DNA transposon expansion drives genome plasticity in *Diutina catenulata*

Supplementary Information

Frédéric BIGEY[1], Marc WESSNER[2], Martine PRADAL[1], Corinne CRUAUD[3], Jean-Marc AURY[2], and Cécile NEUVEGLISE[1,*]

[1]SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France
[2]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France
[3]Genoscope, Institut François Jacob, CEA, Univ Evry, Université Paris-Saclay, Evry, France
[*]corresponding author

June 1, 2026

Contents

1 Supplementary Results

1.1 New genomic resources for *D. catenulata*

Five haploid *D. catenulata* strains (CBS 565^{T}, 43B2s-11, CLIB 1179, 25B2s-01 and 1F1c-01) were selected for genome sequencing and assembly. These strains are representative of the clades identified during the phylogenetic analysis (see manuscript Figure 1). A hybrid sequencing strategy was employed, in which short Illumina reads were coupled with long Oxford Nanopore Technologies (ONT) reads (see manuscript, Methods section). The *D. catenulata* assembled nuclear genome ranges in size from 14.0 Mb ($n = 4$) to 14.7 Mb ($n = 1$; CBS 565) and comprises six near-complete chromosomes ranging in size from 0.8-1.2 Mb for the shortest, to 3.5-4.2 Mb for the largest (Supplementary Data 3), which is compatible with the sizes observed in the karyotypes (see manuscript Figure 4). Telomere repeats of 26 bp were identified at the extremities of some contigs, thereby confirming that at least two telomere-to-telomere chromosomes were obtained for each strain. The rDNA loci were located at three contig extremities that lacked telomere repeats. The 53.5% GC-content nuclear genomes of *D. catenulata* encode approximately 6,300 CDSs. The BUSCO score was found to be 98.1% on average after the annotation was manually curated. In addition to protein-coding genes, approximately 300 tRNA genes were predicted. Summary statistics are presented in Supplementary Data 3.

1.2 Analysis of the mitochondrial genomes

Complete circular mitochondrial genomes (Supplementary Data 3), ranging from 20,934 to 22,153 bp (average GC content is 25.8%), were identified in *D. catenulata* strains CBS 565 (accession: OZ261072), 43B2s-11 (accession: OZ261100) and 25B2s-01 (accession: OZ262066). The annotation revealed 15 CDS and 25 tRNA, all of which are in the same orientation. A putative sequence of an intron-encoded DNA homing endonuclease with a LAGLIDADG domain was found in the gene coding NADH dehydrogenase subunit 5 (nad5) in strains 43B2s-11 and 25B2s-01 only.

1.3 Analysis of antifungal resistance and virulence genes

We surveyed genes associated with antifungal resistance and virulence in pathogenic yeasts (Supplementary Data 4). Core drug targets, including *ERG11* (lanosterol 14-α-demethylase, azoles) and *FKS1* (β-1,3-glucan synthase, echinocandins), are present as single-copy genes. Their presence alone does not confer resistance; specific mutations are required that alter drug binding or target activity. In *Candida albicans*, amino acid substitutions in *ERG11* have been shown to increase azole resistance by reducing drug binding [1]. Similarly, in clinical *Candida auris* isolates, mutations in *ERG11* correlate with fluconazole resistance, and mutations in *FKS1* hot-spot regions are linked to echinocandin resistance [2]. Azole resistance can also arise from copy-number variations (CNVs) in *Candida parapsilosis* [3], and reviews of fungal CNVs highlight that increased gene dosage of efflux pumps and their regulators can enhance drug tolerance [4].

Effector genes mediating multidrug resistance are amplified: *CDR1* occurs in three copies in all strains, *MDR1* in three to four copies, whereas *CDR2* is absent. The majority of regulatory genes are single-copy: *TAC1*, controlling *CDR1/2* and *ERG11* expression, is conserved, whereas *MRR1*, regulating *CDR1/2* and *MDR1*, is duplicated in two cheese strains (43B2s-11, 25B2s-01). Other transcription factors and regulators involved in morphogenesis and filamentous growth (*CPH1*, *EFG1*, *RBF1*, *TUP1*) are primarily single-copy, reflecting conserved essential functions.

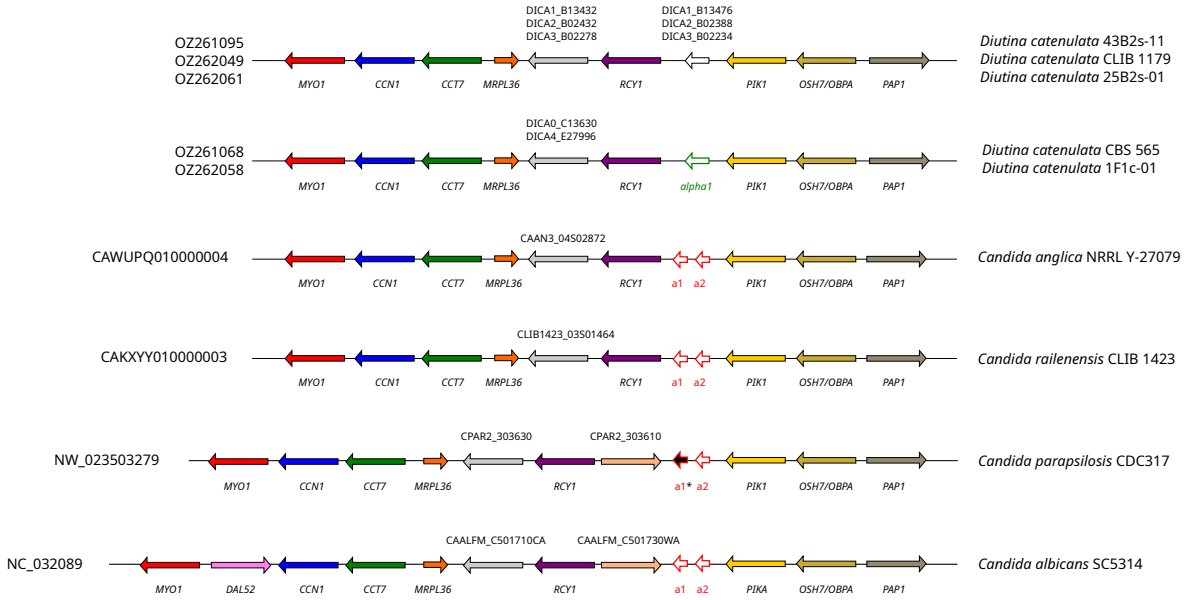
Genes associated with host–pathogen interactions also show CNVs. *PLB1*, encoding phospholipase

B1, is present in three copies. Members of the secreted aspartyl protease (SAP) family, which degrade host proteins, exhibit copy-number variation, with *SAP5* ranging from three to five copies. In *C. albicans*, expansion and expression of *SAP* genes have been linked to enhanced host tissue invasion and virulence [5].

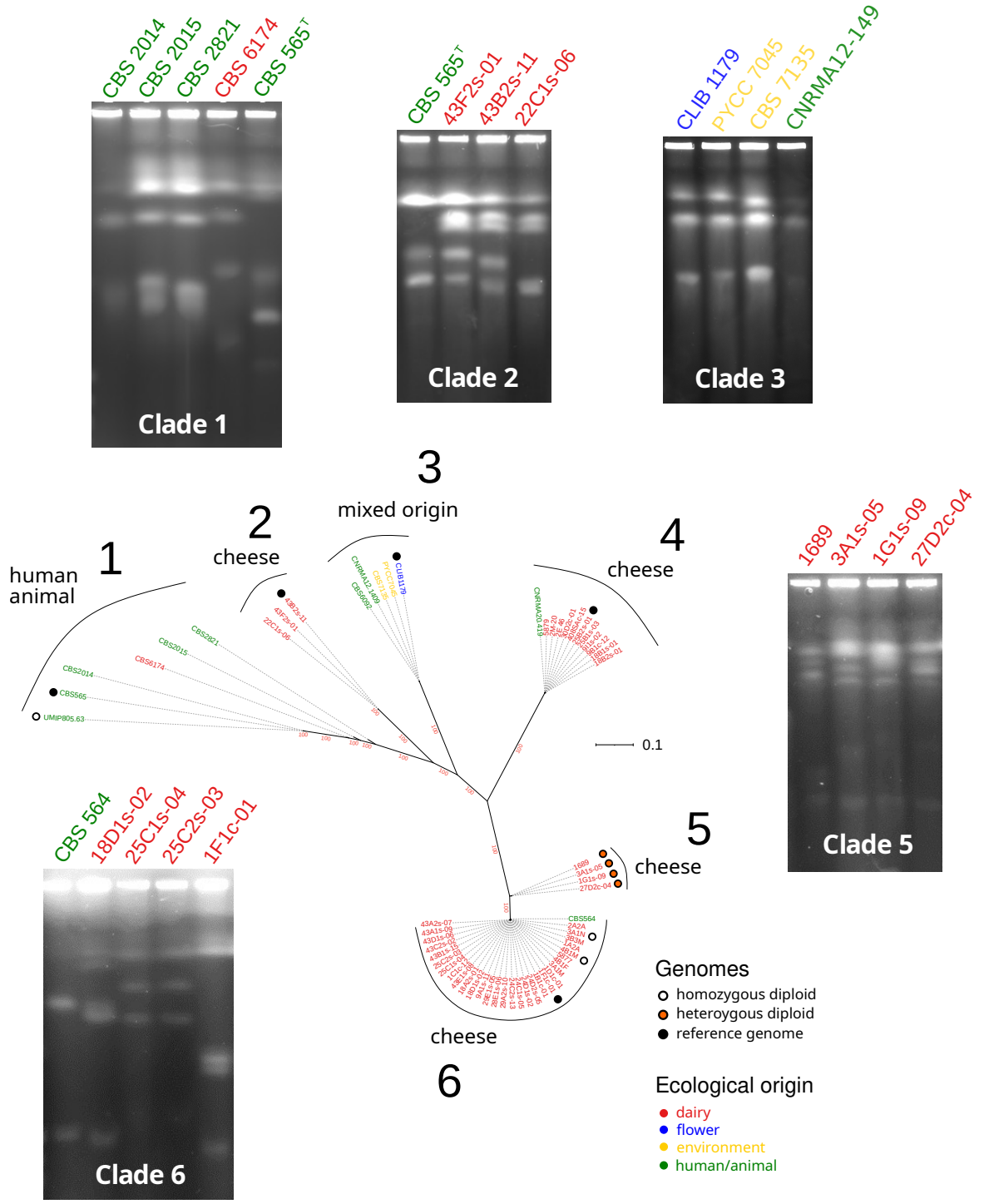
The distribution of resistance and virulence-associated genes does not differ between the clinical isolate CBS 565 and the cheese-derived strains, suggesting that the genetic potential for antifungal resistance and pathogenicity is conserved in environmental and clinical isolates. Overall, these patterns indicate that while the regulatory backbone of resistance and virulence remains predominantly single-copy, the amplification of effector genes — particularly multidrug transporters and hydrolytic enzymes — may enhance the adaptive potential of *D. catenulata*.

2 Supplementary Figures

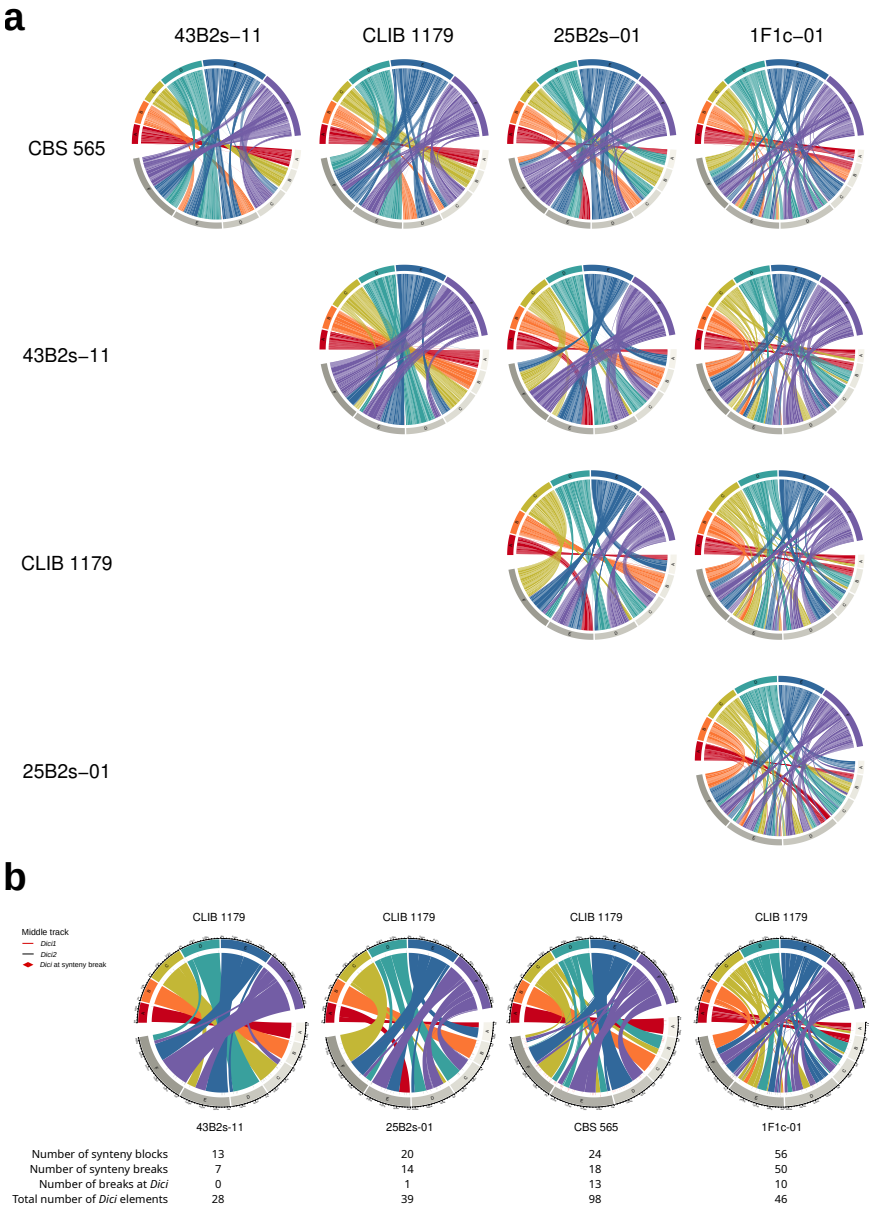
- **Supplementary Figure 1**
  Structure of the mating-type like (MTL) locus of *D. catenulata*.
- **Supplementary Figure 2**
  Karyotype pattern variation within *D. catenulata* species.
- **Supplementary Figure 3**
  Pairwise genome comparisons among five *D. catenulata* strains.
- **Supplemental Figure 4**
  Example of complex chromosomal rearrangements between two strains of *D. catenulata*.
- **Supplementary Figure 5**
  Expression of the three *Dici* transposon genes.
- **Supplementary Figure 6**
  Chromosomal rearrangements observed at *Dici* positions following experimental evolution.
- **Supplementary Figure 7**
  Alignments of fungal transposases of the *Dici* family.
- **Supplementary Figure 8**
  Characteristics of the transposon detected in *Mucor piriformis*.
- **Supplementary Figure 9**
  Phylogenetic tree of the ITm superfamily transposases with uncollapsed clades.



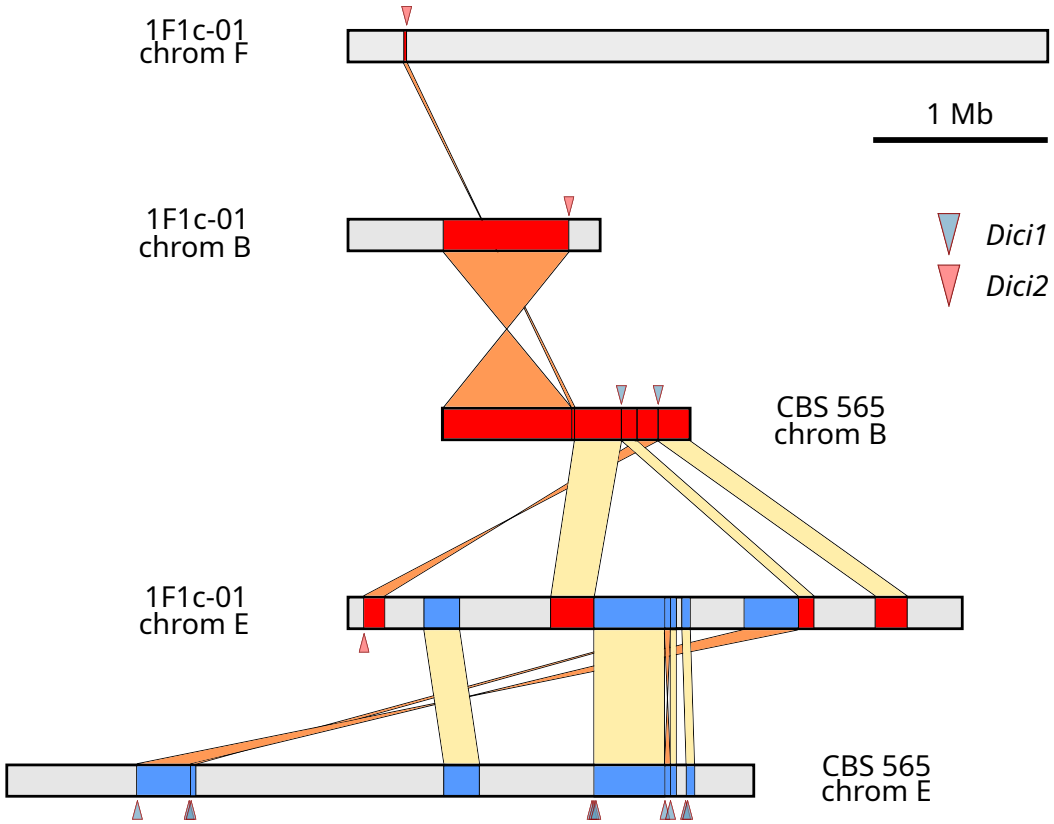
Supplementary Figure 1. Structure of the MTL loci in *D. catenulata*, *Candida anglica*, *C. railenensis*, *C. parapsilosis*, and *C. albicans*. Species and strain names are shown on the right, and scaffold or chromosome accession numbers on the left. Genes conserved in synteny across species are shown in the same color and labeled with their corresponding gene names. Genes encoding proteins of unknown function are labeled with their locus tags. Pseudogenes are indicated by an asterisk.



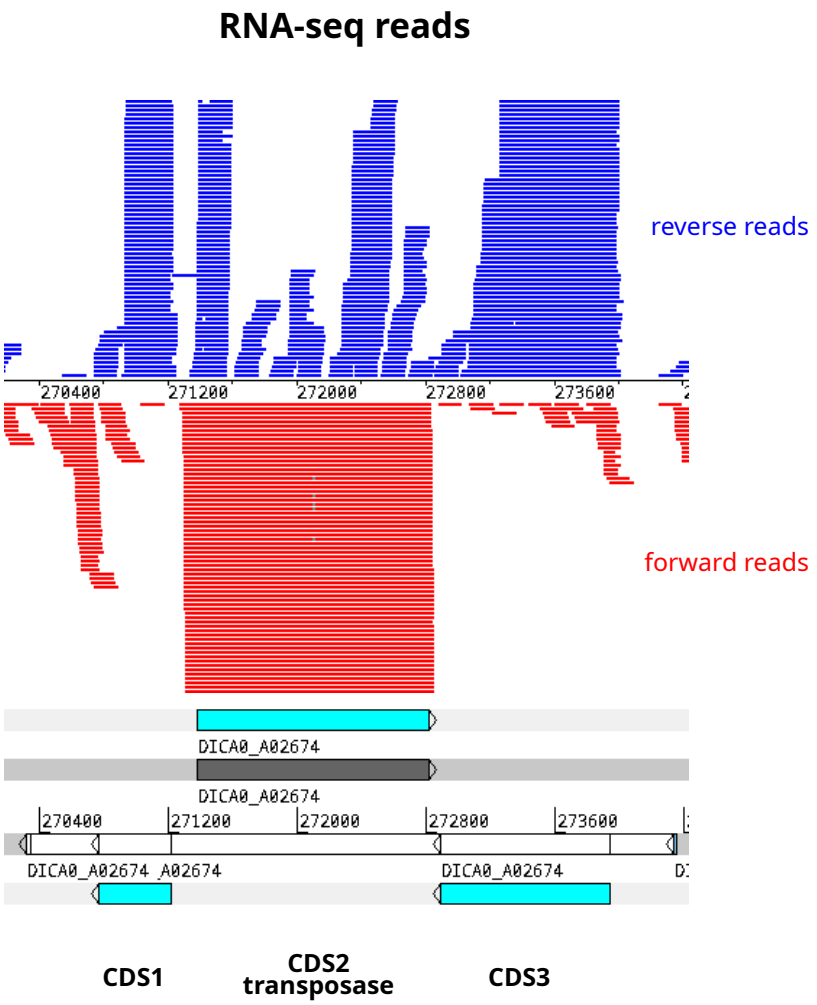
Supplementary Figure 2. Karyotype pattern variation within *D. catenulata* species. For each clade of the phylogenetic tree (identical to the tree of Figure 1), four to five representative strains were selected for karyotype determination by PFGE (see Methods).



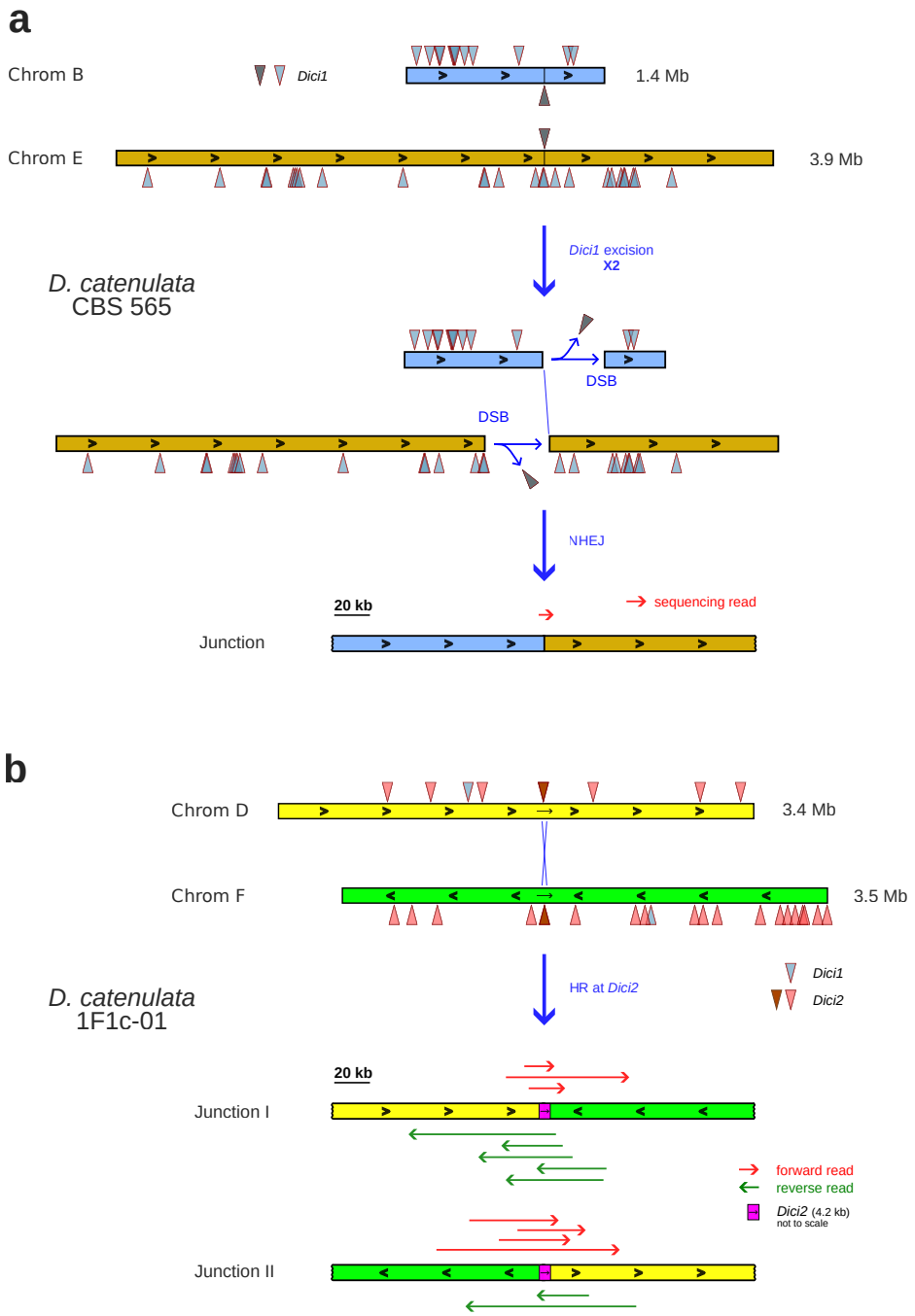
Supplementary Figure 3. Pairwise genome comparisons among *D. catenulata* strains. a Circular plots show pairwise genome alignments generated using MUMmer (see Methods). Each circle represents one pairwise comparison, with the reference genome (row label) displayed at the top and the query genome (column label) at the bottom. Lines connecting chromosomes represent aligned regions longer than 5 kb and are colored according to the chromosomes of the reference strain. **b** Circular plots show synteny blocks generated using SynChro (see Methods). Each circle represents shared synteny between strain CLIB 1179 (top) and one of the four query strains (43B2s-11, 25B2s-01, CBS 565, and 1F1c-01; bottom). CLIB 1179 was used as the reference because it harbors the fewest *Dici* elements (2 *Dici1* and 8 *Dici2*), making it less likely to be affected by ectopic recombination between copies. Chromosomes of CLIB 1179 are color-coded (top outer track). *Dici* elements are indicated as vertical bars on the middle track (*Dici1*, red; *Dici2*, black), and those located at synteny breakpoints are marked with a red diamond. Synteny blocks shared between genomes (inner track) are connected by lines colored according to their positions on CLIB 1179 chromosomes. For each genome, the number of synteny blocks, synteny breakpoints, breakpoints associated with a *Dici* element, and total number of *Dici* elements (*Dici1*+*Dici2*) are reported below each plot.



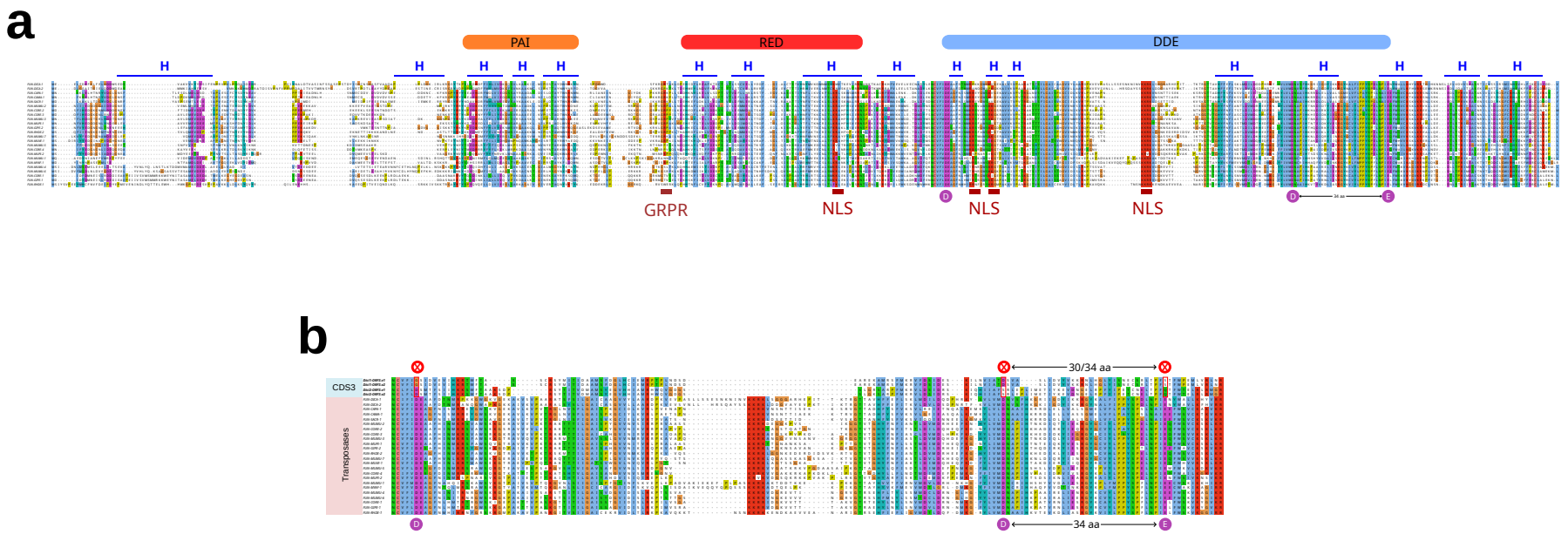
Supplementary Figure 4. Example of complex chromosomal rearrangements between two strains of *D. catenulata*. Recombination events were inferred from shared synteny blocks involving chromosomes B and E of strain CBS 565 and chromosomes B, F, and E of strain 1F1c-01. Chromosomes are drawn to scale (scale bar, 1 Mb). Synteny blocks are colored according to their corresponding chromosomes in CBS 565 (red for chromosome B; blue for chromosome E). Regions shown in gray share no synteny blocks with the other plotted chromosomes. Positions of *Dici* elements are indicated by triangles.



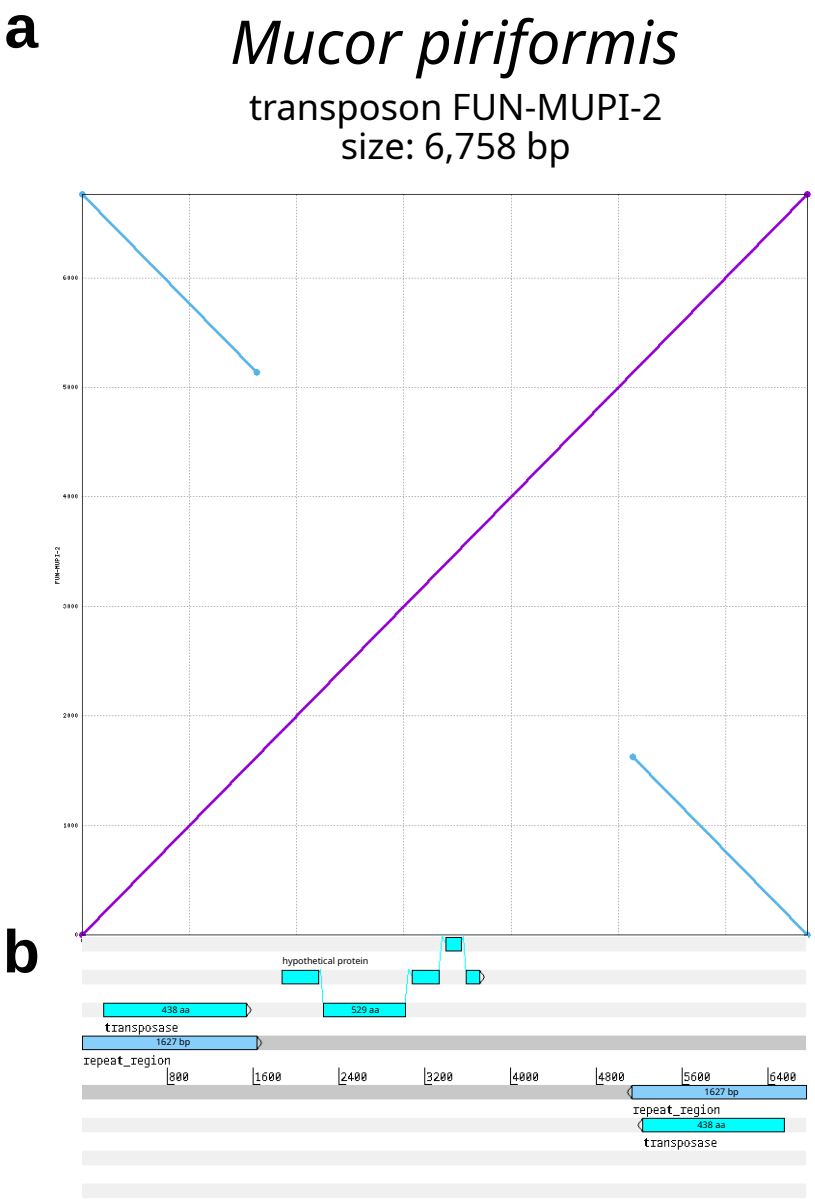
Supplementary Figure 5. Expression of the three CDSs of the *Dici1* transposon in *D. catenulata* strain CBS 565. RNA-seq reads were mapped to the CBS 565 reference genome and visualized using the Artemis software. The genome browser view shows RNA-seq read coverage across the three CDSs of *Dici1* (locus tag: DICA0_A02674), providing evidence for their transcriptional activity. Similar expression patterns were observed for *Dici2* and across the four other sequenced strains.



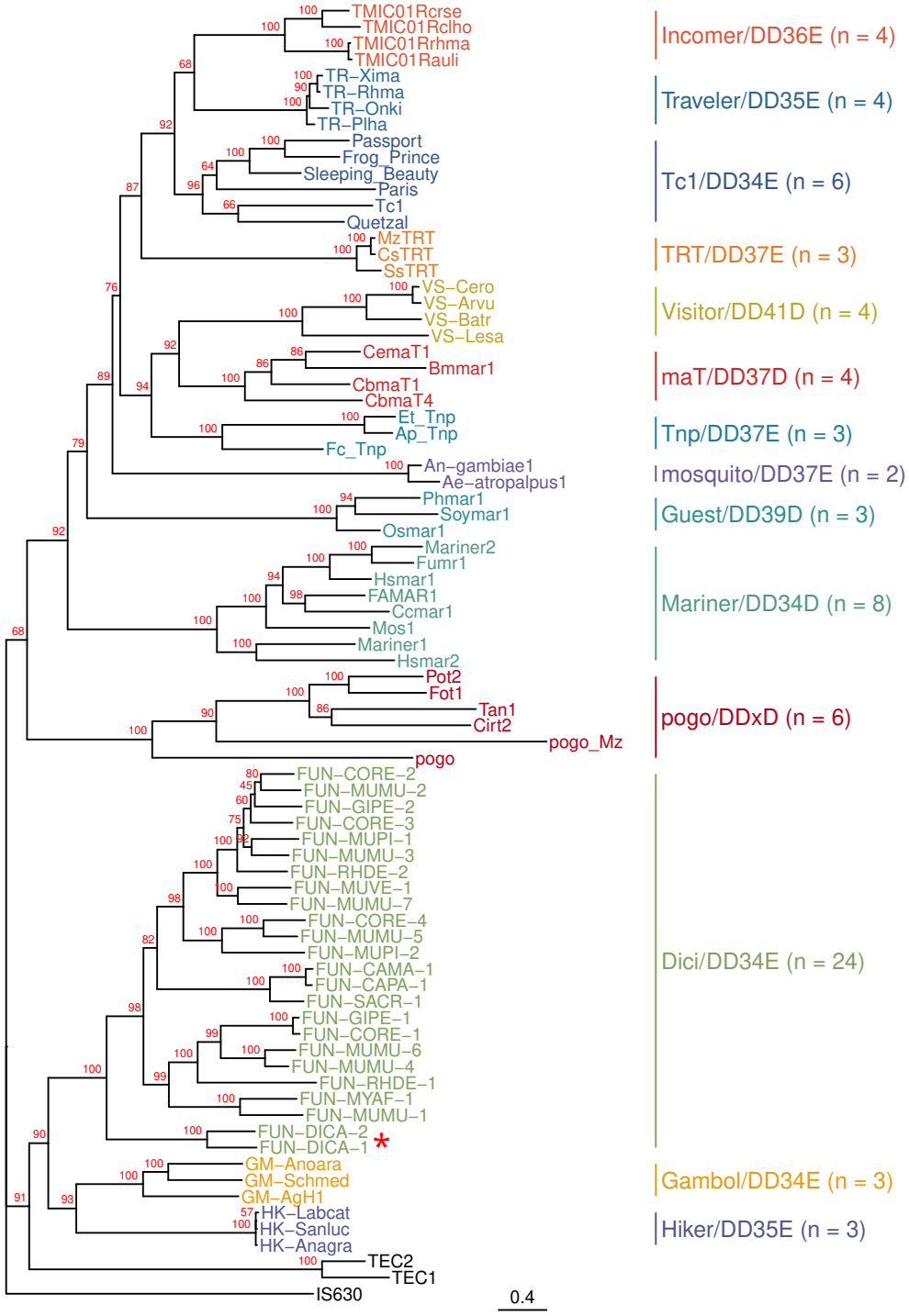
Supplementary Figure 6. Chromosomal rearrangements observed at *Dici* positions following experimental evolution. a In the pooled DNA obtained from the CBS 565 evolved lines, an ONT-sequencing read compatible with a recombination between chromosomes B and E at a *Dici1* insertion site (dark blue) was identified. The absence of *Dici1* copies at the junction suggests simultaneous excision of both elements, the resulting double-strand breaks (DSBs) being repaired by non-homologous end joining (NHEJ) mechanism. This structure was supported by a single read, suggesting that either the event occurred recently, or was counter-selected in the evolved population. **b** In the evolved lines derived from 1F1c-01, 14 read sequences are compatible with a reciprocal translocation, suggesting that homologous recombination occurred between two *Dici2* copies (dark red) located on chromosomes D and F. ONT-sequencing reads are represented above (forward reads in red) or below (reverse reads in green) the junctions. Note that the *Dici2* element is not to scale.



Supplementary Figure 7. Characteristics of the *Dici1* and *Dici2* transposases. a Alignment of fungal transposases from the *Dici* clade highlighting their main structural features. A pair of helix-turn-helix (HTH) motifs forming a PAIRED domain (orange and red regions, top) is located in the N-terminal region, each composed of three α-helices (blue lines, top). A conserved GRPR-like motif is located between the two HTH motifs and is particularly well conserved in the top 14 sequences. Four putative nuclear localization signals (NLS), composed of basic amino acids (K or R), are also conserved. The catalytic DDE domain (blue region, top) contains the catalytic triad (purple circles, bottom), with 34 amino acids separating the second aspartic acid (D) and the glutamic acid (E). **b** Alignment of the DDE-catalytic domain of CDS3 (pale blue rectangle, left) and transposases used in panel **a**. The DDE triad is not conserved in CDS3 (red circles, top), in contrast to the other transposases (purple circles, bottom). Corresponding sequence names, organisms, and accession numbers are provided in Supplementary Data 8.



Supplementary Figure 8. Characteristics of the 6,758-bp transposon detected in *Mucor piriformis*. **a** Dotplot obtained from a reciprocal alignment of the complete transposon sequence using MUMmer. The two inverted-repeat regions of 1,627 bp, observed at both extremities, are drawn in blue. **b** Annotation of the transposon sequence. The repeat regions contain two identical copies of a CDS that encodes the transposase (438 aa). An additional CDS coding for a hypothetical protein (529 aa) was annotated with four spliceosomal introns.



Supplementary Figure 9. Phylogenetic tree of transposases from the *IS630–Tc1–mariner* (ITm) superfamily. The tree was inferred from full-length transposase sequences using the maximum-likelihood method with bootstrap support (in red). The *IS630* transposase was used as an outgroup. A red asterisk indicates the position of *Dici1* and *Dici2*. This tree is identical to the phylogeny shown in the main text (Figure 6a), but with clades expanded. Detailed information on transposase names, catalytic domains, family assignments, and biological origin is provided in Supplementary Data 8.

3 Supplementary Data

Additional spreadsheet documents (XLSX)

- **Supplementary Data 1.** *D. catenulata* strains used in the study. Also includes accession numbers for sequencing and assembled genome data.
- **Supplementary Data 2.** Genetic diversity and differentiation statistics measured in *D. catenulata*.
- **Supplementary Data 3.** Assembly and annotation statistics of *D. catenulata* assembled genomes.
- **Supplementary Data 4.** Genes associated with drug resistance and virulence detected in *D. catenulata*.
- **Supplementary Data 5.** Characteristics and chromosomal positions of the RNA transposons discovered in *D. catenulata* genomes.
- **Supplementary Data 6.** Characteristics and chromosomal positions of the *Dici* transposons discovered in *D. catenulata* genomes.
- **Supplementary Data 7.** Synteny blocks detected between the *D. catenulata* genomes.
- **Supplementary Data 8.** Characteristics of the *Dici* transposons newly identified in Fungi. Also contains the list of the *IS630-Tc1-mariner* transposase sequences used to generate the phylogenetic tree shown in Supplementary Figure 9.
- **Supplementary Data 9.** Sequencing statistics for the five assembled *D. catenulata* genomes, and experimentally evolved populations. Includes short-read mapping metrics for the 61 *D. catenulata* strains.

Supplementary References

[1] Stephanie A. Flowers et al. "Contribution of Clinically Derived Mutations in ERG11 to Azole Resistance in *Candida Albicans*". In: *Antimicrobial Agents and Chemotherapy* 59.1 (Jan. 2015), pp. 450–460. DOI: 10.1128/AAC.03470-14.

[2] Anuradha Chowdhary et al. "A Multicentre Study of Antifungal Susceptibility Patterns among 350 Candida Auris Isolates (2009-17) in India: Role of the ERG11 and FKS1 Genes in Azole and Echinocandin Resistance". In: *The Journal of Antimicrobial Chemotherapy* 73.4 (Apr. 2018), pp. 891–899. DOI: 10.1093/jac/dkx480.

[3] Sean Bergin et al. "Analysis of Clinical *Candida Parapsilosis* Isolates Reveals Copy Number Variation in Key Fluconazole Resistance Genes". In: *Antimicrobial Agents and Chemotherapy* 68.6 (June 2024), e0161923. DOI: 10.1128/aac.01619-23.

[4] Adarsh Jay et al. "The Role of Gene Copy Number Variation in Antimicrobial Resistance in Human Fungal Pathogens". In: *Npj Antimicrobials and Resistance* 3 (2025), p. 1. DOI: 10.1038/s44259-024-00072-1.

[5] Julian R. Naglik, Stephen J. Challacombe, and Bernhard Hube. "*Candida Albicans* Secreted Aspartyl Proteinases in Virulence and Pathogenesis". In: *Microbiology and molecular biology reviews* 67.3 (Sept. 2003), pp. 400–428. DOI: 10.1128/MMBR.67.3.400-428.2003.