

3D organization of synthetic and scrambled chromosomes

Guillaume Mercy, Julien Mozziconacci, Vittore F Scolari, Kun Yang, Guanghou Zhao, Agnès Thierry, Yi Luo, Leslie A Mitchell, Michael Shen, Yue Shen, et al.

▶ **To cite this version:**

Guillaume Mercy, Julien Mozziconacci, Vittore F Scolari, Kun Yang, Guanghou Zhao, et al.. 3D organization of synthetic and scrambled chromosomes. Science, 2017, 355 (6329), pp.eaaf4597. 10.1126/science.aaf4597 . pasteur-01523221

HAL Id: pasteur-01523221

https://pasteur.hal.science/pasteur-01523221

Submitted on 16 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

3D organization of synthetic and scrambled chromosomes

Guillaume Mercy[1,2,3]†, Julien Mozziconacci[4]†, Vittore F. Scolari[1,2], Kun Yang[5], Guanghou Zhao[9], Agnès Thierry[1,2], Yisha Luo[12], Leslie A. Mitchell[6], Michael Shen[6], Yue Shen[10,11,12], Roy Walker[12], Weimin Zhang[9], Yi Wu[7], Ze-xiong Xie[7], Zhouqing Luo[9], Yizhi Cai[12], Junbiao Dai[9], Huanming Yang[8,10], Ying-Jin Yuan[7], Jef D. Boeke[6], Joel S. Bader[5], Héloïse Muller[1,2,13]*, Romain Koszul[1,2]*

[1] Institut Pasteur, Department Genomes & Genetics, Spatial Regulation of Genomes, Paris, 75015, France.

[2] UMR3525, Centre National de la Recherche Scientifique, Paris, 75015, France.

[3] Sorbonne Universités, UPMC Université Paris 6, 75005, Paris, France.

[4] Laboratoire de Physique Théorique de la Matière Condensée, CNRS UMR 7600, Université Pierre et Marie Curie, Sorbonne Universités, Paris, France.

[5] Department of Biomedical Engineering and High-Throughput Biology Center, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[6] Institute for Systems Genetics and Department of Biochemistry & Molecular Pharmacology, NYU Langone Medical Center, New York, NY 10016.

[7] Key Laboratory of Systems Bioengineering (Ministry of Education), SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, PR China.

[8] James D. Watson Institute of Genome Sciences, Hangzhou 310058, China.

[9] Key laboratory for Industrial Biocatalysis (Ministry of Education), Key laboratory of Bioinformatics (Ministry of Education), Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

[10] BGI-Shenzhen, Shenzhen 518083, China.

[11]BGI-Qingdao, Qingdao 266555, China.

[12] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK.

[13] Present address: Institut Curie, UMR3664 Dynamique du Noyau, Paris, France.

*Correspondence to: heloise.muller@curie.fr, romain.koszul@pasteur.fr.

†These authors contributed equally.

Abstract:

While the design of the synthetic yeast genome Sc2.0 is highly conservative regarding gene content, the deletion of several classes of repeated sequences and the introduction of thousands of designer changes may affect genome organization and potentially alter cellular functions. We report here the Hi-C determined 3D conformations of Sc2.0 chromosomes. While the absence of repeats leads to a smoother contact pattern and more precisely tractable chromosome conformation, the large scale genomic organization is generally globally unaffected by the presence of synthetic chromosome(s). An exception is *synIII* which lacks the silent mating type cassettes, and *synXII* from which the rDNA has been moved to another chromosome. We also exploit the contact maps to detect rearrangements induced in SCRaMbLE strains.

Introduction

Genes of the budding yeast *Saccharomyces cerevisiae* genome are non-randomly distributed along its 16 chromosomes. For instance, centromere- and telomere-proximal regions are enriched in families of co-regulated genes (*1–4*). Chromosome 3D organization has been studied through imaging and genetic studies (*5–8*), revealing a characteristic "Rabl-organization" (*9*). Four main factors drive the overall organization: i) 16 centromeres tethered and clustered at the spindle pole body (SPB, the yeast microtubule organizing center), ii) telomeres forming smaller groups tethered at the nuclear envelope (NE), iii) a single nucleolus across the nucleus from the SPB where the rDNA is sequestered from other chromosomes, and iv) chromosomal arm lengths. In addition to these factors, several sequences have been shown or suggested to influence the organization. For instance, tRNA genes were proposed to influence chromosome folding through re-positioning in the vicinity of either the SPB or the nucleolus (*10*, *11*). Telomeric repeats may also play a role in telomere clustering and anchoring to the NE (*8*, *12*). Sc2.0 genome design specifies strong conservation of gene content and arrangement with respect to the native yeast genome (*13*). However, synthetic chromosomes encode thousands of designer changes intended to improve genetic stability and increase genetic flexibility (*13*). For instance, tRNA genes were deleted during design of Sc2.0 chromosomes, to be relocated to a separate 'neochromosome' (*13*), while subtelomeric regions were substantially altered as well, with large repeated sequences corresponding either to Y' or genes families deleted. Further, *loxPsym* sites encoded by Sc2.0 chromosomes enable inducible evolution by 'SCRaMbLE' (*14*, *15*) to generate combinatorial genomic diversity through rearrangements (*16*). Several synthetic chromosomes are now built (*14*, *17–22*); it is thus possible to experimentally address whether or not Sc2.0

modifications affect the overall chromosome organization in strains carrying synthetic chromosomes.

Genomic chromosome conformation capture approaches (e.g. Hi-C; (*23*, *24*)), which provide access to the average 3D chromosome organization from a population of cells, highlighted the Rabl organization of the *S. cerevisiae* and other fungi genomes (Fig. 1A; (*25*, *26*)). In a Hi-C experiment, ligation frequencies between DNA restriction fragments are quantified through deep sequencing, reflecting contacts within the nuclear space and presumably their averaged respective positions in populations of cells. These frequencies are usually represented as 2D heat maps (fig S1) and can be processed and visualized as 3D representations to facilitate their interpretation (Fig. 1A; (*27*)). To investigate Sc2.0 chromosome organization, Hi-C contact maps (bin=5kb) and 3D representations (see Material and Methods for interpretation of such structures) of *synII*, *synIII*, *synV*, *synVI*, *synIXR*, *synX* and *synXII* were generated from daughter cells synchronized through elutriation (Fig. 1B, C and D; fig S2-10; tables S1-2; movies S1-10; (*28*)).

Results

Trajectories of synthetic chromosomes in the nuclear space

To determine whether individual or combinations of Sc2.0 chromosomes influence the overall genome organization, the contact map of the subset of invariant (i.e. native) chromosomes was generated for each strain. Euclidean distances between these maps were then computed and revealed that the presence of one or more Sc2.0 chromosomes did not dramatically affect the

overall structure of the remaining genome (fig S11). The 3D representations also allow for the qualitative comparison of native and *syn* sequences, revealing that the average trajectories of individual Sc2.0 chromosome did not appear significantly altered compare to their native counterparts, with *syn* chromosomes neighboring the same chromosomes as the native ones (Fig. 1 and fig S2-10 for side by side comparisons). To quantitatively compare the internal folding of Sc2.0 chromosomes with their native counterparts, we computed the contact frequency decay as a function of the increasing genomic distance for all pairs of loci. No substantial changes between synthetic and native chromatin were detected (Fig. 1E), suggesting that the redesigned sequence has little if no effect on the internal folding of the chromosome beyond a ~5-10% shortening through the removal of repeats.

Sc2.0 design improves mappability

A comparison of Sc2.0 maps with their native counterparts nevertheless revealed remarkably "smoother" contact patterns in the designer chromosomes, as shown by the quasi-absence of white lines in *syn* contact maps. This stems from the intentional deletion of repeated elements relative to the native sequence, which leads to improved "mappability" along the length of the synthetic chromosome (Fig. 1E and F and fig S2-10). For instance, for chromosome II, eight bins (5% of the chromosome) appeared insufficiently covered in the WT when its synthetic counterpart had none (fig S2 and fig S12). In addition, the normalized contact frequencies between *synVI* subtelomeres appear to be much higher than in the WT counterpart (fig S13). This variation results directly from the improved mappability of subtelomeric regions following the removal of subtelomeric repeats. On the other hand, subtelomeric contacts are underestimated in

native chromosome because of these repetitive sequences. These observations highlight the conceptual advance brought by bottom-up chromosome design and synthesis to investigate fundamental questions, as scientists can now design at unprecedented levels of refinement experimental systems making complex questions easier to address experimentally. For instance, investigating the elusive regulatory influences of repeated sequences in mammalian genomes could benefit from similar techniques.

SynIII conformation is modified by the design

Specific contacts between the duplicated *HML* and *HMR* silent mating-type loci that bridge the two arms of chromosome III have been identified (*23*, *29*). The deletion of these cassettes led to a loss of contacts on *synIII* (Fig. 2A, B). Moreover, chr III shows a mating-type specific conformation that depends on the integrity of a recombination enhancer sequence (RE) (*30*). Despite the presence of the RE in *synIII*, the mating-type specific conformation is lost (fig S14). We speculate that the deletion of the silent loci in synIII underlies this result, but additional experiments are required to discriminate this possibility from other changes in the sequences. Note that *synIII* still exhibits enriched contacts between its subtelomeres, a feature characteristic of the small metacentric chromosomes *I, III,* and *VI* (*12*, *31*).

Re-positioning the rDNA locus affects the global 3D structure of the genome

The rDNA locus on the right arm of chromosome XII consists of a cluster (100-200) of ~9kb rDNA units, each containing two ribosomal precursor genes. This cluster assembles in the nucleolus, a discrete, crescent shaped nuclear compartment that occupies ~1/3rd of the total nuclear space and is positioned opposite to the SPB. Two versions of the *synXII* chromosome

were designed, with and without the rDNA cluster. When not located on *synXII*, the rDNA was positioned either on the multicopy 2µ plasmid (one unit per plasmid) or reintegrated as an array within the right arm of the small chromosome III (*22*). Chromosome organization was investigated in these strains. Because of their repetitive nature, rDNA unit reads cannot be uniquely mapped, and the rDNA cluster on the 3D reconstruction thus appears as an empty space flanked by the closest regions that can be mapped unambiguously (Fig. 1A and Fig. 3A, B and C). In strains without the rDNA locus on *synXII* (JDY446, JDY448 and JDY449), the continuity of the two regions flanking the original insertion site was clearly visible on the contact map and 3D representations of the right arm (Fig. 3A, B and C and fig S5-6). The presence of 2µ plasmids carrying rDNA unit (strain JDY446) did not result in a large reorganization of the genome (fig S5), suggesting that the plasmids do not form large discrete heterochromatic structures, in agreement with the studies reporting the formation of multiple small clusters (*22*, *32*). On the other hand, the insertion of the rDNA array in the middle of the right arm of chromosome III (strains JDY448 and JDY449) resulted in substantial reorganization of the genome (Fig. 3C and fig S6). The right arm of chromosome III now appears split into two non-interacting regions by the rDNA cluster, an arrangement similar to the native chromosome XII structure (*25*). Relocating the rDNA to a short chromosomal arm imposes new constraints on the genome: the nucleolus is now trapped between short and long chromosome arms, forming a physical barrier between them as shown by reduced contacts between these arms (Fig. 3C). In addition, contacts between the rDNA and subtelomeric regions are increased overall (Fig. 3D), as a result of the greater proximity of the nucleolus to the entire set of chromosomal extremities. Also, the IIIR subtelomere in this strain is now contacting long arms subtelomeres (blue arrow on Fig. 3C). No growth rate defects were detected in these strains, suggesting that the position of the rDNA locus

has little effect on fitness in laboratory growth conditions tested here. However, given the observed constraints, it would be informative to run long-term cultures to see if the rDNA locus spontaneously re-localizes, and if so, where.

Hi-C analysis of SCRaMbLE chromosomes

All synthetic chromosomes carry the "synthetic chromosome rearrangement and modification by *LoxP* mediated evolution" or SCRaMbLE system, a chromosome-wide expansion of the Cre/lox site specific recombination assay exploited to measure local DNA concentrations in genomes (*5*, *33*). The Sc2.0 design includes the integration of hundreds of *loxPsym* sites all along chromosomes, at the end of every non-essential gene and at specific landmarks. Upon activation of the Cre recombinase, two *loxPsym* will recombine if they encounter each other in the nuclear space, leading to a structural variant (SV). The hundreds of sites therefore hold a large combinatory potential, leading to rapid and complex structural changes of synthetic chromosomes. Previous analysis of *synIX*R SCRaMbLE strains has revealed a large variety of *cis* events ranging from simple duplications, inversions and deletions to highly reorganized structures (*14*, *16*). We took the opportunity offered by strains carrying two Sc2.0 chromosomes to explore further the power of the SCRaMbLE recombination assay in *trans* (*14*, *15*). The Cre recombinase was induced for 48h in strain yLM539 carrying *synIII* and *synIXR* (Fig. 4A; fig S15). Independent clones were sampled during the time course experiment and two clones, HMSY029 (T_2 = 2h) and HMSY030 (T_8 = 8h), were retained for in-depth analysis. Genome-wide Hi-C contact data were recently shown to provide a convenient way to identify and solve chromosomal rearrangements (fig S16; (*34–37*)). We therefore applied Hi-C to investigate the consequences of genome structural alterations in these two isolates. The contact maps identify

gross chromosomal rearrangements (GCRs) in both SCRaMbLE strains, with each strain exhibiting deletion and translocation events, the latter representing a class of *trans* SCRaMbLE events not previously observed (Fig. 4B). The resolution of the contact map (~2 to 5 kb) allowed identification of relatively large scale rearrangements, but *loxPsym* sites are often more closely spaced. To address this, a refined analysis of the reads coverage was performed (table S3; fig S17), unveiling multiple short-scale modifications of *loxPsym* flanked segments (summarized in Fig. 4C; see fig S18 for details). These results highlight the potential of the SCRaMbLE design to generate GCRs, with more rearrangements being achieved after longer induction times. Incidentally, Hi-C analysis of Sc2.0 strains identified several chromosomal rearrangements that spontaneously arose during chromosome assemblies such as aneuploidy, duplication or translocation events outside of the synthetic chromosome of interest (table S1, fig S19; (*17*)). Hi-C is therefore a convenient way to validate the assembly and subsequent processing of strains carrying synthetic chromosomes.

Conclusion

Overall, this large dataset represents a resource that will be exploited in future studies exploring the power of the SCRaMbLE system. By showing that the spatial organization of the Sc2.0 genome is not significantly altered, this work pioneers future studies addressing the influence of genome-wide engineering approaches on essential features of living systems.

Materials & Methods

Media

Yeast were grown either in YPD [1% (w/v) Bacto peptone (Difco) 1% (w/v) Bacto yeast extract (Difco), 2% (w/v) glucose and 2% (w/v) Bactoagar] or in selective SC media [0.67% Yeast Nitrogen Base without amino acids (Difco), supplemented with an amino acid mix minus those whose prototrophy is selected for, 2% (w/v) glucose and 2% (w/v) Bactoagar].

Strains

All strains used in the study are described in Table S1.

Culture conditions

Cells were inoculated and grown overnight in 10 ml YPD. Overnight cultures were then diluted and grown to exponential phase in 500 ml YPD for ~15h at 30°C. For SCRaMbLE and mating-type specific organization experiments, Hi-C libraries were generated from asynchronous cultures (see Table S2). For all other Sc2.0 strains, Hi-C libraries were generated from daughter cells recovered using elutriation and restarted for 30 min in YPD at 30°C (*38*).

Generation of Hi-C libraries

Hi-C libraries were generated with a protocol adapted from (*23*) with introduction of a biotin-ligation step (*24*). Briefly, aliquots of 1-3 x 10^9 cells were cross-linked for 30 minutes with fresh formaldehyde (3% final concentration) and quenched with glycine for 15 min. Pelleted cells were dissolved in 10 ml sorbitol 1M and incubated for 30 minutes with DTT 5mM and Zymolyase 100T (C_{Final}=1 mg/ml). Spheroplasts were washed first with 5 ml sorbitol 1M,

then with 5 ml 1X restriction buffer (DpnII NEB buffer), and suspended in 3.5 ml of 1X restriction buffer. Cells were split into aliquots (V=500µL) and incubated in SDS (3%) for 20 minutes at 65°C. Crosslinked DNA was digested at 37°C overnight with 150 units of DpnII restriction enzyme (NEB). The digestion mix was subsequently centrifuged 20 minutes at 18000 g and the supernatant discarded. Pellets were suspended and pooled into 400 µL of cold water. DNA ends were repaired in the presence of 14-dCTP biotin (Invitrogen), and crosslinked complexes incubated for 4 hours at 16°C in presence of 250 U of T4 DNA ligase (Thermo Scientific, 12.5 mL final volume). DNA purification was achieved through an overnight incubation at 65°C in presence of 250µg/mL proteinase K in 6.2mM EDTA followed by a precipitation step and RNAse treatment.

The resulting Hi-C libraries were sheared and processed into Illumina libraries according to manufacturer instructions (Paired-End DNA sample Prep Kit – Illumina – PE-930-1001) using custom-made versions of the Illumina PE adapters (*28*). Fragments of sizes between 400 and 800 bp were purified on from 2% agarose gel, PCR amplified, and paired-end (PE) sequenced on an Illumina platform (HiSeq2000 or NextSeq500, Table S2).

Raw sequences are accessible on SRA database through the following accession number: SRP070421.

Generation and normalization of contact maps

Sequencing pair-end data was processed as follow. For each library, PCR duplicates were collapsed using the 6 Ns present on each of the custom-made adapters and trimmed. Reads were then aligned using Bowtie 2 in its most sensitive mode against the *S. cerevisiae* reference genome adapted with the synthetics sequences (*39*). An iterative alignment procedure was used:

for each read the length of the mapped sequence increases gradually from 20 bp until mapping becomes unambiguous (mapping quality > 30). Read pairs were aligned independently and assigned to a restriction fragment (RF). Religation and other unwanted events were filtered out, taking into account the orientation of the reads, as described (*28*, *40*). Contact matrices were built for each strain by binning the aligned reads into units of single RF. Adjacent restriction fragments were then pooled into fixed size bins of either 2,000 or 5,000 base pairs. Bins exhibiting important contact frequencies variations (< or > 2 Standard deviation, SD were filtered out of the data, with the corresponding bin vectors either set to zero (white band on the contact maps) or removed from the picture (see Figure legend). Finally, binned contact maps were normalized following the "Sequential Component Normalization" (SCN) procedure described in (*40*).

3D Representation of contact maps

The 3D representations of the contact maps were generated using ShRec3D (*27*) on the normalized contact maps filtered for low signal bins. These 3D structures are average representations and therefore do not represent the exact structure found in an individual cell. It must also be underlined that they are not polymer models and cannot be interpreted as such. They have to be interpreted in light contact frequencies over a population of cells. For instance, on these 3D representations the telomeres loosely cluster together. In a single nucleus, telomeres would rather form smaller groups scattered all around the nuclear membrane. Since in different cells these group gather different partners, they are regrouped together in the average structure that reflects the population average of contacts. All the 3D structures presented here were rendered using VMD (*41*).

Similarity between contact maps

To assess for the similarity between different datasets, we proceeded as follows. First, normalized matrices of chromosomes I, IV, VII, VIII, XIII, XIV, XV and XVI were binned at 50 kb and quantile normalized. We then computed the Euclidean distance between all pairs of maps and performed a principal component analysis on the resulting distance matrix.

Genomic distance plot

Pairs of reads mapping in cis- (i.e. intrachromosomal) positions along the genome were partitioned by chromosome. Reads oriented towards different directions or separated by less than 1,500 bp were discarded. For each chromosome read pairs were log-binned in function of their genomic distance s (in kb) according to the following formula:

$$bin = \lfloor \log_{1.1}(s) \rfloor.$$

The genomic distance plot is the weighted histogram computed from the sum of read pairs for each bin weighted with both the difference between the length of the chromosome and the genomic distance s (as a proxy for the number of possible events) and the bin-size $1.1^{(1+bin)}$ (because of the log-binning). To compare *syn* and native chromosome both distributions were normalized by their mean computed over the 70% of the curve at lower distances.

SCRaMbLE assay

Strain yLM539 (*synIII, synIXR*) was transformed with the plasmid pSCw11 CRE-EBD_HIS3, and grown in SC-HIS media (*14*). Cre expression was induced by adding 1µM final of estradiol, and samples of induced and non-induced cultures were analyzed at multiple time

point during 48h. For each time point (T_0, 1h, 2h, 4h, 5h, 6h, 24h, 48h), OD_{600} of a sample of each culture was recorded and then adjusted to OD_{600} = 0.05. Aliquots of 5µL of these cells and four 10-fold serial dilutions were spotted on YPD and SC-HIS agar plates. Growth curve and dilution dots experiments are shown in fig S15. Two clones, HMSY029 (T_2=2h) and HMSY030 (T_8=8h), were retained for sequencing and Hi-C analysis.

Genomic analysis of SCRaMbLE strains

Illumina paired-end short-reads were trimmed to remove adapter sequences and filtered to remove reads either shorter than 100 bp, and/or with unknown bases, and/or exhibiting 1 or more bases with a Phred-score under 7. The filtered reads were then mapped to the reference sequence using Bowtie2. The read coverage over the entire genome was computed, with a specific focus on the synthetic regions for which the copy number of each segment in-between *loxPsym* sites was assessed to detect deletions, duplications, and higher amplifications. To account for possible systematic experimental biases in the sequencing depth resulting from library preparation and mapping, we used an iterative algorithm developed previously to refine the copy number estimation and the potential bias (*16*).

We then focused on the unmapped reads carrying a *loxPsym* site to characterize further potential SCRaMbLE rearrangements. The two segments flanking a *loxPsym* site defines a junction. We trisected unmapped reads into a loxPsym site and its two flanking extremities, then mapped each of the latter to the reference genome using EMBOSS water (*42*) to identify novel junction. The average depth of novel junctions for the 2h and 8h Cre-induced strains was 16.3 and 14.3, respectively. We combined novel junctions with a depth higher than 5 and the copy number of segments to characterize the SCRaMbLE induced rearrangements (Table S3).

That no off-target rearrangement occurred elsewhere in the genome was confirmed by analyzing unmapped reads without *loxPsym* site with EMBOSS water to detect potential recombination events.

Analysis of rDNA contacts

Genome-wide rDNA contact frequencies were investigated in Sc2.0 strains carrying the rDNA either at its natural position on chromosome XII (datasets from strains YS031, yXZX538, yLM539, yLM896, HMSY012, yXZX573) on chromosome *synXII* (datasets from strains JDY465, JDY512, JDY452), on chromosome III (datasets from strains JDY448, JDY449), or on the 2μ plasmid (dataset from strain JDY446). For each dataset, pair-end reads were aligned on the reference genome of the corresponding strain lacking rDNA sequences and on an extra contig corresponding to a single rDNA unit (~9kb). Contact maps of trans interactions were generated (bin = 5 kb) and normalized(*40*). To compare the contacts made by the rDNA contig with the rest of the genome between different datasets, the values in the corresponding vector were divided by their median to alleviate differences of coverage and/or number of rDNA units. Chromosomes III and XII, that carries the rDNA cluster in three out of four datasets, were removed from the analysis to allow comparison of the rest of the genome. For each dataset, the contacts made by the rDNA with the 28 subtelomeric regions (30 kb at the extremities of the 14 chromosomes investigated) and a central region of 30 kb in each arm were plotted using violin plot function from the "vioplot" R library.

Distribution of read coverage for *syn* and native chromosomes

The removal of repeated genetic elements in synthetic genomes relative to native counterparts leads to smother contact maps in the designed chromosomes. Indeed, the presence of repeated sequences in the genome prevent the non-ambiguous alignment of Hi-C reads against them, resulting in the corresponding vectors in the contact matrix to display low if not null number of contacts. Those empty vectors that appear as empty rows and columns in the contact maps are a source of noise and spurious contacts in the matrices after normalization. For this reason, these regions are filtered according to a threshold (see the Generation and normalization of contact maps paragraph). To quantify the improvement in the visibility of *syn* chromosomes all along their length, we compared the histogram of their coverage between the synthetic and the native dataset (fig S12). The histograms count the number of contacts made by each vector within the region of interest with the entire genome (histogram bin = 500 reads). The filtering threshold (median – 2SD) was computed over the entire genome (reported on the fig S12 histograms as dotted lines).

References:

1. P. M. Sharp, A. T. Lloyd, Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. *Nucleic Acids Res.* **21**, 179–183 (1993).
2. K. R. Bradnam, C. Seoighe, P. M. Sharp, K. H. Wolfe, G+C content variation along and among Saccharomyces cerevisiae chromosomes. *Mol. Biol. Evol.* **16**, 666–675 (1999).
3. B. A. Cohen, R. D. Mitra, J. D. Hughes, G. M. Church, A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26**, 183–186 (2000).

4. E. J. Louis, A. V. Vershinin, Chromosome ends: different sequences may provide conserved functions. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **27**, 685–697 (2005).

5. S. M. Burgess, N. Kleckner, Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. *Genes Dev.* **13**, 1871–1883 (1999).

6. V. Guacci, E. Hogan, D. Koshland, Centromere position in budding yeast: evidence for anaphase A. *Mol. Biol. Cell*. **8**, 957–972 (1997).

7. P. Therizols, T. Duong, B. Dujon, C. Zimmer, E. Fabre, Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2025–2030 (2010).

8. A. Taddei, S. M. Gasser, Structure and function in the budding yeast nucleus. *Genetics*. **192**, 107–129 (2012).

9. C. Rabl, On Cell Division. *Morphol. Jahrb.* **10**, 214–330 (1985).

10. M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, Nucleolar clustering of dispersed tRNA genes, 1399–1401 (2003).

11. R. A. Haeusler, D. R. Engelke, Genome organization in three dimensions: thinking outside the line. *Cell Cycle Georget. Tex*. **3**, 273–275 (2004).

12. H. Schober *et al.*, Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res.* **18**, 261–271 (2008).

13. S. M. Richardson et al., Design of a synthetic yeast genome (This issue).

14. J. S. Dymond *et al.*, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*. **477**, 471–476 (2011).

15. J. Dymond, J. Boeke, The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioeng. Bugs*. **3**, 168–171 (2012).

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

17. Y. Shen et al., Deep functional analysis of synII, a 770 kb synthetic yeast chromosome (This issue).

18. N. Annaluru *et al.*, Total synthesis of a functional designer eukaryotic chromosome. *Science*. **344**, 55–58 (2014).

19. Z-X. Xie et al., “Perfect” designer chromosome V and behavior of a ring derivative (This issue).

20. L. A. Mitchell, et al., Synthesis, debugging and effects of synthetic chromosome consolidation: synVI and beyond (This issue).

21. Y. Wu et al., Bug mapping and fitness testing of chemically synthesized chromosome X (This issue).

22. W. Zhang et al., Engineering the ribosomal DNA in a megabase synthetic chromosome (This issue).

23. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Capturing chromosome conformation. *Science*. **295**, 1306–1311 (2002).

24. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).

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

26. H. Marie-Nelly *et al.*, Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinforma. Oxf. Engl.* **30**, 2105–2113 (2014).

27. A. Lesne, J. Riposo, P. Roger, A. Cournac, J. Mozziconacci, 3D genome reconstruction from chromosomal contacts. *Nat. Methods*. **11**, 1141–1143 (2014).

28. A. Cournac, M. Marbouty, J. Mozziconacci, R. Koszul, Generation and Analysis of Chromosomal Contact Maps of Yeast Species. *Methods Mol. Biol. Clifton NJ*. **1361**, 227–245 (2016).

29. A. Miele, K. Bystricky, J. Dekker, Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.* **5**, e1000478 (2009).

30. J.-M. Belton *et al.*, The Conformation of Yeast Chromosome III Is Mating Type Dependent and Controlled by the Recombination Enhancer. *Cell Rep.* **13**, 1855–1867 (2015).

31. K. Bystricky, T. Laroche, G. van Houwe, M. Blaszczyk, S. M. Gasser, Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J. Cell Biol.* **168**, 375–387 (2005).

32. M. Oakes *et al.*, Mutational analysis of the structure and localization of the nucleolus in the yeast Saccharomyces cerevisiae. *J. Cell Biol.* **143**, 23–34 (1998).

33. E. R. Hildebrandt, N. R. Cozzarelli, Comparison of recombination in vitro and in E. coli cells: Measure of the effective concentr ation of DNA in vivo. *Cell*. **81**, 331–340 (1995).

34. H. Marie-Nelly *et al.*, High-quality genome (re)assembly using chromosomal contact data. *Nat. Commun.* **5**, 5695 (2014).

35. N. Kaplan, J. Dekker, High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat. Biotechnol.* **31**, 1143–1147 (2013).

36. J.-F. Flot, H. Marie-Nelly, R. Koszul, Contact genomics: scaffolding and phasing (meta)genomes using chromosome 3D physical signatures. *FEBS Lett.* (2015), doi:10.1016/j.febslet.2015.04.034.

37. J. N. Burton *et al.*, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).

38. M. Guidi *et al.*, Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biol.* **16**, 1 (2015).

39. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

40. A. Cournac, H. Marie-Nelly, M. Marbouty, R. Koszul, J. Mozziconacci, Normalization of a chromosomal contact map. *BMC Genomics*. **13**, 436 (2012).

41. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).

42. P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG*. **16**, 276–277 (2000).

Acknowledgments:

We thank Axel Cournac, Martial Marbouty and Luciana Lazar-Stefanita for fruitful discussions and advices. This research was supported by funding to R.K. from the European Research Council under the 7th Framework Program (FP7/2007-2013, ERC grant agreement 260822), from Agence Nationale pour la Recherche (MeioRec ANR-13-BSV6-0012-02), and from ERASynBio and ANR (IESY ANR-14-SYNB-0001-03). H.M. and V.S are partly supported by Pasteur-Roux Fellowships. J.D. was funded by National Science Foundation of China (31471254), Tsinghua University Initiative Scientific Research Program (2011Z02296), Ph.D. Programs Foundation of Ministry of Education of China (20110002120055) and Chinese Ministry of Science and Technology (2012CB725201). Y.J.Y. was supported by Natural Science Foundation of China 21621004 and 21390203. Y.C. was funded by a Chancellor's Fellowship

from the University of Edinburgh, a startup fund from SULSA and BBSRC grants [BB/M005690/1; BB/M025640/1 and BB/M00029X/1]. This research was supported to Y.S. by the research grant from the Shenzhen Engineering Laboratory for Clinical Molecular Diagnostic Promotion (JZF No. [2016]884). This work was supported in part by funding from the United States National Science Foundation [grant MCB-0718846 and MCB-1026068 to J.D.B. and J.S.B., MCB-0546446 and MCB-1445545 to J.S.B.]. J.D.B. and J.S.B. are founders and directors of Neochromosome, Inc. J.D.B. serves as a scientific advisor to Recombinetics, Inc. and Sample6, Inc. These arrangements are reviewed and managed by the committees on Conflict of interest at NYU Langone Medical Center (J.D.B.) and Johns Hopkins University (J.S.B.). FASTQ files of the reads were deposited in the SRA database under the accession number SRP070421.

Fig. 1. **3D genome organization of wild-type and synthetic chromosomes**. (**A**) 3D representation of the wild-type yeast chromosomes inferred from the Hi-C contact map displayed in Fig S1 (*32*). Each bead represents a 5kb chromosome segment. Centromeres, telomeres and rDNA flanking regions are indicated with white, black and red beads, respectively. Each chromosomal arm has been colored according to its length. (**B, C, D**) Contact maps and corresponding 3D representations of synthetic chromosomes in three different strains: (**B**) JDY512 (*synII*, *synXII*), (**C**) yXZX573 (*synV*, *synX*), and (**D**) yLM896 (*synIII*, *synVI*, *synIXR*). Top: normalized contact maps of the seven synthetic chromosomes (bin = 5kb). Normalized contact frequencies (*32*) are indicated in a $\log_{10}$ scale ranging from white (few contacts) to dark red (many contacts). Filtered bins are set to zero (white vectors). Bottom: corresponding 3D representation of the contact maps displayed above, with synthetic chromosome and the set of natural chromosomes represented with colored and gray beads, respectively. (**E**) Distribution of the number of contacts as a function of the genomic distance for *synV* (pink) and native chromosomes V (black). (**F**) *Vis à vis* comparison of *synII* (strain YS031) and native chromosome II (strain BY4742) normalized contact maps. Non-mappable, repeated regions are highlighted with red arrowheads on the wild-type map.

Fig. 2. Loss of subtelomeric contacts in *synIII*. (**A**) Normalized contact maps (bin = 5kb) of native (left; strain BY4742) and *synIII* (right; strain yLM896) chromosome III. Colorscale as in Fig. 1. Annotation of mating type sequences are indicated along the top X-axis: *MAT*: mating type locus, *HML*, *HMR*: silent mating cassettes Left and Right. The positions of deleted loci are indicated with grey doted empty triangles. (**B**) Quantitative analysis of subtelomeric contacts in native chromosome III (dark dots) and *synIII* (yellow dots) using a bait chromosome capture

approach. Contacts of 10kb subtelomeric regions positioned at 20kb from left (top) or right (bottom) telomere (dark gray areas) are shown. Y-axis: normalized contact frequencies. X-axis: distance (in kb) from the left and right telomere. Each point represents the mean contact frequency for the bait region, computed for each chromosome from three independent experiments (on strains BY4742, YS031, and JDY512 for native III, and yLM896, yLM539 and JDY452 for *synIII*). Δ1 and Δ2 point at the normalized contacts discrepancies between both ends of *synIII* compared to native chromosome III.

Fig. 3. Repositioning of rDNA affects the overall genomic organization. (**A** - **C**) Top panel: normalized contact map of strains BY4742 (**A**), JDY465 (**B**) and JDY449 (**C**) carrying the rDNA cluster at different positions. Full gray triangles point at the position of the rDNA cluster in each strain. The position of the deleted rDNA cluster is indicated with a grey doted empty triangle. Bottom panel: 3D representations of the corresponding contact maps. The nucleolus and the rDNA cluster flanking regions are represented on each structure. Color code of chromosomes is the same as in Fig. 1B, C and D. The blue arrow points at the chromosome III right arm displaced following rDNA insertion. (**D**) Violin plot distribution of the contact frequencies between the rDNA cluster and either subtelomeric (light gray) or intra-arm (dark gray) chromosomal regions. ***: P-value < 0.001.

Fig. 4. SCRaMbLE induction of *syn* chromosomes results in multiple types of genome rearrangements. (**A**) Illustration of the time-course induction of the SCRaMbLE system in strain yLM539 carrying two synthetic chromosomes: *synIII* and *synIXR*. (**B**) Normalized contact

maps (bin =2kb) of the parental strain yLM539 (T_0; *synIII*, *synIXR*) and two SCRaMbLE clones isolated after 2h (T_2; HMSY029) and 8h (T_8; HMSY030) of Cre induction. All Hi-C reads are mapped against the reference genome of the parental strain yLM539. (**C**) Schematic representations of the duplications, deletions, inversions and translocations identified from these maps and coverage analysis. DNA segments between two *loxPsym* sites are numbered, from left to right (in blue for *synIII* and pink for *synIX*). The schematic representation shows the rearranged *synIII* and *synIX* by using numbering of the segments from the parental strain. See details in fig S18.

Supplementary Materials:

Figures S1-S19

Tables S1-S3

Movies S1-S10

Fig. 1

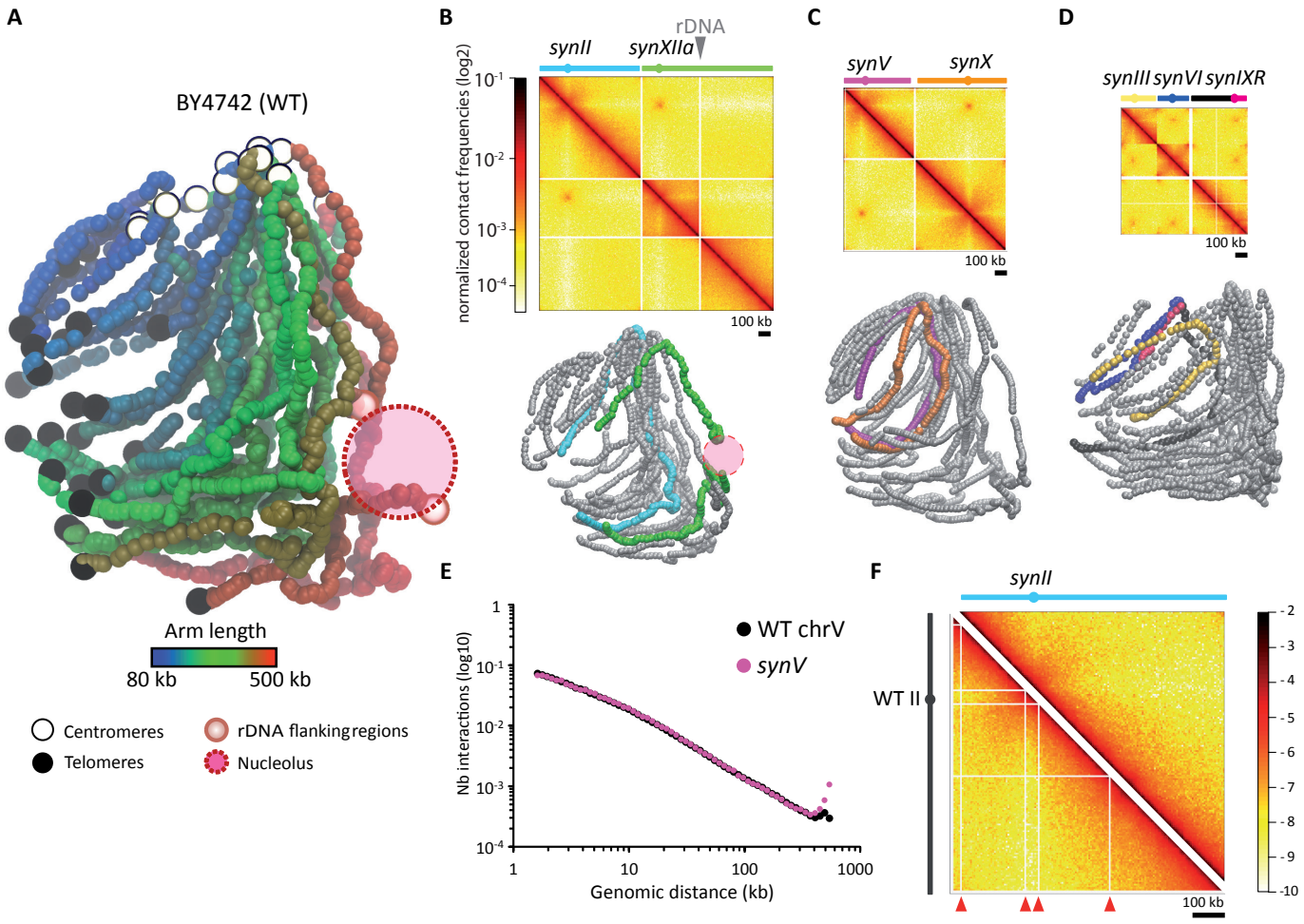

A
BY4742 (WT)
Arm length
80 kb
500 kb
Centromeres
Telomeres
rDNA flanking regions
Nucleolus
B
rDNA
synII
synXIIa
normalized contact frequencies (log2)
100 kb
C
synV
synX
100 kb
D
synIII synVI synIXR
100 kb
E
WT chrV
synV
Nb interactions (log10)
Genomic distance (kb)
F
synII
WT II
100 kb

Fig. 2

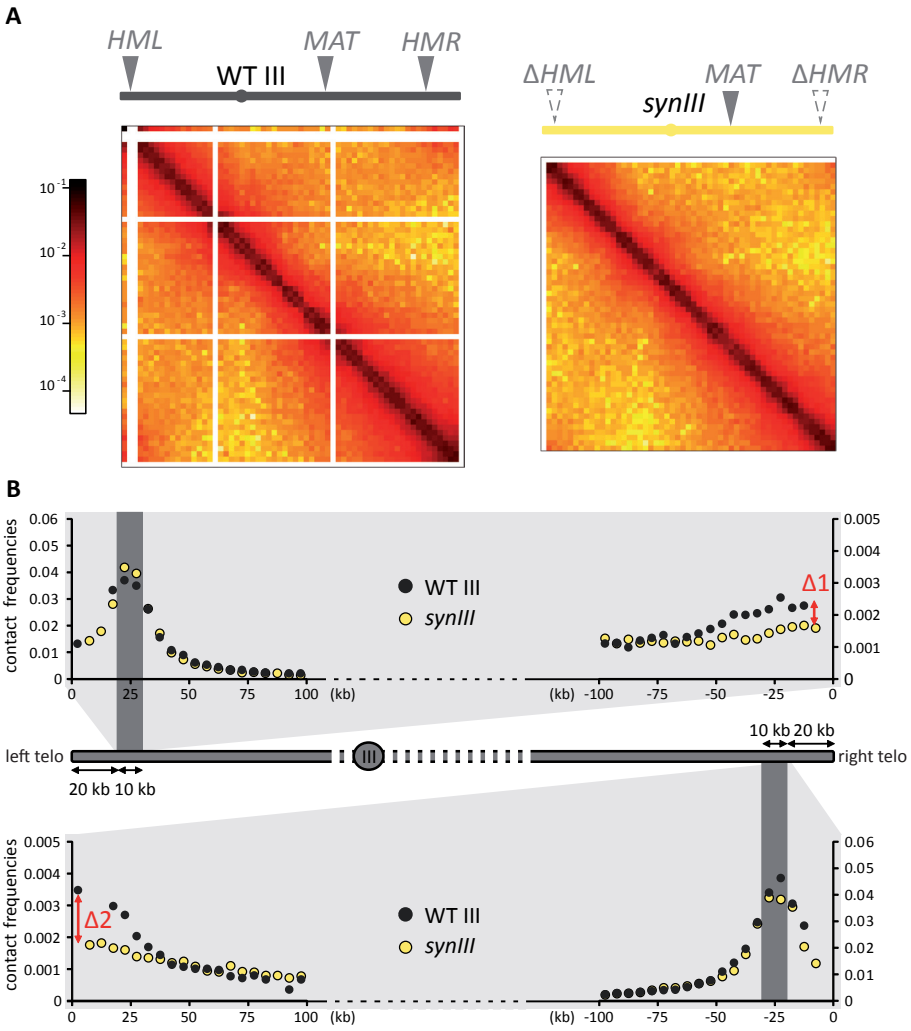
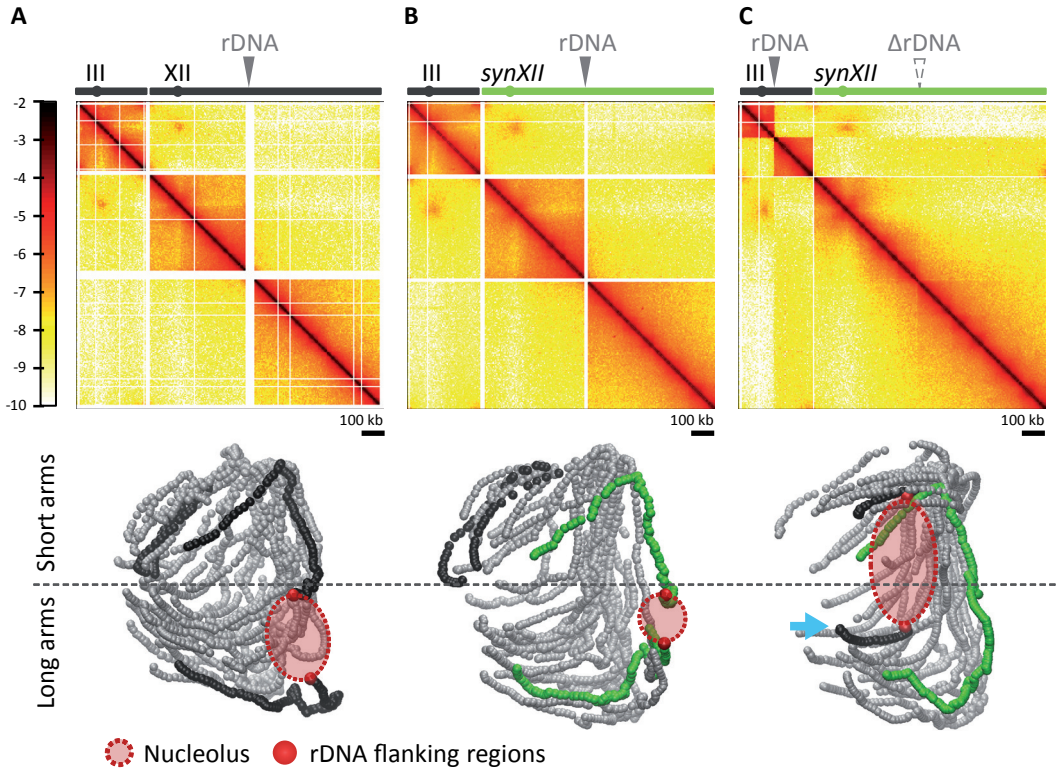

A
HML
MAT
HMR
WT III
ΔHML
MAT
ΔHMR
synIII
B
contact frequencies
WT III
synIII
Δ1
(kb)
left telo
right telo
III
20 kb 10 kb
10 kb 20 kb
Δ2

Fig. 3


A
B
C
rDNA
III
XII
synXII
ΔrDNA
-2
-3
-4
-5
-6
-7
-8
-9
-10
100 kb
Short arms
Long arms
Nucleolus
rDNA flanking regions

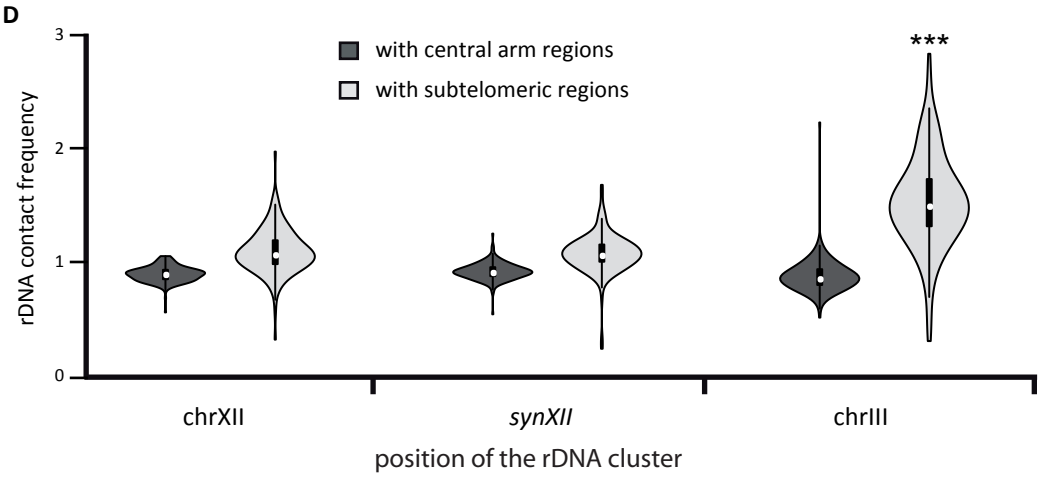
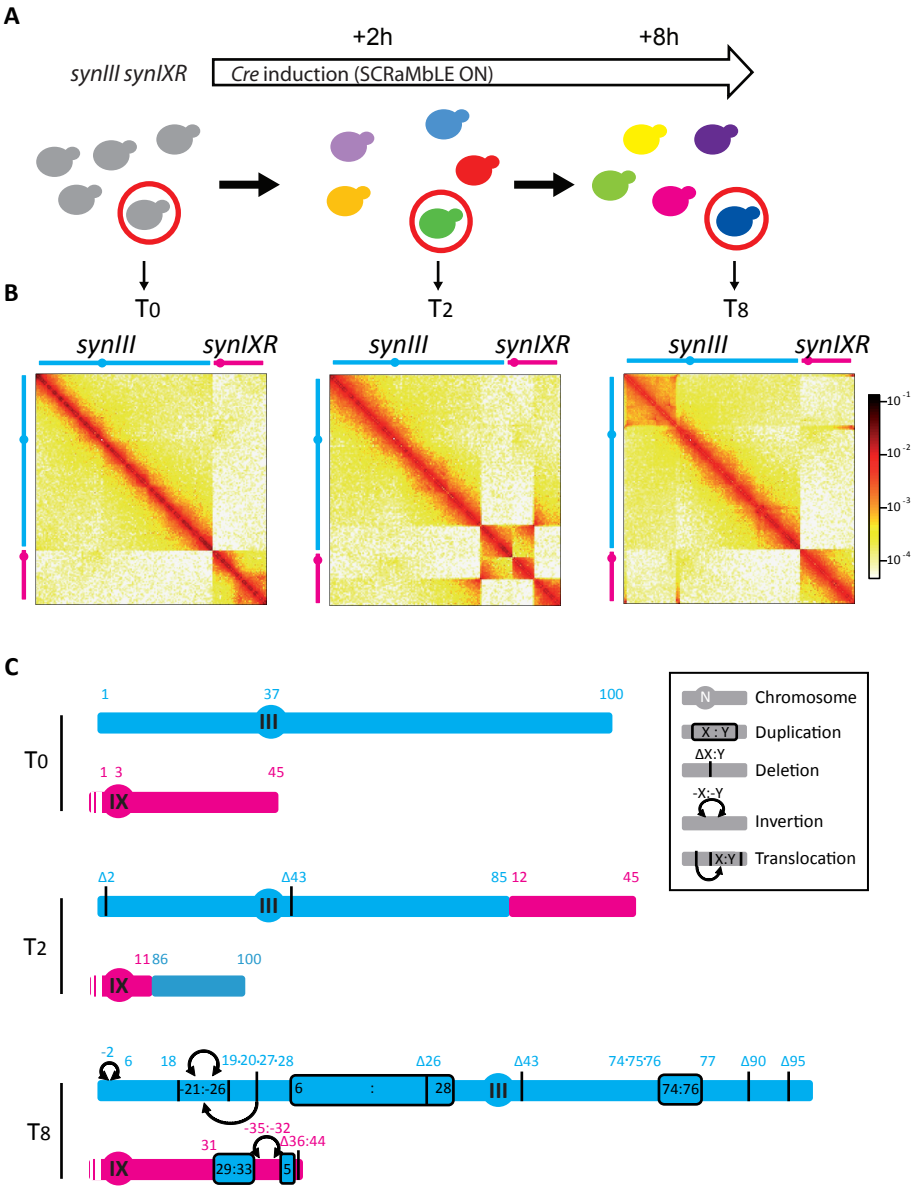

D
with central arm regions
with subtelomeric regions
***
rDNA contact frequency
3
2
1
0
chrXII
synXII
chrIII
position of the rDNA cluster

Fig. 4


A
+2h
+8h
synIII synIXR
Cre induction (SCRaMbLE ON)
T0
T2
T8
B
synIII
synIXR
C
Chromosome
Duplication
Deletion
Invertion
Translocation
T0
T2
T8

Supplementary Materials

3D organization of synthetic and scrambled chromosomes

Guillaume Mercy et al.

Fig. S1. Contact map of the wild-type strain BY4742

Normalized contact map of all chromosomes of WT strain BY4742 (bin = 5 kb). Normalized contact frequencies ((*1*)) are indicated in a log scale from white (few contacts) to dark red (many contacts). Filtered bins were discarded from the map.

Figs. S2-10: Contact maps and 3D conformations of synthetic chromosomes

Normalized contact maps (with filtered bins set to zero) and corresponding 3D representation of the genome using ShRec3D (*27*) of each strain described in this study. Top panel: normalized contact maps of the synthetic chromosome(s) carried by each synthetic strain (right) and of the native chromosome(s) in the WT strain (left). Bottom panel: two views of the 3D representation of the contact maps of the corresponding genomes from the top panel, with chromosomes of interests highlighted. Each synthetic chromosome and its native counterpart in the WT chromosomal set is represented by a specific color (*synII*: cyan, *synIII*: yellow, *synV*: purple, *synVI*: blue, *synIXR*: pink, *synX*: orange, *synXII*: green).

S2: Contact map and 3D conformation of strain YS031

S3: Contact map and 3D conformation of strain yXZX538

S4: Contact map and 3D conformation of strain yXZX573

S5: Contact map and 3D conformation of strains JDY465 (middle) and JDY446 (right)

S6: Contact map and 3D conformation of strain JDY448 (middle) and JDY449 (right)

S7: Contact map and 3D conformation of strain JDY512

S8: Contact map and 3D conformation of strain JDY452

S9: Contact map and 3D conformation of strain yLM539

S10: Contact map and 3D conformation of strain yLM896

Fig. S11: Correlation between contact maps

(**A**) Euclidean distances between all pairs of contact maps for chromosomes I, IV, VII, VIII, XIII, XIV, XV and XVI. (**B**) Principal component analysis of the distance matrix (**A**).

Fig. S12: Quantification of the increased visibility of *syn* chromosomes

The distribution of bin coverage was plotted for the contact matrix of each *syn* chromosome (top histogram) and their native counterpart (bottom histogram). In each of these comparisons, the colored histogram (top histogram) represents the sum of the elements for each column in the contact matrix in the *syn* chromosomes while the dark histogram (bottom histogram) represents the sum of the elements for each column in the native contact matrix, for the same region. The dashed lines represent the filtering threshold used for normalization purposes (M&M). The number on top of the first bin of each histogram corresponds to the number of filtered vectors in the region. The lower the number, the smoother (and complete) the matrix. The histograms illustrate how synthetic matrices are now much less affected by the filtering step compared to native chromosomes. For each chromosome, the number and percentage of filtered bins are the following: II: 8 (5%) *synII* 0 (0%); III: 5 (8%), *synIII*: 1 (2%); IXR: 0 (0%), *synIXR*:1 (5%); V: 7 (6%), *synV*: 2 (2%); VI: 5 (9%), *synVI*: 0 (0%); X: 12 (8%), *synX*: 5 (4%); XII: 20 (9%), *synXII*: 6 (3%).

Fig. S13: Telomere interactions of *synVI*

(**A**) Normalized contact maps (bin = 5 kb) of native (left; strain BY4742) and *synVI* (right; strain yLM896) chromosome VI. Same colorscale as in Fig. 1. (**B**) Quantitative analysis of intrachromosomal contacts made by the subtelomeric regions of the native chromosome VI (dark dots) and *synVI* (blue dots) using a bait chromosome capture approach. The contacts made by a 10kb window positioned either 20kb away from the left (top) or the right (bottom) telomere (dark gray areas) are displayed. Y-axis: contact frequencies. X-axis: distance (in kb) from the left and right telomere. Each point represents the mean contact frequency for the bait region with 5kb windows (bins) over the native and *synVI* chromosome, computed from three independent experiments. The Δ symbol points at the normalized contacts discrepancies between the left arm extremities of *synVI* and native chromosomes. Whereas the native chromosome end is "invisible" to the 3C assay, the telomere proximal region of the synthetic chromosome, which does not harbor repeated elements anymore, is now fully accessible and thus "visible".

Fig. S14: loss of mating-type specific conformation in *synIII*

Normalized contact maps (bin = 10 kb) of chr III (top matrices) and *synIII* (bottom matrices) from *MATa* (**A**) and *MATalpha* (**B**) cells growing in asynchronous conditions. Contact frequencies are displayed in $\log_2$. (**C**) *MATa*/*MATalpha* differential contact maps of WT strains (top) and *syn* strains (bottom). Colorscale reflects contact enrichment in *MATa* (the more red, the more contacts in *MATa*) or in *MATalpha* (the more blue, the more contacts in *MATalpha*).

Fig. S15: SCRaMbLE induction

(**A**) Growth curve of strain yLM539 + pSCw11 in SC-HIS media in presence or absence of estradiol (+/- E). (**B**) Dilution dot experiment on YPD and SC-HIS agar plates during SCRaMbLE induction of strain yLM539 + pSCw11.

Fig. S16: Hi-C contacts patterns resulting from various chromosomal rearrangement

Hi-C contacts maps allow the identification of chromosomal rearrangements compared to the reference genome. Because of DNA polymer nature, two fragments of DNA at close distance on the chromosome will collide more frequently than two fragments separated by a larger distance. In addition, two fragments positioned onto two different chromosomes will display, in general, even lower contacts. (**A**) If no rearrangement is observed, chromosomes appear as discrete domains composed of adjacent DNA segments frequently colliding with each other's within one chromosome. Two chromosomes (DNA molecules) will display less contacts between them than intrachromosomal contacts. (**B**) If a fragment is duplicated and the reference genome not modified to take into account this mutation, the signal will be doubled over the duplicated region, and therefore a cross will appear in the contact map at that position. (**C**) If a fragment is deleted, no reads corresponding to this fragment will be present in the data. A region without signal will therefore appear on the contact map. (**D**) If a fragment is inverted, the continuity of the signal will be altered intrachromosomally, with strong signal appearing at the extremities of the inverted region, away from the diagonal of the contact map. (**E**) If a fragment is translocated from one chromosome to another, a characteristic high interchromosomal contact will appear at the region involved in the translocation.

Fig. S17: Genomic analysis of SCRaMbLE strains

Number of reads remaining at each step of genomic analysis as described in M&M, for the parental strain yLM539 (T_0) and the two SCRaMbLE strains HMSY029 (T_2=2h) and HMSY030 (T_8=8h).

Fig. S18: Genomic structure of SCRaMbLE clones

Detailed genetic maps of *synIII* and *synIXR* of strains yLM539 (**A**), HMSY029 (**B**), and HMSY030 (**C**). (**A**) For each synthetic chromosome carried by the parental strain yLM539, segments between two *loxPsym* sites (and telomeres and a *loxPsym* site) are numbered from left to right (1 to 100 for *synIII*, blue numbers; 1 to 45 for *synIXR*, pink numbers). Genetic elements along the chromosomes are indicated with symbols and colors according to figure legend. Names of genes at rearrangements borders are also indicated. Chromosome segments along the SCRaMbLEd *synIII* and *synIXR* chromosomes from stains HMSY029 (**B**) and HMSY030 (**C**) are numbered and colored according to their original location along the parental strain depicted in (**A**). The

following chromosomal rearrangements are indicated: deletion (crossed number), duplication (bold number), inversion (minus sign in front of number), and translocation (boxed number). (**B**) Strain HMSY029 presents deletions of *synIII* segments 2 and 43, which carry the uncharacterized open reading frames YCL073c and YCR007C respectively. It also bears a translocation between the two right arms of the synthetic chromosomes (breakpoint between segments 85 and 86 for *synIII* (YCR075c to right telomere) and segments 11 and 12 for *synIXR* (YIR013c to right telomere). (**C**) The genome of strain HMSY030 exhibits deletions of *synIII* segments 43, 90 and 95 (both containing no genes) and *synIXR* segments 36-44 (containing the uncharacterized open reading frames YIP038c to YIR044c). It also presents three inversions, *synIII* segment 2 and 79-81, and *synIXR* segments 32-35. *SynIII* segments 74-76 are tandemly duplicated, whereas *synIII* segments 5 and 6-28 (with an additional deletion of 26) are duplicated and translocated on *synIXR*. Finally, *synIXR* segments 21-26 are inverted and translocated internally to the same chromosome.

Fig. S19: Accompanying rearrangements generated during *syn* assembly, and identify through Hi-C

(**A**) Genome wide contact map of S*c2.0 synII* strain (YS031). A duplication and a translocation occurred between chromosomes XIII and XVI of this strain, highlighted with the black arrows. (**B**) Schematic representation of the rearrangements between chromosome XIII (yellow) and XVI (orange). Reciprocal translocation happened between the left arm of the chromosome XIII (coordinates 1-212,400 bp) and the right arm of chromosome XVI (850,341-924,431 bp). Additionally, a segment of chromosome XIII (184,000-212,400 bp; in red) has been duplicated on chromosome XVI. This kind of double-rearrangement is uncommon, but was reported before to occur spontaneously in the yeast genome (*2*). (**C**) Contact map of chromosomes XIII and XVI aligned with corrected reference sequences, including the reciprocal translocation. As duplicated sequences cannot be unambiguously mapped, the reference sequence contains the duplicated segment of chromosome XIII only in chromosome XVI for a clearer visualization. Read coverage shown in the lower panel allows to identify the duplication.

Table S1. Strain description

strain name	description	genotype	karyotype abnormalities	Original description
BY4742	WT	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0		(*3*)
BY4741	WT	*MATa* ura3Δ0 leu2Δ0 his3Δ1 met15Δ0		(*3*)
YS031	*synII*	*MATa* ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 synLYS2, *synII*	XIIItXVI + segmental	(*4*);

			duplications	
yLM422	*synIII MATalpha*	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 synSUP61::HO::ura3, *synIII*		This work
HMSY012	*synIII MATa*	*MATa* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 *syn*SUP61::HO::ura3, *synIII*		This work
yYW0115	*synX*	*MATa* ura3Δ0 leu2Δ0 his3Δ1 met15Δ0, *synX*		(*5*)
yXZX573	*synV synX*	*MATa* ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 LYS2, *synV, synX*		This work
JDY465	*synXII*	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 Oj::HIS3, *synXII*		(*6*)
yLM539	*synIII synIXR*	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 hoΔ::synSUP61_KlURA3, *synIII, synIXR*		This work
yLM896	*synIII synVI synIXR*	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 hoΔ::synSUP61::ura3, *synIII, synVI* SYN-WT.PRE4, IXL-*synIXR*		(*7*)
JDY512	*synII synXII*	*MATa*/*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 *syn*LYS2 *syn*MET15 *syn*OjΔ::HIS3 ZLP101-pRS316-(tL(UAG)L1+tL(UAU)L+tL(UAG)L2)::natNT2, *synII, synXII*	duplication of chr III small rearrangement in *synXIIa*	This work
JDY452	*synIII synXII*	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 *syn*MET15 *syn*OjΔ::HIS3 hoΔ::*syn*SUP61, *synIII, synXII*		This work
JDY446	*synXII* pRDN	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 rdn::NatMX4 + pRDN-wt-U, *synXII_* RDNΔΔ		(*6*)
JDY448	*synXII* ChrIII-rDNA_17	*MATa* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 rdn::NatMX4 ChrIII-rDNA, *synXII*_RDNΔΔ	duplication of chr XI	(*6*)
JDY449	*synXII* ChrIII-rDNA_18	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 rdn::NatMX4 ChrIII-rDNA, *synXII*_RDNΔΔ	duplication of chr XI	(*6*)
HMSY029	*synIII synIXR* - SCRaMbLE_T2	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 hoΔ::*syn*SUP61_KlURA3, *synIII, synIXR*, SCRaMbLE		This work

HMSY030	*synIII synIXR* - SCRaMbLE_T8	*MATalpha* ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 hoΔ::*syn*SUP61_KlURA3, *synIII*, *synIXR*, SCRaMbLE		This work

Table S2. Number of pair-end reads of Hi-C libraries

Strain	Synchronization	Total Reads	Aligned Reads	Aligned reads after filtering
WT BY4742	elutriated	39 064 198	26 992 227	22 312 642
SynII	elutriated	29 338 003	20 242 978	18 503 055
SynIII MATa	elutriated	46 516 034	32 428 055	27 124 106
SynV	elutriated	45 137 272	30 264 624	25 803 212
SynV SynX	elutriated	39 075 196	25 690 540	20 642 354
SynXII	elutriated	24 609 744	16 649 288	10 444 262
***SynXII* pRDN**	elutriated	25 972 574	18 994 222	14 898 107
***SynXII* ChrIII-rDNA_17**	elutriated	32 174 741	24 406 105	14 688 663
***SynXII* ChrIII-rDNA_18**	elutriated	60 793 665	47 785 032	33 547 074
SynII SynXII	elutriated	46 479 888	33 765 768	28 957 424
SynIII SynXII	elutriated	40 325 041	28 353 744	24 331 186
***SynIII SynIX*R**	elutriated	34 986 055	23 192 750	21 223 097
***SynIII SynVI SynIX*R**	elutriated	45 249 989	32 268 208	25 547 568
WT BY4741	asynchronous	50 039 070	34 269 612	3 325 961
WT BY4742	asynchronous	47 799 156	34 494 832	3 229 555
synIII MATa	asynchronous	50 190 359	34 235 518	4 056 044

synIII MATalpha	asynchronous	45 716 748	31 088 161	5 337 666
***synIII synIXR* - SCRaMbLE_T2**	asynchronous	27 973 841	19 529 594	16 859 454
***synIII synIXR* - SCRaMbLE_T8**	asynchronous	29 802 054	21 234 553	18 270 085

Table S3. Depth of novel junctions.

Novel junctions with depth more than 5 were used to identify SCRaMbLE rearrangements (highlighted in blue)**.**

2h strain HMSY029		
Junction 1	**Junction 2**	**Depth**
3.-03	3.99	5
3.01	3.03	37
3.21	3.-05	1
3.-22	3.22	1
3.-44	3.-42	29
3.42	9.01	1
9.-12	3.-85	39
9.11	3.86	48
9.34	9.-01	1
3.01	9.-43	1
8h strain HMSY030		
Junction 1	**Junction 2**	**Depth**
3.-03	3.02	22
3.01	3.-02	10
3.01	3.03	2
3.01	3.05	14
9.-32	3.-05	19
3.28	3.06	26
3.26	3.-18	8
3.-21	3.19	18
3.-27	3.-20	31
3.-27	3.-25	29
9.-01	3.-27	1
9.31	3.29	16
9.35	3.-33	28
3.-44	3.-42	8
3.42	9.01	1

3.-73	3.-72	1
3.76	3.74	13
3.81	3.-78	25
3.-82	3.79	36
9.11	3.86	2
3.-91	3.-89	21
3.-92	3.92	1
3.-96	3.-94	23
3.02	3.99	2
3.-05	3.99	1

Movie S1A. 3D reconstruction of the WT strain BY4742 contact matrix (related to Fig. 3A)

Movie S1B. 3D reconstruction of the synthetic strain JDY465 contact matrix (related to Fig. 3B)

Movie S1C. 3D reconstruction of the synthetic strain JDY449 contact matrix (related to Fig. 3C)

Movies S2A-S10A. 3D reconstruction of the WT strain BY4742 contact matrix (related to fig S2-10)

Movies S2B-S10B. 3D reconstruction of synthetic strains contact matrix (related to fig S2-10)

References

1. A. Cournac, H. Marie-Nelly, M. Marbouty, R. Koszul, J. Mozziconacci, Normalization of a chromosomal contact map. *BMC Genomics*. **13**, 436 (2012).

2. R. Koszul, S. Caburet, B. Dujon, G. Fischer, Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234–243 (2004).

3. C. Baker Brachmann *et al.*, Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*. **14**, 115–132 (1998).

4. Y. Shen et al., Deep functional analysis of synII, a 770 kb synthetic yeast chromosome (This issue).

5. Y. Wu et al., Bug mapping and fitness testing of chemically synthesized chromosome X (This issue).

6. W. Zhang et al., Engineering the ribosomal DNA in a megabase synthetic chromosome (This issue).

7. L. A. Mitchell, et al., Synthesis, debugging and effects of synthetic chromosome consolidation: synVI and beyond (This issue).

fig S1

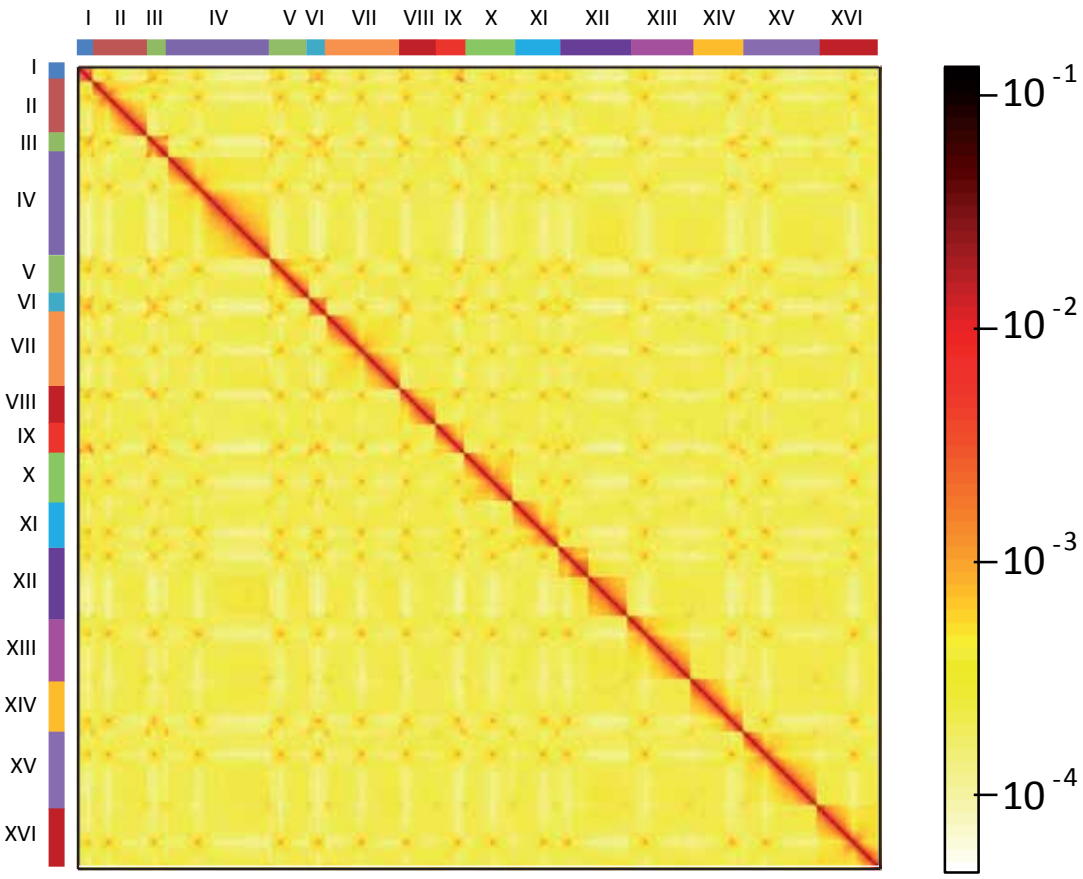
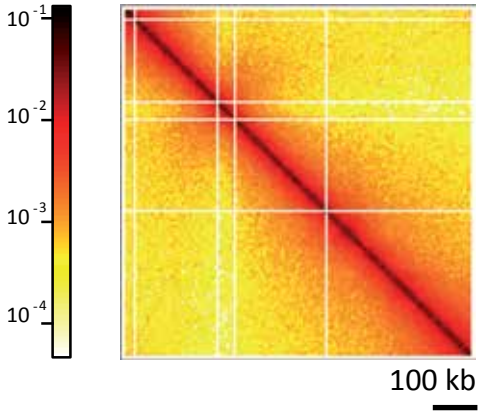

I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
10 -1
10 -2
10 -3
10 -4

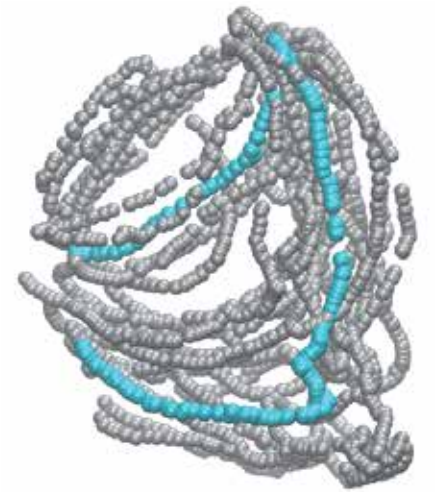
fig S2

synII

BY4742

II

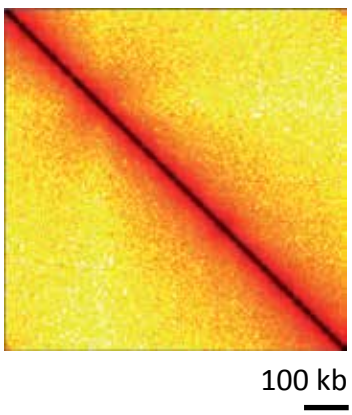


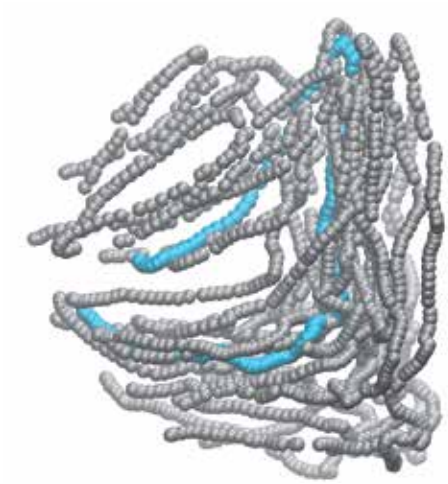




YS031

synII





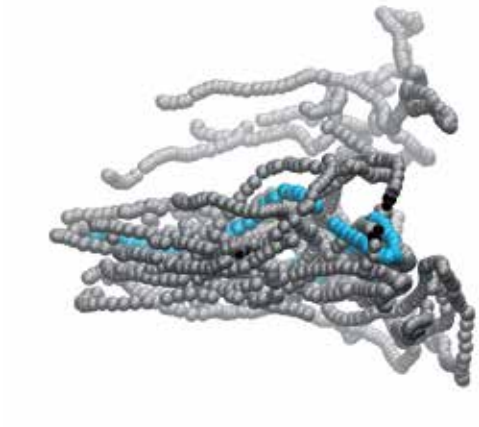
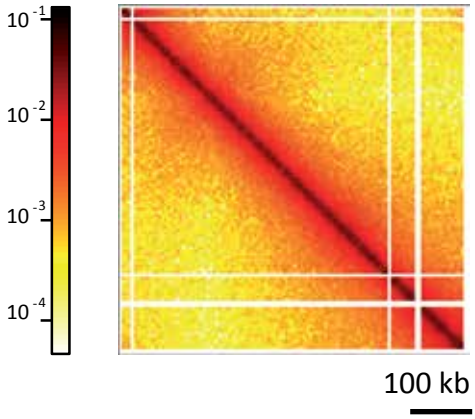


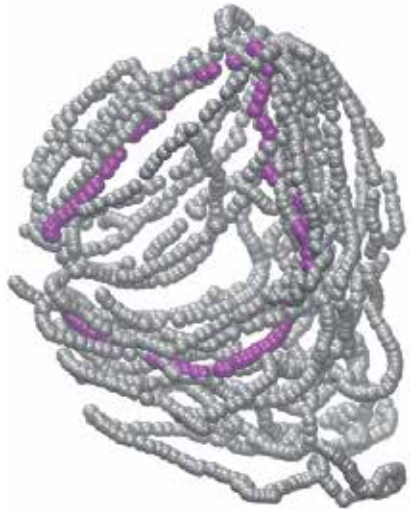
fig S3

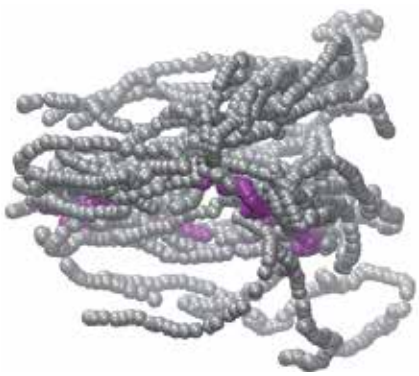
synV

BY4742

V







yXZX538

synV

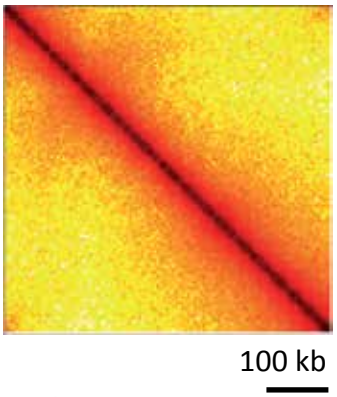


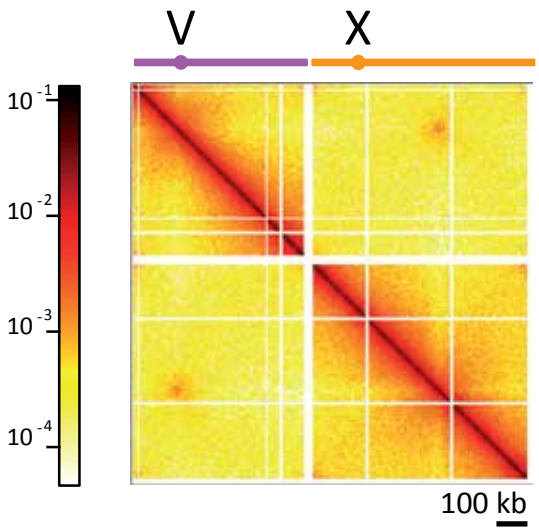


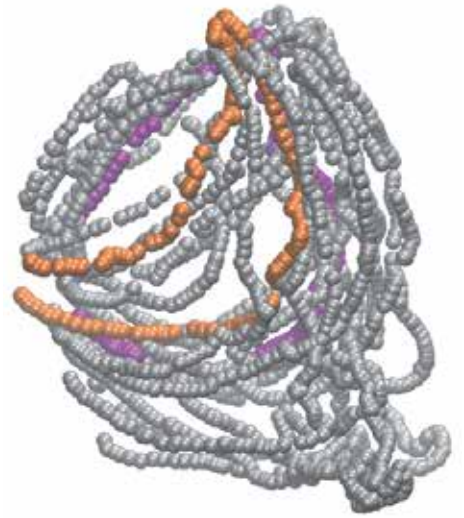


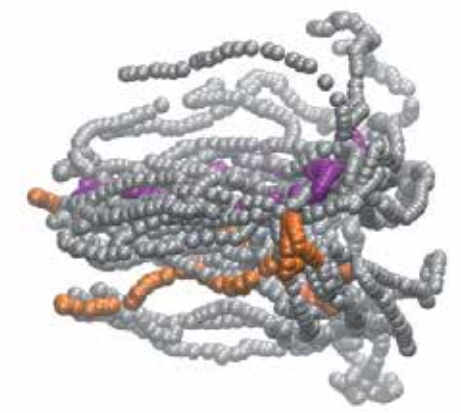
fig S4

synV synX

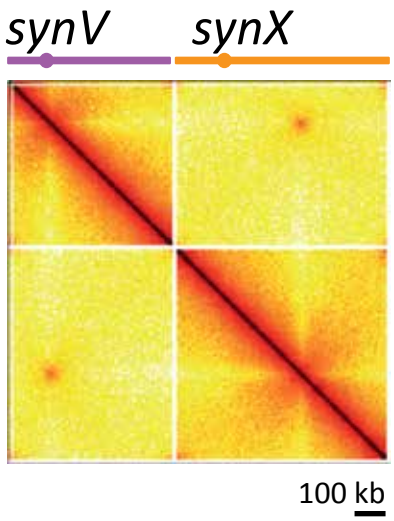
BY4742

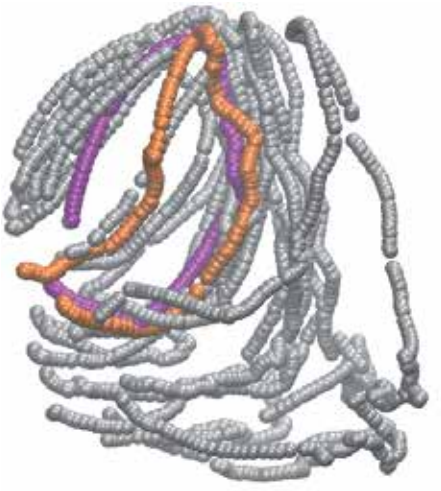






yXZX573





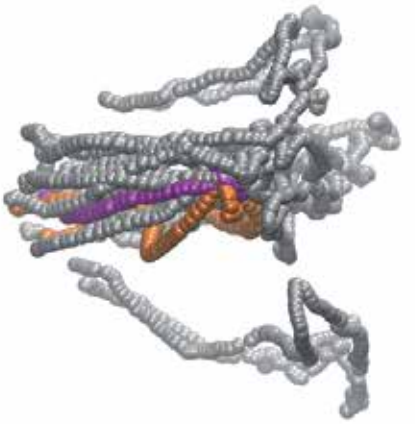
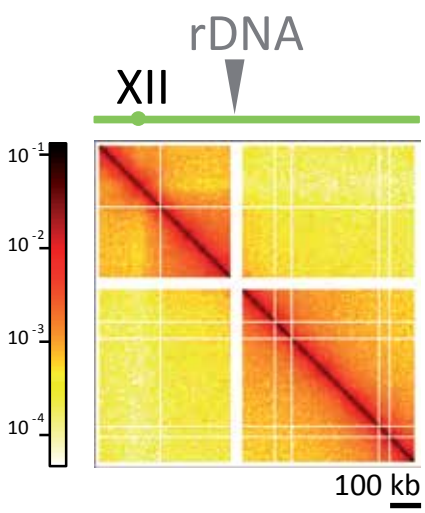
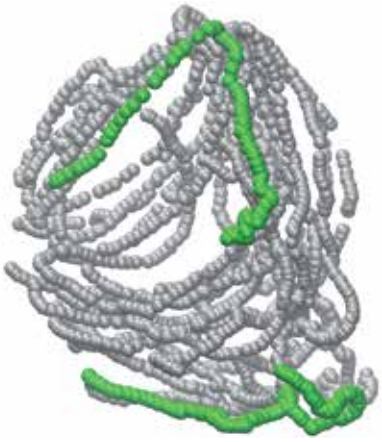


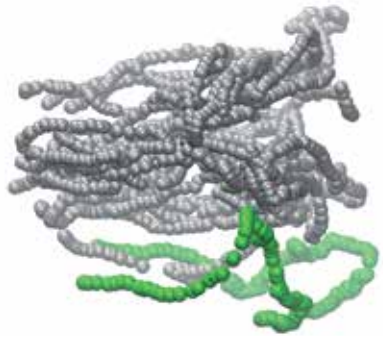
fig S5

synXII +/- rDNA

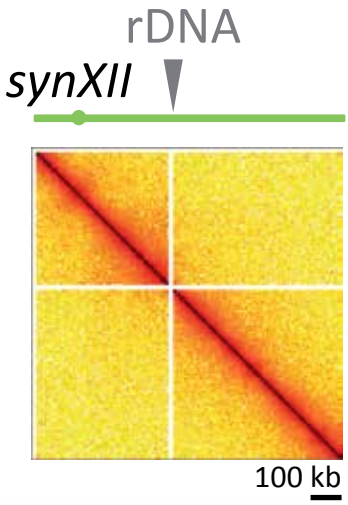
BY4742

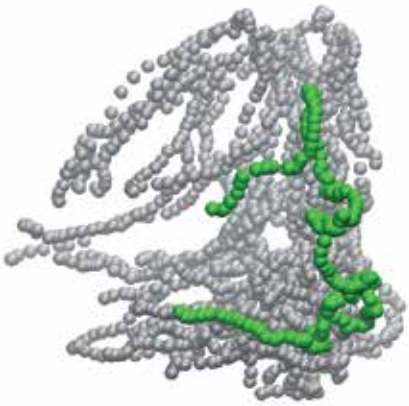


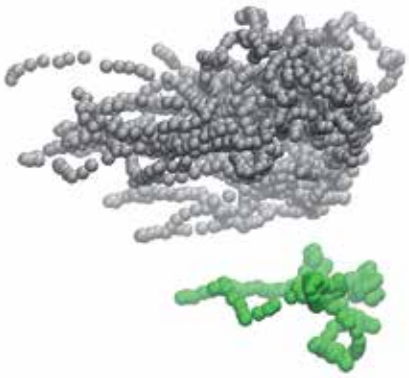




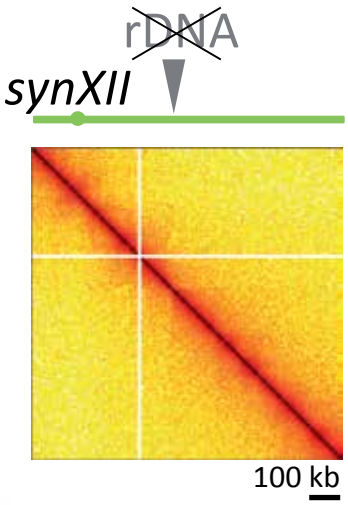
yDY444

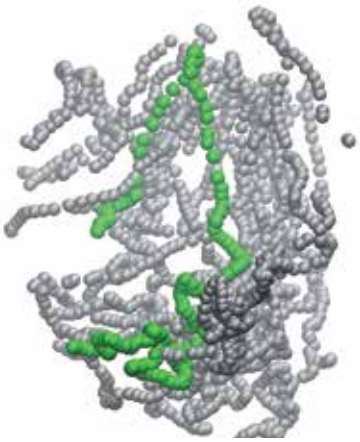






yDY446





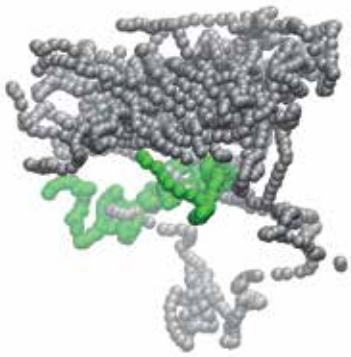
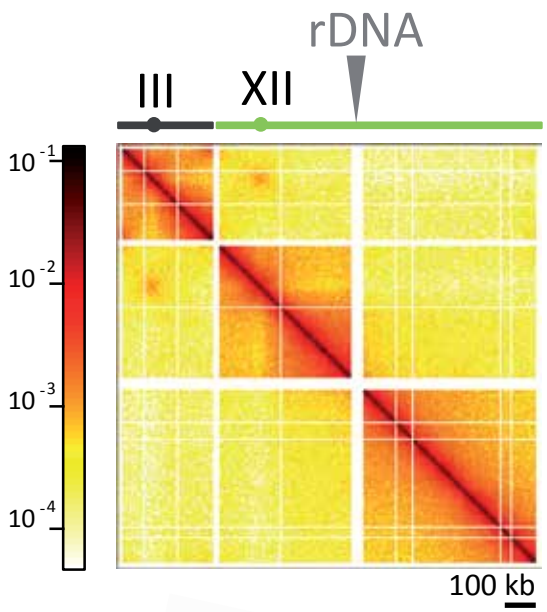
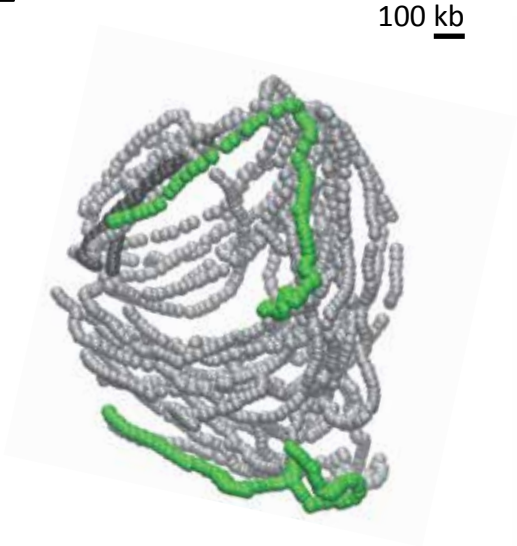


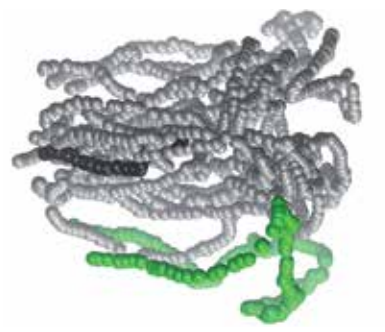
fig S6

synXII chrIII-rDNA

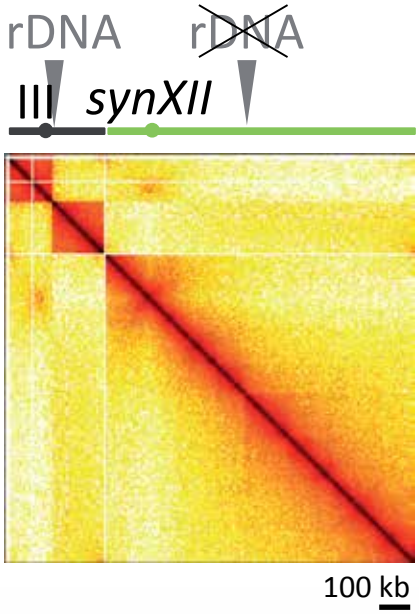
BY4742







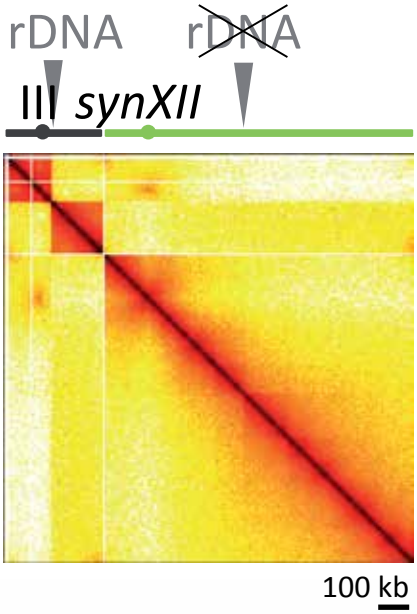
JDY448







JDY449



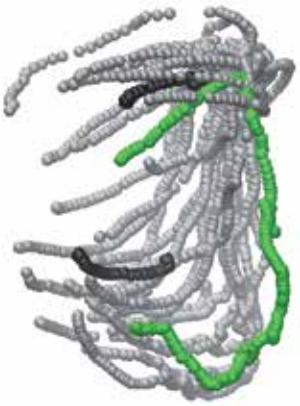
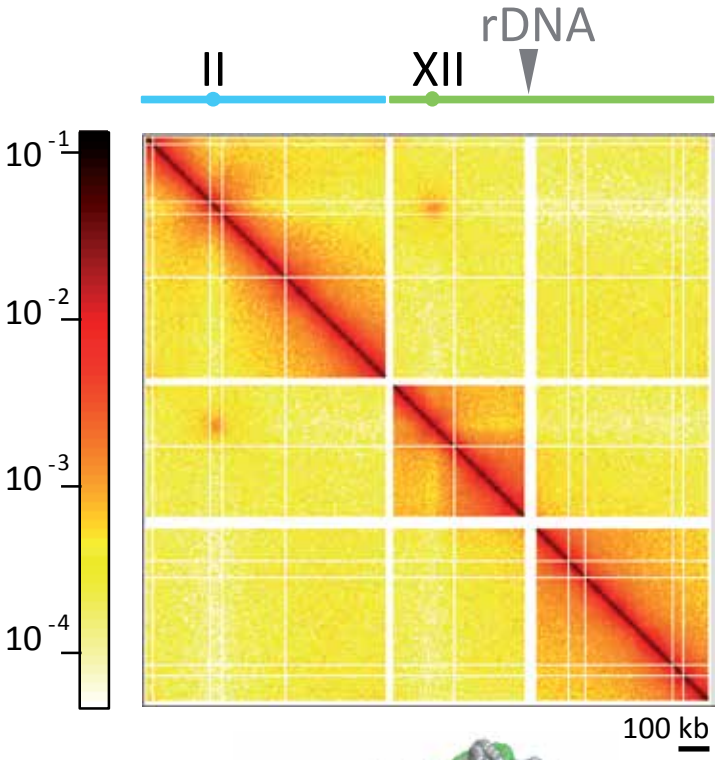


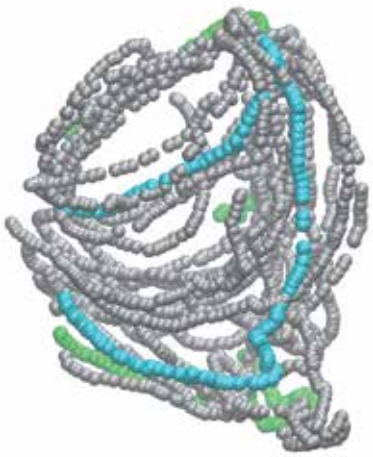


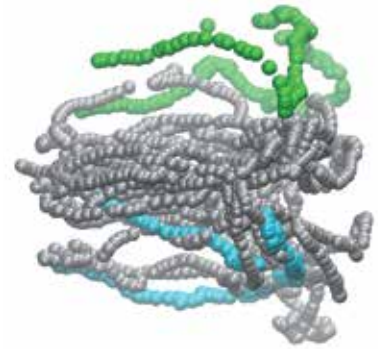
fig S7

synII synXII

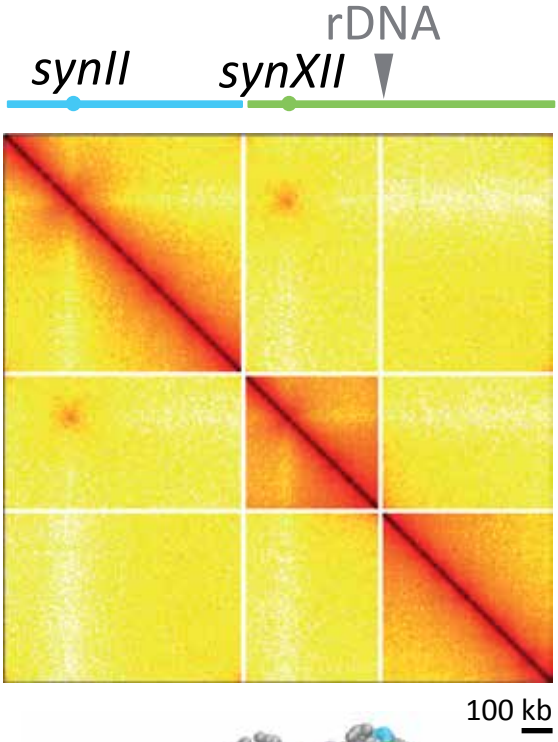
BY4742







JDY450



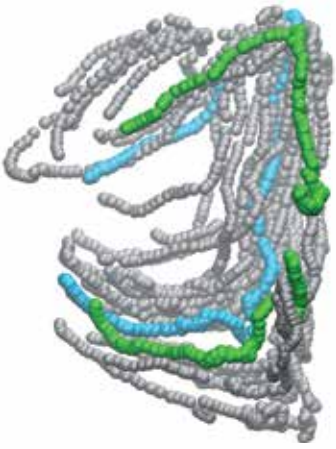
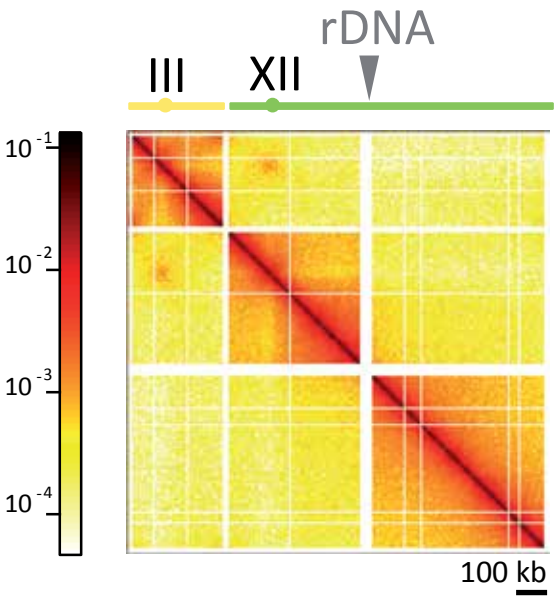


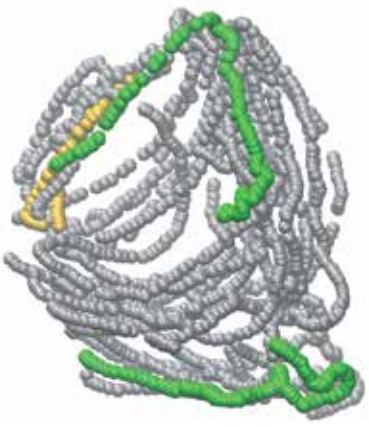


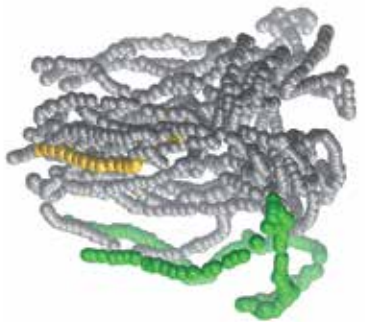
fig S8

synIII synXII

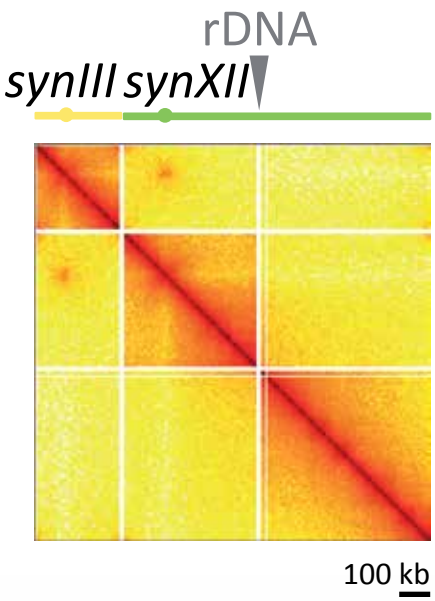
BY4742

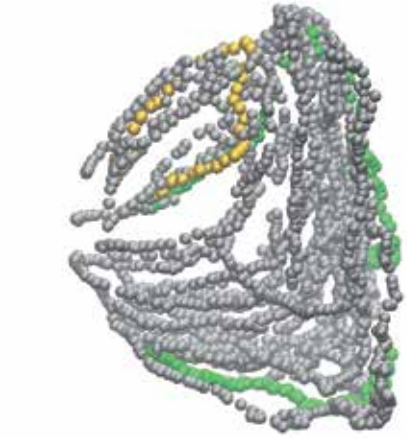






JDY451





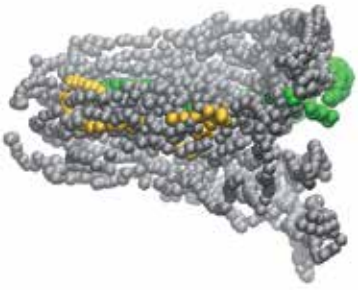


fig S9

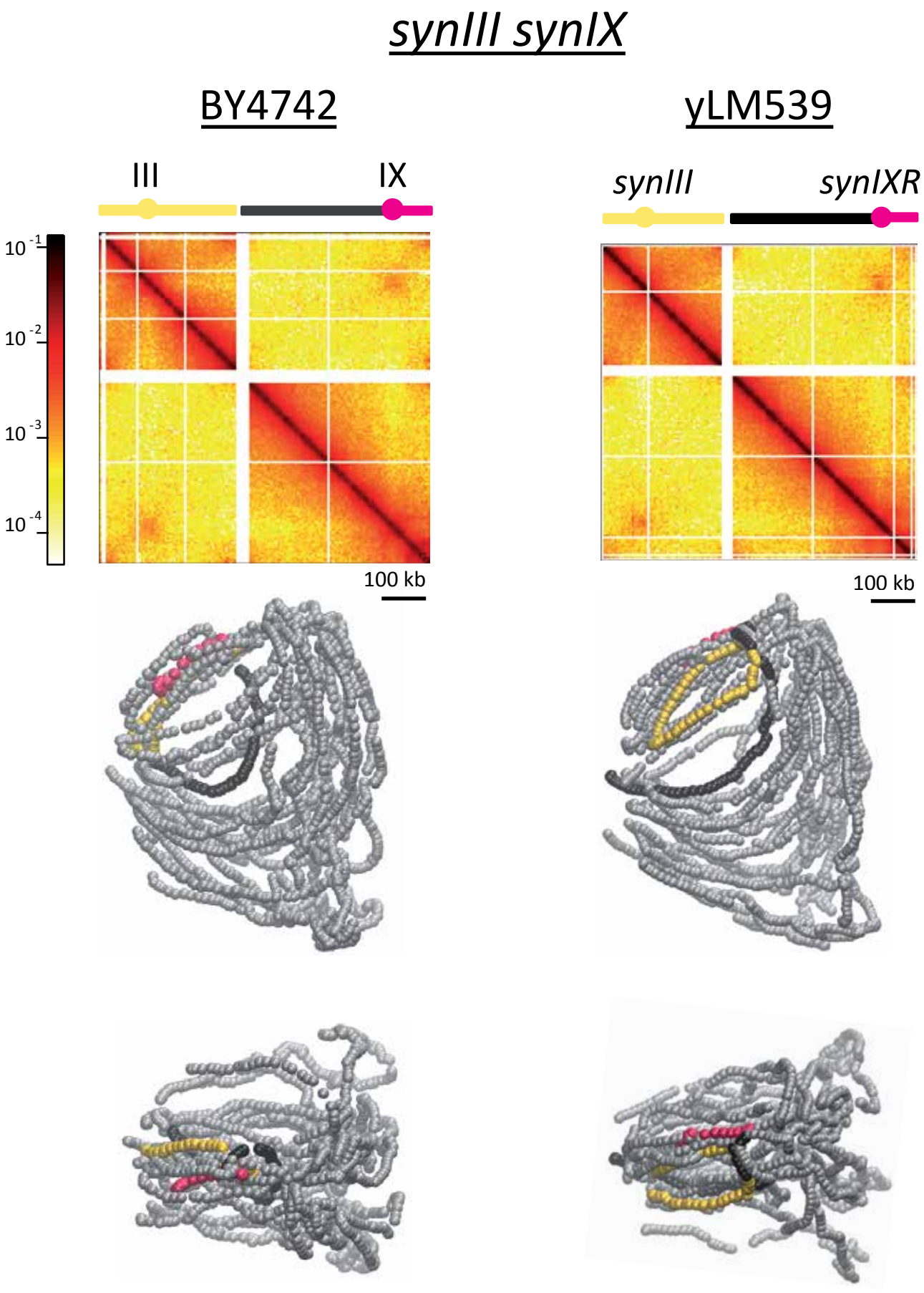
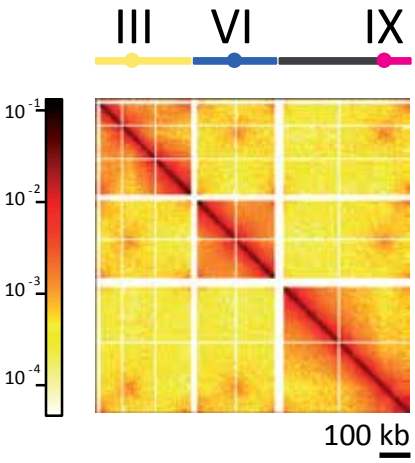
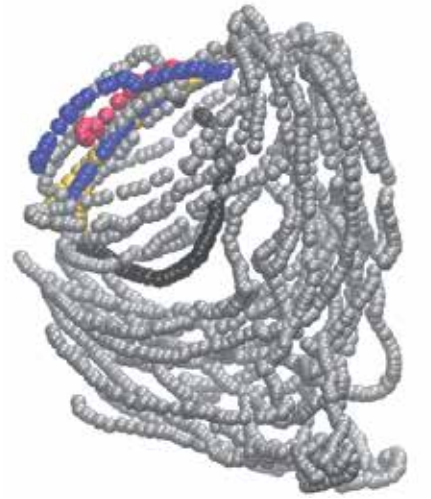

synIII synIX
BY4742
yLM539
III
IX
synIII
synIXR
10 -1
10 -2
10 -3
10 -4
100 kb
100 kb

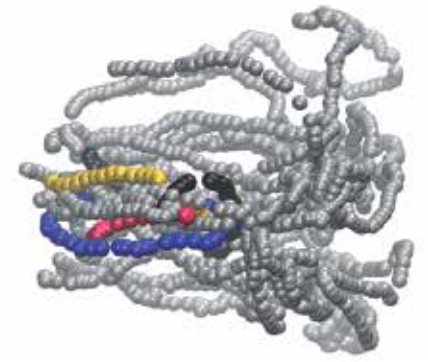
fig S10

synIII synVI synIX

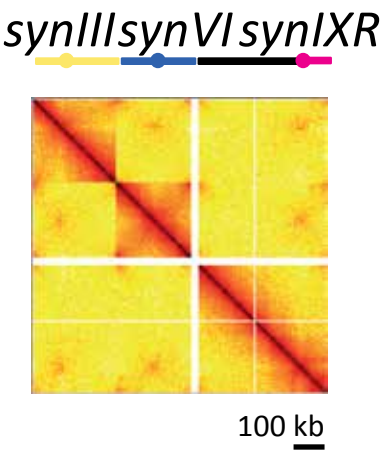
BY4742

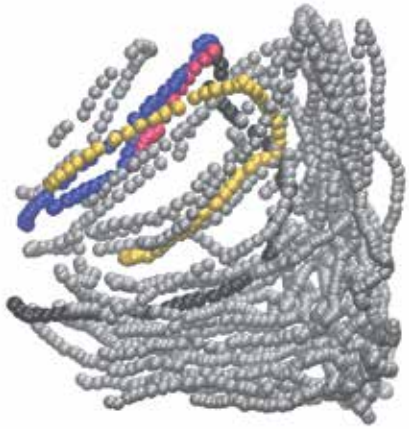






yLM896





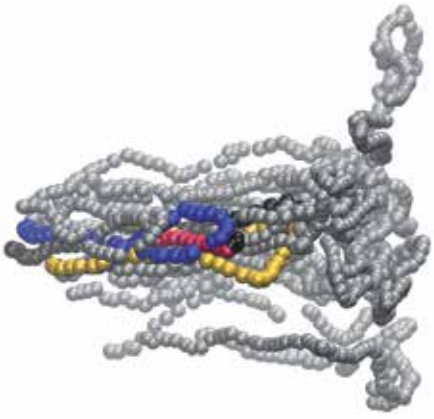
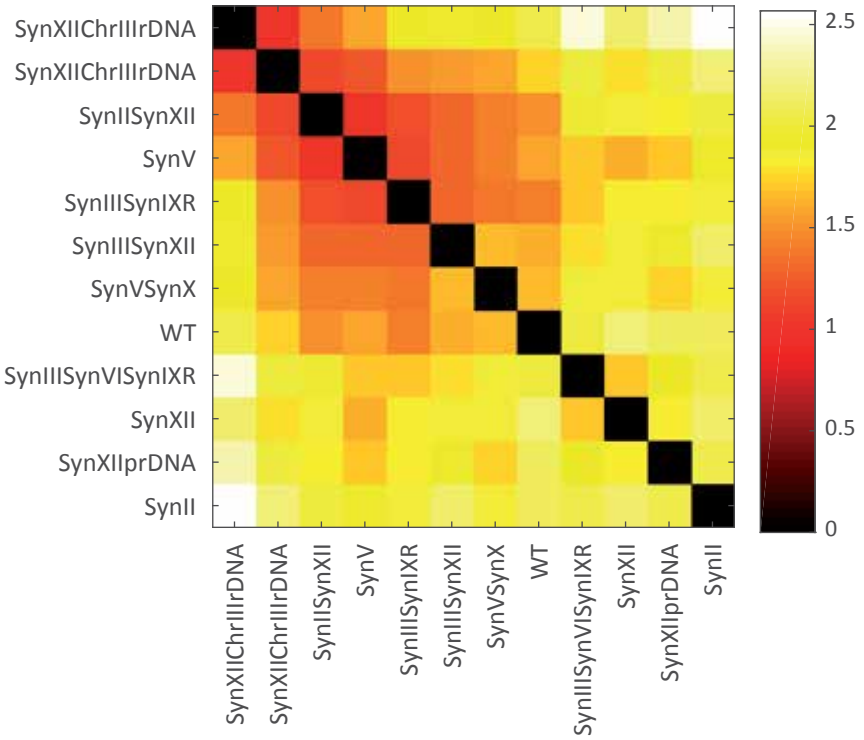


fig S11

A


SynXIIChrIIIrDNA
SynXIIChrIIIrDNA
SynIISynXII
SynV
SynIIISynIXR
SynIIISynXII
SynVSynX
WT
SynIIISynVISynIXR
SynXII
SynXIIprDNA
SynII
2.5
2
1.5
1
0.5
0

B

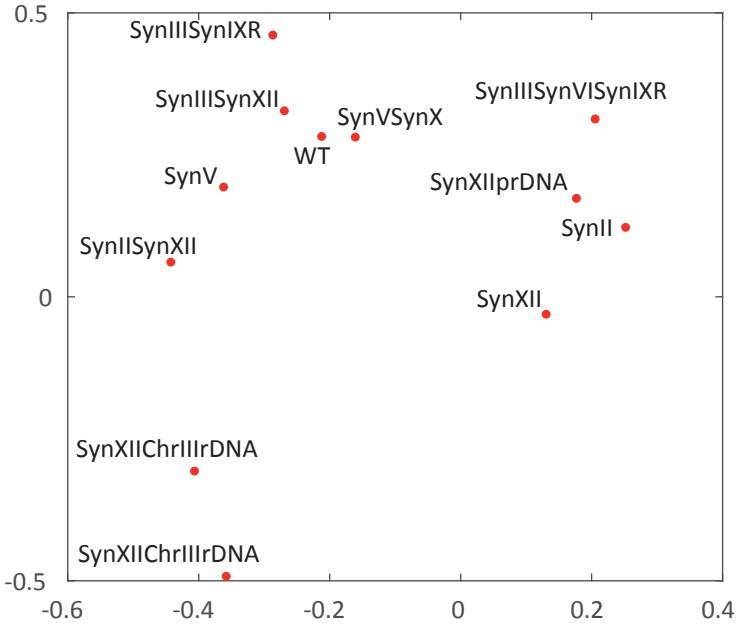

SynIIISynIXR
SynIIISynXII
SynVSynX
WT
SynV
SynIIISynVISynIXR
SynXIIprDNA
SynII
SynIISynXII
SynXII
SynXIIChrIIIrDNA
SynXIIChrIIIrDNA
0.5
0
-0.5
-0.6
-0.4
-0.2
0
0.2
0.4

fig S12

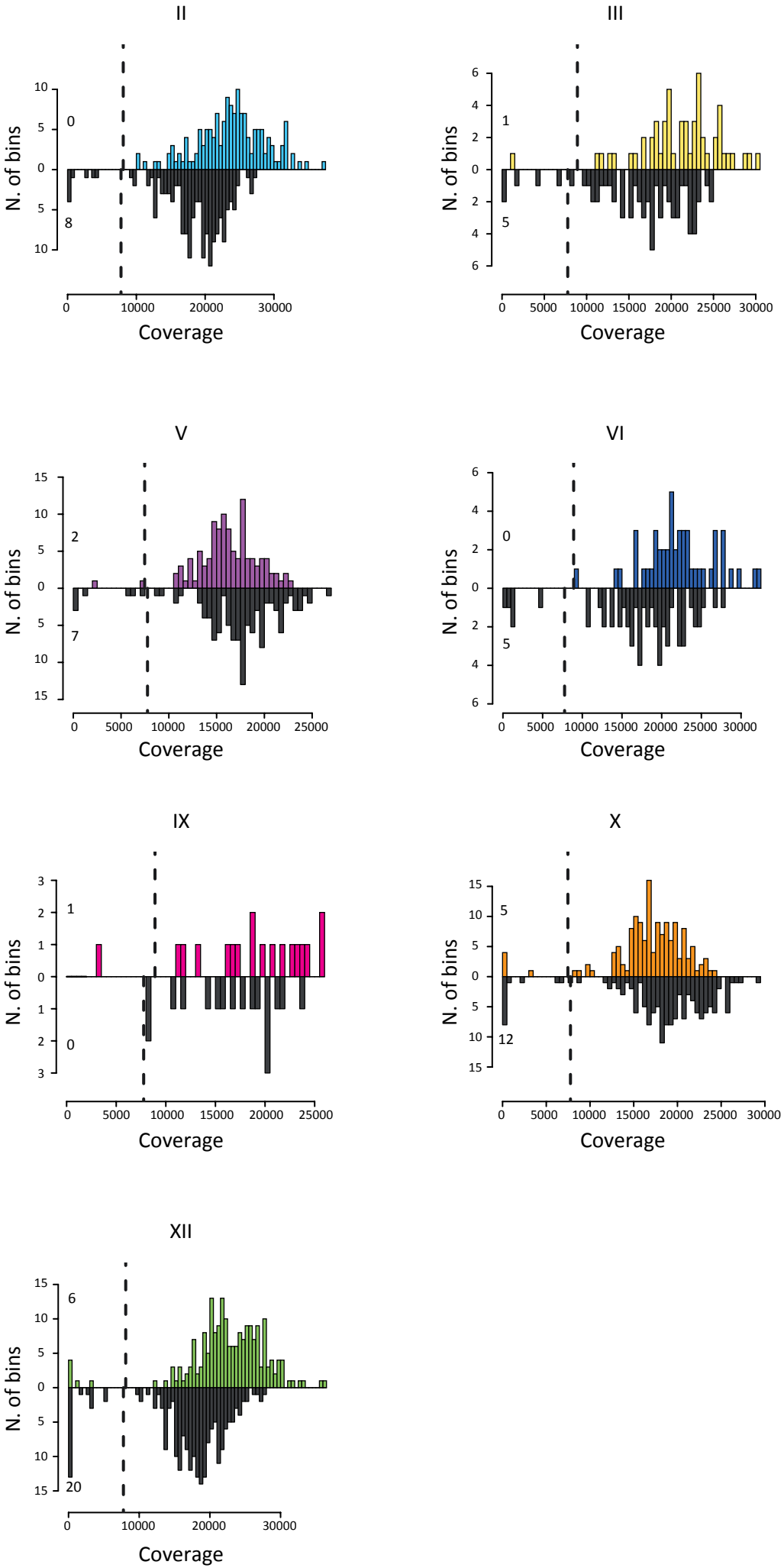


fig S13

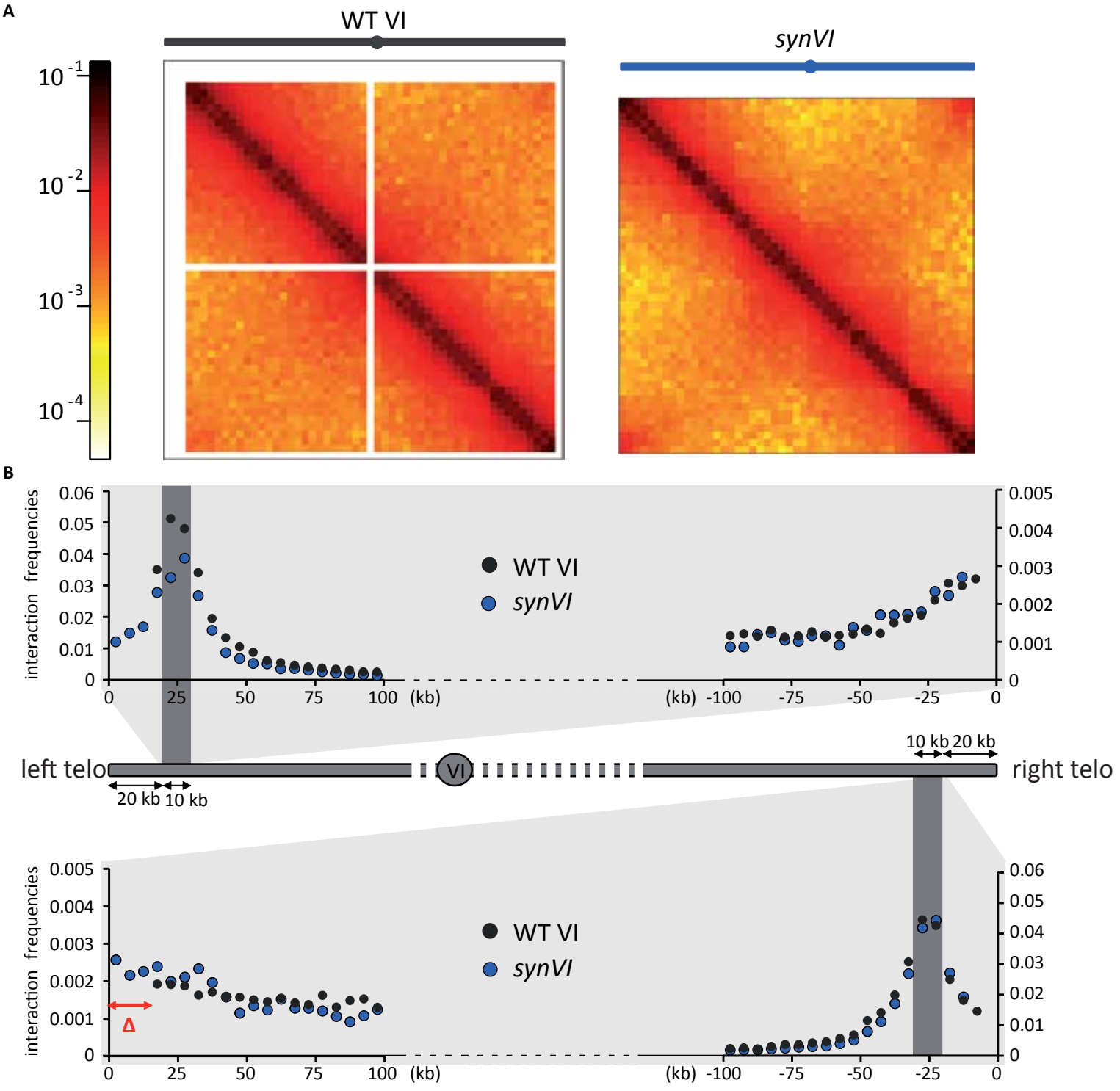
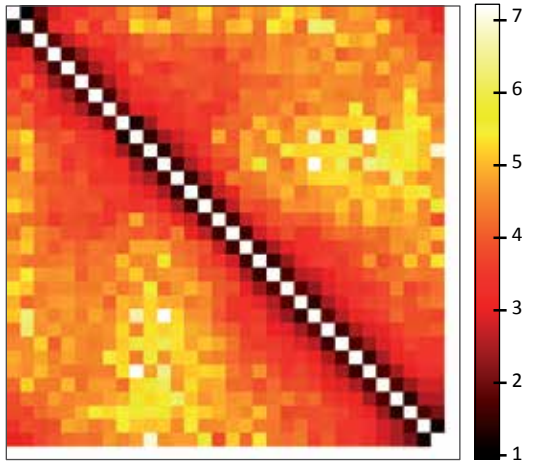
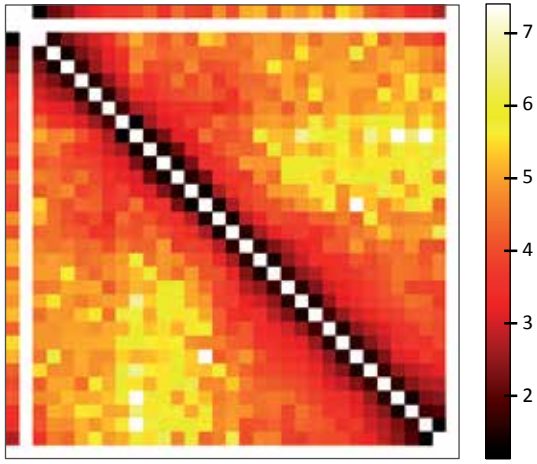

A
WT VI
synVI
10 -1
10 -2
10 -3
10 -4
B
interaction frequencies
0.06
0.05
0.04
0.03
0.02
0.01
0
0.005
0.004
0.003
0.002
0.001
0
25
50
75
100
(kb)
-100
-75
-50
-25
WT VI
synVI
10 kb 20 kb
left telo
VI
right telo
20 kb 10 kb
Δ

fig S14

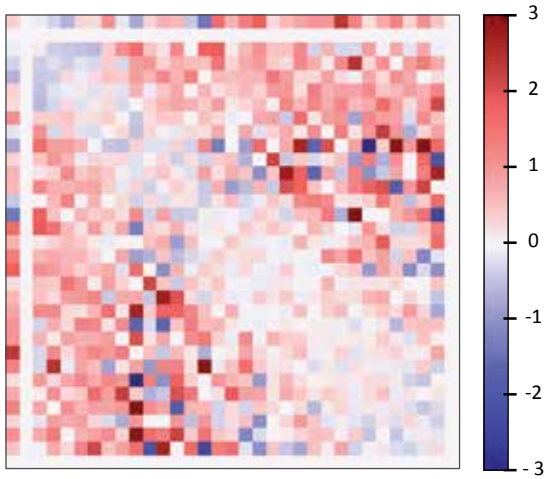
BY4741
WT *MATa*



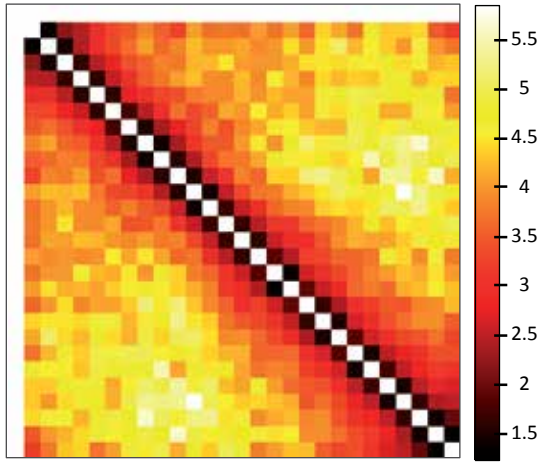
BY4742
WT *MATalpha*



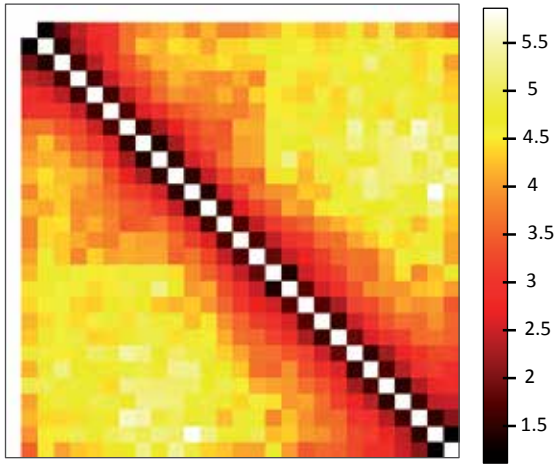
WT
MATa/MATalpha



HMSY012
synIII MATa



yLM422
synIII MATalpha



synIII
MATa/MATalpha

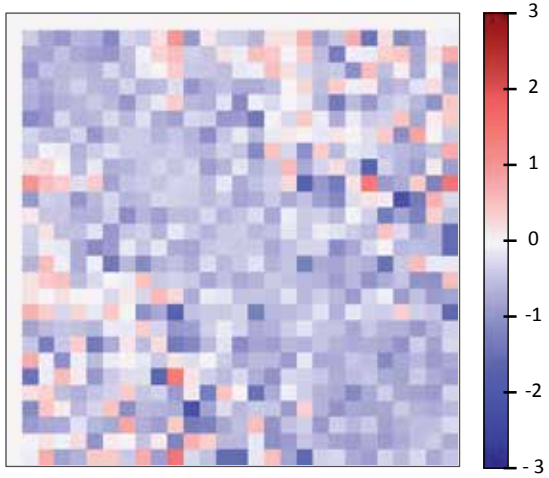
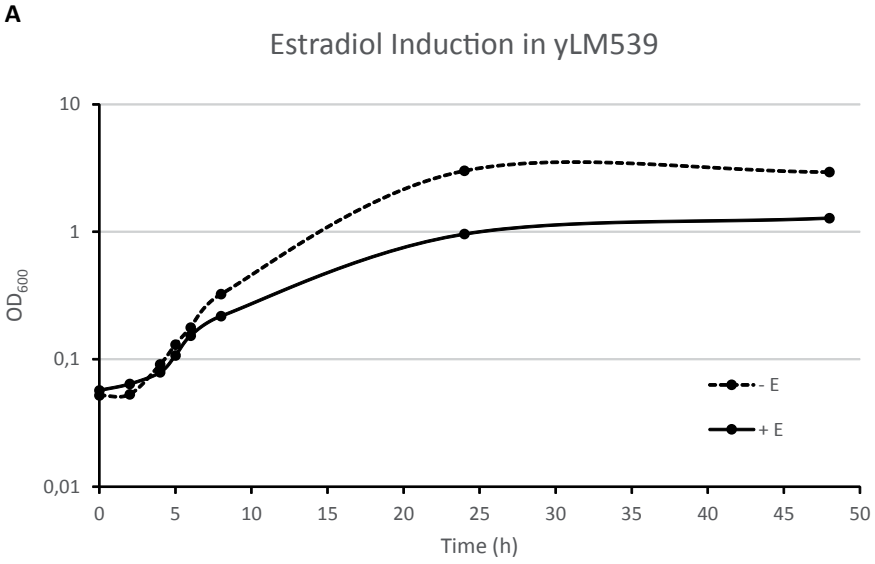


fig S15

A


Estradiol Induction in yLM539
10
1
0,1
0,01
OD600
0
5
10
15
20
25
30
35
40
45
50
Time (h)
- E
+ E

B

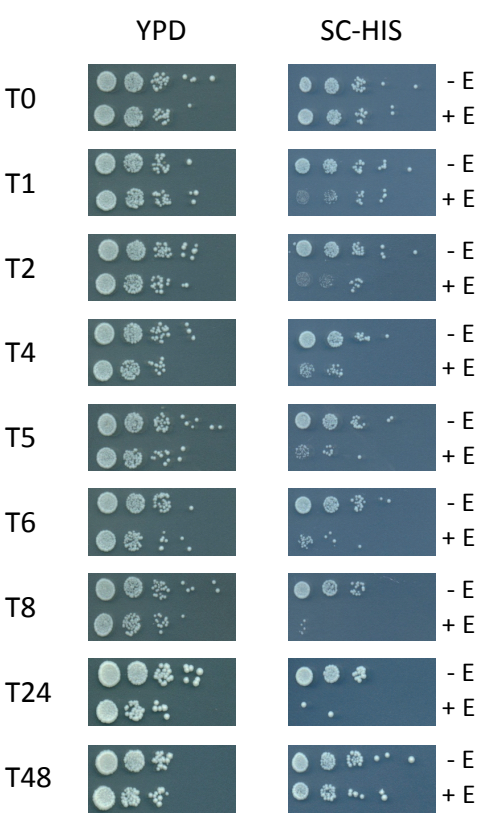

YPD
SC-HIS
T0
T1
T2
T4
T5
T6
T8
T24
T48
- E
+ E

fig S16

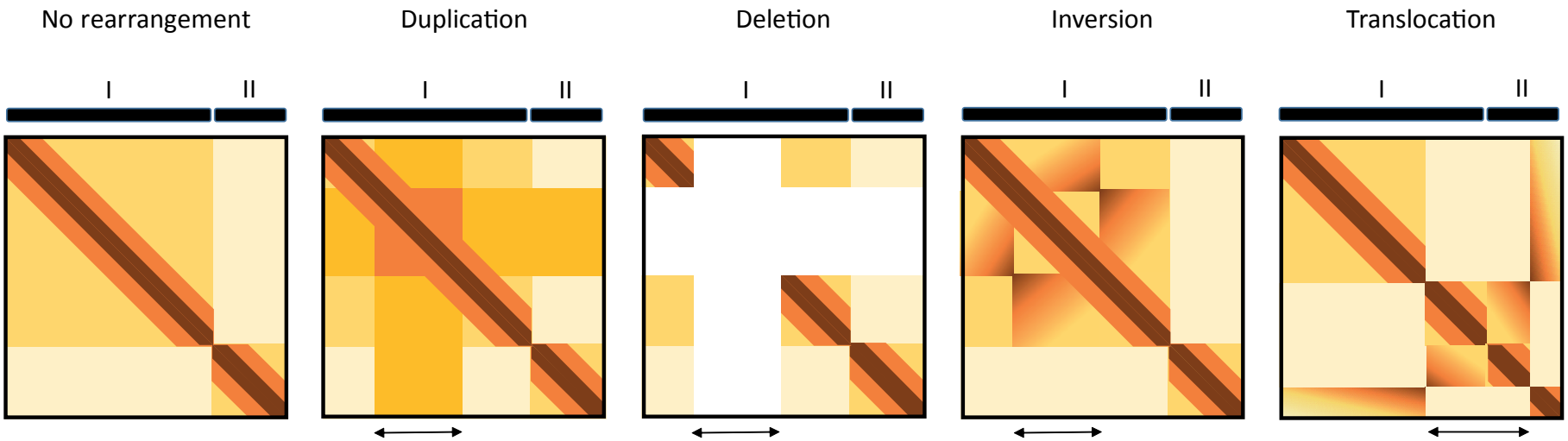
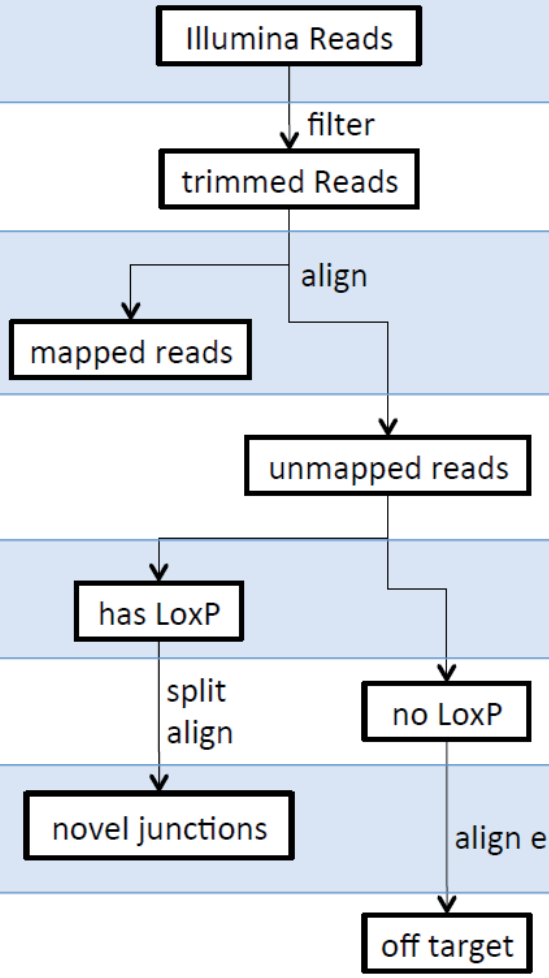

No rearrangement
Duplication
Deletion
Inversion
Translocation
I
II
I
II
I
II
I
II
I
II

fig S17



	0h	2h	8h
Illumina Reads	20,564,748	30,533,244	12,166,236
trimmed Reads	20,006,480	27,907,864	11,711,782
mapped reads	19,960,646	27,821,613	11,674,850
unmapped reads	45,834	86,251	36,932
has LoxP	29	163	358
no LoxP	45,805	86,088	36,574
novel junctions	0	4	17
off target	0	0	0

fig S18

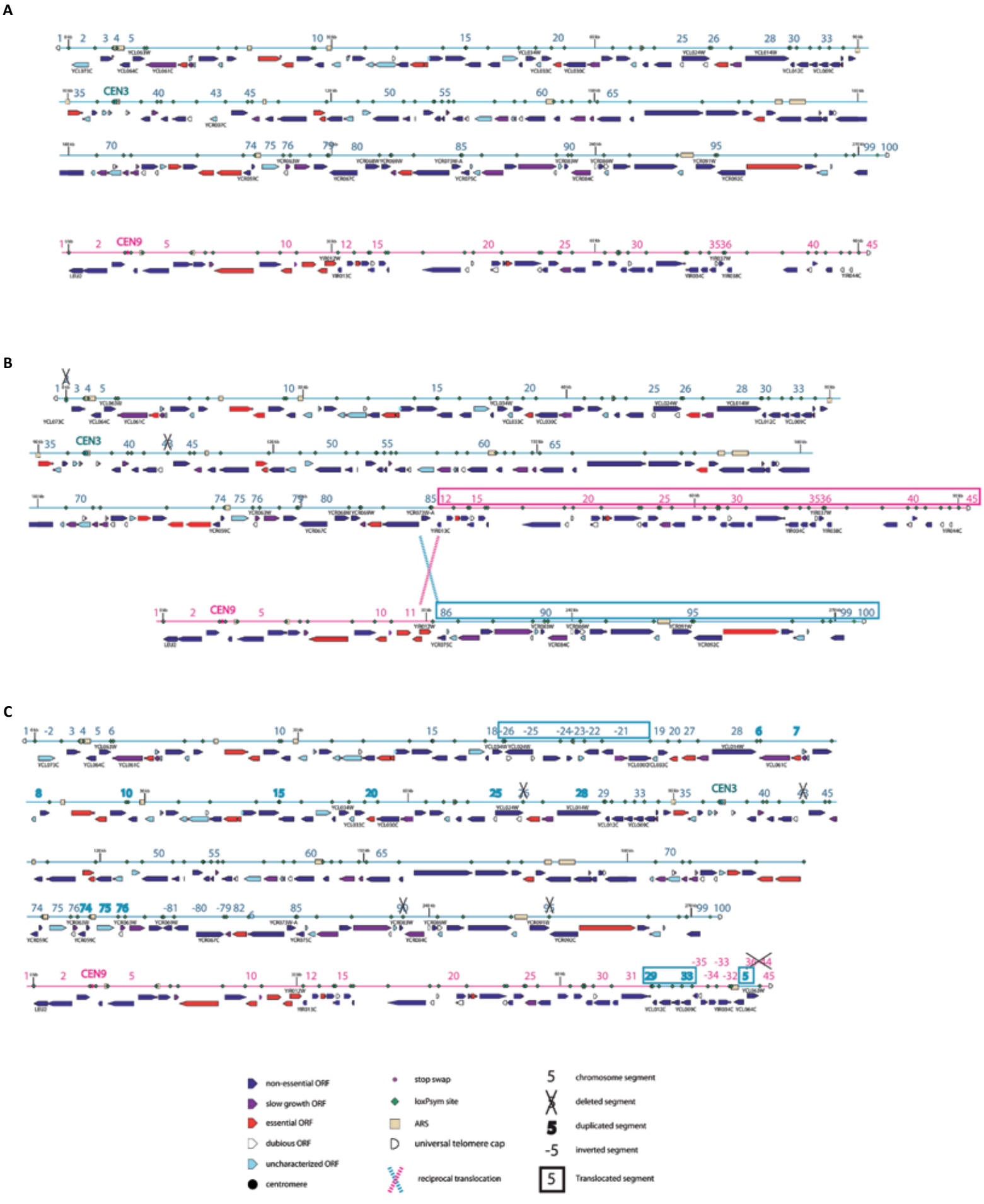
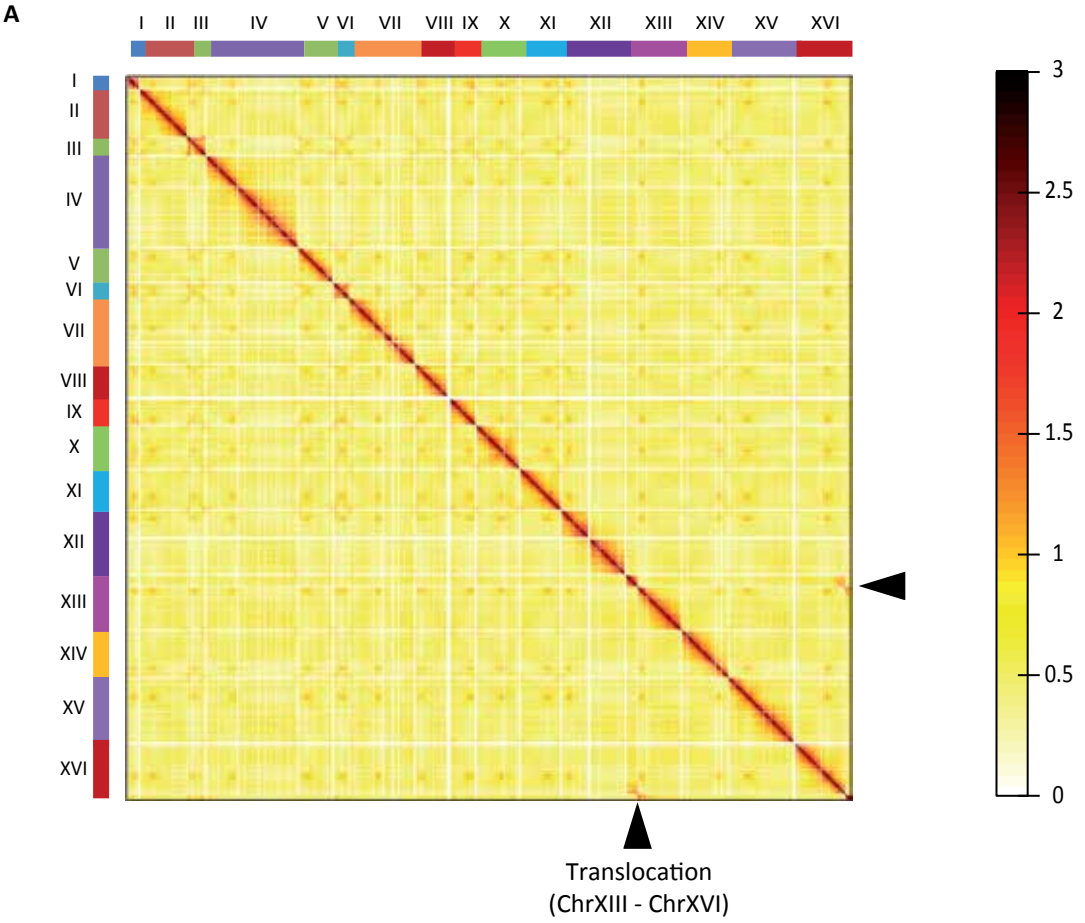
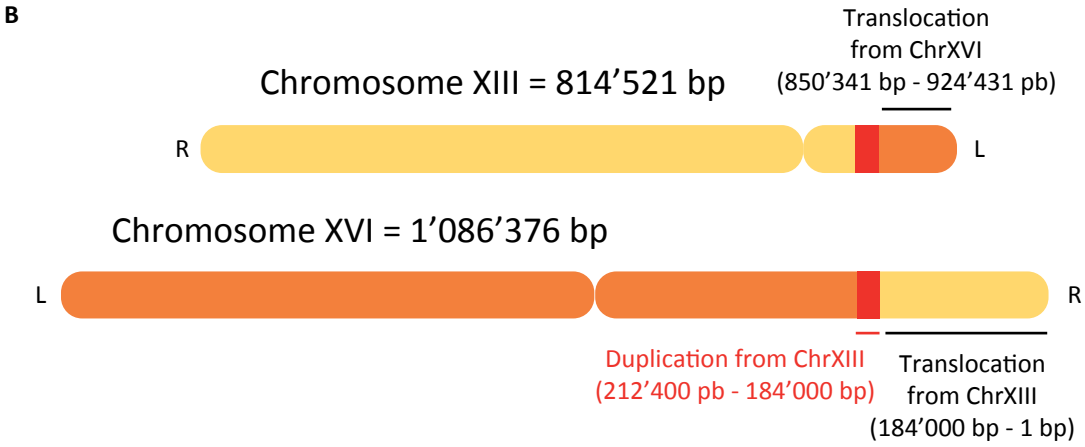
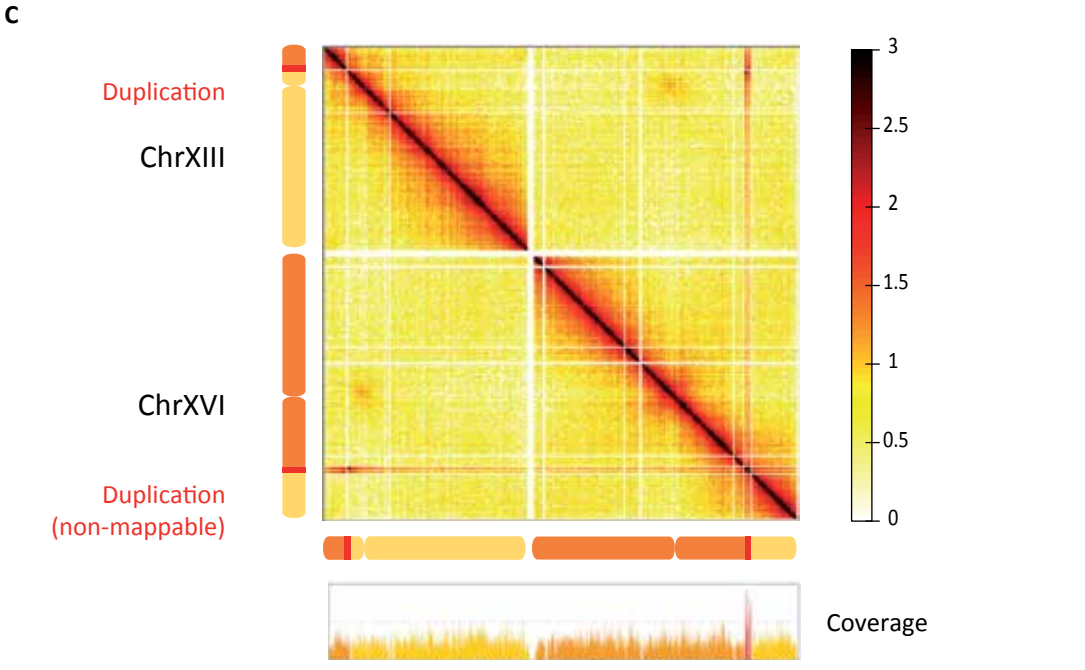

A
B
C
CEN3
CEN9
non-essential ORF
slow growth ORF
essential ORF
dubious ORF
uncharacterized ORF
centromere
stop swap
loxPsym site
ARS
universal telomere cap
reciprocal translocation
chromosome segment
deleted segment
duplicated segment
inverted segment
Translocated segment

fig S19


A
I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
3
2.5
2
1.5
1
0.5
0
Translocation
(ChrXIII - ChrXVI)


B
Translocation
from ChrXVI
(850'341 bp - 924'431 pb)
Chromosome XIII = 814'521 bp
R
L
Chromosome XVI = 1'086'376 bp
L
R
Duplication from ChrXIII
(212'400 pb - 184'000 bp)
Translocation
from ChrXIII
(184'000 bp - 1 bp)


C
Duplication
ChrXIII
ChrXVI
Duplication
(non-mappable)
3
2.5
2
1.5
1
0.5
0
Coverage