nature portfolio

Peer Review File

DNA transposon expansion drives genome plasticity in *Diutina catenulata*

Corresponding Author: Dr Cécile Neuvéglise

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The results of this interesting manuscript are noteworthy in a variety of ways: the exploration of an understudied yeast species' inherent genomic diversity, the assembly of high-quality annotated genomes of said species, and the exploration of a novel transposon class therein which potentially contributes to genomic plasticity.

I think the work is of value to many fields and provides a roadmap for comprehensive genomic, genetic, and population analysis of understudied species, especially ones which have been implicated in emerging clinical contexts.

The analysis is of high quality and done to an impressive standard. In particular the attention to detail and care taken when trimming, filtering, and manually inspecting various aspects of the sequence data to ensure high-quality read mapping and genome assembly, as well as analysis of the transposable elements themselves. The data is analysed very well and the results presented are strongly supported, especially the nature of this new class of TE and its potential link to the genome plasticity and divergence within the species. The methods are very thorough and detailed, sufficient for replication in my opinion.

Disclaimer: I am no expert on the AMOVA test for genetic differentiation. Based on my cursory reading it seems appropriate and well executed.

These are my line-by-line comments:

Line 98 - this and the following sentence seem, even if not, to be at odds with one another. The flow is confusing and it reads slightly repetitively. Would suggest clearer wording akin to 'the same factors can threaten genome integrity'

Line 218 - its not always clear throughout the manuscript exactly what the 'PDO' refers to. When described it sounds like an adjective-acronym, here it is used as a noun in conjunction with a specific numeric label (PDO #43; as elsewhere), and on line 876 the phrase 'PDO cheeses' is used like PDO is again an adjective. I suggest 'PDO cheese #43' be used here, and elsewhere, to clear any confusion.

Line 227 - 'is almost exclusively composed of cheese strains'

Line 271 - 'to differ by no more than 851 SNPs with any of the clades cheese strains'

Line 292 - Here PDO sounds like it refers to strains. I think it does not, but the whole following section, until the paragraph end, reads confusingly. Suggest a new paragraph at the word 'Interestingly,' and wording more like 'PDO cheeses #18 and #25 had representatives/strains in clades #4 and #6, etc.

Line 366 - Here there were two confusions: figure 4 is discussed in the text before figures 2 and 3, so the flow is a little strange. There is no reason figure 4 cannot be figure 2? Furthermore it wasn't clear that the analysis here relied on simply blasting the reads of the 61 illumina strains against the Dici elements. Suggest adding a modest line about how this was an assembly-free estimation.

Line 415 - Figure 2: the lengths in panels A & B are not explained at all: 143 / 146 aa : why is there two? What are they? It should be in the figure legend. I only know from the later text and SuppTable6

Line 461 - Figure 3: Scale bars for chromosomes in panel B?

Line 600 - Suggest removing 'tested condition' or clarifying that RNA was sequenced in plain vegetative rich media for annotation purposes. The RNA-seq is not discussed except in the supplemental text, just to annotate, so saying 'tested condition' reads like an experiment for DE was performed but never discussed?

Line 719 - later = latter?

Line 758 - It is more accurate to say 'The highest conservation, when compared to the full-length Dici transposases, was observed with Gambol and Hiker'

Line 823 - 'with no more collinearity' sounds very awkward and mis-interpretable. Suggest re-wording.

Line 937 - It is unclear what this sentence means: 'Only one strain per cheese was analyzed in this study.', when the results discuss several times there were two strains from the same cheese, but different producers?

Line 1679 - Supplementary figure 6 is not discussed anywhere in the main text, I suppose it is hard to fit it in. I see it is in the supplemental text, and think it is of value. Would suggest very brief mention of the fact that MAT loci were investigated.

Reviewer #2

(Remarks to the Author)
This manuscript characterises a previously undescribed family of IS630–Tc1–mariner DNA transposons, termed Dici, in the opportunistic yeast Diutina catenulata. Using population genomic data from 61 isolates spanning six major clades, the authors identify two related elements (Dici1 and Dici2), describe their domain architecture and insertion signatures, and correlate their presence with extensive chromosomal rearrangements. In addition to transposon characterisation, the work provides detailed genome assemblies for clade-representative isolates and reports ploidy variation in this species. They further argue that (i) D. catenulata is deeply structured into diverged lineages, (ii) Dici has recently and potentially continuously expanded across the species, and (iii) Dici-associated rearrangements may have contributed to reduced gene flow or reproductive isolation among clades.
The dataset is valuable, the identification of a putatively novel transposable element is interesting, and the comparative genomics is clearly presented. However, several central claims—especially those regarding ongoing transposition, causality of genomic rearrangements, and effects on reproductive isolation—are currently not supported by direct evidence.
Although the work is interesting and contains several promising observations, in its current form the manuscript is not ready for publication. Multiple major conclusions rely on inference rather than demonstration, and the strongest claims (recent activity, mobilisation, genome reshaping, contribution to reproductive isolation) require additional analyses or experiments. Below we outline the essential issues that must be addressed to fully support the manuscript's claims.
Major Comments
1. Claims of "recent and ongoing" Dici activity require experimental support
The manuscript repeatedly states or implies that Dici is undergoing a recent or ongoing burst of transposition. However, all evidence provided is indirect:
• high sequence identity among copies,
• presence of intact ORFs,
• detectable RNA expression,
• proximity to breakpoints.
These data are consistent with activity, but they do not demonstrate that the element is currently mobilising.
This could be tested in several ways, for example:
(a) Stress-induced mobilisation assays in D. catenulata
Expose a high-copy strain to physiological stress (e.g., heat, oxidative stress, DNA damage), passage for several dozens of generations, resequence, and detect new insertions via split-read mapping.
(b) A transposition reporter assay
Clone Dici TIRs flanking a marker cassette; express the Dici transposase from its native or inducible promoter; monitor excision and reintegration by PCR or selection.
(c) Heterologous transposition in a tractable yeast
Co-express the Dici transposase and TIR-flanked reporter in Saccharomyces cerevisiae, where transposition systems are well characterised.
Any one of these would greatly strengthen the central assertion that Dici is actively reshaping genomes. Without such data,

the manuscript should tone down all language implying current mobilisation.
2. Inference of reduced gene flow or reproductive isolation is not currently justified
The manuscript implies that the six clades may be reproductively isolated. While population structure is strong, the study:
• does not present evidence of sexual reproduction in D. catenulata,
• does not test inter-clade mating or hybrid viability.
Thus, claims regarding reproductive isolation should be moderated unless supported by formal gene-flow analyses.
3. Causal link between karyotype variation and clade differentiation is not currently supported
The manuscript proposes that karyotype differences may contribute to reproductive isolation and clade divergence. However, the study assesses karyotype structure from only one representative genome per clade, which does not establish whether karyotypes are conserved within clades or variable among strains.
To support this inference, karyotype variation should be evaluated across multiple isolates per clade.
Minor comments
1. Human isolates within the dairy-associated clades are described as commensal, but the manuscript does not present evidence that these isolates are not transient from food/dairy sources.
2. The legend for Figure 3 lacks sufficient methodological detail to allow readers to interpret how results were generated. Discuss hybrid sequence data assembly to support synteny plot. Scale bar on synteny pot would also make validation of assembly against CHEF gel easier.
3. Supplementary Figure 6 is not cited in the main text, and the observed MTL defects are not discussed. The supplementary data indicate defective MTL in D. catenulata, which has major implications for the discussion of potential sexual reproduction. This result should be referenced and integrated into the main text, addressing how MTL disruption affects the interpretation of reproductive capacity.
4. Candida resistance determinants listed include genes such as ERG11, which encodes the azole drug target. The text should clarify that these genes contribute to resistance only when they undergo mutation, overexpression, or copy-number variation, rather than implying that their mere presence confers resistance. Although the authors note multicopy genes in the dataset, the discussion does not connect these findings to established examples of CNV-driven resistance in Candida. Incorporating this comparison would strengthen the interpretation and relevance of the results.
5. Heterozygous diploids are omitted from assembly and karyotype analyses. Incorporating these genomes, and analysing the ancestry of their constituent haplotypes, may be important for fully understanding D. catenulata evolution.

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #4

(Remarks to the Author)
Bigey et al.

This is an interesting and carefully executed study that reports a remarkable discovery about genome organization and population structure in the budding yeast Diutina catenulata. This yeast is associated with dairy environments and appears to be an emerging pathogen. It appears to have undergone recent and dramatic amplifications of two transposon elements, named Dici1 and Dici2. These are DNA transposons (Class I mobile elements), which are quite rare among yeast species, and the level of amplification is unprecedented (up to 96 copies in one isolate). The authors show that different isolates of D. catenulata differ by extensive amounts of chromosomal rearrangement (e.g. there are 60 breaks of synteny between two isolates; Fig 3B) and that the endpoints of these rearrangements tend to lie at Dici elements, which implicates Dici in the mechanism of rearrangement. By a phylogenetic analysis of 61 isolates, they show that this species had a deep clades structure, indicative of low (or zero) genetic exchange between different clades which may indicate reproductive isolation resulting from the genomic rearrangements driven by Dici elements.

Major comments:

L492 and L830: It's very clear that Dici elements are over-represented at the endpoints of genomic rearrangements, but I would have liked to see one (or more) examples drawn in detail in a Figure. Are the rearrangements usually reciprocal translocations? Do they appear to involve ectopic recombination between two Dici elements? Or is there usually only a Dici element at one of the two new junctions formed in a rearrangement? Is it possible that some of the genomic rearrangements are the result of failed attempts to transpose?

L691: This section of text concludes that CDS3 is likely a relic of a transposase gene, i.e. a pseudogene. If this is true, we would not expect to see evolutionary conservation of CDS3 between Dici1 and Dici2, which are only ~52% identical in DNA sequence. Calculating measures of evolutionary conservation, such as the dN/dS ratio (nonsynonymous vs synonymous substitutions) should be informative about whether CDS3 (and also CDS1) has been under selection to conserve its ability to encode a protein, or has been evolving neutrally.

L702: The text in Section 2.6, and Figure 5A, should be more assertive about the proposal that Dici is newly described

transposon family that has members in many other fungal species too. Dici1 and Dici2 in D. catenulata are just 2 out of 24 transposons in this new family, all of which were previously unknown. It wasn't initially obvious to me that the green triangle in Fig 5A called "Dici/DD34E (n=24)" includes transposons from species other than D. catenulata. I suggest adding a panel to Fig 5, to show a zoomed-in view of the Dici clade that is currently shown in Figure S5, including the names of the species that contain members of the Dici family, and highlighting the positions of Dici1 and Dici2 within the Dici family. It is easy to overlook this point, because the name "Dici" is initially used in the manuscript as a collective name for Dici1 and Dici2 (section headings on L507, L608) but then on L702 it is used more broadly as the name of the newly-proposed family that exists in multiple species. I suggest mentioning that a new cross-species transposon family was discovered, in the Abstract, and around L162.

Minor comments:

L043: The Abstract doesn't mention that the number of copies of Dici extends to at least 96 in one isolate, which is one of the most interesting discoveries in this study.

L076: Please provide more information about what a "cut-and-paste" mechanism means. How does the number of copies of the element increase, if it is lost from the original site and gained at a new site?

L218: Please explain what PDO numbers mean, such as "PDO #43". Is it an anonymous designation for a particular type of cheese? Or a particular geographical location, or cheese producer? I am not asking for the name (which might be commercially sensitive), just a description of what the category means.

L307: This section needs to quantify the amount of sequence divergence between the clades, either by stating the average number of SNP differences between clades, or the percent nucleotide divergence in the genome. Some of this information is stated at a later point in the text (L524) but it should be stated earlier, in the context of Fig 1A.

L346: Change to "sequenced using long reads", to differentiate these isolates from the other ~56 isolates that were sequenced only by Illumina short reads.

L366: Figure 4 is cited out of order, before Figures 2 and 3.

L428: Is Figure 2A drawn to scale? Please say this in the legend. It seems to show that there is a lot of non-coding DNA in the Dici elements, which is surprising for a mobile genetic element.

L526: The text implies that Supplementary Table 6 contains information about the sequence divergence between the different sequenced isolates, but this information isn't in that Table. I didn't see it in any Table.

L537 and Figure 4: It isn't clear how the Y-axis in Figure 4 (transposon abundance, expressed in percent) relates to the number of copies of the transposons. The Methods section explains how the percentage was calculated (L1107) but an estimate of the actual copy number would be more informative. Can you not use the data from the 5 completely-sequenced isolates to convert these percentages into an estimated number of copies of the Dici elements? I was left wondering if any of the isolates have more than 96 copies.

L990: I found the division of material between the Methods section and the Supplementary Methods section rather confusing. Supplementary Methods section 2.4 mentions that long-read assemblies were corrected using "Illumina reads", but these are not the Illumina RNA-seq reads that were mentioned in the previous paragraph (section 2.3). Instead, they are the Illumina genomic DNA reads that are not described at all in Supplementary Methods, but in section 4.6 of the main text. It would be more sensible to put all the sequencing details together in one place.

Pages 40-69: Formatting of the Supplementary Tables became messed up during conversion from Excel to PDF.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I find the responses to my comments, as well as the others reviewers comments + responses thereof, satisfactory. I believe the manuscript is stronger now than before, especially due to the clearer wording, rearrangements, and additional experiment. This experiment (and it's simplicity) clearly demonstrates the flexible nature of D. catenulata genomes, and provides an interesting precedent for more of such experiments.

The new Figure 2C & D are very interesting. They are very nice representations of the translocation junctions. As a pedant for figures, I wonder though what the green vertical line is right at the center of figure 2C?

Otherwise, no comments.

I recommend the manuscript for review.

(Remarks on code availability)
NA

Reviewer #2

(Remarks to the Author)
I am satisfied with the responses provided and have no additional comments

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The authors have done an excellent job of revising the manuscript, and they have addressed all of my previous comments to my satisfaction (reviewer #4). The new section on Dici element mobility during laboratory evolution is particularly exciting.

I noticed two minor errors in Figures in the revised manuscript.
In Figure 2A, the key lacks an explanation of the 5th item (blue). Is it Ty5 v1 ?
In Figure 3B, the label GPRR should be GRPR, I think.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Dici: a novel DNA transposon reshaping the genome of the opportunistic yeast *Diutina catenulata*

Response to reviewers

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

1 General response to the reviewers

We thank the reviewers for their careful evaluation of our manuscript and for their constructive and insightful comments, which have helped us significantly to improve the quality and clarity of the work.

To make the review process easier, we have uploaded the supplementary tables and figures to Zenodo (https://doi.org/10.5281/zenodo.17613032), where they can be easily downloaded.

Below, we provide a point-by-point response to their comments.

2 Reviewer #1

The results of this interesting manuscript are noteworthy in a variety of ways: the exploration of an understudied yeast species' inherent genomic diversity, the assembly of high-quality annotated genomes of said species, and the exploration of a novel transposon class therein which potentially contributes to genomic plasticity.

I think the work is of value to many fields and provides a roadmap for comprehensive genomic, genetic, and population analysis of understudied species, especially ones which have been implicated in emerging clinical contexts.

The analysis is of high quality and done to an impressive standard. In particular the attention to detail and care taken when trimming, filtering, and manually inspecting various aspects of the sequence data to ensure high-quality read mapping and genome assembly, as well as analysis of the transposable elements themselves. The data is analysed very well and the results presented are strongly supported, especially the nature of this new class of TE and its potential link to the genome plasticity and divergence within the species. The methods are very thorough and detailed, sufficient for replication in my opinion.

Disclaimer: I am no expert on the AMOVA test for genetic differentiation. Based on my cursory reading it seems appropriate and well executed.

Response: we thank the reviewer for this positive feedback.

2.1 Comments

Line 98 this and the following sentence seem, even if not, to be at odds with one another. The flow is confusing and it reads slightly repetitively. Would suggest clearer wording akin to 'the same factors can threaten genome integrity'

Response: as suggested, the paragraph was revised accordingly (lines 113-118) to avoid repetition and improve clarity and flow.

Line 218 its not always clear throughout the manuscript exactly what the 'PDO' refers to. When described it sounds like an adjective-acronym, here it is used as a noun in conjunction with a specific numeric label (PDO #43; as elsewhere), and on line 876 the phrase 'PDO cheeses' is used like PDO is again an adjective. I suggest 'PDO cheese #43' be used here, and elsewhere, to clear any confusion.

Response: PDO stands for Protected Designation of Origin. We worked on cheeses with 44 PDO labels. The PDOs were anonymized and designated as PDO1 to PDO44. We replaced all occurrences of "PDO #XX" with "PDOXX" throughout the manuscript. We revised the wording to "strains isolated from PDOXX cheeses" where appropriate. Additional details were added to the Methods section (lines 1330-1365) to explain the encoding used to distinguish PDOs, producers and seasonal cheese samples.

In addition, the phrasing on line 1273 of the discussion was corrected to "cheeses from the same PDO" to improve clarity and avoid redundancy.

Line 227 'is almost exclusively composed of cheese strains'

Response: this sentence has been revised as suggested (line 294) and further details about the strain nomenclature have been added.

Line 271 'to differ by no more than 851 SNPs with any of the clades cheese strains'

Response: this sentence was completely rewritten (line 305).

Line 292 Here PDO sounds like it refers to strains. I think it does not, but the whole following section, until the paragraph end, reads confusingly. Suggest a new paragraph at the word 'Interestingly,' and wording more like 'PDO cheeses #18 and #25 had representatives/strains in clades #4 and #6, etc.

Response: this was revised to clarify this issue. As suggested, a new paragraph was created and the wording corrected accordingly (lines 328-337).

Line 366 Here there were two confusions: figure 4 is discussed in the text before figures 2 and 3, so the flow is a little strange. There is no reason figure 4 cannot be figure 2? Furthermore it wasn't clear that the analysis here relied on simply blasting the reads of the 61 illumina strains against the Dici elements. Suggest adding a modest line about how this was an assembly-free estimation.

Response: indeed, Fig. 4 was not in the right place (it becomes Fig. 2). Furthermore, we added to the legend a short sentence about the method used for copy-number estimation: mapping Illumina short-reads sequences with BWA on individual TE sequences.

Line 415 Figure 2: the lengths in panels A & B are not explained at all: 143 / 146 aa : why is there two? What are they? It should be in the figure legend. I only know from the later text and SuppTable6

Response: the size difference between the *Dici1* (4.0 kb) and *Dici2* (4.2 kb) elements is now clearly shown in Fig. 2 (new Fig. 3). The lengths of the coding sequences (CDSs) are indicated for both *Dici* variants. In addition, all three CDSs are now drawn to scale, as requested by the reviewer #4.

Line 461 Figure 3: Scale bars for chromosomes in panel B?

Response: this suggestion was also proposed by reviewer #2. A scale bar was added to panel B of Fig. 3 (new Fig. 4) to help estimate chromosome size. Furthermore, the figure legend was modified to include additional methodological details regarding the analysis of synteny blocks.

Line 600 Suggest removing 'tested condition' or clarifying that RNA was sequenced in plain vegetative rich media for annotation purposes. The RNA-seq is not discussed except in the supplemental text, just to annotate, so saying 'tested condition' reads like an experiment for DE was performed but never discussed?

Response: as suggested, this was corrected to clearly state that the data were obtained solely for annotation purposes (lines 765-767).

Line 719 later = latter?

Response: we corrected this typo.

Line 758 It is more accurate to say 'The highest conservation, when compared to the full-length Dici transposases, was observed with Gambol and Hiker'

Response: the sentence was corrected as suggested by the reviewer (lines 1138-1140).

Line 823 'with no more collinearity' sounds very awkward and mis-interpretable. Suggest re-wording.

Response: as suggested, the sentence was corrected accordingly (lines 1230-1232).

Line 937 It is unclear what this sentence means: 'Only one strain per cheese was analyzed in this study.', when the results discuss several times there were two strains from the same cheese, but different producers?

Response: We agree that the methodological description was unclear. A distinction has to be made between PDO and cheeses. Although several strains originated from the same PDO, they were not isolated from the same cheese or sample of cheese. In addition, in some cases, strains were isolated from the same producer at different times of the year to assess seasonal variation, resulting in different cheese samples being analyzed. Finally, a single *D. catenulata* isolate per cheese sample was considered in this study.

We added clarifications to the Methods section to provide a better description of the sampling strategy (lines 1350-1365).

Line 1679 Supplementary figure 6 is not discussed anywhere in the main text, I suppose it is hard to fit it in. I see it is in the supplemental text, and think it is of value. Would suggest very brief mention of the fact that MAT loci were investigated.

Response: indeed, analysis of the mating-type-like (MTL) locus is valuable for understanding the biological life cycle of *D. catenulata*. The relevant content from the supplementary text has been moved to the main text (lines 390-425), and a characterization of the MTL locus has been performed in the 61 *D. catenulata* strains and is now discussed (Supplementary Table 1).

3 Reviewer #2

This manuscript characterises a previously undescribed family of IS630–Tc1–mariner DNA transposons, termed Dici, in the opportunistic yeast Diutina catenulata. Using population genomic data from 61 isolates spanning six major clades, the authors identify two related elements (Dici1 and Dici2), describe their domain architecture and insertion signatures, and correlate their presence with extensive chromosomal rearrangements. In addition to transposon characterisation, the work provides detailed genome assemblies for clade-representative isolates and reports ploidy variation in this species. They further argue that (i) D. catenulata is deeply structured into diverged lineages, (ii) Dici has recently and potentially continuously expanded across the species, and (iii) Dici-associated

rearrangements may have contributed to reduced gene flow or reproductive isolation among clades.

The dataset is valuable, the identification of a putatively novel transposable element is interesting, and the comparative genomics is clearly presented. However, several central claims—especially those regarding ongoing transposition, causality of genomic rearrangements, and effects on reproductive isolation—are currently not supported by direct evidence.

Although the work is interesting and contains several promising observations, in its current form the manuscript is not ready for publication. Multiple major conclusions rely on inference rather than demonstration, and the strongest claims (recent activity, mobilisation, genome reshaping, contribution to reproductive isolation) require additional analyses or experiments.

Below we outline the essential issues that must be addressed to fully support the manuscript's claims.

3.1 Major Comments

3.1.1 Claims of "recent and ongoing" Dici activity require experimental support

The manuscript repeatedly states or implies that Dici is undergoing a recent or ongoing burst of transposition. However, all evidence provided is indirect:

- high sequence identity among copies,
- presence of intact ORFs,
- detectable RNA expression,
- proximity to breakpoints.

These data are consistent with activity, but they do not demonstrate that the element is currently mobilising.

This could be tested in several ways, for example:

a Stress-induced mobilisation assays in D. catenulata Expose a high-copy strain to physiological stress (e.g., heat, oxidative stress, DNA damage), passage for several dozens of generations, resequence, and detect new insertions via split-read mapping.

b A transposition reporter assay Clone Dici TIRs flanking a marker cassette; express the Dici transposase from its native or inducible promoter; monitor excision and reintegration by PCR or selection.

c Heterologous transposition in a tractable yeast Co-express the Dici transposase and TIR-flanked reporter in Saccharomyces cerevisiae, where transposition systems are well characterised.

Any one of these would greatly strengthen the central assertion that Dici is actively reshaping genomes. Without such data, the manuscript should tone down all language implying current mobilisation.

Response: we thank the reviewer for suggesting experimental methods to test *Dici* mobilization. To address this, we performed an evolution experiment combined with long-read ONT sequencing to detect insertion and excision events in populations (see Methods, starting on line 1841). We identified genomic signatures consistent with *Dici* insertion and excision events (see new section starting on line 776 and Supplementary Table 6). Furthermore, to assess the susceptibility of *D. catenulata* genomes to chromosomal rearrangements, we examined variation in karyotypes among clones isolated from the evolution lines (new Fig. 5). We also observed rearrangements occurring at *Dici* insertion sites (Supplementary Fig. 6), suggesting that *Dici* may be in part responsible for the genome plasticity.

3.1.2 Inference of reduced gene flow or reproductive isolation is not currently justified

The manuscript implies that the six clades may be reproductively isolated. While population structure is strong, the study:

- does not present evidence of sexual reproduction in D. catenulata,
- does not test inter-clade mating or hybrid viability.

Thus, claims regarding reproductive isolation should be moderated unless supported by formal gene-flow analyses.

Response: we thank the reviewer for highlighting the critical distinction between population structure and formal evidence of reproductive isolation. We acknowledge that no experimental mating or hybrid viability assays were performed for *D. catenulata*. Such experiments are technically demanding, highly speculative, and remain beyond the scope of this study, which primarily focuses on comparative genomics and the role of transposable elements.

However, to address this point further, we conducted a complementary genomic analysis of the mating-type-like (MTL) locus across the 61 *D. catenulata* strains (starting on line 390). Our analysis revealed the presence of two incomplete and distinct MTL locus configurations with the apparent absence of the a1/a2 genes, suggesting that the mating pathway of *D. catenulata* may be impaired. While this does not strictly preclude alternative or cryptic modes of reproduction, it is consistent with a lack of canonical sexual recombination in these populations. Interestingly, a single MTL locus configuration per clade was

observed in clades #2, #3, #4, and #6. This provides further support for the hypothesis of clonal propagation.

Furthermore, our population-genetic analyses (F_{ST}, Tajima's D, and AMOVA) independently demonstrate strong genetic structure among the six clades and indicate negligible levels of gene flow (lines 338-361). While we cannot formally conclude reproductive isolation, restricted genetic exchange remains the most parsimonious interpretation of the observed genomic patterns.

In response to the reviewer's concern, we have moderated the language throughout the manuscript. We now frame these conclusions as well-supported genomic hypotheses rather than definitive statements regarding the reproductive biology of *D. catenulata*.

3.1.3 Causal link between karyotype variation and clade differentiation is not currently supported

The manuscript proposes that karyotype differences may contribute to reproductive isolation and clade divergence. However, the study assesses karyotype structure from only one representative genome per clade, which does not establish whether karyotypes are conserved within clades or variable among strains.

To support this inference, karyotype variation should be evaluated across multiple isolates per clade.

Response: we acknowledge the reviewer's concern regarding sampling bias in the karyotype analysis. To address this, we obtained karyotype profiles for three to five strains per clade (Supplementary Fig. 2). The exception is clade #4, for which repeated PFGE experiments failed to yield interpretable results due to DNA degradation. We also included karyotype data for the four diploid heterozygous strains of clade #5, which were not characterized in the original version of the manuscript. Furthermore, karyotype profiles were obtained for clones isolated from the experimental evolution lines (new Fig. 5).

These results reveal a significant degree of karyotype polymorphism, demonstrating that chromosomal architecture is dynamic in *D. catenulata*. While the presence of intra-clade variation confirms that a single karyotype does not define a clade, the significant structural differences observed between clades support our hypothesis that chromosomal rearrangements contribute to the chromosomal architecture evolution and subsequent genetic isolation of these lineages. We have updated the manuscript to reflect this nuanced view, and discussed how karyotype instability is a potential mechanism for limiting intra- and inter-clade crossing, thereby reinforcing clade differentiation (see discussion, starting on line 1285).

3.2 Minor comments

1. Human isolates within the dairy-associated clades are described as commensal, but the manuscript does not present evidence that these isolates are not transient from food/dairy sources.

Response: we agree that the precise origin and the isolation conditions of the strain observed in clade 4 remain unknown. It is difficult to ascertain whether this strain is transiently associated with food or opportunistic. Our analyses indicate that this strain is genetically closely related to cheese-associated strains. While this genetic similarity suggests a shared evolutionary background, it does not allow us to reconstruct its full ecological or epidemiological history. Therefore, we do not comment on it further and instead refer to its original description. It is nevertheless well established that some yeasts typically associated with food niches can exhibit opportunistic behavior in immunocompromised patients. For example, food-associated strains of *Saccharomyces cerevisiae* have been shown to cause opportunistic infections under certain conditions (for example: Morard et al. Comparative genomics of infective *Saccharomyces cerevisiae* strains reveals their food origin. Sci. Rep. 13 (2023). https://doi.org/10.1038/s41598-023-36857-z)

2. The legend for Figure 3 lacks sufficient methodological detail to allow readers to interpret how results were generated. Discuss hybrid sequence data assembly to support synteny plot. Scale bar on synteny pot would also make validation of assembly against CHEF gel easier.

Response: we have expanded the legend of Figure 3 (new Fig. 4) to provide additional methodological details, including a description of the long-read genome assemblies used to generate the synteny plots and how synteny blocks were identified across strains. In addition, as suggested by reviewers #1 and #4, a scale bar has been added to panel B to facilitate direct comparison of chromosome sizes with karyotype from PFGE data.

3. Supplementary Figure 6 is not cited in the main text, and the observed MTL defects are not discussed. The supplementary data indicate defective MTL in D. catenulata, which has major implications for the discussion of potential sexual reproduction. This result should be referenced and integrated into the main text, addressing how MTL disruption affects the interpretation of reproductive capacity.

Response: a similar comment was also raised by reviewer #1. The section describing the characterization of MTL loci in the sequenced *D. catenulata*

genomes has therefore been moved from the supplementary text to the main manuscript. A new paragraph now focuses on the characterization of MTL loci across all *D. catenulata* strains in our collection and discusses the potential implications of the presence of defective loci in the *D. catenulata* life cycle (starting on line 390).

4. Candida resistance determinants listed include genes such as ERG11, which encodes the azole drug target. The text should clarify that these genes contribute to resistance only when they undergo mutation, overexpression, or copy-number variation, rather than implying that their mere presence confers resistance. Although the authors note multicopy genes in the dataset, the discussion does not connect these findings to established examples of CNV-driven resistance in Candida. Incorporating this comparison would strengthen the interpretation and relevance of the results.

Response: we thank the reviewer for pointing to this clarification.

In the main text, the corresponding paragraph was modified to focus on the presence of multicopy genes encoding efflux pumps and hydrolytic enzymes (starting on line 427).

In the Supplementary text, we have explicitly stated that the mere presence of canonical antifungal resistance determinants such as *ERG11* and *FKS1* does not confer resistance (Supplementary text, section 1.3). We now emphasize that resistance in pathogenic *Candida* species arises through specific mutations, altered expression levels, or copy-number variation affecting these genes or associated pathways, and we cite well-established examples for both azole and echinocandin resistance.

In addition, we expanded the discussion to place the observed copy-number variations (CNVs) in the context of known CNV-driven resistance mechanisms in *Candida*. We now explicitly connect the observed amplification of effector genes in *D. catenulata*, particularly the multidrug efflux transporters *CDR1* and *MDR1*, as well as hydrolytic enzymes such as *PLB1* and *SAP5*, to documented cases in which increased gene dosage enhances drug tolerance or virulence in pathogenic yeasts.

Together, these revisions ensure that our interpretation does not imply intrinsic drug resistance, but rather highlights a conserved genetic potential for adaptation that mirrors established resistance mechanisms in related *Candida* species.

5. Heterozygous diploids are omitted from assembly and karyotype analyses. Incorporating these genomes, and analysing the ancestry of their constituent

haplotypes, may be important for fully understanding D. catenulata evolution.

Response: we agree that heterozygous diploid strains raise interesting questions regarding *D. catenulata* biology and genome evolution. We now include karyotypes for diploid strains from clade #5, which confirm that these genomes also display structural variations, especially in large chromosomes (Supplementary Fig. 2). However, we did not incorporate heterozygous diploids into the comparative genomics analysis, as their high heterozygosity substantially increases assembly complexity and would require haplotype-resolved approaches, with associated increases in sequencing depth and cost.

We note that the evolutionary origin of these diploid genomes is partially addressed in the population structure analyses, where we detect contributions from an additional, currently uncharacterised ancestral haplotype in all four strains (lines 312-317; Fig. 1). This raises the hypothesis of a common hybridisation event. While a detailed reconstruction of haplotype ancestry and hybridization history would be highly informative, such analyses fall outside the scope of the present study, which is focused on transposable element dynamics and genome structural plasticity. We therefore view a comprehensive investigation of diploid ancestry and hybrid genome evolution as an important direction for future work.

4 Reviewer #3

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: we thank reviewer #3 for his co-review and for his contribution to the evaluation process. No specific comments requiring a response were raised.

5 Reviewer #4

This is an interesting and carefully executed study that reports a remarkable discovery about genome organization and population structure in the budding yeast Diutina catenulata. This yeast is associated with dairy environments and appears to be an emerging pathogen. It appears to have undergone recent and dramatic amplifications of two transposon elements, named Dici1 and Dici2. These are DNA transposons (Class I mobile elements), which are quite rare among yeast species, and the level of amplification is unprecedented (up to 96 copies in one isolate). The authors show that different isolates of D. catenulata

differ by extensive amounts of chromosomal rearrangement (e.g. there are 60 breaks of synteny between two isolates; Fig 3B) and that the endpoints of these rearrangements tend to lie at Dici elements, which implicates Dici in the mechanism of rearrangement. By a phylogenetic analysis of 61 isolates, they show that this species had a deep clades structure, indicative of low (or zero) genetic exchange between different clades which may indicate reproductive isolation resulting from the genomic rearrangements driven by Dici elements.

5.1 Major comments

L492 and L830 It's very clear that Dici elements are over-represented at the endpoints of genomic rearrangements, but I would have liked to see one (or more) examples drawn in detail in a Figure. Are the rearrangements usually reciprocal translocations? Do they appear to involve ectopic recombination between two Dici elements? Or is there usually only a Dici element at one of the two new junctions formed in a rearrangement? Is it possible that some of the genomic rearrangements are the result of failed attempts to transpose?

Response: we agree that the nature of the rearrangements was not previously discussed. Most rearrangements are consistent with reciprocal translocations, typically involving two *Dici* copies on different chromosomes. We also observed cases compatible with ectopic recombination between two *Dici* elements located on the same chromosome, which leads to inversion. In some cases, no *Dici* copy has been observed at the recombination points. We propose a scenario in which chromosomal rearrangements occur during *Dici* transposition. Excision of the element generates a double-strand break (DSB) that may subsequently be repaired via error-prone pathways.

The corresponding section has been completely rewritten (section 2.5). As requested, we have added a new panel to Fig. 4 and generated a supplementary figure (Supplementary Fig. 4) illustrating different types of recombination events.

L691 This section of text concludes that CDS3 is likely a relic of a transposase gene, i.e. a pseudogene. If this is true, we would not expect to see evolutionary conservation of CDS3 between Dici1 and Dici2, which are only 52% identical in DNA sequence. Calculating measures of evolutionary conservation, such as the dN/dS ratio (nonsynonymous vs synonymous substitutions) should be informative about whether CDS3 (and also CDS1) has been under selection to conserve its ability to encode a protein, or has been evolving neutrally.

Response: we thank the reviewer for this suggestion. In response, we per-

formed an evolutionary analysis of CDS3 for both *Dici1* and *Dici2*, and extended this analysis to include CDS1 and the transposases. Although CDS3 lacks a conserved catalytic triad and was inferred to represent a relic of an ancestral transposase, our analyses revealed signatures of pervasive purifying selection. Specifically, four sites in CDS3 were inferred to be under purifying selection, with no evidence of episodic positive selection. Notably, CDS3 retains a recognizable homeodomain in both *Dici1* and *Dici2*, suggesting that despite the loss of the catalytic domain, selective constraints may be maintaining a residual DNA-binding or structural function, or alternatively reflect limited sequence flexibility in a degenerated remnant. In contrast, CDS1 and the transposase showed distinct evolutionary patterns, including signatures of episodic positive selection in *Dici2*, consistent with more heterogeneous selective pressures. These results have now been incorporated into a new section 2.8 (starting on line 980) to provide a comprehensive view of the evolutionary dynamics of the *Dici* coding sequences (Table 1).

L702 The text in Section 2.6, and Figure 5A, should be more assertive about the proposal that Dici is newly described transposon family that has members in many other fungal species too. Dici1 and Dici2 in D. catenulata are just 2 out of 24 transposons in this new family, all of which were previously unknown. It wasn't initially obvious to me that the green triangle in Fig 5A called "Dici/DD34E (n=24)" includes transposons from species other than D. catenulata. I suggest adding a panel to Fig 5, to show a zoomed-in view of the Dici clade that is currently shown in Figure S5, including the names of the species that contain members of the Dici family, and highlighting the positions of Dici1 and Dici2 within the Dici family. It is easy to overlook this point, because the name "Dici" is initially used in the manuscript as a collective name for Dici1 and Dici2 (section headings on L507, L608) but then on L702 it is used more broadly as the name of the newly-proposed family that exists in multiple species. I suggest mentioning that a new cross-species transposon family was discovered, in the Abstract, and around L162.

Response: this comment was found to be fully justified, and Section 2.6 has been revised to clarify the distinction between *Dici1*/*Dici2* elements and their counterparts in other fungi. The term "*Dici* family" was introduced once it became evident that these elements form a distinct clade (starting on line 1123) and is now used as a collective designation for clarity and consistency. As requested, Fig. 5 (new Fig. 6) was updated to include a new panel (panel B) showing an enlarged view of the *Dici*, Gambol, and Hiker clades, with transposase and species names indicated.

5.2 Minor comments

L043 The Abstract doesn't mention that the number of copies of Dici extends to at least 96 in one isolate, which is one of the most interesting discoveries in this study.

Response: as requested, the abstract now references this finding, and the text has been modified slightly to keep the word count below 200.

L076 Please provide more information about what a "cut-and-paste" mechanism means. How does the number of copies of the element increase, if it is lost from the original site and gained at a new site?

Response: as requested, a few sentences were added to the Introduction (lines 76-90) to describe the "cut-and-paste" transposition mechanism. We also expanded this section to discuss how increases in DNA transposon copy number can occur despite the conservative nature of the transposition mechanism.

L218 Please explain what PDO numbers mean, such as "PDO #43". Is it an anonymous designation for a particular type of cheese? Or a particular geographical location, or cheese producer? I am not asking for the name (which might be commercially sensitive), just a description of what the category means.

Response: a clearer definition of PDO has been added to the Methods section of the manuscript (starting on line 1350). PDO stands for Protected Designation of Origin and refers to a category of cheese produced in a specific geographical area. The coding of PDOs and producers has been clarified. In brief, PDOs were anonymized using numerical identifiers (e.g., PDO43), and producers within each PDO were further anonymized using letters (e.g., producers 43B and 43F are from PDO43). In some cases, cheeses from the same producer were obtained in different seasons; these samples were distinguished by adding a digit to the producer code (e.g., 43B1 stands for season 1, while 43B2 stands for season 2).

L307 This section needs to quantify the amount of sequence divergence between the clades, either by stating the average number of SNP differences between clades, or the percent nucleotide divergence in the genome. Some of this information is stated at a later point in the text (L524) but it should be stated earlier, in the context of Fig 1A.

Response: as requested, we discuss the average number of SNP differences between clades in a new paragraph in the first section of the Results (lines269-276). All values are provided in Supplementary Table 2 (sheet 2A).

L346 Change to "sequenced using long reads", to differentiate these isolates from the other 56 isolates that were sequenced only by Illumina short reads.

Response: as the requested, this was addressed in the text (line 384).

L366 Figure 4 is cited out of order, before Figures 2 and 3.

Response: this problem was also pointed by reviewer #1. The position of Figure 4 in the main text was corrected (new Fig. 2).

L428 Is Figure 2A drawn to scale? Please say this in the legend. It seems to show that there is a lot of non-coding DNA in the Dici elements, which is surprising for a mobile genetic element.

Response: Fig. 2A (new Fig. 3) has been edited to show the three CDSs at their correct scale. This is also indicated in the figure's legend. As requested by reviewer #1, we have also added more details about the size differences between *Dici1* and *Dici2*.

L526 The text implies that Supplementary Table 6 contains information about the sequence divergence between the different sequenced isolates, but this information isn't in that Table. I didn't see it in any Table.

Response: this information is indeed present in Supplementary Table 6 (sheet 6C). However, the conversion to PDF affected the table formatting. To improve accessibility during the revision process, the supplementary tables have been deposited on Zenodo (https://doi.org/10.5281/zenodo.17613032). Nucleotide divergence between the different isolates (% pairwise genome identity) are located above the diagonal (with yellow background on sheet 6C).

L537 and Figure 4 It isn't clear how the Y-axis in Figure 4 (transposon abundance, expressed in percent) relates to the number of copies of the transposons. The Methods section explains how the percentage was calculated (L1107) but an estimate of the actual copy number would be more informative. Can you not use the data from the 5 completely-sequenced isolates to convert these percentages into an estimated number of copies of the Dici elements? I was left wondering if any of the isolates have more than 96 copies.

Response: we have taken this recommendation into account. Genome abundance and estimated copy number are now shown simultaneously in Fig. 4 (new Fig. 2), allowing direct comparison across all 61 *D. catenulata* strains. We additionally comment on the prevalence and distribution of *Dici1* and *Dici2* in the

main text (starting on line 599). Based on these analyses, no isolate was found to harbor more than 96 copies of *Dici* elements. The highest copy numbers were observed in strain CBS 565 and its derivative UMIP805.63 (a homozygous diploid).

L990 I found the division of material between the Methods section and the Supplementary Methods section rather confusing. Supplementary Methods section 2.4 mentions that long-read assemblies were corrected using "Illumina reads", but these are not the Illumina RNA-seq reads that were mentioned in the previous paragraph (section 2.3). Instead, they are the Illumina genomic DNA reads that are not described at all in Supplementary Methods, but in section 4.6 of the main text. It would be more sensible to put all the sequencing details together in one place.

Response: we agree that this separation was confusing. The relevant Methods paragraphs from the supplementary text have been moved to the main text to better complement the main Methods section.

Pages 40-69 Formatting of the Supplementary Tables became messed up during conversion from Excel to PDF.

Response: We agree that converting the XLS tables to PDF format using the journal submission system is problematic. As stated above, we have deposited the supplementary tables on Zenodo (https://doi.org/10.5281/zenodo.17613032).

Dici: a novel DNA transposon reshaping the genome of the opportunistic yeast *Diutina catenulata*

– Second revision –

Response to reviewers

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

General response to the reviewers

We thank the reviewers for their careful reading of the revised manuscript and for the few additional comments, which we have addressed below.

Reviewer #1

Reviewer: I find the responses to my comments, as well as the others reviewers comments + responses thereof, satisfactory. I believe the manuscript is stronger now than before, especially due to the clearer wording, rearrangements, and additional experiment. This experiment (and it's simplicity) clearly demonstrates the flexible nature of D. catenulata genomes, and provides an interesting precedent for more of such experiments.

Response: We thank reviewer #1 for their positive assessment of the revised manuscript and for their encouraging remarks regarding the experimental evolution approach. We agree that this experiment opens interesting perspectives for future investigations of genome plasticity in *D. catenulata*.

Reviewer: The new Figure 2C & D are very interesting. They are very nice representations of the translocation junctions. As a pedant for figures, I wonder though what the green vertical line is right at the center of figure 2C?

Response: We believe that the reviewer #1 may refer to Figure 4C. The line in question was included to highlight the synteny breakpoint. Although it is not essential for the interpretation of the figure, we have removed it in the revised version.

Reviewer #2

Reviewer: I am satisfied with the responses provided and have no additional comments.

Response: We thank reviewer #2 for their positive assessment of our revised manuscript and for the time and effort devoted to reviewing our work.

Reviewer #4

Reviewer: The authors have done an excellent job of revising the manuscript, and they have addressed all of my previous comments to my satisfaction (reviewer #4). The new section on Dici element mobility during laboratory evolution is particularly exciting.

Response: We sincerely thank reviewer #4 for their positive comment of the revised manuscript, and for the constructive comments provided during the review process, which have significantly contributed to improving the quality of our work. We are particularly pleased that the new section on *Dici* element mobility during experimental evolution meets their expectations, as we also consider this a key contribution of the study.

Reviewer: I noticed two minor errors in Figures in the revised manuscript. In Figure 2A, the key lacks an explanation of the 5th item (blue). Is it Ty5 v1?

Response: We thank the reviewer for pointing out this oversight. The graphical legend was indeed incomplete for the *Ty* retrotransposon categories. A corrected version of Figure 2 has been generated, in which all items in the legend are now explicitly labeled.

Reviewer: In Figure 3B, the label GPRR should be GRPR, I think.

Response: We thank the reviewer for catching this typo. The label in Figure 3B has been corrected from GPRR to GRPR in the revised version.