Evolutionary Dynamics of *hAT* DNA Transposon Families in Saccharomycetaceae

Véronique Sarilar[1,2], Claudine Bleykasten-Grosshans[3], and Cécile Neuvéglise[1,2,*]

[1]INRA, UMR 1319 Micalis, Jouy-en-Josas, France

[2]AgroParisTech, UMR Micalis, Jouy-en-Josas, France

[3]CNRS, UMR 7156, Laboratoire de Génétique Moléculaire, Génomique et Microbiologie, Université de Strasbourg, Strasbourg, France

*Corresponding author: E-mail: Cecile.Neuveglise@grignon.inra.fr.

Accepted: December 6, 2014

Data deposition: This project has been deposited at EMBL-ENA under the accession LM651396-LM651399.

Abstract

Transposable elements (TEs) are widespread in eukaryotes but uncommon in yeasts of the Saccharomycotina subphylum, in terms of both host species and genome fraction. The class II elements are especially scarce, but the *hAT* element *Rover* is a noteworthy exception that deserves further investigation.

Here, we conducted a genome-wide analysis of *hAT* elements in 40 ascomycota. A novel family, *Roamer*, was found in three species, whereas *Rover* was detected in 15 preduplicated species from *Kluyveromyces*, *Eremothecium*, and *Lachancea* genera, with up to 41 copies per genome. *Rover* acquisition seems to have occurred by horizontal transfer in a common ancestor of these genera. The detection of remote *Rover* copies in *Naumovozyma dairenensis* and in the sole *Saccharomyces cerevisiae* strain AWRI1631, without synteny, suggests that two additional independent horizontal transfers took place toward these genomes. Such patchy distribution of elements prevents any anticipation of TE presence in incoming sequenced genomes, even closely related ones.

The presence of both putative autonomous and defective *Rover* copies, as well as their diversification into five families, indicate particular dynamics of *Rover* elements in the *Lachancea* genus. Especially, we discovered the first miniature inverted-repeat transposable elements (MITEs) to be described in yeasts, together with their parental autonomous copies. Evidence of MITE insertion polymorphism among *Lachancea waltii* strains suggests their recent activity. Moreover, 40% of *Rover* copies appeared to be involved in chromosome rearrangements, showing the large structural impact of TEs on yeast genome and opening the door to further investigations to understand their functional and evolutionary consequences.

Key words: yeast, MITE, evolution, *Rover*, *Roamer*, horizontal transfer.

Introduction

Transposable elements (TEs) are segments of DNA that can move within the genome, via RNA intermediates (class I elements) or by a cut-and-paste mechanism (class II elements) (Wicker et al. 2007). They are ubiquitous in eukaryotes and sometimes represent up to 85% of the host genome, as in *Zea mays* (Schnable et al. 2009). However, they are uncommon in yeasts of the subphylum Saccharomycotina. Class I elements, also called retrotransposons, are the most widespread elements found in these species (Bleykasten-Grosshans and Neuvéglise 2011). Retrotransposons bearing long terminal repeats (LTRs) were first discovered in *Saccharomyces cerevisiae* several decades ago, in the laboratories of Fink and Davis (Cameron et al. 1979; Greer and Fink 1979; Roeder and Fink 1980). At that time, *Ty1* and *Ty3* in *S. cerevisiae*, along with the *copia* and *gypsy* elements in *Drosophila*, were considered to be the references for two LTR–retrotransposon superfamilies (Wicker et al. 2007). Other yeast LTR–retrotransposons were subsequently characterized in *Schizosaccharomyces pombe* (Levin et al. 1990) and in a few Saccharomycotina species, such as *Candida albicans* (Chen and Fonzi 1992) and *Yarrowia lipolytica* (Schmid-Berger et al. 1994). When comparative genomics emerged

in the early 2000s, TEs were identified in a broader range of species based on sequence similarity, which allowed phylogenetic studies to be conducted and evolutionary scenarios to be reconstructed (Neuvéglise et al. 2002). Around the same time, non-LTR yeast retrotransposons were discovered in *C. albicans* (Goodwin et al. 2001) and *Y. lipolytica* (Casaregola et al. 2002), as were the first class II elements (Goodwin and Poulter 2000). In general, class II elements are very rare in yeasts and so far only a few families have been described. A family of *Tc1/mariner* elements, named *Cirt* or *Fot1_CA*, was discovered in *C. albicans*; it was then also detected in other CTG species by sequence similarity searches (Bleykasten-Grosshans and Neuvéglise 2011). A member of the *Mutator* superfamily, named *Mutyl*, has been found in *Y. lipolytica* (Neuvéglise et al. 2005). Its second coding sequence (CDS) has the surprising capacity to be alternatively spliced and can thus yield four transcript variants. More recently, sequences similar to *hAT* DNA transposons in plants and fungi (Rubin et al. 2001) have been found in several hemiascomycete yeasts, that is, *Lachancea thermotolerans*, *Lachancea kluyveri*, *Kluyveromyces lactis* and *Eremothecium gossypii* (Souciet et al. 2009). These elements have been called *Rover*. The full-length *hAT*-like *Rover* elements reported in these yeasts are characterized by a single CDS and recognizable terminal inverted repeats (TIRs). However, given the limited number of *Rover* copies, the authors concluded that this group of elements probably had limited evolutionary success. Thus far, *Rover* is the sole member of the *hAT* superfamily described in Saccharomycotina that contains all the characteristics of putative active elements.

By moving about in the genome and recombining, TEs can generate chromosomal rearrangements and have mutagenic effects; they thus play a critical role in genome function and evolution. LTR–retrotransposons and solo LTRs have, in several cases, modified gene expression in *S. cerevisiae* by displacing gene regulatory sequences located adjacent to their insertion sites or by bringing a new regulatory sequence with them; they can also directly disrupt coding sequences (Lesage and Todeschini 2005). Class II elements are also involved in genome restructuration at both the structural and the functional level, which may lead to phenotypic variation. For instance, in plants, where *hAT* elements are abundant and have been extensively studied, *hAT* elements have often been found to contribute to phenotypic variation (Oliver et al. 2013; Vitte et al. 2014). They can do so by inserting themselves into exons, 5′ untranslated regions, or sequences nearer to or in promoter regions; they can also use transposase exaptation. Class II elements may employ similar mechanisms in yeasts and thus contribute to genome expression modification. However, as these elements are not found in the model yeast *S. cerevisiae*, nothing is yet known about their impacts.

This study was motivated by our finding that several other newly sequenced *Lachancea* genomes (unpublished results) contain up to 41 *Rover* copies, including potentially active ones. We studied the occurrence and evolution of *hAT* elements in Ascomycota yeast species and characterized intact copies of *Rover*. We also describe *Roamer*, a novel *hAT* family recently found to be co-opted and to promote sexual differentiation in *K. lactis* (Rajaei et al. 2014), and potentially active in *Naumovozyma castellii*; 18 intact or degenerate copies are present in the strain CBS 4309. Because *Rover* is only found in the genera *Lachancea*, *Kluyveromyces*, and *Eremothecium*, we propose that *Rover* originated prior to the diversification of these taxa. However, *Naumovozyma dairenensis* and *S. cerevisiae* strain AWRI1631 also contain a few *Rover* copies, which probably originated from two independent horizontal transfers (HTs). We paid particular attention to the dynamics of *Rover* in the *Lachancea* clade and found evidence for species-specific amplification bursts, chromosomal rearrangements, and the formation of miniature inverted-repeat transposable elements (MITEs) from parental autonomous copies; this is the first time that MITEs have been reported in yeast.

Materials and Methods

Analysis of *hAT* Copies

Annotated genome sequences for 41 Ascomycota yeast species were retrieved from different databases. The complete list is provided in supplementary table S1, Supplementary Material online. We sequenced the seven *Lachancea* genomes that were as yet unavailable, using a Roche 454 sequencer and employing a combination of 400-bp paired ends and 8-kb mate pairs. This sequencing strategy allowed the assembly of repeated sequences shorter than 8 kb in the scaffolds, which is the case of all TEs in the *Lachancea* species. The resulting coverage ranged from 25 to 30×. Genome assembly of the seven genomes was achieved with Celera Assembler v6.1 (Myers et al. 2000) and Newbler v2.7 (454 Life Sciences) with default parameters. As many scaffolds as chromosomes were obtained; for each species, with the exception of *Lachancea mirantina*, both Celera and Newbler assemblies turned out to be collinear and congruent with karyotypes visualized on pulsed field gels (data not shown). In *L. mirantina*, the misassembly involved two chromosomes, only Newbler being able to produce scaffolds of the expected size.

We searched for *Rover*-like transposases in these genomes with BLASTP and TBLASTN, using annotated *Rover* elements from *L. thermotolerans* (Souciet et al. 2009). To extend our analysis to the whole *hAT* superfamily, *hAT*-like transposases were identified using a Pfam v27.0 scan (Finn et al. 2014), with an *e*-value cutoff of 0.01 for the 41 proteomes. We searched for the most characteristic domains (i.e., the BED zinc finger domain PF02892 and the dimerization domain PF5699) that are associated with *hAT* transposases (Aravind 2000; Essers et al. 2000). TIRs were then identified by performing pairwise mapping of sequences adjacent to

transposase CDSs (dotplots were generated using http://www.vivo.colostate.edu/molkit/dnadot/, last accessed December 19, 2014) and by using TBLASTN to search genome sequences. We also used TBLASTN to search the nonredundant NCBI (National Center for Biotechnology Information) database to check if *Rover* and *Roamer* transposases could be found in other species.

Sequence similarity between the transposases was computed using GeneDoc v2.7.0 (Nicholas et al. 1997), following sequence alignment using MAFFT v6.903b (Katoh et al. 2002) and manual corrections. Only transposases longer than 500 aa or 800 aa were used for *Rover* and *Roamer*, respectively. Heat map and similarity-based clustering of the 48 *Rover* transposases carrying at least one TIR was produced with R (R Core Team 2014). We checked for the presence of *hAT*-like conserved domains (Rubin et al. 2001) using a single transposon copy from each *Rover* family in *Lachancea* species and using one transposon copy per species for the other species; we also examined 13 *Roamer* copies. Sequence logos of TIRs were generated by WebLogo v2.8.2 (Crooks et al. 2004). DNA secondary structures of MITEs were determined by SantaLucia free energy rules (SantaLucia 1998) using the program mfold v3.6 with the default settings (http://mfold.rit.albany.edu//?q=mfold/DNA-Folding-Form, last accessed December 19, 2014; Zuker 2003). Synteny analyses for *Rover* and *Roamer* insertion site conservation were performed with CHROnicle/SynChro (Drillon et al. 2014, available at http://www.lgm.upmc.fr/CHROnicle/SynChro.html, last accessed December 19, 2014), with a Delta parameter of 2 or 4, meaning that two adjacent genes are separated by less than two or four genes, respectively, in the absence of orthology. Synteny was visualized using dotplots produced by SynChro. Genomes were visualized using ARTEMIS (Rutherford et al. 2000).

Phylogenetic Analyses

Phylogenetic analyses were conducted on the 41 Ascomycota yeast species where *hAT* elements were searched for, on *hAT* elements from diverse eukaryotic lineages, and on the *Rover* transposases. The general line of analyses was to align sequences with MAFFT v6.903b (Katoh et al. 2002) or MultAlin (Corpet 1988) in case of highly divergent sequences and to draw a phylogeny with the amino acid substitution model recommended by ProtTest v3.4 (Darriba et al. 2011). The maximum-likelihood algorithm PhyML v3.0 (Guindon et al. 2010) and the Bayesian inference program MrBayes v3.2.2 (Ronquist and Huelsenbeck 2003) were both used except when the substitution model was not implemented in MrBayes, as it is the case for LG model.

The species phylogeny was constructed using a set of 104 protein sequences from referenced hemiascomycete genomes (supplementary table S1, Supplementary Material online); these proteins are singletons in all the species considered. Protein sequences were aligned using MAFFT and subsequently cleaned using Gblocks v.0.91b, in which the options −b3 = 20 and −b0 = 5 were applied (Castresana 2000). The 104 alignments were then concatenated, leading to an alignment that was 39,926 residues long. The phylogenetic analysis was performed with PhyML, using the amino acid substitution model LG (Le and Gascuel 2008) with estimated proportion of invariable sites (+I) and shape parameter of the gamma distribution (+G). Confidence was assessed in 100 bootstrap replicates.

The *hAT* phylogeny was constructed using transposase amino acid sequences from 46 full-length and potentially active transposases, which were selected because they are associated with known *hAT* elements or because they were revealed by BLASTP hits in the nonredundant NCBI database (nr) when *Rover* elements were used as queries. Sequences were aligned using MultAlin and manually corrected using GeneDoc v2.7.0 (Nicholas et al. 1997). Only conserved blocks of amino acids were kept, leading to a 468-aa alignment. The phylogenetic analysis was performed with PhyML, using the amino acid substitution model LG + G (Le and Gascuel 2008). Confidence was assessed in 100 bootstrap replicates.

To analyze *Rover* copy clustering in *Lachancea* genomes, sequences of the 110 transposases longer than 500 aa were aligned using MAFFT. Only conserved blocks of amino acids were kept after manual correction using GeneDoc, leading to a 633-aa alignment. The phylogenetic analysis was performed with both MrBayes and PhyML using the amino acid substitution model JTT + G (Jones et al. 1992) and 100 bootstrap replicates. With MrBayes program, 2 million generations were evaluated, sampling the most probable tree every 100 generations and burning 25% of those. Trees resulting from both programs showed the same tree topology with coherent confidence.

Analysis of Selection Pressure and d*N*/d*S* Ratios

To examine the selection pressure acting on the potentially fixed *Rover* copies, we evaluated the strength of purifying selection on their transposases in comparison with the 104 genes used for phylogenetic reconstructions. Pairwise nucleotide alignments were deduced from MAFFT protein alignments using tranalign from the EMBOSS package (Rice et al. 2000). The d*N*/d*S* ratios were calculated from these alignments using CodeML model with a pairwise run mode from PAML4 package version 4.4b (Yang 2007). d*S* values higher than 1 were removed from the analysis to avoid d*S* saturation.

Experimental Detection of MITE Insertion Polymorphisms in Nine *L. waltii* Strains

We looked for the presence of insertion polymorphisms of the five MITE copies found in *L. waltii* CBS6430 (= NCYC 2644) in the eight other *L. waltii* strains available in international culture collections: CBS7703, CBS8527, CBS8528, UWO PS79-161,

UWO PS81-128, UWO PS78-160, UWO PS82-227, and UWO PS82-228. These strains were grown at 28 °C on YPD plates (yeast extract, peptone, and glucose; 10 g/l each) for 72 h. Total DNA was extracted using a standard phenol/chloroform extraction protocol followed by ethanol precipitation. For PCR (Polymerase Chain Reaction) amplification, primers were designed upstream and downstream from each of the five MITE copies found in *L. waltii* strain NCYC2644. The expected sizes of the amplicons, based on the presence or absence of a MITE copy at the locus, as well as the primer sequences are provided in supplementary table S2, Supplementary Material online. PCR reactions were carried out in a total volume of 25 µl, which contained about 100 ng of total DNA, 25 pmol of each primer, 200 µM of dinucleotide triphosphates (dNTPs), and 1 U of Ex-Taq Polymerase (Takara) in a corresponding 1 × buffer. The PCR program consisted of an initial denaturation step at 95 °C for 2 min, which was followed by 30 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min and a final elongation step at 72 °C for 5 min. Amplification products were visualized on a ethidium bromure-stained 0.8% agarose gel.

Results

hAT Elements Are Widespread in Saccharomycetaceae

We investigated whether *hAT* elements are present in 41 sequenced Ascomycota yeast genomes (fig. 1*A*), including 7 newly sequenced *Lachancea* species (our laboratory, unpublished results). An initial BLAST-based search of the previously annotated *Rover* elements found in the *L. thermotolerans* genome led to the identification of a total of 151 *Rover* elements; they were found in 10 *Lachancea* genomes (143 copies), in *E. gossypii* (1 copy), *Eremothecium aceri* (1 copy), *Eremothecium cymbalariae* (1 copy), *K. lactis* (2 copies), *Kluyveromyces marxianus* (2 copies), and *N. dairenensis* (1 copy) (table 1). Two *Rover* copies had been previously annotated by the Génolevures consortium in the *K. lactis* genome (Souciet et al. 2009), but the coordinates of the copy KLLA0F02610t on chromosome Klla0F had to be modified (updated version is available in http://gryc.inra.fr, last accessed December 19, 2014; supplementary table S4, Supplementary Material online).

In the other 25 complete genomes analyzed, no sequences homologous to the *Rover* transposase or its TIRs were found. An additional homology search was performed using the nonredundant NCBI database and revealed the presence of three degenerate copies in the wine yeast *S. cerevisiae* strain AWRI1631 (table 1).

To broaden our investigation of *hAT* elements in ascomycota, we used another approach, in which we systematically searched for the transposase Pfam domains PF02892 and PF5699 in the 41 proteomes studied. Using this technique, other groups of *hAT* elements were identified (supplementary table S3, Supplementary Material online). In *Naumovozyma castellii,* 17 CDSs were found to contain the dimerization domain and 3 of them also contained the BED ZnF domain. Eleven of them are flanked by 28-bp TIRs. These CDSs had already been predicted and were annotated as hypothetical proteins, but the coordinates of one of them, NCAS0F00700, had to be extended. This group differs from *Rover* in TIR length; their transposases are also dissimilar. We named this new group *Roamer*. A TBLASTN search involving the 41 genomes led to the identification of an 18th copy in *N. castellii* and putative homologs in *K. lactis* (KLLA0D05677g, recently named *Kat1*; Rajaei et al. 2014) and *K. marxianus*, which showed up to 15% identity and 34% similarity. In the latter species, the *Roamer*-like copy appears to be the concatenation of the two genes KLMA_70035 and KLMA_70036, which formed a pseudogene of 971 aa in strain DMKU3-1042. Both *Kluyveromyces* copies, which have 68% identity and 82% similarity, show strictly conserved synteny, which suggests that the acquisition occurred in a single shared ancestor. No other *Roamer* elements were found in the 38 other species surveyed in this study.

When the *Roamer*- and *Rover*-containing species were included in the Saccharomycotina phylogeny, it became apparent that the distribution of these elements is biased and restricted to particular species groups, which suggested acquisition by different HTs: Two primary events followed by three secondary HTs (fig. 1*A*). Three criteria are commonly used to infer HT: 1) Patchy taxonomic distribution, 2) incongruence between host and TE phylogenies, and 3) sequence conservation incompatible with a vertical inheritance, that is, higher similarity between TEs of distantly related species than between either the host protein-coding genes or the TE copies from the same genome (Loreto et al. 2008; Schaack et al. 2010). The two primary HTs are only supported by a patchy distribution, as no closely related copies were detected outside the 40 species studied. For the three secondary HTs, the three criteria were investigated, as well as synteny conservation, which may indicate a vertical transmission in case of orthologous position conservation. The patchy distribution and the fact that species and *Rover* phylogenies are discordant are shown in figures 1*A* and 3. The *Rover* copy from *N. dairenensis* is inserted between CDS orthologous to YDR496C (PUF6) and YDR489W (SLD5) from *S. cerevisiae*. No *Rover* copy is found next to these genes in *Lachancea, Eremothecium*, and *Kluyveromyces* studied species (LaKlEr), suggesting that the *N. dairenensis* copy is not an ancestral copy shared with these genomes. Similarly, the *S. cerevisiae* copies are not conserved in synteny with any of the LaKlEr *Rover* and their transposase sequences display a lower degree of identity to each other (less than 15% identity and 34% similarity over 446 aa) than they are to the LaKlEr copies (up to 41% identity and 62% similarity over 773 aa with *Rover1* LAFA0C07074t). The *N. dairenensis* element displays 21% identity and 38% similarity (over 530 aa) with the most closely related *Rover*

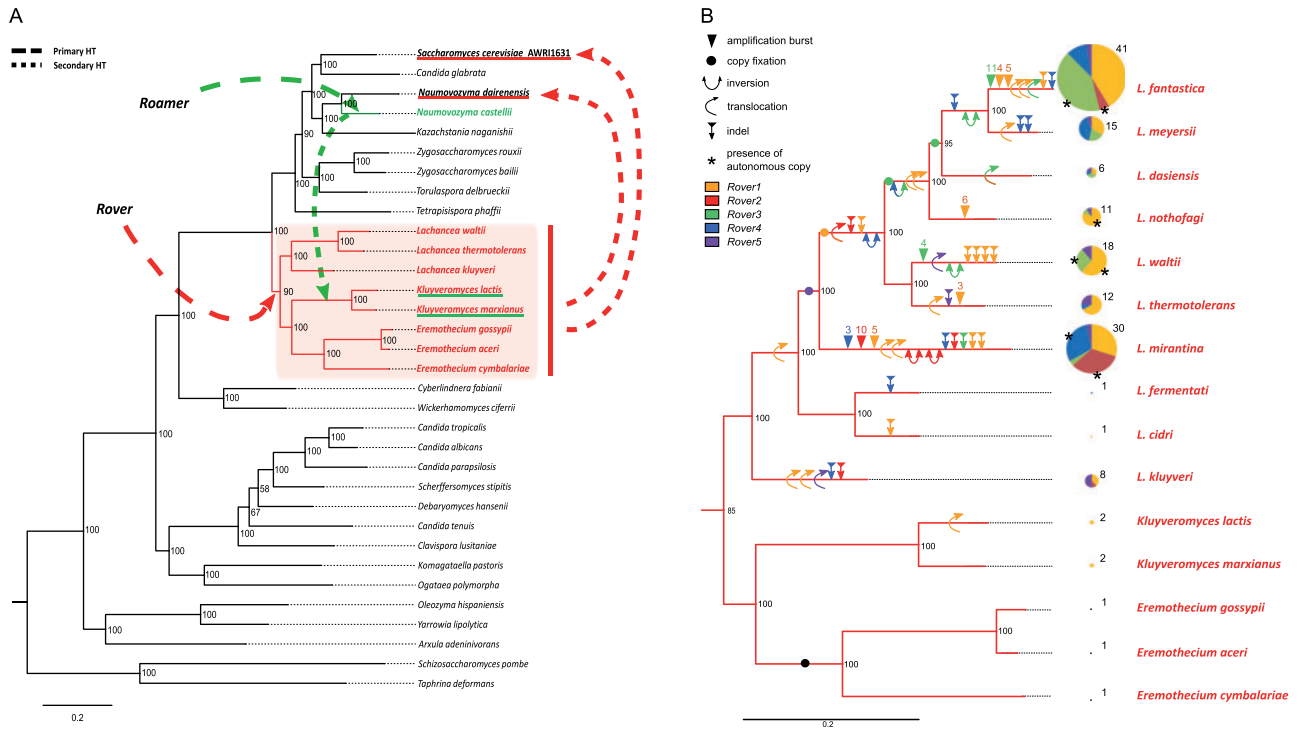

Fig. 1.—Dynamics of *Rover* and *Roamer* elements in Ascomycota yeasts. (*A*) Taxonomic distribution of *Rover* and *Roamer* among Ascomycota yeasts suggesting successive HT events. The maximum-likelihood tree was generated from the concatenation of 104 protein sequence alignments. Bootstrap values supporting each clade are indicated on nodes. *Rover* and *Roamer* elements were found in the species highlighted in bold. Dotted arrows indicate hypothetical HT events that would explain the patchy distribution of *Rover* and *Roamer* elements among Ascomycota. It is hypothesized that a *Rover* element was initially acquired by the ancestor of the *Lachancea/Kluyveromyces/Eremothecium* species group—highlighted in red (red dotted line)—and then subsequently diversified in the descendant lineages. Two *Rover* copies were independently transferred to *Saccharomyces cerevisiae* strain AWRI1631 and *N. dairenensis*. Similarly, *Roamer* seems to have been horizontally transferred to *N. castellii* (green dotted line), and then further spread to the *Kluyveromyces* species we studied. (*B*) Detailed *Rover* dynamics within the *Lachancea/Kluyveromyces/Eremothecium* clade. The amplification bursts, the copy fixations, and the chromosomal rearrangements—translocations, inversions, and indels of genes next to a copy insertion site—involving *Rover* elements are depicted on the maximum-likelihood tree. Amplification bursts are numbered by the count of copies having emerged from the burst according to the transposase phylogeny presented in figure 5. Each event is colored as follows: Orange for *Rover1*, red for *Rover2*, green for *Rover3*, blue for *Rover4*, and purple for *Rover5*. Genome contents in *Rover* families are specified by pie charts of a size proportional to the total *Rover* copy numbers, indicated next to them. Stars notify the presence of autonomous copies from particular families.

element, *Rover1* LAFA0G24652t. These values are higher than those found between some *Lachancea Rover* copies (fig. 4), although lower than the average percentage of similarity between the 104 genes that were used to define the phylogeny (77% between *N. dairenensis* and LaKlEr genes); this comparison being probably biased due to less constraint on the TE sequences. For *Roamer*, the patchy distribution is limited to three species, which prevents phylogenetic comparisons between TEs and hosts. The *Roamer* copies of *K. lactis* and *K. marxianus* are not found at the orthologous position in *N. castellii* (i.e., between YFL045C and YLR075W homologous CDSs), and sequence conservation between *Kluyveromyces* and *N. castellii Roamer* copies (26–35%) is always slightly higher than the lowest conservation between *N. castellii* copies (24–100%). Thus, *Rover* and *Roamer* elements fulfill all applicable criteria to support the hypothesis of secondary HTs.

In addition to *Rover* and *Roamer* elements, 52 putative *hAT*-like elements and other class II elements were found in 35 of the 41 species studied, using the Pfam scan. Eleven are singletons and the others belong to five families that are either species specific or conserved among clades (supplementary table S3, Supplementary Material online). For instance, in *Taphrina deformans*, we identified three *hAT*-like degenerate copies without TIRs that nonetheless contained the dimerization domain PF05699. Once again, the transposases had already been predicted to be hypothetical proteins, but their coordinates had to be extended. Coordinates for the *Rover*, *Roamer*, and *T. deformans* elements are given in supplementary table S4, Supplementary Material online.

Table 1
Distribution of *Rover* Elements in *Rover*-Containing Saccharomycetaceae Species

Species	Strains	Number of *Rover* copies		
		Potentially autonomous	Nonautonomous or degenerate	MITE/ miniature copy
E. gossypii	ATCC 10895	0	1	0
E. aceri		0	1	0
E. cymbalariae	DBVPG#7215	0	1	0
K. lactis	CLIB 210	0	2	0
K. marxianus	DMKU3-1042	0	2	0
L. cidri	CBS 2950	0	1	0
L. dasiensis	CBS 10888	0	6	0
L. fantastica	CBS 6924	8	30	3
L. fermentati	CBS 6772	0	1	0
L. kluyveri	CBS 3082	0	8	0
L. meyersii	CBS 8951	0	15	0
L. mirantina	CBS 11717	5	24	1
L. nothofagi	CBS 11611	1	10	0
L. thermotolerans	CBS 6340	0	12	0
L. waltii	NCYC 2644	5	8	5
N. dairenensis	CBS 421	0	1	0
S. cerevisiae	AWRI1631	0	3	0

NOTE.—References to the genome sequences and annotation are supplied in supplementary table S1, Supplementary Material online.

Survey of *Rover* and *Roamer* Structural Features

Full-length *Rover* copies are about 3-kb long; their transposases, which are 800 aa in length on average, harbor a C-terminal dimerization domain (PF05699) that is characteristic of the *hAT* element superfamily (Essers et al. 2000) and an N-terminal DNA-binding domain (PF02892) that is fairly typical of transposases (Aravind 2000). They have 18- to 20-bp TIRs flanked by 8-bp target-site duplications (TSDs) (fig. 2). Among the 154 copies observed, 19 copies, all belonging to *Lachancea* species, exhibit perfect TIRs (or nearly perfect in the case of one copy with an identity of 17/18 bp) and encode transposases longer than 724 aa that do not contain any interrupting stop codons or frameshifts and are thus considered to be potentially autonomous. Eighteen copies are nonautonomous—interrupted by stop codons or frameshifts—but contain two TIRs. Seventy-one copies are relics with degenerate CDSs and no more than one TIR or have an intact CDS that is associated with a single TIR. Thirty-seven copies are potentially functional transposases that are not flanked by TIRs. Nine of the copies appear to be miniature copies shorter than 500 bp (supplementary table S4, Supplementary Material online). The sequences of these 154 *Rover* copies are provided in supplementary files S1 (nucleotide sequences) and S2 (amino acid transposase sequences), Supplementary Material online.

Target-site preferences for the *Rover* elements were examined, but no conservation motif was found. This result contrasts with what has been found for other groups of *hAT* elements, such as *Buster* and *Ac* (Arensburger et al 2011). The G + C content of the TSDs is around 43%, which is not statistically different from the G + C content of the whole genome of the host species, which is 38.8% for *K. lactis*, 41.5% for *L. kluyveri*, and 47.3% for *L. thermotolerans* (Souciet et al. 2009). Moreover, the distribution pattern of *Rover* elements on the seven or eight chromosomes of *Lachancea* species (supplementary fig. S1*A*, Supplementary Material online) appeared to be random. No wide-scale copy pileup was observed. In a single case, seven copies were found to be clustered together within the first 66 kb of *Lachancea fantastica* chromosome F (supplementary fig. S1*B*, Supplementary Material online).

Full-length *Roamer* copies are about 3-kb long; their transposases are 860-aa long on average and they have 28-bp TIRs flanked by 8-bp TSDs (fig. 2). TIRs of *hAT* are usually shorter than 20 bp (Kempken and Windhofer 2001), but unusual cases of TIRs 25–27 bp in length have been reported in *Caenorhabditis elegans* (Bigot et al. 1996). At 28 bp, *Roamer* TIRs are the longest *hAT* TIRs reported thus far. Half of the 18 *Roamer* copies we observed are autonomous, 2 are nonautonomous, 6 are TIR-free transposases, and 1 is a relic (supplementary table S4, Supplementary Material online). The perfect TIR conservation seen in 11 copies and the presence of 9 intact transposases in full-length elements indicate that members of this family are potentially still active. The sequences of these 18 *Roamer* copies are provided in supplementary siles S3 (nucleotide sequences) and S4 (amino acid transposase sequences), Supplementary Material online.

Rover and *Roamer* Position in the *hAT* Phylogeny

A number of general characteristics differentiate *hAT* DNA transposons from other class II elements, that is, the presence of short TIRs 5–27 bp in length, 8-bp TSDs, and transposases with conserved motifs and domains (Kempken and Windhofer 2001; Wicker et al. 2007; Arensburger et al. 2011). The presence of the six *hAT*-conserved domains described by Rubin et al. (2001) (supplementary fig. S2, Supplementary Material online), as well as the characteristics of the TIRs and TSDs, confirm that the *Roamer* and the *Rover* groups belong to the *hAT* superfamily, as has already been suggested by the Génolevures consortium (Souciet et al. 2009) for *Rover* and by Rajaei et al. (2014) for *Roamer*. On the basis of the primary sequence of the transposases and TSD sequence conservation, Arensburger et al. (2011), Zhang et al. (2013), and Rossato et al. (2014) have divided the *hAT* superfamily into three groups: *Ac*, *Buster*, and *Tip*. To assign the *Rover* and the *Roamer* groups to one of the *hAT* groups, we compared their transposases with those found in other taxa, including plants, animals, and fungi. In our analyses, the *Tip* group did not appear to be monophyletic, which reflects how difficult it is to generate a robust alignment from highly variable

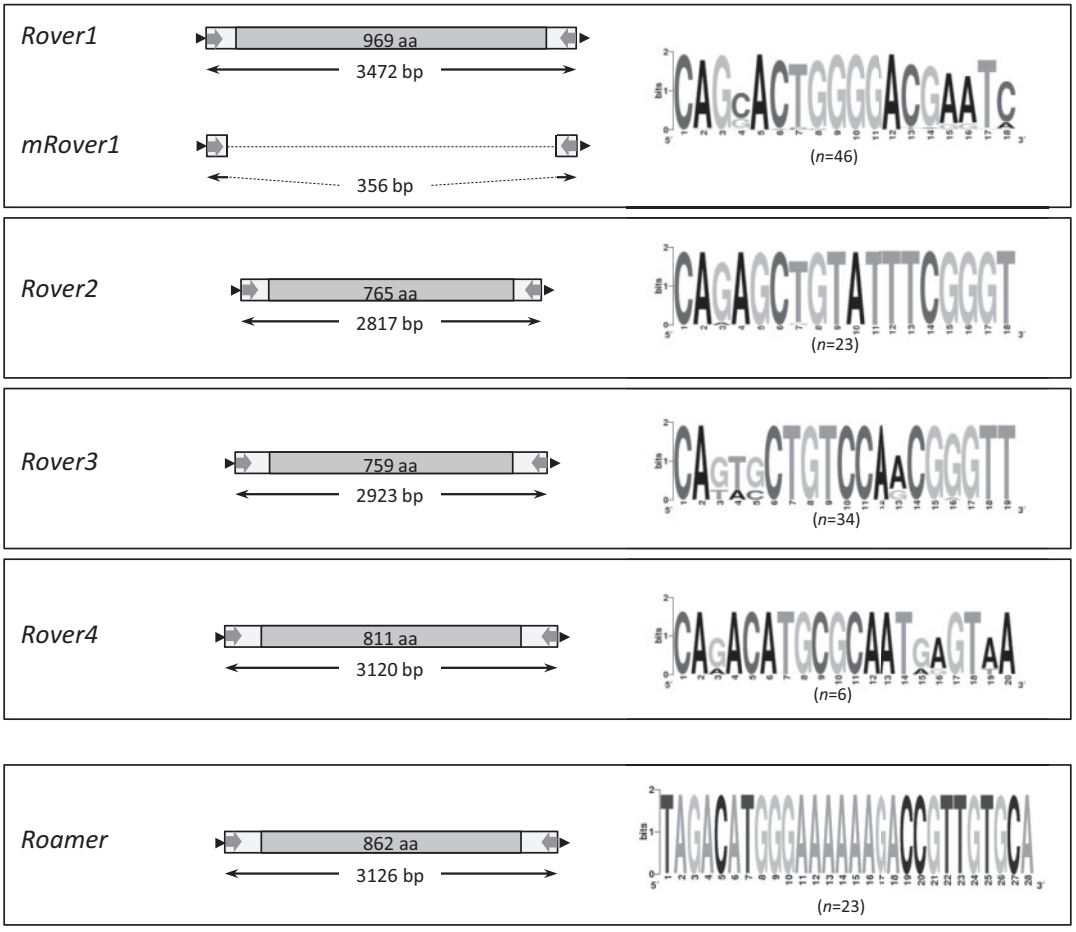

Fig. 2.—Structural features of *Roamer* and the four putative active *Rover* families. The CDSs that encode the transposases are depicted using gray rectangles; their mean length in amino acids (aa) is provided. Mean lengths were determined from full-length copies (3, 8, 5, 9, 1, and 9 copies for *Rover1*, *mRover1*, *Rover2*, *Rover3*, *Rover4*, and *Roamer*, respectively). TIRs and TSDs are represented by gray arrows and black triangles, respectively. The TIR sequences from all the TIRs available for each family (number provided in brackets) are represented as sequence logos located to the right.

sequences. Additional transposase datasets would be needed to infer a more robust phylogeny. However, the phylogeny showed that the *Rover* elements formed a monophyletic group, and that along with *Roamer*, they belong to *Ac* elements (fig. 3). The *Rover* and *Roamer* groups' lack of target-site preferences however distinguishes them from other *Ac* elements. Indeed, only 7 of the 74 *Rover* TSDs (9%) and 3 of the 12 *Roamer* TSDs (25%) have conserved T and A nucleotides at the second and seventh positions, which is a trait that is characteristic of *Ac* elements (Arensburger et al 2011).

Rover Elements Are Subdivided Into Five Families

Roamer copies are homogenous; their transposases have a mean pairwise amino acid similarity of 70%. Similarity ranges from 40% to 100% among the 15 potentially functional transposases (except for the shortest one). All TIRs are identical. These copies are thus considered to belong to the same family.

In contrast, the potentially functional transposases of *Rover* (19 autonomous copies + the 37 TIR-free copies mentioned above) range in size from 629 to 1,170 aa and they are 14–100% similar. This broad degree of sequence conservation led us to consider the possibility of defining different families. To this end, two criteria were used: Transposase sequence similarity and TIR conservation. Four groups of TIRs clearly emerged; their within-group variability varies from two to nine nucleotides (fig. 2). When the *Rover* transposases containing at least one TIR ($n = 48$ elements) were compared, a consistent overlap between the TIR groups and the transposase sequence clusters was observed. The elements in each TIR group had transposases that were monophyletic in origin (fig. 4). Although the delimitation of the phylogenetic groups may be open to debate, four families were nonetheless

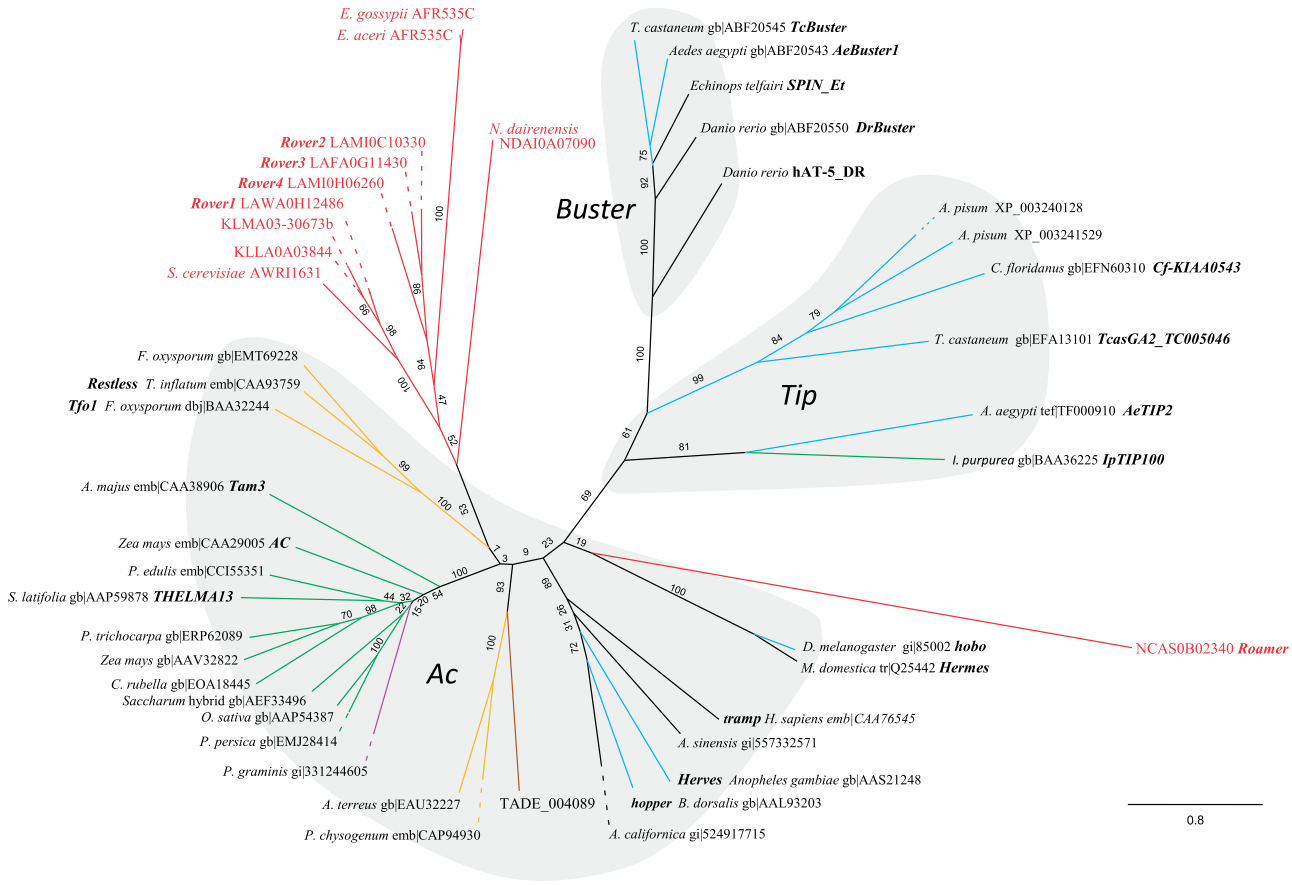

Fig. 3.—Phylogenetic tree of transposase sequences from *Rover*, *Roamer*, and other groups of *hAT* elements. A 468-aa alignment of 46 transposases was used to draw a maximum-likelihood tree. Bootstrap values are indicated above branches. Transposases are clustered in three groups, as defined by Rossato et al. (2014). However, in our tree, the *Buster* group is nested within the *Tip group* with however a moderate support. Branches are colored according to taxonomic group: Red for the hemiascomycetes, brown for Taphrinomycotina, orange for members of Pezizomycotina, purple for the basidiomycetes, green for the plants, blue for the insects, and black for the vertebrates. Sequence accession numbers for the representatives in the international sequence databases are provided: gb and gi, GenBank; emb, EMBL; dbj, DDBJ; and tef, TEfam.

defined and named *Rover1* to *Rover4*. Indeed, transposase conservation is highly variable among the four groups and a different number of groups might have been created if only phylogeny had been taken into account or if the 80-80-80 theory of Wicker et al. (2007) had been applied.

Given the consistency between sequence similarity and TIR conservation, elements without any TIRs were assigned to a given family according to their closest homolog within the reference set of the 48 elements used to define the families (supplementary table S4, Supplementary Material online). This clustering was consistent with the *Rover* transposase phylogeny (fig. 5), with the exception of a small remote group of TIR-free copies, which led to define a fifth family, *Rover5*. Finally, 69 elements were placed in the *Rover1* family, 13 in *Rover2*, 29 in *Rover3*, 22 in *Rover4*, 14 in *Rover5* and 7 elements too distant remained unclassified. The *Rover* families contain variable proportions of element types (autonomous, nonautonomous, relics, transposases, and miniature copies). For instance, no autonomous copy is found in *Rover5*, only 1 autonomous copy is found in *Rover4* (in *L. mirantina*), whereas as many as 10 autonomous copies are found in *Rover3*. Similarly, MITEs are only found in the *Rover1* family (see below).

Fixed Copies and Species-Specific Amplification Bursts

The phylogeny of the *Lachancea Rover* transposases revealed 9 species-specific clusters with 3–12 members (fig. 5). The largest clusters were found in *L. fantastica* (12 members) and *L. mirantina* (10 members) and corresponded to elements in the *Rover3* and *Rover2* families, respectively. The degree of transposase sequence conservation in these clusters is higher than in the family at large, with 79–99% identity in *L. fantastica* and 95–99% identity in *L. mirantina*. These two clusters contain the greatest number of autonomous copies: Six in *L. fantastica* and four in *L. mirantina*. These findings suggest that several amplification bursts took place during *Lachancea* evolution and indicate that *Rover* elements may have remained active after the divergence of each species (fig. 1*B*). Another

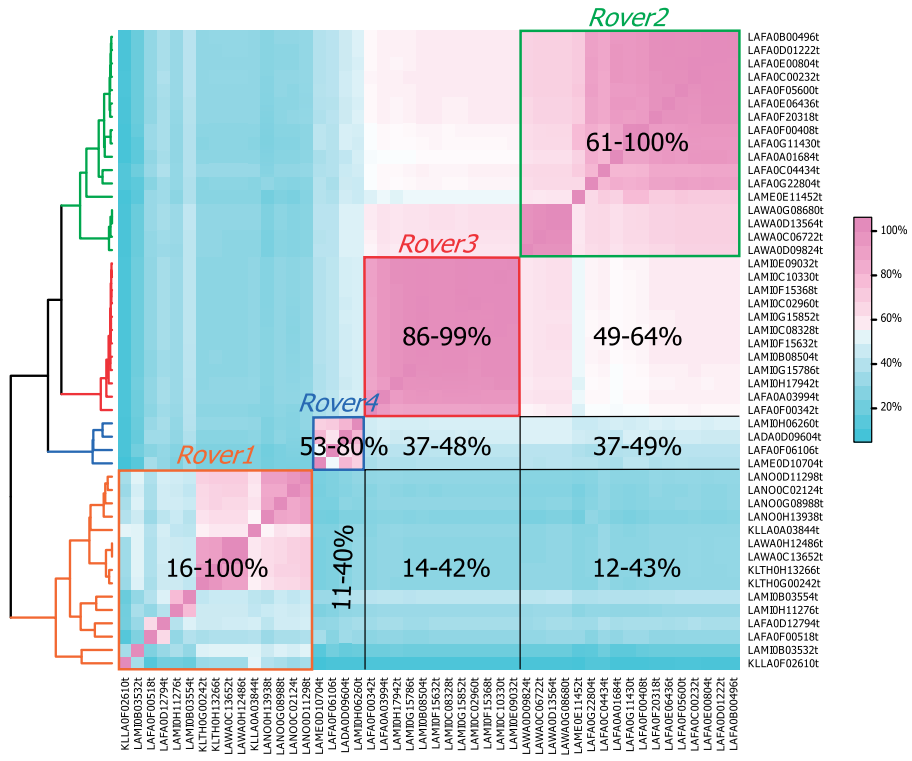

Fig. 4.—Delimitation of four *Rover* families using TIR sequences and transposase similarity. The delimitation is based on the 48 *Rover* copies carrying at least one TIR and a transposase longer than 500 aa. The heat map indicates the transposase pairwise similarity from 0% in blue to 100% in pink. The degenerate copy KLLA0F02610t was excluded from the similarity analysis. The colored rectangles delimit the copies harboring the same TIR types: *Rover1* is in orange, *Rover2* in red, *Rover3* in green, and *Rover4* in blue. The clustering on the left, based on similarity values, shows that transposase sequences cluster accordingly to the TIR sequences.

clear example of an amplification burst can be observed in the *Rover3* family, where a group of four highly conserved (96–100% identity) *L. waltii* copies are differentiated by a three-nucleotide modification in their TIRs: CA*GTG*CTGTCC AACGGGTTG has been replaced by CA*TAC*CTGTCCAGCGG GTT (supplementary table S4, Supplementary Material online). The emergence of this subfamily in *Rover3* after the speciation of *L. waltii* confirms the dynamics of *Rover* elements.

In contrast, three clusters of TIR-free transposases, with one member per species, were observed and appear to be related to the *Rover1*, *Rover3*, and *Rover5* families (fig. 5). The phylogenic relationships of the transposases within these clusters are congruent with the species phylogeny (fig. 1*B*). Moreover, in the three clusters, the position of the copies is fully conserved: Between the homologs of YHR143W-A (RPC10) and YHL048W (COS8) in the case of the *Rover1*-related cluster, upstream of YKL210W (UBA1) in the case of the *Rover3*-related cluster, and upstream of YAL040C (CLN3) in the case of the *Rover5*-related cluster. From the phylogenetic analysis, it is possible to estimate when the copy become immobilized—after the speciation of *L. mirantina* in the case of the *Rover1*-related cluster, before it in the case of the *Rover5*-related cluster, and before the speciation of *Lachancea nothofagi* with a loss in *Lachancea dasiensis* for the *Rover3*-related cluster (fig. 1*B*). An additional pair of fixed copies could be identified; it is composed of a TIR-free transposase and a relic copy in *L. dasiensis* and *Lachancea meyersii*, respectively. These findings suggest that these elements have been trapped at these loci, probably because they have lost their TIRs, and in some cases may have been further co-opted by the host genome. Besides an intact coding sequence and an inability to transpose due to an absence of TIRs, co-option is characterized by purifying selection acting on the transposase sequence. Hence, the d*N*/d*S* ratio was calculated for the 6 *Rover5*-related copies, the 6 *Rover1*-related copies, and 104 singleton host genes. However, all d*S* values for *Rover* pairwise comparisons were higher than 1, revealing a substitution saturation preventing further interpretation.

mRover1 Is a Newly Formed MITE Family

Nine of the *Rover* element copies observed in our study appear to be short nonautonomous derivatives resulting from internal deletion events. They are all found in *Lachancea* species: *L. waltii* (5 copies), *L. fantastica* (3 copies), and *L. mirantina* (1 copy). All of them lack transposases, and only a few

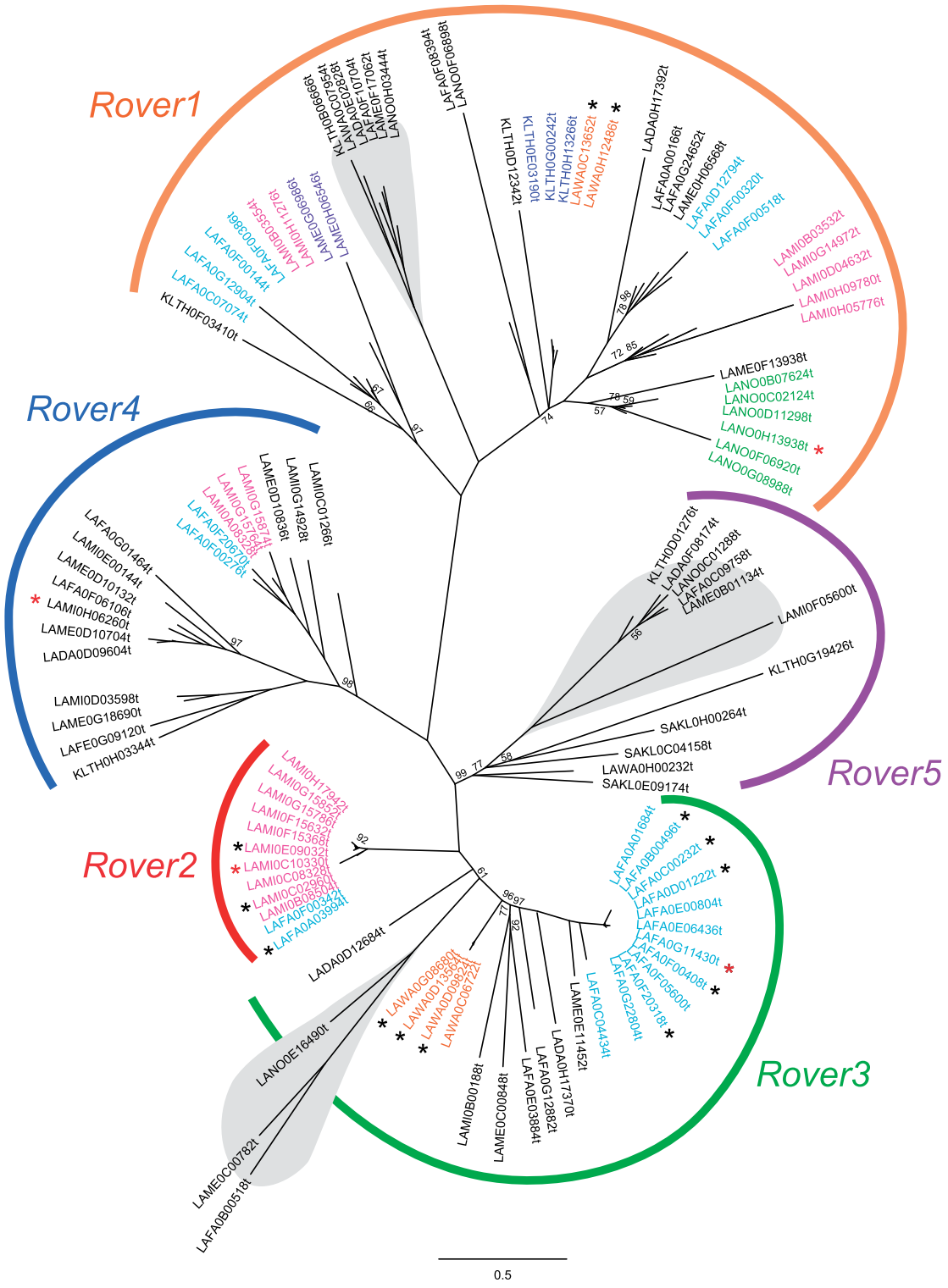

Fig. 5.—Phylogeny of *Rover* transposases in the *Lachancea* clade. The phylogeny was generated by a Bayesian analysis, based on the 110 transposases longer than 500 aa. The maximum-likelihood phylogeny showed the same topology. Posterior probabilities, written as a percentage and lower than 100, are indicated. Species-specific clusters are highlighted, with one color per species (light blue for *L. fantastica*, purple for *L. meyersii*, green for *L. nothofagi*, orange for *L. waltii*, dark blue for *L. thermotolerans*, and red for *L. mirantina*). Gray blocks indicate the three groups of fixed copies. Stars indicate the 18 putative autonomous copies. The red stars correspond to the four putative autonomous copies that were designated as representatives for the four putative active *Rover* families and that were deposited in the ENA database.

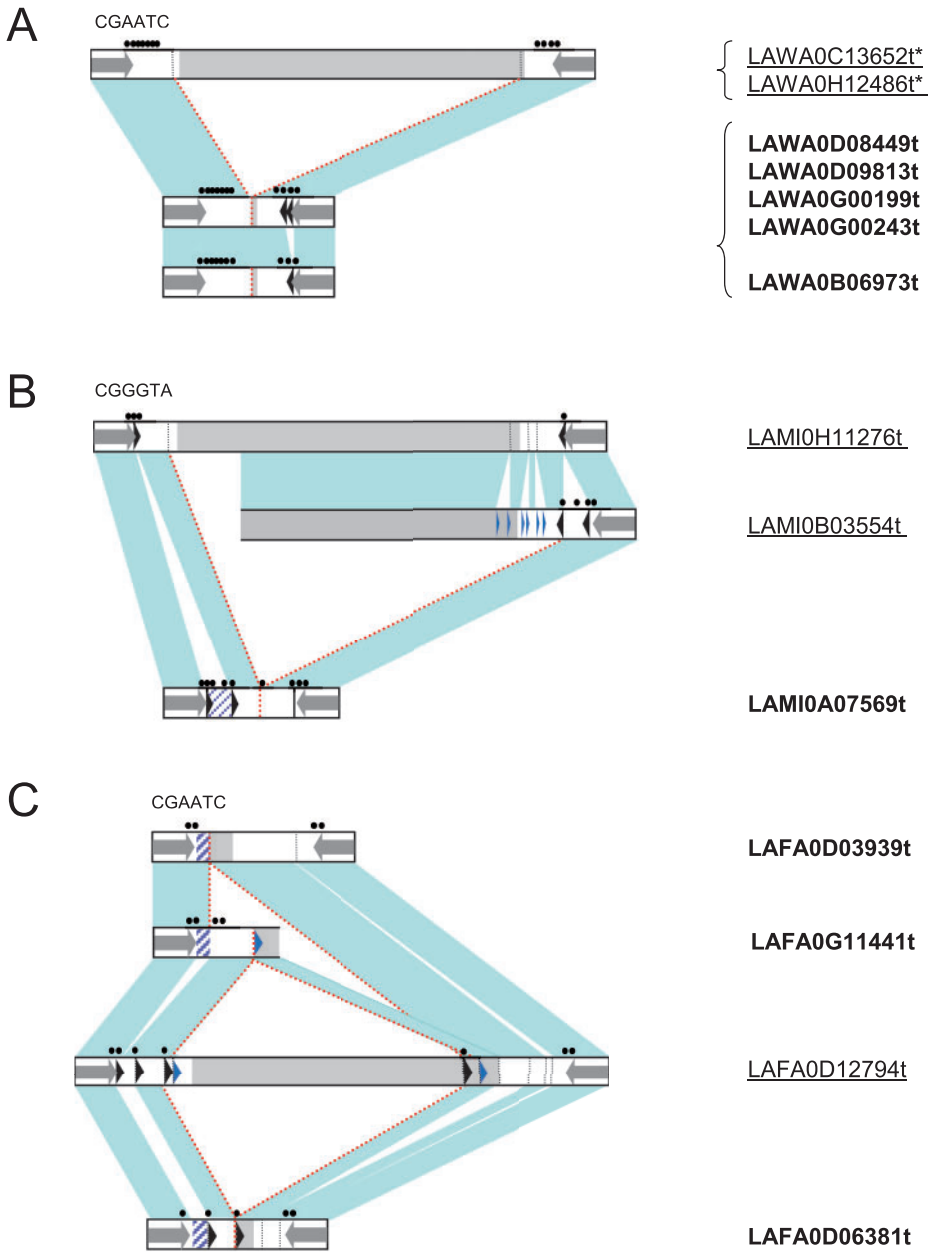

Fig. 6.—Relationships between miniature copies and *Rover1* segments. Miniature copies and *Rover* elements are represented as in figure 2. Gray arrows represent TIRs and shaded gray rectangles depict the transposase sequences. Sizes are not to scale. The blue areas between the elements are conserved homologous segments. Hatched rectangles correspond to additional segments. Thin dotted black lines indicate segment boundaries. Bold dotted red lines highlight junctions of *Rover* segments that resulted in miniature copy formation via internal deletion. Black dots indicate the positions of all the *Rover1*-specific hexameric repeats; their consensus sequences are provided just above the first repeat. Triangles represent microhomologies at junctions: In black for microhomologies involving hexameric repeats and in blue for microhomologies not related to hexameric repeats. Single nucleotide polymorphisms and indels shorter than 10 bp are not represented. The presence of a star after a *Rover* name indicates that it is a potentially autonomous *Rover1* copy.

residues of the C-terminal coding sequence were occasionally found (fig. 6). They have the same TIRs as the *Rover1* elements and similar subterminal segments where identity ranged from 76% to 96%. They have an average length of 356 bp, with the exception of one truncated copy that is 277-bp long and contains only one TIR (LAFA0G11441t). Seven copies have 8-bp TSDs that all differ from each other. These nonautonomous elements have typical DNA secondary structures

(supplementary fig. S3, Supplementary Material online), which may generate small noncoding RNAs involved in gene regulation, as reported for the *Nezha* MITE found in cyanobacteria (Zhou et al. 2008). If such miniature copies are subject to subsequent amplification, this process may give rise to a homogeneous group of elements: A MITE family (Feschotte et al. 2002; Yang et al. 2013). The five *L. waltii* miniature copies are highly conserved: They are collinear and present 93–100% nucleotide identity (fig. 6*A*). Furthermore, they all have different TSDs, which fits with the idea that they result from transposase-mediated amplification. Therefore, the *L. waltii* nonautonomous copies unambiguously constitute a MITE subfamily and have been called *mRover1* (fig. 2). On the basis of the sequence alignments, we found that the parental *Rover1* copy was related to one of the autonomous elements LAWA0C13652t or LAWA0H12486t. The parental element may have undergone an internal 3,077-bp deletion. No similarities have been found at the extremities of the deleted fragment, which indicates that the deletion resulted from a microhomology-independent mechanism. However, in its 3′ subterminal segment, MITE LAWA0B06973t is characterized by an additional 21-bp deletion that is flanked by 8-bp microhomologies (CGGATTCG). Interestingly, this sequence contains a GATTCG motif that is identical to the last six nucleotides of the TIRs. This hexameric repeat is found 16 times (6 forward plus 10 reverse) in the MITEs and in the 200-bp subterminal extremities of the autonomous copies. The three miniature elements present in *L. fantastica* appear to result from single deletion events. They all present different breakpoint junctions that are the result of different internal deletions in the autonomous copies (fig. 6*C*). Microhomology sequences 5 ([T/A]TTGA) and 6 bp ([C/A]GAA[T/A]C) in length are detected at the LAFA0G11441t and LAFA0D06381t breakpoints, respectively. The 6-bp microhomology at the LAFA0D06381t breakpoint corresponds to the hexameric motif CGAATC also present in *L. waltii Rover1* and MITEs. This motif occurs in four copies in the *L. fantastica* elements.

Finally, the single LAMI0A07569t miniature copy has homologies with the segment 1–135 of the full-length element LAMI0H11276t and the segment 2,385–2,484 of the 5′-truncated element LAMI0B03554t (fig. 6*B*). It also contains an additional 77-bp sequence located just after the 5′ TIR, which is absent from LAMI0H11276t. As for the *L. waltii* and *L. fantastica Rover1*-related elements, in their subterminal regions, the *L. mirantina Rover1* elements contain 11 forward and reverse repeats of a particular hexameric sequence of consensus motif ([C/T]G[G/A]GTA), corresponding to the end of the TIR. Interestingly, such repeats flank the additional 77-bp segment in the *L. mirantina* miniature copy as well as the 81-bp segment shared with LAMI0B03554t. This finding suggests once again that these particular hexameric repeats served as microhomologies.

To investigate the activity of *L. waltii mRover1* at the species level, we compared the insertion patterns of the five MITEs found in *L. waltii* strain CBS 6430 with those observed by PCR in the eight other *L. waltii* strains available from international culture collections. Two different patterns were seen: Strains CBS 6430 and CBS 7703 both contained five MITEs, whereas all other strains do not have MITEs at three loci and did not amplify at the two other MITE loci (supplementary fig. S4, Supplementary Material online). The lack of PCR products may be due either to a highly divergent intergenic region that prevents the primers from hybridizing or to rearrangement within the region. The insertion polymorphism between nine strains indicates that *mRover1* MITEs underwent new insertions and/or deletions after *L. waltii* speciation.

Rover and *Roamer* Element Insertion Sites Found at Synteny Breakpoints

We investigated the possibility that *Rover* and *Roamer* elements could be involved in structural rearrangements in chromosomes. To this end, the locations of each element were determined and compared with the pairwise synteny breakpoints mapped with SynChro (Drillon et al. 2014). Indeed, the high quality of *Lachancea* genome assemblies achieved through high coverage sequencing and the use of paired reads (separated by 8 kb) allowed us to circumvent assembly errors. Of the 143 *Rover* copies found in *Lachancea* species, 86 were located in regions without any structural modifications of gene order. Thirty-six copies were located at synteny breakpoints between at least two species—corresponding to 22 different events of either interchromosomal translocations or intrachromosomal inversions. Twenty-one other copies were located in regions with local synteny modifications due to insertions of few genes. Thus, nearly 40% of the *Rover* copies appeared to have been involved in short- or large-scale chromosomal rearrangements in *Lachancea* species; the corresponding events have been situated along the *Lachancea* phylogenetic tree on figure 1*B*. These copies include 3 miniature copies, 3 potentially autonomous copies, 9 nonautonomous copies, 29 relic copies, and 13 transposases from the five families. Even in *Lachancea cidri*, where only one relic copy still exists, the copy is located at the extremity of a synteny block shared with *Lachancea fermentati*, *L. cidri*'s closest relative. In addition, among the 46 copies containing 2 TIRs, 9 copies lack TSDs. The absence of a TSD may indicate a relatively old insertion or may be the signature of a recombination event between two copies. A convincing example is provided by copies LAMI0F15368t and LAMI0F15632t. The 5′ TSD of LAMI0F15368t is identical to the 3′ TSD of LAMI0F15632t and vice versa (supplementary table S4, Supplementary Material online). Gene order comparisons among *L. mirantina* and the nine other *Lachancea* species showed that the region between LAMI0F15368t and LAMI0F15632t has been reverted in

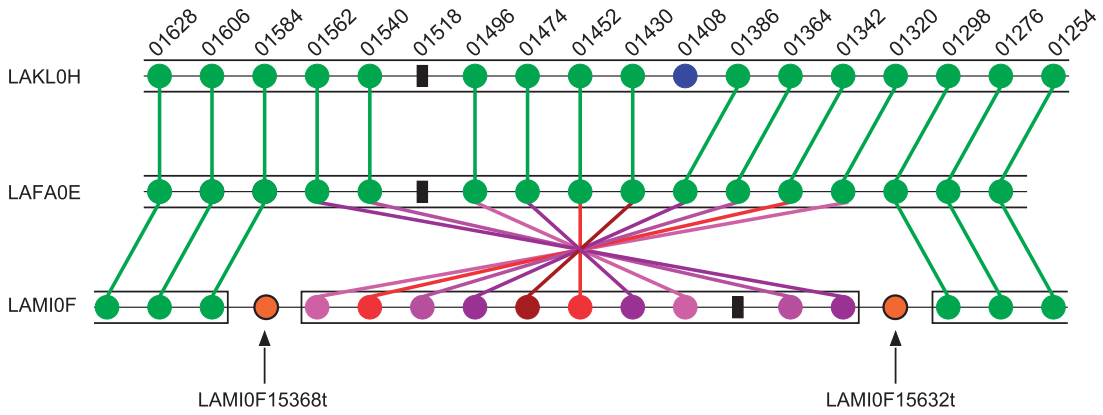

Fig. 7.—Intrachromosomal rearrangements between two *Rover2* copies. A representation of synteny conservation among *L. kluyveri* (LAKL), *L. fantastica* (LAFA) and *L. mirantina* (LAMI) was obtained using a screenshot of the SynChro synteny map output. The horizontal lines depict chromosomes and the vertical lines depict orthology relationships. The colored disks represent the CDSs: The green scale is for blocks with the same orientation and the red scale is for blocks that are inverted relative to the centromere. The blue disks depict genes without any orthologs. The color scale conveys the pairwise degree of similarity: The darker the color is, the more similar the sequences are. The black rectangles depict tRNA genes and the thick black line circles filled with orange depict transposable elements. On the top, the genetic elements are pinpointed by their digit in *L. kluyveri* genome annotation.

L. mirantina (fig. 7). These copies run in opposite directions and have recombined; it is a case of an intrachromosomal rearrangement leading to an inversion. Twenty-one copies containing only a single TIR and consequently lacking an identified 8-bp TSD motif are also potentially involved in reciprocal translocations.

In the *Eremothecium* genus, the single copies found in each species—*E. gossypii*, *E. cymbalariae*, and *E. aceri*—are ancestral and surrounded by genes that demonstrate perfectly conserved synteny among the three species.

In *K. marxianus*, the two copies are located in regions where synteny has been conserved with *K. lactis*, which suggests that these copies have transposed after *K. marxianus* speciation and that they have not been involved in chromosomal rearrangements since. In contrast, in *K. lactis*, the two copies are located at the extremities of synteny blocks shared with *K. marxianus* (fig. 8). In this region, *K. marxianus* still shares the ancestral gene order with *L. kluyveri*; however, a rearrangement occurred in *K. lactis* at the precise location of the two copies. The presence of a degenerate TSD upstream from the 5′ TIR of the Klla0A copy (TTCAAAAAA) and downstream from the 3′ TIR of the Klla0F copy (GTCAAAAAG) confirms that a reciprocal translocation occurred within the *Rover* elements.

In *N. castellii*, 13 of the 18 *Roamer* copies are located at a synteny breakpoint with *N. dairenensis*. The others are located in regions where synteny has not been conserved. These findings illustrate the strong involvement of TEs in the dynamics of *N. castellii*. An example of recombination between elements can be seen in the two copies NCAS0A07480 and NCAS0F00700, as evidenced by the exchange of their TSDs (supplementary table S4, Supplementary Material online).

Discussion

More *hAT* Elements in Saccharomycetaceae than Previously Thought

Our systematic search for *hAT* transposases in the 41 sequenced Ascomycota yeast genomes revealed the existence of two groups that contain potentially active elements: *Roamer* and *Rover*. So far, these are the first members of the *hAT* superfamily to be fully described in Ascomycota yeasts. Our analysis was not exhaustive, given the extremely large number of genomes available. As *hAT* distribution within the Ascomycota yeast tree is very patchy, it is difficult to predict which of the genomes currently being sequenced will contain such elements. As observed in the case of *S. cerevisiae* strain AWRI1631, some individuals may acquire *hAT* elements while the rest of the population does not. As a consequence, population genomics will also contribute to a better understanding of TE propagation and dynamics.

Overall, our results show that class II TEs are likely more common than previously thought. They are not easy to identify during genome annotation because transposases are highly variable and therefore are hard to recognize using BLAST-based approaches alone. Our approach of systematically searching our species dataset using characteristic transposase domains and TIRs ended up being very powerful (supplementary table S3, Supplementary Material online). In addition to hits corresponding to *Rover* and *Roamer* copies, we obtained 52 hits corresponding to singletons or genes in families, in single genomes, or in several closely related species. Given their species distributions, some families are probably very ancient and represented by a few members containing relics of transposase functional domains. One of these families contains homologs of SAKL0H14366g, itself

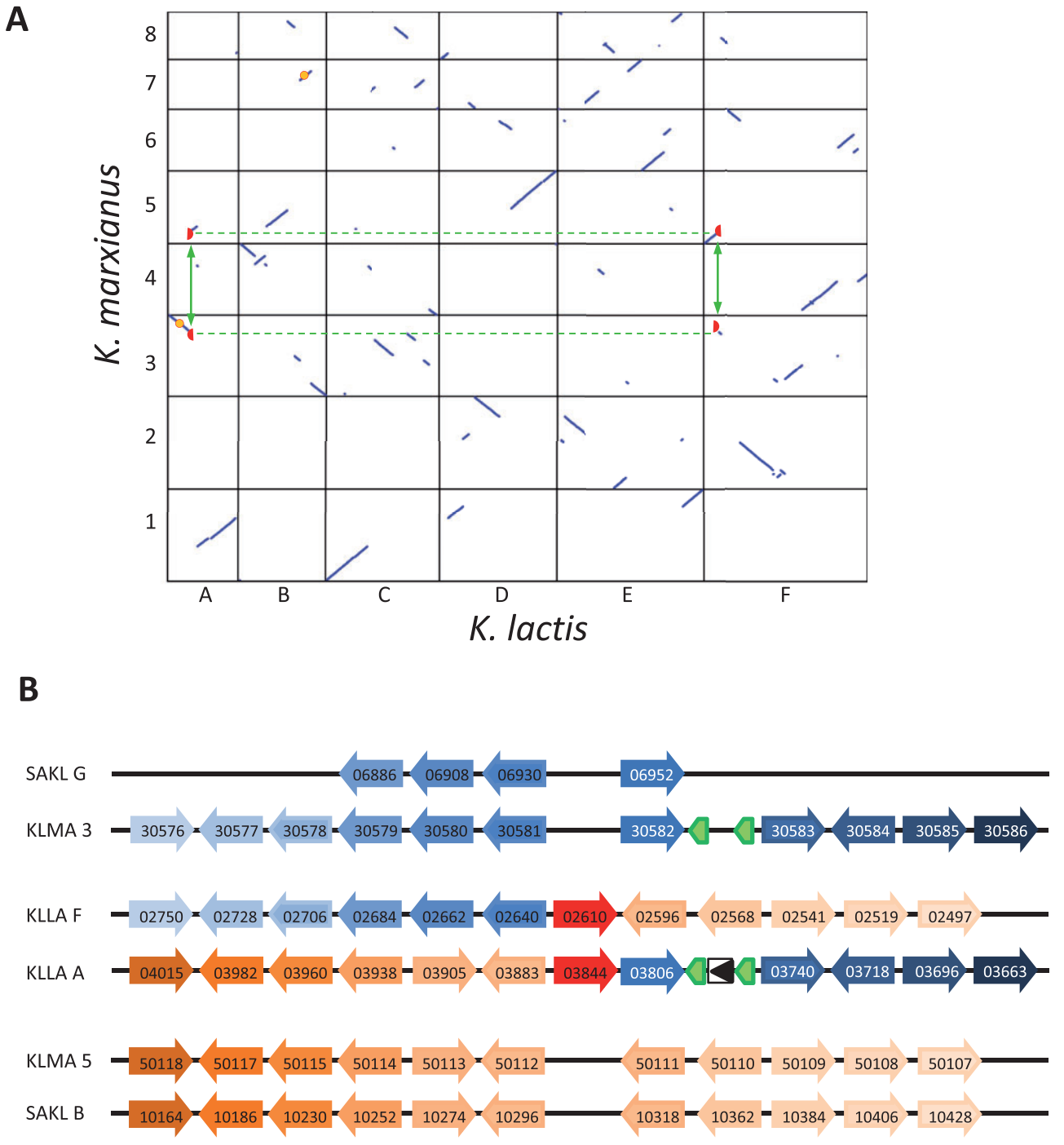

Fig. 8.—Chromosomal rearrangements between *Kluyveromyces lactis* and *Kluyveromyces marxianus*. (*A*) Dotplot of *K. lactis* and *K. marxianus* chromosomes, labeled A–F and 1–8, respectively. The two *Rover* copies found in *K. marxianus* are depicted as orange disks within a red circle. The two copies of *K. lactis* elements involved in chromosomal rearrangement are represented by red half-disks. The physical link between the two parts of each *K. lactis* element is represented by green double arrows and the link between the consecutive chromosomal regions of *K. marxianus* is represented by green dotted lines. (*B*) Synteny conservation around the insertion site of *K. lactis Rover* elements (in red). Chromosome names are on the left and gene names are placed within colored arrows. The *Rover* element on Klla0F, previously known as KLLA0F02568t and located opposite a CDS of 1,191 amino acids, was reannotated a few kb downstream and now defines a full-length element of 2,458 bp that contains perfect 17-bp TIRs and a degenerate transposase. *L. kluyveri* orthologous loci on chromosomes G (arrows—blue color gradient) and B (arrows—brown color gradient) were used as outgroups to assess the direction of the reciprocal translocation. Orthologous genes in the three species are represented with the same color. Green arrows correspond to the tRNA genes and the black triangle in KLLA-A corresponds to a single LTR.

homologous to *S. cerevisiae* RSF1. In a previous study, homologs of *RSF1* found in *Vanderwaltozyma polyspora, N. castellii*, and *L. kluyveri* were classified as domesticated *hAT*-like elements (Gordon et al. 2009). However, members of this *RSF1* family are weakly conserved among species and most of them contain neither the C-terminal dimerization domain characteristic of the *hAT* superfamily (PF05699) nor the BED finger DNA-binding motif PF02892 common to different class II elements (Aravind 2000; Essers et al. 2000). Similarly, YGR071C and its homologs in Saccharomycetaceae might derive from an ancestral *hAT* element, but distinct from the *RSF1 hAT* ancestor. Thus, if all the families and singletons that we detected actually arose from *hAT* elements, they would have come from different *hAT* ancestors, probably acquired independently, and would have diverged until they had lost almost all their *hAT* signatures. These examples illustrate the importance of using Pfam domain prediction or similar approaches to track TEs or their relics and, consequently, to improve our knowledge of TE dynamics in yeasts.

hAT Elements May Have Been Acquired by Horizontal Transfer

In the debate surrounding the evolutionary dynamics of transposons, two main models have been developed to explain the existence of multiple transposon lineages or the emergence of a new subfamily: The vertical diversification model and the HT model. These models are not mutually exclusive. The first model requires that at least one transposon copy remains autonomous until a new amplification burst occurs; otherwise, the transposon family will go extinct. The second model can also explain amplification bursts, but its mechanism results in phylogenetic incongruence between the TE tree and the species tree. Using a large number of yeast taxa, we examined *Rover* and *Roamer* phylogenetic trees, synteny conservation, and similarity in order to reconstruct evolutionary scenarios for both elements. We arrived at the conclusion that the most parsimonious explanation for the presence and distribution of these TE families in Saccharomycetaceae involves five HTs. HTs of TEs are frequent and widespread in some taxonomic groups, such as plants (El Baidouri et al. 2014), fish (Jiang et al. 2012), and insects (Sánchez-Gracia et al. 2005; Zhang et al. 2013). Even if TEs are relatively rare in hemiascomycetes, this is not the first time that their acquisition by HT has been proposed (Carr et al. 2012).

However, for the primary transfers, the question of the donor species is still unresolved. The presence of *hAT* elements has never been observed in other hemiascomycetes, but *Rover* elements share conserved domains with *hAT* elements found in Pezizomycotina, such as *Tfo1* and *Restless* (supplementary fig. S2, Supplementary Material online). Therefore, yeast species of the LaKlEr clade probably acquired a *Rover* element from a Pezizomycotina species, although we cannot exclude that a *Rover* element was already present in the common ancestor for the ascomycetes but was lost in all the Saccharomycotina lineages except for the LaKlEr clade, where the genome content would have allowed its maintenance and diversification.

Rover copies present in *N. dairenensis* and in *S. cerevisiae* strain AWRI1631 have transposases that are not flanked by TIRs, which indicates an ancient acquisition event or the recent HT of transposases devoid of TIRs. In *N. dairenensis*, much remains unknown about the HT mechanism and the transposon acquisition date. The *Rover* copy is inserted in a region that is highly conserved in synteny with *N. castellii*, its closest relative, which does not contain a *Rover* element; this finding suggests either that acquisition took place after the two species diverged or that *Rover* has been completely lost from the *N. castellii* genome. Sequencing additional strains of *N. dairenensis* may yield clues that will clarify this HT event. In *S. cerevisiae*, at least three copies are present in the AWRI1631 genome but have not been detected in any other *S. cerevisiae* strains thus far, which may indicate that this strain experienced a single, recent acquisition event. This acquisition event probably involved a DNA segment that was longer than the *Rover* element itself, as the *Rover* copies are flanked by genes only found in Pezizomycotina and CTG species. The presence of these CDSs in AWRI1631 therefore suggests that the HT mechanism could have involved introgressive hybridization, the authors having excluded the possibility of laboratory contamination (Borneman et al. 2008). Introgressions into a host genome may explain why acquisitions can involve large DNA segments; for instance, the genome of *S. cerevisiae* EC1118 contains a 65-kb fragment from a non-*Saccharomyces* species (Novo et al. 2009). Such events are particularly well documented for many *Saccharomyces sensu stricto* species (Morales and Dujon 2012; Dunn et al. 2013).

Similarly, we hypothesize that *Roamer*, which must still be active in *N. castellii*, spread by HT from this species to the ancestor of *Kluyveromyces* genus, which is devoid of autonomous copies. This includes the sister species *K. lactis* and *K. marxianus* (fig. 1*A*), and also *Kluyveromyces aestuarii*, *Kluyveromyces nonfermentans*, *Kluyveromyces wickerhamii*, *and Kluyveromyces dobzhanskii*, where remnant copies of *Roamer* have been recently detected as TIR-free transposases (Rajaei et al. 2014).

Overall, these proposed examples of HT reveal that such events have occurred more frequently during the evolution of hemiascomycetes than previously thought. However, it remains difficult to identify the donor species involved in the initial acquisition event, despite the development of additional genomic resources. Indeed, the full species diversity of yeast will probably never be characterized and in addition, the donor species may have evolved and lost its TE copies, leaving little chances for this issue to be resolved in the future.

Dynamics of *Rover* Elements in Preduplicated Species

Once a TE has been acquired by HT, its life cycle is comprised of transposition events, which may generate duplicated copies that are inserted into new sites in the genome. These copies then accumulate mutations independently over time, causing them to diverge. This divergence may result in diversification and the formation of new families or produce defective copies such as miniature copies, TIR-free transposases, or relics (Kidwell and Lisch 2001).

It has been proposed that class II TE activity can lead to the formation of internal deletion derivatives. This affects the entire, or part of the, transposase coding sequence while retaining TIRs. Some of these elements, which are shorter than 600 bp and nonautonomous, may subsequently be amplified, giving rise to a group of homogeneous elements called MITEs (Feschotte et al. 2002). Nonautonomous copies such as MITEs need the transposase that is produced by an autonomous copy in order to transpose, and must therefore share identical TIRs with the autonomous copy (Bergemann et al. 2008). MITEs have been described in a wide variety of organisms, including plants and animals such as *Drosophila*, mosquitoes, fish, and humans (Deprá et al. 2012; Fattash et al. 2013). Numerous MITE families have also been observed in fungi and are sometimes classified based on their class II progenitor. For instance, MITEs have been found in the microsporidium *Nosema bombycis* (Xu et al. 2010) and in Pezizomycotina species, such as *Neurospora crassa* (Yeadon and Catcheside 1995), *Botrytis cinerea* (Amselem et al. 2011), *Sclerotinia sclerotiorum* (Amselem et al. 2011), *Epichloë festucae* (Fleetwood et al. 2011), *Penicillium digitatum* (Sun et al. 2013), and *Fusarium oxysporum* (Bergemann et al. 2008). In this latter species, MITEs can be mobilized by autonomous elements that share identical TIRs (Dufresne et al. 2007; Bergemann et al. 2008). However, so far *mRover1* is the first case of MITE reported in yeasts. Five *mRover1* MITEs have been identified in *L. waltii*, where they coexist with two autonomous *Rover1* copies. MITEs derived from *hAT* elements with autonomous copies in the same genome have so far been reported only in plants (Saito et al. 2005; Benjak et al. 2009) and animals (Pace et al. 2008; Zhang et al. 2013). Different molecular mechanisms have been proposed to explain the formation of miniature copies by internal deletions; recombination between microhomologies is often evoked (Negoua et al. 2013; Yang et al. 2013). Such mechanism may also occur in yeasts, as hexameric microhomologies have been characterized in *Rover1*. The *Rover1* microhomologies are unique among *Rover* elements. Indeed, only *Rover1* copies 1) contain one tandem copy of the 6-bp TIR terminal motif just 0–3 bp downstream from the 5′ TIR and upstream of the 3′ TIR and 2) have several additional forward and reverse copies of these repeats in their 200-bp-long subterminal regions (up to 16 repeats). It is therefore tempting to hypothesize that, just as *Mariner* elements have a greater tendency to form MITEs due to their AT-rich subterminal sequences (Negoua et al. 2013), the *Rover1* family is more likely to give rise to miniature copies than the *Rover2*, *3*, and *4* families.

In *Lachancea* species, we observed every step of the TE life cycle described above. In *L. fantastica, Rover* has undergone several amplification bursts. In *L. waltii*, we can see how internal deletion derivatives could lead to MITE formation if amplification occurred. *Lachancea fermentati* contains only a putative functional transposase that is 826-aa long, which illustrates a situation in which the activity of *Rover* elements has been so effectively controlled that only a putative co-opted transposase remains. Finally, *L. cidri* contains only a relic copy: Its genome has removed all the functional copies of *Rover*. Overall, the diversified *Rover* families, the relatively high number of *Rover* elements (including MITEs), and the presence of autonomous copies suggest that these elements have been particularly active in the whole clade until recently (fig. 1*B*). Given the distribution of the five *Rover* families in *Lachancea* species, most of their diversification probably occurred in the clade's ancestor or just after the divergence of *L. kluyveri*. In *L. nothofagi, L. waltii, L. mirantina*, and *L. fantastica*, both autonomous and nonautonomous copies coexist in the genome. In *L. thermotolerans*, one nonautonomous *Rover1* copy with two identical TIRs and a conserved TSD, KLTH0H13266t, seems to have been recently transmobilized by the only *Rover1* transposase copy, KLTH0B06666t, which is a TIR-free transposase. This implies that TIR-free transposase are able to produce efficient proteins that will mediate the transposition of nonautonomous elements of the same family. In contrast, in *L. dairenensis*, *L. cidri*, *L. fermentati*, *L. meyersii* and *L. kluyveri*, only TIR-free transposases and/or relic copies remain. Thus, *Rover* can no longer transpose itself in these species. The different invasion states observed across the different *Lachancea* species probably reflect different elimination and transposition rates (Le Rouzic et al. 2013).

A possible fate of TEs is to be recruited by host genome, providing useful cellular functions, as has been observed in some plants and animals (Volff 2006; Arensburger et al. 2011; Alzohairy et al. 2013). Co-opted TEs are characterized by an intact coding sequence, missing transposition sequences, and a presence at orthologous loci across different organisms; they are also submitted to a purifying selection pressure (Alzohairy et al. 2013). Several TIR-free *Rover* and *Roamer* transposases fulfill the three first criteria, which suggest that they may have been co-opted by host genome. A recent study demonstrated that *Roamer* transposase (named KAT1) in the *Kluyveromyces* clade was a domesticated transposase-derived endonuclease responsible for the sexual differentiation of the host (Rajaei et al. 2014). The frameshift disrupting the coding sequence is actually a way for the host to regulate the transposase expression by programmed ribosomal frameshifting. For *Rover*, the function of the putatively co-opted transposases has not been assessed, but might involve protein domains that are useful for cells: Domains that

mediate DNA binding, help regulate gene transcription, or aid in chromatin organization (Feschotte 2008; Sinzelle et al. 2009).

Besides sharing the ability to move through the genome and to amplify themselves, autonomous, nonautonomous, relic, and TIR-free transposon copies may also be involved in chromosomal rearrangements.

Rover and *Roamer* Elements Are Involved in Chromosomal Rearrangements

Chromosomal rearrangements have previously been associated with transfer RNA genes and LTR–retrotransposon insertion sites in *S. cerevisiae* and related species (Fischer et al. 2000). TEs lead to chromosomal rearrangements in several ways. First, as repeated sequences, they can easily be used as templates for ectopic recombination. This can then lead to reciprocal translocation, deletion of the region between two copies with the same orientation, or inversion of the region between two copies oriented in opposite directions. Second, the "alternative transposition" mechanism described by Gray (2000) may result in the pairing of the terminal repeats from two homologous copies. This pairing leads to the formation of a hybrid element that can excise itself and then may end up reinserting itself at a new target site. Depending on the location of the target site and the orientation of the reinserted excised hybrid elements, this alternative form of transposition can lead to different chromosomal rearrangement configurations (Gray 2000).

Considering the literature and the undeniable examples of TE-induced chromosomal rearrangements we observed, we hypothesized that *Roamer* and *Rover* elements located at synteny breakpoints have been involved in structural rearrangements, even if we cannot exclude that a negligible number of them may be a result of independent and later integration. Our results show that a large part of *Roamer* and *Rover* is concerned, whether they be full length or miniature copies. Chromosomal rearrangements can have functional and evolutionary effects by modifying gene content and collinearity. The potential for deleterious effects, which increases with increasing copy number, may explain why genomes tend to remove TEs, as we observed in *L. cidri* and *L. fermentati*. In addition, because transposons contribute to divergence among homologous chromosomes in populations, they can promote phenomena such as reproductive isolation and speciation, as has been seen in experimental studies of evolution (Hou et al. 2014). In Saccharomycetaceae, *hAT* elements thus appear to be a driving force in genome remodeling; however, their involvement in speciation remains to be evaluated.

Supplementary Material

Supplementary figures S1–S4, table S1, and files S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was funded by the ANR project 2010 BLAN1606. We thank the GB-3G consortium for generating fruitful discussions about *Lachancea* genomics and anonymous reviewers for their constructive comments on the manuscript.

Literature Cited

Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A. 2013. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid 69:1–15.

Amselem J, et al. 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 7: e1002230.

Aravind L. 2000. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem. Sci. 25:421–423.

Arensburger P, et al. 2011. Phylogenetic and functional characterization of the *hAT* transposon superfamily. Genetics 188:45–57.

Benjak A, Boué S, Forneck A, Casacuberta JM. 2009. Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera L.*). Genome Biol Evol. 1:75–84.

Bergemann M, Lespinet O, M'Barek SB, Daboussi MJ, Dufresne M. 2008. Genome-wide analysis of the *Fusarium oxysporum mimp* family of MITEs and mobilization of both native and de novo created *mimps*. J Mol Evol. 67:631–642.

Bigot Y, Augé-Gouillou C, Periquet G. 1996. Computer analyses reveal a *hobo*-like element in the nematode *Caenorhabditis elegans*, which presents a conserved transposase domain common with the *Tc1-Mariner* transposon family. Gene 174:265–271.

Bleykasten-Grosshans C, Neuvéglise C. 2011. Transposable elements in yeasts. C R Biol. 334:679–686.

Borneman AR, Forgan AH, Pretorius IS, Chambers PJ. 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. FEMS Yeast Res. 8:1185–1195.

Cameron JR, Loh EY, Davis RW. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. Cell 16:739–751.

Carr M, Bensasson D, Bergman CM. 2012. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. PLoS One 7: e50978.

Casaregola S, Neuvéglise C, Bon E, Gaillardin C. 2002. *Ylli*, a non-LTR retrotransposon *L1* family in the dimorphic yeast *Yarrowia lipolytica*. Mol Biol Evol. 19:664–677.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17: 540–552.

Chen JY, Fonzi WA. 1992. A temperature-regulated, retrotransposon-like element from *Candida albicans*. J Bacteriol. 174:5624–5632.

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16:10881–10890.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Deprá M, Ludwig A, Valente VL, Loreto EL. 2012. *Mar*, a MITE family of *hAT* transposons in *Drosophila*. Mob DNA. 3:13.

Drillon G, Carbone A, Fischer G. 2014. SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. PLoS One 9:e92621.

Dufresne M, et al. 2007. Transposition of a fungal miniature inverted-repeat transposable element through the action of a *Tc1*-like transposase. Genetics 175:441–452.

Dunn B, et al. 2013. Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. PLoS Genet. 9:e1003366.
El Baidouri M, et al. 2014. Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res. 24:831–838.
Essers L, Adolphs RH, Kunze R. 2000. A highly conserved domain of the maize activator transposase is involved in dimerization. Plant Cell 12: 211–224.
Fattash I, et al. 2013. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. Genome 56:475–486.
Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 9:397–405.
Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: where genetics meets genomics. Nat Rev Genet. 3:329–341.
Finn RD, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222–D230.
Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. 2000. Chromosomal evolution in *Saccharomyces*. Nature 405:451–454.
Fleetwood DJ, et al. 2011. Abundant degenerate miniature inverted-repeat transposable elements in genomes of epichloid fungal endophytes of grasses. Genome Biol Evol. 3:1253–1264.
Goodwin TJ, Ormandy JE, Poulter RT. 2001. *L1*-like non-LTR retrotransposons in the yeast *Candida albicans*. Curr Genet. 39:83–91.
Goodwin TJ, Poulter RT. 2000. Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res. 10:174–191.
Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet. 5:e1000485.
Gray YH. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. 16:461–468.
Greer H, Fink GR. 1979. Unstable transpositions of *his4* in yeast. Proc Natl Acad Sci U S A. 76:4006–4010.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.
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. 24:1153–1159.
Jiang XY, et al. 2012. Goldfish transposase *Tgf2* presumably from recent horizontal transfer is active. FASEB J. 26:2743–2752.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8: 275–282.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Kempken F, Windhofer F. 2001. The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma 110:1–9.
Kidwell MG, Lisch DR. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55:1–24.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol. 25:1307–1320.
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.
Levin HL, Weaver DC, Boeke JD. 1990. Two related families of retrotransposons from *Schizosaccharomyces pombe*. Mol Cell Biol. 10:6791–6798.
Loreto ELS, Carareto CMA, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100:545–554.
Morales L, Dujon B. 2012. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. Microbiol Mol Biol Rev. 76:721–739.
Myers EW, et al. 2000. A whole-genome assembly of *Drosophila*. Science 287:2196–2204.
Negoua A, Rouault JD, Chakir M, Capy P. 2013. Internal deletions of transposable elements: the case of *Lemi* elements. Genetica 141:369–379.
Neuvéglise C, Chalvet F, Wincker P, Gaillardin C, Casaregola S. 2005. *Mutator*-like element in the yeast *Yarrowia lipolytica* displays multiple alternative splicings. Eukaryot Cell. 4:615–624.
Neuvéglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S. 2002. Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res. 12:930–943.
Nicholas KB, Nicholas HB, Deerfield DW II. 1997. GeneDoc: analysis and visualization of genetic variation, EMBNEW.NEWS 4:14.
Novo M, et al. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci U S A. 106:16333–16338.
Oliver KR, McComb JA, Greene WK. 2013. Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol. 5:1886–1901.
Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A. 105:17023–17028.
Rajaei N, Chiruvella KK, Lin F, Aström SU. 2014. Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. Proc Natl Acad Sci U S A. 111:15491–15496.
R Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.
Roeder GS, Fink GR. 1980. DNA rearrangements associated with a transposable element in yeast. Cell 21:239–249.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Rossato DO, et al. 2014. *BuT2* is a member of the third major group of *hAT* transposons and is involved in horizontal transfer events in the genus *Drosophila*. Genome Biol Evol. 6:352–365.
Rubin E, Lithwick G, Levy AA. 2001. Structure and evolution of the *hAT* transposon superfamily. Genetics 158:949–957.
Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944–945.
Saito M, Yonemaru J, Ishikawa G, Nakamura T. 2005. A candidate autonomous version of the wheat MITE *Hikkoshi* is present in the rice genome. Mol Genet Genomics. 273:404–414.
Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. Trends Genet. 21: 200–203.
SantaLucia J Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A. 95:1460–1465.
Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25:537–546.
Schmid-Berger N, Schmid B, Barth G. 1994. *Ylt1*, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. J Bacteriol. 176:2477–2482.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.
Sinzelle L, Izsvák Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci. 66:1073–1093.
Souciet JL, et al. 2009. Comparative genomics of protoploid *Saccharomycetaceae*. Genome Res. 19:1696–1709.
Sun X, et al. 2013. Genomewide investigation into DNA elements and ABC transporters involved in imazalil resistance in *Penicillium digitatum*. FEMS Microbiol Lett. 348:11–18.
Vitte C, Fustier MA, Alix K, Tenaillon MI. 2014. The bright side of transposons in crop evolution. Brief Funct Genomics. 13:276–295.

Volff JN. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28:913–922.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Xu J, et al. 2010. Identification of *NbME* MITE families: potential molecular markers in the microsporidia *Nosema bombycis*. J Invertebr Pathol. 103:48–52.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang G, Fattash I, Lee CN, Liu K, Cavinder B. 2013. Birth of three *stowaway*-like MITE families via microhomology-mediated miniaturization of a *Tc1/Mariner* element in the yellow fever mosquito. Genome Biol Evol. 5:1937–1948.
Yeadon PJ, Catcheside DE. 1995. *Guest:* a 98 bp inverted repeat transposable element in *Neurospora crassa*. Mol Gen Genet. 247:105–109.
Zhang HH, et al. 2013. A novel *hAT* element in *Bombyx mori* and *Rhodnius prolixus:* its relationship with miniature inverted repeat transposable elements (MITEs) and horizontal transfer. Insect Mol Biol. 22: 584–596.
Zhou F, Tran T, Xu Y. 2008. *Nezha*, a novel active miniature inverted-repeat transposable element in cyanobacteria. Biochem Biophys Res Commun. 365:790–794.
Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.

Associate editor: Esther Betran