Recurrent chromosome destabilization through repeat-mediated rearrangements in a fungal pathogen

Simone Fouché¹,², Ursula Oggenfuss¹,⁵, Bruce A. McDonald², Daniel Croll¹,*

¹ Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000, Neuchâtel, Switzerland
² Plant Pathology, Institute of Integrative Biology, ETH Zurich, CH-8092 Zurich, Switzerland

* Current address: Department of Microbiology and Immunology, University of Minnesota Medical School, Minneapolis, Minnesota, United States of America

Correspondence: daniel.croll@unine.ch

Data availability:
The genome assembly and annotation for 1A5, 1E4 and 3D7 genome are available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession numbers PRJEB15648, PRJEB20900 and PRJEB14341. Progeny genomes are available under the project PRJNA645795.
Abstract

Genomic instability caused by chromosomal rearrangements has severe consequences for organismal
fitness and progression of cancerous cell lines. The triggers of destabilized chromosomes remain poorly
understood but are often assumed to be associated with fragile sites. Here, we retrace a runaway
chromosomal degeneration process observed in a fungal pathogen using telomere-to-telomere
assemblies across an experimental pedigree. We show that the same fragile sites triggered reproducible,
large-scale rearrangements through non-allelic recombination. Across the four-generation pedigree,
chromosomal rearrangements were accompanied by non-disjunction events and caused aneuploid
progeny to carry up to four chromosomal copies. We identify a specific transposable element as the
most likely trigger for the repeated chromosomal degeneration. The element is associated with higher
virulence of the pathogen and has undergone a burst increasing copy numbers across the genome.
Chromosome sequences are also targeted by a genome defense mechanism active on duplicated
sequences, which may contribute to decay. Our work identifies the exact sequence triggers initiating
chromosome instability and perpetuating degenerative cycles. Dissecting proximate causes leading to
run-away chromosomal degeneration expands our understanding of chromosomal evolution beyond
cancer lines.
Introduction

Meiosis is a highly conserved process in eukaryotes, whereby homologous chromosomes pair, undergo recombination and separate into daughter cells. Aberrations during the faithful transmission of chromosomes through meiosis can have serious consequences for an organism. Non-disjunction events resulting in additional or fewer chromosomal copies occur frequently in humans when compared to Saccharomyces cerevisiae and Drosophila melanogaster, and is the leading cause of miscarriage in humans. Similarly, non-disjunction and chromosome rearrangements occur in somatic lines leading to genomic instability, which is a hallmark of cancers. Factors that cause fragility of pre-cancerous genomes or trigger meiotic errors are largely unknown though. Chromosomal breakage is a major factor contributing to instability and occurs often at specific loci referred to as fragile sites. Though the location of many mitotic fragile sites has been identified in the human genome, how fragile sites contribute mechanistically to breakage remains poorly understood. Fragile sites for non-allelic homologous or ectopic recombination during meiosis remain largely unknown, possibly because major rearrangements tend to be lethal. Investigations of non-lethal rearrangements offer a promising approach to unravel sequence determinants of fragile sites.

Non-lethal chromosome rearrangements are highly prolific in plant and fungal accessory chromosomes. These chromosome (also called supernumerary or B) are present in a karyotype in addition to the regular chromosomes and show ample presence/absence as well as structural variation within species. The exact mechanisms that generate chromosome rearrangements and whether fragile sites are involved, remains elusive. For example the mini-chromosomes of Magnaporthe oryzae show ample rearrangements and frequent interchromosome translocations between core and mini-chromosomes. Mini-chromosomes are hypothesized to originate from core chromosomes through rearrangements involving the terminal regions of core chromosomes. Zymoseptoria tritici exhibits some of the most extreme degrees of structural variation observed within fungal species, including large differences in terms of chromosome length, transposable element (TE) and gene content, and recombination rate, as well as telomere and centromere structure. Chromosome rearrangements occur frequently during meiosis with higher rates of rearrangements observed in the accessory chromosomes.

A rearranged accessory chromosome was identified in Z. tritici, a haploid fungal pathogen of wheat, but the mechanism whereby this chromosome variant was generated could not be identified. We aim to identify how the rearranged chromosome was generated and whether further rearrangements will occur through subsequent rounds of meiosis. We also analyzed evidence for specific triggers of the chromosome rearrangement and degenerative cycles by identifying sequence features overlapping with breakpoints.
Results

Origin of a complex chromosome rearrangement

The rearrangement of the *Z. tritici* accessory chromosome 17 was first discovered in two F1 progeny identified as A66.2 and A2.2, respectively (fig. 1A). To assess how frequently rearrangements of chromosome 17 occur through meiosis, we screened 48 progenies from the same 1A5x1E4 cross (with and without chromosome 17, respectively) for rearranged chromosome 17 with a segment-specific PCR assay and found that anomalies of chromosome 17 were most likely restricted to progeny A66.2 and A2.2 (fig. 1A, supplementary fig. S1). We screened a further 228 progeny from this cross and found four progeny with rearranged chromosome 17, of which only A66.2 and A2.2 carried the same rearrangement profile. Additionally, we assessed the frequency of rearranged chromosome 17 variants using a whole-genome sequence coverage-based approach in 150 isolates from Switzerland matching the sampling location of the parental isolates. Based on read coverage variation, the population carries likely disomic variants of chromosome 17 or partially duplicated variants (*n* = 3 isolates). The enlarged nature of chromosome 17 in A66.2 and A2.2 was confirmed by pulsed-field gel electrophoresis (PFGE) as well as southern hybridization (fig. 1B).
Figure 1: Progeny pedigree reveals enlarged chromosome 17. (A) Chromosome 17 pedigree of four rounds of meiosis. The colors indicate whether the isolate carries either a parental, unrelated (not carried by a parent), rearranged or no chromosome 17 according to pulsed-field gel electrophoresis (PFGE) analysis. (B) PFGE of the parental chromosomes (1E4 and 1A5) and the progeny A2.2 and A66.2, showing the enlarged chromosome 17 with orange arrows. (C) The coverage and breakpoints of the progeny A66.2 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, >300x coverage were removed). Vertical dashed lines indicate the chromosomal breakpoints identified from mapped reads (at positions B and E, turquoise and continuing in grey). Solid vertical lines indicate the positions of loci amplified by PCR (grey: amplification; red: no amplification). Dotplot of the assembled chromosome for progeny A66.2 compared to the parental chromosome. Inverted regions are indicated in blue. (i) and (ii) show the location and continuation of split reads and below is a schematic of the resulting enlarged chromosome in the progeny. (D) Schematic representation of the breakpoints and rearrangements between the two chromatids of 1A5.
that generated the enlarged chromosome 17 (0-ABCDEA-0) recovered in the progeny A66.2. The hypothetical smaller rearrangement product (FBCDEF) was not observed.

To investigate at base-pair resolution how the enlarged chromosome 17 was generated in the progeny A66.2 and A2.2, we sequenced the isolates with long-read PacBio technology. The recency of the duplication creates a significant challenge for the chromosomal assembly as duplicated regions are not expected to show sequence divergence and therefore could erroneously suggest the presence of only one copy. Therefore, we used a combination of coverage and split long-read mapping approaches. First, we mapped reads for the progeny to the 1A5 parent chromosomes and identified regions of chromosome 17 with higher than the mean coverage of the core chromosomes (1-13). We found sharp transitions in read coverage along chromosome 17 (fig. 1C). We found that region 0-B in progeny A66.2 and A2.2 had approximately double the mean coverage of the core chromosomes, implying that this region was duplicated (fig. 1C, supplementary fig. S2 and S3, supplementary table S1). Further supporting evidence included a breakpoint at the end of segment 0-B, with some reads showing split mapping in two distinct regions of the chromosome (fig. 1C, supplementary fig. S2, supplementary table S2). Reads mapping at this position were evenly split between canonically mapping across positions B-C and split reads mapping from A-B and continue in an inverted orientation from position E towards D (fig. 1C, supplementary fig. S2). When combined, this indicates that the duplicated region (0-B) exists twice in the progeny genome, and that the duplicated sequences are connected to two distinct locations of the chromosome at C and E (fig. 1C, supplementary fig. S2 and S3). Based on coverage, the region between C and E is single-copy. A lack of coverage after position E towards F in the progeny, suggesting that this regions is absent in the rearranged chromosome, is confirmed by the failure of a PCR to amplify in this region (fig. 1B, supplementary fig. S2-3). A lack of coverage between B and C indicates that this region is also missing. Using information on coverage and reads spanning distinct chromosomal regions, the chromosome 17 of progeny A66.2 and A2.2 is composed of region 0-B, followed by region C-E, and a second copy of region 0-B in an inverted orientation (fig. 1C; supplementary fig. S2, supplementary table S2). We identified the putative centromere to be located at position 405,779-415,898 bp based on sequence homology to chromosome 17 in the reference genome IPO323.

Therefore, the resulting enlarged chromosome 17 is expected to carry a single copy of the centromeric region (fig. 1C). Both progenies most likely inherited an identical, enlarged chromosome 17 generated during the first round of meiosis through the same non-allelic recombination between sister chromatids at locations B and E (fig 1D). The other putative product of this event was not recovered, despite being predicted to have a centromere. We reconstructed the sequence of the rearranged chromosome 17 based on the above links using sequences of the parental chromosome 17 from 1A5 (the only parent carrying the chromosome). The enlarged chromosome 17 was 819 kb in length for A66.2 and A2.2, matching the length identified from a PFGE gel separating chromosomes by length (fig. 1B).
Figure 2: Sequence of chromosome 17 rearrangements tracked through four rounds of meiosis. Chromosome rearrangements are shown according to breakpoint positions A-F identified through split read mapping. Arrows linking different letters A-F indicate translocations and deletions for each successive round of
meiosis. Panels on the right show schematic representations of each chromosome 17 variant present in each progeny. If multiple variants were detected, variants are labelled V1-V4.

**Sustained chromosome degeneration in subsequent rounds of meiosis**

To investigate whether the enlarged chromosome 17 experiences instability in further rounds of meiosis, we traced the novel enlarged chromosome through further generations by performing several backcrosses (fig. 1A). New variants of chromosome 17 were already generated in the following round of meiosis (fig. 2B, 3B, supplementary fig. S3). Chromosome 17 of Ztprog11 carried a duplicated region 0-A and the second copy of 0-A is joined to position E in an inverted orientation (fig. 2B, 3B, supplementary fig. S3, S4). Ztprog01 has two copies of chromosome 17 including a small and a large variant consistent with non-disjunction. One copy lacked the segment between B and D (fig 2B, fig 3B, supplementary fig. S3, S5). During the third round of meiosis, ZtProg19 stably inherited a chromosome 17 variant (fig. 2D, fig 3C, supplementary fig. S3, S7). In contrast, based on pedigree and read mapping evidence, the second progeny ZtProg45 carried multiple variants. Variant V2 may have undergone non-disjunction (2xV1 is inherited with 1xV4) or one copy of V1 is present with one copy of V2 and V3 each, indicating rearrangements at breakpoints B and D as well as non-disjunction (fig. 2C, fig. 3C, supplementary fig. S3, S6). Therefore, two small variants are present together with one large variant.

In the last round of meiosis the region between B and D was deleted in Ztprog8 and Ztprog64 (fig. 2D, fig 3D and ii, supplementary fig. S3, S8, S9). Ztprog64 carries four chromosome 17 variants (three small and one large variant) missing this region (fig. 2D, fig 3i and ii, supplementary fig. S3, S9). Ztprog9 most likely inherited the enlarged chromosome 17 without further rearrangements from the previous generation (fig. 2D, fig 3D, supplementary fig. S3, S10). Ztprog30 has the same chromosome complement as Ztprog01 (fig. 2D, fig 3E, supplementary fig. S3, S11). The enlarged chromosome 17 was highly unstable through further rounds of meiosis and degenerative cycles occurred via non-allelic recombination as well as non-disjunction. Rearrangements were triggered recurrently at known breakpoints with the exception of breakpoint A being involved only once.
Figure 3: Reconstructed chromosome 17 variants based on sequence rearrangement breakpoints. For each progeny the mapped read coverage is shown relative to the 1A5 reference genome sequence. Black horizontal dashed lines indicate the mean coverage of the core chromosomes and in thin grey two, three and four times the mean core chromosome coverage (i.e. coverage in duplicated regions). Red dots indicate the mean coverage in 1
kb windows. Vertical dashed lines indicate the chromosomal breakpoints A-F (see fig. 1 and 2). Solid vertical lines indicate the positions of loci amplified by PCR (grey: amplification; red: no amplification). Below each coverage plot, arrows show variants reconstructed for each progeny. If multiple variants were detected, variants are labelled with V1-V4. Variant labels are independent between progeny. Arrows are colored based on the level of duplication (absent, single-copy, multiple-copy regions). The pedigree is represented by (A) the progeny of the first meiotic round, (B) second, (C) third and (D-E) fourth round.

Figure 4: Characterization of repeat-induced point mutation (RIP) acting on duplicated sequences. (A) Coverage plot with read coverage along the chromosome 17 and evidence for a duplicated region 0-B (0-A in Ztprog11). The upper schematic shows how RIP mutations are introduced into duplication regions causing an excess of A and T with RIP acting on Gs or Cs. The lower schematic shows how mapped PacBio reads in duplicated regions were used for SNP calling, revealing potentially RIP-related mutations. (B) Overview of progeny over generations with barplots summarizing the number of RIP-like and other mutations detected per progeny generation in region 0-B (0-A in Ztprog11). Dotplots show the percentage of reads carrying the alternative allele in reads mapped to the reference chromosome 17 (parent 1A5). The identified copy numbers of the region...
Degeneration through a mutation mechanism targeting duplicated sequences

Besides degeneration via rearrangements and non-disjunction, the duplicated region (0-B and 0-A for Ztprog11) of chromosome 17 was also targeted by repeat-induced point (RIP) mutations (fig. 4A). We mapped reads for each progeny to the parental chromosome 17 (isolate 1A5) and identified regions that were targeted by RIP-like mutations (transitions from G or a C to an A or a T; fig. 4B region (0-B) and supplementary fig. S4 (other regions)). We found that most of the mutations localized to the region 0-B were RIP-like. When combining total evidence, we identified multiple copies of the region targeted by RIP-like mutations. Progeny Ztprog08 lacking any duplicated regions also showed no evidence for RIP being active. Interestingly, RIP-like mutations never targeted the putative centromere (405,779-415,898 bp; supplementary fig. S4). RIP-like mutations predominantly target CpA dinucleotides (TpG in the reverse complement) in other model fungi \(^{16}\), and this is also the case in Z. tritici (fig. 3C). Additionally, CpG dinucleotides were frequently targeted. New RIP-like mutations were found after each generation, suggesting that RIP was activated over multiple rounds of meiosis and did not reach saturation. The highest number of RIP mutations was found in Ztprog11 and Ztprog19 (generations 2 and 4) with each carrying two copies of region 0-B on one chromosome. Additionally, both progenies were generated by a backcross with 1A5. Our results show that in addition to rearrangements and non-disjunction, RIP is a mutational mechanism driving degenerative cycles in fungi.
Figure 5: Recent expansion of the *Styx* element underpinning the chromosomal rearrangements. (A) The length and location of coding regions of the long and short copies of *Styx*. Dotplot of the consensus sequence showing duplicated regions (red circle). (B) Copy number, GC-content and length distribution of *Styx* copies recovered from a global panel of *Z. tritici* reference genomes. Dark grey shows genomes of sister species *Z.*
ardabiliae (Za17), Z. passerinii (Zpa63), Z. brevis (Zb87) and Z. pseudotritici (Zp13). (C) Phylogeny of Styx copies in Z. tritici with Z. pseudotritici as an outgroup. The colors scale indicates the GC content. The two full-length copies found on chromosome 17 of the parent 1A5 are indicated by *. (D) Density plot of branch length vs. GC-content of individual Styx copies. Styx copies of the most recent burst show small branch lengths and high GC-content.

Repeated rearrangement triggered by a TE that underwent a recent burst

Through four rounds of meiosis, most of the breakpoints (B-E) were introduced multiple times independently, and one breakpoint at position A, occurred only once. To understand the mechanism triggering the initial chromosome 17 rearrangement and sustaining the degeneration, we examined the breakpoint sequences. We identified a full copy of the DNA transposon Styx at the position B and a partial copy of the same family Styx at position C in the parental chromosome of 1A5 (fig. 5A). Styx was previously described as a negative regulator of virulence 17. Furthermore, Styx was shown to proliferate in the pedigree and pathogen populations across continents 18,19. A second complete copy of Styx was found at position D (~8 kb in length) (fig. 5A). Styx has 21 copies in the 1A5 genome including three copies on chromosome 17 (fig. 5B). Full-length copies of Styx contain four coding regions of which one shows weak homology to RNAse H and an integrase (fig. 5A) 18. We analyzed Styx copy numbers in 19 completely assembled genomes of Z. tritici 13 and in the genomes of the sister species 20. Styx is present in high copy numbers in the sister species Z. passerinii (Zpa63), Z. brevis (Zb87) and Z. pseudotritici. Styx is nearly absent in genomes from the center of origin of Z. tritici 18, suggesting that Styx was originally present in very low numbers but appears to have undergone a recent burst in North and South America, Europe and Australia, resulting in high copy numbers in 1A5 and 1E4 and other isolates (fig. 5B). The long variant of Styx is most abundant in Z. tritici and its sister species. The TE copy on chromosome 17 was introduced following a recent burst (fig. 5C). TE copies created during the burst are characterized by short terminal branch lengths and high GC-content (fig. 5D). The creation of the enlarged chromosome 17 was possibly mediated by non-allelic homologous recombination between Styx copies at positions B and a different sequence with microhomology at position E. We examined the regions 1 kb up- and downstream from position A, B and E for similar repetitive sequences. Positions B and E carried a similar 6 bp repeat that may have ultimately triggered the rearrangement (supplementary files S1-2). At position B, the element is at 582 bp from the breakpoint. Finally, near position E the repeat is 34 bp away from the breakpoint. Similarly, Position A and E had two similar 3 bp repeats (supplementary file S2 and S3).

Discussion

We recapitulated the dynamics of a highly unusual chromosomal rearrangement. Using split long-reads, we identified the exact breakpoints and retraced the degeneration through four rounds of meiosis. We found that the primary degenerative rearrangement was caused by non-allelic recombination between a
copy of a TE family *Styx* and a region with microhomology to the TE. The degenerated chromosome was composed of a large duplicated region connected to a single-copy chromosomal segment near the centromere. The sequences serving as triggers for the rearrangement in the first round of meiosis was likely the *Styx* TE being present in multiple copies on chromosome 17 and regions with microhomology to the *Styx*. Subsequent rounds of meiosis increased the spectrum of rearrangement breakpoints consistent with runaway chromosomal degeneration. Chromosomal segments affected by the rearrangement were repeatedly affected multiple times and three of these segments co-localize with copies of the *Styx* TE. Concurrent with the degeneration, non-disjunction events of chromosome 17 increased along the pedigree, characterized by disomy and trisomy. The frequency of non-disjunction events increased with the presence of chromosome 17 variants in both parents suggesting that opportunities for mispairing of chromatids increases the likelihood of aberrant segregation. The degenerating chromosome was furthermore affected by the genomic defense mechanism RIP that introduced AT-biased mutations into recently duplicated regions. Overall, we found that the identity of paired parental genotypes had a major influence on the degree of rearrangements observed in the progeny.

The copy-number amplification observed in progeny Ztprog64 and Ztprog30 shares hallmarks of breakage-fusion-bridge (BFB) cycles observed in many cancer lines \(^\text{21–26}\). Following the initial degeneration, chromosome 17 was stabilized in pairings of a single rearranged chromosome variant together with a parental chromosomal variant (i.e., Ztprog8 and Ztprog9). This suggests that the progression of the degenerative cycle was interrupted through proper segregation. The chromosome 17 variant in Ztprog8 was generated through non-allelic homologous 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 \(^\text{27}\). The reconstructed chromosome 17 variants pinpointed the most likely sequence triggers for the observed rearrangements. The observation that specific locations on the chromosome were repeatedly involved in creating new rearrangements during chromosome degeneration indicates that these locations are fragile sites (i.e. sites frequently co-localizing with chromosome rearrangements). Three out of four of these fragile sites co-locate with copies of TEs in the *Styx* family. The tendency of *Styx* to trigger chromosomal degeneration is likely to be heterogenous within the species and is also at the origin of an inter-chromosomal rearrangement in the same progeny \(^\text{18}\). Complete genomes from isolates collected near the center of origin of the pathogen carry no *Styx* copies and the expansion to higher copy-numbers appears restricted to European genotypes and their descendants \(^\text{13}\). A separate burst of the TE Styx appears to have occurred in the sister species of *Z. tritici* \(^\text{18,28}\). Hence, the propensity to trigger degenerative cycles on chromosome 17 is an emerging trait within the species. The significant consequences for chromosomal integrity could mean that the activity of *Styx* and the presence of fragile sites are under strong selection.
Taken together, our results show that specific TE sequences trigger runaway chromosome degeneration. Non-allelic homologous recombination drives the deleterious rearrangements at the onset of the process with non-disjunction events in subsequent rounds of meiosis. Specific regions on chromosomal segments are preferentially amplified, consistent with patterns observed during degenerative BFB cycles in cancer cell lines. BFB was first discovered in maize by McClintock in dicentric chromosomes going through cycles of degeneration and are also known to occur in animals and fungi. BFB cycles are initiated via telomere-telomere fusions of chromosomes with degraded or missing telomeres. The centromeres of the dicentric chromosome are pulled in opposite directions during anaphase generating a bridge that breaks apart and results in daughter cells with different lengths of the chromosome that lack telomeres and are susceptible to fusion and therefore further deteriorating cycles perpetuate. At the karyotypic level, chromosome 17 undergoes cycles consistent with BFB cycles. However, the likely absence of centromere duplications is inconsistent with classic BFB cycles. We pinpoint an alternative mechanism driving amplification of chromosomal regions in a pattern resembling BFB involving ectopic recombination. Recombination serves both as the initial trigger to create unstable chromosomes and maintains the degenerative process. Non-disjunction amplifies the process following the initial meiosis producing the aberration. We show that large-scale chromosome rearrangements can spontaneously occur in natural pairings of fungal individuals bearing hallmarks of degenerative processes in somatic cell lines such as observed in cancer.
Materials and Methods

Establishment of a four-generation pedigree

We analyzed four generations of crosses, including several backcrosses of the haploid progeny A66.2 (fig. 1A). The initial cross between 1A5 with 1E4 and subsequent crosses were described earlier 7,40. The genomes of 1A5, 1E4 and 3D7 were sequenced using PacBio high coverage sequencing and assembled into complete chromosomes 9,11. Parental genome assemblies were also validated using high-density genetic maps 7,41. Crosses were performed by co-infecting wheat leaves with asexual conidia from the parental strains of opposite mating types according to an established crossing protocol 42.

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 medium for 6–7 days at 20°C prior to DNA extraction.

Chromosome segment PCR assay

In order to assess 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 7. Detailed information on primer binding sites in the genome of 1A5 is available in 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 for a conserved microsatellite locus in each PCR mix 43. Successful PCRs produced an additional band that was clearly distinguishable from the PCR product associated with the amplified chromosome region. PCR products were analyzed on agarose gels. Data was visualized using the R package gplots (https://github.com/talgalili/gplots).

DNA extraction for PacBio sequencing

Progeny DNA from each cross was extracted using a modified version of the cetyltrimethylammonium bromide (CTAB) DNA extraction protocol developed for plant DNA extractions 44. Fungal spores were grown for 5–7 days in YSB broth and lyophilized overnight. Approximately 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.

**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. We included seven *Z. tritici* isolates confirmed to have inherited chromosome 17 from each of the crosses. Isolates were transferred from stocks maintained in glycerol at -80°C to yeast malt agar plates and 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 50 ml Falcon tubes and incubated at 55°C in a water bath for a few minutes. 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 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. De-staining was performed in water for 5–10 min. Photographs were taken under ultraviolet light with a Molecular Imager (Gel...
As size standards, we used chromosome preparations of *Saccharomyces cerevisiae* (BioRad, Switzerland).

**Southern hybridization for A66.2 and A2.2**

Southern blotting and hybridization were performed following standard protocols. First, hydrolysis was performed in 0.25 M HCl for 30 min, then DNA was transferred onto Amersham Hybond-N+ membranes (GE Healthcare, Switzerland) overnight under alkaline conditions. 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 described earlier (supplementary table S4).

**PacBio 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, Inc.). 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.

**Assembly of chromosome 17 and breakpoint-junction analyses**

We followed the same approach to assemble chromosome 17 in progeny from all four generations (fig. 1A). We mapped the reads to the reference genome 1A5 using minimap2 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 estimate copy numbers of distinct regions. We then identified breakpoints by analyzing regions with >15 reads that either end or start at a specific position using bedtools bamtoBed and extracted split reads in this region. Hereafter, we assembled draft chromosomes by using the structural 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 spanning each established junction point (supplementary table S2). Established chromosome 17 assemblies were error-corrected with Quiver.
Characterization of sequence features of breakpoints

Transposable elements consensus sequences were annotated on the parent 1A5 chromosome 17 sequence using RepeatMasker using previously described TE family consensus sequences. The cut-off value was set to 250 and both simple repeats and low-complexity regions were discarded. We used Flexidot version 1.06 to visualize the Styx TE consensus sequence. To further classify the Styx TE, we analyzed conserved domains in the consensus sequence with BlastX and the nonredundant NCBI protein database. We detected four putative coding regions consistent with previous analyses of the TE. As a proxy for repeat-induced point mutations, we calculated the GC-content for each copy with geecee from EMBOSS version 6.6.0.

We extracted all annotated copies of the Styx TE family in the 19 reference-quality Z. tritici genomes and the sister species with samtools faidx version 1.9. Multiple sequence alignments of all copies were made with MAFFT version 7.453 and the following parameters: --reorder --localpair --maxiterate 1000 --nomemsave --leavegappyregion. We located the putative coding region 4 in the multiple sequence alignment and extracted the sequence with extractalign from EMBOSS. We excluded empty sequences (i.e. not containing coding regions) with trimAl version 1.4rev15 and sequences with more than 20% of gap sites with seqkit version 0.11.0. To remove large blocks of gaps and regions that represent rare insertions, we used gblocks version 0.91b with the parameters: Type of Sequence: DNA, Minimum Length Of A Block: 5, Allowed Gap Positions: With Half. We used RAxML version 8.2.12 to create phylogenetic trees in three rounds. First, we generated 20 ML trees each with a different starting tree and extracted the tree with best likelihood with the following parameters: raxmlHPC-PTHREADS-SSE3 -T 4 -m GTRGAMMA -p 12345 -# 10 --print-identical-sequences. Second, we made a bootstrap search for support values with the following parameters: raxmlHPC-PTHREADS-SSE3 -T 4 -m GTRGAMMA -p 12345 -b 12345 -# 50 --print-identical-sequences. Finally, we drew bipartitions on the best ML tree with bootstrapping: raxmlHPC-PTHREADS-SSE3 -T 4 -m GTRGAMMA -p 12345 -f b --print-identical-sequences. We imported the tree into R with read.tree from the package treeio version 1.10.0 (https://github.com/YuLab/SMU/treeio), converted it to a tibble object with as.tibble from package tibble version 3.0.1 in tidyverse version 1.3.0 and added the GC-content per sequence as a variable with left_join from the package dplyr version 0.8.5 in tidyverse and recreated the tree with as.phylo and as.treedata in treeio. We visualized the tree with GGTree version 2.0.1. We extracted sequences from Z. pseudotritici closest to the Z. tritici subtree as an outgroup to root the Z. tritici tree. A dotplot of the consensus sequence was generated using Flexidot.

To search for microhomology at rearrangement breakpoints, we analyzed tandem repeats within 1000 bp of the breakpoint locations A-E using mreps with the parameters -exp 3 and -res 5 to allow for fuzzy detection of degenerate repeats.

Repeat-induced point mutation analysis

19
We mapped PacBio reads to the reference genome 1A5 using minimap2\textsuperscript{47} as described above. We performed SNP calling with the software longshot and included a minimum mapping quality cutoff of -q 30\textsuperscript{63}. All SNPs with a “dn” tag, indicating mapping issues, were removed. We allowed for 20% errors in the uncorrected PacBio reads. We expected the alternate allele at polymorphic sites to be shared by ≥80% of the mapped reads for single-copy regions, shared by ≥40% of the reads in regions with evidence for duplications (based on read coverage), shared by ≥27%, ≥20% and ≥16% of the reads in regions with three, four and five copies (based on read coverage), respectively. We also excluded regions with ≤50% or ≥150% of the expected read coverage for further analyses to reduce erroneous variant calls due to inconsistent read mapping.

Acknowledgements

We are grateful for technical support in the laboratory by Marcello Zala. Thomas Badet provided advice on genome sequence analyses.

Funding information

This work was supported by grants from the Swiss National Science Foundation to BAM (155955) and DC (173265).

Competing interests

We declare to have no competing interests.
References


A

meiosis 1
meiosis 2
meiosis 3
meiosis 4

B

1A5
1E4

Ztprog08 Ztprog09 Ztprog64
Ztprog01 Ztprog11
Ztprog19 Ztprog45

C

D

Progeny A66.2
coverage
putative centromere

Progeny A66.2
position (kb)
parent 1A5 position (kb)

(i) Translocation
Deletion B-C
(ii) Translocation
Deletion B-C

Size ladder (markers in Mb)

Progeny A66.2 and A2.2

A66.2
A2.2

B-C deletion
E-F deletion

Translocation
Deletion B-C
Deletion E-F

2x mean coverage (core chromosomes)
mean coverage (core chromosomes)
putative centromere

Size: 819 kb
Deletion E-F
Deletion B-C
Translocation

interchromatid non-allelic recombination

original chr17
rearranged chr17
missing chr17
unrelated chr17

800 600 400 200 0

800 600 400 200 0

A B C D E F

A B C D E F
A. Meiosis 1

B. Meiosis 2

C. Meiosis 3

D. Meiosis 4

Parent copy of chromosome 17

2 variants
V1 with V4, or, V2 with V3

V1 and V2 with V3, or 2x V1 with V4

2 possible variants: V1 with V4, or V2 with V3

2 variants V1 with V4 or, V2 with V3

2 possible variants: V1 with V4, or V2 with V3

2 possible variants: V1 with V4, or V2 with V3