The impact of recombination hotspots on genome evolution of a fungal plant pathogen

Daniel Croll*,§, Mark H. Lendenmann*, Ethan Stewart*, Bruce A. McDonald*

Affiliations
* Plant Pathology, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland
§ Michael Smith Laboratories, University of British Columbia, V6T 1Z4 Vancouver BC,

ACCESSION NUMBERS
NCBI BioSample accessions for parental strains are SRS383146, SRS383147, SRS383142 and SRS383143. Progeny sequence data is deposited under NCBI BioProject accession numbers PRJNA256988 (cross ST99CH3D1 x ST99CH3D7) and PRJNA256991 (cross ST99CH1A5 x ST99CH1E4).

SHORT TITLE
Recombination hotspot variability

Copyright 2015.
Corresponding author:
Daniel Croll
LFW B28
Universitätsstrasse 2
CH-8092 Zürich
Switzerland
E-mail: daniel.croll@usys.ethz.ch
Telephone: +41 44 632 33 87; Fax +41 44 632 15 72

Keywords: Recombination hotspots, pathogen evolution, restriction site associated DNA sequencing, population genomics, linkage disequilibrium
Abstract

Recombination impacts genome evolution by maintaining chromosomal integrity, affecting the efficacy of selection and increasing genetic variability in populations. Recombination rates are a key determinant of the co-evolutionary dynamics between hosts and their pathogens. Historic recombination events created devastating new pathogens, but the impact of ongoing recombination in sexual pathogens is poorly understood. Many fungal pathogens of plants undergo regular sexual cycles and sex is considered a major factor contributing to virulence. We generated a recombination map at kilobase-scale resolution for the haploid plant pathogenic fungus *Zymoseptoria tritici*. In order to account for intra-specific variation in recombination rates, we constructed genetic maps from two independent crosses. We localized a total of 10,287 crossover events in 441 progeny and found that recombination rates were highly heterogeneous within and among chromosomes. Recombination rates on large chromosomes were inversely correlated with chromosome length. Short, accessory chromosomes often lacked evidence for crossovers between parental chromosomes. Recombination was concentrated in narrow hotspots that were preferentially located close to telomeres. Hotspots were only partially conserved between the two crosses, suggesting that hotspots are short-lived and may vary according to genomic background. Genes located in hotspot regions were enriched in genes encoding secreted proteins. Population resequencing showed that chromosomal regions with high recombination rates were strongly correlated with regions of low linkage disequilibrium. Hence, genes in pathogen recombination hotspots are likely to evolve faster in natural populations and may represent a greater threat to the host.
Introduction

Recombination is a fundamental process shaping the evolution of genomes. Crossover between homologous chromosomes ensures proper segregation during meiosis (MATHER 1938; BAKER et al. 1976; HASSOLD and HUNT 2001) and the integrity of chromosomal structure over evolutionary time is affected by the frequency of recombination. Sex chromosomes in plants and animals and mating type regions of fungi experience degeneration, including significant gene loss and sequence rearrangements, after cessation of recombination between homologs (BULL 1983; CHARLESWORTH and CHARLESWORTH 2000; BROWN et al. 2005; MENKIS et al. 2008; WILSON and MAKOVA 2009). Recombination breaks up linkage between alleles of different loci and creates novel haplotypes and phenotypic diversity. Through this process, recombination promotes adaptation because selection acting across multiple loci becomes more efficient as linkage between loci decreases (HILL and ROBERTSON 1966; OTTO and BARTON 1997; OTTO and LENORMAND 2002).

Recombination also plays a major role in the co-evolutionary dynamics of hosts and their pathogens. A major driver to maintain sexual reproduction in hosts was proposed to be the constant selection to evade disease caused by pathogens (HAMILTON 1980; LIVELY 2010; MORRAN et al. 2011). However, pathogens are also under strong selection to overcome newly evolved host resistance mechanisms. Recombination in pathogens plays a central role in disease outbreaks caused by viruses, bacteria and eukaryotic microorganisms. Epidemic influenza is driven by annual re-occurring outbreaks of recombined viral strains (NELSON and HOLMES 2007). The highly virulent lineage of Salmonella enterica causing typhoid fever (DIDELOT et al. 2007; HOLT et al. 2008) emerged from recombination with the related Paratyphi A lineage. Recombination between ancestral lineages of Toxoplasma created hypervirulent clones of the protozoan human pathogen T. gondii (GRIGG et al. 2001) and ongoing outcrossing and clonal expansion contributes to extant Toxoplasma
outbreaks (WENDTE et al. 2010). Some of the most threatening fungal pathogen lineages were generated by historic recombination events between lineages. Reshuffling between different genotypes may have contributed to the emergence of the hypervirulent human pathogen Cryptococcus gattii in the Pacific Northwest (FRASER et al. 2005; BYRNEs et al. 2011), where same-sex mating was suggested to have enabled the creation of a recombinant hypervirulent genotype (FRASER et al. 2005). The chytrid fungus Batrachochytrium dendrobatidis is causing a global epidemic on a wide range of amphibians (FISHER et al. 2009). The pathogen is comprised of a complex set of lineages. However, the outbreak is overwhelmingly caused by a single hypervirulent lineage that was created by hybridization and subsequent recombination of two distinct, allopatric Batrachochytrium lineages (ROSENBLUM et al. 2008; FARRER et al. 2011).

In contrast to the evidence that historic recombination events led to the creation of novel fungal pathogens, the impact of frequent recombination on genome evolution in extant sexual pathogens is poorly understood (AWADALLA 2003). As the extent of recombination correlates with the evolutionary potential of a population, the occurrence of sexual reproduction was suggested to be a predictor of a pathogen's capacity to overcome host resistance mechanisms (McDONALD and LINDE 2002; STUKENBROCK and McDONALD 2008). Pathogenic fungi of plants span a broad range of lifestyles from asexual to mainly sexual or mixed reproduction and can include multiple sexual cycles per year (McDONALD and LINDE 2002). Hence, these fungi are ideal models to study the impact of recombination on pathogen evolution. Loci contributing to virulence in pathogens undergo rapid allele frequency changes in response to changes in host populations (BARRETT et al. 2009; THRALL et al. 2012) and fungal pathogenicity is often found to be encoded by a large number of virulence loci distributed throughout the genome (KARASOV et al. 2014).
Recombination enables the emergence of pathogen strains carrying novel combinations of virulence alleles that may increase virulence on specific hosts.

A key determinant of the speed of adaptation of a pathogenic organism may be intragenomic variation in recombination rates that locally increase the efficacy of selection. Strong signatures of selection at linked sites were found in nematodes and humans, however evidence for such signatures was weak or absent in plants and the yeast *S. cerevisiae* (Cutter and Payseur 2013). Nevertheless, recombination rates in a genome can vary dramatically. In some model organisms, a majority of crossovers occurring during meiosis are focused in narrow chromosomal regions termed recombination hotspots (Kong et al. 2002; Jensen-Seaman et al. 2004; Myers et al. 2005; Arnheim et al. 2007).

Hotspots can have a profound effect on the recombination landscape of a genome. In the human genome, hotspots with a mean width of 2.3 kb accounted for 70-80% of all recombination events (1000 Genomes Project Consortium et al. 2010). Hotspots were preferentially located near telomeres in humans, plants and the yeast *S. cerevisiae* (Myers et al. 2005; Drovaud 2005; Tsai et al. 2010). Recombination hotspots are thought to be inherently unstable due to a "hotspot drive" mechanism that continuously replaces alleles favoring higher recombination rates (Webster and Hurst 2012). Genomic signatures of elevated recombination rates comprised increased GC content, likely resulting from the preferential incorporation of C and G during the mismatch repair process initiated by the pairing of homologous sequences (Duret and Galtier 2009).

We aimed to identify recombination rate heterogeneity and its impact on the genome of a pathogenic fungus. *Zymoseptoria tritici* is a globally occurring pathogen of wheat causing severe yield losses (O'Driscoll et al. 2014). The fungus undergoes sexual reproduction at least once per wheat growing season (Kema et al. 1996; Cowger et al. 2002; 2008).
Recombination between isolates was suggested to facilitate the evolution of virulence on previously resistant wheat varieties (Zhan et al. 2007). The genetic basis of virulence was shown to have a strong quantitative component (Zhan et al. 2005). Z. tritici populations undergo rapid allele frequency shifts in response to the application of fungicides (Torriani et al. 2009). The genome is haploid and was assembled into 21 complete chromosomes including telomeres (Goodwin et al. 2011). Eight chromosomes were termed accessory because these chromosomes were not present in all isolates. Meiosis frequently contributed to the loss, fusion or rearrangement of these accessory chromosomes (Wittenberg et al. 2009; Croll et al. 2013).

We generated a high-density SNP-based recombination map of two large progeny populations of Z. tritici. First, we used the map to precisely locate recombination hotspots, identified correlations with genome characteristics and tested whether hotspots were conserved between crosses. We then compared recombination rates between the small, accessory and large, core chromosomes in Z. tritici to determine whether recombination differently affects the two chromosomal classes. Next, we investigated whether hotspots were enriched in specific sequence motifs and whether hotspots were more likely to contain specific functional categories of genes. Finally, we tested whether recombination hotspots identified in the mapping populations predicted local breakdowns in linkage disequilibrium in a resequenced natural pathogen population.
Results

High-throughput progeny genotyping and crossover detection

Four parental isolates (3D1, 3D7, 1A5 and 1E4) collected from Swiss wheat fields in 1999 were used to construct two independent genetic maps. Progeny were genotyped based on restriction-associated DNA sequencing (RADseq) (Etter et al. 2011) which consists of Illumina sequencing DNA fragments immediately adjacent to a defined restriction site in the genome (generated by PstI in this study). Illumina reads obtained from each progeny were aligned to the reference genome of Z. tritici for SNP genotyping and filtering. The reference genome IPO323 is completely assembled including telomeres and centromeres (Goodwin et al. 2011). Whole-genome re-sequencing data from the parental isolates (Torriani et al. 2011) was used to validate segregating SNPs among progeny.

Recombination between homologs can result in either a crossover or a non-crossover. Crossovers result in the exchange of alleles over long chromosomal intervals. Non-crossovers lead to short intervals of unidirectional transfer among homologs (i.e. gene conversion). Distinguishing between non-crossover and crossover events is important as only the latter increase genetic diversity and break up haplotypes. In the model fungus S. cerevisiae, tracts of gene conversions resulting from non-crossover span a median of 1.8 kb and a maximum observed tract length of 40.8 kb (Mancera et al. 2008). Hence, we investigated the interval length between recombination events for all progeny and chromosomes. We observed a bimodal distribution of recombination event interval lengths (Figure 1). In both crosses, we found two modes corresponding to recombination event interval distances of 3-4 kb and ~1000 kb, respectively.
Genetic map construction and recombination rate variation among chromosomes

In order to conservatively estimate genetic map lengths based on crossover events, we required a minimum distance of 50 kb between recombination events. Genetic maps were constructed using the marker order known from physical positions of the SNP markers based on the reference genome sequence. The consistency of the physical and genetic marker distances and marker order was evaluated. A single problematic region was found on chromosome 13, so this chromosome was excluded from analyses requiring physical positions on the chromosome (see Methods and Figure S1).

The total genetic map constructed for cross 3D1x3D7 was based on 227 progeny and spanned 2723.6 cM. For cross 1A5x1E4, the total map was based on 214 progeny and was slightly smaller (2158.9 cM). The number of genotyped SNP markers greatly exceeded (by ∼12-15 X) the number of genetic map positions (the minimal set of markers separated by at least one crossover). This shows that the SNP markers largely saturate the number of genetic map positions in the two crosses (Tables 1 and 2, Figure 2).

Recombination rates were inversely related to the length of the chromosome (Figure 2). The two crosses contained different sets of accessory chromosomes (i.e. chromosomes not present in all isolates of the species). In cross 3D1x3D7, both parents carried a copy of accessory chromosomes 16, 17, 19 and 20. In cross 1A5x1E4, accessory chromosomes 14-16 and 18-21 were present in both parents. Accessory chromosome 17 in the cross 1A5x1E4 was found to undergo spontaneous chromosomal fusion (CROLL et al. 2013), however no other chromosomal rearrangements were documented in the two crosses. Accessory chromosomes are significantly shorter than core chromosomes and this is reflected in genetic map lengths. Chromosomes 14 (33.2 cM), 18 (22.1 cM) and 21 (16.5 cM) showed the shortest genetic maps (Table 2, Figure 2). Chromosomes 18 and 21 were
also found to have the lowest recombination rates of any chromosome (38.5 and 40.4 cM/Mb).

Infrequent crossovers on some accessory chromosomes

We assessed the minimal number of crossover events that generated each progeny genotype. Average crossover counts decreased as expected with chromosome length (Figure S2; Table 3). Crossover counts for a given core chromosomes followed a Poisson-like distribution and for each core chromosome we identified progeny with zero crossovers. We found that accessory chromosomes 16, 17, 19 and 20 of cross 3D1x3D7 had crossover counts of 0.45 - 0.54 per progeny. Average crossover counts were lower in cross 1A5x1E4. The lowest average crossover counts were 0.15 and 0.2 for chromosomes 21 and 18, respectively. The low crossover counts per accessory chromosome resulted in a large proportion of progeny that did not show a single crossover event (Figure S2). Most crossover events on accessory chromosomes occurred in narrow segments of the chromosome as shown by large linked chromosomal segments among progeny (Figure S3). Chromosome 14 was found to recombine only in two narrow subtelomeric chromosomal sections.

Recombination rate heterogeneity on chromosomes

In order to identify variation in recombination rates, we estimated recombination rates in non-overlapping 20 kb segments along chromosomes. Recombination rates in these segments were highly heterogeneous and varied between 0 and 1135 cM / Mb (Figure 3). Central tracts of the chromosomes that were devoid of any recombination signal may indicate centromere locations. Recombination rates estimated per 20 kb segments were significantly correlated between the two crosses (Pearson’s product moment correlation coefficient $r = 0.61; p$-value < 2.2e-16). We tested whether recombination rate
heterogeneity along chromosomes was significantly different than expected from a random
distribution of crossovers. For this, we tested whether the observed crossover counts per
chromosomal segment deviated significantly from a predicted random distribution of
crossovers. We found that crossovers were non-randomly distributed in both crosses and
at two different chromosomal scales (counting crossover in 10 kb and 200 kb segments;
detailed statistics are reported in Figure S4). In addition, we found that crossovers were
non-randomly distributed on all core chromosomes when analyzed individually. On
accessory chromosomes for which statistical analyses could be performed, crossovers
were also non-randomly distributed (see Table S1 for details on the statistical tests).
In order to identify systematic variations in recombination rates depending on
chromosomal position, we summarized recombination rates on core chromosomes
according to the relative distance to the closest telomere (Figure 4). Visual trends suggest
that median recombination rates were highest in subtelomeric regions of chromosomes in
both crosses (5-15% distance to the telomere). The lowest median recombination rates
were found closest to the telomeres (within 5% distance). However, individual core
chromosomes differed significantly for the distribution of recombination rates in relation to
telomere distance (Figure S5).

Localization of recombination hotspots

We aimed to identify narrow sequence tracts with the highest rates of exchange. For this,
we counted the observed number of crossover events in non-overlapping 10 kb sequence
segments in both crosses. The average recombination rate per segment was 1.1 and 1.46
crossover events in the crosses 1A5x1E4 and 3D1x3D7, respectively. In cross 1A5x1E4
we found that 76 segments showed 10 or more crossovers, representing 1.9% of the
genome but 27.8% of all crossovers (Table S2). Similarly, in cross 3D1x3D7, 130 hotspot
segments were found representing 3.3% of the genome and 38.2% of all crossovers
The probability to observe 10 or more crossovers by chance was $p$-value < $4.3 \times 10^{-7}$ (Poisson distribution with lambda = 1.1 and 1.46, respectively for 1A5x1E4 and 3D1x3D7). Hence, these hotspot segments represent highly unusual regions in the genome. The rate of exchange in the hotspots ranged from 4.4% – 17.6% of all progeny showing evidence for a crossover event. Hotspot locations were shared between crosses 1A5x1E4 and 3D1x3D7 in 35 segments. The overlap between hotspots corresponds to 46.1% of all hotspots identified in 1A5x1E4 and to 26.9% of all hotspots identified in 3D1x3D7.

**Association of recombination hotspots with genome characteristics**

The *Z. tritici* genome is gene dense (31.5% coding sequences) and GC-rich (51.7%). Chromosomal sequences contain long sections of high GC-content interspersed with short sections of low GC-content and low gene density (GOODWIN et al. 2011). In order to test for genome characteristics associated with the presence of hotspots, we divided the entire genome into non-overlapping 10 kb sequence segments and assessed differences between 10 kb sequences identified as hotspots (see above) versus all other 10 kb segments. GC-content was slightly but significantly higher in hotspots versus the genomic background (53.6% versus 52.2%, Student’s t test $t = 3.42$ and 5.17 and $p$-value = 0.0009 and 6.2e-7 for the cross 1A5x1E4 and 3D1x3D7, respectively). Coding sequence densities were not significantly different between the hotspots and the genomic background (Student’s t test, $p$-value ~ 0.8; Figure 5A). However, hotspot segments were significantly more polymorphic between parental genomes than the genomic background. The mean SNP density was 1.37% and 1.60% for the cross 1A5x1E4 and 3D1x3D7, respectively, versus the genomic background diversity of 1.01% (Student’s t test $t = 4.50$ and 5.45 and $p$-value = 1.41e-5 and 6.27e-7 for the cross 1A5x1E4 and 3D1x3D7, respectively; Figure 5B). Relationships between recombination rates and GC-content, gene density and
polymorphism between parental genomes, respectively, are shown in Figures S6, S7 and S8.

Recombination hotspots are enriched with GC-rich sequence motifs
Recombination hotspots were found to be enriched in short sequence motifs in animals and protozoa (ARNHEIM et al. 2007; MYERS et al. 2008; JIANG et al. 2011). Evidence for motifs from fungi is species-dependent. A simple sequence motif predicted recombination hotspots in the fission yeast Schizosaccharomyces pombe (STEINER and SMITH 2005), however no association was found for S. cerevisiae (MANCERA et al. 2008). In order to identify potential motifs associated with recombination hotspots in Z. tritici, we analyzed 10 kb sequence segments showing 10 or more crossover events as described above using HOMER. We restricted the search to any possible 8, 10 and 12 bp motifs in order to avoid confounding effects of overlapping motifs. In the cross 3D1x3D7 the 10 bp (GGGGGGGATG) GC-rich motif was found in 53.9% of the hotspots compared to 22.4% of sequence segments not containing hotspots. After accounting for differences in GC-content between hotspots and the genomic background, we found that hotspots were significantly enriched in this motif ($p$-value = 1e-14; Figure 6). Similarly, three GC-rich motifs were found to be significantly enriched in hotspots of the cross 1A5x1E4. The most significant motif was 12 bp (GTGGCTGGGGA) and found in 21.1% of all hotspot sequences compared to 1.6% in the genomic background ($p$-value = 1e-13; Figure 6).

Recombination hotspots are enriched in genes encoding secreted proteins
Genes located in recombination hotspots may be subject to both increased mutation rates and higher rates of reshuffling of allelic variants in the population. Hence, the mutagenic potential and higher rates of adaptive evolution of recombination hotspots may lead to an enrichment in fast-evolving gene categories near hotspots. As above, we defined recombination hotspots as 10 kb segments showing 10 or more crossover events. Protein
secretion is an important component of virulence in plant pathogenic fungi. We tested whether recombination hotspots were more likely to contain genes encoding secreted proteins. We used the dataset of predicted secretory signal peptides to test for enrichment in hotspots. We found that in cross 3D7x3D1, hotspots were significantly enriched for genes encoding secreted proteins (hypergeometric test; \( p\)-value = 0.0076). In cross 1A5x1E4, hotspots were also enriched in genes encoding secreted proteins (\( p\)-value = 0.021).

We screened genes overlapping recombination hotspots for previously identified virulence candidate genes (GOODWIN et al. 2011; MORAIS DO AMARAL et al. 2012; BRUNNER et al. 2013). We identified several categories of genes underlying important components of virulence in plant pathogenic fungi. Three small secreted proteins were found in hotspots overlapping in both crosses (Table 4). In cross 3D1x3D7 hotspots, we identified genes encoding cell wall degrading enzymes: the cutinase MgCUT4, the beta-xylosidase MgXYL4 and an alpha-L-arabinofuranosidase.

**Recombination rates are correlated with population linkage disequilibrium**

The extent of linkage disequilibrium (LD) among markers is a defining characteristic of genetic variation in a population. High LD is indicative of partial selective sweeps or population substructure. Recombination rates are expected to influence the extent of LD in populations because high recombination rates may reduce LD among markers. We tested whether recombination rates were correlated with LD in a natural *Z. tritici* population. For this, we resequenced 25 isolates from the same field population (including the parental isolates in this study). All isolates were genotyped at the same marker positions that were used in the recombination rate analyses. We calculated both the LD in the population and the recombination rate between all adjacent markers that were separated by at least 500
We found that the recombination rate was negatively correlated with LD in both crosses (Spearman’s rank correlation $p < 2.2e-16$ and $\rho = -0.213$ and $-0.184$ for $3D1x3D7$ and $1A5x1E4$, respectively; Figure 7). Adjacent markers with the highest LD in the natural field population were among the markers with the lowest recombination rates in the crosses.
We established a dense map of recombination rates in the genome of a fungal pathogen based on two independent crosses. The SNP marker density was ~12-15 times greater than the number of unique marker positions in the genetic map, hence crossover events could be precisely localized to narrow chromosomal tracts. As genomic analyses of all four meiotic products are currently not feasible in Z. tritici, we used a conservative approach to avoid confounding effects due to non-crossover events. By requiring a minimum distance of 50 kb between recombination events in a progeny, a small number of double crossovers occurring in a narrow interval may have been missed.

Recombination rates were highly heterogeneous and highest in subtelomeric recombination hotspots. The genetic maps produced for the two crosses were substantially larger than previously published maps for Z. tritici (2158.9 - 2723.6 cM versus 1216 - 1946 cM) (KEMA et al. 2002; WITTMERG et al. 2009). The larger map lengths were likely due to denser marker coverage in subtelomeric regions and larger progeny populations. Furthermore, in contrast to the present study, no recombination was previously detected on accessory chromosomes (WITTMERG et al. 2009). The new genetic map of Z. tritici was also substantially longer than the map of the model fungi Coprinus cinereus (948 cM; (STAIJICH et al. 2010)) and Neurospora crassa (1000 cM; (LAMBRECHTS et al. 2009)). The genetic map of Saccharomyces cerevisiae was estimated to be 4900 cM (BARTON et al. 2008). The genetic maps differed between the two analyzed crosses by 565 cM. The genetic distance between the parental genomes, number of analyzed progeny, number of genotyped markers and genotyping rates differed only marginally between the two crosses. Hence, these parameters are unlikely to explain differences in estimated genetic map lengths. However, analyzing physical distances between recombination events showed that the cross 1A5x1E4 had a lower proportion of
long distance (> 50 kb) recombination events compared to cross 3D1x3D7. Unknown recombination modifiers could be responsible for differences in crossover rates between crosses.

**Low recombination rates and non-canonical meiosis on accessory chromosomes**

We found that the frequency of reciprocal recombination (or crossover) per chromosome was positively correlated with chromosomal length. However, the recombination rate expressed per Mb was higher for smaller core chromosomes. The inverse relationship between recombination rate and chromosomal length is well established in *S. cerevisiae* (Mortimer et al. 1989; Cherry et al. 1997; Kaback et al. 1999). Crossover counts per chromosome reached 0.94-1.14 for the smallest core chromosomes. The occurrence of at least one crossover event per chromosome and meiosis promotes proper homologous segregation (Mather 1938; Hassold and Hunt 2001). Non-disjunction of chromosomes leads to aneuploidy (Baker et al. 1976). We found that some *Z. tritici* accessory chromosomes had unusually low crossover rates, as approximately half of the progeny of cross 3D1x3D7 showed no crossovers and in cross 1A5x1E4 per progeny crossover counts ranged from 0.15 to 0.45 per chromosome. Failure to undergo crossovers may contribute to frequent non-disjunction of particular accessory chromosomes (Wittenberg et al. 2009; Croll et al. 2013). Accessory chromosomes were frequently found to be lost or disomic following meiosis, indicative of non-disjunction (Wittenberg et al. 2009; Croll et al. 2013).

Lower than expected recombination rates (in cM per Mb) on some accessory chromosomes are expected to lead to substantial linkage disequilibrium among genetic variants in a population. The purging of deleterious mutations critically depends on sufficient recombination between loci. Therefore, some accessory chromosomes may be
subject to more rapid mutation accumulation and decay. Reduced levels of gene density and shorter transcript lengths (GOODWIN et al. 2011) may be indicative of a degeneration process operating on accessory chromosomes. Degeneration and rearrangements in turn may lead to further recombination suppression; a process similar to the one observed in sex chromosome evolution.

Recombination rate heterogeneity and hotspots

Recombination rates were highly heterogeneous along chromosomes. A large fraction of all crossover events were restricted to narrow chromosomal bands or hotspots of recombination. In the strongest hotspots, 18.5% of all progeny showed a crossover event within a 10 kb region. Recombination rates tended to be highest in subtelomeric regions of the chromosomes. Increased recombination rates in the chromosomal periphery were widely reported in animals (e.g. BACKSTRÖM et al. 2010; BRADLEY et al. 2011; AUTON et al. 2012; ROESTI et al. 2013), and plants (AKHUNOV et al. 2003; ANDERSON et al. 2003). In fungi, hotspot associations with subtelomeric regions were found in C. cinereus and S. cerevisiae (BARTON et al. 2008; STAJICH et al. 2010). Peripheral clustering of chromosomes (bouquet formation) during prophase I of meiosis is thought to favor proper homologous pairing and by this process promote crossovers (BROWN et al. 2005; NARANJO and CORREDOR 2008).

Recombination hotspots showed no strong difference in GC content or coding sequence density compared to the genomic background. However, associations of GC content and recombination rates are ubiquitous in eukaryotes (GERTON et al. 2000; JENSEN-SEAMAN et al. 2004; DURET and ARNDT 2008; BACKSTRÖM et al. 2010; MYULE et al. 2011; AUTON et al. 2012). In S. cerevisiae, experimentally introduced GC-rich regions generate novel recombination hotspots (WU and LICHTEN 1995). Hotspots of Z. tritici were enriched in G/C
mononucleotide repeats compared to the genomic background. Similarly, *S. cerevisiae* hotspots were found to be enriched in 20-41 bp poly(A) stretches (Mancera et al. 2008). *Z. tritici* hotspots were in regions of higher diversity between the parental genomes compared to the genomic background. A potential explanation is that if hotspot locations persist over many generations, frequent gene conversion events may cause a higher mutation rate in hotspots and, hence, higher sequence diversity (Hurles 2005). Reduced diversity in regions of low recombination rates is well supported both by theory and empirical data (Begun and Aquadro 1992; Nachman 2002; Cutter and Payseur 2013).

**Divergent hotspot locations within a species**

Hotspots are thought to be ephemeral in genomes because of their inherent self-destructive properties. A crossover is initiated by a double-strand break on one chromosome. The homologous chromosome serves as a template to repair the crossover initiator sequence. Therefore, the conversion process continually replaces any allele that favors increased recombination with its homolog. Studies of humans and chimpanzees showed that hotspot locations are rarely conserved, consistent with rapid turnover (Coop and Myers 2007; Auton et al. 2012). In contrast, hotspots were frequently conserved among a pair of divergent *Saccharomyces* species (Tsai et al. 2010). The unexpected conservation was attributed to the low frequency of sex, reducing the opportunity for biased gene conversion to erode hotspots. *Z. tritici* undergoes sexual reproduction very frequently (Zhan et al. 2002). Hotspot locations were only partially conserved (27 - 46%) between the two crosses established from isolates of the same regional population. A lack of overlap in hotspot locations may be due to the fact that not all hotspot locations were used during specific rounds of meiosis. However, each of the two analyzed crosses contained a large number of individual meiosis events reflecting the large number of
analyzed ascospore progeny. Hence, stochastic variations in hotspot occupancies should be minimal.

**Functional enrichment of genes in recombination hotspots**

Plant pathogenic fungi rely on secretion to deliver effectors and toxins into the host and genes encoding short secreted polypeptides were found to evolve very rapidly in pathogen genomes (RAFFAELE et al. 2010; PETRE and KAMOUN 2014). Recombination hotspots in *Z. tritici* were enriched in genes encoding secreted proteins. This may indicate that the localization of these genes in hotspots was favored by selection. We also screened hotspots in both crosses for the presence of candidate effector genes likely to play a role in pathogenesis (Table 4). Hotspots overlapping in both crosses contained genes encoding small, cysteine-rich secreted proteins. This gene category was found to be up-regulated during host infection and gene products likely modulate the host resistance response (MORAIS DO AMARAL et al. 2012; MIRZADI GOHARI et al. 2014). Two cell wall degrading enzymes (MgCUT4 and MgXYL4) were specifically up-regulated during host infection and showed evidence for selection in the pathogen population (BRUNNER et al. 2013). However, MgCUT4 and MgXYL4 were located in hotspots exclusive to the 3D1x3D7 cross, suggesting that associations of virulence genes and hotspots may be ephemeral over evolutionary time.

**Consequences of recombination hotspots for genome evolution of pathogens**

Regions of the genome with high recombination rates showed significantly lower levels of linkage disequilibrium in the natural population. This suggests that effects of demography, periods of clonal reproduction and selection were insufficient to affect the inverse relationship between recombination rates and linkage disequilibrium in the population. An emerging paradigm in the study of pathogenic organisms is that the genome organization
reflects the intense selection pressure faced by pathogens (CROLL and MCDONALD 2012; RAFFAELE and KAMOUN 2012). The observed genome compartmentalization was termed the "two-speed" genome (RAFFAELE et al. 2010). The genome of the causative agent of the Irish potato famine, the oomycete Phytophthora infestans, contains fast-evolving, repeat-rich compartments hosting an abundance of pathogenicity genes (RAFFAELE et al. 2010). In contrast, genes encoding essential functions are preferentially located in GC-rich, gene-dense compartments. However, the enrichment of effector genes in repeat-rich regions is not universal among pathogens. Many pathogen genomes are neither repeat-rich nor contain large tracts of low GC-content. Our findings in Z. tritici suggest that genome organization can evolve to compartmentalize gene functions correlated with variations in recombination rates. Recombination hotspots may constitute a previously unrecognized genomic compartment that favors the emergence of fast-evolving virulence genes in pathogens.
Materials and Methods

Establishment of controlled sexual crosses

We performed two sexual crosses between four different *Z. tritici* isolates collected from two wheat fields separated by ~10 km in Switzerland. All isolates were previously characterized phenotypically and genetically (ZHAN et al. 2005; CROLL et al. 2013). The two crosses were performed between the isolates ST99CH1A5 and ST99CH1E4 (abbreviated 1A5 and 1E4) and ST99CH3D1 and ST99CH3D7 (abbreviated 3D1 and 3D7), respectively. The parental isolates were used to co-infect wheat leaves according to a protocol established for *Z. tritici* (Kema et al. 1996). Shooting ascospores were collected, grown in vitro and genotyped by microsatellite markers to confirm that progeny genotypes were recombinants of the parental genotypes.

SNP genotyping based on restriction-associated DNA sequencing (RADseq)

We chose a high-throughput genotyping method originally developed for animal and plant population genomics (BAIRD et al. 2008). We adapted the RADseq protocol (ETTER et al. 2011) for *Z. tritici* by using *PstI* for the digestion of 1.3 µg of genomic DNA per progeny. After DNA digestion, shearing and adapter-ligation, we performed 100 bp Illumina paired-end sequencing on pooled samples. Progeny were identified by combining 24 distinct inline barcodes located in the P1 adapter and 6 index sequences in the P2 adapter.

Illumina data analysis and reference alignment

Raw Illumina reads were quality trimmed with Trimmomatic v. 0.30 (BOLGER et al. 2014) and split into unique sets for each progeny with the FASTX toolkit v. 0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). We used the short-read aligner bowtie version 2.1.0 (LANGMEAD and SALZBERG 2012) to align the read set of each progeny to the reference genome IPO323 (assembly version MG2, Sep 2008; (GOODWIN et al. 2011)). We
used the default settings for a sensitive end-to-end alignment (-D 15; -R 2; -L 22; -i S,1,1.15). The whole genomes of all four parental isolates were previously re-sequenced (TORRIANI et al. 2011) and available under the NCBI SRA accession numbers SRS383146 (3D1), SRS383147 (3D7), SRS383142 (1A5) and SRS383143 (1E4). We used identical trimming and assembly parameters to align the whole genome sequencing dataset to the reference genome IPO323. All Illumina sequence data generated for progeny re-sequencing is available under the NCBI BioProject accession numbers PRJNA256988 (cross 3D1x3D7) and PRJNA256991 (cross 1A5x1E4).

**Variant calling and filtration**

Single nucleotide polymorphisms (SNP) were identified based on the Genome Analysis Toolkit (GATK) version 2.6-4-g3e5ff60 (DEPRISTO et al. 2011). SNPs were called separately for each cross with the GATK UnifiedGenotyper. Each UnifiedGenotyper run combined all progeny and their two respective parents. For UnifiedGenotyper we set the sample-level ploidy to 1 (haploid) and permitted a maximum of 2 alternative alleles compared to the reference genome. The genotype likelihood model was set to the SNP general ploidy model.

We used the GATK filtering and variant selection tools to remove spurious SNP calls from the dataset. We set the following requirements for a SNP to pass the filtering: Overall quality score (QUAL ≥ 100), mapping quality (MQ ≥ 30), haplotype score (HaplotypeScore ≤ 13), quality by depth (QD ≥ 5), read position rank sum test (ReadPosRankSumTest ≥ -8) and Fisher's Exact Test for strand bias (FS ≤ 40). After a joint SNP locus filtering, we filtered genotypes for each parental isolate and progeny at each retained locus. We required that each included isolate had a minimum mean genotyping depth of 5 high-
quality reads. We retained individual genotypes if the phred-scaled genotype quality (GQ) assigned by GATK UnifiedGenotyper was at least 30.

**Genetic map construction and quality assessment**

We constructed a genotype matrix containing all progeny that were genotyped at a minimum of 90% of all SNPs. Subsequently, we removed all SNP markers that were genotyped in less than 90% of the progeny. We assessed the clonal fraction among the progeny in the offspring populations using the r/qtl package in R (ARENDS *et al.* 2010). If two progeny were identical at 90% or more of the SNPs, we randomly excluded one of the two likely clones. We inspected the quality of the progeny genotype set for problematic double crossovers at very closely spaced markers. For this, we calculated error LOD scores (LINCOLN and LANDER 1992) as implemented in the r/qtl package. The error LOD score compares likelihoods for a correct genotype versus an erroneous genotype based on pairwise linkage. We excluded genotypes exceeding an error LOD score of 2 from the dataset. We filtered the progeny genotype data for random genotyping errors. We required that any double crossover event in a progeny must be spanned by at least three consecutive SNPs. Furthermore, any double crossover event had to span at least 500 bp. These filtering steps ruled out that double crossover events could be erroneously produced by a single stack (i.e. restriction site) of RAD sequencing reads. In addition, we visually inspected the genotype grid for potential erroneous read mapping or translocation events indicated by switched genotypes in all progeny of a cross. We identified three potential erroneous locations in the cross of 3D1x3D7, with each spanning a maximum physical distance of 650 bp. In the 1A5x1E4 cross, we found three potential erroneous locations with SNP loci spanning a maximum of 50 bp for each region. SNP loci within an erroneous mapping location were excluded from further analyses.
In summary, we retained 23,563 SNPs in 227 progeny of the cross between parental isolates 3D1 and 3D7. We analyzed 214 progeny at 23,284 SNP markers for the cross between parental isolates 1A5 and 1E4. The SNP marker density was 0.6 SNP per kb. The genotyping rate averaged 96.5% (chromosomal averages 93.7 - 97.3%) for cross 3D1x3D7 and 95.0% (chromosomal averages 90.9 - 98.3%) for cross 1A5x1E4 (Table 1). SNP markers covered the entire length of all core chromosomes with the exception of telomeric regions. Marker densities were similarly distributed along chromosomes in both crosses. We calculated recombination fractions among all pairs of retained markers using the r/qtl package. We set the maximum of iterations to 10,000 and the tolerance for recombination fraction estimates to 0.0001. Using identical settings, we calculated genetic distances for marker pairs on each chromosome.

**Genetic map reconstruction of chromosome 13**

Based on pairwise recombination fractions, we identified one problematic region located on chromosome 13 (Figure S1). In both crosses, the pattern of recombination fractions suggested that the parental isolates differed from the reference genome used for read mapping. In order to generate the correct marker order on chromosome 13, we produced the genetic map using r/qtl. We used 10,000 iterations and a tolerance of 0.0001 for cM estimates between markers. The re-calculated marker order showed that the problematic region contains a large weakly recombining region (Figure S1). As the marker order in weakly recombining regions is difficult to confirm with a high confidence, we excluded chromosome 13 from analyses requiring exact physical locations of markers.
Correlations of recombination rates and genome characteristics

We downloaded the reference genome IPO323 (assembly version MG2, Sep 2008; (GOODWIN et al. 2011)) from http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html (accessed March 2014). Gene annotations of Z. tritici related to the RefSeq assembly ID GCF_000219625.1 were accessed in January 2014. SignalP annotations for gene models of IPO323 were retrieved from http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html (accessed March 2014). Enrichment for signalP secretion signals (probability > 0.9) was tested using a hypergeometric test in R.

Motif search in recombination hotspots

We searched recombination hotspots for enriched sequence motifs with HOMER v 4.3 (HEINZ et al. 2010). For this, we divided the chromosome sequences into non-overlapping 10 kb segments. For each cross, we identified segments showing at least 10 crossover events. We used the HOMER findMotifsGenome module to identify enriched oligo-nucleotides in the hotspot segments compared to the genomic background divided into equally long segments of 10 kb. We used GC-content auto-normalization to correct for bias in sequence composition. The motif length search was restricted to 8, 10 and 12 bp. The top scoring motif was then used to detect chromosomal regions containing the motif using the annotatePeaks tool included in Homer.

Population genomic analyses of linkage disequilibrium

We analyzed Illumina whole-genome sequence data from all four parental isolates and 21 additional isolates (data available under NCBI BioProject PRJNA178194) from the same Swiss population (TORRIANI et al. 2011). We used identical reference genome alignment and SNP quality filtering procedures as described above for the RADseq analyses. In order to correlate recombination rates and linkage disequilibria, we retained SNP positions for the population dataset if these positions were included in the genetic map construction.
Adjacent SNP marker pairs were omitted if markers were separated by less than 500 bp. All linkage disequilibria calculations were made with vcf tools v. 0.1.12a (DANECEK et al. 2011).

**Acknowledgements**

We are grateful to Christine Grossen, Sam Yeaman, Alan Brelsford and Jessica Purcell for helpful comments on previous versions of the manuscript. The Genetic Diversity Center and the Quantitative Genomics Facility at ETH Zurich were used to generate sequence data. Funding was provided by Swiss National Science Foundation grants to DC [PA00P3_145360], BAM [31003A_134755] and an ETH Zurich grant [ETH-03 12] to DC and BAM.
References


DIDELOT X., ACHTMAN M., PARKHILL J., THOMSON N. R., FALUSH D., 2007 A bimodal pattern of
relatedness between the *Salmonella* Paratyphi A and Typhi genomes: Convergence or divergence by homologous recombination. Genome Res **17**: 61–68.


GRIGG M. E., BONNEFOY S., HEHL A. B., SUZUKI Y., BOOTHROYD J. C., 2001  Success and virulence in *Toxoplasma* as the result of sexual recombination between two distinct ancestries. Science **294**: 161–165.


HOLT K. E., PARKHILL J., MAZZONI C. J., ROUMAGNAC P., WEILL F.-X., GOODHEAD I., RANCE R.,
BAKER S., MASKELL D. J., WAIN J., DOLECEK C., ACHTMAN M., DOUGAN G., 2008 High-
throughput sequencing provides insights into genome variation and evolution in Salmonella

HURLES M., 2005 How homologous recombination generates a mutable genome. Hum. Genomics
2: 179–186.

JENSEN-SEAMAN M. I., FUREY T. S., PAYSEUR B. A., LU Y., ROSKIN K. M., CHEN C.-F., THOMAS M.
A., HAUSSLER D., JACOB H. J., 2004 Comparative recombination rates in the rat, mouse, and

JIANG H., LI N., GOPALAN V., ZILVERSMIT M. M., VARMA S., NAGARAJAN V., LI J., MU J., HAYTON K.,

KABACK D. B., BARBER D., MAHON J., LAMB J., YOU J., 1999 Chromosome size-dependent control
of meiotic reciprocal recombination in Saccharomyces cerevisiae: the role of crossover

KARASOV T. L., HORTON M. W., BERGELSON J., 2014 Genomic variability as a driver of plant-

A. J., DE WEERDT M., BONANTS P. J. M., WAALWIJK C., 2002 A combined amplified fragment
length polymorphism and randomly amplified polymorphism DNA genetic kinkage map of
Mycosphaerella graminicola, the septoria triticí leaf blotch pathogen of wheat. Genetics 161:
1497–1505.

KEMA G. H. J., VERSTAPPEN E. C. P., TODOROVA M., WAALWIJK C., 1996 Successful crosses and
molecular tetrad and progeny analyses demonstrate heterothallism in Mycosphaerella

KONG A., GUDBJARTSSON D. F., SAINZ J., JONSDOTTIR G. M., GUDJONSSON S. A., RICHARDSSON B.,
SIGURDARDOTTIR S., BARNARD J., HALLBECK B., MASSON G., SHLIEN A., PALSSON S. T., FRIGGE
M. L., THORGEIRSSON T. E., GULCHER J. R., STEFANSSON K., 2002 A high-resolution
recombination map of the human genome. Nat Genet.

LAMBRECHTS R., SHI M., BELDEN W. J., DECAPRIO D., PARK D., HENN M. R., GALAGAN J. E.,
BAŞTÜRKMEN M., BIRREN B. W., Sachs M. S., Dunlap J. C., LOROS J. J., 2009 A high-density

357–359.

LINCOLN S. E., LANDER E. S., 1992 Systematic detection of errors in genetic linkage data.
Genomics 14: 604–610.

LIVELY C. M., 2010 A Review of Red Queen Models for the Persistence of Obligate Sexual

MANCERA E., BOURGON R., BROZZI A., HUBER W., STEINMETZ L. M., 2008 High-resolution mapping


MCDONALD B. A., LINDE C., 2002 Pathogen population genetics, evolutionary potential, and


Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus Mycosphaerella graminicola. Molecular Ecology 14: 2683–2693.

Sexual reproduction facilitates the adaptation of parasites to antagonistic host environments: Evidence from empirical study in the wheat-Mycosphaerella graminicola system. International Journal for Parasitology 37: 861–870.
**Figure legends**

**Figure 1:** Physical distance between all detected recombination events per progeny summarized for each of the two analyzed crosses (3D1x3D7 and 1A5x1E4). A short distance between recombination events in a progeny may indicate a potential noncrossover. In this case the two recombination events would be boundaries of a gene conversion tract. In order to conservatively identify true crossovers, a minimum distance of 50 kb was required for a recombination event to be considered a crossover (shaded in black) instead of a putative noncrossover event (shaded in grey). All analyses of recombination rates and hotspots are based on recombination events at least 50 kb from the nearest neighboring recombination event.

**Figure 2:** Summary statistics of the genetic maps constructed for the two crosses 3D1x3D7 and 1A5x1E4. A) Genetic map length per chromosome. Horizontal hatches show the positions of genetic markers in the genetic maps. In cross 3D1x3D7, four accessory chromosomes were shared between the parental isolates and could therefore be used for map construction. In cross 1A5x1E4, seven accessory chromosomes were shared between the parental isolates. B) Recombination rates per chromosome expressed as cM per Mb. Genetic maps for the two crosses are distinguished by color. See panel C for color legend. C) Number of genetic map positions per chromosome (minimal set of markers separated by at least one crossover).

**Figure 3:** Recombination landscape of the plant pathogenic fungus *Zymoseptoria tritici*. Recombination rates were estimated in two crosses by calculating genetic map distances (in cM) for non-overlapping segments of 20 kb along chromosomes. Variations in recombination rates are shown for cross 1A5x1E4 (blue) and 3D1x3D7 (red). Recombination rates on accessory chromosomes are shown only for chromosomes that
were found in both parental isolates of a cross. Chromosome 13 contains a region for which recombination rate variations could not be accurately determined (see Methods).

**Figure 4: Recombination rate variations in relation to telomere distance.**

Recombination rates were calculated in non-overlapping 10 kb segments for core chromosomes. Recombination rates were then summarized into 5% bins representing the relative distance to the telomere end. The boxplot shows the median recombination rate (horizontal bar), the 25% and 75% quartiles as a solid box and the 5% and 95% quantiles as vertical lines. Recombination rates were highest in both crosses for subtelomeric distances of 10-15% of the chromosome length. Core chromosomes vary from 1.19 – 6.09 Mb in length.

**Figure 5: Differences in gene density and sequence diversity in the recombination hotspots and non-hotspots of crosses 3D1x3D7 and 1A5x1E4.** For both panels the boxplots show the median of the distribution (horizontal bar), the 25% and 75% quartiles as a solid box and the 5% and 95% quantiles as vertical lines. A) Differences in coding sequence density between hotspot segments (each 10 kb) and segments of the genomic background (“non-hotspots”) divided into 10 kb segments. B) Differences in diversity between the two parental genomes (% SNP) assessed for hotspot segments (each 10 kb) and segments of the genomic background (“non-hotspots”) divided into 10 kb segments.

**Figure 6: The top enriched oligonucleotide motifs that were identified by a de novo motif search in the recombination hotspots of crosses 3D1x3D7 and 1A5x1E4.** The enrichment of sequence motifs in recombination hotspots was tested using a hypergeometric test. The genomic background was segmented and binned according to GC-content. Enrichment tests were performed using weighted background sequences.
reflecting the GC-content distribution found in hotspots. Weighting background sequences avoided spurious enrichments due to differences in GC-content between the genomic background and hotspots.

Figure 7: Recombination rates estimated from linkage analyses in crosses compared to linkage disequilibrium estimated from population resequencing data. Recombination rates between adjacent SNP markers were calculated as cM per Mb in crosses 3D1x3D7 and 1A5x1E4. Resequencing data of field isolates from the same regional population as the parents was used to calculate linkage disequilbria ($r^2$). For this, whole genome sequencing data was analyzed for SNPs segregating at identical positions in the genome as in the SNP dataset generated for the two crosses. Different symbols are used to show data points from core and accessory chromosomes.
Tables

Table 1: Summary of restriction-associated DNA sequencing (RADseq) SNP markers used for the construction of genetic maps. The average genotyping rate of the SNP markers is shown as an average per chromosome and cross. Chromosomes 1-13 are core chromosomes found in all strains of the species, chromosomes 14-21 are accessory.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Cross 3D1x3D7 (227 progeny)</th>
<th>Cross 1A5x1E4 (214 progeny)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of SNPs</td>
<td>Genotyping rate (%)</td>
</tr>
<tr>
<td>1</td>
<td>4697</td>
<td>96.16</td>
</tr>
<tr>
<td>2</td>
<td>2486</td>
<td>97.26</td>
</tr>
<tr>
<td>3</td>
<td>2299</td>
<td>97.31</td>
</tr>
<tr>
<td>4</td>
<td>1614</td>
<td>96.87</td>
</tr>
<tr>
<td>5</td>
<td>1634</td>
<td>95.62</td>
</tr>
<tr>
<td>6</td>
<td>1398</td>
<td>97.03</td>
</tr>
<tr>
<td>7</td>
<td>1379</td>
<td>97.27</td>
</tr>
<tr>
<td>8</td>
<td>1576</td>
<td>96.54</td>
</tr>
<tr>
<td>9</td>
<td>1448</td>
<td>97.08</td>
</tr>
<tr>
<td>10</td>
<td>1013</td>
<td>97.09</td>
</tr>
<tr>
<td>11</td>
<td>920</td>
<td>97.26</td>
</tr>
<tr>
<td>12</td>
<td>923</td>
<td>96.90</td>
</tr>
<tr>
<td>13</td>
<td>806</td>
<td>96.95</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>193</td>
<td>95.49</td>
</tr>
<tr>
<td>17</td>
<td>395</td>
<td>93.72</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>461</td>
<td>97.06</td>
</tr>
<tr>
<td>20</td>
<td>321</td>
<td>94.25</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2: Overview of genetic maps constructed in two crosses of Zymoseptoria tritici. For each chromosome the physical length (in kb), genetic map length (cM) and the number of genetic map positions (minimal number of markers separated by at least one crossover) is shown for each chromosomal map. The recombination rate per chromosome is indicated in cM/Mb.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Physical length (kb)</th>
<th>Genetic map length (cM)</th>
<th>Recombination rate (cM per Mb)</th>
<th>Genetic map positions</th>
<th>Genetic map length (cM)</th>
<th>Recombination rate (cM per Mb)</th>
<th>Genetic map positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6088.8</td>
<td>351.58</td>
<td>57.74</td>
<td>287</td>
<td>255.62</td>
<td>41.98</td>
<td>211</td>
</tr>
<tr>
<td>2</td>
<td>3860.1</td>
<td>250.74</td>
<td>64.96</td>
<td>177</td>
<td>165.97</td>
<td>43.00</td>
<td>148</td>
</tr>
<tr>
<td>3</td>
<td>3505.3</td>
<td>240.19</td>
<td>68.52</td>
<td>176</td>
<td>183.56</td>
<td>52.37</td>
<td>134</td>
</tr>
<tr>
<td>4</td>
<td>2880.0</td>
<td>217.03</td>
<td>75.36</td>
<td>180</td>
<td>147.23</td>
<td>51.12</td>
<td>111</td>
</tr>
<tr>
<td>5</td>
<td>2861.8</td>
<td>201.39</td>
<td>70.37</td>
<td>161</td>
<td>156.52</td>
<td>54.69</td>
<td>117</td>
</tr>
<tr>
<td>6</td>
<td>2675.0</td>
<td>188.22</td>
<td>70.36</td>
<td>124</td>
<td>174.2</td>
<td>65.12</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>2665.3</td>
<td>166.57</td>
<td>62.50</td>
<td>109</td>
<td>143.18</td>
<td>53.72</td>
<td>102</td>
</tr>
<tr>
<td>8</td>
<td>2443.6</td>
<td>165.56</td>
<td>67.75</td>
<td>124</td>
<td>132.77</td>
<td>54.33</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>2142.5</td>
<td>169.94</td>
<td>79.32</td>
<td>120</td>
<td>127.35</td>
<td>59.44</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>1682.6</td>
<td>140.97</td>
<td>83.78</td>
<td>86</td>
<td>108.63</td>
<td>64.56</td>
<td>77</td>
</tr>
<tr>
<td>11</td>
<td>1624.3</td>
<td>154.48</td>
<td>95.11</td>
<td>76</td>
<td>113.57</td>
<td>69.92</td>
<td>76</td>
</tr>
<tr>
<td>12</td>
<td>1462.6</td>
<td>138.58</td>
<td>94.75</td>
<td>89</td>
<td>96.74</td>
<td>66.14</td>
<td>68</td>
</tr>
<tr>
<td>13</td>
<td>1185.8</td>
<td>121.46</td>
<td>102.43</td>
<td>64</td>
<td>97.56</td>
<td>82.28</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>773.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>33.22</td>
<td>42.97</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>639.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>45.98</td>
<td>71.90</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>607.0</td>
<td>49.24</td>
<td>81.11</td>
<td>29</td>
<td>42.26</td>
<td>69.62</td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td>584.1</td>
<td>53.34</td>
<td>91.32</td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>573.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22.08</td>
<td>38.49</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>549.9</td>
<td>58.39</td>
<td>106.19</td>
<td>30</td>
<td>48.36</td>
<td>87.95</td>
<td>30</td>
</tr>
<tr>
<td>20</td>
<td>472.1</td>
<td>55.94</td>
<td>118.49</td>
<td>30</td>
<td>47.58</td>
<td>100.78</td>
<td>24</td>
</tr>
<tr>
<td>21</td>
<td>409.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16.54</td>
<td>40.42</td>
<td>15</td>
</tr>
<tr>
<td>all</td>
<td>39'686.3</td>
<td>2723.6</td>
<td>1892</td>
<td>2158.9</td>
<td>1592</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 3:** Overview of crossover counts per chromosome in progeny of cross 1A5x1E4 and 3D1x3D7.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Cross 3D1x3D7</th>
<th></th>
<th>Cross 1A5x1E4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median crossover count</td>
<td>Mean crossover count</td>
<td>Crossing over count per Mb</td>
<td>Median crossover count</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.34</td>
<td>0.55</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2.42</td>
<td>0.63</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.30</td>
<td>0.65</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2.10</td>
<td>0.73</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1.89</td>
<td>0.66</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.79</td>
<td>0.67</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1.61</td>
<td>0.60</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1.59</td>
<td>0.65</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1.61</td>
<td>0.75</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1.33</td>
<td>0.79</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1.45</td>
<td>0.90</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1.33</td>
<td>0.91</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1.14</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0.45</td>
<td>0.75</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0.48</td>
<td>0.83</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4: Genes with a predicted function in virulence located in recombination hotspots of *Zymoseptoria tritici*. Functional predictions and population genetic analyses of cell wall degrading enzymes and small secreted proteins are summarized where available (Goodwin et al. 2011; Morais Do Amaral et al. 2012; Brunner et al. 2013).

<table>
<thead>
<tr>
<th>Cross</th>
<th>Protein ID</th>
<th>Protein characteristics</th>
<th>Chromosome and position</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D1x3D7</td>
<td>107758</td>
<td>Small secreted tandem repeat protein (TRP18)</td>
<td>1 5652012-5653092</td>
</tr>
<tr>
<td>1A5x1E4</td>
<td>90262</td>
<td>Small, secreted protein</td>
<td>1 5790864-5792208</td>
</tr>
<tr>
<td>3D1x3D7</td>
<td>96331</td>
<td>Cystinase MgCUT4, under diversifying selection in <em>Z. tritici</em> for evasion of host</td>
<td>2 3479087-3479856</td>
</tr>
<tr>
<td>3D1x3D7 &amp; 1A5x1E4</td>
<td>104283</td>
<td>Small, cysteine-rich secreted protein</td>
<td>4 2371447-2372597</td>
</tr>
<tr>
<td>3D1x3D7</td>
<td>42164</td>
<td>Small, secreted protein (ZPS1 family)</td>
<td>5 224006-225198</td>
</tr>
<tr>
<td>1A5x1E4</td>
<td>105351</td>
<td>Small, secreted protein</td>
<td>8 444217-444909</td>
</tr>
<tr>
<td>3D1x3D7</td>
<td>105728</td>
<td>Beta-xylosidase MgXYL4, under purifying selection in <em>Z. tritici</em> for substrate optimization</td>
<td>9 989316-990401</td>
</tr>
<tr>
<td>3D1x3D7 &amp; 1A5x1E4</td>
<td>110790</td>
<td>Small, secreted protein</td>
<td>9 1270903-1271945</td>
</tr>
<tr>
<td>3D1x3D7 &amp; 1A5x1E4</td>
<td>95797</td>
<td>Small, secreted protein</td>
<td>9 1274185-1275470</td>
</tr>
<tr>
<td>3D1x3D7</td>
<td>111130</td>
<td>Hemi-cellulase (ABFa alpha-L-arabinofuranosidase-A)</td>
<td>10 1379647-1381988</td>
</tr>
<tr>
<td>3D1x3D7</td>
<td>111451</td>
<td>Secreted alpha-1,2-mannosyltransferase ALG</td>
<td>12 554075-555902</td>
</tr>
<tr>
<td>1A5x1E4</td>
<td>97702</td>
<td>Aspartic protease</td>
<td>16 71035-72069</td>
</tr>
</tbody>
</table>
Distance between recombination events (kb) vs. Number of events for 3D1x3D7 and 1A5x1E4.
Cross 3D1X3D7

<table>
<thead>
<tr>
<th>Background sequences with motif (%)</th>
<th>Hotspots with motif (%)</th>
<th>Enrichment p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.42</td>
<td>53.85</td>
<td>1e-14</td>
</tr>
</tbody>
</table>

Cross 1A5X1E4

<table>
<thead>
<tr>
<th>Background sequences with motif (%)</th>
<th>Hotspots with motif (%)</th>
<th>Enrichment p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.62</td>
<td>21.05</td>
<td>1e-13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Background sequences with motif (%)</th>
<th>Hotspots with motif (%)</th>
<th>Enrichment p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.94</td>
<td>48.68</td>
<td>1e-12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Background sequences with motif (%)</th>
<th>Hotspots with motif (%)</th>
<th>Enrichment p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.61</td>
<td>23.68</td>
<td>1e-12</td>
</tr>
</tbody>
</table>
[Linkage resequencing]

Population disequilibrium ($r^2$)

Recombination rate between adjacent markers (cM/Mb)

Cross 3D1x3D7

Cross 1A5x1E4

[Linkage analyses in crosses]