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

Simone Fouché[1,2], Ursula Oggenfuss[1, ¶], Bruce A. McDonald[2], Daniel Croll[1,*]

[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 [1]. Similarly, non-disjunction and chromosome rearrangements occur in somatic lines leading to genomic instability, which is a hallmark of cancers [2]. 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 [3]. Though the location of many mitotic fragile sites has been identified in the human genome [4], 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 [5]. 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 [6]. 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 [7–13]. Chromosome rearrangements occur frequently during meiosis with higher rates of rearrangements observed in the accessory chromosomes [7,14].

A rearranged accessory chromosome was identified in *Z. tritici*, a haploid fungal pathogen of wheat [7], 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.

81

82 ## Results

83

84 ***Origin of a complex chromosome rearrangement***

85 The rearrangement of the *Z. tritici* accessory chromosome 17 was first discovered in two F1 progeny
86 identified as A66.2 and A2.2, respectively (fig. 1A) [7]. To assess how frequently rearrangements of
87 chromosome 17 occur through meiosis, we screened 48 progenies from the same 1A5x1E4 cross (with
88 and without chromosome 17, respectively) for rearranged chromosome 17 with a segment-specific PCR
89 assay and found that anomalies of chromosome 17 were most likely restricted to progeny A66.2 and
90 A2.2 (fig. 1A, supplementary fig. S1). We screened a further 228 progeny from this cross and found
91 four progeny with rearranged chromosome 17, of which only A66.2 and A2.2 carried the same
92 rearrangement profile [14]. Additionally, we assessed the frequency of rearranged chromosome 17
93 variants using a whole-genome sequence coverage-based approach in 150 isolates from Switzerland
94 matching the sampling location of the parental isolates. Based on read coverage variation, the
95 population carries likely disomic variants of chromosome 17 or partially duplicated variants (n = 3
96 isolates) [15]. The enlarged nature of chromosome 17 in A66.2 and A2.2 was confirmed by pulsed-field
97 gel electrophoresis (PFGE) as well as southern hybridization (fig. 1B) [7].

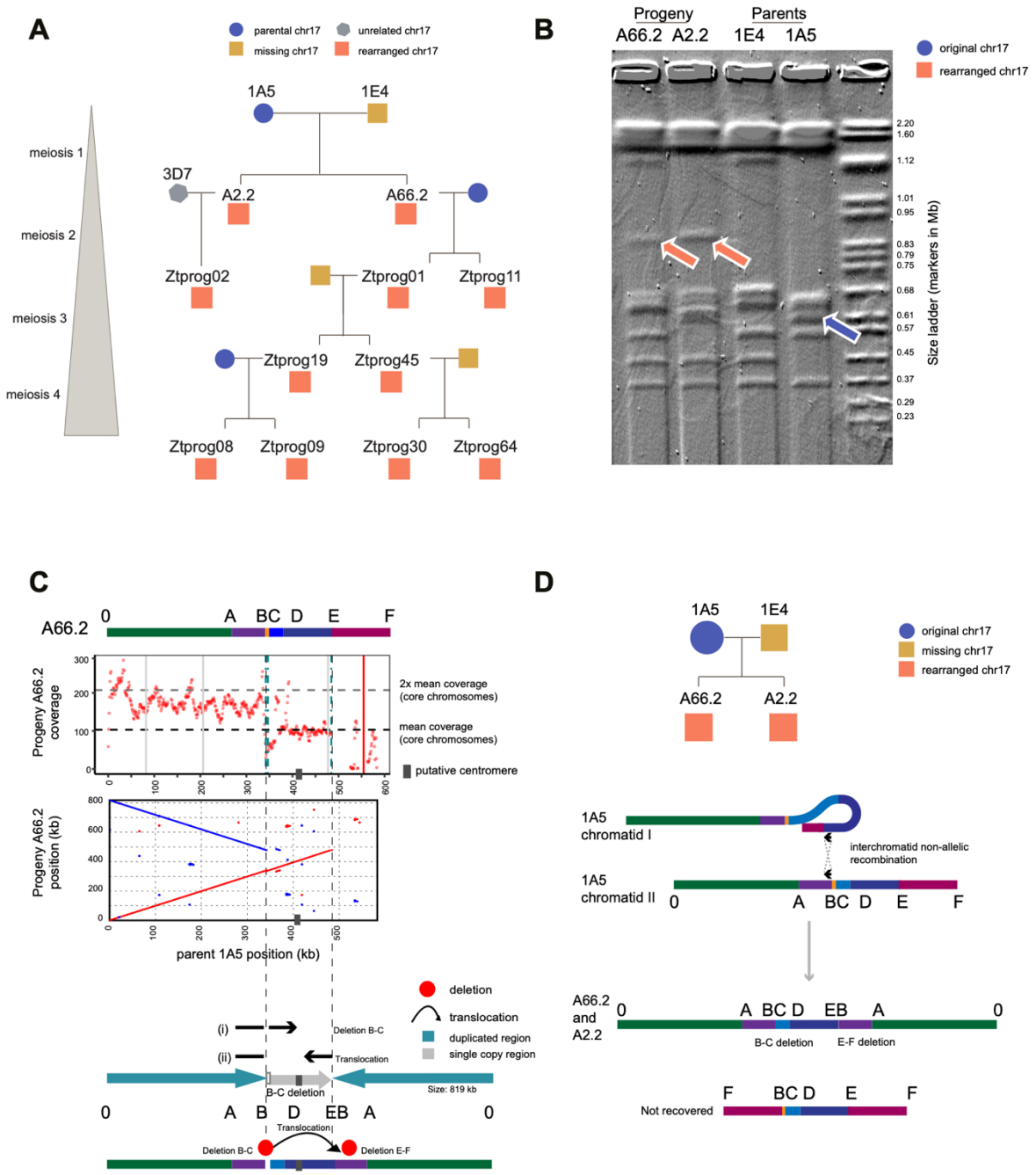


Figure 1: Progeny pedigree reveals enlarged chromosome 17. (A) Chromosome 17 pedigree of four rounds of meiosis. The colors indicate whether the isolate carries either a parental, unrelated (not carried by a parent), rearranged or no chromosome 17 according to pulsed-field gel electrophoresis (PFGE) 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 [7]. (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

112 that generated the enlarged chromosome 17 (0-ABCDEBA-0) recovered in the progeny A66.2. The hypothetical
113 smaller rearrangement product (FBCDEF) was not observed.
114
115 To investigate at base-pair resolution how the enlarged chromosome 17 was generated in the progeny
116 A66.2 and A2.2, we sequenced the isolates with long-read PacBio technology. The recency of the
117 duplication creates a significant challenge for the chromosomal assembly as duplicated regions are not
118 expected to show sequence divergence and therefore could erroneously suggest the presence of only
119 one copy. Therefore, we used a combination of coverage and split long-read mapping approaches. First,
120 we mapped reads for the progeny to the 1A5 parent chromosomes and identified regions of chromosome
121 17 with higher than the mean coverage of the core chromosomes (1-13). We found sharp transitions in
122 read coverage along chromosome 17 (fig. 1C). We found that region 0-B in progeny A66.2 and A2.2
123 had approximately double the mean coverage of the core chromosomes, implying that this region was
124 duplicated (fig. 1C, supplementary fig. S2 and S3, supplementary table S1). Further supporting evidence
125 included a breakpoint at the end of segment 0-B, with some reads showing split mapping in two distinct
126 regions of the chromosome (fig. 1C, supplementary fig. S2, supplementary table S2). Reads mapping
127 at this position were evenly split between canonically mapping across positions B-C and split reads
128 mapping from A-B and continue in an inverted orientation from position E towards D (fig. 1C,
129 supplementary fig. S2). When combined, this indicates that the duplicated region (0-B) exists twice in
130 the progeny genome, and that the duplicated sequences are connected to two distinct locations of the
131 chromosome at C and E (fig. 1C, supplementary fig. S2 and S3). Based on coverage, the region between
132 C and E is single-copy. A lack of coverage after position E towards F in the progeny, suggesting that
133 this regions is absent in the rearranged chromosome, is confirmed by the failure of a PCR to amplify in
134 this region (fig. 1B, supplementary fig. S2-3). A lack of coverage between B and C indicates that this
135 region is also missing. Using information on coverage and reads spanning distinct chromosomal
136 regions, the chromosome 17 of progeny A66.2 and A2.2 is composed of region 0-B, followed by region
137 C-E, and a second copy of region 0-B in an inverted orientation (fig. 1C; supplementary fig. S2,
138 supplementary table S2). We identified the putative centromere to be located at position 405,779-
139 415,898 bp based on sequence homology to chromosome 17 in the reference genome IPO323 [8].
140 Therefore, the resulting enlarged chromosome 17 is expected to carry a single copy of the centromeric
141 region (fig. 1C). Both progenies most likely inherited an identical, enlarged chromosome 17 generated
142 during the first round of meiosis through the same non-allelic recombination between sister chromatids
143 at locations B and E (fig 1D). The other putative product of this event was not recovered, despite being
144 predicted to have a centromere. We reconstructed the sequence of the rearranged chromosome 17 based
145 on the above links using sequences of the parental chromosome 17 from 1A5 (the only parent carrying
146 the chromosome). The enlarged chromosome 17 was 819 kb in length for A66.2 and A2.2, matching
147 the length identified from a PFGE gel separating chromosomes by length (fig. 1B).
148

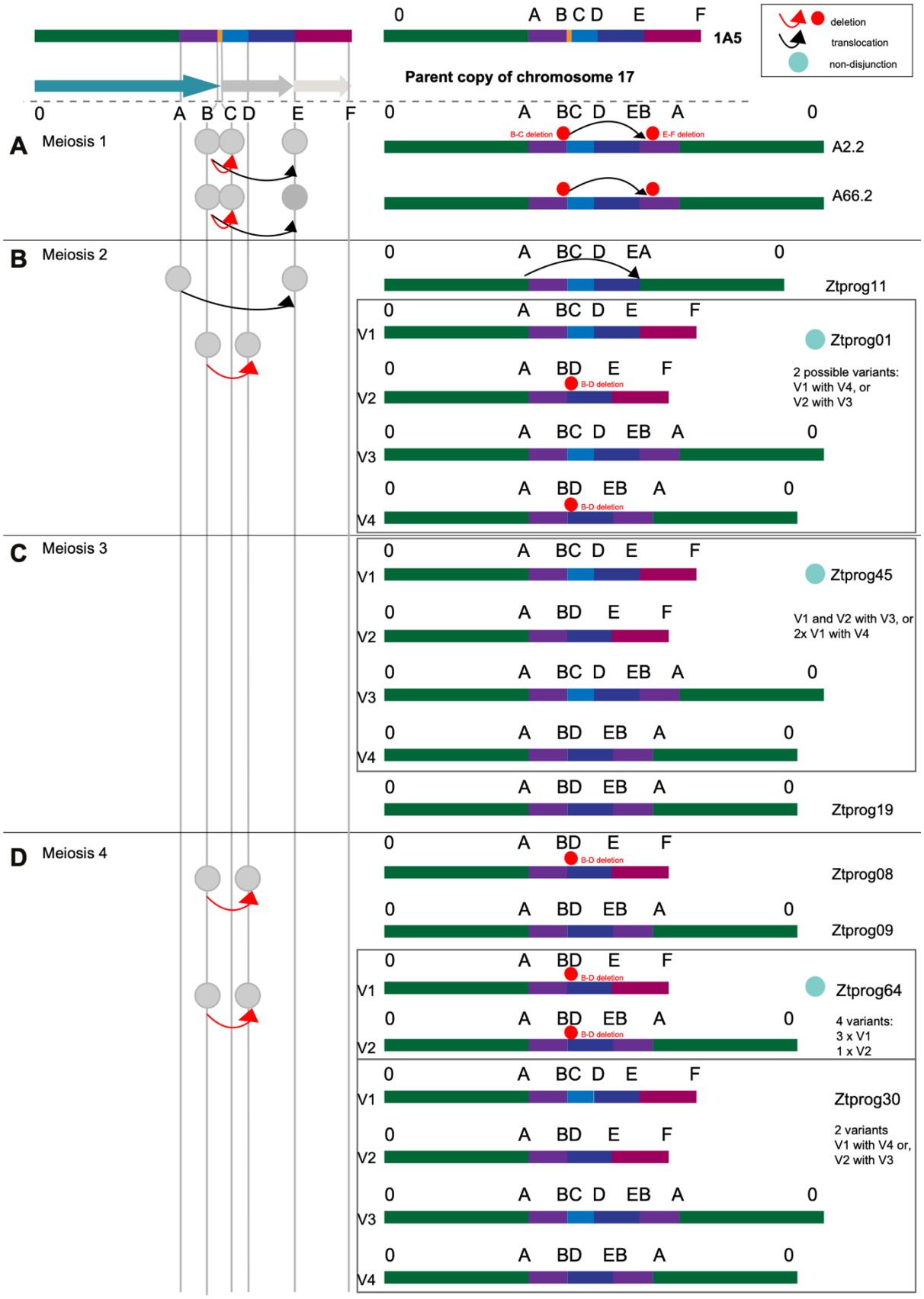


Figure 2: Sequence of chromosome 17 rearrangements tracked through four rounds of meiosis. Chromosome rearrangements are shown according to breakpoint positions A-F identified through split read mapping. Arrows linking different letters A-F indicate translocations and deletions for each successive round of

154 meiosis. Panels on the right show schematic representations of each chromosome 17 variant present in each
155 progeny. If multiple variants were detected, variants are labelled V1-V4.
156

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

158 To investigate whether the enlarged chromosome 17 experiences instability in further rounds of meiosis,
159 we traced the novel enlarged chromosome through further generations by performing several
160 backcrosses (fig. 1A). New variants of chromosome 17 were already generated in the following round
161 of meiosis (fig. 2B, 3B, supplementary fig. S3). Chromosome 17 of Ztprog11 carried a duplicated region
162 0-A and the second copy of 0-A is joined to position E in an inverted orientation (fig. 2B, 3B,
163 supplementary fig. S3, S4). Ztprog01 has two copies of chromosome 17 including a small and a large
164 variant consistent with non-disjunction. One copy lacked the segment between B and D (fig 2B, fig 3B,
165 supplementary fig. S3, S5). During the third round of meiosis, ZtProg19 stably inherited a chromosome
166 17 variant (fig. 2D, fig 3C, supplementary fig. S3, S7). In contrast, based on pedigree and read mapping
167 evidence, the second progeny ZtProg45 carried multiple variants. Variant V2 may have undergone non-
168 disjunction (2xV1 is inherited with 1xV4) or one copy of V1 is present with one copy of V2 and V3
169 each, indicating rearrangements at breakpoints B and D as well as non-disjunction (fig. 2C, fig. 3C,
170 supplementary fig. S3, S6). Therefore, two small variants are present together with one large variant.
171 In the last round of meiosis the region between B and D was deleted in Ztprog8 and Ztprog64 (fig. 2D,
172 fig 3D and ii, supplementary fig. S3, S8, S9). Ztprog64 carries four chromosome 17 variants (three
173 small and one large variant) missing this region (fig. 2D, fig 3i and ii, supplementary fig. S3, S9).
174 Ztprog9 most likely inherited the enlarged chromosome 17 without further rearrangements from the
175 previous generation (fig. 2D, fig 3D, supplementary fig. S3, S10). Ztprog30 has the same chromosome
176 complement as Ztprog01 (fig. 2D, fig 3E, supplementary fig. S3, S11). The enlarged chromosome 17
177 was highly unstable through further rounds of meiosis and degenerative cycles occurred via non-allelic
178 recombination as well as non-disjunction. Rearrangements were triggered recurrently at known
179 breakpoints with the exception of breakpoint A being involved only once.

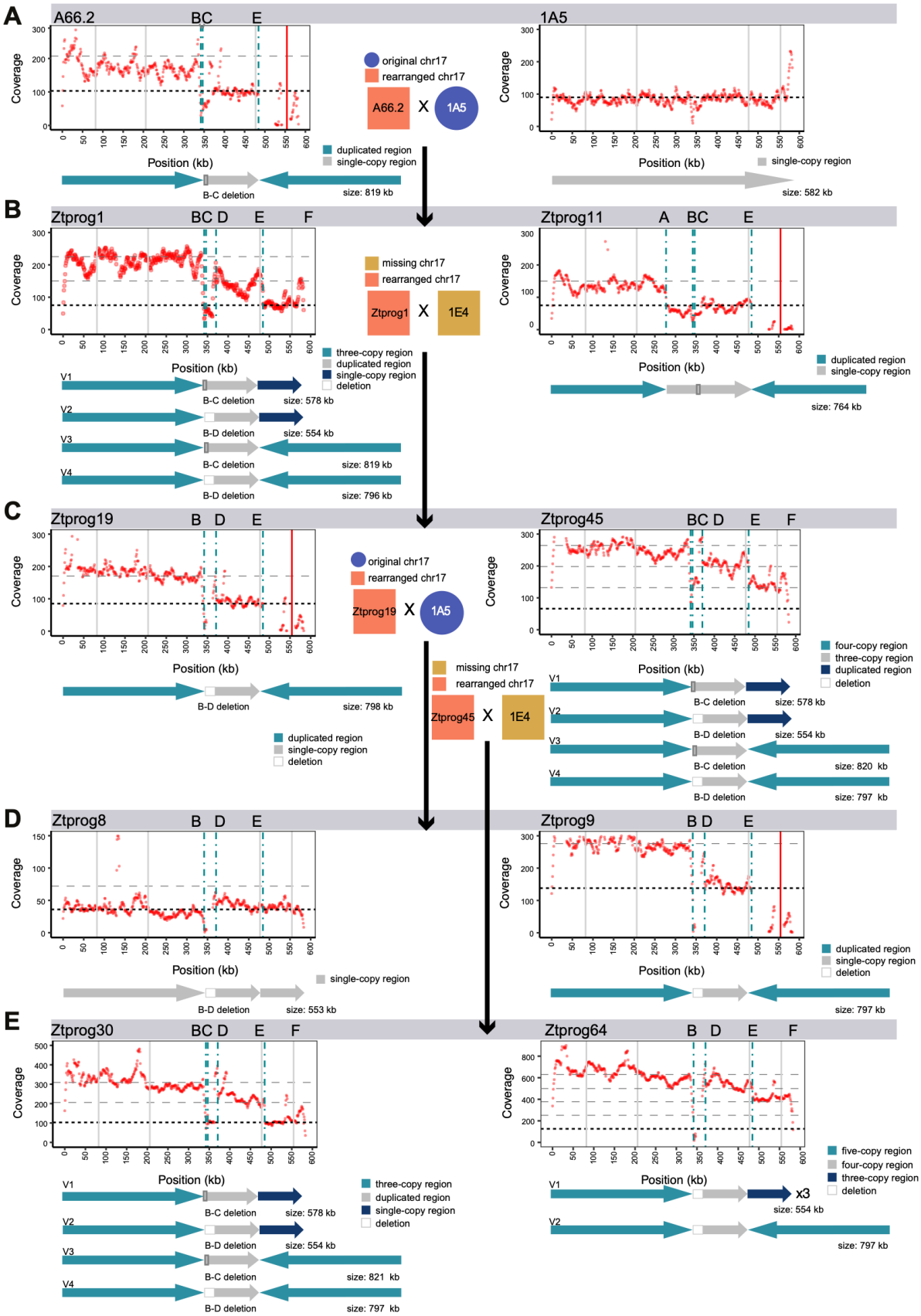


Figure 3: Reconstructed chromosome 17 variants based on sequence rearrangement breakpoints. For each progeny the mapped read coverage is shown relative to the 1A5 reference genome sequence. Black horizontal dashed lines indicate the mean coverage of the core chromosomes and in thin grey two, three and 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.

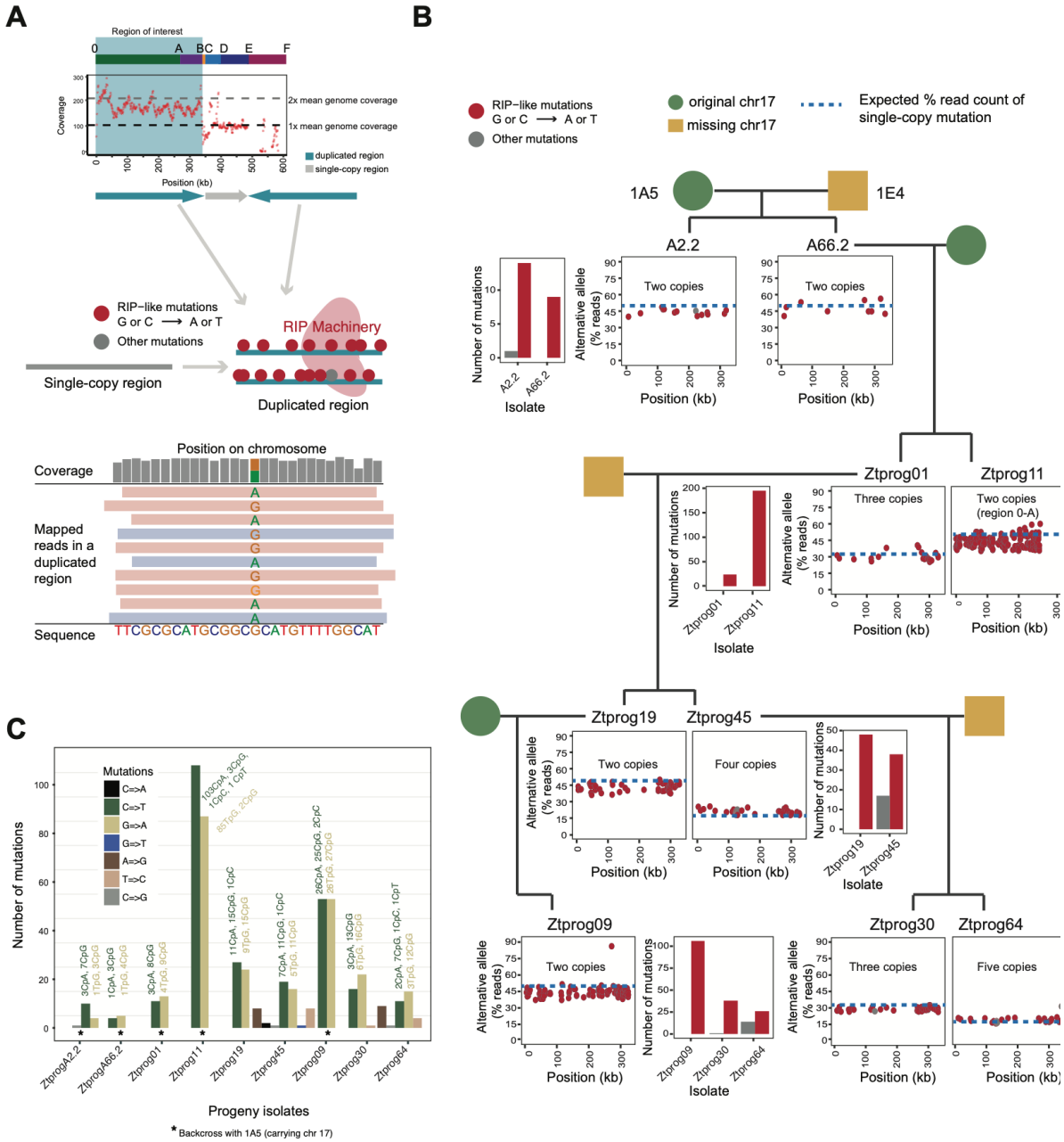


Figure 4: Characterization of repeat-induced point mutation (RIP) acting on duplicated sequences. (A) Coverage plot with read coverage along the chromosome 17 and evidence for a duplicated region 0-B (0-A in Ztprog11). The upper schematic shows how RIP mutations are introduced into duplication regions 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

200 0-B (0-A in Ztprog11) is indicated. Blue dashed lines show the expected % reads confirming a new mutation
201 carried by at least one copy of the duplicated regions. (C) Breakdown of all mutations detected in duplicated
202 regions of each progeny. Mutations at particular dinucleotides are summarized individually per progeny.
203

204 **Degeneration through a mutation mechanism targeting duplicated sequences**

205 Besides degeneration via rearrangements and non-disjunction, the duplicated region (0-B and 0-A for
206 Ztprog11) of chromosome 17 was also targeted by repeat-induced point (RIP) mutations (fig. 4A). We
207 mapped reads for each progeny to the parental chromosome 17 (isolate 1A5) and identified regions that
208 were targeted by RIP-like mutations (transitions from G or a C to an A or a T; fig. 4B region (0-B) and
209 supplementary fig. S4 (other regions)). We found that most of the mutations localized to the region 0-
210 B were RIP-like. When combining total evidence, we identified multiple copies of the region targeted
211 by RIP-like mutations. Progeny Ztprog08 lacking any duplicated regions also showed no evidence for
212 RIP being active. Interestingly, RIP-like mutations never targeted the putative centromere (405,779-
213 415,898 bp; supplementary fig. S4). RIP-like mutations predominantly target CpA dinucleotides (TpG
214 in the reverse complement) in other model fungi [16], and this is also the case in *Z. tritici* (fig. 3C).
215 Additionally, CpG dinucleotides were frequently targeted. New RIP-like mutations were found after
216 each generation, suggesting that RIP was activated over multiple rounds of meiosis and did not reach
217 saturation. The highest number of RIP mutations was found in Ztprog11 and Ztprog19 (generations 2
218 and 4) with each carrying two copies of region 0-B on one chromosome. Additionally, both progenies
219 were generated by a backcross with 1A5. Our results show that in addition to rearrangements and non-
220 disjunction, RIP is a mutational mechanism driving degenerative cycles in fungi.

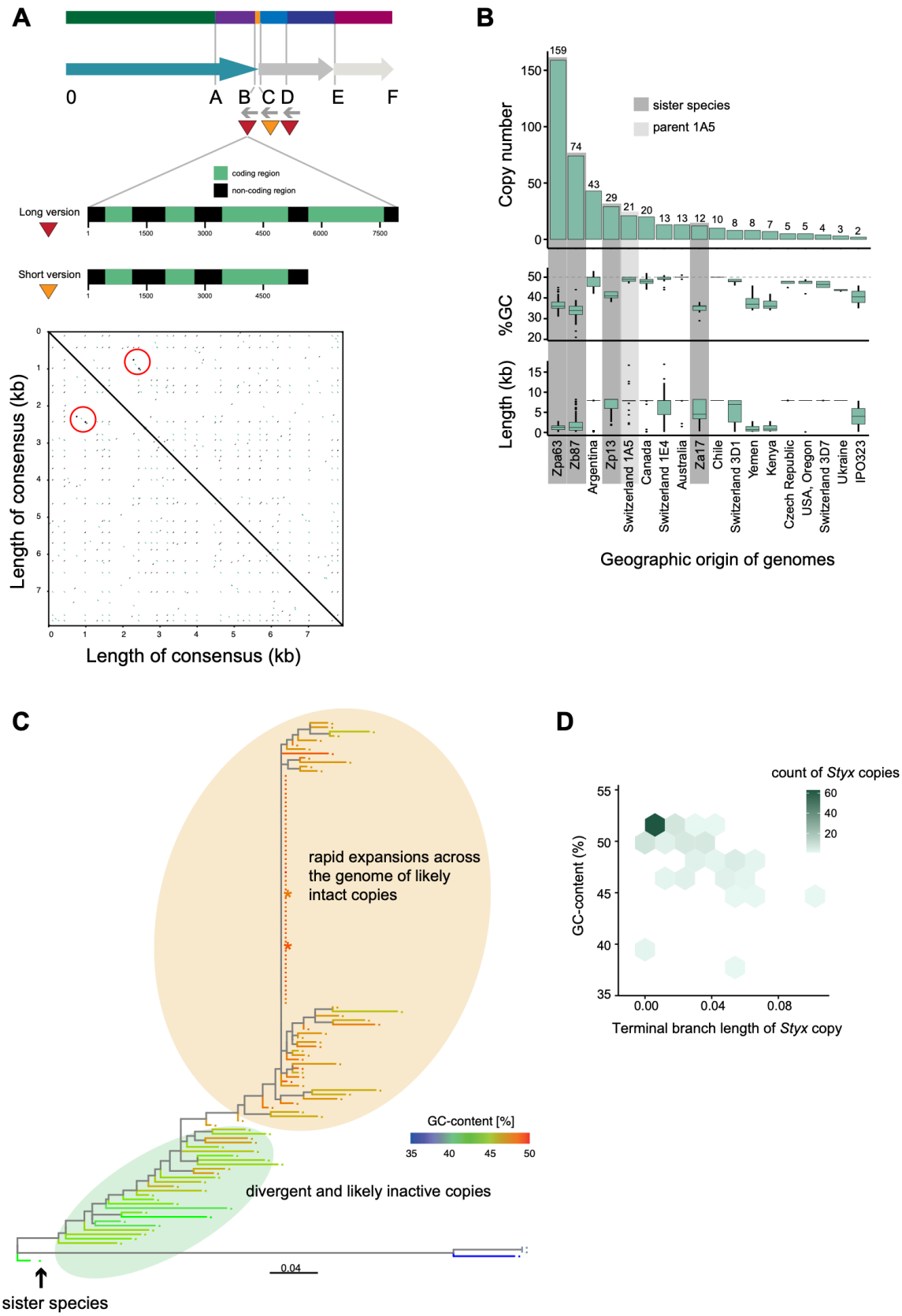


Figure 5: Recent expansion of the *Styx* element underpinning the chromosomal rearrangements. (A) The length and location of coding regions of the long and short copies of *Styx*. Dotplot of the consensus sequence showing duplicated regions (red circle). (B) Copy number, GC-content and length distribution of *Styx* copies recovered from a global panel of *Z. tritici* reference genomes. Dark grey shows genomes of sister species *Z.*

ardabiliae (Za17), *Z. passerini* (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

264 copy of a TE family *Styx* and a region with microhomology to the TE. The degenerated chromosome
265 was composed of a large duplicated region connected to a single-copy chromosomal segment near the
266 centromere. The sequences serving as triggers for the rearrangement in the first round of meiosis was
267 likely the *Styx* TE being present in multiple copies on chromosome 17 and regions with microhomology
268 to the *Styx*. Subsequent rounds of meiosis increased the spectrum of rearrangement breakpoints
269 consistent with runaway chromosomal degeneration. Chromosomal segments affected by the
270 rearrangement were repeatedly affected multiple times and three of these segments co-localize with
271 copies of the *Styx* TE. Concurrent with the degeneration, non-disjunction events of chromosome 17
272 increased along the pedigree, characterized by disomy and trisomy. The frequency of non-disjunction
273 events increased with the presence of chromosome 17 variants in both parents suggesting that
274 opportunities for mispairing of chromatids increases the likelihood of aberrant segregation. The
275 degenerating chromosome was furthermore affected by the genomic defense mechanism RIP that
276 introduced AT-biased mutations into recently duplicated regions. Overall, we found that the identity of
277 paired parental genotypes had a major influence on the degree of rearrangements observed in the
278 progeny.
279
280 The copy-number amplification observed in progeny Ztprog64 and Ztprog30 shares hallmarks of
281 breakage-fusion-bridge (BFB) cycles observed in many cancer lines [21–26]. Following the initial
282 degeneration, chromosome 17 was stabilized in pairings of a single rearranged chromosome variant
283 together with a parental chromosomal variant (*i.e.,* Ztprog8 and Ztprog9). This suggests that the
284 progression of the degenerative cycle was interrupted through proper segregation. The chromosome 17
285 variant in Ztprog8 was generated through non-allelic homologous recombination between two *Styx*
286 copies. The repair of chromosomes through recombination between repeats is known to produce
287 intermediary chromosomes that can ultimately produce stable karyotypes [27]. The reconstructed
288 chromosome 17 variants pinpointed the most likely sequence triggers for the observed rearrangements.
289 The observation that specific locations on the chromosome were repeatedly involved in creating new
290 rearrangements during chromosome degeneration indicates that these locations are fragile sites (*i.e.*
291 sites frequently co-locating with chromosome rearrangements). Three out of four of these fragile sites
292 co-locate with copies of TEs in the *Styx* family. The tendency of *Styx* to trigger chromosomal
293 degeneration is likely to be heterogenous within the species and is also at the origin of an inter-
294 chromosomal rearrangement in the same progeny [18]. Complete genomes from isolates collected near
295 the center of origin of the pathogen carry no *Styx* copies and the expansion to higher copy-numbers
296 appears restricted to European genotypes and their descendants [13]. A separate burst of the TE Styx
297 appears to have occurred in the sister species of *Z. tritici* [18,28]. Hence, the propensity to trigger
298 degenerative cycles on chromosome 17 is an emerging trait within the species. The significant
299 consequences for chromosomal integrity could mean that the activity of *Styx* and the presence of fragile
300 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 [23,29]. BFB was first discovered in maize by McClintock in dicentric chromosomes going through cycles of degeneration [30,31] and are also known to occur in animals [32,33] and fungi [34,35]. BFB cycles are initiated via telomere-telomere fusions of chromosomes with degraded or missing telomeres [36]. 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 [24,32,37–39]. 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.

321 ## Materials and Methods

322

323 ***Establishment of a four-generation pedigree***

324 We analyzed four generations of crosses, including several backcrosses of the haploid progeny A66.2
325 (fig. 1A). The initial cross between 1A5 with 1E4 and subsequent crosses were described earlier [7,40].
326 The genomes of 1A5, 1E4 and 3D7 were sequenced using PacBio high coverage sequencing and
327 assembled into complete chromosomes [9,11]. Parental genome assemblies were also validated using high-
328 density genetic maps [7,41]. Crosses were performed by coinfecting wheat leaves with asexual conidia
329 from the parental strains of opposite mating types according to an established crossing protocol [42].
330 Spores of both parents were sprayed onto wheat plants in equal concentration and incubated outdoors
331 for 40–60 days. Ascospores were isolated over several days by incubating infected wheat leaves on wet
332 filter paper inside Petri dishes. Wheat leaves were covered with upside down Petri dish lids filled with
333 water agar, enabling the capture of vertically ejected ascospores. Ascospores captured on the water agar
334 were left to germinate and inspected for contaminants. Only progeny isolates from single ascospores
335 were selected. Each germinating ascospore was transferred to an individual culture plate for clonal
336 propagation. The mycelium produced by each ascospore was used for DNA extraction. Progeny
337 mycelium was grown in YSB (yeast sucrose broth) liquid medium for 6–7 days at 20°C prior to DNA
338 extraction.

339

340 ***Chromosome segment PCR assay***

341 In order to assess the presence-absence polymorphism of chromosome 17 segments, we used previously
342 designed PCR assays to amplify ca. 500 bp regions of coding sequences at regular intervals along the
343 chromosome 17 of reference strain IPO323 [7]. Detailed information on primer binding sites in the
344 genome of 1A5 is available in supplementary table S3. PCR reactions were performed in 20 µl volumes
345 with 5–10 ng genomic DNA, 0.5 mM of each primer, 0.25 mM dNTP, 0.6 U Taq polymerase
346 (DreamTaq, Thermo Fisher, Inc.) and the corresponding PCR buffer. In order to avoid false negatives,
347 we included a primer pair for a conserved microsatellite locus in each PCR mix [43]. Successful PCRs
348 produced an additional band that was clearly distinguishable from the PCR product associated with the
349 amplified chromosome region. PCR products were analyzed on agarose gels. Data was visualized using
350 the *R* package *gplots* (https://github.com/talgalili/gplots).

351

352 ***DNA extraction for PacBio sequencing***

353 Progeny DNA from each cross was extracted using a modified version of the cetyltrimethylammonium
354 bromide (CTAB) DNA extraction protocol developed for plant DNA extractions [44]. Fungal spores were
355 grown for 5-7 days in YSB broth and lyophilized overnight. Approximately 60-100 mg of dried material
356 were crushed with a mortar and pestle. The phenol-chloroform-isoamyl alcohol extraction step was

357 performed twice and the washing step three times. In the last step, the DNA pellet was resuspended in
358 100 µl of sterile water.
359
360 ***Preparation of fungal material for molecular karyotyping***
361 DNA from intact chromosomes was extracted from conidia embedded in agarose gels by the *in situ*
362 digestion of cell walls, using a modified non-protoplasting method [45]. We included seven *Z. tritici*
363 isolates confirmed to have inherited chromosome 17 from each of the crosses. Isolates were transferred
364 from stocks maintained in glycerol at -80°C to yeast malt agar plates and incubated for 3-4 days in the
365 dark at 18°C. Hereafter, conidia were isolated by washing the plates with sterile water and transferring
366 600–800 µl of suspended conidia to new YMA plates. The plates were again incubated for 2 to 3 days
367 as described above. Conidia were harvested by washing the plates with sterile distilled water and filtered
368 through sterile Miracloth (Calbiochem, La Jolla CA, USA) into 50 ml Falcon tubes. The volume was
369 adjusted to 50 ml by adding more distilled water and the suspension was centrifuged at 3750 rpm at
370 room temperature for 15 min with a clinical centrifuge (Allegra X-12R, Beckman Coulter, Brea CA,
371 USA). The pellets were resuspended in 1–3 ml TE buffer (10 mM Tris-HCL, pH 7.5; 1 mM EDTA, pH
372 8.0) and vortexed gently. The spore concentration of the solution was calculated using a Thoma
373 haematocytometer cell counter. The 1.5 ml spore suspensions with a concentration between $8x10^7$ to
374 $2x10^8$ spores/ml were transferred to 50 ml Falcon tubes and incubated at 55°C in a water bath for a few
375 minutes. We added 1.5 ml pre-warmed (55°C) low-melting-point agarose prepared in TE Buffer (2%
376 w/v; molecular biology grade, Biofinex, Switzerland). The solution was mixed by gentle pipetting. An
377 aliquot of 500 µl was solidified on ice for approximately 10 min in a pre-cooled plug casting mold
378 (BioRad Laboratories, Switzerland). Agarose plugs were incubated in 15 ml Falcon tubes containing 5
379 ml of a lysing solution containing 0.25 M EDTA, pH 8.0, 1.5 mg/mL protease XIV (Sigma, St. Louis
380 MO, USA), 1.0% sodium dodecyl sulfate (Fluka, Switzerland). Plugs were incubated for ~24h at 55°C.
381 The lysing solution was changed once after ~18h and gently mixed every few hours. Plugs with whole
382 chromosomal DNA were washed three times for 15-20 min in ~5ml of a 0.1 M EDTA (pH 9.0) solution
383 and then stored in the same solution at 4°C until used.
384
385 ***Pulsed-field gel electrophoresis***
386 Pulsed-field gel electrophoresis (PFGE) was performed using a BioRad CHEF II apparatus (BioRad
387 Laboratories, Hercules CA, USA). Chromosomal plugs were placed in the wells of a 1.2% (wt/vol)
388 agarose gel (Invitrogen, Switzerland) to separate small chromosomes up to 1Mb. Chromosomes were
389 separated at 13°C in 0.56 x Tris-borate-EDTA Buffer [46] at 200 V with a 60–120 s pulse time gradient
390 for 24–26 h. Gels were stained in ethidium bromide (0.5 mg/ml) for 30 min. De-staining was performed
391 in water for 5–10 min. Photographs were taken under ultraviolet light with a Molecular Imager (Gel

392 Doc XR+, BioRad, Switzerland). As size standards, we used chromosome preparations of
393 *Saccharomyces cerevisia*e (BioRad, Switzerland).
394
395 ***Southern hybridization for A66.2 and A2.2***
396 Southern blotting and hybridization were performed following standard protocols [46]. First, hydrolysis
397 was performed in 0.25 M HCl for 30 min, then DNA was transferred onto Amersham Hybond-N+
398 membranes (GE Healthcare, Switzerland) overnight under alkaline conditions. DNA was heat-fixed
399 onto the membranes at 80°C for 2 h. Membranes were prehybridized overnight with 25 ml of a buffer
400 containing 20% (w/v) SDS, 10% BSA, 0.5 M EDTA (pH 8.0), 1 M sodium phosphate (pH 7.2) and 0.5
401 ml of sonicated fish sperm solution (Roche Diagnostics, Switzerland). Probes were labeled with ^{32}P by
402 nick translation (New England Biolabs, Inc.) following the manufacturer's instructions. Hybridization
403 was performed overnight at 65°C. Blots were subjected to stringent wash conditions with a first wash
404 in 16 X SSC and 0.1% SDS and a second wash with 0.26 X SSC and 0.1% SDS. Both washes were
405 performed at 60°C. Membranes were exposed to X-ray film (Kodak BioMax MS) for 2 to 3 days at -
406 80°C. We used the same probe as described earlier (supplementary table S4) [7].
407
408 ***PacBio library preparation***
409 PacBio SMRTbell libraries were prepared using 15-31 µg of high-molecular-weight DNA. The libraries
410 were size-selected with an 8 kb cutoff on a BluePippin system (Sage Science, Inc.). After selection, the
411 average fragment length was 15 kb. PacBio sequencing was run on a PacBio RS II instrument or Sequel
412 at the Functional Genomics Center, Zurich, Switzerland using P4/C2 and P6/C4 chemistry, respectively.
413
414 ***Assembly of chromosome 17 and breakpoint-junction analyses***
415 We followed the same approach to assemble chromosome 17 in progeny from all four generations (fig.
416 1A). We mapped the reads to the reference genome 1A5 using minimap2 [47] with the parameters --
417 secondary=no -ax map-pb. We compared the coverage of regions of chromosome 17 to the mean
418 coverage of the core chromosomes (1-13) to estimate copy numbers of distinct regions. We then
419 identified breakpoints by analyzing regions with >15 reads that either end or start at a specific position
420 using bedtools bamtobed [48] and extracted split reads in this region. Hereafter, we assembled draft
421 chromosomes by using the structural information from reads showing split alignments and joining
422 individual breakpoints (supplementary table S2). Reads were mapped to the assembled chromosomes
423 with minimap2 and we counted the number of reads spanning each established junction point
424 (supplementary table S2). Established chromosome 17 assemblies were error-corrected with Quiver [49].
425

426 ***Characterization of sequence features of breakpoints***

427 Transposable elements consensus sequences were annotated on the parent 1A5 chromosome 17
428 sequence using RepeatMasker using previously described TE family consensus sequences [13]. The cut-
429 off value was set to 250 and both simple repeats and low-complexity regions were discarded. We used
430 Flexidot version 1.06 to visualize the *Styx* TE consensus sequence [50]. To further classify the *Styx* TE,
431 we analyzed conserved domains in the consensus sequence with BlastX and the nonredundant NCBI
432 protein database. We detected four putative coding regions consistent with previous analyses of the TE
433 [18]. As a proxy for repeat-induced point mutations, we calculated the GC-content for each copy with
434 *geecee* from EMBOSS version 6.6.0.
435
436 We extracted all annotated copies of the *Styx* TE family in the 19 reference-quality *Z. tritici* genomes
437 and the sister species [13,51] with samtools faidx version 1.9 [52]. Multiple sequence alignments of all copies
438 were made with MAFFT version 7.453 and the following parameters: --reorder --localpair --maxiterate
439 1000 --nomemsave –leavegappyregion [53]. We located the putative coding region 4 in the multiple
440 sequence alignment and extracted the sequence with extractalign from EMBOSS. We excluded empty
441 sequences (*i.e.* not containing coding regions) with trimAl version 1.4rev15 and sequences with more
442 than 20% of gap sites with seqkit version 0.11.0 [54,55]. To remove large blocks of gaps and regions that
443 represent rare insertions, we used gblocks version 0.91b with the parameters: Type of Sequence: DNA,
444 Minimum Length Of A Block: 5, Allowed Gap Positions: With Half [56]. We used RAxML version 8.2.12
445 to create phylogenetic trees in three rounds [57]. First, we generated 20 ML trees each with a different
446 starting tree and extracted the tree with best likelihood with the following parameters: raxmlHPC-
447 PTHREADS-SSE3 -T 4 -m GTRGAMMA -p 12345 -# 10 --print-identical-sequences. Second, we
448 made a bootstrap search for support values with the following parameters: raxmlHPC-PTHREADS-
449 SSE3 -T 4 -m GTRGAMMA -p 12345 -b 12345 -# 50 --print-identical-sequences. Finally, we drew
450 bipartitions on the best ML tree with bootstrapping: raxmlHPC-PTHREADS-SSE3 -T 4 -m
451 GTRGAMMA -p 12345 -f b --print-identical-sequences. We imported the tree into R with read.tree
452 from the package treeio version 1.10.0 (https://github.com/YuLab-SMU/treeio), converted it to a tibble
453 object with as.tibble from package tibble version 3.0.1 in tidyverse version 1.3.0 and added the GC-
454 content per sequence as a variable with left_join from the package *dplyr* version 0.8.5 in tidyverse and
455 recreated the tree with as.phylo and as.treedata in *treeio* [58,59]. We visualized the tree with GGTree
456 version 2.0.1 [60,61]. We extracted sequences from *Z. pseudotritici* closest to the *Z. tritici* subtree as an
457 outgroup to root the *Z. tritici* tree. A dotplot of the consensus sequence was generated using Flexidot [50].
458 To search for microhomology at rearrangement breakpoints, we analyzed tandem repeats within 1000
459 bp of the breakpoint locations A-E using *mreps* with the parameters -exp 3 and -res 5 to allow for fuzzy
460 detection of degenerate repeats [62].
461
462 ***Repeat-induced point mutation analysis***

463 We mapped PacBio reads to the reference genome 1A5 using minimap2 [47] as described above. We
464 performed SNP calling with the software longshot and included a minimum mapping quality cutoff of
465 -q 30 [63]. All SNPs with a "dn" tag, indicating mapping issues, were removed. We allowed for 20%
466 errors in the uncorrected PacBio reads. We expected the alternate allele at polymorphic sites to be shared
467 by ≥80% of the mapped reads for single-copy regions, shared by ≥40% of the reads in regions with
468 evidence for duplications (based on read coverage), shared by ≥27%, ≥20% and ≥16% of the reads in
469 regions with three, four and five copies (based on read coverage), respectively. We also excluded
470 regions with ≤50% or ≥150% of the expected read coverage for further analyses to reduce erroneous
471 variant calls due to inconsistent read mapping.
472
473
474 **Acknowledgements**
475 We are grateful for technical support in the laboratory by Marcello Zala. Thomas Badet provided advice
476 on genome sequence analyses.
477
478 **Funding information**
479 This work was supported by grants from the Swiss National Science Foundation to BAM (155955) and
480 DC (173265).
481
482 **Competing interests**
483 We declare to have no competing interests.
484
485

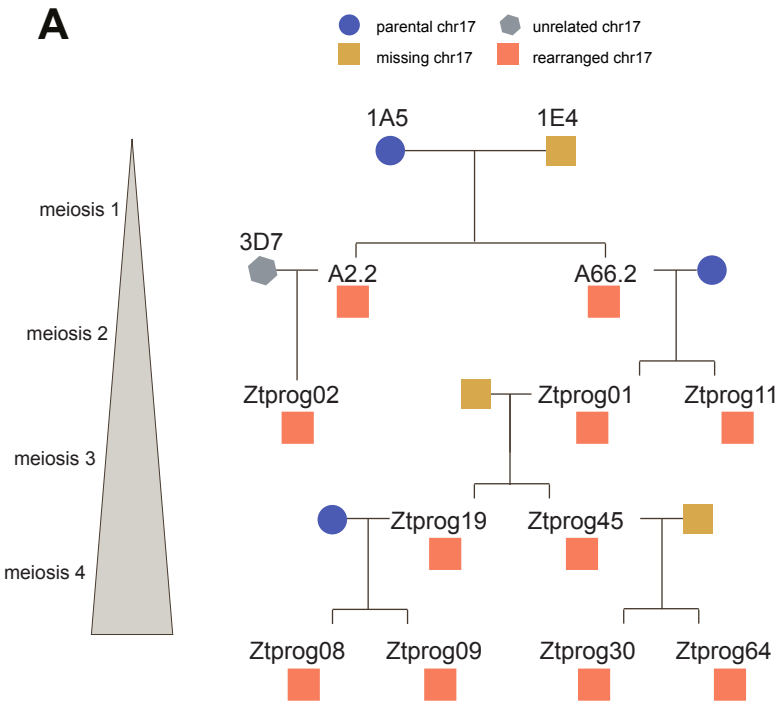
References

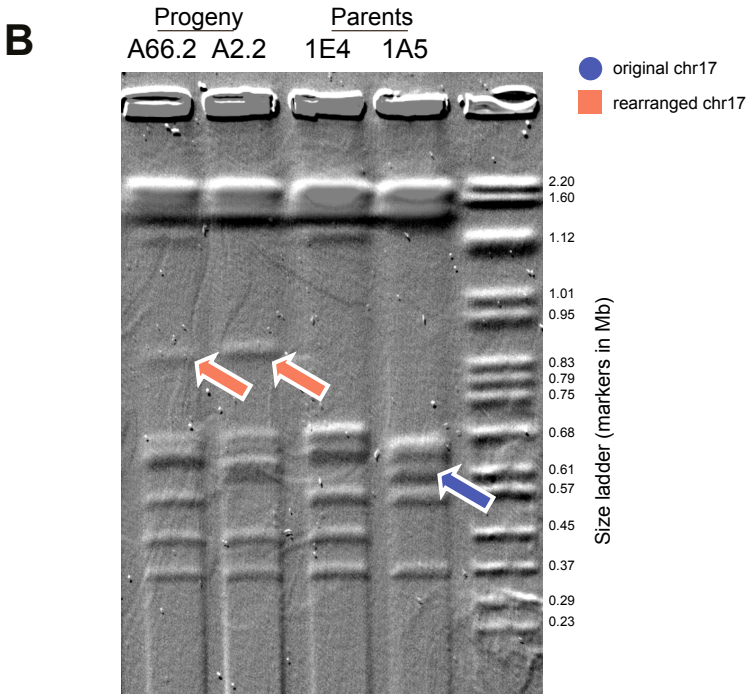
1. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2**, 280–291 (2001).
2. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
3. Durkin, S. G. & Glover, T. W. Chromosome Fragile Sites. *Annu. Rev. Genet.* **41**, 169–192 (2007).
4. Kumar, R. *et al.* HumCFS: a database of fragile sites in human chromosomes. *BMC Genomics* **19**, 985 (2019).
5. Camacho, J. P. M., Sharbel, T. F. & Beukeboom, L. W. B-chromosome evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **355**, 163–178 (2000).
6. Langner, T. *et al.* Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLOS Genet.* **17**, e1009386 (2021).
7. Croll, D., Zala, M. & McDonald, B. A. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLOS Genet* **9**, e1003567 (2013).
8. Schotanus, K. *et al.* Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics Chromatin* **8**, 41 (2015).
9. Plissonneau, C., Stürchler, A. & Croll, D. The evolution of orphan regions in genomes of a fungal pathogen of wheat. *mBio* **7**, e01231-16 (2016).
10. Fouché, S., Plissonneau, C. & Croll, D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* **46**, 34–42 (2018).
11. Plissonneau, C., Hartmann, F. E. & Croll, D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5 (2018).
12. Sánchez-Vallet, A. *et al.* The genome biology of effector gene evolution in filamentous plant pathogens. *Annu. Rev. Phytopathol.* **56**, 21–40 (2018).
13. Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A. & Croll, D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.* **18**, 12 (2020).
14. Fouché, S., Plissonneau, C., McDonald, B. A. & Croll, D. Meiosis Leads to Pervasive Copy-Number Variation and Distorted Inheritance of Accessory Chromosomes of the Wheat Pathogen *Zymoseptoria tritici*. *Genome Biol. Evol.* **10**, 1416–1429 (2018).
15. Singh, N. K., Karisto, P. & Croll, D. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*. *Microb. Genomics* **7**, (2021).
16. Gladyshev, E. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. in *The Fungal Kingdom* (eds. Heitman, J. et al.) 687–699 (ASM Press, 2017). doi:10.1128/9781555819583.ch33.
17. Wang, C., Milgate, A. W., Solomon, P. S. & McDonald, M. C. The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.* **22**, 800–816 (2021).
18. Badet, T., Feurtey, A. & Croll, D. Recent reactivation of a pathogenicity-associated transposable element triggers major chromosomal rearrangements in a fungal wheat pathogen. bioRxiv (2023) doi:10.1101/2023.03.29.534637.
19. Feurtey, A. *et al.* A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nat. Commun.* **14**, 1059 (2023).

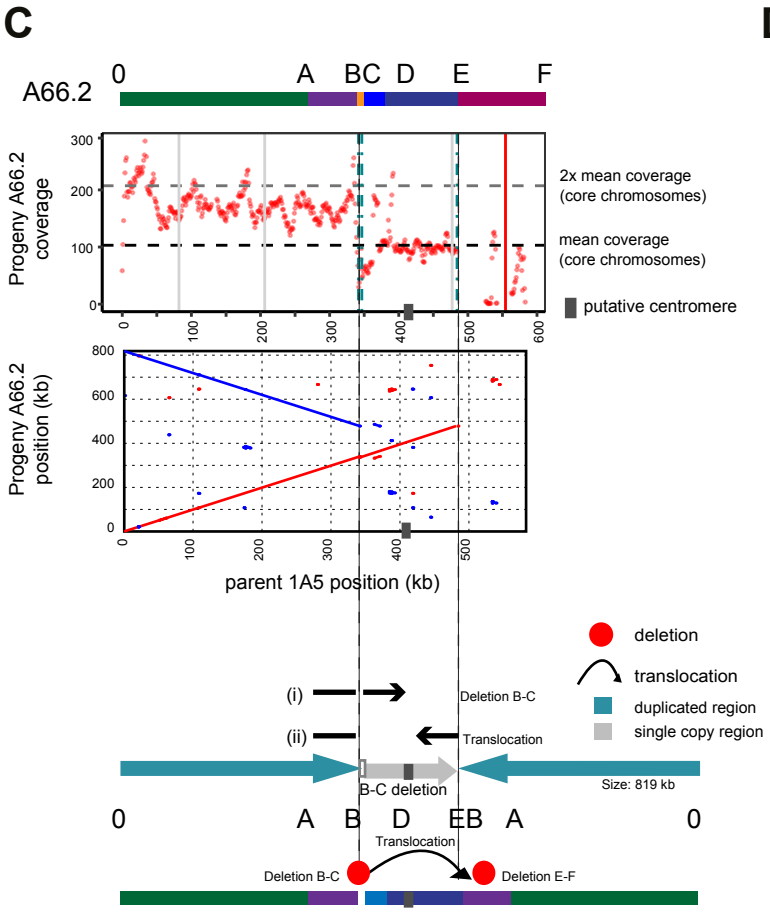
20. Feurtey, A. *et al.* Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria. *BMC Genomics* **21**, 588 (2020).
21. Ciullo, M. *et al.* Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. *Hum. Mol. Genet.* **11**, 2887–2894 (2002).
22. Hellman, A. *et al.* A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell* **1**, 89–97 (2002).
23. Marotta, M. *et al.* A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications. *Breast Cancer Res.* **14**, R150 (2012).
24. Marotta, M. *et al.* Homology-mediated end-capping as a primary step of sister chromatid fusion in the breakage-fusion-bridge cycles. *Nucleic Acids Res.* **41**, 9732–9740 (2013).
25. Marotta, M. *et al.* Palindromic amplification of the ERBB2 oncogene in primary HER2-positive breast tumors. *Sci. Rep.* **7**, 41921 (2017).
26. Bianchi, J. J., Murigneux, V., Bedora-Faure, M., Lescale, C. & Deriano, L. Breakage-Fusion-Bridge Events Trigger Complex Genome Rearrangements and Amplifications in Developmentally Arrested T Cell Lymphomas. *Cell Rep.* **27**, 2847-2858.e4 (2019).
27. Hoang, M. L. *et al.* Competitive Repair by Naturally Dispersed Repetitive DNA during Non-Allelic Homologous Recombination. *PLOS Genet.* **6**, e1001228 (2010).
28. Oggenfuss, U. & Croll, D. Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. *PLOS Pathog.* **19**, e1011130 (2023).
29. Umbreit, N. T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, (2020).
30. McClintock, B. The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes. *Genetics* **23**, 315–376 (1938).
31. McClintock, B. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* **26**, 234–282 (1941).
32. Toledo, F., Buttin, G. & Debatisse, M. The origin of chromosome rearrangements at early stages of AMPD2 gene amplification in Chinese hamster cells. *Curr. Biol.* **3**, 255–264 (1993).
33. Bi, X., Wei, S.-C. D. & Rong, Y. S. Telomere Protection without a Telomerase: The Role of ATM and Mre11 in Drosophila Telomere Maintenance. *Curr. Biol.* **14**, 1348–1353 (2004).
34. Rank, G. H., Xiao, W., Kolenovsky, A. & Arndt, G. FLP recombinase induction of the breakage-fusion-bridge cycle and gene conversion in Saccharomyces cerevisiae. *Curr. Genet.* **13**, 273–281 (1988).
35. Hackett, J. A., Feldser, D. M. & Greider, C. W. Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**, 275–286 (2001).
36. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175–186 (2017).
37. Smith, K. A., Gorman, P. A., Stark, M. B., Groves, R. P. & Stark, G. R. Distinctive chromosomal structures are formed very early in the amplification of CAD genes in Syrian hamster cells. *Cell* **63**, 1219–1227 (1990).
38. Ma, C., Martin, S., Trask, B. & Hamlin, J. L. Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. *Genes Dev.* **7**, 605–620 (1993).
39. Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. & Debatisse, M. Expression of Fragile Sites Triggers Intrachromosomal Mammalian Gene Amplification and Sets Boundaries to Early Amplicons. *Cell* **89**, 215–225 (1997).

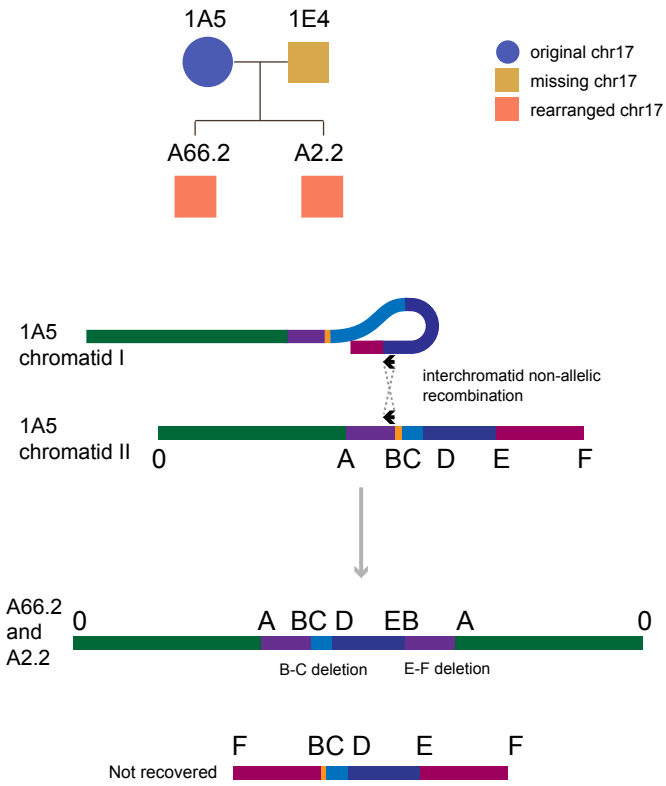
40. Badet, T., Fouché, S., Hartmann, F. E., Zala, M. & Croll, D. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. *Nat. Commun.* **12**, 3551 (2021).
41. Lendenmann, M. H., Croll, D., Stewart, E. L. & McDonald, B. A. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3 GenesGenomesGenetics* **4**, 2519–2533 (2014).
42. Kema, G. H., Verstappen, E. C., Todorova, M. & Waalwijk, C. Successful crosses and molecular tetrad and progeny analyses demonstrate heterothallism in *Mycosphaerella graminicola*. *Curr. Genet.* **30**, 251–258 (1996).
43. Goodwin, S. B. *et al.* Identification and genetic mapping of highly polymorphic microsatellite loci from an EST database of the septoria tritici blotch pathogen *Mycosphaerella graminicola. Fungal Genet. Biol.* **44**, 398–414 (2007).
44. Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S. & Thompson, W. F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1**, 2320 (2006).
45. McCluskey, K., Russell, B. W. & Mills, D. Electrophoretic karyotyping without the need for generating protoplasts. *Curr. Genet.* **18**, 385–386 (1990).
46. Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2001).
47. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
48. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
49. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
50. Seibt, K. M., Schmidt, T. & Heitkam, T. FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. *Bioinformatics* **34**, 3575–3577 (2018).
51. Feurtey, A. *et al.* Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria. *bioRxiv* 864561 (2019) doi:10.1101/864561.
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
54. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
55. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, 1–10 (2016).
56. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
57. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
58. Wang, L. G. *et al.* Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Mol. Biol. Evol.* **37**, 599–603 (2020).
59. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
60. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

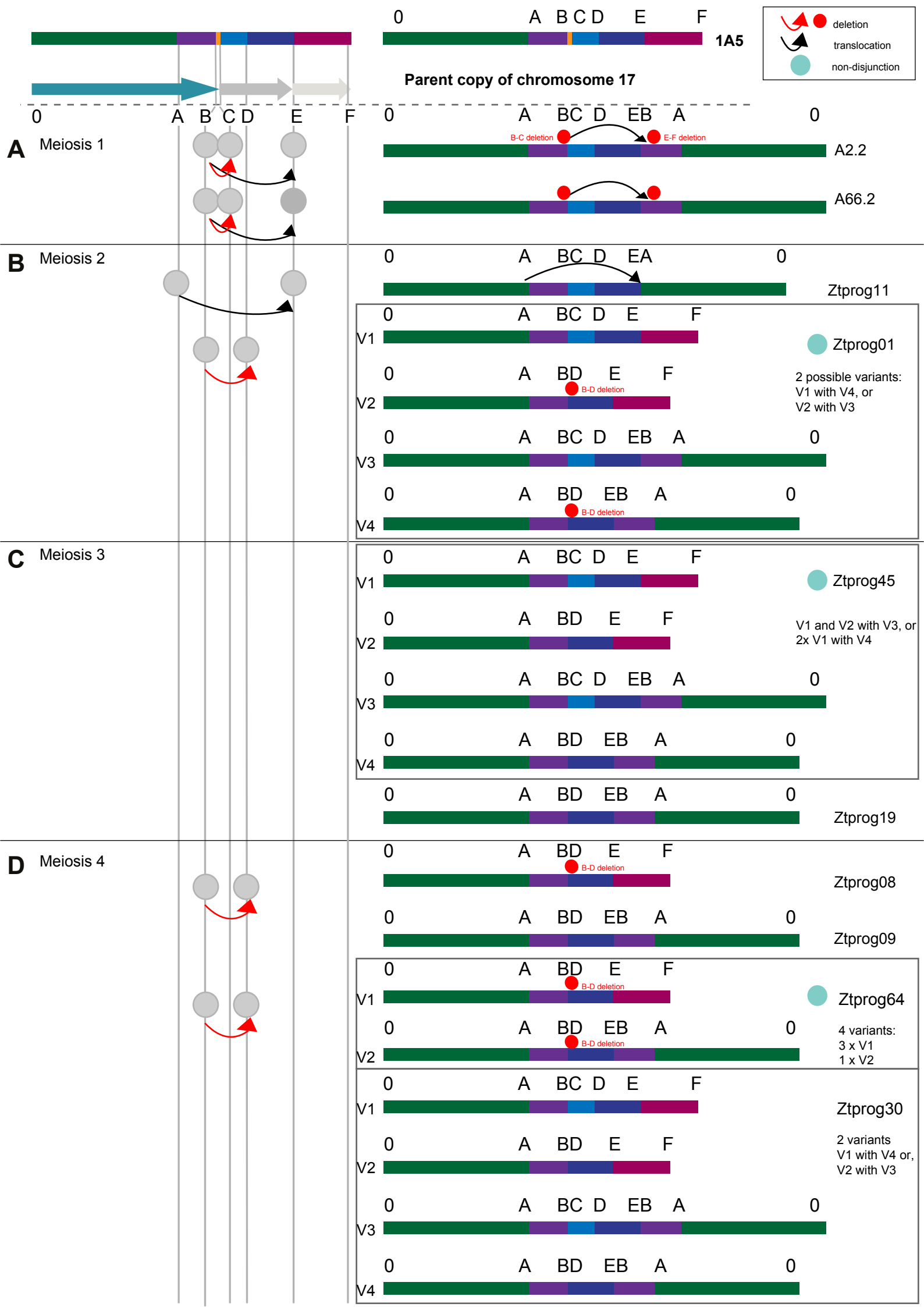
61. Yu, G., Lam, T. T. Y., Zhu, H. & Guan, Y. Two methods for mapping and visualizing associated data on phylogeny using GGTree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).
62. Kolpakov, R., Bana, G. & Kucherov, G. mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.* **31**, 3672–3678 (2003).
63. Edge, P. & Bansal, V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* **10**, 4660 (2019).

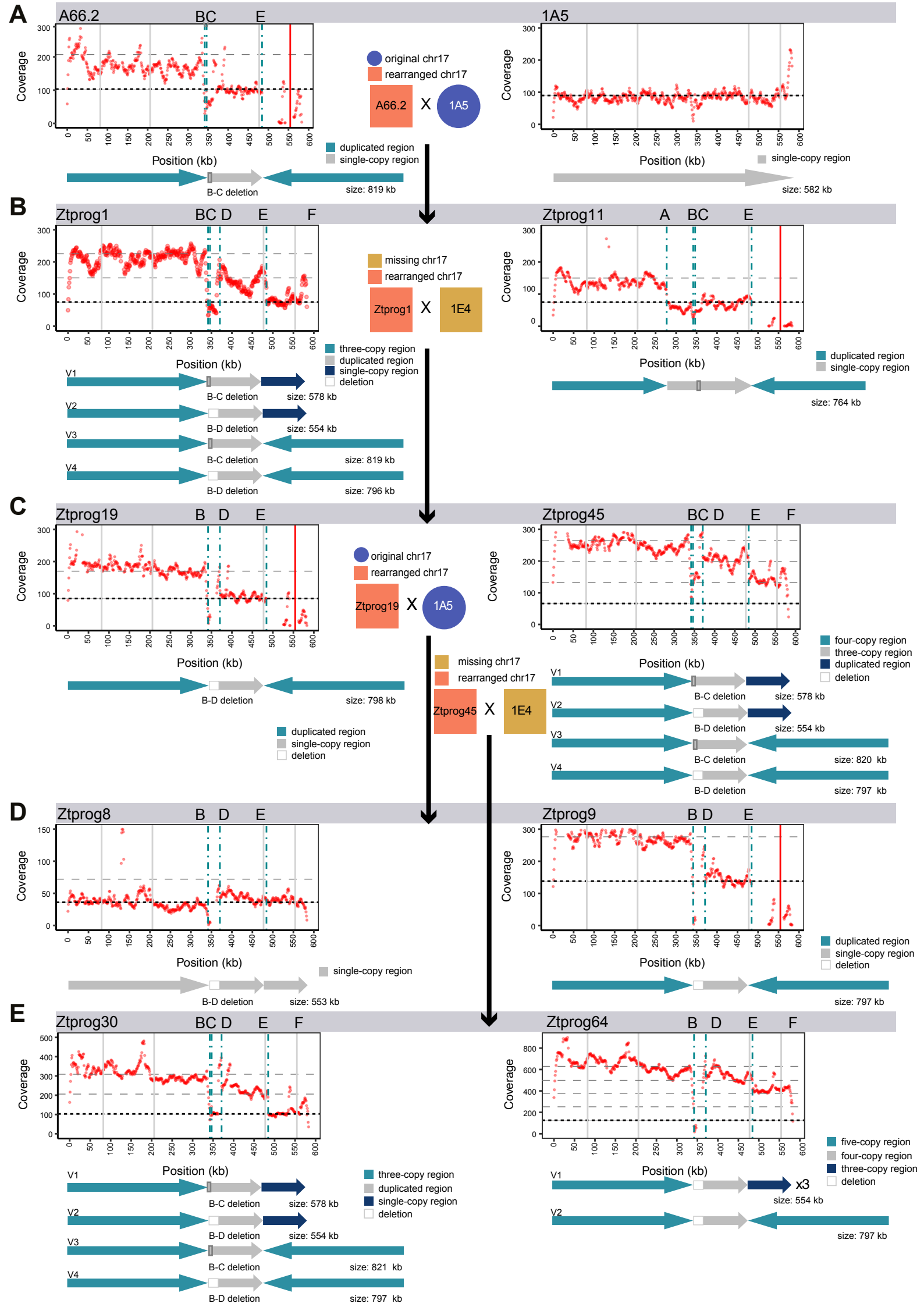

A
parental chr17
unrelated chr17
missing chr17
rearranged chr17
meiosis 1
meiosis 2
meiosis 3
meiosis 4
1A5
1E4
3D7
A2.2
A66.2
Ztprog02
Ztprog01
Ztprog11
Ztprog19
Ztprog45
Ztprog08
Ztprog09
Ztprog30
Ztprog64

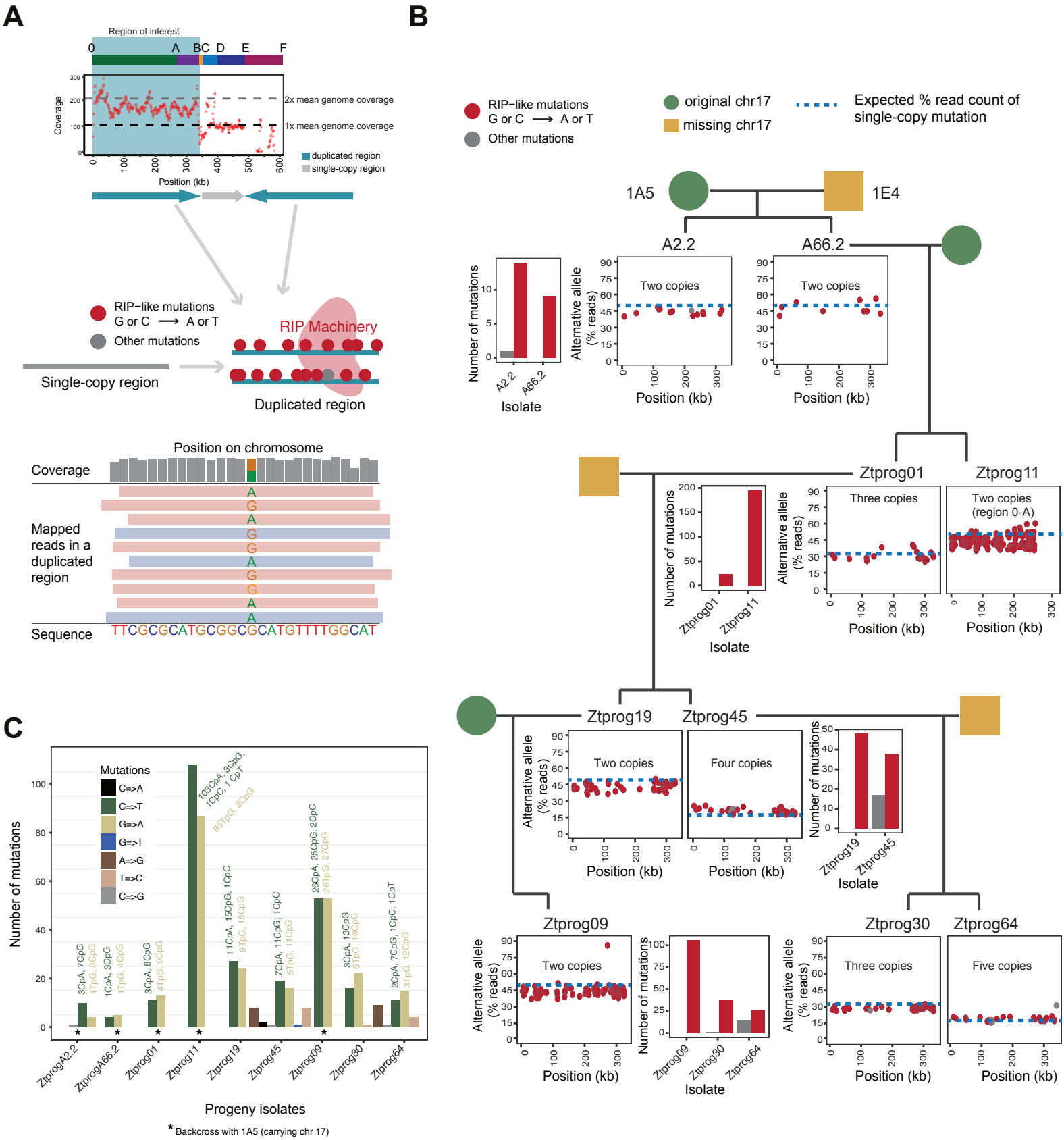

B
Progeny
Parents
A66.2 A2.2
1E4 1A5
original chr17
rearranged chr17
2.20
1.60
1.12
1.01
0.95
0.83
0.79
0.75
0.68
0.61
0.57
0.45
0.37
0.29
0.23
Size ladder (markers in Mb)

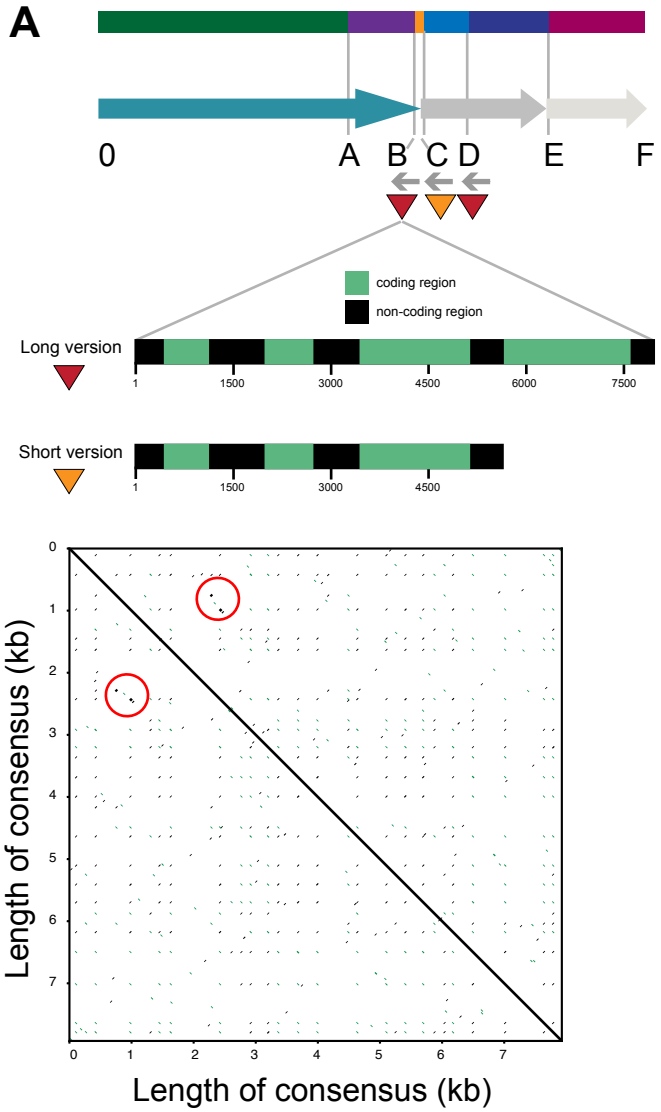

C
A66.2
0
A
BC
D
E
F
Progeny A66.2 coverage
2x mean coverage (core chromosomes)
mean coverage (core chromosomes)
putative centromere
Progeny A66.2 position (kb)
parent 1A5 position (kb)
deletion
translocation
duplicated region
single copy region
(i)
Deletion B-C
(ii)
Translocation
B-C deletion
Size: 819 kb
0
A
B
D
EB
A
0
Translocation
Deletion B-C
Deletion E-F

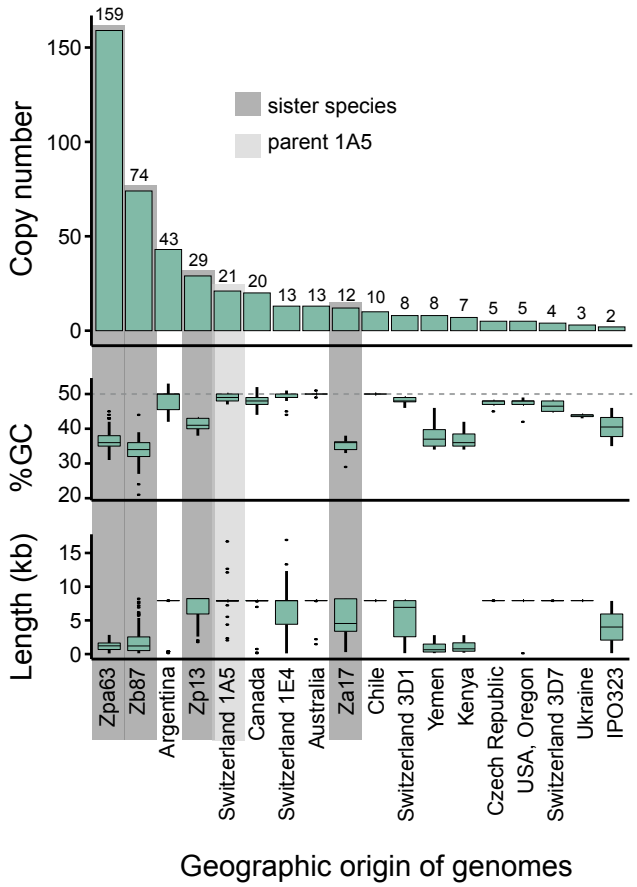

D
1A5
1E4
A66.2
A2.2
original chr17
missing chr17
rearranged chr17
1A5 chromatid I
interchromatid non-allelic recombination
1A5 chromatid II
0
A
BC
D
E
F
A66.2 and A2.2
0
A
BC
D
EB
A
0
B-C deletion
E-F deletion
Not recovered
F
BC
D
E
F

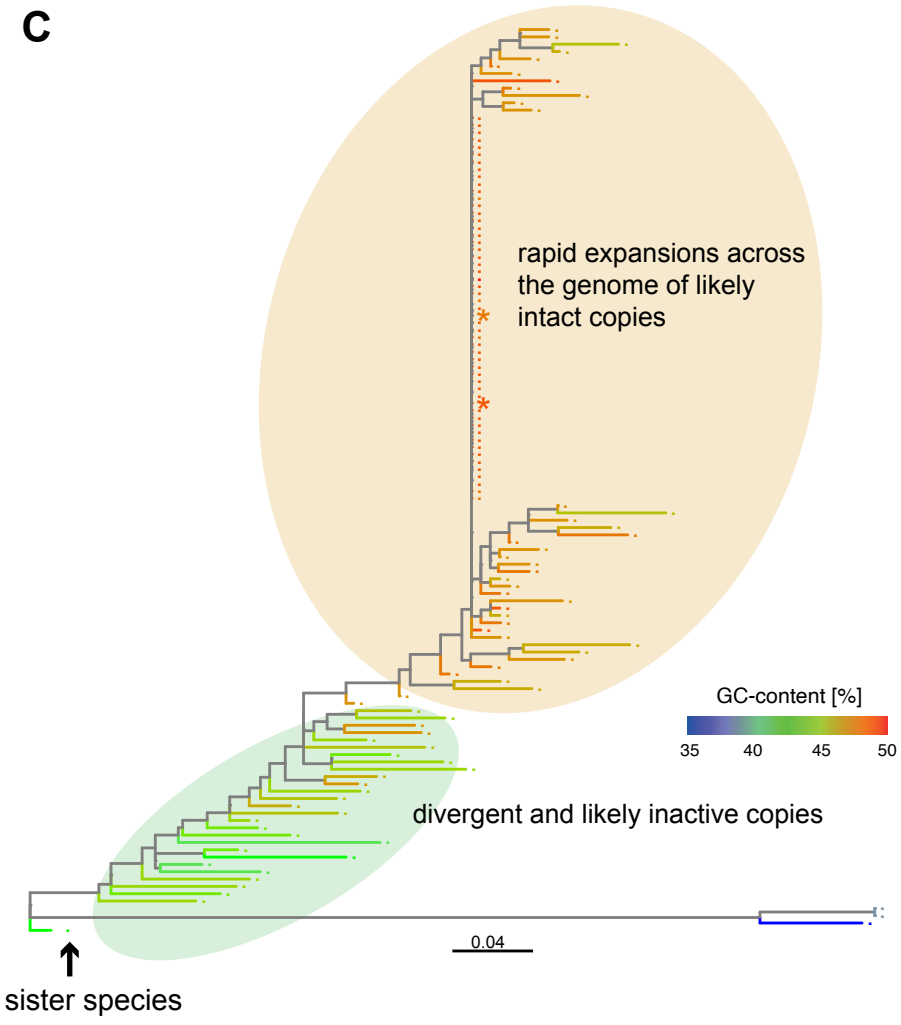

0 A B C D E F
1A5
deletion
translocation
non-disjunction
Parent copy of chromosome 17
A
Meiosis 1
B-C deletion
E-F deletion
A2.2
A66.2
B
Meiosis 2
Ztprog11
V1
V2
V3
V4
Ztprog01
2 possible variants:
V1 with V4, or
V2 with V3
B-D deletion
C
Meiosis 3
Ztprog45
V1 and V2 with V3, or
2x V1 with V4
Ztprog19
D
Meiosis 4
Ztprog08
Ztprog09
Ztprog64
4 variants:
3 x V1
1 x V2
Ztprog30
2 variants
V1 with V4 or,
V2 with V3

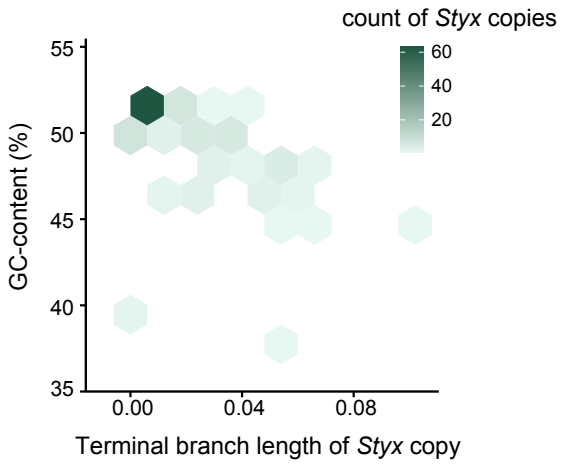

A
A66.2
BC
E
Coverage
Position (kb)
original chr17
rearranged chr17
A66.2
X
1A5
duplicated region
single-copy region
B-C deletion
size: 819 kb
1A5
single-copy region
size: 582 kb
B
Ztprog1
BC D
E
F
missing chr17
rearranged chr17
Ztprog1
X
1E4
three-copy region
duplicated region
single-copy region
deletion
V1
B-C deletion
size: 578 kb
V2
B-D deletion
size: 554 kb
V3
B-C deletion
size: 819 kb
V4
B-D deletion
size: 796 kb
Ztprog11
A
BC
E
duplicated region
single-copy region
size: 764 kb
C
Ztprog19
B
D
E
original chr17
rearranged chr17
Ztprog19
X
1A5
B-D deletion
size: 798 kb
duplicated region
single-copy region
deletion
Ztprog45
BC D
E
F
four-copy region
three-copy region
duplicated region
deletion
missing chr17
rearranged chr17
Ztprog45
X
1E4
V1
B-C deletion
size: 578 kb
V2
B-D deletion
size: 554 kb
V3
B-C deletion
size: 820 kb
V4
B-D deletion
size: 797 kb
D
Ztprog8
B
D
E
single-copy region
B-D deletion
size: 553 kb
Ztprog9
B D
E
duplicated region
single-copy region
deletion
size: 797 kb
E
Ztprog30
BC D
E
F
V1
B-C deletion
size: 578 kb
V2
B-D deletion
size: 554 kb
V3
B-C deletion
size: 821 kb
V4
B-D deletion
size: 797 kb
three-copy region
duplicated region
single-copy region
deletion
Ztprog64
B
D
E
F
five-copy region
four-copy region
three-copy region
deletion
V1
x3
size: 554 kb
V2
size: 797 kb


A
Region of interest
0
A
BC
D
E
F
Coverage
2x mean genome coverage
1x mean genome coverage
duplicated region
single-copy region
Position (kb)
RIP-like mutations
G or C ⟶ A or T
Other mutations
RIP Machinery
Single-copy region
Duplicated region
Position on chromosome
Coverage
Mapped reads in a duplicated region
Sequence
TTCGCGCATGCGGCGCATGTTTTGGCAT
B
RIP-like mutations
G or C ⟶ A or T
Other mutations
original chr17
missing chr17
Expected % read count of single-copy mutation
1A5
1E4
A2.2
A66.2
Number of mutations
Isolate
Alternative allele (% reads)
Two copies
Position (kb)
Ztprog01
Ztprog11
Three copies
Two copies (region 0-A)
Ztprog19
Ztprog45
Four copies
Ztprog09
Ztprog30
Ztprog64
Five copies
C
Mutations
C=>A
C=>T
G=>A
G=>T
A=>G
T=>C
C=>G
Number of mutations
3CpA, 7CpG
1TpG, 3CpG
1CpA, 3CpG
1TpG, 4CpG
3CpA, 8CpG
4TpG, 9CpG
103CpA, 3CpG, 1CpC, 1 CpT
85TpG, 2CpG
11CpA, 15CpG, 1CpC
9TpG, 15CpG
7CpA, 11CpG, 1CpC
5TpG, 11CpG
26CpA, 25CpG, 2CpC
26TpG, 27CpG
3CpA, 13CpG
6TpG, 16CpG
2CpA, 7CpG, 1CpC, 1CpT
3TpG, 12CpG
ZtprogA2.2
ZtprogA66.2
Progeny isolates
* Backcross with 1A5 (carrying chr 17)


A
0
A B C D E F
coding region
non-coding region
Long version
Short version
1
1500
3000
4500
6000
7500
Length of consensus (kb)
Length of consensus (kb)


B
Copy number
sister species
parent 1A5
159
74
43
29
21
20
13
13
12
10
8
8
7
5
5
4
3
2
%GC
Length (kb)
Zpa63
Zb87
Argentina
Zp13
Switzerland 1A5
Canada
Switzerland 1E4
Australia
Za17
Chile
Switzerland 3D1
Yemen
Kenya
Czech Republic
USA, Oregon
Switzerland 3D7
Ukraine
IPO323
Geographic origin of genomes


C
rapid expansions across
the genome of likely
intact copies
GC-content [%]
35
40
45
50
divergent and likely inactive copies
0.04
sister species


D
count of Styx copies
60
40
20
GC-content (%)
55
50
45
40
35
0.00
0.04
0.08
Terminal branch length of Styx copy