The genomic landscape of transposable elements in yeast hybrids is shaped by structural variation and genotype-specific modulation of transposition rate

Mathieu Hénault[1245]*, Souhir Marsit[12345†], Guillaume Charron[1345‡] and Christian R. Landry[12345]

[1]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada

[2]Département de biochimie, microbiologie et bioinformatique, Université Laval, Québec, QC, Canada

[3]Département de biologie, Université Laval, Québec, QC, Canada

[4]Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO), Université Laval, Québec, QC, Canada

[5]Université Laval Big Data Research Center (BDRC_UL), Québec, QC, Canada

[†]Current address: Département de biologie, chimie et géographie, Université du Québec à Rimouski, Rimouski, QC, Canada

[‡]Current address: Centre de foresterie des Laurentides, Ressources naturelles Canada, Québec, QC, Canada

*Corresponding author

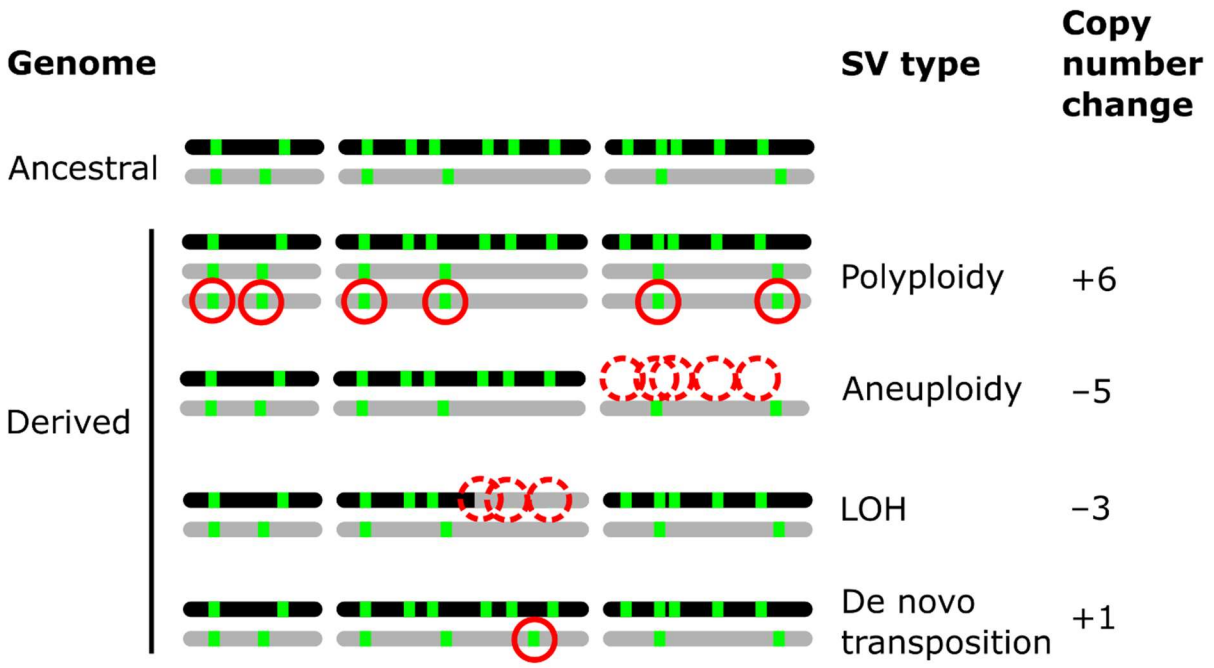
E-mail: mathieu.henault.1@ulaval.ca

Abstract

Transposable elements (TEs) are major contributors to structural genomic variation by creating interspersed duplications of themselves. In return, structural variants (SVs) can affect the genomic distribution of TE copies and shape their load. One long-standing hypothesis states that hybridization could trigger TE mobilization and thus increase TE load in hybrids. We previously tested this hypothesis by performing a large-scale evolution experiment by mutation accumulation (MA) on multiple hybrid genotypes within and between wild populations of the yeasts *Saccharomyces paradoxus* and *Saccharomyces cerevisiae*. Using aggregate measures of TE load with short-read sequencing, we found no evidence for TE load increase in hybrid MA lines. Here, we resolve the genomes of the hybrid MA lines with long-read phasing and assembly to precisely characterize the role of SVs in shaping the TE landscape. Highly contiguous phased assemblies of 127 MA lines revealed that SV types like polyploidy, aneuploidy and loss of heterozygosity have large impacts on the TE load. We characterized 18 de novo TE insertions, indicating that transposition only has a minor role in shaping the TE landscape in MA lines. Because the scarcity of TE mobilization in MA lines provided insufficient resolution to confidently dissect transposition rate variation in hybrids, we adapted an in vivo assay to measure transposition rates in various *S. paradoxus* hybrid backgrounds. We found that transposition rates are not increased by hybridization, but are modulated by many genotype-specific factors including initial TE load, TE sequence variants and mitochondrial DNA inheritance. Our results show the multiple scales at which TE load is shaped in hybrid genomes, being highly impacted by SV dynamics and finely modulated by genotype-specific variation in transposition rates.

Introduction

Structural variants (SVs) are a powerful source of genetic variation that fuels the evolution of genomes and species (Gorkovskiy and Verstrepen, 2021). SVs regroup a wide variety of large-scale sequence alterations that can have profound consequences for organismal phenotypes and genome evolution. For example, aneuploidies can underlie the emergence of antifungal resistance in fungal pathogens (Selmecki et al., 2006) and drive tumorigenesis in mammalian cells (Shoshani et al., 2021; Trakala et al., 2021). Polyploidization is frequently linked to speciation and adaptation, notably in plants (Otto and Whitton, 2000). Mitotic recombination can yield loss of heterozygosity (LOH) events that drive adaptation in heterozygous hybrid genomes (Smukowski Heil et al., 2017), but also the somatic fixation of oncogene mutations (Cavenee et al., 1983). Repeated sequences are known to play a leading role in the generation of SVs (Lower et al., 2019). Notably, transposable elements (TEs) are sequences capable of semi-autonomous replication, producing families of repeats spread within genomes (Bourque et al., 2018). Because of their replicative nature, TEs are major generators of SVs: not only do they create interspersed duplications of themselves, but they can also trigger genome rearrangements through ectopic recombination (Dunham et al., 2002; Petrov et al., 2003). While TEs can cause SVs, various types of SVs can also affect the genomic landscape of TEs by producing gains and losses of individual copies (Figure 1).



60

61 **Figure 1. Examples of how SVs can impact the genomic landscapes of TEs in hybrids.** An
62 ancestral heterozygous genome is shown as an example, with two subgenomes (black and gray)
63 harboring TEs (green boxes) in distinct copy numbers and at different positions. Examples of
64 various types of SVs are shown in derived genomes: polyploidy (gain of haploid sets of
65 chromosomes), aneuploidy (gain or loss of whole chromosomes), loss of heterozygosity (LOH,
66 inter-homolog conversion of chromosome segments), and de novo transposition. The
67 corresponding changes in TE copy numbers are shown at the right.

68 TEs are inherent components of virtually all Eukaryotic genomes and owe their evolutionary
69 success to the fitness advantage of their own replicative ability. However, this selfish propagation
70 conflicts with the fitness of the host, and can have profound deleterious impacts (Bucheton et al.,
71 1984; Pasyukova et al., 2004; Wilke and Adams, 1992). The abundance of TEs per genome (or
72 TE load) is an important population genetic parameter, as it is expected to correlate with the
73 fitness cost for the host. TE load in a population is most often modeled as a quantity reflecting the
74 balance between the gain of novel copies by transposition, and the decrease in copies by loss
75 (e.g. by excision or sequence degeneration) and purifying selection against insertion alleles
76 (Charlesworth et al., 1994; Charlesworth and Langley, 1989). However, TE load is not only
77 defined by the copy number (CN) of TEs in the genome, but also where TE insertions fall and,
78 consequently, how they interact with neighboring host sequences. The multidimensional nature

of TE load makes its full characterization a complex task. Recent advances in long-read sequencing make it increasingly possible to decompose TE load by resolving the specific genomic context in which TE insertions are found (Oggenfuss and Croll, 2023; Rech et al., 2022). Additionally, accurate genome resolution opens unprecedented opportunities for a comprehensive assessment of the role of SVs in shaping TE load.

One long-standing hypothesis postulates that hybridization can cause a genomic shock that triggers TE mobilization and increases TE load in hybrids (McClintock, 1984), with major implications for speciation and hybrid evolution. We previously tested this hypothesis in *Saccharomyces* yeast hybrids by performing a large-scale evolution experiment by mutation accumulation (MA) to investigate the near-neutral evolution of TEs in hybrid genomes (Hénault et al., 2020). *Saccharomyces cerevisiae* genomes harbor five main long terminal repeat (LTR) retrotransposon families named Ty1-Ty5 (Kim et al., 1998). Its undomesticated sister species *Saccharomyces paradoxus* comprises related families Ty1, Ty3 and Ty5 (Yue et al., 2017), in addition to Tsu4 which was horizontally transferred from *Saccharomyces uvarum* (Bergman, 2018). We generated a collection of MA lines comprising a set of 11 hybrid genotypes created by crossing wild strains from North American populations of *S. paradoxus* and *S. cerevisiae* (Charron et al., 2014a; Kuehne et al., 2007; Leducq et al., 2016, 2014) that are separated by a range of evolutionary divergence, and by replicating each cross many dozen times independently (Charron et al., 2019; Hénault et al., 2020). The resulting collection was evolved by subjecting each line to periodical extreme bottlenecks by streaking for single colonies on plates of rich medium for ~770 mitotic generations, thus amplifying the power of random genetic drift. We previously sequenced these MA lines with short reads and measured the changes in Ty load after evolution by producing aggregate estimates of relative abundance for each family, and found no support for the TE reactivation hypothesis (Hénault et al., 2020). However, the use of short reads alone precluded the detailed resolution of hybrid genomes and the decomposition of their Ty load. For instance, different SVs (including de novo transposition) may have changed Ty CNs in opposing directions, yielding no significant change in net CN.

While producing collapsed assemblies for haploid or predominantly homozygous genomes is now relatively straightforward, complex heterozygous genomes are much more difficult to characterize. Recent studies demonstrated the feasibility of the approach (Heasley and Argueso, 2022; O'Donnell et al., 2022), but the task of resolving hybrid genomes across a range of parental divergence under a unified comparative framework remains challenging. Here, we design a long-read sequencing, phasing and assembly approach to resolve hybrid MA lines genomes. We find

112 that de novo transposition events play a minor role in shaping the Ty load in comparison to other
113 SV types. Since the number of de novo insertions gained during MA is low, we further dissect
114 transposition rate variation in *S. paradoxus* hybrids at a higher resolution with in vivo transposition
115 rate measurements. We identify many aspects of hybrid genotype specificity that shape
116 transposition rates, including initial Ty load, Ty sequence variation and mitochondrial DNA
117 (mtDNA) inheritance.

118

Results

The eleven MA crosses (Charron et al., 2019; Hénault et al., 2020) comprise five crosses within *S. paradoxus* populations (CC, within *SpC*; BB, within *SpB*), four crosses between *S. paradoxus* populations (BC, *SpB*×*SpC*; BA, *SpB*×*SpA*), and two interspecific crosses (BSc, *S. paradoxus SpB*×*S. cerevisiae*) (Figure 2A). We previously selected 127 independently evolved MA lines (10-12 per cross) and sequenced their genome with Oxford Nanopore long reads (Hénault et al., 2022). To produce an accurate representation of the two distinct subgenomes of each evolved MA line, we designed a strategy consisting of 1) sorting long-read libraries using variants that discriminate the two parental genomes of each cross, and 2) performing separate de novo assembly on the sorted reads. The proportion of sequenced bases successfully classified into subgenomes varied substantially depending on the cross (Figure 2B). CC crosses had the lowest classification rate (median: 70.8%) due to the scarcity of variants that discriminate *SpC* genomes from each other. BB crosses had a median classification rate of 84.5%, while the remainder (BC, BA and BSc) reached a higher rate (median: 93.7%). Unclassified reads tended to be shorter than classified reads, but the difference was narrower for libraries with lower classification success (Supplemental Figure S1). The scarcity of discriminating variants in low-divergence crosses is likely causing the classification of increasingly longer reads to fail. However, unclassified reads showed no overrepresentation in specific genomic regions (Supplemental Figure S2) and their exclusion is thus unlikely to bias the resulting assemblies.

Despite the variation in classification rates, the classified reads yielded highly contiguous subgenome-level assemblies, with a median of 22 contigs, a median N50 of 783 kb, and 95% of the assemblies having N50 values larger than 200 kb (Figure 2C, Supplemental Table S1). Only six assemblies were more fragmented (>100 contigs) and corresponded to the sorted libraries with the least sequenced bases (Supplemental Figure S3A), which are expected to complicate the assembly. Thus, our approach yielded chromosome-level representations of the two subgenomes of each MA line.

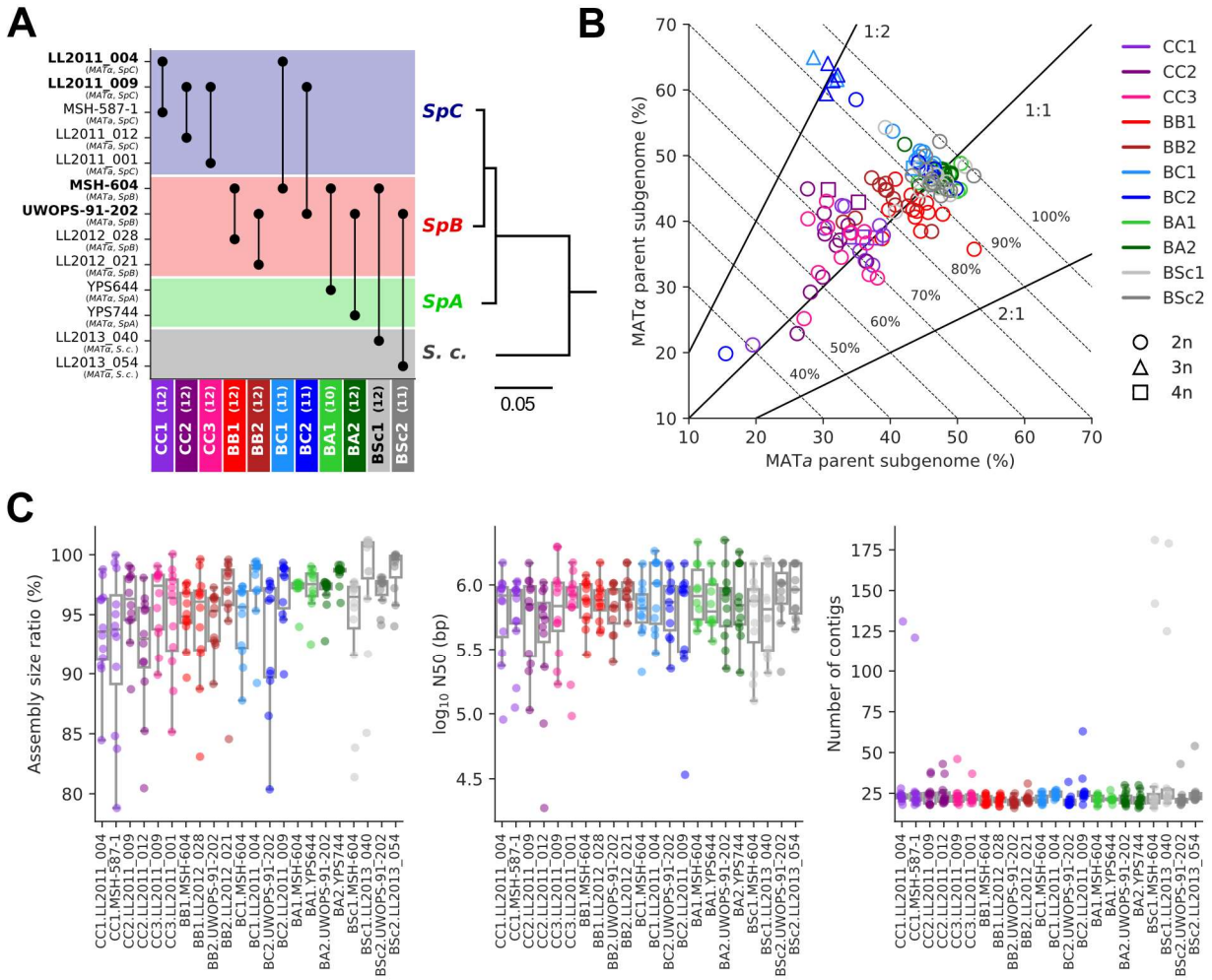


Figure 2. Subgenome-level resolution of hybrid yeast MA lines genomes using a long-read phasing and assembly approach. (A) Design of the MA crosses. Haploid derivatives of wild isolates used as parental strains for the MA crosses are shown on the Y-axis. Strains in bold are shared by multiple crosses. MA crosses are shown on the X-axis (left to right, from the least to the most divergent), with the number of lines randomly sampled for long-read sequencing shown in parentheses. The phylogenetic tree displays the evolutionary relatedness between wild populations of *S. paradoxus* and *S. cerevisiae*, based on genome-wide nucleotide variants (scale bar: substitutions per site) (Hénault et al., 2022). **(B)** Summary of the classification of long reads in two parental subgenomes. For each MA line library, the percent of sequenced bases classified as the *MAT*a and *MAT*α parental subgenomes are shown on the X and Y axes, respectively. Dotted lines indicate global classification rates. Full lines indicate expected subgenome classification ratios of 1:1 for diploid (circles) or tetraploid (squares) lines, and 2:1 or 1:2 for triploid

159 (triangles) lines. **(C)** Summary statistics of the de novo subgenome-level assemblies of the MA
160 lines.

161 We then integrated various data to infer the state of all loci containing Ty elements in the MA lines
162 subgenomes in relation to the corresponding parental genome. First, we produced whole-genome
163 assemblies of the 13 haploid parental strains from long reads. We annotated their Ty elements
164 and used whole-genome alignments to determine the orthology of Ty loci between the parental
165 strains of each cross (Supplemental Figure S4). Only 10 pairs of loci were found to be orthologous
166 in four of the lowest divergence crosses (CC1: 1, CC2: 5, CC3: 3, BB1: 1). We also annotated the
167 Ty elements in the MA lines subgenome assemblies and defined the orthology between Ty loci
168 by aligning subgenome assemblies against the corresponding parental assembly. We also
169 computed the depth of coverage of sorted long-read libraries mapped on the corresponding
170 parental genome assemblies to yield estimates of CN per genomic window. We then applied a
171 decision tree on this data to classify individual Ty loci by type of SV (Supplemental Figure S5).
172 89.8% of Ty loci were successfully classified, with low-contiguity assemblies contributing
173 disproportionately to classification failure (Supplemental Figure S3B). From the successfully
174 classified loci, 79.0% had no change compared to the parental genomes. The remainder had an
175 uneven distribution of SV classes (Figure 3A). The largest class was polyploidy, since many lines
176 within CC and BC crosses have become tetraploid or triploid (Charron et al., 2019; Marsit et al.,
177 2021). Triploidization was exclusive to six BC lines and was likely caused by the diploidization of
178 *SpC* parental cells (creating mating-competent pseudohaploids) prior to hybridization (Charron et
179 al., 2019), meaning that triploidization occurred before MA strictly speaking. LOH and aneuploidy
180 also accounted for large fractions of Ty loci affected by SVs. Notably, there was a substantial
181 impact of LOH events in low-divergence CC crosses. Ty1 is the most abundant family in *SpC*
182 genomes and despite a relatively constant load, Ty1 copies mostly occupy non-orthologous loci
183 (Supplemental Figure S4A), providing the substrate for LOH to affect Ty1 landscapes in CC
184 crosses. While LOH had a bidirectional effect on Ty CNs, causing gains and losses in
185 approximately equal proportions, aneuploidies had a strong bias towards gains (Figure 3B). The
186 distributions of net Ty CN change per MA line showed that most crosses had significant gains
187 (Figure 3C). The post-evolution Ty CNs quantified from relative depth of coverage of short-read
188 libraries (Hénault et al., 2020) and the net CN changes determined here were globally correlated
189 (mixed linear model with random intercepts for MA cross×Ty family combinations, P-
190 value=3.01×10^{-35}). In many individual cases, significant positive correlations were observed
191 (Supplemental Figure S6), indicating that our current analyses recapitulate previous results.

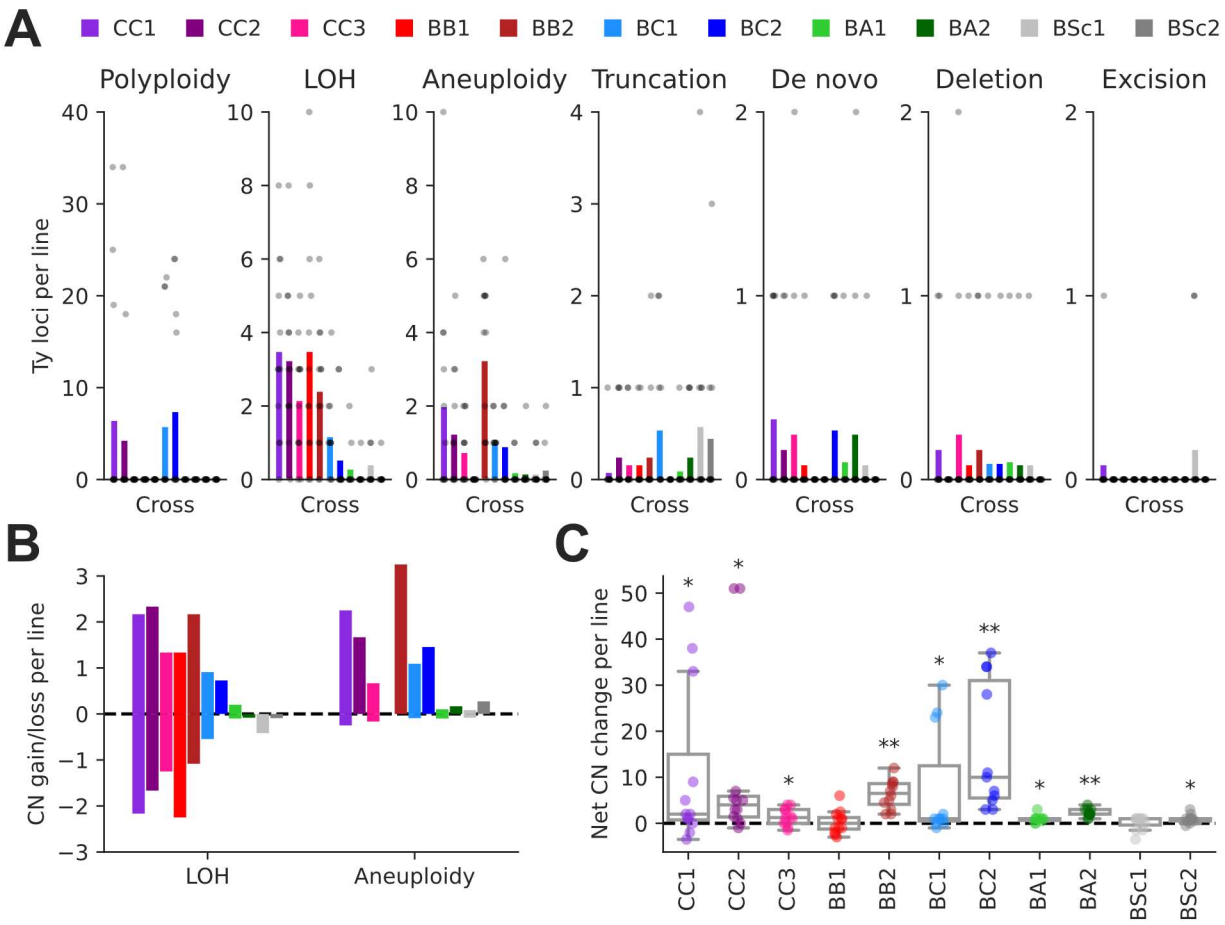


Figure 3. Distribution of the number of Ty loci affected by various SV classes. (A) Number of Ty loci affected by SVs per line for each cross. Dots represent individual MA lines and bars show the average count per cross. Polyploidy refers to increased ploidy level to triploidy or tetraploidy from the expected diploid initial state. LOH refers to the conversion of large chromosome segments (interstitial or terminal) from one parental genotype to the other. Aneuploidy refers to gains or losses of whole chromosomes. Truncation refers to partial deletion of a full-length Ty copy. De novo refers to the presence of a new full-length Ty copy not found in the parental background or any other MA line. Deletion refers to the local deletion of a full-length Ty copy without other SVs. Excision refers to the replacement of a full-length Ty copy with a matching solo LTR, consistent with intra-element LTR-LTR recombination. **(B)** Average gain and loss of Ty CN per line for each cross, for SV classes that can have bidirectional effects on Ty CN. **(C)** Distributions of net Ty CN change for individual MA lines. Symbols indicate FDR-corrected P-values for Wilcoxon signed-rank tests (*: p≤0.05, **: p≤0.01).

206 We identified 23 candidate de novo Ty insertions as loci occupied by one full-length Ty element
207 that was absent from the corresponding parental genome and any other MA line. From these
208 candidates, we used strict filtering criteria to narrow down confident loci that were compatible with
209 retrotransposition. First, we required the presence of characteristic target site duplications (TSDs)
210 of 5-6 nt (Supplemental Figure S7), consistent with the staggered dsDNA cut catalyzed by the
211 Ty1 integrase (Wilhelm et al., 2005). We further confirmed the uniqueness of Ty insertion junction
212 sequences by inspecting whole-chromosome alignments comprising the MA line subgenomes
213 and all parental genomes. We also ensured that all de novo TSD sequences were unique and
214 absent from the parental full-length Ty loci.

215 We determined that 18 de novo loci were compatible with all these criteria (Figure 4). The
216 quantitative impact of de novo insertions on Ty load evolution is thus minor in comparison to other
217 SV types, with 0.14 insertions per line on average (Figure 3A). All de novo loci belonged to the
218 Ty1 family of *S. paradoxus*. Eight out of 11 crosses exhibited de novo Ty1 retrotransposition
219 events, half of them being found in the CC crosses. Eight insertions were found at loci with
220 parental Ty1 elements (full-length or solo-LTR) in close proximity, highlighting the importance of
221 the de novo assembly approach to accurately resolve these complex loci. Two retrotransposition
222 events in BC2 lines inserted into the UWOPS-91-202 *SpB* subgenome that is natively devoid of
223 full-length elements, effectively re-colonizing a genome in which all Ty families had gone extinct.
224 We built a maximum likelihood phylogenetic tree comprising the de novo insertions and all
225 parental Ty1 and Ty2 full-length sequences (Supplemental Figure S8). This analysis revealed
226 that de novo copies originated from major Ty1 clades populating *SpC*, *SpB* and *SpA* genomes,
227 indicating that Ty1 is globally active in *S. paradoxus* populations. However, this tree had generally
228 poor branch support values from the bootstrap analysis. A phylogenetic network built from the
229 same dataset showed substantial signal for reticulate evolution between Ty1 families of
230 *S. cerevisiae* and *S. paradoxus SpA* (Supplemental Figure S9), in agreement with previous
231 studies that showed interspecific horizontal transfers (Bleykasten-Grosshans et al., 2021; Czaja
232 et al., 2020).

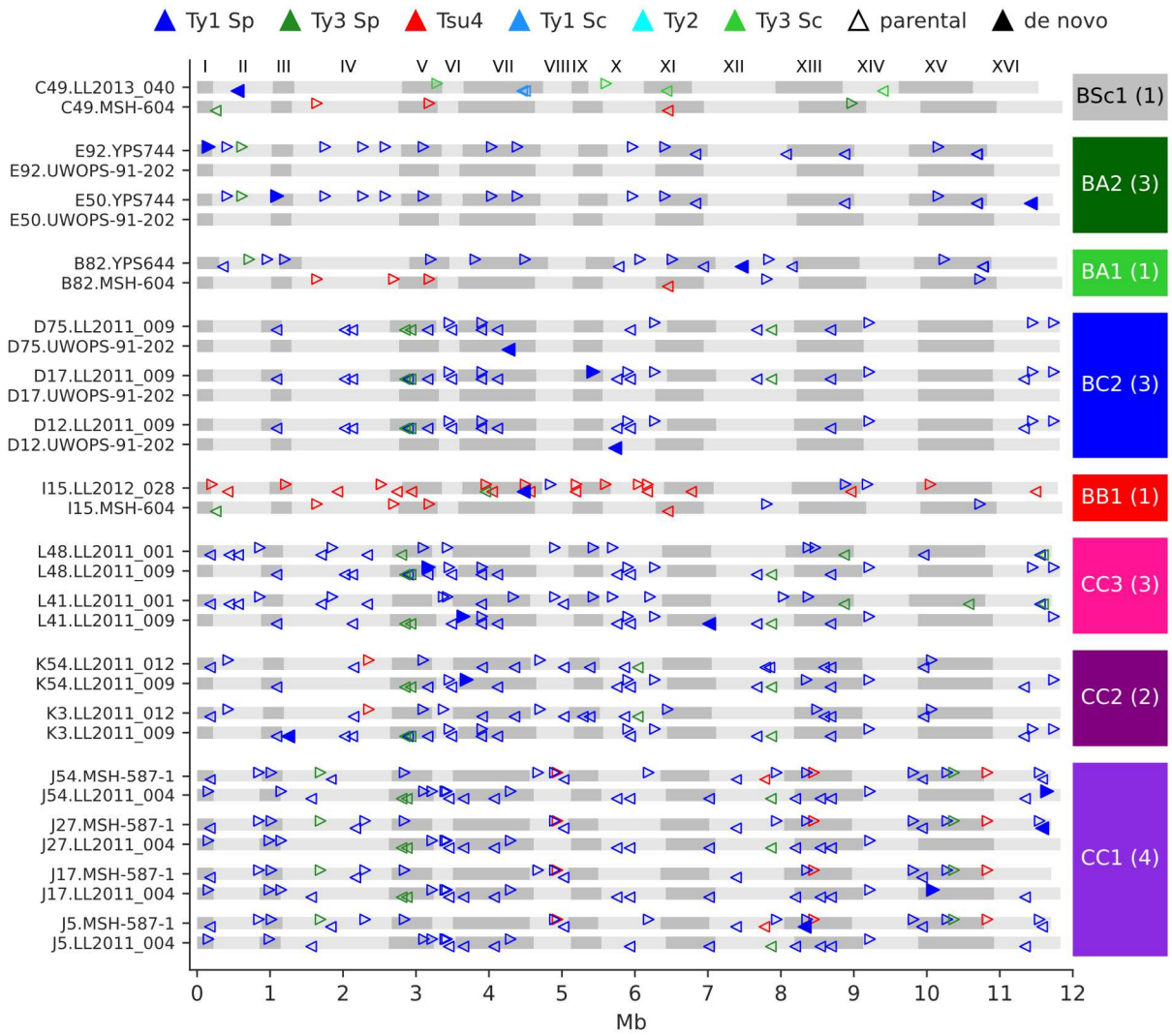


Figure 4. De novo retrotransposition events in the MA lines. Genome maps show the full-length Ty elements annotated in the subgenome assemblies of the MA lines that have de novo insertions. Gray boxes represent chromosomes, which are numbered at the figure top. Empty triangles represent the parental annotations present at a CN greater or equal to one. Full triangles correspond to the confident de novo retrotransposition events. Triangle orientation indicates whether the annotation is on the plus (right pointing) or minus (left pointing) chromosome strand. The cross corresponding to each MA line is labeled at the right, with the total count of de novo retrotransposition events shown between parentheses.

The sequence resolution of de novo loci also enabled the characterization of insertions that disrupted parental genomic features. One de novo Ty1 element inserted into the left LTR of a parental full-length Ty1 copy in the MSH-587-1 subgenome of the MA line J27 (CC1 cross,

245 Supplemental Figure S10A), resulting in a nested insertion. Additionally, in a single case, a de
246 novo Ty1 element inserted 37 bp downstream of the start codon of the host gene *AIM29* in the
247 LL2011_009 subgenome of the MA line L41 (CC3 cross, Supplemental Figure S10B), introducing
248 an early stop codon that yields a peptide of 36 amino acid residues.

249 The frequency of de novo Ty1 retrotransposition events per cross tends to support our previous
250 conclusion that TE mobilization does not scale positively with parental divergence in
251 *Saccharomyces* hybrids (Hénault et al., 2020). A similar conclusion was reached from
252 interspecific hybrids between *S. cerevisiae* and the more distantly related species *S. uvarum*
253 (Smukowski Heil et al., 2021). However, the number of de novo Ty1 insertions characterized by
254 MA are quite low (at most four per cross) making it difficult to confidently test for differences in
255 transposition rate between hybrid genotypes. Multiple aspects of natural genetic variation were
256 shown to impact the mobilization of Ty1 (Czaja et al., 2020; Smukowski Heil et al., 2021),
257 suggesting that some of these aspects could also contribute to modulate transposition rate in
258 *S. paradoxus* hybrids.

259 We investigated how multiple aspects of genotype specificity impact transposition rate in
260 *S. paradoxus* hybrids using an existing genetic assay developed in *S. cerevisiae* to measure the
261 retrotransposition rate of a single tester Ty1 element (Curcio and Garfinkel, 1991). This assay
262 relies on an antisense artificial intron (AI) that interrupts a copy of the *HIS3* gene, which is itself
263 inserted in antisense orientation at the 3' end of the internal ORF of a plasmid-borne Ty1 element.
264 The correct splicing of the AI is enabled by Ty1 transcription, following which the reverse
265 transcription of the Ty1 mRNA and the integration of the resulting dsDNA copy in the genome
266 restores the *HIS3* ORF. Selecting for histidine prototrophy in a fluctuation assay (Luria and
267 Delbrück, 1943) enables the estimation of the transposition rate of the tester element. We adapted
268 this assay to measure retrotransposition rates in multiple *S. paradoxus* backgrounds, and with
269 two Ty1 sequence variants sampled from *SpB* and *SpC* genomes as tester elements
270 (Supplemental Figure S11, Supplemental Tables S2-6).

271 A mechanism of negative copy-number dependent self-regulation of retrotransposition was
272 characterized in the Ty1 family of *S. cerevisiae*, termed copy-number control (CNC) (Garfinkel et
273 al., 2003). This mechanism relies on the expression of an N-truncated Ty1 Gag protein (p22) from
274 two alternative downstream start codons (Nishida et al., 2015; Saha et al., 2015). p22 interferes
275 with the assembly of the viral-like particles essential for Ty1 life cycle completion (Cottee et al.,
276 2021; Saha et al., 2015). CNC yields a steep negative relationship between the retrotransposition

277 rate measured with a tester element and the number of p22-expressing Ty1 copies in the genome
278 (Garfinkel et al., 2003). We previously observed a negative relationship between Ty1 load change
279 after MA and the initial Ty1 load, and hypothesized that CNC-like regulation could be at play
280 (Hénault et al., 2020). While the scarcity of Ty1 mobilization in MA lines now refutes this
281 hypothesis, transposition rates could still be impacted by the initial load. We measured
282 transposition rates for Ty1 elements in eight different haploid backgrounds of *S. paradoxus* (one
283 *SpA*, four *SpB*, and three *SpC*; Supplemental Tables S5,7) that exhibit a wide variation in genomic
284 Ty1 CNs. The transposition rate in *SpB* backgrounds had a strong negative relationship with Ty1
285 CN (Figure 5). The transposition rate in a *SpB* background with a single Ty1 copy (UWOPS-79-
286 140, estimated CN: 1.2) reached 1.73×10^{-6} cell^{-1} gen^{-1}. In comparison, an *SpB* background
287 harboring a single additional Ty1 copy (MSH-604, estimated CN: 2.2) had a transposition rate
288 over an order of magnitude lower ($9{,}77 \times 10^{-8}$ cell^{-1} gen^{-1}). In *SpC* backgrounds, which all have
289 higher Ty1 CNs (20.7-32.4), the transposition rate was below 1×10^{-8} cell^{-1} gen^{-1}.

290

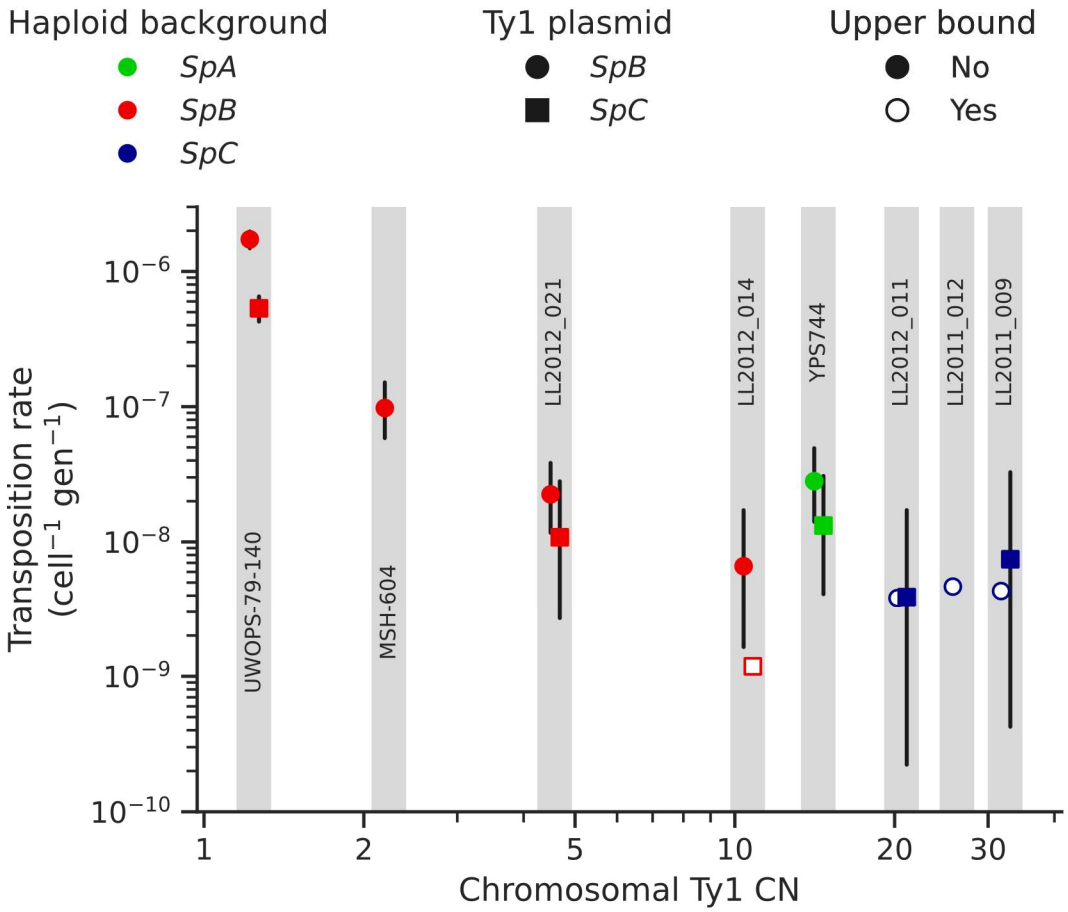


Figure 5. Ty1 transposition rate as a function of chromosomal Ty1 CN in *S. paradoxus*. Transposition rates were measured in eight haploid *S. paradoxus* backgrounds, highlighted by background gray bars at the corresponding CN. Black bars indicate the 95% confidence intervals. Empty symbols represent upper bound transposition rate estimates obtained by artificially inflating the mutant count by 1 when 0 was observed. Marker shapes indicate the sequence variant of the tester Ty1 element, either *SpB* (circles) or *SpC* (squares). The *S. paradoxus* backgrounds and estimated Ty1 CNs are, from left to right: UWOPS-79-140 (*SpB*, 1.2); MSH-604 (*SpB*, 2.2); LL2012_021 (*SpB*, 4.6); LL2012_014 (*SpB*, 10.6); YPS744 (*SpA*, 14.4); LL2012_011 (*SpC*, 20.7); LL2011_012 (*SpC*, 26.3); LL2011_009 (*SpC*, 32.4).

In each background, we tested two sequence variants of the tester Ty1 element: one derived from a *SpB* genome (MSH-604), and one derived from a *SpC* genome (LL2011_012). We found that the *SpC* variant generally yielded lower transposition rates (Figure 6A). In the UWOPS-79-140 *SpB* background, the transposition rate of the *SpB* variant (1.73×10^{-6} cell^{-1} gen^{-1}) was significantly

305 higher than the *SpC* variant, with a difference of over threefold (5.34×10^{-7} $cell^{-1}$ gen^{-1}). We
306 crossed three of the haploid backgrounds, one from each population (*SpB*: UWOPS-79-140, *SpC*:
307 LL2011_012 and *SpA*: YPS744), to generate all the possible homozygous and heterozygous
308 diploid backgrounds (Supplemental Table S6), and measured their transposition rates (Figure 6B,
309 Supplemental Table S8). This experiment confirmed that transposition rates are lower with the
310 *SpC* Ty1 variant, with the difference being significant in three different backgrounds (*SpA*×*SpB,*
311 *SpB*×*SpB* and *SpB*×*SpC*). Additionally, these measurements showed that transposition rates in
312 heterozygous hybrids are intermediate between their respective homozygous parental
313 backgrounds (Supplemental Figure S12), providing further evidence for the robustness of
314 transposition regulation in hybrids (Hénault et al., 2020; Smukowski Heil et al., 2021).

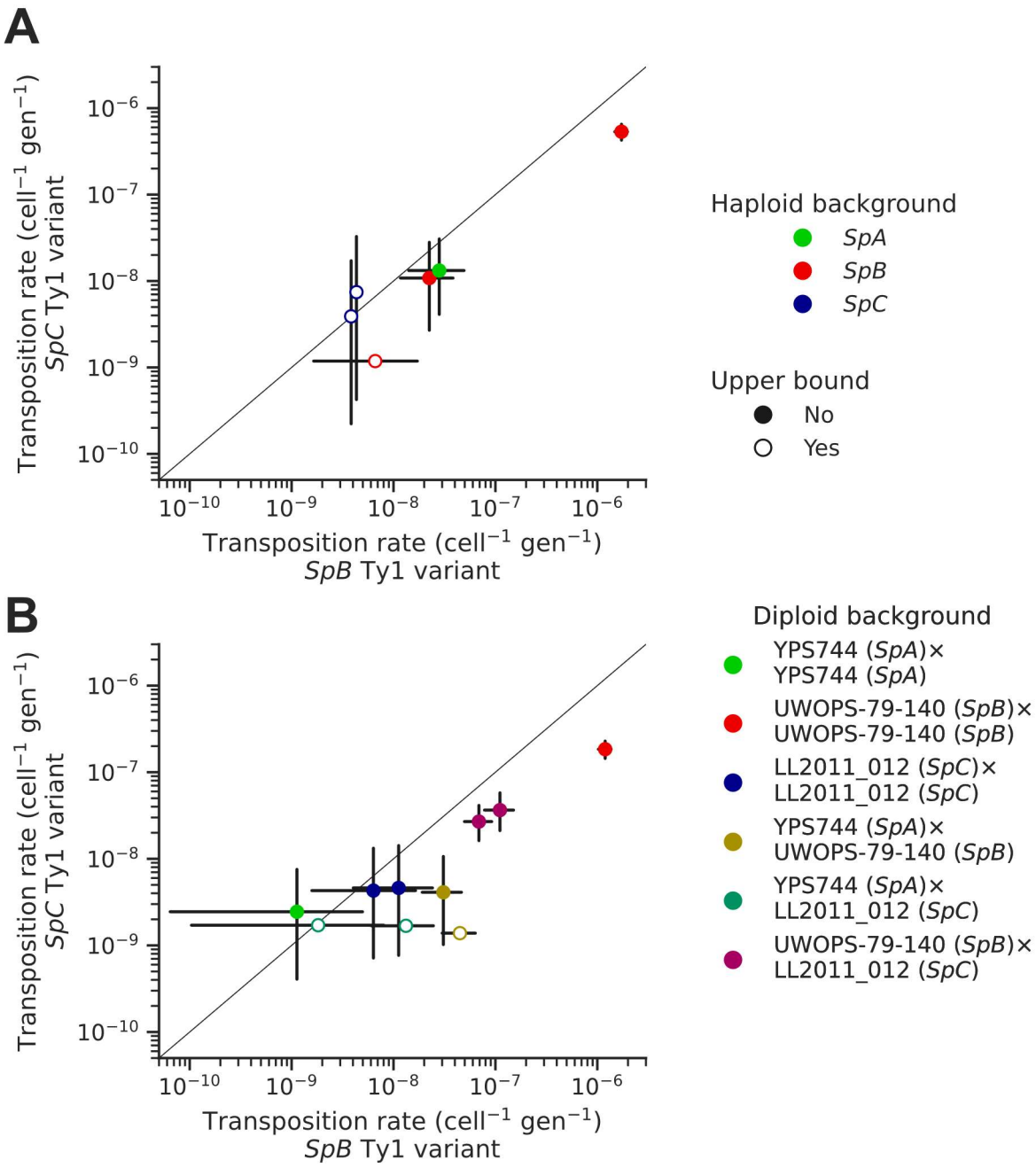


Figure 6. Ty1 transposition rate as a function of the sequence variant of the tester Ty1 element. Transposition rates of the *SpB* (X-axis) and *SpC* (Y-axis) Ty1 variants are shown for haploid backgrounds **(A)** and diploid backgrounds **(B)**. Black bars indicate the 95% confidence intervals. Empty symbols represent upper bound transposition rate estimates obtained by artificially inflating the mutant count by 1 when 0 was observed. The diagonal line indicates a 1:1 relation between transposition rates for the *SpB* and *SpC* Ty1 variants.

322 Like many other fungal species, mtDNA inheritance in *Saccharomyces* yeasts is biparental
323 (Wilson and Xu, 2012), yielding a transient state of heteroplasmy that is resolved with the fixation
324 of a single mtDNA haplotype by vegetative segregation (Birky, 2001). A recent study measured
325 the rate of Ty1 transposition in *S. cerevisiae*×*S. uvarum* hybrids while controlling mtDNA
326 inheritance and showed an effect of the parental mtDNA haplotype (Smukowski Heil et al., 2021).
327 The measured effect was quite strong, the transposition rate linked with the *S. uvarum* mtDNA
328 haplotype being approximately one order of magnitude lower than the *S. cerevisiae* mtDNA
329 haplotype. We investigated whether mtDNA inheritance also affected transposition rates at the
330 intraspecific level in hybrids between *S. paradoxus* natural populations. For each of the diploid
331 crosses described above, we tested reciprocal crosses with either parental haploid background
332 being depleted of its mtDNA (ρ^0). As expected, the reciprocal ρ^0 parents in the homozygous *SpC*
333 background yielded no significant difference in transposition rate (Figure 7). However, many
334 significant differences were observed in heterozygous backgrounds. In both *SpA*×*SpC* and
335 *SpB*×*SpC* hybrids with the *SpB* Ty1 tester variant, the transposition rate was significantly lower
336 when the mtDNA was inherited from the *SpC* (*MAT*α) parent (Figure 7). Although not statistically
337 significant, the same trend was observed for the *SpC* Ty1 tester variant in the *SpB*×*SpC* hybrid.
338 These results indicate a moderate but consistent effect of the *SpC* mtDNA in lowering Ty1
339 transposition rate.

340

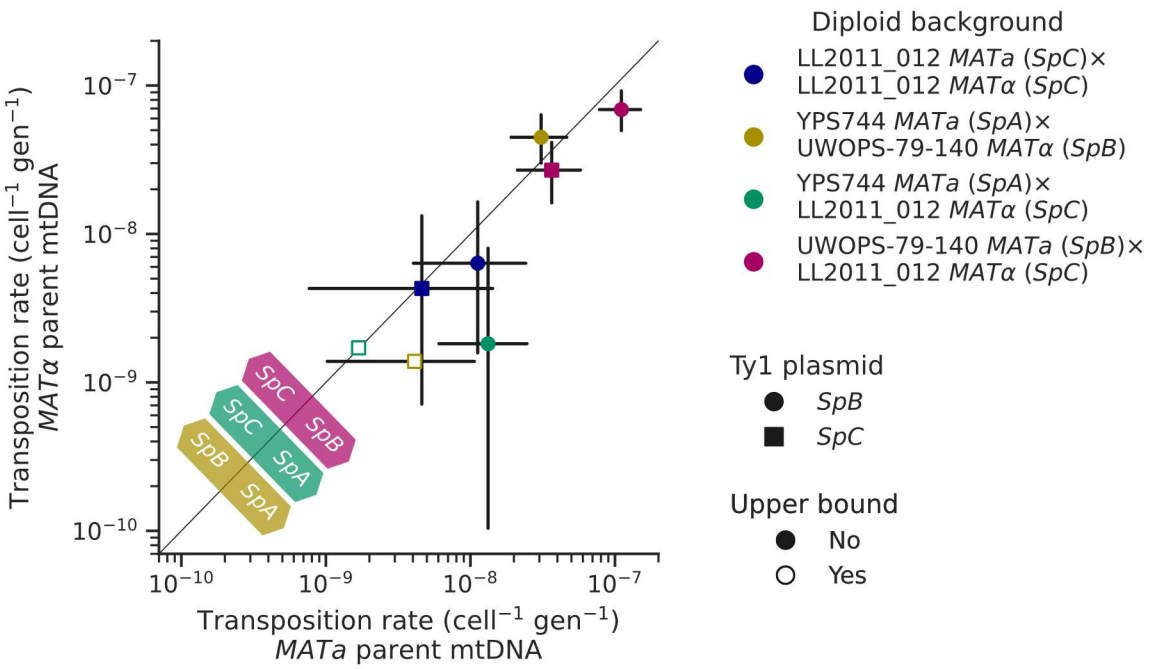


Figure 7. Ty1 transposition rate as a function of the inherited mtDNA haplotype. Transposition rates in diploid backgrounds as a function of whether the inherited mtDNA is that of the *MAT*a parent (X-axis) or the *MAT*α (Y-axis) are shown. Black bars indicate the 95% confidence intervals. Empty symbols represent upper bound transposition rate estimates obtained by artificially inflating the mutant count by 1 when 0 was observed. Marker shapes indicate the sequence variant of the tester Ty1 element, either *SpB* (circles) or *SpC* (squares). The diagonal line indicates a 1:1 relation between transposition rates for reciprocal mtDNA inheritance. Colored arrows at the bottom left indicate the mtDNA haplotype that would correspond to a higher transposition rate on that side of the diagonal for each hybrid background.

352 ## Discussion

353 The repetitive nature of TEs makes them notoriously difficult to resolve in genomes (Goerner-
354 Potvin and Bourque, 2018). As such, studies investigating the TE reactivation hypothesis in
355 hybrids have used a variety of methods for quantifying the activity of TEs at different levels, for
356 instance in terms of genomic abundance (O'Neill et al., 1998; Ungerer et al., 2006) or expression
357 level (Dion-Côté et al., 2014; Renaut et al., 2014). Not all methods enable the decomposition of
358 TE load at the level of sequence-resolved insertion loci (Tusso et al., 2021), which is an especially
359 challenging task in heterozygous genomes. A full mechanistic understanding of how TE load
360 evolves in hybrids requires the ability to characterize how TE copies are gained and lost, through
361 the action of de novo transposition and other types of SVs.

362 Here, we use long-read sequencing to yield accurate representations of the genomes of yeast
363 hybrids experimentally evolved under a near-neutral MA regime, and investigate the mechanisms
364 responsible for changes in the genomic landscape of Ty elements. We find that SV types like
365 polyploidy, LOH and aneuploidy have the largest impact on Ty CN variation. While de novo Ty1
366 insertions compatible with retrotransposition were found in most of the MA crosses, their
367 contribution to CN changes is very modest compared to the transposition-unrelated SVs. The
368 scarcity of de novo insertions in the MA lines provides insufficient resolution to dissect subtle
369 transposition rate variation between hybrid genotypes. We thus complement the genomic analysis
370 of MA lines with in vivo Ty1 transposition assays to measure transposition rate in multiple genetic
371 backgrounds. We find that initial Ty1 load, Ty1 sequence variation and mtDNA haplotype
372 inheritance are factors that modulate transposition rates within *S. paradoxus*. In addition, our
373 results confirm that unlike other yeasts (Tusso et al., 2021), Ty mobilization in *Saccharomyces* is
374 not triggered by hybridization (Hénault et al., 2020; Smukowski Heil et al., 2021) and this, along
375 a wide gradient of parental evolutionary divergence.

376 Our in vivo transposition assays revealed sharp contrasts in the dynamics of Ty1 mobilization.
377 Multiple aspects of the *SpC* genotypes are associated with lower mobilization rates. Ty1
378 transposition occurs at a very low rate in *SpC* backgrounds compared to other *S. paradoxus*
379 populations, and using a *SpC* Ty1 sequence variant as a tester element showed a much lower
380 transposition rate compared to a *SpB* Ty1 variant in matched backgrounds. Moreover, reciprocal
381 crosses with controlled mtDNA inheritance suggested an effect of *SpC* mtDNA in reducing Ty1
382 transposition rates. Despite the clear evidence for lower mobilization in vivo, half of the de novo
383 transposition events in the MA lines genomes happened in intra-lineage *SpC* crosses. The *SpC*

384 population harbors a low level of genetic diversity compared to *SpB* (Eberlein et al., 2019; Leducq
385 et al., 2016), and have a fairly constant load of Ty1 (Hénault et al., 2020). Our new whole-genome
386 assemblies revealed that this apparent CN stability hides major variation in terms of Ty1 insertion
387 position, which explains how LOH events can drastically alter the Ty1 landscape in CC crosses.

388 The contrasted mobilization dynamics suggest that intraspecific variation affects not only the
389 potency of Ty1 transposition, but also the strength of its regulation. *SpB* genomes generally have
390 low Ty1 CNs, potent Ty1 sequence variants, and a steep negative relationship of Ty1
391 transposition rate as a function of standing genomic Ty1 CN. This latter pattern is entirely
392 consistent with the CNC mechanism identified for Ty1 in *S. cerevisiae,* although the functional
393 demonstration of CNC in our system remains to be done. On the other hand, the Ty1 load is
394 higher in *SpC* genomes, but *SpC* Ty1 variants and mtDNAs are linked to markedly lower
395 transposition rates. Still, our data suggest that transposition rates in *SpC* are much higher than
396 what would be predicted by extrapolating the steep CNC-like negative relationship of *SpB*
397 genomes (Figure 5). The less potent *SpC* Ty1 variant seems to have established a state that can
398 maintain substantial polymorphism in the population, perhaps enabled by lower or non-functional
399 CNC regulation. The first alternative start codon from which the p22 protein can be expressed in
400 *S. cerevisiae* is conserved in *SpB* and *SpC* Ty1 variants (M249, orthologous to M253 in the
401 *S. cerevisiae* Gag protein sequence alignment from Czaja et al., 2020), suggesting that both could
402 express p22. Despite recent structural insights into the mechanism of p22-mediated CNC (Cottee
403 et al., 2021), the molecular basis of the difference between CNC^+ and CNC^- p22 variants in
404 *S. cerevisiae* remains unclear (Czaja et al., 2020). Variation outside of p22 is also likely to impact
405 CNC. For instance, the expression of CNC components in *S. cerevisiae* is affected by the
406 Mediator transcriptional regulator complex (Salinero et al., 2018), providing the host genome with
407 a mechanism to modulate the action of CNC (Curcio, 2019). We can hypothesize that intraspecific
408 differences in Ty1 mobilization rate may stem from complex interactions between Ty1 potency,
409 Ty1 self-regulation and host regulation.

410 In conclusion, our work provides a detailed account of the evolution of TE landscapes in the
411 genomes of hybrid yeast MA lines, through an accurate resolution of their subgenomes by phased
412 assembly. We highlight that retrotransposition is only a minor contributor to changes in TE load
413 compared to transposition-unrelated SVs. While transposition rate harbors substantial variation
414 that is underpinned by multiple sources of genetic diversity at the intraspecific level, its potential
415 for shaping the TE landscape in *Saccharomyces* hybrids is likely to be overpowered by other
416 forces of structural genome evolution.

417 # Material and methods

418 *Experimental evolution by MA and long-read sequencing*

419 The MA evolution experiments used in this study were described previously (Charron et al., 2019;
420 Hénault et al., 2020). Briefly, crosses were made between 13 haploid derivatives of wild yeast
421 isolates (11 *S. paradoxus* isolates and two *S. cerevisiae* isolates). A total of 11 crosses were
422 designed to span a range of evolutionary divergence between the parental strains, from very low
423 divergence at the intraspecific level to very high divergence at the interspecific level. Each cross
424 was replicated between 48 and 96 times with independent matings. The resulting collection was
425 evolved under an MA regime for approximately 770 mitotic generations by streaking for single
426 colonies on YPD agar medium (Supplemental Table S9) at 3-day intervals. Glycerol stock
427 archives were made of the entire collection at regular intervals, and a subset of 10-12 MA lines
428 per cross were randomly selected for long-read sequencing. Long-read library preparation and
429 sequencing were described in (Hénault et al., 2022). Briefly, DNA was extracted using a standard
430 phenol-chloroform protocol. Native (PCR-free) genomic DNA libraries were prepared using the
431 SQK-LSK109 kit (Oxford Nanopore Technologies [ONT]) and multiplexed with the EXP-NBD104
432 barcoding kit (ONT). Sequencing was performed on a MinION sequencer (MIN-101B, ONT) with
433 FLOMIN106 (revC) flowcells. The sequencing was run using MinKNOW v3.3.2 (ONT) and the
434 basecalling was run using guppy v3.0.3 (ONT). Demultiplexing was done using guppy_basecaller
435 v3.1.5.

436 *Haploid genome assembly and annotation of the parental strains*

437 De novo haploid genome assemblies for the 13 parental strains of the MA crosses were produced
438 from basecalled long reads using wtdbg2 v2.5 (Ruan and Li, 2020) with parameters –p 0 –k 15 –
439 AS 2 –s 0.05 –L 8000 –g 12m –X 120. Basecalled long reads were aligned to the draft assemblies
440 using Minimap2 v2.17 (Li, 2018) with preset map–ont. Draft assemblies were polished with the
441 long-read alignments and the linked raw nanopore signal data using nanopolish v0.13.3 (Loman
442 et al., 2015) with parameter —min–candidate–frequency 0.1. Short-read libraries from the
443 corresponding 13 parental strains were retrieved from NCBI (accession number PRJNA515073).
444 Reads were trimmed using Trimmomatic v0.33 (Bolger et al., 2014) with parameters
445 ILLUMINACLIP:{custom adapters file}:6:20:10 MINLEN:40. The short-read libraries were mapped
446 on the signal-level polished draft assemblies using BWA-MEM v0.7.17 (Li, 2013) and SAMtools

447 v1.15 (Li et al., 2009). The signal-level polished draft assemblies were polished with the short-
448 read alignments using Pilon v1.22 (Walker et al., 2014).

449 Polished de novo genome assemblies were scaffolded at the chromosome level using the
450 following procedure. Reference genomes for *S. paradoxus* strains LL2012_001 (*SpA*), MSH-604
451 (*SpB*) and LL2011_012 (*SpC*) (Eberlein et al., 2019) were scaffolded using RaGOO v1.1 (Alonge
452 et al., 2019), using the genome of *S. cerevisiae* strain S288c (Yue et al., 2017) as a reference.
453 For each of the MA parental strains, polished contigs were aligned against the scaffolded
454 reference of the corresponding species or population using the nucmer (with preset —mum) and
455 show-coords (with parameters –r –d –T) tools from the MUMmer v3.23 suite (Kurtz et al., 2004).
456 Contigs were sorted and merged into chromosomes using a custom Python v3.9.10 (Van Rossum
457 and Drake, 2009) script. We manually corrected for misassemblies that would yield dicentric or
458 acentric chromosomes, or that were incompatible with karyotypes determined by CHEF-PFGE
459 (Charron et al., 2014b). Chromosome-level assemblies were masked for rDNA repeats as follows.
460 Repeats from the S288c reference genome (release R64-2-1, downloaded from
461 www.yeastgenome.org) were extracted from chromosome XII (NC_001144, 451417-468931) and
462 searched against the chromosome-level assemblies using BLASTN v2.11.0 (Camacho et al.,
463 2009). Sequences corresponding to hits with alignment lengths over 200 bp were hard-masked
464 using a custom Python v3.9.10 script.

465 Final assemblies were annotated for families of Ty LTR retrotransposons using RepeatMasker
466 v4.1.2-p1 (Smit et al., 2015) with parameters –s –nolow –no_is –gccalc and a custom database
467 of reference Ty sequences from *S. paradoxus* (*lib_Sp.fasta*) or *S. paradoxus* and *S. cerevisiae*
468 combined (*lib.fasta*). RepeatMasker outputs were defragmented using REannotate version
469 26.11.2007 (Pereira, 2008) with parameters -s 15000 -d 10000. Final assemblies for the two
470 parents of each MA cross were aligned to each other using the nucmer (with preset —mum) and
471 show-coords (with parameters –r –d –T) tools from the MUMmer v3.23 suite. The orthology
472 between the full-length Ty annotations for each cross was scored from visual inspection of the
473 alignments.

474 *Sorting of reads per MA line subgenome*

475 For each of the 11 crosses of the MA experiment, genomic variants that distinguish the two
476 parental genomes were obtained from a variant call dataset previously generated (Hénault et al.,
477 2020) from short reads mapped on the reference genome of *S. paradoxus SpB* strain MSH-604

478 (Eberlein et al., 2019). Informative variants were identified with bcftools v1.16 (Li, 2011) using a
479 custom bash script. Long-read libraries were mapped against the MSH-604 genome (Eberlein et
480 al., 2019) and a custom Python v3.9.10 script was used to classify individual reads as one parental
481 genotype or the other. Each read had to have at least two informative variants. For a read to be
482 classified as originating from the subgenome of one parent, it had to be supported by a count of
483 informative variants at least twice as high as the count of informative variants supporting the
484 opposing parental subgenome. Lists of read IDs for reads successfully classified were used to
485 split MA lines libraries using Seqtk v1.3-r106 (https ://github.com/lh3/seqtk).

486 *Subgenome-level assembly and annotation of MA line subgenomes*

487 Basecalled long reads split by subgenome for each MA line were assembled separately using
488 wtdbg2 v2.5 with parameters –p 0 –k 15 –AS 2 –s 0.05 –g 12m. Draft assemblies were polished
489 with basecalled reads using medaka v1.4.4 (https://github.com/nanoporetech/medaka) with
490 model r941_min_high_g303. Polished assemblies were annotated for families of Ty LTR
491 retrotransposons using the same methodology as for haploid parental genomes (see above). To
492 enable the comparison of Ty annotations between the different assemblies, the following
493 procedure was designed to establish a common coordinate system and to define orthology
494 relationships. Subgenome-level assemblies were aligned against the corresponding haploid
495 parental genome using Minimap2 v2.20-r1061 with preset –x asm10. Ty annotation coordinates
496 were exported into BED format using a custom Python v3.9.10 script and lifted over to haploid
497 parental genome coordinates using paftools liftover v2.20-r1061 (Li, 2018). The resulting
498 annotation coordinates were merged in clusters (roughly corresponding to Ty orthogroups) in a
499 custom Python v3.9.10 script using the DBSCAN algorithm from the scikit-learn v1.1.3 Python
500 package (Pedregosa et al., 2011) with parameters eps=500 min_samples=1. Annotations within
501 one cluster were required to share the same Ty family and strand orientation. Clusters were
502 classified as complex or non-complex, depending on whether they contained more than one
503 annotation in any genome.

504 *Classification of orthologous Ty clusters*

505 For each cluster of orthologous Ty annotations in each MA line subgenome assembly, we
506 assigned an SV class, if applicable. Unless otherwise stated, all the steps of the classification
507 procedure were performed using custom Python v3.9.10 scripts.

508 Basecalled long reads split by subgenome were mapped against the corresponding parental
509 assembly using Minimap2 v2.20-r1061 with preset map–ont, and alignments were filtered using
510 SAMtools v1.16.1 to remove secondary alignments. Depth of coverage values from these
511 alignments were extracted using SAMtools v1.16.1. Tracts of discrete relative depth of coverage
512 along each genome were defined as follows. Assemblies were subdivided into 30 kb-wide non-
513 overlapping windows, and the median depth of coverage value of each window was divided by
514 the genome-wide median. Adjacent windows having the same value were merged to generate
515 tracts that reflect major CN changes along the genome (i.e. polyploidy, aneuploidies and LOH
516 events).

517 In addition to the discrete estimates of relative depth of coverage along each subgenome, the
518 existence of a subgenome-level assembly was inferred for each parental Ty cluster using a
519 reverse liftover procedure as follows. The parental genomes were aligned against the
520 corresponding subgenome-level assemblies using Minimap2 v2.20-r1061 with preset –x asm10.
521 For each Ty cluster, the two 500 bp-wide windows immediately flanking the annotation were
522 exported into BED format and lifted over to subgenome-level assembly coordinates using paftools
523 liftover v2.20-r1061. For the downstream steps, the clusters having inconsistent depth of
524 coverage-based CN estimates and assembly alignment-based presence (i.e. having a CN of zero
525 and assembly presence, or a CN greater than zero and assembly absence) were excluded.

526 These data were organized as boolean variables as follows. At the level of the individual
527 orthologous cluster, the variables included were: whether the cluster is complex (comprising more
528 than a single annotation per genome); and whether the cluster contains an annotation in the
529 corresponding parental genome. At the level of the individual orthologous cluster in individual MA
530 line subgenomes, the variables included were: whether the cluster contains an annotation;
531 whether there is the presence of an assembly based on assembly alignment; whether the content
532 of the cluster is the same as the parent; and whether the CN of the cluster is the same as the
533 parent. The CN of the cluster and the types of annotations in the cluster are non-boolean variables
534 that were also included at this level.

535 Using these variables, a decision tree (Supplemental Figure S5) was used to attribute one of the
536 following SV classes for each orthologous cluster in each MA line subgenome. The majority of
537 clusters were identical to the corresponding cluster in the parental genome, and were called "No
538 change". The clusters which had identical contents to the parental genome, but in a different CN,
539 were called "Polyploidy/aneuploidy/LOH". The clusters which had a full-length annotation in the

540 parent and a truncated annotation in the MA line subgenome were called "Truncation". The
541 clusters which had a full-length annotation in the parent and a solo LTR in the MA line subgenome
542 were called "Excision". The clusters which had a unique additional full-length annotation in the
543 MA line subgenome compared to the parent were called "De novo". The clusters which had an
544 annotation in the parental genome, but no annotation in the MA line subgenome despite having
545 support for MA line subgenome assembly presence, were called "Deletion".

546 We manually classified loci from the class "Polyploidy/aneuploidy/LOH" using the mappings of
547 long read libraries against parental assemblies. Genome-wide relative depth of coverage profiles
548 were generated by binning values in non-overlapping windows of 10 kb. Profiles were individually
549 examined to determine if CN changes were due to polyploidy, aneuploidy or large tracts of LOH.

550 *Identification of de novo transposition events*

551 From the classification produced by the decision tree, the loci identified as "De novo" were curated
552 to identify confident loci corresponding to retrotransposition events. We extracted the sequences
553 immediately flanking the insertion loci and looked for short repeated motifs in direct orientation,
554 corresponding to TSDs. TSDs of five or six nucleotides were found, and loci with no identifiable
555 TSD were discarded. We produced chromosome alignments comprising the MA line
556 subgenome(s) with de novo loci and all parental strains using Mauve v2015-02-26 (Darling et al.,
557 2010) with the progressiveMauve algorithm and default parameters. Alignments were inspected
558 to confirm that the junctions of de novo loci were unique to the corresponding MA line subgenome.
559 Finally, we compared the TSD motifs of candidate de novo retrotransposition to TSD motifs of all
560 parental full-length copies to filter out non-unique ones. The remaining loci (all from the Ty1 family)
561 were classified as de novo retrotransposition events.

562 A phylogenetic analysis of the de novo Ty1 insertions was performed by extracting the sequences
563 of de novo loci and of all full-length Ty1 and Ty2 parental elements. We added the sequences of
564 representative Ty1 and Ty2 elements identified by (Bleykasten-Grosshans et al., 2021) as
565 references. We included the 50 bp flanking sequences in 5' and 3' of all extracted sequences and
566 ran a multiple sequence alignment using MUSCLE v3.8.31 (Edgar, 2004). The alignment was
567 trimmed manually using MEGA v11.0.11 (Tamura et al., 2021). A maximum likelihood
568 phylogenetic tree was computed on the trimmed alignment using RAxML-NG v1.1.0 (Kozlov et
569 al., 2019) with the GTR+G model initiated with 10 parsimony trees and 200 bootstraps. The
570 phylogenetic tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

571 Additionally, we built a phylogenetic network from the same trimmed alignment using SplitsTree
572 v4.18.3 (Huson and Bryant, 2006) with default parameters.

573 *Plasmids used and constructed in this study*

574 Plasmids used and constructed in this study are listed in Supplemental Table S2. Oligonucleotide
575 sequences are detailed in Supplemental Table S3. PCR reactions and cycles are detailed in
576 Supplemental Table S4. Media recipes are detailed in Supplemental Table S9. All the PCRs used
577 for cloning were performed using the KAPA HiFi HotStart DNA polymerase (Roche). All the PCRs
578 used for cloning confirmations were performed using the Taq DNA Polymerase (BioShop).
579 Plasmid minipreps were performed on overnight cultures in 5 mL 2YT+Amp or 2YT+Kan medium,
580 depending on the plasmid, using the Presto Mini Plasmid Kit (GeneAid).

581 Gibson cloning reactions were performed as follows. A Gibson assembly master mix was
582 prepared with the following components (for 1.2 mL): 360 µL of 5X ISO buffer (500 mM Tris-Cl,
583 50 mM MgCl2, 4 mM dNTPs mix, 50 mM DTT, 250 mg mL^{-1} PEG 8000, 5 mM NAD^+), 0.64 µL
584 10 U μL^{-1} T5 exonuclease (New England Biolabs), 20 of 2 U μL^{-1} Phusion High-Fidelity DNA
585 Polymerase (New England Biolabs), 160 µL of 40 U μL^{-1} Taq DNA ligase (New England Biolabs)
586 and 700 µL of PCR-grade H_2O. Gibson reactions were assembled by adding all the components
587 to 7.5 µL of Gibson master mix, and completing the volume to 10 µL with PCR-grade H_2O. Gibson
588 reactions were incubated at 50°C for 60 min.

589 Competent cells of *Escherichia coli* were prepared as detailed in Green and Rogers (Green and
590 Rogers, 2013). *E. coli* transformations were performed by combining 100 µL of competent cells
591 with 5 µL of mutagenesis or Gibson assembly reaction and incubating the mixture on ice for
592 15 min. Heat shock was done at 42°C for 1 min, followed by incubation on ice for 5 min. 900 µL
593 of 2YT medium was added and the mixture was incubated for 45 min at 37°C with agitation. Cells
594 were plated on 2YT+Amp agar or 2YT+Kan agar medium, depending on the plasmid, and
595 incubated at 37°C for 24 hrs.

596 *Construction of Ty1 retrotransposition marker plasmids*

597 Retrotransposition rate measurement assays were developed based on the plasmid-borne assay
598 designed by Curcio and Garfinkel (Curcio and Garfinkel, 1991) with several modifications. First,
599 we replaced the *S. cerevisiae* Ty1 sequence with variants sampled from wild *S. paradoxus*
600 genomes. To measure transposition rates at a transcription rate closer to the natural genomic

601 context and prevent potential biases arising from galactose induction in multiple wild genetic
602 backgrounds, we opted for the native promoter of Ty1 variants instead of the *GAL1* promoter.
603 Finally, we opted for a low CN centromeric plasmid backbone instead of a high CN plasmid
604 backbone in order to increase the mitotic stability and reduce the CN variance of the system.

605 Plasmid pRS31N was constructed as follows. Using a pFA6-natNT2 (Janke et al., 2004) miniprep
606 as a template, we amplified the natNT2 resistance cassette against nourseothricin (Nat). Using a
607 pRS316 (Sikorski and Hieter, 1989) miniprep as a template, we amplified the backbone excluding
608 the *URA3* selection marker. A Gibson assembly mix was prepared by adding 1 µL of natNT2
609 cassette and 1 µL of backbone to the Gibson master mix. After incubation for 60 min at 50°C, the
610 Gibson reaction was digested with DpnI (New England Biolabs). Competent cells of *E. coli*
611 MC1061 (Casadaban and Cohen, 1980) were transformed with 5 µL of the Gibson reaction.
612 Cloning was confirmed by PCR by testing the 5' junction. A miniprep was migrated by agarose
613 gel electrophoresis and transformed into yeast, selecting on both YPD+Nat agar medium and SC-
614 ura agar medium.

615 Plasmids pRS31N-Ty1-*SpB*-*HIS3*AI and pRS31N-Ty1-*SpC*-*HIS3*AI were constructed as follows.
616 The plasmid pGTy1-H3*HIS3*AI was a gift from David Garfinkel (Addgene plasmid # 62228 ;
617 http://n2t.net/addgene:62228 ; RRID:Addgene_62228) (Curcio and Garfinkel, 1991). Using a
618 pGTy1-H3m*HIS3*AI miniprep as a template, we amplified the retrotransposition indicator gene
619 (RIG) *HIS3*AI. We treated the *HIS3*AI RIG amplicon with DpnI followed by purification using solid-
620 phase reversible immobilization (SPRI) beads. We digested the plasmid pRS31N with SacII (New
621 England Biolabs), treated the digestion with calf intestinal alkaline phosphatase (CIP, New
622 England Biolabs) and purified the digested plasmid using SPRI beads. We extracted genomic
623 DNA from two *S. paradoxus* wild diploid strains (MSH-604, *SpB*; and LL2011_012, *SpC*) by
624 following a standard phenol-chloroform extraction protocol on overnight cultures in 5 mL YPD
625 medium incubated at 30°C with agitation. A genomic region of ~10 kb containing a single full-
626 length Ty1 element was amplified from the MSH-604 (chrXV) and LL2011_012 (chrVII) genomic
627 DNA extractions using unique primers. The full-length element chosen from the LL2011_012
628 genome was identical to five other full-length Ty1 elements in the same genome, and one of the
629 two full-length Ty1 elements in MSH-604 was chosen at random. From each genomic region
630 amplicon, the full-length Ty1 element was amplified in two segments: a first segment spanning
631 the left LTR and complete internal ORF (Ty1 5' segment), and a second segment spanning the
632 right LTR (Ty1 3' segment). The Ty1 segment amplicons were purified with SPRI beads. A first
633 Gibson assembly reaction was performed by combining 40 fmol each of the digested pRS31N

634 backbone, the Ty1 5' segment amplicon and the *HIS3*AI RIG amplicon. Competent cells of
635 *Escherichia coli* MC1061 were transformed with 5 µL of the Gibson assembly reaction. Cloning
636 was confirmed by PCR by testing the 5' and 3' junctions. We prepared a miniprep of the first
637 Gibson assembly and validated its restriction profile with EcoRV (New England Biolabs). The
638 miniprep of the first Gibson assembly was digested with NotI (New England Biolabs), and the
639 digestion was treated with CIP and purified using SPRI beads. A second Gibson assembly
640 reaction was performed by combining 40 fmol each of the first Gibson assembly digestion and
641 the Ty1 3' segment amplicon. Cloning was confirmed by PCR by testing the 5' and 3' junctions.
642 We prepared a miniprep of the second Gibson assembly and validated its restriction profile with
643 EcoRV. We confirmed the final clones with long-read whole-plasmid sequencing at
644 Plasmidsaurus (http://www.plasmidsaurus.com).

645 *Yeast strains used and constructed in this study*

646 *S. paradoxus* strains used and constructed in this study are listed in Supplemental Table S5.
647 Oligonucleotide sequences are detailed in Supplemental Table S3. PCR reactions and cycles are
648 detailed in Supplemental Table S4. Media recipes are detailed in Supplemental Table S9. All the
649 PCRs for plasmid mutagenesis and cassette amplification were performed using the KAPA HiFi
650 HotStart DNA polymerase. All the PCRs used for confirmations or diagnostics were performed
651 using the Taq DNA Polymerase.

652 *S. paradoxus* competent cells were prepared from 25 mL cultures in YPD medium incubated at
653 30°C with agitation until they reached an optical density at 600 nm (OD_{600} mL^{-1}) of 0.5-0.7.
654 Cultures were centrifuged at 500×*g* for 5 min. Cells were washed once in 5 mL in H_2O and once
655 in 5 mL SORB solution (1 M sorbitol, 100 mM lithium acetate, 10 mM Tris-Cl, 1 mM EDTA). Cells
656 were resuspended in 180 µL of SORB solution and 20 µL of salmon sperm DNA that was
657 previously boiled for 5 min and cooled on ice. Competent cells were stored at -70°C until use.

658 *S. paradoxus* transformations were performed by combining 20 µL of competent cells, 250 ng of
659 plasmid or 8 µL cassette amplicon, and 100 µL of Plate Mixture solution (40% PEG 3350, 100 mM
660 lithium acetate, 10 mM Tris-Cl, 1 mM EDTA). The mixture was incubated at room temperature for
661 30 min. 7.5 µL of DMSO were added and a heat shock was performed at 37°C for 30 min. Cells
662 were centrifuged at 2000 rpm for 3 min and resuspended in 100 µL of YPD medium for 2 hrs in
663 the case of plasmid transformations, or 5 hrs in all other cases.

664 *Generation of histidine auxotroph strains*

665 Short placeholder sequences (hereafter referred to as stuffers) (Dionne et al., 2021) to be used
666 as donor DNA in CRISPR-Cas9 mediated deletion of *HIS3* (YOR202W) were amplified using the
667 following reactions: 2 µL of Taq 10X Buffer (BioShop), 1.2 µL of 25 mM $MgCl_2$, 0.4 µL of 10 mM
668 dNTPs mix, 0.4 µL of 10 µM 5' primer, 0.4 µL of 10 µM 3' primer, 2 µL of 0.01 µM template primer,
669 0.12 µL of Taq DNA Polymerase and 13.5 µL of PCR-grade H_2O. The following PCR cycle was
670 run: 3 min at 94°C; 35 cycles of 30 sec at 94°C, 30 sec at 68°C, and 15s at 72°C; 25 sec at 72°C;
671 hold at 10°C.

672 A plasmid encoding CRISPR-Cas9 and a guide RNA (gRNA) sequence targeting the *HIS3* coding
673 sequence in *S. paradoxus* was designed as follows. The plasmid pCAS was a gift from Jamie
674 Cate (Addgene plasmid # 60847 ; http://n2t.net/addgene:60847 ; RRID:Addgene_60847) (Ryan
675 et al., 2014). pCAS was used as a template to mutagenize the gRNA sequence to target the
676 coding sequence of the *PEX11* gene (YOL147C). Mutagenesis PCRs were treated with DpnI and
677 transformed in *E. coli* MC1061. The mutagenesis was confirmed by Sanger sequencing of the
678 gRNA locus from a miniprep. pCAS-*PEX11* was used as a template to mutagenize the gRNA
679 sequence to target the coding sequence of *HIS3*. Mutagenesis PCRs were treated with DpnI and
680 transformed in *E. coli* MC1061. The gRNA locus was amplified for confirmation, and the
681 mutagenesis was confirmed by digesting the amplicon with HinfI (New England Biolabs) (the
682 *PEX11* gRNA sequence contains a HinfI restriction site, while the *HIS3* gRNA doesn't) and by
683 Sanger sequencing.

684 CRISPR-Cas9 mediated deletion of *HIS3* was done following the method of Ryan et al. (Ryan et
685 al., 2016) with pCAS-*HIS3* and the corresponding stuffer amplicon in eight different haploid
686 *S. paradoxus* strains derived from wild isolates (Charron et al., 2014b; Leducq et al., 2016). Gene
687 deletion was confirmed by PCR and growth on SC-his agar medium.

688 To enable crosses among three of the *his3*::stuffer strains, one per population (*SpA*: YPS744,
689 *SpB*: UWOPS-79-140, *SpC*: LL2011_012), the hphMX resistance cassette against hygromycin B
690 (HygB) at the *HO* locus was replaced by the kanMX resistance cassette against geneticin (G418),
691 and the mating types were switched.

692 The plasmid pUG6 (Güldener et al., 1996) was obtained from Euroscarf
693 (http://www.euroscarf.de). pUG6 was used as a template to amplify the kanMX cassette with

694 flanking homology to the HO locus. Competent cells were transformed with the cassette and
695 plated on YPD+G418 agar medium. Cassette switch was confirmed by PCR.

696 The plasmid pHS3 was a gift from John McCusker (Addgene plasmid # 81038 ;
697 http://n2t.net/addgene:81038 ; RRID:Addgene_81038). Competent cells from the mating type
698 switched strains were transformed with pHS3 and plated on YPD+G418+Nat agar medium.
699 Colonies were grown in 3 mL YPD medium at room temperature with agitation for two days to
700 allow for pHS3 segregation, diluting cultures ~1000X after 24 hrs. Cultures were streaked on YPD
701 plates and incubated at room temperature for five days. Single colonies were checked for the loss
702 of pHS3 by making patches on YPD+Nat agar medium and for autodiploidization by PCR (Huxley
703 et al., 1990). Strains were sporulated as described in Plante and Landry (Plante and Landry,
704 2020) and tetrads were dissected on YPD agar medium using a SporePlay microscope (Singer
705 Instruments). Dissection plates were incubated at room temperature for five days. The mating
706 types of spores were confirmed by PCR (Huxley et al., 1990).

707 To enable crosses with reciprocal mtDNA inheritance, we generated strain derivatives devoid of
708 mtDNA (ρ^0) as described in Fox et al. (Fox et al., 1991). The loss of respiration ability was
709 confirmed by spotting the strains on YPEG agar medium.

710 *Crosses with reciprocal mtDNA inheritance*

711 Histidine auxotroph strains and their *HO* cassette-switch, mating type-switch and ρ^0 derivatives
712 were crossed as detailed in Supplemental Table S6. For each parental strain, an overnight
713 preculture was grown in 1 mL of YPD medium at room temperature. 100 µL of each preculture
714 were combined in 1 mL YPD medium and incubated at room temperature without agitation for
715 4 hrs. 10 µL of mating cultures were streaked on YPD+G418+HygB agar medium and incubated
716 for three days at room temperature. Two single colonies per streaking were used to inoculate
717 1 mL of YPD medium and the strains were tested for respiratory proficiency by spotting on YPEG
718 agar medium. The strains were also tested for diploidy by cellular DNA staining and fluorescence
719 measurement by flow cytometry using a Guava easyCyte 8HT flow cytometer (EMD Millipore) as
720 described in Marsit et al. (Marsit et al., 2021).

721 *Measurement of retrotransposition rates*

722 Fluctuation assays were performed to measure retrotransposition rates. Competent cells of the
723 corresponding haploid strains and diploid crosses were transformed with minipreps of either

724 pRS31N-Ty1-*SpB-HIS3*AI or pRS31N-Ty1-*SpC-HIS3*AI, and plated on YPD+Nat agar medium.
725 Transformation plates were incubated at 30°C for three days. From the transformation plates,
726 single colonies were picked and resuspended in 10 mL of YPD medium. Cell densities were
727 estimated by measuring the OD_{600} of the resuspensions, and dilutions at 2×10^{-3} OD_{600} mL^{-1} were
728 prepared in YPD medium. 96-well plates were filled with 225 µL of YPD+Nat 1.1X medium. 25 µL
729 of cell dilutions were added to each well for a final concentration of 2×10^{-4} OD_{600} mL^{-1} (~500 cells
730 per well). Each combination of background and Ty1 variant was replicated in 48 independent
731 cultures (half a plate). Plates were covered with breathable rayon films and incubated at 20°C for
732 three days. Final cultures were centrifuged at 500×*g* for 5 min and resuspended in 200 µL of
733 H_2O. The entire resuspensions were spotted in groups of 7 or 9 on SC-his agar plates that were
734 previously dried at 37°C for three days. Plates were incubated at room temperature and His^+
735 colonies were counted after five days.

736 For each haploid background, 42 cultures were spotted on SC-his agar plates and six were
737 pooled. Serial dilutions were prepared by diluting pools twice 200X and 100 µL of final dilutions
738 were plated on YPD agar plates. Plates were incubated at room temperature for 3 days and
739 colonies were counted to estimate final population sizes. Colony counts are detailed in
740 Supplemental Table S7. For each diploid background, 36 cultures were spotted on SC-his agar
741 plates and 12 were pooled. Serial dilutions were prepared by diluting pools twice 12X and cell
742 densities of the final dilutions were measured by flow cytometry using a Guava easyCyte 8HT
743 flow cytometer to estimate final population sizes. Colony counts are detailed in Supplemental
744 Table S8.

745 The number of mutations (retrotransposition events) in each culture and associated 95%
746 confidence intervals were estimated using the newtonLD and confintLD functions from the Python
747 implementation of the R package rSalvador (Zheng, 2017) (https://github.com/eeeeeric/
748 rSalvador/blob/master/pysalvador/pysalvador.py). Retrotransposition rates were computed by
749 dividing mutation count estimates by the estimates of final population sizes.

750

Data accessibility

Nanopore long-read sequencing data (basecalled reads) and parental genome assemblies are available at NCBI under accession number PRJNA828354. Illumina short-read sequencing data of the MA lines is available at NCBI under accession number PRJNA515073. Custom scripts, custom databases of reference Ty sequences and subgenome-level assemblies of MA lines used in this study are available at GitHub (https://github.com/Landrylab/lrma).

Competing interests

We declare no competing interest.

Acknowledgements

We thank Alexandre Dubé for helpful comments on the manuscript and contributions to plasmid constructions. We thank Anna Fijarczyk and Hélène Martin for contributions to sequencing data preprocessing. We thank the members of the Landry laboratory for helpful discussions on the project. This project was supported by funding to C.R.L. from an NSERC discovery grant (RGPIN-2020-04844) and an FRQNT team grant (2019-PR-254415), and an NSERC Alexander Graham Bell doctoral scholarship to M.H. C.R.L. holds the Canada Research Chair in Cellular Systems and Synthetic Biology.

Author contributions

C.R.L. and M.H. designed the project. M.H. performed the analyses and experiments. M.H., S.M., and G.C. performed the MA evolution experiment and short-read sequencing as cited. M.H. wrote the manuscript with input from all authors.

771 ## References

772 Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz
773 MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes.
774 *Genome Biol* **20**:224.

775 Bergman CM. 2018. Horizontal transfer and proliferation of Tsu4 in Saccharomyces paradoxus.
776 *Mob DNA* **9**:18.

777 Birky CW Jr. 2001. The inheritance of genes in mitochondria and chloroplasts: laws,
778 mechanisms, and models. *Annu Rev Genet* **35**:125–148.

779 Bleykasten-Grosshans C, Fabrizio R, Friedrich A, Schacherer J. 2021. Species-Wide
780 Transposable Element Repertoires Retrace the Evolutionary History of the Saccharomyces
781 cerevisiae Host. *Mol Biol Evol* **38**:4334–4345.

782 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence
783 data. *Bioinformatics* **30**:2114–2120.

784 Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák
785 Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know
786 about transposable elements. *Genome Biol* **19**:199.

787 Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. 1984. The molecular basis of I-R
788 hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the
789 I factor. *Cell* **38**:153–163.

790 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
791 BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421.

792 Casadaban MJ, Cohen SN. 1980. Analysis of gene control signals by DNA fusion and cloning in
793 Escherichia coli. *J Mol Biol* **138**:179–207.

794 Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong
795 LC, White RL. 1983. Expression of recessive alleles by chromosomal mechanisms in
796 retinoblastoma. *Nature* **305**:779–784.

797 Charlesworth B, Langley CH. 1989. The population genetics of Drosophila transposable
798 elements. *Annu Rev Genet* **23**:251–287.

799 Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA
800 in eukaryotes. *Nature* **371**:215–220.

801 Charron G, Leducq JB, Bertin C, Dubé AK, Landry CR. 2014a. Exploring the northern limit of
802 the distribution of Saccharomyces cerevisiae and Saccharomyces paradoxus in North
803 America. *FEMS Yeast Res* **14**:281–288.

804 Charron G, Leducq JB, Landry CR. 2014b. Chromosomal variation segregates within incipient
805 species and correlates with reproductive isolation. *Mol Ecol* **23**:4362–4372.

806 Charron G, Marsit S, Hénault M, Martin H, Landry CR. 2019. Spontaneous whole-genome
807 duplication restores fertility in interspecific hybrids. *Nat Commun* **10**:4126.

808 Cottee MA, Beckwith SL, Letham SC, Kim SJ, Young GR, Stoye JP, Garfinkel DJ, Taylor IA.
809 2021. Structure of a Ty1 restriction factor reveals the molecular basis of transposition copy
810 number control. *Nat Commun* **12**:5590.

811 Curcio MJ. 2019. Border collies of the genome: domestication of an autonomous retrovirus-like
812 transposon. *Curr Genet* **65**:71–78.

813 Curcio MJ, Garfinkel DJ. 1991. Single-step selection for Ty1 element retrotransposition. *Proc*
814 *Natl Acad Sci U S A* **88**:936–940.

815 Czaja W, Bensasson D, Ahn HW, Garfinkel DJ, Bergman CM. 2020. Evolution of Ty1 copy
816 number control in yeast by horizontal transfer and recombination. *PLoS Genet*
817 **16**:e1008632.

818 Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene
819 gain, loss and rearrangement. *PLoS One* **5**:e11147.

820 Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic
821 shock involving transposable elements reactivation in hybrids of young lake whitefish
822 species. *Mol Biol Evol* **31**:1188–1199.

823 Dionne U, Bourgault É, Dubé AK, Bradley D, Chartier FJM, Dandage R, Dibyachintan S,
824 Després PC, Gish GD, Pham NTH, Létourneau M, Lambert J-P, Doucet N, Bisson N,
825 Landry CR. 2021. Protein context shapes the specificity of SH3 domain-mediated
826 interactions in vivo. *Nat Commun* **12**:1597.

827 Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002.
828 Characteristic genome rearrangements in experimental evolution of Saccharomyces
829 cerevisiae. *Proc Natl Acad Sci U S A* **99**:16144–16149.

830 Eberlein C, Hénault M, Fijarczyk A, Charron G, Bouvier M, Kohn LM, Anderson JB, Landry CR.
831 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nat*
832 *Commun* **10**:923.

833 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
834 throughput. *Nucleic Acids Res* **32**:1792–1797.

835 Fox TD, Folley LS, Mulero JJ, McMullin TW, Thorsness PE, Hedin LO, Costanzo MC. 1991.
836 Analysis and manipulation of yeast mitochondrial genes. *Methods Enzymol* **194**:149–165.

837 Garfinkel DJ, Nyswaner K, Wang J, Cho JY. 2003. Post-transcriptional cosuppression of Ty1
838 retrotransposition. *Genetics* **165**:83–99.

839 Goerner-Potvin P, Bourque G. 2018. Computational tools to unmask transposable elements.
840 *Nat Rev Genet* **19**:688–704.

841 Gorkovskiy A, Verstrepen KJ. 2021. The Role of Structural Variation in Adaptation and Evolution
842 of Yeast and Other Fungi. *Genes* **12**. doi:10.3390/genes12050699

843 Green R, Rogers EJ. 2013. Chapter Twenty Eight - Transformation of Chemically Competent E.
844 coli In: Lorsch J, editor. Methods in Enzymology. Academic Press. pp. 329–336.

845 Güldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH. 1996. A new efficient gene

846 disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**:2519–2524.

847 Heasley LR, Argueso JL. 2022. Genomic characterization of a wild diploid isolate of
848 Saccharomyces cerevisiae reveals an extensive and dynamic landscape of structural
849 variation. *Genetics* **220**. doi:10.1093/genetics/iyab193

850 Hénault M, Marsit S, Charron G, Landry CR. 2022. Hybridization drives mitochondrial DNA
851 degeneration and metabolic shift in a species with biparental mitochondrial inheritance.
852 *Genome Res* **32**:2043–2056.

853 Hénault M, Marsit S, Charron G, Landry CR. 2020. The effect of hybridization on transposable
854 element accumulation in an undomesticated fungal species. *Elife* **9**.
855 doi:10.7554/eLife.60474

856 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol*
857 *Biol Evol* **23**:254–267.

858 Huxley C, Green ED, Dunham I. 1990. Rapid assessment of S. cerevisiae mating type by PCR.
859 *Trends Genet* **6**:236.

860 Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A,
861 Doenges G, Schwob E, Schiebel E, Knop M. 2004. A versatile toolbox for PCR-based
862 tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution
863 cassettes. *Yeast* **21**:947–962.

864 Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable elements and
865 genome organization: a comprehensive survey of retrotransposons revealed by the
866 complete Saccharomyces cerevisiae genome sequence. *Genome Res* **8**:464–478.

867 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and
868 user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**:4453–
869 4455.

870 Kuehne HA, Murphy HA, Francis CAA, Sniegowski PD. 2007. Allopatric Divergence, Secondary
871 Contact, and Genetic Isolation in Wild Yeast Populations. *Curr Biol* **17**:407–411.

872 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004.
873 Versatile and open software for comparing large genomes. *Genome Biol* **5**:R12.

874 Leducq J-B, Charron G, Samani P, Dubé AK, Sylvester K, James B, Almeida P, Sampaio JP,
875 Hittinger CT, Bell G, Landry CR. 2014. Local climatic adaptation in a widespread
876 microorganism. *Proc Biol Sci* **281**:20132472.

877 Leducq JB, Nielly-Thibault L, Charron G, Eberlein C, Verta JP, Samani P, Sylvester K, Hittinger
878 CT, Bell G, Landry CR. 2016. Speciation driven by hybridization and chromosomal plasticity
879 in a wild yeast. *Nature Microbiology* **1**:1–10.

880 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–
881 3100.

882 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM
883 **00**:1–3.

884 Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and
885 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–
886 2993.

887 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
888 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map
889 format and SAMtools. *Bioinformatics* **25**:2078–2079.

890 Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using
891 only nanopore sequencing data. *Nat Methods* **12**:733–735.

892 Lower SE, Dion-Côté A-M, Clark AG, Barbash DA. 2019. Special Issue: Repetitive DNA
893 Sequences. *Genes* **10**. doi:10.3390/genes10110896

894 Luria SE, Delbrück M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance.
895 *Genetics* **28**:491–511.

896 Marsit S, Hénault M, Charron G, Fijarczyk A, Landry CR. 2021. The neutral rate of whole-
897 genome duplication varies among yeast species and their hybrids. *Nat Commun* **12**:3126.

898 McClintock B. 1984. The significance of responses of the genome to challenge. *Science*
899 **226**:792–801.

900 Nishida Y, Pachulska-Wieczorek K, Błaszczyk L, Saha A, Gumna J, Garfinkel DJ, Purzycka KJ.
901 2015. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic
902 acid chaperone functions. *Nucleic Acids Res* **43**:7414–7431.

903 O'Donnell S, Yue J-X, Saada OA, Agier N, Caradec C, Cokelaer T, De Chiara M, Delmas S,
904 Dutreux F, Fournier T, Friedrich A, Kornobis E, Li J, Miao Z, Tattini L, Schacherer J, Liti G,
905 Fischer G. 2022. 142 telomere-to-telomere assemblies reveal the genome structural
906 landscape in Saccharomyces cerevisiae. *bioRxiv*. doi:10.1101/2022.10.04.510633

907 Oggenfuss U, Croll D. 2023. Recent transposable element bursts are associated with the
908 proximity to genes in a fungal plant pathogen. *PLoS Pathog* **19**:e1011130.

909 O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement
910 activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*
911 **393**:68–72.

912 Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* **34**:401–437.

913 Pasyukova EG, Nuzhdin SV, Morozova TV, Mackay TFC. 2004. Accumulation of transposable
914 elements in the genome of Drosophila melanogaster is associated with a decrease in
915 fitness. *J Hered* **95**:284–290.

916 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
917 P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,
918 Duchesnay E. 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**:2825–
919 2830.

920 Pereira V. 2008. Automated paleontology of repetitive DNA with REANNOTATE. *BMC*
921 *Genomics* **9**:1–21.

922 Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR
923 retrotransposable elements and ectopic recombination in Drosophila. *Mol Biol Evol* **20**:880–
924 892.

925 Plante S, Landry CR. 2020. Purification of Yeast Spores to Investigate Their Dynamics of
926 Activation. *Curr Protoc Microbiol* **59**:e123.

927 Rech GE, Radío S, Guirao-Rico S, Aguilera L, Horvath V, Green L, Lindstadt H, Jamilloux V,
928 Quesneville H, González J. 2022. Population-scale long-read sequencing uncovers
929 transposable elements associated with gene expression variation and adaptive signatures
930 in Drosophila. *Nat Commun* **13**:1–16.

931 Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid
932 speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in
933 hybrid sunflowers. *Philos Trans R Soc Lond B Biol Sci* **369**:20130345–20130345.

934 Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**:155–
935 158.

936 Ryan OW, Poddar S, Cate JHD. 2016. Crispr–cas9 genome engineering in Saccharomyces
937 cerevisiae cells. *Cold Spring Harb Protoc* **2016**:525–533.

938 Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, Lee ME, DeLoache W, Dueber JE,
939 Arkin AP, Cate JHD. 2014. Selection of chromosomal DNA libraries using a multiplex
940 CRISPR system. *Elife* **3**. doi:10.7554/eLife.03703

941 Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, Garfinkel DJ. 2015. A
942 trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number
943 control. *J Virol* **89**:3922–3938.

944 Salinero AC, Knoll ER, Zhu ZI, Landsman D, Curcio MJ, Morse RH. 2018. The Mediator co-
945 activator complex regulates Ty1 retromobility by controlling the balance between Ty1i and
946 Ty1 promoters. *PLoS Genet* **14**:e1007232.

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

949 Shoshani O, Bakker B, de Haan L, Tijhuis AE, Wang Y, Kim DH, Maldonado M, Demarest MA,
950 Artates J, Zhengyu O, Mark A, Wardenaar R, Sasik R, Spierings DCJ, Vitre B, Fisch K,
951 Foijer F, Cleveland DW. 2021. Transient genomic instability drives tumorigenesis through
952 accelerated clonal evolution. *Genes Dev* **35**:1093–1108.

953 Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for
954 efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**:19–27.

955 Smit AFA, Hubley R, Green P. 2015. RepeatMasker.

956 Smukowski Heil C, Patterson K, Shang-Mei Hickey A, Alcantara E, Dunham MJ. 2021.
957 Transposable element mobilization in interspecific yeast hybrids. *Genome Biol Evol* **13**.
958 doi:10.1093/gbe/evab033

959 Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. 2017. Loss of

960 Heterozygosity Drives Adaptation in Hybrid Yeast. *Mol Biol Evol* **34**:1596–1612.

961 Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis
962 Version 11. *Mol Biol Evol* **38**:3022–3027.

963 Trakala M, Aggarwal M, Sniffen C, Zasadil L, Carroll A, Ma D, Su XA, Wangsa D, Meyer A,
964 Sieben CJ, Zhong J, Hsu P-H, Paradis G, Ried T, Holland A, Van Deursen J, Amon A.
965 2021. Clonal selection of stable aneuploidies in progenitor cells drives high-prevalence
966 tumorigenesis. *Genes Dev* **35**:1079–1092.

967 Tusso S, Suo F, Liang Y, Du L-L, Wolf JBW. 2021. Reactivation of transposable elements
968 following hybridization in fission yeast. *Genome Res* **32**:324–336.

969 Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species
970 is associated with retrotransposon proliferation. *Curr Biol* **16**:R872–3.

971 Van Rossum G, Drake FL. 2009. Python 3 reference manual. Scotts Valley, CA: CreateSpace.

972 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
973 Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive
974 Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**:e112963.

975 Wilhelm F-X, Wilhelm M, Gabriel A. 2005. Reverse transcriptase and integrase of the
976 Saccharomyces cerevisiae Ty1 element. *Cytogenet Genome Res* **110**:269–287.

977 Wilke CM, Adams J. 1992. Fitness effects of Ty transposition in Saccharomyces cerevisiae.
978 *Genetics* **131**:31–42.

979 Wilson AJ, Xu J. 2012. Mitochondrial inheritance: Diverse patterns and mechanisms with an
980 emphasis on fungi. *Mycology* **3**:158–166.

981 Yue JX, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergström A, Coupland P, Warringer J,
982 Lagomarsino MC, Fischer G, Durbin R, Liti G. 2017. Contrasting evolutionary genome
983 dynamics between domesticated and wild yeasts. *Nat Genet* **49**:913–924.

984 Zheng Q. 2017. rSalvador: An R Package for the Fluctuation Experiment. *G3* **7**:3849–3856.

985