Comparison of fine-scale recombination maps in fungal plant pathogens reveals dynamic recombination landscapes and intragenic hotspots

Authors:

Eva H. Stukenbrock[1+2)] and Julien Y. Dutheil[3+4)]

[1)] Environmental Genomics, Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Plön, Germany, [2)] Christian-Albrechts University of Kiel, Environmental Genomics, Am Botanischen Garten 1-11, 24118 Kiel, Germany, [3)] Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Plön, Germany and [4)] Institut des Sciences de L'Évolution – Montpellier, CNRS – Université Montpellier 2, 34095 Montpellier, France

Corresponding authors:

Eva H. Stukenbrock and Julien Y. Dutheil

Max Planck Institute for Evolutionary Biology

August-Thienemann-Str. 2, 24306 Plön, Germany

stukenbrock@evolbio.mpg.de

dutheil@evolbio.mpg.de

Running title: Recombination rate evolution in fungal plant pathogens

Keywords: genome evolution, recombination analyses, recombination hotspots, fungal plant pathogens, effectors, *Zymoseptoria*

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 recombination hotspot positions can be studied using population genomics data and statistical methods. In this study, we applied 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 *Zymoseptoria 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 co-localize between the two species, suggesting that hotspots dynamics contribute to the overall pattern of fast evolving recombination in these species.

Introduction

Meiotic recombination is a fundamental process, which in many eukaryotes shapes genetic variation in populations and drives evolutionary changes. 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. Studies based on experimental and empirical data have demonstrated that recombination in sexual organisms plays a crucial role in defining genome-wide neutral and non-neutral nucleotide variation patterns (Begun and Aquadro 1992; Spencer et al. 2006), rates of protein evolution (Betancourt et al. 2009), transposable elements distribution (Rizzon et al. 2002), GC content (Meunier and Duret 2004), and codon bias (Marais et al. 2003).

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 (Jeffreys et al. 1998; Broman et al. 1998; McMullen et al. 2009). However, pedigree studies rely on a 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 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; Singhal et al. 2015; Hunter et al. 2016; Horton et al. 2012). A general finding from these studies is that recombination events are non-uniformly distributed across chromosomes. Furthermore, 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, more than 25.000 recombination hotspots have been identified, with a number of these hotspots showing a more than hundred-fold increase in recombination rates and exhibiting a strong impact on the overall recombination landscape and genome evolution in general (Myers et al. 2005; Jeffreys and Neumann 2009; Winckler et al. 2005).

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 in chimpanzee and human 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 non-overlapping 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 (Myers et al. 2010). In other taxa recombination hotspots are formed by other mechanisms. In some species, including species without PRDM9, including 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 (Lam and Keeney 2015; Singhal et al. 2015; Smeds et al. 2016; Horton et al. 2012; Choi et al. 2013). 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. 2011). Indeed, in the fission yeast *Schizosaccharomyces pombe* and the Brassicaceae plant *Arabidopsis thaliana* meiotic recombination hotspots were shown to co-localize with nucleosome–depleted regions supporting a link between chromatin structure and recombination in these species (de Castro et al. 2011; Wijnker et al. 2013).

Although many pathogens and parasites are sexual, the impact of recombination on the evolution of their genome has been rarely addressed (Awadalla 2003). As recombination can be an important driver of overall genome evolution in pathogen species, we here set out to investigate patterns of recombination in plant pathogenic fungi. We focused on the 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 co-evolved and dispersed with its host since early wheat domestication (Stukenbrock et al. 2007). A close relative of *Z. tritici*, *Zymoseptoria ardabiliae*, has been isolated from wild grass species in the Middle East (Stukenbrock et al. 2012). The two pathogen species diverged recently but have non-overlapping host ranges and show some differences in morphology and host infection patterns (Stukenbrock et al. 2011, 2012). Both species undergo frequent sexual recombination, which result in the formation of ascospores that serve as a mean of long distance wind dispersal and primary infection of new hosts (Stukenbrock et al. 2011). The co-linear genomes of *Z. tritici* and *Z. ardabiliae* share 90% nucleotide similarity on average, thus

131 providing an excellent resource for comparative analyses of genome evolution (Stukenbrock
132 et al. 2011). The 40Mb haploid genome of the reference *Z. tritici* isolate comprises 21
133 chromosomes of which eight are accessory chromosomes (Goodwin et al. 2011b). The
134 accessory chromosomes represent a highly variable genome compartment characterized by
135 presence/absence variation of entire chromosomes, high repeat content and low gene
136 densities (Goodwin et al. 2011a; Grandaubert et al. 2015). The accessory chromosomes are
137 partly conserved among several species in the genus *Zymoseptoria*, suggesting that these
138 small chromosomes have been maintained over long evolutionary times predating the
139 divergence of species (Stukenbrock et al. 2011). In a previous study, we applied a whole-
140 genome coalescence approach to generate a genetic map of the ancestral species of *Z. tritici*
141 and another closely related species, *Z. pseudotritici* (Stukenbrock et al. 2011). We found
142 evidence of a high recombination rate in the ancestral species (genome average 46cM/Mb)
143 and showed a significantly higher proportion of sites showing incomplete lineage sorting in
144 regions with high recombination rate. The existence of high recombination rates in the genus
145 *Zymoseptoria* was recently supported by experimental data. Croll and colleagues generated a
146 linkage map of *Z. tritici* from two independent crosses of Swiss field isolates (Croll et al.
147 2015). This map based on actual crossing-over events along the 40Mb genome, confirms the
148 high recombination rates (genome average 66 cM/Mb, measured in windows of 20 kb) in the
149 present-day pathogen species. Interestingly, the study also reported large variation between
150 the two independent crosses of *Z. tritici*, suggesting that recombination is highly dynamic in
151 this pathogen (Croll et al. 2015).

152 In this study we addressed the evolution of recombination rate in fungal pathogens. We
153 applied a population genomics approach to generate a fine-scale recombination map of the
154 two recently diverged species *Z. tritici* and *Z. ardabiliae*. This allowed us to infer and compare
155 fine-scale genome-wide patterns of recombination rates in the two species and investigate
156 the dynamics of recombination landscapes. We confirm the exceptionally high
157 recombination rates as also observed in a previous coalescence-based genome analysis and
158 shown by experimental crosses (Stukenbrock et al. 2011; Croll et al. 2015). Furthermore, we
159 identify 2,578 and 862 recombination hotspots in *Z. tritici* and *Z. ardabiliae* respectively.
160 Intriguingly, detailed analyses of the recombination hotspots show not only a comparatively
161 higher hotspot frequency in the wheat pathogen but also the occurrence of stronger
162 hotspots in *Z. tritici*. Our findings confirm that recombination rate landscapes are highly
163 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.

Results and Discussion

Genome alignments and SNP calling

A total of 30 whole genome sequences were used to infer the recombination landscapes of the two haploid 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 (Supplemental Table S1). The haploid genomes, including 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 40.8Mb for *Z. tritici* and 32.4Mb for *Z. ardabiliae*.

Recombination analyses rely on single nucleotide polymorphism (SNP) data. However, erroneously called SNPs or alignment errors can greatly bias linkage disequilibrium (LD) inference in genomes. To generate high-quality SNP datasets we therefore extensively 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 filter 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 repeat elements, including tranposable elements. 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, repeats were 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 correspond to 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 program packages Ldhat (Auton and McVean 2007) and Ldhelmet (Chan et al. 2012). Both methods estimate the local population recombination

rates based on the LD between SNPs in a given genome dataset 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 $\rho = 2N_e * r$ for haploid individuals, where N_e is the effective population size and *r* is the per site rate of recombination across the region. As θ substantially varies along genomes, we generated recombination maps using three scaled effective population size values as inputs (θ = 0.05, 0.005 and 0.0005). For both methods, we find that the three different input θ values only have a marginal influence on the recombination rate estimates obtained from Ldhat and Ldhelmet (Fig. 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* (θ = 0.0139) and in *Z. ardabiliae* (θ = 0.0087) (Table 1).

To assess the performance of the two methods and the input parameters for the fungal dataset, 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, 3D7x3D1 and SW5xSW39 (Croll et al. 2015). The two 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 (Fig. 1B). We find a significant correlation between the Ldhat map and the two genetic maps (3D7x3D1, Kendall's rank correlation test, $\tau = 0.27$, p-value < 2.2e-16 and SW5xSW39, Kendall's rank correlation test, $\tau = 0.23$, p-value < 2.2.e-16). Using an average recombination rate of the 3D7x3D1 and SW5xSW39 crosses the correlation coefficient further increases (Kendall's rank correlation test, $\tau = 0.29$, p-value < 2.2.e-16) (Fig. 1B). While correlated, the new recombination maps of *Z. tritici* encompasses more than 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 ca 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 3D7x3D1, and 0.20 for the cross SW5xSW39 and 0.25 using the average of the two crosses; all p-values < 2.2e-16). These correlations, although highly significant, have relatively small size effects. However, it is noteworthy that also the correlation between the two Swiss crosses 3D7x3D1 and SW5xSW39 only is 0.43 (Kendall's rank correlation test, p-value < 2.2e-

16) supporting a high variability in recombination even between individual crosses of *Z. tritici*.

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 have not been tested. Therefore, we next assessed the robustness of the composite likelihood approach with respect to sample size and SNP density. We conducted simulations to assess the power of LD-based recombination estimators under such conditions. 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 size decrease the variance in estimate, we show that Ldhat reliably infers recombination when as few as 10 haploid genomes are used (Fig. 2). We observe that ρ generally tends to be underestimated and its estimation variance larger for small sample sizes. Yet better estimates can be obtained by discarding all estimates with a 95% confidence interval at least 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 suppress this bias (Fig. 2). We also note that the inference bias is stronger for low recombination rates, and that this effect is independent of the sample size (Fig. 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-value < 2.2e-16). Interestingly, correlations between the genetic map and the linkage disequilibrium (LD) map inferred here increases with increased window size: using 500 kb windows, the correlation becomes 0.43 (Kendall's tau, p-value = 0.000206) comparable to the correlation between the two genetic maps of 3D7x3D1 and SW5xSW39.

Recombination inference based on patterns of linkage disequilibrium 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 non-coding regions. To this end, we compared the previously published genetic map with estimates of ρ exclusively in the intergenic regions (excluding coding sequences and 500-bp

up and downstream of the annotated genes). These analyses based on non-coding sequences resulted in correlations of 0.22 for the Ldhat map and the average of the two genetic crosses (Kendall's rank correlation test, p-value < 2.2e-16) and 0.24 for the Ldhelmet map (Kendall's rank correlation test, p-value < 2.2e-16). Thus, the best correlations of LD based on the recombination maps and genetic crosses are obtained by complete genome data that include coding regions. 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 = 0.005 and filtered according to confidence intervals, for both *Z. tritici* and *Z. ardabiliae*.

A five fold 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 (Fig. 3 and Supplementary Data 1). We find a five-fold 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. This five-fold difference might reflect differences in actual recombination rates as well as differences in effective population sizes. 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. We further note that r represents the recombination rate per generation per nucleotide. Therefore, a putative difference in number of generations per year between the two pathogens also cannot account for the observed difference. The higher value of ρ estimated in *Z. tritici* thus likely reflects a higher actual recombination rate in the wheat pathogen compared to *Z. ardabiliae*.

Recombination on small arms of acro-centric chromosomes

Physical factors, such as chromosome length, chromosome arm length or distance to the centromere have been reported to impact broad-scale 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

294 reference genome sequence of *Z. tritici* consists of 21 fully sequenced chromosomes,
295 including eight so-called accessory chromosomes (Goodwin et al. 2011a). Furthermore, the
296 exact positions of the centromeres for all chromosomes have been characterized
297 experimentally using a chromatin immunoprecipitation assay targeting the centromere
298 specific protein CenH3 (Schotanus et al. 2015). An interesting finding is that the
299 chromosomes in *Z. tritici* are either acro-centric or near-acrocentric, and every chromosome
300 consequently consists of one long and one short chromosome arm (Schotanus et al. 2015).
301 Because a complete chromosome assembly is not available for *Z. ardabiliae*, we mapped the
302 recombination estimates of *Z. ardabiliae* on the genome of *Z. tritici* to assess the impact of
303 the karyotype structure on recombination rate variation. Similar to findings from other
304 species (Jensen-Seaman et al. 2004; Munch et al. 2014), we observe a negative correlation
305 between recombination rate and the size of the thirteen core chromosomes (Kendall's τ =
306 -0.59 with p-value = 4.29e-3 for *Z. tritici* and τ = -0.72 with p-value = 2.84e-4 for *Z. ardabiliae*;
307 Fig. 4A). This pattern is generally explained by the necessity of one crossing over to occur per
308 chromosome or chromosome arm per generation, resulting in a higher recombination rate
309 on smaller chromosomes (e.g., (2004; Smeds et al. 2016; Kong et al. 2002)). The significant
310 correlation of the recombination map of *Z. ardabiliae* with the genome structure of *Z. tritici*
311 is an indication of a conserved karyotype of the ancestral species of *Z. tritici* and *Z.*
312 *ardabiliae*.

313 Given the acro-centric nature of the *Z. tritici* chromosomes we considered to which extent
314 recombination also occurs on the short chromosome arms. If meiosis involves one crossover
315 event per chromosome, then the recombination rate should be correlated with the
316 chromosome size and not the chromosome arm length. However, if meiosis involves one
317 crossover event per chromosome arm, then a higher frequency of recombination should
318 occur on shorter chromosome arms. Correlations between recombination rates and
319 chromosome arm lengths also show negative values, yet only significant in *Z. ardabiliae*
320 (Kendall's τ = -0.14 with p-value = 0.3356 for *Z. tritici* and τ = -0.42 with p-value = 2.16e-3 for
321 *Z. ardabiliae*, Fig. 4B). The negative correlation observed at the chromosome arm level
322 suggests that meiosis in the *Zymoseptoria* pathogens requires at least one crossing over per
323 chromosome arm and that the small chromosome arms consequently also recombine. The
324 weaker correlations and lack of significance in *Z. tritici* could be due to a fast evolution of
325 centromere positions, erasing the signal of arm-specific recombination rates.

326

No association between recombination rate and GC content in *Z. tritici* and *Z. ardabiliae*

In many species recombination strongly impacts 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 (Duret and Galtier 2009; Piganeau et al. 2002), birds (Weber et al. 2014), plants (Serres-Giardi et al. 2012) and even bacteria (Lassalle et al. 2015). However, gBGC has never been assessed 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 non-homogeneous, non-stationary model of substitution in 10 kb windows in intergenic regions allowing us to estimate the equilibrium GC content (frequency of GC towards 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* (Supplemental Fig. 1, Kendall's rank correlation test, τ = 0.69 and 0.45, p-values < 2.2e-16 for the observed and equilibrium GC content, respectively, essential chromosomes only). However, although both species show similar observed GC content (mean 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* (mean equilibrium GC content on autosomes of 53.2, significantly higher that the observed GC content, Wilcoxon paired rank test, p-value = 0.04712) while decreasing in *Z. tritici* (mean equilibrium GC content of 51.6%, which is significantly lower than the observed GC content, Wilcoxon paired rank test, p-value = 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 weakly or non-significant correlations between GC content and recombination rate (Supplemental Fig. S1), both for observed (Kendal's tau = -0.05, p-value = 0.0404 for *Z. tritici* and tau = -0.05, p-value = 0.02253 for *Z. ardabiliae*) and equilibrium GC content (Kendal's tau = -0.02, p-value = 0.5082 for *Z. tritici* and tau = 0.01, p-value = 0.7128 for *Z. ardabiliae*).

Together these results do not support GC-biased gene conversion 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 crossing-over rate and GC content in *Z. tritici*, (Kendall's rank test, τ = 0.006 and p-value = 0.7035 for observed GC and τ = -0.024, p-value = 0.1149 for equilibrium GC content) supporting an absence or little effect of GC-biased gene conversion in *Z. tritici* and *Z. ardabiliae*.

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 *Drosophila melanogaster* and *Drosophila 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 kb 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 centeromere 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-value = 0.5771) (Table 2, Fig. 3). The centromeres of *Z. tritici* exhibit several features common to neocentromeres such as a short length (approx. 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. 2011a). They differ considerably from the core chromosomes as they display a higher repeat content, lower gene density, overall lower transcription rate and are enriched

391 with different chromatin modifications (Stukenbrock et al. 2010; Kellner et al. 2014;
392 Grandaubert et al. 2015; Schotanus et al. 2015). Electrophoretic separation of accessory
393 chromosomes from several isolates of *Z. ardabiliae* have shown this species also comprises
394 accessory chromosomes (Stukenbrock et al. 2011). In this study we used sequence homology
395 to define the accessory components of the *Z. ardabiliae* genome. We find that the aligned
396 fragments of the accessory chromosomes show very low recombination rates in both species
397 (median ρ = 0.0059 in *Z. tritici* and median ρ = 0.0001 in *Z. ardabiliae* over 13 10-kb windows
398 where both genomes could be aligned, which is 25% and 2% of the autosomal rates,
399 respectively) (Fig. 4C). The lower recombination rates reflect the lower effective population
400 size of accessory chromosomes that are present at lower frequencies in populations of *Z.*
401 *tritici* and *Z. ardabiliae* compared to the core chromosomes. Furthermore we speculate that
402 frequent structural rearrangements on accessory chromosomes can prevent homologous
403 chromosomes pairings and also contribute to the low recombination rates. Our findings add
404 further evidence to support different evolutionary modes of the two sets of chromosomes
405 (core and accessory chromosomes) contained in the same genome. As observed on the
406 accessory chromosomes, suppression of recombination is also found on mating-type
407 chromosomes in other fungi including species of *Neurospora* and *Microbotryum* (Hood et al.
408 2013; Petit et al. 2012; Whittle and Johannesson 2011). These regions are characterized by
409 an increased accumulation of transposable elements and structural variants as well as non-
410 adaptive mutations in coding sequences (Badouin et al. 2015; Whittle et al. 2011; Whittle
411 and Johannesson 2011).

412 We also observe a remarkable drop in the recombination rate on the right arm of
413 chromosome 7 (Supplemental Data 1). The right arm of chromosome 7 displays several
414 similarities to the DNA of the accessory chromosomes including a lower gene density, higher
415 repeat content and less gene transcription (Grandaubert et al. 2015). Furthermore, the
416 entire chromosome arm is enriched with the heterochromatic mark H3K27me3, which is
417 similarly enriched on the accessory chromosomes (Schotanus et al. 2015). We previously
418 proposed that this particular chromosome region represents a recent translocation of an
419 accessory chromosome to a core chromosome (Schotanus et al. 2015). This hypothesis is
420 consistent with the observation that the recombination rate of the chromosome arm
421 resembles the overall reduced recombination rate of the accessory chromosomes
422 (Supplemental Data 1).

423

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 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 with a minimum of 3 filtered SNPs (500-bp upstream and downstream CDS regions, respectively, Fig. 5A). Overall, we observe significant differences but with small size effects in fine-scale rates of recombination across different genome regions (Kruskal-Wallis test with post-hoc comparisons, FDR 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 (Fig. 5A). A lower value of $\rho = 2.N_e.r$ 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; Scally et al. 2012; Hobolth et al. 2011). The highly similar observed recombination rates in coding and non-coding sequences in *Z. tritici* and *Z. ardabiliae* suggests that r is not suppressed in these regions in the same way as observed in other organisms. The pattern indicates that different mechanisms define fine-scale recombination rates in these fungi leading 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 co-evolution. 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 co-evolution 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 1,122 and *Z. ardabiliae*.

By comparing the recombination rates in different genetic regions for effector and non-effector encoding genes, we show a significantly lower recombination rate in exons and introns for effector proteins in *Z. ardabiliae* (Wilcoxon rank test, p-value = 1.305e-4 (exons) and 2.534e-5 (introns), p-values corrected for multiple testing) (Fig. 5B). These differences are mostly driven by an excess of zero estimates in these regions, as visible on the distribution of measures (Fig. 5B). Discarding these regions with a mean recombination of zero leads to non-significant differences between effector and non-effector genes. A recombination rate estimated to zero can either be due to suppression of recombination in the region or to an estimation error. Intron and exons with a recombination estimate of zero in *Z. ardabiliae* are found to be shorter and to have a higher SNP density (Supplementary Data 3). 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. tritici* therefore appears as a biological signal which origin remains to be elucidated.

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. Broad-scale 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, broad-scales 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 results 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 showed that the genomes of the two species show a high extent of co-linearity 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 (Fig. 6, see Materials and Methods). Next, we calculated the average recombination rate in non-overlapping windows with at least 100 SNPs in each species, which resulted in 3,851 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, τ = 0.2327, p-value < 2.2e-16, Fig. 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 with the scale at which the comparison is performed. We find that the correlations, consistently inferred with different correlation measures, peak at 0.5-1 Mb scale (Fig. 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* (Fig. 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 of the most significant hotspots in *Z. tritici* and *Z. ardabiliae* (Fig. 8A). Interestingly, our approach revealed a considerably greater number of recombination hotspots in *Z. tritici* (2,578 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 base pairs) than in *Z. ardabiliae* (66 base pairs, Wilcoxon ranked test p-value < 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-value < 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 hotspots architecture.

While the differences to some extent can mirror the larger density of SNPs in *Z. tritici* that enables a finer resolution of the hotspot distribution and structure, 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, hotspot positions are highly conserved in some species (Singhal et al. 2015), and highly variable in other species (Myers et al. 2010). We mapped *Z. ardabiliae* hotspots on the *Z. tritici* genomes and counted the numbers of co-localizing hotspots in the two species. We considered that a hotspot in *Z. tritici* as co-localizing with a hotspot in *Z. ardabiliae* if the distance between the two hotspots is less than 1kb and if not other hotspot is present in between. We report that only 149 hotspots are co-localizing (6% of hotspots in *Z. tritici* and 20% of hotspots in *Z. ardabiliae*). This number is however significantly more than expected by chance (p-value < 9.99e-5, permutation test, Fig. 8B). These results are consistent with the previously reported genetic maps of *Z. tritici*, which also show little overlap of hotspots 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 2,578 *Z. tritici* hotspots, 132 are located in introns and 1,435 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 (Fig. 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 site (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 non-exclusive 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 re-arrangements including changes in ploidy and full chromosome gains or losses (e.g., (Hickman et al. 2013, 2015; Ma et al. 2010; Croll et al. 2013)). So far the importance of meiotic recombination in rapid evolution of pathogens has been poorly addressed. Our analyses demonstrate extraordinary high recombination rates in two fungal plant pathogens and thereby suggest that sexual recombination also can 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 propagate only in managed ecosystems. Agricultural management strategies, dense host populations and increased gene flow between geographically distant populations are factors that contribute to a 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.

Materials and methods

Genome data

The lifecycle 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. 2011a). Two Iranian *Z. tritici* isolates and four Iranian *Z. ardabiliae* isolates were sequenced in a previous study using Illumina sequencing (Table S1) (Stukenbrock et al. 2011). We used genome data from an additional ten isolates of *Z. tritici* that originate from wheat fields in Denmark, France and Germany (Grandaubert, Dutheil and Stukenbrock, in prep). In this study, we report the genome sequences of hirteen 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. Sequence data has been deposited under the NCBI BioProject IDs PRJNA277174.

The thirteen *Z. ardabiliae* re-sequenced 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 RNAseq based annotation to distinguish the parameter estimates for coding and non-coding 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 sequence predicted to be secreted by SignalP (Petersen et al. 2011)

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: IPO123 for *Z. tritici* and STO4IR-1.1.1 for *Z. ardabiliae* (Goodwin et al. 2011a; Stukenbrock et al. 2011). The projected alignments in MAF format were filtered using the MafFilter program [27] with the following filters: 1) each syntenic block was realigned using Mafft (Katoh et al. 2009), and blocks with more than 10 kb were split for computer 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 one position with gaps in at least 2 species 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 more than 100 gaps 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',

621 which resulted in one masked alignment per chromosome for *Z. tritici* and one masked
622 alignment per contig for *Z. ardabiliae*. The chromosome and contig alignments were further
623 divided in non-overlapping windows of 1 Mb (data set 1) or 100 kb (data set 2). The MafFilter
624 program was further used to estimate statistics on the alignments at each filtering step, and
625 to compute the nucleotide diversity (Watterson's θ) from the final filtered genome
626 alignments.

627

628 **Estimating recombination**

629 Filtered alignments (1-Mb windows, data set 1) were exported as fasta files for the Ldhat and
630 Ldhelmet packages. The program *convert* from the Ldhat package was used to convert fasta
631 files into input loci files for the program *interval* (Auton and McVean 2007). Only fully
632 resolved biallelic positions were exported (see Table 1 for the details of SNP numbers).
633 Likelihood tables were generated for θ values of 0.0005, 0.005 and 0.05. The *interval*
634 program was run with 10,000,000 iterations and sampled every 5,000 iterations with a burn-
635 in of 100,000 iterations. Ldhelmet was run with the parameters suggested in the user manual
636 ((Chan et al. 2012) and https://sourceforge.net/projects/ldhelmet/). We calculated average
637 recombination rates in windows and regions by taking the average of recombination
638 estimates between every pairs of SNPs, weighted by the physical distance between the SNPs.
639 Pairs of SNPs for which the confidence interval of the recombination estimate was higher
640 than two times the mean were discarded and therefore not used in the average
641 computation. Using the gene annotations available for the two reference species
642 (Grandaubert et al. 2015), we calculated the following information for each gene: 1) the
643 average recombination rate in exons, 2) the average recombination rate in introns, and 3) the
644 average recombination rate in the 500 bp flanking 5' region and 4) in the 500 bp flanking 3'
645 region. We also calculated the average recombination rate for each intergenic region (500 bp
646 from / to genes). GFF3 files from (Grandaubert et al. 2015) were retrieved and processed
647 using the "genometools" package to add intron annotations (Gremme et al. 2013). The
648 resulting gene annotations were analyzed in R together with recombination maps (R Core
649 Team 2013).

650

651 **Assessment of LD-based recombination estimates by simulation**

652 We used the SCRM coalescent simulator (Staab et al. 2015) in order to simulate
653 polymorphism data with a constant mutation rate but variable recombination rate.

Recombination rates were drown randomly from an exponential distribution with mean 0.02. Segments with piecewise constant recombination rate were taken randomly from an exponential distribution with mean 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 exact 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 real 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 "lift-over" 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. Non -overlapping windows containing at least 100 analysed 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 window 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, 1,000 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 1,024 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 bi-normality 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 to assess 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 and their confidence intervals (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 [28]. 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 1,000 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 9,133 hotspots in *Z. tritici* and 1,287 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 ten times higher than the flanking regions. Thus 2,578 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 co-localizing with a hotspot in *Z. Ardabiliae* if the distance between them was less than 1kb, 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. In order to do so, we used the following procedure:

1) compute the total "inter-hostpots" 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, then expand each interval to match its original size and compute the corresponding end positions. Correct the coordinates in order to account for previous intervals.

We assessed the significance of the number of co-localizing hotspots using 10,000 permutations. The corresponding R scripts are available as Supplementary Data 3.

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 block longer than 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 more than 29 gaps 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 to 13) were retained. The final alignment contained 2,052 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 non-homogeneous, non-stationary model of substitution, allowing us to estimate three distinct GC contents for *Z. tritici*, *Z. ardabiliae* and their common ancestor. Other parameters were consider 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 4 classes 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 windows containing at least 1% polymorphic position (leaving 1,642 windows).

As similar analysis was conducted using recombination rate estimated from (Croll et al. 2015) which were calculated in 20 kb windows. The corresponding pairwise alignment regions were extracted and filtered, and coding regions from both species were discarded, which resulted in 1,948 windows of at least 1 kb where a non-homogeneous, non-stationary model of substitution could be fitted.

Acknowledgements

EHS is supported by intramural funding from the Max Planck Society, Germany and a personal grant from the State of Schleswig-Holstein, Germany. JYD is supported by intramural funding from the Max Planck Society, Germany. The authors thank Daniel Croll for providing genetic data from experimental crosses of *Z. tritici*.

Literature

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. 2012. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. *Science (80-)* **336**: 193–198.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. *Genome Res* **17**: 1219–1227. http://genome.cshlp.org/content/17/8/1219.abstract.

Auton A, Myers S, McVean G. 2014. Identifying recombination hotspots using population genetic data. *arXiv Prepr arXiv14034264*.

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

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus Microbotryum lychnidis-dioicae. *Genetics* **200**: 1275–1284.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature* **356**: 519–520.

Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced Effectiveness of Selection Caused by a Lack of Recombination. *Curr Biol* **19**: 655–660.

783 Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K,
784 Clawson H, Green ED, et al. 2004. Aligning Multiple Genomic Sequences With the
785 Threaded Blockset Aligner. *Genome Res* **14**: 708–715.
786 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive Human
787 Genetic Maps: Individual and Sex-Specific Variation in Recombination. *Am J Hum Genet*
788 **63**: 861–869.
789 Brunner PC, Stefanato FL, McDonald BA. 2008. Evolution of the CYP51 gene in
790 Mycosphaerella graminicola: evidence for intragenic recombination and selective
791 replacement. *Mol Plant Pathol* **9**: 305–316.
792 Chan AH, Jenkins PA, Song YS. 2012. Genome-Wide Fine-Scale Recombination Rate Variation
793 in <italic>Drosophila melanogaster</italic>. *PLoS Genet* **8**: e1003090.
794 Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA,
795 Copenhaver GP, Franklin FCH, et al. 2013. Arabidopsis meiotic crossover hot spots
796 overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet* **45**: 1327–1336.
797 Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The Impact of Recombination
798 Hotspots on Genome Evolution of a Fungal Plant Pathogen. *Genetics* **201**: 1213–1228.
799 Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions
800 contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS*
801 *Genet* **9**: e1003567.
802 de Castro E, Soriano I, Marín L, Serrano R, Quintales L, Antequera F. 2011. Nucleosomal
803 organization of replication origins and meiotic recombination hotspots in fission yeast.
804 *EMBO J* **31**: 124 LP-137.
805 Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the
806 human genome. *PLoS Genet* **4**.
807 Duret L, Galtier N. 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic
808 Landscapes. *Annu Rev Genomics Hum Genet* **10**: 285–311.
809 Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++
810 suite of libraries and programs. *BMC Evol Biol* **8**: 1.
811 Dutheil JY, Gaillard S, Stukenbrock EH. 2014. MafFilter: a highly flexible and extensible
812 multiple genome alignment files processor. *BMC Genomics* **15**: 53.
813 Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der
814 Lee TAJ, Grimwood J, Aerts A, et al. 2011a. Finished Genome of the Fungal Wheat
815 Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome
816 Plasticity, and Stealth Pathogenesis. *PLoS Genet* **7**: e1002070.

817 Goodwin SB, M'barek S Ben, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der
818 Lee T a J, Grimwood J, Aerts A, et al. 2011b. Finished genome of the fungal wheat
819 pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome
820 plasticity, and stealth pathogenesis. *PLoS Genet* **7**: e1002070.
821 Grandaubert J, Bhattacharyya A, Stukenbrock EH. 2015. RNA-seq Based Gene Annotation and
822 Comparative Genomics of Four Fungal Grass Pathogens in the Genus Zymoseptoria
823 Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements.
824 *G3 Genes|Genomes|Genetics* 5.7: 1323-1333
825 Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for
826 Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans Comput Biol*
827 *Bioinforma* **10**: 645–656.
828 Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock
829 of mitochondrial DNA. *J Mol Evol* **22**: 160–174.
830 Hickman MA, Paulson C, Dudley AM, Berman J. 2015. Parasexual Ploidy Reduction Drives
831 Population Heterogeneity Through Random and Transient Aneuploidy. *Genet* .
832 Hickman MA, Zeng G, Forche A, Hirakawa MP, Abbey D, Harrison BD, Wang Y-M, Su C,
833 Bennett RJ, Wang Y, et al. 2013. The /`obligate diploid/' Candida albicans forms mating-
834 competent haploids. *Nature* **494**: 55–59.
835 Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting
836 patterns among human, chimpanzee, and orangutan suggest recent orangutan
837 speciation and widespread selection. *Genome Res* **21**: 349–356.
838 Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type chromosomes
839 of the anther-smut fungus. *Genetics* **193**.
840 Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A,
841 Sperone FG, Vilhjalmsson BJ, et al. 2012. Genome-wide patterns of genetic variation in
842 worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nat Genet* **44**: 212–
843 216.
844 Hunter CM, Huang W, Mackay TFC, Singh ND. 2016. The Genetic Architecture of Natural
845 Variation in Recombination Rate in <italic>Drosophila melanogaster</italic>. *PLoS*
846 *Genet* **12**: e1005951.
847 Jeffreys AJ, Murray J, Neumann R. 1998. High-Resolution Mapping of Crossovers in Human
848 Sperm Defines a Minisatellite-Associated Recombination Hotspot. *Mol Cell* **2**: 267–273.
849 Jeffreys AJ, Neumann R. 2009. The rise and fall of a human recombination hot spot. *Nat*
850 *Genet* **41**: 625–629.

851 Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D,
852 Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human
853 genomes. *Genome Res* **14**: 528–538.
854 Jones JD, Dangl JL. 2006. The plant immune system. *Nature* **444**: 323-329
855 Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond.
856 *Nat Rev Genet* **13**: 484–492.
857 Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes
858 TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a
859 eukaryotic genome. *Nature* **458**: 362–366.
860 Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT.
861 *Bioinforma DNA Seq Anal* 39–64.
862 Kellner R, Bhattacharyya A, Poppe S, Hsu TY, Brem RB, Stukenbrock EH. 2014. Expression
863 Profiling of the Wheat Pathogen Zymoseptoria tritici Reveals Genomic Patterns of
864 Transcription and Host-Specific Regulatory Programs. *Genome Biol Evol* **6**: 1353–1365.
865 Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B,
866 Sigurdardottir S, Barnard J, Hallbeck B, Masson G. 2002. A high-resolution
867 recombination map of the human genome. *Nat Genet* **31**: 241–247.
868 Lam I, Keeney S. 2015. Nonparadoxical evolutionary stability of the recombination initiation
869 landscape in yeast. *Science (80-)* **350**: 932–937.
870 Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, Daubin V. 2015. GC-Content Evolution in
871 Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLoS Genet* **11**:
872 e1004941.
873 Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S,
874 Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol* **66**:
875 513–545.
876 Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-JM-J, Di Pietro A, Dufresne
877 M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals
878 mobile pathogenicity chromosomes in Fusarium. *Nature* **464**: 367–373.
879 Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of
880 meiotic crossovers and non-crossovers in yeast. *Nature* **454**:479-485
881 Marais G, Mouchiroud D, Duret L. 2003. Neutral effect of recombination on base composition
882 in Drosophila. *Genet Res (Camb)* **81**: 79–87.
883 McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J,
884 Acharya C, Bottoms C, et al. 2009. Genetic Properties of the Maize Nested Association

885 Mapping Population. *Science (80-)* **325**: 737 LP-740.
886 Meunier J, Duret L. 2004. Recombination Drives the Evolution of GC-Content in the Human
887 Genome. *Mol Biol Evol* **21**: 984–990.
888 Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination
889 landscape and demography to genomic base composition. *BioEssays* **37**: 1317–1326.
890 Munch K, Mailund T, Dutheil JY, Schierup MH. 2014. A fine-scale recombination map of the
891 human–chimpanzee ancestor reveals faster change in humans than in chimpanzees and
892 a strong impact of GC-biased gene conversion. *Genome Res* **24**: 467–474.
893 Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A Fine-Scale Map of
894 Recombination Rates and Hotspots Across the Human Genome. *Science (80-)* **310**:
895 321–324.
896 Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P.
897 2010. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic
898 Recombination. *Science (80-)* **327**: 876–879.
899 Nordborg M, Charlesworth B, Charlesworth D. 1996. The effect of recombination on
900 background selection. *Genet Res (Camb)* **67**: 159–174.
901 Petersen TN, Brunak S, Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides
902 from transmembrane regions. *Nat Methods* **8**.
903 Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nat Rev Genet* **2**: 360–369.
904 Petit E, Giraud T, Vienne DM, Coelho MA, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory
905 F, Wincker P, et al. 2012. Linkage to the mating-type locus across the genus
906 Microbotryum: insights into nonrecombining chromosomes. *Evol Int J Org Evol* **66**.
907 Piganeau G, Mouchiroud D, Duret L, Gautier C. 2002. Expected relationship between the
908 silent substitution rate and the GC content: implications for the evolution of isochores.
909 *J Mol Evol* **54**.
910 R Core Team. 2013. R: A language and environment for statistical computing.
911 Rizzon C, Marais G, Gouy M, Biémont C. 2002. Recombination Rate and the Distribution of
912 Transposable Elements in the Drosophila melanogaster Genome. *Genome Res* **12**: 400–
913 407.
914 Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T,
915 Mailund T, Marques-Bonet T. 2012. Insights into hominid evolution from the gorilla
916 genome sequence. *Nature* **483**: 169–175.
917 Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M,
918 Stukenbrock EH. 2015. Histone modifications rather than the novel regional

centromeres of Zymoseptoria tritici distinguish core and accessory chromosomes. *Epigenetics Chromatin* **8**: 41.

Serres-Giardi L, Belkhir K, David J, Glémin S. 2012. Patterns and Evolution of Nucleotide Landscapes in Seed Plants. *Plant Cell* **24**: 1379–1397.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science (80-)* **350**: 928–932.

Smeds L, Mugal CF, Qvarnström A, Ellegren H. 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 KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. 2011. Heterochromatin is required for normal distribution of Neurospora crassa CenH3. *Mol Cell Biol* **31**: 2528–2542.

Spencer CCA, Deloukas P, Hunt S, Mullikin J, Myers S, Silverman B, Donnelly P, Bentley D, McVean G. 2006. The influence of recombination on human genetic diversity. *PLoS Genet* **2**: e148.

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytol* **210**: 743–761. http://dx.doi.org/10.1111/nph.13794.

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

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

Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonald BA, Wang J, Schierup MH. 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 EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. 2010. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen mycosphaerella graminicola. *PLoS Genet* **6**: 1–13.

Stukenbrock EH, McDonald BA. 2009. Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Mol Plant Microbe Interact* **22**: 371–380.

Stukenbrock EH, Quaedvlieg W, Javan-Nikhah M, Zala M, Crous PW, McDonald BA. 2012.

953 Zymoseptoria ardabiliae and Z. pseudotritici, two progenitor species of the septoria
954 tritici leaf blotch fungus Z. tritici (synonym: Mycosphaerella graminicola). *Mycologia*
955 **104**: 1397–407.
956 Stumpf MPH, McVean GAT. 2003. Estimating recombination rates from population-genetic
957 data. *Nat Rev Genet* **4**: 959–968.
958 Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE. 2015. Clonal reproduction in fungi.
959 *Proc Natl Acad Sci* **112**: 8901–8908.
960 True JR, Mercer JM, Laurie CC. 1996. Differences in Crossover Frequency and Distribution
961 Among Three Sibling Species of Drosophila. *Genetics* **142**: 507–523.
962 http://www.genetics.org/content/142/2/507.abstract.
963 Tsai IJ, Burt A, Koufopanou V. 2010. Conservation of recombination hotspots in yeast. *Proc*
964 *Natl Acad Sci* **107**: 7847–7852. http://www.pnas.org/content/107/17/7847.abstract.
965 Wallberg A, Glémin S, Webster MT. 2015. Extreme Recombination Frequencies Shape
966 Genome Variation and Evolution in the Honeybee, <italic>Apis mellifera</italic>. *PLoS*
967 *Genet* **11**: e1005189.
968 Wang Y, Rannala B. 2014. Bayesian Inference of Shared Recombination Hotspots Between
969 Humans and Chimpanzees. *Genetics* **198**: 1621–1628.
970 Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. 2014. Evidence for GC-biased gene
971 conversion as a driver of between-lineage differences in avian base composition.
972 *Genome Biol* **15**: 1–16.
973 Whittle CA, Johannesson H. 2011. Evidence of the accumulation of allele-specific non-
974 synonymous substitutions in the young region of recombination suppression within the
975 mating-type chromosomes of Neurospora tetrasperma. *Heredity* **107**: 305-314
976 Whittle CA, Sun Y, Johannesson H. 2011. Degeneration in codon usage within the region of
977 suppressed recombination in the mating-type chromosomes of Neurospora
978 tetrasperma. *Eukaryot Cell* **10**: 594-603
979 Wijnker E, Velikkakam James G, Ding J, Becker F, Klasen JR, Rawat V, Rowan BA, de Jong DF,
980 de Snoo CB, Zapata L, et al. 2013. The genomic landscape of meiotic crossovers and
981 gene conversions in Arabidopsis thaliana ed. G. McVean. *Elife* **2**: e01426.
982 Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GAT,
983 Gabriel SB, Reich D, Donnelly P, et al. 2005. Comparison of Fine-Scale Recombination
984 Rates in Humans and Chimpanzees. *Science (80-)* **308**: 107–111.
985 A language and environment for statistical computing. http://www.R-project.org
986 2004. Sequence and comparative analysis of the chicken genome provide unique

perspectives on vertebrate evolution. *Nature* **432**: 695–716.

Tables

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	39,686,251 bp		31,546,591 bp	
Number of exonic sites in reference genome	17,296,247 bp (43.6%)		15,570,421 bp (49.4%)	
Number of haplotypes	13		17	
Summary genome alignment	**Total alignment length**	**Number of alignment blocks**	**Total alignment length**	**Number of alignment blocks**
MultiZ alignment	40.8 Mb	21,500	32.4 Mb	22,296
Splitting in max 10 Kb	40.8 MB	21,904	32.4 Mb	23,001
MAFFT Realignment	40.5 Mb	21,904	32.2 Mb	23,001
Keep blocks with all strains	27.7 Mb	6,455	28.2 Mb	7,117
Filter 1	27.5 Mb	15,703	28.0 Mb	18,402
Filter 2	27.3 Mb	18,785	27.7 Mb	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**
Min. number of SNPs	143	0	0	0
Median number of SNPs	43,680	3,556	1,598	634
Max. number of SNPs	102,400	15,170	33,680	20,110
Diversity (median of Watterson's theta in windows of 10 kb)	0.0139		0.008663	

994

995 **Table 2:** Recombination and repeat content in centromeres of *Z. tritici*.

	Chromosome	Start	Stop	Length	Mean rho	Nb. of SNPs in centromere	Mean rho for Full chromosome	Repeat density	TE density
Essential	1	3839299	3851749	12450	0.229	20	0.021	0.94%	31.33%
	2	512901	521916	9015	0.053	77	0.024	0.00%	32.39%
	3	3348307	3356535	8228	0.097	269	0.025	0.00%	0.00%
	4	217113	226545	9432	0.033	421	0.028	0.00%	9.88%
	5	2604117	2614736	10619	0.104	47	0.027	0.94%	28.19%
	6	625186	637601	12415	NA	0	0.026	3.10%	37.46%
	7	255824	266207	10383	0.006	79	0.044	0.32%	0.00%
	8	213892	227444	13552	0.059	62	0.029	0.45%	39.99%
	9	2067589	2076063	8474	0.015	106	0.040	0.50%	0.00%
	10	99716	109365	9649	0.016	77	0.049	0.00%	15.32%
	11	365130	373557	8427	NA	0	0.049	0.00%	46.30%
	12	180233	188209	7976	0.001	150	0.052	2.48%	7.10%
	13	236993	242558	5565	0.015	156	0.037	0.50%	0.00%
Dispensable	14	59960	70870	10910	0.000	785	0.000	0.00%	35.86%
	15	382500	394754	12254	0.001	1098	0.001	0.86%	20.04%
	16	332004	342592	10588	0.099	83	0.023	0.00%	35.97%
	17	406958	418893	11935	NA	0	0.000	0.24%	46.85%
	18	159000	171999	12999	NA	0	0.159	0.00%	46.62%
	19	148227	159387	11160	0.001	4	0.000	0.76%	1.38%
	20	94677	105169	10492	NA	0	0.008	0.30%	11.86%
	21	340264	346657	6393	NA	0	NA	0.31%	2.33%

997

998 # Figure legends

999 **Figure 1: Correlations among recombination maps in *Z. tritici* show highly correlated**
1000 **estimates from two composite likelihood methods.** A) Correlation circle of the six
1001 population genomic recombination maps based on the two first principal components. The
1002 programs Ldhat interval (Auton and McVean 2007) and Ldhelmet (Chan et al. 2012) were
1003 both used with three distinct input scaled effective population sizes (Θ) of 0.0005, 0.005 and
1004 0.05. B) Correlation of the Ldhat and Ldhelmet maps with Θ = 0.005. The Ldhat map was
1005 discretized into 10 categories with equal number of points. The points represent the mean
1006 +/- the standard error for each category. C) To assess the quality of the inferred
1007 recombination maps, genome-wide estimates of recombination were correlated with a
1008 genetic map obtained by experimental crossing of *Z. tritici* isolates. Correlations between
1009 population genomic maps (obtained by Ldhat and Ldhelmet) with a scaled population size of
1010 0.005 and the average recombination map from two independent crosses (Croll et al. 2015).
1011
1012 **Figure 2: effect of sample size and diversity on the estimation of recombination rate by**
1013 **Ldhat.** 10 Mb regions (1Mb for regions with θ = 0.05) were simulated using a coalescent
1014 model with variable recombination rate. Dots are average of point estimates of local
1015 recombination rate inferred using Ldhat ("interval" program). Each dot corresponds to a
1016 region with constant recombination rate in the simulated alignment. Bars indicate the 1st and
1017 3rd quartiles of Ldhat estimates for the region. Grey points are raw estimates; black points are
1018 computed from filtered estimates (see Methods). The red diagonal line shows the 1:1 ratio.
1019 Columns indicate distinct population mutation rate (θ = 4 Ne u) and rows distinct sample
1020 sizes (number of haploid genomes).
1021
1022 **Figure 3: Variation in recombination rate across chromosomes.** Based on the population
1023 genomics data of *Z. tritici* and *Z. ardabiliae*, genome-wide patterns of recombination are
1024 estimated. Patterns of variation across chromosome 1 of *Z. tritici* is shown as example. A)
1025 SNP density in 10 kb windows with corresponding smoothing curve. B) Distribution of called
1026 sites along the chromosome in black, corresponding to the regions that were included in the
1027 analyses. C) Estimates of the population recombination rate ρ show a highly heterogeneous
1028 recombination landscape across the chromosomes. D) Observed GC content. The position of
1029 the centromere of chromosome 1 is marked over the chromosome plots as a vertical stippled
1030 line.
1031

Figure 4: Broad-scale recombination rates in *Z. tritici* and *Z. ardabiliae*. Broad-scaled patterns of recombination rate in *Z. tritici* and *Z. ardabiliae* demonstrate a strong effect of chromosome size and chromosome type. A) Mean recombination rate in *Z. tritici* and *Z. ardabiliae* per essential chromosome as a function of the chromosome size. B) Mean recombination rate per essential chromosome arm as a function of the arm size. C) Distribution of mean recombination rate per chromosome in Z. tritici as a function of chromosome type (essential or accessory).

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 non-coding, 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 effector and non-effector 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 effector genes compared to non-effector genes (Wilcoxon rank test corrected for multiple testing, NS: non significant, *: 5% level, ***: below 0.1% level).

Figure 6: Recombination maps of *Z. tritici* and *Z. ardabiliae* plotted along the chromosome 1 of *Z. tritici*. A) Recombination map in 100 kb windows plotted together with smoothing curves. B) Cumulative curves of the recombination maps, scaled in order to be comparable. Figures for other chromosomes are available as Supplementary Data.

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 number of windows, error bars represent the mean +/- standard error. 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 1,000 samples and bars shows the standard errors of the means. Lines correspond to local regression smoothing (LOES).

Figure 8: Distribution of hotspots in the genomes of *Z. tritici* and *Z. ardabiliae*. A: example 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, width and strength 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.

Supplementary Material

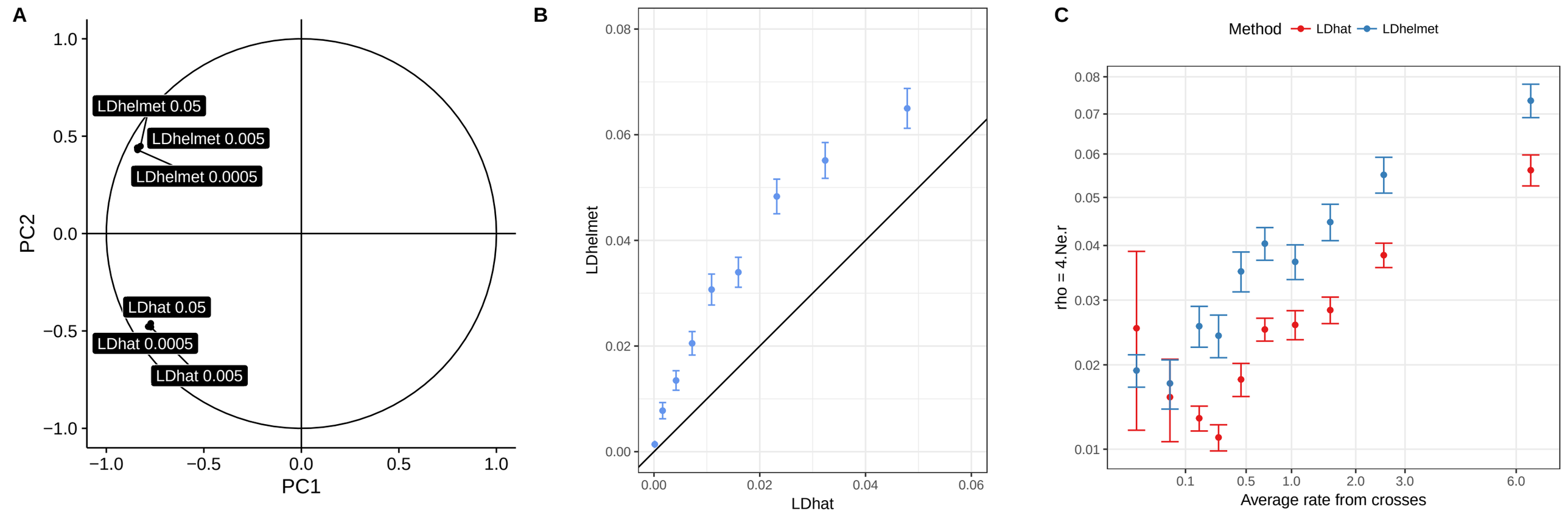
Table S1: Summary information of *Z. tritici* and *Z. ardabiliae* isolates used in the study.

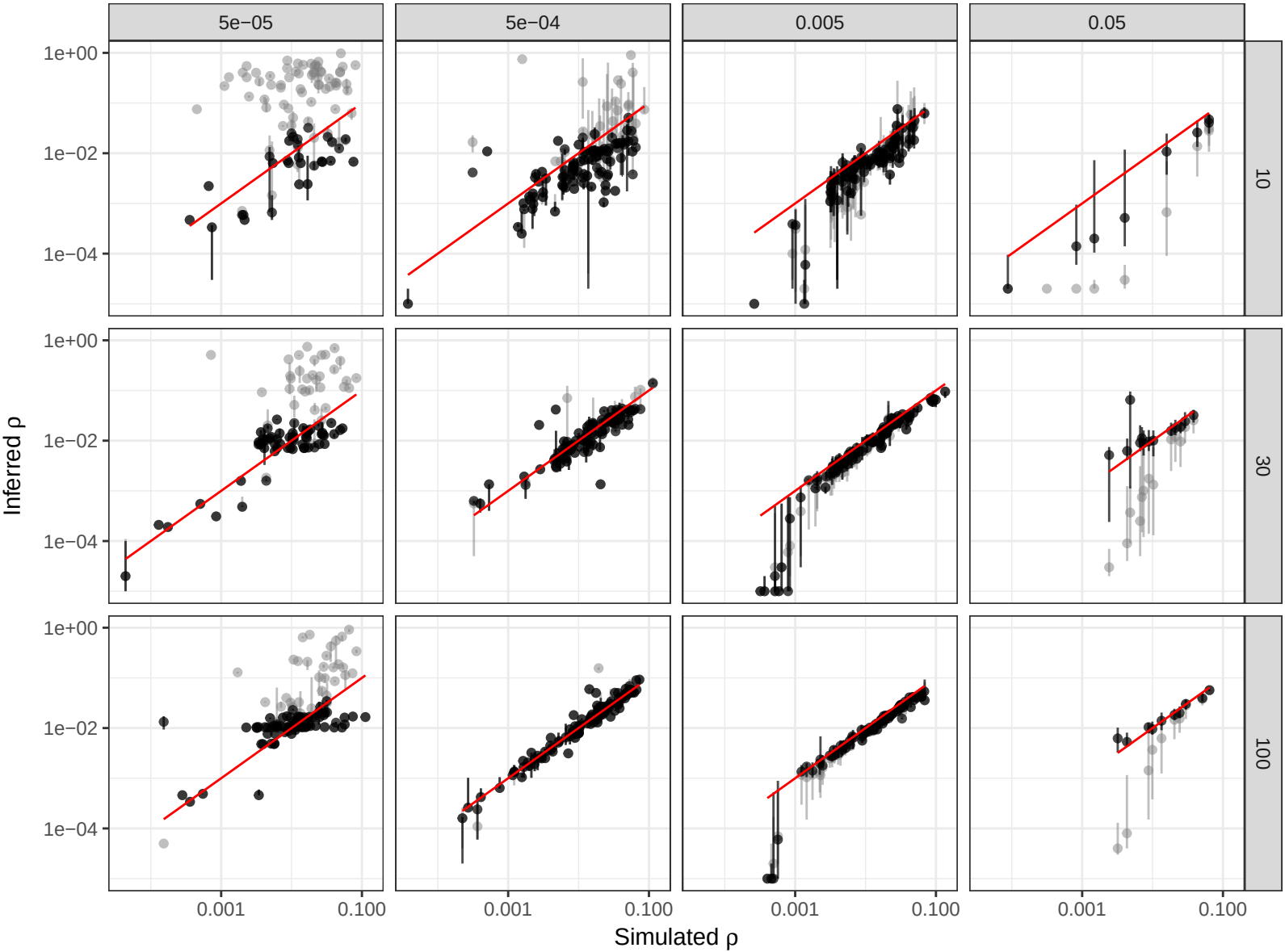
Figure S1: Genome-wide recombination rate and GC content. A) Observed GC content in *Z. tritici* plotted against observed GC content in *Z. ardabiliae*. B) Equilibrium GC content in *Z. tritici* plotted against equilibrium GC content in *Z. ardabiliae*. C) GC content as a function of recombination rate. Recombination rate was discretized in 10 categories with the same amount of points. Points indicate the mean GC content in each category; and bars correspond to standard errors of the means.

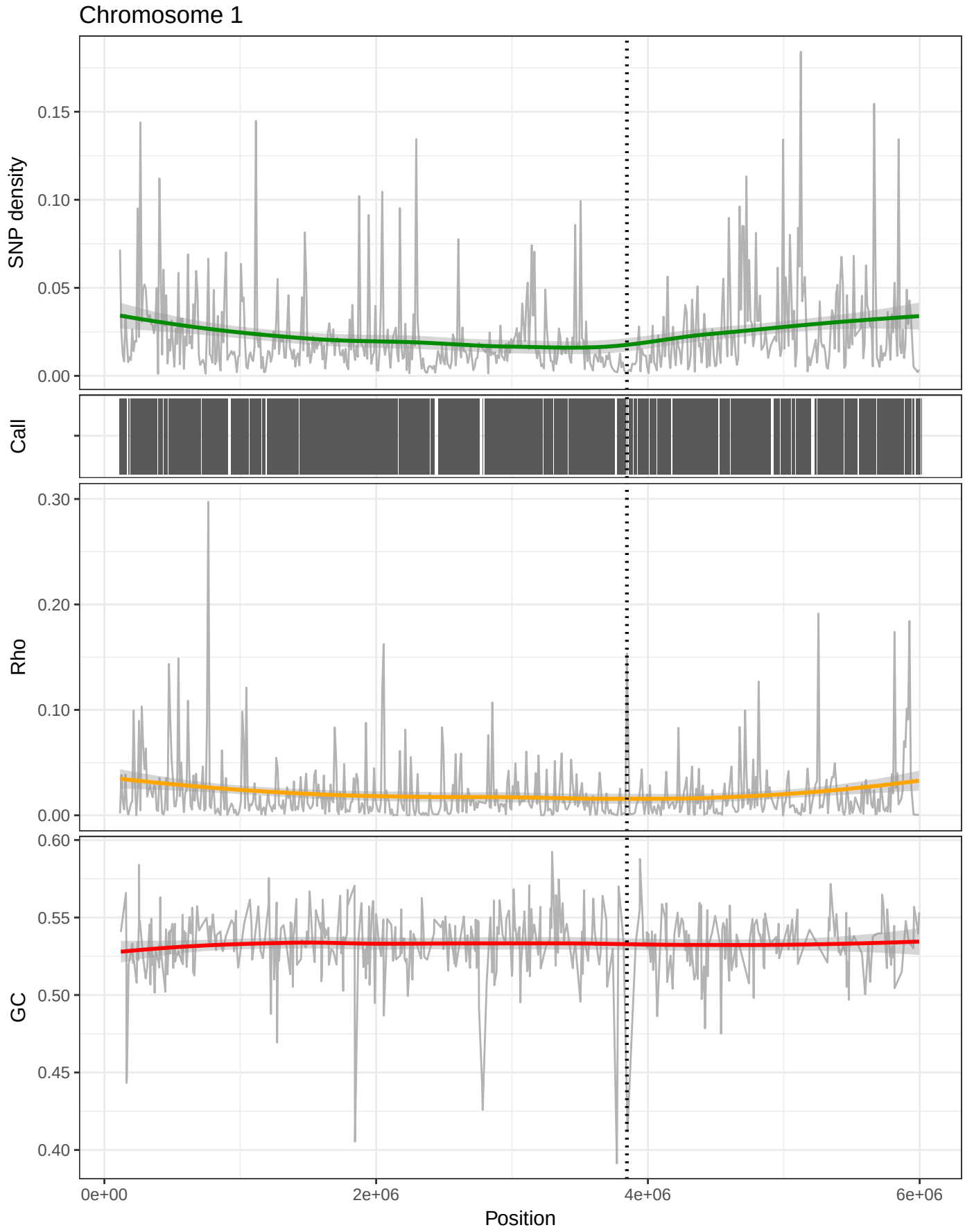
Supplementary Data 1: Chromosomal patterns for every chromosomes. Legends as in Figure 3.

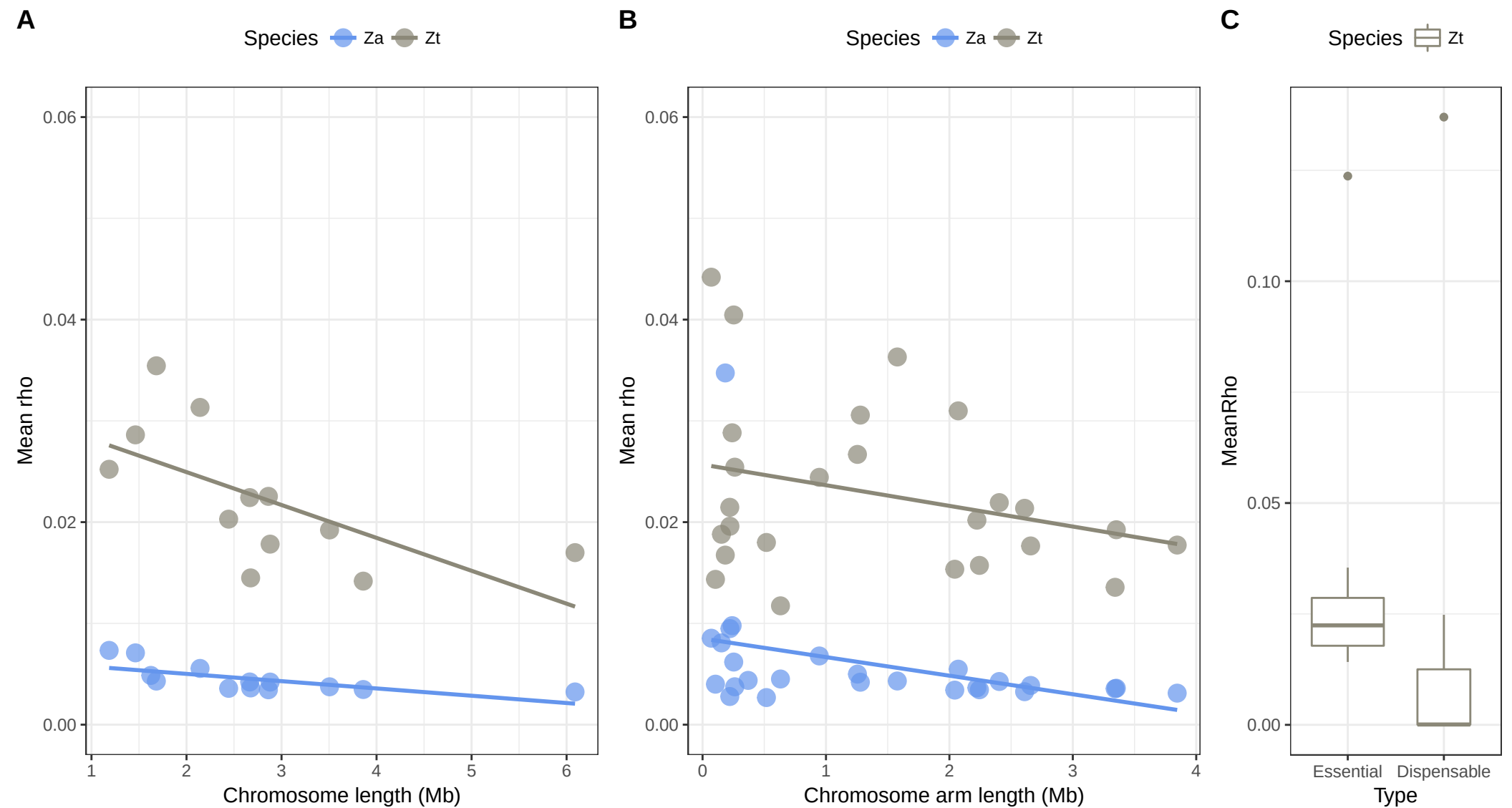
Supplementary Data 2: Correlation of recombination maps for every chromosomes. Legends as in Figure 6.

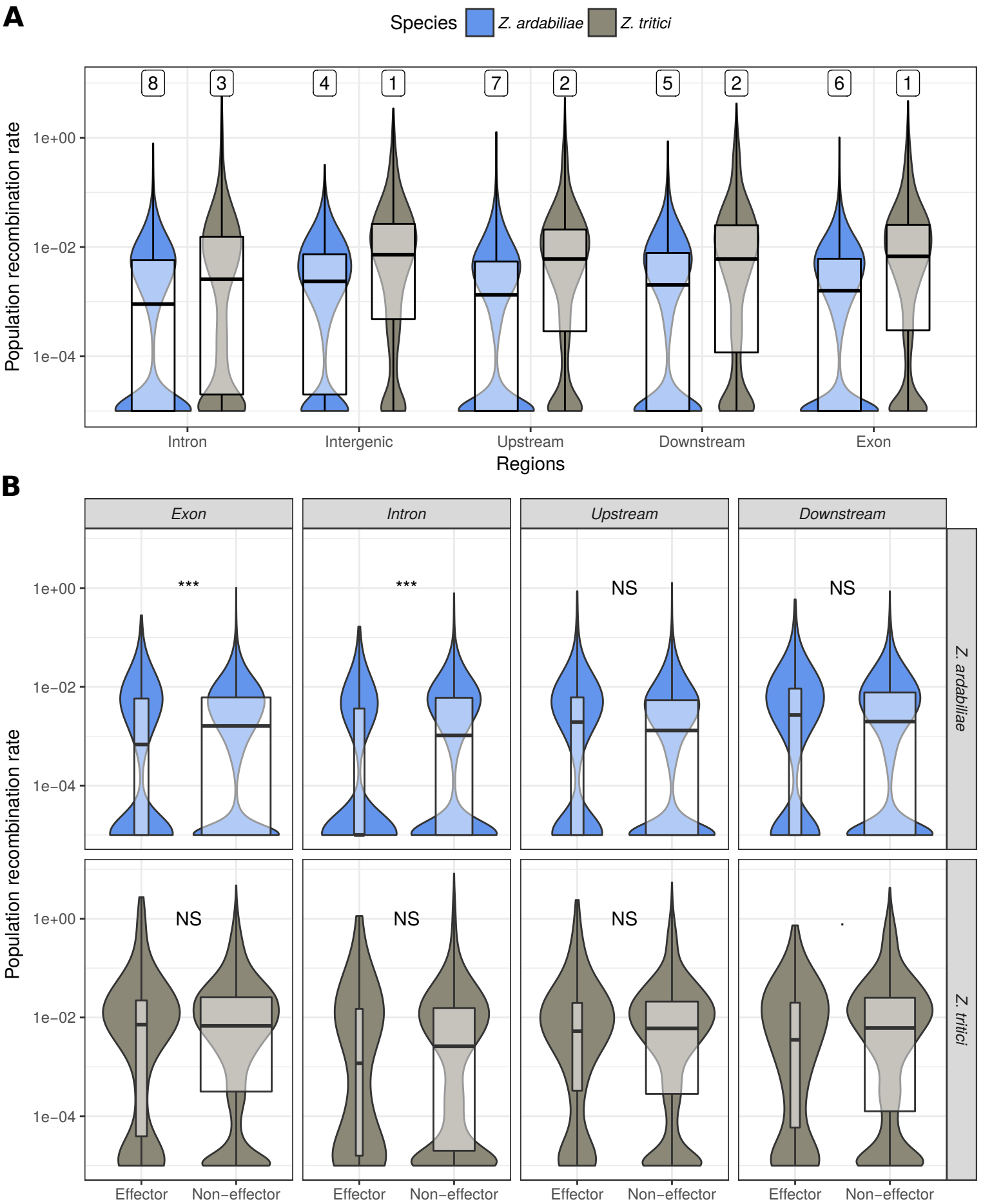
Supplementary Data 3: All scripts and data allowing reproducing results and figures in this manuscript (deposited on FigShare).

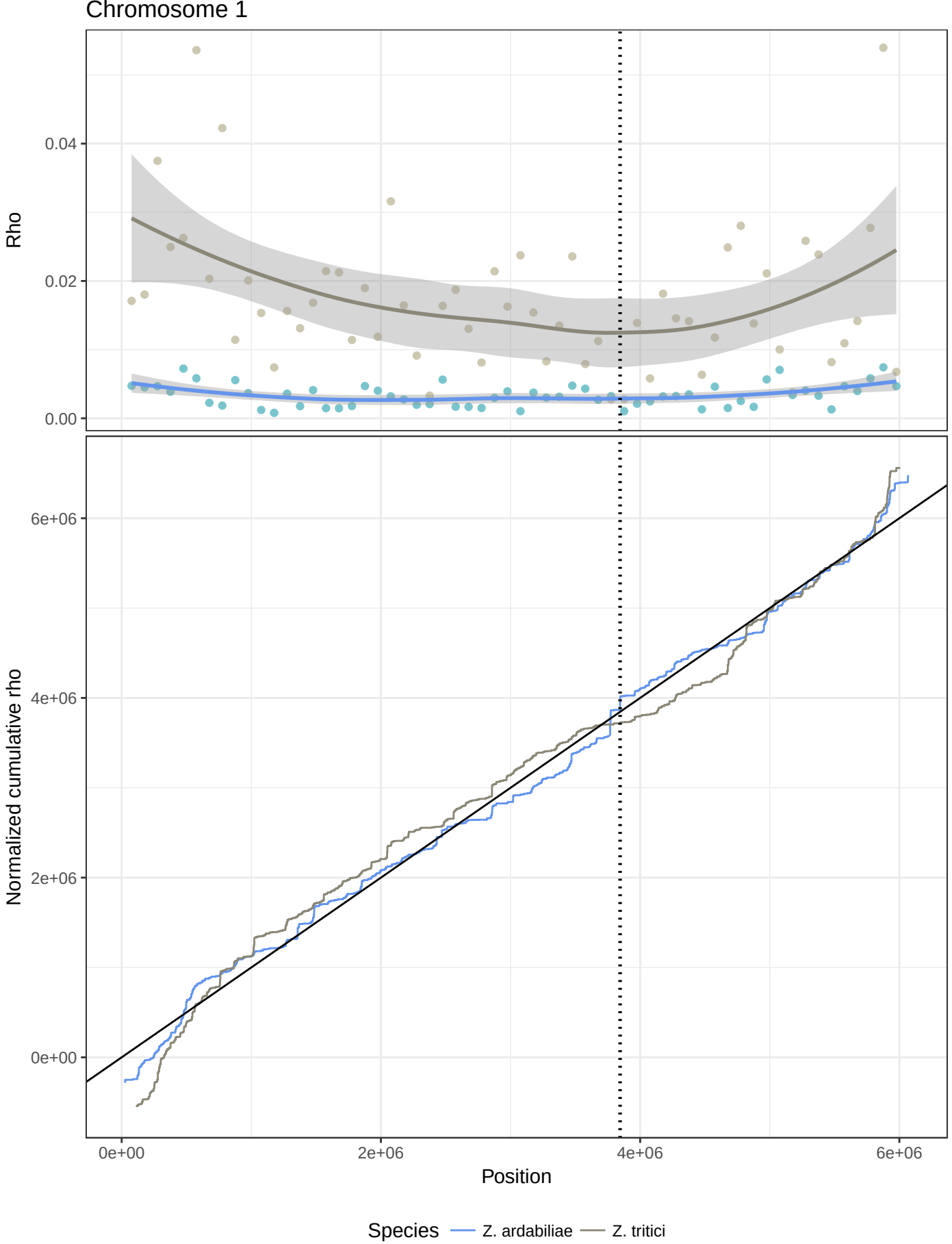

A
LDhelmet 0.05
LDhelmet 0.005
LDhelmet 0.0005
LDhat 0.05
LDhat 0.0005
LDhat 0.005
PC1
PC2
B
LDhelmet
LDhat
C
Method
LDhat
LDhelmet
rho = 4.Ne.r
Average rate from crosses

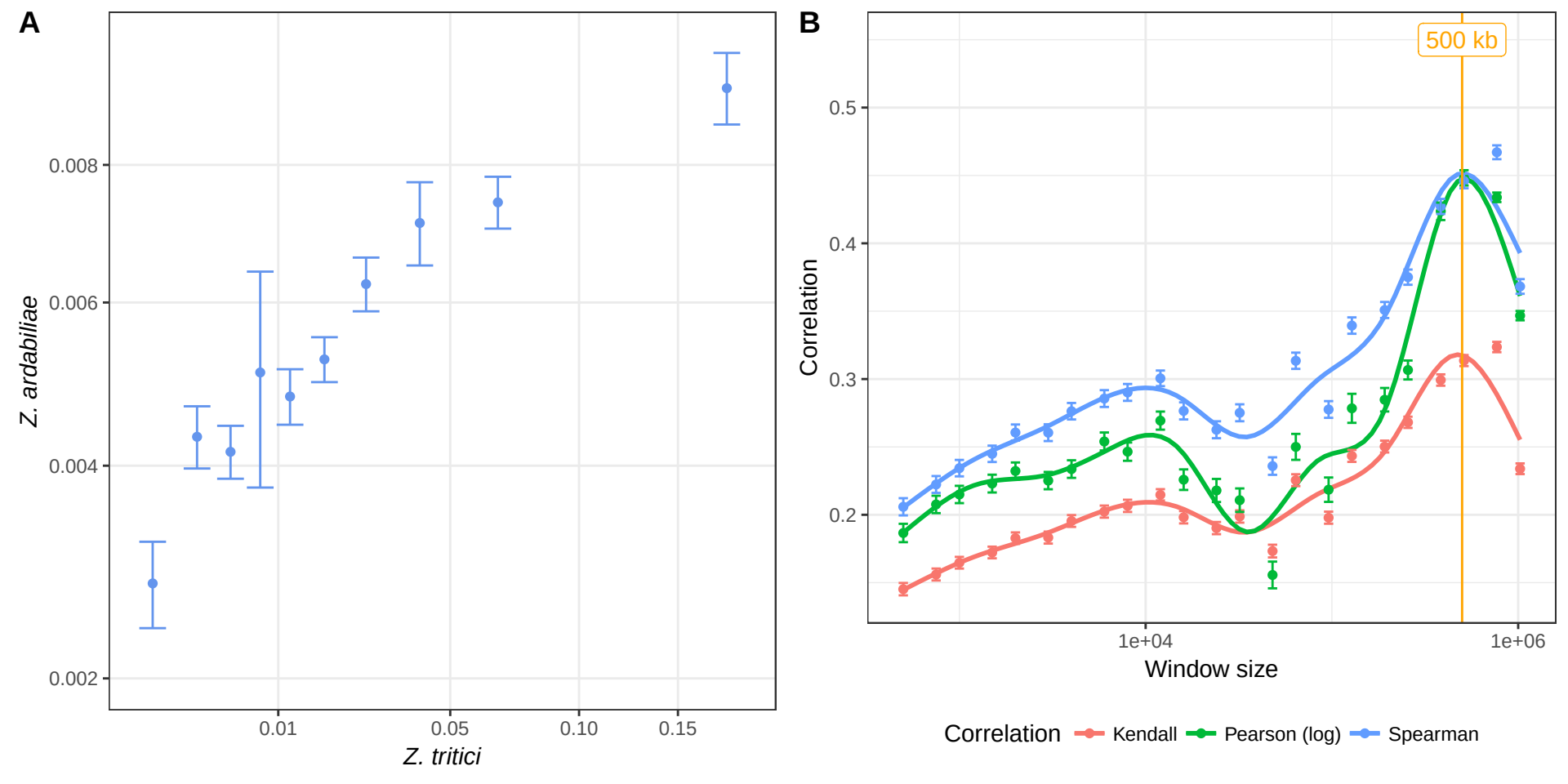

5e−05
5e−04
0.005
0.05
10
30
100
1e+00
1e−02
1e−04
Inferred ρ
0.001
0.100
Simulated ρ

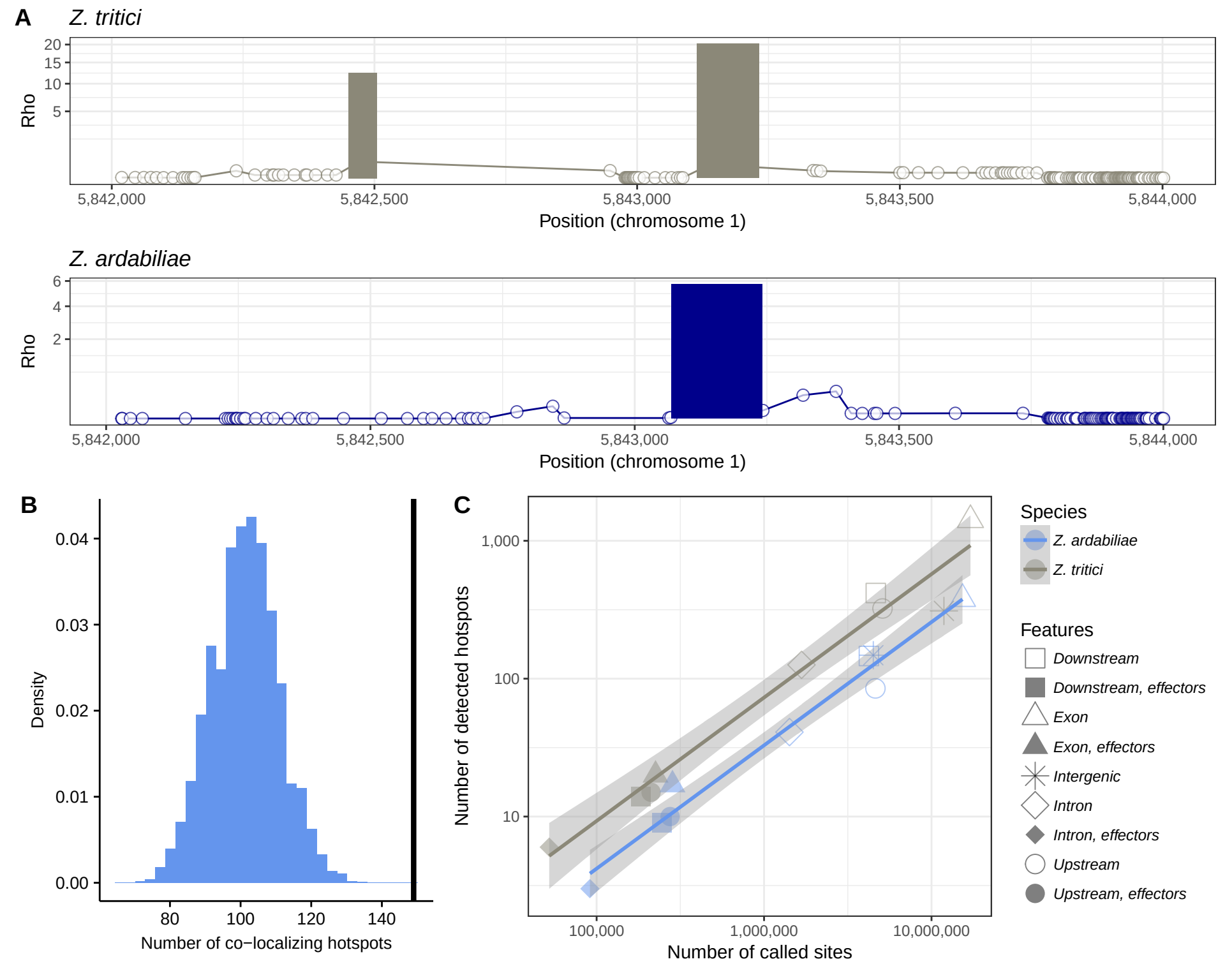

Chromosome 1
SNP density
0.15
0.10
0.05
0.00
Call
Rho
0.30
0.20
0.10
0.00
GC
0.60
0.55
0.50
0.45
0.40
0e+00
2e+06
4e+06
6e+06
Position


A
Species
Za
Zt
Mean rho
0.06
0.04
0.02
0.00
1
2
3
4
5
6
Chromosome length (Mb)
B
Species
Za
Zt
Mean rho
0.06
0.04
0.02
0.00
0
1
2
3
4
Chromosome arm length (Mb)
C
Species
Zt
MeanRho
0.10
0.05
0.00
Essential
Dispensable
Type


A
Species
Z. ardabiliae
Z. tritici
8
3
4
1
7
2
5
2
6
1
1e+00
1e−02
1e−04
Population recombination rate
Intron
Intergenic
Upstream
Downstream
Exon
Regions
B
Exon
Intron
Upstream
Downstream
***
***
NS
NS
NS
NS
NS
.
Z. ardabiliae
Z. tritici
Effector
Non−effector


Chromosome 1
Rho
0.04
0.02
0.00
Normalized cumulative rho
6e+06
4e+06
2e+06
0e+00
0e+00
2e+06
4e+06
6e+06
Position
Species
Z. ardabiliae
Z. tritici


A
0.008
0.006
0.004
0.002
Z. ardabiliae
0.01
0.05
0.10
0.15
Z. tritici
B
500 kb
0.5
0.4
0.3
0.2
Correlation
1e+04
1e+06
Window size
Correlation
Kendall
Pearson (log)
Spearman


A
Z. tritici
Rho
20
15
10
5
5,842,000
5,842,500
5,843,000
5,843,500
5,844,000
Position (chromosome 1)
Z. ardabiliae
Rho
6
4
2
5,842,000
5,842,500
5,843,000
5,843,500
5,844,000
Position (chromosome 1)
B
Density
0.04
0.03
0.02
0.01
0.00
80
100
120
140
Number of co−localizing hotspots
C
Number of detected hotspots
1,000
100
10
100,000
1,000,000
10,000,000
Number of called sites
Species
Z. ardabiliae
Z. tritici
Features
Downstream
Downstream, effectors
Exon
Exon, effectors
Intergenic
Intron
Intron, effectors
Upstream
Upstream, effectors