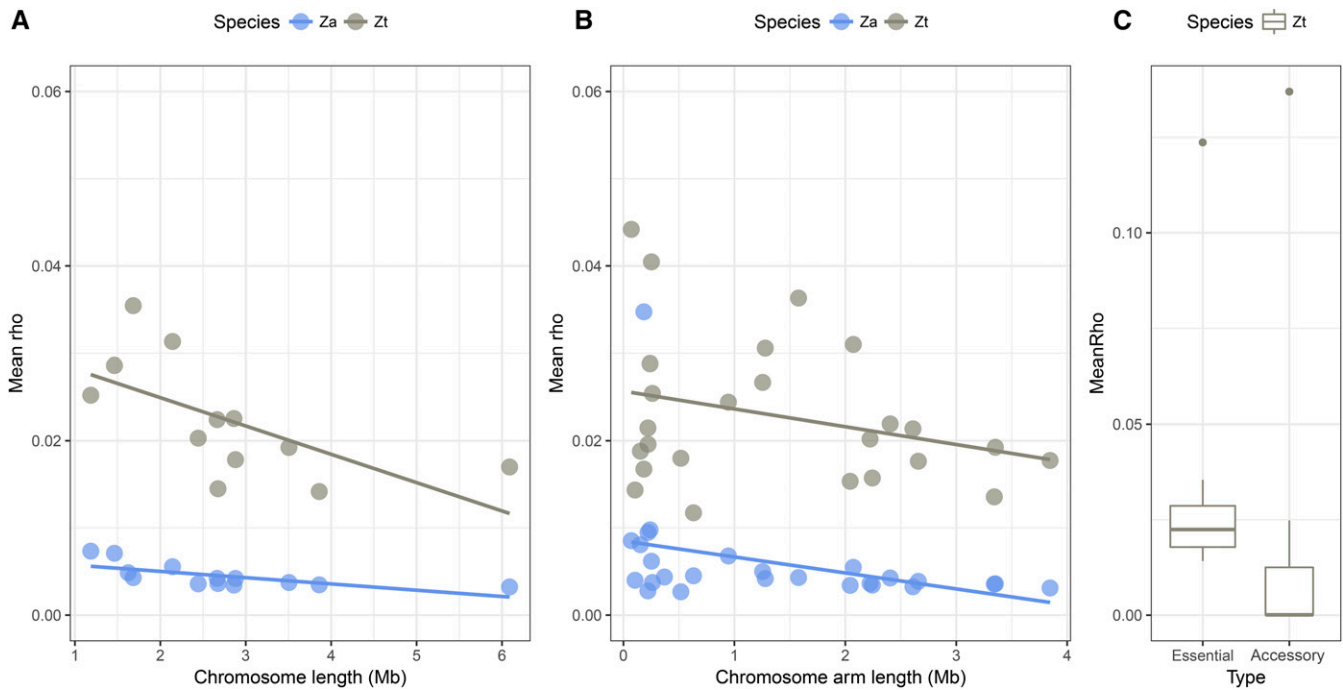
Figure 4 Broadscaled patterns of recombination rate in *Z. tritici* and *Z. ardabiliae* demonstrate a strong effect of chromosome length and type. (A) Mean recombination rate in *Z. tritici* and *Z. ardabiliae* per essential chromosome as a function of the chromosome length. (B) Mean recombination rate per essential chromosome arm as a function of the arm length. (C) Distribution of mean recombination rate per chromosome in *Z. tritici* as a function of type (essential or accessory). Za, *Z. ardabiliae*; Zt, *Z. tritici*.

to affect broadscale recombination patterns in eukaryotes (Jensen-Seaman *et al.* 2004). To investigate the rate and distribution of crossover events along the genomes of the two *Zymoseptoria* species, we correlated the inferred recombination maps with features of the well-characterized karyotype of *Z. tritici*. The reference genome sequence of *Z. tritici* consists of 21 fully sequenced chromosomes, including 8 so-called accessory chromosomes showing presence/absence polymorphisms between individuals (Goodwin *et al.* 2011). Furthermore, the exact positions of the centromeres for all chromosomes have been characterized experimentally using a chromatin immunoprecipitation assay targeting the centromere-specific protein CenH3 (Schotanus *et al.* 2015). An interesting finding is that the chromosomes in *Z. tritici* are either acrocentric or near acrocentric, and every chromosome consequently consists of one long and one short chromosome arm (Schotanus *et al.* 2015). Because a complete chromosome assembly is not available for *Z. ardabiliae*, we mapped the recombination estimates of *Z. ardabiliae* on the genome of *Z. tritici* to assess the impact of the karyotype structure on recombination rate variation. Similar to findings from other species (Jensen-Seaman *et al.* 2004; Munch *et al.* 2014), we observe a negative correlation between recombination rate and the size of the 13 core chromosomes (Kendall's $\tau = -0.59$ with $P = 4.29e^{-3}$ for *Z. tritici* and $\tau = -0.72$ with $P = 2.84e^{-4}$ for *Z. ardabiliae*; Figure 4A). This pattern is generally explained by the necessity of one crossing over to occur per chromosome or chromosome arm per generation, resulting in a higher

recombination rate on smaller chromosomes (*e.g.*, Kong *et al.* 2002; Smeds *et al.* 2016). The significant correlation of the recombination map of *Z. ardabiliae* with the genome structure of *Z. tritici* is an indication of a conserved karyotype of the ancestral species of *Z. tritici* and *Z. ardabiliae*.

Given the acrocentric nature of the *Z. tritici* chromosomes, we considered to what extent recombination also occurs on the short chromosome arms. If meiosis involves one crossover event per chromosome, then the recombination rate should be correlated with the chromosome size and not the chromosome arm length. However, if meiosis involves one crossover event per chromosome arm, then a higher frequency of recombination should occur on shorter chromosome arms. Correlations between recombination rates and chromosome arm lengths also show negative values, yet they are only significant in *Z. ardabiliae* (Kendall's $\tau = -0.14$ with $P = 0.3356$ for *Z. tritici* and $\tau = -0.42$ with $P = 2.16e^{-3}$ for *Z. ardabiliae*; Figure 4B). The negative correlation observed at the chromosome-arm level suggests that meiosis in the *Zymoseptoria* pathogens requires at least one crossing over per chromosome arm and that the small chromosome arms consequently also recombine. The weaker correlations and lack of significance in *Z. tritici* could be due to a fast evolution of centromere positions, erasing the signal of arm-specific recombination rates.

Extremely weak or absent GC-biased gene conversion in *Z. tritici* and *Z. ardabiliae*

In many species, recombination strongly affects evolution of GC content by a mechanism called GC-biased gene conversion

(gBGC) (Duret and Galtier 2009; Mugal *et al.* 2015). The effect of gBGC has been demonstrated in mammals (Piganeau *et al.* 2002; Duret and Galtier 2009), birds (Weber *et al.* 2014), plants (Serres-Giardi *et al.* 2012), and even bacteria (Lassalle *et al.* 2015). However, gBGC has been poorly addressed in fungal species beyond the yeast model, which represents one of the rare organisms for which gBGC was experimentally demonstrated (Mancera *et al.* 2008). To study the possible occurrence and impact of gBGC in the *Z. tritici* and *Z. ardabiliae* genomes, we studied the patterns of GC content along the genomes of the two species. We fitted a nonhomogeneous, nonstationary model of substitution in 10-kb windows in intergenic regions, allowing us to estimate the equilibrium GC content (frequency of GC toward which the sequences evolve) in the extant species. We inferred the dynamics of GC content by comparing the actual GC content of the sequence (observed GC content) with the equilibrium GC content (Duret and Arndt 2008). We find that both the observed and equilibrium GC are highly correlated between *Z. tritici* and *Z. ardabiliae* (Kendall's rank correlation test, $\tau = 0.69$ and 0.45, $P < 2.2e^{-16}$ for the observed and equilibrium GC content, respectively, essential chromosomes only; Figure S3). However, although both species show similar observed GC content (median of 53.3% for *Z. tritici* and 53.6% for *Z. ardabiliae*) they also show contrasting patterns, with the GC content found to be slightly increasing in *Z. ardabiliae* (median equilibrium GC content on autosomes of 53.8, significantly higher that the observed GC content, Wilcoxon paired rank test, $P = 0.04712$) while it is decreasing in *Z. tritici* (median equilibrium GC content of 51.6%, which is significantly lower than the observed GC content, Wilcoxon paired rank test, $P = 2.728e^{-15}$).

To assess the impact of recombination on GC evolution, we correlated the equilibrium GC content in *Z. tritici* and *Z. ardabiliae* to the recombination maps in the two species. We find overall negative, yet weak or nonsignificant, correlations between GC content and recombination rate (Figure S3), both for observed (Kendal's $\tau = -0.047$, $P = 0.04304$ for *Z. tritici* and $\tau = -0.054$, $P = 0.02253$ for *Z. ardabiliae*) and equilibrium GC content (Kendal's $\tau = -0.02$, $P = 0.5082$ for *Z. tritici* and $\tau = 0.01$, $P = 0.7128$ for *Z. ardabiliae*). These results do not support gBGC as a major mechanism shaping GC content in the two fungal pathogen genomes. To test whether this conclusion could be an artifact of recombination rates estimated from population data, we also correlated the equilibrium GC content with the two previously published genetic maps (Croll *et al.* 2015). Consistent with our finding from the LDhat-based recombination map, we confirm an absence of correlation between the equilibrium GC content and the crossover rate and GC content in *Z. tritici*, (Kendall's rank test, $\tau = 0.006$ and $P = 0.7035$ for observed GC; and $\tau = -0.024$, $P = 0.1149$ for equilibrium GC content).

The absence of correlation between GC content and recombination could also be because of a lack of statistical power due to the overall very homogeneous GC content and

Table 3 Segregation patterns at AT/GC biallelic sites

Species	Frequency of GC alleles (%)	Number of alleles with GC >50%	Number of alleles with GC <50%	Ratio GC/ (AT + GC) (%)
Z. tritici	50.41	2,74,517	2,68,589	50.55
Z. ardabiliae	51.86	2,61,777	2,36,232	52.56

large-scale recombination landscapes (recall Figure 3), and the notable absence of isochores that characterize genome composition in other organisms, *e.g.*, in mammals (Galtier *et al.* 2001). As a complementary line of evidence, we investigated the segregation patterns of AT and GC alleles at AT/GC biallelic sites in intergenic regions of both *Z. tritici* and *Z. ardabiliae*, as gBGC is expected to increase the frequency of GC alleles (Escobar *et al.* 2011). We find that the frequency of GC alleles is virtually identical to the frequency of AT alleles in *Z. tritici* and only slightly higher in *Z. ardabiliae* (Table 3), supporting an absence or only weak effect of gBGC in *Z. tritici* and *Z. ardabiliae*, respectively.

No suppression of recombination in centromeres

Recombination is normally found to be absent in centromeric regions where spindles attach during chromosome segregation (see review by Petes 2001). A known exception is *Drosophila mauritiana*, which, in contrast to *D. melanogaster* and *D. simulans*, shows no suppression of recombination in centromeres (True *et al.* 1996). The centromeres of core and accessory chromosomes in *Z. tritici* range from 5.5 to 14 kb in size and do not locate in AT-rich regions (Schotanus *et al.* 2015) as is otherwise observed for centromeres of other species such as *Neurospora crassa* (Smith *et al.* 2011). Correlating the recombination map of *Z. tritici* with centromere positions, we observe (as in *D. mauritiana*) no significant suppression in recombination rate across the centromeric chromosome regions (Wilcoxon signed rank test on 11 chromosomes for which recombination rate in the centromeric region could be inferred, $P = 0.5771$) (Figure 3 and Table 4). The centromeres of *Z. tritici* exhibit several features common to neocentromeres such as a short length ($\sim$10,000 bp in length), lack of enriched repetitive DNA, and weakly transcribed genes (Schotanus *et al.* 2015). We hypothesize that recombination in centromeric sequences has additional implications for evolution of the centromeres in these fungi. A more detailed characterization of chromosome structures and centromere locations in *Z. ardabiliae* is necessary to better understand karyotype evolution in these grass pathogens.

Absence of recombination on accessory chromosomes

The small accessory chromosomes have previously been well characterized in *Z. tritici* (Goodwin *et al.* 2011). They differ considerably from the core chromosomes as they display a higher repeat content, lower gene density, overall lower transcription rate, and are enriched with different chromatin modifications (Stukenbrock *et al.* 2010; Kellner *et al.* 2014;

Table 4 Recombination and repeat content in centromeres of *Z. tritici*

	Chromosome	Start	Stop	Length	Mean ρ	No. of SNPs	Mean ρ for full chromosome	Repeat density (%)	TE density (%)
Essential	1	3,839,299	3,851,749	12,450	0.229	20	0.021	0.94	31.33
	2	512,901	521,916	9,015	0.053	77	0.024	0.00	32.39
	3	3,348,307	3,356,535	8,228	0.097	269	0.025	0.00	0.00
	4	217,113	226,545	9,432	0.033	421	0.028	0.00	9.88
	5	2,604,117	2,614,736	10,619	0.104	47	0.027	0.94	28.19
	6	625,186	637,601	12,415	NA	0	0.026	3.10	37.46
	7	255,824	266,207	10,383	0.006	79	0.044	0.32	0.00
	8	213,892	227,444	13,552	0.059	62	0.029	0.45	39.99
	9	2,067,589	2,076,063	8,474	0.015	106	0.040	0.50	0.00
	10	99,716	109,365	9,649	0.016	77	0.049	0.00	15.32
	11	365,130	373,557	8,427	NA	0	0.049	0.00	46.30
	12	180,233	188,209	7,976	0.001	150	0.052	2.48	7.10
	13	236,993	242,558	5,565	0.015	156	0.037	0.50	0.00
Dispensable	14	59,960	70,870	10,910	0.000	785	0.000	0.00	35.86
	15	382,500	394,754	12,254	0.001	1098	0.001	0.86	20.04
	16	332,004	342,592	10,588	0.099	83	0.023	0.00	35.97
	17	406,958	418,893	11,935	NA	0	0.000	0.24	46.85
	18	159,000	171,999	12,999	NA	0	0.159	0.00	46.62
	19	148,227	159,387	11,160	0.001	4	0.000	0.76	1.38
	20	94,677	105,169	10,492	NA	0	0.008	0.30	11.86
	21	340,264	346,657	6,393	NA	0	NA	0.31	2.33

Grandaubert *et al.* 2015; Schotanus *et al.* 2015). Electrophoretic separation of accessory chromosomes from several isolates of *Z. ardabiliae* has shown that this species also comprises accessory chromosomes (Stukenbrock *et al.* 2011). In this study, we used sequence homology to define the accessory components of the *Z. ardabiliae* genome. We find that the aligned fragments of the accessory chromosomes show very low recombination rates in both species (median ρ = 0.0059 in *Z. tritici* and median ρ = 0.0001 in *Z. ardabiliae* over 13 10-kb windows where both genomes could be aligned, which is 25 and 2% of the autosomal rates, respectively) (Figure 4C). The lower recombination rates reflect the lower effective population size of accessory chromosomes that are present at lower frequencies in populations of *Z. tritici* and *Z. ardabiliae* compared to the core chromosomes. Furthermore, we speculate that frequent structural rearrangements on accessory chromosomes can prevent homologous chromosome pairings and also contribute to the low recombination rates. Our findings add further evidence to support different evolutionary modes of the two sets of chromosomes (core and accessory chromosomes) contained in the same genome. Suppression of recombination is also found on mating-type chromosomes in other fungi including species of *Neurospora* and *Microbotryum* (Whittle and Johannesson 2011; Petit *et al.* 2012; Hood *et al.* 2013). These regions are characterized by an increased accumulation of TEs and structural variants as well as nonadaptive mutations in coding sequences as a consequence of suppressed recombination (Whittle and Johannesson 2011; Whittle *et al.* 2011; Badouin *et al.* 2015).

We also observe a remarkable drop in the recombination rate on the right arm of chromosome 7 (File S1). The right arm of chromosome 7 displays several similarities to the DNA of the accessory chromosomes, including a lower gene density, higher repeat content, and less gene transcription (Grandaubert *et al.* 2015). Furthermore, the entire chromosome arm is enriched with the heterochromatic mark H3K27me3, which is similarly enriched on the accessory chromosomes (Schotanus *et al.* 2015). We previously proposed that this particular chromosome region represents a recent translocation of an accessory chromosome to a core chromosome (Schotanus *et al.* 2015). This hypothesis is consistent with the observation that the recombination rate of the chromosome arm resembles the overall reduced recombination rate of the accessory chromosomes (File S1).

High recombination rates in coding sequences of *Z. tritici*

In primates and birds, recombination increases at CpG islands and around transcription start and end sites (Auton *et al.* 2012; Singhal *et al.* 2015; Smeds *et al.* 2016). In the honeybee, recombination rates in introns and intergenic regions are significantly higher than recombination rates in 3′ and 5′ UTRs and coding sequences (Wallberg *et al.* 2015). It has been proposed that altered chromatin structures, such as destabilized nucleosome occupancy at CpG islands and promoters contribute to this fine-scale variation in recombination rate (Jones 2012). To determine whether specific sequence features in the fungal pathogen genomes similarly affect the overall recombination landscape, we inferred and compared the mean recombination rates in exons, introns, intergenic regions, and 5′ and 3′ flanking regions (500-bp upstream and downstream coding DNA sequence regions, respectively) with a minimum of three filtered SNPs (Figure 5A). Overall, we observe significant differences but with
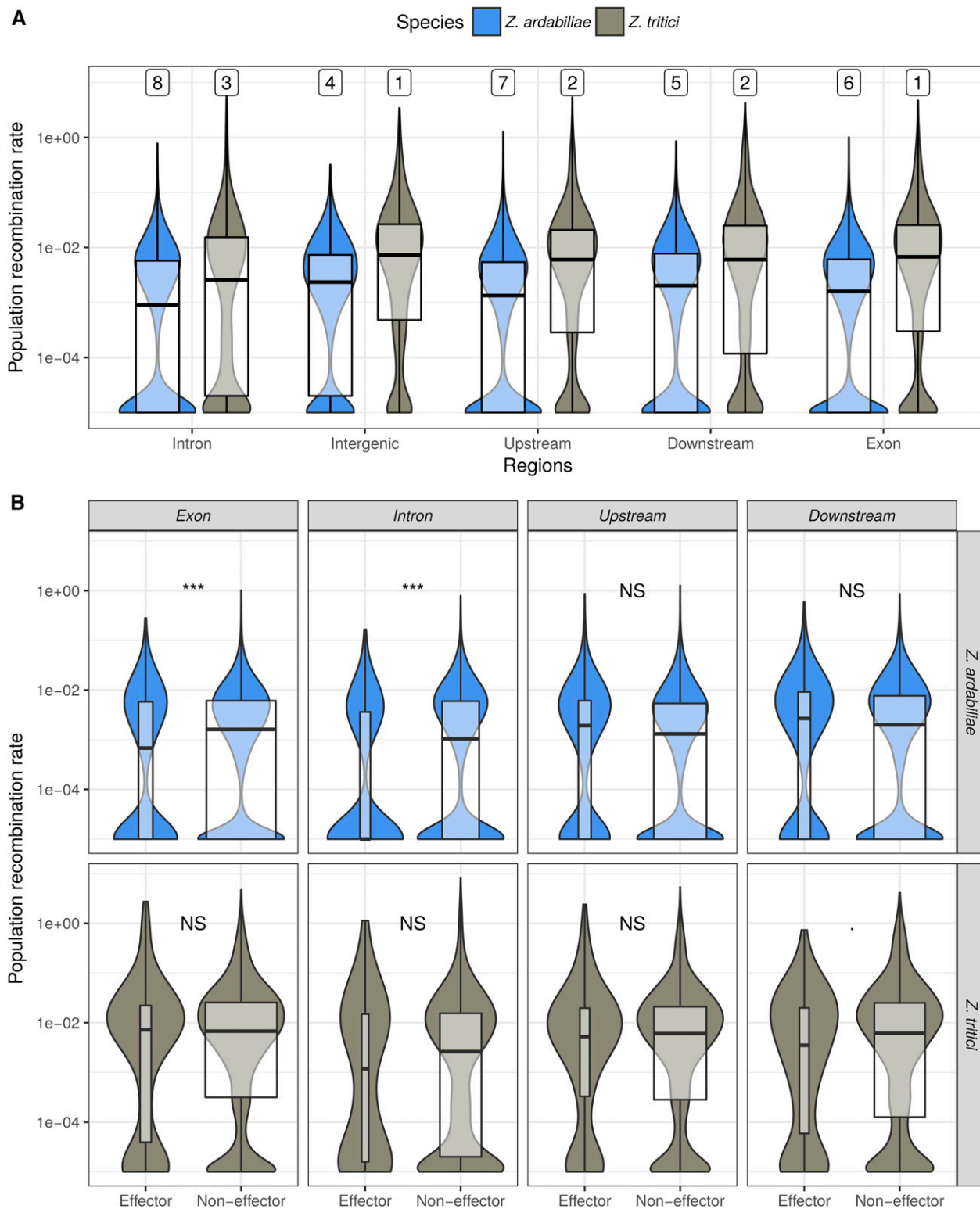
Figure 5 Fine-scale recombination patterns within chromosomes. (A) The distribution of recombination rate estimates in different sequence features in *Z. tritici* and *Z. ardabiliae* reveals small, but significant, differences among the noncoding, coding, and UTR sequences in both species. Top line numbers indicate significance groups by decreasing value of recombination rate. Categories with identical numbers are not significantly different at the 1% level. (B) Distribution of recombination rate estimates in exons, introns, and UTRs of candidate effector and noneffector genes is shown. Bow widths are proportional to the sample sizes. For *Z. ardabiliae*, the recombination rate in exons and introns is significantly lower in candidate effector genes compared to noneffector genes. Wilcoxon rank test corrected for multiple testing, *** $P < 0.1\%$. NS, nonsignificant.

small size effects in fine-scale rates of recombination across different genome regions (Kruskal–Wallis test with *post hoc* comparisons, false discovery rate set to 1%). In both *Z. tritici* and *Z. ardabiliae,* we find the lowest recombination rates in introns and the highest rates in intergenic sequences (Figure 5A). A lower value of $\rho = 2N_e r$ can result from a reduced N_e, a reduced r, or both. N_e in the proximity of genes is expected to be lower due to the presence of background selection (Nordborg *et al.* 1996; Hobolth *et al.* 2011; Scally *et al.* 2012). The highly similar observed recombination rates in coding and noncoding sequences in *Z. tritici* and *Z. ardabiliae* suggests that *r* is not suppressed in these regions in the same way as is observed in other organisms. The pattern indicates that other mechanisms define fine-scale recombination rates in these fungi which lead to high recombination frequencies in protein-coding sequences.

Because of the relatively high rates of recombination in exons of *Z. tritici* and *Z. ardabiliae,* we sought to determine whether recombination could play a particular role in plant–pathogen coevolution. Plant pathogens interfere with host defenses and manipulate the host metabolism by the secretion of so-called effector proteins produced to target molecules from the host (Lo Presti *et al.* 2015). Antagonistic coevolution of these interacting proteins is often reflected in accelerated evolution and signatures of positives selection (Stukenbrock and McDonald 2009). To assess the role of recombination on effector evolution, we first predicted effector proteins computationally in the secretomes of both species using the EffectorP software (Sperschneider *et al.* 2016). This approach identified 868 putative effector proteins in *Z. tritici* and 1122 and *Z. ardabiliae*.

By comparing the recombination rates in different genomic regions encoding effector and noneffector genes, we show a significantly lower recombination rate in exons and introns of effector proteins in *Z. ardabiliae* (Wilcoxon rank test, $P = 1.305e^{-4}$ for exons and $2.534e^{-5}$ for introns, *P*-values corrected for multiple testing) (Figure 5B). The differences are mostly driven by an excess of zero estimates in effector-encoding regions in *Z. ardabiliae,* as visible on the distribution of measures (Figure 5B). Discarding these regions with a mean recombination of zero leads to nonsignificant differences between effector and noneffector genes. A recombination rate estimated to zero can either be due to suppression of recombination in the region or due to an estimation error. Introns and exons with a recombination estimate of zero in *Z. ardabiliae* are found to be shorter and to have a higher SNP density. While these differences are significant, they are of a small size and are unlikely to be a cause of estimation error, and the suppression of recombination in some effector genes of *Z. ardabiliae* therefore appears to be a biological signal whose origin remains to be elucidated by detailed analysis of these regions.

Large-scale but not fine-scale correlation of recombination landscapes in Z. tritici and Z. ardabiliae

Recombination landscapes have been compared in different model species to assess the extent of conservation of recombination rate variation. Broadscale recombination rates in zebra finches and long-tailed finches have similar levels and present correlation factors as high as 0.82 and 0.86 at the 10-kb and 1-Mb scales, respectively (Singhal *et al.* 2015). Similarly, broadscale recombination rates in human and chimpanzee tend to be conserved with few exceptions, such as the human chromosome 2 which originates from a chromosome fusion in the human lineage (Auton *et al.* 2012). However, when comparing the recombination rates of more distantly related mammal species, the correlation of recombination rates decreases even when comparing homologous syntenic blocks (Jensen-Seaman *et al.* 2004). In studies of mammals and fruit flies, it is considered that the recombination landscape evolves as a result of evolution of other sequence variables (Jensen-Seaman *et al.* 2004) and the dynamics of fine-scale recombination rates, including the positions of hotspots (Winckler *et al.* 2005; Chan *et al.* 2012).

To address the evolution of recombination landscapes in *Z. tritici* and *Z. ardabiliae,* we compared the genome-wide recombination maps of the two species. We previously reported that the genomes of the two species show a high extent of colinearity and we found a mean sequence divergence of $d_{xy} = 0.13$ substitutions per site (Stukenbrock *et al.* 2011). Here, we first aligned the two reference genomes of *Z. tritici* and *Z. ardabiliae* to compare recombination rates in homologous genome regions (Figure 6; see *Materials and Methods*). Next, we calculated the average recombination rate in nonoverlapping windows with at least 100 SNPs in each species, which resulted in 3851 windows for which recombination in both species could be averaged. The two maps show a moderate yet highly significant correlation (Kendall's rank correlation test, $\tau = 0.2327$, $P < 2.2e^{-16}$; Figure 7A), which suggests certain similarities in the recombination landscape of the two fungi. To determine the scale at which the maps are most correlated (broad- or fine-scale recombination rates), we further investigated how the correlations vary when various window sizes are used. We find that the correlations, consistently inferred with different correlation measures, peak at the 0.5–1 Mb scale (Figure 7B), suggesting that the recombination landscape is conserved at large scales but shows rapid evolution at smaller scales. These results mirror findings from other eukaryotic species (*e.g.,* Winckler *et al.* 2005; Singhal *et al.* 2015) and suggest that distinct mechanisms determine the recombination landscape at fine and broad scales in these two species.

Frequency and intensity of recombination hotspots is higher in Z. tritici

The fine-scale LDhat recombination maps clearly reveal the presence of distinct peaks of recombination in both *Z. tritici* and *Z. ardabiliae* (Figure 3). We used the program LDhot to call positions of statistically significant recombination hotspots (Auton *et al.* 2014) and applied highly stringent selection criteria (see *Materials and Methods*) to obtain positions
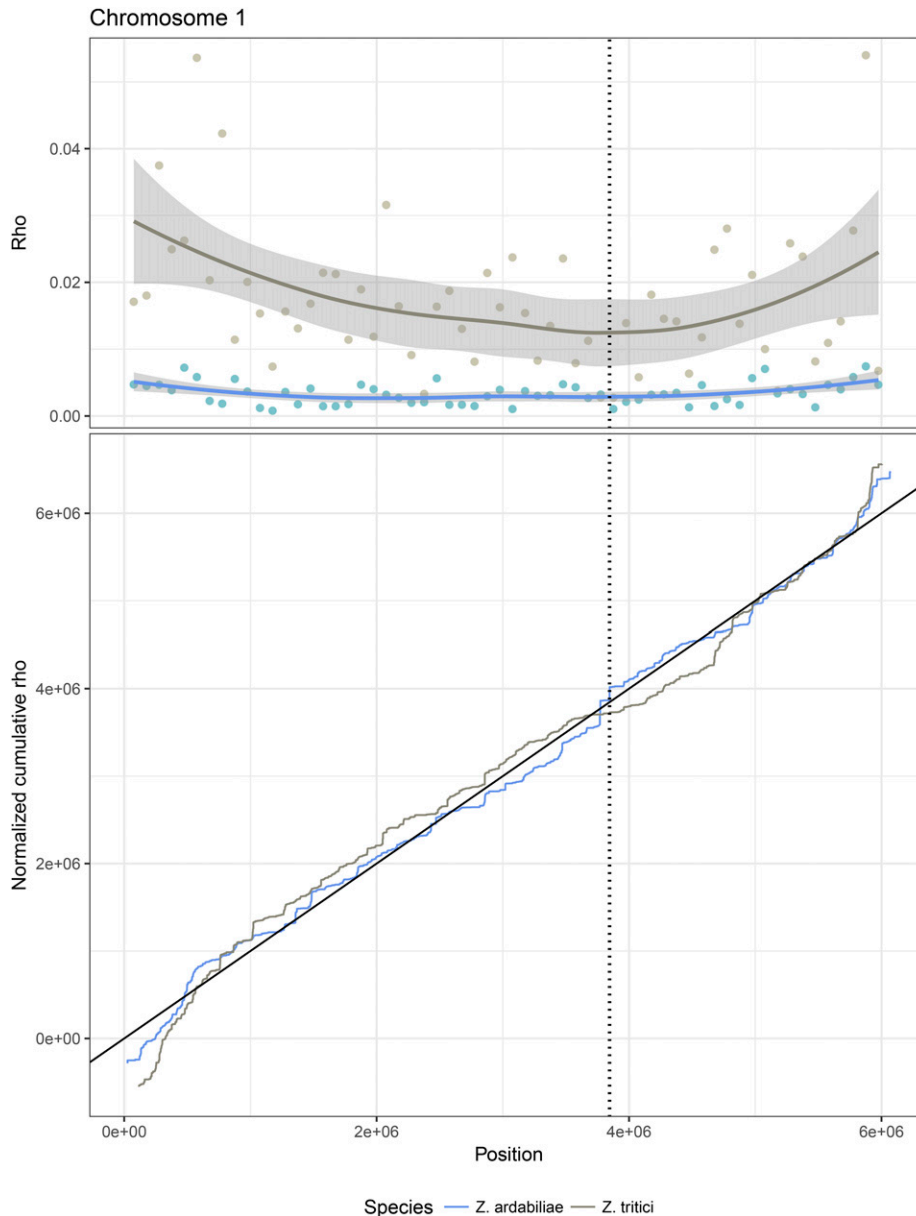
Figure 6 Recombination maps of *Z. tritici* and *Z. ardabiliae* plotted along the chromosome 1 of *Z. tritici*. (Upper panel) Recombination map in 100-kb windows plotted together with smoothing curves. (Lower panel) Cumulative curves of the recombination maps, scaled to be comparable. The position of the centromere is marked over the chromosome plots as a vertical stippled line. Figures for other chromosomes are available in File S2.

of the most significant hotspots for which the within-hotspot rate was at least 10 times higher than the flanking regions (Figure 8A). Interestingly, our approach revealed a considerably greater number of recombination hotspots in *Z. tritici* (2578 hotspots) than in *Z. ardabiliae* (862 hotspots). Furthermore, we find a significant difference in the size of the hotspot regions between the two species. In general, the recombination hotspots span significantly shorter regions in *Z. tritici* (median 39 bp) than in *Z. ardabiliae* (66 bp, Wilcoxon ranked test $P < 2.2e^{-16}$). We also compared the intensity of the recombination hotspots, as estimated by LDhot (ρ across hotspot) and also find the median value of ρ in hotspots to be significantly higher in *Z. tritici* (median of 16.44 compared with 8.42 for *Z. ardabiliae*, Wilcoxon rank test $P < 2.2e^{-6}$). The higher frequency of more intense hotspots in *Z. tritici* not only reveals a different hotspot

landscape in the wheat pathogen, it also suggests that the overall higher recombination rate we observe in *Z. tritici* partly is explained by the different recombination hotspot architecture. These differences to some extent mirror the larger density of SNPs in *Z. tritici* that enables a finer resolution of the hotspot distribution and structure, and could potentially be affected by a different demography and population structure in the two species. We also speculate that recombination hotspots in these fungi have evolved since the divergence of *Z. tritici* and *Z. ardabiliae*. To address the extent of conservation in hotspot positions, we correlated the hotspot maps of the two species.

The position of recombination hotspots is defined by different mechanisms in different taxa, *e.g.*, PRDM9 in primates and transcription start and end sites in other species such as birds (Myers *et al.* 2005; Singhal *et al.* 2015). Consequently,
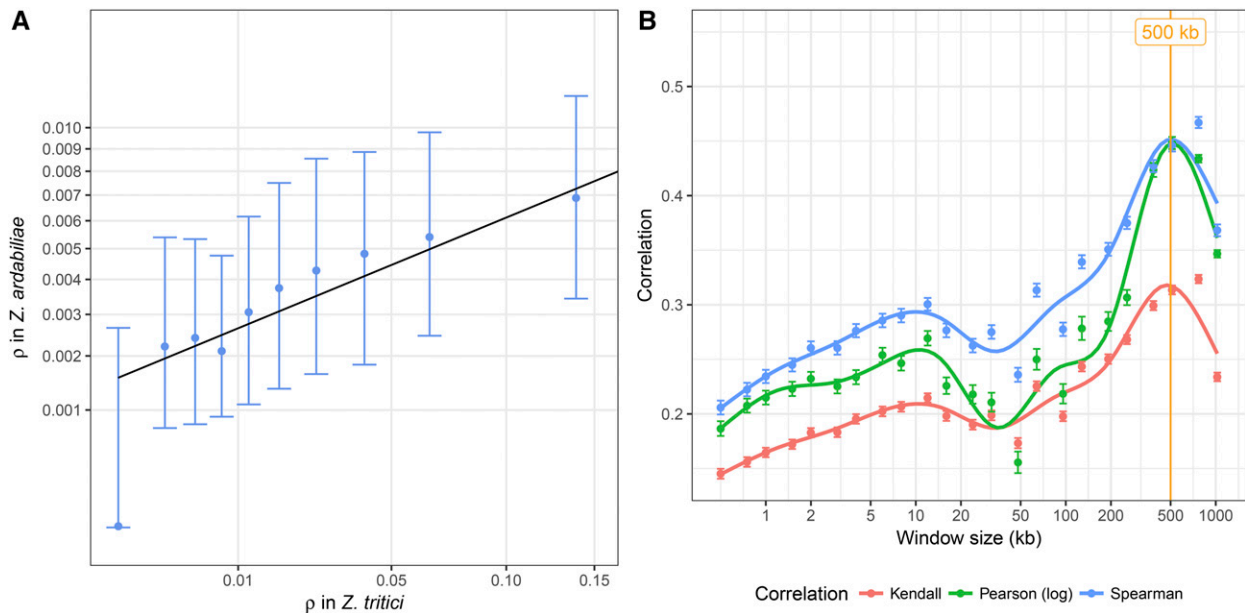
Figure 7 Correlation of recombination maps of *Z. tritici* and *Z. ardabiliae*. (A) Comparison of the two recombination maps based on average recombination rates in windows of at least 100 SNPs in each species. Points represent averages in 10 classes with equal numbers of windows; points and error bars represent the median and first and third quartile of the distribution for each category. (B) Correlation of recombination maps in sliding windows of different sizes. Three distinct correlation coefficients are plotted against recombination rates averaged in different window sizes (see *Materials and Methods*). Points indicate the averages of 1000 samples and bars shows the SEM. Lines correspond to local regression smoothing.

hotspot positions are highly conserved in some species (Tsai *et al.* 2010; Singhal *et al.* 2015), while highly variable in others (Myers *et al.* 2010). We mapped *Z. ardabiliae* hotspots on the *Z. tritici* genomes and counted the number of colocalizing hotspots in the two species. We considered a hotspot in *Z. tritici* as colocalizing with a hotspot in *Z. ardabiliae* if the distance between the two hotspots is <1 kb and if no other hotspot is present in between. We report that only 149 hotspots are colocalizing (6% of hotspots in *Z. tritici* and 20% of hotspots in *Z. ardabiliae*). This number is however significantly more than expected by chance ($P < 9.99e^{-5}$, permutation test; Figure 8B). These results are consistent with the previously reported genetic maps of *Z. tritici*, which also show little overlap of hotspot positions between two Swiss crosses (Croll *et al.* 2015). Conversely, the patterns are highly different from *Saccharomyces* species in which hotspot positions are highly conserved and associated with functional elements across the yeast genomes (Tsai *et al.* 2010).

Given the dense genomes of *Z. tritici* and *Z. ardabiliae*, we assessed the number of hotspots mapped to coding sequences. Of the 2578 *Z. tritici* hotspots, 132 are located in introns and 1435 are located in exons. Interestingly, in *Z. ardabiliae*, we find 44 hotspots in introns and only 396 in exons. We plotted the number of hotspots as a function of the number of called sites in each region (Figure 8C). We observe a general trend in which the number of detected hotspots increases with the number of called sites as a power law (linear relationship in log space), and with

more hotspots detected in *Z. tritici*. In contrast to patterns of previously studied species, this reveals the presence of hotspots in all parts of the genome, including coding regions. We do not observe a significant enrichment close to transcription start sites (upstream regions) like in yeast (Lam and Keeney 2015). We further note that comparatively fewer hotspots locate in intergenic regions of *Z. tritici*, these regions displaying a density of hotspots similar to what is expected in *Z. ardabiliae* for the observed number of callable sites. We hypothesize two nonexclusive possible origins for this result: (1) the number of callable sites is higher in *Z. tritici* intergenic regions than in *Z. ardabiliae*, due to the lack of telomere-to-telomere assembly of a reference genome for this species. The missing regions could potentially bias our estimate of hotspot densities in intergenic regions. (2) Another possible explanation is that the comparatively larger number of hotspots in *Z. tritici* is due to an increased hotspot density in protein-coding genes in this species, which raises the question whether intragenic recombination hotspots represent a selected feature during evolution of the wheat-infecting lineage.

Conclusions

Pathogens need to adapt rapidly to overcome immune responses in their host (Jones and Dangl 2006). Several examples from animal and plant pathogens document exceptionally high rates of genome rearrangements, including changes in ploidy and full chromosome gains or losses (*e.g.*, Ma *et al.* 2010; Croll *et al.* 2013; Hickman *et al.* 2013, 2015).
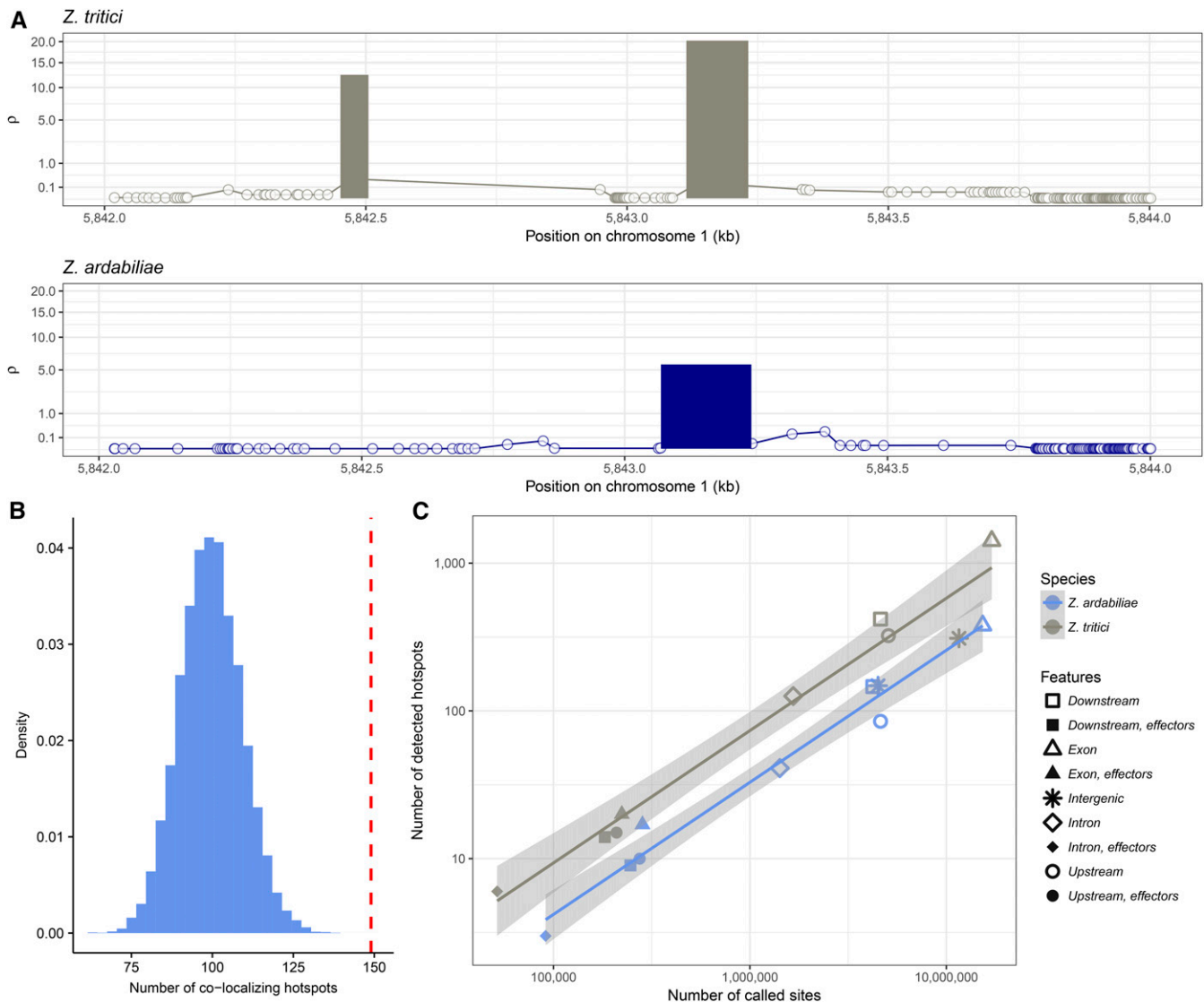
Figure 8 Distribution of hotspots in the genomes of *Z. tritici* and *Z. ardabiliae*. (A) Example of mapped hotspot in a homologous region in *Z. tritici* and *Z. ardabiliae*. Lines indicate the background recombination rate as estimated by LDhat. Bars indicate the positions, widths, and strengths of hotspots detected by LDhot in the region, after filtering (see *Materials and Methods*). (B) Number of hotspots in *Z. tritici* in the direct 1-kb range of a hotspot in *Z. ardabiliae* (vertical line) and the corresponding distribution under the null hypothesis of a random distribution of hotspots. (C) Frequencies of hotspots in distinct regions of the genome. Number of detected hotspots in each region as a function of the number of called sites. Lines correspond to ordinary least-square regressions.

So far, the importance of meiotic recombination in rapid evolution of pathogens has been poorly addressed. Our analyses demonstrate extraordinarily high recombination rates in two fungal plant pathogens and thereby suggest that sexual recombination can also be a major driver of rapid pathogen evolution.

The overall higher recombination rate and the increased density of recombination hotspots in the crop pathogen *Z. tritici* are remarkable. *Z. tritici* and *Z. ardabiliae* share a recent common ancestor, but exist and evolve in highly different environments. While *Z. ardabiliae* infects wild grasses in a natural ecosystem, *Z. tritici* infects a crop host and propagates only in managed ecosystems. Agricultural management strategies, dense host populations, and increased gene flow between geographically distant populations are

factors that contribute to the different population structure of *Z. tritici*. We hypothesize that an increased rate of recombination in coding sequences of *Z. tritici* was selected as it favored the rapid generation of new alleles and allele combinations (Brunner *et al.* 2008). The exceptionally high recombination rate in *Z. tritici* allows the pathogen to rapidly overcome new host resistances and explains the current difficulties of controlling this important wheat pathogen.

Acknowledgments

Literature Cited

Auton, A., and G. McVean, 2007   Recombination rate estimation in the presence of hotspots. Genome Res. 17: 1219–1227.

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Ségurel *et al.*, 2012   A fine-scale chimpanzee genetic map from population sequencing. Science 336: 193–198.

Auton, A., S. Myers, and G. McVean, 2014   Identifying recombination hotspots using population genetic data. arXiv: 1403.4264.

Awadalla, P., 2003   The evolutionary genomics of pathogen recombination. Nat. Rev. Genet. 4: 50–60.

Badouin, H., M. E. Hood, J. Gouzy, G. Aguileta, S. Siguenza *et al.*, 2015   Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus Microbotryum lychnidis-dioicae. Genetics 200: 1275–1284.

Begun, D. J., and C. F. Aquadro, 1992   Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature 356: 519–520.

Betancourt, A. J., J. J. Welch, and B. Charlesworth, 2009   Reduced effectiveness of selection caused by a lack of recombination. Curr. Biol. 19: 655–660.

Blanchette, M., W. J. Kent, C. Riemer, L. Elnitski, A. F. A. Smit *et al.*, 2004   Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14: 708–715.

Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White, and J. L. Weber, 1998   Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. 63: 861–869.

Brunner, P. C., F. L. Stefanato, and B. A. McDonald, 2008   Evolution of the CYP51 gene in Mycosphaerella graminicola: evidence for intragenic recombination and selective replacement. Mol. Plant Pathol. 9: 305–316.

Chan, A. H., P. A. Jenkins, and Y. S. Song, 2012   Genome-wide fine-scale recombination rate variation in Drosophila melanogaster. PLoS Genet. 8: e1003090.

Choi, K., X. Zhao, K. A. Kelly, O. Venn, J. D. Higgins *et al.*, 2013   Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. 45: 1327–1336.

Croll, D., M. Zala, and B. A. McDonald, 2013   Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLoS Genet. 9: e1003567.

Croll, D., M. H. Lendenmann, E. Stewart, and B. A. McDonald, 2015   The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201: 1213–1228.

Daverdin, G., T. Rouxel, L. Gout, J.-N. Aubertot, I. Fudal *et al.*, 2012   Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. PLoS Pathog. 8: e1003020.

de Castro, E., I. Soriano, L. Marín, R. Serrano, L. Quintales *et al.*, 2012   Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. EMBO J. 31: 124–137.

de Jonge, R., M. D. Bolton, A. Kombrink, G. C. M. van den Berg, K. A. Yadeta *et al.*, 2013   Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23: 1271–1282.

Dray, S., and A.-B. Dufour, 2007   The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22: 1–20.

Duret, L., and P. F. Arndt, 2008   The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet. 4: e1000071.

Duret, L., and N. Galtier, 2009   Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet. 10: 285–311.

Dutheil, J., and B. Boussau, 2008   Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. BMC Evol. Biol. 8: 255.

Dutheil, J. Y., S. Gaillard, and E. H. Stukenbrock, 2014   MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics 15: 53.

Dutheil, J. Y., G. Mannhaupt, G. Schweizer, C. M. K. Sieber, M. Münsterkötter *et al.*, 2016   A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol. Evol. 8: 681–704.

Escobar, J. S., S. Glémin, and N. Galtier, 2011   GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. Mol. Biol. Evol. 28: 2561–2575.

Galtier, N., G. Piganeau, D. Mouchiroud, and L. Duret, 2001   GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159: 907–911.

Goodwin, S. B., S. Ben M'barek, B. Dhillon, A. H. J. Wittenberg, C. F. Crane *et al.*, 2011   Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7: e1002070.

Grandaubert, J., A. Bhattacharyya, and E. H. Stukenbrock, 2015   RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus Zymoseptoria identify novel orphan genes and species-specific invasions of transposable elements. G3 (Bethesda) 5: 1323–1333

Grandaubert, J., J. Y. Dutheil, and E. H. Stukenbrock, 2017   The genomic rate of adaptation in the fungal wheat pathogen Zymoseptoria tritici. bioRxiv. DOI: https://doi.org/10.1101/176727.

Gremme, G., S. Steinbiss, and S. Kurtz, 2013   GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans. Comput. Biol. Bioinform. 10: 645–656.

Hasegawa, M., H. Kishino, and T. Yano, 1985   Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22: 160–174.

Hickman, M. A., G. Zeng, A. Forche, M. P. Hirakawa, D. Abbey *et al.*, 2013   The 'obligate diploid' Candida albicans forms mating-competent haploids. Nature 494: 55–59 [corrigenda: Nature 530: 242 (2016)].

Hickman, M. A., C. Paulson, A. M. Dudley, and J. Berman, 2015   Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in Candida albicans. Genetics 200: 781–794.

Hobolth, A., J. Y. Dutheil, J. Hawks, M. H. Schierup, and T. Mailund, 2011   Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21: 349–356.

Hood, M. E., E. Petit, and T. Giraud, 2013   Extensive divergence between mating-type chromosomes of the anther-smut fungus. Genetics 193: 309–315.

Horton, M. W., A. M. Hancock, Y. S. Huang, C. Toomajian, S. Atwell *et al.*, 2012   Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat. Genet. 44: 212–216.

Hunter, C. M., W. Huang, T. F. C. Mackay, and N. D. Singh, 2016 The genetic architecture of natural variation in recombination rate in Drosophila melanogaster. PLoS Genet. 12: e1005951.

Jeffreys, A. J., and R. Neumann, 2009 The rise and fall of a human recombination hot spot. Nat. Genet. 41: 625–629.

Jeffreys, A. J., J. Murray, and R. Neumann, 1998 High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol. Cell 2: 267–273.

Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin et al., 2004 Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14: 528–538.

Jones, J. D., and J. L. Dangl, 2006 The plant immune system. Nature 444: 323–329.

Jones, P. A., 2012 Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13: 484–492.

Kaplan, N., I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo et al., 2009 The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458: 362–366.

Katoh, K., G. Asimenos, and H. Toh, 2009 Multiple alignment of DNA sequences with MAFFT. Methods Mol. Biol. 537: 39–64.

Kellner, R., A. Bhattacharyya, S. Poppe, T. Y. Hsu, R. B. Brem et al., 2014 Expression profiling of the wheat pathogen Zymoseptoria tritici reveals genomic patterns of transcription and host-specific regulatory programs. Genome Biol. Evol. 6: 1353–1365.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson et al., 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31: 241–247.

Lam, I., and S. Keeney, 2015 Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science 350: 932–937.

Lassalle, F., S. Périan, T. Bataillon, X. Nesme, L. Duret et al., 2015 GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet. 11: e1004941.

Linde, C. C., J. Zhan, and B. A. McDonald, 2002 Population structure of Mycosphaerella graminicola: from lesions to continents. Phytopathology 92: 946–955.

Lo Presti, L., D. Lanver, G. Schweizer, S. Tanaka, L. Liang et al., 2015 Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66: 513–545.

Ma, L.-J., H. C. van der Does, K. A. Borkovich, J. J. Coleman, M.-J. M.-J. Daboussi et al., 2010 Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464: 367–373.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454: 479–485.

Marais, G., D. Mouchiroud, and L. Duret, 2003 Neutral effect of recombination on base composition in Drosophila. Genet. Res. 81: 79–87.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li et al., 2009 Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Meunier, J., and L. Duret, 2004 Recombination drives the evolution of GC-content in the human genome. Mol. Biol. Evol. 21: 984–990.

Möller, M., and E. H. Stukenbrock, 2017 Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15: 756–771.

Mugal, C. F., C. C. Weber, and H. Ellegren, 2015 GC-biased gene conversion links the recombination landscape and demography to genomic base composition. BioEssays 37: 1317–1326.

Munch, K., T. Mailund, J. Y. Dutheil, and M. H. Schierup, 2014 A fine-scale recombination map of the human–chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. Genome Res. 24: 467–474.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321–324.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman et al., 2010 Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327: 876–879.

Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. 67: 159–174.

Petersen, T. N., S. Brunak, G. Heijne, and H. Nielsen, 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8: 785–786.

Petes, T. D., 2001 Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. 2: 360–369.

Petit, E., T. Giraud, D. M. Vienne, M. A. Coelho, G. Aguileta et al., 2012 Linkage to the mating-type locus across the genus Microbotryum: insights into nonrecombining chromosomes. Evolution 66: 3519–3533.

Piganeau, G., D. Mouchiroud, L. Duret, and C. Gautier, 2002 Expected relationship between the silent substitution rate and the GC content: implications for the evolution of isochores. J. Mol. Evol. 54: 129–133.

Raffaele, S., and S. Kamoun, 2012 Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10: 417–430.

R Core Team, 2013 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rizzon, C., G. Marais, M. Gouy, and C. Biémont, 2002 Recombination rate and the distribution of transposable elements in the Drosophila melanogaster genome. Genome Res. 12: 400–407.

Scally, A., J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead et al., 2012 Insights into hominid evolution from the gorilla genome sequence. Nature 483: 169–175.

Schotanus, K., J. L. Soyer, L. R. Connolly, J. Grandaubert, P. Happel et al., 2015 Histone modifications rather than the novel regional centromeres of Zymoseptoria tritici distinguish core and accessory chromosomes. Epigenetics Chromatin 8: 41.

Serres-Giardi, L., K. Belkhir, J. David, and S. Glémin, 2012 Patterns and evolution of nucleotide landscapes in seed plants. Plant Cell 24: 1379–1397.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn et al., 2015 Stable recombination hotspots in birds. Science 350: 928–932.

Smeds, L., C. F. Mugal, A. Qvarnström, and H. Ellegren, 2016 High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. PLoS Genet. 12: e1006044.

Smith, K. M., P. A. Phatale, C. M. Sullivan, K. R. Pomraning, and M. Freitag, 2011 Heterochromatin is required for normal distribution of Neurospora crassa CenH3. Mol. Cell. Biol. 31: 2528–2542.

Spencer, C. C. A., P. Deloukas, S. Hunt, J. Mullikin, S. Myers et al., 2006 The influence of recombination on human genetic diversity. PLoS Genet. 2: e148.

Sperschneider, J., D. M. Gardiner, P. N. Dodds, F. Tini, L. Covarelli et al., 2016 EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210: 743–761.10.1111/nph.13794

Staab, P. R., S. Zhu, D. Metzler, and G. Lunter, 2015 Scrm: efficiently simulating long sequences using the approximated coalescent with recombination. Bioinformatics 31: 1680–1682.

Stukenbrock, E. H., and B. A. McDonald, 2009 Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. Mol. Plant Microbe Interact. 22: 371–380.

Stukenbrock, E. H., S. Banke, M. Javan-Nikkhah, and B. A. McDonald, 2007 Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. Mol. Biol. Evol. 24: 398–411.

Stukenbrock, E. H., F. G. Jørgensen, M. Zala, T. T. Hansen, B. A. McDonald *et al.*, 2010   Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen mycosphaerella graminicola. PLoS Genet. 6: e1001189.

Stukenbrock, E. H., T. Bataillon, J. Y. Dutheil, T. T. Hansen, R. Li *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.

Stukenbrock, E. H., W. Quaedvlieg, M. Javan-Nikhah, M. Zala, P. W. Crous *et al.*, 2012   Zymoseptoria ardabiliae and Z. pseudotritici, two progenitor species of the septoria tritici leaf blotch fungus Z. tritici (synonym: Mycosphaerella graminicola). Mycologia 104: 1397–1407.

Stumpf, M. P. H., and G. A. T. McVean, 2003   Estimating recombination rates from population-genetic data. Nat. Rev. Genet. 4: 959–968.

Taylor, J. W., C. Hann-Soden, S. Branco, I. Sylvain, and C. E. Ellison, 2015   Clonal reproduction in fungi. Proc. Natl. Acad. Sci. USA 112: 8901–8908.

True, J. R., J. M. Mercer, and C. C. Laurie, 1996   Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics 142: 507–523.

Tsai, I. J., A. Burt, and V. Koufopanou, 2010   Conservation of recombination hotspots in yeast. Proc. Natl. Acad. Sci. USA 107: 7847–7852.

Wallberg, A., S. Glémin, and M. T. Webster, 2015   Extreme recombination frequencies shape genome variation and evolution in the honeybee, Apis mellifera. PLoS Genet. 11: e1005189.

Wang, Y., and B. Rannala, 2014   Bayesian inference of shared recombination hotspots between humans and chimpanzees. Genetics 198: 1621–1628.

Weber, C. C., B. Boussau, J. Romiguier, E. D. Jarvis, and H. Ellegren, 2014   Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. Genome Biol. 15: 549.

Whittle, C. A., and H. Johannesson, 2011   Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of Neurospora tetrasperma. Heredity (Edinb) 107: 305–314.

Whittle, C. A., Y. Sun, and H. Johannesson, 2011   Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of Neurospora tetrasperma. Eukaryot. Cell 10: 594–603.

Wickham, H., 2016   *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

Wijnker, E., G. Velikkakam James, J. Ding, F. Becker, J. R. Klasen *et al.*, 2013   The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana (G McVean, Ed.). *Elife* 2: e01426.

Winckler, W., S. R. Myers, D. J. Richter, R. C. Onofrio, G. J. McDonald *et al.*, 2005   Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308: 107–111.

Zhan, J., R. E. Pettway, and B. A. McDonald, 2003   The global genetic structure of the wheat pathogen Mycosphaerella graminicola is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. Fungal Genet. Biol. 38: 286–297.

Communicating editor: M. Hahn