This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.

Author(s):   Nikhil Kumar Singh, Petteri Karisto & Daniel Croll

Title: Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*

Year: 2021

Version: Published version

Copyright: The Author(s) 2021

Rights: CC BY-NC 4.0

Rights url: http://creativecommons.org/licenses/by-nc/4.0/

Please cite the original version:

Singh N.K., Karisto P., Croll D. (2021)l. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici.* Microbial Genomics 7(10) article id 000678. https://doi.org/10.1099/mgen.0.000678.

MICROBIOLOGY
SOCIETY

OPEN DATA   OPEN ACCESS

Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*

Nikhil Kumar Singh[1], Petteri Karisto[2] and Daniel Croll[1,*]

Abstract

Pathogens cause significant challenges to global food security. On annual crops, pathogens must re-infect from environmental sources in every growing season. Fungal pathogens have evolved mixed reproductive strategies to cope with the distinct challenges of colonizing growing plants. However, how pathogen diversity evolves during growing seasons remains largely unknown. Here, we performed a deep hierarchical sampling in a single experimental wheat field infected by the major fungal pathogen *Zymoseptoria tritici*. We analysed whole genome sequences of 177 isolates collected from 12 distinct cultivars replicated in space at three time points of the growing season to maximize capture of genetic diversity. The field population was highly diverse with 37 SNPs per kilobase, a linkage disequilibrium decay within 200–700 bp and a high effective population size. Using experimental infections, we tested a subset of the collected isolates on the dominant cultivar planted in the field. However, we found no significant difference in virulence of isolates collected from the same cultivar compared to isolates collected on other cultivars. About 20% of the isolate genotypes were grouped into 15 clonal groups. Pairs of clones were disproportionally found at short distances (<5 m), consistent with experimental estimates for per-generation dispersal distances performed in the same field. This confirms predominant leaf-to-leaf transmission during the growing season. Surprisingly, levels of clonality did not increase over time in the field although reproduction is thought to be exclusively asexual during the growing season. Our study shows that the pathogen establishes vast and stable gene pools in single fields. Monitoring short-term evolutionary changes in crop pathogens will inform more durable strategies to contain diseases.

DATA SUMMARY

All Illumina sequence data are available from the NCBI SRA BioProject number PRJNA596434 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596434). Tables S1–S3 (available in the online version of this article) list the exact strain names, collection location, genotype and genetic diversity indices.

INTRODUCTION

Infectious diseases have a major impact on plant and animal populations including humans. Designing effective strategies for disease prevention requires knowledge of the evolutionary dynamics of pathogens. The evolutionary responses of pathogen populations to novel resistance factors and changing environmental conditions are critical [1]. Responses are often a function of genetic diversity in pathogens with consequences for disease severity and spread in heterogeneous host populations [2–4]. Highly resistant host populations tend to harbour the most virulent pathogens, whereas less virulent pathogens dominate susceptible populations [3]. In addition to selection by host genotypes, pathogen populations are subject to selection at life cycle stages outside of the host [5–7] and other environmental factors including the application of chemical compounds such as fungicides [8, 9]. Pathogens of annual plants must survive periods of time in the absence of a host and succeed in re-infecting new plants [10, 11]. During the growing season of the host,

pathogens are selected to colonize additional plants. Hence, evolutionary trajectories are expected to vary in space and time depending on selective pressures imposed by both the host and the environment [12, 13]. Beyond selection, small pathogen populations may experience genetic bottlenecks or gene flow. Hence, it remains largely unknown how pathogen populations evolve over short periods [14].

In agricultural crop–pathogen systems, infections begin with the exposure of a susceptible host to pathogen strains at the beginning of the growing season. The intensity of an infection depends on the level of host resistance, the amount of pathogen inoculum and environmental factors [15, 16]. If the infection is successful, pathogens complete their life cycle within the host and then transmit to another host. The genetic architecture of traits to successfully passage between infection stages is likely to be distinct [17, 18]. Entering a host plant requires manipulation of the host immune system [19–21]. Proliferation and reproduction inside the plant depend probably on optimal resource exploitation and continued suppression of immunity [22, 23]. Dispersal between hosts is governed by propagule properties [24, 25]. Therefore, pathogen populations probably experience complex selection pressures over the life cycle. Depending on the mode of between-season transmission and environmental conditions, plant pathogen populations with annual infection cycles may experience severe demographic bottlenecks during the off-season, restricting the transmission of genetic variation between growing seasons [14, 26]. After a filtering of genotypes for persistence in the off-season environment, the surviving genotypes undergo selection for infectivity and spread among hosts during the subsequent epidemic phase. In any such scenario, pathogens must re-infect crops annually and successfully spread among plants in a field. Therefore, considering the time-scale of pathogen population turnover and demographic effects is important.

The mode of dispersal is often tightly associated with specific reproductive modes. Pathogens often have mixed reproductive strategies, meaning that sexual and asexual reproduction can alternate over the course of the life cycle adding to the complexity of recognizing species and challenging genetic diversity analyses [27, 28]. Fungal crop pathogens often reproduce asexually during the growing season to maximize dispersal at low population density. In turn, asexual reproduction during the epidemic phase probably has a strong influence on the genetic structure of pathogen populations [14]. On senescent plants, sexual reproduction can lead to favourable allele combinations for long-distance dispersal and survival [29]. An important pathogen with a mixed reproductive strategy is the fungal wheat pathogen *Zymoseptoria tritici*. This fungus is a haploid ascomycete and one of the most destructive pathogens of wheat with yield losses of 5–30% [30]. Populations worldwide harbour significant variation in pathogenicity and diversity at loci underlying host specificity [31–34]. The pathogen follows typically a sequential regime of sexual and asexual cycles of reproduction over the growing season. During the sexual stage, the pathogen produces asci containing four pairs of genetically distinct ascospores, which

Impact Statement

Microbial pathogens threaten crop production worldwide. A large body of work has investigated genetic factors contributing to the success of a pathogen attacking plants. Similarly, pathogen species have been investigated for levels of genetic diversity across their distribution range. However, how individual crop fields are colonized from pathogen inoculum and how this process influences genetic diversity at the field level remains poorly understood. Here, we performed a deep sampling and whole-genome sequencing of 177 isolates of the wheat pathogenic fungus *Zymoseptoria tritici* collected within the span of weeks in a single wheat field. The pathogen population is highly diverse, showing extensive signatures of recombination and a high effective population size. The pathogen managed to spread through asexual reproduction at the metre scale within the field without noticeably leading to reduced genetic diversity overall. Our study provides a detailed picture of how individual crop fields become colonized, with implications for how the spread of the pathogen could be more sustainably managed in the future.

are thought to be the source of primary infection in the field and can disperse over long distances [35, 36]. Once infections are established in a field, *Z. tritici* produces asexual pycnidiospores, which can disperse with rain splash and initiate secondary infections nearby. Spread within fields during the growing season is thought to be predominantly clonal [35]. Population genomics analyses of worldwide pathogen samples showed high genetic diversity within fields and extensive gene flow [32, 37–39]. Multi-year studies of *Z. tritici* populations have revealed trade-offs between the intra- and interannual scales in the evolution of aggressiveness [40] with re-infections at the beginning of the growing season imposing no detectable genetic bottleneck [41]. Consistent with the maintenance of high levels of diversity, individual wheat leaves are frequently infected by multiple different genotypes [42]. However, whole genome analyses of hierarchical field collections to resolve the spatial and temporal changes in genotypic diversity have been lacking.

In this study, we used whole-genome sequencing of 177 distinct isolates of *Z. tritici* from a single wheat field covering three time points over the growing season. We analysed 12 distinct cultivars replicated in space to maximize capture of genetic diversity. Based on genome-wide polymorphisms, we analysed standing levels of genetic diversity based on linkage disequilibrium and chromosomal presence/absence variation. Using information about the spatial organization in the field and experimental assessments of dispersal distances, we quantified the extent of clonal spread in space and time.
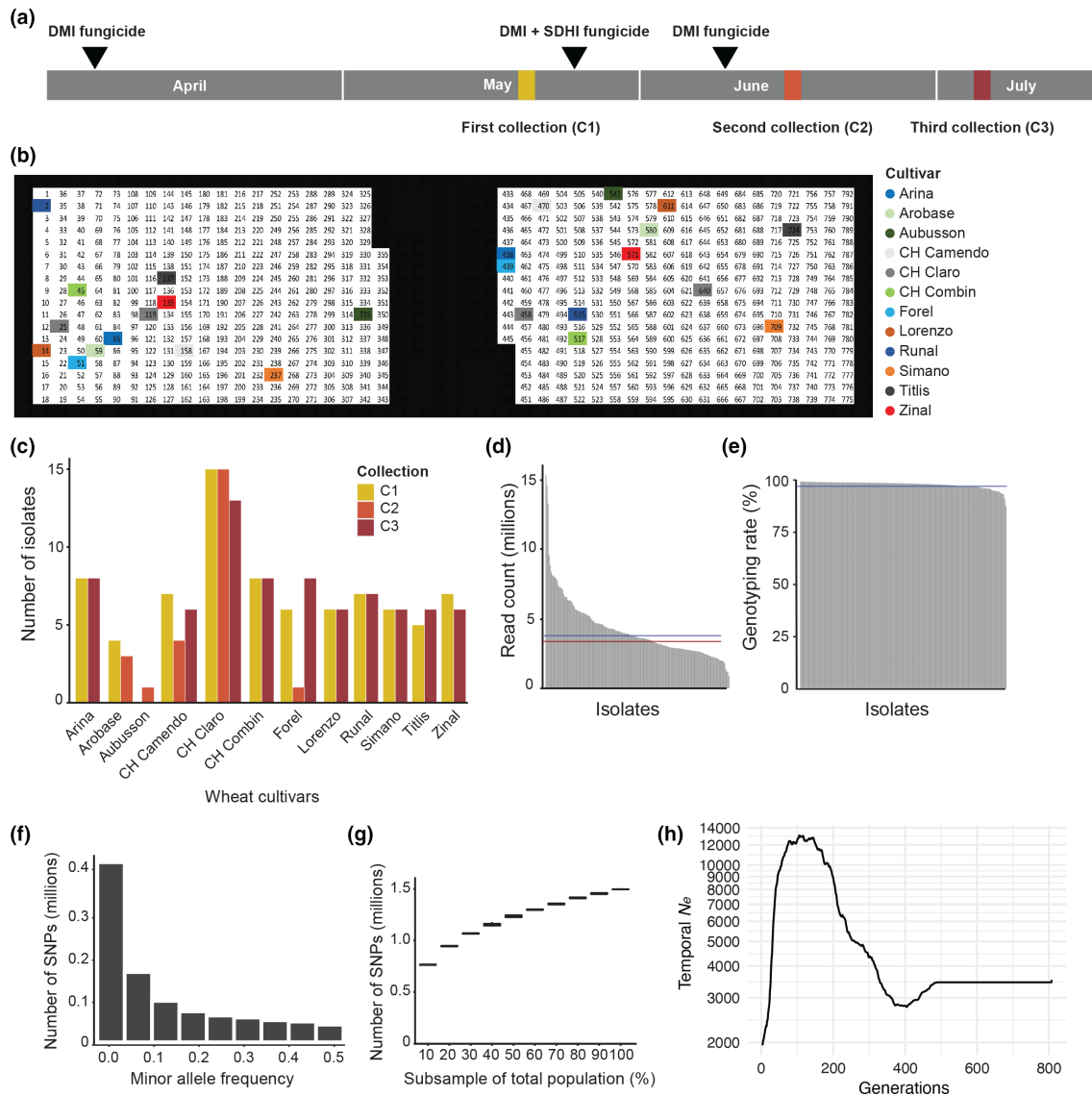
Fig. 1. Genome sequencing of 177 *Zymoseptoria tritici* isolates from a single field. (a) Time scale showing the three collection points from May to July. The experimental field was treated with fungicides three times (see Methods for details). (b) Graphical representation of the experimental wheat field showing the sampled cultivars (coloured background). Each number indicates a plot of 1.2×1.7 m. Wheat cultivar colour codes are maintained throughout the paper. Sampled cultivar numbers are also shown in Table S4. (c) Number of isolates collected from each of the 12 cultivars at each collection time point. (d) Trimmed read count for each isolate. Lines represent the mean (blue) and median (red) read counts. (e) Genotyping rate of all the 177 isolates (% loci genotyped). The blue line represents the mean genotyping rate. (f) Minor allele frequency (frequency at which the second most common allele occurs in a given population) spectrum of 1 496 037 SNPs genotyped in 177 isolates. (g) Boxplot showing the number of segregating SNPs in subsets of the 177 isolates. Error bars show 10 replicates of resampled isolates. (h) Estimates of temporal changes in effective population size (N_e) over time. The simulated generations range from the present (0) to 800 generations in the past.

METHODS

Field collection and storage

Z. tritici isolates were collected from the Field Phenotyping Platform (FIP) site of the ETH Zurich, Switzerland (47.449° N 198 8.682° E) [43]. A total of 335 European winter wheat varieties were grown in two replicates (Fig. 1b) during the 2015–16 growing season separated each by approximately 100 m. During the collection season in 2016 the following

fungicide applications were made: 4–8 April (300 g l⁻¹ Spirox-amin, 160 g l⁻¹ Prothioconazole), 25 May (Aviator Xpro with 75 g l⁻¹ Bixafen + 150 g l⁻¹ Prothioconazol), 6 June (Osiris with 56.25 g l⁻¹ epoxiconazole and 41.25 g l⁻¹ metconazole). Wheat was also subjected to regular applications of fertilizers and herbicides. A total of 177 isolates of *Z. tritici* were isolated from 12 winter wheat cultivars that are commonly grown in Switzerland [44] over three collection time points (Fig. 1a).

To isolate strains, individual cirri from pycnidia identified on infected leaves were streak-plated on a yeast sucrose broth (YSB) solid media plate and incubated at 18 °C for a week. Media were supplemented with 50 µg ml⁻¹ kanamycin to prevent bacterial growth. After the incubation period, a single colony from the media plate was inoculated into 35 ml YSB liquid medium and incubated on a shaking incubator at 18 °C for 8 days (140–180 r.p.m.). After the incubation period, a dense culture of blastospores was obtained. The culture was washed twice with sterile distilled water. The spores were stored as both glycerol or silica stock for future use. For the preparation of glycerol stocks, equal amounts of washed blastospores and sterile glycerol were mixed in a cryotube and stored at −80 °C for future use. For the preparation of silica stocks, 200 µl of washed blastospores was poured onto 1 ml of sterile silica powder in a cryotube. Silica stocks were stored at 80 °C.

Culture preparation and seedling infection assay

Isolates were revived from glycerol stock by adding 50 µl stock solution to a 50 ml conical flask containing 35 ml liquid YSB medium. The inoculated flasks were incubated in the dark at 18 °C and 140–180 r.p.m. After 8 days of incubation, the cultures were passed through four layers of meshed cheese-cloth and washed twice with sterile water to remove traces of media. The filtering step also largely eliminated hyphal biomass but retained spores. We analysed pathogenicity profiles for a subset of 120 isolates out of the total number of isolates analysed in this study. The pathogenicity assay dataset was generated to perform genome-wide association studies [45]. In summary, seedlings of the cultivar Claro were grown under controlled conditions as follows: 16/8 h day/night periods at 18 °C throughout the experiment. The growth chamber was maintained at 70% humidity. Plants were grown for 3 weeks before infection with *Z. tritici*. To establish infections, washed spores were diluted to 2×10^5 spores ml⁻¹ in 15 ml of sterile water containing 0.1% TWEEN20. Twenty-At 21 days post-inoculation, the second leaf of each plant was cut and fixed on a barcoded white paper. Leaves were scanned immediately using a flatbed scanner at 1200 dpi. The scanned images were batch-processed using automated image analysis [46]. We recorded the percentage leaf area covered by lesions and pycnidia counts.

Whole-genome sequencing and variant calling

Approximately 100 mg of lyophilized spores was used to extract high-quality genomic DNA using the Qiagen DNeasy Plant Mini Kit, according to the manufacturer's protocol. We sequenced paired-end reads of 100 bp each with an insert size of ~550 bp on the Illumina HiSeq 4000 platform. Raw reads are available on the NCBI Short Read Archive under the BioProject PRJNA596434 [47]. We performed sequencing quality checks using FastQC v.0.11.9. [48] and extracted read counts. Sequencing reads were then trimmed for adapter sequences and sequencing quality using Trimmomatic v.0.39 [49] using the following settings: illuminaclip=TruSeq3 PE.fa:2:30:10, leading=10, trailing=10,

sliding-window=5:10 and minlen=50. Trimmed sequencing reads were aligned to the reference genome IPO323 [50] using Bowtie2 v.2.4.1 [51]. Multi-sample joint variant calling was performed using the HaplotypeCaller and GenotypeGVCF tools of the GATK package v.4.0.1.2 [52]. We retained only SNP variants (excluding indels) and proceeded to hard-filtering using the GATK VariantFiltration tool based on the following cutoffs: QD <5.0; QUAL<1000.0; MQ <20.0; −2>ReadPosRankSum>2.0; −2>MQRankSum>2.0; −2>BaseQRankSum>2.0. After filtering for locus-level genotyping rate (>80%) and minor allele count (MAC) of 1 using VCFtools v.0.1.15 [53], we analysed genotyping rate at the isolate level using the --missing function of PLINK v.1.07 [54]. PLINK analyses were possible only for bi-allelic sites.

Population genetic analyses

We performed a down-sampling analysis to estimate the effect of sample size on the total number of segregating SNPs in the population. We randomly created subsets of 177 isolates from the population and counted the number of segregating SNPs in the subset. Presence–absence polymorphism of accessory chromosomes was analysed based on chromosome-level coverage using BAMstats v.1.25 (https://sourceforge.net/projects/bamstats/). We calculated the standardized coverage of all chromosomes by normalizing to the mean coverage of the core chromosomes. We categorized a chromosome as present in an isolate if the standardized coverage was within the range of 0.625–1.5. Otherwise, the chromosome was categorized as partially deleted (coverage of 0.25–0.625), absent (coverage <0.25) or duplicated (coverage >1.5). We estimated SNP density on each accessory chromosome by adjusting the '--max-missing' filter for SNPs to 80% of the total number of observed chromosomes. We used the loess model implemented in the geom_smooth function from the R package ggplot2 [55].

Linkage disequilibrium and population structure analyses

We analysed the decay in linkage disequilibrium for each chromosome separately. For this, we used all SNPs with a minor allele frequency >5% in the population. We calculated the linkage disequilibrium r^2 between marker pairs using the option –hap-r2 in VCFtools v.0.1.15 [53] with --LD-window-bp of 10 000. The decay of linkage disequilibrium with physical distance was estimated using a non-linear regression model [56, 57]. To perform PCAs, we processed the filtered SNPs using the R packages vcfR v.1.8.0 [58] and PCs were calculated using the function dudi.pca() of the R package ade4 v.1.7-16 [59]. The graph was visualized using ggplot2 v.3.1.0 [55]. We generated an unrooted phylogenetic network using SplitsTree v.4.14.6 using uncorrected *p* distances [60]. We performed a pairwise homoplasy index (Phi) test for recombination using SplitsTree v.4.14.6 [60]. File format conversions were performed using PGDSpider v.2.1.1.5 (Lischer & Excoffier, 2011). To identify groups of clonal genotypes, we calculated the pairwise genetic distances between all genotypes using the function dist.dna() included in the R package ape v.5.3 [61]. Isolate pairs with a

pairwise genetic distance below 0.01 were considered as clones for further analyses. The Simpson diversity index was calculated using the R package vegan v.2.5 [62].

We estimated effective population sizes (N_e) using a linkage disequilibrium approach. The genetic optimization for N_e estimation (GONE) approach infers the recent demographic history of a population based on contemporary individuals and dense polymorphism data [63]. The genetic algorithm searches for a sequence of historical N_e values that best explains the observed linkage disequilibrium spectrum. We defined the data as known phase because *Z. tritici is* haploid and defined the recombination rate as 50 cM Mb^{-1} approximating the average recombination rate assessed for the species using crosses [64]. We used Haldane distance corrections and performed the simulations for 2000 generations and a bin number of 400 following recommended settings [63]. The SNP dataset was subset with a minor allele frequency filter of 0.05 and contained only genetically distinct isolates (a single isolate per clone group was randomly selected). The simulations were replicated 100 times. Following recommendations, we capped the maximum permissible recombination rate (*hc*) at 0.005 to reduce artefacts from recent immigrants and physical gaps in the SNP panel [63]. We focused on core chromosome SNPs only and capped the maximum number of analysed SNPs per chromosome to 50 000. GONE was retrieved from https://github.com/esrud/GONE (version from 21 June 2021).

Spatial analyses and dispersal kernel estimation

The spatial coverage of isolate sampling and distances between the observed clonal isolate pairs were analysed. First, distances between all possible isolate pairs were calculated based on distance between plots: 2 m along the columns and 1.5 m along the rows (plot dimensions plus gaps between plots). The expected distribution of distances between isolates was calculated assuming balanced sampling of isolates from each plot. The distribution of detected clonal pairs was compared to expected dispersal distances in successive generations of asexual propagation. The dispersal distance distribution was calculated assuming an exponential dispersal location kernel parametrized with dispersal kernel estimate from Karisto *et al.* [65]. In brief, the dispersal kernel estimates were obtained by analysing an experimentally set up infection focus in the same experimental wheat field at the beginning of the growing season. The infection was started from a previously genotyped isolate. Over the course of the growing season, infected leaves in immediate proximity (metre scale) of the infection focus were genotyped to determine the presence of the isolate introduced during the experiment. The shape parameter of the exponential dispersal location kernel was estimated to be α=13.5 cm. The expected total dispersal distance distribution was calculated as convolutions of the dispersal distance kernel neglecting dispersal direction. The analyses were implemented in Python 3.7, using packages NumPy v.1.17.3 [66], SciPy v.1.3.1 [67], pandas v.0.25.3 [68] and Matplotlib v.3.2.1 [69].

RESULTS

Pathogen field population shows extensive sequence variation

A total of 177 isolates of *Z. tritici* were collected from an experimental wheat field in Switzerland planted with 335 wheat cultivars in 1.2×1.7 m plots (Table S1 [46]). The wheat field was planted with elite European winter wheat to assess cultivar performance using an automated phenotyping platform. We collected isolates at three time points during the wheat growing season representing different phases in the infection life cycle of the pathogen (Fig. 1a) [46]. The isolates from the first collection (C1) were exposed to the application of demethylation inhibitor (DMI) fungicides. Isolates from the second and third collection were exposed additionally to a succinate dehydrogenase inhibitor (SDHI) and additional DMIs (Fig. 1a). We collected isolates from 12 winter wheat cultivars (*n*=1–43 isolates per cultivar) which were planted in two replicate plots separated by ~100 m (Fig. 1b, c). Cultivar Claro was sampled on two additional plots. Collection was performed when the wheat plants were in growth stages GS 41 (C1), GS 75 (C2) and GS 85 (C3) [46, 70]. We used Illumina sequencing of whole genomes for all 177 isolates, producing 0.8–15 million reads per isolate with an average read count of 4.2 million (Fig. 1d). The isolates were sequenced to a depth of 5–77 and a mean coverage of 21.4×. We detected a total of 1496037 high-confidence SNPs. The average genotyping rate was 97.85% across loci (Fig. 1e). We detected on average 37 SNPs per kb, showing that the field population is highly polymorphic. In line with this, we found that the minor allele (i.e. the less frequent allele at a locus) frequency spectrum showed a strong skew towards rare alleles in the population, which suggests that the population did not experience any recent bottlenecks (Fig. 1f). Population bottlenecks are expected to generate an excess of high-frequency alleles. Based on the finding that our population shows substantial genetic variation, we aimed to estimate the total genetic diversity in the field. For this, we performed random down-sampling analysis and we found that subsets of 10 and 50% of the population harboured on average 763 765 and 1 234'846 segregating SNPs representing 51 and 82.5% of the total detected SNPs, respectively (Fig. 1g). We observed only a weak plateau effect for subsets close to 100% indicating that the total polymorphism in the field population is likely substantially higher. We simulated historic trends in the effective population size (N_e) using a linkage disequilibrium based approach. We found that the N_e fluctuated between 1968 and 13 067 over the past 800 generations (Fig. 1h). The simulations indicated a drop from the maximum N_e to the contemporary minimum within ~100 generations.

Accessory chromosome polymorphism

Z. tritici harbours the most extensive set of accessory chromosomes known for plant pathogens [71]. We analysed chromosome-wide read coverage to identify chromosomal presence–absence variation. All core chromosomes (1–13) were consistently detected in all isolates as expected. In contrast, we found 139 isolates (~79%) missing one or more
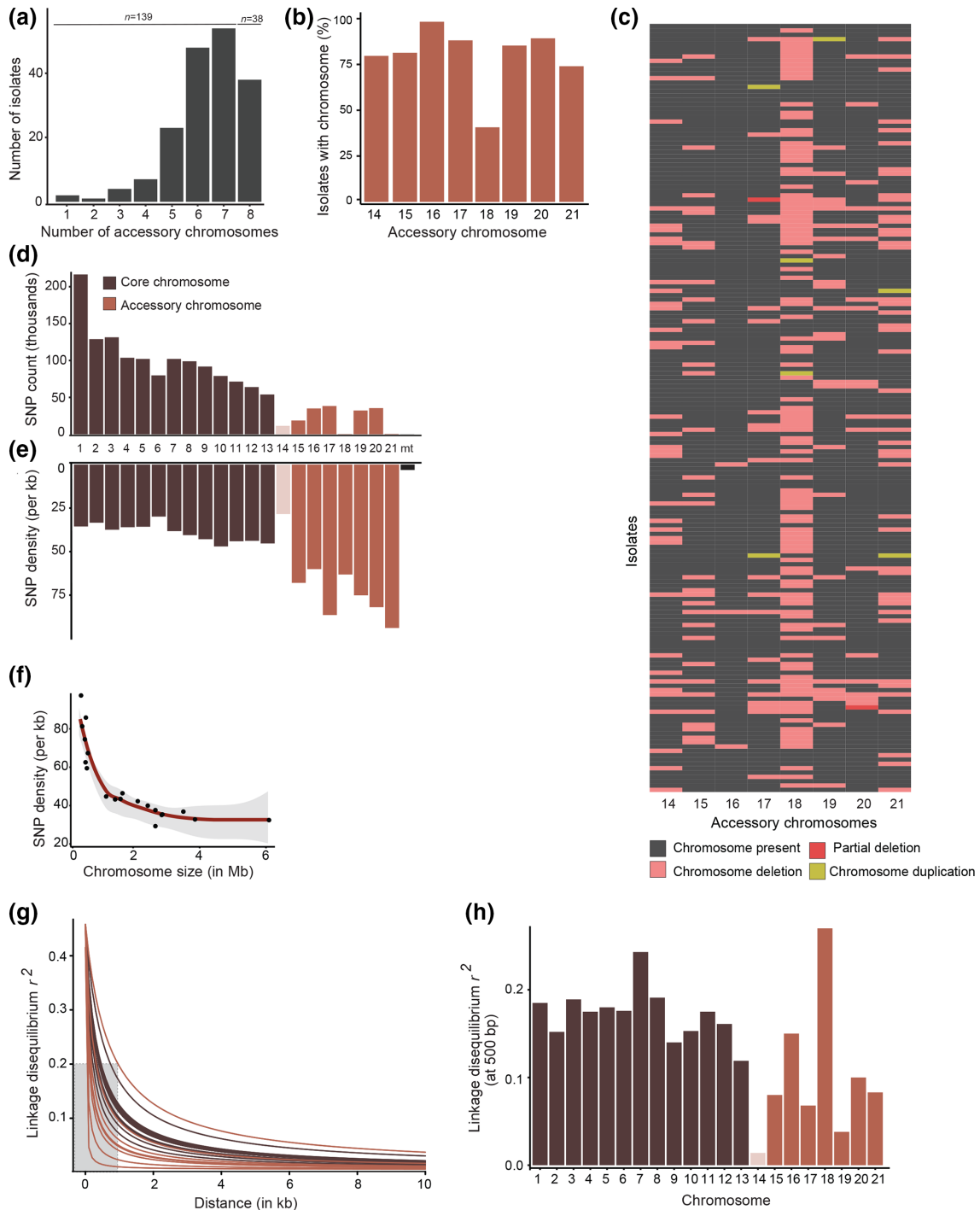
Fig. 2. Genome-wide polymorphism and accessory chromosomes. (a) Total number of accessory chromosomes per isolate. The numbers above the bars indicate the isolates missing at least one accessory chromosome or carry all chromosomes. (b) Percentage presence of all the accessory chromosomes within the population calculated based on normalized chromosome coverage. (c) Heat map showing the presence–absence variation in accessory chromosomes across isolates assessed by normalized read depth. (d) SNP count and (e) SNP density per chromosome. (f) Correlation plot of SNP density and chromosome size. (g) Pairwise linkage disequilibrium decay among all pairs of SNPs within a fixed window size of 10000 bp for each chromosome. A non-linear model was fitted using the equation of Ingvarsson [56]. The grey shading indicates the maximum distance needed for all chromosomes to reach r^2=0.2. (h) Linkage disequilibrium r^2 for each chromosome at 500 bp. Light shade: chromosome 14 was omitted from per-chromosome genetic diversity analyses due to the heterogeneous distribution of regions with robust SNP calls.

accessory chromosomes (Fig. 2a). We found no isolate that lacked all accessory chromosomes, but we identified two isolates with only a single accessory chromosome out of eight. Accessory chromosome 16 was the most frequent (174/177 isolates) and chromosome 18 was the rarest being present in fewer than half of the isolates (Fig. 2b). We identified several accessory chromosomes showing evidence for partial deletions or duplications based on normalized read coverage (Fig. S1). Specifically, we identified two partial deletions of chromosomes 17 and 20 (Fig. 2c), as well as chromosomal duplications of chromosomes 17, 18, 19 and 21 (Fig. 2c). In a next step, we analysed how polymorphism was structured among core and accessory chromosomes. We found that the accessory chromosomes show higher SNP densities but lower total SNP counts compared to core chromosomes (Fig. 2d, e). Overall, smaller chromosomes tend to have higher SNP densities (Fig. 2f). This is consistent with relaxed selection leading to accumulation of mutations on these chromosomes [72]. It is important to note, however, that short read-based SNP calling is not feasible in highly repetitive chromosomal regions. Consistent with this, we found gaps in high-confidence SNP calls along all chromosomes, but the effect was particularly pronounced on accessory chromosomes (Fig. S2). Hence, our polymorphism analyses probably underestimate SNP densities on accessory chromosomes.

Another important component of genetic variation is the extent of linkage disequilibrium. We estimated linkage disequilibrium decays for each chromosome separately and found that most accessory chromosomes showed a faster decay compared to core chromosomes (r^2=0.2 within 100 bp, Fig. 2g), except chromosomes 16 and 18 showing a decay of r^2 to 0.2 within 300 and 900 bp. Interestingly, chromosome 7 showed the slowest decay of all core chromosomes (r^2=0.2 at~700 bp, Fig. 2h). This chromosome was proposed to have originated at least partially from an accessory chromosome [73]. Chromosome 18 shows a particularly low degree of collinearity among genomes of the same species [74] (Fig. S3). It remains to be investigated whether structural variation had an impact on linkage disequilibrium assessments on accessory chromosomes. Together, we show that the accessory chromosomes accumulated more sequence variation compared to core chromosomes, but the level of conservation varies substantially among accessory chromosomes.

Population clonality in space and time

To track the evolution of genotypic diversity in space and time, we first constructed an unrooted phylogenetic network based on SplitsTree and found that most pairs of genotypes are at a similar genetic distance (Fig. 3a). The predominantely star-like shape with little reticulation is in principle consistent with a clonally reproducing population. However, the high effective population size, the rapid decay in linkage disequilibrium and the pairwise homoplasy index (Phi; P<0.0001) show strong signatures of recombination. Hence, the star-like shape is most likely explained by extensive recombination generating nearly identical distances between genotypes. Interestingly, a subset of genotypes showed much shorter

genetic distances, suggesting recent ancestry or clonal reproduction. We also performed a principal components analysis (PCA) to determine the degree of differentiation in the population (Fig. 3b). We detected no meaningful population subdivision except for seven isolates, which clustered into three groups (Fig. 3b). Interestingly, all seven isolates were collected from cultivar CH Combin (arrows in Fig. 3b). To clarify how clonal reproduction impacts the genetic structure within the field, we identified groups of clonal genotypes based on pairwise genetic distances. We found that most isolates were at a relative pairwise distance of more than 0.15, where a value of 1 corresponds to a genetic difference at every analysed SNP position (Fig. 3c). We defined isolate pairs with a genetic distance of <0.01 as being clones (Fig. 3d). We assigned groups of clones into four categories according to the distance between field plots and collection time point: clones originating from the same plot and same collection time point (category A), different plot but same collection time point (category B), same plot but different collection time point (category C), and different plot and different collection time point (category D; Fig. 4a, Table S2).

We identified a total of 15 distinct groups of clones (~8.5% of the isolates; Fig. 4b, Table S2). The proportion of clones ranged from 8.9% in first collection to 12.1% in third collection, but the difference was not statistically significant (Fisher's exact test, P=0.244; Fig. 4b). The majority of the clone groups (n=9) were sampled from the same plot and same collection time point (category A; Fig. 4c, Table S2). Interestingly, isolates collected from the cultivar CH Combin showed a high degree of clonality (Table S2). We also found that three clonal groups each were either collected from the same plot at different time points or different plots at the same time point (categories B and C; Fig. 4c). The maximum distance at which we identified clonal pairs was 20 ms (Table S2) although distances between plots ranged up to 120 m (Fig. 5a).

Estimating dispersal distances of clonal genotypes

We analysed all pairwise distances of isolate sampling locations across the field. Given the spatial arrangement of the analysed plots, we found a bimodal distribution of pairwise distances (Fig. 5a). Next, we used data collected from an asexual spore dispersal experiment conducted in the same field [65]. In this experiment, a focal infection with a known genotype was established and the radiation of the infection focus was monitored. Disease intensity decreased substantially already at the scale of 1 m (Fig. 5c). Based on this information, we compared the distribution of the detected clonal pairs against the experimentally estimated dispersal distances in successive generations of asexual propagation (Fig. 5b) [65]. We could not re-estimate the dispersal kernel of the pathogen due to low number of clonal isolates and low spatial resolution of the sampling. Comparing the distances between the observed clonal isolates to the expected dispersal distance distributions suggests that most clonal pairs are probably originating from recent asexual dispersal. However, the longest distance of 20 m between clones is very unlikely to have occurred by
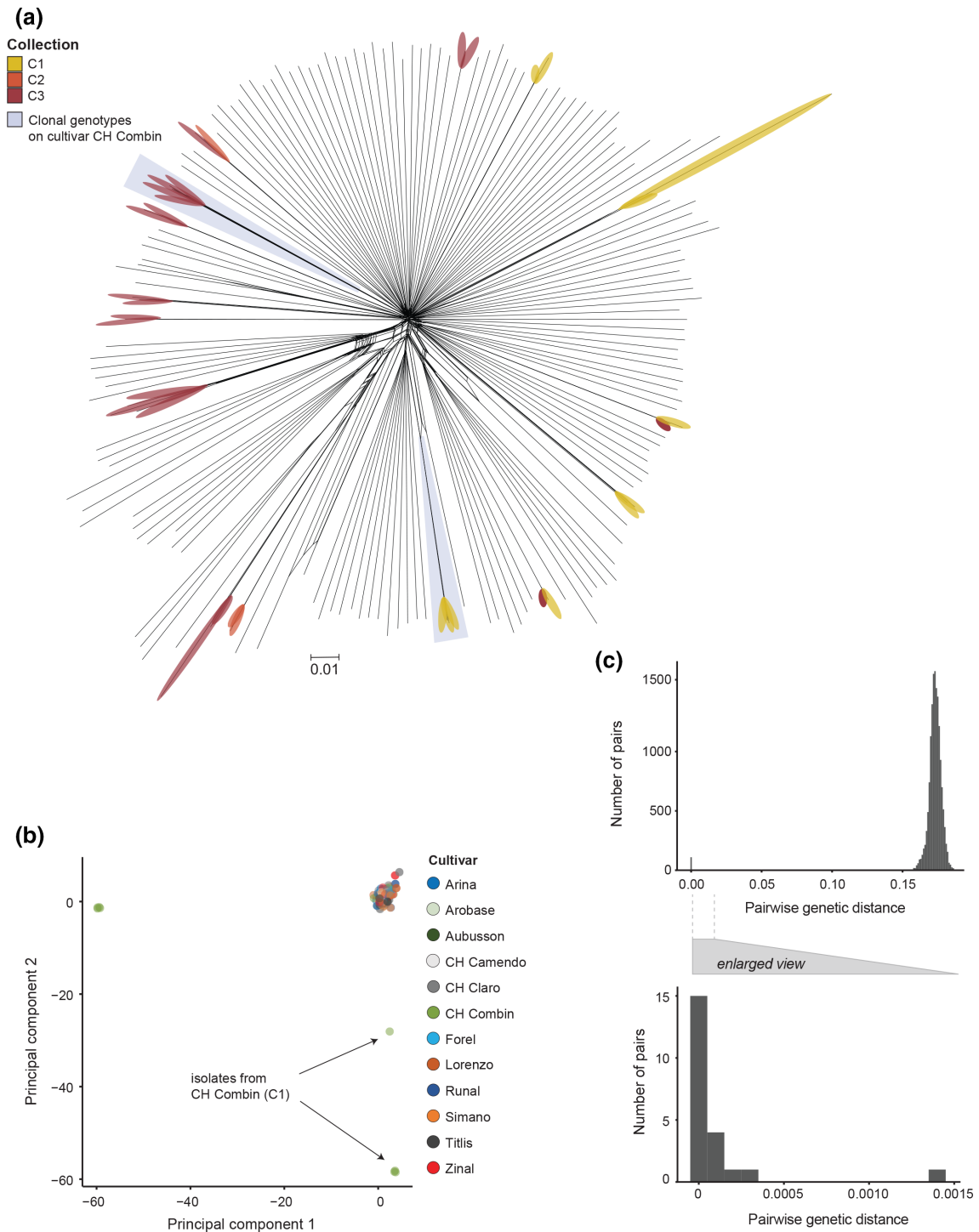
Fig. 3. Genetic differentiation and clonality within the field. (a) Phylogenetic network constructed using SplitsTree. Groups of clonal genotypes are marked with highlights identifying the collection of origin. (b) The first two principal components (PC) from a PCA. Isolates are colour coded by the cultivar of origin. (c) Histograms showing the distribution of pairwise genetic distances among genotypes. A distance of 1 corresponds to the total number of SNPs. (d) Histogram of pairwise genetic distances <0.01.
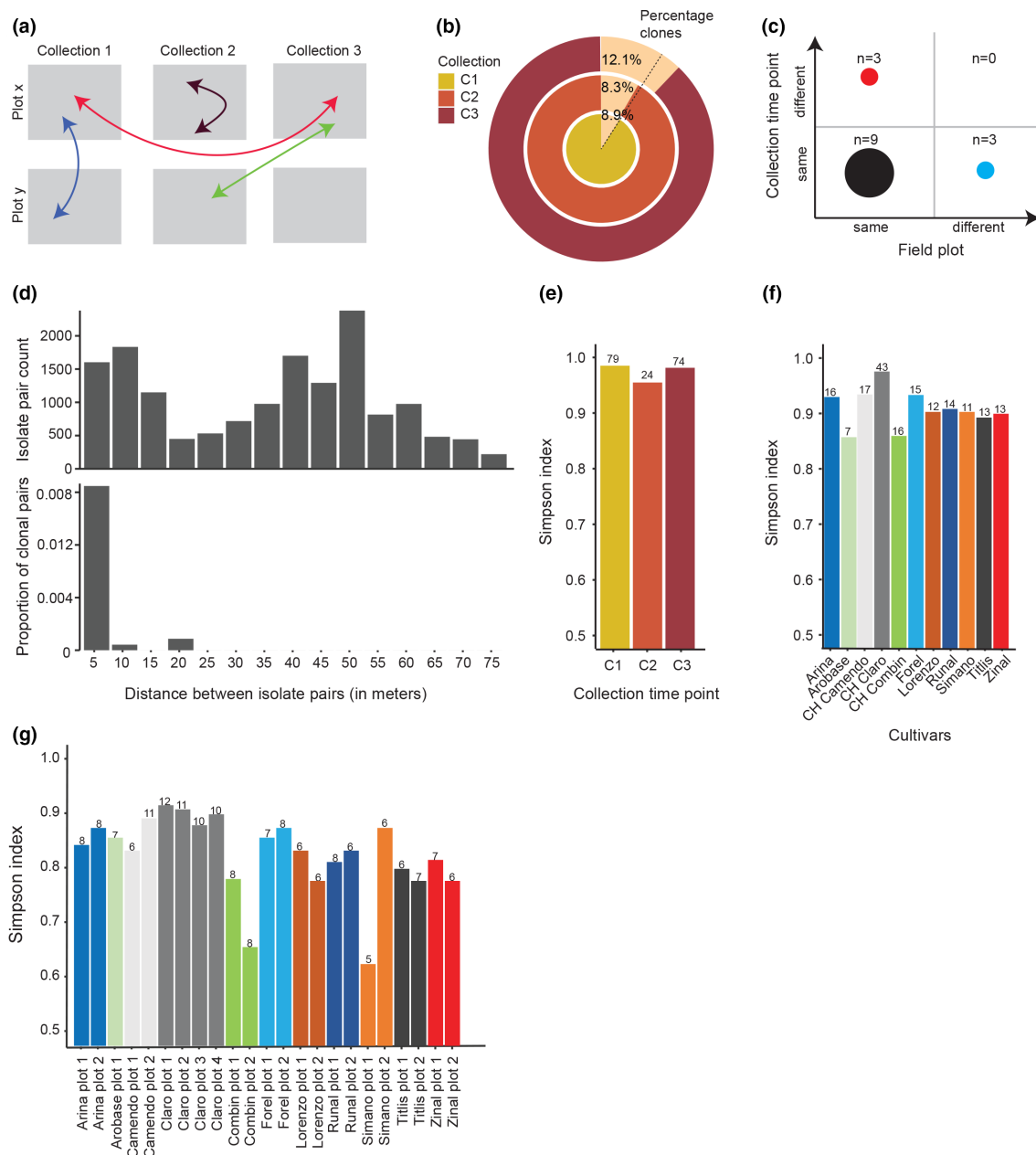
Fig. 4. Genetic diversity through space and time. (a) Schematic representation of different clone categories with pairs identified from the same or different plots and time points. (b) Pie chart showing the change in clonality over time. The dotted line indicates the average clonality of the whole population. (c) Number of clone groups categorized based on differences in space and time. The numbers indicate the clone groups in each category. (d) Analyses of all possible pairwise comparisons of isolates with the distance between collection points shown as a histogram. The proportion of clone pairs is shown as a proportion of the total number of isolate pairs. (e) Simpson diversity index between collections C1 and C3, (f) between cultivars and (g) between plots.

splash dispersal. Human or animal movements within the field are the most likely explanations.

Genotypic diversity assessments

We analysed the genotypic diversity in the field across time points using diversity indices. Overall, we found less diversity in the third collection compared to the first collection (Fig. 4e, Table S3). The second collection contained too few

isolates for comparison. Genotypic diversity was highest for cultivar Claro (0.97, Fig. 4f) constituting also the best sampled cultivar in the field. Isolates from cultivar Combin had the lowest diversity (0.86, Fig. 4f) because almost all recovered genotypes were clones. Both field plots of cultivar Combin showed low diversity with a Simpson index of 0.78 and 0.66, respectively (Fig. 4g). We did not consider the wheat cultivar Aubusson due to low sample size (*n*=1). Overall, clonal

(a)

(b)   **(c)**
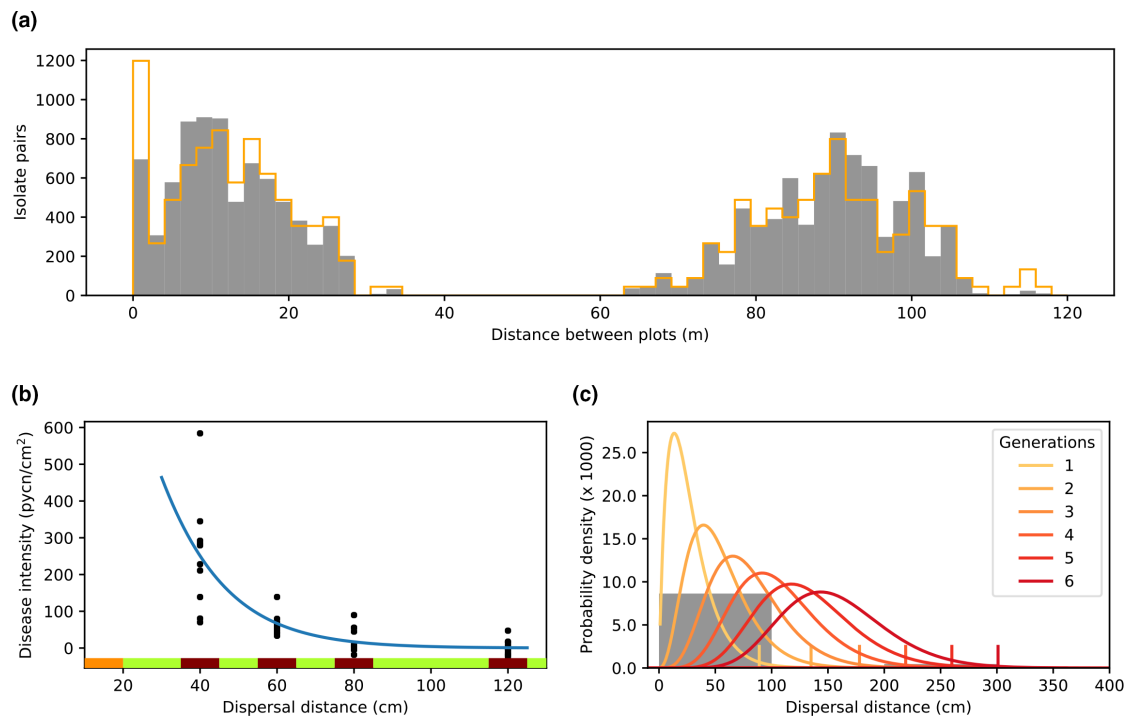

Fig. 5. Estimation of dispersal distances. (a) Observed distribution of distances between all isolates (grey) and the expected distribution assuming a uniform sampling (orange). The bimodal distribution results from the two-block design of the experiment. (b) Experimental assessment of dispersal distances based on data generated by Karisto *et al.* [65] showing the observed primary disease gradient from an infected source (orange) to measurement lines (red) at specific distances from the source. (c) Spatial distribution of clonal pairs (grey) compared to distributions of the expected dispersal distances after successive generations.

genotypes are an important constituent of the genetic structure in the field. However, the exclusive asexual reproduction during the growing season does not meaningfully affect the trajectory of genetic diversity.

Experimental analyses of virulence on cultivar Claro

Using greenhouse assays we tested whether isolates sampled from different time points or cultivars showed differences in virulence on the most frequent cultivar (CH Claro) planted in the field. For this, we analysed data generated for a genome-wide association study from a subset (*n*=120) of the isolates [45] (Table S5). We found that isolates from the third collection produced significantly more pycnidia (a proxy for pathogen reproduction) compared to earlier collection (one-way ANOVA, *P*=0.002) (Fig. 6a, b). Clonal isolates recovered from the cultivar CH Combin were found to be causing significantly less lesion damage (i.e. percentge leaf area covered by lesions; one-way ANOVA, *P*=0.003) and produced fewer pycnidia (one-way ANOVA, *P*=0.046; Fig. 6c, d). However, we found no significant difference in virulence traits between isolates sampled from CH Claro compared to other wheat cultivars (Fig. 6e, f). We found no significant difference in virulence traits between isolates found in clone groups or not (Fig. 6g, h)

DISCUSSION

The mode of reproduction and levels of genetic diversity play an important role in the rapid evolution of plant pathogens. Here we analysed a large collection of *Z. tritici* genomes obtained from a single wheat field and found that the population was highly diverse with a rapid decay in linkage disequilibrium. We also found a minor degree of clonal structure probably driven by splash dispersal among neighbouring wheat plants.

Genome-wide levels of genetic diversity in a single field

The analysed field population shows a dominant sexual reproduction regime and is highly polymorphic. Using hierarchical sampling from the same field plots and wheat cultivars across different time points, we found that ~80% of the isolates were genetically distinct. The allele frequency spectrum and the rapid decay in linkage disequilibrium suggest that there has been no obvious recent population bottleneck. Linkage disequilibrium decayed to $r^2 < 0.2$ within 1.5 kb across all chromosomes and often at a much shorter distance. Since the average distance between genes in the genome of *Z. tritici* is approximately 1 kb [50], polymorphism in any gene is expected to evolve largely independently. Hence, polymorphism in genes playing roles in adaptation
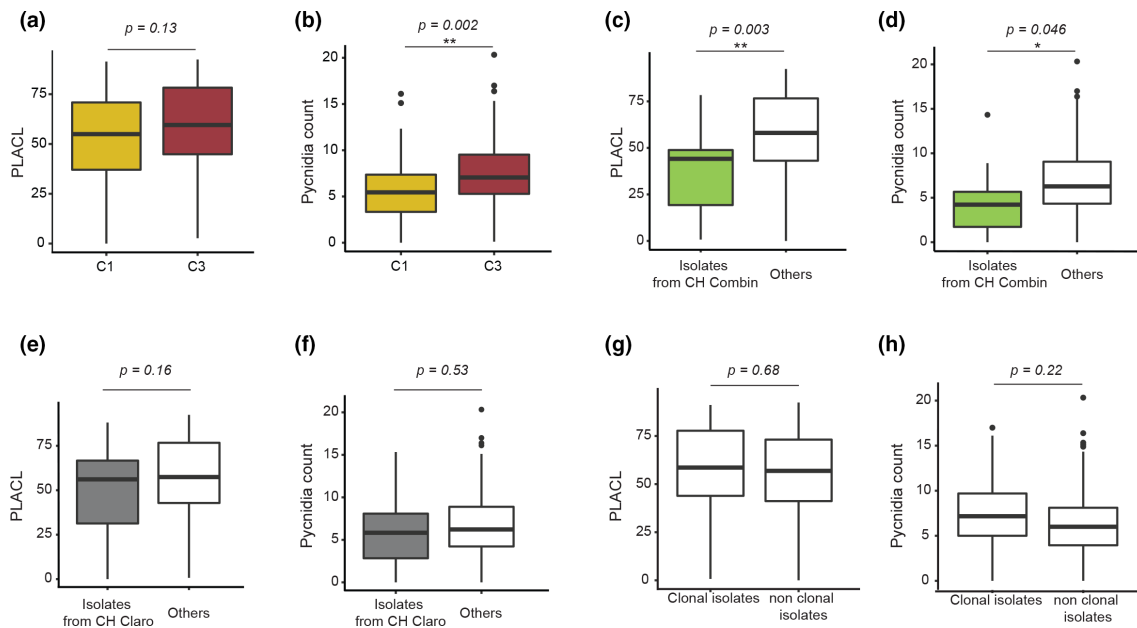
Fig. 6. Experimental assessment of virulence on the wheat cultivar CH Claro. Percentage of leaf area covered by lesions (PLACL) and pycnidia counts are shown. (a, b) Isolates from different collection time points. (c, d) Isolates originally collected from the wheat cultivar CH Combin compared to isolates collected from other cultivars. (e, f) Isolates originating the from wheat cultivar CH Claro compared to isolates collected from other cultivars. (g, h) Isolates with a genotype grouped into a clone group compared to other isolates.

to either host resistance or environmental adaptation should respond independently to selection pressures, allowing the population to adapt very rapidly [75, 76]. In addition, the effective population size estimated for the field was high with the contemporary N_e ~2000. This estimate is comparable with estimates from other field populations of *Z. tritici* around the world using a different algorithm to estimate population size from linkage disequilibrium patterns [31]. The historical population size simulations suggest that a gradual reduction in population size started approximately 100 generations ago. N_e estimates are sensitive to recent immigration generating linkage disequilibrium, and hence some of the reduction may be attributable to the genetic algorithm used here [63]. In *Z. tritici,* there is typically a single sexual generation per year. Hence, beyond effects of immigrants, the fluctuations in the field population size rather reflect historical bottleneck events at the scale of dozens to hundreds of years. Environmental drivers may be changes in agriculture [77] or climate change rather than annual bottleneck events due to the recolonization of wheat fields.

We have also shown that the rate of linkage disequilibrium decay is not the same in all chromosomes, with accessory chromosomes showing faster decays. In *Z. tritici*, accessory chromosomes are small, low in G+C and gene content, and have higher degrees of repetitive DNA [50]. In line with this, accessory chromosomes have been shown to have higher recombination rates than core chromosomes [64]. Analyses of polymorphism and linkage disequilibria may have been affected by the low degree of synteny (or collinearity) among accessory chromosomes and the high repetitive sequence

content [74] [78]. Genotyping using short read data is generally only possible in the least repetitive sections of accessory chromosomes. Substantial insertion and deletion polymorphism changes the physical distances among SNP loci, which could also have affected estimates of linkage disequilibrium. Rapid decay in linkage disequilibrium is an important property also in other rapidly evolving pathogens such as *Leptosphaeria maculans* causing blackleg disease in oilseed rape. Rapid allelic diversification of avirulence genes driven by sexual reproduction can lead to resistance break-downs of the host [79]. Similarly, the poplar rust *Melampsora larici-populina* shows signs of rapid sweeps of virulent alleles associated with population replacements [80].

Clonal genotype contributions in space and time

Sexual ascospores originating from crop residues of the previous cropping season are thought to be the source of primary infections in the field [35, 81]. As a result of this, a wheat field infected by *Z. tritici* is expected to carry essentially no clonal genotypes at its initial colonization stage. Clonal genotypes are expected to increase in frequency because of splash-dispersed asexual pycnidiospores. However, we found no significant increase in clonality even though asexual reproduction should dominate over the course of a growing season. In addition to the sampling period, we found that the likelihood of observing clonal genotypes did not meaning-fully change with the cultivar or physical distance. This means that even at the level of individual plots of ~2 m², wheat was colonized nearly entirely by genetically distinct isolates. It is important to note that we defined clonal genotypes as having

differences at less than 1% of all polymorphic sites detected among all isolates in the field. Polymorphism levels below this threshold could well be the outcome of spontaneous mutations arising within the growing season [15, 26, 82]. Sequencing errors are unlikely to explain a meaningful amount of the detected differences as SNP candidates were stringently filtered based on previous method validations [38]. However, we cannot exclude the possibility that some divergence among clone pairs is older than the onset of the wheat field colonization. The persistence of clone pairs between years would imply though that the primary inoculum of the field was not purely a set of recombined offspring genotypes generated in previous years.

The pairs of clonal isolates found at the shortest distance (0 and 2–3 m distance classes) are probably the result of splash dispersal. This contrasts with wind-dispersed sexual spores of *Z. tritici* and other pathogens travelling much longer distances. The observed distances between clone pairs were consistent with our experimental assessment of splash dispersal distances in the same field. The experiments indicated a per-generation dispersal distance of well under 1 m. Notable exceptions included the clonal genotypes associated with the cultivar CH Combin and clonal genotypes on different cultivars physically separated by as much as 20 m. Given our estimates of splash dispersal distances, human (or animal) movement within the field is a more plausible explanation for these unusual dispersal events. We also identified clone groups with isolates collected in the same plot but from different collection time points. Hence, successful genotypes can persist in individual field plots at high enough levels through time to be resampled.

It is currently unknown how adaptation to environmental factors (including host genotypes) influences the evolution of genetic diversity in infected fields. Selection due to seasonal changes can lead to short-term increases in aggressiveness at the annual scale [83], but this effect can disappear at the interannual scale [40]. Genotypes carrying beneficial alleles to overcome host immunity should increase in frequency. However, evolutionary constraints may slow down the rise of such beneficial alleles [84]. In the absence of sexual recombination during the growing season, this should lead to the expansion of successful clonal genotypes. We found overall no signature of clonal expansion except for clones associated with cultivar CH Combin. The expansion of clones on CH Combin may indeed be the result of highly successful genotypes rising in frequency. Host damage (lesions) caused by these isolates, however, is lower compared to other isolates when tested on a different cultivar. Neutral processes such as population bottlenecks could well underlie the observed clonal expansion associated with the cultivar CH Combin. Consistent with selection on (asexual) pathogen reproduction over the course of the growing season is the significantly increased pycnidia production of isolates collected at C3 compared to C1. Matching changes in genotype frequencies with changes in frequencies of beneficial alleles for host colonization over the course of a growing season will help to disentangle neutral from selective processes.

The high degree of genotypic diversity originating from sexual reproduction is consistent with recent studies based on microsatellite markers showing spatial genetic structure but no meaningful loss in diversity over consecutive growing seasons [85, 86]. Uniformity in host genotypes in agricultural ecosystems enables clonal pathogen populations to establish infections [87]. However, the introduction of new host resistance genes can trigger the local extinction of such pathogen populations [88]. Similarly, the application of fungicides can lead to local extinctions of clonal populations if populations lack resistance mutations. In contrast, in sexually reproducing populations recombination produces a multitude of new genotypes, increasing the likelihood for adaptive traits to evolve [89]. Recent work on *Z. tritici* shows that pathogen populations maintain substantial polymorphism underlying fungicide and virulence traits despite years of directional selection [84]. The maintenance of such adaptive genetic variation is probably because of trade-offs among traits. This may also explain why no 'super genotypes' arise that are virulent on all cultivars or are fully resistant to fungicides. Managing virulence and fungicide resistance emergence in pathogens remains a pressing issue to ensure sustainable agricultural production [90]. Hence, understanding how pathogen populations respond in the short term (i.e. over the scale of weeks during the growing season) to challenges becomes a critical area of investigation. Detecting changes in pathogen populations *in situ* may become a powerful approach to identify the emergence of previously unknown adaptive mutations, changes in reproductive modes or the introduction of foreign genotypes. Deep population sequencing approaches revealing rapid changes in allele frequencies and population structure will become key tools in such endeavours.

References

1. **Gandon S**, **Hochberg ME**, **Holt RD**, **Day T**. What limits the evolutionary emergence of pathogens? Philosophical transactions of the Royal Society B: biological sciences 2013;368:20120086.

2. **Read AF**, **Taylor LH**. The ecology of genetically diverse infections. *Science* 2001;292:1099–1102.

3. **Thrall PH**, **Burdon JJ**. Evolution of virulence in a plant host-pathogen metapopulation. *Science* 2003;299:1735–1737.

4. **Vignuzzi M**, **Stone JK**, **Arnold JJ**, **Cameron CE**, **Andino R**. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006;439:344–348.

5. **Hamelin FM**, **Castel M**, **Poggi S**, **Andrivon D**, **Mailleret L**. Seasonality and the evolutionary divergence of plant parasites. *Ecology* 2011;92:2159–2166.

6. Linde CC. Population genetic analyses of plant pathogens: new challenges and opportunities. *Australasian Plant Pathology* 2010;39:23–28.

7. Walther BA, Ewald PW. Pathogen survival in the external environment and the evolution of virulence. *Biol Rev Camb Philos Soc* 2004;79:849–869.

8. He DC, Zhan JS, Xie LH. Problems, challenges and future of plant disease management: from an ecological point of view. *J Integr Agric* 2016;15:705–715.

9. Sundin GW, Castiblanco LF, Yuan X, Zeng Q, Yang C-H. Bacterial disease management: challenges, experience, innovation and future prospects: challenges in bacterial molecular plant pathology. *Mol Plant Pathol* 2016;17:1506–1518.

10. Laine A.-L., Barrès B. Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. *Plant Pathol* 2013;62:96–105.

11. Mariette N, Mabon R, Corbière R, Boulard F, Glais I et al. Phenotypic and genotypic changes in French populations of *Phytophthora infestans* : are invasive clones the most aggressive? *Plant Pathol* 2016;65:577–586.

12. Becheler R, Xhaard C, Klein EK, Hayden KJ, Frey P, et al. Genetic signatures of a range expansion in natura: when clones play leapfrog. *Ecol Evol* 2016;6:6625–6632.

13. Tack AJM, Thrall PH, Barrett LG, Burdon JJ, Laine AL. Variation in infectivity and aggressiveness in space and time in wild host-pathogen systems: Causes and consequences. *J Evol Biol* 2012;25:1918–1936.

14. Barrett LG, Thrall PH, Burdon JJ, Linde CC. Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends Ecol Evol (Amst)* 2008;23:678–685.

15. Burdon JJ, Zhan J, Barrett LG, Papaix J, Thrall PH. Addressing the challenges of pathogen evolution on the world's arable crops. *Phytopathology* 2016;106:1117–1127.

16. Suffert F, Delestre G, Gélisse S. Sexual reproduction in the fungal foliar pathogen *Zymoseptoria tritici* is driven by antagonistic density dependence mechanisms. *Microb Ecol* 2019;77:110–123.

17. Dodds P, Thrall P. Recognition events and host-pathogen co-evolution in gene-for-gene resistance to flax rust. *Funct Plant Biol* 2009;36:395–408.

18. Jones JDG, Dangl JL. The plant immune system. *Nature* 2006;444:323–329.

19. Nomura K, DebRoy S, Lee YH, Pumplin N, Jones J, et al. A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science* 2006;313:220–223.

20. Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF. Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. *Plant Cell* 2006;18:764–779.

21. De Torres M, Mansfield JW, Grabov N, Brown IR, Ammouneh H, *et al*. *Pseudomonas syringae* effector AvrPtoB suppresses basal defence in Arabidopsis. *Plant J* 2006;47:368–382.

22. Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* 2006;124:803–814.

23. Steinberg G. Cell biology of *Zymoseptoria tritici*: Pathogen cell organization and wheat infection. *Fungal Genet Biol* 2015;79:17–23.

24. Parratt SR, Numminen E, Laine A-L. Infectious disease dynamics in heterogeneous landscapes. *Annu Rev Ecol Evol Syst* 2016;47:283–306.

25. Peay KG, Bruns TD. Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant-fungal interactions. *New Phytol* 2014;204:180–191.

26. Susi H, Burdon JJ, Thrall PH, Nemri A, Barrett LG. Genetic analysis reveals long-standing population differentiation and high diversity in the rust pathogen melampsora lini. *PLoS Pathog* 2020;16:e1008731.

27. Billiard S, López-Villavicencio M, Hood ME, Giraud T. Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. *J Evol Biol* 2012;25:1020–1038.

28. McDonald BA. The population genetics of fungi: tools and techniques. *Phytopathology* 1997;87:448–453.

29. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* 2002;40:349–379.

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

31. Hartmann FE, Croll D. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Mol Biol Evol* 2017;34:2808–2822.

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

33. Krishnan P, Ma X, McDonald BA, Brunner PC. Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. *BMC Evol Biol* 2018;18:7.

34. Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, *et al*. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. *New Phytol* 2017;214:619–631.

35. Chen R-S, Mcdonald BA. Sexual reproduction plays a major role in the genetic structure of populations of the fungus *Mycosphaerella graminicola* 1996.

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

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

38. Hartmann FE, McDonald BA, Croll D. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Mol Ecol* 2018;27:2725–2741.

39. McDonald MC, Renkin M, Spackman M, Orchard B, Croll D, *et al*. Rapid parallel evolution of azole fungicide resistance in Australian populations of the wheat pathogen *Zymoseptoria tritici*. *Appl Environ Microbiol* 2019;85:e01908-18.

40. Suffert F, Goyeau H, Sache I, Carpentier F, Gélisse S, *et al*. Epidemiological trade-off between intra- and interannual scales in the evolution of aggressiveness in a local plant pathogen population. *Evol Appl* 2018;11:768–780.

41. Morais D, Duplaix C, Sache I, Laval V, Suffert F *et al*. Overall stability in the genetic structure of a *Zymoseptoria tritici* population from epidemic to interepidemic stages at a small spatial scale. *Eur J Plant Pathol* 2019;154:423–436.

42. Linde CC, Zhan J, McDonald BA. Population structure of *Mycosphaerella graminicola*: From lesions to continents. *Phytopathology* 2002;92:946–955.

43. Kirchgessner N, Liebisch F, Yu K, Pfeifer J, Friedli M *et al*. The ETH field phenotyping platform FIP: a cable-suspended multi-sensor system. *Functional Plant Biol*. 2017;44:154–168.

44. Levy L, Courvoisier N, Rechsteiner S, Herrera J, Brabant C. Winterweizen: Bilanz AUS 15 Jahren Sortenprüfung unter extensiven Anbaubedingungen. *Agrarforschung Schweiz* 2017;8:300–309.

45. Singh NK, Badet T, Abraham L, Croll D. Rapid sequence evolution driven by transposable elements at a virulence locus in a fungal wheat pathogen. *BMC Genomics* 2021;22:393.

46. Karisto P, Hund A, Yu K, Anderegg J, Walter A, *et al*. Ranking quantitative resistance to septoria tritici blotch in elite wheat cultivars using automated image analysis. *Phytopathology* 2018;108:568–581.

47. Oggenfuss U, Badet T, Wicker T, Hartmann FE, Singh NK. A population-level invasion by transposable. *bioRxiv* 2020.

48. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. www.bioinformatics.babraham.ac.uk/projects/fastqc

49. Bolger AM, Lohse M, Usadel B. *Trimmomatic: a Flexible Trimmer for Illumina Sequence Data*. Bioinformatics (Oxford, England), 2014, pp. 2114–2120.

50. Goodwin SB, Ben MS, Dhillon B, Wittenberg AHJ, Crane CF, *et al*. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis (HS Malik, Ed.). *PLoS Genet* 2011;7:e1002070.

51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.

52. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.

53. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al*. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–2158.

54. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al*. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.

55. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016.

56. Ingvarsson PK. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (Populus tremula L., salicaceae). *Genetics* 2005;169:945–953.

57. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, *et al*. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A* 2001;98:11479–11484.

58. Knaus BJ, Grünwald NJ. *VCFR: A Package to Manipulate and Visualize Variant Call Format Data in R*. John Wiley & Sons, Ltd, 2017, pp. 44–53.

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

60. Huson DH. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* 1998;14:68–73.

61. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528.

62. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P. *Package 'vegan' Title Community Ecology Package Version 2.5-6*; 2019.

63. Santiago E, Novo I, Pardiñas AF, Saura M, Wang J, *et al*. Recent demographic history inferred by high-resolution analysis of linkage *Disequilibrium*. *Mol Biol Evol* 2020;37:3642–3653.

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

65. Karisto P, Suffert F, Mikaberidze A. Measuring splash-dispersal of a major wheat pathogen in the field. PhytoFrontiers™ 2021.

66. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, *et al*. Array programming with NumPy. *Nature* 2020;585:357–362.

67. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, *et al*. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–272.

68. Mckinney W. Data structures for statistical computing in python 2010.

69. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci* Eng 2007;9:90–95.

70. Meier U. Growth stages of mono-and dicotyledonous plants BBCH monograph edited by Uwe Meier federal biological research centre for agriculture and forestry 2001.

71. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS* Pathog 2012;8:e1002608.

72. Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, *et al*. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola* (HS Malik, Ed.). *PLoS Genet* 2010;6:e1001189.

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

74. Badet T, Fouché S, Hartmann FE, Croll D. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. *bioRxiv* 2020:2020.

75. Barton NH. Genetic linkage and natural selection. Philosophical transactions of the Royal Society B: biological sciences 2010;365:2559–2569.

76. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 2008;9:477–485.

77. Stukenbrock EH, McDonald BA. The origins of plant pathogens in agro-ecosystems. *Annu Rev* Phytopathol 2008;46:75–100.

78. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. *BMC Biol* 2020;18:12.

79. Daverdin G, Rouxel T, Gout L, Aubertot JN, Fudal I, *et al*. Genome Structure and Reproductive Behaviour Influence the Evolutionary Potential of a Fungal Phytopathogen (A Andrianopoulos, Ed.). *PLoS Pathog* 2012;8:e1003020.

80. Persoons A, Hayden KJ, Fabre B, Frey P, De Mita S, *et al*. The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Mol Ecol* 2017;26:1902–1918.

81. Kerdraon L, Laval V, Suffert F. Microbiomes and pathogen survival in crop residues, an ecotone between plant and soil. *Phytobiomes J* 2019;3:246–255.

82. Barrès B, Carlier J, Seguin M, Fenouillet C, Cilas C, *et al*. Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity (Edinb)* 2012;109:269–279.

83. Suffert F, Ravigné V, Sachec I. Seasonal changes drive short-term selection for fitness traits in the wheat pathogen *Zymoseptoria tritici*. *Appl Environ Microbiol* 2015;81:6367–6379.

84. Dutta A, Hartmann FE, Francisco CS, McDonald BA, Croll D. Mapping the adaptive landscape of a major agricultural pathogen reveals evolutionary constraints across heterogeneous environments. *ISME J* 2021:1–18.

85. Hassine M, Siah A, Hellin P, Cadalen T, Halama P, *et al*. Sexual reproduction of *Zymoseptoria tritici* on durum wheat in Tunisia revealed by presence of airborne inoculum, fruiting bodies and high levels of genetic diversity. *Fungal Biol* 2019;123:763–772.

86. Siah A, Bomble M, Tisserant B, Cadalen T, Holvoet M, *et al*. Genetic structure of *Zymoseptoria tritici* in northern France at region, field, plant, and leaf layer scales. *Phytopathology* 2018;108:1114–1123.

87. Drenth A, McTaggart AR, Wingfield BD. Fungal clones win the battle, but recombination wins the war. *IMA Fungus* 2019;10:18.

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

89. Burdon JJ, Silk J. Sources and patterns of diversity in plant-pathogenic fungi. *Phytopathology* 1997;87:664–669.

90. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, *et al*. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 2012;484:186–194.