DISS. ETH NO.26555

DRIVERS OF GENOME EVOLUTION IN A FUNGAL PATHOGEN OF WHEAT

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SIMONE FOUCHÉ

Master of Science in Microbiology, University of Pretoria

Born on 02.01.1989

Citizen of South Africa

accepted on the recommendation of

Prof. Dr. Bruce McDonald

Prof. Dr. Daniel Croll

Prof. Dr. Alexander Suh

2020

TABLE OF CONTENTS

SUMMARY

Chromosome rearrangements involve duplication, deletions, inversions and translocations. Breakpoints of chromosome rearrangements are frequently in close proximity to transposable elements (TEs). TEs are known to mediate chromosome rearrangements through their own activity or through ectopic recombination. During this PhD we aimed to better understand the causes and consequences of chromosome rearrangements in *Zymoseptoria tritici*, an important pathogen of wheat. To study the origins of chromosome rearrangements the first chapter focusses on the de-repression of TEs, which is stress induced during a wheat infection cycle as well as in nutrient limited media. Stress was shown to drive epigenetic changes and trigger TE de-repression in multiple organisms. We find that TEs respond differently to stresses. Furthermore, effector genes in close proximity to TEs show a de-repression during early infection suggesting that TEs and effectors may be under the same epigenetic control. De-repressed TEs can place a mutational burden on the genome. Therefore, in the second chapter we aimed to quantify the number of chromosome rearrangements occurring in all 21 chromosomes in hundreds of progeny through a single round of meiosis. We find that the fidelity with which chromosomes go through meiosis differs between chromosomes. Chromosomes with a higher repeat content and lower synteny were less stable. In the final chapter we focused on a single rearranged chromosome that was generated by a self-fusion. We hypothesized that such a fused chromosome would go through degenerative breakage-fusion-bridge (BFB) cycles. Here we show the exact process whereby the highly unstable fused chromosome was created through ectopic recombination between a specific repeat family. We trace the fate of the novel chromosome through five rounds of meiosis and show that degenerative cycles occur through repeated ectopic recombination and non-disjunction. The ability of *Z. tritici* to tolerate chromosome duplications, losses and rearrangements makes this species a great

model to observe and investigate the interplay between TE dynamics and chromosome rearrangements.

ZUSAMMENFASSUNG

Chromosomenumlagerungen beinhalten Duplizierungen, Deletionen, Inversionen und Translokationen. Bruchstellen von Chromosomenumlagerungen liegen häufig in unmittelbarer Nähe zu transponierbaren Elementen (TEs). TEs können Chromosomenumlagerungen durch ihre eigene Aktivität oder durch ektopische Rekombination vermitteln. In dieser Arbeit wollten wir die Ursachen und Folgen von Chromosomenumlagerungen bei *Zymoseptoria tritici*, einem wichtigen Erreger von Weizen, besser verstehen. Um die Ursprünge der Chromosomenumlagerungen zu untersuchen, konzentrierten wir uns im ersten Kapitel auf die Derepression von TEs, die während des Infektionszyklus von Weizen und in nährstoffarmen Medien induziert werden. In mehreren Organismen konnte gezeigt werden, dass Stress epigenetische Veränderungen und TE Derepression auslöst. Wir stellen fest, dass TEs unterschiedlich auf Belastungen reagieren. Darüber hinaus zeigen Effektorgene in unmittelbarer Nähe zu TEs eine Derepression während einer frühen Infektion, was darauf hindeutet, dass TEs und Effektoren unter der gleichen epigenetischen Kontrolle stehen. Derepressive TEs sind stark mutagen und stellen deshalb eine große Belastung dar für das Genom. Wir haben deshalb im zweiten Kapitel versucht, die Anzahl der Chromosomenumlagerungen zu quantifizieren, die in allen 21 Chromosomen in einer einzigen Runde Meiose in Hunderten von Nachkommen entstanden sind. Wir stellten fest, dass die Genauigkeit, mit der Chromosomen durch die Meiose gehen, sich zwischen Chromosomen unterscheidet. Chromosomen mit einem höheren Gehalt von repetitiven Elementen und einer niedrigeren Syntenie waren instabiler. Im letzten Kapitel konzentrierten wir uns auf ein einzelnes umgelagertes Chromosom, welches durch eine Selbstfusion erzeugt wurde. Wir haben angenommen, dass ein solches fusioniertes Chromosom degeneriert sein würde aufgrund von sogenannten „Bruch-Fusion-Verbindungs-Zyklen“ (engl. breakage-fusion-bridge cycles). Wir konnten den genauen Prozess aufzeigen, bei dem das hochstabile fusionierte Chromosom durch ektopische Rekombination zwischen

einer bestimmten Familie von repetitiven Elementen erzeugt wurde. Wir verfolgten das Schicksal des neuartigen Chromosoms durch fünf Runden Meiose und zeigten auf, dass degenerative Zyklen durch wiederholte ektopische Rekombination und durch fehlende Disjunktion entstehen. Die Fähigkeit von *Z. tritici*, Chromosomenduplikationen, Chromosomenverluste und Chromosomenumlagerungen zu tolerieren, macht diese Spezies zu einem großartigen Modell, um das Zusammenspiel zwischen TE-Dynamik und Chromosomenveränderungen zu beobachten und zu untersuchen.

GENERAL INTRODUCTION

Genome evolution and transposable elements

The majority of the eukaryotic genome is organized into linear chromosomes (Timmis et al., 2004). Chromosomes despite the presence of stabilizing centromeres and telomeres are dynamic genetic elements, contributing to genome plasticity through duplications, deletions, rearrangements, translocations and inversions (Coghlan et al., 2005). Hence, genomes are continuously evolving in populations. Most chromosome aberrations are deleterious, but in rare cases they can be beneficial. Aneuploidy or the presence of an additional copy of an additional chromosome has enabled adaptive evolution in some rare cases (Chen et al., 2012; Yona et al., 2012). Breakpoints of chromosome rearrangement are generally found in close proximity to transposable elements (TEs) (Coghlan et al., 2005). TEs are subdivided into two major categories according to their mechanism of transposition. Class I TEs that retrotranspose through an RNA intermediate (*i.e.* RNA transposons) according to a 'copy-and-paste' mechanism (Boeke et al. 1985) and class II TEs that transpose as DNA through a 'cut-and-paste' mechanism (*i.e.* DNA transposons) (Greenblatt and Brink 1963; Rubin et al. 1982) or a 'peel and paste' replicative mechanism in the case of *Helitrons* (Grabundzija et al. 2016). Both classes of TEs are associated with chromosome rearrangements. Transposable elements are major components of eukaryotic genomes even though the percentage of TEs varies greatly between species.

TEs affect genomes by inserting into and disrupting genes, by introducing chromosomal rearrangements and by altering the epigenetic state and expression profiles of adjacent genes (Hollister & Gaut, 2009; Lim, 1988; Oliver, McComb, & Greene, 2013; Petrov, Aminetzach, Davis, Bensasson, & Hirsh, 2003; Slotkin & Martienssen, 2007). The uncontrolled spread and proliferation of mobile elements can have a fitness cost by bloating the genome size and increasing the likelihood of deleterious ectopic recombination (Chuong et al. 2017; Mita & Boeke 2016). TEs can promote chromosome rearrangements long after

they are no longer able to transpose through recombination between highly similar regions (Carvalho and Lupski 2016). TE insertions rarely provide any fitness advantage to the host and so a major question is why TEs persist through evolution? Hosts have evolved a plethora of mechanisms to control their TEs (Slotkin and Martienssen 2007). Furthermore, natural selection and drift are two of the major forces shaping the TE content in genomes, where strongly deleterious TEs are rapidly removed from the population (Lynch 2007). Insertions that are close to neutral in terms of their effect on genome function or host fitness can reach fixation by drift. Therefore, the fate of a TE in a genome also depends on some intrinsic properties of the elements, such as insertion site preferences (Bourque et al. 2018).

Plant pathogenic fungi are great models to study TEs and chromosome rearrangements. Plant pathogenic fungi and oomycetes pose a major risk to food security (Fisher et al. 2012) because of the dynamic nature of their genomes and subsequent ability to rapidly evolve virulence to crops. Fungal genomes are highly polymorphic in terms of karyotype, likely as a result of occurring in such a wide range of habitats (Kistler and Miao, 1992; Zolan, 1995). Many fungal pathogens have so-called accessory chromosomes that have presence-absence polymorphism within a species and typically accumulate more mutations and structural variations than other regions of the genome (Bertazzoni et al. 2018). Not surprisingly filamentous pathogens show great variability in terms of genome size. This size variation is explained by the accumulation of TEs in some genomes (Raffaele and Kamoun, 2012). TEs are drivers of chromosome length polymorphisms through ectopic recombination during meiosis (Bzymek and Lovett 2001; Argueso et al. 2008; Raskina et al. 2008). However, the extent to which chromosome rearrangements happens through one round of meiosis and the factors affecting the transmission of chromosomes from one generation to the next are not well characterized.

TEs and chromosome rearrangements have played a role in shaping pathogen genomes. Many fungal pathogens have a bi-partite genome architecture (Raffaele and Kamoun 2012; Dong et al. 2016), where so-called effector genes are located in the repeat rich regions of the genome. Effectors are small secreted proteins that play a role in manipulating the host cell and/or interfere with or protect the pathogen from the host's defenses. Furthermore, effectors have a very characteristic and tightly regulated expression profile during infection, because of playing a highly specific role at a critical point of time (Skibbe et al. 2010; Rouxel et al. 2011; Hacquard et al. 2012; Kleemann et al. 2012; Gervais et al. 2016; Palma-Guerrero et al. 2016). TE rich regions are hotspots for mutations, gene deletions and alterations to the expression profiles of existing effectors as well as the emergence of new effectors which can contribute to gains in virulence (Presti et al. 2015). Furthermore, RIP (repeat-induced point mutation) is a fungal specific premeiotic genomic defense mechanism that targets and mutates TEs and other repeated sequences and has been shown to contribute to effector diversification in *Leptosphaeria maculans* (Galagan and Selker 2004; Rouxel et al. 2011). In addition, effector genes can benefit from their co-localization with TEs because there appears to be a relationship between the de-repression of TEs and the expression of effectors. Specifically, the expression of effectors is often governed by de-repression of facultative heterochromatin (Connolly et al. 2013; Qutob et al. 2013; Chujo and Scott 2014; Soyer et al. 2014; Schotanus et al. 2015; Soyer et al. 2015; Studt et al. 2016). Regions of facultative heterochromatin encoding effectors overlap with TEs (Soyer et al. 2015; Seidl and Thomma 2017). TE transcription levels are influenced by epigenetic silencing of the host genome and environmental stimuli such as stress, but the underlying mechanisms are poorly understood.

Zymoseptoria tritici is an important pathogen of wheat with a highly dynamic genome containing 13 core and up to eight accessory chromosomes (Goodwin et al., 2011). The pathogen has a highly plastic genome consisting of 13 core and up to eight accessory

chromosomes that have presence-absence polymorphism within the species (Goodwin et al. 2011). Accessory chromosomes frequently undergo chromosomal rearrangements with breakpoints overlapping with TE insertions (Croll et al. 2013; Plissonneau et al. 2016, 2018; Hartmann et al. 2017). Furthermore, *Z. tritici* reproduces sexually at least once per wheat growing season (Kema et al., 1996), providing the means for losses, duplications, disomy and other rearrangements to occur during either meiosis or mitosis (Wittenberg et al., 2009; Croll et al., 2013). *Z. tritici* tolerates extensive chromosome length and number polymorphisms, including whole chromosome losses and disomy (Wittenberg et al., 2009), making it an ideal model for the study of different aspects of genome plasticity and transposable element dynamics.

Main questions and outlines for this PhD thesis:

In this PhD project, I aim to identify the causes and consequences of chromosomal rearrangements in the fungal pathogen model organism *Z. tritici*.

To understand the origins of TE-mediated chromosomal rearrangements, in my first chapter I focus on factors driving the activation or repression of TEs in the genome. Stress has been shown to trigger epigenetic changes which result in TE de-repression (Miousse et al. 2015). My **first chapter** focuses on the relationship between TE de-repression dynamics and stress. TE responsiveness to stress can increase TE transcription and transposition levels, driving bursts of activity of certain TE families and many new TE insertions (Dubin et al. 2018). Plant pathogens such as *Z. tritici* experience distinct stressors both on the host and off the host. Host defense mechanisms significantly alter the biotic environment for an attacking pathogen. Pathogens also likely face starvation stress after depleting resources on the host or during resting stages outside of the host. Hence, I investigated how TE expression is shaped by these specific stress factors.

The responsiveness of TEs to stress places as mutational burden on the host genome. TEs and their activity are drivers of chromosome rearrangements such as duplications, deletions, inversion and translocations. Hence, my **second chapter** focuses on quantifying the number of whole chromosome losses, duplications and rearrangements occurring in a single round of meiosis focusing on all of the 21 chromosomes in hundreds of progeny isolates. We also identify mechanisms affecting the fidelity with which chromosomes go through meiosis.

Having established general patterns that shape the faithful transmission of chromosomes, I focus in my **third chapter** on one massive, serendipitously discovered chromosome rearrangement. The chromosome appeared to be a self-fusion resulting in the amplification of a large segment of the chromosome. We hypothesized that such a fused chromosome would go through degenerative (breakage-fusion-bridge (BFB) cycles). Here we wanted to understand the exact process whereby highly unstable fused chromosomes are created and degenerate.

Given the focus on plant pathogens in this work, I wanted to provide a broader context how TE activity and chromosomal rearrangements can ultimately contribute to virulence evolution. **Chapter four** constitutes a short review focusing on how effector genes emerge and are lost in pathogen genomes by linking chromosome rearrangements to effector gene evolution as a consequence of chromosome rearrangements. I conclude with a **general conclusion and perspective**.

For my three main experimental chapters, I am now addressing the following research questions in more detail:

1. What is the impact of stress on transposable element de-repression on an infection cycle in wheat?

2. What are the factors that affect the fidelity of chromosomal inheritance through meiosis?

3. What are the mechanisms underlying the formation of an enlarged chromosome 17 and how does the new chromosome degenerate through breakage-fusion-bridge cycles?

Chapter 1 describes how I used transcriptome profiling to test for the impact of two major stress factors in the life-cycle of the pathogen on TE de-repression, namely nutrient deprivation stress and infection stress. My setup included transcriptomes of *Z. tritici* from a nutrient-rich culture medium as a non-stress environment and from a nutrient-deprived medium that simulates starvation. I also analyzed the fungal transcriptome at four distinct stages during the infection of wheat spanning the early symptomless stage, the peak of lesion formation and the saprotrophic stage. I replicated the two stress experiments with four genetically distinct strains of *Z. tritici* to identify how the genetic background influences TE responsiveness. TEs showed the highest de-repression under nutrient stress, but the expression differed significantly between TE families and between genetic backgrounds. During infection stress, a large number of TE families were de-repressed at the peak of symptom development on wheat leaves. Hereafter, I determined how the genomic location affected the de-repression dynamics of TEs and identified distinct patterns depending on the type of stress, the distance to the closest genes and the impact of genomic defense mechanisms. Finally, I analyzed a locus segregating variation at a key effector gene involved in host adaptation for the impact of TE de-repression. I show that the insertion of specific TEs led to silencing and in turn promoting virulence.

Chapter 2: In this study, I analyze the mechanisms that affect the fidelity of chromosomal inheritance through meiosis. For this I identified chromosomal rearrangements, losses, and duplications by screening hundreds of progeny genotypes generated from two independent crosses. I determined the rate of aneuploidy, patterns of rearrangement and distortions in

transmission rates. Chromosomal duplications (disomy) rarely occurred in the core chromosomes, while accessory chromosomes showed high frequencies of disomy. However, chromosomal rearrangements were found to occur only on accessory chromosomes and were even more frequent than disomy. Accessory chromosomes present in only one of the parents in one cross were inherited significantly more frequently than the expected 1:1 segregation ratio. Hereafter, I investigated whether factors such as length similarity, synteny, recombination rate, and repetitive element content affected the accuracy with which chromosomes were inherited. Both the chromosome and the parental background had affected the rates of disomy, losses, rearrangements, and distorted inheritance. Chromosomes with higher sequence similarity and lower repeat content were inherited more faithfully.

Chapter 3: In this study, I analyze the mechanism whereby a novel enlarged chromosome 17 is created and how the chromosome degenerates through four rounds of meiosis. We sequence twelve progeny genomes to track the sequence changes as this chromosome goes through degenerative cycles. Crosses were performed between isolates selected specifically to determine the effect of sequence and length similarity on the fidelity with which the novel chromosome goes through meiosis. We performed the fourth generations of crosses to determine if the chromosome 17 variants would stabilize or continue to degenerate. This study provides insights into the mechanisms generating a highly unstable chromosome through ectopic recombination involving transposable elements. Degeneration progresses through a combination of non-homologous recombination and non-disjunction. The process is fairly random as progeny from the same cross had different derivatives of chromosome 17, but also to some extent predictable as the same TE family and chromosome regions were almost always involved.

References

Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. U. S. A. 105:11845–11850.

Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, Hane JK. 2018. Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. Mol. Plant-Microbe Interact. MPMI 31:779–788.

Boeke JD, Garfinkel DJ, Styles CA, Fink GR. 1985. Ty elements transpose through an RNA intermediate. Cell 40:491–500.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. Genome Biol. 19:199.

Bzymek M, Lovett ST. 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U. S. A. 98:8319–8325.

Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. 17:224–238.

Chen G, Rubinstein B, Li R. 2012. Whole chromosome aneuploidy: Big mutations drive adaptation by phenotypic leap. BioEssays 34:893–900.

Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates fungal alkaloid biosynthesis in a fungal endophyte-plant symbiosis. Mol. Microbiol. 92:413–434.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. 18:71–86.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 21:673–682.

Connolly LR, Smith KM, Freitag M. 2013. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLOS Genet 9:e1003916.

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

Dong Y, Li Y, Qi Z, Zheng X, Zhang Z. 2016. Genome plasticity in filamentous plant pathogens contributes to the emergence of novel effectors and their cellular processes in the host. Curr. Genet. 62:47–51.

Dubin MJ, Mittelsten Scheid O, Becker C. 2018. Transposons: a blessing curse. Curr. Opin. Plant Biol. 42:23–29.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186–194.

Galagan JE, Selker EU. 2004. RIP: the evolutionary cost of genome defense. Trends Genet. TIG 20:417–423.

Gervais J, Plissonneau C, Linglin J, Meyer M, Labadie K, Cruaud C, Fudal I, Rouxel T, Balesdent M-H. 2016. Different waves of effector genes with contrasted genomic location are expressed by *Leptosphaeria maculans* during cotyledon and stem colonisation of oilseed rape. Mol. Plant Pathol.:n/a-n/a.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Lee TAJV der, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat

pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLOS Genet 7:e1002070.

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, et al. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat. Commun. 7:1–12.

Greenblatt IM, Brink RA. 1963. Transpositions of Modulator in Maize into Divided and Undivided Chromosome Segments. Nature 197:412–413.

Hacquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, Legué V, Kohler A, Tanguay P, Petre B, et al. 2012. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). Mol. Plant-Microbe Interact. MPMI 25:279–293.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11:1189–1204.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 19:1419–1428.

Kema GH, Verstappen EC, Todorova M, Waalwijk C. 1996. Successful crosses and molecular tetrad and progeny analyses demonstrate heterothallism in *Mycosphaerella graminicola*. Curr. Genet. 30:251–258.

Kistler HC, Miao VPW. 1992. New Modes of Genetic Change in Filamentous Fungi. Annu. Rev. Phytopathol. 30:131–153.

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Themaat EVL van, Does HC van der, Hacquard S, Stüber K, Will I, Schmalenbach W, et al. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLOS Pathog. 8:e1002643.

Lim JK. 1988. Intrachromosomal rearrangements mediated by hobo transposons in Drosophila melanogaster. Proc. Natl. Acad. Sci. 85:9153–9157.

Lynch M. 2007. The origins of genome architecture. Sunderland, Mass.: Sinauer Associates Available from: http://catalog.hathitrust.org/api/volumes/oclc/77574049.html

Miousse IR, Chalbot M-CG, Lumen A, Ferguson A, Kavouras IG, Koturbash I. 2015. Response of transposable elements to environmental stressors. Mutat. Res. Rev. Mutat. Res. 765:19–39.

Mita P, Boeke JD. 2016. How retrotransposons shape genome regulation. Curr. Opin. Genet. Dev. 37:90–100.

Oliver KR, McComb JA, Greene WK. 2013. Transposable Elements: Powerful Contributors to Angiosperm Evolution and Diversity. Genome Biol. Evol. 5:1886–1901.

Palma-Guerrero J, Stefano F. F. Torriani, Zala M, Carter D, Courbot M, Rudd JJ, McDonald BA, Croll D. 2016. Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intraspecific variability in transcription profiles. Mol. Plant Pathol. 17:845–859.

Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila. Mol. Biol. Evol. 20:880–892.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16:5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio 7:e01231-16.

Presti LL, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. 66:513–545.

Qutob D, Patrick Chapman B, Gijzen M. 2013. Transgenerational gene silencing causes gain of virulence in a plant pathogen. Nat. Commun. 4:1349.

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

Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet. Genome Res. 120:351–357.

Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, et al. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202.

Rubin GM, Kidwell MG, Bingham PM. 1982. The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. Cell 29:987–994.

Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock EH. 2015. Histone modifications rather than the novel regional

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

Seidl MF, Thomma BPHJ. 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. [Internet] 0. Available from: http://www.cell.com/trends/genetics/abstract/S0168-9525(17)30112-9

Skibbe DS, Doehlemann G, Fernandes J, Walbot V. 2010. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. Science 328:89–92.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8:272–285.

Soyer JL, Ghalid ME, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent M-H, Connolly LR, Freitag M, Rouxel T, et al. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. PLOS Genet. 10:e1004227.

Soyer JL, Rouxel T, Fudal I. 2015. Chromatin-based control of effector gene expression in plant-associated fungi. Curr. Opin. Plant Biol. 26:51–56.

Studt L, Rösler SM, Burkhardt I, Arndt B, Freitag M, Humpf H-U, Dickschat JS, Tudzynski B. 2016. Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. Environ. Microbiol.:n/a-n/a.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123–135.

Wittenberg AHJ, Lee TAJ van der, M'Barek SB, Ware SB, Goodwin SB, Kilian A, Visser RGF, Kema GHJ, Schouten HJ. 2009. Meiosis drives extraordinary genome plasticity

in the haploid fungal plant pathogen *Mycosphaerella graminicola*. PLOS ONE 4:e5863.

Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, Pilpel Y, Dahan O. 2012. Chromosomal duplication is a transient evolutionary solution to stress. Proc. Natl. Acad. Sci. U. S. A. 109:21010–21015.

Zolan ME. 1995. Chromosome-length polymorphism in fungi. Microbiol. Rev. 59:686–698.

CHAPTER 1:

Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen

Simone Fouché, Thomas Badet, Ursula Oggenfuss, Clémence Plissonneau, Carolina Sardinha Francisco, Daniel Croll

Accepted for publication in *Molecular Biology and Evolution* (September 2019), in press, doi: https://doi.org/10.1093/molbev/msz216

The genome assembly and annotation for 1A5, 1E4, 3D1, and 3D7 genomes are available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession numbers PRJEB15648, PRJEB20900, PRJEB20899, and PRJEB14341. The *in planta* RNA-sequencing raw data sets are available at the NCBI Short Read Archive under accession number SRP077418. The in vitro RNA-sequencing raw sequencing data were deposited into the NCBI Short Read Archive under the accession number SRP152081.

Abstract

Transposable elements (TEs) are drivers of genome evolution and affect the expression landscape of the host genome. Stress is a major factor inducing TE activity; however, the regulatory mechanisms underlying de-repression are poorly understood. Plant pathogens are excellent models to dissect the impact of stress on TEs. The process of plant infection induces stress for the pathogen, and virulence factors (*i.e.,* effectors) located in TE-rich regions become expressed. To dissect TE de-repression dynamics and contributions to virulence, we analyzed the TE expression landscape of four strains of the major wheat pathogen *Zymoseptoria tritici*. We experimentally exposed strains to nutrient starvation and host infection stress. Contrary to expectations, we show that the two distinct conditions induce the expression of different sets of TEs. In particular, the most highly expressed TEs, including a miniature inverted-repeat transposable element and long terminal repeat-*Gypsy* elements, show highly distinct de-repression across stress conditions. Both the genomic context of TEs and the genetic background stress (*i.e.* different strains harboring the same TEs) were major predictors of de-repression under stress. Gene expression profiles under stress varied significantly depending on the proximity to the closest TEs and genomic defenses against TEs were largely ineffective to prevent de-repression. Next, we analyzed the locus encoding the Avr3D1 effector. We show that the insertion and subsequent silencing of TEs in close proximity likely contributed to reduced expression and virulence on a specific wheat cultivar. The complexity of TE responsiveness to stress across genetic backgrounds and genomic locations demonstrates substantial intraspecific genetic variation to control TEs with consequences for virulence.

Introduction

Transposable elements (TEs) are mobile genetic elements that were first discovered in maize (McClintock 1950) and propagate in genomes without apparent benefit to the host (Doolittle and Sapienza 1980). Uncontrolled spread of TEs is thought to have a fitness cost to the host due to the increased genome size and higher likelihood of deleterious, non-homologous recombination events (Chuong et al. 2017; Mita & Boeke 2016). TEs are subdivided into two major categories according to their mechanism of replication, namely, class I TEs that transpose through an RNA intermediate (*i.e.* RNA transposons) and class II TEs that transpose through a cut-and-paste mechanism (*i.e.* DNA transposons). Both classes of TEs are expressed. Host genomes have co-evolved with their TEs to suppress their expression (Slotkin and Martienssen 2007). These mechanisms include epigenetic silencing through histone modifications or DNA methylation, targeted mutagenesis and small RNA interference. In order to autonomously replicate in the genome, some TEs evolved or co-opted regulatory sequences to ensure their own transcription. As a consequence, the dispersed nature of TE regulatory sequences shapes the expression landscape of the genome (Mita and Boeke 2016; Chuong et al. 2017). Epigenetic silencing of the host genome and environmental triggers are major factors influencing TE transcription levels, although the underlying mechanisms are poorly understood.

Most TEs are transcriptionally and transpositionally quiescent (Yoder et al. 1997; Zilberman et al. 2007). However, environmental stimuli and stress, in particular, have been shown to trigger epigenetic de-repression of TEs resulting in the activation of insertional mutagenesis (Miousse et al. 2015). TE de-repression in response to stress is widely shared across eukaryotes (Bundo et al. 2014; Van Meter et al. 2014; Voronova et al. 2014; Romero-Soriano and Guerreiro 2016; Ryan et al. 2016; Zovoilis et al. 2016; Huang et al. 2017; Hummel et al. 2017; Shpyleva et al. 2018). De-repression of TEs under stress usually

impacts TE transcription levels and can increase transpositional activity (Dubin et al. 2018). The impact of stress on TEs is often mediated through changes in the epigenetic state of the genome (*i.e.* de-repression) (Horváth et al. 2017) or the activation by a transcription factor (Capy et al. 2000). Some TEs have stress response elements (SRE) that are regulatory sequences activated in response to stress (Bucher et al. 2012; Casacuberta and González 2013). SREs are most common in long terminal repeat (LTR) retrotransposons and have been identified in one family of miniature inverted-repeat transposable elements (MITE) (Yasuda et al. 2013). The relationship between stress and TE activation is complex with some studies showing TEs being upregulated, some show TE repression and yet other studies show transient upregulation and then downregulation following exposure to a stress (Horváth et al. 2017). Stress mostly impacts facultative heterochromatin (Trojer and Reinberg 2007), while constitutive heterochromatin is typically associated with gene-poor, TE rich regions that maintain repression (Dillon 2004; Saksouk et al. 2015). The distribution of TE families or specific copies of a TE can be strongly correlated with the local chromatin state (Lanciano and Mirouze 2018). The epigenetic landscape influencing TE de-repression dynamics is a highly dynamic trait among closely related species (Niederhuth et al. 2016) but also showing significant variation within species (Barah et al. 2013).

TE responsiveness to stress potentially constitutes a major compound cost to the deleterious impact of stress on an organism. However, stress can induce both the activation and repression of TEs as was shown for different ecotypes of *A. thaliana* exposed to cold stress (Barah et al. 2013). In yeast and human cells TEs were found to be repressed in response to stress (Menees and Sandmeyer 1996; Trivedi et al. 2014). Another example in *Arabidopsis thaliana,* the *ONSEN* (LTR) retrotransposon is activated in response to heat stress due to heat response factors recognizing a regulatory sequence in the promoter of the ONSEN transposon (Ito et al. 2011; Cavrak et al. 2014). As a consequence, ONSEN insertions into genic regions were shown to induce the transcriptional upregulation of

neighboring genes in response to heat stress (Ito et al. 2011). Therefore, TEs are frequently re-activated in response to stress and their activation can introduce new TE copies into the genome with *cis*-regulatory elements or associated chromatin states that are responsive to stress, thereby rewiring the stress response network of the genome (Cowley and Oakey 2013; Galindo-González et al. 2017). Hence, the stress activation of TEs likely depends on the type of stress, the identity of the TE and the genetic background of the host. Furthermore, TE activation may generate adaptive genetic variation and accelerate host stress adaptation.

TE de-repression dynamics in pathogens of plants show the hallmarks of a conflict between TE proliferation and host control. Insertions of TEs in pathogen genomes generate significant adaptive genetic variation through gene inactivation, gene copy-number variation and altered gene expression, and have been shown to play a role in the evolution of genes encoding proteins involved in host interaction (Croll and McDonald 2012; Seidl and Thomma 2017; Fouché et al. 2018). In fungi, TEs can also lead to genetic variation through repeat-induced point mutation (RIP), a genome defense mechanism that targets and mutates repetitive sequences (Selker 2002). In *Leptosphaeria maculans* for instance, leakage of RIP into neighboring regions contributes to the diversification of effector genes (Rouxel et al. 2011). During the infection of a host plant, the pathogen must overcome a number of severe stresses (Ferreira et al. 2006; Hernández-Chávez et al. 2017). Initially the pathogen is exposed to nutrient stress on the surface of the plant (Derridj 1996). Once the pathogen enters the plant, host defenses stimulate the accumulation of toxic reactive oxygen species (Shetty et al. 2007). To face plant-induced stresses and promote disease, pathogens express virulence factors (*i.e.* effectors). The expression of effectors is often governed by de-repression of facultative heterochromatin (Connolly et al. 2013; Qutob et al. 2013; Chujo and Scott 2014; Soyer et al. 2014; Schotanus et al. 2015; Soyer et al. 2015; Studt et al. 2016). Hence, infection stress incidentally serves as an epigenetic trigger for adaptive

upregulation of effectors (Sánchez-Vallet et al. 2018). Importantly, regions of facultative heterochromatin encoding effectors overlap with TEs (Soyer et al. 2015; Seidl and Thomma 2017). This raises the possibility that the de-repression of TEs interacts with the expression of effectors.

Zymoseptoria tritici is the most important pathogen of wheat in Europe (Fones and Gurr 2015; Torriani et al. 2015). The pathogen's ability to infect host plants is largely determined by a complement of small proteins, most of them effectors, that manipulate the host physiology upon contact. Effector genes are frequently located in proximity to TEs and are highly up-regulated during early, stressful conditions of the host infection (Rudd et al. 2015; Palma-Guerrero et al. 2016; Haueisen et al. 2017; Palma-Guerrero et al. 2017; Fouché et al. 2018; Plissonneau et al. 2018). Effectors are thought to become upregulated by de-repression of facultative heterochromatin (Soyer et al. 2015; Soyer et al. 2019). Both facultative and obligate heterochromatin is highly enriched in TEs in *Z. tritici* (Schotanus et al. 2015). *Z. tritici* has a very plastic genome consisting of 13 core and up to eight accessory chromosomes that are not fixed within the species (Goodwin et al. 2011). Accessory chromosomes frequently undergo chromosomal rearrangements with breakpoints co-localized with TE insertions (Croll et al. 2013; Plissonneau et al. 2016; Hartmann et al. 2017; Plissonneau et al. 2018). Genes involved in pathogenicity and stress tolerance are frequently located in close proximity to TEs (Hartmann et al. 2017; Krishnan et al. 2018; Meile et al. 2018). Populations segregate over a thousand gene presence-absence polymorphisms and gene deletions are preferentially located in proximity to TEs (Hartmann et al. 2017; Plissonneau et al. 2018). Adaptation to specific wheat cultivars is governed by either the deletion or mutation of effector genes (Hartmann et al. 2017; Zhong et al. 2017; Meile et al. 2018). Importantly, some effector genes were shown to have undergone concurrent reductions in expression raising the possibility that the observed reconfigurations in TE content close to effectors made critical contributions to host adaptation.

In this study, we used transcriptome profiling to test for the impact of two major stress factors in the life-cycle of the pathogen on TE de-repression. We used a nutrient-rich culture medium as a non-stress environment and transferred the fungus to a nutrient-deprived medium that simulates starvation. Independently, we analyzed the fungal transcriptome at four distinct stages during the infection of wheat spanning the early symptomless stage, the peak of lesion formation and the saprotrophic stage. The early stages expose the pathogen to substantial nutrient and host defense stress factors. We replicated the two stress experiments with four genetically distinct strains of *Zymoseptoria tritici* to identify how the genetic background influences TE responsiveness. All strains have fully assembled genomes and have experimentally confirmed virulence differences. TEs showed the highest expression under nutrient stress, but the expression differed significantly between TE families and between genetic backgrounds. Infection stress led to a large number of TE families to be upregulated at the peak of the symptom development on wheat leaves. Next, we determined how the genomic location affected the expression of TEs and identified distinct de-repression patterns depending on the type of stress, the distance to the closest genes and the impact of genomic defense mechanisms. Finally, we analyzed a locus segregating variation at a key effector gene involved in host adaptation for the impact of TE de-repression. We show that the insertion of specific TEs led to silencing and in turn promoting virulence.

Results

TE landscape and transcriptomic response to stress conditions

We analyzed four strains of *Z. tritici* that differed significantly in the progression of infection and response to stress (Lendenmann et al. 2014; Lendenmann et al. 2016; Palma-Guerrero et al. 2017). The most virulent strain (3D7) developed visible symptoms within 12 days after infection (Palma-Guerrero et al. 2016). Strains 1A5 and 1E4 developed symptoms on average with a two-day delay and strain 3D1 showed the slowest symptom progression (Palma-Guerrero et al. 2017; Stewart et al. 2017). Each strain has a fully assembled and annotated genome (Plissonneau et al 2018) with similar percentages of TEs 16.0-18,1% (fig. 1A). Of the 111 TE families identified previously in the reference genome of the species (Grandaubert et al. 2015), all families were present in 1A5, 110 were identified in 1E4 and 3D1, and 108 were found in 3D7. LTR-*Gypsy* elements were the most abundant in all of the strains making up between 5-7% of the genomes, followed by LTR-*Copia* and LINE-*1* elements (fig. 1B). The TE content was highest in accessory chromosomes (14-21; fig. 1C). Chromosome 14 of strain 3D1 had the highest TE content (>40%). Despite the similarity in overall TE content between strains, TE superfamilies showed marked differences in their distribution across chromosomes (fig. 1C). SINE elements were only present on chromosome five for strain 1A5, 1E4 and 3D1 and on chromosome three for strain 3D7.

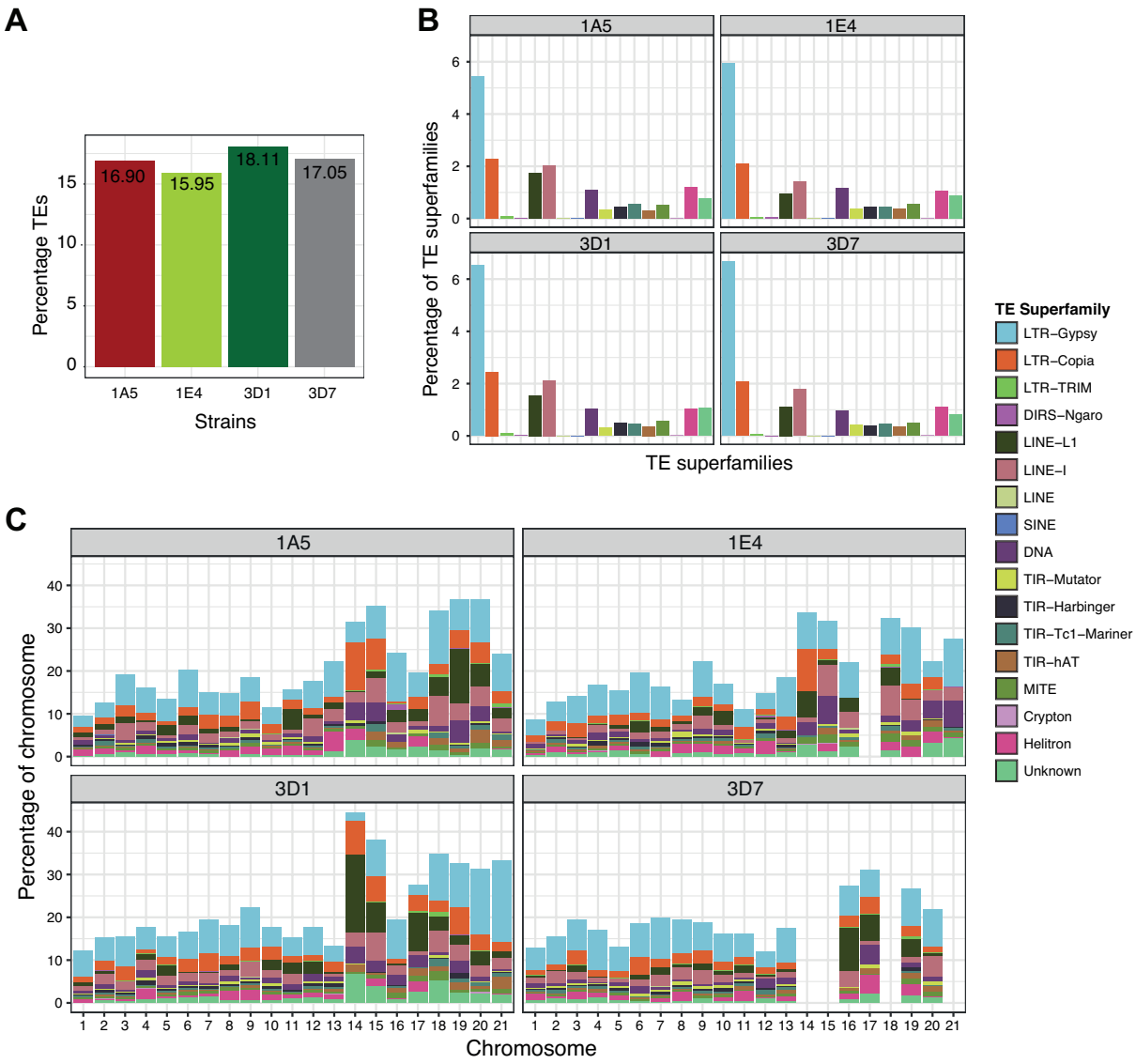


Figure 1: TE composition of the completely assembled genomes of *Zymoseptoria tritici*. (*A*) The percentage TE in each genome. (*B*) The percentage of TE superfamilies in each strain. (*C*) The distribution of TE superfamilies as a percentage of the chromosomes for each strain.

We analyzed the transcriptomic response to specific stress conditions by culturing the fungi first in nutrient-rich conditions, then analyzed the same strains growing in a minimal carbon source medium (*i.e.* starvation stress; fig. 2A). In parallel, we passaged all strains through an infection cycle on a wheat host (*i.e.* infection stress; fig. 2A). Infection stages were sampled at four time points (7, 12, 14 and 28 days post-infection). Across all conditions, we found that biological replicates clustered tightly together showing high stress reproducibility (supplementary fig. S2). Gene expression profiles clustered mainly according to condition

with early and late infection phases resembling nutrient starvation (fig. 2C). We analyzed the expression of putative virulence factors (*i.e.* effectors) and carbohydrate-active enzymes (CAZymes) in strain 3D7 in order to recapitulate the progression of the infection and impact of starvation. Overall, effector genes were upregulated during early infection stages (7-14 days post-infection; dpi) followed by downregulation at the final infection time point (supplementary fig. S3). CAZymes are enzymes that digest carbohydrates and digest the plant cell walls, releasing nutrients for the pathogen. CAZymes differed widely in expression profiles with subsets showing upregulation during early infection stages, in nutrient-rich conditions and nutrient starvation, respectively (supplementary fig. S4). Overall, stress conditions impose major gene expression profile changes consistent with the lifestyle transitions of the pathogen.

Differential stress response dynamics of TEs across environments

We analyzed TE expression across all conditions and strains using TETranscripts, which quantifies the expression abundance of a TE across all copies in the genome allowing for multiple read mapping. TE families differed substantially in expression profiles depending on the imposed stress condition (fig. 2B). A principal component analysis showed that the expression of TEs clustered according to the host genotype rather than stress condition (fig. 2D). Most TEs were expressed in most backgrounds and stress conditions (fig. 2E). The lowest percentage of expressed TEs was observed during early infection. We analyzed the relative expression of TEs versus genes expression and found that the highest relative TE expression occurred under nutrient-rich and starvation stress conditions (fig. 2F). The relative expression decreased with the progression of infection (fig. 2F).

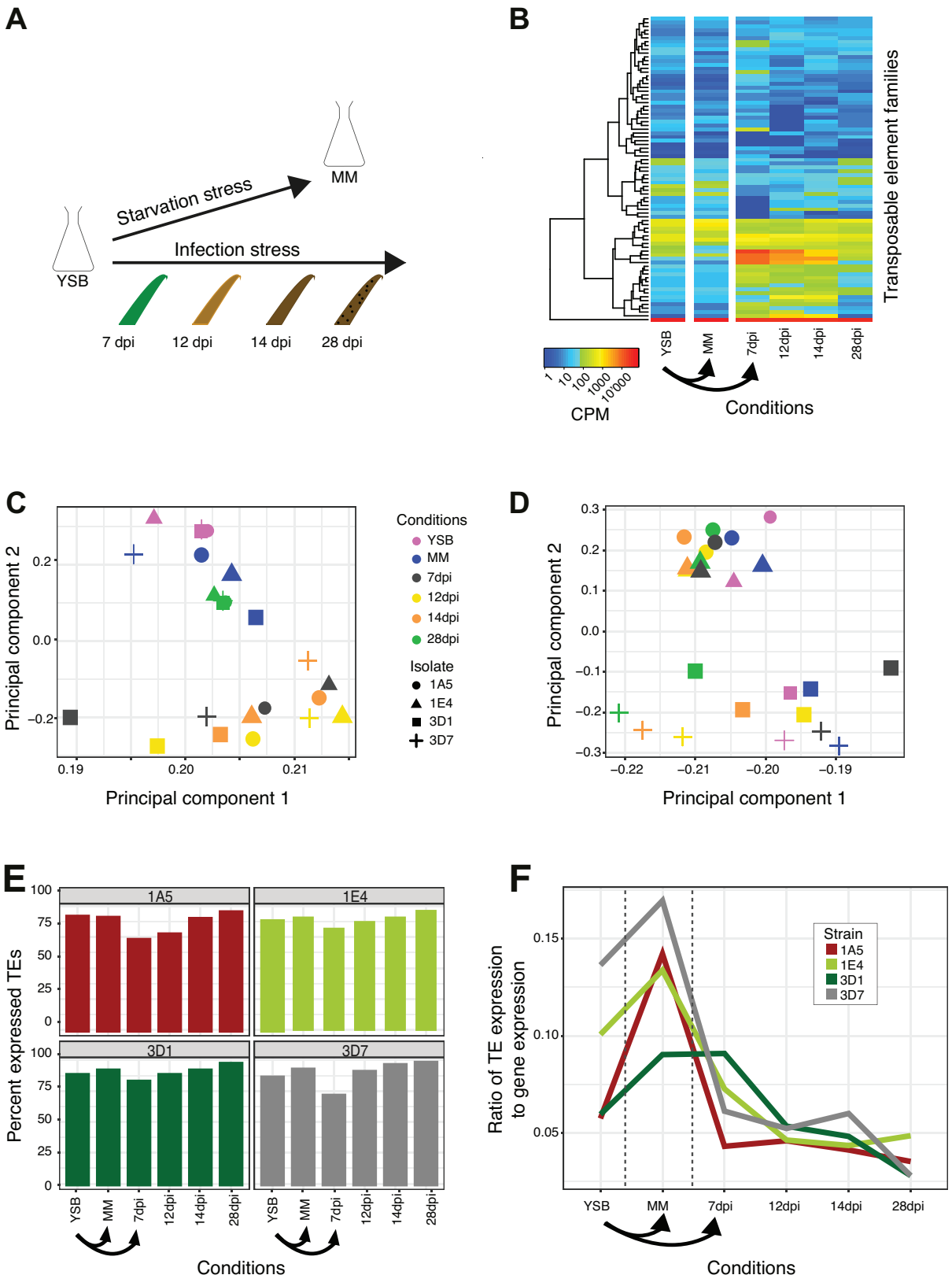


Figure 2: Responsiveness of TE to different stresses. (*A*) Transcriptomes analyzed in response to nutrient and infection stress. (*B*) Heatmap showing expression of TE families in the strain 3D1. Principal component analyses of gene expression (*C*) and TEs (*D*) in the four strains. (*E*) The percentage of expressed TEs compared with all TEs for each condition. (*F*) Ratio of TE expression to gene expression. MM, nutrient poor media; YSB, nutrient rich media.

The response of TEs to stress conditions was highly specific to individual TE families. In strain 3D7, two LTR-*Gypsy* element families and a TIR-*Tc1-mariner* element family, were only upregulated during early infection (7-14 days post-infection; supplementary fig. S5B). Similarly, in strain 3D1 four LTR-*Gypsy* element families were mainly upregulated early during infection (supplementary fig. S5A). Two of these upregulated families were shared between the strains (LTR-*Gypsy* element families 6 and 9) and are the most infection stress responsive elements. In 1A5 and 1E4, a shared TE element family (TIR-*hAT* element 1) was most highly expressed during starvation and mostly repressed during infection (supplementary fig. S5C and D). Some TE families showed consistently high expression across all conditions suggesting generally weak genomic defenses against expression of this specific TE in comparison to other TEs that are only responsive to specific stress conditions (supplementary fig. S5A-D). TE expression in all four strains was dominated by a MITE-Undine family (fig. 3), which is the most highly upregulated TE under nutrient starvation stress in all strains. The exception is strain 3D7 where the family was similarly expressed under nutrient rich conditions and starvation stress (fig. 3A). MITE-Undine is a non-autonomous element lacking coding regions. We were unable to identify the helper autonomous element with the same terminal-inverted repeats (TIR). MITE-Undine was also the most abundant element in any of the four genomes (fig. 3B) with a copy number of 250-296. The mean number of TE copies per family in each genome was 29-32. The average distance of MITE-Undine to the nearest gene was 17.6-33.8 kb compared to 19.5-21.1 kb for all TEs (supplementary table S2). The element was present on all chromosomes (fig. 3C for isolate 3D1) and contains target site sequences, TIR and low-complexity regions (palindromes and tandem repeats; fig. 3D, supplementary file S1). We found no evidence for the element in the genomes of the most closely related *Z. pseudotritici* and the more distantly related *Z. brevis*. However, *Z. ardabiliae* which has an intermediate divergence time from *Z. tritici* harbors eight copies of the MITE-Undine.

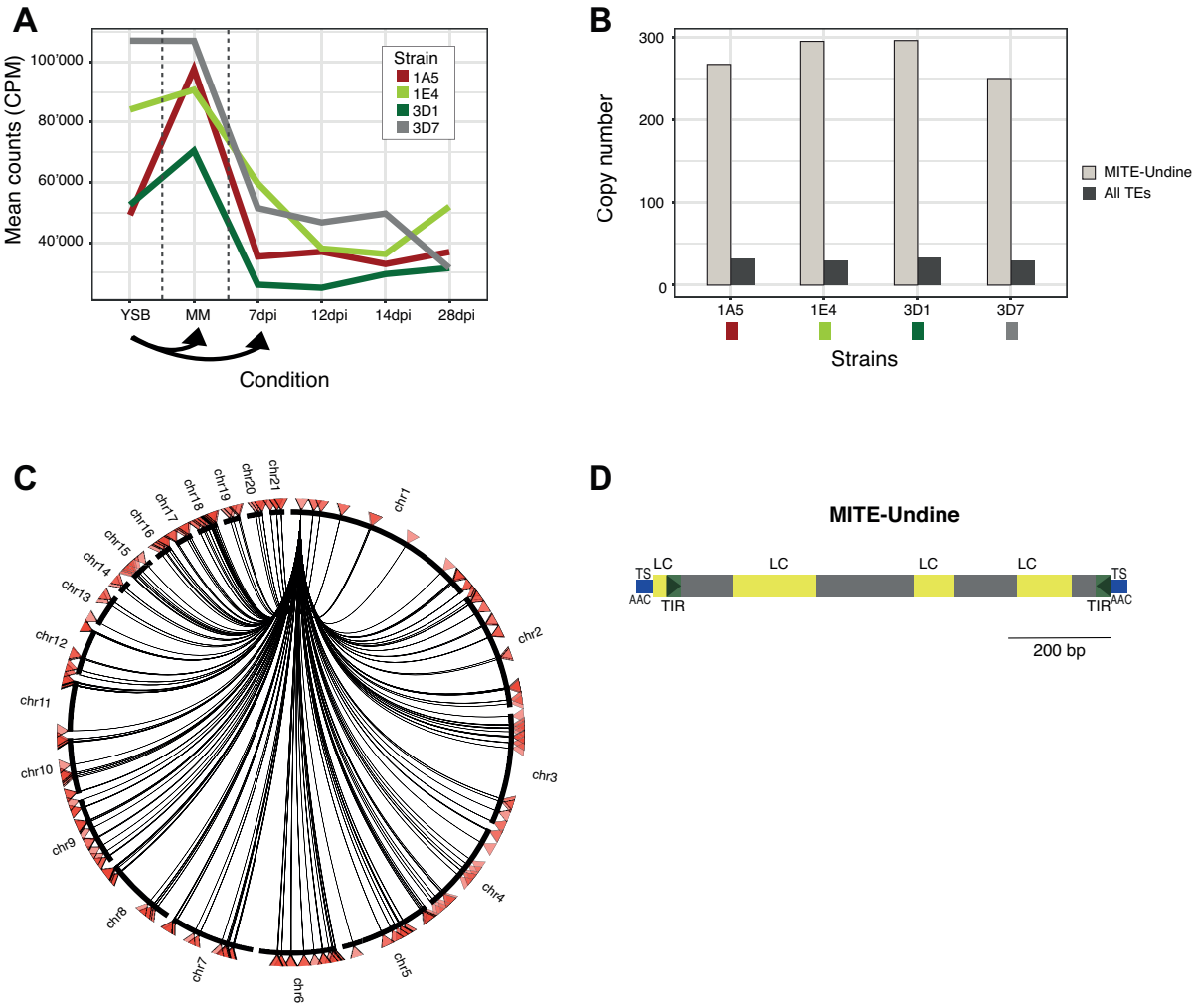


Figure 3: Characterization of the high-copy and highly expressed MITE-Undine element. (*A*) The expression of MITE-Undine in CPM in all four strains. (*B*) The copy number of MITE-Undine compared with the mean copy number of all TEs. (*C*) Identification of MITE-Undine copies in the genome of 3D1. (*D*) Schematic of MITE-Undine with target site duplications (TS), TIR, and low complexity regions (LC). MM, nutrient poor media; YSB, nutrient rich media.

TE and gene expression dynamics as a function of the genomic localization

To address the impact of physical proximity to TEs on gene expression patterns, we analyzed gene expression across stress conditions as a function of distance to the closest TE. Genes within 1 kb upstream or downstream of TEs were upregulated early during infection (7-14 dpi depending on the strain) compared to genes >1 kb away from TEs (fig. 4A and supplementary fig. S6). Genes >1 kb away from TEs showed higher expression than genes <1 kb of TEs late in the infection (28 dpi). Genes with TE insertions showed

consistently low levels of expression. Effector genes were overall closer to TEs than CAZymes, genes encoding secreted proteins or genes overall (fig. 4B and supplementary fig. S7). Consistent with the proximity to TEs, effector genes were strongly upregulated early during infection compared to other gene categories (fig. 4C and supplementary fig. S8). Notably, the increase occurred first for 3D7, the strain with the most rapid infection progression (Palma-Guerrero et al. 2017) (supplementary fig. S8). The increase occurred at 12 dpi and peaked at 14 dpi for the other three strains (supplementary fig. S8).

Next, we analyzed TE expression responses to stress as a function of the distance to the closest genes. TE expression generally peaked for TEs at a mean distance of 15-20 kb to the closest gene for 3D1 and 3D7 and at 20-45kb away from the closest gene for 1A5 and 1E4 (fig. 4D and E and supplementary fig. S9). In strain 3D1 (fig. 4D) and 1E4 (supplementary fig. S9B), TEs with a mean distance of within 1 kb of genes were upregulated during early infection. In 3D1, the upregulation at 7-14 dpi was primarily due to the expression of an LTR-Gypsy element, but also due to the expression of a LTR-Copia element family at 14 dpi (fig. 4D). In 1E4, TEs with a mean distance within 1 kb of genes were also upregulated under nutrient stress, primarily due to the expression of an unknown element and an LTR-Gypsy element (supplementary fig. S9B). In 3D7 and 1A5, TE families further away from genes were more strongly expressed (fig. 4E and supplementary fig. S9D). Exceptions in 1A5 include TE families with a mean distance of 1-2 kb to the nearest gene, which were upregulated under all of the conditions due to the expression of two unknown element families and two LTR-Gypsy families (supplementary fig. S9D). TE families with a mean distance of >45 kb away from the nearest gene showed upregulation under all conditions in strain 3D7, with a major peak at 12dpi (fig 4E). The most highly expressed families falling in this category were an LTR-Copia family at mean distance of 45-75 kb and MITE-Undine at 75-95 kb from the nearest gene.

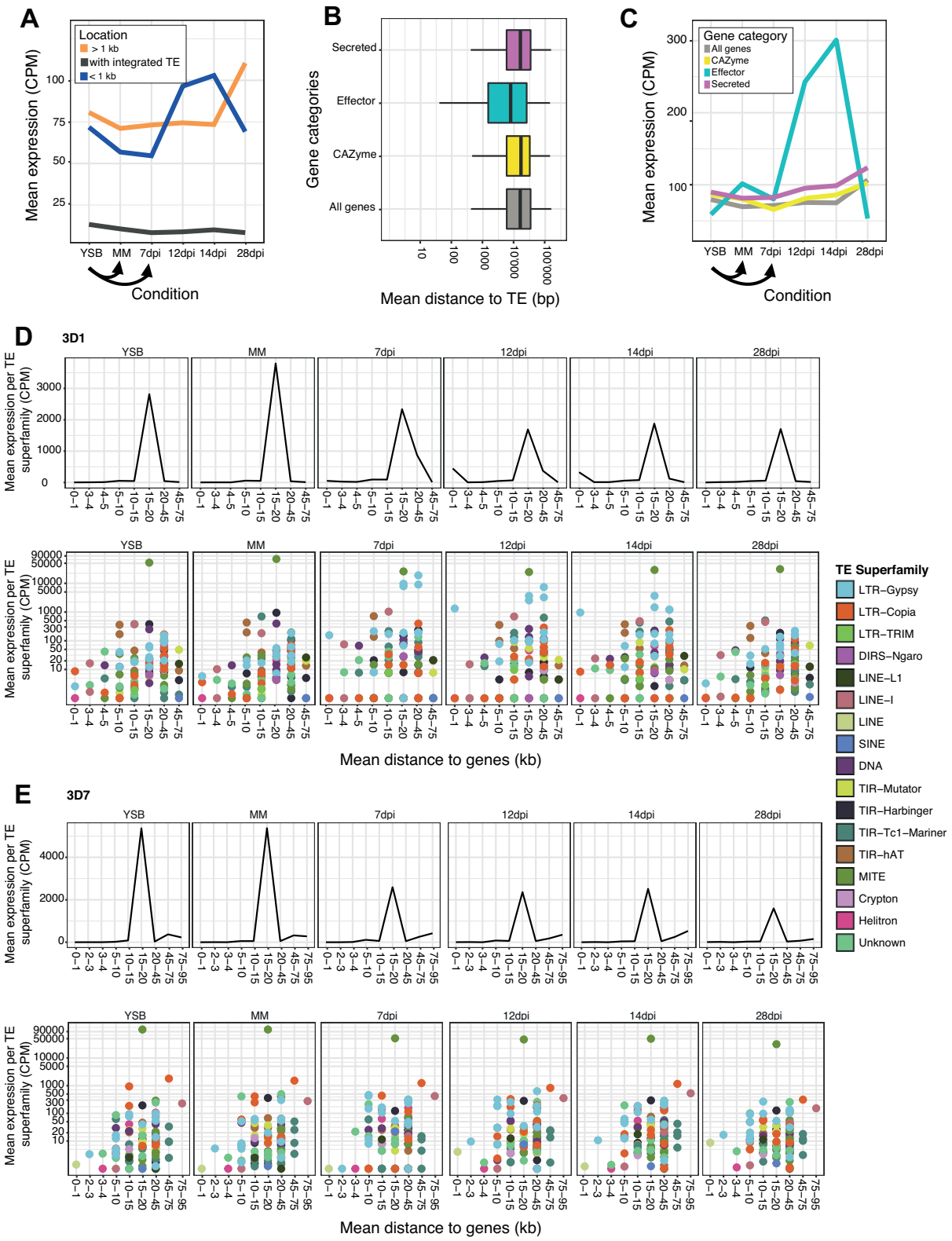


Figure 4: Gene expression as a function of proximity to TEs in strain 3D1. (*A*) Expression of genes with inserted TEs, within 1 kb of the nearest TE or more than 1 kb from the nearest TE. (*B*) Mean distance between genes grouped by functional category. (*C*) Mean expression of genes grouped by functional category. TE expression responses to stress as a function of the distance to the closest genes in 3D1 (*D*) and 3D7 (*E*). Distance segments lacking TEs were omitted. MM, nutrient poor media; YSB, nutrient rich media.

Co-expression networks of genes and TEs across stress conditions

Many TEs in the *Z. tritici* genome are in close physical proximity to genes and may, hence, converge on similar epigenetic de-repression dynamics across stress conditions. To infer synchronicity of TE and gene expression, we performed clustering analyses to define profiles of TE and gene co-regulation under stress. The analysis identified a total of twenty co-expression profiles, of which six were shared by all four strains and six were strain-specific. Eighteen co-expression profiles contained TEs, but were not identified in all isolates (fig. 5A). Expression profiles included different kinetics of upregulation upon infection stress (see profiles 16-17-18-27-28-29-31-39) but also downregulation (see profiles 8-9-13-15-19-20-21-32-35) with various intermediary profiles (fig. 5A). Co-expression profiles included on average >98% of genes and up to 5% of TEs in each genome (fig. 5A). To infer the biological relevance of different co-expression clusters, we performed enrichment analyses of gene ontology (GO) terms. In total, 11, 7, 14 and 12 co-expression profiles showed significant enrichment for GO terms in strains 1A5, 1E4, 3D1 and 3D7, respectively (p-value < 0.05). Four co-expression profiles were found consistently enriched for GO terms in the four strains (profiles 8, 18, 32 and 36; fig. 5A; supplementary table S3). These profiles included enrichment for hydrolase, phosphorylation and transcriptional activity, as well as carbohydrate metabolism, kinases, and DNA replication functions (fig. 5A; supplementary table S3).

Both DNA and retrotransposons were co-expressed with genes. LTR-*Gypsy* elements were dominating co-expression profiles reflecting the abundance of the elements in the genomes. Among all four strains, the co-expression profiles displaying higher numbers of TE superfamilies were consistently those with a peak of expression early in the infection process (*e.g*. profile 18). TE and gene co-regulation could be driven by shared epigenetic

environments due to physical proximity and/or transcriptional leakage. To test for the effect of physical proximity, we analyzed the physical distance between co-expressed TEs and their closest co-expressed genes. In concordance with the previous global analysis, co-expressed genes with a peak of expression early upon infection are found closer to TEs. However, the closest distance between co-expressed genes and TEs within an expression profile is on average ten times longer than the distance of the closest co-expressed genes and TEs not in the same expression profile (fig. 5B). Therefore, TEs and co-expressed genes are not closer than TEs and genes that do not share an expression profile.

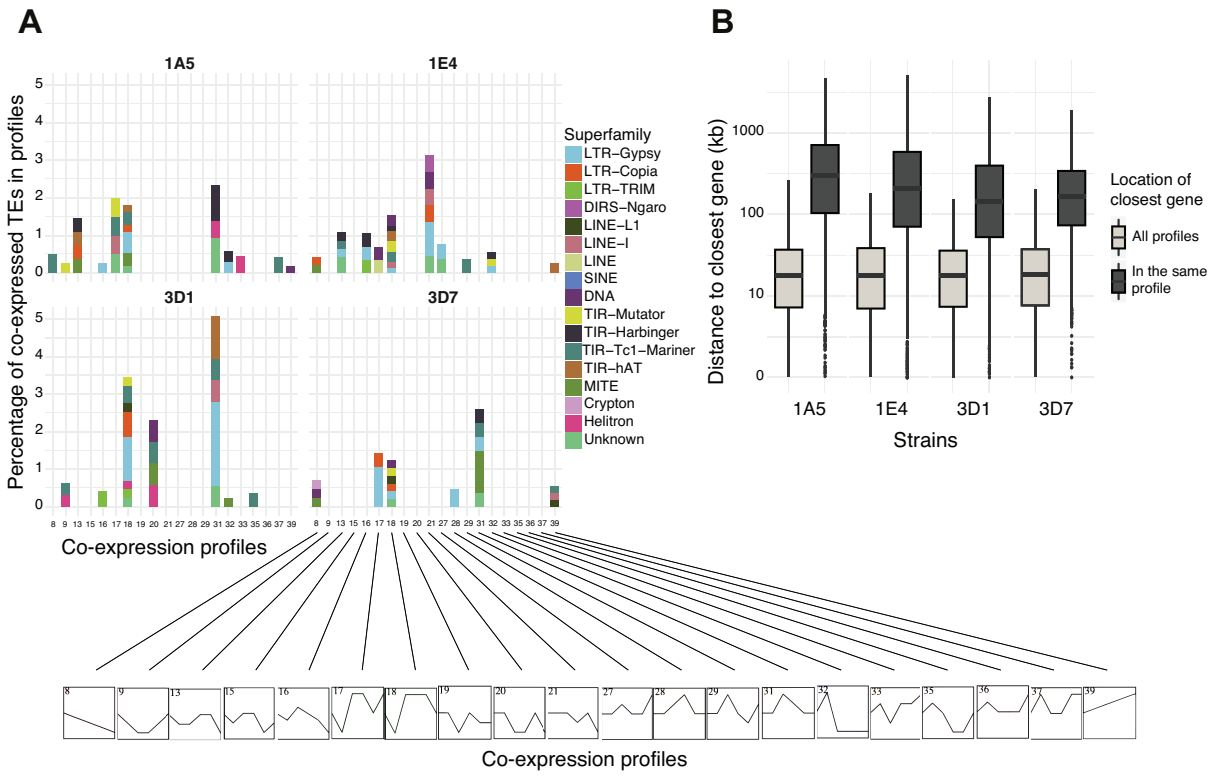


Figure 5: Co-expression profiles of TEs and genes. (*A*) The percentage of TE families in each co-expression profile. (*B*) The distribution between TEs and co-expressed genes. TEs and co-expressed genes are not closer than TEs and genes that do not share an expression profile.

Impact of genomic defenses on TE expression under stress

Fungi evolved sophisticated genomic defenses that inactivate TE copies through the introduction of RIP mutations (Selker 2002). In order to determine how RIP may impact TE

expression under stress, we analyzed mutational biases among genomic TE copies. Most TE families in all four genomes were affected by RIP (fig. 6A and supplementary fig. S11). Only TE families in the LTR-TRIM superfamily and one family in the TIR-*Tc1-mariner* superfamily were not affected by RIP. In all strains, TEs in the TRIM family were among the most highly affected TEs. In 1A5, a family belonging to the TIR-*Tc1-mariner* superfamily is consistently expressed under all conditions (supplementary fig. S10A). Similarly, LTR-*Gypsy* elements with strong RIP signatures were upregulated upon infection in the strains 1A5, 1E4 and 3D1 (supplementary fig. S11A). Hence, most TE families affected by RIP are still expressed under at least some stress conditions.

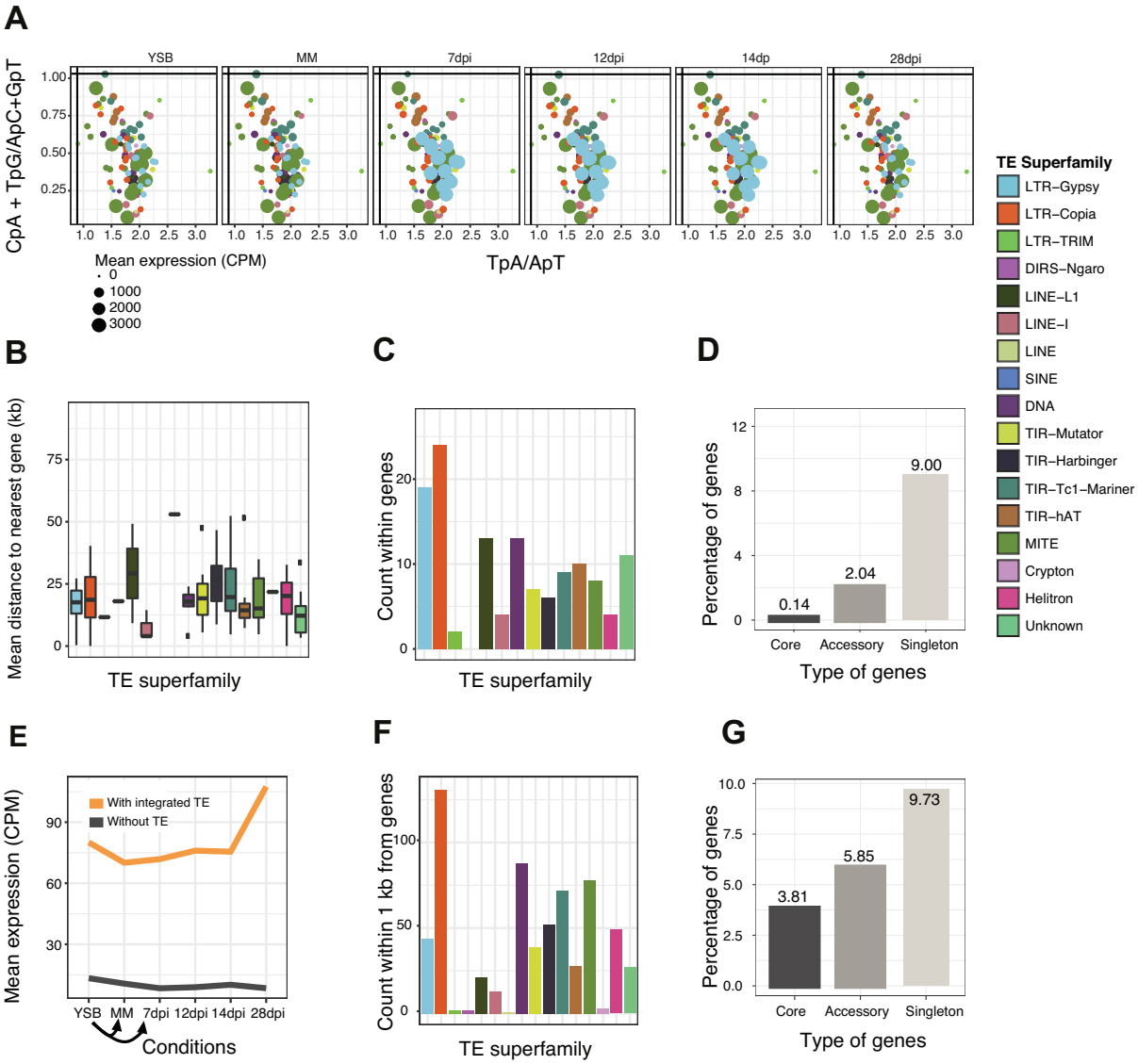


Figure 6: Genomic defenses and TE expression. (*A*) RIP indices for each TE family and mean expression of the family under all stress conditions for strain 3D1. Vertical and horizontal lines represent commonly used thresholds to detect RIP (Hane and Oliver 2008). Colors indicate the superfamily and size the expression at the family level in CPM. (*B*) The mean interval of TE superfamilies to the nearest gene. (*C*) TE superfamilies within genes in 3D1. (*D*) Gene categories with inserted TEs in 3D1. The percentages are given as the number of genes of that category (core accessory or strain specific) within genes as a fraction of the total number of genes in a category. (*E*) Expression of genes with or without inserted TEs. (*F*) TE superfamilies within 1 kb of the closest gene in strain 3D1. (*G*) Gene categories within 1 kb of TEs. The percentages are given as the number of genes of that category (core, accessory or strain specific) as a fraction of the total number of genes in a category. MM, nutrient poor media; YSB, nutrient rich media.

TE insertion dynamics in proximity to genes

TE superfamilies showed substantial variation in their mean distance to the closest gene with most having a mean distance to the nearest gene of less than 25 kb (fig. 6B and supplementary fig S12). The closest TE superfamilies to genes were DIRS-*Ngaro* (2867 bp) in 1A5, LINE (415 bp) in 1E4, LINE-*1* (7457 bp) in 3D1 and LINE (832 bp) in 3D7 (supplementary table S4). Next, we analyzed coding sequence disruptions across the genome and LTR-*Copia* elements were the most frequently inserted TEs into genes in all strains except 1A5 (fig. 6C; supplementary fig. S12A; supplementary table S5). Singleton genes defined as present in only one of the four strains were the most frequently disrupted genes (fig. 6D and supplementary fig. S12B). Genes with inserted TEs had a lower expression than genes without TEs (fig. 6E and supplementary fig. S12E). Hereafter, we analysed TE insertions in close proximity to genes. We found again that LTR-*Copia* elements were the most abundant elements with 108-132 copies (fig. 6F; supplementary fig. S12C; supplementary table S6). Singleton genes most frequently had an integrated TE or were located within 1 kb from a TE, followed by accessory genes and core genes (fig. 6D and G and supplementary fig. S12B and D).

Impact of TE de-repression dynamics on virulence

Effectors are frequently among the closest genes to TEs and play a key role in virulence. To dissect the role of TE de-repression in virulence, we first analyzed all 1381 predicted effector orthologs across the four genomes (Plissonneau et al. 2018). We found that 320 effector orthologs were within 1 kb of a TE and 447 were within 2 kb of a TE. Effector genes with a TE insertion within 1 kb showed higher expression at the peak of symptom development on the host (12-14 dpi) compared to other effectors (fig. 7, supplementary fig. S13). This suggests that effector genes sharing genomic compartments with TEs benefit from the epigenetic landscape to optimize up-regulation during the critical period of infection.

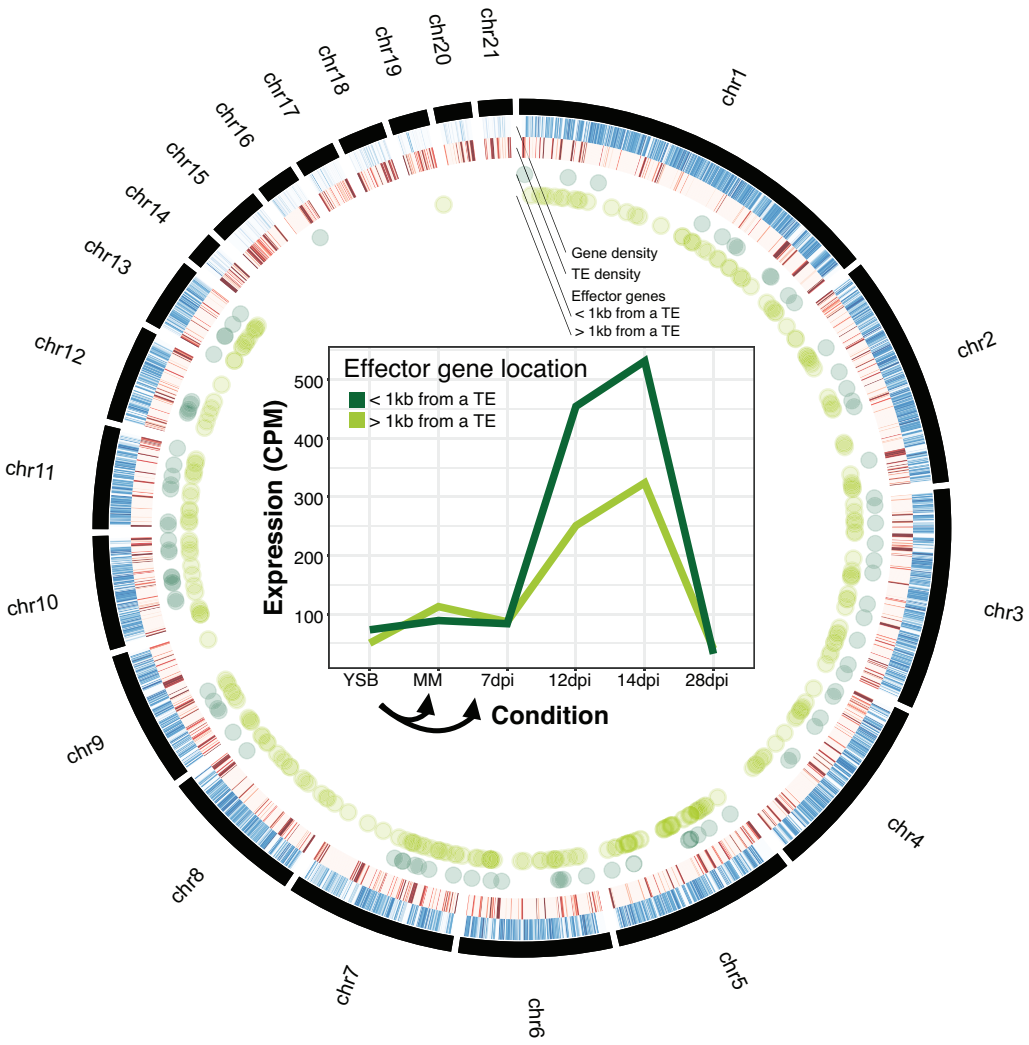


Figure 7: Effector gene expression according to the presence or absence of nearby TEs. Circular representation of the 3D1 genome with gene and TE density in 10-kb windows, as well as the position of predicted effectors. MM, nutrient poor media; YSB, nutrient rich media.

Given the synchronicity in the expression of effectors and de-repression of TEs over the course of a plant infection, we investigated potential causal links between TE de-repression and virulence. Mapping populations generated for two pairs of the four strains analyzed here revealed a major effect locus in each of the strain pairings (Zhong et al. 2017; Stewart et al. 2018). In progeny populations of the cross 3D1x3D7, the locus on chromosome 7 encoding the effector Avr3D1 explains nearly all variation in virulence on the wheat cultivar Runal (Meile et al. 2018). Analysis of deletion and ectopic insertion mutants revealed that the strain

3D7 carries 12 amino acid substitutions and one indel in Avr3D1 and at least a subset of these mutations are critical for successfully avoiding recognition and infecting the host (Meile et al. 2018). Interestingly, *Avr3D1* shows regulatory variation, which may also contribute to differences in virulence. Indeed, *Avr3D1* shows much stronger but delayed expression in the avirulent strain 3D1 compared to earlier and lower expression in the virulent strain 3D7 (fig. 8B).

In *Z. tritici*, heterochromatin remodeling plays a major role in effector expression during the switch to necrotrophy (*i.e.* the appearance of lesions) (Soyer et al. 2019). The necrotrophic infection period corresponds to the peak expression of *Avr3D1* in both 3D1 and 3D7 (fig. 8B). Hence, we investigated evidence for epigenetic remodeling of the locus driven by TEs. *Avr3D1* is located at the boundary of a gene-rich region in the avirulent strain 3D1 (fig. 8A). We used uniquely mapped RNAseq reads to assess expression variation at the level of individual TE copies. The closest TE to *Avr3D1* in 3D1 is the TIR-*Mutator* element 2 at 12.3 kb. This TIR-*Mutator* copy next to *Avr3D1* shows expression nearly exclusively at 12 dpi while other copies in the genome were mostly expressed under different conditions and infection stages (fig. 8A, supplementary fig. S14A). The unknown TE element 8 has three copies close to *Avr3D1* showing similar expression profiles (fig. 8A, supplementary fig. S14B, C and E). The second copy is silenced at 12 dpi but is most expressed at 14 dpi. This is in contrast to the other copies outside of the locus showing the opposite expression profile across conditions. Hence, nearby TEs show expression profiles matching the *Avr3D1* expression in the 3D1 strain across infection stages.

Next, we analyzed how the epigenetic landscape of TEs evolved in the strain 3D7. This strain gained virulence on the cultivar Runal and the effector is expressed earlier during infection (7-12 dpi; fig. 8B). The Avr3D1 locus experienced a drastic reconfiguration with the

insertion of two large TE clusters (Meile et al. 2018). The closest TE, a copy of a Crypton element 1 as well as TE copies in the same cluster were only expressed at 14 dpi but silenced during the peak of symptom developments of the strain 3D7 (fig. 8A, supplementary fig. S15). Other copies of TEs present near Avr3D1 showed variable de-repression patterns peaking either under nutrient starvation (MM) or during the late infection stage (28 dpi). These atypical de-repression profiles are characterized by a resilience to de-repression during early stages of host infection. Interestingly, the large TE cluster inserted in 3D7 impacted a boundary region of euchromatin and facultative heterochromatin in the reference genome of IPO323 (fig. 8C). The presence of large TE clusters leads to obligate heterochromatin and strong silencing (Schotanus et al. 2015). This would be consistent with the TE silencing observed near Avr3D1 in the 3D7 genome. The Avr3D1 locus is furthermore located at a major epigenetic boundary region of chromosome 7 splitting off a chromosomal arm with nearly uniform H3K27m3 facultative heterochromatin (fig. 8C) (Schotanus et al. 2015). Taken together, our analyses show that the reduced expression of Avr3D1 in the virulent strain 3D7 is most consistently explained by the presence of strongly silenced TEs, which do not respond to stress triggers caused by host infection.

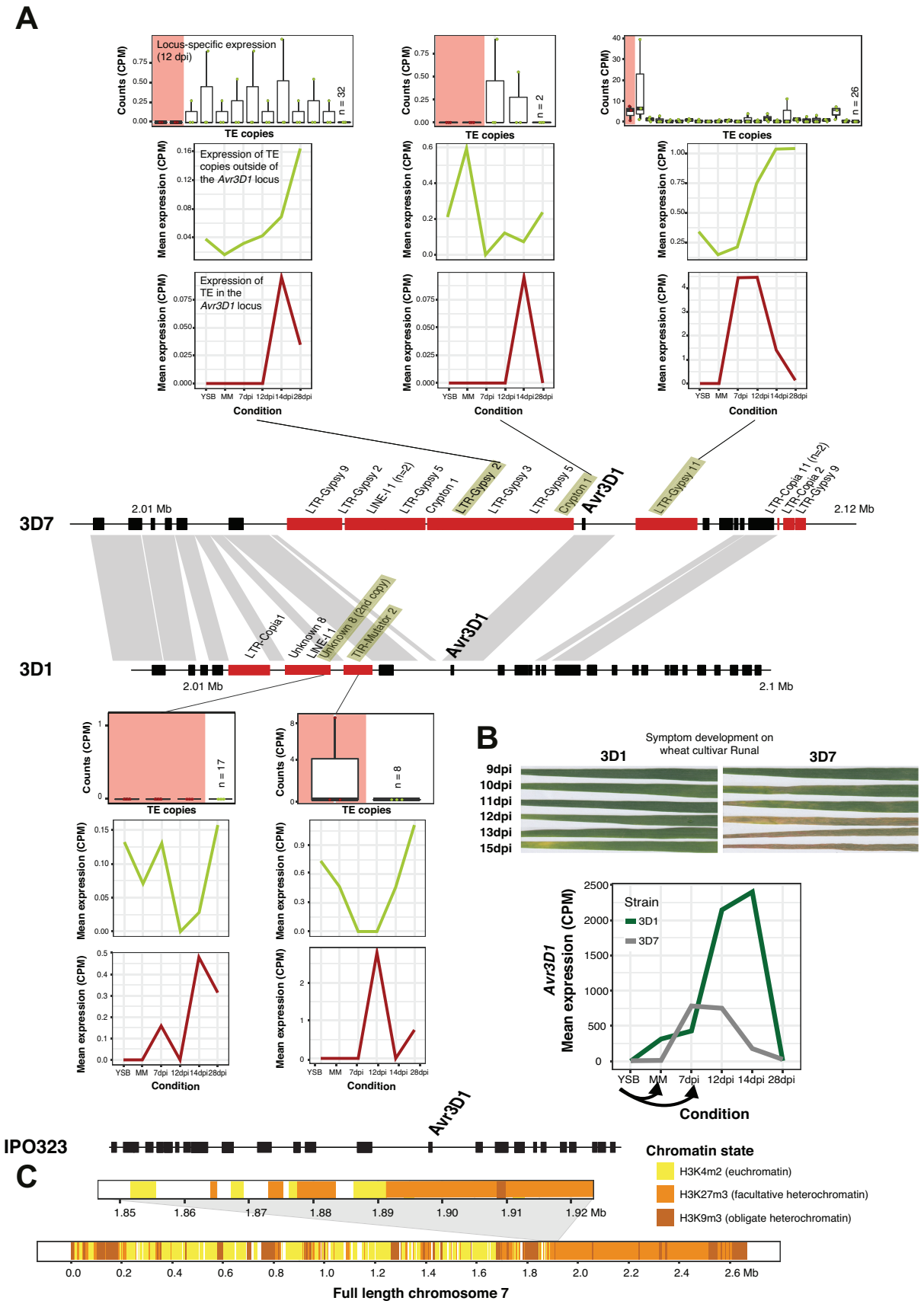


Figure 8: TE de-repression of individual TEs inserted in proximity to the effector gene *Avr3D1*. (*A*) Synteny plots of the chromosomal regions encoding Avr3D1 in strains 3D7, 3D1, and IPO323. Genes are shown in black and inserted TEs in red. Expression levels of five TEs in close proximity

of *Avr3D1* are shown in more detail. The top plot shows expression levels at 12 days post-infection (dpi) of all identified copies across the genome using uniquely mapped reads (expression levels of some copies were summarized). Variation is expressed by showing uniquely mapped reads from each individual replicate. The TE copy highlighted in red corresponds to the copy inserted nearby *Avr3D1.* The middle plot shows averaged expression levels of all TE copies outside of the *Avr3D1* locus. The bottom plot shows the expression levels of the TE copy found nearby *Avr3D1.* (*B*) Disease progress and symptom development by 3D1 and 3D7 infecting the wheat cultivar Runal (Meile et al., 2018). The expression variation of *Avr3D1* is shown below in wheat cultivar Drifter. (*C*) Histone methylation marks assessed for the reference genome IPO323 (Schotanus et al. 2015). Both the region of *Avr3D1* as well as the entire chromosome 7 are shown. MM, nutrient poor media; YSB, nutrient rich media.

Discussion

TEs are major drivers of genome evolution due to their transpositional activity. Repression of TEs is largely governed through epigenetic control and is, hence, susceptible to external stress. Using transcriptome profiling, we show that two distinct stress conditions induce the expression of distinct sets of TEs. By replicating the analyses across four genetic backgrounds, we show that the major expression dynamics of TEs are conserved. However, some of the most highly expressed TEs including MITE and LTR-*Gypsy* elements showed highly distinct de-repression across stress conditions. The genomic context of TEs was a major predictor of de-repression dynamics during stress. Consistent with TE de-repression being governed by epigenetic effects, we found that gene expression profiles under stress varied significantly depending on the proximity to the closest TEs. The evolution of virulence was most likely due to TE-driven epigenetic reconfigurations impacting expression profiles across a major effector locus encoding Avr3D1.

Stress-dependent TE de-repression dynamics

The completely assembled genomes of *Z. tritici* display substantial variability in chromosome-level TE content despite highly similar overall repetitive element proportions. The TE content variation is striking given the fact that all four strains were collected from nearby wheat fields, interfertile and from populations with a rapid decay in linkage disequilibrium (Croll et al. 2015). LTR-*Gypsy* were the most abundant elements consistent with their abundance in many other fungal genomes (Muszewska et al. 2011). Members of the LTR-*Gypsy* superfamily in conjunction with a MITE showed among the strongest de-repression under stress. MITEs and LTR-retrotransposons are also most frequently associated with stress responsiveness in other organisms (Yasuda et al. 2013; Negi et al. 2016). However, the impact of stress on TE expression is highly variable among TE families,

copies and species. Some TEs are expressed and potentially mobilized in response to stress, while other TEs are suppressed after an initial stress-induced activation, and some TEs are downregulated in response to stress (Horváth et al. 2017).

Nutrient starvation and host infection constitute the major stress factors in the life cycle of filamentous plant pathogens (Ferreira et al. 2006; Hernández-Chávez et al. 2017). We exposed *Z. tritici* to two stress conditions. Growth in a carbon source depleted culture medium (MM) exposed the fungus to nutrient starvation. Early infection stages induce stress due to host immune responses targeted at the pathogen and imposes growth under limited nutrient conditions. Carbohydrate-active enzymes (CAZymes) showed highly distinct profiles depending on the stress condition. Hence, starvation and infection stress have distinct impacts on gene expression consistent with the biological context of the stress (Palma-Guerrero et al. 2016). The majority of TE families showed some degree of de-repression in at least one of the stress conditions. Elements in the LTR-*Gypsy* superfamily were upregulated during early infection, which corresponds to the most stressful period on the host (Rudd et al. 2015; Palma-Guerrero et al. 2016). Infection stress first causes the upregulation of effector genes and later cell-wall degrading enzymes (Skibbe et al. 2010; Wang et al. 2011; Kleemann et al. 2012; Hacquard et al. 2013; Soyer et al. 2014; Palma-Guerrero et al. 2016). Effector gene expression is known to be epigenetically regulated in plant pathogens and timed to maximize exploitation of the host (Qutob et al. 2013; Schotanus et al. 2015; Sánchez-Vallet et al. 2018; Soyer et al. 2019). We found that a different set of TEs showed the highest expression under starvation stress including a MITE, which was the most strongly expressed TE in the genomes. The regulatory framework governing stress responses is largely unknown in *Z. tritici,* but distinct epigenetic regulation in response to stress is likely playing a key role (Schotanus et al. 2015; Soyer et al. 2019).

TE and gene co-expression dependent on the genomic environment

Genes and TEs in close physical proximity likely undergo joint epigenetic regulation in response to stress. We found that genes close to TEs were upregulated during early infection consistent with the de-repression observed for TEs. LTR-*Copia* elements were the most frequently found TEs close to genes and one LTR-*Copia* family showed upregulation during early infection. In contrast, genes far from TEs were upregulated towards the end of the infection cycle, which is after the transition to a less stressful saprophytic lifestyle (Palma-Guerrero et al. 2016). Interestingly, we found no association of gene-to-TE distance with expression under nutrient starvation stress. This distinction may be due to the fact that epigenetic control of TEs is less pronounced under nutrient starvation stress. To understand TE de-repression dynamics as a function of the genomic environment, we also analyzed mean distances of TE families to the closest genes. Due to the repetitive nature of TEs, most transcriptome-derived short sequences cannot reliably be assigned to a single TE copy. Hence, our distance analyses were performed using summary statistics per TE family and not per individual TE copy. Copies of the most highly expressed TE, a MITE-Undine, are 17.6-33.8 kb away from genes across all genetic backgrounds. This activation could still be affecting the expression of genes as was shown for the Hopscotch TE in maize. This TE influences the expression of the TB1 locus at a distance of ~60kb (Studer et al. 2011).

Based on our co-expression clustering analyses, we found that TEs were not physically closer to co-expressed genes than other genes, suggesting that co-regulation is occurring in *trans* rather than in *cis*. Alternatively, this may reflect the epigenetic landscape of the genome with a multitude of distal chromosomal regions showing concerted de-repression dynamics. Interestingly, in other fungi such as *Coccidioides* (Kirkland et al. 2018) and *Pleurotus* (Borgognone et al. 2018) species, genes within 1 kb of some TE families were more repressed than genes overall and these genes were enriched for kinase function in

Coccidioides species. In other organisms, the influence of a TE on nearby genes is largely determined by the chromatin state of the TE (Saze and Kakutani 2007; Martin et al. 2009; Zeng and Cheng 2014; Lei et al. 2015; Williams et al. 2015; Williams et al. 2015; Hirsch and Springer 2017). This is most evident for stress responsive genes that carry TE insertions in the promoter sequences leading to upregulation upon TE demethylation (Le et al. 2014). Whether TE silencing through chromatin modification can spread to adjacent genes is not well understood (Sienski et al. 2012; Le Thomas et al. 2013; Lee 2015). TE stress responsiveness can be governed by epigenetic de-repression (Slotkin and Martienssen 2007) or the loss of a repressive mechanism under stress (Van Meter et al. 2014). If all TEs showed a correlated response to stress, this would suggest that activation is mostly due to epigenetic effects alone. However, different TE families show different stress responsiveness according to the stress condition and the host genotype, suggesting distinct epigenetic environments between the strains.

The role of the genetic background in TE expression dynamics

The set of four completely assembled con-specific genomes enabled us to analyze de-repression dynamics of the same TEs across different genetic backgrounds. We found that TE family expression differed between the strains indicating that the genetic background plays a role in the ability of TEs to respond to stress. Several LTR-*Gypsy* elements were upregulated early during infection in one genetic background but not in all. TE families in close proximity to genes were upregulated during early infection in strains 3D1 and 1E4, while families with the longest average distance to genes were upregulated during infection in strain 3D7. Our evidence for TE family expression by genetic background interactions suggests high degrees of polymorphism for TE control within the species. In *A. thaliana,* TE responsiveness to cold stress was found to differ among ecotypes and this was largely explained by differences in the genomic locations of specific TEs (Barah et al. 2013). Such

variation in the ability to control TE expression provides selectable genetic variation for the host genome to evolve more efficient control mechanisms.

Stress de-repression of TEs and the evolution of TE control mechanisms

De-repressed TEs are mutagenic (Le Rouzic and Capy 2005) and can lead to genome expansions (Lonnig and Saedler 2002). Hence, host genomes evolved to suppress TE proliferation. Stress responsiveness of TE families is indicative of the ability by the host to control proliferation. The MITE with the highest expression is consistently expressed in all conditions, suggesting that the host genome has not yet evolved effective control mechanisms. RIP is a genomic defense mechanism that hypermutates duplicated DNA sequences in fungi and counteracts TE proliferation (Selker 2002). We found that TE families with signatures of RIP were still responsive to stress. In particular, the highly responsive MITE and LTR-*Gypsy* elements under starvation and infection stress, respectively, display strong signatures of RIP. This suggests that point mutations introduced by RIP may well introduce loss-of-function mutations and disable *e.g.* transposase functions. However, RIP in *Z. tritici* seems ineffective at preventing TE de-repression under stress.

Pathogens of plants are exposed to unique stress conditions upon entering their host. The challenges mounted by the plant immune system are designed to effectively contain a pathogen's deployment of its infection program. Specialized pathogens evolved to time the expression of pathogenicity factors with the onset of stress by localizing the underlying genes in epigenetically silenced chromosomal regions (Sánchez-Vallet et al. 2018). We show here that the co-localization of epigenetically silenced TEs and effector genes can underlie major adaptations to successfully circumvent detection by the host. While the localization of pathogenicity factors in epigenetically silenced regions is most likely adaptive for the pathogen, the localization of TEs in the same compartment is likely only adaptive in

absence of stress. Hence, the co-localization of pathogenicity factors and TEs creates a complex selection regime on the pathogen. Selection for more effective TE control under infection stress may actually be deleterious for the coordinated gene expression during infection. We identified unexpected complexity in both the genomic localization of TEs across genetic backgrounds and in the TEs response to stress. This suggests that there is standing variation for the ability to control TEs within the species. Hence, host genomes and TEs may be engaged in rapid co-evolutionary arms races to maintain effective control and escape repression, respectively.

Materials and methods

Strains and growth conditions

Strains 1A5, 1E4, 3D1 and 3D7 were isolated from two fields in Switzerland in 1999 and have been phenotypically and genotypically characterized (Zhan et al. 2005; Croll et al. 2013). These strains were used in previous studies for quantitative loci mapping (Lendenmann et al. 2014; Stewart and McDonald 2014; Lendenmann et al. 2016; Stewart et al. 2017). The genomes of all four strains have been sequenced and assembled into complete chromosomes using high-coverage PacBio sequencing (Plissonneau et al. 2016; Plissonneau et al. 2018). High-density genetic maps (Lendenmann et al. 2014; Croll et al. 2015) were used to validate each assembly.

Gene and repetitive element annotation

We used pangenome gene annotation generated by Plissonneau et al. (2018). Genes were predicted by using splicing evidence from the *in planta* RNA-seq data from the same time points and strains as described above. Repetitive elements were annotated in all four genomes using RepeatMasker 4.0.5 (Smith, 1996) and a repeat element library for the reference genome (IPO323) produced by Grandaubert et al. (2015). This library was created using the REPET pipeline (Flutre et al. 2011). Repeat families in each species were clustered with BLASTClust from the NCBI-BLAST package (Altschul et al. 1990) and aligned with MAFFT (Katoh and Standley 2013) to create new consensus sequences. This process was repeated with lower identity percentages (from 100% to 75% identity) and lower coverage (from 100% to 30%) until sequences did not form new clusters anymore. The repetitive sequences were classified with the TEClassifier.py script from REPET using tBLASTx and BLASTx against the Giri Repbase Update database (Jurka et al. 2005) and by the identification of characteristic TE features such as LTRs. The sequences were also

translated into the six reading frames in order to identify protein domains in the conserved domain database (Marchler-Bauer et al. 2011) using RPS-BLAST. Identified repetitive sequences were finally named according to the three-letter nomenclature defined by Wicker et al. (2007). The single TE library enables the comparison of TEs between the four strains as all the elements have exactly the same naming. RepeatMasker was used with the following parameters: pa 2, -s and –a for using two parallel processors, in slow mode for increased sensitivity and generating an alignment output file. Additional elements were identified using MITEtracker with default parameters (Crescente et al. 2018).

RNA extractions, library preparation and sequencing

Seedlings of the wheat cultivar 'Drifter' were infected with the four strains 3D1, 3D7, 1E4 and 1A5, on the same day and in the same greenhouse chamber (Palma-Guerrero et al. 2016; Palma-Guerrero et al. 2017). Total RNA was extracted from inoculated second leaves at time points 7, 12 , 14 and 28 days post-infection (dpi) using a TRIzol (Invitrogen) extraction protocol (Palma-Guerrero et al. 2017). The time points were selected to include asymptomatic (biotrophic), necrotrophic and the saprophytic stages of infection. In addition, a TRIzol RNA extraction was performed for all four strains in a nutrient limited, defined salts medium without sucrose (Minimal Medium – MM pH 5.8) and a nutrient rich YSB media (10g/L sucrose and 10g/L yeast extract , pH 6.8) (Vogel et al. 1956; Francisco et al. 2018). Cells were recovered in YSB medium and then transferred to either YSB or MM, incubated for four days at 18°C, prior to harvest for RNA extraction. For the infection experiments, cells were harvested from three leaf samples from each time point for the *in planta* samples and three *in vitro* samples for each condition. Samples with the highest RNA quality as determined with a Bioanalyzer 2100 (Agilent), were selected as biological replicates for each time point for library preparation and sequencing. RNA quantity was assessed with a Qubit fluorometer (Life Technologies) and libraries were prepared using the TruSeq stranded

mRNA sample prep kit (Illumina Inc.) according to the provided protocol. Total RNA samples were ribosome depleted by using PolyA selection and reverse-transcribed into double-stranded cDNA. Actinomycin was added during the first strand synthesis. The cDNA was then fragmented, end-paired and a A-tail was added before the ligation of the TruSeq adapters. A selective enrichment for fragments with TruSeq adapters on each end was performed by polymerase chain reaction. The quality and quantity of the enriched libraries were verified with a Qubit (1.0) fluorometer and a Tapestation (Agilent). Paired-end libraries were sequenced on an Illumina HiSeq 2500 with read lengths of 2 x 125 bp (Illumina Inc.) for *the in planta* samples and with read lengths of 4 x 100 bp for the *in vitro* conditions.

Transcription mapping and quantification

Raw sequencing reads were quality-trimmed and filtered for adapter contamination and low-quality reads using Trimmomatic 0.36 (Bolger et al. 2014) using the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:5:10 MINLEN:50. Trimmed and filtered reads were mapped to the reference genome sequence of the specific strain (Plissonneau et al. 2018) using STAR 2.6.0 (Dobin and Gingeras 2016) allowing multiple mapped reads with the following settings: --outFilterMultimapNmax 100 --winAnchorMultimapNmax 200 --outSAMtype BAM Unsorted --outFilterMismatchNmax 3, according to the recommended parameters for TE analyses (Jin et al. 2015; Jin and Hammell 2018). We performed a saturation analysis to determine the cut-off to be used for the optimal number of reported alignments of a specific read, where increasing the threshold did not increase the number of mapped reads significantly as recommended (Jin and Hammell 2018) (supplementary fig. S1, supplementary table S1). The resulting unsorted bamfiles were sorted by read name with SAMtools 1.9 and the expression levels of TEs and genes were quantified using TEtranscripts 2.0.3 (Jin et al. 2015) with the following parameters: --stranded no --mode multi -p 0.05 -i 10. TEtranscripts counts uniquely and

multiple-mapped reads that align to genes and TE regions to determine TE- and gene-level transcript abundance. The software assumes that transcribed TEs will have reads mapping along the entire length of the element (Jin and Hammell 2018). Elements with reads mapping to only a fraction of the length were assumed to be non-transcribed as these subregions may not be unique enough in the genome compared to *e.g.* other TEs.

TE and gene read counts were normalized between replicates and time points for the *in planta* and *in vitro* samples using the R/Bioconductor package EdgeR 3.8 (Robinson and Smyth 2008; Robinson et al. 2010; Robinson and Oshlack 2010). Genes and TEs without at least one read in all the samples were excluded for the normalization step (Anders et al. 2013) and were assumed for the rest of the analyses to have zero expression. Library sizes were normalized with the TMM method (Robinson and Oshlack 2010). CPM (counts per million) were generated with the EdgeR CPM function using TMM-normalized libraries.

Locus-specific TE expression between 3D1 and 3D7

We analyzed transcriptomic reads for unique mapping (*i.e.* to a single genomic region) and extracted reads with the samtools view -q 255 (the flag assigned by the STAR aligner). Reads were quantified with htseq_count 0.8.0 and normalized as described above to generate CPM for each copy of a TE with locus specific expression information. The genomic region encoding the effector Avr3D1 was compared between the genomes of 3D7, 3D1 and IPO323 using pairwise blastn on repeat masked genomes. Hits were filtered for a minimum identity of 95%, e-values reported as effectively 0. Synteny blocks were visualized using the R package genoPlotR (Guy et al. 2010).

Genomic localization of TEs and co-expression analyses

In order to investigate the association of the genomic environment with TE expression, we identified the nearest gene to each TE using bedtools 2.27 command *closest* with the option to report only the closest TE to each gene and to allow overlaps to include genes that have been disrupted by intragenic TEs (Quinlan and Hall 2010). Co-expression clusters were computed using the Short Time-Series Expression Miner (STEM) software 1.3.11, designed to analyze time series with 3-8 time points (Ernst et al. 2005; Ernst and Bar-Joseph 2006). STEM software uses a non-parametric clustering method to assign genes to predefined expression profiles. It considers expression profiles to be significant if the number of genes assigned to a cluster departs from random. We used all three individual replicates per condition and isolate (option: repeat data) and transformed all data using log normalization. The analysis identified a total of 20 co-expression profiles, namely 8-9-13-15-16-17-18-19-20-21-27-28-29-31-32-33-35-36-37 and 39. The statistical significance of the number of genes assigned to each profile was computed by applying a Bonferroni correction with alpha = 0.05. The biological relevance of co-expression profiles was assessed by gene ontology (GO) term enrichment analysis (Ernst et al. 2005; Ernst and Bar-Joseph 2006). STEM software implements a GO term enrichment method that uses the hypergeometric distribution based on the number of genes assigned to the co-expression profile, the number of genes assigned to the GO category and the number of unique genes in the experiment. Enrichment significance was corrected by using randomization tests.

RIP analysis

We identified RIP by aligning each copy of a given TE with the consensus sequence using MAFFT 7.407 (Katoh et al. 2002; Katoh and Standley 2013). The consensus sequence could be more affected by RIP than individual TE copies in the genome because mutations occurring in a given sequence are likely to be removed by the “base-pair majority rule” used

to build the consensus. In this case, the copy with the highest GC content (*i.e*. the least affected by RIP) is used as the RIPCAL 1.0 input (Hane and Oliver 2008). All TE families (individual TE copies and the consensus) were aligned and processed by RIPCAL using default parameters (Hane and Oliver 2008). RIPCAL output provides the number of transition and transversions, single mutations and dinucleotide targets used in all possible transition mutations for each genomic TE copy. The RIPCAL output can be used to determine whether individual TE copies are "RIPped" based on two indices: (CpA+TpG)/(ApC+GpT) indicating a decrease in RIP targets and TpA/ApT, indicating an increase in RIP products. We used the default criteria where a (CpA+TpG)/(ApC+GpT) ratio of below 1.03 is indicative of RIP and a TpA/ApT of higher than 0.89 is indicative of RIP (Hane and Oliver 2008). In general TE families with a lower (CpA+TpG)/(ApC+GpT) value and a higher TpA/ApT are more affected by RIP (Hane and Oliver 2008). We excluded unknown elements from this analysis. R (R Core Team, 2017) was used to generate graphics from RIPCAL outputs. These outputs were parsed to search for RIP signatures in TE copies and the dinucleotide targets used in the transition type mutations that are usually associated with RIP.

Acknowledgements

We thank Javier Palma-Guerrero for providing access to transcriptomic datasets and Emilie Chanclud for helpful comments on a previous version of the manuscript. Data produced in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Center Zurich. SF was supported by the Swiss National Science Foundation (http://www.snf.ch) through the grant (SNF 31003A_155955) awarded to Bruce A. McDonald. DC receives support from the Swiss National Science Foundation (grants 31003A_173265 and IZCOZO_177052).

Author contributions

S.F., C.P., and D.C.: conceived the study; S.F., T.B., U.O., and C.P.: performed analyses; C.S.F.: contributed data set; S.F. and D.C.: wrote the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3):403–410.

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat. Protoc. 8(9):1765–1786.

Barah P, Jayavelu ND, Rasmussen S, Nielsen HB, Mundy J, Bones AM. 2013. Genome-scale cold stress response regulatory networks in ten *Arabidopsis thaliana* ecotypes. BMC Genomics 14:722.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Borgognone A, Castanera R, Morselli M, López-Varas L, Rubbi L, Pisabarro AG, Pellegrini M, Ramírez L. 2018. Transposon-associated epigenetic silencing during *Pleurotus ostreatus* life cycle. DNA Res. 25:451–464.

Bucher E, Reinders J, Mirouze M. 2012. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. Curr. Opin. Plant Biol. 15(5):503–510.

Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, Sunaga F, Toritsuka M, Ikawa D, Kakita A, et al. 2014. Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron 81(2):306–313.

Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity 85(2):101–106.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. Mol. Ecol. 22(6):1503–1517.

Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Scheid OM. 2014. How a retrotransposon exploits the plant's heat stress response for its activation. PLOS Genet. 10(1):e1004115.

Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates fungal alkaloid biosynthesis in a fungal endophyte-plant symbiosis. Mol. Microbiol. 92(2):413–434.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. 18(2):71–86.

Connolly LR, Smith KM, Freitag M. 2013. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLOS Genet 9(10):e1003916.

Cowley M, Oakey RJ. 2013. Transposable elements re-wire and fine-tune the transcriptome. PLOS Genet. 9(1):e1003234.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics 19(1):348.

Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201(3):1213–1228.

Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. PLOS Pathog 8(4):e1002608.

Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLOS Genet 9(6):e1003567.

Derridj S. 1996. Nutrients on the leaf surface. In: Morris CE, Nicot PC, Nguyen-The C, editors. Aerial plant surface microbiology. Boston, MA: Springer US. p. 25–42. Available from: https://doi.org/10.1007/978-0-585-34164-4_2.

Dillon N. 2004. Heterochromatin structure and function. Biol. Cell 96(8):631–637.

Dobin A, Gingeras TR. 2016. Optimizing RNA-seq mapping with STAR. Methods Mol. Biol. 1415:245–262.

Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284(5757):601–603.

Dubin MJ, Mittelsten Scheid O, Becker C. 2018. Transposons: a blessing curse. Curr. Opin. Plant Biol. 42:23–29.

Ernst J, Bar-Joseph Z. 2006. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 7:191.

Ernst J, Nau GJ, Bar-Joseph Z. 2005. Clustering short time series gene expression data. Bioinforma. Oxf. Engl. 21 (Suppl) 1:i159-168.

Ferreira RB, Monteiro S, Freitas R, Santos CN, Chen Z, Batista LM, Duarte J, Borges A, Teixeira AR. 2006. Fungal Pathogens: The Battle for Plant Infection. Crit. Rev. Plant Sci. 25(6):505–524.

Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in *de novo* annotation approaches. PLOS ONE 6(1):e16526.

Fones H, Gurr S. 2015. The impact of *Septoria tritici* Blotch disease on wheat: An EU perspective. Fungal Genet. Biol. 79:3–7.

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46:34–42.

Francisco CS, Ma X, Zwyssig MM, McDonald BA, Palma-Guerrero J. 2019. Morphological changes in response to environmental stresses in the fungal plant pathogen *Zymoseptoria tritici*. Sci. Rep. 9:1–18.

Galindo-González L, Mhiri C, Deyholos MK, Grandbastien M-A. 2017. LTR-retrotransposons in plants: Engines of evolution. Gene 626:14–25.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Lee TAJV der, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLOS Genet 7(6):e1002070.

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

Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26(18):2334–2335.

Hacquard S, Kracher B, Maekawa T, Vernaldi S, Schulze-Lefert P, Themaat EVL van. 2013. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. Proc. Natl. Acad. Sci. 110(24):E2219–E2228.

Hane JK, Oliver RP. 2008. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9:478.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11(5):1189–1204.

Haueisen J, Möller M, Eschenbrenner CJ, Grandaubert J, Seybold H, Adamiak H, Stukenbrock EH. 2019. Highly flexible infection programs in a specialized wheat pathogen. Ecol Evol. 9(1):275–294.

Hernández-Chávez MJ, Pérez-García LA, Niño-Vega GA, Mora-Montes HM. 2017. Fungal strategies to evade the host immune recognition. J. Fungi 3:51.

Hirsch CD, Springer NM. 2017. Transposable element influences on gene expression in plants. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1860(1):157–165.

Horváth V, Merenciano M, González J. 2017. Revisiting the relationship between transposable elements and the eukaryotic stress response. Trends Genet. 33(11)832-841.

Huang J, Wang Y, Liu W, Shen X, Fan Q, Jian S, Tang T. 2017. *EARE-1*, a transcriptionally active Ty1/Copia-like retrotransposon has colonized the genome of *Excoecaria agallocha* through horizontal transfer. Front. Plant Sci. 8:45.

Hummel B, Hansen EC, Yoveva A, Aprile-Garcia F, Hussong R, Sawarkar R. 2017. The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. Nat. Struct. Mol. Biol. 24(3):234–242.

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472(7341):115–119.

Jin Y, Hammell M. 2018. Analysis of RNA-Seq Data Using TEtranscripts. In: Wang Y, Sun M, editors. Transcriptome data analysis: methods and protocols. Methods in molecular biology. New York, NY: Springer New York. p. 153–167. Available from: https://doi.org/10.1007/978-1-4939-7710-9_11

Jin Y, Tam OH, Paniagua E, Hammell M. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics 31(22):3593–3599.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110(1-4):462–467.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4):772–780.

Kirkland TN, Muszewska A, Stajich JE. 2018. Analysis of Transposable Elements in *Coccidioides* Species. J. Fungi 4:13..

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Themaat EVL van, Does HC van der, Hacquard S, Stüber K, Will I, Schmalenbach W, et al. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLOS Pathog. 8(4):e1002643.

Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA, Sánchez-Vallet A. 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biol. 16(1):78.

Lanciano S, Mirouze M. 2018. Transposable elements: all mobile, all different, some stress responsive, some adaptive? Curr. Opin. Genet. Dev. 49:106–114.

Le T-N, Schumann U, Smith NA, Tiwari S, Au PCK, Zhu Q-H, Taylor JM, Kazan K, Llewellyn DJ, Zhang R, et al. 2014. DNA demethylases target promoter transposable elements

to positively regulate stress responsive genes in Arabidopsis. Genome Biol. 15(9):458.

Le Rouzic A, Capy P. 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 169(2):1033–1043.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 27(4):390–399.

Lee YCG. 2015. The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. PLOS Genet. 11(6):e1005269.

Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K. 2015. Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. 112(11):3553–3557.

Lendenmann MH, Croll D, Palma-Guerrero J, Stewart EL, McDonald BA. 2016. QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. Heredity 116(4):384–394.

Lendenmann MH, Croll D, Stewart EL, McDonald BA. 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. G3 (Bethesda). 4(12):2519–2533.

Lonnig W-E, Saedler H. 2002. Chromosome rearrangements and transposable elements. Annu. Rev. Genet. 36:389–410.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39:D225-229.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A. 2009. A transposon-induced epigenetic change leads to sex determination in melon. Nature 461(7267):1135–1138.

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. 36(6):344–355.

Meile L, Croll D, Brunner PC, Plissonneau C, Hartmann FE, McDonald BA, Sánchez-Vallet A. 2018. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. New Phytol. 219(3):1048–1061.

Menees TM, Sandmeyer SB. 1996. Cellular stress inhibits transposition of the yeast retrovirus-like element Ty3 by a ubiquitin-dependent block of virus-like particle formation. Proc. Natl. Acad. Sci. U. S. A. 93(11):5629–5634.

Miousse IR, Chalbot M-CG, Lumen A, Ferguson A, Kavouras IG, Koturbash I. 2015. Response of transposable elements to environmental stressors. Mutat. Res. Rev. Mutat. Res. 765:19–39.

Mita P, Boeke JD. 2016. How retrotransposons shape genome regulation. Curr. Opin. Genet. Dev. 37:90–100.

Muszewska A, Hoffman-Sommer M, Grynberg M. 2011. LTR Retrotransposons in Fungi. PLoS ONE 6(12):e29425

Negi P, Rai AN, Suprasanna P. 2016. Moving through the stressed genome: emerging regulatory roles for transposons in plant stress response. Front. Plant Sci. 7:1448

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al. 2016. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 17(1):194.

Palma-Guerrero J, Ma X, Torriani SFF, Zala M, Francisco CS, Hartmann FE, Croll D, McDonald BA. 2017. Comparative transcriptome tnalyses in *Zymoseptoria tritici* reveal significant differences in gene expression among strains during plant infection. Mol. Plant. Microbe Interact. 30(3):231–244.

Palma-Guerrero J, Stefano F. F. Torriani, Zala M, Carter D, Courbot M, Rudd JJ, McDonald BA, Croll D. 2016. Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intraspecific variability in transcription profiles. Mol. Plant Pathol. 17(6):845–859.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16(1):5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio 7(5):e01231-16.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.

Qutob D, Patrick Chapman B, Gijzen M. 2013. Transgenerational gene silencing causes gain of virulence in a plant pathogen. Nat. Commun. 4:1349.

R Core Team. 2017. R: a language and environment for statistical com- puting. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/, last accessed April 15, 2019.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139-140

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11(3):R25.

Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9(2):321–332.

Romero-Soriano V, Guerreiro MPG. 2016. Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids. PLOS ONE 11:e0147903.

Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, et al. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202.

Rudd JJ, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, Lysenko A, Saqi M, Desai NM, Powers SJ, et al. 2015. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. Plant Physiol. 167:1158–1185.

Ryan CP, Brownlie JC, Whyard S. 2016. Hsp90 and physiological stress are linked to autonomous transposon mobility and heritable genetic change in nematodes. Genome Biol. Evol. 8(12):3794–3805.

Saksouk N, Simboeck E, Déjardin J. 2015. Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin 8:3.

Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56:21–40.

Saze H, Kakutani T. 2007. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. EMBO J. 26(15):3641–3652.

Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock EH. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin 8:41.

Seidl MF, Thomma BPHJ. 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. 33(11):842-851.

Selker EU. 2002. Repeat-induced gene silencing in fungi. Adv. Genet. 46:439–450.

Shetty NP, Mehrabi R, Lütken H, Haldrup A, Kema GHJ, Collinge DB, Jørgensen HJL. 2007. Role of hydrogen peroxide during the interaction between the hemibiotrophic fungal pathogen *Septoria tritici* and wheat. New Phytol. 174(3):637–647.

Shpyleva S, Melnyk S, Pavliv O, Pogribny I, Jill James S. 2018. Overexpression of LINE-1 retrotransposons in autism brain. Mol. Neurobiol. 55(2):1740–1749.

Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151(5):964–980.

Skibbe DS, Doehlemann G, Fernandes J, Walbot V. 2010. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. Science 328(5974):89–92.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8(4):272–285.

Soyer JL, Ghalid ME, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent M-H, Connolly LR, Freitag M, Rouxel T, et al. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. PLOS Genet. 10(3):e1004227.

Soyer JL, Grandaubert J, Haueisen J, Schotanus K, Stukenbrock EH. 2019. *In planta* chromatin immunoprecipitation in *Zymoseptoria tritici* reveals chromatin-based regulation of putative effector gene expression. bioRxiv:544627.

Soyer JL, Rouxel T, Fudal I. 2015. Chromatin-based control of effector gene expression in plant-associated fungi. Curr. Opin. Plant Biol. 26:51–56.

Stewart E l., Croll D, Lendenmann MH, Sanchez-Vallet A, Hartmann FE, Palma-Guerrero J, Ma X, McDonald BA. 2017. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Mol. Plant Pathol. 19(1):201–216.

Stewart EL, McDonald BA. 2014. Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. Phytopathology 104(9):985–992.

Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat. Genet. 43(11):1160–1163.

Studt L, Rösler SM, Burkhardt I, Arndt B, Freitag M, Humpf H-U, Dickschat JS, Tudzynski B. 2016. Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. Environ. Microbiol. 18(11):4037-4054.

Torriani SFF, Melichar JPE, Mills C, Pain N, Sierotzki H, Courbot M. 2015. *Zymoseptoria tritici*: A major threat to wheat production, integrated approaches to control. Fungal Genet. Biol. FG B 79:8–12.

Trivedi M, Shah J, Hodgson N, Byun H-M, Deth R. 2014. Morphine induces redox-based changes in global DNA methylation and retrotransposon transcription by inhibition of excitatory amino acid transporter type 3–mediated cysteine uptake. Mol. Pharmacol. 85(5):747–757.

Trojer P, Reinberg D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? Mol. Cell 28(1):1–13.

Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V. 2014. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat. Commun. 5:5011.

Vogel HJ, Vogel HJ, Vogel H. 1956. A convenient growth medium for *Neurospora crassa* (medium N). Microbial Genetics Bulletin 13:42-43.

Voronova A, Belevich V, Jansons A, Rungis D. 2014. Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris L.*) genome. Tree Genet. Genomes 10(4):937–951.

Wang Q, Han C, Ferreira AO, Yu X, Ye W, Tripathy S, Kale SD, Gu B, Sheng Y, Sui Y, et al. 2011. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. Plant Cell 23(6):2064.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973–982.

Williams BP, Pignatta D, Henikoff S, Gehring M. 2015. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet. 11(3):e1005142.

Yasuda K, Ito M, Sugita T, Tsukiyama T, Saito H, Naito K, Teraishi M, Tanisaka T, Okumoto Y. 2013. Utilization of transposable element *mPing* as a novel genetic tool for modification of the stress response in rice. Mol. Breed. New Strateg. Plant Improv. 32:505–516.

Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13(8):335–340.

Zeng F, Cheng B. 2014. Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of FAE1 in *Sinapis alba*. Plant Cell 26(6):2648–2659.

Zhan J, Linde CC, Jürgens T, Merz U, Steinebrunner F, McDonald BA. 2005. Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus *Mycosphaerella graminicola*. Mol. Ecol. 14(9):2683–2693.

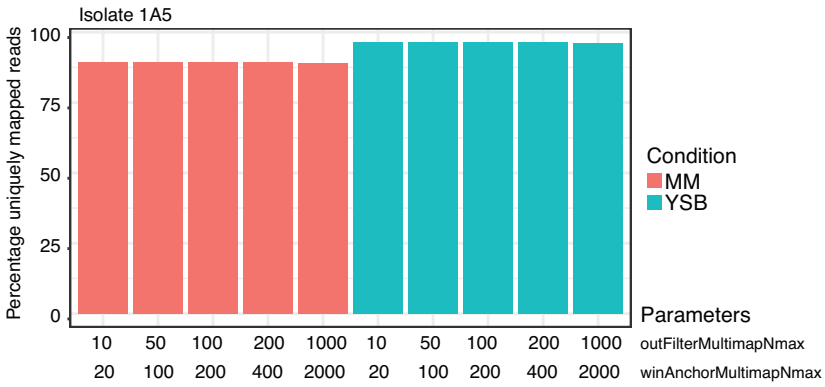
Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, Zala M, Ducasse A, Confais J, Compain J, Lapalu N, et al. 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. New Phytol. 214(2):619–631.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 39(1):61–69.

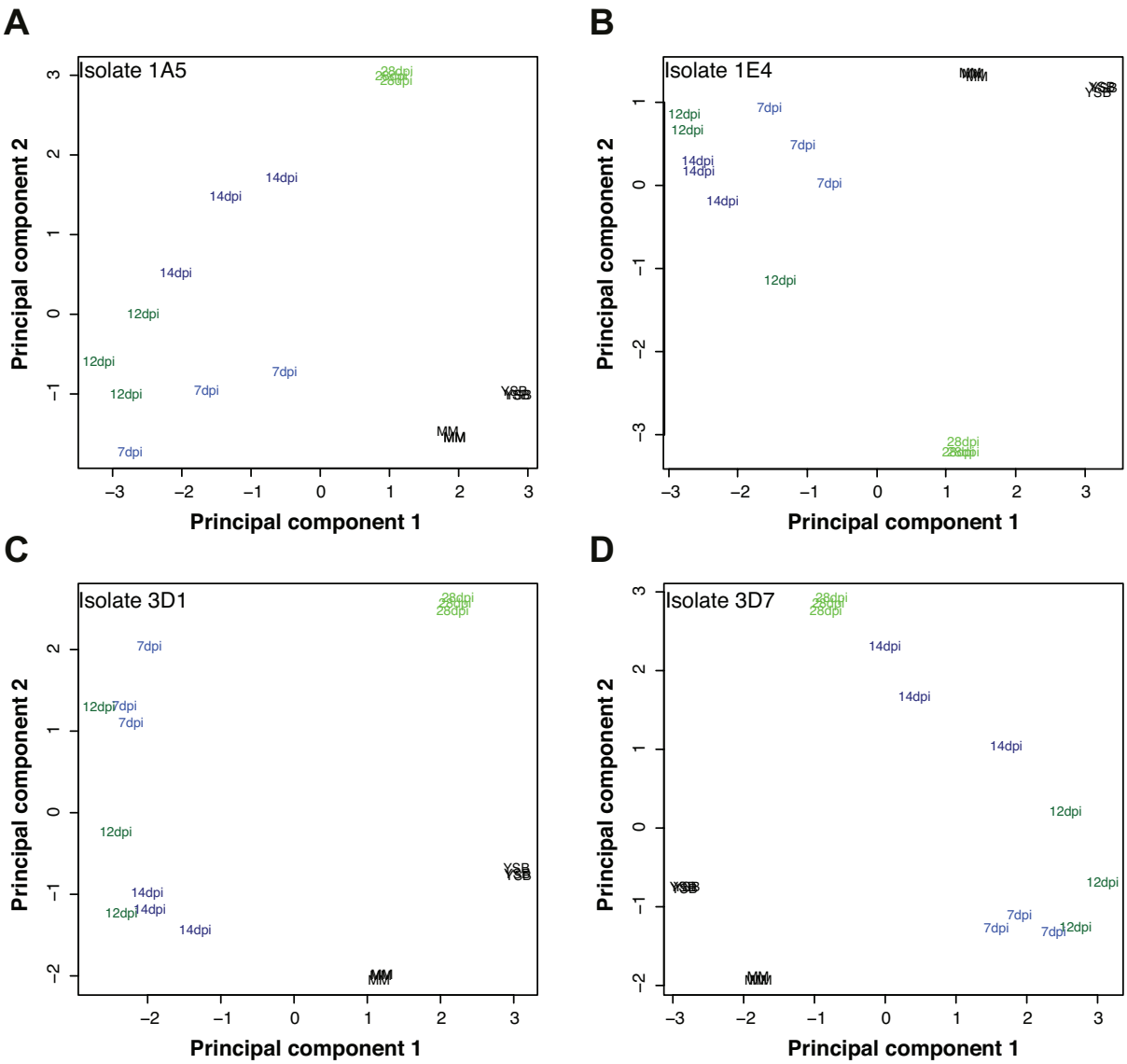
Zovoilis A, Cifuentes-Rojas C, Chu H-P, Hernandez AJ, Lee JT. 2016. Destabilization of B2 RNA by EZH2 activates the stress response. Cell 167(7):1788-1802.e13.

Supplementary Information

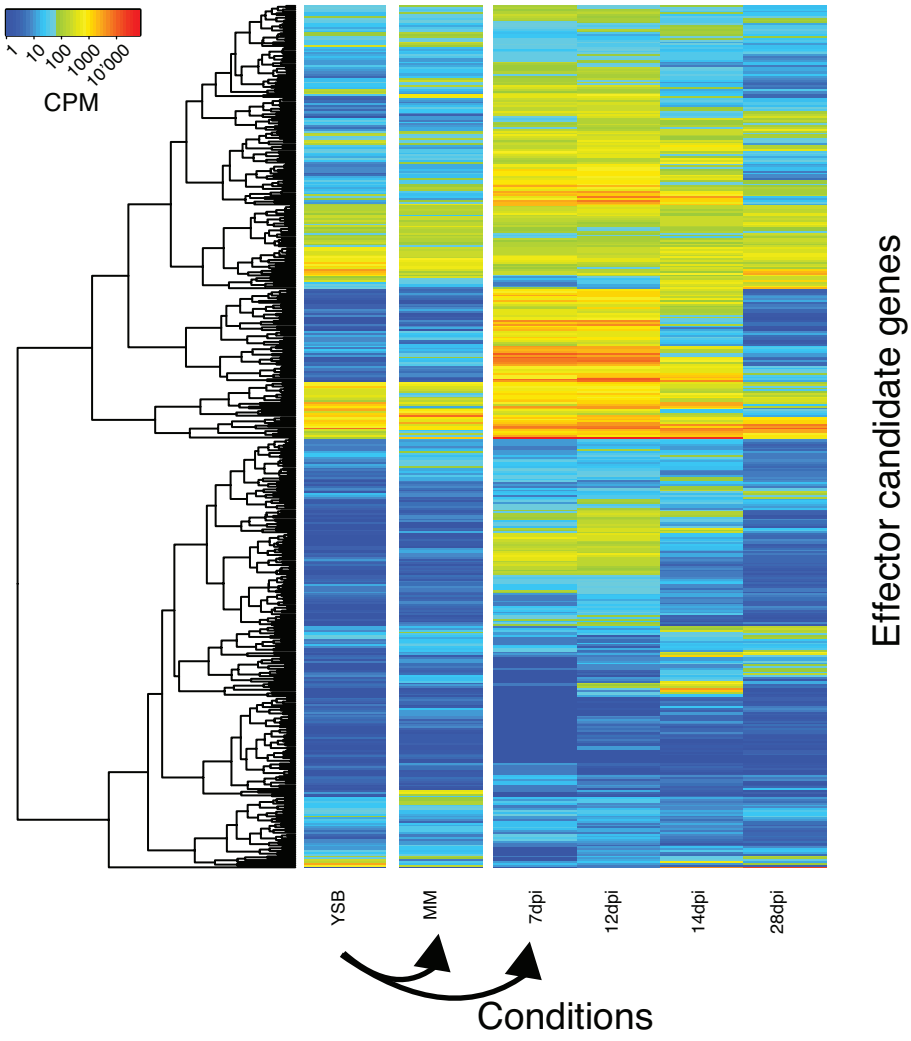
Supplementary Figures



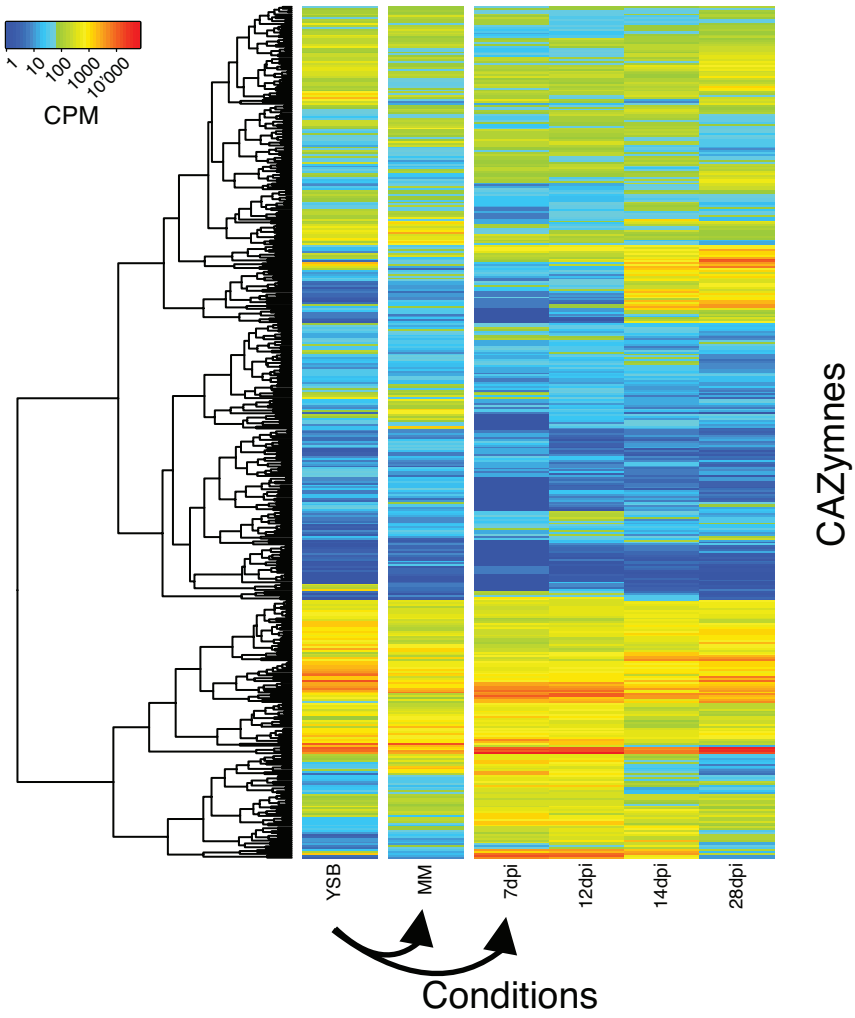
Supplementary Figure S1: Saturation analysis to determine the cut-off to be used for the optimal number of reported alignments of a specific read.



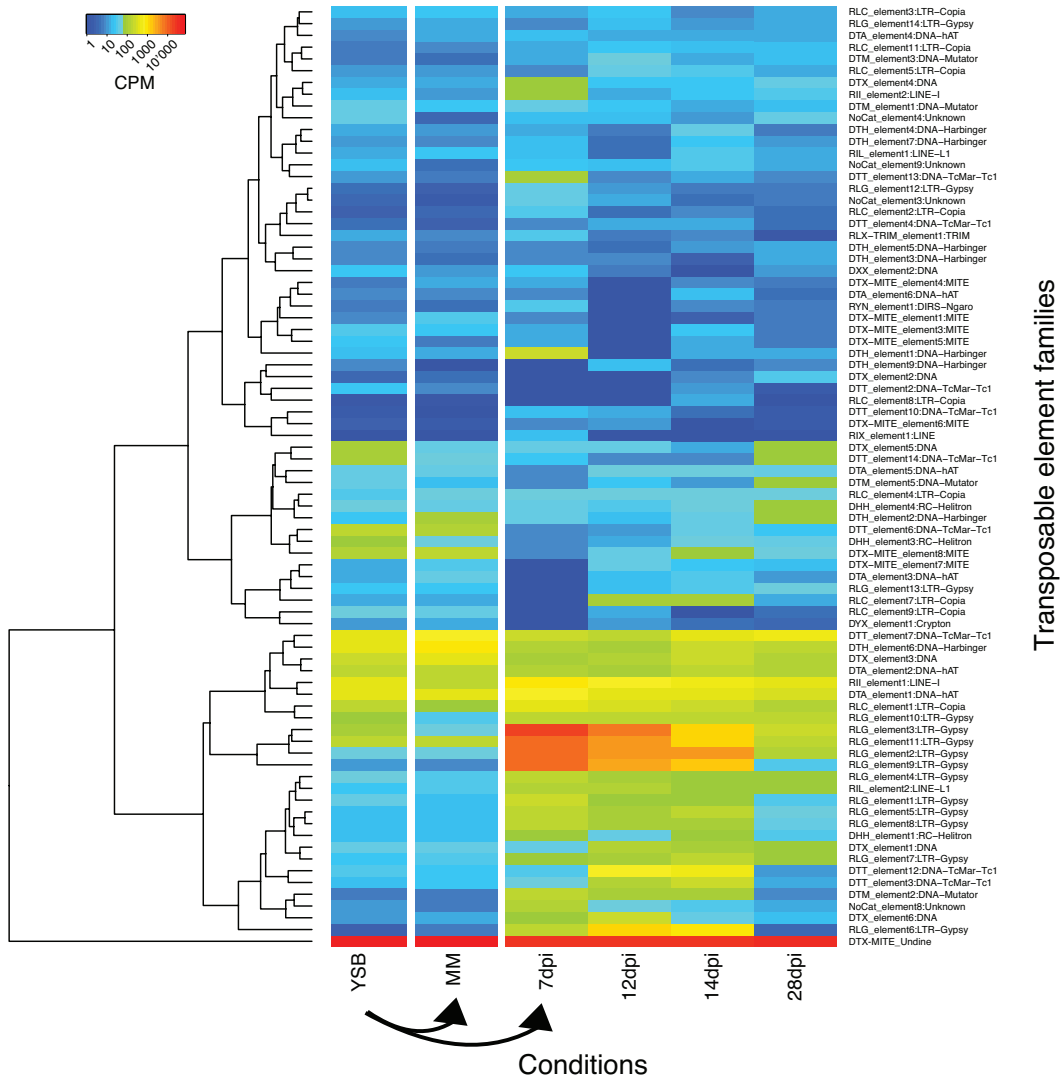
Supplementary Figure S2: MDS plot of genes and TEs in all four strains. Biological replicates are shown for each condition and time point.



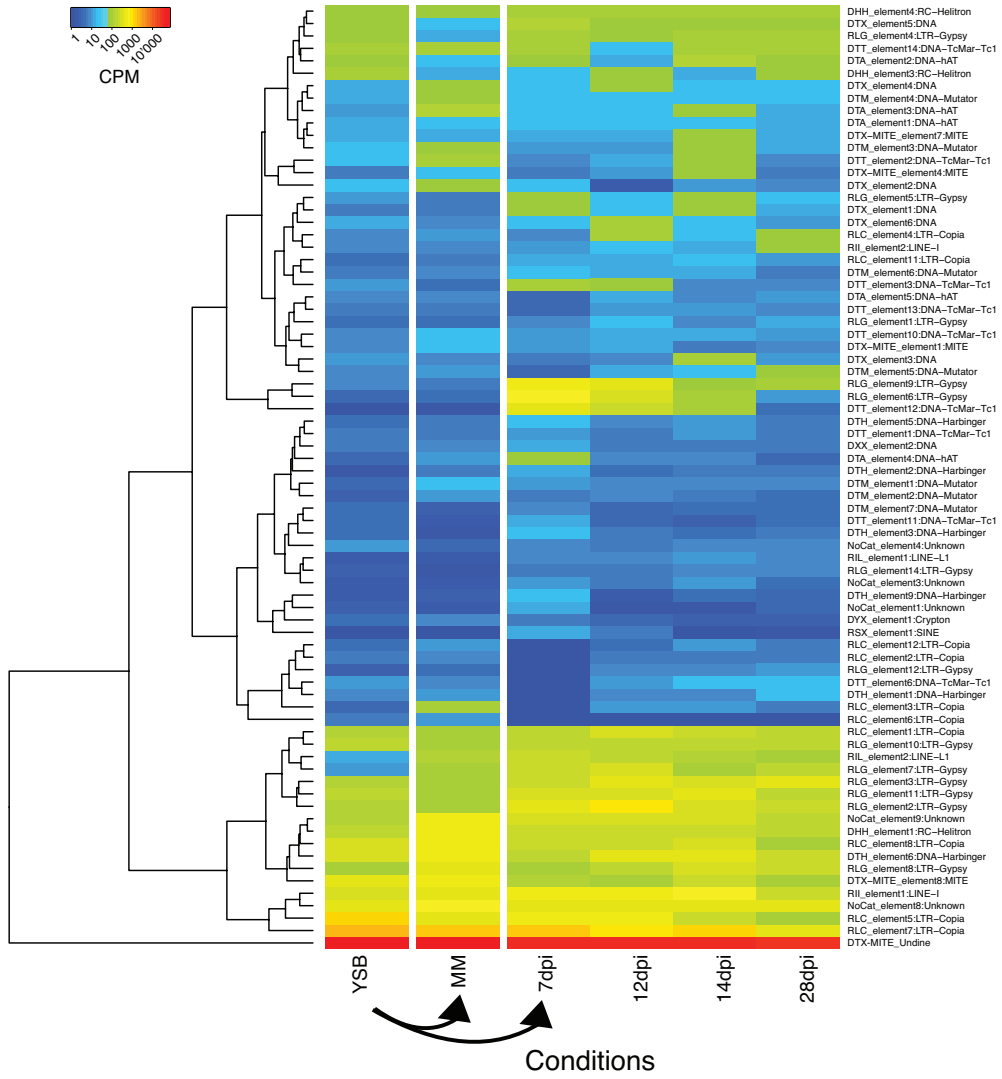
Supplementary Figure S3: Heatmap of putative effector gene expression in 3D7. CPM: counts per million.



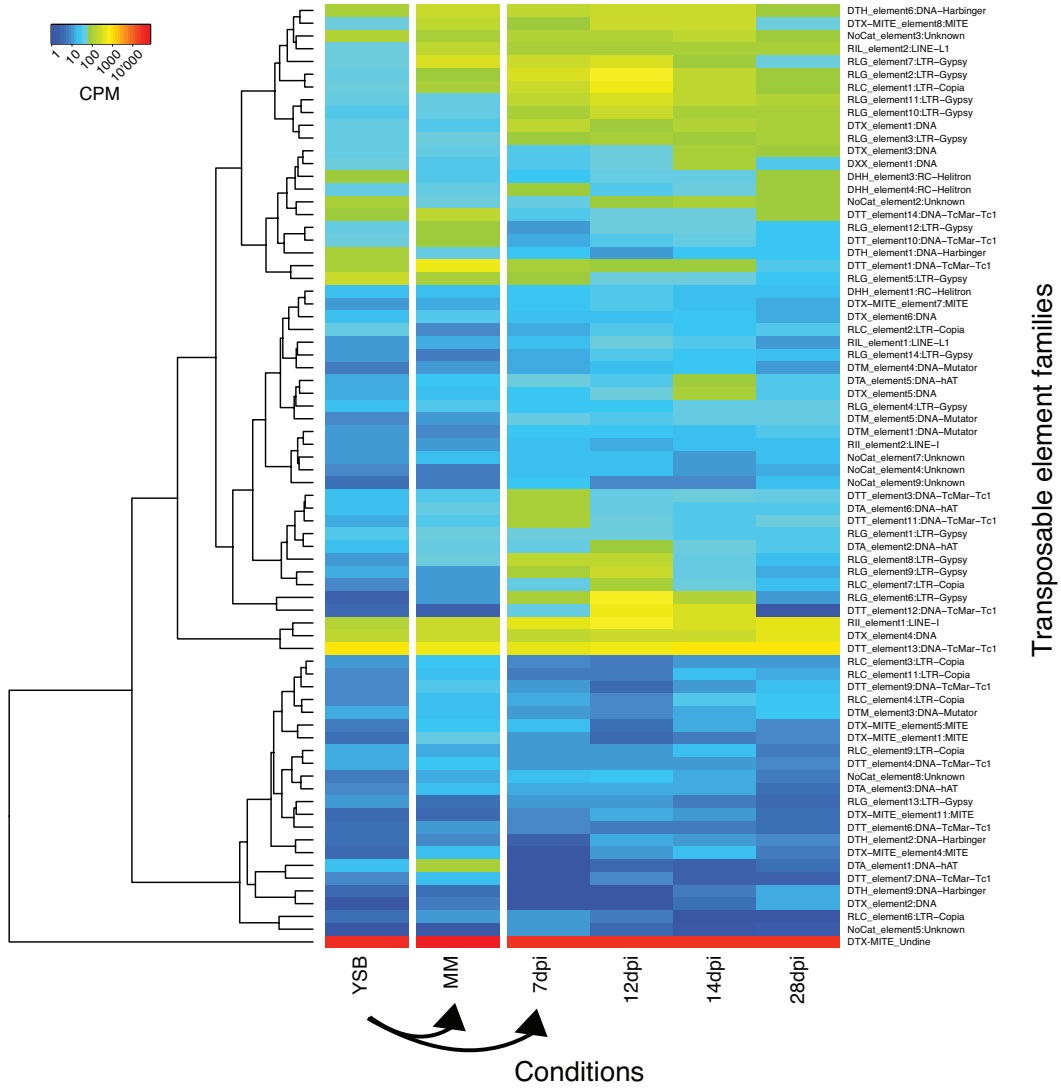
Supplementary Figure S4: Heatmap of carbohydrate-active enzyme (CAZyme) expression in strain 3D7. CPM: counts per million.



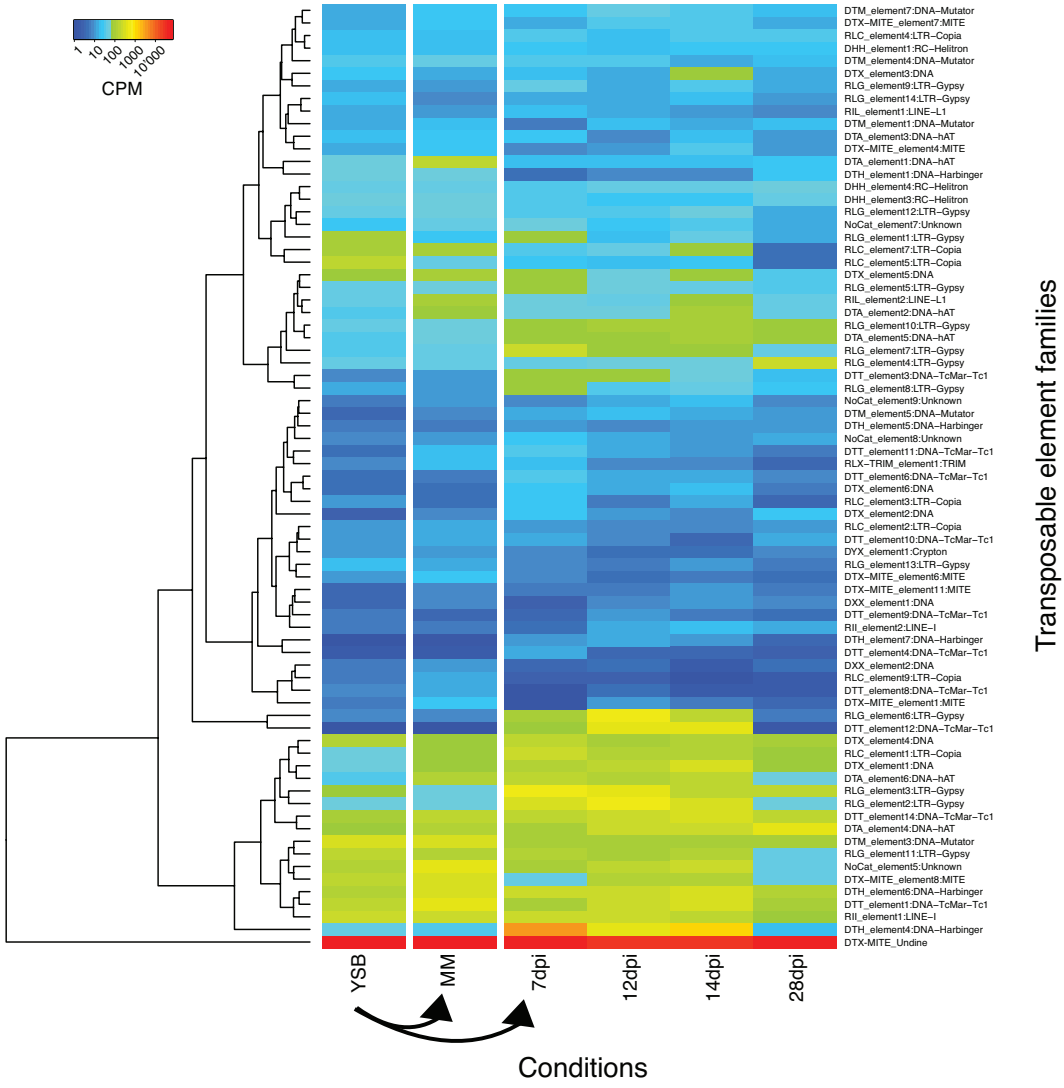
Supplementary Figure S5A: Heatmap of expression of transposable elements in 3D1. CPM: counts per million.



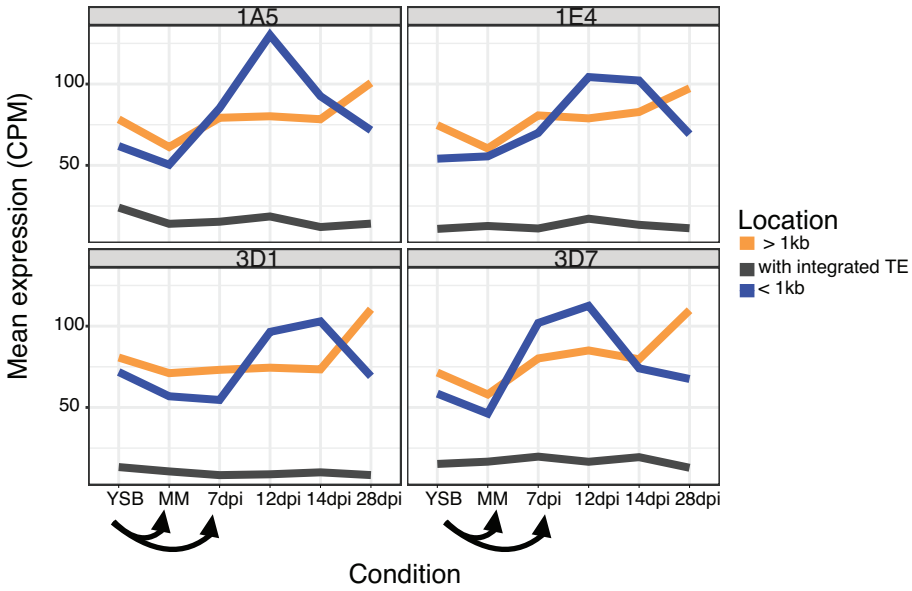
Supplementary Figure S5B: Heatmap of expression of transposable elements in 3D7. CPM: counts per million.



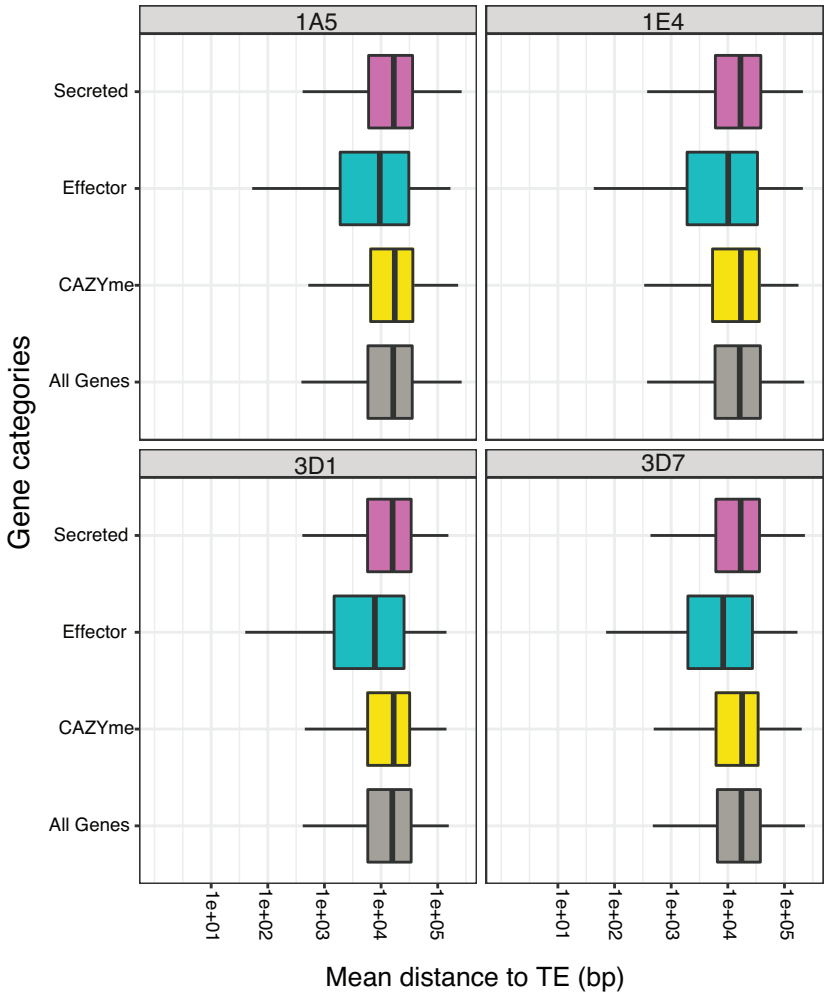
Supplementary Figure S5C: Heatmap of expression of transposable elements in 1A5. CPM: counts per million.



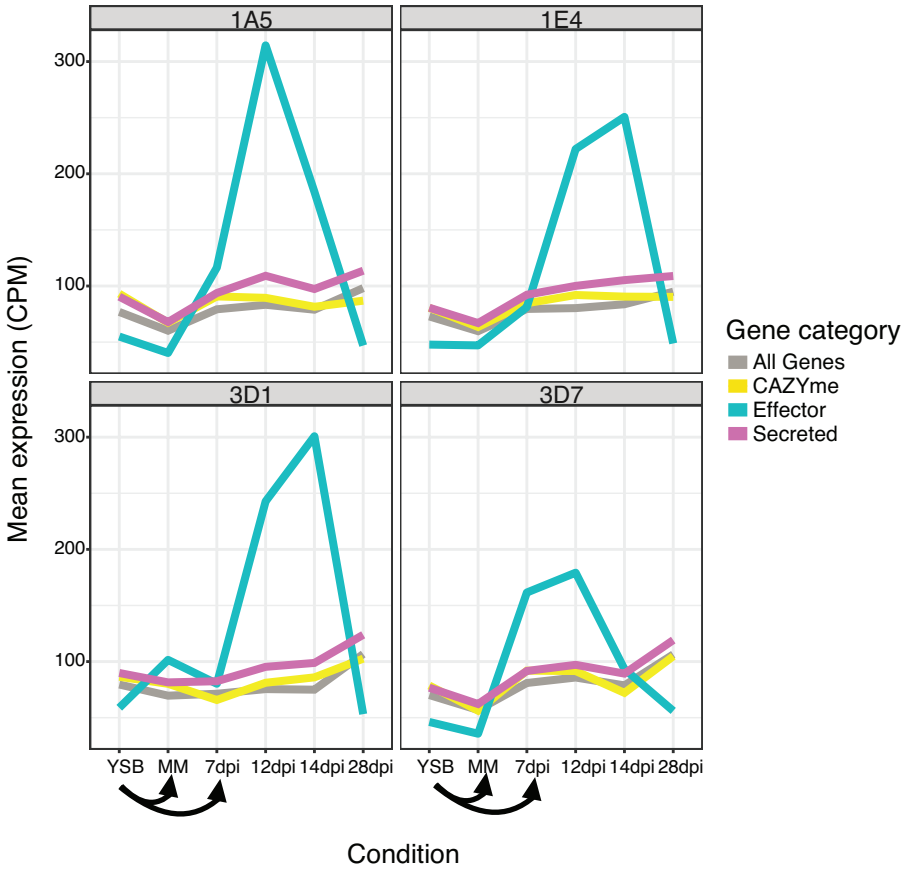
Supplementary Figure S5D: Heatmap of expression of transposable elements in 1E4. CPM: counts per million.



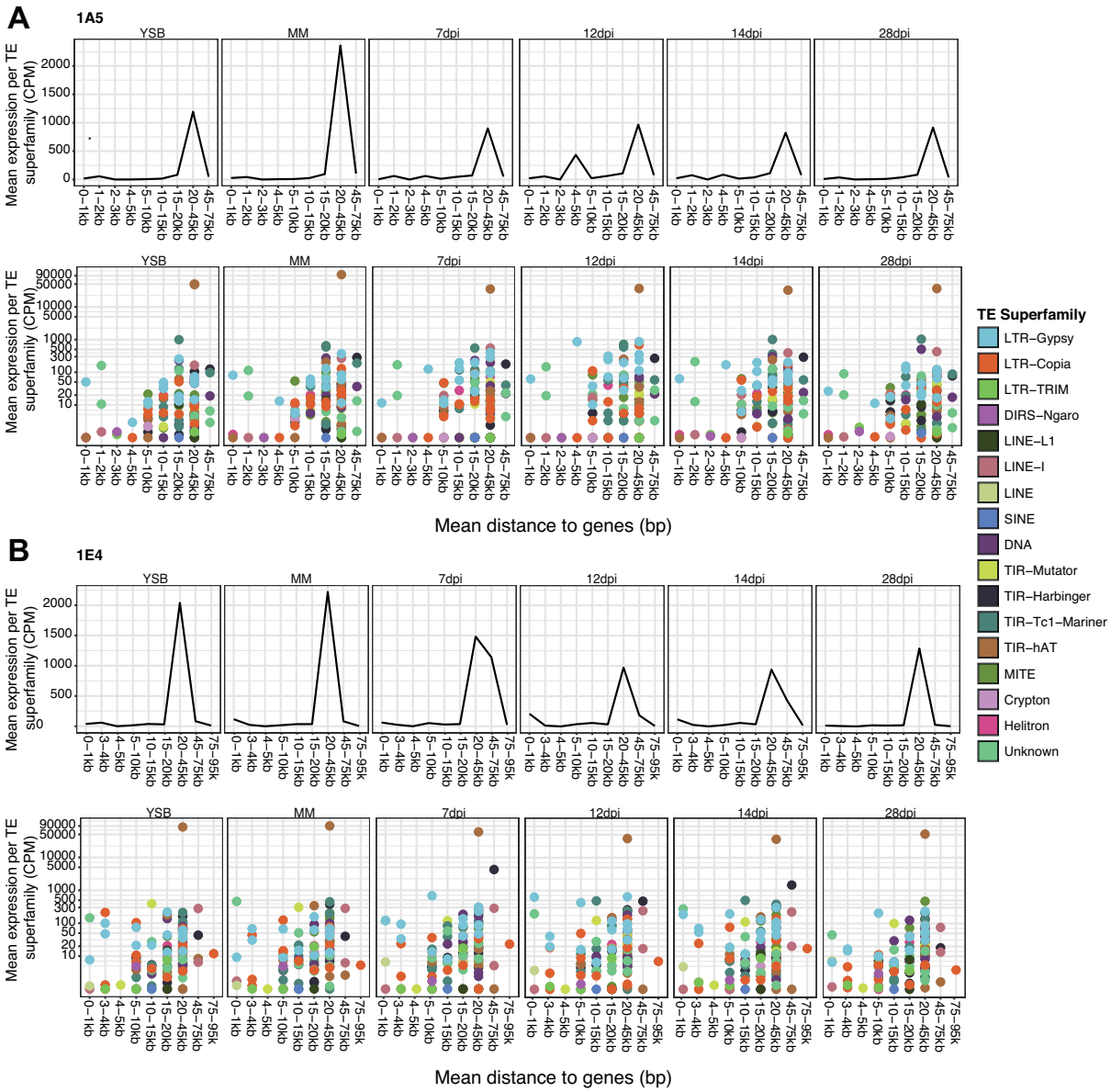
Supplementary Figure S6: Gene expression as a function of the proximity to transposable elements in all four strains. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.



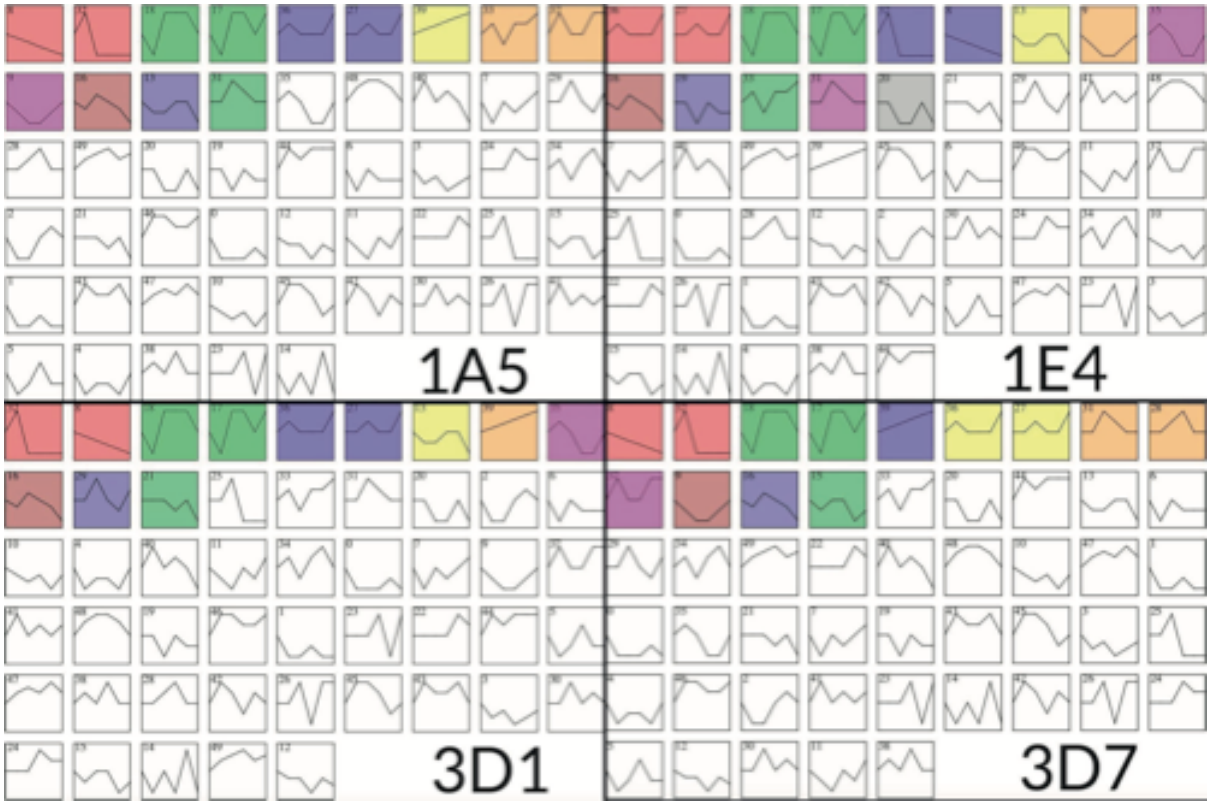
Supplementary Figure S7: Mean distance between genes grouped by functional category.



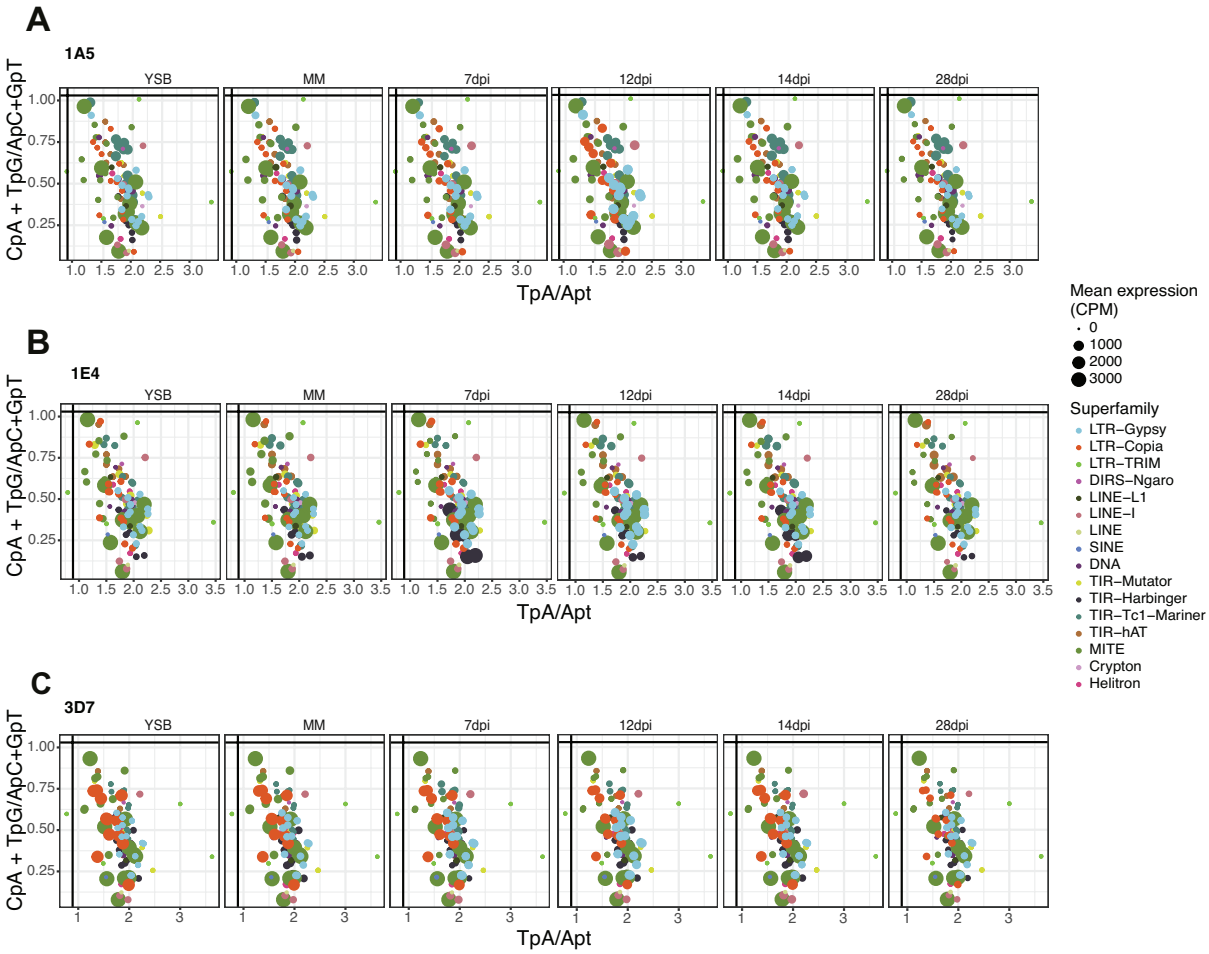
Supplementary Figure S8: Mean expression of genes grouped by functional category. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.



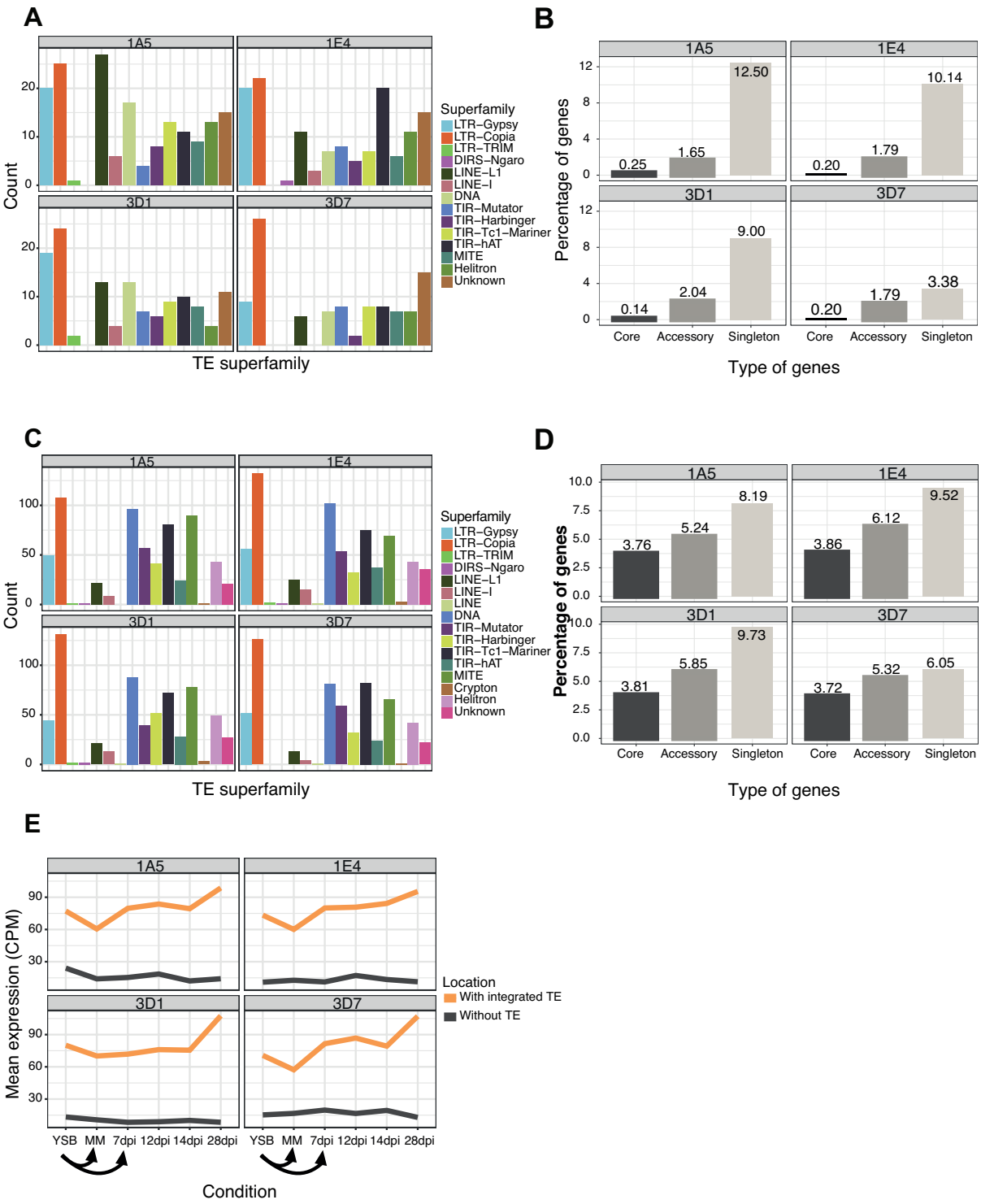
Supplementary Figure S9: Mean expression of transposable element families at intervals from the nearest gene in 1A5 and 1E4. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.



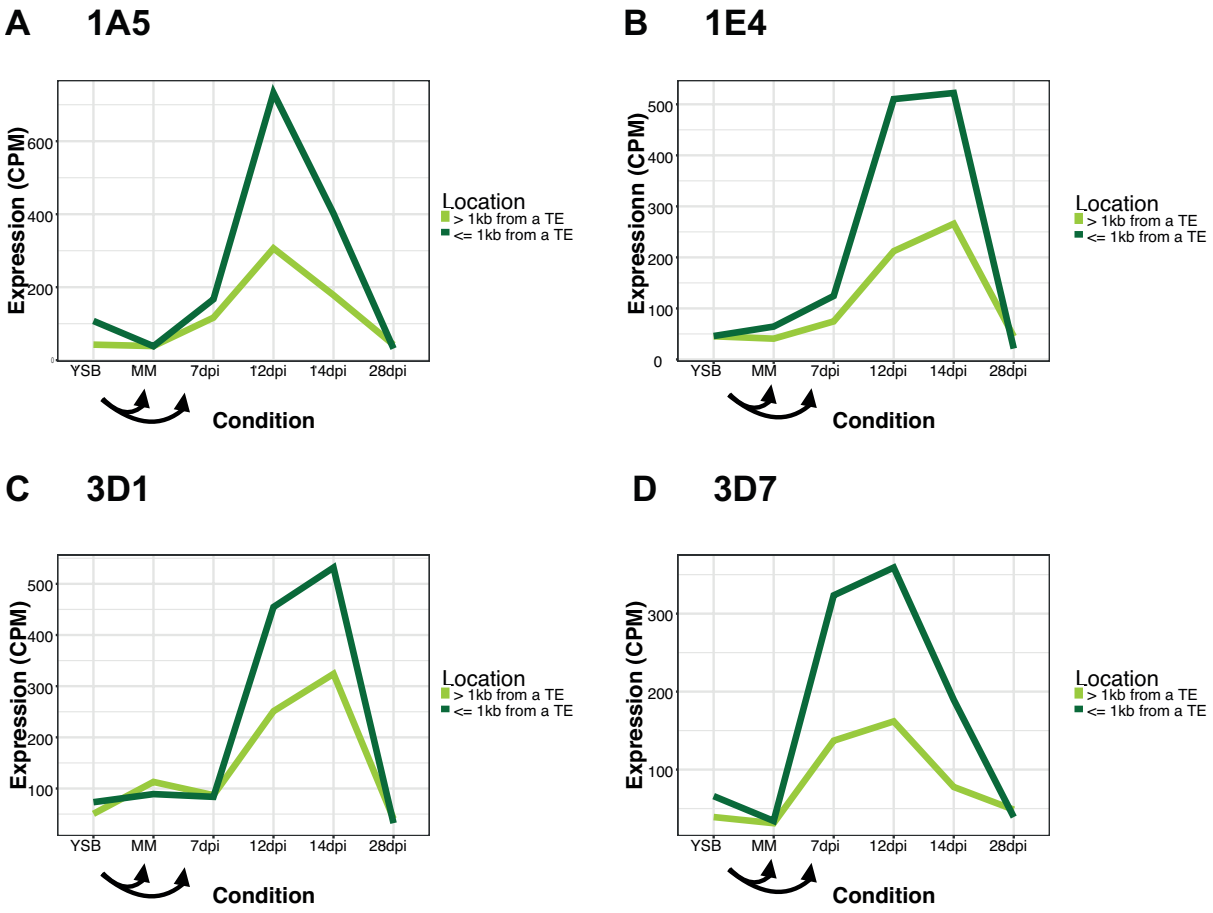
Supplementary Figure S10: Co-expression profiles for all four strains. Colored profiles are statistically significant. See methods for details.



Supplementary Figure S11: The RIP indices for each transposable element family and mean expression of the family under all stress conditions for strains 1A5, 1E4 and 3D7. Vertical and horizontal lines represent commonly used thresholds to detect RIP (Hane & Oliver 2008). Colors indicate the superfamily and size the expression at the family level in CPM.

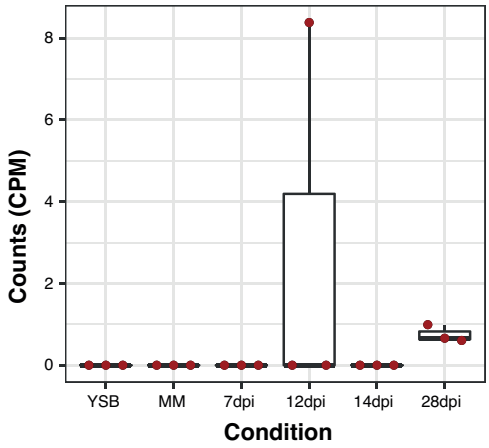


Supplementary Figure S12: (A) Transposable element (TE) superfamilies within genes in the four strains. (B) Categories genes with inserted TEs. The percentages are given as the number of genes of that category (core, accessory or strain specific) within genes as a fraction of the total number of genes in a category. (C) TE superfamilies within 1kb of the closest gene. (D) Categories of genes within 1 kb of TEs. The percentages are given as the number of genes of that category (core accessory or strain specific) within genes as a fraction of the total number of genes in a category. (E) Expression of genes with or without inserted TEs. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.

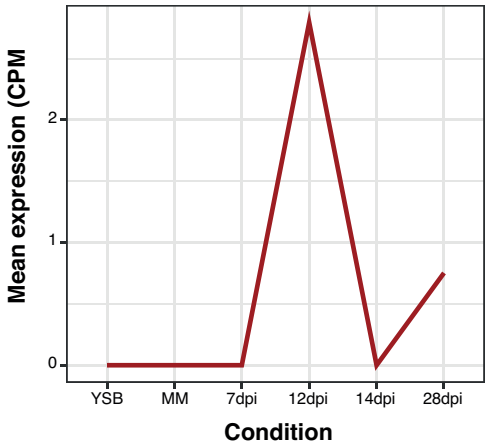


Supplementary Figure S13: Effector gene expression and proximity to the closest TE. Effectors within 1 kb of a TE (dark green) or more than one kb from the closest TE (light green) and the expression of the two categories is shown. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.

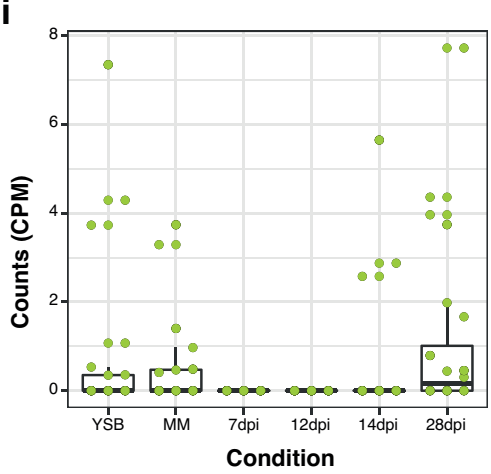
Ai



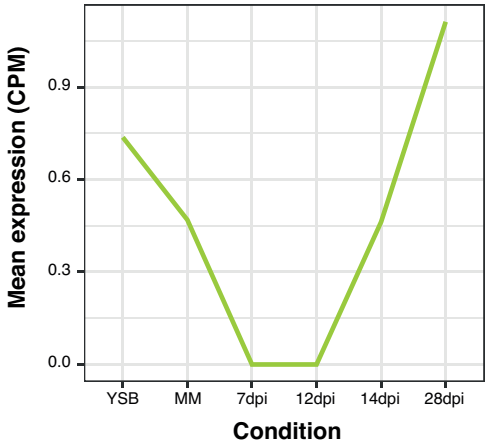
Aii



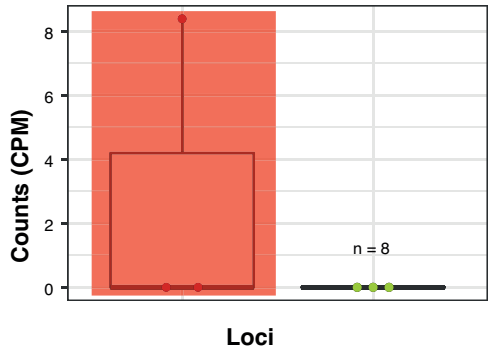
Aiii



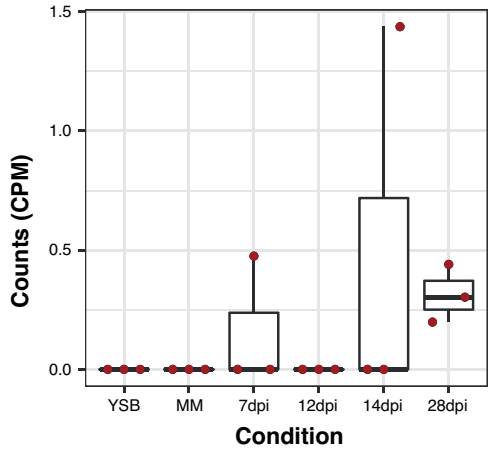
Aiv



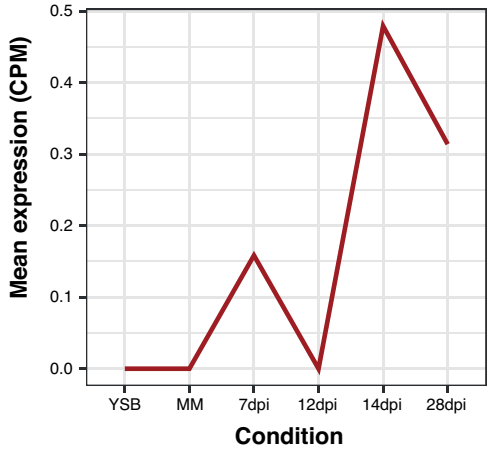
Av



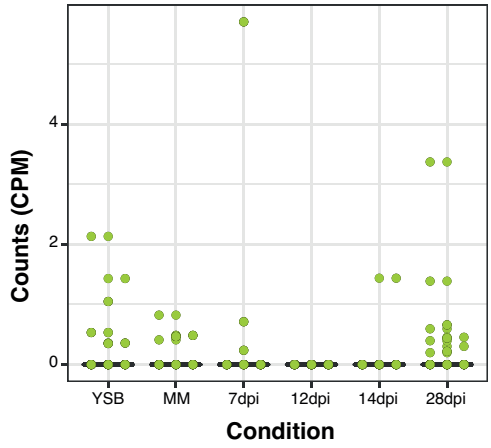
Bi



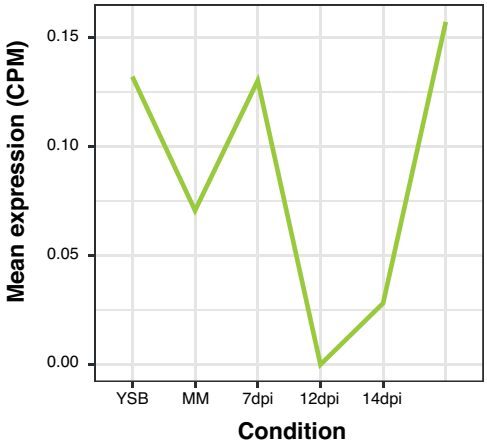
Bii



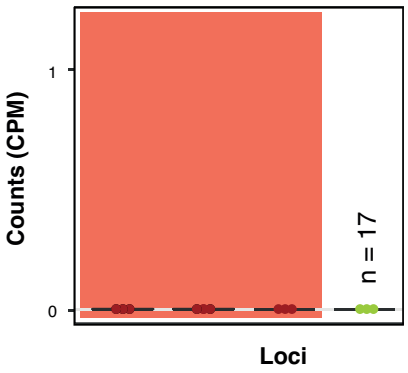
Biii



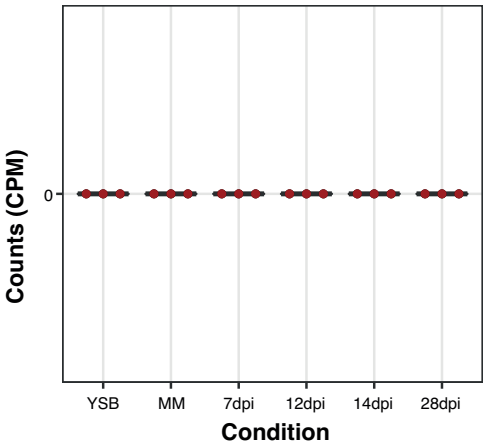
Biv



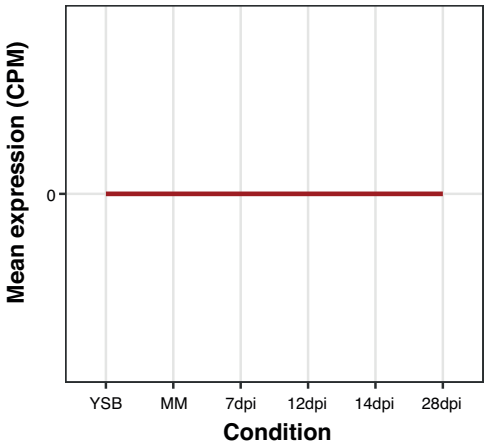
Bv



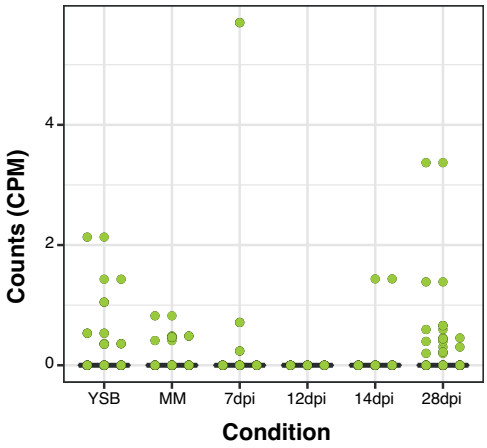
Ci



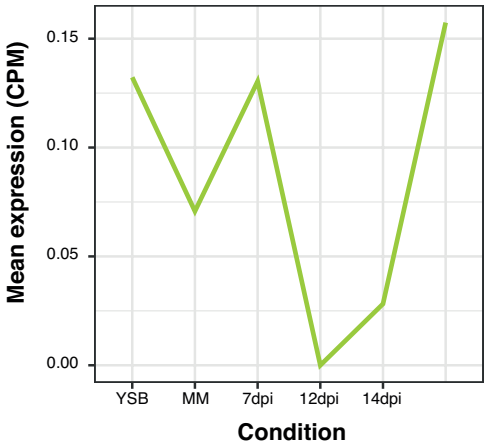
Cii



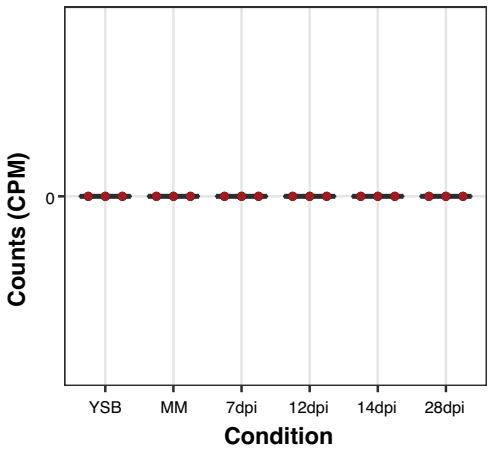
Ciii



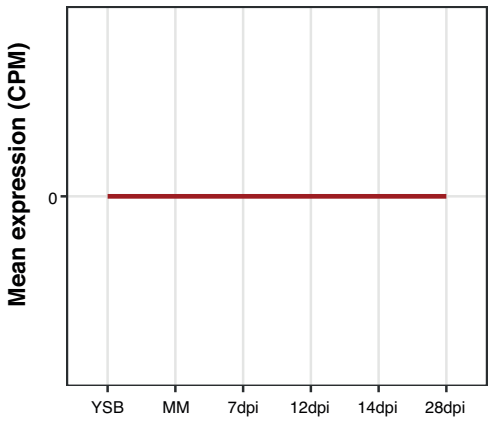
Civ



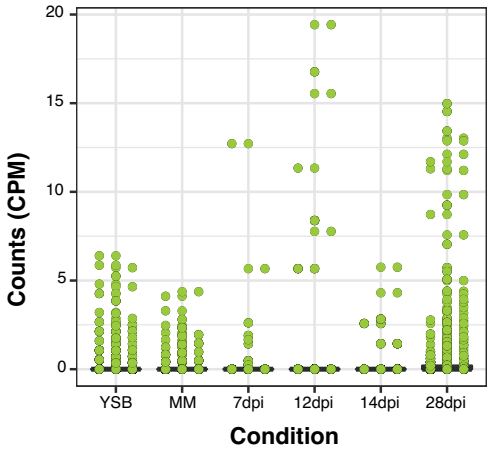
Di



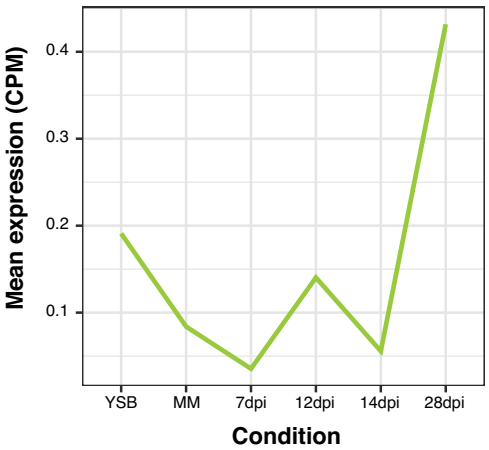
Dii

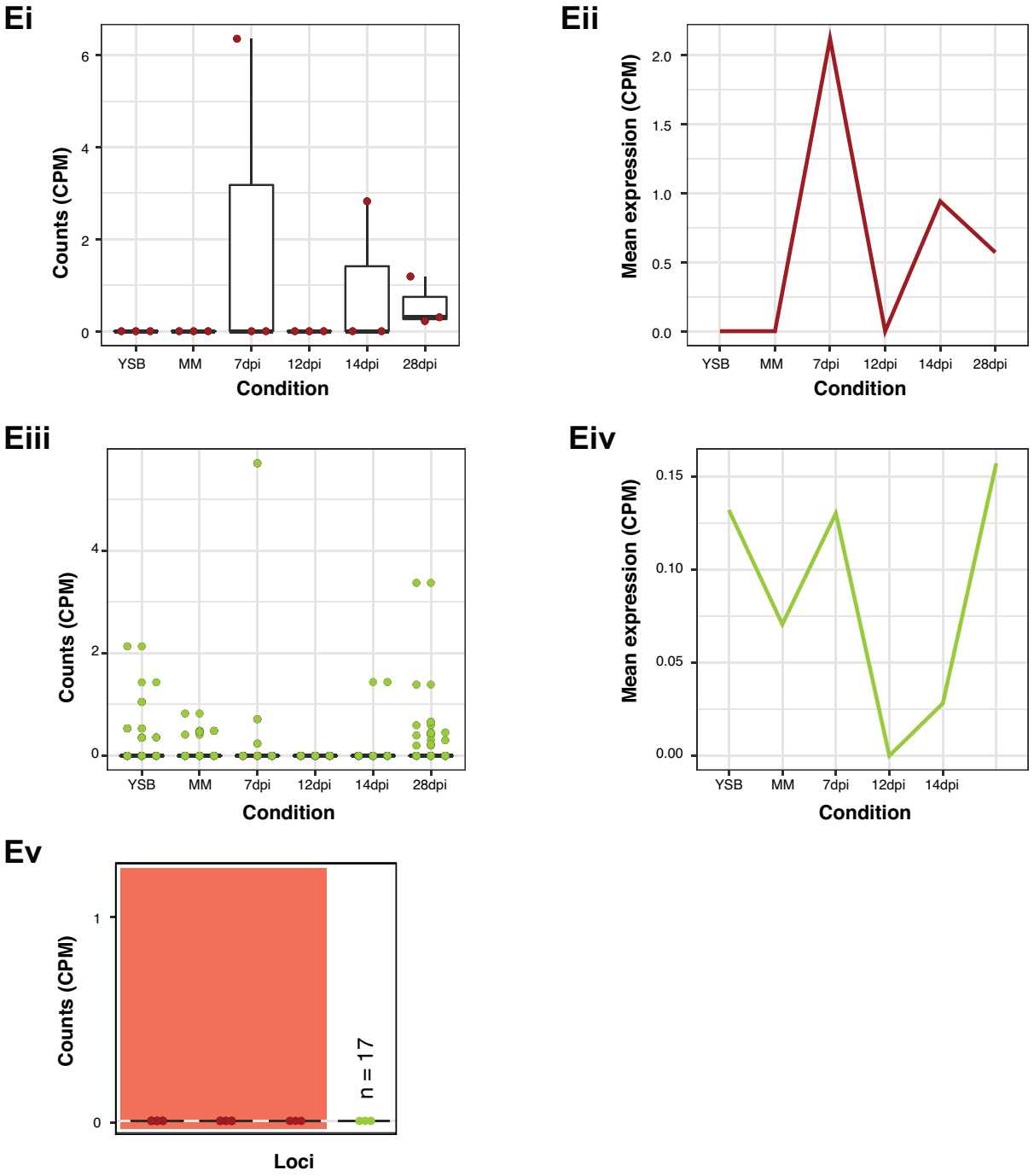


Diii



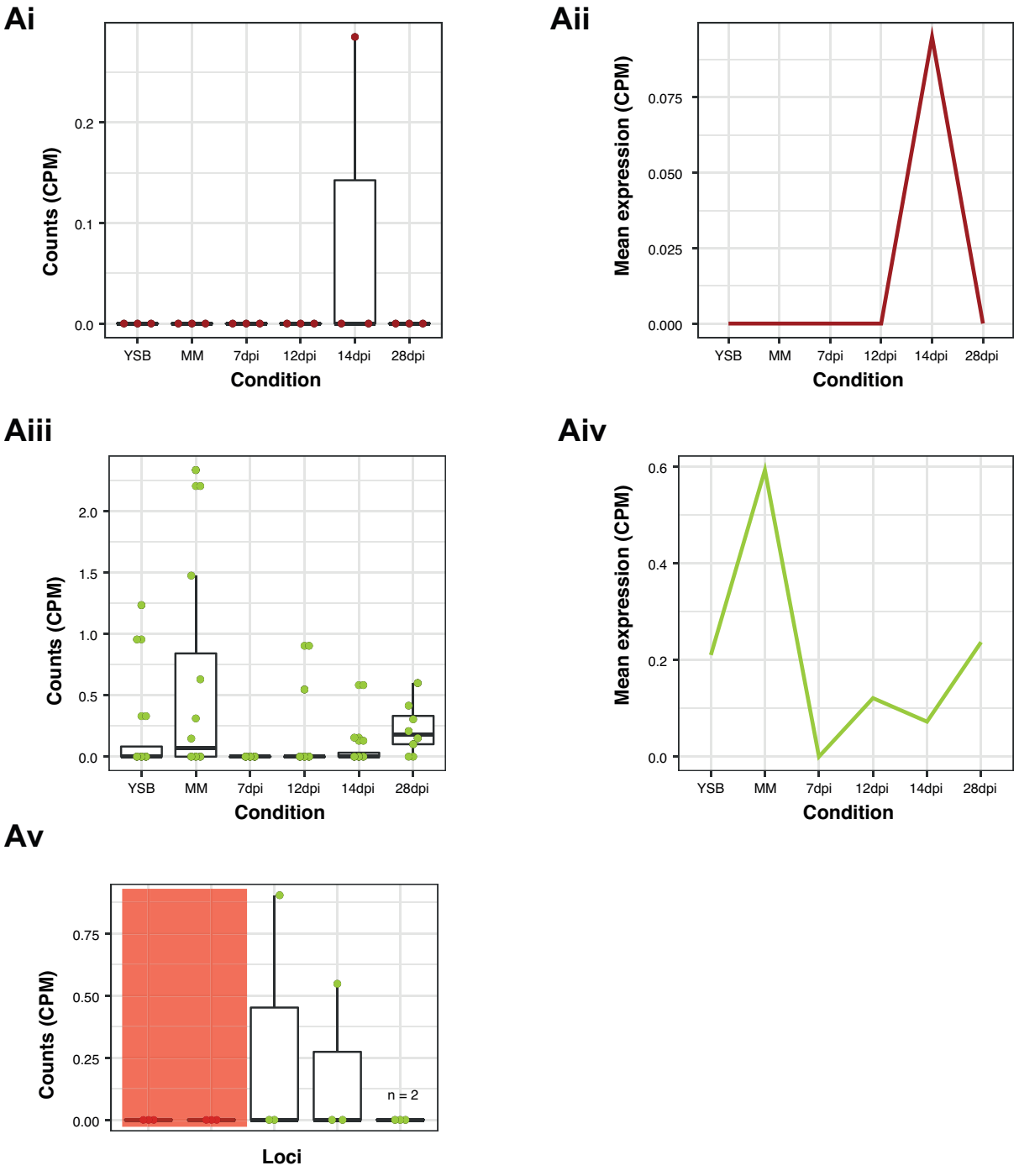
Div



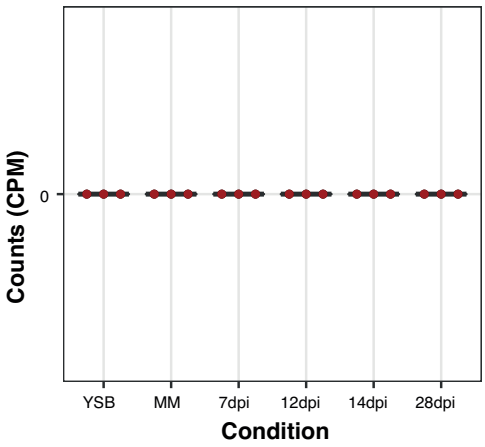


Supplementary Figure S14: TE de-repression of individual TEs inserted in proximity to the effector gene *Avr3D1* in 3D1. (A) TRIM-*Mutator* element 2, (B) Unknown element 8, (C) Unknown element 8, (D) LINE-*I* element 1, (E) Unknown element 8, (F) LTR-*Copia* element 1 and showing, (i.) Locus-specific expression of the TE found near Avr3D1. (ii.) Mean locus-specific expression of the TE found near Avr3D1. (iii.) Expression of all TE copies outside of the *Avr3D1* locus. (iv.) Averaged expression levels of all TE copies outside of the *Avr3D1* locus. (v.) Expression levels at 12 days post-infection (dpi) of all identified copies across the genome using uniquely mapped reads (expression levels of some copies were summarized). Variation is expressed by showing uniquely mapped reads from each individual replicate. The TE copy highlighted in red corresponds to the copy inserted

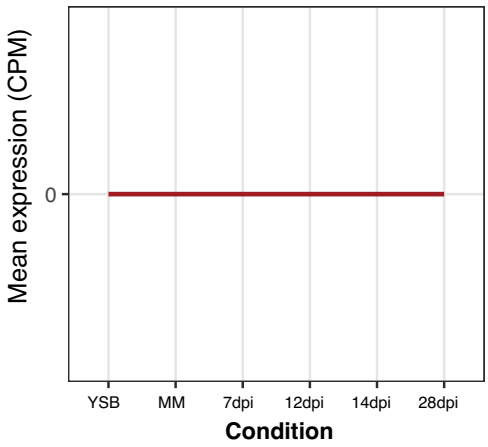
nearby *Avr3D1*. dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.



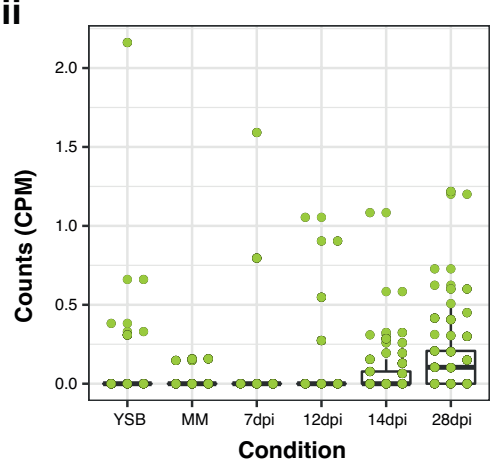
Bi



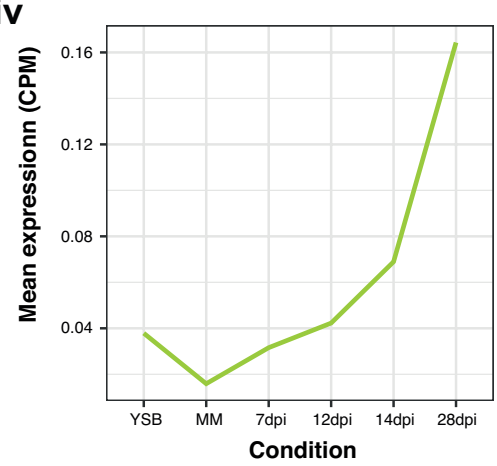
Bii

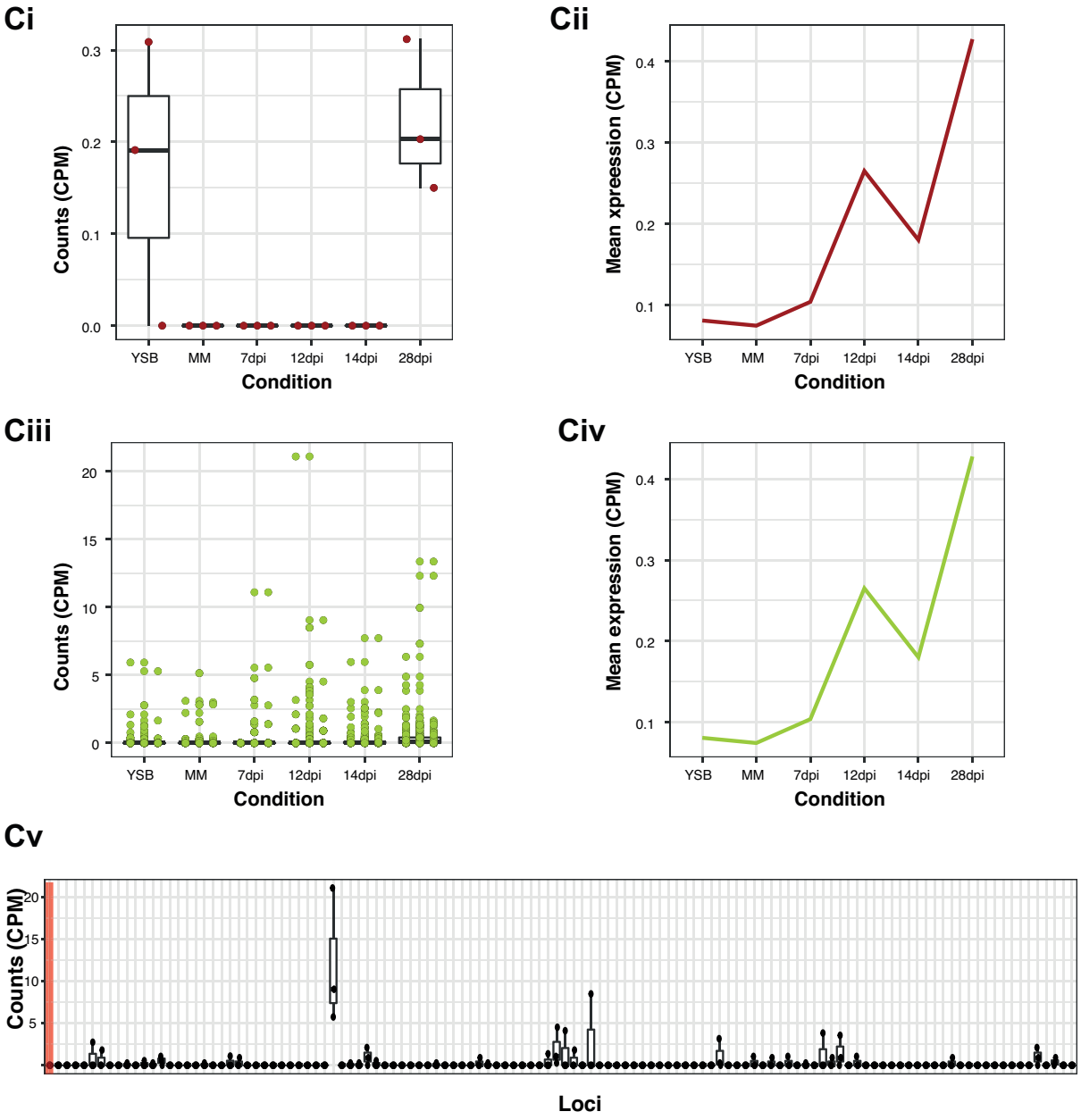


Biii

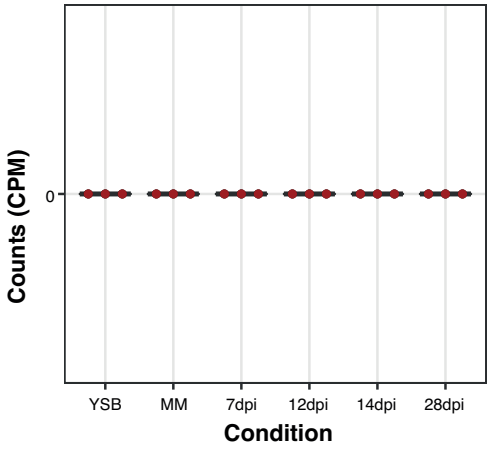


Biv

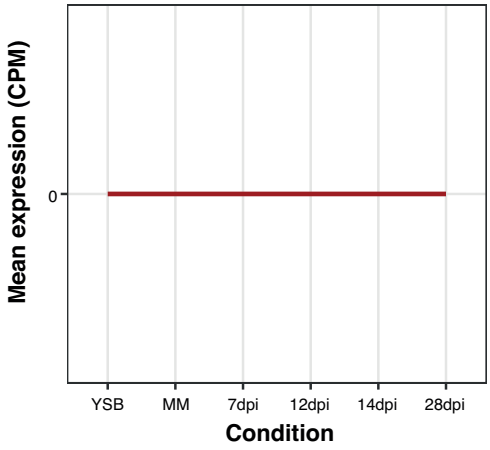



Ci
Counts (CPM)
0.0
0.1
0.2
0.3
YSB
MM
7dpi
12dpi
14dpi
28dpi
Condition
Cii
Mean xpression (CPM)
0.1
0.2
0.3
0.4
YSB
MM
7dpi
12dpi
14dpi
28dpi
Condition
Ciii
Counts (CPM)
0
5
10
15
20
YSB
MM
7dpi
12dpi
14dpi
28dpi
Condition
Civ
Mean expression (CPM)
0.1
0.2
0.3
0.4
YSB
MM
7dpi
12dpi
14dpi
28dpi
Condition
Cv
Counts (CPM)
5
10
15
20
Loci

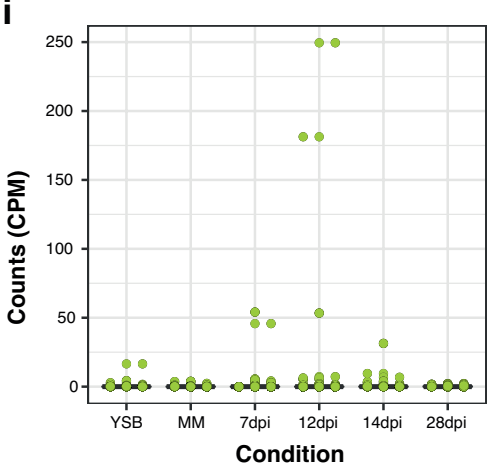
Di



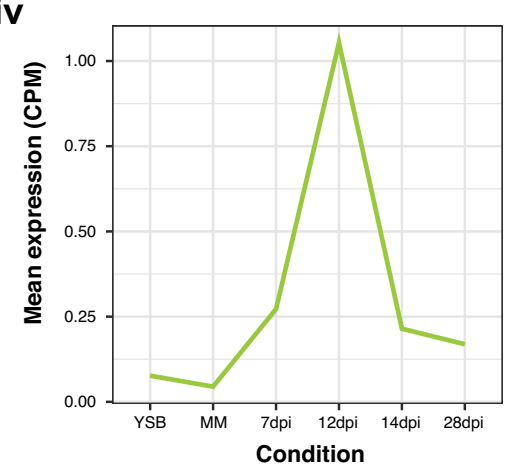
Dii



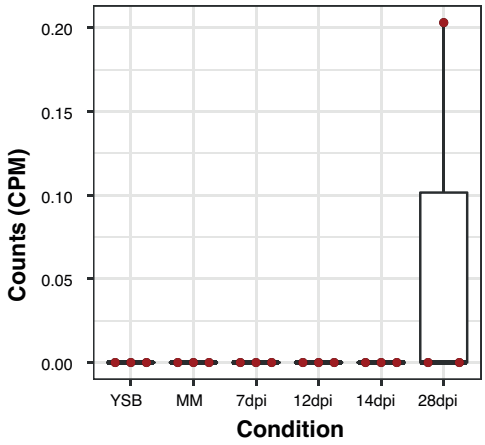
Diii



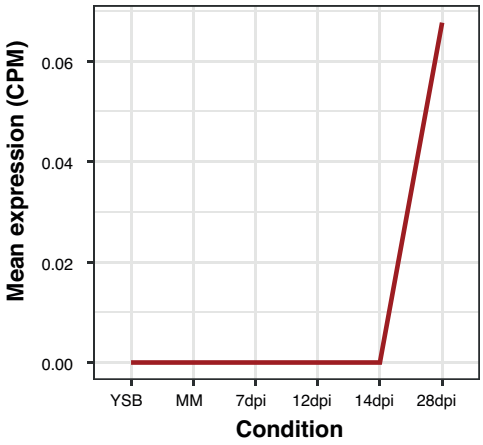
Div



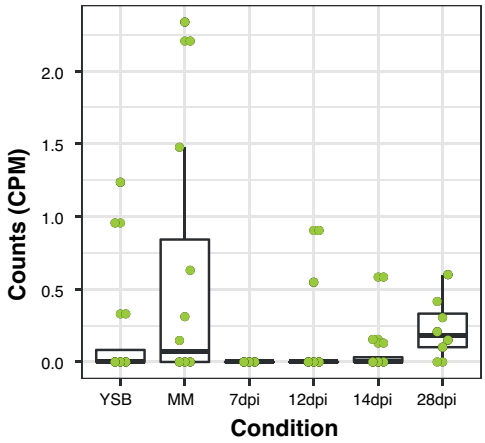
Ei



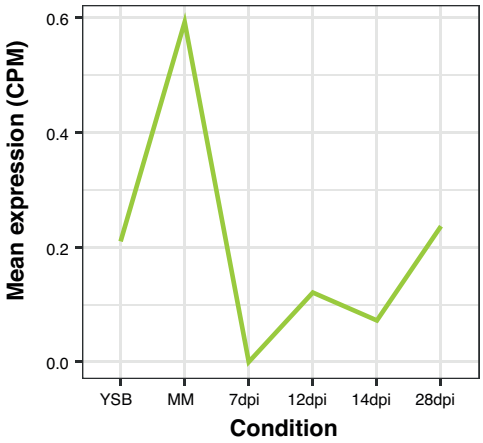
Eii



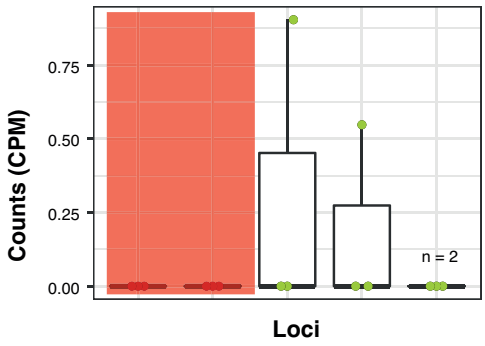
Eiii



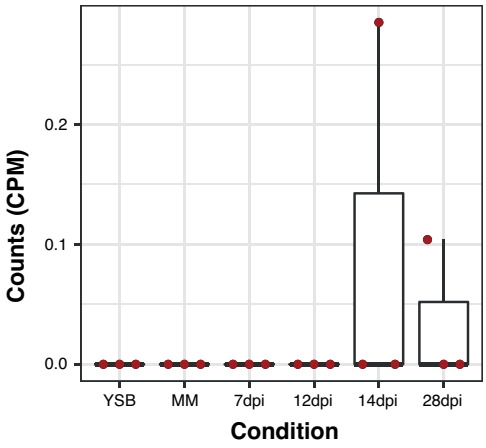
Eiv



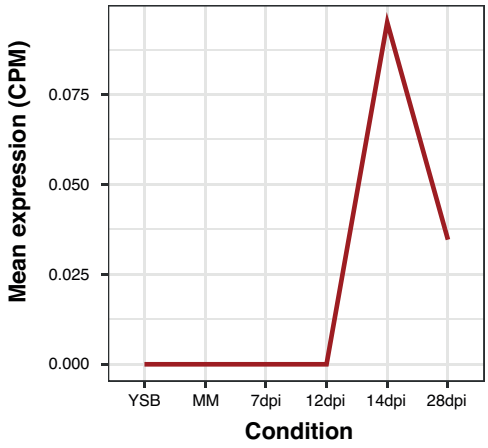
Ev



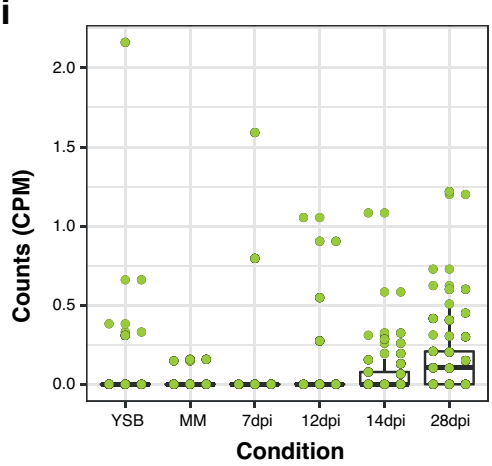
Fi



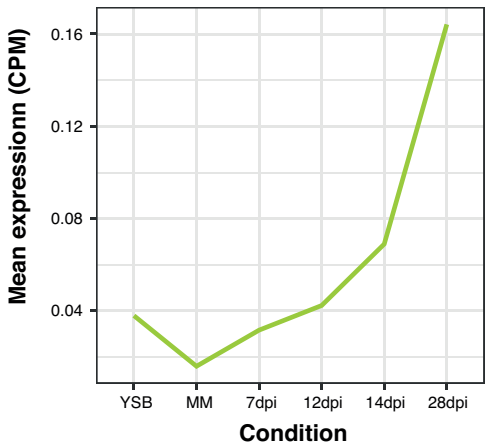
Fii



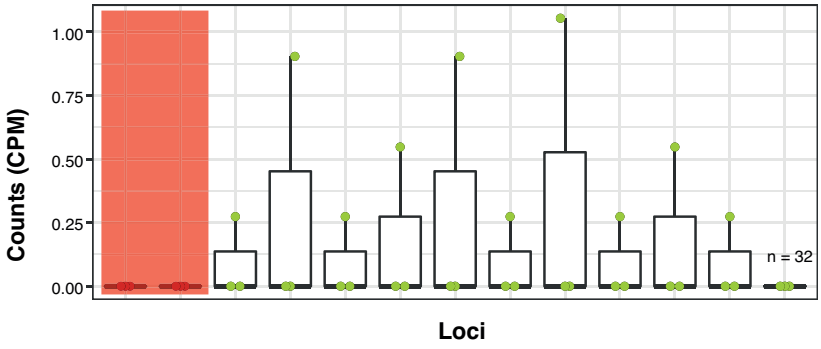
Fiii

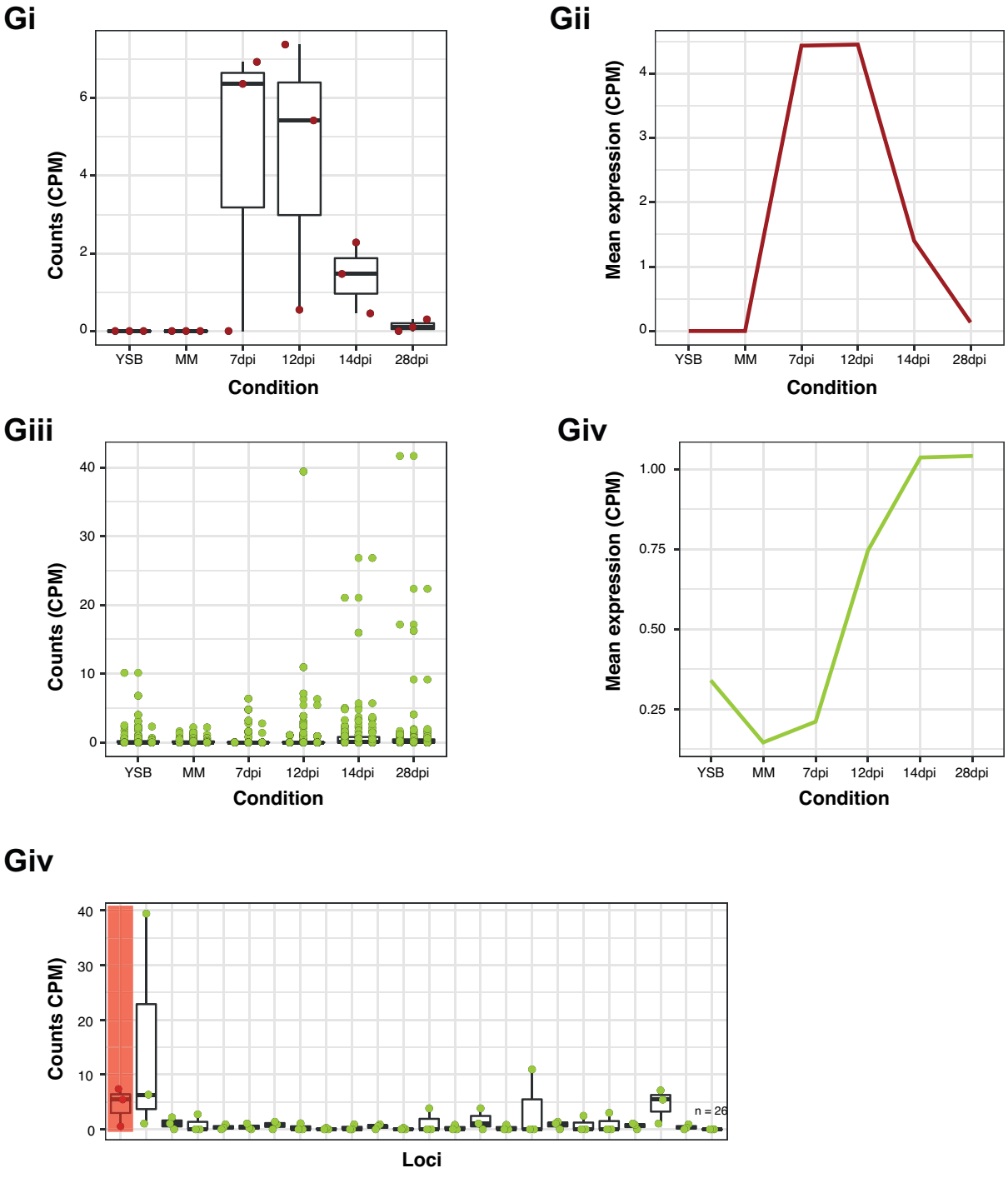


Fiv



Fv





Supplementary Figure S15: TE de-repression of individual TEs inserted in proximity to the effector gene *Avr3D1* in 3D1. (A) Crypton element 1, (B) LTR-*Gypsy* element 5, (C) LTR-*Gypsy* element 3, (D) LTR-*Gypsy* element 2, (E) Crypton element 1, (F) LTR-*Gypsy* element 5, (G) LTR-*Gypsy* element 11 and showing (i.) Locus-specific expression of the TE found near Avr3D1. (ii.) Mean locus-specific expression of the TE found near Avr3D1. (iii.) Expression of all TE copies outside of the *Avr3D1* locus. (iv.) Averaged expression levels of all TE copies outside of the *Avr3D1* locus. (v.) Expression levels at 12 days post-infection (dpi) of all identified copies across the genome using uniquely mapped reads (expression levels of some copies were summarized). Variation is expressed

by showing uniquely mapped reads from each individual replicate. The TE copy highlighted in red corresponds to the copy inserted nearby *Avr3D1.* dpi: days post-infection, MM: nutrient poor media, YSB: nutrient rich media, CPM: counts per million.

Supplementary Files

Supplementary File S1: Consensus sequence of the MITE-*Undine* (Grandaubert et al. 2015).

>MITE-Undine

TTGTTTGTTTGTGTTTGTTTGTTTGTTTGTTCCATTTACGTCCTTTCGGAGTCCAAGACT

ATGTAGGACGGCTTTGCCATAAAGGCAAAGCAGTAGATCTAGTTACAATTAGATTCTTT

CGGTATTCTTTGACCTTAGGGGAGACCCCGTGCCTAAGGCAAGACCTGCAAGCAGAAT

CTGCTAAAGGACTTTAAACAAATGCGAGACAAAGGAAATTAAACGAGTGTTGCTGCAGC

CTGCAGCCGGGGCTGCAGAAGTCTACAGCCGGGGCTGCAGAAGTCTACAGCCAAGGC

TGCAGAATTTTCGCGCCAAGCTTTTTGCAGAATTTTCAGCGGATCCTTCCGCTTCGCAC

ACACGCTGCGAATTCGAAAGAGATGTAGCGCAAAAGCAGTGCAAAATGCCGTGCACAA

GCAGTGCAAAATGCAGTGCAAAAGCAGTGCAAAAGCAGAAGCAGCTGTCTAGACTCGC

TGCCTTCTCTTGTCCTTAATATGCCTTTGTAGAACTATATGCCGCTTTCTAACGCATATA

GCTACGCAGGTTAATCGCAGAGTTGTAGATTCTACCGACTGTTCCAGCTCTTTAAAGTG

CATGTATGCTTCCTTTCGAGGGCGACGAAGTCTCTACTAGAGAAGGGACATGCGATTTC

GATTCGGATTTCGAGGAGGTGCGCGTATAGGGCAGAGTAGACTGCTTCTTCTTCTTCTT

CTTCTCGTTCTCGTTCGTAGCGTAGAGGGCTATTCGTTCGAATCGTGCTGTTGTTCTGT

TGTAGCTTGTATTTCGATGGGAGGCGGATTTCGAGAAGGATTTCGGGCGGATGTCGAG

AGGTTTTCGAGGAGGTTTCTTCCAGACTGTCTTAGTTTCCCTTCTGAGGAAGGTTGGCT

CTGCAACGACGATAGTATCGCAGACTTACGACGGCCACCGCGTAAAGGTGCAGTCTGT

CGTCGGCCGTCT

Supplementary Tables

Supplementary Table S1: Saturation analysis of multiple mapped reads

Sample	OutFilterMultimapNmax	winAnchorMultimapNmax	PercUniqueMappedReads	PercMultiMappedReads	Condition	Isolate
MM_1	10	20	89.4	8.53	MM	1A5
MM_2	50	100	89.4	8.75	MM	1A5
MM_3	100	200	89.4	8.75	MM	1A5
MM_4	200	400	89.39	8.76	MM	1A5
MM_5	1000	2000	89.07	8.6	MM	1A5
YSB_1	10	20	96.57	2.56	YSB	1A5
YSB_2	50	100	96.57	2.7	YSB	1A5
YSB_3	100	200	96.56	2.7	YSB	1A5
YSB_4	200	400	96.55	2.71	YSB	1A5
YSB_5	1000	2000	96.18	2.68	YSB	1A5

Supplementary Table S2: The average distance of the most highly expressed DTX-MITE-*Undine* element family from the nearest gene in all four strains.

Isolate	**Mean distance of DTX-MITE-dragon to nearest gene**	**Mean distance of TE families to the nearest gene**	**Mean distance of TE families to genes in all genomes combined**
1A5	21153.07	19879.56	20255.75
1E4	33795.44	20577.36	20255.75
3D1	17652.62	19464.37	20255.75
3D7	19737.04	2114. 39	20255.75

Supplementary Table S3: Gene ontology terms of genes in co-expression profiles.

Available online: https://doi.org/10.1093/molbev/msz216

Supplementary Table S4: The mean distance of transposable element superfamilies to the closest gene.

Available online: https://doi.org/10.1093/molbev/msz216

Supplementary Table S5: Numbers of transposable element superfamilies within genes.

Available online: https://doi.org/10.1093/molbev/msz216

Supplementary Table S6: Numbers of transposable element superfamilies within 1 kb of genes.

Available online: https://doi.org/10.1093/molbev/msz216

Supplementary Table S7: Normalized locus-specific TE expression in 3D1.

Available online: https://doi.org/10.1093/molbev/msz216

Supplementary Table S8: Normalized locus-specific TE expression in 3D7.

Available online: https://doi.org/10.1093/molbev/msz216

CHAPTER 2:

Meiosis leads to pervasive copy-number variation and distorted inheritance of accessory chromosomes of the wheat pathogen *Zymoseptoria tritici*

Simone Fouché, Clémence Plissonneau, Bruce A. McDonald , Daniel Croll

Accepted for publication in *Genome Biology and Evolution* (May 2018), in press, doi: https://doi.org/10.1093/gbe/evy100

Data deposition: This project has been deposited at NCBI under the SRA accession numbers SRS383146 (3D1), SRS383147 (3D7), SRS383142 (1A5), and SRS383143 (1E4) and the NCBI BioProject accession numbers PRJNA256988 and PRJNA256991.

Abstract

Meiosis is one of the most conserved molecular processes in eukaryotes. The fidelity of pairing and segregation of homologous chromosomes has a major impact on the proper transmission of genetic information. Aberrant chromosomal transmission can have major phenotypic consequences, yet the mechanisms are poorly understood. Fungi are excellent models to investigate processes of chromosomal transmission, because many species have highly polymorphic genomes that include accessory chromosomes. Inheritance of accessory chromosomes is often unstable and chromosomal losses have little impact on fitness. We analyzed chromosomal inheritance in 477 progeny coming from two crosses of the fungal wheat pathogen *Zymoseptoria tritici*. For this, we developed a high-throughput screening method based on restriction site-associated DNA sequencing that generated dense coverage of genetic markers along each chromosome. We identified rare instances of chromosomal duplications (disomy) in core chromosomes. Accessory chromosomes showed high overall frequencies of disomy. Chromosomal rearrangements were found exclusively on accessory chromosomes and were more frequent than disomy. Accessory chromosomes present in only one of the parents in an analyzed cross were inherited at significantly higher rates than the expected 1:1 segregation ratio. Both the chromosome and the parental background had significant impacts on the rates of disomy, losses, rearrangements, and distorted inheritance. We found that chromosomes with higher sequence similarity and lower repeat content were inherited more faithfully. The large number of rearranged progeny chromosomes identified in this species will enable detailed analyses of the mechanisms underlying chromosomal rearrangement.

Introduction

Sexual reproduction requires chromosomes to undergo meiosis, whereby homologous chromosomes pair, recombine, and finally separate and migrate to opposite poles of the meiotic cell. Meiosis is a highly conserved process initiated by the pairing of homologous chromosomes that first recognize one another and then establish recombination-dependent links between homologs to form the synaptonemal complex (reviewed in Roeder 1997). This is followed by two divisions, first to separate homologous chromosomes and then to separate sister chromatids. While accurate pairing of homologs is essential for the faithful segregation of chromosomes (Naranjo 2012), chromosomes can pair along their entire length or in a segment-specific manner where only some regions align (Roeder 1997). This suggests that the length and degree of sequence similarity can affect homolog identification and pairing. After pairing, recombination produces crossovers that physically link homologs, mediate proper segregation, and thereby preserves chromosomal integrity (Mather 1938; Baker et al. 1976; Hassold and Hunt 2001). Recombination between misaligned repetitive sequences can generate length variation among the daughter chromosomes (Montgomery et al. 1991). After pairing and recombination, segregation occurs via centromeres that bind to chromosome proteins and mediate accurate segregation to the opposite poles of the cell.

Aberrant transmission of chromosomes from one generation to the next, including partial and whole chromosome duplications or losses, are caused largely by erroneous pairing during meiosis. Such duplication and loss events can affect a large number of genes and alter gene expression across the genome (Harewood and Fraser 2014). The most dramatic copy-number variation is aneuploidy. Unequal sets of chromosomes result from nondisjunction and are the leading genetic cause of miscarriages in humans (Hassold and Hunt 2001). Atypical phenotypes associated with aneuploid states are caused by gene

dosage imbalances that can cause severe defects (Torres et al. 2008). In general, aneuploidy and chromosomal rearrangements are associated with lower fitness (Torres et al. 2008), but in rare circumstances, errors during meiosis can provide adaptive genetic variation. For example, in the human pathogenic fungi *Cryptococcus neoformans* and *Candida albicans*, specific aneuploidies contribute to drug resistance (Selmecki et al. 2006, 2008; Sionov et al. 2010; Ngamskulrungroj et al. 2012). Adaptive aneuploidy is frequently associated with response to stressful environments (Chen et al. 2012). The dosage imbalance and altered stoichiometry due to additional copies of genes on a duplicated chromosome may not be beneficial under normal conditions, but can become beneficial under stress (Pavelka, Rancati, and Li 2010; Pavelka, Rancati, Zhu, et al. 2010). In pathogenic fungi, aneuploidy often occurs for only a restricted number of chromosomes, however the mechanisms determining the rate of aneuploidy generation and its maintenance are poorly understood.

Aneuploidy also plays an important role in several plant pathogenic fungi. Several important plant pathogens have highly dynamic genomes with chromosomes that show significant length and number polymorphisms within the species. This chromosomal plasticity is often restricted to a well-defined set of accessory chromosomes. This bipartite genome structure, characterized by an accessory genome region that is rapidly diversifying and a core genome region that remains conserved, can be associated with the trajectory of pathogen evolution (Croll and McDonald 2012; Dong et al. 2015). The accessory region is often rich in transposable elements that drive chromosomal rearrangements, deletions, and duplications (Zhang et al. 2011). Accessory chromosomes are not shared among all members of a species, therefore these chromosomes can contribute significantly to polymorphism within a species. Importantly, many plant pathogens have been shown to harbor pathogenicity loci on accessory chromosomes (Möller and Stukenbrock 2017). In contrast, the core regions

encode essential functions required for survival and reproduction. Plant pathogenic fungi provide particularly powerful models to investigate factors affecting the transmission of chromosomes through meiosis because of their extreme chromosomal plasticity, the ubiquity of sexual reproduction, and their experimental tractability.

The fungal wheat pathogen *Zymoseptoria tritici* provides a striking example of genome plasticity. The bipartite genome consists of 13 core and up to eight accessory chromosomes that exhibit significant length polymorphism within and among field populations (Goodwin et al. 2011; Croll and McDonald 2012). Chromosomal rearrangements played an important role in adaptation to different host genotypes (Hartmann et al. 2017). The accessory chromosomes are highly unstable through meiosis and were shown to undergo rearrangements, segregation distortion, and nondisjunction (Wittenberg et al. 2009; Croll et al. 2013). *Z. tritici* reproduces sexually when hyphae originating from two haploid spores of opposite mating type fuse to produce a transient diploid stage that undergoes two rounds of meiosis followed by one round of mitosis to produce eight ascospores in an ascus. The pathogen tolerates aneuploidy, so chromosomal rearrangements generated through this process in both the core and accessory genomes can remain viable (Wittenberg et al. 2009; Croll et al. 2013; Schotanus et al. 2015). Hence, this species is an ideal model to analyze patterns of aberrant chromosomal transmission.

In this study, we analyze the mechanisms that affect the fidelity of chromosomal inheritance through meiosis, including identification of chromosomal rearrangements, losses, and duplications. For this, we screened hundreds of progeny genotypes generated from two independent crosses and determined the rate of aneuploidy, patterns of rearrangement and distortions in transmission rates. Finally, we investigated whether factors such as length

similarity, synteny, recombination rate, and repetitive element content affected the fidelity of chromosomal inheritance.

Materials and Methods

Generation of sexual crosses

Two crosses were performed between four parental *Z. tritici* isolates collected from two Swiss wheat fields separated by ~10 km. Isolate ST99CH3D1 was crossed with isolate ST99CH3D7 (hereafter abbreviated 3D1 and 3D7) and isolate ST99CH1A5 was crossed with isolate ST99CH1E4 (abbreviated 1A5 and 1E4), producing 359 and 341 haploid ascospore progeny, respectively. The genomes of all four parental isolates were sequenced using Illumina technology (Torriani et al. 2011) and are available under the NCBI SRA accession numbers SRS383146 (3D1), SRS383147 (3D7), SRS383142 (1A5), and SRS383143 (1E4). The parental isolates were already genetically characterized and have been phenotyped for virulence and many other traits (Zhan et al. 2005; Croll et al. 2013). Full sib families were produced by coinfecting wheat leaves with asexual conidia from the parental strains of opposite mating types using the crossing protocol described by Kema et al. (1996). Briefly, spores of a pair of parents were sprayed onto wheat plants and incubated outdoors for 40–60 days until well-developed symptoms including pseudothecia were observed. Ascospores were isolated from pseudothecia over a period of several days by placing the infected wheat leaves on wet filter paper inside Petri dishes. Leaves were covered with upside down Petri dish lids that were previously filled with water agar. This setup allowed us to capture ascospores that were vertically ejected from mature pseudothecia. Released ascospores were left to germinate on the water agar to enable inspection for potential contaminants and to ensure that only progeny resulting from single ascospores were selected. Germinating ascospores were transferred to individual culture plates for clonal propagation. The mycelium produced by each success- fully germinated ascospore was used for DNA extraction and plant infection experiments. Offspring mycelium was produced in YSB (yeast sucrose broth) liquid media for 6–7 days prior to DNA extraction.

Reference alignment using restriction site-associated DNA sequencing

We used Restriction Site-Associated DNA Sequencing (RADseq) (Baird et al. 2008) for large-scale sequence genotyping as described previously (Croll et al. 2015). Briefly, the RADseq protocol (Etter et al. 2011) was applied to *Z. tritici* by using the *PstI* restriction enzyme to digest 1.3 μg of DNA extracted with the DNAeasy plant mini kit (QIAGEN Inc., Basel, Switzerland) for each offspring. After digestion and adapter annealing, the pooled DNA was sequenced on an Illumina HiSeq2000 using a paired- end 100-bp library. Pools contained ~132 progeny, six different Illumina TruSeq compatible P2 adapters and 22 P1 adapters with unique barcodes. Progeny DNA with the same P2 adapter were distinguishable by using the unique barcodes ligated to the P1 adapters.

Illumina reads were quality trimmed using Trimmomatic v. 0.30 (Bolger et al. 2014) and separated into distinct sets for each progeny based on the P1 adapter using FASTX toolkit v 0.13 (http://hannonlab.cshl.edu/fastx_toolkit/; last accessed March 2015). Reads were aligned to the gapless telomere to telomere IPO323 reference genome (assembly version MG2, September 2008) (Goodwin et al. 2011) with the short-read aligner version of bowtie 2.1.0 (Langmead and Salzberg 2012) using the default parameters for sensitive end-to-end alignment (-D 15; -R 2; L- 22; -I S, 1, 1.15). The same parameters from trimming and reference assembly were used to align the four parental genome sequences (Croll et al. 2013) to the reference genome (IPO323). RADseq aligned reads are available under the NCBI BioProject accession numbers PRJNA256988 and PRJNA256991. Potential clones were identified as genotypes sharing >90% identity based on single nucleotide polymorphism (SNP) analyses as previously described (Lendenmann et al. 2014). Only one randomly selected progeny per clonal group was kept for further analyses, reducing the number of progeny to 263 in the 3D1x3D7 cross and to 261 in the 1A5x1E4 cross.

Determining chromosome number and length polymorphisms based on coverage

Restriction sites cut by *PstI* were identified in silico using the EMBOSS restrict program (http://www.bioinformatics.nl/cgi-bin/emboss/restrict; last accessed September 2016). Thereafter, the coverage of RADseq reads mapping to the restriction sites was determined using the BEDtools v. 2.25.0 intersectBed and coverageBed commands (Quinlan and Hall 2010). Reads were counted if the map- ping quality score was >= 20. The coverage of the sequenced parent genomes was determined following the same procedure. Progeny with a median read coverage of <20x were excluded from further analyses to avoid biases introduced by low-coverage data, resulting in fewer isolates being included in this analysis than in previous studies (Lendenmann et al. 2014, 2016; Stewart et al. 2018). We retained 249 progeny in the 3D1x3D7 cross and 228 isolates in the 1A5x1E4 cross. We used normalized read counts to detect chromosomal anomalies, where those with a normalized coverage close to zero (<0.3) were classified as missing, those with a normalized coverage close to one (>=0.7 and <1.3) were classified as present and those with a normalized coverage close to two (>=1.7) were classified as disomic (fig. 1A). Partially deleted and partially duplicated chromosomes were identified based on a normalized coverage ratio of >=0.3 and <0.7 or >1.3 and <1.7, respectively. Deviations from Mendelian inheritance for accessory chromosomes present in only one of the parents were determined using a chi-squared (χ^2) test.

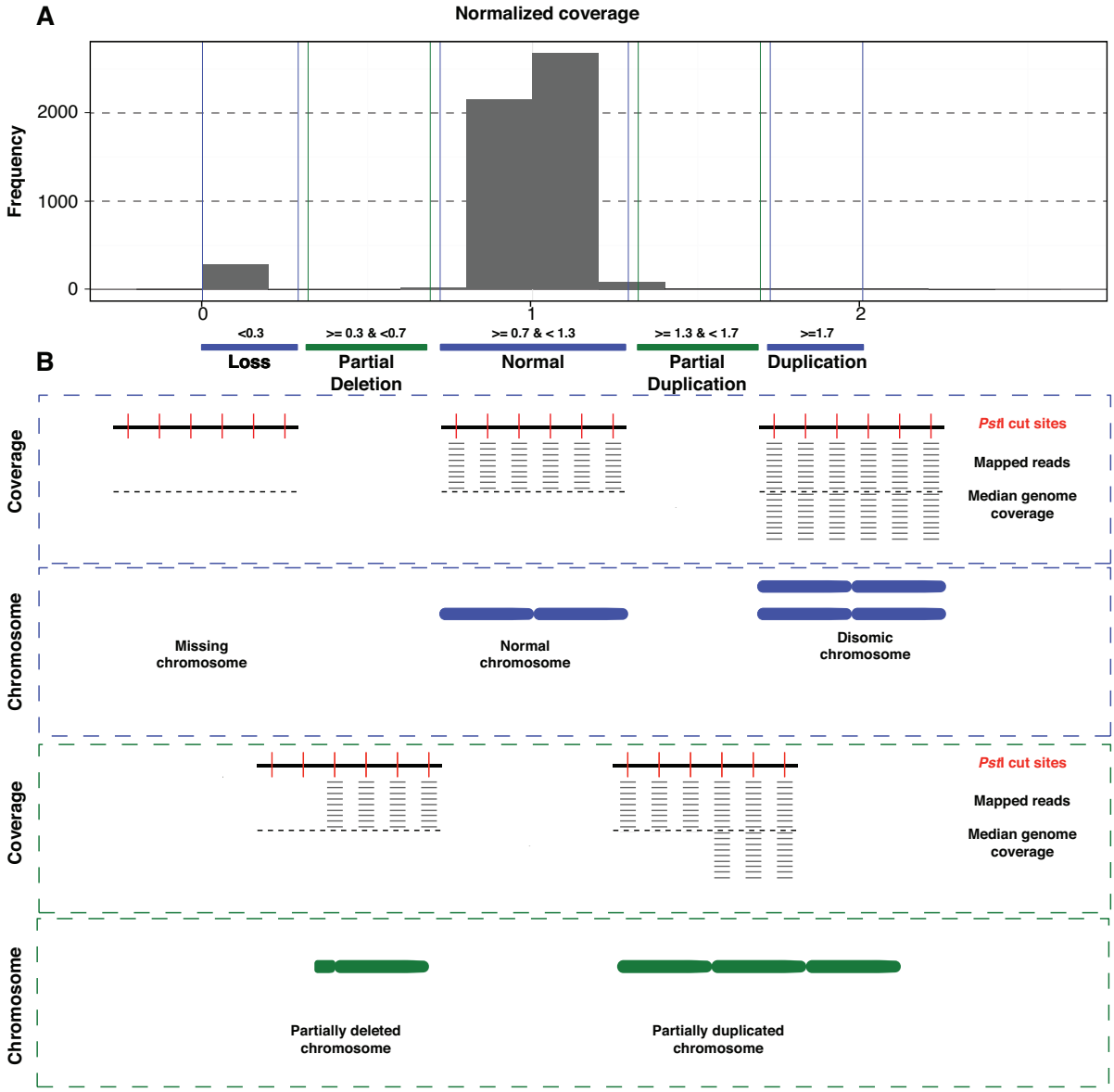


Figure 1: Procedure to detect chromosomal anomalies. (A) Reads mapped to the *PstI* restriction sites were used to analyze coverage across the genome. Sequencing data were generated by restriction site-associated DNA sequencing (RADseq). The normalized coverage represents the coverage of each chromosome normalized to the median coverage of all chromosomes of the same progeny. The normalized coverage distribution of progeny from cross 3D1x3D7 is shown with the cutoffs used to detect a whole chromosome loss (ratio < 0.3), partial deletion (ratio 0.3–0.7), normal transmission (0.7–1.3), partial duplication (1.3–1.7), and whole chromosome duplication (>1.7). (B) Schematic overview of read coverage expected for complete chromosome losses and duplications (in blue). Partial deletions and duplications are shown in green.

Distinguishing between homozygous and heterozygous disomy

SNP calling was performed using Freebayes (Version 1.0.2_1 1.1.0) (Garrison and Marth 2012) using the bam- files of each isolate mapped to the IPO reference genome. We used the parameters no-indels, no-mnps, no-complex, and ploidy 2. Then we filtered for sites that differed be- tween the parents (maf 0.2) and considered only these regions to determine whether disomic chromosomes originated from one or both parents. We also filtered for depth (minDP 30) and quality (minQ 30). The VCF tools - het function was used to determine the number of homozygous sites and the total number of sites. We determined the ratio of homozygous sites to the total number of sites and defined those with a ratio >0.6 as homozygous while those with a ratio <0.4 were defined as heterozygous. All other cases were considered to be ambiguous.

Chromosome instability and recombination rate, chromosome length, synteny and transposable element content of the parent chromosomes

We correlated chromosome instability with the percentage length difference in homologs among the parents and recombination rates based on the recombination rates reported in Croll et al. (2015). We also correlated synteny and the fidelity with which chromosomes were inherited using the NUCmer pipeline from MUMmer (version 3.23) software (Kurtz et al. 2004) to determine the sequence similarity between two homologous chromosomes. The minimum cluster length was set to 50 and we used the –mum option to anchor matches that were unique in both the reference and query sequence. The transposable elements (TEs) in the parent genomes were an- notated using RepeatMasker (http://www.repeatmasker.org; last accessed May 2017) and the TE library compiled for *Z. tritici* and its sister species (Grandaubert et al. 2015). The percentage of TEs on a chromosome was compared with the

likelihood of being inherited with high fidelity. We also compared the frequency of disomic chromosomes with the frequency of rearrangements for all chromosomes in both crosses.

Analyses of progeny phenotypes

Clonally propagated mycelium from each germinated ascospore was previously used to infect wheat plants in the framework of a QTL mapping study (Stewart et al. 2018). Progeny from both crosses were phenotyped for percentage of leaf area covered by lesions (PLACL), pycnidia density (pycnidia/ cm^2 leaf area), pycnidia size (mm^2), and pycnidia melanization on seedlings of the wheat cultivars Runal and Titlis in a previously described glasshouse-based assay (Stewart and McDonald 2014). Gray values were previously shown to be a good measure for melanization (Lendenmann et al. 2014). Replication of the infection assays was made possible by inoculating replicate wheat plants with a fixed concentration of blastospores from each progeny mycelium. The assay was repeated three times over three consecutive weeks, resulting in three biological replicates and six total replicates per progeny-cultivar pair. Automated image analysis of the second leaf was performed at 23dpi as previously described (Stewart and McDonald 2014). Progeny were also phenotyped for temperature sensitivity, growth morphology, and fungicide sensitivity (Lendenmann et al. 2015, 2016). Phenotypes were compared in normal progeny and progeny with “abnormal” (partially deleted, partially duplicated, disomic, or absent) chromosomes to determine if particular chromosome genotypes were associated with outlier virulence, fungicide resistance, temperature sensitivity, or growth rate phenotypes. These analyses were performed in R version 3.4.0.

Results

Mapping RADseq reads to the reference genome

The chromosome state (absent, present, or duplicated) was determined for each chromosome of the four haploid parental isolates (3D1, 3D7, 1A5, 1E4) and 477 progeny, using RADseq reads generated for each progeny mapped to the IPO323 reference genome. The 3D1 and 1A5 parents had all 21 chromosomes, while the 3D7 and 1E4 parents were missing four and one accessory chromosomes, respectively (Croll et al. 2013). None of the four parental strains carried additional chromosomes beyond the 21 chromosomes identified in IPO323 (Plissonneau et al. 2018). We selected the parental isolate from each cross that carried all 21 chromosomes (3D1 and 1A5) as a reference. We mapped whole- genome sequencing data of the two selected parents against the IPO323 reference genome and identified regions missing in the parental genomes. Missing regions were not expected to show coverage in any of the progeny chromosomes and were excluded from further analyses. RADseq loci genotyped in progeny showed an even distribution across all 21 chromosomes, with no apparent differences between core (1–13) and accessory chromosomes (14–21; supplementary fig. S1). Similarly, RADseq loci showed homogeneous read coverage across the genome for progeny with high and low overall sequence coverage in both crosses (supplementary fig. S2A–D). For each progeny, we calculated the coverage for each chromosome and compared this to the median coverage of all chromosomes for that isolate (fig. 1). The normalized coverage per chromosome was close to 1 for the large majority of the chromosomes (supplementary fig. S3). The mean normalized coverage ratio was 0.96 and 0.95 for the progeny from cross 3D1x3D7 and cross 1A5x1E4, respectively.

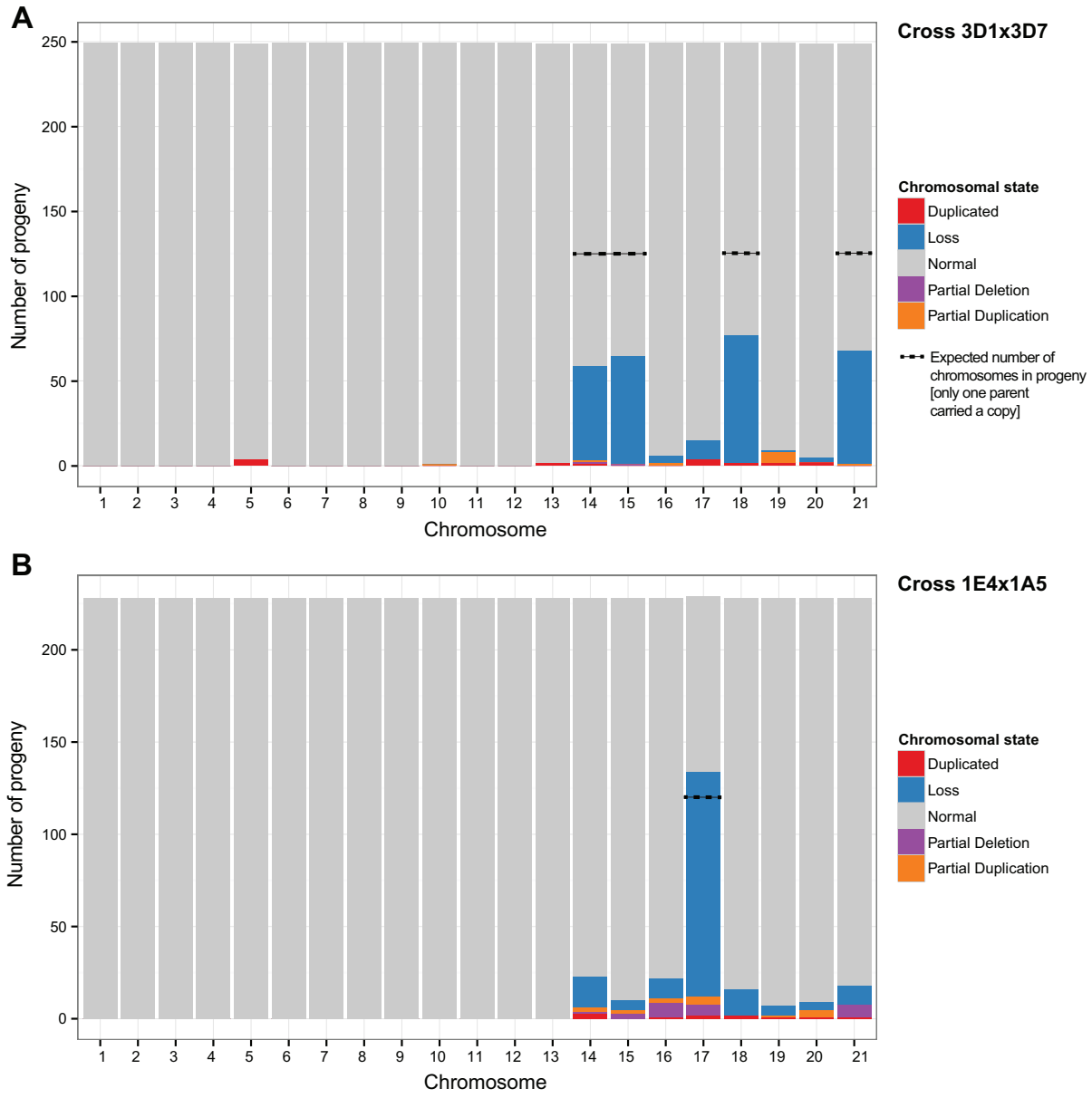


Figure 2: Summary of the total chromosome anomalies in the progeny of two crosses. Normal, disomic, lost, and rearranged (partially duplicated or deleted) chromosomes are shown separately for cross 3D1x3D7 and 1E4x1A5. Dotted lines show the expected number of progeny for chromosomes that were present in only one of the two parental isolates.

Patterns of chromosome transmission in the two crosses

Analyzing normalized read coverage among progeny revealed high rates of chromosome losses in both crosses. In cross 3D1x3D7, accessory chromosomes 16, 17, 19, and 20 were

present in both parents but were missing in 1.6% (4/ 249), 4.4% (11/249), 0.4% (1/249), and 1.2% (3/249) of the progeny, respectively (fig. 2A). In the 1E4x1A5 cross, accessory chromosomes 14, 15, 16, 18, 19, 20, and 21 were present in both parents but were absent in 7.5% (17/228), 2.2% (5/228), 4.8% (11/228), 6.1% (14/228), 2.2% (5/228), 1.8% (4/228), and 4.4% (10/228) of the progeny, respectively (fig. 2B). We found no progeny lacking a core chromosome in either of the crosses.

We also identified numerous instances of disomy in progeny accessory chromosomes. In cross 3D1x3D7, chromosomes 17, 19, and 20 were present in two copies in 1.6% (4/249), 0.8% (2/249), and 0.8% (2/249) of the progeny, respectively (fig. 2A). Interestingly, 2.4% (6/249) of the progeny were disomic for a core chromosome, with 1.6% (4/249) of the progeny disomic for chromosome 5 and 0.8% (2/249) disomic for chromosome 13. No disomic core chromosomes were identified in cross 1E4x1A5 (fig. 2B), but 1.3% (3/228) of the progeny were disomic for chromosome 14, 0.9% (2/ 228) were disomic for chromosome 18 and chromosomes 16, 19, 20, and 21 were each disomic in 0.4% (1/228) of the progeny.

Chromosomal inheritance that differed from the expected 1:1 ratio was observed for several chromosomes that were present in only one of the two parents of a cross. In the 3D1x3D7 cross, chromosomes 14, 15, 18, and 21 were absent in the 3D7 parent, hence we expected these chromosomes to be absent in half of the progeny. Instead, chromosomes 14, 15, 18, and 21 were absent in only 22.5% (56/249), 25.7% (64/249), 30.1% (75/249), and 26.9% (67/249) of the progeny, respectively (fig. 2A). The inheritance of these chromosomes are significant departures from the canonical Mendelian ratio (chromosome 14: χ^2=37.7, P < 0.001, chromosome 15: χ^2=29.4, P < 0.001, chromosome 18: χ^2=19.7, P < 0.001, and chromosome 21: χ^2=26.6, P < 0.001). We also tested whether chromosomes 14, 15, 18,

and 21 occurred independently from one another in progeny. We found that progeny lacking one or four chromosomes did not deviate significantly from expectations (χ^2 = 0.01, *P* = 0.9; χ^2 = 0.82, *P* =0.3, respectively). However, we found that progeny having all four chromosomes occurred much more frequently than expected (χ^2= 622.65, *P* < 0.001), while having one or two of the four chromosomes also occurred more frequently than expected (χ^2=14.7, *P* < 0.001; χ^2 = 37.76, *P* < 0.001, respectively). In the 1E4x1A5 cross, chromosome 17 was missing in 53.5% (112/228) of the progeny and did not exhibit distorted inheritance (χ^2=0.56, *P* 1⁄4 0.3) (fig. 2B). Disomy was also found for several accessory chromosomes that were present in only one of the parents. In cross 3D1x3D7, additional copies of chromosome 14 and 18 were identified in 0.4% (1/249) and 0.8% (2/249) of the progeny, respectively (fig. 2A). In cross 1E4x1A5, chromosome 17 was disomic in 0.9% (2/228) of the progeny (fig. 2B).

Disomic chromosomes can either be heterozygous, carrying one of each parental chromosomal copy, or homozygous if the disomy arose from a single parental chromosome (fig. 3). To distinguish these scenarios, we analyzed disomic progeny chromosomes and restricted the analyses to cases where both parents were carrying a chromosomal copy. In the 3D1x3D7 cross, 59% (10/17 cases) of the disomic isolates were heterozygous, with a chromosome originating from each parent and 29% (5/17 cases) of the disomic isolates were homozygous, with both chromosomes originating from one parent (fig. 4A). In the case of chromosomes 14 and 18, the chromosomes could only originate from one parent. In cross 1E4x1A5, 5 of the 11 disomic isolates were homozygous, three disomic isolates had chromosomes originating from both parents and the other three cases were ambiguous. As indicated earlier, chromosome 17 could only have originated from one of the parents.

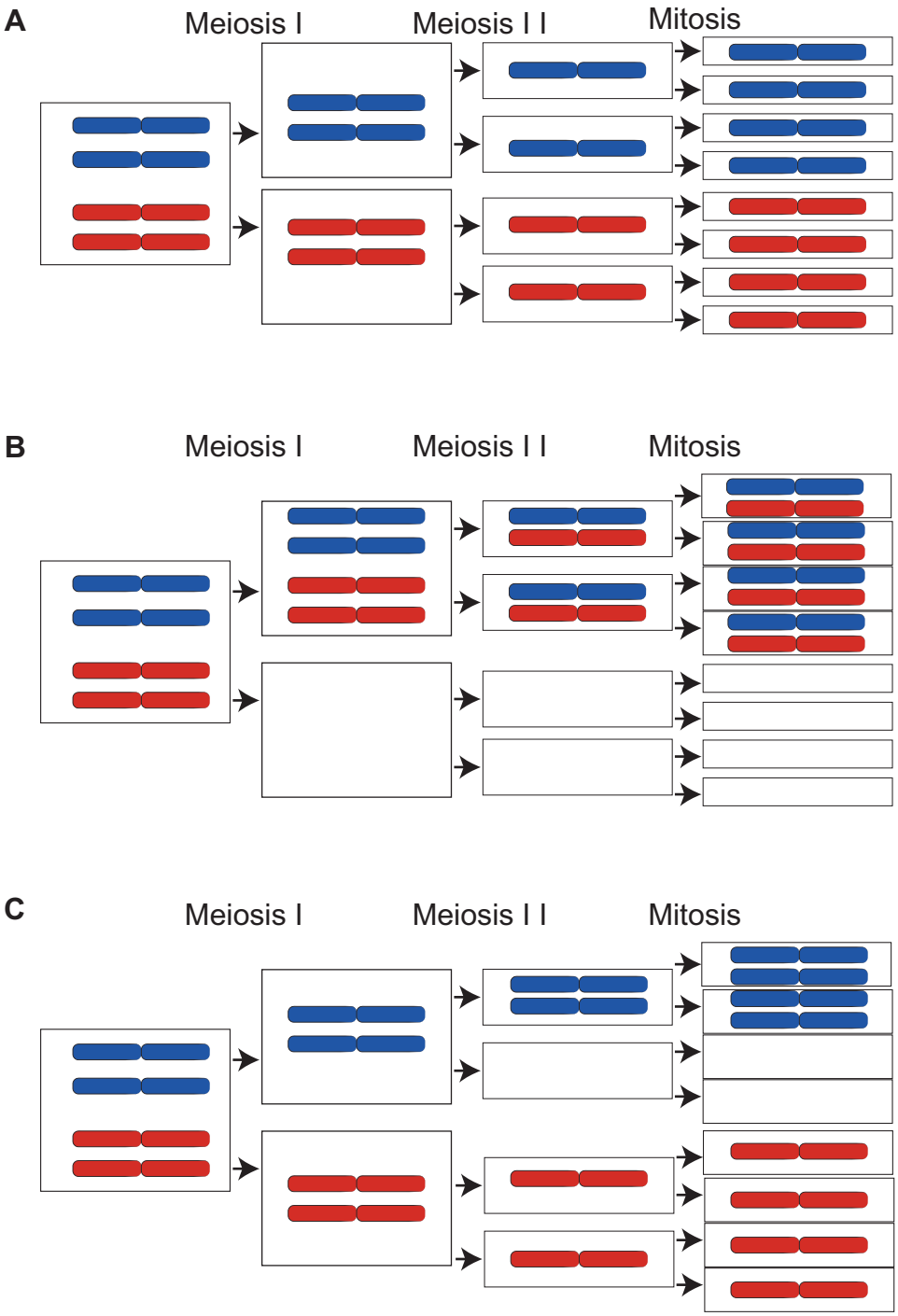


Figure 3: Schematic overview of how chromosomal nondisjunction can result in chromosome loss or disomy. (A) During canonical meiosis, the haploid nuclei from the two parents fuse resulting in a single diploid nucleus. Parental chromosomes are shown with distinct colors. Chromosomes go through meiosis I and II, followed by mitosis, resulting in eight haploid ascospores. Chromosome loss or disomy can occur as a result of homologous chromosomes failing to segregate during meiosis I (B), resulting in heterozygous disomy with one chromosome originating from each of the parents. The alternative is the failure of sister chromatid segregation during meiosis II (C), generating homozygous disomic progeny with both copies of the chromosome originating from the same parent.

Figure 4: Identification of heterozygous and homozygous disomic chromosomes in cross 3D1x3D7 and cross 1E4x1A5. Single nucleotide polymorphism (SNP) loci were screened on progeny chromosomes that showed evidence for disomy. SNPs were genotyped as either homozygous, containing only one of the parental alleles, or heterozygous if both parental alleles were found. The ratio represents the number of homozygous SNPs compared with the total number of genotyped SNPs. Individual dots represent each of the disomic progeny chromosomes identified in the two crosses. Due to uncertainties in SNP calling, we used cut-offs to assign progeny chromosomal states. Chromosomes with a ratio <0.4 were assigned as heterozygous disomic, likely resulting from nondisjunction at meiosis II, >0.6 as homozygous disomic, likely resulting from nondisjunction at meiosis I, and ratios between 0.4 and 0.6 were assigned as ambiguous.

Meiosis generates novel chromosome length polymorphism

In order to identify partially deleted or duplicated chromosomes in the progeny, we investigated chromosomes which had a normalized coverage between 0.3 and 0.7, and between 1.3 and 1.7 (fig. 1). In cross 3D1x3D7 (fig. 2A), partial deletions were identified for chromosomes 14 (0.4% of offspring, 1/249) and 15 (0.4%, 1/249). Partial duplications were detected for chromosomes 14 (0.4%, 1/249), 16 (0.8%, 2/249), 19 (2.4%, 2/249), and 21 (0.4%, 1/249). We also identified one isolate which may have a partially duplicated core chromosome 10. In cross 1E4x1A5 (fig. 2B), partial duplications were detected in the

progeny for chromosomes 14 (0.9%, 2/228), 15 (0.9%, 2/228), 16 (0.9%, 2/228), 17 (1.8%, 4/228), 19 (0.4%, 1/228), and 20 (1.8%, 4/228). Partial losses were identified for chromosomes 14 (0.4%, 1/ 228), 15 (1.3%, 3/228), 16 (3.5%, 8/228), 17 (2.6%, 6/228), and 21 (3.1%, 7/228).

We identified some progeny with multiple chromosomal anomalies, however these associations did not deviate significantly from a random expectation. In cross 3D1x3D7, isolate 89.1 was disomic for chromosome 13 and had a large, partial duplication of chromosome 10 while isolate 137.2 had partial duplications of chromosomes 16, 19, and 21. In cross 1E4x1A5, isolate B23.1 was disomic for chromosome 20 and had partial deletions of chromosomes 17 and 21. This isolate also had a partially duplicated chromosome 14. Isolate B24.2 also had partial deletions of chromosomes 17 and 21. Isolate C44.2 had partially deleted chromosomes 16 and 21. Isolate B50.1 was disomic for chromosome 17 and had a partially deleted chromosome 21. Isolate A57.1 was disomic for chromosome 14 and had a partially duplicated chromosome 16.

In cross 3D1x3D7, we found twelve progeny isolates with partial deletions and duplications. Seven of these partial aneuploidies affected chromosomal segments near the telomeric ends (supplementary fig. S4). Isolate 89.1 had a normalized coverage ratio for chromosome 10 of 1.63 suggesting a partial duplication. However, the coverage along the chromosome was homogeneous, with no apparent duplicated chromosomal regions when compared with the parent chromosomes (supplementary fig. S5). We considered such cases as ambiguous duplications. In cross 1A5x1E4, we found 40 partial deletions and duplications, of which 19 were ambiguous and 15 occurred in chromosomal segments near the telomeric ends (fig. 5).

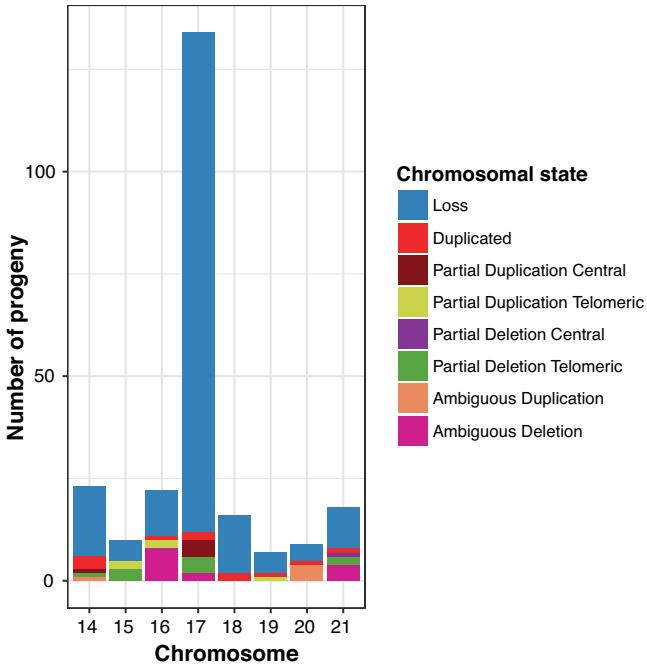


Figure 5: Identification of partial chromosome losses or duplications in cross 1E4x1A5. A summary of all the chromosome number and length polymorphisms in the progeny of cross 1E4x1A5, as well as the location where the length polymorphism occurred. Most of the rearrangements were ambiguous (19), 15 were located toward the ends of chromosomes and 6 rearrangements occurred in the central region of the chromosomes.

Correlation of chromosomal features with the fidelity of transmission

During meiosis, chromosomes pair prior to recombination and therefore length similarity could play a role in homolog identification and enable chromosomes to pair and recombine. However, we found no correlation between the length similarity of the parent chromosomes and the fidelity with which chromosomes were inherited (fig. 6A). In general accessory chromosomes were more unstable than core chromosomes. Interesting exceptions were a disomic core chromosome 13 (length difference 5% between the parents) and a disomic core chromosome 5 (length difference of 8.4% between the parents). The rate of disomy for these core chromosomes was 1.6% (4/249 progeny). We found no significant correlation between the recombination rate and chromosome transmission fidelity (fig. 6B). However, in cross 1A5x1E4, most of the chromosome losses and disomies occurred in accessory

chromosomes with a low recombination rate (fig. 6B). Next, we analyzed sequence similarities between parental chromosomes and correlated this with the chromosome transmission fidelity. For this, we compared whole chromosome sequences and calculated the percentage of syntenic regions between homologous chromosomes. The accessory chromosomes in the parents for both crosses had a much lower synteny than the core chromosomes and had substantially lower transmission fidelity (fig. 6C). Accessory chromosomes had overall a higher content of repetitive elements, which was similarly correlated with lower transmission fidelity (fig. 6D).

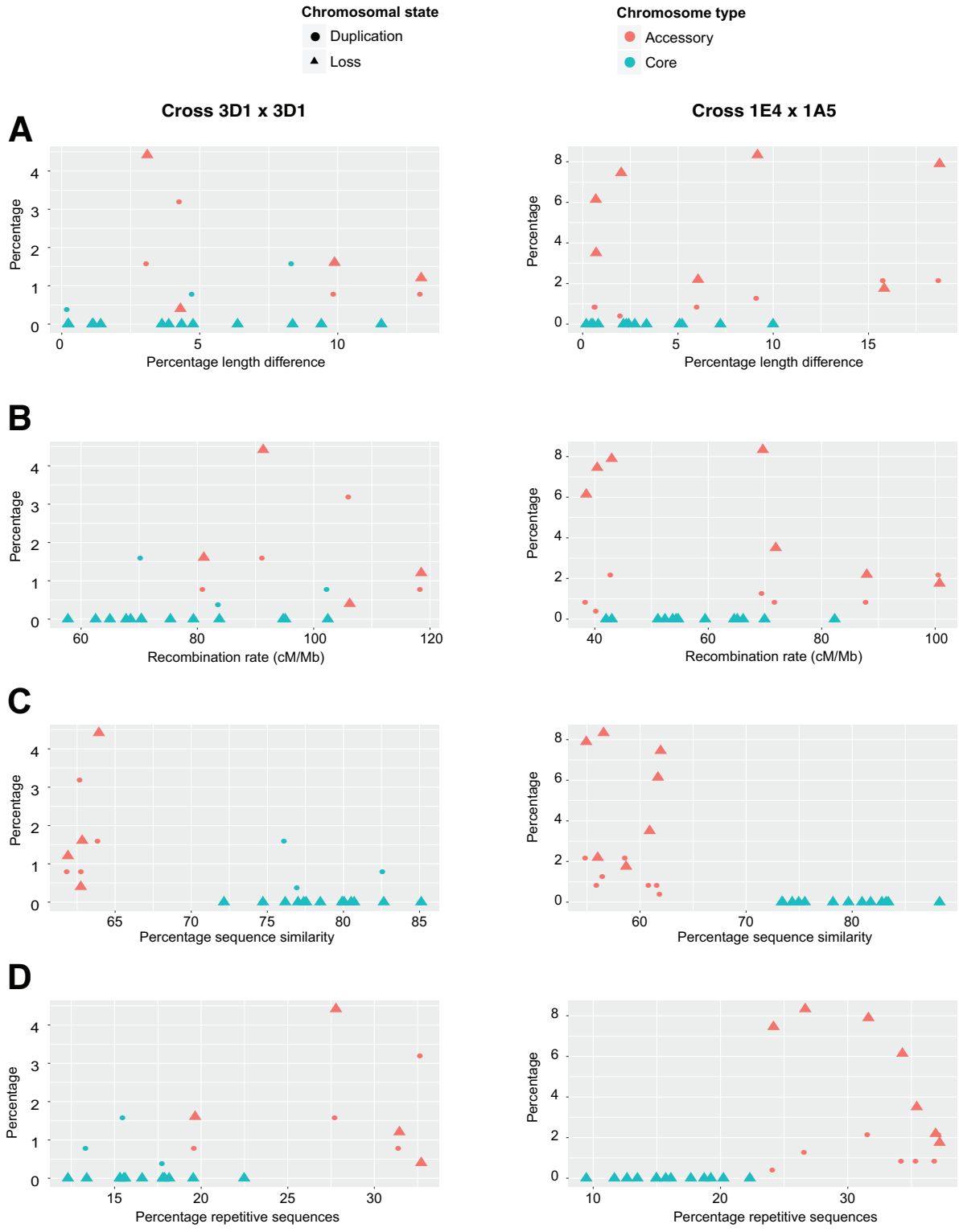


Figure 6: Correlations between chromosome length similarity, recombination rate, percent sequence similarity, fraction of repetitive sequences and the inheritance of chromosomes. Complete and partial chromosome losses and duplications were correlated with length similarity (A), recombination rate (B), sequence similarity (C), and repeat content (D) of the parental chromosomes. Correlations are shown separately for crosses 3D1x3D7 and 1E4x1A5.

Association between accessory chromosomes and phenotypic traits

We analyzed whether the chromosome states in progeny were correlated with variation in phenotypic traits. For this, we considered first only two chromosome states: normal (haploid) or abnormal (any loss, duplication, or rearrangements). We tested for an association with phenotypic traits using two-tailed *t*-tests (multiple testing significance threshold at $P < 0.002$). We first tested for associations with virulence on two wheat cultivars (Runal and Titlis) using data from a previous study (Stewart and McDonald 2014; Stewart et al. 2018). Progeny from cross 3D1x3D7 with a normal chromosome 17 had a higher pycnidia count on the cultivar Runal than isolates with an abnormal chromosome 17 ($P = 0.0019$; fig. 7A, supplementary fig. S6). Isolates missing chromosome 17 had a lower pycnidia count than isolates that were disomic for chromosome 17. On cultivar Titlis, progeny from cross 3D1x3D7 with a normal chromosome 18 had significantly darker pycnidia ($P=0.0018$; supplementary fig. S7). Progeny with an abnormal chromosome 19 had a marginally higher percent leaf area covered by lesions (PLACL; $P = 0.0024$; supplementary fig. S7). For progeny from cross 1E4x1A5, we found a correlation of the PLACL produced on Titlis with chromosome 21 ($P = 0.00002$; fig. 7B; supplementary fig. S8). Isolates with a partially deleted or lost chromosome 21 had a higher PLACL. For progeny of cross E4x1A5, we found that isolates with an abnormal chromosome 20 showed higher PLACL on Runal. We found no significant correlations for phenotypes related to growth, fungicide resistance, or temperature sensitivity.

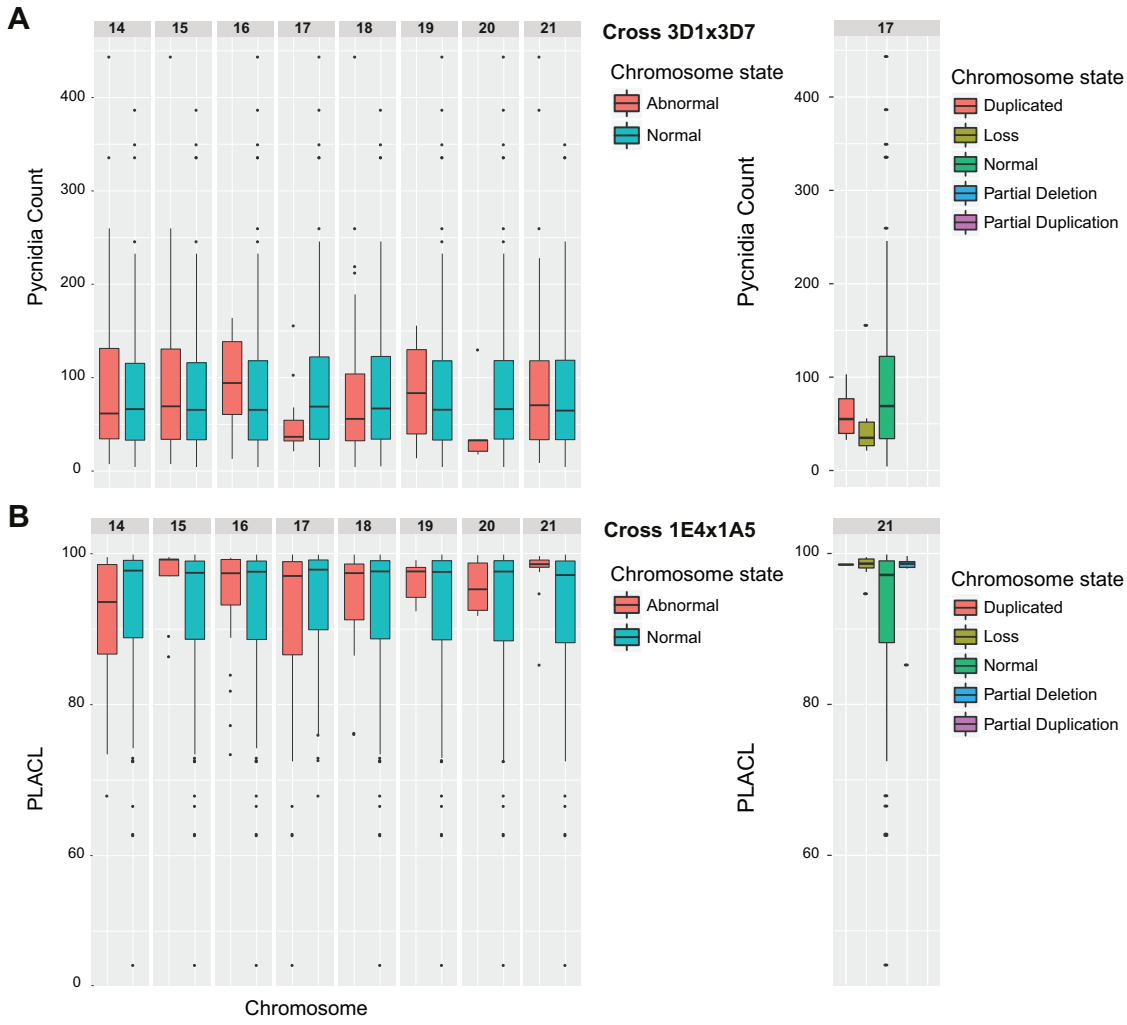


Figure 7: Association between accessory chromosomes and phenotypes. Accessory chromosome states, normal or abnormal (duplicated, lost, partially duplicated, or partially lost), were compared with virulence traits using a two-sample t-test (multiple testing correction threshold of $P < 0.002$). (A) In the progeny of 3D1x3D7, isolates with a normal chromosome 17 had a significantly higher pycnidia count on the wheat cultivar Runal that isolates with a duplicated or lost chromosome (P =0.0019). (B) In cross 1E4x1A5, isolates with a lost or partially deleted chromosome 21 had a higher percent leaf area covered by lesions (PLACL) on Titlis than isolates with a normal chromosome 21 (P =0.000024).

Correlation between disomy and chromosomal rearrangements

We analyzed whether rates of disomy were correlated with rates of rearrangements. Nondisjunction results in the loss of a chromosome in one progeny and a chromosome gain

in the corresponding twin spore from the same ascus. Core chromosomes generally showed only very rare cases of disomy or rearrangements (fig. 8). Accessory chromosome 14 was more frequently disomic and rearranged in progeny from cross 1A4x1E5. Chromosome 15 underwent partial duplications and deletions, but we found no evidence for nondisjunction. Chromosome 16 was both frequently rearranged (4.4%, 10/ 228) and disomic (0.4% 1/228) among the progeny in 1E4x1A5. In cross 3D1x3D7, chromosome 17 was disomic in 1.6% (4/249) of the progeny, while in cross 1E4x1A5 chromosome 17 was more rarely disomic (0.9%, 2/228). Chromosome 17 showed even stronger differences in rearrangements among crosses, with 4.4%, (10/228) in cross 1E4x1A5 versus 0.0% in cross 3D1x3D7. Chromosome 19 was both more likely to undergo rearrangements and to be inherited as a disomic chromosome in cross 3D1x3D7. In contrast, chromosome 21 was both more likely to be rear- ranged and to be inherited in a disomic state in cross 1E4x1A5.

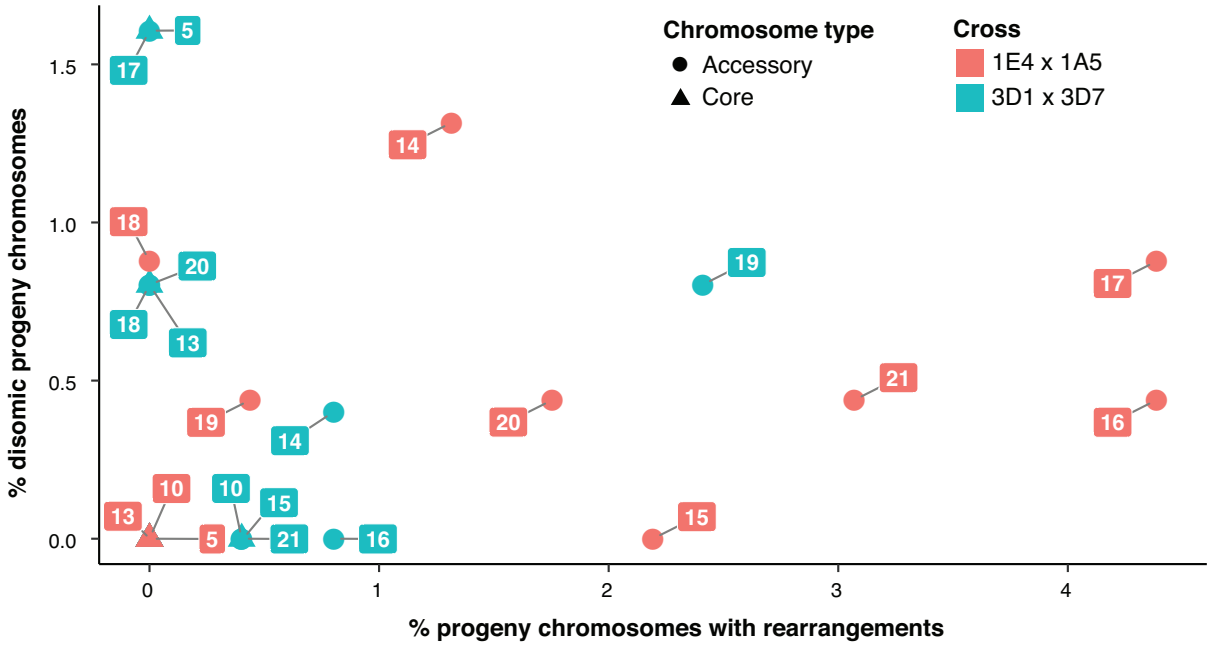


Figure 8: Correlation between number of disomic progeny and chromosomal rearrangements. Circles and triangles represent accessory chromosomes and core chromosomes, respectively. Chromosomes from cross 3D1x3D7 are represented in blue, and chromosomes from cross 1E4x1A5 are in red.

Discussion

We used RADseq data generated for several hundred progeny from two crosses of *Z. tritici* to identify aberrations in chromosomal transmission through meiosis. We found extensive chromosome number and length variation among the progeny in both crosses. The rates of disomy and rearrangements differed greatly between chromosomes and crosses. Nearly all aberrant chromosomal transmission events affected accessory chromosomes with the rare exception of core chromosome disomies. Several accessory chromosomes showed strongly distorted chromosomal inheritance.

Chromosome number polymorphism in *Z. tritici* has previously been linked to errors occurring during meiosis (Wittenberg et al. 2009; Croll et al. 2013). In our study, we generated a substantially more dense marker coverage using the Illumina-based sequencing technique RADseq and were able to screen more isolates (477 isolates compared with 144 and 216 isolates, respectively; Wittenberg et al. 2009; Croll et al. 2013). Because RADseq generated a high coverage of ~100-bp sequences at defined restriction sites, we could precisely map sequences to chromosomal positions without having to rely on genetic map constructions. Physical marker positions are particularly important for analyzing accessory chromosomes of *Z. tritici* because of their very low rates of recombination (Croll et al. 2015). In contrast to previous studies, our use of RADseq markers allowed us to directly detect duplicated chromosomal segments by analyzing variations in sequencing coverage.

Our analyses revealed that all eight accessory chromosomes underwent chromosome loss during meiosis. The rate of chromosomal loss depended on the chromosome and varied between the crosses. This confirms the findings of Croll et al. (2013) except that a loss of

chromosome 15 had not previously been detected. We found that 5 progeny (2.1%) had lost this chromosome. No isolate was found lacking a core chromosome despite screening 477 progeny. This indicates that all 13 core chromosomes are likely encoding essential functions for the growth and survival of the fungus.

Chromosome loss most likely occurred as a result of errors during chromosome segregation, specifically nondisjunction of sister chromatids during either meiosis I or II. In accordance with previous studies, we found that the loss of accessory chromosomes during meiosis is common. In natural populations, this may lead to the complete loss of an accessory chromosome in the absence of counteracting mechanisms that maintain these chromosomes.

Wittenberg et al. (2009) proposed that distorted segregation of accessory chromosomes could serve as a mechanism to prevent their complete loss from a population. Chromosomes present in only one parent are expected to segregate into 50% of the daughter cells. However, we found that in cross 3D1x3D7 chromosomes 14, 15, 18, and 21 from parent 3D1 were significantly overrepresented in the progeny. The transmission advantage resulting from unequal segregation is referred to as “meiotic drive” and is frequently associated with accessory or B chromosomes (Jones 1991). In our study, distorted inheritance was not universal, for example chromosome 17 in cross 1E4x1A5 segregated normally. The distorted inheritance pattern in cross 3D1x3D7 could be explained if parent 3D1 already had disomic accessory chromosomes. But our coverage analysis did not detect disomic chromosomes in any of the parents. It is possible that a small fraction of the clonal cell pool of a parental mycelium might have harbored disomic chromosomes, but this is not likely to explain the observed rates of disomic accessory chromosomes. The

overrepresentation of progeny carrying a specific accessory chromosome could be due to selection favoring progeny carrying this chromosome. Such viability selection could not be tested in this experiment because we were un- able to generate full tetrad sets of offspring and quantify genotype-specific survival rates. However, if loci located on accessory chromosomes encoded strongly deleterious variants for growth on culture media, quantitative trait mapping studies performed on the same progeny sets would most likely have identified QTLs linked to accessory chromosomes. However, no such evidence was found (Lendenmann et al. 2014, 2016).

Additional explanations for the observed distortion in inheritance may include a meiotic drive mechanism such as selective spore killing. The distortion could also be linked to “sticky” centromeres similar to those found in rye B chromosomes where the transmission at higher than Mendelian frequencies was explained by the presence of particular centromeres that ensure that B chromosomes migrate to the generative pole that will be transmitted to the next generation of plants (Banaei-Moghaddam et al. 2012). In order to distinguish among the possible mechanisms leading to distorted inheritance, all meiotic products from individual tetrads would have to be analyzed. However, experimental limitations in the generation of large numbers of individual tetrads prevented us from making more detailed investigations.

We found that an average of 5.9% of the progeny isolates were disomic for one or more chromosomes. This number is similar to what was found for *Saccharomyces cerevisiae*, where 8% of the lab strains were estimated to be aneuploid (Hughes et al. 2000). Disomy is generated when chromosomes undergo nondisjunction during meiosis, resulting in one daughter cell with two copies of a chromosome and one daughter cell with no copies of that chromosome (fig. 3). Therefore, for each disomic offspring, we expect a corresponding offspring that is missing the same chromo- some. As expected, we found that chromosomal

loss was often accompanied by disomy. However, contrary to expectations, there was no symmetry in the loss and disomy rates. For example, despite finding many progeny lacking chromo- some 15, no isolate disomic for chromosome 15 was recovered. The rates of nondisjunction also differed between chromosomes and between crosses, suggesting that the loss or disomy of specific chromosomes may be counter selected. In addition, chromosomes differed in their composition of repetitive elements. Repetitive elements are likely to play an important role by influencing the likelihood of faithful disjunction. We also found that nondisjunction was happening during both meiosis I and II. We found heterozygous disomic chromosomes, which were created as a result of nondisjunction in meiosis I. Heterozygous disomic chromosomes were most frequent in cross 3D1x3D7. In cross 1E4x1A5, homozygous disomy resulting from nondisjunction in meiosis II occurred more frequently. Aneuploidy can play an important role in the adaptive evolution of fungal pathogens. In human pathogens, aneuploidy is often associated with drug resistance (Hu et al. 2008; Selmecki et al. 2010). Over 50% of the fluconazole-resistant strains isolated from patients had whole or partial chromosome duplications (Selmecki et al. 2006). Correlations between disomic states and phenotypic traits in *Z. tritici* suggests that selection could also be affecting rates of disomy, albeit with less drastic impacts than in human pathogens selected for drug resistance.

Aneuploidy typically causes a dosage imbalance, which could explain why accessory chromosome aneuploidies are tolerated more frequently than core chromosome aberrations. Alternatively, gene expression or dosage compensation could have evolved on frequently disomic chromosomes, which may explain the tolerance for additional copies of certain chromosomes, but not others (Torres et al. 2008). Chromosomes that have a higher rate of disomy could have shorter or nonfunctional telomeres. Telomere defects were found to explain mitotic instability in human mammary epithelial cells (Pampalona et al. 2010).

Chromosomes with shorter telomeres are more likely to undergo nondisjunction. Furthermore, chromosomes with higher degrees of synteny are more likely to pair correctly, resulting in fewer nondisjunction events. We found indications that sequence similarity in the parent chromosomes indeed leads to higher fidelity of chromosomal inheritance.

Homologous chromosomes of *Z. tritici* segregate significant structural variation in populations, differing in repeat and gene content, chromosomal length, and recombination rate, as well as telomere and centromere composition (Croll et al. 2013, 2015; Schotanus et al. 2015; Plissonneau et al. 2016). Synteny breakpoints are commonly associated with repetitive sequences or transposable element clusters that can misalign during recombination, thereby generating length polymorphism. Such a mechanism was thought to generate a novel chromosome 17 in the progeny of cross 1A5x1E4 (Croll et al. 2013). In our study, we found no correlation between length similarity and recombination rate of the parent chromosomes, and the fidelity of chromosome inheritance. However, chromosomes with higher synteny between the parents and fewer repeats were transmitted more faithfully.

Selection favoring the presence or absence of specific accessory chromosomes would require that accessory chromosomes directly or indirectly influence phenotypic traits. However, accessory chromosomes carry few genes and none are thought to perform a specific function during the life cycle of the fungus (Goodwin et al. 2011). Interestingly, we found a correlation between the presence of chromosomes 15, 18, and 21 and higher levels of virulence in cross 3D1x3D7 (Stewart et al. 2018). In addition, we found a correlation between the presence of a normal chromosome 17 and an abnormal chromosome 19 and higher levels of pycnidia and PLACL, respectively. In a separate study, whole-chromosome deletion mutants of a different *Z. tritici* strain were generated by blocking b-tubulin assembly

during mitosis using carbendazim (Habig et al. 2017). A comparison of isogenic lines lacking individual accessory chromosomes showed that the loss of chromosomes 14, 16, 18, 19, and 21 resulted in increased virulence on the wheat cultivar Runal. This finding is in opposition to our own study that showed that the presence of some of the same accessory chromosomes increased virulence on Runal. The effect sizes on virulence were similar albeit of the opposite sign. Habig et al. (2017) found no effect of chromosome loss on cultivars Obelisk and Titlis. In contrast, our study showed that abnormal chromosome 21 was associated with higher levels of PLACL on Titlis. It should be noted though that chromosome 21 (as all other accessory chromosomes) shows substantial sequence variation among homo- logs within the species (Croll et al. 2013). In conjunction, the two studies suggest that the identity of accessory chromosomes, the genetic background and the host genotype interact to affect the phenotypic consequences of accessory chromosomes. Even though the individual effect sizes were relatively small, the observed differences in virulence traits may be significant under natural conditions. The production of lesions (expressed as PLACL) and pycnidia counts can increase the survival and reproductive potential of the pathogen in the field. If isolates harboring specific accessory chromosomes gain a fitness advantage in at least some strain-by-host genotype combination, then accessory chromosomes may be maintained in the species pool by a selection-drift balance.

Most chromosome rearrangements are thought to be deleterious and therefore counter-selected. The ability of *Z. tritici* to tolerate a large number of disomies and chromosomal rearrangements makes this species an excellent model for detailed analyses of rearrangements and nondisjunction events. Despite the fact that the meiotic machinery is highly conserved, the strength of selection against erroneous chromosomal transmission can differ widely among species. Relaxed selection on chromosomal transmission can lead to highly polymorphic chromosomal sets observed in some eukaryotic pathogens.

Determining the trade-offs involved in maintaining chromosomal integrity and generating chromosomal polymorphism will elucidate how selection operates to maintain the fidelity of meiotic processes.

Acknowledgments

We thank Marcello Zala for providing access to progeny collections and helpful discussions. C.P. was supported by an INRA Young Scientist grant. S.F. and B.A.M. are supported by the Swiss National Science Foundation (grant 31003A_155955). D.C. is supported by the Swiss National Science Foundation (grant 31003A_173265).

References:

Baird NA et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLOS ONE. 3(10):e3376.

Baker BS, Carpenter AT, Esposito MS, Esposito RE, Sandler L. 1976. The genetic control of meiosis. Annu. Rev. Genet. 10(1):53–134.

Banaei-Moghaddam AM et al. 2012. Nondisjunction in favor of a chromosome: the mechanism of rye B chromosome drive during pollen mitosis. Plant Cell 24(10)4124–4134.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2114-2120.

Chen G, Rubinstein B, Li R. 2012. Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. BioEssays. 34(10):893–900.

Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics. 201(3):1213–1228.

Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. PLOS Pathog. 8(4):e1002608.

Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLOS Genet. 9(6):e1003567.

Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opini Genetics Dev. 35:57-65

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. Clifton NJ. 772:157–178.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio. [cited 2016 August 31]. Avalable from: http://arxiv.org/abs/1207.3907.

Goodwin SB et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLOS Genet. 7(6):e1002070.

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

Habig M, Quade J, Stukenbrock EH. 2017. Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen *Zymoseptoria tritici*. mBio 8(6):e01919-17.

Harewood L, Fraser P. 2014. The impact of chromosomal rearrangements on regulation of gene expression. Hum. Mol. Genet. 23(R1):R76–R82.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11(5):1189–1204.

Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. Nat. Rev. Genet. 2(4):280–291.

Hu G et al. 2008. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. Genome Biol. 9(2):R41.

Hughes TR et al. 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet. 25(3):333–337.

Jones RN. 1991. B-chromosome drive. Am. Nat. 137(3):430–442.

Kurtz S et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12..

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9(4):357–359.

Lendenmann MH, Croll D, McDonald BA. 2015. QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. Fungal Genet. Biol. 80:53–67.

Lendenmann MH, Croll D, Palma-Guerrero J, Stewart EL, McDonald BA. 2016. QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. Heredity. 116(4):384–394.

Lendenmann MH, Croll D, Stewart EL, McDonald BA. 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. G3 (Bethesda) 4:2519–2533.

Mather K. 1938. Crossing-over. Biol. Rev. 13(3):252–292.

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

Montgomery EA, Huang SM, Langley CH, Judd BH. 1991. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. Genetics. 129(4):1085–1098.

Naranjo T. 2012. Finding the correct partner: The meiotic courtship. Scientifica. 2012:1.

Ngamskulrungroj P, Chang Y, Sionov E, Kwon-Chung KJ. 2012. The primary target organ of *Cryptococcus gattii* is different from that of *Cryptococcus neoformans* in a murine model. mBio. 3(3):e00103-12.

Pampalona J, Soler D, Genescà A, Tusell L. 2010. Whole chromosome loss is promoted by telomere dysfunction in primary cells. Genes. Chromosomes Cancer. 49:NA–378.

Pavelka N, Rancati G, Zhu J, et al. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 468(7321) :321–325.

Pavelka N, Rancati G, Li R. 2010. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. Curr. Opin. Cell Biol. 22(6):809–815.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16(1):5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio. 7(5):e01231-16.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842.

Roeder GS. 1997. Meiotic chromosomes: it takes two to tango. Genes Dev. 11(20):2600–2621.

Schotanus K et al. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 8(1):41.

Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science. 313(5785):367–370.

Selmecki A, Forche A, Berman J. 2010. Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot. Cell. 9(7):991–1008..

Selmecki A, Gerami-Nejad M, Paulson C, Forche A, Berman J. 2008. An isochromosome confers drug resistance *in vivo* by amplification of two genes, *ERG11* and *TAC1*. Mol. Microbiol. 68:(3)624–641.

Sionov E, Lee H, Chang YC, Kwon-Chung KJ. 2010. *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. PLOS Pathog. 6:e1000848. doi: 10.1371/journal.ppat.1000848.

Stewart E l. et al. 2017. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Mol. Plant Pathol. 19(1):201

Stewart EL, McDonald BA. 2014. Measuring Quantitative Virulence in the Wheat Pathogen Zymoseptoria tritici Using High-Throughput Automated Image Analysis. Phytopathology. 104(9):985–992.

Torres EM, Williams BR, Amon A. 2008. Aneuploidy: Cells losing their balance. Genetics. 179(2):737–746.

Torriani SFF, Stukenbrock EH, Brunner PC, McDonald BA, Croll D. 2011. Evidence for extensive recent intron transposition in closely related fungi. Curr. Biol. 21(23):2017–2022.

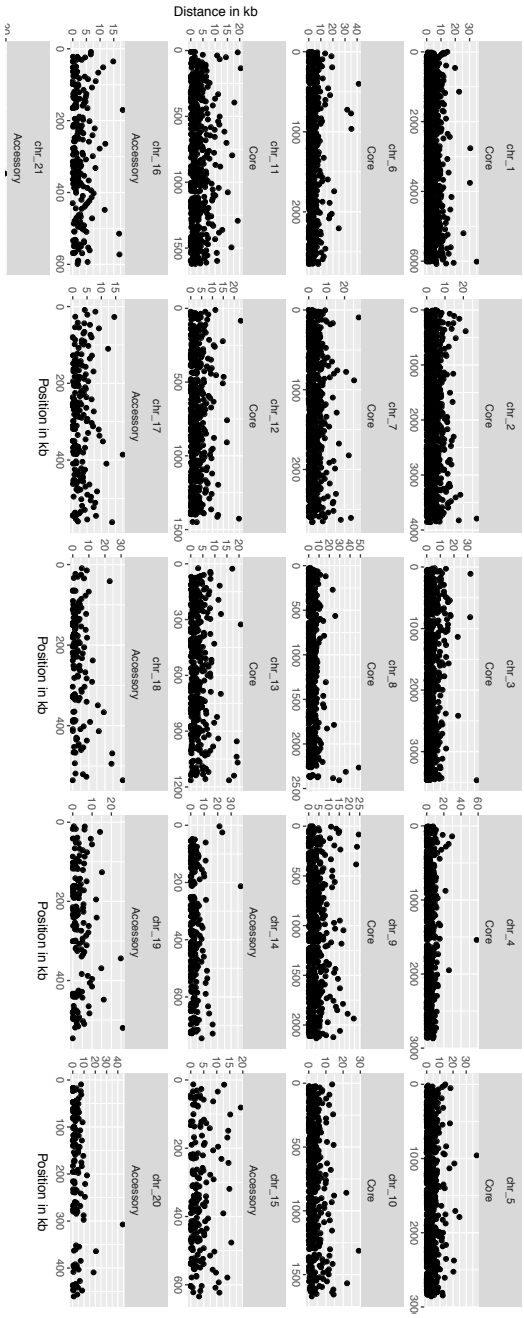
Wittenberg AHJ et al. 2009. Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. PLOS ONE. 4(6):e5863.

Zhan J et al. 2005. Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus *Mycosphaerella graminicola*. Mol. Ecol. 14(9):2683–2693.

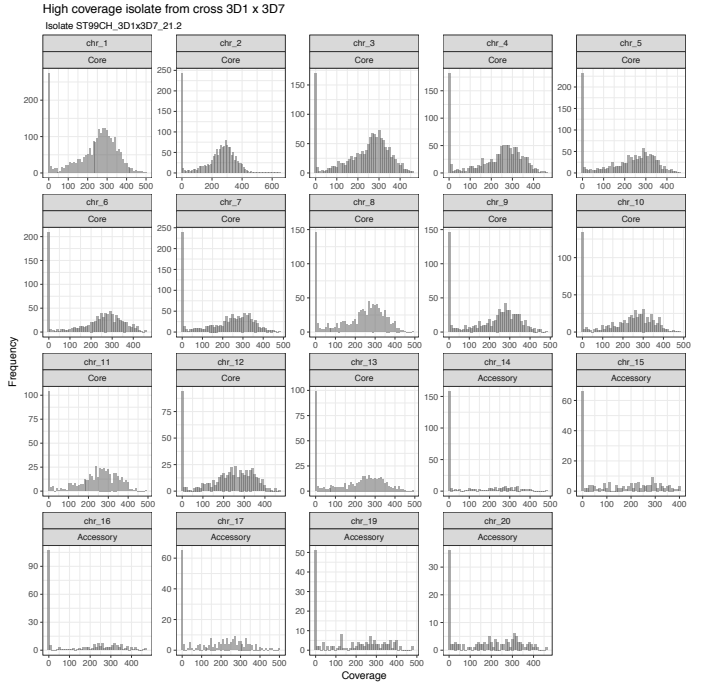
Zhang J, Yu C, Krishnaswamy L, Peterson T. 2011. Transposable elements as catalysts for chromosome rearrangements. Methods Mol. Biol. Clifton NJ. 701:315–326.

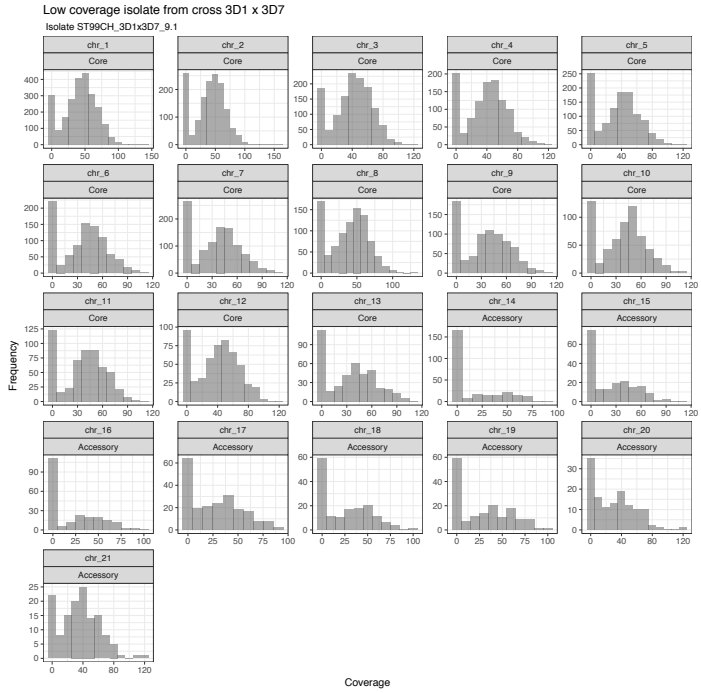
Supplementary Information

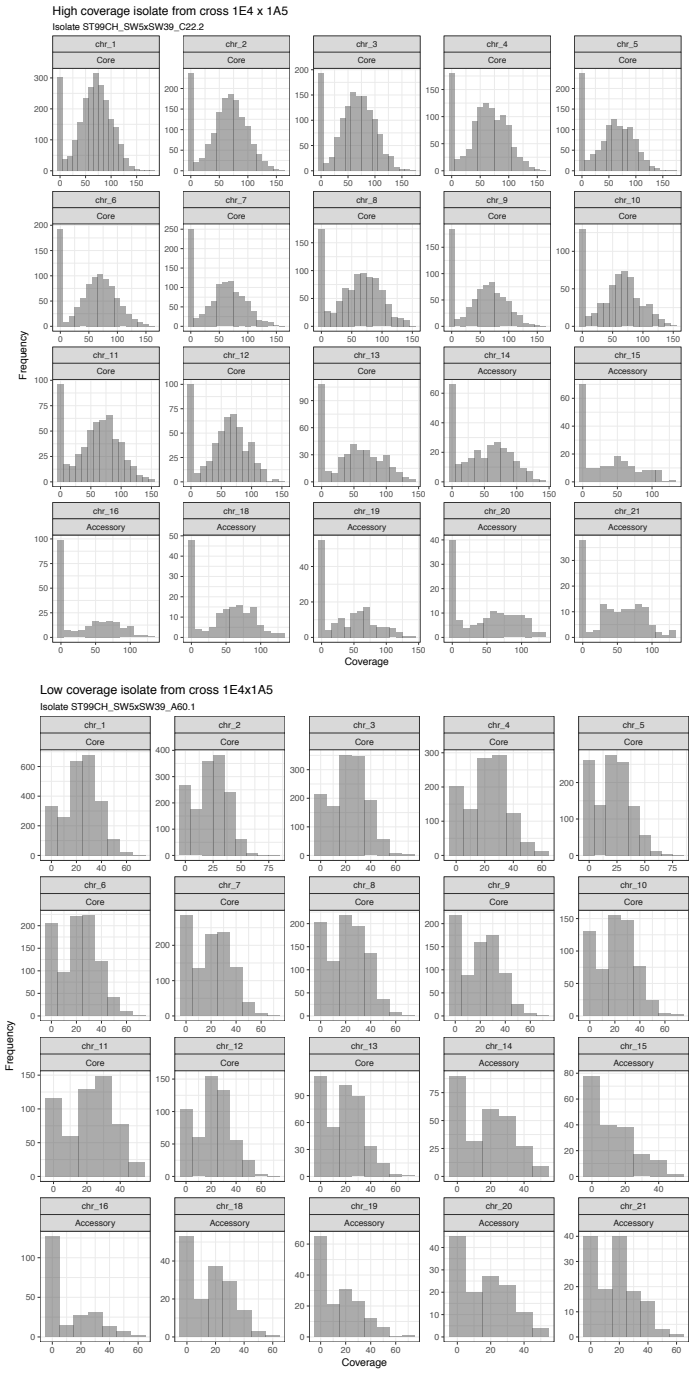
Supplementary Figures



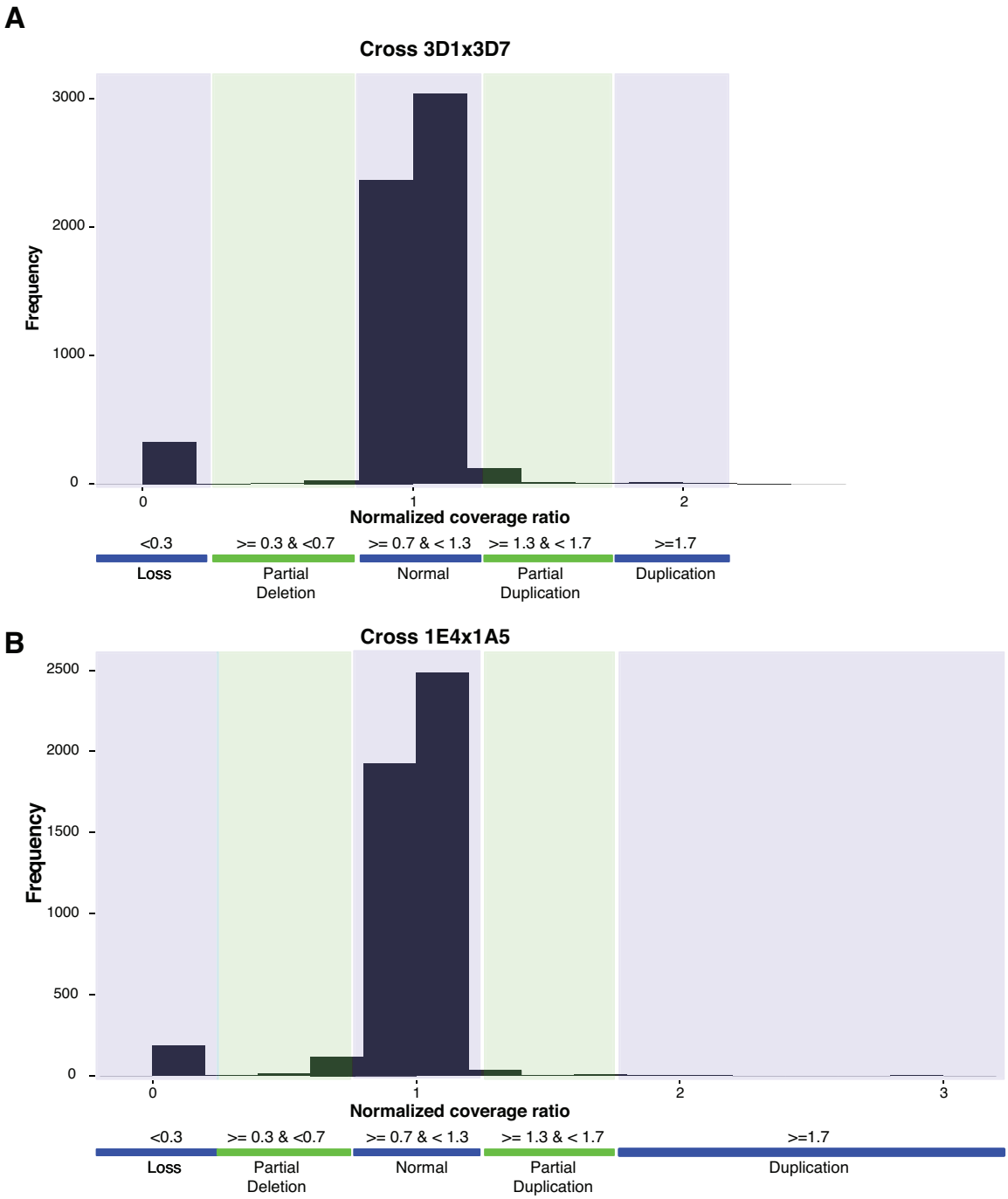
Supplementary Figure S1: The spacing of RADseq markers on chromosomes. The distance between adjacent *Pst*I cut sites are shown for each chromosome separately.


High coverage isolate from cross 3D1 x 3D7
Isolate ST99CH_3D1x3D7_21.2
chr_1
chr_2
chr_3
chr_4
chr_5
chr_6
chr_7
chr_8
chr_9
chr_10
chr_11
chr_12
chr_13
chr_14
chr_15
chr_16
chr_17
chr_19
chr_20
Core
Accessory
Frequency
Coverage

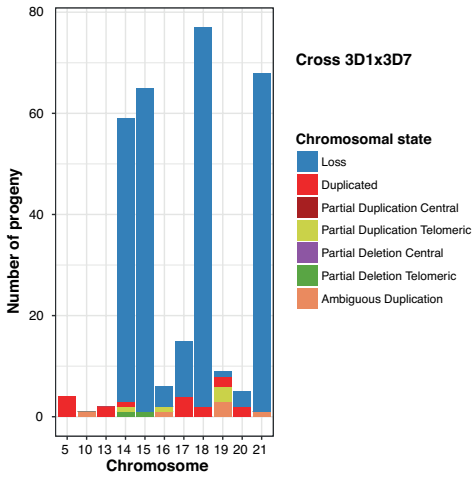

Low coverage isolate from cross 3D1 x 3D7
Isolate ST99CH_3D1x3D7_9.1
chr_1
chr_2
chr_3
chr_4
chr_5
chr_6
chr_7
chr_8
chr_9
chr_10
chr_11
chr_12
chr_13
chr_14
chr_15
chr_16
chr_17
chr_18
chr_19
chr_20
chr_21
Core
Accessory
Frequency
Coverage



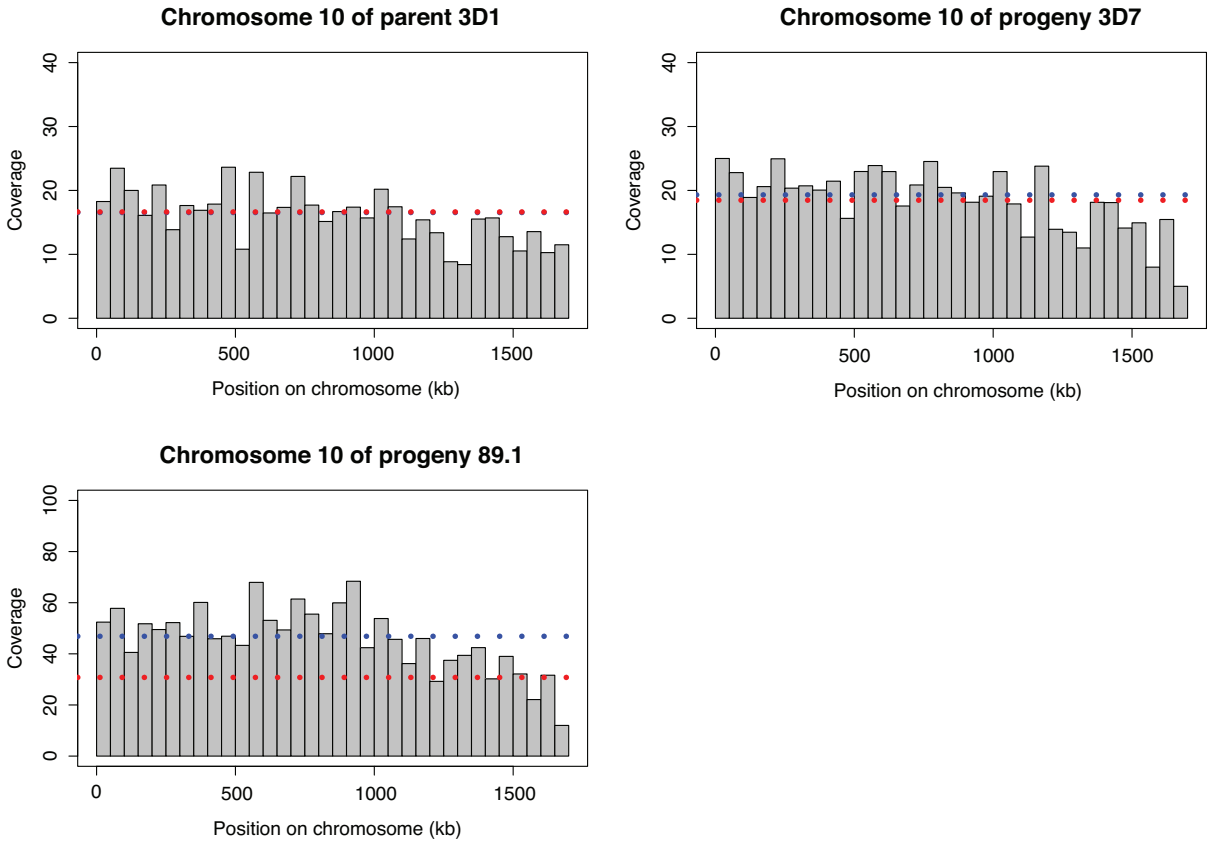
Supplementary Figure S2: Sequencing read coverage at RADseq loci across chromosomes. Two progeny each from the low end and high end of the coverage spectrum, respectively, are shown. Progeny ST99CH_3D1x3D7_21.2 (A) and ST99CH_3D1x3D7_9.1 (B) were from cross 3D1x3D7 and provide examples for high and low coverage samples, respectively. Progeny ST99CH_SW5xSW39_C22.2 (C) and ST99CH_SW5xSW39_A60.1 (D) were from the cross 1E4x1A5 and provide examples for high and low coverage samples, respectively.



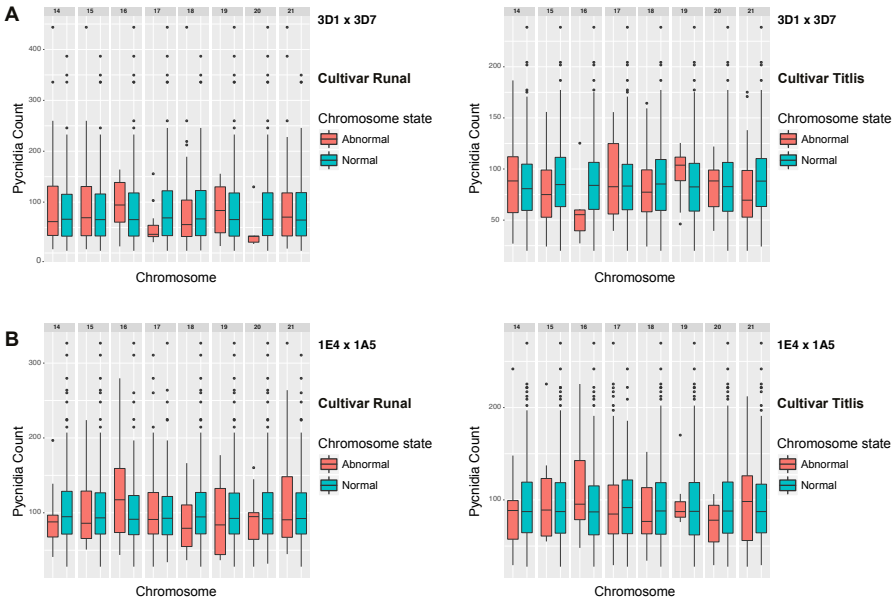
Supplementary Figure S3: The normalized coverage distribution of progeny from cross 3D1x3D7 (A) and 1E4x1A5 (B). Both normalized coverage distributions indicate the cutoffs used to detect a whole chromosome loss (ratio < 0.3), partial deletion (ratio 0.3-0.7), regular transmission (0.7-1.3), partial duplication (1.3-1.7) and whole chromosome duplication (>1.7). The normalized coverage was calculated by determining the coverage for each chromosome and comparing this to the median coverage of all chromosomes for that isolate.



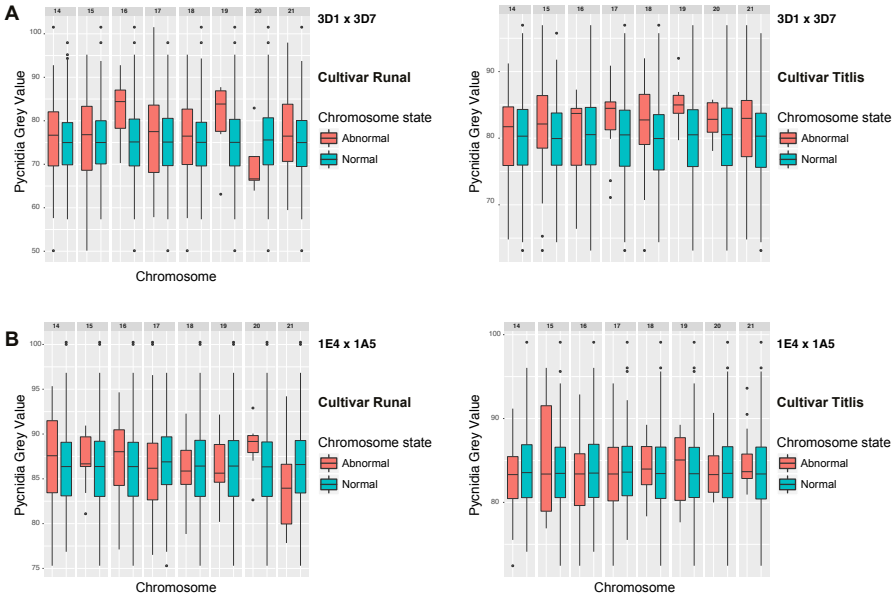
Supplementary Figure S4: Identification of partial chromosome losses or duplications in cross 3D1x3D7. A summary of all the chromosome number and length polymorphisms in the progeny of cross 3D1x3D7.



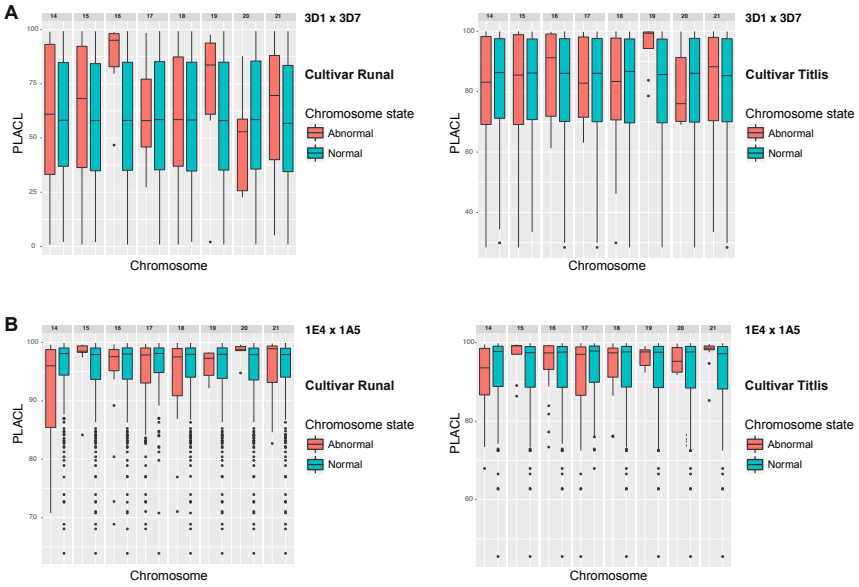
Supplementary Figure S5: The coverage distribution on chromosome 10. Isolates 3D1 and 3D7 are parental isolates. Progeny 89.1 displays a putative partial duplication of chromosome 10. Red and blue dotted lines show the genome-wide and chromosome-wide median coverage, respectively.



Supplementary Figure S6: Association between accessory chromosomes and pycnidia counts. Accessory chromosome states, normal or abnormal (duplicated, lost, partially duplicated or partially lost), were compared to pycnidia counts using a two-sample *t*-test (multiple testing correction threshold of $p < 0.002$). (A) Progeny of 3D1x3D7 isolates with a normal chromosome 17 had a significantly higher pycnidia count on the wheat cultivar Runal that isolates with an abnormal chromosome ($p = 0.0019$). This was not the case on Titlis. (B) In cross 1E4x1A5, no significant correlations between chromosome state (normal or abnormal) and pycnidia count were found on either wheat cultivar.



Supplementary Figure S7: Association between accessory chromosomes and pycnidia grey value. Accessory chromosome states, normal or abnormal (duplicated, lost, partially duplicated or partially lost), were compared to pycnidia grey value using a two-sample t-test (multiple testing correction threshold of $p < 0.002$). (A) Progeny of 3D1x3D7 isolates with a normal chromosome 18 had significantly lighter pycnidia on the wheat cultivar Titlis than isolates with an abnormal chromosome ($p = 0.0018$). This was not the case on Runal. (B) In cross 1E4x1A5, no significant correlations between chromosome state (normal or abnormal) and pycnidia grey value were found on either wheat cultivar.



Supplementary Figure S8: Association between accessory chromosomes and percentage leaf area covered by lesions (PLACL). Accessory chromosome states, normal or abnormal (duplicated, lost, partially duplicated or partially lost), were compared to PLACL using a two-sample *t*-test (multiple testing correction threshold of $p < 0.002$). (A) Progeny of 3D1x3D7 isolates with an abnormal chromosome 19 had a marginally higher PLACL on the wheat cultivar Titlis (PLACL; $p = 0.0024$). This was not the case on Runal. (B) In cross 1E4x1A5, isolates with an abnormal chromosome 21 had a higher PLACL on Titlis ($p = 0.00002$). On Runal isolates with an abnormal chromosome 20 had a higher PLACL ($p < 0.002$).

Supplementary Tables:

Supplementary Table S1: Phenotypic trait data available for progeny. PLACL: Percent leaf area covered by lesions. MC: Mean centered.

Available online: https://doi.org/10.1093/gbe/evy100

CHAPTER 3:

Repeated chromosomal degeneration through repeat-mediated chromosomal rearrangements in a fungal pathogen

Simone Fouché, Marcello Zala, Ursula Oggenfuss, Bruce A. McDonald, Daniel Croll

Manuscript in preparation.

Abstract

Genomic instability is often triggered by chromosomal rearrangements and can have severe consequences for fitness. Breakage-fusion-bridge (BFB) cycles are a broadly observed mechanism of genome instability and found in human cancer lines, plants and fungi. However, the exact mechanisms that initiate and maintain BFB are poorly understood. Fungal plant pathogens have highly dynamic genomes showing extensive chromosomal polymorphism within the same species making fungi excellent models to study the causes and consequences of chromosomal instability. In this study, we recapitulate the progressive degeneration of a chromosome of *Zymoseptoria tritici* across four rounds of meiosis and twelve progeny genomes. Using long-read sequencing and pulsed-field electrophoresis, we show that the unstable chromosome was first generated by ectopic recombination on repeats of a recently expanded transposable element. Subsequent rounds of meiosis perpetuated the cycle of chromosomal degeneration by additional ectopic recombination. Amplification of chromosomal sequences and non-disjunction led to progeny carrying up to five copies after four rounds of meiosis. Using PCR genotyping of over 380 progeny, we identified that the level of chromosomal polymorphism segregating among progeny was predictable from parental genotypes. Our work identifies the exact sequence triggers initiating and perpetuating degenerative BFB cycles. Dissecting proximate causes leading to run-away chromosomal degeneration will improve our understanding of chromosomal evolution in cancer lines and beyond.

Introduction

Genomic instability is defined by an increase in the rate at which cells acquire genomic alterations and is a hallmark feature of cancers. Genomic alterations include mutations, the loss or gain of chromosome segments or complete chromosomes. A major class of genomic instability is defined by breakage-fusion bridge cycles (BFB). BFB cycles involve a series of chromosome breaks, duplications and deletions that can alter the copy number of chromosome segments. BFB was first discovered in maize by McClintock in dicentric chromosomes going through cycles of degeneration (McClintock 1938; McClintock 1941). The centromeres of the dicentric chromosome are pulled in opposite directions during anaphase. The bridge breaks and generates daughter cells with different lengths of the chromosome that lack telomeres and are susceptible to fusion. Fusion of sister chromatids result in a duplicated chromatid with two centromeres that again forms a bridge that can rupture and drive further degenerative cycles. The chromosome breakage can occur at any position between the two centromeres and thus chromatids inherit unequal amounts of genetic material where one inherits an inverted duplicated region and the other has a deletion of this region (Smith et al. 1990; Ma et al. 1993; Toledo et al. 1993; Coquelle et al. 1997; Marotta et al. 2013). Through subsequent degenerative cycles, a region between the two centromeres can get amplified. BFB can have wide-ranging phenotypic consequences due the large-scale duplication or loss of genes (Lo et al. 2002; Maser and DePinho 2002; Tanaka et al. 2002; Narayanan et al. 2006; Tanaka and Yao 2009). However, chromosome-level reconfigurations occurring during BFB cycles remain poorly understood.

After the initial discovery in maize, BFB cycles have also been identified in animals (Toledo et al. 1993; Bi et al. 2004) and fungi (Rank et al. 1988; Hackett et al. 2001). Chromosome structure can be changed by the mis-repair of double stranded breaks (Pastink et al. 2001;

Schubert 2007) or through the non-homologous recombination between repeats (Bzymek and Lovett 2001; Mieczkowski et al. 2006; Argueso et al. 2008; Raskina et al. 2008). Fusions are particularly likely when chromosomes are missing telomeres. In addition, cancerous cell lines often show the amplification of very specific genes (*i.e.* oncogenes) adding palindromic sequences to the amplified region, especially at the breakpoint between duplicated regions and normal-copy regions (Ciullo et al. 2002; Hellman et al. 2002; Marotta et al. 2012; Marotta et al. 2013; Marotta et al. 2017; Bianchi et al. 2019). This suggests that certain genomic features (e.g. the presence of repeats or a propensity to break) make some regions more likely to be involved in BFB cycles than others. Palindromic fusions can also be composed of inverted primate *Alu* elements or enabled through microhomology-mediated events (Marotta et al. 2013). Characterizing the sequence features of regions where breakage and subsequent fusions happen is critical for our understanding of mechanisms underlying degenerative cycles.

The products of BFB cycles depend on the region of breakage of the fused sister chromatids which leads to losses, duplication or novel rearrangements. Sustained cycles can generate highly amplified regions (Lo et al. 2002; Maser and DePinho 2002; Tanaka et al. 2002; Narayanan et al. 2006; Tanaka and Yao 2009), enter a state of stasis (Ciullo et al. 2002) or chromosomes are 'healed' by definitive breakage at the location of the initial fusion event (Pobiega and Marcand 2010; Lopez et al. 2015). Because BFB chromosomes are often dicentric and lack telomeres during the initial stages, chromosomes can gain more stability through centromere inactivation (Pennaneach and Kolodner 2009; MacKinnon and Campbell 2011; Sato et al. 2012; Song et al. 2013) and *de novo* telomere formation (Pennaneach and Kolodner 2009; Murnane 2012). Chromosome repair after BFB can occur through repeat-induced non-homologous recombination events (Hoang et al. 2010; Song et al. 2013). The mechanisms and sequence characteristics determining why BFB chromosomes either

undergo progressive degenerative cycles or the cycle becomes arrested and chromosomes 'cured' are not known.

Fungal pathogens have highly dynamic genomes with substantial variation in chromosome length and number within a species. Many fungal pathogens have so-called accessory chromosomes that have presence-absence polymorphism within a species and typically accumulate more mutations and structural variations than other regions of the genome (Bertazzoni et al. 2018). Interestingly, genes determining interactions with the host are often found in highly polymorphic compartments of the genome (Sánchez-Vallet et al. 2018). BFB has been speculated to play a role in generating novel chromosomes from a highly polymorphic accessory chromosome in the wheat pathogen *Zymoseptoria tritici* (Croll et al. 2013). *Z. tritici* also shows among the most extreme degrees of structural variation observed within fungal species including large differences in terms of chromosome length, TE and gene content, recombination rate as well as telomere and centromere structure (Croll et al. 2013; Schotanus et al. 2015; Plissonneau et al. 2016; Fouché, Plissonneau, and Croll 2018; Plissonneau et al. 2018; Sánchez-Vallet et al. 2018; Badet et al. 2019). Chromosome rearrangements occur frequently during meiosis with higher rates of rearrangements observed in more polymorphic chromosomes (Croll et al. 2013; Fouché, Plissonneau, McDonald, et al. 2018). A chromosomal rearrangement causing the near doubling of the original chromosome was hypothesized to have triggered the onset of BFB (Croll et al. 2013). The novel enlarged chromosome should hence be highly unstable due to the presence of two centromeres and enter degenerative cycles in future rounds of meiosis.

In this study we sequenced the genomes of twelve progeny of *Z. tritici* using long reads in order to track the onset of BFB cycles. We experimentally set up four consecutive cycles of

meiosis to test multiple hypotheses about how parental combinations of chromosome variants would affect degenerative cycles. First, we aimed to generate a base-pair resolution assembly of the primary occurrence of the chromosomal rearrangement event. Hereafter, we performed three crosses to determine the stability of the BFB chromosome when paired either with a homologous chromosome variant from either before or after the onset of the BFB. We analyzed a third and fourth round of meiosis to determine whether BFB chromosomes would stabilize or continue to degenerate. Our study provides insights into the exact sequence rearrangements underlying entry into highly unstable BFB during meiosis through ectopic recombination and multiple non-disjunction events.

Materials and Methods

Performing crosses

We performed four generations of crosses, including several backcrosses (fig. 1). During the first meiotic cycle, isolate ST99CH1A5 was crossed with isolate ST99CH1E4 (abbreviated 1A5 and 1E4) isolated from fields ~10 km apart in Switzerland, producing two progeny isolates A2.2 and A66.2 (Croll et al. 2013). Subsequently, during the second meiotic event A2.2 was crossed to an unrelated isolate obtained from the same field 3D7 and A66.2 was backcrossed to 1A5. A2.2 and A66.2 were also crossed with one another. During the third meiotic round, a progeny isolate from the cross A66.2x1A5, namely Ztprog1, was backcrossed to 1E4 and in the final meiotic cycle two progeny from this cross Ztprog19 and Ztprog45 were backcrossed to 1A5 and 1E4, respectively. The genomes of 1A5, 1E4 and 3D7 have been sequenced using PacBio high coverage sequencing and have been assembled into complete chromosomes (Plissonneau et al. 2016; Plissonneau et al. 2018). These assemblies were also validated using high-density genetic maps (Croll et al. 2013; Lendenmann et al. 2014).

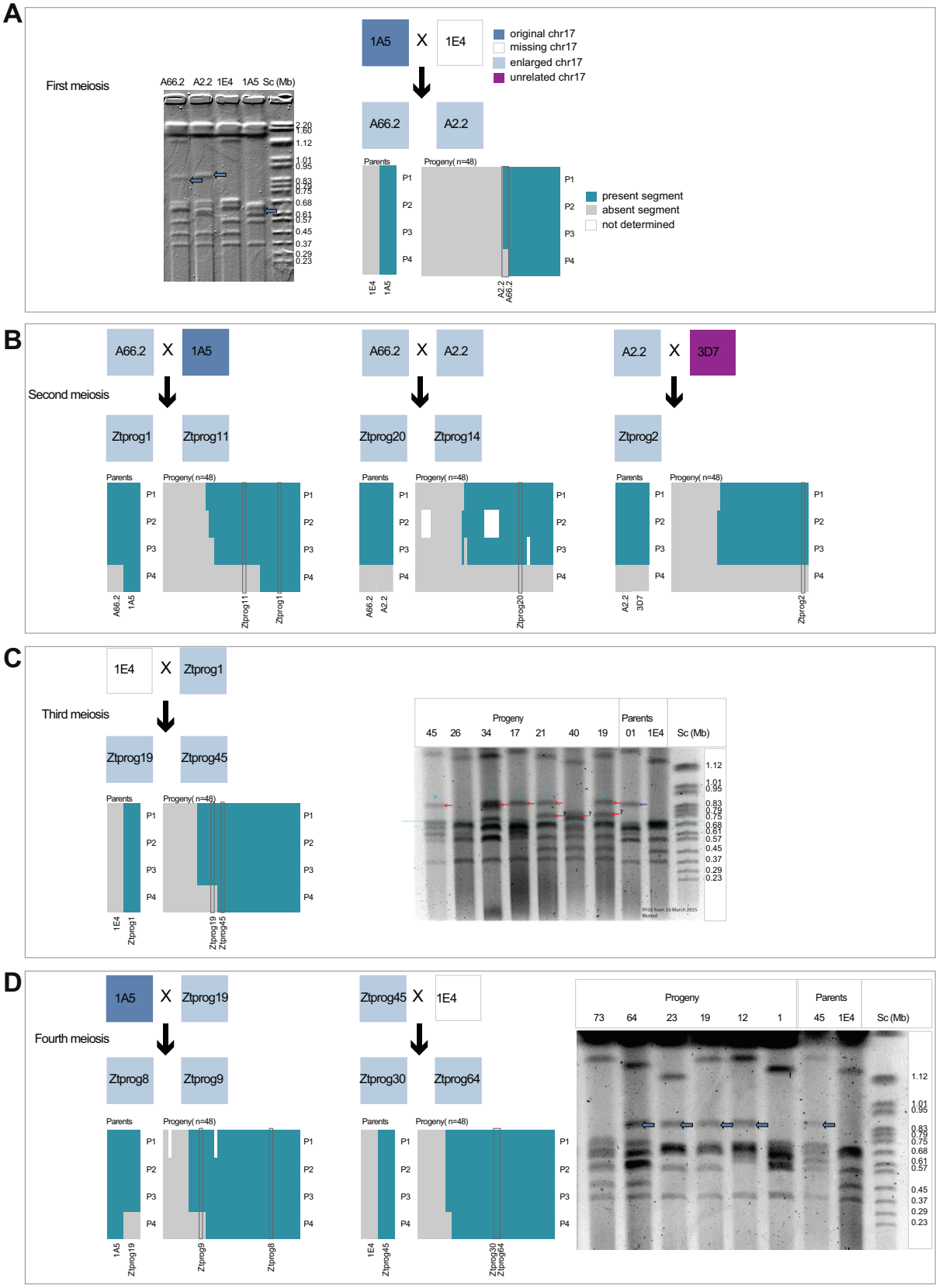


Figure 1: Chromosome pedigree of progeny and parents of four rounds of meiosis (A-D). The colors of the block indicate whether the isolate carries either an original, enlarged or no chromosome

17 according to pulsed-field gel electrophoresis (PFGE) analysis. Barplots below each cross summarize the presence (turquoise) or absence (grey) of four chromosomal segments assayed by PCR in parents and progeny (n=48) from each cross. Electrophoretic karyotype diversity of chromosome 17 among a selection of progeny from each cross are shown along parental karyotypes. The size marker is *Saccharomyces cerevisiae* chromosomes (Sc). Arrows indicate the most likely band representing chromosome 17.

We performed crosses by coinfecting wheat leaves with asexual conidia from the parental strains of opposite mating types (as described above) according to the crossing protocol described by Kema et al. (1996). Spores of both parents were sprayed onto wheat plants in equal concentration and incubated outdoors for 40–60 days. Ascospores were isolated over several days by incubating infected wheat leaves on wet filter paper inside Petri dishes. Wheat leaves were covered with upside down Petri dish lids filled with water agar enabling the capture of vertically ejected ascospores. Ascospores captured on the water agar were left to germinate and inspected for contaminants. Only progeny isolates from single ascospores were selected. Each germinating ascospore was transferred to an individual culture plate for clonal propagation. The mycelium produced by each ascospore was used for DNA extraction. Progeny mycelium was grown in YSB (yeast sucrose broth) liquid media for 6–7 days at 20°C prior to DNA extraction.

DNA extraction

DNA from the progeny from each cross was extracted using a modified version of the cetyltrimethylammonium bromide (CTAB) DNA extraction protocol that was developed for plant DNA extraction (Allen et al. 2006). Fungal spores were grown for 5-7 days in YSB broth and lyophilized overnight. 60-100 mg of dried material were crushed with a mortar and pestle. The phenol-chloroform- isoamyl alcohol extraction step was performed twice and the

washing step three times. In the last step, the DNA pellet was resuspended in 100 µl of sterile water.

Library preparation

PacBio SMRTbell libraries were prepared using 15-31 µg of high-molecular-weight DNA. The libraries were size selected with an 8 kb cutoff on a BluePippin system (Sage Science). After selection, the average fragment length was 15 kb. PacBio sequencing was run on a PacBio RS II instrument or Sequel at the Functional Genomics Center, Zurich, Switzerland using P4/C2 and P6/C4 chemistry, respectively.

Genome assembly using self-corrected PacBio long reads.

PacBio read assembly was performed using HGAP version 4 of the SMRTanalysis suite (version 6, release 6.0.0.47841) (Chin et al. 2013). HGAP was run with the default parameters, except for the minimum seed read length, to initiate the self-correction. First, we produced assemblies for all chromosomes except chromosome 17 using the cutoffs automatically chosen by HGAP (supplementary table S1). Isolates with a too low coverage overall (<30x) or with highly fragmented assemblies (172 contigs) were discarded. In order to improve the contiguity of the assembly, we tested minimum seed read lengths of 8000, 10000, 12000, 15000, 20000 (supplementary table S1). Chromosomes which could not be assembled into a single contig were scaffolded into chromosomes using Ragout (Reference-Assisted Genome Ordering UTility) version 2.2 with default parameters and Sibelia for synteny block decomposition (Kolmogorov et al. 2014; Kolmogorov et al. 2018). Unusual assemblies were analyzed using Quiver 6.0.0.47835 with default settings as implemented in the SMRTanalysis suite and mapped reads were inspected visually in IGV (2.4.10). The synteny of assembled progenies was compared to the parent isolate 1A5 using Nucmer

(mummer 3.23) with the parameters delta-filter -i 80 -l 1000 -1 -q. A mummerplot was generated with the parameters --filter --large --fat --layout-t postscript to visually inspect assembly contiguity. We also performed a homology search using Blastn to the reference genome 1A5 and hits were filtered for minimum identity of 98% and a bitscore of 1000 with a minimum alignment length of 1 kb.

Annotation of coding sequences and repetitive elements

Repetitive elements in the genomes of the isolates were annotated using RepeatMasker version 4.0.7_4 (Smit AFA. "RepeatMasker." URL: http://www.repeatmasker.org (1996-2005)) using default parameters and the repeat library from Badet et al., 2019. Coding sequences were predicted using braker version 1.9 (Hoff et al. 2016). Braker uses intron hints identified based on mapping RNA-seq reads. We used RNA-seq reads from isolates 1E4 and 1A5 for four time points 7, 12, 14 and 28 days post inoculation of an infection cycle on the wheat cultivar Drifter (Palma-Guerrero et al. 2017) and in two *in vitro* conditions one in minimal medium and one in nutrient rich medium (Francisco et al. 2019). Reads were aligned to the genome using Hisat2 (v2.1.0) (Pertea et al. 2016) using a maximum intron length cutoff of 500. To obtain the predicted intron junction sites, we used bam2hints from Augustus v3.2.1 (Stanke et al. 2006). Intron hints were used in braker to predict genes in each assembled genome.

Assembly of chromosome 17

We followed the same approach to assemble chromosome 17 in progeny from all four generations (fig. 1). We mapped the reads to the reference genome 1A5 using minimap2 (Li 2018) with the parameters --secondary=no -ax map-pb. We compared the coverage of regions of chromosome 17 to the mean coverage of the core chromosomes (1-13) to

determine the copy number of distinct regions. We then identified breakpoints by asking for >15 reads to either end or start at a specific position using bedtools bamtobed (Quinlan and Hall 2010) and extracted split reads in this region. Hereafter, we assembled draft chromosomes by using the information from reads showing split alignments and joining individual breakpoints (supplementary table S2). Reads were mapped to the assembled chromosomes with minimap2 and we counted the number of reads going through each established junction point (supplementary table S2). Established chromosome 17 assemblies were error-corrected with Quiver (Chin et al. 2013).

Chromosome 17 segment PCR assay

In order to survey the presence-absence polymorphism of chromosome 17 segments, we used previously designed PCR assays to amplify ca. 500 bp regions of coding sequences at regular intervals along the chromosome 17 of reference strain IPO323 (Croll et al. 2013). For detailed information where these primers amplify on the genome of 1A5, see supplementary table S3. PCR reactions were performed in 20 µl volumes with 5–10 ng genomic DNA, 0.5 mM of each primer, 0.25 mM dNTP, 0.6 U Taq polymerase (DreamTaq, Thermo Fisher, Inc.) and the corresponding PCR buffer. In order to avoid false negatives, we included a primer pair of a microsatellite locus in each PCR mix (Goodwin et al. 2007). Successful PCRs produced an additional band that was clearly distinguishable from the PCR product associated with the amplified chromosome region. PCR products were analyzed agarose gels. We used the R graphics package heatmap2 to visualize the presence-absence dataset.

Preparation of fungal material for molecular karyotyping

DNA from intact chromosomes was extracted from conidia embedded in agarose gels by the *in situ* digestion of cell walls, using a modified non-protoplasting method (McCluskey et al. 1990). We included seven *Z. tritici* isolates that were confirmed to have a chromosome 17 from each of the crosses. Isolates were transferred from stocks maintained in glycerol at -80°C to Yeast Malt Agar (YMA) plates and were incubated for 3-4 days in the dark at 18°C. Hereafter, conidia were isolated by washing the plates with sterile water and transferring 600–800 μl of suspended conidia to new YMA plates. The plates were again incubated for 2 to 3 days as described above. Conidia were harvested by washing the plates with sterile distilled water and filtered through sterile Miracloth (Calbiochem, La Jolla CA, USA) into 50 ml Falcon tubes. The volume was adjusted to 50 ml by adding more distilled water and the suspension was centrifuged at 3750 rpm at room temperature for 15 min with a clinical centrifuge (Allegra X-12R, Beckman Coulter, Brea CA, USA). The pellets were resuspended in 1–3 ml TE buffer (10 mM Tris-HCL, pH 7.5; 1 mM EDTA, pH 8.0) and vortexed gently. The spore concentration of the solution was calculated using a Thoma haematocytometer cell counter. The 1.5 ml spore suspensions with a concentration between $8x10^7$ to $2x10^8$ spores/ml were transferred to new 50 ml falcon tubes and incubated at 55°C in a water bath for a few minutes. Hereafter, we added 1.5 ml pre-warmed (55°C) low-melting-point agarose prepared in TE Buffer (2% w/v; molecular biology grade, Biofinex, Switzerland). The solution was mixed by gentle pipetting. An aliquot of 500 μl was solidified on ice for approximately 10 min in a pre-cooled plug casting mold (BioRad Laboratories, Switzerland). Agarose plugs were incubated in 15 ml falcon tubes containing 5 ml of a lysing solution containing 0.25 M EDTA, pH 8.0, 1.5 mg/mL protease XIV (Sigma, St. Louis MO, USA), 1.0% sodium dodecyl sulfate (Fluka, Switzerland). Plugs were incubated for ~24h at 55°C. The lysing solution was changed once after ~18h and gently mixed every few hours. Plugs with whole chromosomal

DNA were washed three times for 15-20 min in ~5ml of a 0.1 M EDTA (pH 9.0) solution and then stored in the same solution at 4°C until used.

Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) was performed using a BioRad CHEF II apparatus (BioRad Laboratories, Hercules CA, USA). Chromosomal plugs were placed in the wells of a 1.2% (wt/vol) agarose gel (Invitrogen, Switzerland) to separate small chromosomes up to 1Mb. Chromosomes were separated at 13°C in 0.56 x Tris-borate-EDTA Buffer (Sambrook & Russell 2001) at 200 V with a 60–120 s pulse time gradient for 24–26 h. Gels were stained in ethidium bromide (0.5 mg/ml) for 30 min. Destaining was performed in water for 5–10 min. Photographs were taken under ultraviolet light with a Molecular Imager (Gel Doc XR+, BioRad, Switzerland). As size standards, we used chromosome preparations of *Saccharomyces cerevisia*e (BioRad, Switzerland).

Southern transfer and hybridization of pulsed-field gels

Southern blotting and hybridization were performed following standard protocols (Sambrook & Russell 2001). First, hydrolysis was performed in 0.25 M HCl for 30 min, then DNA was transferred onto Amersham HybondTM-N+ membranes (GE Healthcare, Switzerland) overnight under alkaline conditions Sambrook & Russell 2001). DNA was heat-fixed onto the membranes at 80°C for 2 h. Membranes were prehybridized overnight with 25 ml of a buffer containing 20% (w/v) SDS, 10% BSA, 0.5 M EDTA (pH 8.0), 1 M sodium phosphate (pH 7.2) and 0.5 ml of sonicated fish sperm solution (Roche Diagnostics, Switzerland). Probes were

labeled with ^{32}P by nick translation (New England Biolabs, Inc.) following the manufacturer's instructions. Hybridization was performed overnight at 65°C. Blots were subjected to stringent wash conditions with a first wash in 16 X SSC and 0.1% SDS and a second wash with 0.26 X SSC and 0.1% SDS. Both washes were performed at 60°C. Membranes were exposed to X-ray film (Kodak BioMax MS) for 2 to 3 days at -80°C. We used the same probe as in Croll et al., 2013 (supplementary table S4).

Results

Assembly and annotation of progeny genomes across the pedigree

In order to quantify exhaustively rearrangements occurring throughout a pedigree, we assembled eleven genomes of the progeny of several different crosses using high coverage PacBio sequencing (fig. 1). The isolates 1A5 and 1E4 were collected from the same field in Switzerland in 1999 and experimentally crossed to obtain progeny A66.2 and A2.2. Progeny A66.2 was crossed with 1A5 to obtain Ztprog1 and Ztprog11. Progeny A66.2 was also crossed to A2.2 to obtain Ztprog20. Progeny A2.2 was crossed to the unrelated isolate 3D7 from a nearby wheat field in Switzerland to produce progeny Ztprog2. In a third round of meiosis Ztprog1 was crossed to 1E4 to produce Ztprog19 and Ztprog45 which were backcrossed to 1A5 and 1E4, respectively. We sequenced Ztprog8 and Ztprog9 from cross Ztprog19x1A5, and Ztprog30 and Ztprog64 from cross Ztprog45x1E4 (fig. 1). The progeny recovered from the pedigree all had 20 or 21 chromosomes, which was similar to the parent isolates 1A5 (21 chromosomes) and 1E4 (missing only 17). The sizes of the assembled progeny genomes ranged from 38.51 Mb to 39.3 Mb (supplementary table S5 and supplementary fig. S1a). We obtained no high contiguity for the Ztprog2 and only looked at chromosome 17. Except for chromosome 17, the read depth across chromosomes suggests no disomic chromosomes appearing in any of the crosses (i.e. double the coverage for a chromosome). Hereafter, we annotated the genomes using splicing evidence from RNA datasets collected *in planta* and in culture medium. The total number of genes identified in each of the eleven progeny genomes ranged from 11,556 to 12,092 (supplementary table S1 and supplementary fig. S1b). We evaluated the completeness of assembled genomes by locating BUSCO genes and found that all genomes were highly complete, ranging from 98.4% to 98.7% (supplementary table S5).

Identification of a major reciprocal translocation event

We identified two novel chromosomes in the isolate Ztprog8 from parents Ztprog19 and 1A5 from the fourth round of meiosis. In order to identify the origin of the novel chromosomes, we performed a chromosome length dotplot of the reference strain 1A5 against the assembly of Ztprog8 and found that the contigs mapped to chromosome 6 and 12 of the parental genomes (supplementary fig. S2a). The length of chromosome 6 was 2.7 Mb in 1A5 and 2.5 Mb in 1E4 and the length of chromosome 12 was 1.5 Mb in both 1A5 and 1E4. The total length of chromosome 6 and 12 is 4.2 Mb in 1A5 and 4 Mb in 1E4. The two novel chromosomes in progeny Ztprog8 are 3 Mb (chr_6_12.03) and 1 Mb (chr_6_12.12), totalling 4 Mb which suggests a reciprocal translocation event between parental chromosomes 6 and 12. Both novel chromosomes had telomeric repeats present on both termini. To better understand the rearrangements that generated the novel chromosomes, we identified the breakpoints of the reciprocal translocations by performing a homology search. Chr_6_12.03 sequences mapped to both chromosome 6 and 12 between positions 672 kb and 680 kb, indicating that rearrangement breakpoints are located in this region (supplementary fig. S2b and c). The breakpoints on chr_6_12.12 are located at positions 228 kb and 233 kb (supplementary fig. 2b and c). We mapped long-reads reads to the assembled chromosomes chr_6_12.03 and chr_6_12.12 using Quiver and found no anomaly in coverage indicative of a mis-assembly (supplementary fig. S2d and e). The sum of genes identified on both rearranged chromosomes together is 1192, while the sum of the genes on chromosomes 6 and 12 of 1A5 and 1E4 is 1223 and 1207, respectively.

Chromosomal rearrangements and initiation of breakage-fusion-bridge cycles

A partially duplicated chromosome 17 was previously discovered in the progeny A66.2 and A2.2 from the cross 1A5x1E4 (fig. 1A) (Croll et al. 2013). We wanted to determine how frequently such chromosome rearrangements occurred in this cross. For this we performed a segment-specific PCR assay for 48 progeny (fig. 1A, supplementary fig. S3). We found that anomalies of chromosome 17 were most likely restricted to progeny A66.2 and A2.2 (fig. 1A, supplementary fig. S3). Chromosome 17 was found in 28/48 progeny. Chromosome 17 from A66.2 and A2.2 had 1.5 times the mean coverage of the core chromosomes, suggesting a partial duplication (supplementary table S6). Analyzing read mapping along the parental chromosome 17, we identified a breakpoint at position 341 kb (supplementary table S2). Reads mapping at this position were evenly split between reads continuing at position 346 kb in the same orientation and reads continuing at position 484 kb in an inverted orientation. This suggests that the duplicated region indeed exists twice in the progeny genome and that the duplicated sequences are connected to two distinct locations of the chromosome. Coverage on the parental chromosome 17 strongly suggests that the entire 1-341 kb region is duplicated as indicated (fig. 2A). Based on coverage, the region 346-484 kb is single copy. A lack of coverage after position 484 kb and the failure of a PCR to amplify in this region jointly indicate that this region is missing in the progeny (fig. 2A). Using information on coverage and reads spanning distinct chromosomal regions, the chromosome 17 of progeny A66.2 and A2.2 was is composed of region 1-341 kb followed by region 346-484 kb and a second copy of region 1-341 kb in an inverted orientation (fig. 2A and B; supplementary table S2). We reconstructed the progeny chromosome 17 based on the above links and sequences of the parental chromosome 17 from 1A5. We used long-reads of the progeny sequencing to polish the reconstructed chromosome 17. The enlarged chromosome 17 was 819 kb in length for A66.2 and A2.2 matching the length identified from a PFGE gel separating the chromosomes by length (fig. 2D).

To understand the mechanism triggering the rearrangement and size increase of chromosome 17, we analyzed breakpoint sequences. We identified a full copy of the TE *Styx* at the position 341 kb and a partial copy of *Styx* at position 346 kb in the parental chromosome of 1A5 (fig. 2C). *Styx* is an unclassified TE with a copy number of 32 in the 1A5 genome including three copies on chromosome 17 (fig. 2C). In addition to position 341 kb, a second complete copy of *Styx* was found at 370 kb (each ~8 kb in length). Full-length copies of *Styx* contain four unknown coding regions of which one shows weak homology to RNAse H and an integrase (supplementary fig. S4A). We analyzed *Styx* copy numbers among 19 complete genomes of *Z. tritici* (Badet et al. 2019). *Styx* is present in high copy numbers in the sister species *Z. pseudotritici* but only the short variant of *Styx* was found. *Styx* is at low copy numbers in the isolates from the center of origin of *Z. tritici* (supplementary fig. S4B). *Styx* appears to have undergone a recent burst in Europe resulting in high copy numbers in 1A5 and 1E4 and an isolate from Argentina (supplementary fig. S4B). The long variant of *Styx* is most abundant in *Z. tritici* (supplementary fig. S4C). The generation of the enlarged chromosome 17 was possibly mediated by ectopic recombination between *Styx* copies at positions 341 kb and a different sequence at 484 kb, but we identified no apparent homology (fig. 2 C). We found a copy of the MITE *Areion* adjacent to the breakpoint at 484 kb. In general, TEs are largely conserved between parental and progeny chromosome 17 copies with the exception of deleted RLG-Pluto and RLG-Pan copies (fig. 2E).

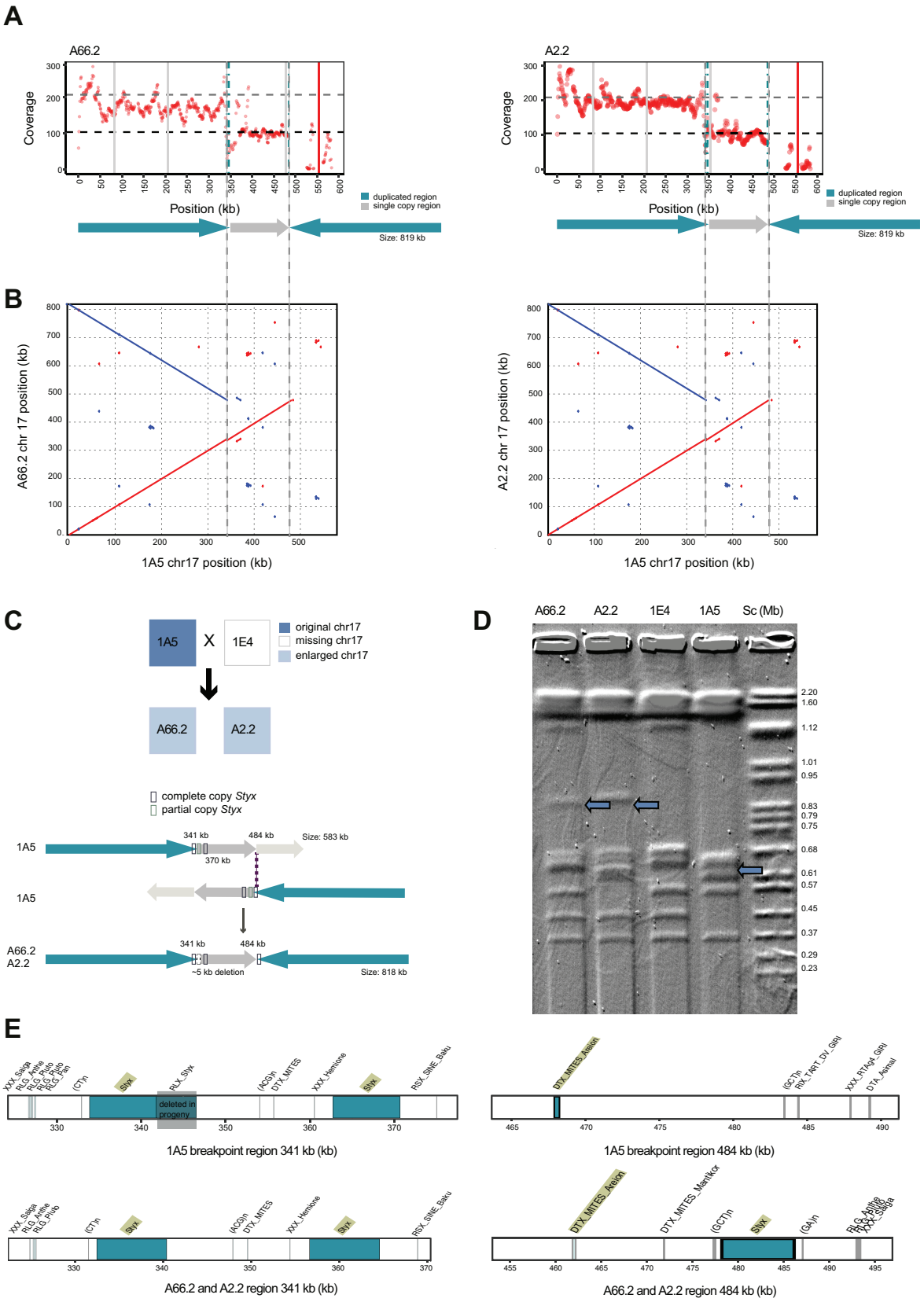


Figure 2: Generation of the enlarged chromosome 17 in A2.2 and A66.2. (A) The coverage and breakpoints of the progeny A66.2 and A2.2 reads mapped to the parent 1A5, horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two times the mean core

chromosome coverage. Red dots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage are removed). Vertical dashed lines indicate the chromosomal breakpoints identified from mapped reads. Solid vertical lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). (B) Dotplots of the assembled chromosome for progeny A66.2 and A2.2, respectively, compared to the parental chromosome. Inverted regions are indicated in blue. (C) Schematic representation of the breakpoints and rearrangement between two copies of the 1A5 chromosome 17 generating the enlarged chromosome 17 in the progeny. Copies of *Styx* are not to scale. (D) Pulsed-field gel electrophoresis of the parental chromosomes (1E4 and 1A5) and the progeny A2.2 and A66.2, showing the enlarged chromosome (arrows; Croll et al., 2013) (E) TEs in the vicinity of the breakpoint 341 kb (left) and 484 kb (right) and the corresponding regions in parent and progeny chromosomes.

Sustained chromosome degeneration in subsequent rounds of meiosis

We wanted to track how the novel elongated chromosome 17 degenerates through further rounds of meiosis and crossed progeny A66.2 back to the parent 1A5. We examined the chromosome 17 segment-specific absence-presence profiles of 48 progenies with a PCR assay (fig. 3B). There was substantial length diversity among the 48 progenies from this cross (fig. 3B, supplementary fig. S5). Fifteen progenies, including Ztprog11, were missing one segment of chromosome 17 and in rare cases progeny were missing two segments or three segments. We sequenced Ztprog1 and Ztprog11 from this cross and identified a new breakpoint in Ztprog1 at position 341 kb were reads mapped to position 370 kb in the same orientation (fig. 3B, supplementary table S2). As in the parent A66.2, reads mapping to position 341 kb also mapped to position 346 kb and position 484 kb in the same and inverted orientation, respectively (fig. 3B, supplementary table S2). Interestingly, the third, full-length copy of *Styx* was located at position 370 kb in the 1A5 parental chromosome (fig. 2C). Read mapping shows that the region 1-341 kb has approximately three times the mean core chromosome coverage, the region 341-370 kb and region 484-580 kb have one time the mean core chromosome coverage and region 370-480 kb has two times the mean core chromosome coverage (fig. 3B). All regions were confirmed to be present by PCR. The coverage and breakpoints suggest that there are four potential versions of this chromosome

(V1-V4) (fig. 3, supplementary fig. S6A and C). One small (either V1 or V2) and one large chromosome variant (either V3 or V4) is present, as indicated. Furthermore, coverage data suggests that one chromosome 17 variant has a deletion in region 341-346 kb (as in parent A66.2, where the partial copy of *Styx* is missing). The other chromosome 17 variant has a deletion in the region 341-370 kb (supplementary fig. S6D). Therefore, either variant V1 is present together with V4, or variant V2 is present together with V3 (fig. 3B). These sequence rearrangements were supported by reads mapping through all three of these connection points (supplementary table S2). The sizes of the large variants are 819 kb (V3) or 796 kb (V4), respectively (fig. 3B). We confirmed by PFGE that the chromosome is ~0.83 Mb (supplementary fig. S6B), suggesting that the larger variant (V3) is more likely to be present together with V2. The small chromosome 17 variants are 578 kb (V1) and 554 kb (V2) depending on the deletion, respectively. The PFGE suggests a size of ~0.57 Mb (supplementary fig. S6B). Variant V3 may be identical to the chromosome 17 of A66.2 or recombination between *Styx* in 1A5 at position 341 kb and the copy of *Styx* in A66.2 where the second partial copy is already deleted (supplementary fig. S6D). V2 was most likely generated by ectopic recombination between the copy of *Styx* at position 341 kb in A66.2 and the *Styx* copy at position 370 kb in 1A5 (supplementary fig. S6D). There is only one copy of Styx where these chromosomes fused in the progeny Ztprog1.

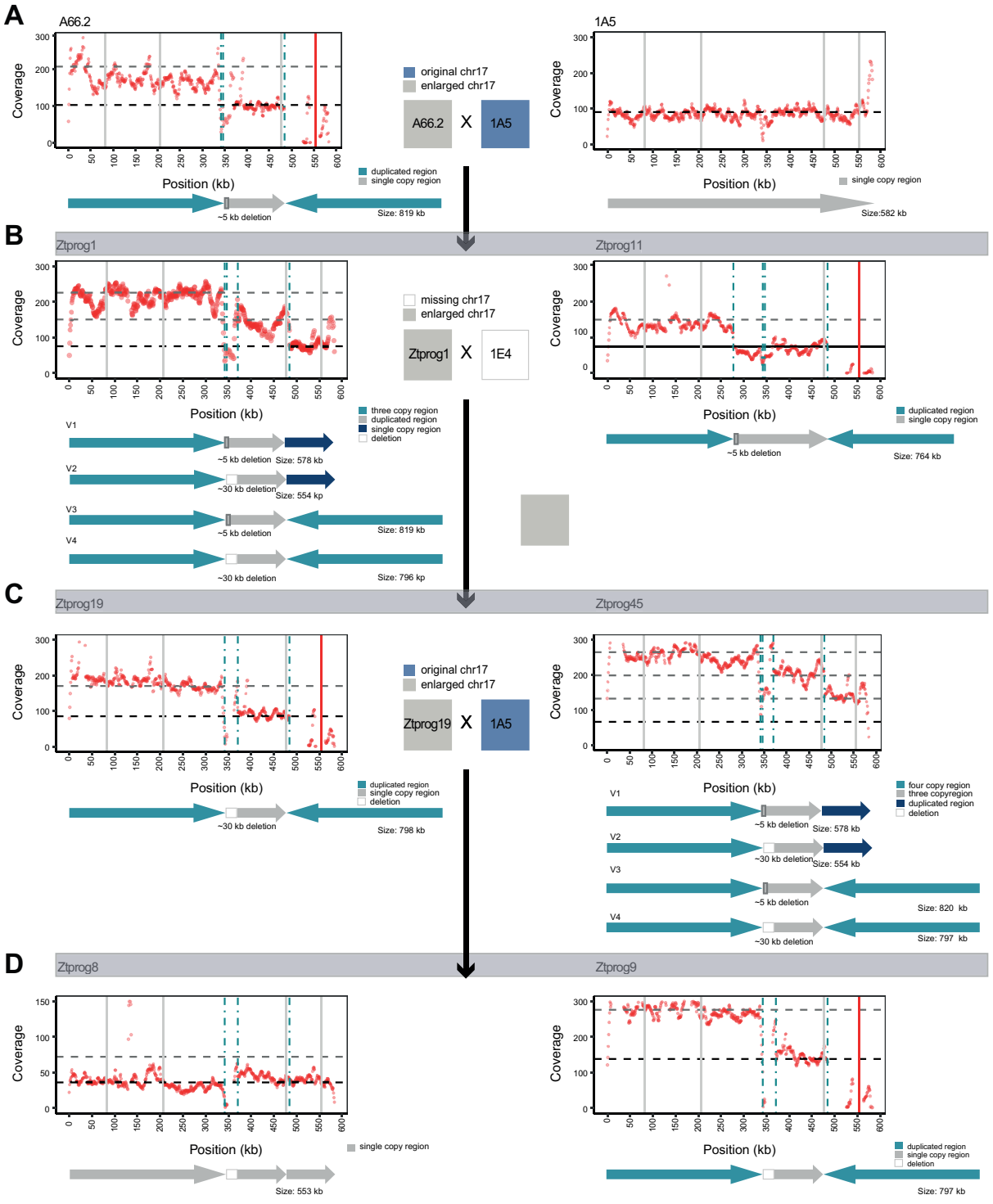


Figure 3: Reconstruction of chromosome 17 variants based on breakpoint analyses (A-D). The coverage and breakpoints of the progeny over four rounds of meiosis. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two, three and four times the mean core chromosome coverage. Red dots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage are removed). Vertical dashed lines indicate the chromosomal breakpoints identified from mapped reads. Solid vertical lines indicate the positions of

loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be present in this isolate.

The sister progeny Ztprog11 obtained from the second round of meiosis shows a new rearrangement breakpoint. We identified reads connecting positions 277 kb and 484 kb in an inverted orientation (supplementary table S2; fig. 3B). The breakpoint is close to a partial copy of the MITE *Pegasos*. We also found support for the previously known breakpoint as before, connecting positions 341 kb and 346 kb shared with the parent A66.2. In contrast to the parent A66.2, only the region 1-277 kb appears duplicated though (fig. 3B). The missing sequence beyond position 484 kb is confirmed by PCR (fig. 3B). The chromosome 17 of this progeny appears to have inhered region 1-484 kb from A66.2 with the ~5 kb deletion, followed by region 1-277 kb in an inverted orientation. This novel variant of the original rearrangement was likely generated through ectopic recombination with the regular chromosome 17 variant present in parent 1A5 (supplementary fig. 7C and D). The error-corrected chromosomal assembly is 764 kb in length which matches with the expected size from PFGE (~0.79 Mb) (supplementary fig. S7B).

We performed two additional second rounds of meiosis in order to track the stability of the enlarged chromosome 17 (fig. 1B). Progeny from cross 1A5 with A66.2 show more length diversity in the absence-presence profile of chromosome 17 and the PFGE (supplementary fig. S8A) than progeny from A2.2 crossed with 3D7 (supplementary fig. S8C). The assembled chromosome 17 variants of Ztprog1 and Ztprog11 from the cross 1A5 with A66.2 were both different from each other and from either parent, showing the polymorphic nature of chromosome 17 variants from this cross. We found that the chromosome 17 from Ztprog2 from a cross of A2.2 with 3D7 has exactly the same breakpoints as A2.2 and was most likely inherited from the parent without further rearrangements. The assembled chromosome is 818 kb in length and matches the size of A2.2 in the PFGE (supplementary fig. S8C).

We next tested whether chromosome 17 was faithfully transmitted through meiosis if not paired with a second copy of chromosome 17 in the mating partner. We did this by crossing Ztprog1 (the progeny of 1A5 crossed with A66.2) with 1E4 (lacking chromosome 17). We found that 29/48 progeny were positive for chromosome 17 suggesting an overrepresentation of the chromosome among progeny (supplementary fig. S9). We sequenced two progenies Ztprog19 and Ztprog45 from this cross displaying similarly sized chromosome 17 but different PCR profiles (fig. 1C, supplementary fig. S9), to identify the nature of the chromosome 17 variants. We found that chromosome 17 of Ztprog19 harbored the breakpoint at position 341 kb and reads connecting the position to both position 370 kb and position 484 kb in an inverted orientation (supplementary table S2). The region 1-341 kb was duplicated and two regions (341-370kb and >480 b) were missing (fig. 3C). Read mapping analyses show that this chromosome is composed of region 1-341 kb followed by region 370-484kb and ending on region 1-341kb in an inverted orientation (supplementary fig. S10C; supplementary table S2). The assembled chromosome is 798 kb in length matching the expected size (~0.83 Mb; supplementary fig. S10B) resembling the chromosome 17 variant V4 predicted to occur in parent Ztprog1. The second band is most likely chromosome 14 which assembled into a 742 kb contig (supplementary fig. S10B). The occurrence of V4 in Ztprog19 suggests that this variant was inherited Ztprog01, but we cannot rule out that the parent Ztprog01 harbors rather V2 and V3 (see above).

Ztprog45 showed identical breakpoints to the parent Ztprog1 (supplementary table S2). However, the coverage profile was different from Ztprog1 (fig. 3C). Reads mapped to the original chromosome 17 show that the region 1-341kb has four times the mean core chromosome coverage, region 341-370 kb (with the small deletion between position 341-346kb) has double the mean core chromosome coverage, region 370-484 kb has three

times the mean core chromosome coverage, while region 484-580 kb has two times the mean core chromosome coverage (fig. 3C). Based on breakpoint and coverage information, the progeny carries most likely two small chromosome variants similar to the original chromosome 17 and one large chromosome 17 variant with region 1-341kb inverted and joined to position 484 kb (fig. 3C, supplementary fig. S11). The parent Ztprog1 carried only one copy of a small chromosome 17 suggesting that non-disjunction played a role in creating the variants present in Ztprog45. Two additional chromosome 17 variants likely have region 341-370 kb (including the ~5kb deletion) and another chromosome 17 variant is likely missing this region. Therefore, Ztprog45 chromosomes are a combination of V1, V2, V3 or V4 as, previously described, but with sizes 578, 554, 520 and 797 kb, respectively. This interpretation matches the PFGE showing a bands at 0.83 Mb and 0.57 Mb (supplementary fig. S11B).

In the final round of meiosis, we tested again for the stability of the rearranged chromosome 17 through backcrosses. For this, we crossed Ztprog19 with 1A5 and Ztprog45 with 1E4 lacking chromosome 17 (fig. 3D, fig. 4D). In cross Ztprog19x1A5, 6/48 progeny in the cross were missing a terminal segment of chromosome 17 (supplementary fig. S12). We sequenced one progeny missing this terminal segment (Ztprog9) and one progeny with all the chromosomal segments (Ztprog8; fig. 3D, supplementary fig. S12). We found that Ztprog8 only has one breakpoint connecting position 341 kb with 370 kb, while Ztprog9 has two breakpoints connecting both position 341 kb to 370 kb and to position 484 kb in an inverted orientation (supplementary table S2). Ztprog9 has a duplicated region 1-341kb, lacks the regions 341-370kb and beyond position 484 kb consistent the PCR assay (fig. 1D and 3D). The chromosome was likely inherited unmodified from parent Ztprog19 (Supplementary fig. S13C and D). The assembled chromosome is 797 kb in length and matches the expected size of the PFGE (~0.83 Mb) (Supplementary fig. S13B). In contrast, chromosome 17 of Ztprog8 is likely the a product of ectopic recombination. The

chromosome 17 variant in Ztprog8 has a uniform coverage suggesting no duplicated regions (fig. 3D). The chromosome is hence likely of short size similar to the original chromosome 17 but carrying a deletion at 341-370 kb (fig. 3D, supplementary fig. S14C; supplementary table S2). The most likely scenario for generating this chromosome is ectopic recombination between the two copies of the TE *Styx* at 341 kb and 370 kb of the chromosome 17 found in Ztprog19 and 1A5. We could not confirm the size of chromosome 17 in the progeny PFGE and Southern hybridization suggesting a size of 553 kb (fig. 3D, supplementary fig. S14B).

We analyzed segregation of chromosome in the cross between Ztprog45 (with multiple chromosome 17 variants) and 1E4 (lacking chromosome 17). We found that chromosome 17 was present in 36/48 progeny indicative of distorted segregation (fig. 1D, supplementary fig S15). We sequenced Ztprog30 and Ztprog64 that were positive for all chromosome segments by PCR. Both progenies showed a breakpoint at position 341 kb connected to both position 370 kb and position 484 kb in an inverted orientation. In addition, Ztprog30 showed reads connecting position 341 with 346 kb (supplementary table S2). We found approximately five times the mean core chromosome coverage of region 1-341kb in Ztprog64 (fig. 4D). Region 341-370 kb was missing, region 370-484 kb had four times the mean core chromosome coverage and region 480-580 kb had three times the mean core chromosome coverage (fig. 4D). PCR assays confirmed the presence of these regions in the progeny (fig. 4D). Taken together, the progeny likely carries one large variant (V1) and three small variants (V2) of the chromosome. Using read mapping data, we propose that the large variant (V1) consists most likely of region 1-34 kb, followed by region 370-484 kb, followed by an inverted version of region 1-341 kb (fig. 4D, supplementary fig. S16C). The small variants (V2) most likely consists out of region 1-341 kb followed by region 370-580 kb (fig. 4D, supplementary fig. S16C; supplementary table S2). The assembled chromosomes were 797 and 554 kb in length, respectively, matching the expected sizes on a PFGE (supplementary fig. S16B). The region 341-370 kb was present in two chromosome 17

variants in the parent Ztprog45, which suggests that ectopic recombination must have occurred between the TE *Styx* copies at 341 kb and *Styx* at 370 kb deleting the region in the progeny. This is at least the second independent observation that ectopic recombination occurred between these two copies of *Styx* (supplementary fig. S6D, V2 and V4).

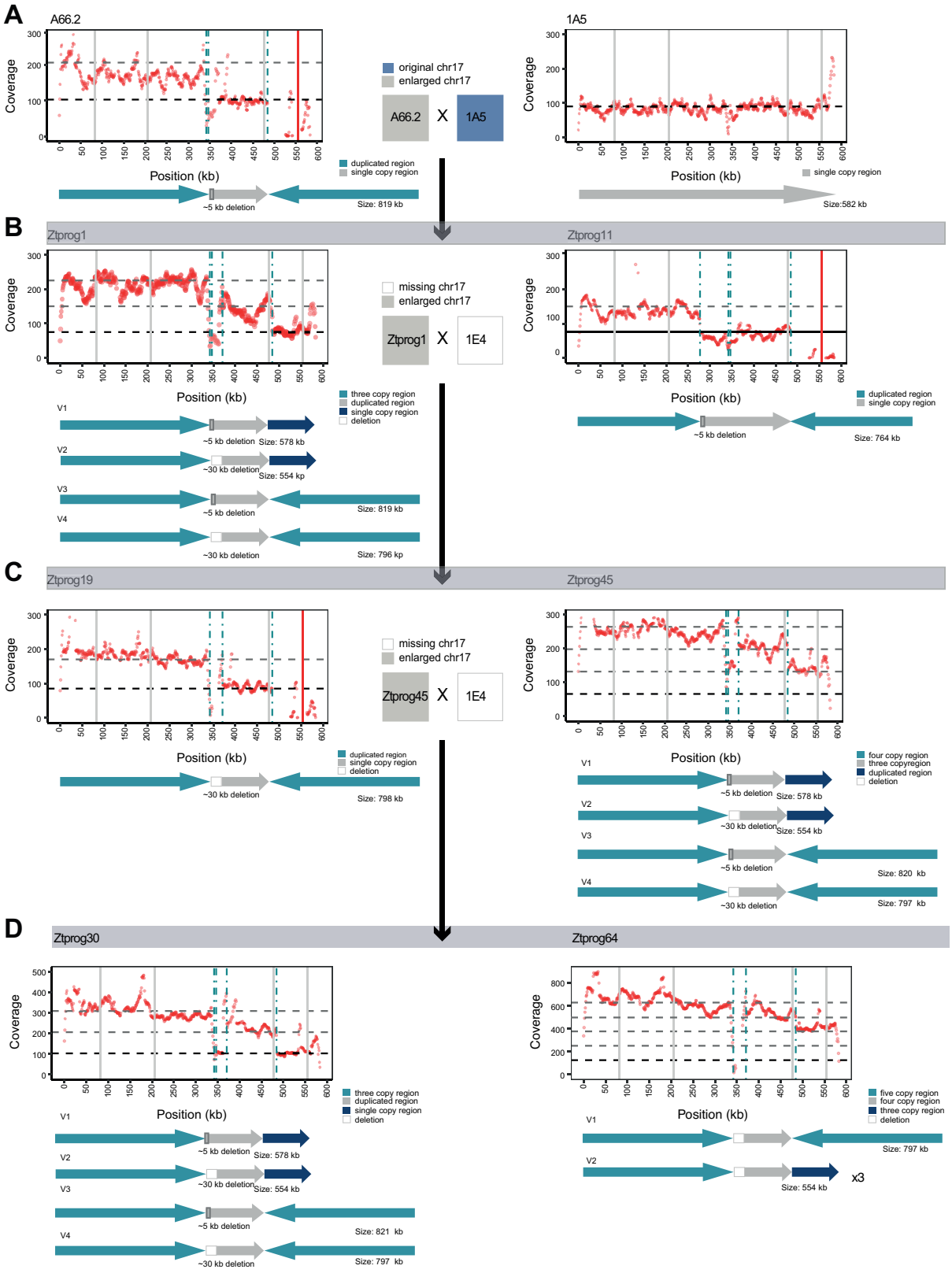


Figure 4: Reconstruction of chromosome 17 variants based on breakpoint analyses (A-D). The coverage and breakpoints of the progeny over four rounds of meiosis. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two, three, four and five times the mean core chromosome coverage. Red dots indicate the mean coverage in 1 kb windows

(regions with excessive, >300, >500 and >900 x coverage are removed). Vertical dashed lines indicate the chromosomal breakpoints identified from mapped reads. Solid vertical lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be present in this isolate.

The chromosome 17 variants in Ztprog30 were inherited without additional ectopic recombination events, in contrast to the situation in Ztprog64. In Ztprog30 region 1-341 kb has three times the mean core chromosome coverage, region 370–484 kb has double the mean core chromosomes coverage and region 341-370 kb with the small (~ 5 kb deletion) and region 484-580 kb is likely single-copy matching evidence from PCR (fig. 4D). Breakpoint analyses suggest that one small and one large version of the chromosome are present in the parent with one lacking region 341-370 kb. All four possible chromosomes (fig. 3D; supplementary fig. S17C) were assembled, error-corrected and read mapping was verified at the breakpoints (supplementary table S2). V1-V4 are 578, 554, 821 and 797 kb, respectively, which matches the chromosome sizes from the PFGE gel (~0.83 Mb and ~0.57Mb) (supplementary fig. S17B). Ztprog45 has different combinations of chromosome 17 variants, two that contain the region between 341kb-370kb and one lacking this region. Hence, Ztprog30 likely inherited chr17 variants without modification from Ztprog45 (fig. 4D).

Summary of the rearrangements throughout four rounds of meiosis

The enlarged chromosome 17 was generated during the first round of meiosis in both progeny through the same rearrangements, namely two deletions (between position b and c; and from position e to f) and a duplication of region 0-b (fig. 5A) of which the second copy is located in an inverted orientation joined to position e. In the next round of meiosis, new variants of chromosome 17 were generated (fig. 5B). Chromosome 17 of Ztprog11 had a duplicated region 0-a and the second copy is joined to position e in an inverted orientation. Ztprog1 has two copies of chromosome 17 and in one copy the region between b and d is

deleted. During the third round of meiosis no new rearrangements occurred (fig. 5C). Multiple chromosome 17 variants are present in Ztprog45 showing non-disjunction of these variants. In the last round of meiosis the region between b and d was deleted in Ztprog8 and Ztprog64 (fig. 5D). Ztprog64 has four chromosome 17 variants missing this region. Ztprog9 and 30 inherited the chromosome 17 complements without further rearrangement from the previous generation.

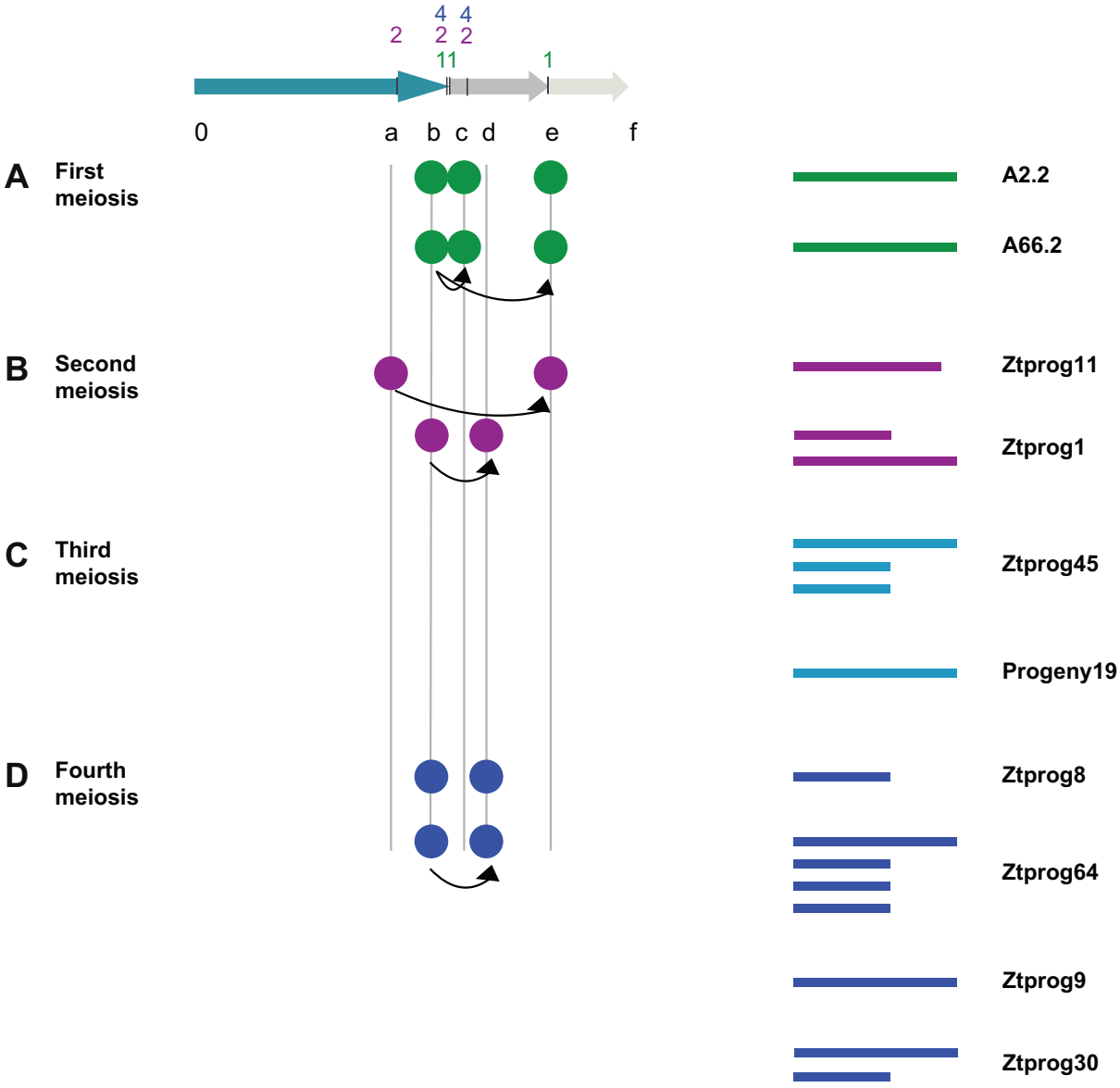


Figure 5: Summary of the rearrangements of chromosome 17 through four rounds of meiosis. Chromosome rearrangements involving positions 0,a-f are indicated and the number of chromosome 17s and in each progeny and whether they are large or small variants.

DISCUSSION

We retraced the degeneration of a chromosome through four rounds of meiosis. Using long-read sequencing, we reconstructed complete homologous chromosomes across a pedigree comprising eleven individuals. We found that the primary degenerative rearrangement was caused by ectopic recombination between copies of the TE family *Styx*. The degenerated chromosome was composed of a large duplicated region connected to a single-copy center core. Subsequent rounds of meiosis increased the spectrum of rearrangement breakpoints suggesting runaway chromosomal degeneration. We also found that non-disjunction events of chromosome 17 increased along pedigree together with disomy and trisomy. Overall, we found that the identity of paired parental genotypes had a major influence on the degree of rearrangements observed in the progeny.

The exact sequence trigger causing the enlarged chromosome in the first round of meiosis was likely the TE *Styx* being present in multiple copies on chromosome 17. In the second round of meiosis non-homologous recombination between the only other copy of Styx generated a new variant of chromosome 17 in Ztprog1. Two chromosome 17s are present in this progeny so non-disjunction of the two chromosome 17 variants occurred together with ectopic recombination. Chromosome 17 variants in the third and fourth round of meiosis were independently generated at the same breakpoints showing that ectopic recombination repeatedly targeted the TE family *Styx*. This shows how specific sequences can act as reliable triggers for the onset of chromosomal degeneration. In the second and third round of meiosis, multiple progeny showed copy-number increases with chromosome 17 becoming disomic or trisomic. Finally, in the fourth round of meiosis, when Ztprog19 is crossed with the original chromosome 17 variant, progeny Ztprog8 and 9 both inherit only one chromosome 17 variant, suggesting that chromosomes undergo fewer non-disjunction events when there is a partner to pair with. Non-disjunction or the maintenance of multiple copies is much more

prevalent in the fourth round of meiosis when crossed with 1E4 with no chromosome 17. The copy-number amplification observed in progeny Ztprog64 and Ztprog30 is similar to what is seen in BFB in many cancers (Ciullo et al. 2002; Hellman et al. 2002; Marotta et al. 2012; Marotta et al. 2013; Marotta et al. 2017; Bianchi et al. 2019). We also identified a likely stabilization of the degeneration in crosses of pairing a single rearranged with an original chromosomal variant (*i.e.* Ztprog8 and Ztprog9). This suggests that the BFB-cycle progression was interrupted through proper segregation. Both chromosomes are single copy and rearranged derivatives of the original chromosome 17 in 1A5. The chromosome 17 variant in Ztprog8 was generated through non-allelic recombination between two *Styx*-copies. The repair of chromosomes through recombination between repeats is known to produce intermediary chromosomes that can ultimately produce stable karyotypes (Hoang et al. 2010).

The completely reconstructed chromosome 17 variants throughout the degenerative process allows to identify sequence triggers for the observed rearrangements. The first generation chromosomal rearrangement was created by an apparently identical ectopic recombination event at the *Styx* locus (position 341 kb) suggesting that this is a fragile site, or a site-frequently co-locating with chromosome rearrangements. In the second round of meiosis two new ectopic recombination events took place. The first recombination event was triggered by a region at 277 kb in close proximity to a MITE and the region at position 484 kb. The second recombination event involved the third copy of *Styx* present on the original chromosome 17 (370 kb). Both non-homologous recombination events caused the duplication of a large chromosomal segment. Amplification of chromosomal sequences is a defining feature of BFB cycles together with the loss of some regions (Lo et al. 2002; Maser and DePinho 2002; Tanaka et al. 2002; Narayanan et al. 2006; Tanaka and Yao 2009). We observed convergent non-homologous rearrangements in the third and fourth round of meiosis triggered by copies of *Styx* at position 341 kb and 370 kb. The tendency of *Styx* to

trigger chromosomal degeneration is likely to be heterogenous within the species. Genomes analyzed from near the center of origin of the pathogen have low *Styx* copy numbers and the expansion to higher copy-numbers appears restricted to European genotypes and their descendants (Badet et al. 2019). Hence, the propensity of triggering BFB cycles on chromosome 17 is may be a derived trait within the species and could be under selection.

Taken together, our results show that specific TE sequences trigger runaway chromosome degeneration. Non-homologous recombination drives the deleterious rearrangements at the onset of the process with non-disjunction events following with the second round of meiosis. Previous studies have shown that BFB cycles are initiated via telomere-telomere fusions of chromosomes with degraded or missing telomeres (Maciejowski and de Lange 2017). In our study, we pinpointed that ectopic recombination provided the initial trigger to create unstable chromosomes. The degenerative cycles were perpetuated by further chromosomal rearrangements and non-disjunction.

References

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. 1:2320.

Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. U. S. A. 105:11845–11850.

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2019. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. bioRxiv:803098.

Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, Hane JK. 2018. Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. Mol. Plant-Microbe Interact. MPMI 31:779–788.

Bi X, Wei S-CD, Rong YS. 2004. Telomere Protection without a Telomerase: The Role of ATM and Mre11 in Drosophila Telomere Maintenance. Curr. Biol. 14:1348–1353.

Bianchi JJ, Murigneux V, Bedora-Faure M, Lescale C, Deriano L. 2019. Breakage-Fusion-Bridge Events Trigger Complex Genome Rearrangements and Amplifications in Developmentally Arrested T Cell Lymphomas. Cell Rep. 27:2847-2858.e4.

Bzymek M, Lovett ST. 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U. S. A. 98:8319–8325.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10:563–569.

Ciullo M, Debily M-A, Rozier L, Autiero M, Billault A, Mayau V, El Marhomy S, Guardiola J, Bernheim A, Coullin P, et al. 2002. Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. Hum. Mol. Genet. 11:2887–2894.

Coquelle A, Pipiras E, Toledo F, Buttin G, Debatisse M. 1997. Expression of Fragile Sites Triggers Intrachromosomal Mammalian Gene Amplification and Sets Boundaries to Early Amplicons. Cell 89:215–225.

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

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46:34–42.

Fouché S, Plissonneau C, McDonald BA, Croll D. 2018. Meiosis Leads to Pervasive Copy-Number Variation and Distorted Inheritance of Accessory Chromosomes of the Wheat Pathogen Zymoseptoria tritici. Genome Biol. Evol. 10:1416–1429.

Francisco CS, Ma X, Zwyssig MM, McDonald BA, Palma-Guerrero J. 2019. Morphological changes in response to environmental stresses in the fungal plant pathogen Zymoseptoria tritici. Sci. Rep. 9:1–18.

Goodwin SB, van der Lee TAJ, Cavaletto JR, te Lintel Hekkert B, Crane CF, Kema GHJ. 2007. Identification and genetic mapping of highly polymorphic microsatellite loci from an EST database of the septoria tritici blotch pathogen Mycosphaerella graminicola. Fungal Genet. Biol. 44:398–414.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Lee TAJV der, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLOS Genet 7:e1002070.

Hackett JA, Feldser DM, Greider CW. 2001. Telomere dysfunction increases mutation rate and genomic instability. Cell 106:275–286.

Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, Trakhtenbrot L, Kerem B. 2002. A role for common fragile site induction in amplification of human oncogenes. Cancer Cell 1:89–97.

Hoang ML, Tan FJ, Lai DC, Celniker SE, Hoskins RA, Dunham MJ, Zheng Y, Koshland D. 2010. Competitive Repair by Naturally Dispersed Repetitive DNA during Non-Allelic Homologous Recombination. PLOS Genet. 6:e1001228.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinforma. Oxf. Engl. 32:767–769.

Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, Odom D, Flicek P, Keane TM, Thybert D, et al. 2018. Chromosome assembly of large and complex genomes using multiple references. Genome Res. [Internet]. Available from: http://genome.cshlp.org/content/early/2018/10/19/gr.236273.118

Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout-a reference-assisted assembly tool for bacterial genomes. Bioinforma. Oxf. Engl. 30:i302-309.

Lendenmann MH, Croll D, Stewart EL, McDonald BA. 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. G3 GenesGenomesGenetics 4:2519–2533.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.

Lo AWI, Sabatier L, Fouladi B, Pottier G, Ricoul M, Murnane JP. 2002. DNA Amplification by Breakage/Fusion/Bridge Cycles Initiated by Spontaneous Telomere Loss in a Human Cancer Cell Line. Neoplasia N. Y. N 4:531–538.

Lopez V, Barinova N, Onishi M, Pobiega S, Pringle JR, Dubrana K, Marcand S. 2015. Cytokinesis breaks dicentric chromosomes preferentially at pericentromeric regions and telomere fusions. Genes Dev. 29:322–336.

Ma C, Martin S, Trask B, Hamlin JL. 1993. Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. Genes Dev. 7:605–620.

Maciejowski J, de Lange T. 2017. Telomeres in cancer: tumour suppression and genome instability. Nat. Rev. Mol. Cell Biol. 18:175–186.

MacKinnon RN, Campbell LJ. 2011. The Role of Dicentric Chromosome Formation and Secondary Centromere Deletion in the Evolution of Myeloid Malignancy. Genet. Res. Int. [Internet]. Available from: https://www.hindawi.com/journals/gri/2011/643628/

Marotta M, Chen X, Inoshita A, Stephens R, Thomas Budd G, Crowe JP, Lyons J, Kondratova A, Tubbs R, Tanaka H. 2012. A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications. Breast Cancer Res. 14:R150.

Marotta M, Chen X, Watanabe T, Faber PW, Diede SJ, Tapscott S, Tubbs R, Kondratova A, Stephens R, Tanaka H. 2013. Homology-mediated end-capping as a primary step of sister chromatid fusion in the breakage-fusion-bridge cycles. Nucleic Acids Res. 41:9732–9740.

Marotta M, Onodera T, Johnson J, Budd GT, Watanabe T, Cui X, Giuliano AE, Niida A, Tanaka H. 2017. Palindromic amplification of the ERBB2 oncogene in primary HER2-positive breast tumors. Sci. Rep. 7:41921.

Maser RS, DePinho RA. 2002. Connecting Chromosomes, Crisis, and Cancer. Science 297:565–569.

McClintock B. 1938. The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes. Genetics 23:315–376.

McClintock B. 1941. The Stability of Broken Ends of Chromosomes in Zea Mays. Genetics 26:234–282.

McCluskey K, Russell BW, Mills D. 1990. Electrophoretic karyotyping without the need for generating protoplasts. Curr. Genet. 18:385–386.

Mieczkowski PA, Lemoine FJ, Petes TD. 2006. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. DNA Repair 5:1010–1020.

Murnane JP. 2012. Telomere dysfunction and chromosome instability. Mutat. Res. 730:28–36.

Narayanan V, Mieczkowski PA, Kim H-M, Petes TD, Lobachev KS. 2006. The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell 125:1283–1296.

Palma-Guerrero J, Ma X, Torriani SFF, Zala M, Francisco CS, Hartmann FE, Croll D, McDonald BA. 2017. Comparative transcriptome tnalyses in *Zymoseptoria tritici* reveal significant differences in gene expression among strains during plant infection. Mol. Plant. Microbe Interact. 30:231–244.

Pastink A, Eeken JC, Lohman PH. 2001. Genomic integrity and the repair of double-strand DNA breaks. Mutat. Res. 480–481:37–50.

Pennaneach V, Kolodner RD. 2009. Stabilization of Dicentric Translocations through Secondary Rearrangements Mediated by Multiple Mechanisms in S. cerevisiae. PLOS ONE 4:e6389.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11:1650–1667.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16:5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio 7:e01231-16.

Pobiega S, Marcand S. 2010. Dicentric breakage at telomere fusions. Genes Dev. 24:720–733.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rank GH, Xiao W, Kolenovsky A, Arndt G. 1988. FLP recombinase induction of the breakage-fusion-bridge cycle and gene conversion in Saccharomyces cerevisiae. Curr. Genet. 13:273–281.

Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet. Genome Res. 120:351–357.

Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56:21–40.

Sato H, Masuda F, Takayama Y, Takahashi K, Saitoh S. 2012. Epigenetic Inactivation and Subsequent Heterochromatinization of a Centromere Stabilize Dicentric Chromosomes. Curr. Biol. 22:658–667.

Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock EH. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin 8:41.

Schubert I. 2007. Chromosome evolution. Curr. Opin. Plant Biol. 10:109–115.

Smith KA, Gorman PA, Stark MB, Groves RP, Stark GR. 1990. Distinctive chromosomal structures are formed very early in the amplification of CAD genes in Syrian hamster cells. Cell 63:1219–1227.

Song W, Gawel M, Dominska M, Greenwell PW, Hazkani-Covo E, Bloom K, Petes TD. 2013. Nonrandom Distribution of Interhomolog Recombination Events Induced by Breakage of a Dicentric Chromosome in Saccharomyces cerevisiae. Genetics 194:69–80.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62.

Tanaka H, Tapscott SJ, Trask BJ, Yao M-C. 2002. Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. Proc. Natl. Acad. Sci. 99:8772–8777.

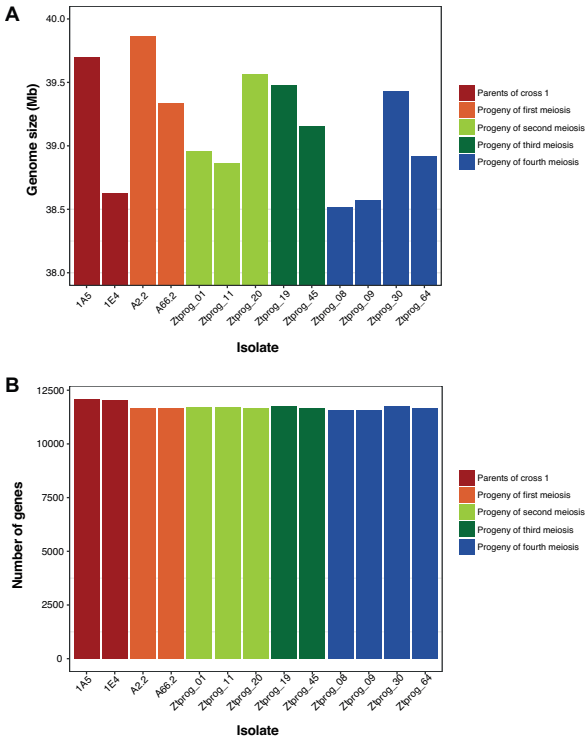
Tanaka H, Yao M-C. 2009. Palindromic gene amplification — an evolutionarily conserved role for DNA inverted repeats in the genome. Nat. Rev. Cancer 9:216–224.

Toledo F, Buttin G, Debatisse M. 1993. The origin of chromosome rearrangements at early stages of AMPD2 gene amplification in Chinese hamster cells. Curr. Biol. 3:255–264.

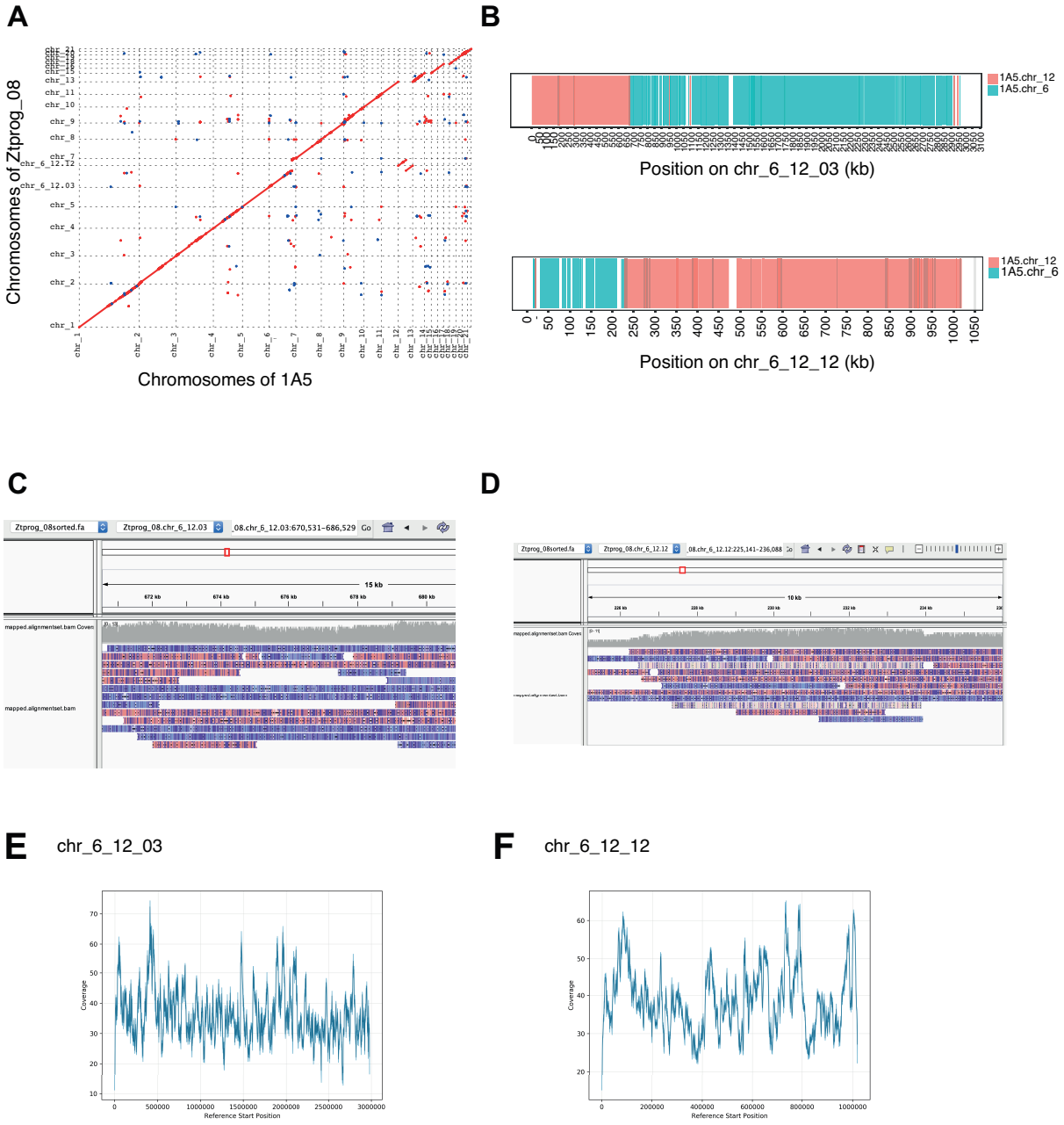
Supplementary Information

Supplementary Figures

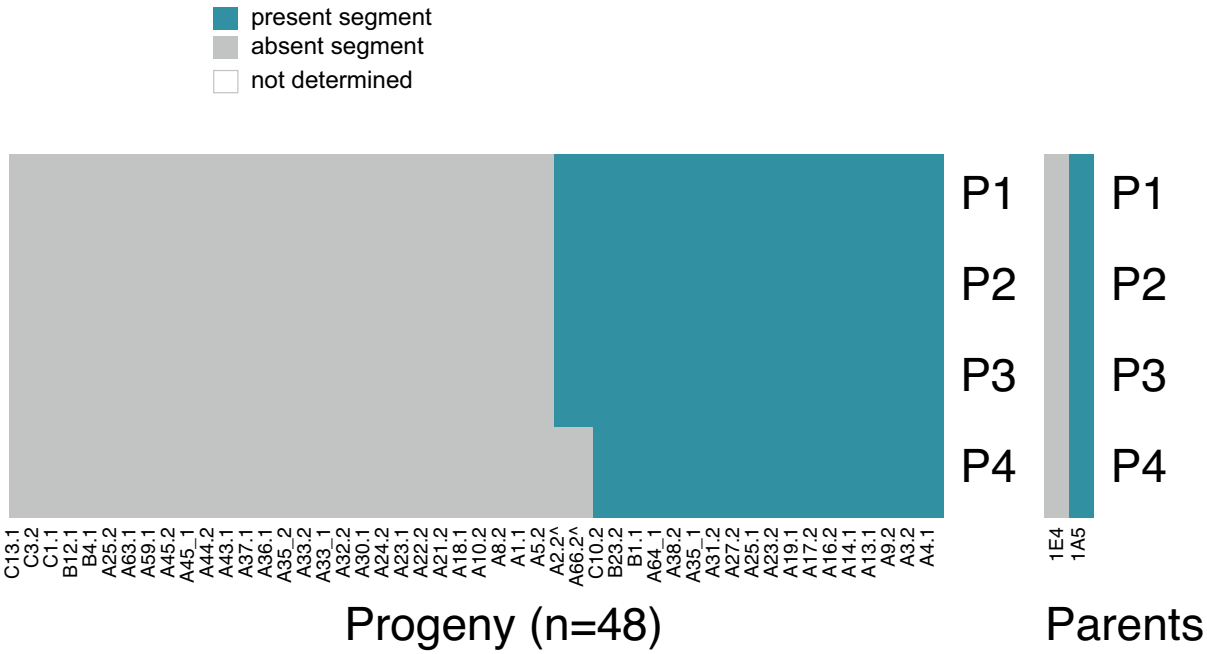
Supplementary Figures:



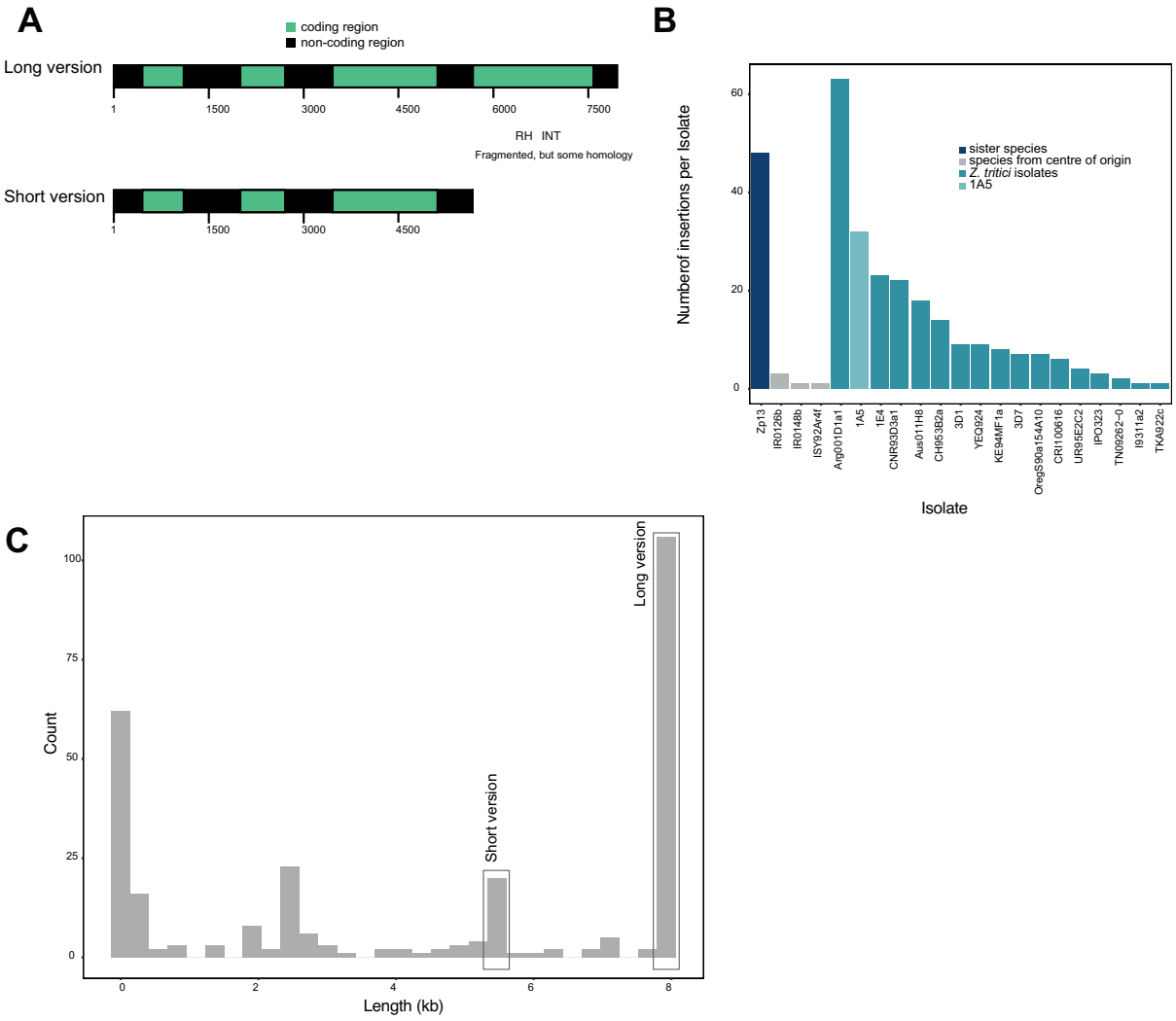
Supplementary Figure S1: Characteristics of the genomes of the parents and progeny of four rounds of meiosis. (A) Genome sizes of parents and progeny from four rounds of meiosis. (B) The number of genes in the parents and progeny from four rounds of meiosis.



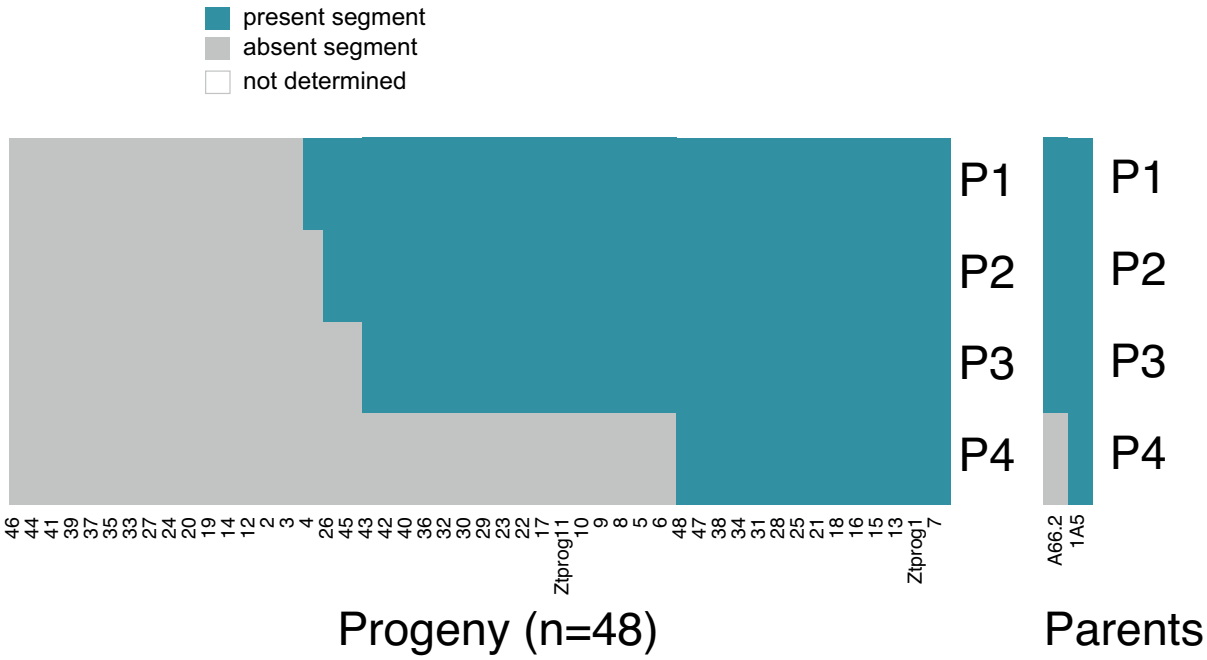
Supplementary Figure S2: Characterization of two reciprocally translocated chromosomes chr_6_12_03 and chr_6_12_12. A) A dotplot of Ztprog8 chromosomes mapped to the genome of 1A5. (B) The regions from chromosome 6 and 12 in the reference genome. (C) and (D) Reads mapping to the breakpoint locations connecting the chromosomes 6 and on chr_6_12.03 and chr_6_12.12, respectively. (D) and (E) The coverage along the new chromosome in 10 kb windows.



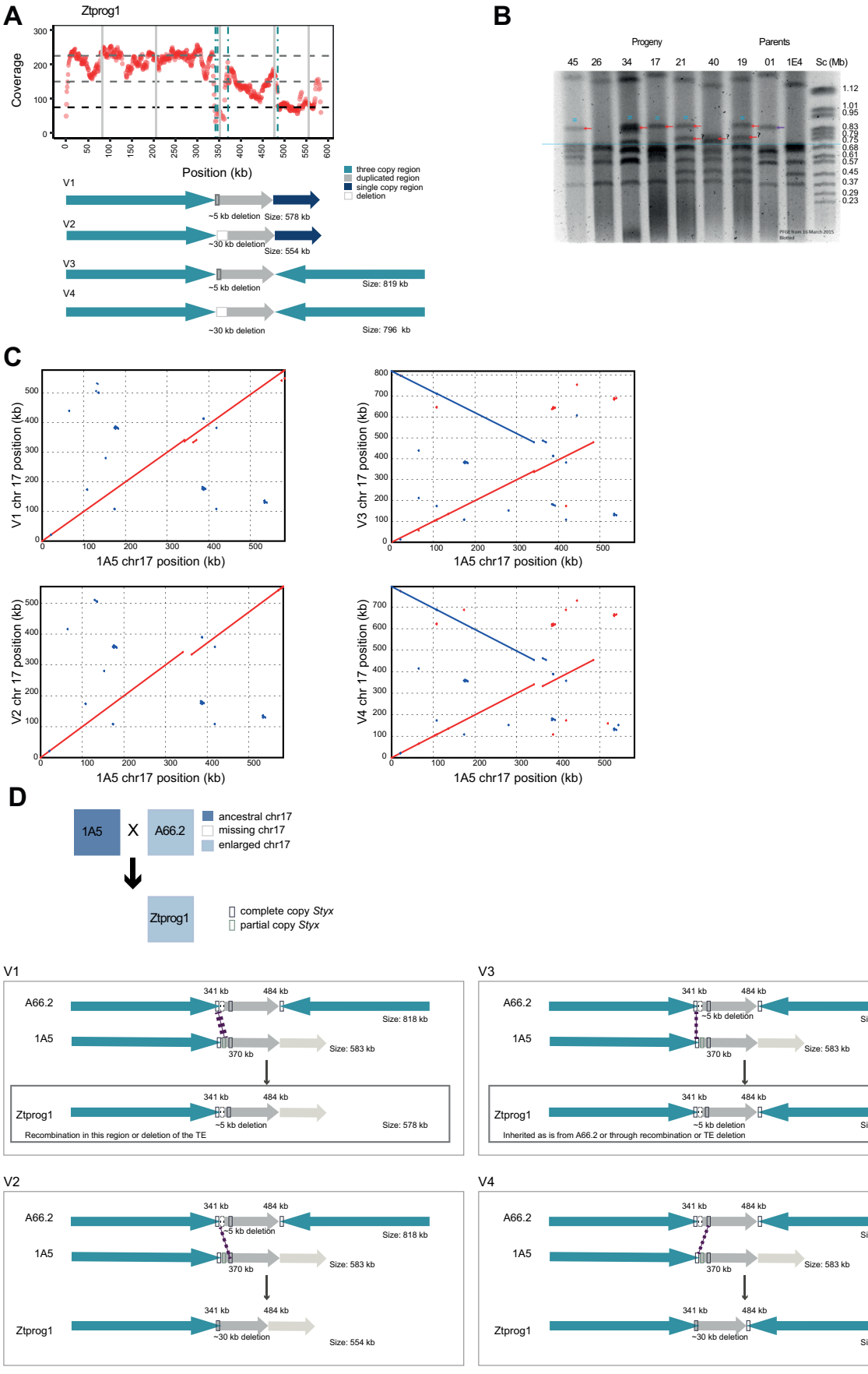
Supplementary Figure S3: Presence or absence of chromosomal segments assayed by PCR for 48 progeny of the cross 1A5 with 1E4. Turquoise represents present segments, and grey absent segments.



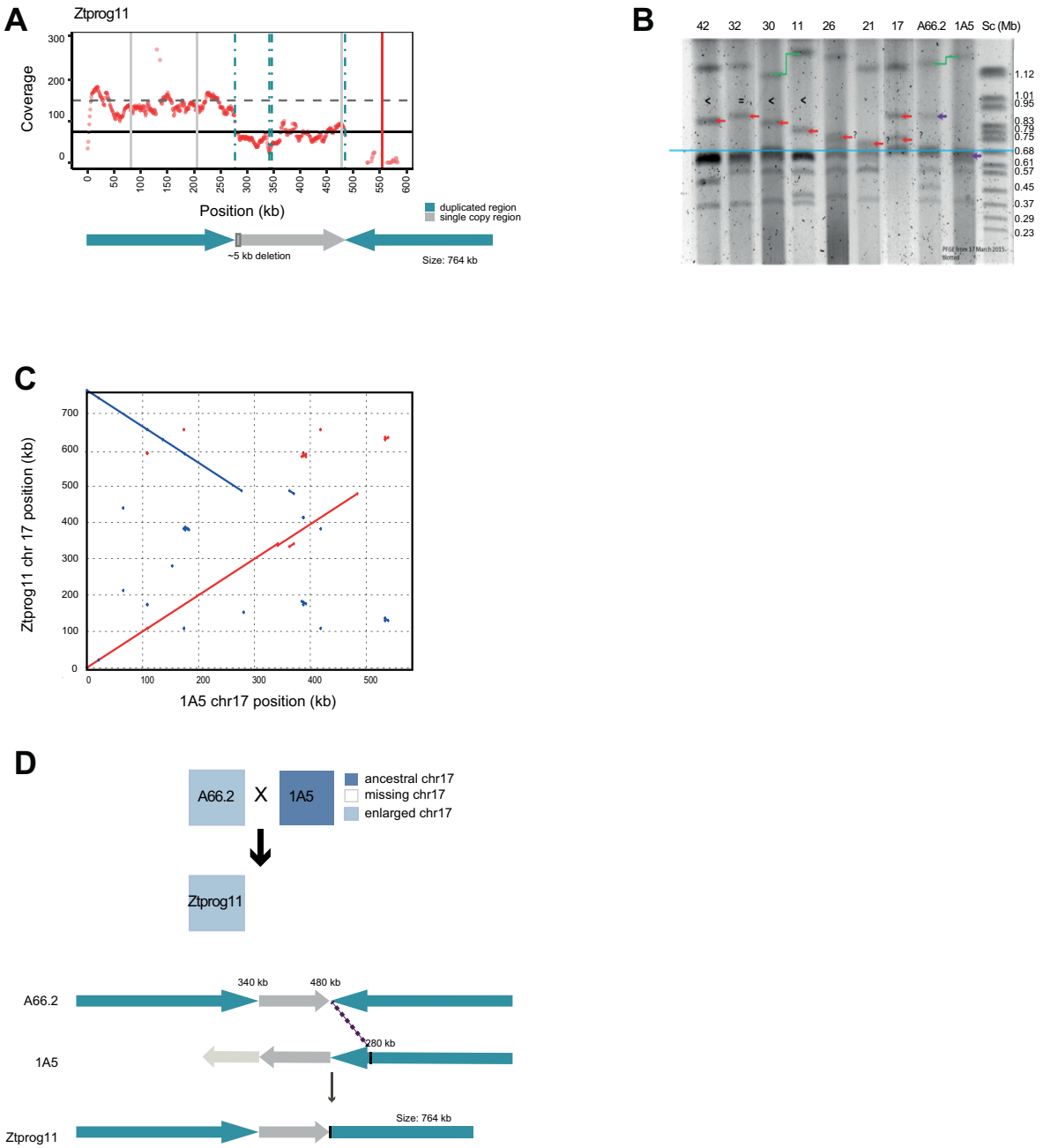
Supplementary Figure S4: Characterization of a TE of unknown family *Styx*. (A) The length and location of coding regions on the long and short copy of *Styx*. (B) Copy number of Styx in isolates from a global population of *Z. tritici*, including isolates from the center of original and an isolate of the sister species *Z. pseudotritici*. (C) The length distribution of copies of *Styx* in the same isolates from a global population of *Z. tritici*.



Supplementary Figure S5: Presence or absence of chromosomal segments assayed by PCR for 48 progeny of the cross A66.2 with 1A5. Turquoise represents present segments, and grey absent segments.

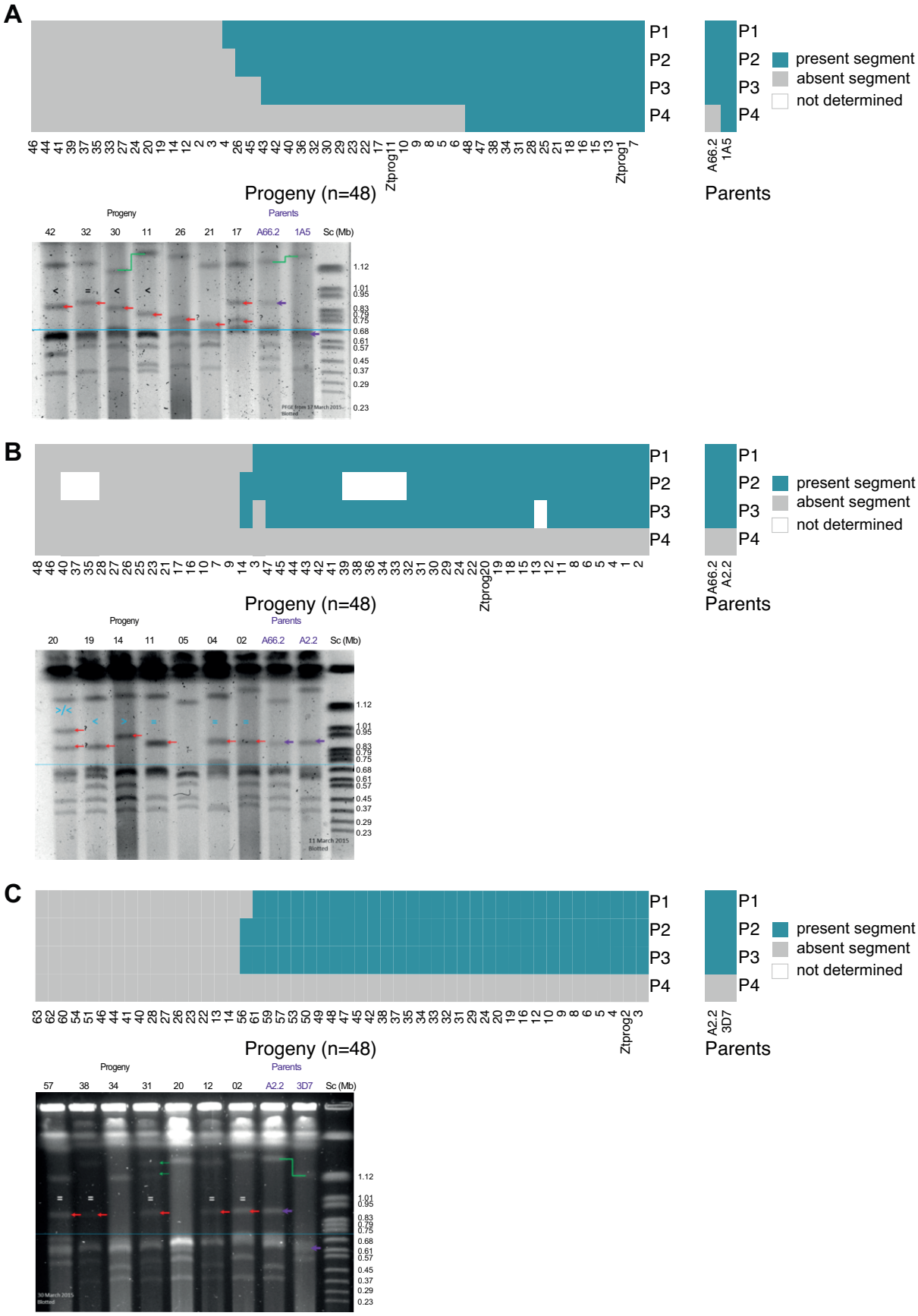

A
Ztprog1
Coverage
Position (kb)
three copy region
duplicated region
single copy region
deletion
V1
~5 kb deletion
Size: 578 kb
V2
~30 kb deletion
Size: 554 kb
V3
~5 kb deletion
Size: 819 kb
V4
~30 kb deletion
Size: 796 kb
B
Progeny
Parents
45 26 34 17 21 40 19 01 1E4 Sc (Mb)
1.12
1.01
0.95
0.83
0.79
0.75
0.68
0.61
0.57
0.45
0.37
0.29
0.23
C
V1 chr 17 position (kb)
1A5 chr17 position (kb)
V3 chr 17 position (kb)
1A5 chr17 position (kb)
V2 chr 17 position (kb)
1A5 chr17 position (kb)
V4 chr 17 position (kb)
1A5 chr17 position (kb)
D
1A5
X
A66.2
ancestral chr17
missing chr17
enlarged chr17
Ztprog1
complete copy Styx
partial copy Styx
V1
A66.2
1A5
Ztprog1
341 kb
484 kb
370 kb
~5 kb deletion
Size: 818 kb
Size: 583 kb
Size: 578 kb
Recombination in this region or deletion of the TE
V3
Size: 819 kb
Inherited as is from A66.2 or through recombination or TE deletion
V2
~30 kb deletion
Size: 554 kb
V4
Size: 796 kb

Supplementary Figure S6: Characterization of the chromosome 17 variants in Ztprog1. (A) The coverage and breakpoints of the progeny Ztprog1 reads mapped to the parent 1A5, horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two and three times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog1 showing the enlarged band (~0.83Mb) and a smaller band at (~0.57 Mb). (C) Dotplots of the assembled chromosome variants (V1-V4) for progeny Ztprog1 compared to the parent (1A5) chromosome. Inverted regions are in blue. (D) Schematic representation of the breakpoints and rearrangements between chromosome 17 in 1A5 and A66.2. Copies of Styx are not to scale.

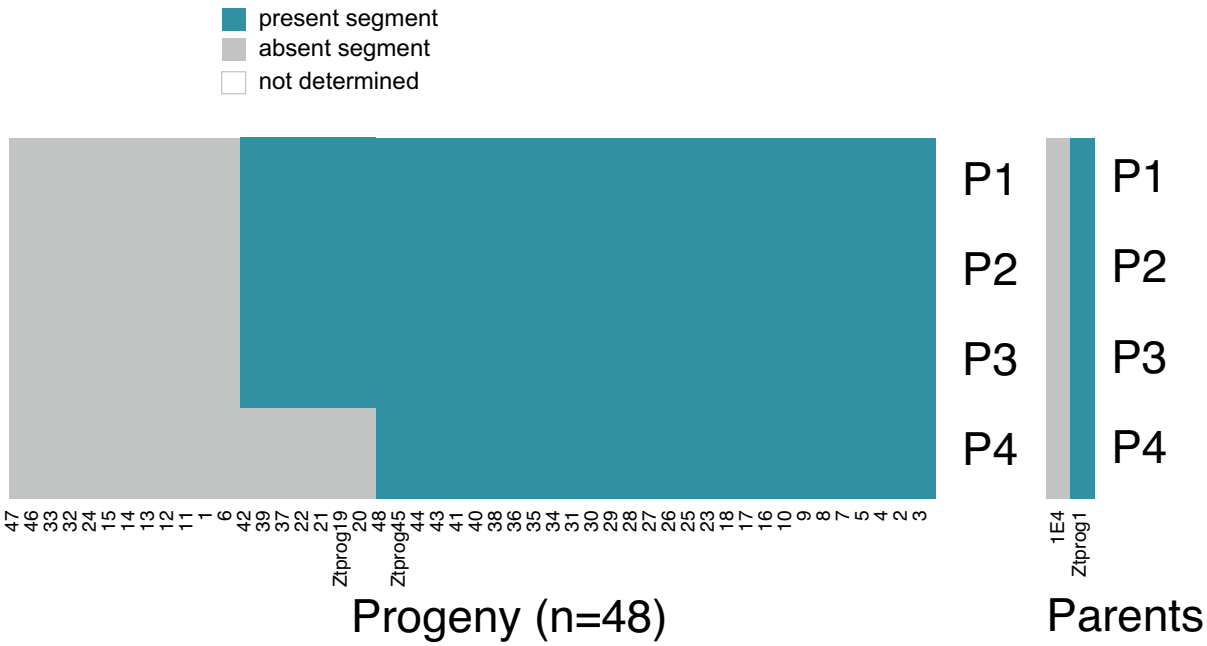


Supplementary Figure S7: Characterization of the chromosome 17 in Ztprog11. (A) The coverage and breakpoints of the progeny Ztprog11 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog11 showing a smaller chromosome than inn A66.2. (C) Dotplots of the assembled chromosome for progeny Ztprog11 compared to the parent (1A5) chromosome. Inverted regions are indicated in blue. (D)

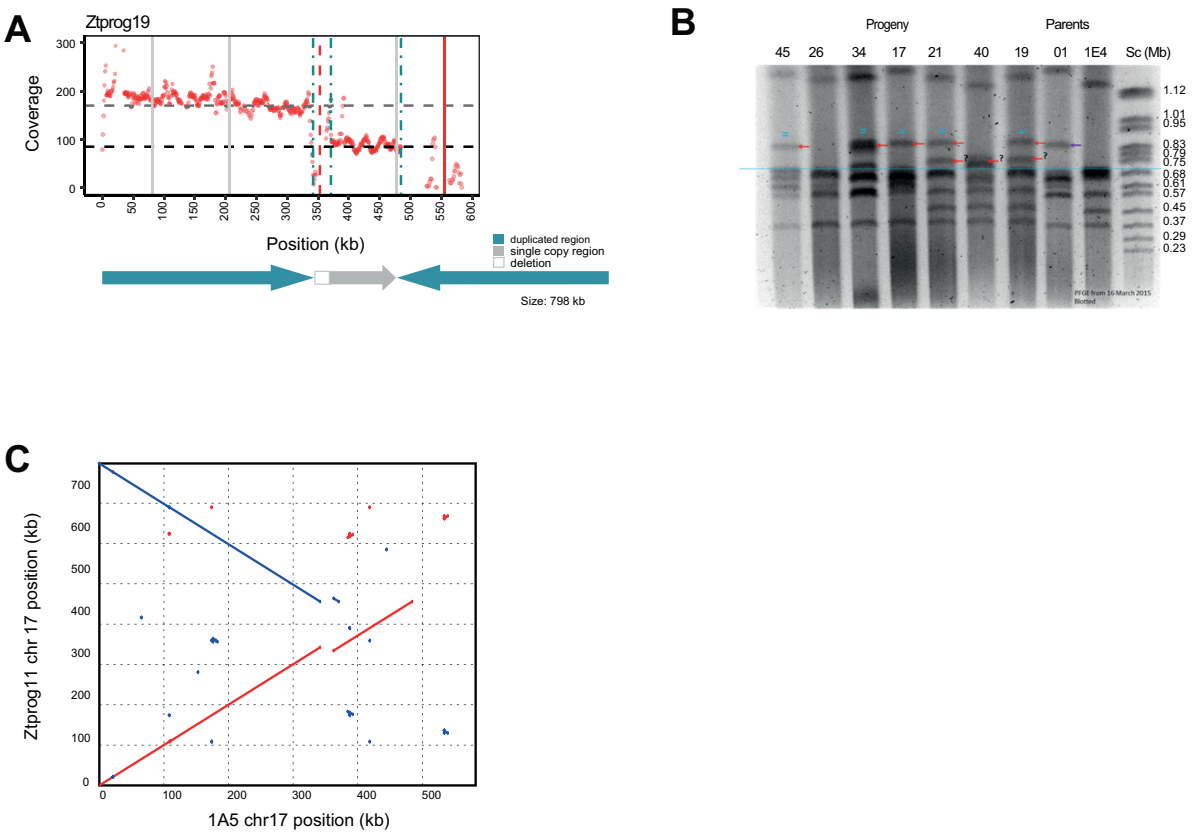
Schematic representation of the breakpoints and rearrangements between chromosome 17 in 1A5 and A66.2.


A
P1
P2
P3
P4
Ztprog11
Ztprog1
Progeny (n=48)
present segment
absent segment
not determined
A66.2
1A5
Parents
Progeny
Parents
Sc (Mb)
B
Ztprog20
A66.2
A2.2
C
Ztprog2
3D7

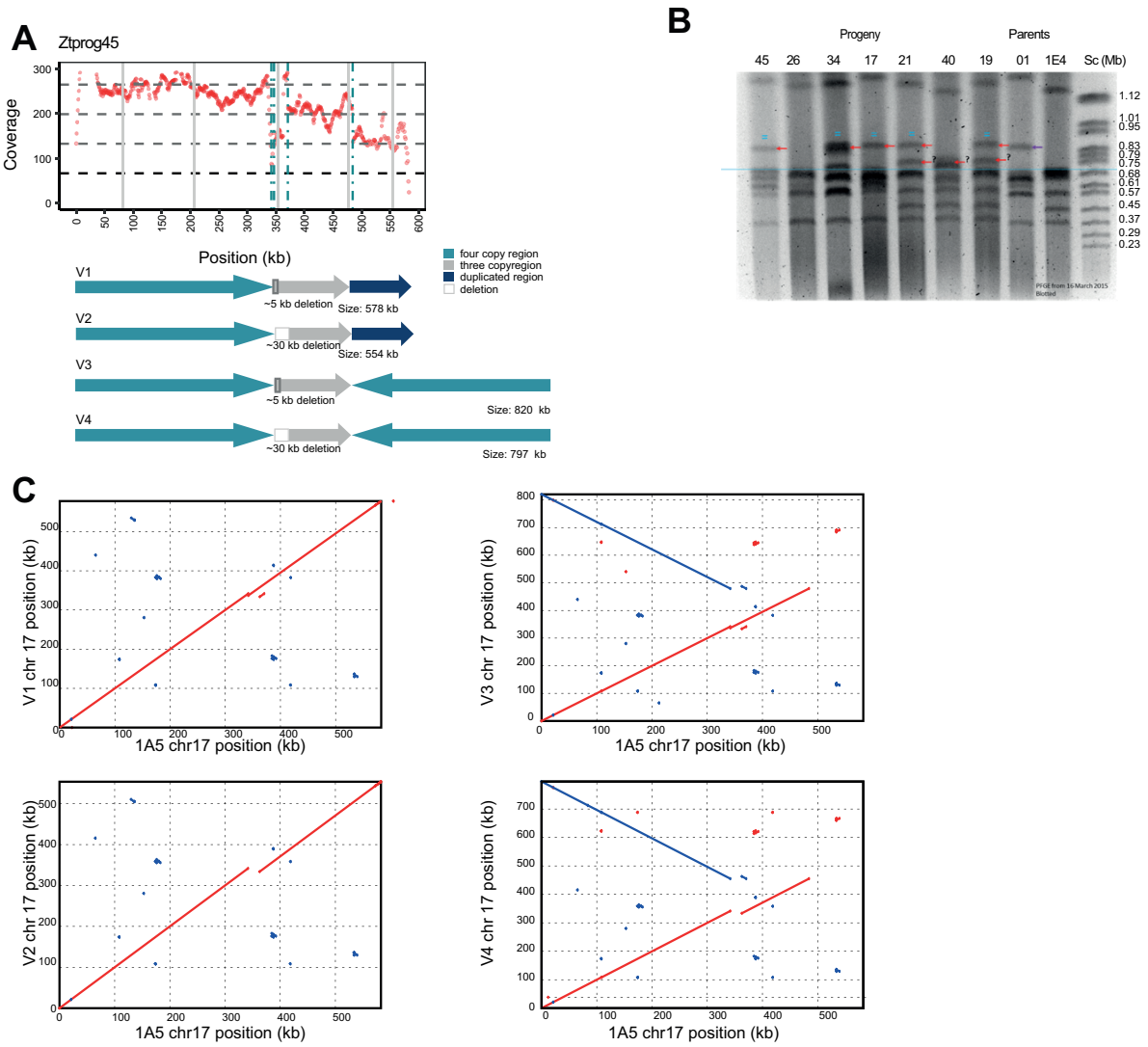
Supplementary Figure S8: Chromosome 17 variant diversity as assayed by PCR and karyotype for three crosses from the second round of meiosis. (A) Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross A66.2 with 1A5 and the karyotype of progeny (including Ztprog11) from this cross. (B) Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross A66.2 with A2.2 and the karyotype of progeny from this cross. (C) Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross A2.2 with and the karyotype of progeny (including Ztprog2) from this cross. Turquoise represents present segments, and grey absent segments.



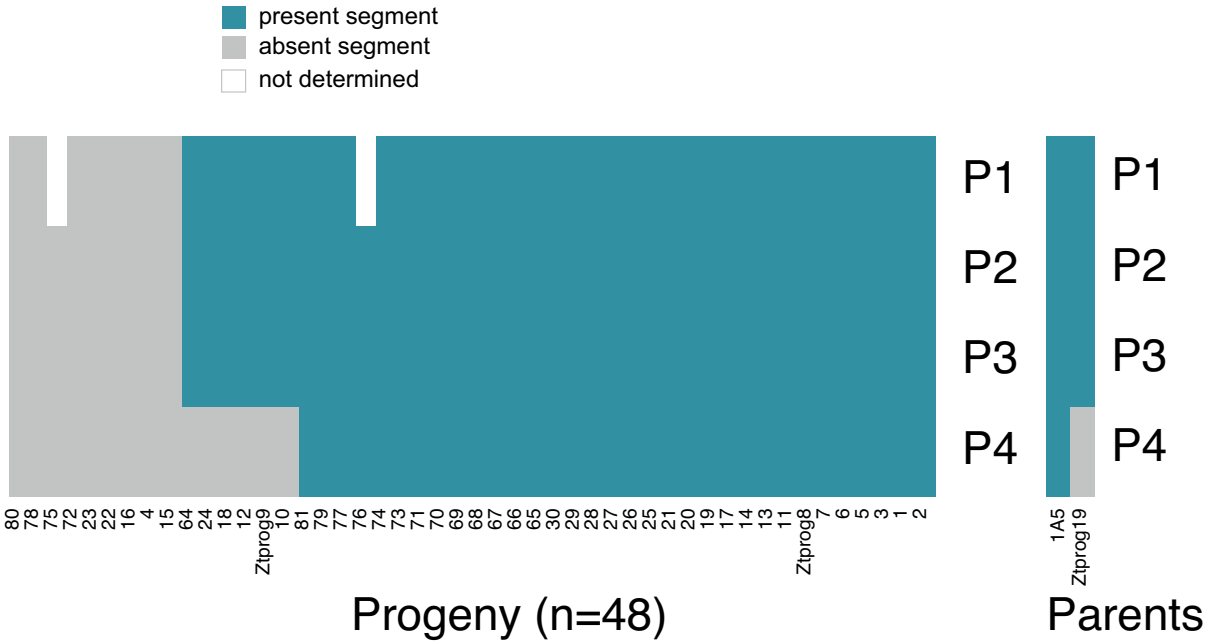
Supplementary Figure S9: Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross Ztprog1 with 1E4. Turquoise represents present segments, and grey absent segments.



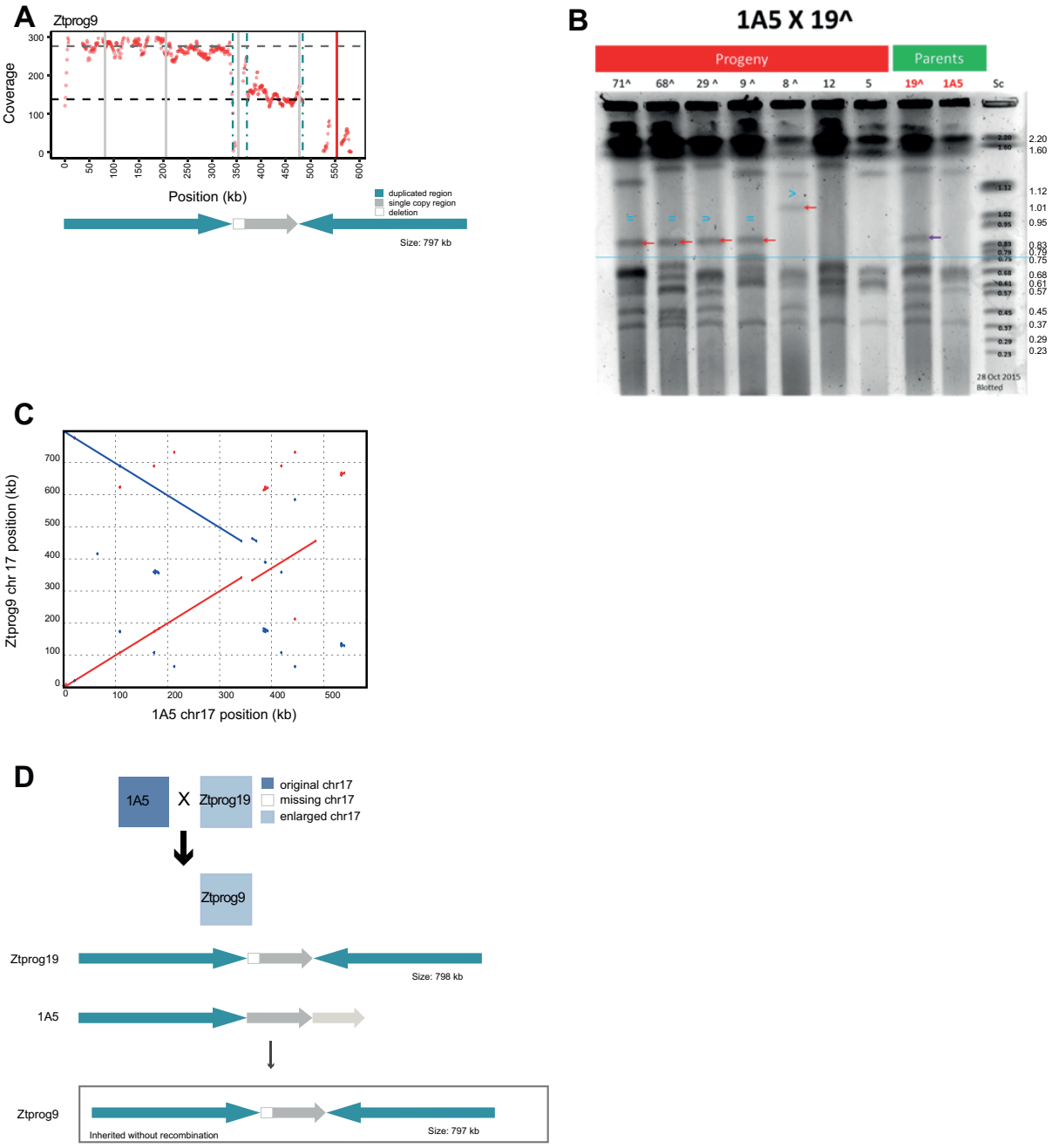
Supplementary Figure S10: Characterization of the chromosome 17 in Ztprog19. (A) The coverage and breakpoints of the progeny Ztprog19 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog19 showing the enlarged band that is similar to the chromosome in the parent Ztprog1. (C) Dotplots of the assembled chromosome for progeny Ztprog19 compared to the parent (1A5) chromosome.



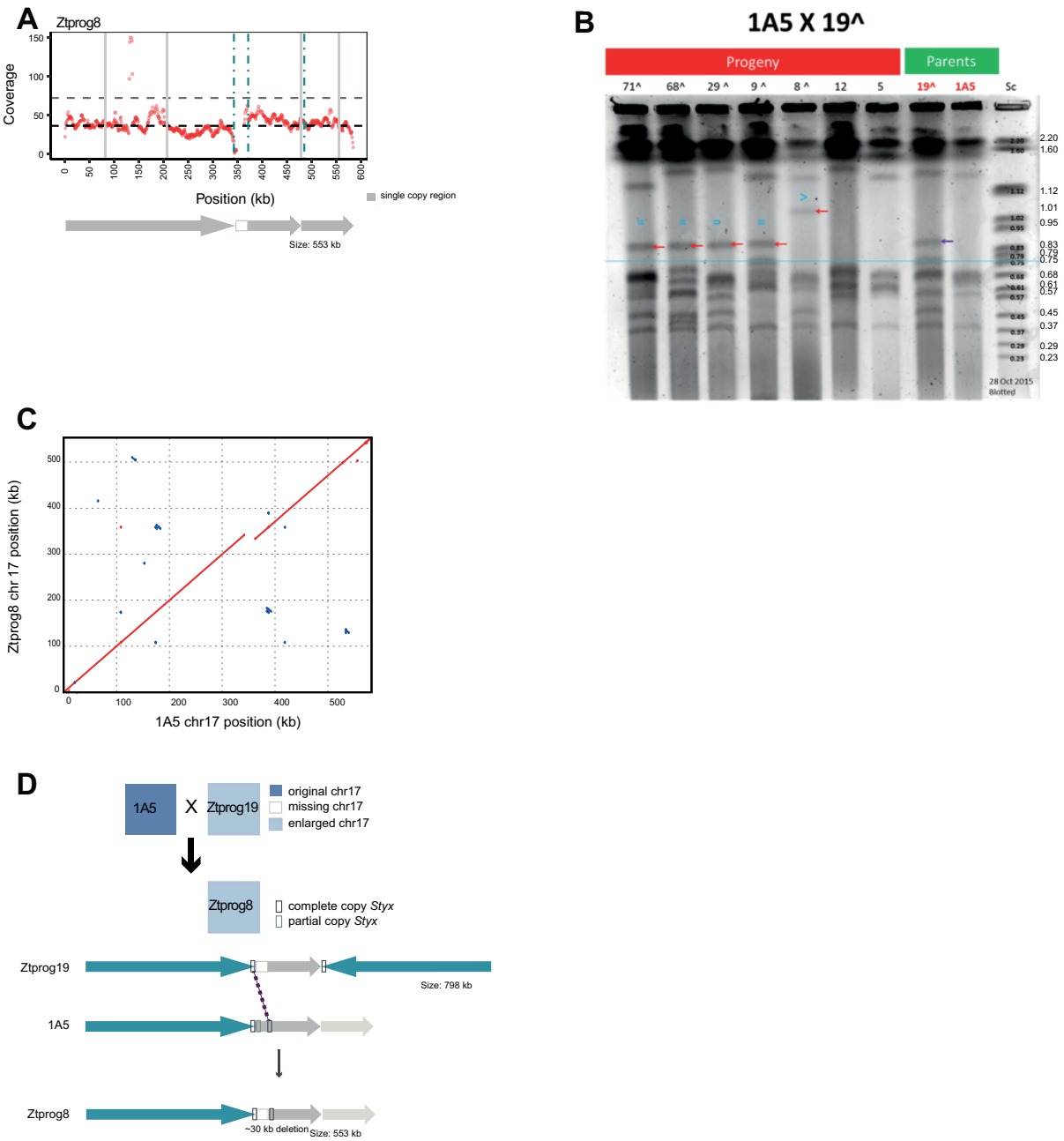
Supplementary Figure S11: Characterization of the chromosome 17 variants in Ztprog45. (A) The coverage and breakpoints of the progeny Ztprog45 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two, three and four times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog45 showing the enlarged band (~0.83Mb) and a smaller band at (~0.57 Mb). (C) Dotplots of the assembled chromosome variants (V1-V4) for progeny Ztprog45 compared to the parent (1A5) chromosome.



Supplementary Figure S12: Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross Ztprog19 with 1A5. Turquoise represents present segments, and grey absent segments.

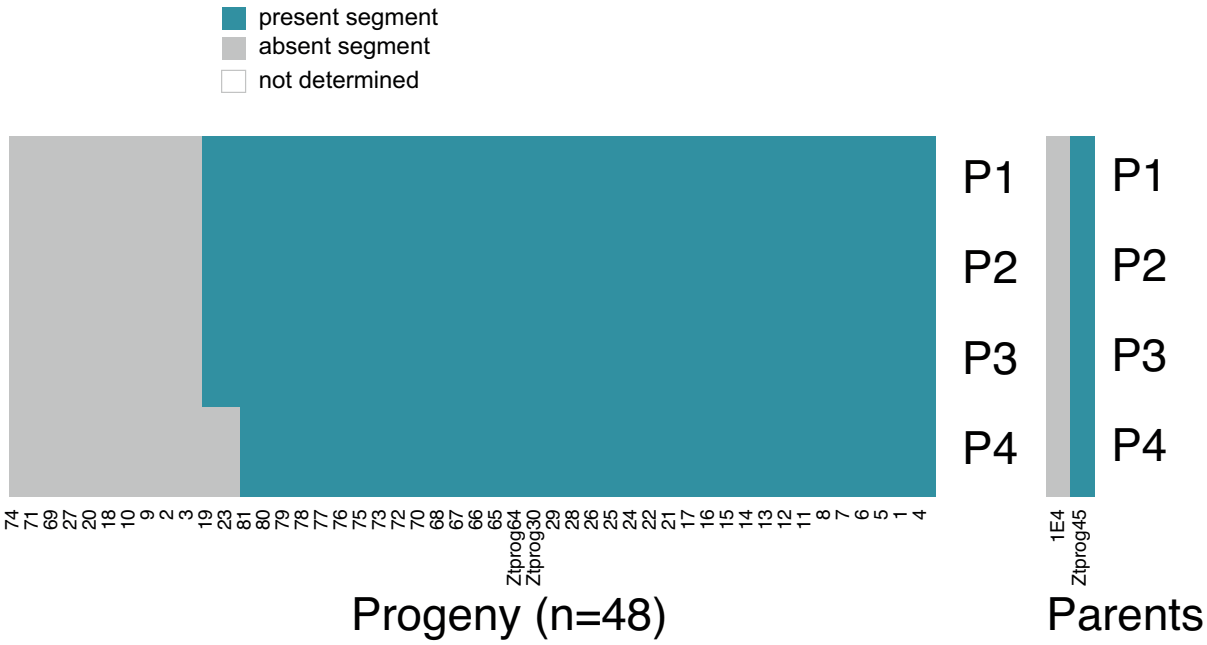


Supplementary Figure S13: Characterization of the chromosome 17 in Ztprog9. (A) The coverage and breakpoints of the progeny Ztprog9 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog9 showing the enlarged band that is similar to the band in Ztprog19. (C) Dotplots of the assembled chromosome for progeny Ztprog9 compared to the parent (1A5) chromosome. (D) Schematic representation of the events resulting in the chromosome that is most likely present in Ztprog9.

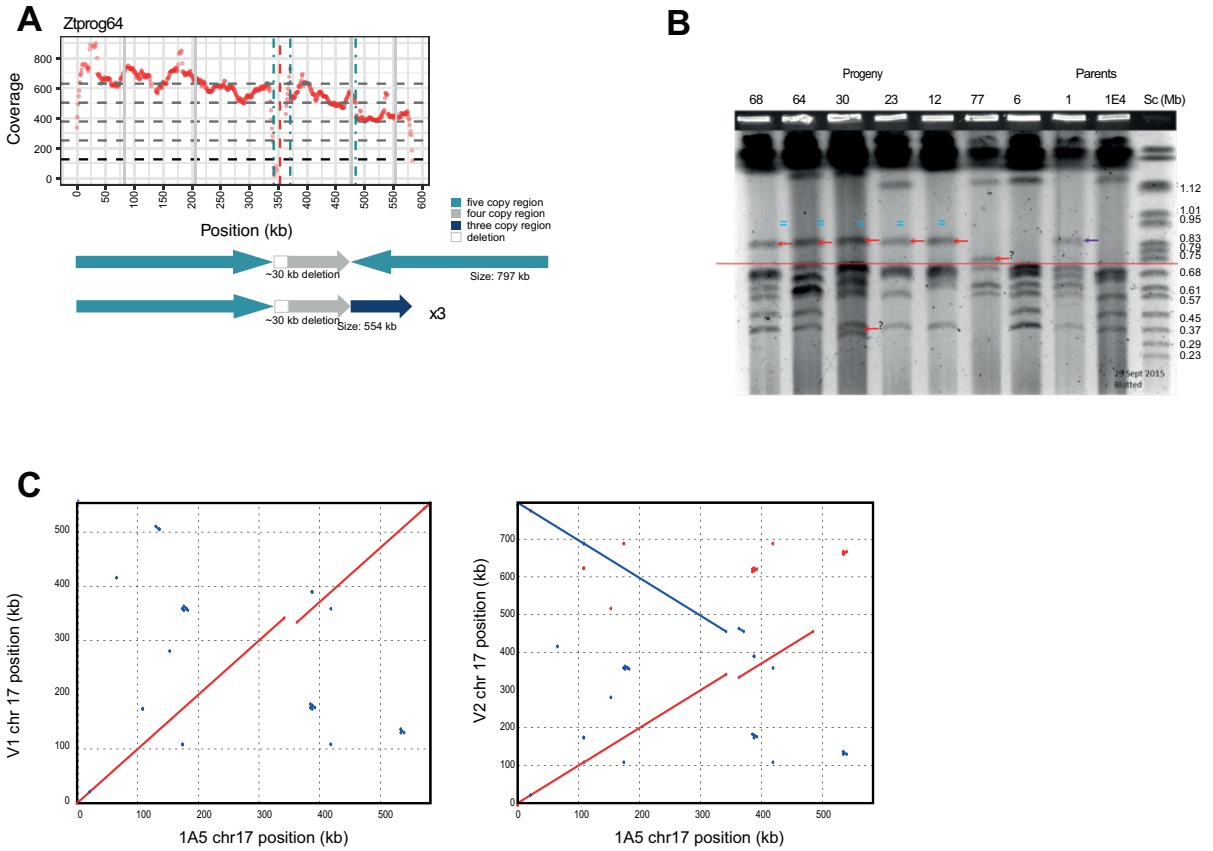


Supplementary Figure S14: Characterization of the chromosome 17 in Ztprog8. (A) The coverage and breakpoints of the progeny Ztprog8 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >300 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog8. (C) Dotplots of the assembled chromosome for progeny Ztprog8 compared to the parent (1A5) chromosome. (D)

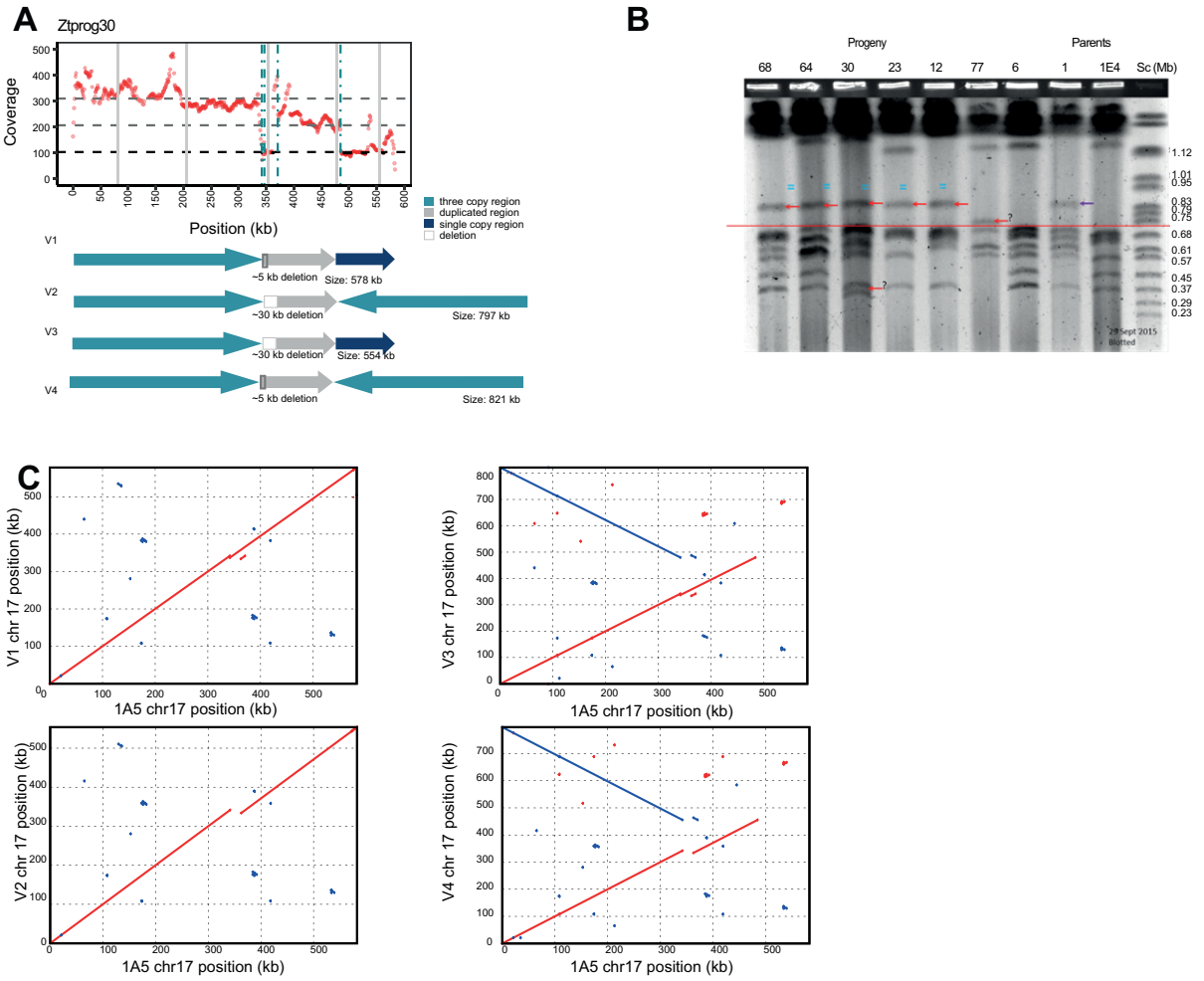
Schematic representation of the events resulting in the chromosome that is most likely present in Ztprog8.



Supplementary Figure S15: Absence or presence of chromosomal segments assayed by PCR for 48 progeny of the cross Ztprog45 with 1E4. Turquoise represents present segments, and grey absent segments.



Supplementary Figure S16: Characterization of the chromosome 17 variants in Ztprog64.. (A) The coverage and breakpoints of the progeny Ztprog64 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two, three, four and times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >900 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog64 showing a similar size chromosome as in the parent Ztprog1. (C) Dotplots of the assembled chromosome variants (V1-V2) for progeny Ztprog64 compared to the parent (1A5) chromosome.



Supplementary Figure S17: Characterization of the chromosome 17 variants in Ztprog30. (A) The coverage and breakpoints of the progeny Ztprog30 reads mapped to the parent 1A5. Horizontal dashed lines indicate the mean coverage of the core chromosomes (black) and in grey two and three times the mean core chromosome coverage. Red spots indicate the mean coverage in 1 kb windows (regions with excessive, >500 x coverage have been removed). Vertical dashed lines indicate chromosomal breakpoints identified from mapped reads. Vertical solid lines indicate the positions of loci amplified by PCR (grey=positive and red=negative). Below each coverage plot: variants of chromosome 17 that are likely to be in this progeny isolate. (B) Pulsed-field gel electrophoresis of the Ztprog30 showing the enlarged band (~0.83Mb) and a smaller band at (~0.57 Mb). (C) Dotplots of the assembled chromosome variants (V1-V4) for progeny Ztprog30 compared to the parent (1A5) chromosome.

Supplementary Tables

Supplementary Table S1: Genome assembly and annotation statistics of the parents and progeny isolates of four rounds of meiosis

	1A5	1E4	A2.2	A66.2	Ztprog1	Ztprog11	Ztprog20	Ztprog19	Ztprog45	Ztprog8	Ztprog9	Ztprog30	Ztprog64
Genomes Assembly													
Genome Size (Mb)	39.7	38.6	39.9	39.3	39	38.9	39.6	39.5	39.2	38.5	38.6	39.4	38.9
Chromosomes, n	21	20	20	20	20	20	20	20	19	19	19	20	19
GC %	52.2	52.26	52.17	52.26	52.24	52.24	52.21	52.25	52.17	52.31	52.29	52.19	52.18
Seed cut-off default (bp)					12328	14736	11439	15976	19323	12818	26861	24174	26446
Seed cut-offs used (bp)			12000	12000	10000	10000	10000	10000	15000	10000	20000	20000	20000
						8000	15000						15000
							8000						
Coverage Analysis													

Disomic Chromosomes	0	0	0	0	0	0	0	0	0	0	0	0	0
Large rearrangements										chr6_12			
Genome Annotation													
Genes, n	12092	12033	11666	11675	11701	11709	11671	11749	11646	11556	11590	11737	11644
Average gene length	1520.2	1524.2	1558.1	1559.4	1548.3	1549.9	1566.1	1547.9	1556.1	1565.9	1567.6	1547.9	1557.5
Average protein length	467.5	468.3	467.9	468	465.3	466.5	469.5	466.1	468	469.1	470.3	465.4	468.4
Number of exons	29716	30015	29101	29039	29023	29017	29273	29144	28933	28988	29027	29226	29225
Average exons per gene	2.46	2.49	2.49	2.49	2.48	2.48	2.51	2.48	2.48	2.51	2.5	2.49	2.51
Average exon length	570.7	563.3	562.9	564.6	563	564.9	561.7	564	565.3	561.2	563.6	561	560.1
Number of introns	17628	17984	17435	17364	17322	17309	17603	17395	17287	17432	17438	17490	17581
Average introns per gene	2.18	2.21	1.49	1.48	1.48	1.48	1.51	1.48	1.48	1.51	1.5	1.49	1.51
Average intron length	80.7	79.7	101	102.2	100.6	99.5	102.2	98.6	100.1	102.9	101.8	99.4	98.5

%TE per Genome	18.68%	17.68%	18.93%	18.06%	18.57%	18.25%	18.28%	18.24%	19.06%	18.06%	17.94%	18.94%	18.47%
** Ztprog2 was not included													

Supplementary Table S2: Split reads and alternative mapping positions used in assembly of chromosome 17

Isolate	Breakpoint identified	Mapping position of split read	Other mapping position of split read	Exact breakpoint	Exact connections	Reads going through	How many reads have a breakpoint here	Size after quiver polishing	Read ID (example) with breakpoint
A2.2	341804	332897-341804 #left to right	479016-484342 #right to left	341804	484342	108	31	818486	m150821_022619_42164_c100867842550000001823193603031662_s1_p0/120028/4603_26526
		325.720-341.799 #left to right	341.804-352.906 #left to right)	341799	346524	105			m150821_110414_42164_c100867842550000001823193603031664_s1_p0/108771/0_21362
A66.2	341803	332869-	476736-	341804	484342	61	23		m150808_084318_42243_c100868202550000001823193603031670_s

		341803 #left to right	484344 ###right to left						1_p0/118649/1693_19710
		331701 - 341804 #left to right	341807-352715 #left to right	341799	346524	64		818482	m150821_152350_42164_c100867842550000001823193603031665_s1_p0/148835/0_17377
Ztprog_01	341803	341807-348585 #left to right	333429-341804 #left to right	341799	346524	65	24	795814	m54073_170426_151000/74056042/0_10347
		325719-341803 #left to right	483163-484341 #right to left	341804	484342	52		578009	m54073_170426_151000/42861198/10371_28130
		328088-341804 #	362762-385840	341803	370700	40		55369	m54073_170426_151000/46006477/0_15306

		right to left	#right to left					8	
								819455	
Ztprog_11	341803	322483-341803 #left to right	341809-350955 #left to right	341799	346524	56	14	763841	m54073_170427_010916/17891446/0_24390
	277129	264719-277129 # right to left	472871-484340 #left to right so an inversion	277129	484342	55	31		m54073_170427_010916/19792181/0_32967
Ztprog_19	341803	330691-341802 #left to	476069-484330 #right to	341803	484342	48	21	798451	m54073_180615_120826/68289362/0_19201

		right	left						
		328677-341796 #left to right	362762-377377 #left to right	341803	370700	56			m54073_180615_120826/20841441/566_23906
Ztprog_45	341803	312922-341799 #right to left	362771-378180 #right to left	341803	370700	150	66	577931	m54073_180616_085253/28443343/0_36904
		315039-341803 # right to left	341804-350902 # right to left	341799	346524	145		553766	m54073_180616_085253/36176564/0_31046
		329338-341790 #left to right	469504-484342 # right to left	341804	484342	75		796590	m54073_180616_085253/41157509/0_28044

								820341	
Ztprog_08	341804	321674-341803 #left to right	362766-385238 #left to right	341803	370700	47	7	553084	m54073_180614_153141/42926737/197_34634
Ztprog_09	341804	316445-341803 # right to left	362767-380708 #right to left	341803	370700	144	63	797012	m54073_180615_014737/60031844/0_35555
		319612-341803 #right to left	464892-484343 #left to right	341804	484342	138			m54073_180615_014737/74318441/0_41397
Ztprog_30	341804	305097-341801	362762-376451	341803	370700	178	60	577974	m54073_180615_222853/13959818/0_42599

		#right to left	#right to left						
		314844-341803 #left to right	478253-484341 #right to left	341804	484342	105		553774	m54073_180615_222853/23724931/29601_64768
		316612-341803 #right to left	341803-351881 #right to left	341799	346524	94		821137	m54073_180615_222853/71631068/0_31138
								797435	
Ztprog_64	341804	303179-341803 #left to right	362766-380087 #left to right	341803	370700	564	143		m54073_180616_190851/19399517/0_48574

		313705-341801 #left to right	469753-484343 #right to left	341804	484342	148			m54073_180616_190851/67633556/0_42735

Supplementary Table S3: Binding positions of primers used in the segment-specific PCR

Primer	Gene in 1A5	Start position of gene/start of best blast hit	Name in Figure 1
97793	1A5.g11754	82047	P1
87809	1A5.g11764	206048	P2
97838	No gene	477232	P3
111767	No gene	554418	P4

Supplementary Table S4: Primer sequences used for the Southern hybridization probes specific for chromosome 17

Chromosome	Probe	Gene ID (IPO323)	Forward Primer (5'-3')	Reverse Primer (5'-3')
chr 17	probe 4	97838	CCA ATC CCA AGA AAA CCG	GAC CTT TTG TGA GCT TCT CAA GTA

Supplementary Table S5: Assessment of genome completeness using BUSCO. The total number of Ascomycota BUSCO groups searched was 1315

Isolate	Complete	Complete%	Complete and single copy	Complete and duplicated	Fragmented	Missing
A2.2	1297	98.70%	1296	1	4	14
A66.2	1296	98.50%	1295	1	4	15

Ztprog1	1297	98.70%	1296	1	4	14
Ztprog11	1297	98.70%	1296	1	3	15
Ztprog20	1293	98.40%	1292	1	5	17
Ztprog19	1294	98.40%	1293	1	5	16
Ztprog45	1297	98.60%	1297	0	5	13
Ztprog8	1297	98.70%	1296	1	3	15
Ztprog9	1297	98.70%	1296	1	4	14
Ztprog30	1296	98.60%	1296	0	5	14
Ztprog64	1296	98.60%	1296	0	6	13

Supplementary Table S6: Mean core chromosome coverage and the mean coverage of chromosome 17 parents and progenies

Isolate	Mean genome coverage chr 1-13	Mean coverage chr17
1A5	89.8814	85.620
A2.2	103.733	165.783
A66.2	103.132	154.037
Ztprog1	75.184	172.531
Ztprog11	75.081	105.519
Ztprog20	61.939	134.383
Ztprog19	85.064	158.996

Ztprog45	66.074	227.409
Ztprog8	35.984	38.9182
Ztprog9	138.063	237.788
Ztprog30	102.936	268.502
Ztprog64	125.878	589.906

CHAPTER 4:

The birth and death of effectors in rapidly evolving filamentous pathogen genomes

Simone Fouché[1], Clémence Plissonneau[1,2], Daniel Croll[3, *]

Accepted for publication in *Current Opinion in Microbiology* (February 2018), in press, doi: https://doi.org/10.1016/j.mib.2018.01.020

Abstract

Plant pathogenic fungi and oomycetes are major risks to food security due to their evolutionary success in overcoming plant defenses. Pathogens produce effectors to interfere with host defenses and metabolism. These effectors are often encoded in rapidly evolving compartments of the genome. We review how effector genes emerged and were lost in pathogen genomes drawing on the fascinating links between effector evolution and chromosomal rearrangements. Some new effectors entered pathogen genomes via horizontal transfer or introgression. However, new effector functions also arose through gene duplication or from previously non-coding sequences. The evolutionary success of an effector is tightly linked to its transcriptional regulation during host colonization. Some effectors converged on an epigenetic control of expression imposed by genomic defenses against transposable elements. Transposable elements were also drivers of effector diversification and loss that led to mosaics in effector presence-absence variation within species. Such effector mosaics within species was the foundation for rapid pathogen adaptation.

Introduction

Infectious diseases are one of the main threats to securely sustaining a growing world population (Fisher et al. 2012). Crops are frequently attacked by a vast range of fungal and oomycete pathogens that feed on plant tissues. Filamentous pathogens extract photosynthates and damage host tissues, which can severely stunt growth or kill crops if untreated. New lineages of pathogens arise with a worrisome frequency. Such lineages often evolved the ability to attack previously resistant crop varieties and are, hence, difficult to contain. The recent outbreak of stem rust lineage called Ug99 was triggered by a breakdown of a major wheat resistance gene and is now considered a global threat to wheat production (Singh et al. 2011). The pathogen causing wheat blast recently spread from its endemic distribution range in South America to South Asia and is devastating wheat production (Islam et al. 2016). Other pathogens such as *Leptosphaeria maculans* and *Zymoseptoria tritici,* infecting oilseed rape and wheat, respectively, continuously adapted to newly deployed control measures. Crop resistance and fungicides typically failed within a few years after deployment (Fudal et al. 2009; Daverdin et al. 2012; Estep et al. 2015).

Our understanding of emerging filamentous pathogens has been revolutionized by analyses of effectors. Effectors are often small secreted proteins that promote the colonization of the pathogen by manipulating the host cell, interfering or protecting the pathogen from host defenses. The mechanisms by which effectors promote host colonization can be wide-ranging. Effectors can alter the host cell's metabolism or hormone homeostasis, lead to necrosis or mask the presence of the pathogen (Cook et al. 2015). Some hosts evolved the ability to detect the presence of effectors and subsequently trigger strong defense reactions (*i.e.* effector triggered immunity). In this case, the literature often refers to these effectors as avirulence factors. Carrying an effector can have dramatically different effects for a pathogen depending on whether the pathogen encounters a host that is able to recognize the

particular effector. Therefore, effector genes often undergo rapid evolutionary change in pathogen populations (Möller & Stukenbrock 2017).

Effectors are often lineage-specific, have no conserved protein domains and may not even be shared among the most closely related pathogen species. Yet, effectors often share properties such as an abundance in cysteine residues that confer a compact structure to the protein and remain intact under stress conditions induced by the host (Saunders et al. 2012). Unrelated effectors can share structural similarities (Guillen et al. 2015) and many effectors show typical expression profiles with a peak during the establishment of the pathogen on or inside the host (Hacquard et al. 2013; Kleemann et al. 2012; Palma-Guerrero et al. 2016; Skibbe et al. 2010; Soyer et al. 2014; Wang et al. 2011). Analyses of pathogen genomes revealed that effectors are often located in repeat-rich genomic compartments (Suomeng Dong et al. 2015). A striking feature of these compartments is that they segregate substantial structural variations within species. Such polymorphism can accelerate the evolution of effectors by dramatically increasing the frequency of non-homologous recombination and, hence, effector gene duplications and deletions. Hence, these chromosomal regions provide unique niches for the evolution of effectors.

Here, we review how recent genome-scale analyses revolutionized our understanding of how effectors are gained and lost in pathogen genomes. We discuss the evolutionary life cycle of an effector gene, starting with its first emergence in a pathogen's gene pool and ending with its potential loss, and describe how chromosomal rearrangements are intimately linked to every step of the process. Finally, we suggest how causality in the emergence of effector genes and genome evolution can be disentangled in future research.

Mechanism of effector gene emergence in pathogens

The evolutionary age of an effector is reflected in its degree of conservation among species. LysM effectors are shared by most filamentous pathogens and prevent plant receptors from recognizing chitin in cell walls (Akcapinar et al. 2015; Sánchez-Vallet et al. 2015). Some shared effectors have evolved novel functions enabling pathogenicity on different hosts such as the protease inhibitor effectors of *Phytophthora infestans* and *P. mirabilis* (Dong et al. 2014). But many effectors are recent additions to a pathogen's gene pool.

Some pathogens gained effectors via horizontal transfer (Figure 1). The best studied case is the effector gene *ToxA* that was transferred from the wheat pathogen *Parastagnospora nodurum* to *Pyrenophora tritici-repentis* (Friesen et al. 2006). The acquisition of *ToxA* enabled *P. tritici-repentis* to gain virulence on wheat. *ToxA* has subsequently been identified in a third species, *Bipolaris sorokiniana* (McDonald et al. 2017). Horizontal acquisition of effectors such as *ToxA*, which acts as a host specific toxin, is particularly attractive for necrotrophic pathogens that proliferate by killing and feeding on host cells. In all species known to carry *ToxA*, the gene is found in a chromosomal region rich in repetitive transposable element (TE) sequences (Friesen et al. 2006; McDonald et al. 2017). The gene encoding the effector Ave1 of the vascular wilt pathogen *Verticillium dahliae* is likely to be of plant origin (de Jonge et al. 2012). *Ave1* is located in a chromosomal region that underwent large rearrangements and is heavily impacted by epigenetic silencing and RIP. RIP is a premeiotic mechanism that rapidly mutates copies of TEs and other near identical sequences. Leakage of RIP can introduce mutations into neighbouring genes. In the asexual pathogen *Fusarium oxysporum,* the horizontal transfer of an entire chromosome introduces host-specific effectors to recipient strains (Ma et al. 2010). Horizontal acquisition also made major contributions to the effector content of the *Phytophthora* and *Pythium* oomycete pathogens (Savory et al. 2015). Horizontally acquired effectors had profound effects on

pathogen lifestyles and evolutionary success, yet we often lack a mechanistic understanding of why effector genes were inserted into repeat-rich compartments.

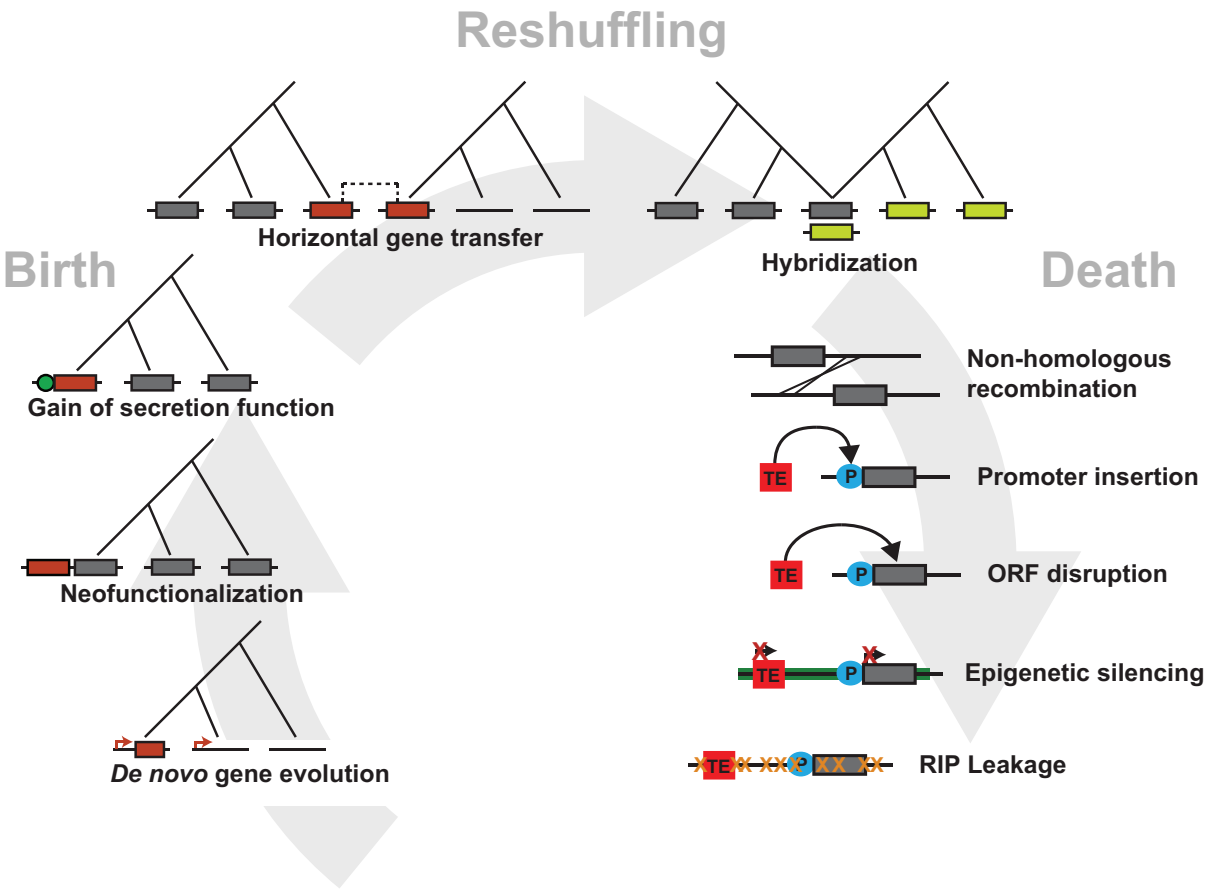


Figure 1: The life cycle of filamentous pathogen effector genes. From the bottom left clock-wise, effector genes could emerge in pathogen genomes *de novo* from non-coding sequences, from duplication and neofunctionalization or through the gain of a secretion signal. Effector genes can be reshuffled among pathogens via horizontal transfer or hybridization. Non-homologous recombination can lead to progeny lacking a copy of the effector. Effector genes can also be lost or inactivated by the insertion of a transposable element (TE) that disrupts the promoter (P) or open reading frame (ORF) sequence. Effector genes can also be inactivated through the leakage of repeat-induced point mutations or through epigenetic silencing triggered by the presence of TEs.

Hybridization is among the best-understood routes to introduce foreign genes into a pathogen's gene pool (fig. 1). Usually, a transient hybrid stage that is sexually compatible with the parent species is formed. Genetic material can then be introgressed from a donor to

a recipient species through repeated backcrossing (Baack & Rieseberg 2007; Stukenbrock 2016b). A fascinating example of pathogen hybridization was recently discovered in the powdery mildew of triticale, which is a hybrid between wheat and rye. Mirroring the host's hybridization, *Blumeria graminis* f. sp. *triticale* is a hybrid of the wheat powdery mildew *B. graminis* f. sp. *tritici* and the rye powdery mildew *B. graminis* f. sp *secalis* (Menardo et al. 2016). A sister species of the major wheat pathogen *Zymoseptoria tritici*, *Z. pseudotritici*, is a hybrid of *Z. tritici* and the closely related pathogen *Z. ardabiliae* found on wild grasses (Stukenbrock et al. 2012). Hybridization and introgression are major routes for pathogen adaptation because they create a mosaic of parental sequences that is optimally adapted to the new host species and environment (Stukenbrock 2016a).

Despite extensive genomic analyses, the evolutionary origins of many effector genes are unknown. Hence, such effector genes are described as orphan or taxonomically restricted genes (Plissonneau et al. 2017). The lack of orthologs in closely related species or the donor if the gene was acquired horizontally makes reconstructing the origins of an effector very challenging. Uncertainty about orthology can be caused by multiple factors. First, the rapid evolution of effector genes due to the selection pressure exerted by the host can quickly weaken evidence for orthology. The major effector *AvrStb6* in *Z. tritici* showed almost exclusively non-synonymous substitutions within populations (16 out of 18 SNPs for a gene of 189 bp) (Zhong et al. 2017). Second, effector genes tend to be lost at significantly higher rates than more conserved genes (fig. 2) (Hartmann & Croll 2017). Such gene loss can be in response to selection imposed by the host (Hartmann et al. 2017; Inoue et al. 2017). Third, effector genes are likely candidates for "young genes" that arose from recent gene duplication events or spontaneously from non-coding sequences in the recent evolutionary past (Plissonneau et al. 2017).

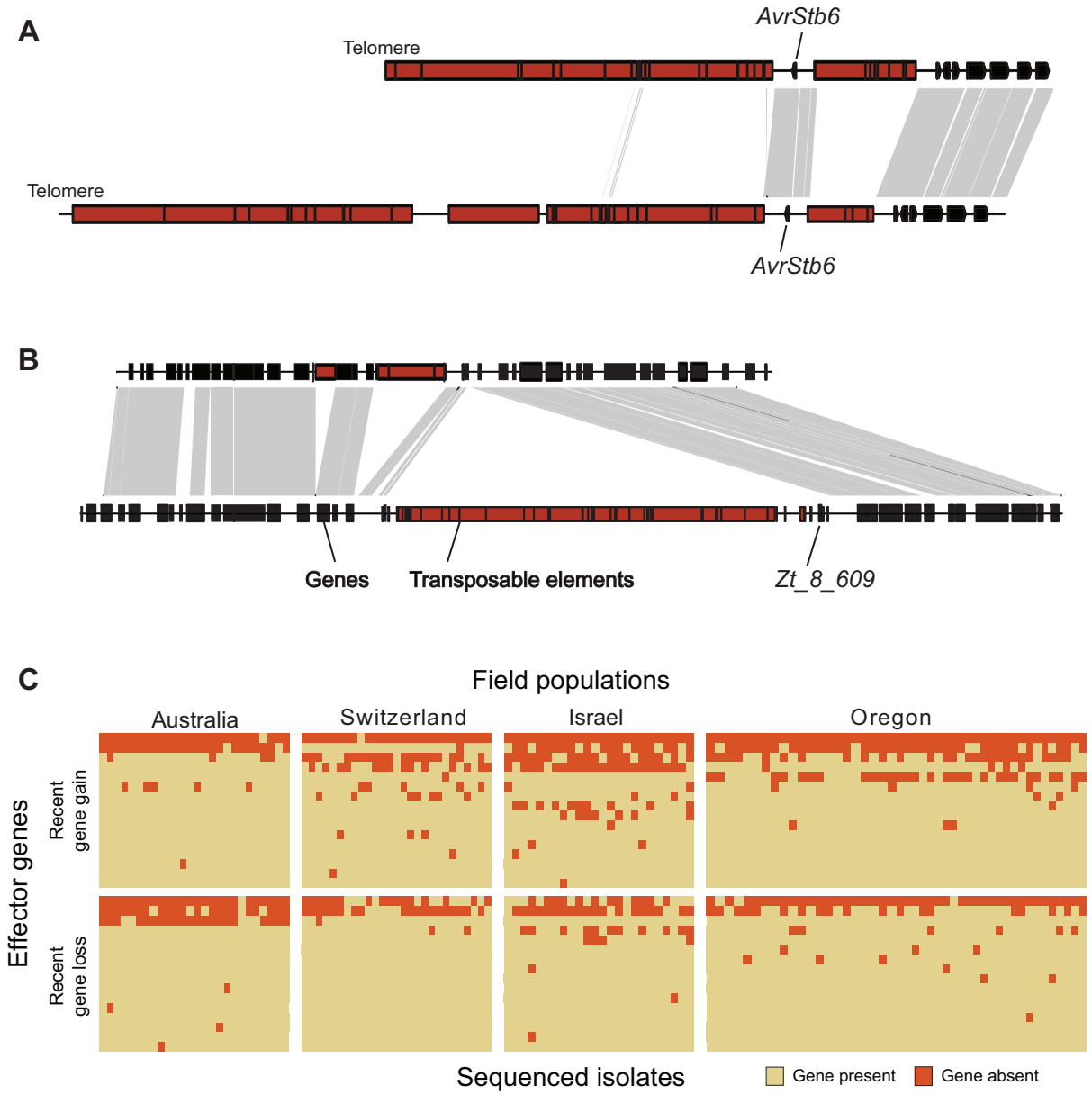


Figure 2: The association of effector genes and structural variation. The genomic environment of the two avirulence effectors (A) *AvrStb6* (Zhong et al. 2017) and (B) *Zt-8-609* (Hartmann et al. 2017) of the wheat pathogen *Zymoseptoria tritici*. Transposable elements are shown in red and genes are shown in black. Grey areas indicate syntenic regions between chromosomes. The subtelomeric *AvrStb6* is present in both isolates but is surrounded by transposable element clusters of highly variable length. *Zt-8-609* is only present in one of the two compared strains and was deleted by non-homologous rearrangements in the adjacent transposable element cluster. C) Mosaic in effector gene presence and absence among field populations of *Z. tritici.* Effector genes (different rows) are grouped according to whether the genes were recently gained (no ortholog in closely related species) or lost (the effector gene was found in closely related species). Isolates (different columns) were recovered from four different fields and analyzed for effector gene presence or absence using whole genome sequencing (Hartmann & Croll 2017).

The birth of new effector genes

In smut fungi, tandem duplications created paralogous effector genes within gene clusters (Dutheil et al. 2016). These gene duplications were favored by neighbouring TEs and the duplicates rapidly accumulated mutations leading to divergence. A *Z. tritici* effector evolved its function by gaining a secretion signal (fig. 1) (Poppe et al. 2015). Effector genes could also emerge *de novo* by fixing mutations that transform non-coding DNA into a functional sequence with an open reading frame (ORF) and a *cis*-regulatory element (fig.1) (McLysaght & Guerzoni 2015; McLysaght & Hurst 2016; Plissonneau et al. 2017). *De novo* gene evolution was once considered to be rare, but recent analyses showed that new genes are an important source of functional diversity among lineages (Carvunis et al. 2012). The analysis of pathogen genomes revealed astonishing numbers of candidate effector genes that resemble *bona fide* effector genes in their structure and regulation during infection (Dong et al. 2015, 2014; Hartmann et al. 2017; Rouxel et al. 2011; Zhong et al. 2017). Yet, functional studies of candidate effector genes often failed to detect an obvious role during the infection process (either due to effector redundancy or a true lack of function). The lack of a conserved protein domain suggests that these effector-like genes form a pool of young and largely non-functional genes. True, functional effectors could then readily evolve from this pool through additional mutations. It is important to note though that most functional studies were restricted to specific stages of the pathogen's life cycle and more comprehensive studies can reveal yet unknown effector functions (see e.g. Gervais et al., 2017).

Epigenetic regulation of effector gene expression

Effector genes are often the most highly expressed genes during infection (Hacquard et al. 2013; Wang et al. 2011). The upregulation at specific stages of the host invasion is carefully attuned to minimize the risk of detection by the host immune system while maximizing host exploitation (Kleemann et al. 2012; Palma-Guerrero et al. 2016; Soyer et al. 2014). The genomes of *L. maculans, Z. tritici* and *P. sojae* regulate the expression of effector epigenetically (Soyer et al. 2014; Schotanus et al. 2015; Qutob et al. 2013). Eukaryotes evolved the ability to prevent TE proliferation in the genome through epigenetic silencing of TE-rich regions (Chujo & Scott 2014; Feng et al. 2010; Zemach et al. 2010). The main mechanisms of silencing are the methylation of DNA or the methylation of histone lysine residues. Specific histone lysine methylations lead to active or repressive chromatin regions (Mikkelsen et al. 2007).

In a case of convergent evolution among plant pathogens, stress during the infection likely leads to changes in chromatin states and deprepression of silenced genes. For example, chromatin states governed by histone methylation are linked to the regulation of secondary metabolite clusters in *Epichloë festucae* (Chujo & Scott 2014), *F. graminearum* (Connolly et al. 2013) and *F. fujikuroi* (Studt et al. 2016). In *Z. tritici* (Schotanus et al. 2015) and *L. maculans* (Soyer et al. 2014) repeat-rich regions show distinct methylation marks that are repression the effector gene expression before host colonization. In *F. oxysporum* f. sp. *lycopersici*, the association of TEs and effector genes is even more direct. Miniature inverted-repeat transposable elements (MITE) are inserted in the promoter of most known effector genes (Schmidt et al. 2013). However, whether MITEs impact effector gene expression is unknown. The epigenetic control of effectors is a mechanism for pathogens to efficiently repress effector recognition by the host. In addition, novel effectors could be

recruited into the same regulatory framework simply by translocating the gene into a TE-rich region.

Generating the raw material for within-species effector evolution

Plants and pathogens are locked in arms races to detect invasion and to disable host resistance, respectively. Hence, the evolvability of effectors is a crucial component of a pathogen's success. One of the most fascinating discoveries about the organization of pathogen genomes is that effector genes are often over-represented in TE-rich, gene-sparse regions of the genome (Dong et al. 2015; Faino et al. 2016; Gibriel et al. 2016; Möller & Stukenbrock 2017; Raffaele & Kamoun 2012). The term “two-speed genome" was coined to conceptually distinguish these regions (Dong et al. 2015). Effector-rich regions can span entire chromosomes such as the accessory chromosomes in *F. oxysporum* lineages (Ma et al. 2010), AT isochores as in *L. maculans* (Rouxel et al. 2011) or subtelomeric regions in *Magnaporthe oryzae* (Chuma et al. 2011). The accelerated evolutionary rates in effector-rich genomic compartments stem from a combination of factors. As an evolutionary response to the proliferation of TEs in the genome, fungi evolved repeat induced point mutations (RIP) (Selker 1990). RIP is a premeiotic mechanism that rapidly mutates copies of TEs and other near identical sequences. Leakage of RIP can introduce mutations into neighbouring genes. The repetitive nature of TEs also increases the rate of ectopic recombination and generates copy-number variation. In asexual species, mitotic rearrangements in TE-rich regions can occur if repeats are misaligned during double stranded break repair (Seidl & Thomma 2014).

Repeat-rich genomic compartments had a profound impact on the diversification of effector genes during pathogen evolution (Daverdin et al. 2012; Klosterman et al. 2011; Raffaele et al. 2010; Rouxel et al. 2011; Spanu et al. 2010). Effector genes in *L. maculans* accumulated mutations within the span of a few years (Daverdin et al. 2012). In *Z. tritici* and *P. infestans*, effector genes are among the most polymorphic genes in the genome. Interestingly, some

globally distributed pathogens, such as *L. biglobosa* and *Ustilago maydis* have genomes without a pronounced genome compartmentalization (Grandaubert et al. 2014; Kämper et al. 2006). Nonetheless, in many of these pathogens virulence genes are still associated with TEs and are located in regions associated with signatures of rapid evolution (Dutheil et al. 2016; Kämper et al. 2006; Möller & Stukenbrock 2017).

Generating variability in effectors through mutation accumulation exposes effector genes to both beneficial and deleterious mutations. At high mutation rates, each genotype is likely to simultaneously harbor multiple mutations that positively or negatively affect effector functions. A crucial mechanism for rapid adaptive evolution is the reshuffling of variants through recombination, which breaks beneficial mutations free from a background of deleterious mutations (Hill & Robertson 2007; Otto & Barton 1997; Otto & Lenormand 2002). Hence, recombination at effector loci should speed up the fixation of beneficial mutations. Recent studies on the recombination landscape of pathogens revealed interesting associations between recombination rates and virulence functions. In *Z. tritici*, genes encoding secreted proteins were overrepresented in recombination hotspots (Croll et al. 2015) and recombination rates increased compared to the sister species *Z. ardabiliae* infecting wild grasses (Stukenbrock & Dutheil 2017). Similar associations of recombination rates and genes putatively involved in host interactions were also found in *F. graminearum* (Laurent et al. 2017).

End-of-life stages of an effector in pathogen populations

Strong selection on pathogen populations to escape recognition led to the fixation of non-synonymous substitutions in effector genes such as *AvrStb6* (Zhong et al. 2017).The analyses of large pathogen collections revealed additional mechanisms of escape from

recognition. Many effector gene loci were found to segregate presence-absence polymorphisms leading to a mosaic of effector genes within species, including *M. oryzae* (Dong et al. 2015; Huang et al. 2014; Yoshida et al. 2016, 2009), *P. infestans* (Dong et al. 2014), *Z. tritici* (Hartmann & Croll 2017), *L. maculans* (Rouxel & Balesdent 2017) and *Melanopsichium pennsylvanicum* (Sharma et al. 2014). However, gene deletion is not the only mechanism for a pathogen population to generate loss-of-function variation at an effector locus. In *L. maculans*, RIP deactivated functional copies of the AvrLm6 or AvrLm4-7 effector genes leading to the breakdown of the oilseed rape resistance genes Rlm6 and Rlm7 (Daverdin et al. 2012; Fudal et al. 2009). An isolate of *M. oryzae* likely gained virulence on rice carrying the Pi9 resistance gene through the disruption of the AvrPi9 effector gene by a TE (Wu et al. 2015).The Ave1 gene of *V. dahliae* is flanked by TEs, which mediated frequent gene losses in populations through non-homologous recombination (de Jonge et al. 2012; Faino et al. 2016; de Jonge et al. 2013). Segregating loss-of-function variation at an important effector locus is strongly indicative that the loss confers a fitness benefit to the pathogen. Given sufficiently strong selection, the effector gene will disappear entirely from a pathogen's gene pool. Such selection can be particularly strong for Avr genes that 'tripped the wire' and are recognized by a cognate host (Bialas et al. 2017).

A consequence of frequent effector gene losses is that the gene content among pathogen strains should vary. Analyses of complete genome sequences of *Z. tritici* showed that homologous chromosomes were highly polymorphic in gene content and length (Croll et al. 2013; Plissonneau et al. 2016; Schotanus et al. 2015). Individual genomes harbored hundreds of orphan genes that were not fixed within the species and these often clustered in blocks of non-syntenic chromosomal segments. As a first for any pathogenic eukaryote, a pangenome was constructed based on five complete genome sequences (Plissonneau et al. 2018). The set of distinct genes discovered among all the genomes was significantly larger

than the gene set described in the reference genome of the species. In particular, the effector gene complement was signifi-cantly expanded in the pangenome (Plissonneau et al. 2018). With the aim of unravelling mechanisms leading to variation in gene content, a population genomic study analyzed over a hundred worldwide *Z. tritici* strains (Hartmann & Croll 2017). Some isolates recently lost primarily pathogenicity related genes and gene clusters encoding secondary metabolite production pathways (Figure 2) (Hartmann & Croll 2017)d. Both recent gene gains and losses more likely occurred in proximity to TEs. Genes that segregate presence-absence variation within the species are less likely to encode a conserved protein domain and are more weakly expressed (Hartmann & Croll 2017). The dynamics of the pathogen genome driving the rise and fall of effector genes was particularly well illustrated by the *Z. tritici* avirulence gene Zt_8_609. Although the gene is of recent origin as it lacks any homologs in closely related species, the gene was recently lost again in some strains by rearrangements in a nearby TE cluster (Figure 2b) (Hartmann et al. 2017).

Conclusions and outlook

The evolution of plant pathogens is intimately associated with the emergence and loss of effectors. The sequencing and assembly of complete pathogen genomes revolution- ized our understanding of effector gene localization. Through convergent evolution of distant pathogen lineages, effectors often emerged in highly repetitive regions rich in TEs. This ‘two-speed’ genome architecture affected nearly every aspect of effector evolution, including transcriptional control, mutation rates, loss-of- function and deletions. However, the strong association raises the question about causality. The presence of TEs can demonstrably impact sequence composition through RIP or cause the silencing of nearby genes. However, pathogens that were highly successful in evolving effectors in fast-evolving compartments may also have carried mutations that prevented the efficient containment of TEs in the genome. Hence, the association of effectors and fast-evolving compartments may not be ‘by design’ but rather be an unintended consequence of TE proliferation. A similar scenario may arise if strong selection favors the deletion of effector genes. Gene deletions can happen through chromosomal rearrangements and as a consequence the pathogen population segregates structural variation. Such polymorphism can cause additional chromosomal rearrangements through non-homologous recombination. As with the invasion of TEs, the observed structural variation may not directly benefit effector gene evolution but rather be a consequence of strong selection for gene losses. Dissecting adaptive versus neutral processes in pathogen genome evolution will be a rewarding future challenge.

The ever-expanding genomic datasets of plant pathogens are creating unique opportunities to disentangle the individual factors driving effector evolution. The conundrum of why so many effectors are taxon-specific can be addressed by identifying suitable species complexes where rapidly evolving genes can be tracked across species and time. Large-

scale transcriptomic and epigenetic analyses will provide additional insight into how effector gene regulation was implemented over evolutionary time. In particular, identifying the earliest evidence for a functional effector that is regulated in synchronicity with the progression of an infection and that is influencing the outcome of it will help address how pathogens gained functional novelty. Hence, an interdisciplinary effort that considers both effector function and evolutionary origins will be necessary.

Acknowledgements

We are grateful for helpful comments from F. E. Hartmann, B. A. McDonald, J. Palma-Guerrero, A. Sanchez-Vallet and U. Oggenfuss on an earlier version of this manuscript. SF is supported by the Swiss National Science Foundation grant 31003A_155955. CP was supported by an INRA Young Scientist grant. DC is supported by the Swiss National Science foundation grant 31003A_173265.

References

Akcapinar GB, Kappel L, Sezerman OU, Seidl-Seiboth V. 2015. Molecular diversity of LysM carbohydrate-binding motifs in fungi. Curr. Genet. 61:103–113. doi: 10.1007/s00294-014-0471-9.

Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. Curr. Opin. Genet. Dev. 17:513–518. doi: 10.1016/j.gde.2007.09.001.

Bialas A et al. 2017. Lessons in effector and NLR biology of plant-microbe systems. bioRxiv. 171223. doi: 10.1101/171223.

Carvunis A-R et al. 2012. Proto-genes and *de novo* gene birth. Nature. 487:370–374. doi: 10.1038/nature11184.

Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates fungal alkaloid biosynthesis in a fungal endophyte-plant symbiosis. Mol. Microbiol. 92:413–434. doi: 10.1111/mmi.12567.

Chuma I et al. 2011. Multiple Translocation of the *AVR-Pita* Effector Gene among Chromosomes of the Rice Blast Fungus *Magnaporthe oryzae* and Related Species. PLOS Pathog. 7:e1002147. doi: 10.1371/journal.ppat.1002147.

Connolly LR, Smith KM, Freitag M. 2013. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLOS Genet. 9:e1003916. doi: 10.1371/journal.pgen.1003916.

Cook DE, Mesarich CH, Thomma BPHJ. 2015. Understanding plant immunity as a surveillance system to detect invasion. Annu. Rev. Phytopathol. 53:541–563. doi: 10.1146/annurev-phyto-080614-120114.

Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics. 201:1213–1228. doi: 10.1534/genetics.115.180968.

Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLOS Genet. 9:e1003567. doi: 10.1371/journal.pgen.1003567.

Daverdin G et al. 2012. Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. PLOS Pathog. 8:e1003020. doi: 10.1371/journal.ppat.1003020.

Dong et al. 2015. Global genome and transcriptome analyses of *Magnaporthe oryzae* epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and lose dynamics in genome evolution. PLOS Pathog. 11:e1004801. doi: 10.1371/journal.ppat.1004801.

Dong S et al. 2014. Effector Specialization in a Lineage of the Irish Potato Famine Pathogen. Science. 343:552–555. doi: 10.1126/science.1246300.

Dong Suomeng, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. bioRxiv. 021774. doi: 10.1101/021774.

Dutheil JY et al. 2016. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol. Evol. evw026. doi: 10.1093/gbe/evw026.

Estep LK et al. 2015. Emergence and early evolution of fungicide resistance in North American populations of *Zymoseptoria tritici*. Plant Pathol. 64:961–971. doi: 10.1111/ppa.12314.

Faino L et al. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. gr.204974.116. doi: 10.1101/gr.204974.116.

Feng S et al. 2010. Conservation and divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. U. S. A. 107:8689–8694. doi: 10.1073/pnas.1002720107.

Fisher MC et al. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature. 484:186–194. doi: 10.1038/nature10947.

Friesen TL et al. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. 38:953–956. doi: 10.1038/ng1839.

Fudal I et al. 2009. Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. Mol. Plant-Microbe Interact. MPMI. 22:932–941. doi: 10.1094/MPMI-22-8-0932.

Gibriel HAY, Thomma BPHJ, Seidl MF. 2016. The age of effectors: Genome-based discovery and applications. Phytopathology. 106:1206–1212. doi: 10.1094/PHYTO-02-16-0110-FI.

Grandaubert J et al. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. BMC Genomics. 15. doi: 10.1186/1471-2164-15-891.

Guillen K de et al. 2015. Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. PLOS Pathog. 11:e1005228. doi: 10.1371/journal.ppat.1005228.

Hacquard S et al. 2013. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. Proc. Natl. Acad. Sci. 110:E2219–E2228. doi: 10.1073/pnas.1306807110.

Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. Mol. Biol. Evol. doi: 10.1093/molbev/msx208.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11:1189–1204. doi: 10.1038/ismej.2016.196.

Hill WG, Robertson A. 2007. The effect of linkage on limits to artificial selection. Genet. Res. 89:311–336. doi: 10.1017/S001667230800949X.

Huang J, Si W, Deng Q, Li P, Yang S. 2014. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. BMC Genet. 15:45. doi: 10.1186/1471-2156-15-45.

Inoue Y et al. 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 357:80–83. doi: 10.1126/science.aam9654.

Islam MT et al. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. BMC Biol. 14:84. doi: 10.1186/s12915-016-0309-7.

de Jonge R et al. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23:1271–1282. doi: 10.1101/gr.152660.112.

de Jonge R et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. 109:5110–5115. doi: 10.1073/pnas.1119623109.

Kämper J et al. 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature. 444:97–101. doi: 10.1038/nature05248.

Kleemann J et al. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLOS Pathog. 8:e1002643. doi: 10.1371/journal.ppat.1002643.

Klosterman SJ et al. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLOS Pathog. 7:e1002137. doi: 10.1371/journal.ppat.1002137.

Laurent B et al. 2017. High-resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. Mol. Plant Pathol. n/a-n/a. doi: 10.1111/mpp.12524.

Ma L-J et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464:367–373. doi: 10.1038/nature08850.

McDonald MC, Ahren D, Simpfendorfer S, Milgate A, Solomon PS. 2017. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. Mol. Plant Pathol. doi: 10.1111/mpp.12535.

McLysaght A, Guerzoni D. 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Phil Trans R Soc B. 370:20140332. doi: 10.1098/rstb.2014.0332.

McLysaght A, Hurst LD. 2016. Open questions in the study of *de novo* genes: what, how and why. Nat. Rev. Genet. 17:567–578. doi: 10.1038/nrg.2016.78.

Menardo F et al. 2016. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nat. Genet. N. Y. 48:201-205B. doi: http://dx.doi.org/10.1038/ng.3485.

Mikkelsen TS et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 448:553–560. doi: 10.1038/nature06008.

Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. advance online publication. doi: 10.1038/nrmicro.2017.76.

Otto SP, Barton NH. 1997. The evolution of recombination: Removing the limits to natural selection. Genetics. 147:879–906.

Otto SP, Lenormand T. 2002. Evolution of sex: Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3:nrg761. doi: 10.1038/nrg761.

Palma-Guerrero J et al. 2016. Comparative transcriptomic analyses of Zymoseptoria tritici strains show complex lifestyle transitions and intraspecific variability in transcription profiles. Mol. Plant Pathol. 17:845–859. doi: 10.1111/mpp.12333.

Plissonneau C et al. 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8. doi: 10.3389/fpls.2017.00119.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16:5. doi: 10.1186/s12915-017-0457-4.

Plissonneau C, Stürchler A, Croll D. 2016. The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. mBio. 7:e01231-16. doi: 10.1128/mBio.01231-16.

Poppe S, Dorsheimer L, Happel P, Stukenbrock EH. 2015. Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). PLOS Pathog. 11:e1005055. doi: 10.1371/journal.ppat.1005055.

Qutob D, Patrick Chapman B, Gijzen M. 2013. Transgenerational gene silencing causes gain of virulence in a plant pathogen. Nat. Commun. 4:1349. doi: 10.1038/ncomms2354.

Raffaele S et al. 2010. Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. Science. 330:1540–1543. doi: 10.1126/science.1193070.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10:417–430. doi: 10.1038/nrmicro2790.

Rouxel T et al. 2011. Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202. doi: 10.1038/ncomms1189.

Rouxel T, Balesdent M-H. 2017. Life, death and rebirth of avirulence effectors in a fungal pathogen of Brassica crops, *Leptosphaeria maculans*. New Phytol. 214:526–532. doi: 10.1111/nph.14411.

Sánchez-Vallet A, Mesters JR, Thomma BPHJ. 2015. The battle for chitin recognition in plant-microbe interactions. FEMS Microbiol. Rev. 39:171–183. doi: 10.1093/femsre/fuu003.

Saunders DGO et al. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLOS ONE. 7:e29847. doi: 10.1371/journal.pone.0029847.

Savory F, Leonard G, Richards TA. 2015. The role of horizontal gene transfer in the evolution of the oomycetes. PLOS Pathog. 11:e1004805. doi: 10.1371/journal.ppat.1004805.

Schmidt SM et al. 2013. MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. BMC Genomics. 14:119. doi: 10.1186/1471-2164-14-119.

Schotanus K et al. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 8:41. doi: 10.1186/s13072-015-0033-5.

Seidl MF, Thomma BPHJ. 2014. Sex or no sex: evolutionary adaptation occurs regardless. BioEssays News Rev. Mol. Cell. Dev. Biol. 36:335–345. doi: 10.1002/bies.201300155.

Selker EU. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. 24:579–613. doi: 10.1146/annurev.ge.24.120190.003051.

Sharma R, Mishra B, Runge F, Thines M. 2014. Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the Smut Fungus *Melanopsichium pennsylvanicum*. Genome Biol. Evol. 6:2034–2049. doi: 10.1093/gbe/evu148.

Singh RP et al. 2011. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. Annu. Rev. Phytopathol. 49:465–481. doi: 10.1146/annurev-phyto-072910-095423.

Skibbe DS, Doehlemann G, Fernandes J, Walbot V. 2010. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. Science. 328:89–92. doi: 10.1126/science.1185775.

Soyer JL et al. 2014. Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus Leptosphaeria maculans. PLOS Genet. 10:e1004227. doi: 10.1371/journal.pgen.1004227.

Spanu PD et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 330:1543–1546. doi: 10.1126/science.1194573.

Studt L et al. 2016. Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. Environ. Microbiol. n/a-n/a. doi: 10.1111/1462-2920.13427.

Stukenbrock EH. 2016a. Hybridization speeds up the emergence and evolution of a new pathogen species. Nat. Genet. 48:113–115. doi: 10.1038/ng.3494.

Stukenbrock EH. 2016b. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology. 106:104–112. doi: 10.1094/PHYTO-08-15-0184-RVW.

Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc. Natl. Acad. Sci. 109:10954–10959. doi: 10.1073/pnas.1201403109.

Stukenbrock EH, Dutheil JY. 2017. Comparison of fine-scale recombination maps in fungal plant pathogens reveals dynamic recombination landscapes and intragenic hotspots. bioRxiv. 158907. doi: 10.1101/158907.

Wang Q et al. 2011. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. Plant Cell Online. tpc.111.086082. doi: 10.1105/tpc.111.086082.

Wu J et al. 2015. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector *AvrPi9* that triggers *Pi9*-mediated blast resistance in rice. New Phytol. 206:1463–1475. doi: 10.1111/nph.13310.

Yoshida K et al. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 17:370. doi: 10.1186/s12864-016-2690-6.

Yoshida Kentaro et al. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell. 21:1573–1591. doi: 10.1105/tpc.109.066324.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 328:916–919. doi: 10.1126/science.1186366.

Zhong Z et al. 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. New Phytol. 214:619–631. doi: 10.1111/nph.14434.

GENERAL DISCUSSION

General discussion and perspectives

In chapter one I looked at the role of transposable elements de-repression dynamics during stress induced by wheat infection and under starvation stress across four different genetic backgrounds. I found that transposable elements responded differently to the stresses, with some being upregulated during early infection and some were upregulated under starvation stress. Although most of the TE families showed some de-repression in response to at least one of the stress conditions, the stress responsiveness varied between different transposable element families and also depended on the genetic background. I also show that genes in close proximity to TEs show upregulation during early infection. Genes and TEs that are in close proximity to one another are likely to be affected by the same epigenetic control. Furthermore, effector genes in close proximity to TEs show a higher peak of expression during early infection than effector genes that are not near to TEs. The co-location of effectors and TEs and their subsequent joint epigenetic control may underlie key adaptations to avoid detection by the host. However, this co-localization is only beneficial in the absence of stress. Stress de-repression places a mutational burden on the genome.

Chapter two focused on quantifying the total number of chromosome rearrangements occurring through a single round of meiosis in hundreds of progeny isolates from two crosses and found that the rates of disomy and rearrangements differed greatly between chromosomes and crosses. Most aberrant inheritance events affected accessory chromosomes, but some core chromosomes were also affected. Several accessory chromosomes that were present in only one of the parents showed strong distorted inheritance. Progeny disomic for core and accessory chromosomes were found. Finally, I found a correlation between repetitive element content and the fidelity with which chromosomes go through meiosis, where chromosomes with a higher repeat content and

lower synteny in the parents were more frequently rearranged, disomic or lost in the progeny.

Chapter three zooms in on a serendipitously discovered major chromosome rearrangement in a cross. During this project I assembled the enlarged chromosome 17 and show that it formed through ectopic recombination of a TE belonging to the family *Styx*. I track the chromosome through another three rounds of meiosis and show that it goes through cycles of degeneration involving two processes, namely ectopic recombination and non-disjunction. I show that the degenerative cycles are fairly random as progeny from the same cross had different derivatives of chromosome 17, but also to some extent predictable as the same TE family (*Styx*) and soft spots for double stranded breakage and fusion were repeatedly involved.

Transposable elements de-repression dynamics, the Dr. Jekyll and Mr. Hyde of pathogen genomes

We know that transposable elements are often found in close proximity to effector genes and that TEs and chromosome rearrangements play a role in effector gene birth and death cycles, which serves as a foundation for rapid pathogen evolution (Fouché et al. 2018). Simultaneously, pathogens have evolved to time the expression of effectors during infection stress by co-localizing genes with TEs in epigenetically silenced regions (Sánchez-Vallet et al. 2018). Simultaneously, de-repressed TEs can be highly mutagenic and can drive chromosome rearrangement either through ectopic recombination or through their activity (Le Rouzic and Capy 2005; Belyayev 2014). Both processes are outcomes of the same fundamental properties of TEs, making TEs the Dr. Jekyll and Mr. Hyde of pathogen genomes.

Interestingly, in our study, some of the most highly expressed TEs during infection were located in highly conserved regions of the genome, while some of the least expressed TEs are located on accessory chromosomes. It seems counterintuitive that accessory chromosomes are highly rearranged when compared to these core regions with highly de-repressed TEs. This can be explained by the two processes having very different evolutionary scales. TE de-repression occurs under stress induced in a single infection cycle. Chromosome rearrangements occur over longer evolutionary scales, where rearrangements in the core chromosomes that have a negative impact on pathogen fitness and are counter-selected. Purifying selection on accessory chromosomes is much weaker as they do not encode essential functions (Cooper et al. 2010) and therefore TEs accumulate in these regions and chromosome rearrangements are far more prolific.

A side effect of co-locating effector genes in TE dense regions and regulating the expression of effectors epigenetically, is that TEs in the vicinity are also de-repressed. Opting for effectors that are regulated epigenetically places a mutational burden on the host genome and creates a tenuous equilibrium between allowing TE de-repression during infection and counter-selection against insertions that decrease host fitness. Furthermore, TE de-repression dynamics may underlie recent genome expansions within some *Z. tritici* populations (Badet et al. 2019, Oggenfuss unpublished). De-repressed TEs can only transpose into other de-repressed regions and therefore regions that are epigenetically silenced during infection could be sheltered from TE activity. At present we do not know to what extent the active TEs that are let loose by the continuous de-repression during infection are placing a burden on the genome.

Can TE activity and structural variation ultimately lead to genome compartmentalization?

The opposing forces of on the one hand selecting for effective TE control under infection stress while selecting for coordinated expression of genes required for successful infection suggest that there should be standing genetic variation amongst different *Z. tritici* strains for the ability to control TEs. The resulting permissive TE control in *Z. tritici* strains, allowed highly active TEs to undergo recent bursts in some populations. Cycles of TE de-repression during infection can lead to genome expansion and also create “TE-islands” in other de-repressed regions and may have played a significant role in shaping the genome in this fungal pathogen. *Z. tritici* has clear genome compartmentalization with 13 core chromosomes and up to eight accessory chromosomes. However, the genome structure is quite unique because the core chromosomes contain islands that harbor nearly all of the TEs in the genome (Plissonneau et al. 2018). These islands show absence-presence polymorphism between different isolates of *Z. tritici* (Plissonneau et al. 2016; Hartmann et al. 2017; Hartmann and Croll 2017; Plissonneau et al. 2018).

TE-driven genome compartmentalization, or the so called two-speed genome structure, was proposed to be to the pathogen’s advantage (Dong et al. 2015). This type of organization could enable essential functions to be conserved in the core regions where there are very few TEs, while new functions can evolve in the rapidly diversifying “TE-islands”. However, it is not clear whether such a genome structure can be selected for. Is the genome structure not just a consequence of TE insertions being limited to other de-repressed regions, and selection on TE insertions being more relaxed in TE-rich regions? Similarly, it is not clear if regions that have a higher propensity to undergo rearrangements can be selected for? Perhaps TEs are driving genome evolution based on the strength of counter-selection on new insertion sites. Highly deleterious TE insertions are strongly counter selected, while

selection on neutral insertions such as in the TE-rich regions is more relaxed. Similarly, chromosome rearrangements that are highly deleterious will have strong counter-selection and thereby retain the integrity of core chromosome regions, while neutral rearrangements, such as the highly degenerated chromosome 17, are tolerated. In this model, what we observe in terms of TE insertions, chromosome rearrangements and the resulting genome structure would be events that are neutral or near neutral and therefore passed the selection filter. The ability of *Z. tritici* to tolerate chromosome duplications, losses and rearrangements and multiple TE insertion makes this species a great model to observe and investigate the interplay between TE dynamics and chromosome rearrangements.

References

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2019. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. bioRxiv:803098.

Belyayev A. 2014. Bursts of transposable elements as an evolutionary driving force. J. Evol. Biol. 27:2573–2584.

Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. 2010. Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. PLOS Comput Biol 6:e1000732.

Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr. Opin. Genet. Dev. 35:57–65.

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46:34–42.

Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. Mol. Biol. Evol.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11:1189–1204.

Le Rouzic A, Capy P. 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 169:1033–1043.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16:5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio 7:e01231-16.

Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56:21–40.

ACKNOWLEDGEMENTS

I am sincerely grateful to all the people who contributed to my work and supported me during my PhD research and writing. In particular I want to thank:

- My PhD supervisors, Prof. Dr. Daniel Croll and Prof. Dr. Bruce McDonald for their support and advice throughout my PhD. Thank you Daniel, for guiding me on this journey. I have learnt most of what I know about Bioinformatics and evolutionary genomics from you. Thank you, Bruce, for your support and for hosting me in your group for the major period of my PhD.

- Prof. Dr. Alexander Suh for agreeing to be the external examiner and for travelling to Zurich.

- My project collaborators and co-authors, especially Ursula Oggenfuss and Dr. Thomas Badet. I have learnt so much from both of you.

- Marcello Zala for your help in the acquisition of the data used in this research. Thank you for your tips and assistance with DNA extractions.

- ETH Zurich and the Swiss National Fund for funding my PhD project and their continued commitment to science, research and education. It has been a privilege. Thank you to the Plant-Science Center (PSC) for contributing to my PhD training.

- ALL of the current and former members of the Plant Pathology group. Thank you for the many discussions and excursions.

- My new family at the University of Neuchâtel. Without your environment of collaboration and support I would have been hard pressed to complete my doctoral thesis. Thank you all for the discussions, ideas, support and for all the fun. Thanks to Vinciane and Nikhil for

feeding me. In particular, thank you to my office mates. I want to thank Leen, for the amazing jungle we work in, Sabina for always initiating fun activities and for spending the final weekends in the lab with me, and Luzia for her truly infectious laugh and enthusiasm for life!

- My TE-sister, Ursula. It has been a journey! Thank you for reminding me daily why I love science with your never wavering enthusiasm for all topics. Thank you for being my sister in science! I do not know how I would have survived the last weeks without your special tea blends, all the motivational gifts and your unique sense of humour. Thank you for your support!

- Nikhil and Thomas. Thank you for the beers and the laughter, for celebrating the victories and helping me forget the failures.

- Last, but not the least, sincere thanks to my friends, family and Nigel, for their support, enthusiasm and passion and always making me feel as if I can do anything I set my mind to doing.

SCIENTIFIC CONTRIBUTIONS

Published Work

Fouché S., Plissonneau C., Croll D. A. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Current Opinion in Microbiology. January 2018.

Fouché S., Plissonneau C., McDonald B. A, Croll D. A. Meiosis Leads to Pervasive Copy-Number Variation and Distorted Inheritance of Accessory Chromosomes of the Wheat Pathogen Zymoseptoria tritici. Genome Biology and Evolution. May 2018.

Fouché S., Badet T., Oggenfuss U., Plissonneau C., Francisco C. S., Croll D. A. Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. Molecular Biology and Evolution. September 2019.

Sánchez-Vallet A., **Fouché S.,** Fudal I., Hartmann F. E., Aurelien T., Croll D. A. The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. Annual Review of Phytopathology. August 2018.

Plissonneau C., Benevenuto J., Assaad. N. B. M., **Fouché S**., Hartmann F. E., Croll. D. A. Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. Frontiers in Plant Science. January 2017.

Yates, S., Mikaberidze A., Krattinger S. G., Abrouk M., Hund A., Yu K., Studer B., **Fouché S.,** Meile L., Pereira D., Karisto P., McDonald B. A. Precision Phenotyping Reveals Novel Loci for Quantitative Resistance to Septoria Tritici Blotch. September 2019.

Unpublished Work

Fouché S., Zala M., McDonald B. A, Croll D. Repeated chromosomal degeneration through repeat-induced chromosomal rearrangements in a fungal pathogen

Conferences and Presentations

Biology 16, Switzerland, Lausanne, 11.02.2016

Type of contribution: Poster
Title of article or contribution: Large-scale detection of chromosomal abnormalities following meiosis in a fungal pathogen of wheat

Population genomics of oomycete and fungal pathogens, Switzerland, Ascona, 08.05.2017

Type of contribution: Poster
Title of article or contribution: Pervasive segregation distortion and copy-number variation of accessory chromosomes in *Zymoseptoria tritici*

Dothideomycete workshop, Israel, Haifa, 25.02.2018

Type of contribution: Talk given at a conference
Title of article or contribution: Meiosis drives chromosome instability in *Zymoseptoria tritici*

European Fungal genetics Conference, Israel, Haifa, 27.02.2018

Type of contribution: Poster
Title of article or contribution: Differential expression of transposable elements through the course of a *Zymoseptoria tritici* infection

2nd Uppsala Transpososn Symposium, Uppsala, Sweden, 4.10.2018

Type of contribution: Talk
Title of article or contribution: Differential expression of transposable elements through the course of a *Zymoseptoria tritici* infection

CURRICULUM VITAE

Simone Fouché

Rue du Râteau 4a
Neuchâtel
2000
Switzerland
simone.fouche@usys.ethz.ch | simfou@gmail.com

Education and Professional Development

2015 – Present (November 2019): Doctoral candidate in Plant Pathology, ETH Zurich, Switzerland

Thesis Title: Drivers of genome evolution of a fungal pathogen of wheat

Supervisors: Prof. BA McDonald, Prof. D. Croll

The aim of this project was to study the causes and consequences of chromosome rearrangements. To understand the origins of TE-mediated chromosomal rearrangements, in my first chapter I focus on factors driving the activation or repression of TEs in the genome. The second chapter quantifies the number of whole chromosome losses, duplications and rearrangements occurring in a single round of meiosis in hundreds of progeny. I also identified mechanisms affecting the fidelity with which chromosomes go through meiosis. The third chapter focuses on a newly discovered rearranged chromosome and tracks the exact processes by which such a unstable chromosome is created and degenerates through five rounds of meiosis.

2013 – 2015: MSc in Microbiology (with distinction) **University of Pretoria, SA**

Master's Project: Molecular characterization of an accessory chromosome of *Fusarium circinatum*

Supervisors: Prof. ET Steenkamp, Prof. BD Wingfield, Dr. MPA Coetzee

My thesis focused on an accessory chromosome found in *Fusarium circinatum*, an important pathogen of pine trees. The first chapter characterized the accessory chromosome found in this species. The chromosome was sequenced, assembled and compared to the accessory chromosomes of two other closely related *Fusarium* species. A large part of the project involved optimization of the pulsed-field gel electrophoresis technique to separate chromosomes, as this technique had not previously been used at the University of Pretoria. The second chapter focused on comparing the accessory chromosome in a global population of *F. circinatum* in terms of length and presence-absence polymorphism.

2012: BSc (Hons) Microbiology (with distinction) **University of Pretoria, SA**

Core Modules: Dissertation; Research methods; Seminar course (Scientific presentation skills); Trends in microbiology; Molecular and cellular biology

Research Project: Bluetongue virus and African horse sickness virus: plasmid-based reverse genetics

Supervisor: Prof. J. Theron

Worked towards establishing a plasmid-based reverse genetics system for bluetongue virus and African horse sickness virus. Gained experience in cloning, designing inserts and cell culture.

2008 – 2011 **BSc Microbiology (with distinction) (2011)** **University of Pretoria, SA**

Core Modules BSc: Biochemistry of proteins, Immunobiology, Structure and diversity of viruses, Environmental microbiology, Vertebrate-microbe interaction, Trends in microbiology, Food microbiology, Molecular biology of prokaryotes, Genetic manipulation of microbes.