bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dynamics of synthetic yeast chromosome evolution shaped by hierarchical chromatin organization

Authors: Sijie Zhou[1,2]†, Yi Wu[1,2]†, Yu Zhao[3]†, Zhen Zhang[1,2], Limin Jiang[4], Lin Liu[6], Yan Zhang[1], Jijun Tang[4,5], and Ying-Jin Yuan[1,2]*

Affiliations:

[1] Frontier Science Center for Synthetic Biology (Ministry of Education), Tianjin University, Tianjin, 300072, China

[2] Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, China

[3] Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

[4] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, 300350, China

[5] Department of Computer Science, University of South Carolina, SC 29208, USA

[6] Frasergen Bioinformatics Co., Ltd., Wuhan 430000, China

* Corresponding author: Ying-Jin Yuan, E-mail: yjyuan@tju.edu.cn.

† Equal contribution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract: Synthetic genomics provides a distinct approach to systematically explore the process of genome evolution in a dynamic manner. SCRaMbLE is an evolutionary system intrinsic to the synthetic yeast genome that can rapidly drive structural variations. Here, we detected over 260,000 rearrangement events after SCRaMbLEing of a synthetic strain harboring 6 synthetic yeast chromosomes. Remarkably, it was found that the rearrangement events exhibited a specific landscape of rearrangement frequency. We further revealed that the landscape was shaped by coordinated effects of chromatin accessibility and spatial contact probability. The rearrangements tend to occur in 3D spatially proximal and chromatin-accessible regions. Enormous numbers of rearrangements by SCRaMbLE provide a driving force to potentiate directed genome evolution and investigation of the population genomic resource offer mechanistic insights into the dynamic genome evolution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main Text:

Studying the processes and mechanisms of genome evolution is critical to understanding genetic diversity and species diversity at the genomic levels (*1-4*). *S. cerevisiae*, a powerful model organism for eukaryotic genome evolution (*5-7*) has been subjected to comparative genomic studies, which provided mechanistic insights underlying genome evolution (*8-12*). However, these studies relied on relatively static genomic sequences, and could have missed many details of dynamic processes. Synthetic genomes by *de novo* design and synthesis of genomic sequences incorporating a variety of strategies such as genome minimization (*13*), genetic codon recoding (*14*, *15*) and introduction of synthetic parts (*16*) have greatly facilitated the study of genome evolution. SCRaMbLE with symmetrical loxP recombination sites positioned downstream of the 3′ untranslated regions (3′UTRs) of all nonessential genes as part of the Synthetic Yeast Genome Project (Sc2.0) (*16*, *17*) has also been used for the study of genome evolution recently. Induced Cre recombinase activity quickly triggers rearrangements between loxPsym sites and generates structural variations, including deletions, inversions, duplications, and translocations (*18-25*), providing missing snapshots of dynamic genome evolution events.

Here we consolidated six synthetic chromosomes (synII (*26*), synIII (*27*), synV (*28*), synVI (*29*), synIXR (*17*), and synX (*30*)) in a single haploid strain based on an orthogonal site-specific recombination system, enabling eliminations of the entirely counterpart wild-type chromosomes. We comprehensively analyzed a SCRaMbLEd pool obtained using the poly-synthetic strain and detected over 260,000 rearrangement events via a loxPsym junction analysis method. By analyzing these rearrangement events, we uncovered a stable rearrangement landscape of the synthetic chromosomes that could be correlated to local chromatin structures and the three-dimensional genome architecture via assay for transposase-accessible chromatin using sequencing (ATAC-seq) and genome-wide chromosome conformation capture (Hi-C), respectively. Our results reveal that the rearrangements tend to occur in 3D spatially proximal and chromatin-accessible regions, which provides insights into the effect of hierarchical chromatin organization on the dynamic evolution of the genome.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Consolidation and SCRaMbLE of six synthetic chromosomes

We used a stepwise method to construct a single yeast haploid strain that harbors six synthetic chromosomes (synII, synIII, synV, synVI, synIXR, synX) (*17, 26-30*). We crossed two haploid strains of opposite mating types harboring synV&X (yYW169) and synII, III, VI&IXR (yZY192) respectively, resulting in a diploid strain (yYW268) (Fig. S1A). Next, we employed the Vika/vox, an orthogonal site-specific recombination system with Cre/loxP (*31, 32*), to excise the centromere from chromosomes and realize the elimination of a whole chromosome. The homologous wild type chromosomes (wtII, wtIX, wtV, wtX, wtVI, and wtIII) were eliminated successively. (Fig. S1B) (*33*). Following sporulation and tetrad dissecting, the haploid strain (yYW394), with six aforementioned wild type chromosomes replaced by synthetic chromosomes, comprising ~2.61 Mb (~22.0% of the yeast genome) (Table S1) was constructed (Fig. S1A), and confirmed by pulsed-field gel electrophoresis (Fig. S1C, Table S1) and genome sequencing (Fig. S2B). The yYW394 strain grew robustly at 30°C but exhibited growth defects at 37°C (Fig. S2A). Adaptive laboratory evolution of yYW394 was performed to recover the growth fitness generating a new strain (yZSJ025) (Fig. S1A).

A Cre recombinase expression plasmid pYW085 (pRS413-pCLB2-Cre-EBD) was transformed into yZSJ025, comprising 894 loxPsym sites in the 3′UTRs of nonessential genes (Fig. S3). SCRaMbLE was then induced by β-estradiol. The SCRaMbLEd cells were recultured in fresh YPD liquid medium without β-estradiol (Fig. 1A), and subjected to deep sequencing (~600,000×). All reads (150 bp each) were screened for the presence of loxPsym. Those with flanking sequences different from the references could be identified and hereafter called rearrangement reads, which were then used to identify and classify rearrangement events (*25*) (Fig. S4). Identical reads were considered as a result of a single rearrangement event. In total, 263,520 rearrangement events, including 124,499 (47.24%) (Fig. 1B) intra- and 139,021 (52.76%) (Fig. 1C) inter- chromosomal events were detected. We further analyzed the intra-chromosomal rearrangement events and found 62,106 (23.57%) inversions, 22,526 (8.55%) deletions and 39,867 (15.13%) complex rearrangement events (Fig. 1D left panel) (Fig. S4). The

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

numbers of identical reads represent frequencies of corresponding rearrangement events. When percentages of different types of rearrangements were calculated by read numbers (Fig. 1D right panel), the percentage of inter-chromosomal events reduced to 9.87%, indicating that these are relatively low frequency events compared to intra-chromosomal events. We next investigated whether there is chromosomal preference for rearrangement to occur. The total number of rearrangement reads for each chromosome was plotted against the number of loxPsym sites in that chromosome. The linear curve suggests no such preference is evident ($r = 0.96$, $p = 0.002$) (Fig. 1E).

A specific rearrangement pattern of the synthetic yeast chromosomes

In theory of Cre/loxPsym reactions, SCRaMbLE can generate random rearrangements between any two loxPsym sites on synthetic chromosomes. However, for different chromosomal loci, the results in Fig. 1B and 1C show high or low rearrangement frequencies. To estimate the local rearrangement frequencies along the synthetic chromosomes in our SCRaMbLEd pool, we mapped the number of rearrangement reads to each loxPsym site, generating a landscape of the rearrangement frequency (Fig. 2A). Rearrangement reads were identified in all 877 loxPsym sites. Clearly, loxPsym sites differed from each other in rearrangement frequencies, which range from 1,027 to 44,353 per site (Fig. 2A). The loxPsym sites with the highest and lowest rearrangement frequencies (90 each) were defined as hotspots and coldspots, and selected for further analyses (Fig. 2A, Table S4). The rearrangement frequencies of these spots are statistically different from the average frequency of the loxPsym sites (Fig. S5). The rearrangement frequency landscape of independent SCRaMbLE experiments is highly reproducible with biological replicates. ($r = 0.99$, $p = 1.7\times10^{-232}$) (Fig. 2B, Fig. S6A-F).

We also compared rearrangement patterns of the same synthetic chromosome in different strains containing varied numbers of incorporated synthetic chromosomes. Similar intra-chromosomal rearrangement patterns were observed for synV in the three synV-containing strains tested (yXZX846, yYW169 and yZSJ025) (Fig. 2C, Fig. S7A-C), as well as for synX in the two synX-containing strains tested (yYW169 and yZSJ025) (Fig. S7D, E). Overall, our results show that the patterns of SCRaMbLE are specific on each synthetic yeast chromosome and that synthetic chromosomes exhibit

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

relatively stable rearrangement hotspots and coldspots.

Rearrangement frequency correlates with chromatin accessibility

We then want to investigate the mechanisms underlying the rearrangement patterns observed in the previous subsection. SCRaMbLE requires physical interaction between Cre recombinase and loxPsym sites, suggesting the importance of chromatin accessibility in determining rearrangement frequency in SCRaMbLE cells. To test this hypothesis, genome-wide chromatin accessibility of the yZSJ025 strain was measured by ATAC-seq (*34*). The ATAC-seq signals of a window containing 400 bp upstream and downstream of each loxPsym site were collected and processed. The ATAC-seq signals of the rearrangement hotspots and coldspots were individually normalized by the average signals of 877 loxPsym sites. The statistical results show that the average ATAC-seq signals of hotspots are significantly strong, while signals of the coldspots are significantly weak (Fig. 3A). The rearrangement frequencies and ATAC-seq signals of a typical loxPsym coldspot and hotspot located in a region of synX were displayed in Fig. 3B. The ATAC-seq signals reveal that the coldspot at the 3′UTR of SET4 has a weak ATAC-seq signal, while the hotspot at the 3′UTRs of PRY3 has strong ATAC-seq signals. In addition, nucleosome occupancy of the yZSJ025 strain showed that nucleosome occupancy is low in the hotspots and high in the coldspots (Fig. S8). Both ATAC-seq and nucleosome occupancy data are consistent with our hypothesis that chromatin accessibility is critical to rearrangement frequency.

We then mapped the positions of the hotspots and coldspots to the wild type genome based on neighboring sequences. We extracted and processed the ATAC-seq for the wild type yeast as we did for the synthetic chromosomes (*34*) (Fig. S9). Results indicate that ATAC-seq signals peak at the positions of the hotspots, and remain weak at the coldspots in the six corresponding wild type chromosomes, suggesting that sequence modifications do not perturb chromatin accessibility.

Rearrangement frequency correlates with 3D chromatin organization

Next, we aimed to explore whether, besides chromatin accessibility, the rearrangement frequency is correlated with the spatial proximity of loxPsym sites. Our SCRaMbLE system provides a unique platform to statistically evaluate the role of

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spatial proximity in chromosomal rearrangements. A genomic chromosome conformation capture approach (Hi-C) was thus carried out for yZSJ025, generating a contact map that displays the frequencies of spatial contacts between any two genomic loci (Fig. S10). We then extracted the information of contact frequencies regarding solely the synthetic chromosomes (Fig. 3C). Rearrangement frequency was then plotted in a similar heatmap format to facilitate direct comparison (Fig. 3D). The diagonal regions represent loci close in primary sequence in each chromosome. Consistent with this, these intra-chromosomal regions are the "hottest" in the Hi-C map (Fig. 3C), and they also exhibit the highest rearrangement frequency (Fig. 3D). That rearrangements tend to occur most frequently between adjacent loxPsym sites for all synthetic chromosomes is shown by statistical analysis of all rearrangement reads (Fig. S11).

Notably, the inter-chromosomal contact probability in the regions of centromeres and telomeres is obviously higher than that in other regions (Fig. 3C), in consistence with centromeres being clustered around the spindle pole body, telomeres being clustered with the nuclear envelope for both wild type (*35*) and synthetic chromosomes (*36*) (Fig. 3E). Compared to other regions, the pericentromeric regions exhibit higher inter-chromosomal rearrangement frequency (Fig. 1C, 3D, 3F). Similar results were found for the peritelomeric regions of the synthetic chromosomes (Fig. 1C, 3D, 3G). Taken together, our results suggest that rearrangement events are generally more likely to occur between genomic loci with spatial proximity.

As both chromatin accessibility and spatial proximity affect rearrangement frequency, we seek to distinguish and weigh these two factors on individual events. Typical intra-chromosomal rearrangement events were selected and schematically demonstrated in Fig. 4A-C, based on actual experimental data. It is showed that: (i) The difference of the frequencies of two rearrangement events at loci with similar chromatin accessibility was determined by spatial proximity (Fig. 4A). (ii) The difference of the frequencies of two rearrangement events at loci with similar spatial proximity was determined by chromatin accessibility (Fig. 4B). (iii) Two rearrangement events with similar frequency differed in chromatin accessibility, which was compensated by the difference of spatial proximity (Fig. 4C).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

In this study, we detected over 260,000 rearrangement events in the SCRaMbLEd pool, revealing the tremendous plasticity of the yeast genome. Previous mechanistic studies of SCRaMbLE were mainly focus on the correlation of chromosomal rearrangement with nucleotide sequence (*19, 21, 23, 25, 37*). Here we revealed that the rearrangement frequency landscape is molded by chromatin accessibility and spatial proximity from an epigenetic perspective. The biochemical essence of SCRaMbLE is recombination between pairs of different loxPsym sites in the synthetic chromosomes catalyzed by Cre recombinase (*38*), we speculate that local chromatin accessibility affects rearrangement by affecting the ability of Cre to access loxPsym sites, while 3D chromatin organization affects the contact probability of two loxPsym sites. For the effect of spatial contact probability on rearrangement, previous studies also showed that intra-chromosomal structural variations in cancer are related to 3D genomic organization (*39-41*). Our SCRaMbLE system may provide a platform to enable the experimental characterization and mechanistic study of the role of genome rearrangement in disease.

The use of synthetic chromosomes and SCRaMbLE allowed us to statistically reveal that the inter-chromosomal rearrangement hotspots strikingly clustered at the peritelomeric and pericentromeric regions. Our results are consistent with previous reports that increased inter-chromosomal reshuffling occurred in the peritelomeric regions of domesticated and wild yeast isolates during genome evolution (*8, 11*). The peritelomeric regions are functionally enriched for secondary metabolisms and stress responses that contribute to environmental adaptation (*8, 11, 42-44*), high frequencies of rearrangement in these regions could thus be important driving force of evolution. As for the pericentromeric regions, inter-chromosomal rearrangements would result in swapping of chromosomal arms between two chromosomes. Recent genomic studies demonstrated that such rearrangements can cause reproductive isolation and promote incipient speciation (*45-47*). We speculate that chromatin structures play important roles in genome evolution in terms of effects to the rearrangements of the peritelomeric and pericentromeric regions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The flexible and controllable synthetic yeast genome provides a unique model to systematically interrogate and explore the dynamic process of eukaryotic genome evolution. The investigation of the large number of rearrangements indicated the tremendous plasticity of the yeast genome and the importance of hierarchical chromatin organization in regional rates of chromosomal variations for genome evolution. These findings give crucial insights that the variability and complexity of synthetic genome design can be further increased; meanwhile, influence of chromatin organization needs to be considered for the design and engineering of higher organism genomes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References and Notes

1. W. Jasinska *et al.*, Chromosomal barcoding of *E. coli* populations reveals lineage diversity dynamics at high resolution. *Nature Ecology & Evolution* **4**, 437-452 (2020).
2. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490-493 (2019).
3. A. N. Nguyen Ba *et al.*, High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494-499 (2019).
4. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45-50 (2017).
5. J. Peter, J. Schacherer, Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* **33**, 73-81 (2016).
6. B. Gallone *et al.*, Domestication and Divergence of *Saccharomyces Cerevisiae* Beer Yeasts. *Cell* **166**, 1397-1410 (2016).
7. D. Botstein, G. R. Fink, Yeast: An Experimental Organism for 21st Century Biology. *Genetics* **189**, 695-704 (2011).
8. J. Peter *et al.*, Genome evolution across 1,011 *Saccharomyces Cerevisiae* isolates. *Nature* **556**, 339-344 (2018).
9. X. Shen *et al.*, Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533-1545 (2018).
10. S. Marsit *et al.*, Evolutionary biology through the lens of budding yeast comparative genomics. *Nat. Rev. Genet.* **18**, 581-598 (2017).
11. J. Yue *et al.*, Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* **49**, 913-924 (2017).
12. B. Dujon, Yeast evolutionary genomics. *Nat. Rev. Genet.* **11**, 512-524 (2010).
13. C. A. Hutchison *et al.*, Design and synthesis of a minimal bacterial genome. *Science* **351**, d6253 (2016).
14. N. Ostrov *et al.*, Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819-822 (2016).
15. J. Fredens *et al.*, Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514-518 (2019).
16. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).
17. J. S. Dymond *et al.*, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
18. J. Wang *et al.*, Ring synthetic chromosome V SCRaMbLE. *Nat. Commun.* **9**, 3783 (2018).
19. Y. Wu *et al.*, *In vitro* DNA SCRaMbLE. *Nat. Commun.* **9**, 1935-1939 (2018).
20. W. Liu *et al.*, Rapid pathway prototyping and engineering using *in vitro* and *in vivo* synthetic genome SCRaMbLE-in methods. *Nat. Commun.* **9** (2018).
21. Z. Luo *et al.*, Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. *Nat. Commun.* **9** (2018).
22. B. A. Blount *et al.*, Rapid host strain improvement by *in vivo* rearrangement of a synthetic yeast chromosome. *Nat. Commun.* **9** (2018).
23. B. Jia *et al.*, Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nat. Commun.* **9** (2018).
24. M. J. Shen *et al.*, Heterozygous diploid and interspecies SCRaMbLEing. *Nat. Commun.* **9** (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

25. Y. Shen *et al.*, SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36-49 (2016).

26. Y. Shen *et al.*, Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, f4791 (2017).

27. N. Annaluru *et al.*, Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science* **344**, 55-58 (2014).

28. Z. Xie *et al.*, "Perfect" designer chromosome V and behavior of a ring derivative. *Science* **355**, f4704 (2017).

29. L. A. Mitchell *et al.*, Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355** (2017).

30. Y. Wu *et al.*, Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355** (2017).

31. Q. Lin, H. Qi, Y. Wu, Y. Yuan, Robust orthogonal recombination system for versatile genomic elements rearrangement in yeast *Saccharomyces Cerevisiae*. *Sci. Rep.-UK* **5**, 15249 (2015).

32. M. Karimova *et al.*, Vika/vox, a novel efficient and specific Cre/loxP-like site-specific recombination system. *Nucleic Acids Res.* **41**, e37 (2013).

33. H. Xu *et al.*, Chromosome drives via CRISPR-Cas9 in yeast. *Nat. Commun.* **11** (2020).

34. A. N. Schep *et al.*, Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757-1770 (2015).

35. Z. Duan *et al.*, A three-dimensional model of the yeast genome. *Nature* **465**, 363-367 (2010).

36. G. Mercy *et al.*, 3D organization of synthetic and scrambled chromosomes. *Science* **355**, f4597 (2017).

37. P. Wang *et al.*, SCRaMbLEing of a Synthetic Yeast Chromosome with Clustered Essential Genes Reveals Synthetic Lethal Interactions. *ACS Synth. Biol.* **9**, 1181-1189 (2020).

38. F. Guo, D. N. Gopaul, G. D. van Duyne, Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-46 (1997).

39. K. Hadi *et al.*, Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell* **183**, 197-210 (2020).

40. Y. Zhang *et al.*, Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. *Cell* **148**, 908-921 (2012).

41. G. Fudenberg, G. Getz, M. Meyerson, L. A. Mirny, High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat. Biotechnol.* **29**, 1109-1113 (2011).

42. A. Bergström *et al.*, A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. *Mol. Biol. Evol.* **31**, 872-888 (2014).

43. C. A. Brown, A. W. Murray, K. J. Verstrepen, Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts. *Curr. Biol.* **20**, 895-903 (2010).

44. R. M. Ames *et al.*, Gene Duplication and Environmental Adaptation within Yeast Populations. *Genome Biol. Evol.* **2**, 591-601 (2010).

45. K. Guin *et al.*, Spatial inter-centromeric interactions facilitated the emergence of evolutionary new centromeres. *ELife* **9** (2020).

46. S. R. Sankaranarayanan *et al.*, Loss of centromere function drives karyotype evolution in closely related Malassezia species. *ELife* **9** (2020).

47. V. Yadav, S. Sun, M. A. Coelho, J. Heitman, Centromere scission drives chromosome shuffling and reproductive isolation. *Proc. Natl. Acad. Sci. USA.* **117**, 7917-7928 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments: We thank Jef D. Boeke from New York University for fruitful discussions and advice. We are grateful to Jef D. Boeke and Leslie A. Mitchell from New York University, Huanming Yang, Yue Shen, Yun Wang and Tai Chen from BGI-Shenzhen, Yizhi Cai from University of Edinburgh, Srinivasan Chandrasegaran, Joel Bader, Narayana Annaluru, Héloïse Muller and Jessica S. Dymond from Johns Hopkins University for sharing the synthetic yeast strains and technical support. This work is part of the Synthetic Yeast Genome Project, http://syntheticyeast.org/sc2-0/.

Funding: This work was funded by the National Natural Science Foundation of China (21621004, 31861143017, 31971351), National Key R&D Program of China, Synthetic Biology Research (2019YFA0903800), New Drug Creation Manufacturing Program (No. 2019ZX09J19105) and Young Elite Scientist Sponsorship Program by CAST (YESS) (2018QNRC001).

Author contributions:

Conceptualization: YJY, YW; experiments: SZ, YZ (construction of strain yZY192), ZZ; analysis: YJY, YW, SZ, ZZ, LJ, LL; writing: YJY, YW, SZ, YZ, JT.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Genome sequencing data have been submitted to NCBI Sequence Read Archive (SRA) under accession number PRJNA705059. The ATAC-seq and Hi-C sequencing data have been submitted to NCBI Gene Expression Omnibus (GEO) under accession number GSE168182. The ATAC-seq data of wild type *S. cerevisiae* is from GEO under accession number GSE66386.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S11

Tables S1 to S6

References (48–51)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

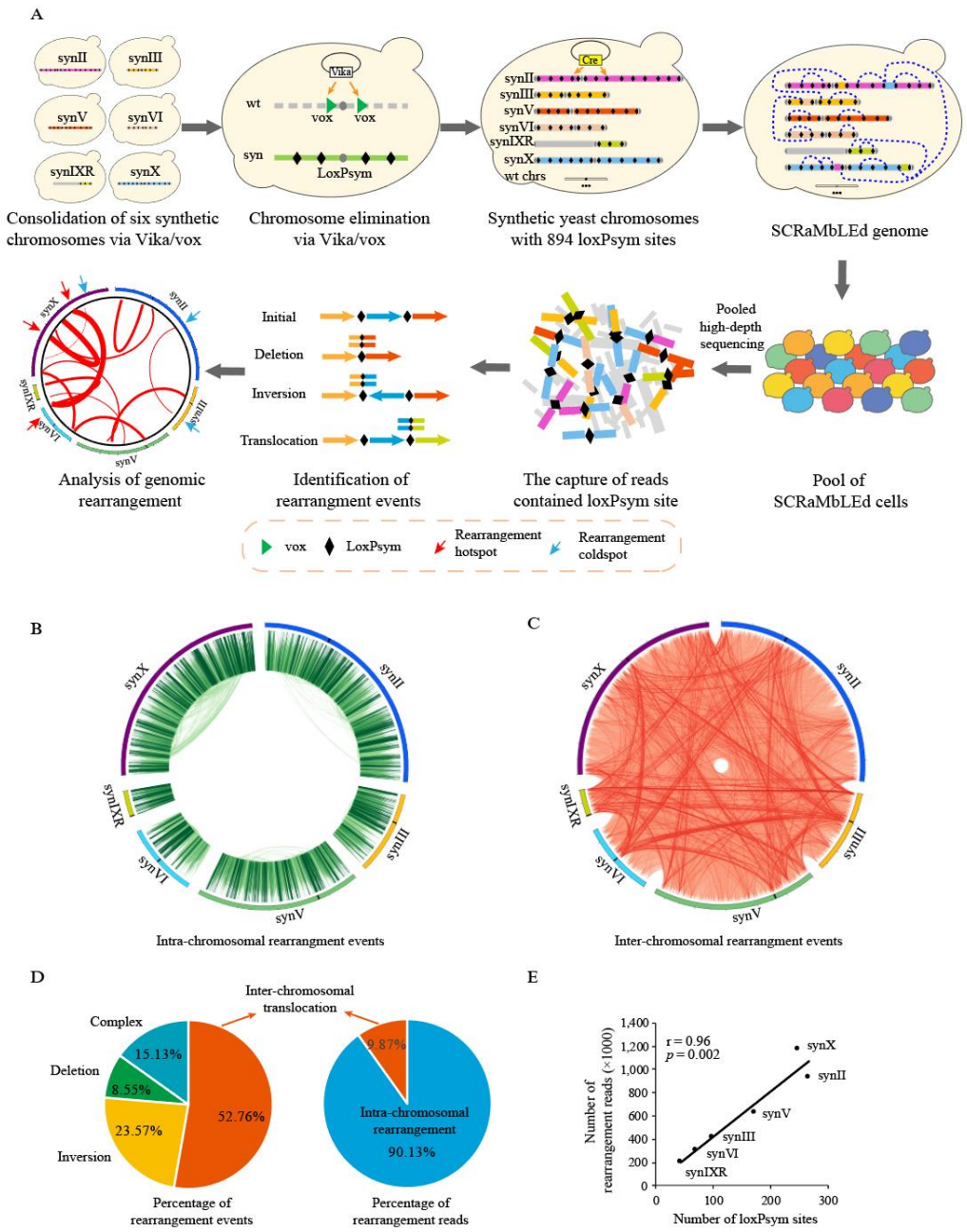

Fig. 1. Consolidation and SCRaMbLE of six synthetic chromosomes. (**A**) Schematic diagram showing the construction of the synthetic yeast strain yZSJ025 and the analyses of rearrangement events following SCRaMbLE. (**B**) Circos plot of intra-chromosomal rearrangement events in yZSJ025. (**C**) Circos plot of inter-chromosomal rearrangement events in yZSJ025. For (B) and (C), each line connects two loxPsym sites, where a rearrangement event occurs with line intensities representing the frequencies of the event. (**D**) Classifications of rearrangement events with percentages of different groups calculated by event numbers and read numbers respectively. Different groups of events were labeled in different colors as indicated. (**E**) Correlation analyses between the rearrangement frequency and the number of loxPsym sites on

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

each chromosome. Pearson correlation analysis was applied to determine the correlation coefficient and associated p values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

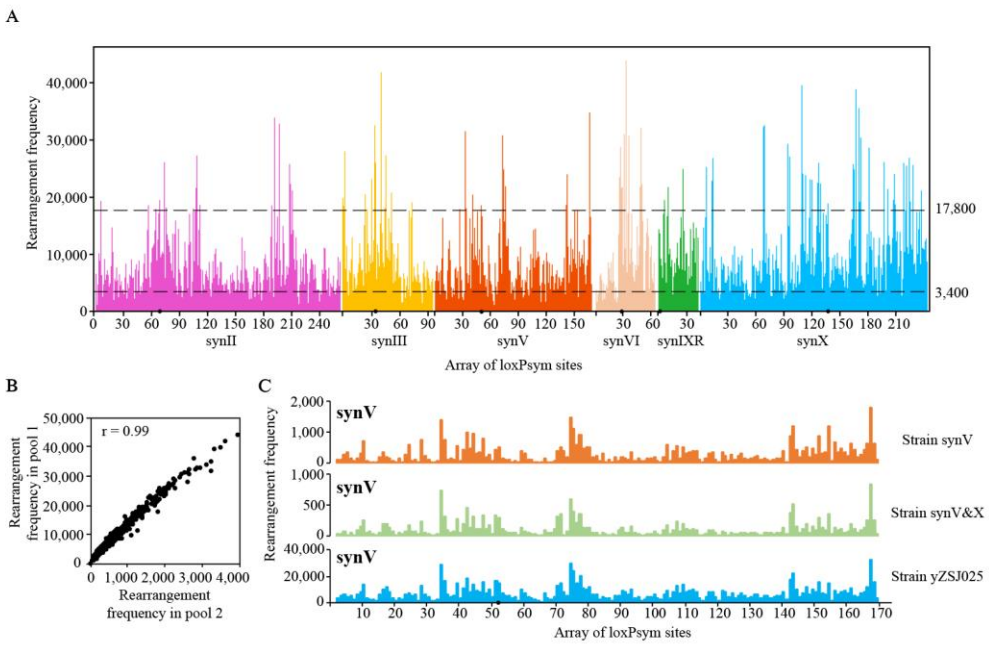

Fig. 2. Investigating rearrangement frequencies reveals a specific rearrangement pattern among the synthetic yeast chromosomes. (**A**) Landscape of the rearrangement frequency along synII, synIII, synV, synVI, synIXR, and synX. Rearrangement frequencies of hotspots were higher than 17,800 per site and coldspots lower than 3,400 per site. 17 loxPsym sites were initially inserted at the synthetic telomere regions, with flanking sequences not distinguishable from each other, and were excluded from the identification of rearrangements reads. (**B**) Comparison of rearrangement frequencies in two replicate pools of SCRaMbLEd cells (r = 0.99, $p = 1.7\times10^{-232}$). (**C**) Intra-chromosomal rearrangement patterns of synV following SCRaMbLE in three different yeast strains, synV (yXZX846), synV&X (yYW169) and yZSJ025 respectively. Pearson correlation analysis was applied to determine the correlation coefficient and associated p values. For (A) and (C), Black dots indicate centromeres.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

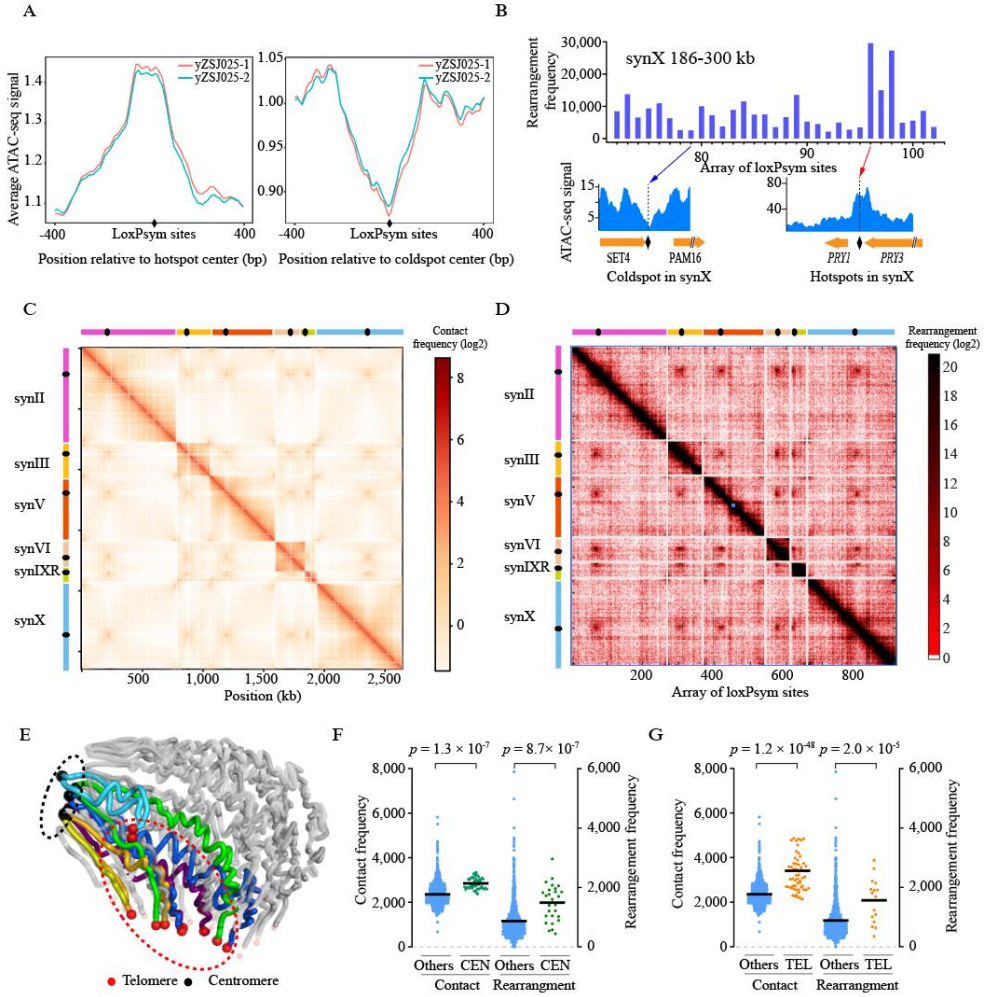

Fig. 3. Rearrangement frequency in correlation with chromatin accessibility and 3D chromatin organization (A) Average ATAC-seq signals of the hotspot- and coldspot- centered 800 bp regions respectively. **(B)** ATAC-seq signals for a typical rearrangement hotspot and a coldspot in the region 186-300 kb of synX. (**C**) Hi-C heatmap. The heatmap value for a site (*i, j*) is the contact probability (log2) between genomic loci *i* (horizontal axis) and *j* (vertical axis). Both axes were displayed at a 1 kb resolution. Spots with different heatmap values from low to high were colored from light yellow to red as indicated. (**D**) Rearrangement frequency heatmap. The heatmap value for a site (*i, j*) is the rearrangement frequency (log2) of the event occurring at the loci *i* (vertical axis) and *j* (horizontal axis). Spots with different heatmap values were labeled in red with different intensities as indicated. (**E**) 3D structures of synthetic chromosomes inferred from the Hi-C contact map displayed in panel (C). Synthetic chromosomes were labeled with different colors and the native chromosomes were shown in grey. Centromeres and telomeres were represented by black and red spheres respectively. (**F**) Comparisons of the contact probability and frequency of rearrangement events at loci in pericentromeric regions ('CEN') and in other regions ('Others'). The CEN were centromere-centered 10 kb regions by definition. (**G**)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comparisons of the contact probability and frequency of rearrangement events at loci in peritelomeric regions ('TEL') and in other regions ('Others'). The TEL were $\leqslant$5 kb apart from telomeres by definition. For (E) and (F), each data point represents 1 kb bin in regions as indicated. Horizontal lines indicate weighted means. The *p* values were calculated using two-tailed paired two-sample *t*-tests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.19.453002; this version posted October 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

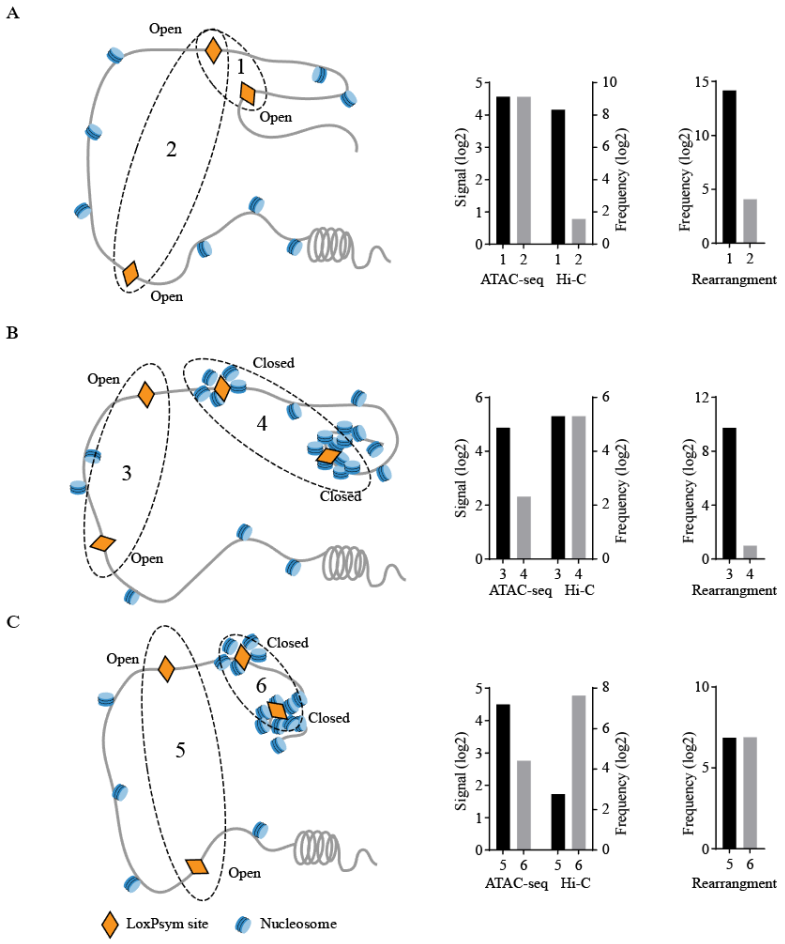

Fig. 4. Mechanistic models of the effects of hierarchical chromatin organization on chromosomal rearrangement. (A) A schematic model interpreting the difference of the frequencies of two rearrangement events with different contact probability. **(B)** A schematic model interpreting the difference of the frequencies of two rearrangement events with different chromatin accessibility. **(C)** A schematic model interpreting two rearrangement events with similar frequencies as a result of counter-effects by contact probability and chromatin accessibility. The loci of loxPsym sites involved in rearrangement events 1-6 (circled) were at synX 183 kb and synX 184 kb, synX 184 kb and synX 342 kb, synII 186 kb and synII 204 kb, synV 346 kb and synV 384 kb, synVI 125 kb and synVI 202 kb, and synII 504 kb and synII 507 kb respectively.