

Species-Wide Transposable Element Repertoires Retrace the Evolutionary History of the *Saccharomyces cerevisiae* Host

Claudine Bleykasten-Grosshans,^{*1} Romeo Fabrizio,¹ Anne Friedrich,¹ and Joseph Schacherer^{*,1,2}

¹Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France

²Institut Universitaire de France (IUF), Paris, France

***Corresponding authors:** E-mails: bleykasten@unistra.fr; schacherer@unistra.fr.

Associate editor: Jeffrey P. Townsend

Abstract

Transposable elements (TE) are an important source of genetic variation with a dynamic and content that greatly differ in a wide range of species. The origin of the intraspecific content variation is not always clear and little is known about the precise nature of it. Here, we surveyed the species-wide content of the Ty LTR-retrotransposons in a broad collection of 1,011 *Saccharomyces cerevisiae* natural isolates to understand what can stand behind the variation of the repertoire that is the type and number of Ty elements. We have compiled an exhaustive catalog of all the TE sequence variants present in the *S. cerevisiae* species by identifying a large set of new sequence variants. The characterization of the TE content in each isolate clearly highlighted that each subpopulation exhibits a unique and specific repertoire, retracing the evolutionary history of the species. Most interestingly, we have shown that ancient interspecific hybridization events had a major impact in the birth of new sequence variants and therefore in the shaping of the TE repertoires. We also investigated the transpositional activity of these elements in a large set of natural isolates, and we found a broad variability related to the level of ploidy as well as the genetic background. Overall, our results pointed out that the evolution of the Ty content is deeply impacted by clade-specific events such as introgressions and therefore follows the population structure. In addition, our study lays the foundation for future investigations to better understand the transpositional regulation and more broadly the TE–host interactions.

Key words: Ty elements, *Ty1*, transposon activity, intraspecific variation, introgression, yeast.

Introduction

Transposable elements (TEs) are interspersed repetitive DNA sequences that are able to move from one genomic location to another via a process of intragenomic propagation, called transposition. Several decades of studies, facilitated by whole-genome sequence analyses, have shown that the genome of almost all species is populated by TEs. However, TEs are incredibly diverse, both in terms of structure and mode of transposition, as illustrated by the TE classification (Wicker et al. 2007; Arkhipova 2017; Bourque et al. 2018; Kojima 2019). TEs can be classified either into class I retrotransposons or class II DNA transposons, depending on their mechanism of transposition. In each class, TEs are grouped into orders, depending on the molecular mechanisms involved in transposition, and then divided into superfamilies, according to their structural characteristics. Finally, DNA sequence conservation and phylogenetic data define myriads of TE families, subfamilies, and sequence variants.

Even if TE elements are ubiquitous components, the content in terms of type, number, and prevalence is very variable across the genomes. This variation was observed at different time scales that is between species from the same phylum as

well as between individuals from the same species (Huang et al. 2012; Thomas-Bulle et al. 2018; Fonseca et al. 2019; Gaiero et al. 2019; Ray et al. 2019). In the most recent studies, the comparison of TE contents has allowed a better understanding of the events that shaped populations, such as the domestication of rice (Carpentier et al. 2019), silkworm (Han et al. 2018), and sunflower (Mascagni et al. 2018). Biotic or abiotic stresses have also been shown to impact the TE content of plant and insect species (Quadrana et al. 2016; Stritt et al. 2018; Baduel et al. 2019; Lerat et al. 2019; Rogivue et al. 2019).

The *Saccharomyces cerevisiae* species has a low TE diversity content with only a handful of families, called *Ty1* to *Ty5* (Lesage and Todeschini 2005). All Ty elements belong to the class I and more precisely to the order of LTR retrotransposons, divided into two superfamilies, namely *copia* (for the *Ty1*, *Ty2*, *Ty4*, and *Ty5* elements) and *gypsy* (for the *Ty3* element). The common feature of their coding region is the presence of two genes similar to the *gag* and *pol* genes of retroviruses. As the Ty elements were among the first LTR retrotransposons to be discovered, *S. cerevisiae* emerged as a good model for TE biology (Cameron et al. 1979; Clark et al. 1988; Hansen et al. 1988). The *Ty1* and *Ty3* elements are the most characterized transposons and have since been

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

benchmarks for retroelement studies (Curcio et al. 2015; Sandmeyer et al. 2015). The in-depth knowledge of the Ty biology stands in contrast with the exploration of its species-wide diversity. This aspect remains overlooked although a large number of *S. cerevisiae* isolates have been sequenced in recent years (Strope et al. 2015; Gallone et al. 2016; Zhu et al. 2016; Peter et al. 2018). In the S288C reference genome, the Ty fraction is modest, around 3%, and in addition to the full-length elements, it includes a large fraction of solo-LTR resulting from the loss of the internal coding region by inter-LTR recombination. Ty elements have developed strong insertion preferences believed to generate neutral alleles (Lesage and Todeschini 2005; Bridier-Nahmias et al. 2015; Patterson et al. 2019). Previous studies have already shown the variability of the Ty contents, both in terms of types and copy number (Carr et al. 2012; Bleykasten-Grosshans et al. 2013; Czaja et al. 2020). However, these studies focused on very limited sets of isolates, preventing a global view of the evolution of the TE repertoire.

The collection of 1,011 *S. cerevisiae* natural isolates that were recently sequenced is a valuable sample to study the TE content repertoires within the species (Peter et al. 2018). This population has a wide geographical distribution and their ecological origins are highly diverse. Its evolutionary history has delimited 26 different subpopulations as well as three groups of mosaic isolates characterized by admixture from different lineages. In addition, a few subpopulations are characterized by introgressed regions, which are the signatures of ancient hybridization events with the closest-related species to *S. cerevisiae*, namely *Saccharomyces paradoxus* (Peter et al. 2018; D'Angiolo et al. 2020). These two species diverged from a common ancestor ~ 5 Ma (Tirosh et al. 2009).

Here, we sought to survey the Ty content variation of *S. cerevisiae* at a species level. We first characterized all the Ty sequence variants present in the 1,011 *S. cerevisiae* genomes by identifying new and undescribed Ty sequence variants. We then explored the repertoire (i.e., the type and number of TEs) in each isolate and subpopulation. Our results clearly showed that the population structure could be defined by the variation of the Ty repertoires. In fact, each defined *S. cerevisiae* clade is characterized by its unique and specific Ty repertoire. In addition, our results highlight that introgression events had a significant impact on the appearance of new sequence variants and therefore on the variation of repertoires between different subpopulations. Finally, we also extended our study by a functional analysis of the permissive behavior with respect to the transpositional activity, which revealed a broad variability of it across genetic backgrounds.

Results

Species-Wide Overview of the Full-Length Ty Sequence Variants in the *S. cerevisiae* Species

To explore the Ty diversity within *S. cerevisiae*, we have established an exhaustive catalog of the sequence variants of each Ty family (Ty1 to Ty5) present in the 1,011 genomes with unprecedented high resolution. To this end, the first step was to determine a set of query sequences consisting of the

internal *gag-pol* coding sequences to accurately capture full-length Ty diversity. We used one query for each of the five Ty families of the S288C reference genome (supplementary file S1, Supplementary Material online) to perform a first BLASTn search (supplementary fig. S1, Supplementary Material online and see Materials and Methods). The search was conducted on both *S. cerevisiae* and *S. paradoxus* genome assemblies to obtain the most representative set of query sequences and to further resolve potential interspecies Ty flows. We have sorted hundreds of sequences based on their similarity patterns (see Materials and Methods) and obtained a final set of 12 Ty representative sequences distributed across the five families, allowing to cover *gag-pol* diversity across the whole *S. cerevisiae* species (supplementary file S2, Supplementary Material online).

To detect already identified and even new chimeric TEs, we performed a competitive mapping of the sequencing reads of the 1,011 genomes on the 12 *gag-pol* representative sequences (supplementary fig. S1, Supplementary Material online). This strategy allowed us to distinguish the different Ty sequence variants present in *S. cerevisiae*. We examined the coverage profiles along the 12 sequences for each isolate (supplementary fig. S1, Supplementary Material online and see Materials and Methods) and we defined a precise catalog of Ty sequence variants in *S. cerevisiae*. Out of the 12 queries, 2 of the *S. paradoxus* queries did not reveal significant coverage (Ty3p and Ty56p). Partial or complete coverage of the ten other queries resulted in the definition of a total of 13 sequence variants (fig. 1A). Our results clearly illustrate a varying degree of diversification among the five Ty families of *S. cerevisiae* and most families do not have a large set of sequence variants. The Ty2 and Ty3 families are only represented by a single sequence variant (fig. 1A). The Ty4 family consists of two sequence variants: one specific to *S. cerevisiae* (Ty4c) and another one coming from its sister species *S. paradoxus* (Ty4p). Ty4p is similar to *Tsu4* that likely came via horizontal transfer from *Saccharomyces uvarum* to *S. paradoxus* (Bergman 2018). The Ty5 family is also composed of two sequence variants: a full-length element (Ty5f) is present in a limited number of isolates ($n = 78$) whereas a majority of isolates carry a truncated version with a 1,500 bp internal deletion (Ty5Δ).

By contrast, the Ty1 family has a complex composition due to a high number of sequence variants (fig. 1A). Few of these sequence variants have already been described (Jordan and McDonald 1998; Kim et al. 1998; Bleykasten-Grosshans et al. 2013; Czaja et al. 2020) and our results provide an exhaustive catalog as well as a detailed view of their mosaic structure. The variants can be organized in three different subfamilies according to the origin of the *gag* coding sequence. The divergence of the *gag* gene reaches 8–15% between subfamilies, with the first subfamily (Ty1') having a *gag* sequence specific to the *S. cerevisiae* species whereas the two other subfamilies (Ty1 and Ty101) exhibit distinct *gag* sequences that are similar to *S. paradoxus* elements. In the first subfamily (Ty1'), there is only one sequence variant, which is composed of *gag* and *pol* sequences both specific to the *S. cerevisiae* species. By contrast, the variants present in the second subfamily (Ty1)

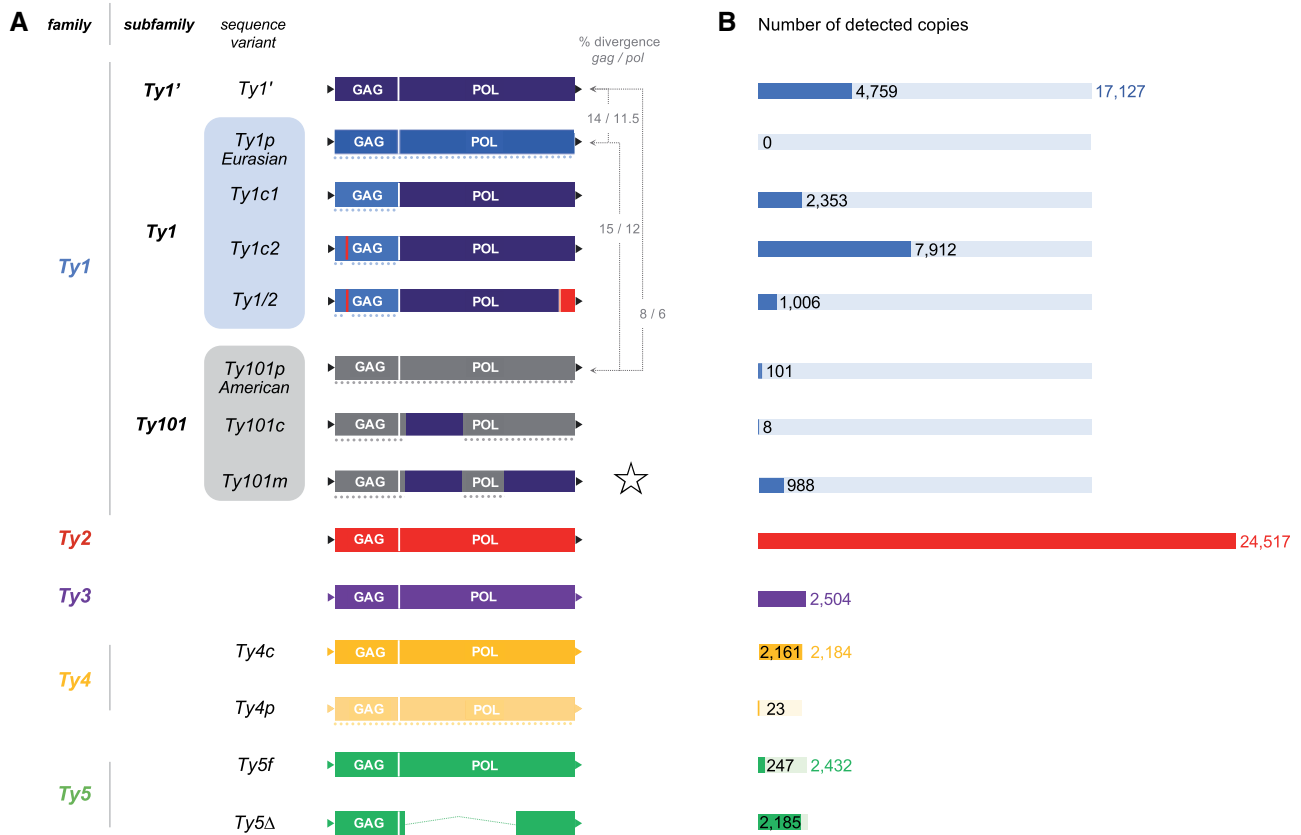


Fig. 1. Mosaic structure and species-wide prevalence of the Ty sequence variants in *S. cerevisiae*. (A) The Ty sequence variants are listed according to their classification in families and subfamilies. The color or shaded color of the *gag* and *pol* coding regions (boxes) and LTRs (triangles) reflect segment identity or high similarity. Dotted lines underline the Ty segments highly similar to *S. paradoxus* Ty elements. The divergence values of *gag* and *pol* sequences correspond to the nucleotide sequence divergence between the *gag* and *pol* sequences of the Ty1', Ty1p, and Ty101p elements. The star highlights the newly described Ty101m variant. The Ty elements are not represented to scale. (B) The size of the bars represents the total number of elements per Ty family (transparent color) and per Ty sequence variant (solid color) detected among the 1,011 genomes. The exact number of copies is indicated for each Ty family (colored numbers) and for each sequence variant, if different from the family copy number (black numbers).

exhibit a *gag* coding segment similar to Ty1p from the Eurasian *S. paradoxus* subpopulation, whereas its *pol* coding segment corresponds to an *S. cerevisiae* sequence (Czaja et al. 2020, fig. 1A and supplementary fig. S3B, Supplementary Material online). The variants of this subfamily differ by the presence (Ty1/2 and Ty1c2) or not (Ty1c1) of small Ty2 segments (fig. 1A). Jordan and McDonald (1998) have already described the Ty2 *pol* segment present in the Ty1/2 element, and here we have identified the Ty1c2 *gag* version, characterized by the presence of a Ty2 segment of 20 bp in the *gag* coding sequence (supplementary fig. S3A, Supplementary Material online). The sequence divergence between Ty1c1 and Ty1c2 reaches 28% in this segment and results in important changes on five out of six consecutive amino acids in the corresponding proteins (supplementary fig. S3C, Supplementary Material online). In the S288C reference genome, 24 out of the annotated Ty1 elements have the Ty1c2 *gag* version whereas the remaining 6 elements have the Ty1c1 *gag* version (supplementary fig. S2B, Supplementary Material online). Finally, in the last and third subfamily (Ty101), we identified an additional interspecies transfer involving the American *S. paradoxus* lineage. This latest subfamily consists

of Ty elements sharing an American *S. paradoxus* *gag* sequence similar to the Ty101p element (fig. 1A). The Ty101c and Ty101m variants differ from Ty101p in that they have a mosaic *pol* sequence involving *S. cerevisiae* Ty1' *pol* segments of different lengths (fig. 1A and supplementary fig. S3B, Supplementary Material online). The Ty101m variant is a newly described element. We extracted the complete sequence of these Ty101m elements present in three isolates for which de novo assemblies were generated using a long-read sequencing strategy (HE015, CBS7962, and RP11.4.11) (Istace et al. 2017) and we were therefore able to confirm the mosaic structure. These new sequence variants are inserted in solo-LTR-rich regions commonly located near tRNA genes, indicating a comparable integration site preference than the other Ty1 sequence variants.

Population-Scale Ty Content Highlights Prevalence of the Ty2 Family

A total of 48,765 full-length Ty elements were detected across the 1,011 genomes. The number of Ty elements per genome is highly variable with an average of about 24 Ty elements per isolate and ranging from none to 100 elements in the PW5

and SJ5L12 isolates, respectively (supplementary file S3, Supplementary Material online). The Ty families were found to be very unequal in terms of the number of elements detected (fig. 1B). As previously shown in a small subset of strains (Kim et al. 1998; Gabriel et al. 2006; Carr et al. 2012; Bleykasten-Grosshans et al. 2013) Ty1 and Ty2 are the most represented families with a total of 17,127 (i.e., 35.1%) and 24,517 elements (i.e., 50.3%), respectively. By contrast, the Ty3, Ty4, and Ty5 families are in considerably lower abundance with approximately 2,000 to 2,500 elements in each family. The distribution among isolates indicates that Ty3 and Ty4 are absent in a large proportion of them (absent in 511 and 593 isolates, respectively) whereas Ty5 is more widely distributed (absent in 209 isolates) (supplementary file S4, Supplementary Material online). Ty3, Ty4, and Ty5 solo-LTR were detected in almost all these isolates (not shown), suggesting that their transpositional activity did not balance Ty loss by inter-LTR recombination.

As mentioned previously, the Ty1 family is characterized by a large number of sequence variants and we observed a high variability regarding their distribution. Two sequence variants from distinct subfamilies are prevalent: the Ty1' and Ty1c2 elements. Although the Ty1' variant, specific to *S. cerevisiae*, is present in a large number of isolates ($n = 887$) with a low copy number, the Ty1c2 variant is the most abundant element but limited to certain isolates ($n = 459$) (supplementary file S4, Supplementary Material online). The other Ty1 sequence variants all show a restricted distribution in a very small subset of isolates, as well as a high degree of variability in terms of copy number. As an example, the Ty101c, Ty101p, Ty1c1, Ty1/2, and Ty101m elements are present in 3–82 isolates with an average number of one (Ty101c) to 20.4 (Ty1c1) elements per genome.

As already mentioned, the Ty2 element appears to be the predominant family in the *S. cerevisiae* species. Only 55 out of the 1,011 strains do not have full-length Ty2 elements, indicating its wide distribution throughout the species. In addition, this element is also prevalent in terms of copy number with 12.7 elements per haploid genome on average (supplementary file S4, Supplementary Material online). It is also present in the most divergent Asian *S. cerevisiae* isolates and, taken together, it can be considered as the ubiquitous element of the species. Regarding the origin of Ty2, it was reported that a very similar sequence variant is present in some *Saccharomyces mikatae* strains but absent in the closely related *S. paradoxus* species. This observation raised the hypothesis of a horizontal transfer of Ty2 between *S. cerevisiae* and *S. mikatae* without addressing the direction of the transfer (Liti et al. 2005). Because Ty2 is widespread and has a homogeneous structure, Ty2 seems to be a recent TE element with a very active transposition in the *S. cerevisiae* species. The Ty2 insertion polymorphism studied in isolates of various origins provides additional support for this hypothesis (Bleykasten-Grosshans et al. 2013). Nevertheless, its prevalence as well as its presence in Asian isolates strongly suggest that the Ty2 family predates the diversification of the species (Bendixsen et al. 2021).

The Ty Repertoires Reveals the *S. cerevisiae* Population Structure

To have a global overview of the evolution of the Ty contents at a species-wide level, we sought to explore the conservation and variation of the Ty repertoires (i.e., the type and number of TEs) across the different subpopulations, which were defined based on the 1,011 sequenced genomes (Peter et al. 2018). By clustering the strains based on the genetic diversity, we found that most of the subpopulations are defined by a unique and specific Ty repertoire (fig. 2). The generated heatmap highlights a variation of the transposon content at two levels: 1) the presence/absence of given Ty sequence variants or elements as well as 2) the variation in terms of copy number across the different subpopulations (fig. 2). In fact, isolates from the same subpopulation exhibit a very similar and specific repertoire, characterized by a precise type and number of TEs. These clear Ty patterns could have only been highlighted via our previous and precise characterization of the Ty elements in the large sample.

Our analysis shows that some Ty sequence variants are only present in specific subpopulations, which is for example the case for the sequence variants of the Ty1 and Ty101 subfamilies (fig. 2). With the exception of the Ty1' and Ty1c2 variants, the other Ty1 sequence variants are restricted to a few subpopulations and we will detail this aspect later. Another example corresponds to the Ty5f variant, which is specific to the subclade 4 of the European wine as well as to the African palm wine subpopulations. Beside the clade-specific TE sequence variants, some subpopulations are also characterized by the absence of pervasive elements. As an example, the Ty2 element is absent in the French Guiana and African beer subpopulations. By contrast, the isolates from the sake clade as well as the M1 and M2 mosaic subpopulations show a very high number of Ty2 copies. We can also notice that the TE repertoires are characterized by specific patterns of different TE families. The most obvious example of the pattern corresponds to the repertoire of the French Guiana isolates. Their Ty content is defined by the presence of Ty101p and Ty101m variants, whereas the other Ty1 variants and Ty families are completely absent. Overall, our observations emphasize the fact that the TE content is primarily shaped at the subpopulation level in *S. cerevisiae*.

Finally, the generated heatmap also highlights an interesting aspect of the evolution of the Ty1 family (fig. 2). As mentioned previously, Ty1' is the *S. cerevisiae*-specific and prevalent element of the Ty1 family. Ty1' segments are almost ubiquitous across the 1,011 isolates as they are present in the mosaic Ty1 variants, such as the variants from the Ty1 and Ty101 subfamilies. Nevertheless, we observed an enrichment of the full-length Ty1' variant in some specific subpopulations and more precisely in the Asian isolates (supplementary fig. S5, Supplementary Material online). This is an interesting observation in respect to the single “out of China” origin of the *S. cerevisiae* species (Peter et al. 2018). The geographic distribution as well as its omnipresence provide a strong support to the fact that Ty1' is the ancestral representative element of the Ty1 family, as recently hypothesized (Czaja et al. 2020).

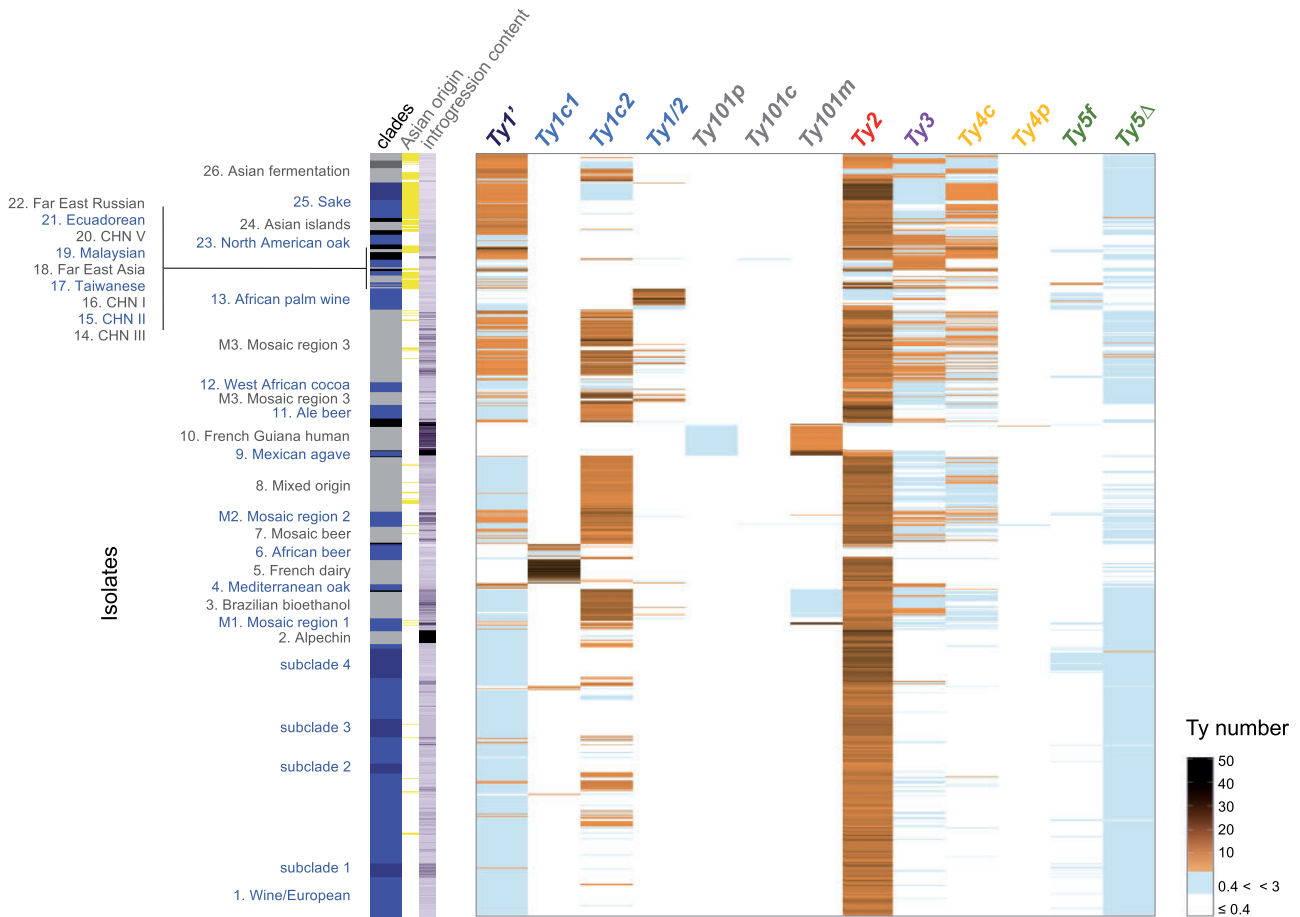


Fig. 2. Correspondence between the Ty contents and the isolate phylogenetic relationships. The main heatmap represents the haploid-equivalent copy number of the Ty sequence variants for each of the 1,011 studied isolates, by using white (no Ty detected), light blue (low Ty copy number <3), and shades of orange color. The related clades are represented (left panel) by alternating of blue and gray colors. Subpopulations are labeled darker and black is associated to strains with no clade. The yellow bars mark the isolates of Asian geographic origin. The violet shaded bars indicate the content in *S. paradoxus* ORFs ranging from 33 (light violet) to 319 (dark violet).

Contrasting Outcomes of Independent Interspecies Hybridization Events on the Ty Contents

Our population genomic study focusing on the 1,011 *S. cerevisiae* genomes highlighted the presence of introgressed ORFs (open reading frames) coming from the closely related species, *S. paradoxus* (Peter et al. 2018). In fact, all studied isolates carry at least one introgressed ORF, with a mean of 32, indicating a ubiquitous gene flow between these yeast species. In this context, it was interesting to find a large number of TE sequence variants of the *Ty1* family, being chimeric elements between *S. cerevisiae* and *S. paradoxus*. As previously described, the *Ty1* diversity was shaped by introgression events with Eurasian and American *S. paradoxus* lineages, leading to the *Ty1* and *Ty101* subfamilies, respectively (fig. 2). These introgression events correspond to secondary contacts with *S. paradoxus* and occurred after the out-of-China dispersal (Peter et al. 2018). They consequently had a major impact on the genome evolution in terms of gene content (introgressed ORFs) but also clearly shaped the Ty repertoire at the subpopulation level (fig. 2). Whereas the *Ty1'* variant is probably the ancestral element of this subfamily, the *Ty1* and *Ty101* variants are the result of Ty flow between these sister species.

Regarding the introgression events involving the American *S. paradoxus* lineage, the Mexican agave and French Guiana subpopulations were found to have a large number of introgressed ORFs ($n = 207$ and $n = 86$ on average, respectively). These two subpopulations were impacted by this major event as they retained copies of the *S. paradoxus* *Ty1* element (*Ty101p*) and contain mosaic elements (*Ty101m*) (fig. 3). Whereas the two subpopulations probably lost the *Ty1'* element, only the Mexican agave subpopulation kept the *Ty2* element, leading to a slightly different repertoire. Finally, we had the opportunity to determine the genomic location of the *Ty101m* elements in three isolates (HE015, CBS7962, and RP11.4.11) for which the genomes were completely assembled using Oxford Nanopore sequencing (Istace et al. 2017). By comparing the location of the introgressed genes and the TEs, it was interesting to observe that there is no overlap between the *Ty101m* elements and this set of genes. This observation provides an evidence of the existence of active *Ty101m* elements.

The introgression events involving the Eurasian *S. paradoxus* lineage led to a contrasting evolution of the Ty repertoires. The Alpechin subpopulation, with the highest number of introgressed ORFs ($n = 287$ on average), exhibits a Ty landscape

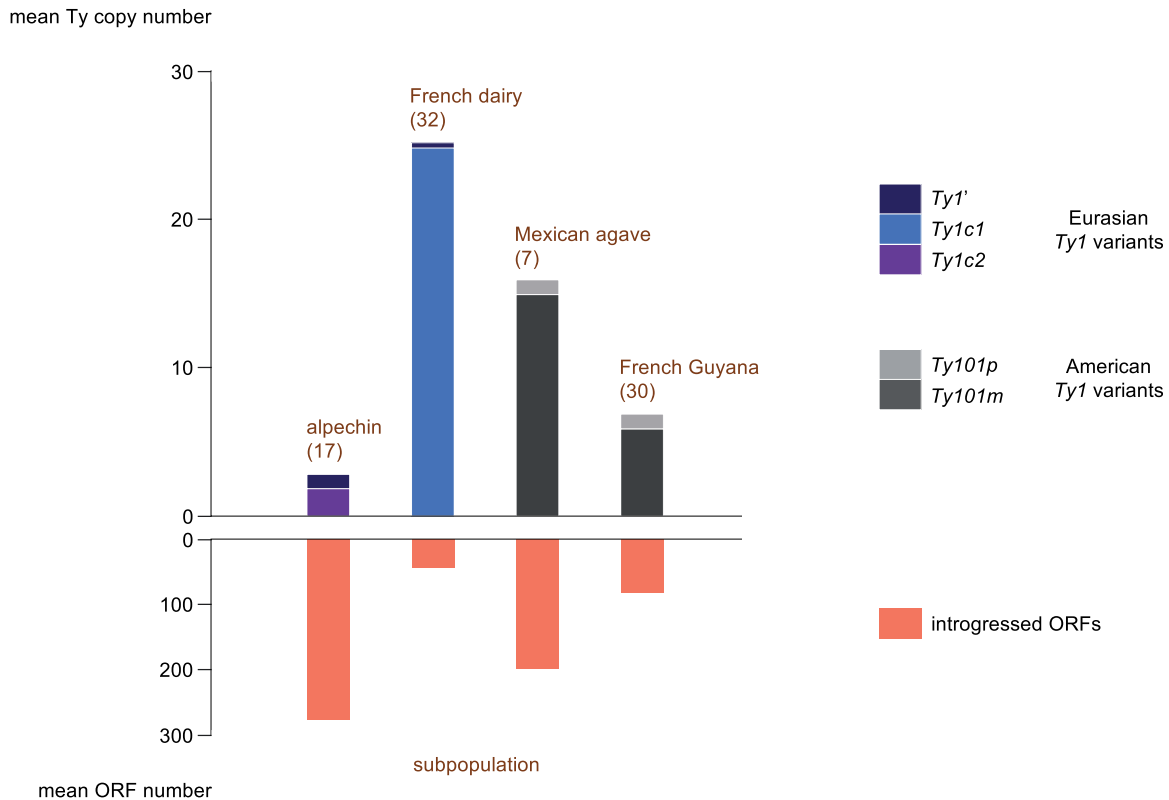


Fig. 3. Outcomes of ancestral hybridization in terms of *Ty1* and introgressed ORF contents in four subpopulations. In the bar plot, the upper bars represent the mean number of the *Ty1* variants. The lower bars represent the number of introgressed *S. paradoxus* ORFs. The number of isolates in the subpopulations is indicated in brackets.

with almost no *S. cerevisiae*/*S. paradoxus* mosaic elements (fig 3). Only 5 out of the 17 isolates carry some *Ty1c2* elements. We then analyzed the *Ty* content of the hybrid ancestor of the Alpechin subpopulation, which was recently described (D'Angiolo et al. 2020). By using our set of *Ty* queries, we detected six *Ty1p* elements in this genome. This observation indicates that the *S. paradoxus Ty1p* is present in this ecological niche and was present in the founder population of the current *S. cerevisiae* Alpechin subpopulation. It is therefore clear that neither this original element nor a mosaic version of it did play a major role in shaping the *Ty* landscape. By contrast, the French dairy clade with a small of introgressed ORFs ($n = 48$ on average) is characterized by the presence of a high number of *Ty1c1* elements. This sequence variant is almost exclusive to this subpopulation and is only found in a few isolates, such as the African beer strains.

Altogether, these results highlight the clade-specific and differential impacts of the introgression events on the *Ty* repertoire. They also reveal the origin of the broad diversity of *Ty1* variants. With the exception of the *Ty1c2* variant, the other sequence variants are mostly private or specific to few subpopulations. This high diversity of *Ty1* variants stands in contrast with the low number of sequence variants present in the other subfamilies.

Variable Transpositional Activity across Genetic Backgrounds

Beside the *Ty* content diversity, we also sought to address transpositional activity variation across this natural

population. In fact, the evaluation of permissive or restrictive transposition behaviors with respect to the *Ty* content could shed light on their mobilization. With this objective in mind, we determined the transpositional activity in a large set of 92 natural isolates as well as 79 stable haploid derivatives (supplementary file S5, Supplementary Material online). To determine the transpositional activity, previous studies used a reporter toolbox based on an element, called *Ty1his3-AI*, that carries an antisense inactive *his3-AI* allele disrupted by an artificial intron (Curcio and Garfinkel 1991). The reverse-transcription of this element based on the transposition process can restore a functional *HIS3* allele after the *AI* intron splicing. This system is commonly used in haploid strains carrying a resident *his3* defective allele (Curcio et al. 2015). To assess the transpositional activity of natural isolates with variable ploidy levels, we constructed and validated a new reporter that allows a direct selection (see Materials and Methods and supplementary file S6, Supplementary Material online). This reporter, called *Ty1hygro-AI*, carries the same *Ty1* variant than the original construction (a *Ty1/2* element) (Lee et al. 1998; Czaja et al. 2020) but the *his3-AI* cassette was replaced by a *hygro-AI* cassette, allowing a direct selection without any genetic manipulation of the natural isolates (see Material and Methods).

We then used this construction to investigate the transpositional activity of 92 natural isolates, which were chosen to cover the broad genetic diversity of the *S. cerevisiae* species. We found that the frequency of the *Ty* activity across the natural isolates is variable, ranging from 0.45 to 273×10^{-6}

(supplementary file S5, Supplementary Material online and fig. 4B). In addition, we also investigated a collection of 79 generated stable haploid strains (Fournier et al. 2019). For the set of haploid strains, we found that the frequency of the Ty activity exhibits a higher variability, ranging from 0.35 to 1870×10^{-6} (supplementary file S5, Supplementary Material online and fig. 4A and B). The distribution of the activity is bimodal with two groups including a set of strains ($n = 20$) that display a strong restrictive behavior toward transposition and another one ($n = 40$) encompassing permissive isolates.

All these strains were also selected to lead to a significant overlap between the sets of haploid and natural isolates. Consequently, we compared the transpositional activity at two ploidy levels (n and $2n$) for a total of 45 genetic backgrounds (supplementary fig. S7, Supplementary Material online). Our results show an overall higher transpositional activity in the haploid strains compared with the diploid isolates (fig. 4A, P -value = 2.1×10^{-5} , Wilcoxon). This is in agreement with previous studies where it was found that *Ty1* transcription is repressed in an *MATa/MAT α* diploid context (Elder et al. 1981; Morillon et al. 2000; Garfinkel et al. 2005). Even if this observation is a general trend, we still observe a variability across the different genetic backgrounds. Interestingly, most of the strongly restrictive haploid strains ($n = 10$) display a lower transpositional activity compared with their corresponding diploid (supplementary fig. S7, Supplementary Material online). This observation therefore highlights that *Ty1* transposition repression in diploids cannot be generalized and is dependent of the genetic background.

Finally, we explored the correlation between the transpositional activity and the Ty contents for both the haploid and the natural isolates. Our results show an absence of correlation between the Ty activity and the number of copies of the main families or subfamilies (*Ty1*, *Ty1'*, and *Ty2*) in both sets of strains (supplementary fig. S8, Supplementary Material online). This observation is somewhat at odds with previous studies showing that the *Ty1* activity is decreased when additional copies of the *Ty1* element are present in the genome, a mechanism termed Copy Number Control (CNC) (Saha et al. 2015; Tucker et al. 2015; Garfinkel et al. 2016; Błaszczyk et al. 2017; Czaja et al. 2020). If this mechanism would have been a predominant mechanism acting on pervasive transposition at a population-scale, we would expect to observe an anti-correlation between the Ty content and activity for such a large sample. CNC has been studied in a very limited number of isolates and this mechanism might be genetic background specific. In addition, we cannot exclude that CNC may be a preponderant response of the host during a burst of transposition but that additional regulation mechanisms might act in isolates already carrying *Ty1* elements as well as various compositions in *Ty1* sequence variants.

Discussion

To investigate the composition of the repetitive genomic fraction represented by the full-length Ty retrotransposons, we developed a strategy to dissect this aspect across the

whole-genome sequences of 1,011 *S. cerevisiae* isolates. We first determined the catalog of Ty sequence variants present in this species with an unprecedented level of precision. Among the five Ty families (*Ty1* to *Ty5*), we have shown that the *Ty1* family is the more diversified one as three subfamilies were identified, each including one to four sequence variants. Our analyses therefore broaden our knowledge of the diversity of *Ty1* sequences, which will constitute a useful reference in the in-depth understanding of the control of TE transposition, a subject of great importance in the understanding of the TE–host relationship. Additionally, the combination of these potential sequence variant effects with the yet to be identified background host effects could explain why we found such a wide range of transposition activity across the large sample of isolates we tested.

Our results highlighted that the Ty landscape retraces the population structure with each subpopulation having a unique and specific TE repertoire. The distribution pattern of the Ty families and subfamilies in the entire species makes it possible to distinguish between the widespread TEs which originate from the ancestral Ty repertoire of *S. cerevisiae* and the TEs which appeared more recently during the diversification of the species. Whereas the *Ty1'* subfamily most likely arises from the ancestral *Ty1* representative, the *Ty1* and *Ty101* subfamilies are the results of subpopulation events, namely hybridizations with distinct lineages of the *S. paradoxus* species.

The diversification of the panoply of Ty elements in *S. cerevisiae* comes from the generation of mosaic elements that combine segments originating from several species-specific TEs. The process that can lead to such mosaic elements is known as interelement recombination and it was already proposed to explain the existence of the *Ty1/2* hybrids (Jordan and McDonald 1998). Although ectopic homologous recombination between Ty copies can occur (Roeder and Fink 1980; Rachidi et al. 1999; Hou et al. 2014), these mosaic elements can also be the result of the reverse-transcriptase jumping between distinct RNA templates that are encapsidated in the same virus-like particle, like it was described for recombinant retroviruses (Hu and Temin 1990; Negroni et al. 1995; Anderson et al. 1998). The latest mechanism was already shown to be highly efficient in the generation of *Ty1* sequence diversity in the S288C reference strain (Bleykasten-Grosshans et al. 2011).

In fact, *Ty1* is the only family presenting mosaic *S. cerevisiae/S. paradoxus* variants. By contrast, the most preponderant element in *S. cerevisiae*, *Ty2*, is not present in *S. paradoxus* and therefore does not present such a mode of diversification. The apparent homogeneity of the *Ty2* structure across the population calls for additional analyses to assess the origin and age of this element. It would eventually address the already long-term issue of the direction of the transfer between *S. cerevisiae* and *S. mikatae* (Liti et al. 2005; Carr et al. 2012). To perform analyses that discriminate between vertical and horizontal transposon transfer like the VHICA method for example (Wallau et al. 2016), additional data on the *Ty2* sequences of *S. mikatae* would be necessary.

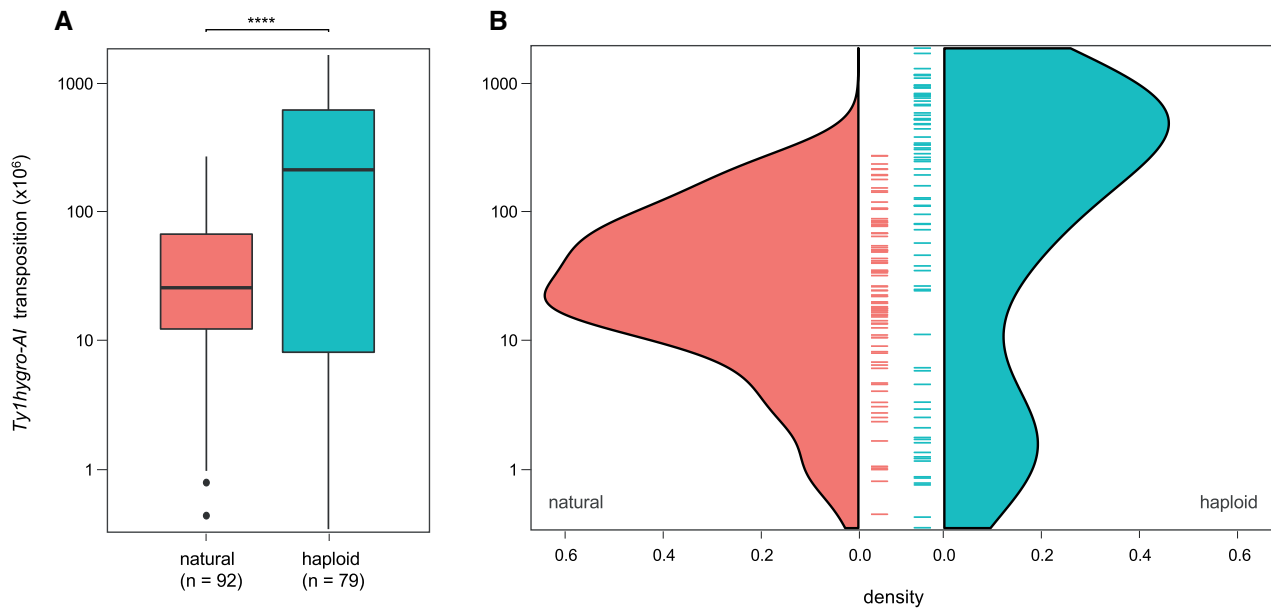


Fig. 4. Variation of the transpositional activity of a set of natural isolates and haploid strains. For each strain, the value of the transpositional activity is the mean of three independent replicates of the ratio between the hygromycin-resistant colonies and the total cells (see Materials and Methods). Panel (A) compares the boxplot representations of the distributions for the two sets of strains. Panel (B) shows the distribution of the values (central panel) and the corresponding probability density functions.

Overall, it is of particular interest to see the rise of new Ty repertoires generated by the successful interspecific hybridization of distant genomes and associated Ty elements. By contrast to the case of hybrid dysgenesis syndrome in some *Drosophila* species (Mérel et al. 2020), it is not possible to predict if these TEs will lead or not to a genomic shock (McClintock 1984) and detrimental TE proliferation that can ultimately result in reproductive isolation. Furthermore, there is growing evidence that hybridization events and their associated TE contents are not obligatory subjected to massive TE deregulation. Such outcomes were described in yeast species (Hénault et al. 2020; Smukowski Heil et al. 2021), as well as in other groups including plants (Kawakami et al. 2011; Heyduk et al. 2021) and insects (Coyne 1989; Vela et al. 2014). Our results show that these hybridization events can lead to different and unpredictable evolutionary trajectories as well as Ty repertoires. Interestingly, hybridization events can lead to new TE sequence variants and this reveals some possible long-term impact of natural hybridization on the TE landscape. By analogy with the transgressive phenotypic traits that arise through hybridization (Gabaldón 2020a, 2020b), these new TE repertoires can therefore being shaped during the so-called reticulated evolution. Exploration of species that result from hybridization events, exhibit admixed populations or contain recursive hybrids (Pryszcz et al. 2015) will therefore be insightful to dissect the formation, the evolution, and the impact of these newly emerging TE repertoires.

Materials and Methods

Yeast Isolates

The 1,011 isolates investigated for Ty content were described in Peter et al. (2018) and are provided in supplementary file

S3, Supplementary Material online. We used the raw Illumina reads generated in this project to perform our analyses. Reads are 102-bp long and the coverage information is in supplementary file S3, Supplementary Material online. The strains and isolates used for the transposition assays are listed in supplementary file S5, Supplementary Material online.

Set-up of the *gag-pol* Query Collection to Detect Full-Length Ty Elements

For a single representative of each of the *Ty1* to *Ty4* families from the *S. cerevisiae* S288C reference genome, we used the internal *gag-pol* segment (without the 5' and 3' LTR segments), as query sequence (supplementary file S1, Supplementary Material online). For the *Ty5* family, as the single *Ty5* copy in S288C carries a 1,500-bp long internal deletion, we used the *gag-pol* segment sequence of the full-length *Ty5* from the YJM178 strain (Strope et al. 2015). In order to take into account of the maximum of variants already present in the sequenced and assembled *S. cerevisiae* and *S. paradoxus* isolates, BLASTn searches were conducted on the NCBI nonredundant nucleotide database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; last accessed July 6, 2018), using the blastn option with default parameters. Target sequences with complete coverage along each query were sampled and aligned (Clustal Omega, <https://www.ebi.ac.uk/Tools/msa/clustalo/>; last accessed June 14, 2021), exact duplicates were excluded. Among sequences sharing more than 95% identity, a single representative was retained. Additional manual inspections of the alignments were performed, leaving a final set of 12 sequences that we considered representative of the Ty diversity in both *S. cerevisiae* and *S. paradoxus* species (supplementary file S2, Supplementary Material online). This

set of *gag-pol* sequences was used as a reference to detect potentially full-length Ty elements in the 1,011 genomes.

Detection of Sequence Reads Mapping on Ty Segments

Read mapping was performed with *bwa* (Li and Durbin 2009) against a reference consisting in the above-described set of 12 *gag-pol* segment sequences. *Samtools mpileup* (Li et al. 2009) was used to generate pileup files that were processed to estimate the read depth along the reference sequences (supplementary file S2, Supplementary Material online) using 10 bp nonoverlapping windows.

Assignment to Ty Families or Ty Sequence Variants and Estimation of the Corresponding Copy Number

Coverage profiles were obtained by plotting the reads depth along the corresponding *gag-pol* Ty query (examples in supplementary fig. S2, Supplementary Material online). The coverage plots were manually inspected to precisely define the structure of the mosaic Ty sequence variants (fig. 1 and supplementary fig. S3, Supplementary Material online). This allows us to 1) delimit the appropriate segments for accurately assigning the detected elements to a Ty category (family, subfamily, or sequence variant) and 2) set up the most appropriate computing scheme to assess its copy number. The corresponding Ty copy number per haploid-equivalent genome was estimated by determining the ratio between the average read depth over the defined region and the total depth of the nuclear genome (supplementary file S3, Supplementary Material online). Determination of the number of copies per haploid-equivalent genome allows further comparisons between isolates of varying ploidy levels (supplementary file S3, Supplementary Material online). Finally, the total number of elements by type and isolate was obtained by multiplying the number of copies by the ploidy of the isolates.

A summary of the defined regions used to assess the copy number is provided hereafter.

- *Ty1'*: region from 100 to 1,000 of the *Ty1'* query that corresponds to the *gag* segment specific to this subfamily;
- *Ty1c2*: region from 100 to 1,000 of the *Ty1c2* query that corresponds to the *gag* segment specific to this subfamily;
- *Ty2*: region from 100 to 4,500 of the *Ty2* query;
- *Ty3c/p* and *Ty4c/p*: the whole respective query;
- *Ty5f*: region from 1,300 to 3,400 of the *Ty5* query. *Ty5Δ* was obtained by the difference of coverage between the 3,450–4,800 and the 1,300–3,400 segments on the *Ty5* query.

The *Ty1/2* elements deserved specific attention. Indeed, they were characterized by an excess of coverage at the 4,900–5,100 region of the *Ty2* query, balanced by a drop in the coverage of the 4,900–5,300 region of the *Ty1c2* query (supplementary fig. S2A, Supplementary Material online, strain S288C). When *Ty1/2* is the sole representative of the *Ty1* subfamily (i.e., the coverage on the *Ty1c2* 4,900–5,300 segment resulted in a value less than 0.5), we inferred the

value computed on the 100–1,100 segment of the *Ty1c2* query as the actual *Ty1/2* copy number. When *Ty1c2* (and/or *Ty1c1*) and *Ty1/2* are altogether present, the *Ty1/2* copy number was assessed by the difference of coverage between the *Ty2* 4,900–5,100 and 100–4,500 segments. The corresponding value was then subtracted to the value computed on the 100–1,100 segment of the *Ty1c2* query, in order to adjust the *Ty1c2* copy number.

Among the *Ty1* family, the following sequence variants were carefully considered:

- *Ty1c1* differs from *Ty1c2* by a very short segment in the *gag* region. *Ty1c1* was revealed by an excess of coverage of this short segment on the *Ty1p* query visualized during manual inspection of the coverage profiles (supplementary fig. S2A, Supplementary Material online, strain S288C). We could only assess its copy number when *Ty1c1* is the only representative of the *Ty1* subfamily within a strain. Its copy number was inferred from the coverage of the *gag* segment (100–1,100 segment) of the *Ty1c2* query. Otherwise, they are in a minority and were identified by default as *Ty1c2* elements;
- *Ty101m* variant was revealed by balanced coverage and absence of coverage of adjacent segments of the *Ty101c*, *Ty101p*, and *Ty1'* reference sequences (supplementary fig. S2A, Supplementary Material online, strain HE015). Its copy number was computed from the coverage along the *Ty101c* 500–1,000 segment;
- *Ty101c* and *Ty101p* variants were detected during manual inspection of the coverage plots and their copy number was computed from the coverage along the 500–1,000 segment of the corresponding query.

LTR Detection

We aimed to detect the presence/absence of the Ty LTR, in order to identify isolates where solo-LTRs are the sole representatives of a given a Ty family. We performed the read mapping against a set of queries including only LTR segments: three LTR sequences that are representative of the very similar *Ty1* and *Ty2* LTRs and a single LTR sequence for each of the *Ty3* to *Ty5* families (supplementary file S7, Supplementary Material online).

Yeast Methods

Yeast cells were grown on YPD (yeast extract 1%, peptone 2%, dextrose 2%) at 30 °C. When required, the temperature was adjusted to 34 °C or to 22 °C that are respectively restrictive or permissive to *Ty1* transposition (Paquin and Williamson 1986; Lawler et al. 2002). Geneticin (Euromedex) to a final concentration of 200 µg/ml was added to obtain the selective YPD-G418 media. The solid selective media YPD-hygro was supplemented with hygromycin (Euromedex) to a final concentration of 200 µg/ml and with 2% agar (Bacto agar Difco). The composition of the nitrogen-depleted media was as follows: 0.67% Yeast Nitrogen Base without amino acids and ammonium sulfate (MP Biomedicals), 2% D-glucose, 0.05 mM ammonium sulfate. Yeast strains were transformed

by the pCeTyX plasmid carrying *Ty1hygro-AI* with the EZ-Yeast Transformation Kit (MP Biomedicals).

Ty1hygro-AI Plasmid Construction and Validation

The *Ty1hygro-AI* carrying plasmid called pCeTyX was constructed by assembling four overlapping DNA fragments thanks to the Gibson Assembly Cloning Kit (NEB). The 5,472 bp plasmid backbone was obtained after the restriction of p41Neo 1-F GW (<https://www.addgene.org/58545/>; last accessed June 14, 2021) by *XhoI* and *XbaI*. We designed and purchased the *hygro-AI* cassette at Genescript and further amplified it by PCR using Iproof high fidelity DNA polymerase (BioRad). The 5'-LTR-*gag-pol* and the 3'-LTR segments were obtained by PCR amplification from the pOY1 template (Lee et al. 1998). The sequence of the final construction was verified by Sanger sequencing (Eurofins). Details on pCeTyX cloning and primer sequences are available in [supplementary file S6, Supplementary Material](#) online.

We first validated the new *Ty1hygro-AI* reporter by testing its activity in different conditions in the S288C genetic background. The results obtained with the *Ty1hygro-AI* reporter recapitulates the available results of *Ty1his3-AI* reporters, as illustrated in [supplementary fig. S6A, Supplementary Material](#) online. Transposition was observed to be higher in the haploid state (FY5) compared with the diploid state (FY3) (Morillon et al. 2000), optimally at 22 °C. Furthermore, the transposition activity was undetectable at 34 °C (Lawler et al. 2002). We also have recorded the transpositional activity of the *Ty1hygro-AI* reporter in the CLQCA_20-259 background ([supplementary file S5, Supplementary Material](#) online), where no copia elements were detected. Importantly, this shows that the *Ty1hygro-AI* reporter is autonomous.

Ty1hygro-AI Mobility Assay

To reliably assess the *Ty1hygro-AI* transpositional activity in a large number of strains, we designed the following protocol. This protocol was carried out on three independent cultures for each of the tested strains. Cells of each of the cultures were transformed with pCeTyX, and bulk G418 resistant transformants were selected in 1.8 ml YPD-G418, at 34 °C (restrictive temperature for *Ty1* transposition) during 48 h. Subsequently, 10 µl of culture were transferred to 150 µl fresh YPD-G418 and grown to saturation for 36 h at 34 °C. To allow for *Ty1* transposition, 100 µl of the saturated culture were inoculated in 1.8 ml YPD and incubated at the permissive temperature of 22 °C for 24 h. A 50 µl aliquot of the induced cultures was saved for further cell counting. Cell counting was performed by counting of the events with $FCS \geq 1.10^4$ and $SSC \geq 5.10^5$ by flow cytometry (BD-Accuri-C6). Variable fractions ranging from 0.1% to 95% of the cells were harvested, plated on solid YPD-hygro, and incubated for 48–72 h at 34 °C. Hygromycin-resistant colonies were counted. The transposition activity was represented as the ratio of the hygromycin-resistant colonies to the total cells. Depending on the efficiency of the transformation step, this protocol ensures that at least three independent transformants were assessed for each of the strains. In parallel, to control for the absence of preexisting transposition events that might have

occurred prior to the induction step at 22 °C, 20 µl of the aforementioned saturated cultures at 34 °C were deposited on solid YPD-hygro and incubated during 48 h at 34 °C; if hygromycin-resistant colonies grew on this control, the corresponding assay was discarded.

Some genetic backgrounds display very low ratio of hygromycin-resistant colonies in the standard assay conditions. In order to rule out issues with the assay, two additional controls were performed. In the first control, we subjected the corresponding strains to conditions of nitrogen starvation during the step of *Ty1* induction. Nitrogen starvation is known to stimulate the *Ty1* transposition ([supplementary fig. S6A, Supplementary Material](#) online and Morillon et al. 2000) and all the tested strains show substantial increases of the ratio in hygromycin-resistant colonies in this condition (examples in [supplementary fig. S6B, Supplementary Material](#) online). This indicates that the *Ty1hygro-AI* reporter is actually functional even for the strains that display low transpositional activity very close to the detection threshold in the usual assay conditions. In the second control, we checked if pCeTyX may be lost during the step of *Ty1* induction. This control was performed in several strains with variable transpositional activity. The results indicate that at least 40% of the cells retained pCeTyX after 24 h of growth at 22 °C in non-selective media. The tested strains display variable rates of pCeTyX plasmid loss; however, the strains with the lowest activity have the highest rates of plasmid retention and *vice-versa* ([supplementary fig. S6C, Supplementary Material](#) online).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank Joan Curcio for the pOY1 plasmid. They thank Casey Bergman for critical reading of the manuscript. They also thank members of the Schacherer lab at the University of Strasbourg for helpful suggestions throughout the project. This work was supported by the European Research Council (ERC Consolidator Grant 772505). J.S. is a Fellow of the University of Strasbourg Institute for Advanced Study (USIAS) and a member of the Institut Universitaire de France.

References

- Anderson JA, Teufel RJ, Yin PD, Hu W-S. 1998. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol.* 72(2):1186–1194.
- Arhipova IR. 2017. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mob DNA.* 8(1):19.
- Baduel P, Quadrana L, Hunter B, Bombles K, Colot V. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun.* 10(1):5818.

- Bendixsen DP, Gettle N, Gilchrist C, Zhang Z, Stelkens R. 2021. Genomic evidence of an ancient East Asian divergence event in wild *Saccharomyces cerevisiae*. *Genome Biol Evol.* 13(2):evab001.
- Bergman CM. 2018. Horizontal transfer and proliferation of Tsu4 in *Saccharomyces paradoxus*. *Mob DNA.* 9:18.
- Błaszczyk L, Biesiada M, Saha A, Garfinkel DJ, Purzycka KJ. 2017. Structure of Ty1 internally initiated RNA influences restriction factor expression. *Viruses* 9(4):74.
- Bleykasten-Grosshans C, Friedrich A, Schacherer J. 2013. Genome-wide analysis of intraspecific transposon diversity in yeast. *BMC Genomics* 14(1):399.
- Bleykasten-Grosshans C, Jung PP, Fritsch ES, Potier S, de Montigny J, Souciet J-L. 2011. The Ty1 LTR-retrotransposon population in *Saccharomyces cerevisiae* genome: dynamics and sequence variations during mobility. *FEMS Yeast Res.* 11(4):334–344.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol.* 19(1):199.
- Bridier-Nahmias A, Tchalikian-Cosson A, Baller JA, Menouni R, Fayol H, Flores A, Saïb A, Werner M, Voytas DF, Lesage P. 2015. Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. *Science* 348(6234):585–588.
- Cameron JR, Loh EY, Davis RW. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* 16(4):739–751.
- Carpentier M-C, Manfroi E, Wei F-J, Wu H-P, Lasserre E, Llauro C, Debladis E, Akakpo R, Hsing Y-I, Panaud O. 2019. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat Commun.* 10(1):24.
- Carr M, Bensasson D, Bergman CM. 2012. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLoS One.* 7(11):e50978.
- Clark DJ, Bilanchone VW, Haywood LJ, Dildine SL, Sandmeyer SB. 1988. A yeast sigma composite element, TY3, has properties of a retrotransposon. *J Biol Chem.* 263(3):1413–1423.
- Coyne JA. 1989. Mutation rates in hybrids between sibling species of *Drosophila*. *Heredity* 63(2):155–162.
- Curcio MJ, Garfinkel DJ. 1991. Single-step selection for Ty1 element retrotransposition. *Proc Natl Acad Sci U S A.* 88(3):936–940.
- Curcio MJ, Lutz S, Lesage P. 2015. The Ty1 LTR-retrotransposon of budding yeast. *Microbiol Spectr.* 3(2):1–35.
- Czaja W, Bensasson D, Ahn HW, Garfinkel DJ, Bergman CM. 2020. Evolution of Ty1 copy number control in yeast by horizontal transfer and recombination. *PLoS Genet.* 16(2):e1008632.
- D'Angiolo M, Chiara MD, Yue J-X, Irizar A, Stenberg S, Persson K, Llored A, Barré B, Schacherer J, Marangoni R, et al. 2020. A yeast living ancestor reveals the origin of genomic introgressions. *Nature* 587(7834):420–425.
- Elder RT, John TPS, Stinchcomb DT, Davis RW. 1981. Studies on the transposable element Ty1 of yeast I. RNA homologous to Ty1. *Cold Spring Harb Symp Quant Biol.* 45:581–591.
- Fonseca PM, Moura RD, Wallau GL, Loreto ELS. 2019. The mobilome of *Drosophila incompacta*, a flower-breeding species: comparison of transposable element landscapes among generalist and specialist flies. *Chromosome Res.* 27(3):203–219.
- Fournier T, Abou Saada O, Hou J, Peter J, Caudal E, Schacherer J. 2019. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. *Elife* 8:e49258.
- Gabaldón T. 2020a. Patterns and impacts of nonvertical evolution in eukaryotes: a paradigm shift. *Ann N Y Acad Sci.* 1476(1):78–92.
- Gabaldón T. 2020b. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res.* 20(5):foaa040.
- Gabriel A, Daprich J, Kunkel M, Gresham D, Pratt SC, Dunham MJ. 2006. Global mapping of transposon location. *PLoS Genet.* 2(12):e212.
- Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR. 2019. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. *Ann Bot.* 123(3):521–532.
- Gallone B, Steensels J, Prah T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, et al. 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166(6):1397–1410.e16.
- Garfinkel DJ, Nyswaner KM, Stefanisko KM, Chang C, Moore SP. 2005. Ty1 copy number dynamics in *Saccharomyces*. *Genetics* 169(4):1845–1857.
- Garfinkel DJ, Tucker JM, Saha A, Nishida Y, Pachulski-Wieczorek K, Błaszczyk L, Purzycka KJ. 2016. A self-encoded capsid derivative restricts Ty1 retrotransposition in *Saccharomyces*. *Curr Genet.* 62(2):321–329.
- Han M-J, Xu H-E, Xiong X-M, Zhang H-H. 2018. Evolutionary dynamics of transposable elements during silkworm domestication. *Genes Genom.* 40(10):1041–1051.
- Hansen LJ, Chalker DL, Sandmeyer SB. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol Cell Biol.* 8(12):5245–5256.
- Hénault M, Marsit S, Charron G, Landry CR. 2020. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. *eLife* 9:e60474.
- Heyduk K, McAssey EV, Grimwood J, Shu S, Schmutz J, McKain MR, Leebens-Mack J. 2021. Hybridization history and repetitive element content in the genome of a homoploid hybrid, *Yucca gloriosa* (Asparagaceae). *Front Plant Sci.* 11:573767.
- Hou J, Friedrich A, de Montigny J, Schacherer J. 2014. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol.* 24(10):1153–1159.
- Hu WS, Temin HM. 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A.* 87(4):1556–1560.
- Huang CRL, Burns KH, Boeke JD. 2012. Active transposition in genomes. *Annu Rev Genet.* 46:651–675.
- Istace B, Friedrich A, d'Agata L, Faye S, Payen E, Beluche O, Caradec C, Davidas S, Cruaud C, Liti G, et al. 2017. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1–13.
- Jordan IK, McDonald JF. 1998. Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. *J Mol Evol.* 47(1):14–20.
- Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable element proliferation and genome expansion are rare in contemporary Sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. *Genome Biol Evol.* 3:156–167.
- Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* 8(5):464–478.
- Kojima KK. 2019. Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet Syst.* 94(6):233–252.
- Lawler JF, Haeussler DP, Dull A, Boeke JD, Keeney JB. 2002. Ty1 defect in proteolysis at high temperature. *J Virol.* 76(9):4233–4240.
- Lee BS, Lichtenstein CP, Faiola B, Rinckel LA, Wysock W, Curcio MJ, Garfinkel DJ. 1998. Posttranslational inhibition of Ty1 retrotransposition by nucleotide excision repair/transcription factor TFIIH subunits Ssl2p and Rad3p. *Genetics* 148(4):1743–1761.
- Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A-B, Vieira C, González J. 2019. Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol Ecol.* 28(6):1506–1522.
- Lesage P, Todeschini AL. 2005. Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet Genome Res.* 110(1–4):70–90.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

- Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* 22(3):177–192.
- Mascagni F, Vangelisti A, Giordani T, Cavallini A, Natali L. 2018. Specific LTR-retrotransposons show copy number variations between wild and cultivated sunflowers. *Genes* 9(9):433.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226(4676):792–801.
- Mérel V, Boulesteix M, Fablet M, Vieira C. 2020. Transposable elements in *Drosophila*. *Mob DNA*. 11:23.
- 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(15):5766–5776.
- Negrini M, Ricchetti M, Nouvel P, Buc H. 1995. Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. *Proc Natl Acad Sci U S A*. 92(15):6971–6975.
- Paquin CE, Williamson VM. 1986. Ty insertions at two loci account for most of the spontaneous antimycin A resistance mutations during growth at 15 degrees C of *Saccharomyces cerevisiae* strains lacking *ADH1*. *Mol Cell Biol*. 6(1):70–79.
- Patterson K, Shavarebi F, Magnan C, Chang I, Qi X, Baldi P, Bilanchone V, Sandmeyer SB. 2019. Local features determine Ty3 targeting frequency at RNA polymerase III transcription start sites. *Genome Res*. 29(8):1298–1309.
- Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339–344.
- Pryszcz LP, Németh T, Saus E, Ksiezopolska E, Hegedúsová E, Nosek J, Wolfe KH, Gacser A, Gabaldón T. 2015. The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. *PLoS Genet*. 11(10):e1005626.
- Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddalo JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* 5:e15716.
- Rachidi N, Barre P, Blondin B. 1999. Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. *Mol Gen Genet*. 261(4–5):841–850.
- Ray DA, Grimshaw JR, Halsey MK, Korstian JM, Osmanski AB, Sullivan KAM, Wolf KA, Reddy H, Foley N, Stevens RD, et al. 2019. Simultaneous TE analysis of 19 Heliconiine butterflies yields novel insights into rapid TE-based genome diversification and multiple SINE births and deaths. *Genome Biol Evol*. 11(8):2162–2177.
- Roeder GS, Fink GR. 1980. DNA rearrangements associated with a transposable element in yeast. *Cell* 21(1):239–249.
- Rogivue A, Choudhury RR, Zoller S, Joost S, Felber F, Kasser M, Parisod C, Gugerli F. 2019. Genome-wide variation in nucleotides and retrotransposons in alpine populations of *Arabidopsis alpina* (Brassicaceae). *Mol Ecol Resour*. 19(3):773–787.
- 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(7):3922–3938.
- Sandmeyer S, Patterson K, Bilanchone V. 2015. Ty3, a position-specific retrotransposon in budding yeast. *Mob DNA*. 11:965–996.
- Smukowski Heil C, Patterson K, Shang M, Hickey A, Alcantara E, Dunham MJ. 2021. Transposable element mobilization in interspecific yeast hybrids. *Genome Biol Evol*. 13(3):evab033.
- Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018. Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of the Mediterranean grass *Brachypodium distachyon*. *Genome Biol Evol*. 10(1):304–318.
- Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS, McCusker JH. 2015. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res*. 25(5):762–774.
- Thomas-Bulle C, Piednoël M, Donnart T, Filée J, Jollivet D, Bonnivard É. 2018. Mollusc genomes reveal variability in patterns of LTR-retrotransposons dynamics. *BMC Genomics* 19(1):821.
- Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324(5927):659–662.
- Tucker JM, Larango ME, Wachsmuth LP, Kannan N, Garfinkel DJ. 2015. The Ty1 Retrotransposon restriction factor p22 targets gag. *PLoS Genet*. 11(10):e1005571.
- Vela D, Fontdevila A, Vieira C, García Guerreiro MP. 2014. A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. *PLoS One* 9(2):e88992.
- Wallau GL, Capy P, Loreto E, Le Rouzic A, Hua-Van A. 2016. VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the Mariner family within *Drosophila*. *Mol Biol Evol*. 33(4):1094–1109.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 8(12):973–982.
- Zhu YO, Sherlock G, Petrov DA. 2016. Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)*. 6(8):2421–2434.