The effect of hybridization on transposable element accumulation in an undomesticated fungal species

Author affiliations

Mathieu Hénault[1245]*, Souhir Marsit[12345], Guillaume Charron[1345#a] 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

[#a]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 mobile genetic elements that can profoundly impact the evolution of genomes and species. A long-standing hypothesis suggests that hybridization could deregulate TEs and trigger their accumulation, although it received mixed support from studies in plants and animals. Here, we tested this hypothesis in fungi using incipient species of the undomesticated yeast *Saccharomyces paradoxus*. Population genomic data revealed no signature of higher transposition in natural hybrids. As we could not rule out the elimination of past transposition increase signatures by natural selection, we performed a laboratory evolution experiment on a panel of artificial hybrids to measure TE accumulation in the near absence of selection. Changes in TE copy numbers were not predicted by the level of evolutionary divergence between the parents of a hybrid genotype. Rather, they were highly dependent on the individual hybrid genotypes, showing that strong genotype-specific deterministic factors govern TE accumulation in yeast hybrids.

2

Introduction

Hybridization is increasingly recognized as an important component of species evolution (Mallet, 2005; Mallet et al., 2016; Stukenbrock, 2016). Abundant empirical evidence shows that hybridization frequently leads to transgressive phenotypes in plants, animals and fungi (Langdon et al., 2019; Rieseberg et al., 1999). Mutation rates are traits for which transgressive segregation can have a major impact on the evolutionary fate of hybrids, with potential implications for the emergence of reproductive barriers and speciation. Insertions caused by transposable elements (TEs) are a class of mutations that could be particularly responsive to hybridization. TEs are diverse and ubiquitous DNA sequences that can duplicate and spread within genomes, forming families of dispersed repeated sequences (Bourque et al., 2018). They can either replicate through “copy-and-paste” (retrotransposons) or “cut-and-paste” (DNA transposons) mechanisms. TEs were first identified in cytological studies of chromosome instability in *Zea mays* (McClintock, 1950). In her seminal 1984 perspective, McClintock summarized evidence suggesting that TEs respond to various types of “genome shocks”, one of which she suspected was hybridization (McClintock, 1984). Since McClintock’s early insights, many studies investigated the association between TEs and hybridization across taxonomic groups and documented increased TE activity in hybrids. We refer to this increase as TE reactivation, but our use of this term implies no mechanistic basis and other terms like derepression are used in the literature. Reactivation was evidenced by increased TE abundance, novel TE insertions, TE epigenetic changes or higher TE transcript levels (Dion-Côté et al., 2014; Kidwell et al., 1977; Labrador et al., 1999; Liu and Wendel, 2000; O’Neill et al., 1998; Picard, 1976; Ungerer et al., 2006). However, hybrid reactivation is not universal, as many studies across a similar taxonomic range found no evidence for it (Coyne, 1989; Göbel et al., 2018; Hey, 1988; Kawakami et al., 2011). In fungi, it was recently shown that no TE

reactivation occurs in experimental interspecific hybrids between the yeasts *Saccharomyces cerevisiae* and *Saccharomyces uvarum* (Smukowski Heil et al., 2020). Thus, TE reactivation exhibits no clear taxonomic distribution, and the conditions favorable for its manifestation are still unclear.

Studies of hybrids in controlled crosses revealed genetic bases to TE reactivation, but less focus was put on cases of natural hybrids (but see (Kawakami et al., 2011; Ungerer et al., 2009)). In addition to the genetic structure of populations, environmental pressures and demographic factors can influence the action of evolutionary forces on TEs, thus impacting the genomic landscapes of TEs in natural populations, including hybrids. Population genetic models have investigated the action of the major evolutionary forces on mean TE copy number (CN) per individual (Charlesworth and Charlesworth, 1983; Charlesworth and Langley, 1989). Three such forces are deterministic (Charlesworth et al., 1994). First, TE CNs increase at a certain transposition rate, which incorporates the various forms of transposition regulation (either by the host or by TEs themselves). Second, TEs are removed from genomes at a certain excision rate. Third, most TE insertions are deleterious, such that individual fitness is assumed to be a decreasing function of CN and natural selection acts to purge TE insertions from populations. The fourth important force is genetic drift, a stochastic force that impacts the efficiency of natural selection by increasing the probability of fixation of deleterious mutations (Lynch, 2007). The intensity of drift is inversely proportional to the effective population size N_e. In populations with small N_e, drift renders natural selection inefficient, such that deleterious mutations are more likely to fix by chance. In principle, any of those evolutionary forces could affect hybrids in a way that differs from their parents, leading to hybrid-specific dynamics in the accumulation of TEs.

Population data in *Drosophila* indicate that deterministic forces have a predominant influence on the allele frequency spectrum of TE insertions (Charlesworth and Langley, 1989), consistent

with a minimal contribution from drift. Among the deterministic forces, transposition rate is arguably the most likely to be impacted by hybridization, for instance if genetic incompatibilities impair transposition regulation mechanisms. A classic illustration of this is hybrid dysgenesis in *Drosophila*, a syndrome caused by mismatches between paternal TEs and egg-loaded pools of Piwi-interacting RNAs (piRNAs) that lead to inefficient TE repression (Brennecke et al., 2008; Bucheton et al., 1984; Erwin et al., 2015; Rubin et al., 1982). Similarly, hybrids between *Arabidopsis thaliana* and *Arabidopsis arenosa* show overexpression of a TE family from the *A. arenosa* subgenome that is dependent on the ploidy level of the parental accessions involved, suggesting that dosage imbalance of epigenetic regulators causes reactivation (Josefsson et al., 2006). It can also be expected that the number of genetic incompatibilities increases with genetic divergence, yielding a positive relationship between evolutionary divergence and TE deregulation (Orr, 1995; Parhad and Theurkauf, 2019). Additionally, evidence from *Drosophila* suggests that hybridization may impact excision rates for specific TE families (Coyne, 1989). Along with these deterministic factors, reductions in natural selection efficiency should also be considered. Evidence from *Arabidopsis* (Lockton et al., 2008; Lockton and Gaut, 2010), *Caenorhabditis* (Dolgin et al., 2008) and *Daphnia* (Ye et al., 2017) suggests that long-term differences in N_e caused by demography or mating systems are associated with variation in TE content (Lynch and Conery, 2003). A reduction in N_e can be expected when a novel hybrid lineage is formed. Indeed, it has been hypothesized that population bottlenecks inherent with hybrid speciation in *Helianthus* might have played a role in the accumulation of TEs through a decreased efficiency of purifying selection (Ungerer et al., 2009).

An emerging eukaryotic model system in ecological and evolutionary genomics is the yeast *Saccharomyces paradoxus*, the undomesticated sister species of the budding yeast S. *cerevisiae*. *S. paradoxus* has a complex admixture history in North America, making it an ideal

101 model system to test hybridization-related evolutionary hypotheses in natural populations and in
102 the laboratory (Figure 1) (Hénault et al., 2017). North American *S. paradoxus* lineages comprise
103 the Eurasian lineage *SpA*, the endemic incipient species *SpB* and *SpC* (hereafter termed pure
104 lineages), and the incipient hybrid species *SpC**, SpD_1 and SpD_2 (hereafter termed hybrid
105 lineages; Figure 2a). *SpC** is an introgressed lineage that arose from a post-glacial secondary
106 contact between the endemic lineages *SpB* and *SpC* and consists of mostly *SpC*-like genomes
107 with ~5% introgressions from *SpB* (Leducq et al., 2016), while SpD_1 and SpD_2 arose more
108 recently from hybridization between *SpC** and *SpB* and have roughly equal genomic
109 contributions from both parents (Eberlein et al., 2019) (Figure 2b).

110 Yeasts of the *Saccharomyces* genus are valuable model species to investigate the interplay
111 between hybridization and transposition. *Saccharomyces* genomes contain many long terminal
112 repeats (LTR) retrotransposons of the Ty1/copia (Ty1, Ty2, Ty4 and Ty5) and Ty3/gypsy (Ty3)
113 families. Ty1/copia and Ty3/gypsy retrotransposon superfamilies (Wicker et al., 2007)
114 respectively correspond to the Pseudoviridae and Metaviridae virus families (Krupovic et al.,
115 2018) and have a broad taxonomic distribution. Ty elements comprise two overlapping open
116 reading frames (ORFs) within a ~5 kb internal sequence, which is flanked by two ~300 bp LTRs
117 in direct orientation. The two LTRs of full-length elements are prone to intra-element
118 recombination, leaving behind solo LTRs in the genome (Parket et al., 1995). Ty1 in the budding
119 yeast *S. cerevisiae* is among the best characterized eukaryotic TEs, benefiting from decades of
120 research on its molecular and cellular biology and, more recently, on its population genomics
121 (Bleykasten-Grosshans et al., 2013; Curcio et al., 2015). Although *Saccharomyces* genomes
122 harbor a relatively modest fraction of Ty elements (~1-3% in size), their streamlined genomes
123 (~12 Mbp) and short generation time (~2h) provide great power to perform TE comparative
124 genomics and laboratory evolutionary experiments.

125 Here, we use population genomic data and laboratory evolution experiments on *S. paradoxus*
126 and its sibling species *S. cerevisiae* to investigate the hybrid reactivation hypothesis and, more
127 generally, the factors governing TE accumulation in natural and experimental lineages. Our
128 results show that natural and experimental hybrids exhibit no reactivation of TEs. Rather, we
129 show that deterministic factors like population structure and the properties of individual hybrid
130 genotypes are major determinants of TE content evolution in hybrid genomes.

7

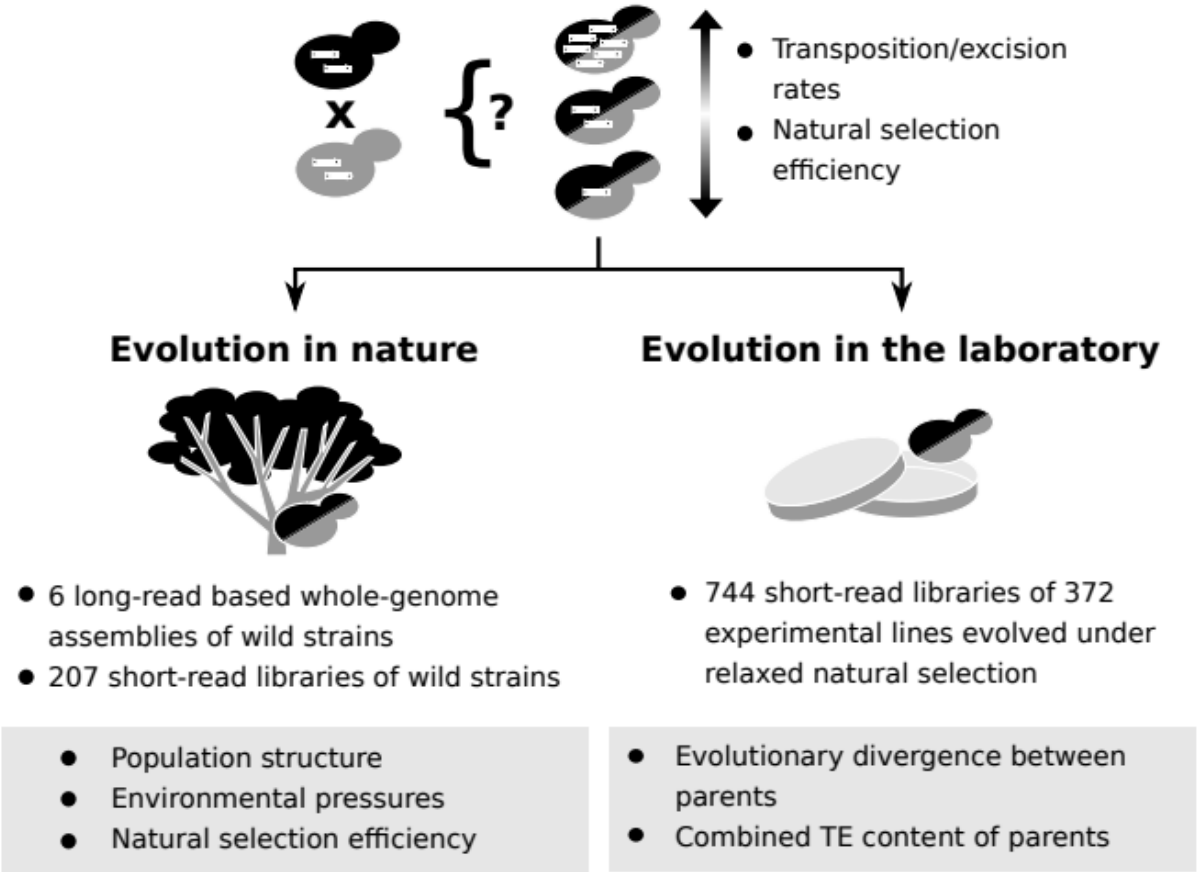


Figure 1. Population genomic data and laboratory evolution experiments allow to uncover the effects of various factors on TE accumulation in hybrids. Upon hybridization, the TE content of a hybrid is determined by equal contributions from the TE contents of its parents. Subsequent evolution can either leave the hybrid TE content unchanged, or drive it to increase or decrease. These changes can be driven by biases in transposition or excision rates or by variation in natural selection efficiency. Two complementary approaches can be used to investigate which factors drive TE dynamics in hybrids. First, genomic data from natural populations can be harnessed to understand how population structure, environmental pressures and natural selection efficiency shaped natural variation in TE content. Second, evolution experiments in laboratory controlled conditions (constant environment and relaxed natural

8

141 selection) allow to test the effect of properties of artificial hybrid genotypes on TE accumulation,
142 namely evolutionary divergence between parents and initial TE content.

143 Results

144 *S. paradoxus lineages exhibit contrasted TE evolutionary dynamics but no reactivation in*
145 *hybrids*

146 To investigate the TE content and dynamics in North American *S. paradoxus* pure and hybrid
147 lineages, we annotated six highly contiguous whole-genome assemblies previously generated
148 from Oxford Nanopore long reads (PRJNA514804) (Eberlein et al., 2019). We used a database
149 comprising reference sequences for *S. cerevisiae* Ty1, Ty2, Ty3, Ty4 and Ty5 families, along
150 with *S. paradoxus* Ty3-related Ty3p (Carr et al., 2012) and Ty4-related Tsu4, the latter
151 originating from a horizontal transfer from the *Saccharomyces eubayanus*/*Saccharomyces*
152 *uvarum* clade to *S. paradoxus* (Bergman, 2018). These annotations revealed counts of full-
153 length elements an order of magnitude lower than solo LTRs (Figure 2c, Figure 2—figure
154 supplement 1), consistent with previously reported data on *S. cerevisiae* and *S. paradoxus* (Carr
155 et al., 2012; Goffeau et al., 1996; Kim et al., 1998; Yue et al., 2017). To overcome the difficulty of
156 comparing multiple families of short and often degenerated LTR sequences within a
157 phylogenetic framework, we built intra-genome LTR sequence similarity networks. Sequences
158 within the Ty1-Ty2 and Ty3p-Ty3 LTR family pairs were indistinguishable (Figure 2—figure
159 supplement 2), in agreement with Ty2 and Ty3 being specific to *S. cerevisiae* (Carr et al., 2012;
160 Naumov et al., 1992). Consequently, Ty2 and Ty3 LTR annotations were merged with Ty1 and
161 Ty3p, respectively. Although they were clearly distinct, Ty4 and Tsu4 LTRs showed a close
162 relatedness, which justified their combination into the Ty4/Tsu4 family. No Tsu4 LTR was

annotated in the *SpA* genome, consistent with this family being specific to North American *S. paradoxus* and absent from Eurasian *S. paradoxus* (Bergman, 2018).

10

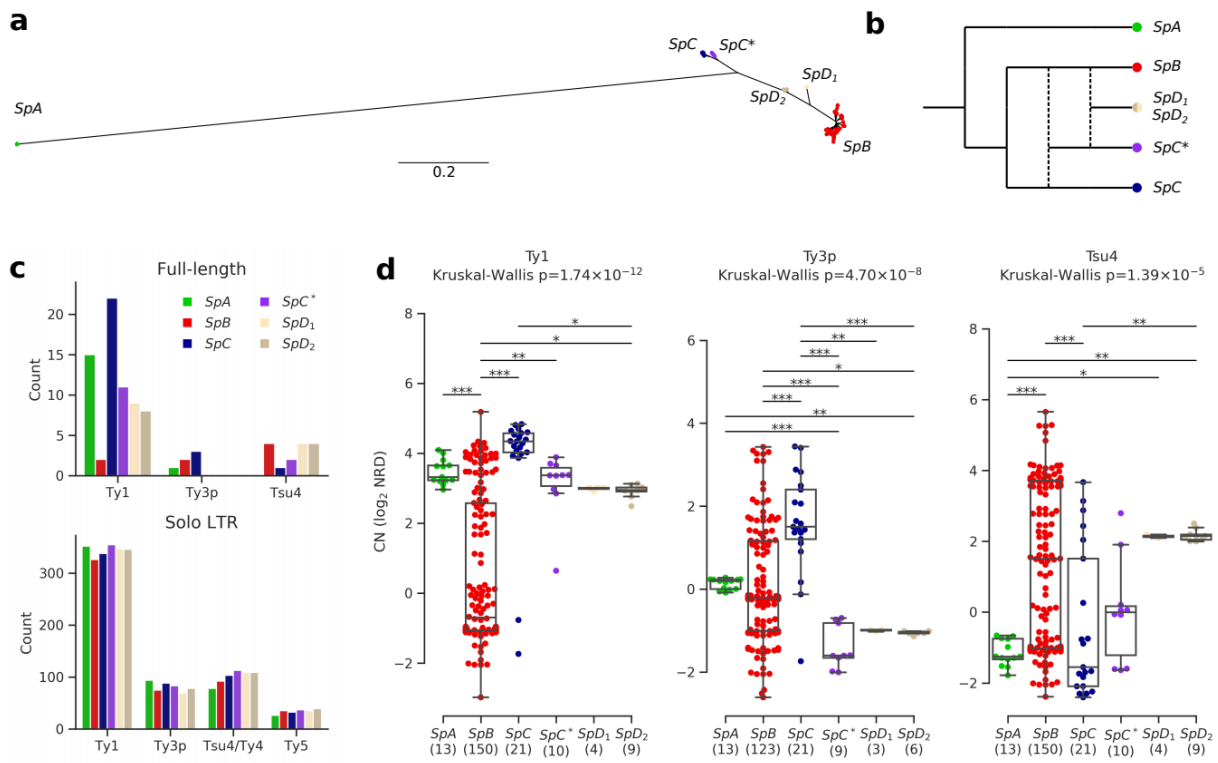


Figure 2. Ty elements showed no reactivation in hybrid lineages and contrasted abundances among pure lineages. a. Bayesian phylogenetic tree of the North American *S. paradoxus* lineages based on genome-wide SNPs. **b.** Schematic admixture graph showing the evolutionary relationships between pure and hybrid lineages as inferred by (Eberlein et al., 2019; Leducq et al., 2016). Dotted lines represent admixture events. **c.** Counts of full-length elements (top) and solo LTRs (bottom) in six whole-genome assemblies based on long reads. No full-length Ty4 or Ty5 were found in any assembly, thus these families were considered extinct. **d.** CN variation in active Ty families (i.e. families which have full-length elements) measured as $\log_2$ normalized read depth (NRD) over Ty reference internal sequences for 207 diploid wild strains. Sample size for each lineage (excluding strains for which there was no coverage) is indicated between parentheses. Whiskers span 1.5 times the interquartile range.

11

176 P-values of Kruskal-Wallis tests are shown. Stars on bars show the results of Conover post-hoc
177 tests. *:p<0.05, **:p<0.01, ***:p<0.001.

178 Based on the assembly annotations, Ty1 was the most abundant family and exhibited the
179 largest CN variation among lineages, with a striking difference between *SpB* and *SpC* (Figure
180 2c). The generality of CN patterns revealed by genome assemblies was confirmed by
181 normalized short-read depth (NRD) over Ty reference sequences for 207 natural diploid isolates
182 (Figure 2d, Figure 2—figure supplement 3, Table S1; PRJNA277692, PRJNA324830,
183 PRJNA479851). The hybrid lineages *SpC**, SpD_1 and SpD_2 had intermediate full-length Ty1 and
184 Tsu4 CNs compared to their parental lineages, indicating no apparent persistent reactivation or
185 maintenance of Ty elements in these lineages. Notably, Ty3p appeared extinct in all hybrid
186 lineages, with no signal for internal sequences. The most outstanding profile was that of the
187 pure lineage *SpC*, which had significantly higher full-length Ty1 and Ty3p CNs compared to the
188 other lineages. For Tsu4, *SpB* was the lineage with the highest CN. Thus, lineage-wide CNs
189 were consistent with an absence of reactivation in hybrids, the most contrasted profiles being
190 found among pure lineages.

191 CNs alone may not reflect the actual dynamics of a TE family within a genome. For instance, a
192 recently expanding family may not have had the time to reach the CN of another family (or the
193 same family in a different lineage) that would be more abundant but at equilibrium. We thus
194 explored the evolutionary dynamics of Ty families within each genome. If high CNs were caused
195 by recent family expansions, they should be associated with a signature of low divergence
196 (Bergman and Bensasson, 2007). We used the sequence divergence to the most closely related
197 LTR for each LTR annotation to approximate evolutionary divergence since transposition. The
198 importance of low-divergence peaks (≤1%) for Ty1 was significantly higher in *SpC* and *SpA*
199 compared to *SpB* (Fisher's exact test, FDR corrected, Figure 3a; see Figure 3—figure

supplement 1 for other divergence thresholds). *SpA* also had a significant underrepresentation of low-divergence LTRs for the composite Ty4/Tsu4 family due to the absence of Tsu4. The height of low-divergence peaks was consistent with CNs of full-length elements (Figure 2d). Hybrid lineages *SpC**, SpD_1 and SpD_2 had Ty1 peaks similar to *SpC*, although less important. They also had no peak for Ty3p, consistent with the absence of full-length elements. Overall, this result indicated that CN variation reflected variation in family dynamics and confirmed that Ty elements exhibited no reactivation in hybrid lineages.

13

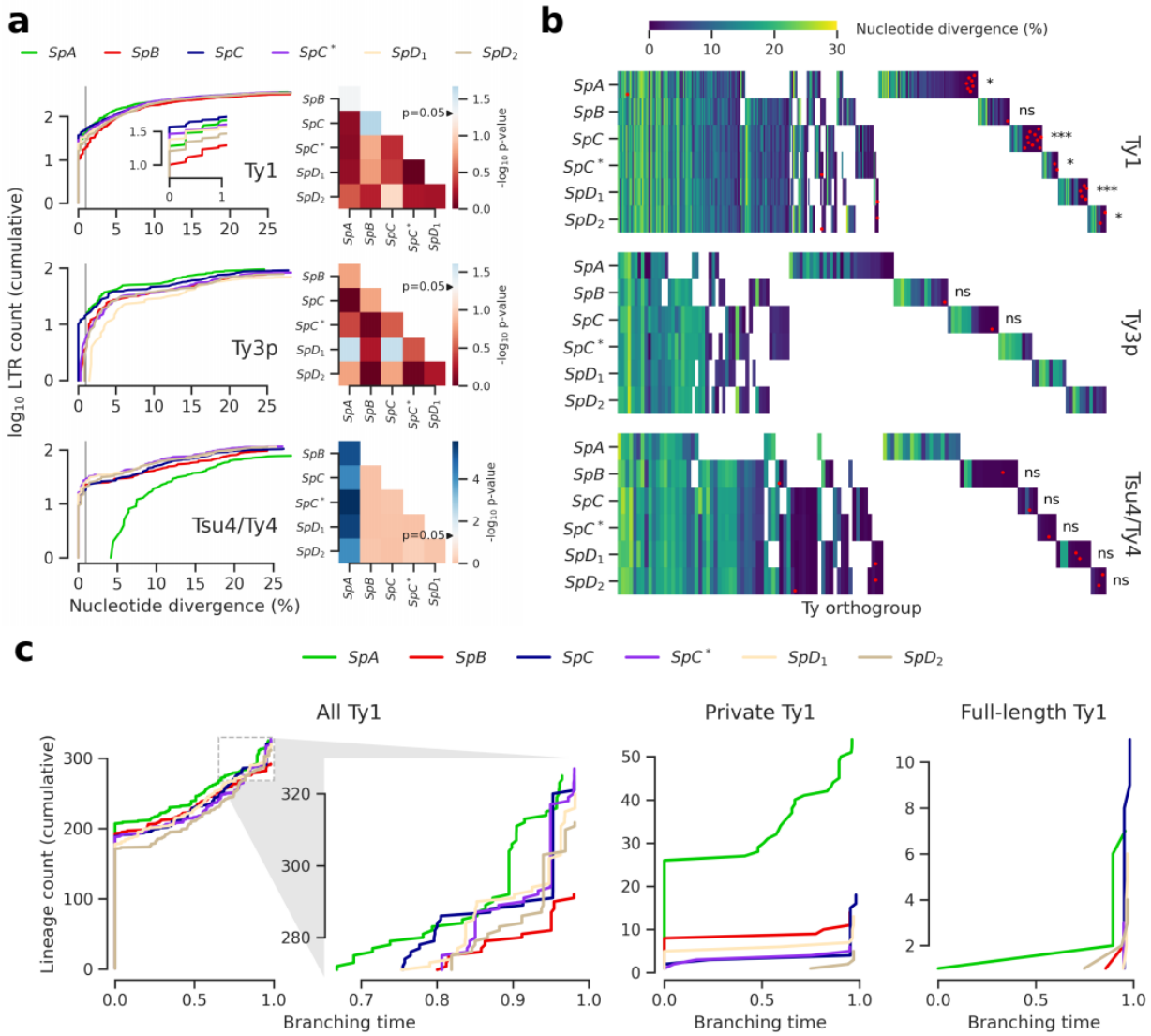


Figure 3. Evolutionary dynamics of Ty families in *S. paradoxus*. a. Distributions of minimum nucleotide divergence between LTR sequences. Heatmaps show FDR-corrected p-values for pairwise Fisher's exact tests between ratios of elements ≤1% and >1% nucleotide divergence. Color maps are centered at the significance threshold of 0.05. Blue cells depict statistically significant tests. See Figure 3—figure supplement 1 for other divergence thresholds. **b.** LTR divergence and conservation levels. Horizontal positions depict Ty orthogroups clustered from conserved (left) to private (right). Color map shows minimum nucleotide divergence. Scattered

14

red dots show LTRs that belong to full-length elements. FDR-corrected p-values of Fisher's exact tests for the ratio of full-length elements in private orthogroups are shown. ns: $p \geq 0.05$, *:$p<0.05$, **:$p<0.01$, ***:$p<0.001$. **c.** Lineages through time curves for Ty1 ultrametric bayesian phylogenetic trees, using a conserved Ty4 LTR orthogroup as an outgroup. Branching time is scaled such that values of one correspond to the present. From left to right, distributions show all LTRs (with a close-up on recent dynamics), LTRs from private elements and LTRs from full-length elements.

Next, we tested whether orthology between genome-wide Ty insertion loci was consistent with transposition activity. We defined Ty orthogroups from a whole-genome multiple alignment of the six assemblies (Figure 3b). Most orthogroups were either private to one genome or conserved across the six (Figure 3—figure supplement 2), suggesting that a major fraction of Ty insertion polymorphisms segregated at low frequency in North American lineages. This was consistent with studies in *S. cerevisiae* showing that insertions are mostly fixed or at low frequency (Bleykasten-Grosshans et al., 2013; Carr et al., 2012). Private orthogroups comprised LTRs that had low divergence, consistent with recent transposition activity (Figure 3—figure supplement 2). Private orthogroups were significantly enriched in full-length Ty1 elements (Figure 3b), including hybrid lineages. This indicated that all lineages had active transposition for at least some Ty families. It also suggested that an absence of transposition in hybrid lineages could not explain the absence of reactivation. We called lineage-wide insertion polymorphisms from discordant mappings of the 207 short read libraries to estimate the distribution of insertion allele frequencies. This calling method is noisy and only allows to call non-reference alleles, but should nevertheless provide rough estimates of the insertion allele frequency spectra. Principal component analysis (PCA) on non-reference insertions yielded the expected grouping of strains into well-defined lineages (Figure 3—figure supplement 3). The allele frequency distributions

15

238 showed a similar excess of very low frequency alleles for the six lineages (Figure 3—figure
239 supplement 3), suggesting that Ty elements were subject to largely uniform evolutionary forces
240 across lineages.

241 We focused on the most abundant Ty1 to investigate the transposition dynamics in a
242 phylogenetic framework (Le Rouzic et al., 2013). Bayesian phylogenetic trees of Ty1 LTRs were
243 converted into ultrametric phylogenies to approximate the timing of branching events using a
244 conserved Ty4 LTR orthogroup as an outgroup. Branching times revealed that lineages had
245 largely similar Ty1 dynamics (Figure 3c). Due to the shortness of LTR sequences, most
246 branches collapsed with the outgroup, indicating an expected saturation of the molecular
247 substitution signal. A focus on recent dynamics revealed a clear expansion signature in *SpC* that
248 was shared by *SpC** and, to a lesser extent, by SpD_1 and SpD_2 (Figure 3c), which was reflected
249 by the peaks of young private and full-length elements. This suggests that any recent expansion
250 signature in hybrid lineages can be largely explained by inheritance from their *SpC*-related
251 parent. The rapid expansion in *SpC* and *SpC** can be contrasted with the more gradual
252 expansion in *SpA*, which harbored the second highest Ty1 CN. This contrasted with the
253 assumption that TE abundances can be interpreted as equilibrium copy numbers under the
254 transposition-selection balance hypothesis (Charlesworth and Charlesworth, 1983) and rather
255 suggested a recent transposition rate increase in *SpC*-related lineages.

256 *Population structure best explains TE content variation, with minor contributions from climatic*
257 *variation*

258 Our results showed that hybrid lineages exhibited no outstanding Ty CN profiles or sequence
259 dynamics. However, potential hybrid-specific differences in transposition rate might be masked
260 by other factors that can be expected to impact Ty dynamics. To exclude this possibility, we
261 investigated which factors best explain the variation in Ty CN in natural lineages. First,

262 population structure incorporates both host-encoded (Curcio et al., 2015) and Ty-encoded
263 (Czaja et al., 2020; Saha et al., 2015) determinants of Ty transposition rates. Second,
264 environmental factors can influence Ty transposition rate, as demonstrated by temperature
265 (Paquin and Williamson, 1984) and nitrogen starvation (Morillon et al., 2000) for Ty1 in *S.*
266 *cerevisiae*. If hybrid lineages experience distinct environmental conditions in nature, this may
267 impact their transposition rates. Finally, the efficiency at which natural selection purges
268 deleterious Ty insertions (which depends on N_e) could vary between lineages, ultimately
269 affecting the transposition-selection balance. To help discriminate global from lineage-specific
270 patterns arising from these factors, analyses with all lineages aggregated were complemented
271 by lineage-specific analyses. However, hybrid lineages were represented by relatively few
272 strains (≤10 in each case), which made it hard to reach meaningful conclusions for some
273 analyses. Thus, some of the lineage-specific analyses were focused on the more abundant *SpB*
274 and *SpC* pure lineages.

275 We first investigated the relative contribution of population structure and environmental factors.
276 We combined genome-wide single nucleotide variants from our population genomic data with
277 climatic (temperature, precipitation, solar radiation, vapor pressure, wind) and water balance
278 data (Abatzoglou et al., 2018) and reduced the dimensionality of both datasets by PCA, using
279 coordinates on the major principal components (PCs) to describe population structure (Figure 4
280 —figure supplement 1) and climatic variation (Figure 4—figure supplement 2, Figure 4—figure
281 supplement 3). We then fitted linear models to compare how genotypes-based PCs (gPCs) and
282 environment-based PCs (ePCs) can explain CN variation in Ty1, Ty3p and Tsu4 (Figure 4a).
283 Globally, gPCs reached higher levels of statistical significance as linear predictors of CN
284 compared to ePCs. gPC1 both explained a high proportion of variance (52%, 9% and 36% for
285 all lineages, *SpB* and *SpC*, respectively; Figure 4—figure supplement 1) and was a highly

286 significant linear predictor of Ty1 CN. Linear models that explained Ty CNs with both gPCs and
287 ePCs had generally lower Akaike information criterion (AIC) values than models that included
288 only gPCs (Table S2), suggesting that climatic variables have modest although significant
289 contributions in explaining CN variation. Unsupervised clustering of strains along gPCs
290 consistently yielded more homogeneous clusters than ePCs in terms of Ty CN (Figure 4b,
291 Figure 4—figure supplement 4). Overall, these results show that the genetic structure of North
292 American *S. paradoxus* is generally a much better predictor of Ty CNs across lineages than
293 climatic variables. Since locations where hybrid lineages were sampled have climatic profiles
294 similar to pure lineages (Figure 4—figure supplement 2), exposure to specific environmental
295 conditions appears unlikely to explain the absence of observable Ty reactivation in hybrid
296 lineages.

18

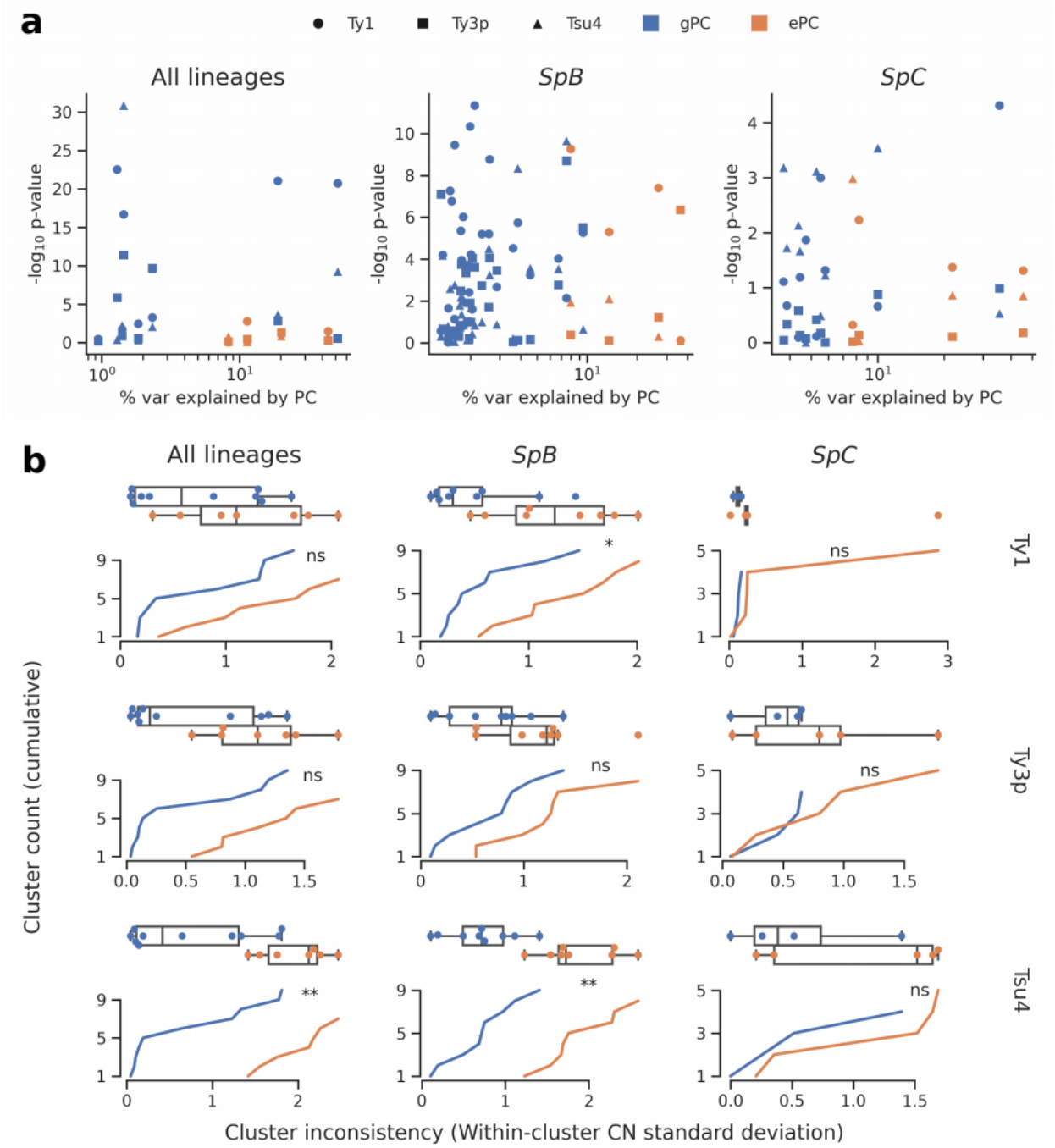


Figure 4. Associations between population structure, climatic variables and Ty CNs for all lineages (left), *SpB* only (middle) or *SpC* only (right). a. Linear models fitted on Ty CNs with gPCs and ePCs as independent variables. For each PC, the p-value in the linear model is plotted against the proportion of variance explained in the respective PCA. **b.** Unsupervised

19

clustering performed on either gPC or ePC coordinates. Cumulative distributions show the standard deviation in Ty CN within each gPC-cluster or ePC-cluster. Distributions are also shown as boxplots, with whiskers spanning 1.5 times the interquartile range. FDR-corrected p-values of two-tailed Mann-Whitney U tests between gPC-cluster and ePC-cluster CN distributions are shown. ns: $p \geq 0.05$, *:$p<0.05$, **:$p<0.01$.

Variation in natural selection efficiency cannot explain TE abundance variation in wild lineages

Most TE insertions are expected to have deleterious effects on the individual in which they occur (Wilke and Adams, 1992), although some can be advantageous (Aminetzach et al., 2005; Filteau et al., 2015; Gresham et al., 2008; Hof et al., 2016). Natural selection acts to purge deleterious insertions, and its efficiency depends on the strength of genetic drift opposing it (which is inversely proportional to N_e). Ty reactivation in hybrid lineages might be counterbalanced by higher efficiency of natural selection in these lineages, although there is no *a priori* reason to expect higher N_e in hybrids. We thus sought to investigate if CN variation across lineages can be explained by variation in N_e. Genome-wide nucleotide diversity at four-fold degenerate codon positions (π_S) is related to long-term N_e and was used to approximate this parameter across diverse taxa (Lynch and Conery, 2003). Nucleotide diversity was also used at the intraspecific level in *S. paradoxus* to estimate N_e (Tsai et al., 2008). While pure lineages did exhibit variation in π_S (Table S3), there was no obvious negative relationship between Ty CNs and π_S (Figure 5—figure supplement 1), as would be expected under a transposition-selection equilibrium (assuming negligible variation in mutation rate). For instance, nucleotide diversity was higher in *SpC* than in *SpA*, while *SpC* exhibited the highest Ty1 and Ty3p CN. While these rough N_e estimates should be interpreted with caution, they suggested that natural selection efficiency is likely not the main determinant in CN variation. However, our data indicated that some Ty families experienced recent expansions and thus may not be at

20

325 equilibrium. Recent hybrid lineages also cannot be meaningfully characterized by long-term N_e,
326 further limiting the usefulness of π_S as a measure of the efficiency of natural selection.

327 We can expect that most of the fitness effects associated with TE insertions are caused by
328 alterations of host functions encoded in *cis.* Perhaps the TE insertions with the largest functional
329 alteration potential are those which interrupt ORFs of host genes. Across the six genome
330 assemblies, no Ty insertion was found to interrupt an ORF. Lineage-wide Ty insertion calls
331 revealed four Ty insertions predicted to be within host ORFs, three of which are found in *SpC*
332 strains and one in *SpB*. All the ORF-disrupting insertions were private alleles, and in all cases
333 the strain harboring them had a CN higher than the median of its lineage (Figure 5—figure
334 supplement 2). Examination of depth of coverage at these four loci revealed narrow peaks in the
335 corresponding strains that were consistent with target site duplications (TSDs) arising from Ty
336 integration (Figure 5—figure supplements 3-6). The absence of ORF-disrupting Ty insertions in
337 hybrid lineages and their scarcity in pure lineages suggests that natural selection is globally very
338 efficient at purging these mutations.

339 The number of ORF-disrupting Ty insertion alleles in *SpC* (three) compared to *SpB* (one)
340 suggested a slight overrepresentation in *SpC*, which could be explained by a bias towards more
341 deleterious insertion alleles in this lineage. To test this hypothesis, we examined the functional
342 annotations of the corresponding genes. There were striking patterns of functional relatedness
343 among genes harboring an ORF-disrupting insertion. Two *SpC* strains had independent Ty1
344 insertions at different loci within YDL024C (*DIA3*, which encodes a protein of unknown function
345 (Palecek et al., 2000)), as evidenced by the corresponding TSDs (Figure 5—figure supplement
346 3, Figure 5—figure supplement 4). The two other genes were involved in resistance to arsenic
347 compounds: 16_037 (*SpB*) had a Tsu4 insertion in YPR200C (*ARR2*, which encodes an
348 arsenate reductase (Mukhopadhyay and Rosen, 1998), Figure 5—figure supplement 5), while

349 LL2011_004 (*SpC*) had a Ty1 insertion disrupting YPR201W (*ARR3*, which encodes an
350 arsenite/H+ antiporter (Wysocki et al., 1997), Figure 5—figure supplement 6). This functional
351 relatedness could be explained by positive selection; however, this scenario is incompatible with
352 the extremely low frequency at which these insertions segregate. Thus, the most parsimonious
353 explanations for these patterns appears to be either a mutational bias favoring Ty insertions at
354 these loci or fitness effects too small to be visible to natural selection.

355 We investigated the fitness effect of the ORF-disrupting Ty1 insertion in *ARR3* by measuring the
356 growth of a collection of 123 *S. paradoxus* strains in culture medium supplemented with arsenite
357 (Figure 5a, Figure 5—figure supplement 7, Figure 5—figure supplement 8). Growth in 0.8 mM
358 sodium meta-arsenite ($NaAsO_2$) showed significant variation between lineages (mixed design
359 ANOVA, $p=9.54\times10^{-39}$), and more total variance was explained by variation between lineages
360 (lineages: 86.58%, strains: 10.20%, residuals: 3.22%). *SpC* and *SpC** lineages exhibited poor
361 growth in 0.8 mM arsenite compared to *SpB* and *SpA* (Figure 5a). This phenotype was
362 generalized to the whole lineages, and while strain LL2011_004 was in the lower part of the
363 *SpC* distribution, it was not significantly different from many *SpC* strains not harboring an
364 insertion (Figure 5b, Figure 5—figure supplement 9). This indicated that the disruption of a
365 major gene involved in arsenite resistance had no further deleterious effect on the poor growth
366 ability of *SpC*-related strains in presence of arsenite. The absence of growth defect in arsenite
367 for LL2011_004, combined with the extremely low frequencies of all the ORF-disrupting alleles,
368 suggests that these insertions are effectively neutral and are not observed because of an
369 important reduction in purifying selection efficiency.

22

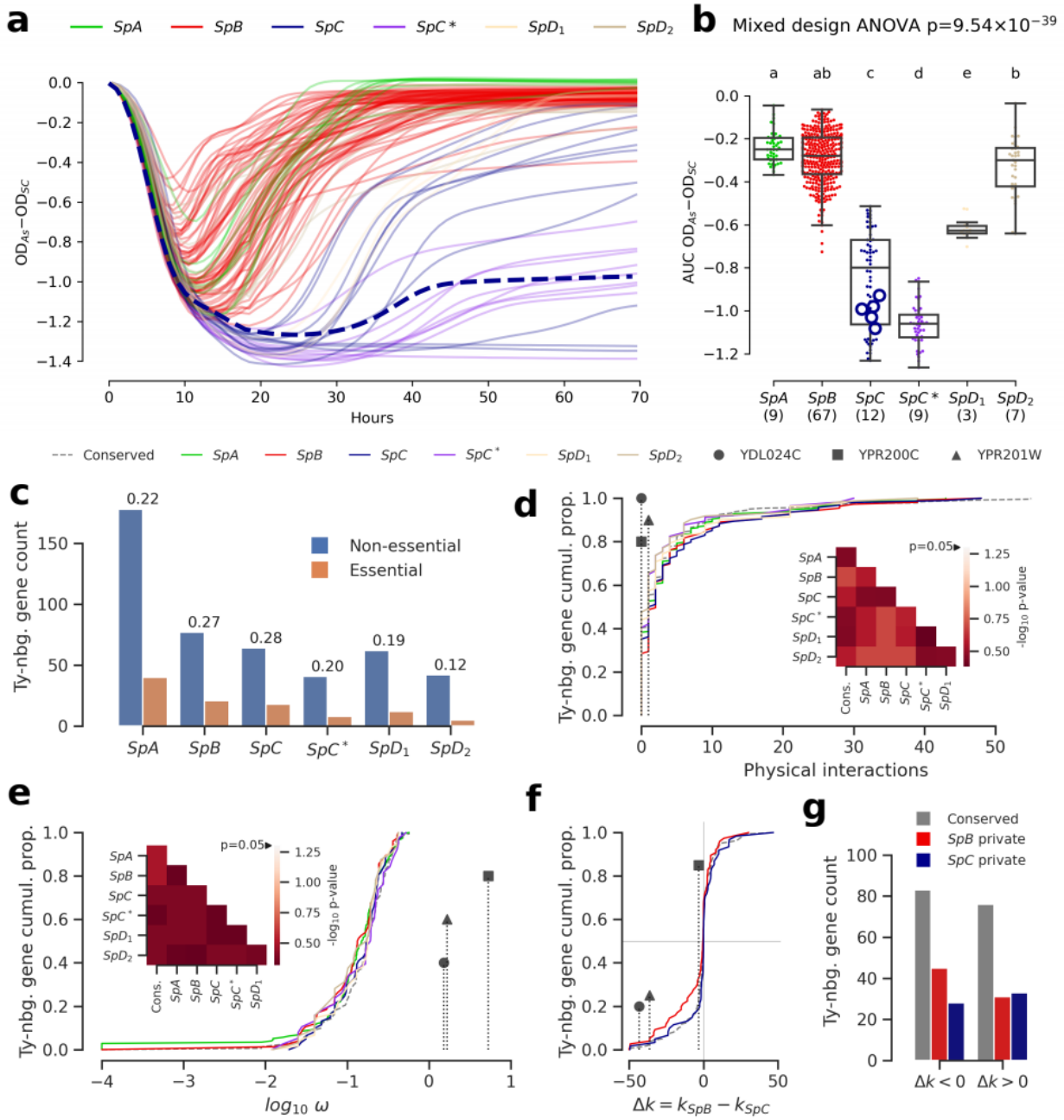


Figure 5. Genomic patterns of Ty insertions indicate that natural selection efficiency cannot explain CN variation in natural lineages. a. Difference between growth curves in 0.8 mM $NaAsO_2$ (As) and control medium (synthetic complete, SC). Curves depict the difference in culture optical density (OD) through time, averaged across replicates and bins of four timepoints

23

374 (one hour). 123 strains were assayed. *SpC* strain LL2011_004, which harbors a Ty1 insertion in
375 *ARR3*, is highlighted with a dashed line. **b.** Area under the curve (AUC) of the difference
376 between growth curves in As and SC. Replicates of 107 strains that passed the criteria of
377 number of replicates and homoscedasticity are shown. The number of strains per lineage is
378 shown between parentheses. Replicates of the *SpC* strain LL2011_004 are highlighted with
379 large empty circle symbols. Lowercase letters indicate groups that are not significantly different
380 following Tukey's HSD test. **c.** Counts of essential and non-essential genes in the immediate
381 vicinity of Ty insertions across the six whole-genome assemblies. Essential to non-essential
382 ratios are shown on top of the bars. **d-e.** Cumulative distributions of number of physical
383 interactions (d) and ω (e) for genes neighboring conserved (dashed line) or private (solid lines)
384 Ty insertions. Genes containing ORF-disrupting insertions are labeled with symbols. Heatmaps
385 show FDR-corrected p-values for pairwise Mann-Whitney U tests between distributions. Color
386 maps are centered at the significance threshold of 0.05. **f.** Cumulative distributions of Δ*k*
387 between *SpB* and *SpC*. **g.** Counts of neighboring genes with positive or negative Δk values.

388 To gain a more comprehensive view of the strength of selection acting on Ty insertions, we
389 compared the properties of genes in their neighborhood. Ty elements contain many transcription
390 factor binding sites and can affect the expression levels of adjacent genes (Morillon et al., 2000;
391 Coney and Roeder, 1988; Todeschini et al., 2005; Lesage and Todeschini, 2005). Dosage-
392 sensitive and functionally central genes should on average be subject to stronger selection
393 against mutations that affect their regulation. For instance, in *S. cerevisiae*, genes that are
394 essential or that have many protein-protein interactions were shown to be depleted in eQTLs,
395 i.e. variants that affect expression levels (Albert et al., 2018). However, the association of Ty
396 insertions with regulatory effects on adjacent genes is expected to be highly variable, as
397 transcriptional effects are known to depend on the location, orientation and nucleotide sequence

398 of the inserted element (Coney and Roeder, 1988; Lesage and Todeschini, 2005). Nevertheless,
399 if selection efficiency explains a significant part of Ty CN variation, we reason that when CNs
400 are high, Ty insertions should be more frequently observed near genes which properties reflect
401 stronger dosage and functional constraints, compared to when CNs are low. Focusing on the
402 Ty insertions private to each lineage, we found no significant difference in the proportion of
403 neighboring genes annotated as essential in *S. cerevisiae* ($p>0.05$, pairwise Fisher's exact
404 tests, FDR correction), although *SpC* did exhibit the highest essential to non-essential ratio
405 (Figure 5c). Distributions of physical distance to neighboring genes were similar across
406 lineages, as were distributions of relative orientation (Figure 5—figure supplement 10).
407 Additionally, genes neighboring Ty insertions private to *SpC* show no bias towards higher
408 number of physical interactions (Figure 5d, $p>0.05$, pairwise two-tailed Mann-Whitney U tests,
409 FDR correction). We fitted the ratio of non-synonymous to synonymous substitutions (ω) to
410 each neighboring gene to estimate the strength of purifying selection on protein-coding
411 sequences. The ω distributions for genes neighboring private Ty insertions exhibited no
412 significant bias ($p>0.05$, pairwise two-tailed Mann-Whitney U tests, FDR correction). In
413 particular, the distribution in *SpC* exhibited no skew towards lower values, suggesting no
414 association with genes under strong purifying selection (Figure 5e). Notably, genes harboring
415 ORF-disrupting insertions show little or no physical interactions and no signature of purifying
416 selection (Figure 5d-e).

417 We next investigated whether lineage-specific selection signatures on genes are associated
418 with Ty insertions. We used the RELAX branch-site codon model (Wertheim et al., 2015) to
419 derive estimates of selection intensity (k) along individual gene trees. Within the RELAX
420 framework, $k<1$ indicates relaxed selection and $k>1$ indicates strengthened (positive or
421 purifying) selection. We define $\Delta k = k_{SpB}-k_{SpC}$, corresponding to the difference between the

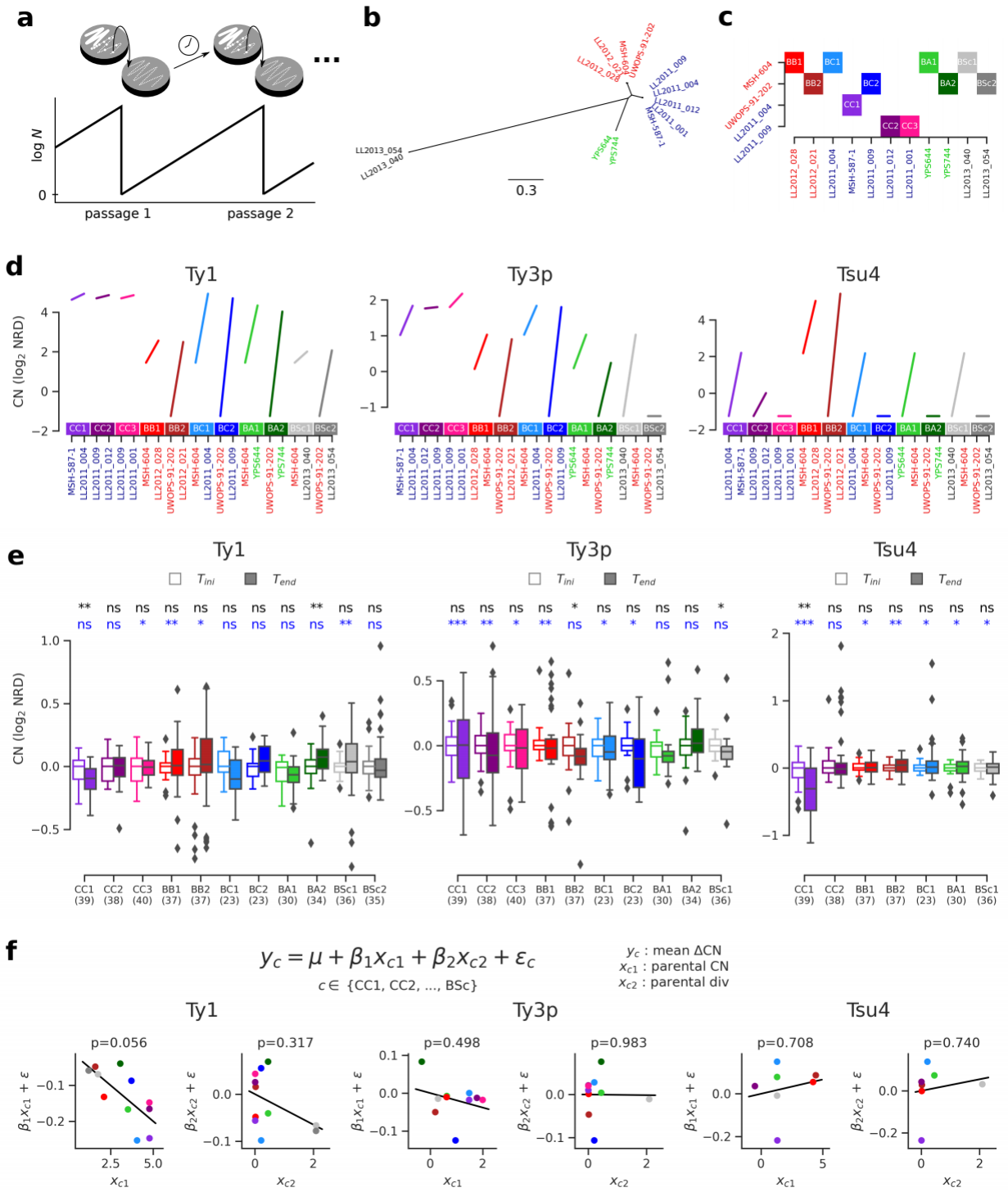
422 average *k* values for clades comprising *SpB* and *SpC* strains, respectively. Overall, the
423 distribution of Δ*k* values for genes neighboring private *SpC* insertions was very similar to that for
424 conserved insertions (Figure 5f), while the distribution for private *SpB* insertions was slightly
425 biased towards genes relaxed in *SpB*. For both *SpB* and *SpC*, there were more private
426 insertions adjacent to genes experiencing relaxed selection in the corresponding lineage,
427 although comparisons with conserved insertions were not statistically significant ($p>0.05$,
428 Fisher's exact tests, Figure 5g). This is consistent with efficient purifying selection acting against
429 insertions near genes in both lineages. While genes harboring ORF-disrupting insertions have
430 signatures of stronger selection in *SpC* (Figure 5e, Figure 5—figure supplement 11), this
431 observation is mitigated by their far-right position in the ω distribution (Figure 5d). Overall, these
432 results show that variation in natural selection intensity is unlikely to explain a large part of the
433 variation in Ty CNs in natural lineages.

434 *TE content variation in experimental evolution under relaxed selection reveal no systematic*
435 *effect of hybridization but major effects of individual genotypes*

436 Thus far, our results showed that no reactivation is apparent in natural hybrid lineages, that
437 population structure over environmental variables best explains variation in Ty CNs and that
438 natural selection efficiency is unlikely to explain much variation in Ty genomic landscapes.
439 However, natural selection seems equally efficient at purging deleterious Ty insertions in all
440 lineages, leaving the possibility that strong purifying selection could mask any reactivation in
441 hybrids. Transient reactivation may also be followed by rapid re-establishment of host
442 repression and leave little to no signature in contemporary genomes. Moreover, natural hybrids
443 are somewhat limited in their ability to uncover the interplay between hybridization and
444 transposition, as they are relatively recent introgressed lineages that resulted from admixture
445 between closely related lineages. To systematically explore the effect of hybridization on Ty

446 transposition rates, we performed a large-scale mutation accumulation (MA) experiment on 11
447 artificial hybrid backgrounds, eight of which were previously published (Charron et al., 2019). A
448 MA experiment with yeast consists of subjecting a population to periodic single-cell bottlenecks
449 (typically during hundreds to thousands of mitotic generations) by streaking a randomly chosen
450 single colony, thus amplifying genetic drift and making natural selection inefficient (Lynch et al.,
451 2008) (Figure 6a). *S. paradoxus* was inferred to reproduce mitotically 99.9% of the time in the
452 wild (Tsai et al., 2008), making MA experiments a reasonably realistic approximation of its
453 natural life cycle. Pairs of haploid derivatives of natural *S. paradoxus* and *S. cerevisiae* strains
454 were chosen to span a range of evolutionary divergence (Figure 6b), yielding four categories of
455 intraspecific crosses (from the least to the most divergent): *SpC*×*SpC* (CC), *SpB*×*SpB* (BB),
456 *SpB*×*SpC* (BC) and *SpB*×*SpA* (BA); and one category of interspecific crosses: *SpB*×*S.*
457 *cerevisiae* (BSc; Figure 6c, Tables S4-6). The crosses used also spanned a range of variation in
458 parental Ty CNs (Figure 6d). For each cross, 48 to 96 MA lines derived from independent
459 matings were propagated for ~770 mitotic divisions.

27



460 **Figure 6. Ty CN variation in MA lines reveal no effect of parental divergence and major**

461 **genotype-specific effects. a.** Principle of a MA experiment on a yeast population. N

462 corresponds to population size. **b.** Bayesian phylogenetic tree of the parental haploid strains
463 used to generate artificial hybrids. **c.** Design of the crosses for the MA experiment. **d.** Ty CNs in
464 the haploid parental strains of each cross in c. **e.** CN at the onset (T_{ini}, empty boxes) and at the
465 end (T_{end}, filled boxes) of the MA experiment. Whiskers span 1.5 times the interquartile range.
466 FDR-corrected p-values of Wilcoxon signed-rank tests and Brown–Forsythe tests for the
467 equality of variances are shown in black and blue, respectively. ns: $p \geq 0.05$, *:$p<0.05$, **:$p<0.01$,
468 ***:$p<0.001$. **f.** Linear models explaining the mean variation in CN ($\Delta CN = CN_{Tend}-CN_{Tini}$) for a
469 given cross as a function of the joint CN and genetic divergence of its parental strains. The p-
470 value for each term in the models is shown. Scatter plots (CCPR plots) show the partial
471 residuals for each independent variable, along with the fitted regression.

472 We selected a random subset of 23 to 40 MA lines per cross and performed whole-genome
473 resequencing on stocks frozen at T_{ini} and T_{end}, totaling 372 strains sequenced at each timepoint
474 (PRJNA515073). All these strains were confirmed to be diploid at T_{ini} and T_{end} by quantifying
475 total DNA content using flow cytometry ((Charron et al., 2019), Marsit *et al.* in prep). Focusing
476 on full-length Ty1, Ty3p and Tsu4, no global trend emerged from the variation of NRD before
477 and after MA (Figure 6e, Figure 6—figure supplement 1). Mixed effects linear models fitting
478 cross-specific intercepts and slopes showed no statistically significant support for a global effect
479 of time on CNs. Nevertheless, specific combinations of crosses and Ty families showed
480 significant differences in either central tendency or variance. Often, crosses within a category of
481 parental strain divergence showed contrasting patterns. For example, although CC1, CC2 and
482 CC3 are genetically very similar *SpC*×*SpC* crosses, only CC1 exhibits a significant decrease in
483 Ty1 and Tsu4 CNs. There were also cases of opposite trends within the same category of cross,
484 for instance Ty1 CN in BC and BA crosses. This showed that individual parental genotypes,

485 even if they are evolutionarily closely related, can have very different dynamics, both in sign and
486 magnitude.

487 We next sought to understand how these dynamics are determined by properties of the different
488 hybrid genotypes tested. We fitted linear models to the variation in CN between T_{ini} and T_{end}
489 (ΔCN), using the sum of CNs of both parental strains of a cross and the evolutionary divergence
490 between them as independent variables. Although these models explain little global variance
491 (R^2=0.40, 0.08 and 0.06 for Ty1, Ty3p and Tsu4 respectively), we found that neither independent
492 variable was a significant predictor of ΔCN (Figure 6f). Only Tsu4 showed a slightly positive
493 regression slope with respect to evolutionary divergence, and none of them was statistically
494 significant, indicating that divergence between parents of a hybrid has no large influence on its
495 CN dynamics. On the other hand, Ty1 ΔCN had a negative relationship with parental Ty1 CN
496 and was marginally significant (p=0.056), suggesting that the initial CN of a hybrid genotype can
497 dictate the sign and magnitude of its CN change. However, this result should be interpreted with
498 care as some of the crosses shared common parental strains and were thus not entirely
499 independent. This relatedness may in fact explain part of the pattern seen for Ty1. Crosses that
500 shared the *SpB* parental strain UWOPS-91-202 (BB2, BC2, BA2, BSc2) seemed to have
501 systematically higher Ty1 ΔCN values than the corresponding crosses that shared the *SpB*
502 parental strain MSH-604 (BB1, BC1, BA1, BSc1). UWOPS-91-202 appeared to be a naturally
503 Ty-less strain, as no full-length element was found for any family in its genome (Figure 6d).
504 Previous reports of such Ty-less strains were shown to be permissive to the accumulation of Ty
505 elements when assayed in the laboratory using a reporter assay (Czaja et al., 2020; Garfinkel et
506 al., 2003). This observation also supports the hypothesis that genotype-specific characteristics
507 can be major determinants of variation in Ty CN.

508

Discussion

The literature on TE reactivation in hybrids demonstrates its relevance to the field of speciation (Serrato-Capuchina and Matute, 2018). Reactivation can be a source of postzygotic reproductive isolation by compromising genome stability (Dion-Côté and Barbash, 2017), which translates into a spectrum of deleterious hybrid phenotypes spanning sterility, developmental abnormalities and inviability. Many studies investigated the TE reactivation hypothesis at the interspecific level and yielded equivocal evidence. For instance, mixed evidence for TE reactivation was found within the *Drosophila* genus (Coyne, 1989; Hey, 1988; Labrador et al., 1999; Vela et al., 2014), where reactivation seems to depend on the species pair involved. On the other hand, three well-established hybrid species that emerged from parallel hybridization between the same species pair in the genus *Helianthus* all exhibit important reactivation of LTR retrotransposons, both at the genomic (Ungerer et al., 2009, 2006) and transcriptomic (Renaut et al., 2014) levels. However, a study of hybrid zones between the two parental species revealed that no reactivation is apparent in recent hybrid populations, casting doubt on the idea that hybridization alone triggered reactivation and suggesting the involvement of other factors like abiotic stress (Kawakami et al., 2011). Overall, this data nuances the hypothesis that hybridization could act as a systematic trigger of TE proliferation and reproductive isolation.

In this study, we used population genomic data on North American lineages of the yeast *S. paradoxus* to test whether TEs were reactivated in natural hybrids. It was previously shown that these lineages experienced two hybridization events that resulted in the establishment of admixed lineages that have partial postzygotic reproductive isolation with their respective parental lineages (Eberlein et al., 2019; Leducq et al., 2016). We found no evidence of TE reactivation in any of these natural hybrid lineages. Our data showed that while some hybrid lineages may have relatively high CNs and exhibit evolutionary dynamics consistent with recent

31

533 expansions, these signatures were never more pronounced than that of their parental lineages
534 (*SpC* in the case of *SpC** and *SpC** in the case of SpD_1 and SpD_2). Moreover, our analyses
535 suggest that environmental and demographic factors are unlikely to be major factors affecting
536 TE CNs in natural lineages, which therefore likely reflect the underlying transposition rates. Our
537 observations would seem consistent with a role of TE reactivation in speciation, as natural
538 hybrids with high levels of reactivation would likely be unobserved. However, they contrast with
539 long-established hybrid species in *Helianthus*, showing that reactivation and long-term
540 maintenance can be compatible.

541 We note that the analysis of contemporary genomes is inherently limited in the ability to uncover
542 transient reactivation that would immediately follow hybridization and rapidly faint, perhaps as a
543 result of a re-establishment of TE repression. However, this scenario may be unlikely in yeast,
544 given the absence of many epigenetic and post-transcriptional repression mechanisms (see
545 discussion below) that can yield such a rapid re-establishment in other species. For instance,
546 the re-establishment of piRNA-mediated P element repression in *Drosophila* P-M dysgenic
547 hybrids occurs in single female individuals because paternally-inherited piRNA clusters become
548 expressed as they age (Khurana et al., 2011). If a hypothetical re-establishment of Ty repression
549 was rather driven by selection for genomic variants, this should result in a gradual improvement
550 of repression with time. Our data provides no support for this hypothesis, as younger hybrid
551 lineages SpD_1 and SpD_2 do not show more pronounced signatures of recent Ty expansion
552 compared to the older *SpC** lineage. Additionally, Ty CNs of artificial hybrids at the initial time
553 point of our MA experiment (i.e. early after hybridization) generally show good agreement with
554 the expectation from their parental haploid strains, which is incompatible with an early
555 reactivation of Ty elements.

32

556 Many factors are susceptible to determine whether a hybrid genotype exhibits reactivation or
557 not. According to the Bateson-Dobzhansky-Muller (BDM) model, genetic incompatibilities should
558 accumulate as populations diverge (Orr, 1995). Additionally, the evolutionary conflict between
559 host genomes and TEs results in an arms race that can accelerate the evolution of the involved
560 genes (Blumenstiel et al., 2016; Van Valen, 1973). Thus, it can be expected that
561 incompatibilities affecting transposition rates are more likely to be observed in hybrids between
562 highly diverged genotypes compared to closely related ones. Here, we report data from
563 mutation accumulation (MA) experiments on artificial hybrids that probe the relationship
564 between evolutionary divergence and transposition rates. We show that contrary to the
565 expectation from the BDM model, hybrids between more divergent genotypes exhibit no
566 deregulation of TEs. This conclusion is consistent with recent results from Smukowski Heil and
567 colleagues, who showed that no Ty1 reactivation occurs in experimental interspecific hybrids
568 between *S. cerevisiae* and *S. paradoxus* (Smukowski Heil et al., 2020). Instead of reactivation,
569 many hybrid genotypes exhibited marked decreases in TE CNs after evolution. This trend is
570 especially important for Ty3p, as eight out of ten genotypes exhibit a negative change in median
571 CN. Ty3p is also the least abundant Ty family in natural lineages, as it is found in rarely more
572 than ~5 copies and extinct in all hybrid strains sampled. Such low abundance families are likely
573 more prone to extinction, as stochastic fluctuation in CN would more frequently lead to the loss
574 of all copies. Stochastic loss of Ty copies may come from intra-element LTR recombination or
575 loss of heterozygosity at a locus where a Ty element is present in one parent only, although the
576 mechanistic causes of CN decrease remain to be investigated. This result also seems
577 consistent with the reported Ty CN decrease in mitotically propagated experimental populations
578 of *S. cerevisiae* (Bast et al., 2019), suggesting a potential intrinsic bias towards TE loss.
579 However, it should be noted that the latter study used haploid populations in a regime with
580 efficient natural selection (McDonald et al., 2016), thus mitigating comparisons with our results.

33

581 Transposition is a complex cellular process, with multiple steps regulated by both TEs and host
582 genomes. Genetic screens in *S. cerevisiae* revealed that genes involved in multiple cellular
583 functions (chromatin organization, transcription, translation, RNA and protein metabolism) affect
584 Ty1 transposition rate, either positively or negatively (reviewed in (Curcio et al., 2015)). Thus, it
585 is possible that divergent hybrids suffer from misregulation at different steps of the transposition
586 process that have antagonistic effects, such that they cancel out at the level of transposition
587 rates. It is also possible that such antagonistic interactions exist between transposition and
588 excision rates, yielding no net change in CN. Answering these questions will require a thorough
589 investigation of novel Ty insertion loci in the MA lines, which is beyond the scope of this study.
590 Transposition regulation being a polygenic trait with a large mutational target is also consistent
591 with the strong genotype specificity we observed in our artificial hybrids. Some of the contrasted
592 dynamics between crosses that share high genetic similarity might originate from cryptic genetic
593 variation in genes involved in Ty regulation, whose effect on transposition rate would be
594 uncovered in hybrids.

595 TEs are structurally very diverse and populate the genomes of a vast taxonomic range, and the
596 control mechanisms in response to their mobilization may exhibit a correspondingly large
597 diversity. In many eukaryotes, TEs are repressed epigenetically through mechanisms that
598 include DNA methylation, chromatin modification and mRNA degradation, processes in which
599 small RNAs play central roles (Lisch, 2009; Slotkin and Martienssen, 2007). In many fungi, TE
600 repression acts via multiple mechanisms involving RNA interference (RNAi) or cytosine
601 methylation, including repeat-induced point mutation (RIP) (Gladyshev, 2017). Notably,
602 introgressed species of the fungal genus *Coccidioides* harbor genomic signatures of efficient
603 RIP (Neafsey et al., 2010), suggesting that this mechanism may remain active in the face of
604 hybridization. Unlike other eukaryotes, *S. cerevisiae* and its closely related species lost the

genes of the RNAi pathway (Drinnenberg et al., 2009) and have no cytosine methylation (Proffitt et al., 1984). Instead, *S. cerevisiae* Ty1 has a potent self-repression mechanism, which strength depends on the actual Ty1 CN in the host genome (thus termed CN control, CNC) (Garfinkel et al., 2003). CNC is a post-transcriptional mechanism that relies on p22, an N-truncated form of the Gag (p49) Ty1-encoded capsid protein which is translated from an alternative transcript of Ty1 (Saha et al., 2015). p22 inhibits transposition by interfering with the assembly of the cytoplasmic virus-like particles that are essential to the replication cycle of Ty1 (Saha et al., 2015). This unique repression mechanism may explain some of the patterns we observed in the artificial hybrids. In particular, our data suggests a negative relationship between the initial Ty1 CN in hybrid genotypes and the change in Ty1 CN between the onset and the end of the MA experiment. This pattern seems consistent with an efficient action of CNC, although our results would then suggest that CNC efficiency is largely independent of the degree of evolutionary divergence between the subgenomes of a hybrid. This contrasts with the recent finding that interspecific sequence divergence in p22 has a dramatic impact on the action of CNC (Czaja et al., 2020), although such impact is likely mitigated by the generally much smaller divergence levels in our crosses. Thus, further experimental work is needed to disentangle the mechanistic contributions of Ty1 CN and sequence divergence in the regulation of transposition rates in *Saccharomyces* hybrids.

In conclusion, our work illustrates how the use of natural model systems that are amenable to experimental investigations can bring novel insights on long-standing evolutionary hypotheses. Our results show that natural hybrid lineages of the wild yeast *S. paradoxus* exhibit no TE reactivation and suggest that dynamics of TE content evolution in hybrids depend on deterministic factors that evolutionary divergence between parents can hardly predict.

35

629 ## Materials and Methods

630 *Key Resources Table*

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	Forward primer to construct the deletion cassette for the *ADE2* gene in *SpC*	Charron, Marsit et al., 2019	CLOP97-F5	acaattaaggaatcaagaaaccgtgataaaaaattcaagtCAGCTGAAGCTTCGTACGC
Sequence-based reagent	Reverse primer to construct the deletion cassette for the *ADE2* gene in *SpC*	Charron, Marsit et al., 2019	CLOP97-F6	gtaattgttcgctggccaagtatattaatacatttatataGCATAGGCCACTAGTGGATC
Strain	Haploid parent for MA lines	*Charron et al., 2014*	LL2011_001 *MATa hoΔ::kanMX4*	
Strain	Haploid parent for MA lines	*This study*	LL2011_001 *MATa hoΔ::kanMX4 ade2Δ::hphNT1*	
Strain	Haploid parent for MA lines	*Leducq et al., 2016*	LL2011_012 *MATa hoΔ::kanMX4*	
Strain	Haploid parent for MA lines	*This study*	LL2011_012 *MATa hoΔ::kanMX4 ade2Δ::hphNT1*	
Strain	Haploid parent for MA lines	*Charron et al., 2014*	MSH-587-1 *MATa hoΔ::natMX4*	
Strain	Haploid parent for MA lines	*This study*	MSH-587-1 *MATa hoΔ::natMX4 ade2Δ::hphNT1*	
Strain	Haploid parent for MA lines	Charron, Marsit et al., 2019	LL2011_004 *MATα hoΔ::kanMX4 ade2Δ::hphNT1*	
Strain	Haploid parent for MA lines	Charron, Marsit et al., 2019	LL2011_009 *MATα*	

			hoΔ::natMX4 *ade2Δ::hphNT1*	

General tools

GNU Parallel version 20170922 (Tange, 2011) and SnakeMake version 5.5.4 (Köster and Rahmann, 2012) were used extensively in the following analyses. Unless otherwise stated, Python version 3.7.6 (Van Rossum and Drake, 2009) was used for the custom scripts used in the following analyses. These scripts were run on JupyterLab version 2.0.1 (Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, Jupyter Development Team, 2016) and made extensive use of the following packages: BioPython version 1.76 (Cock et al., 2009), DendroPy version 4.4.0 (Sukumaran and Holder, 2010), Matplotlib version 3.2.0 (Hunter, 2007), NetworkX version 2.4 (Hagberg et al., 2008), NumPy version 1.18.1 (van der Walt et al., 2011), pandas version 1.0.2 (McKinney, 2010), SciPy version 1.4.1 (Virtanen et al., 2020), seaborn version 0.10.0 (Waskom et al., 2020).

Protein-coding genes annotation and orthogroup definition

Whole-genome *de novo* assemblies were previously generated for six North American *S. paradoxus* strains using a combination of long (Oxford Nanopore) and short (Illumina) reads (PRJNA514804). Protein-coding genes were predicted *ab initio* using AUGUSTUS version 3.3.1 with parameters species=saccharomyces genemodel=partial singlestrand=false. Predicted ORF sequences for the six assemblies were combined with ORF annotated from the *S. cerevisiae* strains S288c and YPS128 from (Yue et al., 2017) as references. Protein-coding genes orthogroups were inferred from sequence similarity and synteny using the program SynChro

37

653 version Jan2015 (Drillon et al., 2014) with parameters a=0 DeltaRBH=1. Orthogroups were
654 defined as genes with one-to-one relationships across all pairwise comparisons between the
655 eight genomes. Protein-coding genes systematic names were propagated from S288c to the
656 corresponding orthogroups.

657 *Annotation of Ty LTR retrotransposons*

658 Ty LTR retrotransposons were annotated from the six whole-genome assemblies using
659 RepeatMasker version open-4.0.7 (Smit et al., 2015) with parameters -s -low -gccalc and the
660 library of reference LTR and internal sequences assembled by (Carr et al., 2012) (containing
661 Ty1, Ty2, Ty3, Ty4 and Ty5 from *S. cerevisiae* and Ty3p from *S. paradoxus*), additionned with
662 Tsu4 from *S. uvarum* (Bergman, 2018). RepeatMasker output was defragmented using
663 REannotate version 26.11.2007 (Pereira, 2008) with parameters -s 15000 -d 10000, yielding
664 defragmented elements classified as full-length, solo LTR, truncated or orphan internal
665 sequences. The -f parameter was used to allow defragmentation of Ty1/Ty2 hybrid elements
666 following (Carr et al., 2012). The defragmented annotations were manually curated to split
667 defragmented elements spanning entire protein-coding genes or forming concatemers.

668 *Network-based Ty LTR sequence similarity analysis*

669 We analyzed the sequence diversity of LTR annotations within each of our six *S. paradoxus*
670 assemblies using a network approach (Levy et al., 2017). The sequences of all LTR annotations
671 within a genome were extracted and all-against-all search was performed between LTR
672 sequences in a genome using BLASTN version 2.7.0 (Camacho et al., 2009) with parameters
673 word_size 7 dust no. Networks were built using the Python package Networkx version 2.4
674 (Hagberg et al., 2008) by keeping hits with a bit score higher or equal to the 10th percentile of
675 all bit scores for a given LTR. Visualization of the resulting networks was done using Cytoscape

38

676 version 3.7.1 (Shannon et al., 2003) using the Prefuse Force Directed layout algorithm.
677 Inspection of these networks justified the merging of the following LTR families: Ty1-Ty2, Ty3p-
678 Ty3 and Tsu4-Ty4.

679 *Estimation of evolutionary divergence since transposition*

680 Evolutionary divergence since transposition was estimated from inter-element LTR sequence
681 identity, assuming that the LTRs of a newly inserted element are identical to those of the source
682 element. From the sequence similarity graphs, the closest relative of each element was
683 retrieved by keeping the hit with the highest bit score. Hits between the two LTRs of full-length
684 elements were excluded, and only one LTR per full-length element was considered. For each
685 element, the evolutionary divergence since transposition was extracted as the percent
686 sequence identity to the closest relative.

687 *Definition of Ty orthogroups from genome assemblies*

688 Orthology of LTR retrotransposons in the six whole-genome assemblies was assessed as
689 follows. Whole-genome multiple sequence alignment (MSA) with rearrangements was
690 performed using Mugsy version 1r2.3 (Angiuoli and Salzberg, 2011) with parameters -d 20000
691 -nucmeropts "-b 1000 -l 15" --fullsearch --refine mugsy. MSA blocks for which at least one
692 genome had no corresponding sequence were trashed. The remaining alignment blocks were
693 scanned in 10kb sliding, non-overlapping windows in both directions and windows were trimmed
694 until at least 80% of sites were informative in a MSA window. Ty annotation coordinates were
695 transformed into an MSA coordinates system and clustered using the DBSCAN algorithm
696 implemented in the Python package scikit-learn version 0.22.2 with parameters min_samples=1
697 eps=500. This one-dimensional clustering approach allowed to define orthologous clusters
698 containing at least one Ty annotation in at least one genome.

699 We defined a set of confident orthogroups to derive an empirical distribution of physical distance
700 between start and end coordinates of orthologs. Clusters were selected if they contained at
701 most one annotation per genome and if all annotations were on the same strand and belonged
702 to the same family. Pairwise BLASTN searches within each cluster were used to keep elements
703 that produced at least one hit with ≥80% nucleotide identity and an alignment length between
704 90% and 110% of both query and subject lengths. From the 38 resulting confident orthogroups,
705 we extracted the distribution of differences between annotation start and end coordinates and
706 used the 95th percentile of that distribution (17 bp) as threshold for the following orthogroup
707 definition procedure. Within each cluster, pairs of annotations meeting the following two criteria
708 were considered orthologous: the pair had to include at least one best BLASTN hit, and either
709 start or end coordinates (or both) had to be within the defined threshold of 17 nt. This method
710 allowed to define 713 Ty orthogroups, 20 of which (3%) had >1 element for at least one genome
711 and were discarded.

712 *Phylogenetic analysis of Ty1 evolution*

713 For each of our six genome assemblies, we extracted the LTR sequence of all Ty1 annotations,
714 keeping only one LTR at random for full-length elements. As an outgroup, a conserved Ty4 LTR
715 orthogroup was added on the basis of minimal variance in start and end coordinates across
716 genomes and minimal difference between the length of the reference Ty4 sequence and the
717 average length of the annotations. MSAs were computed using MUSCLE version 3.8.31 (Edgar,
718 2004) and filtered to maximize the proportion of informative sites while keeping the outgroup,
719 first discarding sequences with >25% gaps and then removing sites with ≥25% gaps. Bayesian
720 phylogenetic trees were computed using MrBayes version 3.2.5 (Ronquist et al., 2012) with
721 model parameters Nst=6 Rates=invgamma. Markov chain Monte Carlo parameters were set to
722 10 million generations for two runs of four chains each with 25% burn-in fraction and

723 temperature parameter of 0.03 to improve the efficiency of Metropolis coupling. Bayesian
724 phylogenetic trees were converted into ultrametric phylogenetic trees using PATHd8 (Britton et
725 al., 2007) and rooted at the Ty4 LTR outgroup. The branching time of each internal node in the
726 ultrametric trees was computed as the distance from the root. Branching events were counted
727 as the number of children nodes for each internal node, as the bayesian phylogenetic
728 reconstructions yielded many polytomies at the base and tips of the trees.

729 *Construction of a library of reference Ty sequences for S. paradoxus*

730 We combined the Ty LTR sequences of all families and full-length sequences of Ty1, Ty3p and
731 Tsu4 from the defragmented annotations of our six *S. paradoxus* assemblies and aligned the
732 sequences of each family and type using MUSCLE. The resulting MSAs were filtered to remove
733 sequences and sites that had ≥70% gaps. Consensus sequences from those MSAs were
734 computed using EMBOSS cons version 6.6 (Rice et al., 2000) and the annotation sequence
735 with the lowest BLASTN E value against the resulting consensus was included in the reference
736 library. LTRs of the full-length reference elements were manually trimmed using MEGA version
737 7.0.26 (Kumar et al., 2016). The *S. paradoxus* sequences were complemented with *S.*
738 *cerevisiae* internal sequences for Ty2, Ty3, Ty4 and Ty5 (Carr et al., 2012).

739 *Ty insertion allele calling and population structure analysis*

740 Illumina PE 150 libraries for 207 wild *S. paradoxus* diploid strains were downloaded from NCBI
741 (PRJNA277692, PRJNA324830, PRJNA479851) and trimmed using Trimmomatic version 0.36
742 (Bolger et al., 2014) with parameters ILLUMINACLIP:combined_adapters.fa:2:30:10
743 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 and a custom library of Illumina adapter
744 sequences. Trimmed reads were mapped against the masked version of our *SpB* genome
745 assembly (strain MSH-604, from the RepeatMasker output) using bwa mem version 0.7.16a-

41

746 r1181 (Li, 2013) and sorted using samtools sort version 1.9 (Li, 2011). Duplicates were removed
747 using Picard RemoveDuplicates version 2.18.5-SNAPSHOT (Broad Institute, n.d.) with
748 parameter REMOVE_DUPLICATES=true. RetroSeq version 1.5 (Keane et al., 2013) was used
749 to call Ty insertion alleles from discordant pairs of mapped reads. The -discover command was
750 used with parameter -eref and our custom reference Ty sequences library. The -call command
751 was run with parameter -soft. Visual inspection of the resulting calls revealed several instances
752 of multiple, tightly clustered calls of the same family, likely reflecting a single true positive call. To
753 minimize false positives, the calls were transformed into MSA coordinates and clustered using
754 DBSCAN with parameters min_samples=1 eps=500. Resulting clusters were considered true
755 positives if they contained at least one original call of high-quality (flag 8). This strategy provided
756 the best balance between false positive rates and false negative rates in an analysis made with
757 various simulated short-read datasets (data not shown). Insertion allele frequency spectra were
758 computed within each lineage. Principal component analysis (PCA) was performed on Ty alleles
759 using the Python package scikit-learn version 0.22.2.post1.

760 *Population structure based on genome-wide SNPs*

761 Trimmed reads from the 207 wild strains were mapped against the unmasked version of our
762 *SpB* genome assembly (strain MSH-604) using bwa mem version 0.7.16a-r1181 and sorted
763 using samtools sort. Duplicates were removed using Picard RemoveDuplicates with parameter
764 REMOVE_DUPLICATES=true. Variants were called using freebayes version 1.3.1-17-gaa2ace8
765 (Garrison and Marth, 2012). Variants were filtered using bcftools view version 1.9 (Li, 2011) with
766 parameter -e ‘QUAL<=20’ -g ^het, along with the tools vcfallelicprimitives, vcfbreakmulti and
767 vcffilter with parameter -f ‘TYPE = snp’ from vcflib version 1.0.0_rc2 (Garrison, 2018). PCA on
768 genome-wide SNPs was performed using the Python package scikit-allel version 1.2.1 (Miles et
769 al., 2019), keeping only biallelic SNPs and excluding singletons. From the filtered SNPs of the

42

207 wild strains, 10000 SNPs were randomly selected and converted to NEXUS format using vcf2phylip version 2.0 (Ortiz, 2019). A bayesian phylogenetic tree was computed using MrBayes with model parameters Nst=6 Rates=invgamma. Markov chain Monte Carlo parameters were set to 1 million generations for two runs of four chains each with 25% burn-in fraction and temperature parameter of 0.1.

Depth of coverage of wild strains on Ty reference sequences

Trimmed reads from the 207 wild strains were mapped against our custom reference Ty sequences library using bwa mem version 0.7.16a-r1181 and sorted using samtools sort version 1.9. Only internal sequences extracted from full-length reference sequences were used. Depth of coverage was extracted using samtools depth. Median depth values of Ty elements were transformed in $\log_2$ and normalized by the $\log_2$ genome-wide depth computed from the mapped bases on the genome-wide mappings previously described. Kruskal-Wallis tests were performed on normalized $\log_2$ depth values for families with full-length elements in natural lineages (Ty1, Ty3p, Tsu4). Conover post-hoc tests were performed to test for pairwise differences between lineages using the Python package scikit-posthocs version 0.6.3 (Terpilowski, 2019).

Climatic data

We used the TerraClimate climatic dataset (Abatzoglou et al., 2018) (accessed in October 2019), which incorporates high spatial and temporal resolution climatic interpolation data from primary and derived climatic variables. For the sampling locations of our 207 wild strains, we aggregated the 12 climatic variables over the time period from 1958 to 2015 and computed the mean, standard deviation, minimum and maximum values over time for a given variable variable at a given location. The resulting dataset was normalized into z-scores and summarized by PCA using scikit-learn.

43

Association analysis between Ty CNs, population structure and climatic variables

SNP-based population structure PCAs and climatic variable PCAs were combined in linear models as predictors of CNs for the three active families Ty1, Ty3p and Tsu4. To avoid spurious effects arising from log transformation of very sparse coverage, CN values lower than -1.25 were replaced by -1.25 as for MA lines (see below). In each case, PCs explaining a cumulative sum of 80% of the total variance in the dataset were kept and named gPCs (from PCAs on SNP data) or ePCs (from PCAs on climatic variables). Ordinary least squares models were fitted using the Python package statsmodels version 0.10.1 (Seabold and Perktold, 2010). Two types of models were fitted: OLS using both gPCs and ePCs as predictors, and OLS using only gPCs as predictors. The quality of the two model types was compared using their AIC values. Strains were clustered according to their coordinates on gPCs using the AffinityPropagation algorithm from scikit-learn with parameters max_iter=10000 damping=0.8, and subsequently according to their coordinates on ePCs using the KMeans algorithm from scikit-learn with the same number of clusters fitted for gPCs. The inconsistency of each cluster was computed as the standard deviation of CN for each Ty family.

Nucleotide diversity at synonymous sites

Long-term effective population size for pure lineages was estimated by computing nucleotide diversity at synonymous sites in protein-coding genes. Only SNPs at four-fold degenerate sites in protein-coding genes (third positions of codons CTN, GTN, TCN, CCN, ACN, GCN, CGN and GGN) were considered. Nucleotide diversity was computed using the mean_pairwise_difference function from scikit-allel, excluding SpD_2 strain WX21 since it is known to be heterozygous (Eberlein et al. 2019).

44

816 *Determination of the optimal $NaAsO_2$ concentration for growth assay*

817 Three strains were chosen at random in each of the lineages *SpA*, *SpB* and *SpC** and two *SpC*
818 strains were chosen at random, totalling 12 strains with the *SpC* strain LL2011_004. Precultures
819 were grown in synthetic complete medium (SC; 2% glucose, 0.174% yeast nitrogen base
820 [without amino acids without ammonium sulfate], 0.5% ammonium sulfate, 0.134% complete
821 drop-out) until they reached a 595 nm optical density (OD_{595}) of ~0.5-1.0. Precultures were
822 diluted at an OD_{595} of 0.1 ml^{-1} in SC medium supplemented with 0, 0.4, 0.6, 0.8 and 1 mM
823 $NaAsO_2$ in a single 96-well microplate, with border columns and rows filled with sterile SC
824 medium. OD_{595} was monitored during 60 hours at 15 minutes intervals at 25°C in a Tecan Spark
825 microplate reader (Tecan Group Ltd, Männedorf, Switzerland). OD_{595} values were normalized by
826 subtracting the initial value for each well.

827 *Growth assay in $NaAsO_2$*

828 From the 207 wild strains used for whole genome sequence analyses, 123 strains had glycerol
829 stocks available. Five replicates of each strain were randomly arranged into four arrays of 14x11
830 positions, totalling 615 replicates with one empty position, with the constraint of having at most
831 three replicates of any given strain within an array. For each array, glycerol stocks of the 123
832 strains were spotted on YPD plates (1% yeast extract, 2% tryptone, 2% glucose, 2% agar) and
833 incubated at room temperature for three days. Cells from fresh spots were diluted in 200 µl SC
834 medium. OD_{595} was monitored and 500 µl precultures were initiated at 0.08 OD_{595} ml^{-1} and
835 incubated at room temperature for 10 hours. OD_{595} was monitored and precultures were diluted
836 at 0.2 OD595 ml^{-1} in 1 ml SC medium. BRAND™ BRANDplates™ pureGrade™ S 384-well
837 microplates (BRAND GMBH + CO KG, Wertheim, Germany) were filled with 40 µl of SC
838 medium (even column numbers) or SC medium supplemented with 1.6 mM $NaAsO_2$ (odd
839 column numbers). 40 µl of cultures at 0.2 OD_{595} ml^{-1} were added to the plates using an

840 Eppendorf epMotion 5075 liquid handling platform (Eppendorf, Hamburg, Germany). Replicates
841 were positioned according to the predefined randomized scheme in side-by-side pairs, yielding
842 adjacent replicates diluted at 0.1 OD_{595} ml^{-1} in SC and SC with 0.8 mM $NaAsO_2$. Border rows
843 and columns of the 384-well microplates were filled with 80 µl sterile medium. Microplate lids
844 were treated with 0.05% Triton X-100 in 20% ethanol for 1 minute and air dried (Brewster,
845 2003). OD_{595} was monitored during 70 hours at 15 minutes intervals at 25°C in a Tecan Spark
846 microplate reader.

847 OD_{595} values were normalized by subtracting the initial value for each well, and values in SC
848 were subtracted from values in $NaAsO_2$ for each replicate. The area under the curve (AUC) of
849 the resulting curves was obtained by summing all timepoints. Replicates arising from
850 precultures that failed to grow were excluded, and strains which had less than 3 replicates left
851 were excluded. The standard deviation of replicates was computed for each strain, and strains
852 which fell outside the [-1, 1] z-score range were filtered out to meet the homoscedasticity
853 requirements of analysis of variance (ANOVA). To assess the significance of lineage effects, a
854 mixed effects linear model (with strains as random variable) was fitted in R version 3.4.3 (R
855 Core Team, 2017) using the lmer function from the package lme4 version 1.1-15 (Bates et al.,
856 2015) and an ANOVA table was produced using the package lmerTest version 3.1-2
857 (Kuznetsova et al., 2017). Variance components were estimated using the R package VCA
858 version 1.3.4 (Schützenmeister and Piepho, 2012). Post-hoc Tukey's HSD test was performed
859 using scikit-posthocs.

860 *Properties of genes neighboring Ty insertions*

861 The analysis of properties of neighboring genes considered Ty-centric immediate adjacency
862 relationships with protein-coding genes, meaning that a single gene could be represented more
863 than once if many Ty annotations are immediately adjacent to it. From the whole-genome

MSAs, protein-coding genes adjacent to Ty annotations were retrieved on the basis of synteny with the *S. cerevisiae* S288c genome. Genes annotated as essential in *S. cerevisiae* were extracted from YeastMine (accessed in August 2019). Physical interaction data for *S. cerevisiae* S288c were retrieved from the BioGRID Multi-Validated Physical interaction dataset version 3.5.188 (Oughtred et al., 2019) and the interactome was built using Networkx.

The strength of purifying selection acting on individual genes was assessed at two levels: at the species-wide level (i.e. including all North American *S. paradoxus* lineages) by fitting a single ω ratio (non-synonymous to synonymous substitution rates) per gene; and at the lineage level, by evaluating the relative strength of natural selection in the main pure lineages *SpB* and *SpC*. A randomly selected subset of 34 strains from the 207 wild strains were chosen. The VCF matrix from the SNP calling step above was used to derive the DNA sequences of protein-coding genes in these 34 strains. Neighbor-joining gene trees were produced using the nj function from the R package ape version 5.0 (Paradis and Schliep, 2019) and R version 3.4.3 (R Core Team, 2017). ω ratios were fitted using codeml from the PAML suite version 4.8 (Yang, 2007) with parameters seqtype=1 model=0. Selection strength at the lineage level was evaluated using the RELAX framework from the HyPhy software suite version develop-0d598ad (Wertheim et al., 2015). The RELAX general descriptive model was fitted using the RELAX-scan script (https://github.com/veg/hyphy-analyses/tree/master/RELAX-scan). From the model output, the fitted value of the *k* parameter (describing the strength of selection) was extracted for each branch of the gene trees. For *SpB* and *SpC*, an average *k* value was computed by selecting the clade containing all the strains of the corresponding lineage (from the random subset) and weighting *k* values by the length of the corresponding branch. The difference in selection intensity between *SpB* and *SpC* was defined as $\Delta k = k_{SpB}-k_{SpC}$. For the three genes with ORF-

47

887 disrupting Ty insertions, the RELAX framework was fitted on gene trees comprising all 207 wild
888 strains.

889 *Mutation accumulation*

890 Crosses BB, BC, BA and BSc were previously published (Charron et al., 2019). CC crosses
891 were performed following the same methodology. The *ADE2* gene was deleted in haploid *MATa*
892 *hoΔ::kanMX4* or *MATa hoΔ::natMX4* derivatives of the *SpC* strains LL2011_001, LL2011_012
893 and MSH-587-1 (Charron et al., 2014) using a *hphNT1* cassette that was amplified from plasmid
894 pFA6a-hphNT1 (Janke et al., 2004) using primers CLOP97-F5 and CLOP97-F6 (Tables S3 and
895 S4). The three strains were paired with the previously published *SpC* strains LL2011_004 *MATα*
896 *hoΔ::kanMX4* and LL2011_009 *MATα hoΔ::natMX4* to generate three CC crosses by selection
897 on medium containing 100 µg ml^{-1} geneticin and 10 µg ml^{-1} nourseothricin. For each cross, 64
898 MA lines were initiated by selecting single colonies from independent mating cultures that were
899 spotted on selective medium. Diploid hybrids were propagated mitotically on Petri dishes
900 containing solid YPD medium (1% yeast extract, 2% tryptone, 2% glucose, 2% agar) by
901 streaking for single colonies. Every three days, a single colony was randomly chosen to
902 inoculate a new plate by picking the colony that grew the closest to a predefined mark on the
903 Petri plate. Colonies that had an obvious difference in red pigmentation were excluded to avoid
904 the propagation of mitochondrial petite mutants. MA lines were propagated for 35 passages
905 (approx. 770 mitotic generations), keeping glycerol stocks of all the lines every three passages.

906 *Illumina library preparation for MA lines*

907 For each cross, 23 to 40 MA lines were randomly chosen among the subset for which diploidy
908 was confirmed at T_{ini} and T_{end} (Charron et al., 2019) (Marsit *et al.* in prep.). The glycerol stocks
909 from passage 1 (T_{ini}) and passage 35 (T_{end}) were streaked on solid YPD medium. Single

48

910 colonies were used to inoculate liquid YPD cultures. DNA was extracted from saturated cultures
911 using QIAGEN Blood and Tissue kit with modifications for yeast cell wall digestion. DNA was
912 treated with RNase A and purified on Axygen AxyPrep Mag PCR Clean-up SPRI beads.
913 Genomic DNA libraries were prepared using the Illumina Nextera Tagmentase and Nextera
914 index kits following a modified version of the protocol described in (Baym et al., 2015). The
915 quality of randomly selected libraries was controlled using an Agilent BioAnalyzer 2100
916 electrophoresis system. Sequencing was performed at the CHU Sainte-Justine and Génome
917 Québec Integrated Centre for Pediatric Clinical Genomics (Montreal, Canada) on an Illumina
918 NovaSeq 6000 sequencer with SP flow cells.

919 *Depth of coverage of MA lines on Ty reference sequences*

920 Illumina PE 150 libraries for 372 MA lines and their 13 parental haploid strains (deposited at
921 NCBI Sequence Read Archive under PRJNA515073) were trimmed using Trimmomatic version
922 0.33 with parameters ILLUMINACLIP:nextera.fa:6:20:10 MINLEN:40 and a library of Illumina
923 Nextera adapter sequences. Trimmed reads were mapped against our *SpB* genome assembly
924 (strain MSH-604) using bwa mem version 0.7.17 and sorted using samtools sort version 1.8.
925 Duplicates were removed using Picard RemoveDuplicates version 2.18.29-SNAPSHOT with
926 parameter REMOVE_DUPLICATES=true. We extracted the list of *S. cerevisiae* genes with
927 paralogs from YeastMine (accessed in December 2018) (Balakrishnan et al., 2012) and
928 randomly selected 50 genes that didn't figure in this list. The protein sequences of these genes
929 were used as TBLASTN queries against our *SpB* genome assembly and the best hit loci were
930 extracted. Depth of coverage over these loci was extracted using samtools depth version 1.9
931 and depth values were transformed in $\log_2$. Depth values were transformed in $\log_2$ and
932 normalized by the $\log_2$ genome-wide depth computed from the mapped bases on the genome-
933 wide mappings.

934 The trimmed reads of the 372 MA lines and their 13 parental haploid strains were mapped
935 against our custom reference Ty sequences library using bwa mem version 0.7.16a-r1181 and
936 sorted using samtools sort version 1.9. Only internal sequences extracted from full-length
937 reference sequences were used. Median depth values of Ty elements were transformed in $\log_2$
938 and normalized by the $\log_2$ genome-wide computed from the mapped bases on the genome-
939 wide mappings. Normalized $\log_2$ depth values for a given cross and Ty family were discarded if
940 the median was lower than the 10th percentile of the normalized $\log_2$ depth values for non-
941 paralogous genes in the corresponding cross, thus considering the family absent from the initial
942 MA genotypes. Similarly, the 10th percentile of normalized $\log_2$ depth values for non-paralogous
943 genes in all crosses aggregated (-1.25) was used as a lower threshold for depth values of the
944 parental haploid strains, replacing lower values by -1.25.

945 *Phylogenetic analysis of the MA lines haploid parents*

946 From the mappings of the MA lines parental haploid strains, variants were called using
947 freebayes. Variants were filtered using bcftools view with parameter -e 'QUAL<=20 -g ^het',
948 along with the tools vcfallelicprimitives, vcfbreakmulti and vcffilter with parameter -f 'TYPE = snp'
949 from vcflib. From the filtered SNPs, 10000 SNPs were randomly selected and converted to
950 NEXUS format using vcf2phylip. A bayesian phylogenetic tree was computed using MrBayes
951 with model parameters Nst=6 Rates=invgamma Ploidy=haploid. Markov chain Monte Carlo
952 parameters were set to 1 million generations for two runs of four chains each with 25% burn-in
953 fraction and temperature parameter of 0.1. From the resulting consensus tree, the evolutionary
954 divergence between parental strains was measured as the sum of branch lengths connecting
955 each pair.

956

Association analysis between Ty ΔCN, evolutionary divergence between parents and initial Ty CN

For each cross, the average ΔCN was analysed in a linear model incorporating parental evolutionary divergence and initial Ty CN of the hybrid genotypes as predictors. For each cross, the Ty CNs of parental haploid strains were summed and used as a measure of the initial Ty CN in all MA lines of the cross. An OLS model was fitted using statsmodels.

Acknowledgements

We thank Hélène Martin and Anna Fijarczyk for contributing to sequencing data analyses and Philippe Després, Anne-Marie Dion-Côté, Marika Drouin, Alexandre Dubé, Clara Dubernard and Anna Fijarczyk for reading and providing useful comments on draft versions of this manuscript.

Competing interests

We declare no competing interest.

References

Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC. 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Sci Data 5:170191.

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. 2018. Genetics of trans-regulatory variation in gene expression. Elife 7. doi:10.7554/eLife.35471

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science 309:764–767.

Angiuoli SV, Salzberg SL. 2011. Mugsy: Fast multiple alignment of closely related whole genomes. Bioinformatics 27:334–342.

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM. 2012. YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database 2012:bar062.

51

982 Bast J, Jaron KS, Schuseil D, Roze D, Schwander T. 2019. Asexual reproduction reduces
983 transposable element load in experimental yeast populations. Elife 8.
984 doi:10.7554/eLife.48548

985 Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4.
986 Journal of Statistical Software. doi:10.18637/jss.v067.i01

987 Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. 2015. Inexpensive
988 multiplexed library preparation for megabase-sized genomes. PLoS One 10:1–15.

989 Bergman CM. 2018. Horizontal transfer and proliferation of Tsu4 in Saccharomyces paradoxus.
990 Mob DNA 9:18.

991 Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves
992 of non-LTR insertion since speciation in Drosophila melanogaster. Proc Natl Acad Sci U S A
993 104:11340–11345.

994 Bleykasten-Grosshans C, Friedrich A, Schacherer J. 2013. Genome-wide analysis of
995 intraspecific transposon diversity in yeast. BMC Genomics 14:399.

996 Blumenstiel JP, Erwin AA, Hemmer LW. 2016. What Drives Positive Selection in the Drosophila
997 piRNA Machinery? The Genomic Autoimmunity Hypothesis. Yale J Biol Med 89:499–512.

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

1000 Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z,
1001 Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about
1002 transposable elements. Genome Biol 19:199.

1003 Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An
1004 epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322:1387–
1005 1392.

1006 Brewster JD. 2003. A simple micro-growth assay for enumerating bacteria. J Microbiol Methods
1007 53:77–86.

1008 Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K. 2007. Estimating divergence times in
1009 large phylogenetic trees. Syst Biol 56:741–752.

1010 Broad Institute. n.d. Picard.

1011 Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. 1984. The molecular basis of I-R
1012 hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the
1013 I factor. Cell 38:153–163.

1014 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
1015 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

1016 Carr M, Bensasson D, Bergman CM. 2012. Evolutionary Genomics of Transposable Elements in

1017 Saccharomyces cerevisiae. PLoS One 7:e50978.

1018 Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements.
1019 Genet Res 42:1–27.

1020 Charlesworth B, Langley CH. 1989. The population genetics of Drosophila transposable
1021 elements. Annu Rev Genet 23:251–287.

1022 Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA
1023 in eukaryotes. Nature 371:215–220.

1024 Charron G, Leducq JB, Landry CR. 2014. Chromosomal variation segregates within incipient
1025 species and correlates with reproductive isolation. Mol Ecol 23:4362–4372.

1026 Charron G, Marsit S, Hénault M, Martin H, Landry CR. 2019. Spontaneous whole-genome
1027 duplication restores fertility in interspecific hybrids. Nat Commun 10. doi:10.1038/s41467-
1028 019-12041-8

1029 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff
1030 F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for
1031 computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.

1032 Coney LR, Roeder GS. 1988. Control of yeast gene expression by transposable elements:
1033 maximum expression requires a functional Ty activator sequence and a defective Ty
1034 promoter. Mol Cell Biol 8:4009–4017.

1035 Coyne JA. 1989. Mutation rates in hybrids between sibling species of Drosophila. Heredity 63
1036 (Pt 2):155–162.

1037 Curcio MJ, Lutz S, Lesage P. 2015. The Ty1 LTR-Retrotransposon of Budding Yeast,
1038 Saccharomyces cerevisiae. Microbiology Spectrum 3:MDNA3–0053–2014.

1039 Czaja W, Bensasson D, Ahn HW, Garfinkel DJ, Bergman CM. 2020. Evolution of Ty1 copy
1040 number control in yeast by horizontal transfer and recombination. PLoS Genet
1041 16:e1008632.

1042 Dion-Côté A-M, Barbash DA. 2017. Beyond speciation genes: an overview of genome stability
1043 in evolution and speciation. Curr Opin Genet Dev 47:17–23.

1044 Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic
1045 shock involving transposable elements reactivation in hybrids of young lake whitefish
1046 species. Mol Biol Evol 31:1188–1199.

1047 Dolgin ES, Charlesworth B, Cutter AD. 2008. Population frequencies of transposable elements
1048 in selfing and outcrossing Caenorhabditis nematodes. Genet Res 90:317–329.

1049 Drillon G, Carbone A, Fischer G. 2014. SynChro: A Fast and Easy Tool to Reconstruct and
1050 Visualize Synteny Blocks along Eukaryotic Chromosomes. PLoS One 9:e92621.

1051 Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in

1052 budding yeast. Science 326:544–550.

1053 Eberlein C, Hénault M, Fijarczyk A, Charron G, Bouvier M, Kohn LM, Anderson JB, Landry CR.
1054 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nat
1055 Commun 10:923.

1056 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
1057 throughput. Nucleic Acids Res 32:1792–1797.

1058 Erwin AA, Galdos MA, Wickersheim ML, Harrison CC, Marr KD, Colicchio JM, Blumenstiel JP.
1059 2015. piRNAs Are Associated with Diverse Transgenerational Effects on Gene and
1060 Transposon Expression in a Hybrid Dysgenic Syndrome of D. virilis. PLoS Genet
1061 11:e1005332.

1062 Filteau M, Hamel V, Pouliot MC, Gagnon-Arsenault I, Dube AK, Landry CR. 2015. Evolutionary
1063 rescue by compensatory mutations is constrained by genomic and environmental
1064 backgrounds. Mol Syst Biol 11:832.

1065 Garfinkel DJ, Nyswaner K, Wang J, Cho JY. 2003. Post-transcriptional cosuppression of Ty1
1066 retrotransposition. Genetics 165:83–99.

1067 Garrison E. 2018. vcflib.

1068 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing 1–9.

1069 Gladyshev E. 2017. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms
1070 in Fungi. Microbiol Spectr 5. doi:10.1128/microbiolspec.FUNK-0042-2017

1071 Göbel U, Arce AL, He F, Rico A, Schmitz G, De Meaux J. 2018. Robustness of transposable
1072 element regulation but no genomic shock observed in interspecific Arabidopsis hybrids.
1073 Genome Biol Evol 10:1403–1415.

1074 Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Hoheisel JD, Jacq C,
1075 Johnston M, Louis EJ, Mewes HW, Murakami Y, Tettelin H, Oliver SG. 1996. Life with 6000
1076 Genes. Science 274:563–567.

1077 Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D,
1078 Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled
1079 nutrient-limited environments in yeast. PLoS Genet 4. doi:10.1371/journal.pgen.1000303

1080 Hagberg A, Swart P, S Chult D. 2008. Exploring network structure, dynamics, and function using
1081 NetworkX In: Gäel Varoquaux, Travis Vaught, and Jarrod Millman, editor. Proceedings of
1082 the 7th Python in Science Conference (SciPy2008). Presented at the SciPy2008. pp. 11–
1083 15.

1084 Hénault M, Eberlein C, Charron G, Durand É, Nielly-Thibault L, Martin H, Landry CR. 2017.
1085 Yeast Population Genomics Goes Wild: The Case of Saccharomyces paradoxus. Springer,
1086 Cham. pp. 1–24.

1087 Hey J. 1988. Speciation via hybrid dysgenesis: negative evidence from the Drosophila affinis

1088 subgroup. Genetica 78:97–103.

1089 Hof AEV t., Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri
1090 IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable
1091 element. Nature 534:102–105.

1092 Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. Computing in Science Engineering
1093 9:90–95.

1094 Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A,
1095 Doenges G, Schwob E, Schiebel E, Knop M. 2004. A versatile toolbox for PCR-based
1096 tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution
1097 cassettes. Yeast 21:947–962.

1098 Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during
1099 interspecies hybridization. Curr Biol 16:1322–1328.

1100 Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable
1101 element proliferation and genome expansion are rare in contemporary sunflower hybrid
1102 populations despite widespread transcriptional activity of LTR retrotransposons. Genome
1103 Biol Evol 3:156–167.

1104 Keane TM, Wong K, Adams DJ. 2013. RetroSeq: Transposable element discovery from next-
1105 generation sequencing data. Bioinformatics 29:389–390.

1106 Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD,
1107 Weng Z, Theurkauf WE. 2011. Adaptation to P element transposon invasion in Drosophila
1108 melanogaster. Cell 147:1551–1563.

1109 Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid Dysgenesis in DROSOPHILA
1110 MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male
1111 Recombination. Genetics 86:813–833.

1112 Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable elements and genome
1113 organization: a comprehensive survey of retrotransposons revealed by the complete
1114 Saccharomyces cerevisiae genome sequence. Genome Res 8:464–478.

1115 Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine.
1116 Bioinformatics 28:2520–2522.

1117 Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering AD, Gifford R, Harrach B, Hull
1118 R, Johnson W, Kreuze JF, Lindemann D, Llorens C, Lockhart B, Mayer J, Muller E,
1119 Olszewski NE, Pappu HR, Pooggin MM, Richert-Pöggeler KR, Sabanadzovic S, Sanfaçon
1120 H, Schoelz JE, Seal S, Stavolone L, Stoye JP, Teycheney P-Y, Tristem M, Koonin EV, Kuhn
1121 JH. 2018. Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing
1122 Viruses. J Virol 92. doi:10.1128/JVI.00515-18

1123 Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
1124 Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software. doi:10.18637/jss.v082.i13

Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol 16:931–937.

Langdon QK, Peris D, Baker EP, Opulente DA, Nguyen H-V, Bond U, Gonçalves P, Sampaio JP, Libkind D, Hittinger CT. 2019. Fermentation innovation through complex hybridization of wild and domesticated yeasts. Nat Ecol Evol 3:1576–1586.

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

Le Rouzic A, Payen T, Hua-Van A. 2013. Reconstructing the evolutionary history of transposable elements. Genome Biol Evol 5:77–86.

Lesage P, Todeschini AL. 2005. Happy together: The life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res 110:70–90.

Levy O, Knisbacher BA, Levanon EY, Havlin S. 2017. Integrating networks and comparative genomics reveals retroelement proliferation dynamics in hominid genomes. Science Advances 3:1–13.

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

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

Lisch D. 2009. Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60:43–66.

Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880.

Lockton S, Gaut BS. 2010. The evolution of transposable elements in natural populations of self-fertilizing Arabidopsis thaliana and its outcrossing relative Arabidopsis lyrata. BMC Evol Biol 10:10.

Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. Proc Natl Acad Sci U S A 105:13965–13970.

Lynch M. 2007. The Origins of Genome Architecture. Sinauer Associates.

Lynch M, Conery JS. 2003. The Origins of Genome Complexity. Science 302:1401–1404.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K,

Kulkarni S, Hartl DL, Thomas WK. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A 105:9272–9277.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends in Ecology and Evolution. doi:10.1016/j.tree.2005.02.010

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? Bioessays 38:140–149.

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

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 36:344–355.

McDonald MJ, Rice DP, Desai MM. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature 531:233–236.

McKinney W. 2010. Data Structures for Statistical Computing in Python In: van der Walt S, Millman J, editors. Proceedings of the 9th Python in Science Conference. pp. 56–61.

Miles A, Ralph P, Rae S, Pisupati R. 2019. scikit-allel.

Morillon A, Springer M, Lesage P. 2000. Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol 20:5766–5776.

Mukhopadhyay R, Rosen BP. 1998. Saccharomyces cerevisiae ACR2 gene encodes an arsenate reductase. FEMS Microbiol Lett 168:127–136.

Naumov GI, Naumova ES, Lantto RA, Louis EJ, Korhola M. 1992. Genetic homology between Saccharomyces cerevisiae and its sibling species S. paradoxus and S. bayanus: Electrophoretic karyotypes. Yeast 8:599–612.

Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, Hung C-Y, McMahan C, White J, Sykes S, Heiman D, Young S, Zeng Q, Abouelleil A, Aftuck L, Bessette D, Brown A, FitzGerald M, Lui A, Macdonald JP, Priest M, Orbach MJ, Galgiani JN, Kirkland TN, Cole GT, Birren BW, Henn MR, Taylor JW, Rounsley SD. 2010. Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. Genome Res 20:938–946.

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

Orr HA. 1995. The population genetics of speciation: The evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Ortiz EM. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.

1195 Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung
1196 G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A,
1197 Dolinski K, Tyers M. 2019. The BioGRID interaction database: 2019 update. Nucleic Acids
1198 Res 47:D529–D541.

1199 Palecek SP, Parikh AS, Kron SJ. 2000. Genetic analysis reveals that FLO11 upregulation and
1200 cell polarization independently regulate invasive growth in Saccharomyces cerevisiae.
1201 Genetics 156:1005–1023.

1202 Paquin CE, Williamson VM. 1984. Temperature effects on the rate of Ty transposition. Science
1203 226:53–55.

1204 Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
1205 analyses in R. Bioinformatics 35:526–528.

1206 Parhad SS, Theurkauf WE. 2019. Rapid evolution and conserved function of the piRNA
1207 pathway. Open Biol 9:180181.

1208 Parket A, Inbar O, Kupiec M. 1995. Recombination of Ty elements in yeast can be induced by a
1209 double-strand break. Genetics 140:67–77.

1210 Pereira V. 2008. Automated paleontology of repetitive DNA with REANNOTATE. BMC Genomics
1211 9:1–21.

1212 Picard G. 1976. Non-mendelian female sterility in Drosophila melanogaster: hereditary
1213 transmission of I factor. Genetics 83:107–123.

1214 Proffitt JH, Davie JR, Swinton D, Hattman S. 1984. 5-Methylcytosine is not detectable in
1215 Saccharomyces cerevisiae DNA. Mol Cell Biol 4:985–988.

1216 R Core Team. 2017. R: A Language and Environment for Statistical Computing.

1217 Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid
1218 speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in
1219 hybrid sunflowers. Philos Trans R Soc Lond B Biol Sci 369:20130345–20130345.

1220 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software
1221 Suite. Trends Genet 16:276–277.

1222 Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and
1223 speciation. Heredity 83 (Pt 4):363–372.

1224 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard
1225 MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and
1226 model choice across a large model space. Syst Biol 61:539–542.

1227 Rubin GM, Kidwell MG, Bingham PM. 1982. The molecular basis of P-M hybrid dysgenesis: the
1228 nature of induced mutations. Cell 29:987–994.

1229 Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, Garfinkel DJ. 2015. A trans-

dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. J Virol 89:3922–3938.

Schützenmeister A, Piepho H-P. 2012. Residual analysis of linear mixed models using a simulation approach. Comput Stat Data Anal 56:1405–1416.

Seabold S, Perktold J. 2010. Statsmodels: Econometric and statistical modeling with pythonProceedings of the 9th Python in Science Conference. Scipy. p. 61.

Serrato-Capuchina A, Matute DR. 2018. The role of transposable elements in speciation. Genes 9. doi:10.3390/genes9050254

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

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

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

Smukowski Heil C, Patterson K, Hickey AS-M, Alcantara E, Dunham MJ. 2020. Transposable element mobilization in interspecific yeast hybrids. doi:10.1101/2020.06.16.155218

Stukenbrock EH. 2016. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology 106:104–112.

Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Tange O. 2011. GNU Parallel: the command-line power tool. ;login: The USENIX Magazine 36:42–47.

Terpilowski M. 2019. scikit-posthocs: Pairwise multiple comparison tests in Python. JOSS 4:1169.

Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, Jupyter Development Team. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows In: Fernando Loizides BS, editor. Proceedings of the 20th International Conference on Electronic Publishing. Presented at the 20th International Conference on Electronic Publishing. IOS Press. pp. 87–90.

Todeschini A-L, Morillon A, Springer M, Lesage P. 2005. Severe Adenine Starvation Activates Ty1 Transcription and Retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol 25:7459–7472.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proceedings of the National

Academy of Sciences 105:4957–4962.

Ungerer MC, Strakosh SC, Stimpson KM. 2009. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biol 7:1–13.

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

van der Walt S, Colbert SC, Varoquaux G. 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science Engineering 13:22–30.

Van Rossum G, Drake FL. 2009. Python 3 Reference Manual: (Python Documentation Manual Part 2). CreateSpace Independent Publishing Platform.

Van Valen L. 1973. A new evolutionary law. Evolutionary Theory 1:1–30.

Vela D, Fontdevila A, Vieira C, Pilar García Guerreiro M. 2014. A genome-wide survey of genetic instability by transposition in Drosophila hybrids. PLoS One 9. doi:10.1371/journal.pone.0088992

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0 Contributors. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272.

Waskom M, Botvinnik O, Ostblom J, Lukauskas S, Hobson P, MaozGelbart, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J, de Ruiter J, Pye C, Hoyer S, Vanderplas J, Villalba S, Kunter G, Quintero E, Bachant P, Martin M, Meyer K, Swain C, Miles A, Brunner T, O'Kane D, Yarkoni T, Williams ML, Evans C. 2020. seaborn. doi:10.5281/zenodo.3629446

Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: Detecting relaxed selection in a phylogenetic framework. Mol Biol Evol 32:820–832.

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

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

Wysocki R, Bobrowicz P, Ułaszewski S. 1997. The Saccharomyces cerevisiae ACR3 gene encodes a putative membrane protein involved in arsenite transport. J Biol Chem 272:30061–30066.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS, Lopez J, Harker B, Raborn RT, Thomas WK, Ramsdell J, Pfrender ME, Lynch M. 2017. A New Reference Genome Assembly for the Microcrustacean Daphnia pulex. G3 7:1405–1416.

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

61