Tethering of molecular parasites on inactive chromatin in eukaryote nucleus

Fabien Girard

Sorbonne Université

ED 515 Compléxité du Vivant

Institut Pasteur de Paris, Unité Régulation spatiale des génomes

Tethering of molecular parasites on inactive chromatin in eukaryote nucleus

Par Fabien Girard

Thèse de Doctorat en génétique

Dirigée par Pr Romain Koszul et Dr Axel Cournac

Présentée et soutenue publiquement le 18 Décembre 2023

Devant un jury composé de:

Dr Stéphane Marcand, Chef d'équipe CEA	Président du jury
Dr Héloïse Muller, CR CNRS	Examinatrice
Dr Gianni Liti, DR CNRS	Rapporteur
Pr Fabienne Malagnac, PU UPsaclay	Rapportrice
Pr Romain Koszul, DR CNRS, Pr Institut Pasteur	Directeur de thèse
Dr Axel Cournac, CR Pasteur	Co-directeur de thèse

Summary

I. Illustration table

II. Abbreviations table

DNA : DesoxyriboNucleic Acid

ecDNA: Extra-Chromosomal DNA

eccDNA: Extra-Chromosomal Circular DNA

RNA : Ribonucleic Acid

tRNA : Transfer RNA

mRNA : messenger RNA

miRNA : MicroRNA

siRNA : Small Interfering RNA

piRNA : PIWI interacting RNA

T2T : Telomere to telomere

Mb : Mega base pair

Kb : Kilo base pair

bp : DNA base pair

cM : Centi Morgan

FRT : Flippase Recognition Target

MGE : Mobile Genetic Element

HBV : Hepatitis B Virus

KSHV : Kaposi's Sarcoma-associated HerpesVirus

EBV : Epstein-Barr Virus

ORF : Open Reading Frame

HGT : Horizontal Gene Transfer

NLS : Nuclear Localisation Signal

cGAS : cyclic GMP-AMP Synthase

cGAMP : cyclic GMP-AMP

H-NS : Histone-like Nucleoid-Structuring protein

NAP : Nucleoids Associated Proteins

RISC : RNA-Induced Silencing Complex

RM : Restriction-Modification

RE : Restriction Enzyme

PRR : Pattern Recognition Receptor

PAMPs : Pathogen-Associated Molecular Pattern

NPC : Nuclear Pore Complex

Ty : Transposons of Yeast

3C : Chromosome Conformation Capture

III. Remerciements

Ce travail n'aurait pas vu le jour sans le soutien et l'aide de nombreuses personne.

Je remercie les membres de mon comités de suivis de thèse qui ont suivis mes travaux. Je remercie Gianni Liti et Fabienne Malagnac qui ont acceptés d'être rapporteur de ce manuscrit ainsi que Stéphane Marcand et Héloïse Muller pour avoir accepté de faire partie de mon jury de thèse.

Je remercie l'Ecole Normale Supérieure Paris-Saclay pour m'avoir accordé un financement CDSN.

Merci Stéphane pour ton soutien dans ma quête pour faire une thèse et ta recommandations auprès de Romain.

Je remercie mes directeurs de thèse, Romain Koszul et Axel Cournac qui m'ont permis de travailler sur des sujets aussi passionnant que le plasmide 2-micron et la mise au point technique du « multi-contact Hi-C ». J'ai énormément appris au cours de ces trois (quatre) dernières années et j'espère continuer à apprendre. Je voudrais aussi vous remercier d'avoir respecté ma vie de jeune père qui parfois devait rester à la maison, et tout le temps partir tôt.

J'ai eu la chance d'effectuer mes travaux dans un formidable laboratoire dans lequel j'ai trouvé plus que des collègues de travail et une convivialité si rafraichissante.

Je voudrais remercier non pas le directeur de thèse mais le chef d'équipe. Merci Romain de m'avoir engagé après mon passage à l'éducation nationale. Pour avoir cru en moi pour rebondir et réussir ma reconversion professionnelle.

Je remercie Agnès pour sa bienveillance et tous ses conseils, sans qui je saurais faire beaucoup moins de choses. Léa et Amaury mes compagnons de thèses et de galères. Les membres du vrai bureau de l'ambiance pour avoir supporté mes gouts musicaux douteux et avec qui il est très plaisant de rire aux éclats. Martial et Axel pour nos discussion scientifiques et celles moins scientifiques. Et globalement, tous les membres du laboratoire RSG avec qui j'ai pu décompresser et échanger.

Je remercie plus globalement les personnes des services de préparation, d'entretien, logistique etc. qui m'ont permis de travailler dans de bonnes conditions. Je remercie toutes les personnes qui se sont occupés de mon fils et notamment ma mère et ma grand-mère, qui ont sauvé de nombreuses manips sans vraiment le savoir.

Je remercie tout mes enseignants qui m'ont permis d'élargir mes horizons et d'osez me lancer dans de longues études.

Je voudrais remercier ma famille qui ne comprends pas toujours tout ce que je fais au travail et m'a soutenu dans les temps difficiles.

Enfin, je remercie ma Femme, qui est à mes côtés depuis quelques années maintenant et m'a toujours soutenue dans mes choix. Même lorsque je lui ai annoncé que je rendais l'agreg en démissionnant et que je voulais faire une thèse. Grace à toi, *je n'ai plus peur du dimanche soir* car *la réponse à toutes mes questions s'endort à mes côtés.*

Je te remercie mon fils d'être ce que tu es. Si un jour il te prend l'envie de comprendre ce que papa a fait pendant tes premières années, je serais heureux de tout t'expliquer.

« Bah ouais c'est sûr c'est la merde, c'est pas trop ça qui était prévu

Nos ambitions sont en berne et notre avenir en garde à vue

[…]

Y'avait surement plusieurs options mais finalement on a opté

Pour accepter cette position, trouver un espoir adapté »

Espoir Adapté, Grand Corps Malade

IV. Abstract

Natural plasmids are common in prokaryotes, but few have been documented in eukaryotes. The natural 2µ plasmid present in budding yeast *Saccharomyces cerevisiae* is one of the most well characterized. This highly stable genetic element coexists with its host for millions of years, efficiently segregating at each cell division through a mechanism that remains poorly understood. Using proximity ligation (Hi-C, Micro-C) to map the contacts between the 2µ and yeast chromosomes under dozens of different biological conditions, we found that the plasmid tether preferentially on regions with low transcriptional activity, often corresponding to long inactive genes. Common players in chromosome structure such as members of the structural maintenance of chromosome complexes (SMC) are not involved in these contacts which depend instead on a nucleosomal signal associated with a depletion of RNA Pol II. These contacts are stable throughout the cell cycle and can be established within minutes. This strategy may involve other types of DNA molecules and species other than *S. cerevisiae*, as suggested by the binding pattern of the natural plasmid along the silent regions of the chromosomes of *Dictyostelium discoideum*.

V. Résumé

Les plasmides naturels sont courants chez les procaryotes, mais peu ont été documentés chez les eucaryotes. Le plasmide naturel 2µ présent dans la levure bourgeonnante *Saccharomyces cerevisiae* est l'un des mieux caractérisés. Cet élément génétique très stable coexiste avec son hôte depuis des millions d'années, ségrégeant efficacement à chaque division cellulaire par un mécanisme qui reste mal compris. En utilisant la ligature de proximité (Hi-C, Micro-C) pour cartographier les contacts entre le plasmide 2µ et les chromosomes de levure dans des dizaines de conditions biologiques différentes, nous avons constaté que le plasmide 2µ se fixe préférentiellement sur des régions à faible activité transcriptionnelle, correspondant souvent à de longs gènes inactifs. Les acteurs communs de la structure des chromosomes, tels que les membres des complexes de maintenance structurale des chromosomes (SMC), ne sont pas impliqués dans ces contacts qui dépendent plutôt d'un signal nucléosomique associé à une déplétion de l'ARN Pol II. Ces contacts sont stables tout au long du cycle cellulaire et peuvent être établis en quelques minutes. Cette stratégie peut aussi être trouvée dans d'autres types de molécules d'ADN et d'autres espèces que *S. cerevisiae*, comme le suggère le schéma de liaison du plasmide naturel le long des régions silencieuses des chromosomes de *Dictyostelium discoideum*.

1 Introduction

1.1 Extra chromosomal DNA, welcome guest, or undesirable alien

1.1.1 Defining ecDNA and chromosomal DNA: making the best of the available data

1.1.1.1 Exploring the DNA content of a cell and building its genome

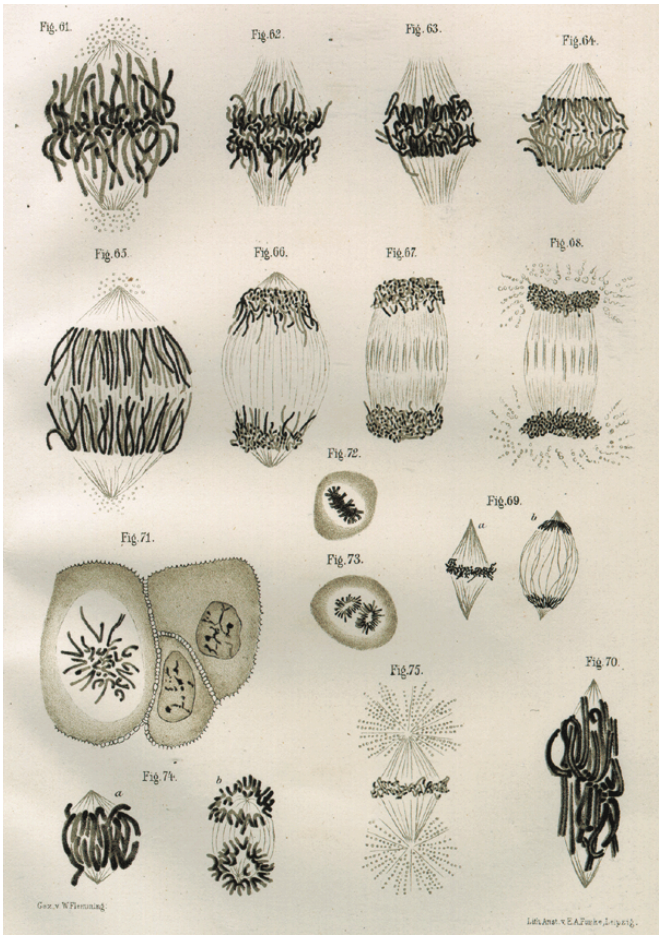


Figure 1. Drawing by Walther Fleming depicting metaphase chromosomes in eukaryotic cells (Flemming Zellsubstanz, Kern und Zelltheilung*, 1882)*

The first use of "genome" is recent compared to the human history, dating back around 100 years ago. However, the questioning about heredity that led to the first definition of the genome dates back the ancient Greek. Hippocrates[1] proposed that fluids coming from all organs converge to genital organs forming the seeds that will be exchanged during sexual intercourse. Hippocrates latter inspired Darwin for the pangenesis theory. The first observation of DNA in a cell was made by Walther Flemming in 1880's with the first observation[2] of chromosomes in eucaryotic cells dyed with anillin (Figure 1). It gave rise to multiple cytology studies that allowed to follow chromosomes in the nucleus during cell division. The large, compact metaphase chromosomes could be extracted from the nucleus and sorted by their size and coloration intensuty, leading to define karyotypes. Presumably consisting in the full chromosomal set. Before the DNA -sequencing technologies and even the demonstration that DNA is the support of genetic information[3] in 1943. The phenotypic analysis of an organism and its progeny i.e., study of all the observable trait of an organism such as the flower colour or the shape of the seed, was the first genetic experiment conducted by Gregor Mendel during XIXth century on edible pea, *Pisum sativum*. However, Mendel's work rose controversy and the link between the phenotype, genes and chromosomes was established in the 1900's by Carl Oven.

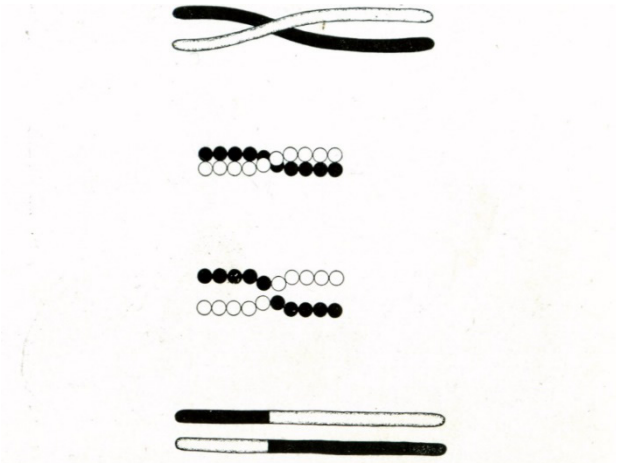


Fig. 64. Scheme to illustrate a method of crossing over of the chromosomes.

Figure 2. Illustration of crossing over from TH morgan, 1916 A Critique of the Theory of Evolution

Thanks to the pioneer work of Thomas Hunt Morgan on *Drosophilia melanogaster*, the phenotypic analysis of the progeny following meiosis, the specialized cell division that generates haploid gametes from a diploid cell through a series of carefully regulated structural and genetic processes, provided an important tool to map genes on the chromosomes by their genomic distance. TH morgan took advantage of the crossing over happening between homologous chromosomes during meiosis (Figure 2). If genes are on the same chromosome, then they are more often on the same exchanged chromosomal portion. Thus, the closer they are on the chromosome the more they will co-segregate in the progeny. The genomic distance was then measured in centimorgan (cM)[4]. After the publication of Morgan and Haldane, the first genetic map of chromosomes from *D. melanogaster* was published in 1913 by Alfred Strurtevant[5]. The findings discussed above led Hans Winkler[6] to define the genome as the chromosomal set and protoplast that are the material foundation of a specie. This simple definition became more and more complex over time with discoveries in eukaryotes, prokaryotes, archaea, and viruses. Nowadays, the term genome is used to refer to the molecules (DNA or RNA) supporting genetic information in a living organism or a virus.

Since the late 1990's, the advent of whole genome sequencing, consisting in characterizing base per base the entire DNA content of a cell, became the ultimate technique to characterise the genome of a specie. The genome of *Saccharomyces cerevisiae* was totally sequenced in 1996 and the complete sequence of the 16 chromosomes, containing 6275 genes was published [7]. The chromosomal DNA of *S. cerevisiae* was the first of many to be sequenced

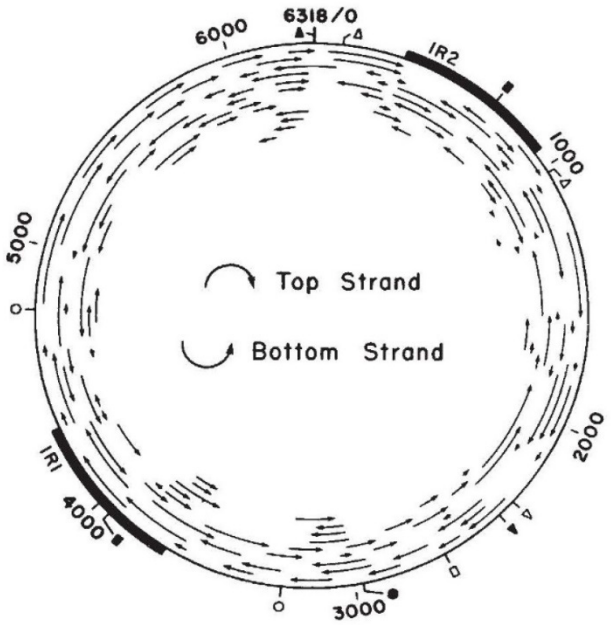


Figure 3. Scheme of the first assembly of the 2µ plasmid genome. The circle represents the final sequence, each arrow represents a sequenced portion of the 2µ plasmid by Sanger technology. Note that every portion overlap each other, Hartley, and Donelson 1980

and was the results of years of collaboration between many laboratories. This study took 7 years and was done using Sanger sequencing. The genome was split in fragments of 5-10 kb. The long fragments were separated in smaller (~500 bp) overlapping fragments (Figure 3). The overlaps allow to reconstitute the genetic puzzle, the size was due technical limitations of the Sanger sequencing technology. Fast-forwarding 30 years or so of technical advances, high throughput sequencing now made it possible today to sequence the entire DNA content of almost any organism or virus within a day. This can be done using different technologies that together allow to precisely sequence the DNA content of an organism. Every technology has it perks and flaws. The Illumina technology gives high quality (Sequencing quality is about 40, meaning that the probability that a base pair is wrong is bellow 10^{-4}) short reads (up to 300 bp), Pacific Bioscience gives longer reads (up to 20 kb) with sequencing quality near 30 and Oxford nanopore technology output ultra-long (up to 300 kb) noisy (quality score is about 20) reads. The combination of those three technologies allowed the Telomere to Telomere (T2T) initiative to publish the first gapless human genome in 2022[8]. The tool developed by the T2T members[9] clearly show that the difficult part, that was not a problem with Sanger sequencing, is to assemble the little DNA chunks to build the genome. For living organism, the support of genetic information is DNA, and the genome refers to the chromosomal set plus mitochondrial, chloroplastic genomes when those are present, with chromosomes being defined as the longest DNA molecules in a cell which are part of the core genome. But what is a long or a short DNA molecule? In bacteria for instance, multiple DNA molecules, such as plasmids which size can range from 800bp to 2.5 Mb[10], can coexist and coevolve with the chromosome(s). In eukaryotes, mitochondrial or chloroplastic DNA is essential but is not included into the chromosomal set, hence the term "nuclear genome" to refer to nuclear DNA. Overall, eukaryotic nuclear genomes typically consist in the chromosomal DNA, but over time peculiar observation such as extra-

chromosomal DNA (ecDNA) further complexify this general definition. I will focus on the 2µ plasmid found in the model organism, *S. cerevisiae.*

1.1.1.2 What is a plasmid?

The term "plasmid" was first proposed in 1952 by Joshua Lederberg[11] to include any extra-chromosomal genetic material in a bacterium. This definition is more and more refined with new discoveries and now only includes extra-chromosomal DNA that are able to autonomously replicate and found predominantly as circular ecDNA[12] in any living organism without any discrimination whether it is essential or not since this feature depend on environmental conditions[13].

As for all definition in biology there is borderline cases[13]. Some bacterial plasmids, bacterial episomes exist as autonomously replicated ecDNA and are also able to integrate in the genome [14]. Whereas an episome in a eukaryotic cell[15–17] is closer to the prokaryotic plasmid because in eukaryotes episomes often refer to viral DNA that persists in the host cell by replicating autonomously and hitchhiking on mitotic chromosomes, meaning that it attaches itself to the host chromosome and takes advantage of its segregation to be transferred vertically. In eukaryotes, a plasmid relates to an extra chromosomal DNA that is not viral DNA and autonomously replicated.

1.1.1.3 The 2µ plasmid, history of a serendipity

As discussed above[18], the study of genetics began with Gregor Mendell's work on peas. He focused on the vertical transfer of a phenotype, laying the basis for the concept of the gene. As for many research fields, model organism tends to emerge because of their ease to be studied. In fact, the studies carried out by the pioneers of genetics were based on difficult model organisms. The sexual life cycle of plants or insect is rather long. Luckily, since the work of Louis Pasteur[19] on the isolation and the first pure culture of "lactic yeast" in 1858's, pure cultures of ascomycetes (such as *S. cerevisiae*, *Neurospora crassa*, *Schizosaccharomyces pombe*) could be easily obtained. Thanks to the pioneer work of Lindegren, Winge, and others, the gametes regrouped in an ascus could be isolated by microdissection and each spore could be phenotypically analysed[20]. *S. cerevisiae* became a major model organism for genetic, cell cycle and cell biochemistry study. This can be explained by the many advantages *S. cerevisiae* had. First of all, many different strains had already been isolated due to their ancestral use in

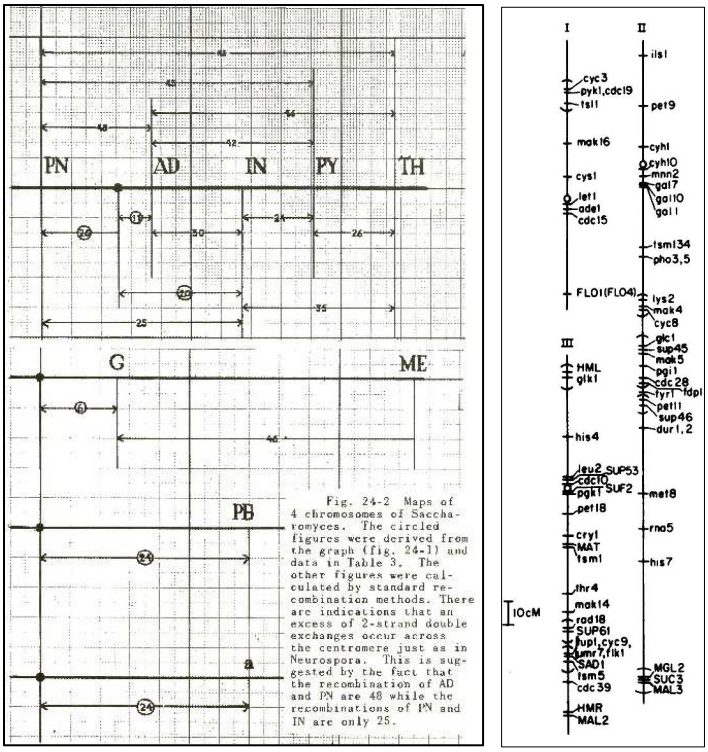


*Figure 4. (**left**)Drawing from C. Lindegren of the first genetic map of S. cerevisiae build with 9 genes (Lindegren ,1949), (**right**) genetic map drawn in 1980 by Mortimer and Schild.*

breweries and bakeries[21], which enabled Winge and Lindegren, who were working on the fermentative capacity of yeast, to identify the first genes of *S. cerevisiae*. The first genetic map of *S. cerevisiae* was produced by Lindegren using the available set of genes[22] (Figure 4,left). Second, the vegetative growth and sexual cycle of *S. cerevisiae* made experiments shorter. Finally, *S. cerevisiae* is not a pathogenic micro-organism. The extensive study of *S. cerevisiae* is a consequence of its ease to be used as a model and several important discoveries were made, such as the non-mendelian heredity of the “petite” phenotype by Boris Ephrussi and Piotr Slonimski[23], opening the way for the human mitochondrial disease; the generation of thermosensitive mutant led to the first understanding of cell cycle[24]; the genome of *S. cerevisiae* proved to be a convenient eukaryotic genome to work with. Its genome is small (12 Mb) and contains few repetitive sequences, making it easy to identify genes and other genomic features. Finally, its genome is very compact and intergenic regions are small. That is why *S. cerevisiae* was the first eukaryote entirely sequenced upon the impulsion by André Goffeau and Bernard Dujon[7] opening the way for the functional genetic studies and giving information on the requisite for eukaryotic life.

The first report of the 2µ plasmid dates back from 1967[25]. It was discovered in DNA preparation purified with caesium chloride gradient; a common technique used to separate the different DNA from a cell based their density. The 2µ plasmid was the most abundant circular satellite DNA and its name came from the contour length of the circle observed using electronic microscopy. This report was later confirmed by other teams[26]. Originally the community thought that the 2µ plasmid was in the mitochondria[27,28] but this was rapidly challenged , with teams claiming it was in the nucleus[26–29]. The main argument for the localisation of the 2µ plasmid in the mitochondria was its non-Mendelian inheritance that resembled the one from the

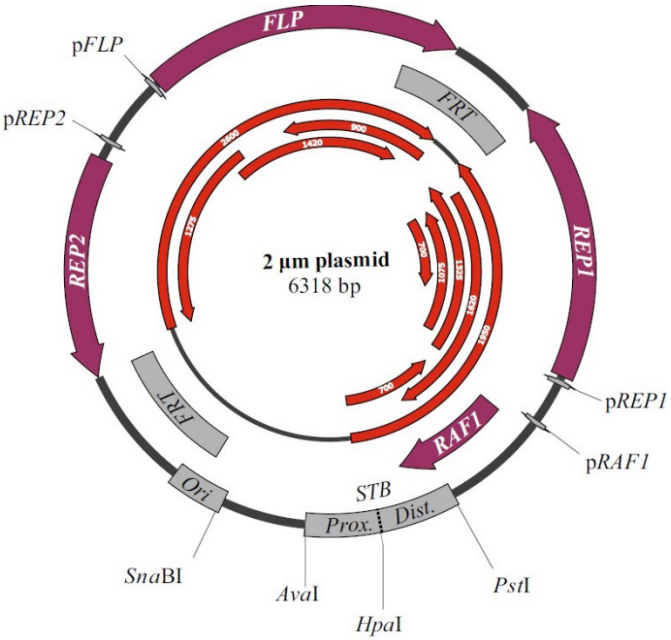


Figure 5. Genetic map of the 2µ plasmid depicting the 4 ORFs, loci of interest (STB, ori, FRT) and the long ncRNAs (Rizvi et al. 2018)

mitochondrial associated "petite" phenotype. Upon the accumulation of evidence of its nuclear localisation, this debate ended near 1980. The same year its complete sequence was published[30]. It revealed that the 2µ plasmid has a peculiar genetic structure (Figure 5). An autonomous replicated sequence, two inverted repeats(FRT) and a specific recombinase Flp1, required for its copy number homeostasis[31] (~60 per nucleus), the STB locus (formerly REP3) required for its stability in its host[32] and three other ORFs (REP1, REP2 and RAF1)[31,33]. Since plasmid free cells ([cir^0]) could be isolated from culture of plasmid positive cells ([cir$^+$]) without displaying any phenotype[34], the idea that the 2µ plasmid was a molecular parasite first appeared in the title of the publication from Mead et al. in 1986[35] with the help of the selfish gene theory of Richard Dawkins[36] and publications of W. Ford Doolittle & Carmen Sapienza[37].

The 2µ plasmid genetic systems has made it easier to study *S. cerevisiae* and genetics in general. Its backbone was the basis of the first stable bio-engineered plasmid in *S. cerevisiae*[38], which eased the transfer of genetic material in *S. cerevisiae*. Furthermore, The FRT/Flp1 is a genome editing strategy that was extensively used before the advent of CRISPR/Cas system. If *S. cerevisiae* had not been a model organism such extensively studied, the 2µ plasmid would most certainly be left unnoticed. However, the 2µ plasmid represents a minor part of all described extra chromosomal DNA in living organisms.

1.1.1.4 Circular extra chromosomal DNA diversity

1.1.1.4.1 Transposable elements

Transposable elements are genetic material found in genomes from both eukaryotes[39] and prokaryotes[40]. They can change their location in the genome thanks to their integrase. Transposable elements are divided in two groups. First, the transposons can move in the genome

by a cut/paste mechanism. Enzymes encoded by the transposon can excise the transposable cassette from the genome and integrate it at a different locus which can be found either in the host genome or on a cDNA element. Second, the retrotransposons are characterized by their RNA intermediate. They originate from retrovirus that co-evolved with their host. Retrotransposons are transcribed in RNA then retro-transcribed into DNA by a retro transposase encoded by the retrotransposon. The DNA is then integrated in the host genome or on ecDNA (especially in prokaryotes where ecDNA is abundant). The retrotransposons are theoretically able to increase their number with each transcription/retro-transcription cycle (copy/paste) and invade genomes. However, defence systems against retro-transposons often prevent genome invasion (discussed below)[41].

1.1.1.4.2 Non transposable ecDNA in prokaryotes

Non transposable ecDNA is often smaller than the main chromosome and is part of accessory genome. It encodes accessory functions that can be useful for the host cell (antibiotic resistance[42], toxic secreted compound[43] or even the ability to use a nutrient) or involved in the replication/transfer cycle of the ecDNA itself. Two types of non-transposable ecDNA are described in prokaryotes: the plasmid and the genome of bacteriophage. The two types of ecDNA can either be linear or circular and the distinction is made based on their genetic structures, encoded functions, and propagation strategies. Bacteriophages are characterised by the de-novo production of viral particles. There is a distinction between lytic phages which are always active and, in most cases, kill their host and tempered/satellite phages that tend to be latent and persist in their host. Latent phages can be re-activated in certain environmental conditions and enter in a lytic cycle. However, many borderline cases are reported now and the distinction between those two classes can be tedious. For instance, some bacteriophages are only active and produces viral capsid in given environmental conditions. They remain as free ecDNA in the nucleus thus looking like a plasmid. With transposable elements, the ecDNA in prokaryotes are the Mobile Genetic Elements (MGE), i.e., genetic material able to move within a genome or be transferred from one organism to another.

In eukaryotes, ecDNA is less abundant. However, *S. cerevisiae* is not the only eukaryote with stable nuclear plasmids and recently extra chromosomal circular DNA(eccDNA) has been discovered.

1.1.1.4.3 Other persistent nuclear plasmids

Over the year following the discovery of the 2µ plasmid, the community focused on finding other nuclear plasmid in other yeasts. Such plasmid could be the basis of simple and stable DNA vectors. Related plasmids were found in *Zygosaccharomyces rouxii*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces fermentati*, *Torulaspora delbrueckii*, *Lachencea lactis* and *Lachencea waltii*[44]. All those species are closely related to *S. cerevisiae*[45] and we will discuss below about the evolution of the 2µ plasmid and its derivatives.

More surprisingly, another eukaryotic model organism, the amoeba *Dictyostelium discoideum*, also have persistent nuclear plasmids. The Ddp plasmid family was discovered by Metz et al. in 1983[46] who were screening for genes involved in the cobalt resistance phenotype of *D. discoideum* (as Lindegren and Winge did with fermentation ability of *S. cerevisiae*). In fact, the Ddp1 plasmid gives the cobalt resistance phenotype, but several other plasmids were found in other isolates without the same phenotypic outcome. Those plasmids will later be called Ddp2, Ddp3, Ddp4, Ddp5.

1.1.1.4.4 Other persistent nuclear, circular DNA in eukaryotes.

Extra-chromosomal DNA is not restricted to plasmid for eukaryotes. As mentioned above, some vertebrate viruses are episomal, meaning that they don't integrate in the genome as other viruses, such as Hepatitis B Virus (HBV) can do[47]. They remain outside of the chromosome throughout their cycle. Episomal viruses such as Epstein Barr Virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV) are either active or latent in their host[15]. Latent cycle is characterised by the few expressed genes required for the maintenance of the viral episome and no production of viral particle. The active or lytic cycle is characterised by the expression of viral genetic material in response to the cellular environment (oxidative stress, aging, decrease in immune system efficiency) and lead to de-novo production of viral particles.

Second, upon DNA metabolic activity (such as replication, transcription and DNA damage repair), parts of the host chromosomes can be excised and persist as extra chromosomal circular DNA (eccDNA)[48,49]. The presence of eccDNA is reported in cells coming from healthy human patients[50] and also in pathological cells such as cells coming from tumors[51]. In most cases, cccDNA does not contain replication origins and is therefore not transferred vertically in a stable manner[52]. However some of them are able to replicate on their own and can be transmitted to the progeny by hitchhiking on host chromosomes.

In the case where multiple DNA molecules uses the same resource, the interactions between the chromosomal and extra-chromosomal DNA can really shape the phenotype of a cell. The first step before establishing a relationship between genomes is the physical meeting between host genome and ecDNA.

1.1.2 Origin of extra chromosomal DNA

1.1.2.1 DNA from exogenous source enter the cell.

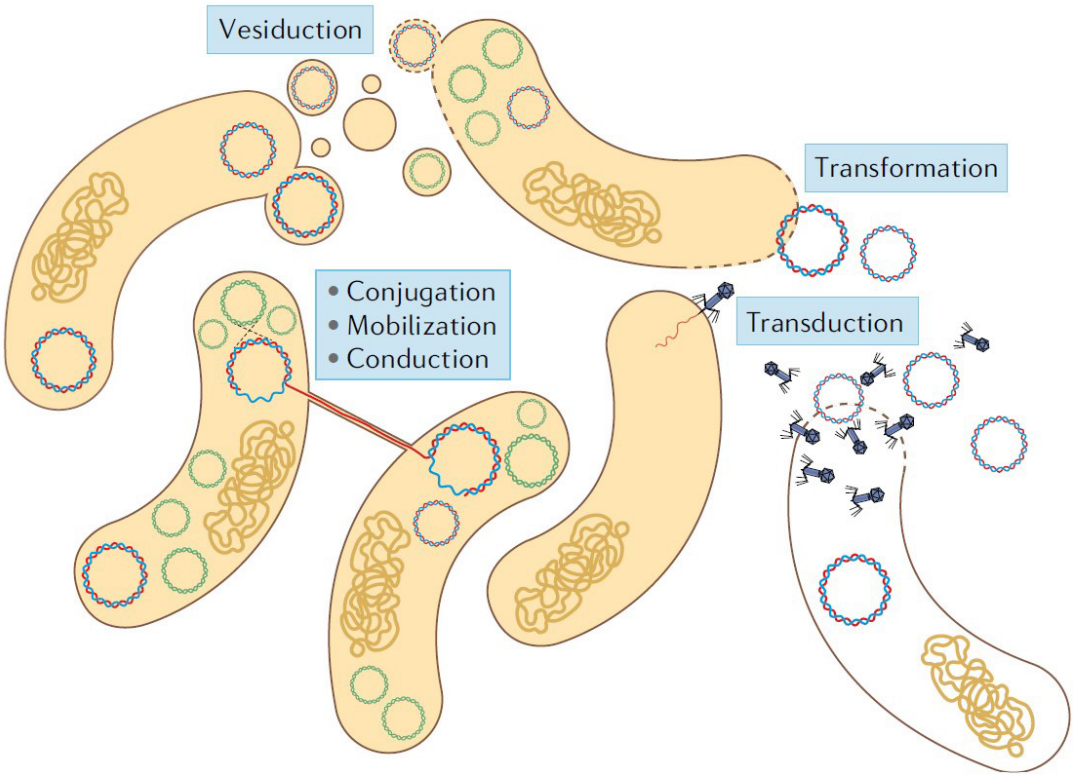


Figure 6. Scheme depicting the different possible mechanisms for DNA to enter in a cell (Rodríguez-Beltrán et al. 2021)

The exchange of DNA between a cell and its environment is the basis of horizontal gene transfer (HGT). Four types of intakes are described[53] transformation, conjugation, infection/transduction and vesiduction (Figure 6).

First, transformation is the uptake of free DNA from the environment. Free DNA outside of the cell can originate from death and lysis of another cell or the active secretion of DNA from another cell[54]. This mechanism is widely used in laboratory to import a recombinant DNA in cells (bacteria, yeasts, plants, cells in cultures). It is also a natural process described for bacteria[54] and yeast[55,56]. The mechanism is not well understood, at least DNA has to be in contact with plasma membrane or cell wall and will be later imported in the cell[54,56].

Conjugation [57,58] is the exchange of DNA between a donor bacterium which will produce a pili to contact recipient cell and inject DNA through the tunnel structured by the pili with the help of a type IV secretion system. Conjugation machinery comes from plasmids (thus called conjugative plasmids) that brings ORFs coding for all the components necessary for the conjugation. Interestingly there are also plasmids that are mobilizable by conjugation but not conjugative. Such plasmids can pass through the channel created by conjugative plasmid but do not generate the tunnel by themselves. Conjugation is thought to be the most important way of HGT since it does not require much phylogenetic homology between donor and recipient[59]. Conjugation between *E. coli* and *S. cerevisiae* was reported in 1989[60] and since then in a wider range of eukaryotes[61].

Infection is the "penetration in a living organism of a foreign entity often pathogenic able to reproduce in its host"[62], the term "entity" is used to regroup living pathogens (such as bacteria, fungi, protozoa …) and non-living ones such as virus. Here we will focus on infection by virus since other cases are not directly related to nuclear ecDNA. Infection is associated with transduction especially for prophages. The transduction is the exchange of genetic material between two organisms mediated by a viral vector. Many biotechnology technologies rely on transduction such as gene therapy. Transduction occurs in natural conditions when prophages are excised from there host genome in a non-precise manner. Thus, along with the phage genome part of host genome is incorporated in the capsid.

Recently it has been reported that free vesicles can also drive DNA exchange by vesiduction. The term "extra-cellular vesicle" refers to a vast population of observed vesicles that were mainly studied as a communication channel between cells from an organism. Among the molecules carried by those vesicles, DNA can be transported. Leading to the exchange of virulence factors and other genomic content for prokaryotes, unicellular eukaryotes[63] and multicellular eukaryotes[64].

For eukaryotes, ecDNA which does not bring their own replication/transcription machinery requires access to the nucleus to insure their own multiplication and propagation. Eukaryotic viruses can enter the nuclear compartment through nuclear pore complexes[65]. Another way to enter the nucleus is to be recognised by cytoplasmic proteins with a nuclear localisation signal (NLS) and then transferred in the nucleus via the importin pathway[66]. Interestingly, it has been recently reported that in mammalian cells exogenous DNA is often segregated in a particular compartment whose membrane derivate from the endoplasmic reticulum called the exclusome[67]

1.1.2.2 Endogenous DNA circularisation: extra chromosomal circular DNA (eccDNA)

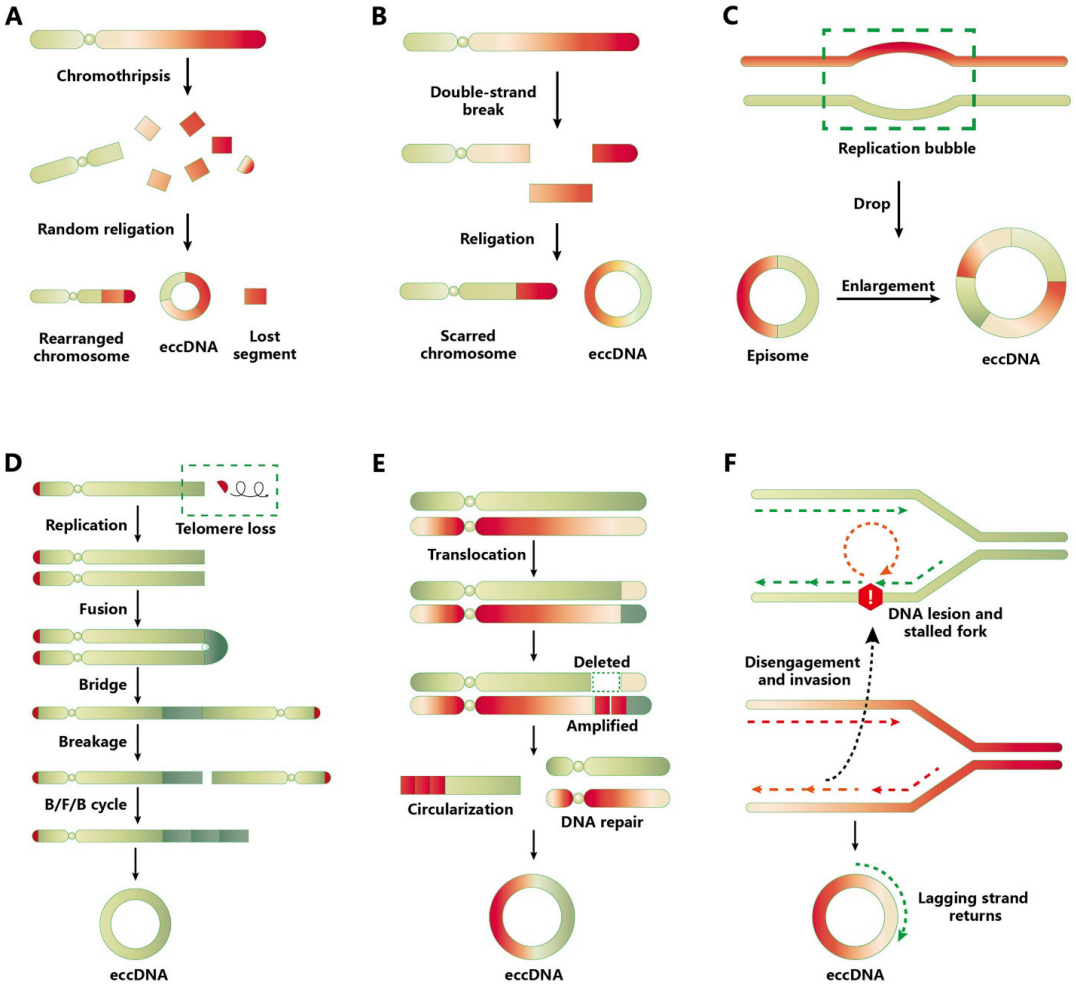


Figure 7. Drawing depicting the different ways for endogenous DNA circularisation and biogenesis of eccDNA in eukaryotes (Zhao et al. 2022)

The first report of eccDNA was made in 1965[68] following the analysis of DNA content from wheat embryos and boar sperm cells. They have been described in a wide variety of organisms from microorganisms to humans with different biological importance[69].

The origin of eccDNA (Figure 7) in the cell remain unclear but many models and hypothesis are proposed, such as chromothripsis, DNA damage repair and abnormal DNA replication[70]. Intrinsic (aging, cancer transformation) and extrinsic (starving, hypoxia …) factors promote eccDNA. It is shown that the eccDNA are common in cancers and normal cells.

ecDNA is an extra DNA in the nucleus. It often comes with a cost and is targeted by host cell defence systems. The stability of an ecDNA relies in its ability to overcome challenges in its host.

1.1.3 Challenges faced by ecDNA in its host.

1.1.3.1 Extra chromosomal DNA comes with a cost and is often a threat.

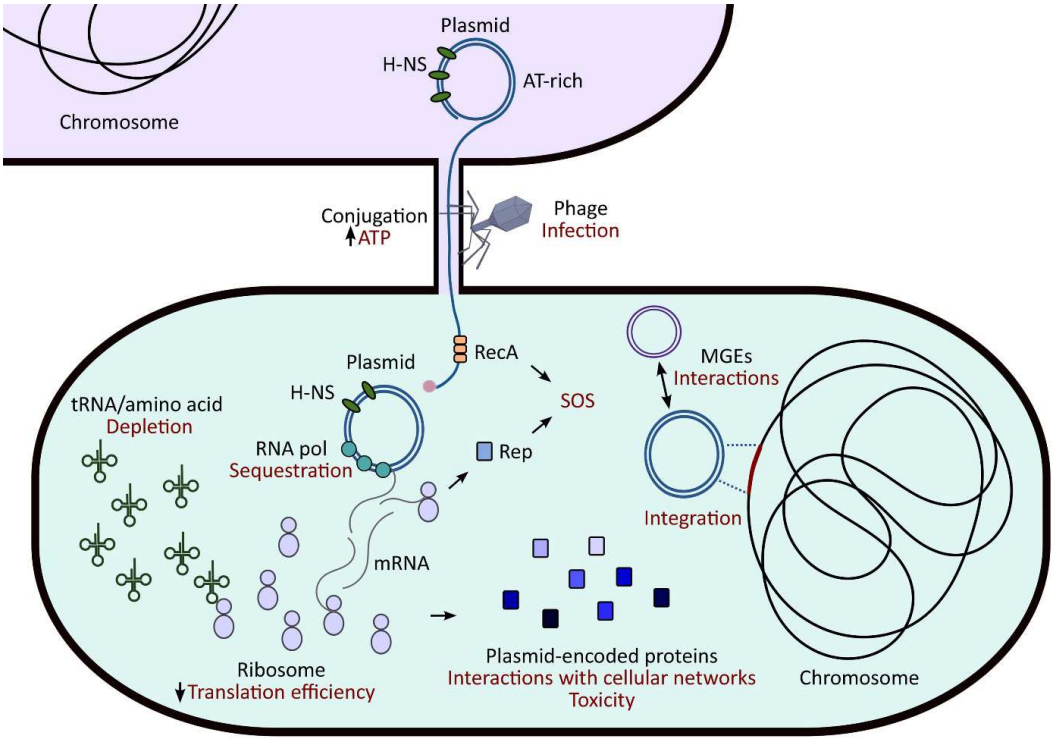


Figure 8. Scheme depicting the different source of fitness cost induced in a host by ecDNA (San Millan and McLean 2017)

ecDNA represents an extra genetic material that relies on host cells for the energy and components needed for its cycle (DNA replication, ORFs expression, capsid synthesis for virus …). This cost for the host cell is called the fitness cost (Figure 8) and regroups all the negative effects induced by ecDNA on the host cell. The fitness cost depends on the considered ecDNA/host couple and can range from minor growth defect to the cell death. The main explanation for the growth defect is the difference of codon usage between host genome and ecDNA genome[59], creating a disequilibrium in tRNA pool and reducing the overall translation speed. The death of the host can be induced by host defence (discussed below) or ecDNA actively kill its host as for lytic bacteriophages. Thus, new DNA coming in the cell is often targeted by many host factors that aims to disrupt ecDNA replication cycle in the host cell and its horizontal transfer.

1.1.3.2 Host defence and its modulation.

1.1.3.2.1 Detection of foreign DNA and intermediates.

For vertebrates, free cytosolic DNA activate many sensors such as “absent in melanoma 2” (AIM2) and more predominantly cyclic GMP–AMP synthase (cGAS)[71]. cGAS binds to DNA

in a sequence independent manner. The activation of cGAS leads to the synthesis of cyclic GMP-AMP (cGAMP) which acts as a secondary messenger. cGAMP activates STING which will activate NFκB pathway, interferon/cytokine signalling and autophagy. These effects will limit the spread of the threat. Whereas in prokaryotes, genomic DNA is not separated by an envelope thus prokaryotic-defence systems must be able to discriminate between self and non-self-DNA. It has been described that Histone-like Nucleoid-Structuring (H-NS), a nucleoid associated protein (NAP)[72], binds preferentially to foreign DNA because they are often richer in AT compared to host genome. This later activates defence strategies by creating a depletion of H-NS protein from the host genome[73]. Moreover, Restriction-Modification systems (RM)[74] distinguishes between host DNA and foreign DNA thanks to host DNA modification. Host DNA has a specific methylation pattern. Foreign DNA that does not match the host modification pattern is restricted. Finally, small RNAs are part of the detection mechanism in both prokaryotes and eukaryotes to recognise invasive DNA and its transcriptional intermediates. In prokaryotes, Clustered regularly interspaced short palindromic repeats/CRISPR associated protein[75] (CRISPR/Cas) system recognise foreign DNA thanks to small guide RNA. Those guide RNA are generated by the transcription of peculiar genomic loci which function is to

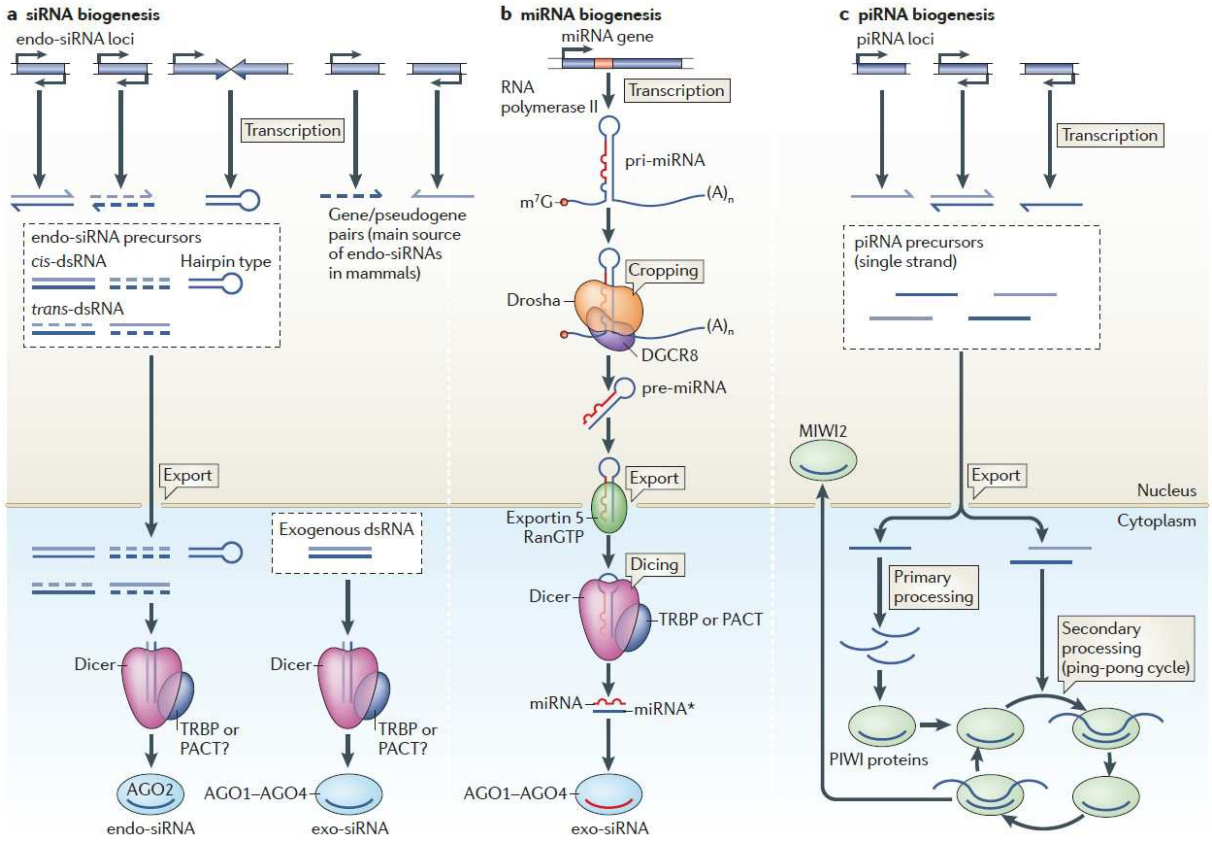


Figure 9. Scheme depicting the different RNA interference pathways (siomi et al. 2011)

store DNA fragments originating from past infection event. In eukaryotes (but not in *S. cerevisiae)* invading DNA transcripts are detected by small noncoding RNAs. They are divided in three classes[76] (Figure 9). The small interfering RNA (siRNA) that comes from dsRNA processed by the Dicer complex. The dsRNA can originate from many sources, from retro-transposable elements or pathogens. The siRNA is then loaded on the RNA induced silencing complex (RISC) guiding it towards its target. The Micro RNA (miRNA) which originate from miRNA genes. Those genes transcript forms a hairpin structure recognised by Dicer and activate RISC. Finally, PIWI interacting RNA (piRNA) originating from piRNA clusters in the genome that binds PIWI protein. PIWI and RISC both silences RNA that has a homology to the loaded siRNA.

1.1.3.2.2 Host defence strategies

1.1.3.2.2.1 Destruction of foreign DNA

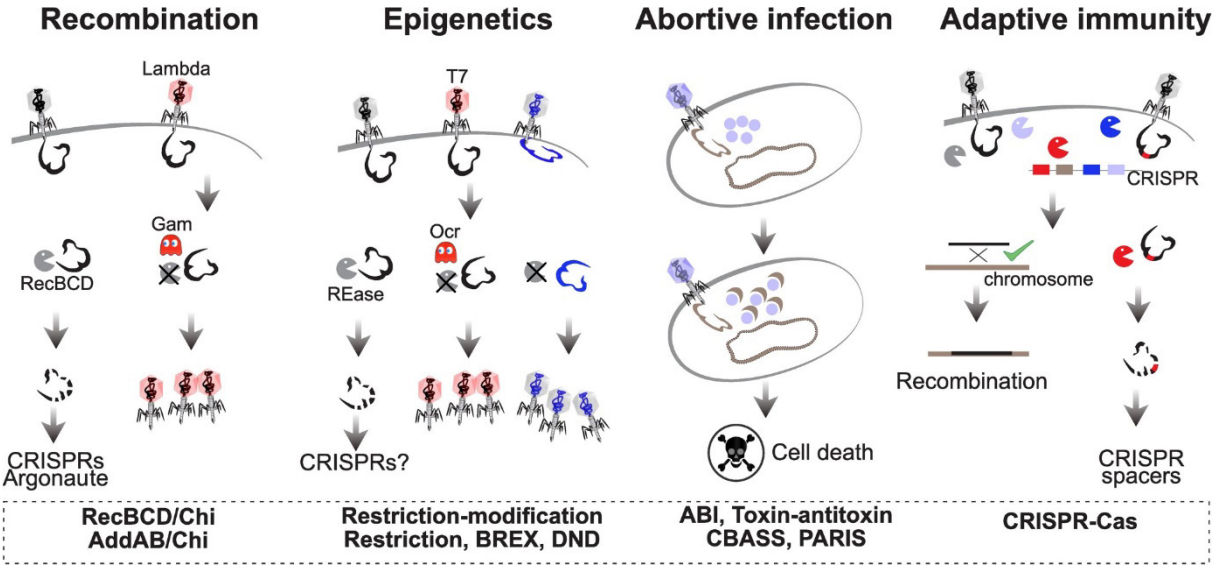


Figure 10. The different known defence mechanism against MGEs in a bacterium (Rocha and Bikard, 2022)

The most straight forward strategy of defence against an invading DNA element is its destruction. This can be achieved either in a sequence dependant or independent manner. In eukaryotes, cytoplasmic DNAse hydrolyse free dsDNA. In prokaryotes (Figure 10), RecBCD degrades linear dsDNA until it meets a Chi sites. Chi sites are found in host and are entry point for RecA. If linear DNA does not have Chi site, it will be totally degraded[77]. The produced small DNA fragments can be archived in the genome and later feed CRISPR/Cas systems. RISC[78] and CRISPR/Cas[79] systems guides endonuclease to foreign RNA/DNA respectively. If a genetic material has a homology with the nucleic acid loaded in the nuclease, then the enzyme will produce a nick in the foreign genetic material. Finally, in prokaryotes, restriction enzymes (RE) are sequence specific endonuclease that degrades DNA when they recognise a DNA motif.

They also are sensitive to DNA modification such as methylation. As discussed above, the host protected from the action of its own RM defence system.

1.1.3.2.2.2 Silencing of foreign DNA

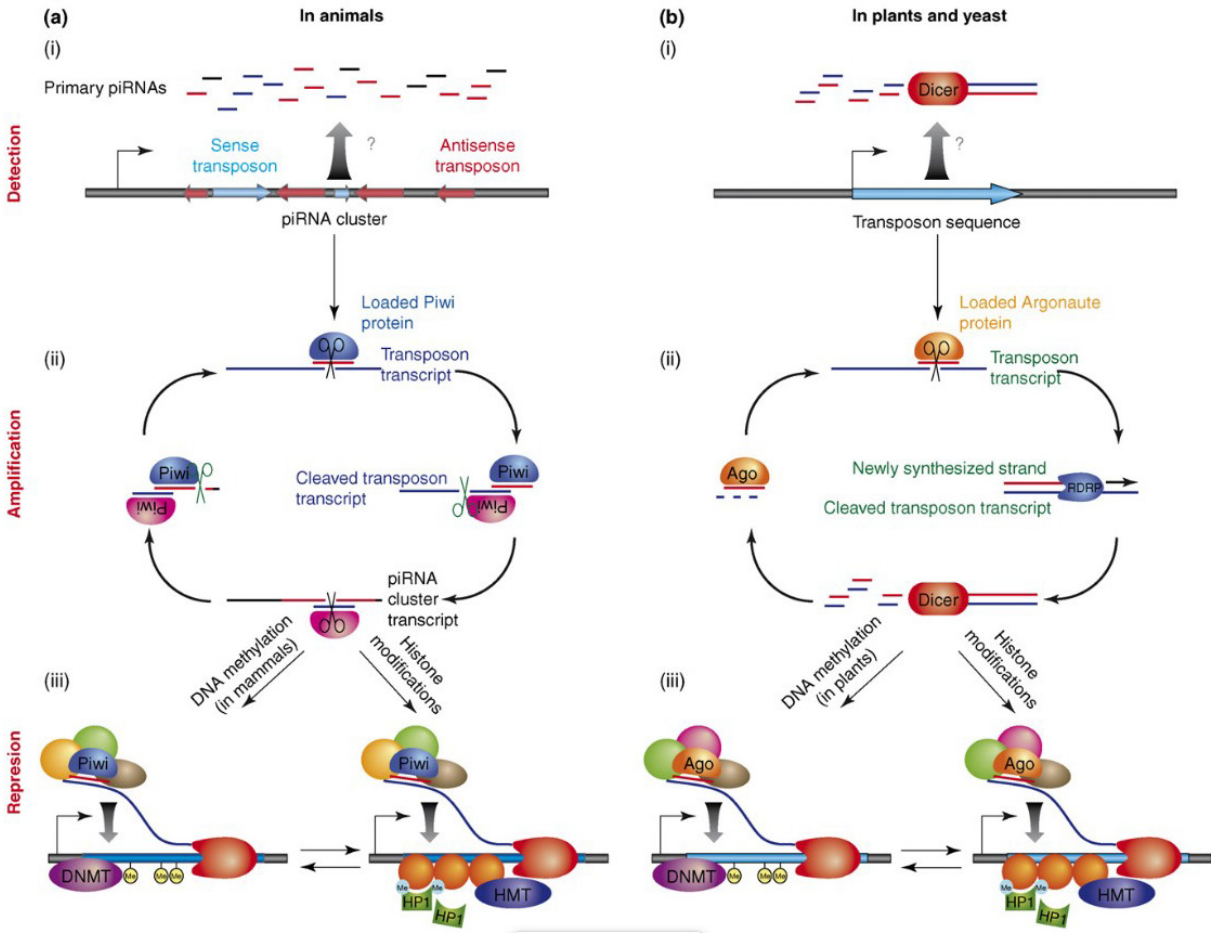


Figure 11. Scheme showing the silencing pathways of RNA interference ion eukaryotes (Girard and Hannon,2008)

Another way to inhibit armful effect of invading DNA is to supress its expression. This is especially the case for transposable elements in eukaryotes. The dsRNA intermediates of retro-transposable elements will activate RISC[80] (discussed above) which will degrade targeted mRNAs but also promote DNA methylation or histone modification (Figure 11). Those epigenetic modification will decrease the level of expression of invading DNA locus. In prokaryotes, MGE genes are under a strict genetic expression control called the xenogeneic silencing notably by H-NS[59].

1.1.3.2.2.3 Host suicide

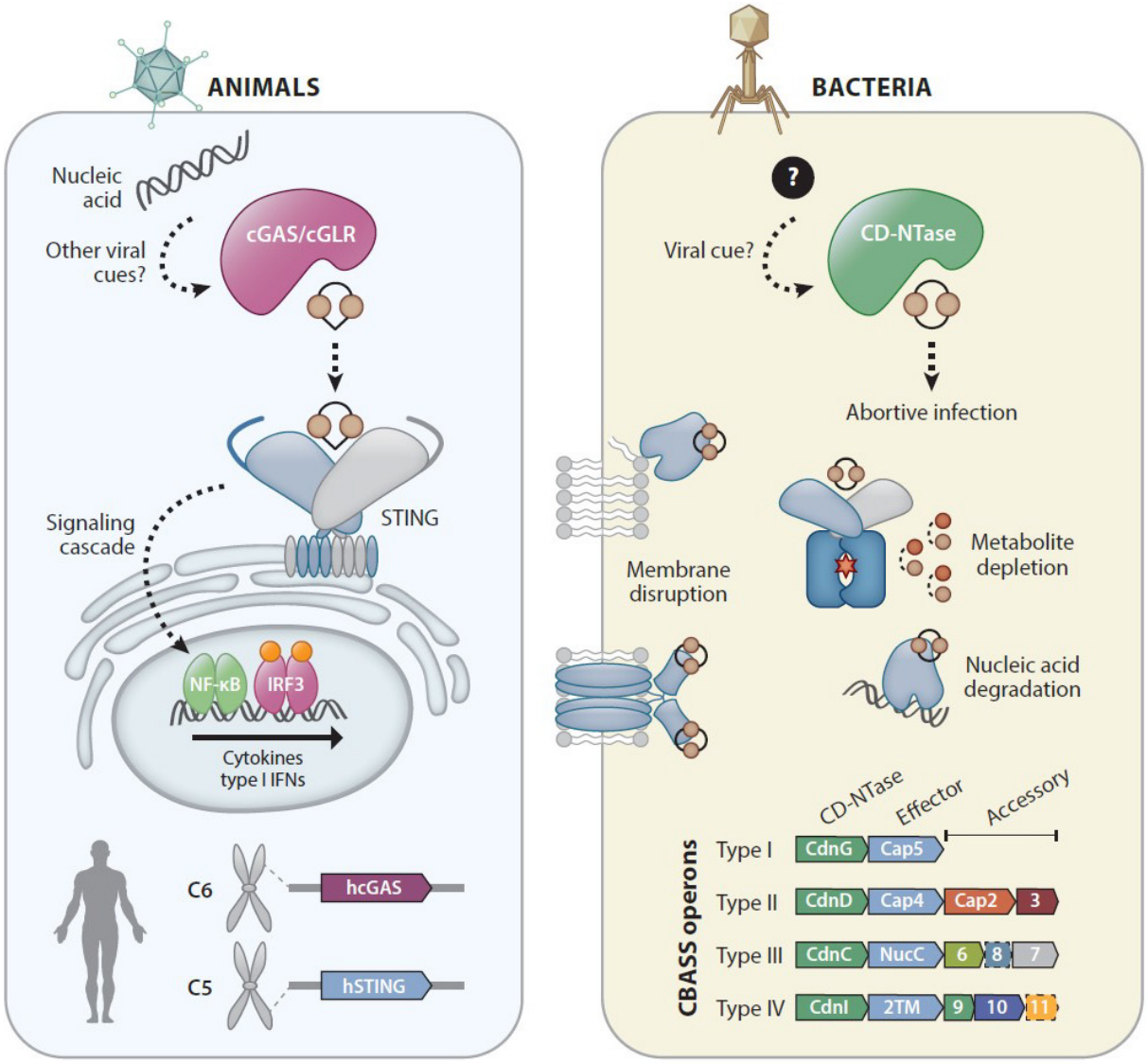


Figure 12. Cyclic di-nucleotides pathway is present in both prokaryotes and vertebrates. (Slavik and Kranzsuch 2023)

As a last resort, host cell can activate cellular death pathway. This is reported in both eukaryotes and prokaryotes[81] (Figure 12) and known as abortive infection. The abortive infection will allow to contain ecDNA thus not allowing its propagation to other cells both vertically and horizontally. Many pathways of abortive infection have been reported. The common feature is that certain infection will be recognised by host factor and will trigger cellular death. Interestingly, in both prokaryotes[82] and vertebrates[71], cyclic di-nucleotides (such as cGAMP) can lead to abortive infection. In prokaryotes, the oligonucleotides cyclase is activated by phage component whereas in vertebrates it is activated by cytoplasmic DNA. In vertebrates the activation of STING by cGAMP also activate innate immune response thought interferon and NFκB pathways.

1.1.3.2.2.4 Immune response

Here we will focus on the immune response from a pluricellular organism, we already discussed above the bacterial "immune system" with the CRISPR/Cas system. The immune system in pluricellular organism protects the organism from infections. We can distinguish two types of immunity. The innate immunity which relies on Pattern Recognition Receptors (PRR) which recognises Pathogen-Associated Molecular Pattern (PAMPs). The innate response depends on the type on activated PRR but is not specific to the threat. The adaptative immunity relies on different cells that will process molecules coming from the pathogens. The processed pathogens, named epitopes are presented to other immune cells via the CMH complex (CMH I if it is an intracellular epitope, CMH II if it is an extracellular epitope). The presentation of an epitope will guide the immune response towards the epitope presented. Overall, the immune system aims to eradicate the infectious agent. This is done by reducing its propagation from cell to cell and by killing the cells hosting the pathogens.

My aim here is not to thoroughly described the whole immune system and focus more on ecDNA. As seen above, the DNA in the cytoplasm is a PAMP recognised by a PRR, cGAS. Beside the cellular effect of the activation of STING by cGAMP (autophagy and apoptosis). The host cell will also present more epitopes with its CMH I complex and promote local inflammation via the secretion of interferons and cytokines. The inflammation allows the migration of immune cells from the blood stream toward the inflamed area.

Overall, we have shown that all organisms have defence (at different level) systems against ecDNA invasion. The efficiency of an infection often relies on the overall ability of ecDNA to shrug host defence systems.

1.1.3.2.3 MGEs counter host defence systems.

1.1.3.2.3.1 Avoid detection by host defence.

If the host detection systems are not activated then the host defence strategy is much more permissive, thus only relying on detection-independent defence systems such as RecBCD in prokaryotes. In prokaryotes, MGE's encoded H-NS are often described as stealth proteins[83,84]. As discussed above, host H-NS binds preferentially foreign DNA. This titrates H-NS from host genome and activate host defence mechanism[59]. This also decreases host fitness because H-NS action on host genome is essential. By bringing their own H-NS, MGEs support their host and avoid the activation of defence mechanism. In vertebrates, cGAS is a target of

choice for pathogenic virus[85]. Moreover, virus often keeps their capsids in the cell. This prevents their genetic material to be exposed in the host cytosol, thus preventing its detection and degradation by host factors. The viral genetic material will only be freed in the nucleoplasm.

1.1.3.2.3.2 Avoid the activation of effectors and/or inhibit them.

The layer above the detection system is the effectors RecBCD, restriction enzyme or the RISC complex for instance. In prokaryotes, Gam is an inhibitor of RecBCD encoded by virulent phages[86]. Virus can also target Dicer complex by protecting the extremities of dsRNA intermediate with proteins inhibiting Dicer[87].

1.1.3.2.3.3 Defence mechanisms are ineffective.

The action of defence strategies relies on the action of enzymes, and enzymes substrate must meet requirement for the enzyme to work. For instance, the RM prokaryotic defence systems is often counter-acted by phages because they are able to extensively modify their genome[88] (methylation for T2 phage or acetimidation for the Mu phage) or even incorporate non canonical bases in their genomes. Thymine can be replaced by uracil or 5-hydroxymethyluracil in certain phages. Those modifications totally inhibit the action of restriction enzyme. Moreover, for the RecBCD defence systems, certain phages have multiple Chi sites in their genome thus are insensitive to RecBCD action[89]. Finally, if the host organism is naïve toward the pathogen, adaptative defence strategies such as CRISPR/Cas or miRNA/piRNA will not work.

1.1.3.2.3.4 Host anti-counter defence mechanism.

More recently, it has been reported that the host defence is not only constituted by a unique line of defence[90]. More and more anti-counter defence systems are described. For instance, RecBCD constitute a first layer of defence. RecBCD can be inactivated by Ocr. The activity of RecBCD can be monitored by retrons in the bacteria[91]. A retron is an RNA-DNA hybrid able to sense the activity of RecBCD in the cell. When the retron is activated, it triggers abortive infection mechanism and host suicide.

1.1.3.2.4 Host select and filters MGEs.

In eukaryotes, the defence mechanism is restrictive towards ecDNA. In prokaryotes, on the other hand, MGEs often brings important accessories function that improves its fitness (discussed below). Thus, bacterial host defence intensity can depend on the nature of the MGEs. The fact that certain MGEs escape bacterial host defence can be seen from two different points of view. First, as discussed above, it is part of the arm race between host and MGE. Second, MGE is tolerated by the host. For instance, MGE matching the epigenetics modification of the host genome are insensitive to RM defence systems and may be the results from the co-evolution between the MGE and their host. Moreover, Baharoglu et al.[92] have shown that the level of activation of the SOS response in the bacterial host, is different between different conjugative plasmids. The one that belongs to a family of conjugative plasmids often found in *E. coli* less activate the SOS response compared to non-naturally present conjugative plasmids.

1.1.3.3 Co-habitation between different MGEs in the same host.

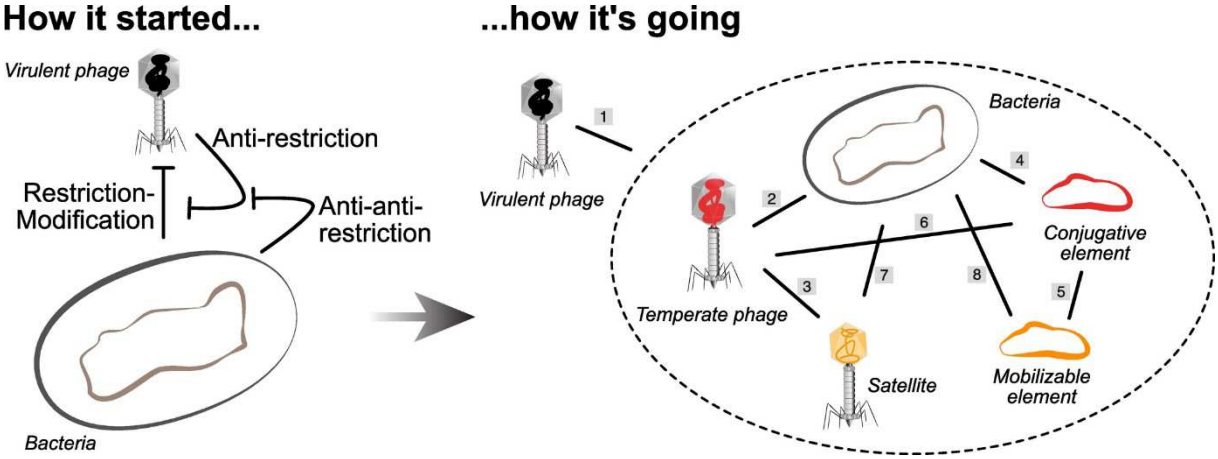


Figure 13. Scheme showing the evolution of known interactions between host and MGEs. Starting from a simple view that a virulent phage and host cell fight with each other but, all MGEs interact with each other and their host cell (Rocha and Bikard, 2022)

Some ecDNA rely on their host and on other ecDNA for their replication and transfer. This is particularly the case in prokaryote. This is a very dynamic field of research, more and more interactions pathways between MGEs and their host are described[90] (Figure 13). For instance, a mobilizable plasmid will benefit from the conjugation machinery of a conjugative plasmid because it cannot transfer itself by conjugation. Whereas a virulent phage will be deleterious for other MGE's since the phage lytic cycle will kill their host. In this case, Rousset et al.[93] showed that anti-virulent phage systems are encoded on satellite phages protecting the

host cell from virulent phage infection. Thus, satellite phage and its host benefit from each other (if the threatening virulent phage is a threat).

The host of an ecDNA is probably not its last host, otherwise it will be lost when its host eventually dies. Thus, ecDNA must transfer to another organism (horizontal transfer) or to the progeny of its host (vertical transfer).

1.1.3.4 Transfer of ecDNA

1.1.3.4.1 Horizontal transfer

The mean of horizontal transfer is the same one discussed above, about the exogenous source of ecDNA. Not every ecDNA is able to transfer from one organism to another, it depends on the genetic material present on an ecDNA.

1.1.3.4.2 Vertical transfer

The chromosomal DNA of the host is transmitted to the progeny thanks to active mechanism ensuring an equal repartition of the duplicated genetic information to both daughter cells. This is the basis of genome integrity and inheritance. Whereas segregation of ecDNA could be envisioned as a facultative event since it is not essential for the host cell. The simplest mechanism would be the diffusion of ecDNA to the progeny, but this has been shown to be tedious.

1.1.3.4.2.1 A not so simple diffusion.

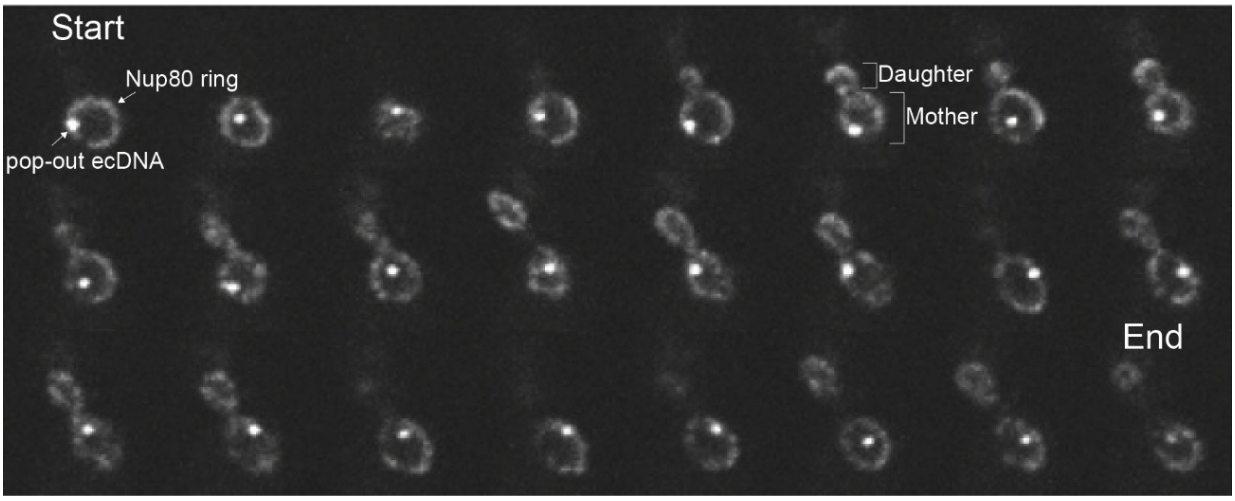


Figure 14. Result adapted from Gehlen et al. 2011 showing how episomal DNA is retained in the mother cell throughout the mitosis in S. cerevisiae.

All chemical entity is animated by Brownian movement in a solution and DNA is not an exception. Diffusion is a passive way for ecDNA to be partitioned to daughter and mother cell and it is thought to be the strategy employed by high copy number plasmids in prokaryotes[94]. This hypothesis is more and more challenged because plasmids tend to form foci in living cells and the viscosity of cytoplasm doesn't allow DNA molecules to diffuse freely in cytoplasm of both prokaryotes and eukaryotes[95,96]. Moreover, it has been reported in *S. cerevisiae* that diffusion of DNA from mother to daughter can be impound by diffusion barrier. DNA molecules lacking a centromere are captured by Nuclear Pore Complex (NPC) and stay in the mother cell[97,98], this is called the mother-biased segregation.

1.1.3.4.2.2 Make your own way to the progeny.

The segregation of plasmid in prokaryotes can be divided in two strategies[99]. The most common strategy uses a nonspecific DNA walker using NTP hydrolysis energy to walk along the host DNA towards the cell poles. The second known strategy is to use the motion force of ATP/GTP dependant polymerisation of actin[100] or tubulin analogues towards the cell poles. Finally, a mechanism which does not use walker protein has been reported. It is based on the repartition of plasmid copies in the whole nucleus. Upon septation, plasmid copies will end up in both daughter cells. This mechanism was described for a single plasmid, R388[101] but could be more abundant. For mammalian persistent DNA viruses[102] (EBV and KSHV for instance), their strategy is to hitchhike as episomes on host chromosomes with the help of latency proteins bridging viral DNA and host DNA. Thus, benefiting from mitotic host chromosome motion to segregate to daughter cells. This is thought to be the strategy employed by the 2µ plasmid of *S. cerevisiae*[97].

Finally, if ecDNA is transiently integrated in host genome, then the segregation question is solved, as the host machinery will insure it. For instance, the adeno associated virus (AAV) can either be integrated in the genome or found as circular episome in host cell[103]. Many phages can integrate in the host genome, entering a dormant stage. Upon reactivation, the phage genome will be excised often with flanking host genomic regions and enter a productive cycle.

1.1.3.4.2.3 Be the poison and the cure.

As discussed above, the issue for the ecDNA vertical transfer is its optional presence in the progeny. If its presence is required for its host survival, then its vertical transfer is insured.

Certain plasmids encode for toxin/antitoxin systems. One well described system is the hok/sok pair encoded by the parB locus of plasmid R1[104]. The hok ORF encode for a small peptide that disrupt the electro chemical gradient of the bacteria thus impeding the production of energy in the cell. The expression of this gene is silenced by the interfering RNA encoded by the sok locus. The hok peptide is stable thus daughter cells will obligatorily inherit the killing peptide, but the silencing RNA is unstable and the only way to survive is to incorporate the whole plasmid. Since R1 is a conjugative plasmid, it can transfer from cell to cell and the hok/sok ensure its stability in its new host.

This kind of poison/cure persistence strategy is also described in yeasts. For instance, in *L. lactis*[105] , two linear double stranded DNA molecules (k1 and k2) located in the cytoplasm. Toxins encoded from the extra chromosomal genome are secreted in the environment, killing organism lacking k1 and k2. This strategy imposes a burden on the cell but improves host ability to colonize a biotope. The secretion of toxic compound giving the ability to eliminate other competing cells [106]. It is worth mentioning that most anti-fungal molecules are secreted killer-peptide[107]. This example and the protection of host cells from virulent phage infection by MGEs discussed above clearly show the potential benefit of ecDNA for the host cell. Thus, ecDNA can be a welcomed help for its host.

1.1.4 Extra chromosomal DNA can improve host fitness and adaptation ability.

1.1.4.1 Relationship between extra chromosomal DNA and host.

When two organisms share the same resource within a community, biological interactions are established. Whether this association benefit, is neutral or harm an organism, several terms are used. Extending this concept to the community formed by the nuclear genome and extra-chromosomal DNA makes it possible to characterise the relationship between them. Clearly, cDNA benefits from its interaction with the nuclear genome, since the host cell provides the energy and components necessary for its replication and expression. However, the consequences for the host cell depend on the ecDNA. The host cell can be harmed (for instance when a lytic phage infects a bacteria) (parasitism), is unaffected by the presence of an extra chromosomal DNA (This case occurs for bacteria hosting tempered phage for instance), and finally benefit from extra chromosomal genetic material improving its fitness (mutualism). The biological interaction between nuclear genome and ecDNA is not written in stone. Upon

environmental changes, a commensal relationship could become parasitism or commensalism. Moreover, genomes are constantly evolving so are the interaction between two genomes.

1.1.4.2 ecDNA can bring new functions.

Since ecDNA is often mobile and can be transferred to an organism from a different species, bringing its genetic material. The fact that ecDNA brings new traits to its host is particularly described for multi-drug resistance bacteria[108,109]. Prokaryotic MGEs are known to encode for drug resistance cassettes. When ecDNA has an advantage for its host, the fitness cost induced by the ecDNA is compensated by the benefits. Thus, true commensal relationship can be established between ecDNA and its host. Moreover, ecDNA stability and integrity is less critical for the host cell. This flexibility can be a true advantage for the host to adapt and overcome environmental stress[110].

1.1.4.3 ecDNA is constantly evolving and reshuffled.

The characterisation of ecDNA pool gives intel on how an organism overcomes environmental stress[111]. For instance, in eukaryotes, eccDNA can be associated with gene amplification. Gene amplification is the multiplication of a genetic cassette in the genome of an organism. It can happen within the chromosomal set with duplication events[112] or with eccDNA that is present is multiple copy. The eccDNA gene amplification is for instance associated with chemotherapy resistant tumors[113] or crop weed herbicide resistance[114]. Moreover, in prokaryotes the MGEs pool is very dynamic and cassette are exchanged between MGEs[40,92,115] and between MGEs and host genome. Finally, since ecDNA integrity is not as critic as the one from the core genome. It is a pool of polyploidy with genes coding for the same trait but each one following different evolutionary track. This can eventually lead to genetic innovation[53] and improve overall genetic plasticity.

1.1.4.4 ecDNA can increase host genome plasticity.

We discussed above about the possible invasion of genome by retrotransposons; however, they can be handy when it comes to adaptive ability and stress response of their host. Many reports show that the level of transcription and integration of Transposons of yeast 1 (Ty1), the major retro-transposable element in *S. cerevisiae* greatly increases in response to

environmental stress such as starvation[116]. Many reports showed that Ty1 are often fragile sites and promote genomic recombination such as duplication, inversions, and translocations. The genomic rearrangement mediated by Ty1 are the key component of the rapid adaptability of *S. cerevisiae* to various environmental stress[112,117,118] such as nutrient deprivation or exposure to toxic compounds.

We could assume the relationship between ecDNA, and its host can last for a long time. Over time, ecDNA and chromosomal genome will become more and more adapted to each other, and genes will be exchanged between them[119], making it more and more difficult to distinguish between each other.

1.1.5 Blurring the line between extra-chromosomal DNA and genomic DNA

Genes that bring the same function in the cell often impose an unnecessary cost to the host cell and events leading to the loss of redundant genes are selected by natural selection since it would improve host fitness. This phenomenon is seen with obligatory intra-cellular bacteria in insects[120] and is thought to have happened in Vibrionaceae family[121].

In most cases, the bacterial genome consists of one chromosome coding for all essential functions needed for the life cycle of the cell. But in a small proportion of known prokaryotes, the genome is separated in multiple (often) circular chromosomes[122]. The *Vibrio cholerae* genome is the case study of bacteria with multiple chromosome and upon the report that its genome was composed of two circular chromosomes in 1990's[123] debate arose whether to call the secondary chromosome a chromosome or merely a very large plasmid. But with the help of genome sequencing[121] and further analysis[124] the chromosome appellation stands. Interestingly, it is thought that the secondary chromosome of Vibrionaceae family come from a plasmid that upon co-evolution and gene transfer became part of the core genome of its host In 2010 the term chromid (**chro**mosome and plas**mid** at the same time)[125] was proposed to class the secondary chromosomes of Vibrionaceae bacteria

The differentiation between ecDNA and genomic DNA sometimes can be difficult if we strictly apply classification criteria moreover, we are able to study a limited part of the tree of life (given we are capable to characterize what is cultivable) without knowing exactly all the evolutionary steps that led to our study subject. One of many interesting ecDNA/host genome couple is the persistence strategies employed by the 2µm plasmid which does not improve its host fitness and yet have a chromosome-like stability.

1.2 The 2µ plasmid, a perfect parasite?

1.2.1 The 2µ plasmid relationship with its host

1.2.1.1 *S. cerevisiae* defence mechanism against invading ecDNA.

As mentioned above, *S. cerevisiae* have lost the RNA interference pathways. Whereas other eukaryotes[126], "heterochromatin" in *S. cerevisiae* is limited to HML/HMR locus on chromosome III and telomeric regions. The inactivation of chromatin uses a peculiar mechanism independent from RNA interference mechanism. Abf1, Rap1 and Orc1 binds silencer sequences. They allow the recruitment of Sir2/Sir4 that spreads on the chromatin and silence it.

However, *S. cerevisiae* is not invaded by retrotransposons. Only 3% of *S. cerevisiae* genome is identified as retro-transposable elements (50% in Human genome and 80% in maize genome)[80] but the Ty1 RNA can represent nearly 1% of total RNAs[127] and up to 10% on total mRNAs[118]. Even if there is no control via the RNA interference system as in most eukaryotes, the retro-transposable elements are under a strict control. In *S. cerevisiae* the nuclear envelope remains intact all along the vegetative growth and meiosis. This alone is a protection against invading genetic material since the only way to enter the nuclear comportment is througth the Nuclear Pore Complex (NPC). Thus, the NPC filters what comes in or out of the nucleus. Moreover, some host factors associate with the NPC and regulate act on the genome organisation and regulate the transcription of retrotransposons. Bonnet et al. showed that a subunit of the NPC regulate retro-transposons transcription and propagation by Ulp1 mediated deSUMOylation[128]. Moreover, as discussed above, NPC retains non centromeric DNA circles in the aging mother cell[50,98]. A recent report from our laboratory[129] has shown that when a 17th chromosome coming either from *Mycoplasma mycoides* or *Mycoplasma pneumoniae* is introduced in *S. cerevisiae* its fate is different. The *M. pneumoniae* genome which GC content is close to the one of *S. cerevisiae* is transcribed and reparteed in the nucleus whereas the *M. mycoides* genome which is AT rich compared to *S. cerevisiae* genome is not transcribed and constitutes a dense DNA blob at the periphery of the nucleus. This looks like the xenogeneic silencing discussed above where H-NS proteins move towards incoming ecDNA. The defence strategies of *S. cerevisiae* against invading DNA are most certainly yet to be discovered and interestingly, the 2µ relationship with its host is quite ancient and it was able to counter defence strategies against invading DNA.

1.2.1.2 2µ plasmid evolutive origin

As discussed above, the 2µ plasmid was discovered in 1967[25] and was thought to be a mitochondrial plasmid because of its non mendelian inheritance. We will discuss below about the structure and functions of the 2µ plasmid and focus more precisely on evolution considerations.

One could wonder if the interaction of the 2µ plasmid with its host *S. cerevisiae* was recent or only restricted to some laboratory strains[34]. In fact, the sequencing of 1,011 genomes[130] revealed that a majority (64%) of the sequenced strains, coming either from laboratory, industrial processes or natural isolates, hosted 2µ plasmid (this number most likely does not reflect the incidence of 2µ in *S. cerevisiae* since many strains derivate from each other). Interestingly, we can identify 4 groups of the 2 µ plasmid based on their sequence divergence. This indicate that the interaction between 2µ started one hundred million years ago.

Along with the 2µ plasmid, several other plasmids were identified in species related to *S cerevisiae*[44]. Interestingly, the 2µ plasmid homologs were found in *L. lactis* and *L. waltii*, two yeast that diverged with the *S. cerevisiae* lineage approximately 114 My ago, before the whole genome duplication event that took place 100 million years ago. Given the rarity of (known) natural, nuclear plasmid we could hypothesize that the meet-up between the 2µ plasmid ancestor and its host happened at least 114 million years ago then evolution acted and generated several taxa (such as *saccharomyces* and *Lachancea*) and different nuclear plasmids. When we compare the GC content of the different 2µ plasmid family and the GC content of their respective host we observe a nice corelation[44] (Figure 15), most likely the results of compensatory mutations that lowered fitness costs of the plasmid.

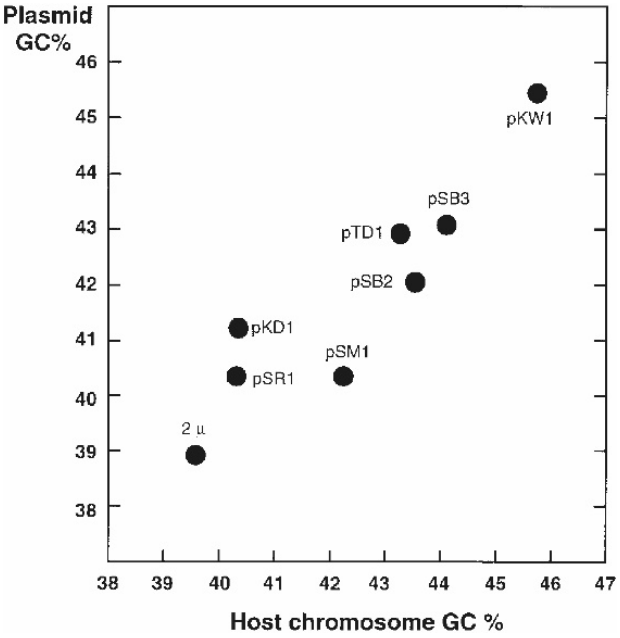


Figure 15. Scatter plot of the correlation between host and parasitic plasmid GC content

All nuclear plasmid in the 2µ plasmid family shares common genetic features. Among them, the Flp1 protein is a site-specific recombinase and part of the Tyrosine-recombinase family. Interestingly, it shares structural homology with P1 phage Cre recombinase[131] and a common catalytic mechanism. Given that conjugation between bacteria and *S. cerevisiae* is

possible[132,133] and certain metabolic activity in *S. cerevisiae* are inherited from prokaryotes[134]. We could hypotheses that 2µ plasmid is in fact a derivative of a phage genome that have been mobilized in its ancestral host, common to *Lachancea* and *Saccharomyces*, thanks to a conjugative plasmids that have been lost as it is the case in laboratory conditions[60]. This doesn't mean that we should find nuclear plasmid in every species of taxon sharing the same common ancestor[45] (namely, *Zygosaccharomyces*, *Candida* and *Eremothecium*), as discussed above, ecDNA stability is rather difficult

1.2.1.3 2µ plasmid fitness cost and potential host fitness improvement

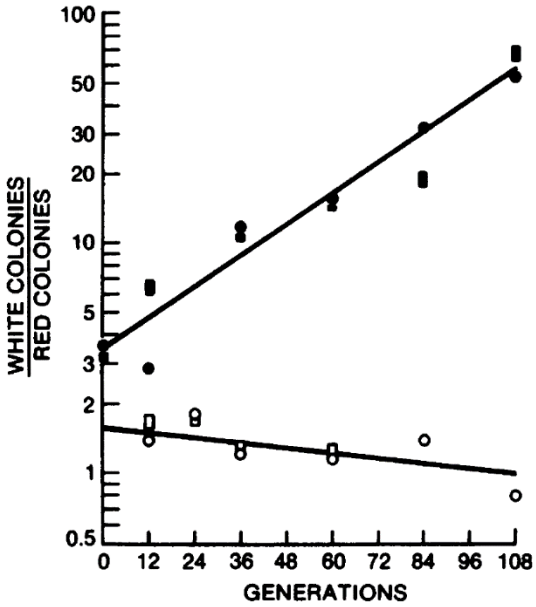


Figure 16. Curves showing the number of white (with a 2µ plasmid:ADE2) colonies and red colonies (ade2, without the 2µ plasmid over generations (adapted from Futcher and Cox, 1983)

The only known functional proteins encoded by the 2µ plasmid, are only useful for the 2µ plasmid replicative cycle and stability. Although it represents 300 kb of DNA (merely the size of a small chromosome) that replicates at each S phase, the fitness cost is barely measurable. The only reported phenotype is a minor growth defect. Cells with the 2µ plasmid grow 1% slower compared to cells lacking the 2µ plasmid[35] (Figure 16). But we must keep in mind that the phenotypic analysis of this study was done in laboratory conditions. If we look at the Ty1 sites, they improve their host fitness only in stress conditions. They give more plasticity to the genome. Maybe the advantage conferred by the 2µ plasmid would only be measurable in peculiar conditions only met in wild biotope or at least in growth conditions that are never met in laboratory conditions. A single report shows that the Flp1 recombinase could increase the survival rate of *S. cerevisiae* lacking Sgs1 and Mus81 (two enzymes involved in DNA damage repair) exposed to hydroxy-urea[135].

1.2.2 Horizontal transfer and spontaneous loss of the 2µ plasmid

1.2.2.1 The 2µ plasmid stability depends on growth conditions.

In 1983, Futcher and Cox measured that the probability of spontaneously losing the 2µ was of $3{,}1.10^{-4}$ close to the loss rate of a chromosome [34,136]. They also showed that cells lacking 2µ plasmid arose spontaneously in cells grown in exponential phase for 28 days, in agitated rich liquid media. Those data suggest that 2µ plasmid can be cured naturally from its host in certain conditions that are very different from the conditions met in wild biotopes by *S. cerevisiae*[137]. In this study cells are haploid, originating from a domesticated strain (derivate from S288C) and mating was abolished thanks to constant shaking and isogenic MAT sexual locus. This observation suggest that plasmid free cells can occur during vegetative growth (thus taking over the plasmid positive cells) and conflicts with observations of wild yeast strains.

In wild biotopes, S. cerevisiae is often diploid. When nutrients are scarce, diploid cells slow down their metabolism and end up producing an asca encompassing the 4 haploid spores resulting from meiosis. Upon more favourable conditions, haploid spores will grow again and mate. This produce a diploid cell. In 1986, Mead et al. showed that the 2µ plasmid is more stable in diploid cells entering the stationary growth state[35] and can only transmit through cell division (mitosis or meiosis) or mating. Given that exponential growth phase and rich environment are rare in natural biotope. We could assume that the rarity of natural plasmid free isolates is explained by the metabolic state of yeast host. Moreover, in laboratory condition, domesticated strains tend to have lost certain ability, notably regarding sporulation[137] so we

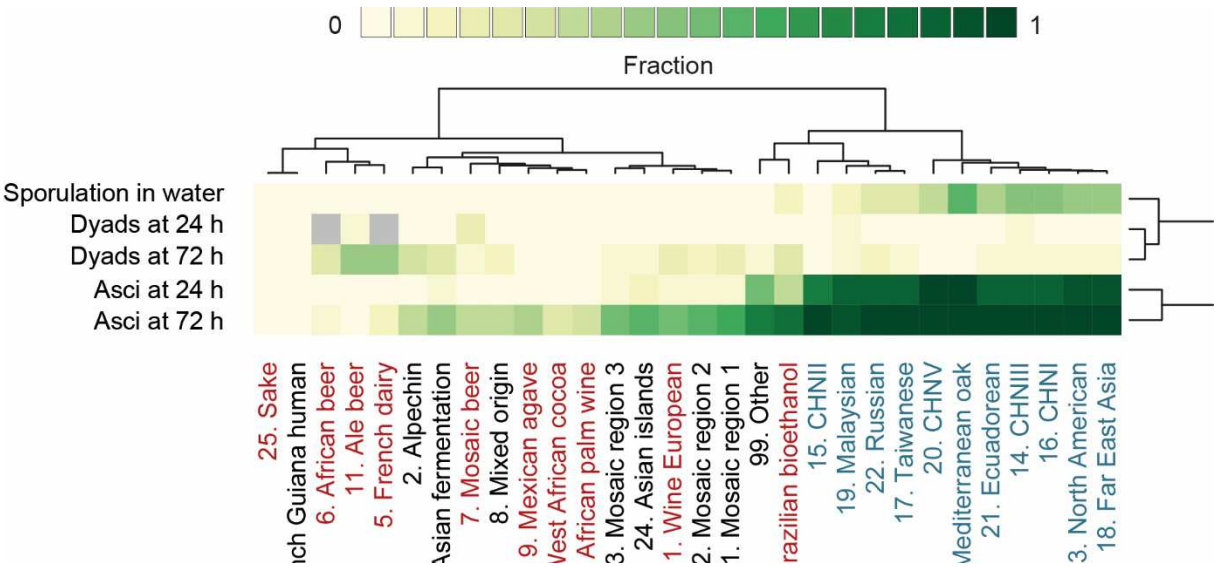


Figure 17. Diagram adapted from De Chiara et al. comparing different S. cerevisiae strains for their ability to produce an ascus(sporulate) when grown in standard KAc sporulation media. Wild isolates are in blue.

could wonder if the 2µ plasmid is more a reminiscence of the wild *S. cerevisiae* and if in different conditions the 2µ plasmid would be an advantage as the killer phenotype[138].

1.2.2.2 Mating is the only horizontal transfer.

The mating of spores will produce a diploid. It has been shown that it is impossible to isolate a 2µ plasmid free spore from an ascus produced by the sporulation of a diploid cell hosting the 2µ plasmid. This means that the 2µ plasmid can propagate horizontally by mating since it can happen between two spores from the same ascus or between unrelated spores.

1.2.3 2µ maintenance strategy

1.2.3.1 The 2µ plasmid in the nucleus

The 2µ plasmid (6318 bp) faces many challenges to be maintained. First, the 2µ plasmid ORFs (REP1, REP2, FLP1, RAF1) are non-essential for *S. cerevisiae*[35]. Second, the 2µ plasmid lack the centromeric consensus sequence thus It cannot benefit from the host segregation machinery. Third, non-centromeric circular DNA molecules have a strong mother biased segregation [50,52,98]. However, the probability to lose 2µ plasmid after mitosis is close to losing a host chromosome (10^{-5} per cell per division)[35]. Moreover, shuttle vector based on the 2µ plasmid are among the most stable *S. cerevisiae* plasmids. We will describe the very optimized plasmid copy number control mechanism and segregation strategy both on plasmid and host sides. Besides the 4 ORFs coding functional proteins, it is worth pointing that several long transcripts coming from the 2µ plasmid can be observed but their function remain unclear, they may result from unstopped RNA-polymerase thus not having any function or be involved in the 2µ plasmid maintenance[139]. FISH imagery[140] revealed that as high copy number plasmid in bacteria, the 2µ plasmid tend to oligomerize and in average reparteed in 5 fluorescent foci in the centre of the nucleus. The efficient vertical transfer of the 2µ plasmid cannot be explained by simple diffusion[52] and several plasmid and host factors have been reported to be crucial for plasmid stability.

1.2.3.2 Minimalist yet very efficient partitioning system

The exact mechanism by which the 2µ plasmid overcome the mother-biased segregation is not yet fully understood neither how does the plasmid copies are split between the mother

and daughter cells. There are two hypotheses to explain the chromosome-like stability of 2µ plasmid. First, the STB sequence would function as a centromere-like region, recruiting a kinetochore and using the cellular mitotic spindle force for its vertical transfer. The uneven repartition between mother and daughter cells of the 2µ plasmid in cells lacking either the histone H3 centromeric variant Cse4[141] and the dynein Kip1[142] is in favour of this hypothesis. Those factors are both recruited to the SPB with the help of Rep1-Rep2 complex. This hypothesis has been challenged by recent data[143] in favour of another hypothesis that propose an association between the 2µ plasmid and host chromatin (as for EBV or KSHV episomes in human cells) and the formation of a pseudo centromere on the plasmid during mitosis. The association of the 2µ plasmid and host chromatin was first described in 1980 using sucrose gradient[29] and confirmed later with for instance live fluorescence microscopy[143] or FISH[140]. To explore the mechanism of segregation of the 2µ plasmid, microscopy offers the ability to track the position of reporter plasmids thus exploring the conditions where the reporter plasmid is properly segregated or not will give insights on the 2µ plasmid segregation machinery. Interestingly, the localisation of the 2µ plasmid in the nucleus is different when using live microscopy and a reporter plasmid compared to FISH experiment (Figure 18). It has been shown that a reporter plasmid (tracked with a lacO array and lacI-GFP fusion protein) with the STB sequence showed a strong mother biased segregation when the 2µ plasmid is absent. The same results were observed for the same plasmid lacking STB but in cells hosting the 2µ plasmid[144]. Moreover, when sister chromatids cohesion or mitotic spindle is impaired, the 2µ plasmid miss segregate. Interestingly, in those conditions, the 2µ plasmid does not segregate more often in the mother or in the daughter cell[144]. It also have been shown that Cse4 recruitment is impaired when the mitotic spindle is depolymerized by nocodazole treatment[141]. The unequal repartitions of the 2µ plasmid copy when the mitotic spindle is not functional indicate that the partitioning systems deals with two challenges. First, plasmid DNA must be tethered to host chromatin with the help of plasmid encoded proteins otherwise it would be segregated in the mother cell. Second, the plasmid

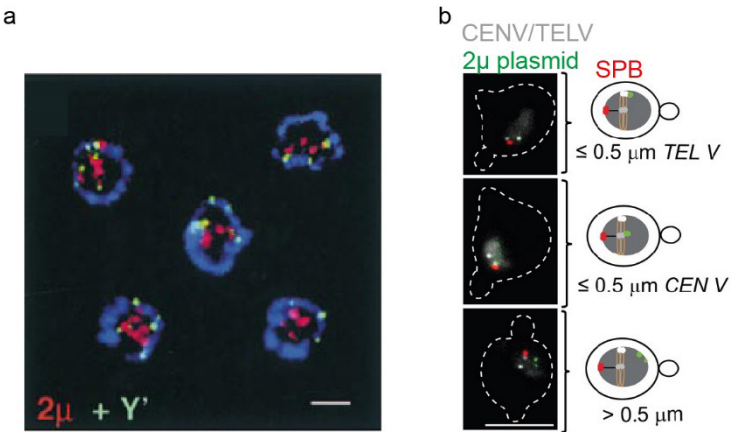


*Figure 18. (a)Results from Heun et al showing the 2µ plasmid localisation in the nucleus of S. cerevisiae by FISH. (Y' corresponds to telomeres) (**b**) Live microscopy images of a 2µ plasmid reporter plasmid adapted from Kumar et al. 2021.*

copies must be reparteed between the two sister chromatids of host chromosomes for the equal segregation of plasmid DNA. This could be mediated by Cse4 and cohesins (recruited with the help of Rsc2 complex)[145]. Along with Kip1, several other Microtubule Associated Proteins (MAPs) (but no kinetochore component[136]) have been shown to be involved in the segregation machinery of the 2µ plasmid. Namely, Bik1 and Bim1[146]. In this report Prajapati and al. showed that the deletion or either one of those proteins increased the loss rate of the 2µ plasmid in exponentially growing cells and as Cse4 recruitment it requires mitotic spindle integrity. It is though that the association between the 2µ plasmid and microtubules is more related to the proper repartition of plasmid copy between the two sisters since cells lacking Kip1 do not lose 2µ plasmid over time[147]. We could assume that the core component of the 2µ plasmid segregation machinery is the tethering to the host chromosome and the microtubules helps the positioning and equal repartition of the plasmid between mother and daughter cells. Given that plasmid copy number can be increased by rolling circle replication, even if the microtubule part of the segregation machinery is defective, the plasmid will be stable at the population level and plasmid copy number would be heterogenous.

1.2.3.3 Plasmid copy number homeostasis.

The copy number of the 2µ plasmid is approximately 60 in exponentially growing cells[130]. In association with miss-segregation events, this suggest that the replication of the 2µ plasmid is very efficient and is able to produce more than 2 copies of the 2µ plasmid per cell cycle[34]. Thanks to its origin of replication that is fired early in the S phase and the recruitment of host DNA replication machinery, the 2µ plasmid genome is replicated. We discussed above that the repartition between the mother and daughter cell is often not equal. Thus, the 2µ plasmid copy number homeostasis is crucial for the stability of 2µ plasmid. If it is too low the plasmid has a higher probability to be lost upon mitosis and if it is too high, the cell will die[148]. The 2µ plasmid amplification need the action of Flp1 recombinase. The FLP1 expression is inhibited by Rep1, Rep2 and Raf1. If the 2µ plasmid copy number is low, then the FLP1 repressors are not at a sufficient level to inhibit FLP1 transcription. Flp1 will induce the rolling circle replication thanks to multiple recombination event between the 2µ plasmid FRT sites. This leads to an increase in the number of plasmid copies[148,149]. However, no mechanism lowering the 2µ plasmid copy number are reported. We could assume that 2µ plasmid tend to accumulate in old mother cells as eccDNA[50].

To our knowledge, no genomic studies have been done on the 2µ plasmid to better understand its relationship with its host. Using genomic approach and more precisely technique to access the conformation of DNA in the nucleus could help understand the interaction between the 2µ plasmid and its host genome. Moreover, such techniques can be used to identify novel nuclear plasmids.

1.3 Unbiased genomic approach to identify and characterize novel ecDNA.

The bloom of next generation DNA sequencing allowed high-throughput approach to be developed. It can be used to identify novel ecDNA and their physical relationships with host genome.

1.3.1 Identifying extra chromosomal DNA, importance of unbiased approach.

1.3.1.1 Whole genome sequencing and scaffolding

Building the reference genome of an organism is an especially major step toward genomic analysis. The quality of future analysis depends on this process. In genomic analysis, we try to assign each read to a genomic locus. Alignment algorithm (such as Bowtie or minimap) search for the best match between a read and the reference genome. If a read is not found on the reference it is discarded and often tagged as a contaminant. This explains why the 2µ plasmid is often not analysed in most genomic studies on *S. cerevisiae* since its sequence is not included in *S. cerevisiae* reference genome. As discussed above, there is a dogma on genomes. For eukaryotes, we expect multiple linear chromosomes and mitochondria/chloroplast small circular genome. For prokaryotes, we expect one circular large chromosome and several significantly smaller ecDNA corresponding to MGEs. However, several examples discussed above have genomic structures outside of this dogma (nuclear plasmids or *V. cholerae*). Once the DNA content of a supposedly pure sample of an organism is sequenced, scaffolding, and assembling software try to solve the puzzle and reconstruct the genome from the sequenced chunks. In details, the assembly of the genome is done by assembling overlapping regions of reads. Ideally, each read overlaps with one unique read at its extremities. However, and especially for eukaryotes, a percentage of the genome is present in multiple copy creating ambiguity. When assembly is over, often small contigs remain. They are often discarded because they often come from contaminants. But do they? If we were de-novo and naively assembling the genome of *S. cerevisiae* only using shotgun reads, the 2µ plasmid

would most certainly tag as a prokaryotic MGE The use of the 3D organisation of the DNA gives information on the physical proximity of DNA molecules and allow to remove ambiguity.

1.3.1.2 Using 3D organisation of genomes to polish genome scaffolding and detect cellular compartment.

1.3.1.2.1 Access the 3D organisation of genome: the Chromosome Conformation Capture method.

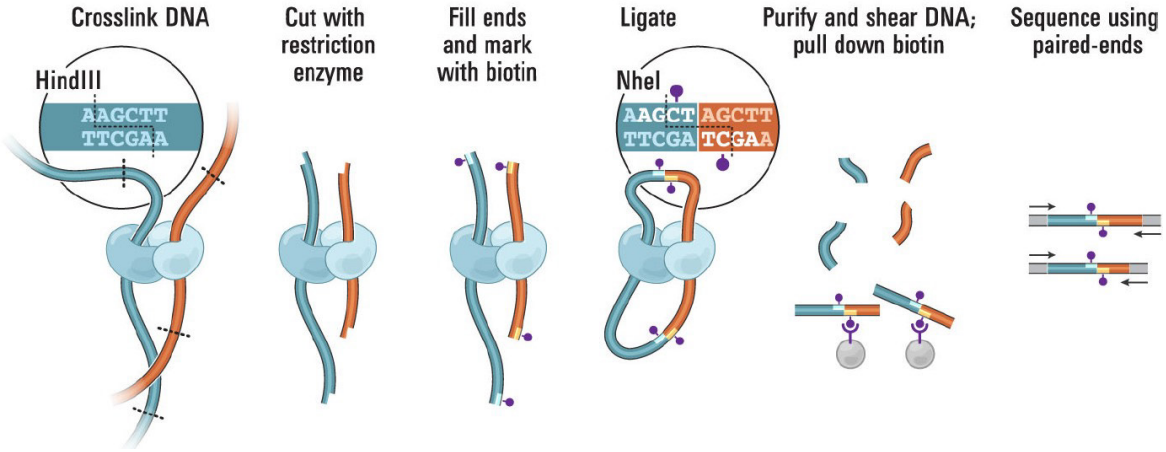


Figure 19. Hi-C experimental procedures with the biotinylation represented in Lieberman-Aiden et al.

When we extract DNA from a sample, we release DNA from its physical constraint, and we lose the 3D organisation information. The use of fixing agent that bridge covalently DNA with proteins allows the 3D organisation to be conserved upon constraints release. The aim of the Hi-C protocol is to create chimeric DNA molecules by bounding together the DNA molecules trapped in protein-DNA complexes. Often, the fixing agent is formaldehyde, but fixation protocol can be adapted based on the peculiar interactions we want to study. For instance, the double fixation using formaldehyde and EGS make long range contact easier to study. Then the protein-DNA complex are separated from each other using endonuclease (restriction enzyme[150] or Micrococcal nuclease for the micro-C[151]) that will produce small protein-DNA complex encompassing DNA molecule that were physically close in the nucleus. The action of a DNA ligase fuses together the DNA molecule trapped in a protein complex. After reversing the crosslink and protein digestion, we end up with chimeric DNA molecule.

Chimeric compared to genomic DNA. The identification of the genomic loci found in the chimeric DNA molecule gives a pair of contact. From the list of all the contact pairs, we can determine contact frequencies between two loci. This process can be done on the whole genome[152]. The most common graphical output of Hi-C technique is the contact map (Figure 20). The contact map is a square which sides represent the genome. If we draw a line starting from one side on the coordinate "a" and another line from an adjacent side on the coordinate "b," the intersection of those two lines correspond to the contact frequency of "a" with "b." The

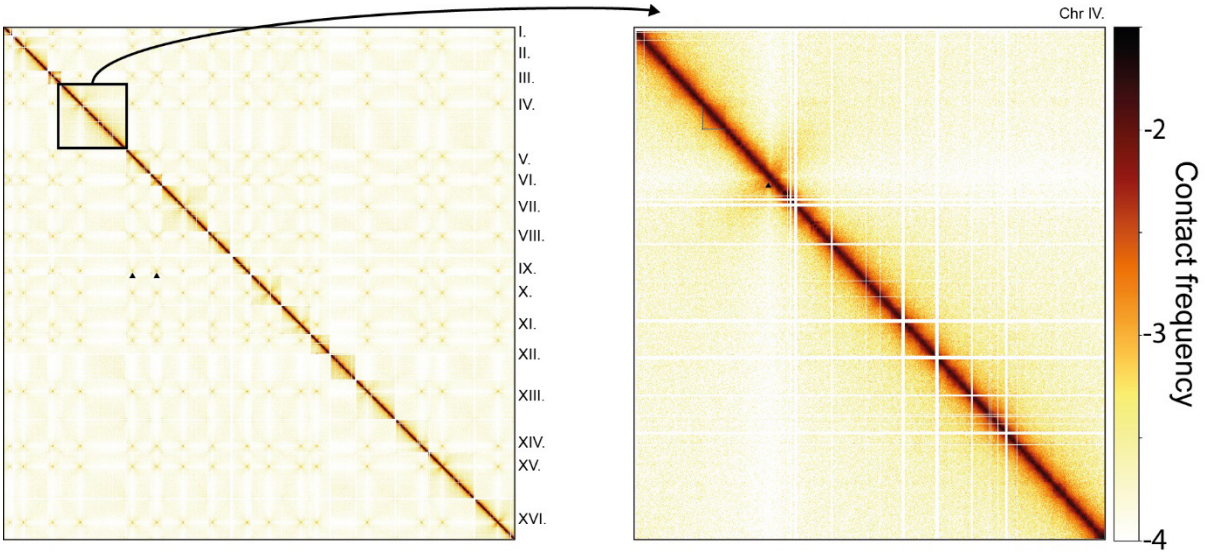


Figure 20. Example of Hi-C contact map representing the whole genome(left) and only the chromosome IV. (right). We can distinguish the clustering of centromeres on the left panel (marked by black arrowhead); the peculiar cis signal of centromere of chromosome IV. And small domains on chromosome IV. (Delimited by black lines)

genome is divided in bins of a given genomic size. It means that we analyse the contact frequency between two genomic bins. The size of the bin depends on the Hi-C technique, the quality of the Hi-C library, the organism and how much DNA molecules have been sequenced. The smaller the bin is the more resolutive the contact map is. 3D organisation of genomes can also be used to better understand biological mechanism such as the condensation of DNA throughout the cell cycle[153] or the homology search in the homologous recombination repair pathway[154]. It also has been used to locate episomal DNA in nucleus[155]

1.3.1.2.2 Physical proximity as a tool to polish genome assembly and assign DNA to compartment.

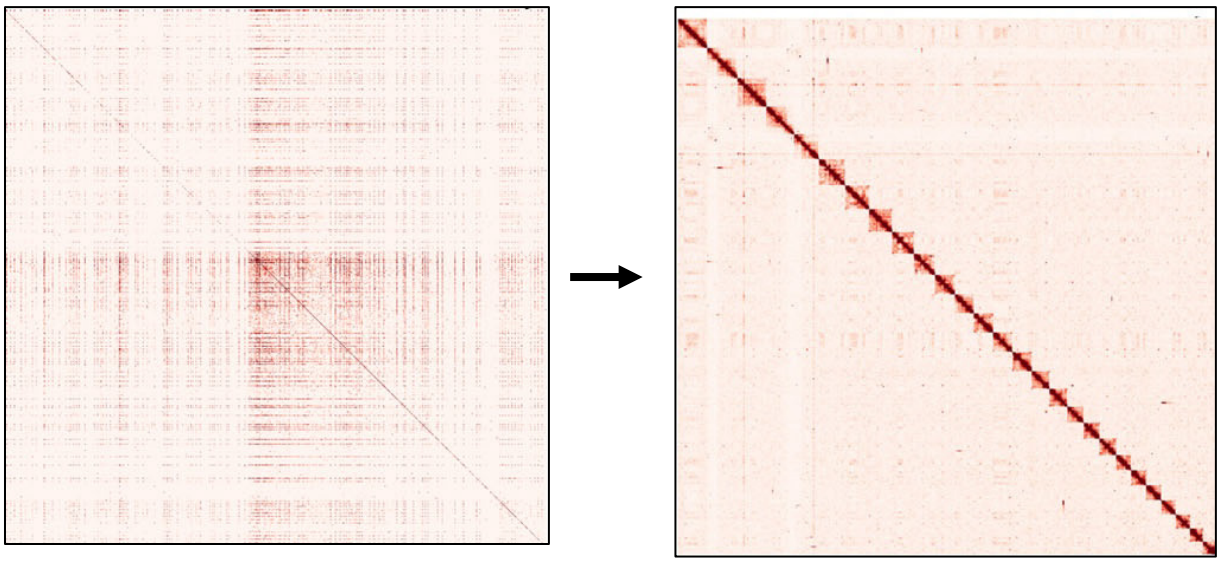


Figure 21. Contact maps adapted from Baudry at al. Contact maps generated using initial reference genome(left) and polished reference genome (right) with the help of Hi-C data.

Given that the DNA is a polymer, colinear sequences have higher contact frequency than sequences distant on the genome. 3D organisation of the genome can solve many ambiguity regarding the assembled genome[156] (Figure 21). Moreover, DNA that is within the same compartment have a higher contact frequency than if they were in different compartment. If we plot the contact map of the nuclear genome and mitochondrial genome, two clear domains are visible [157]. Such analysis can also be used to discriminate between a random contaminant and an ecDNA element. For instance, in S. cerevisiae the contact frequency between the 2µ plasmid and the nuclear genome is higher compared to the contact frequency between mitochondrial genome and nuclear genome. Moreover, ecDNA topology and organisation is accessible with Hi-C. The use of Whole Genome Sequencing and Hi-C together would allow de novo screening. for novel episomes in eukaryotes. It requires to recover what would be discarded in the first place.

1.3.2 Locating ecDNA in the nucleus.

The Hi-C data is an all vs all technique, meaning that it quantifies physical contacts between all DNA molecules The Hi-C data can be analysed in a way to focus on the physical contacts made by a given ecDNA with its host genome. We could assume that ecDNA is randomly distributed in the nucleus but studies on EBV episomal DNA showed that EBV contact more frequently active region of the genome[155]. Moreover, the interaction between host and guest can have consequences on host genome organisation[158]. However, the signal of the extra chromosomal DNA can be diluted by the host genome. DNA sequencing is like a random pick

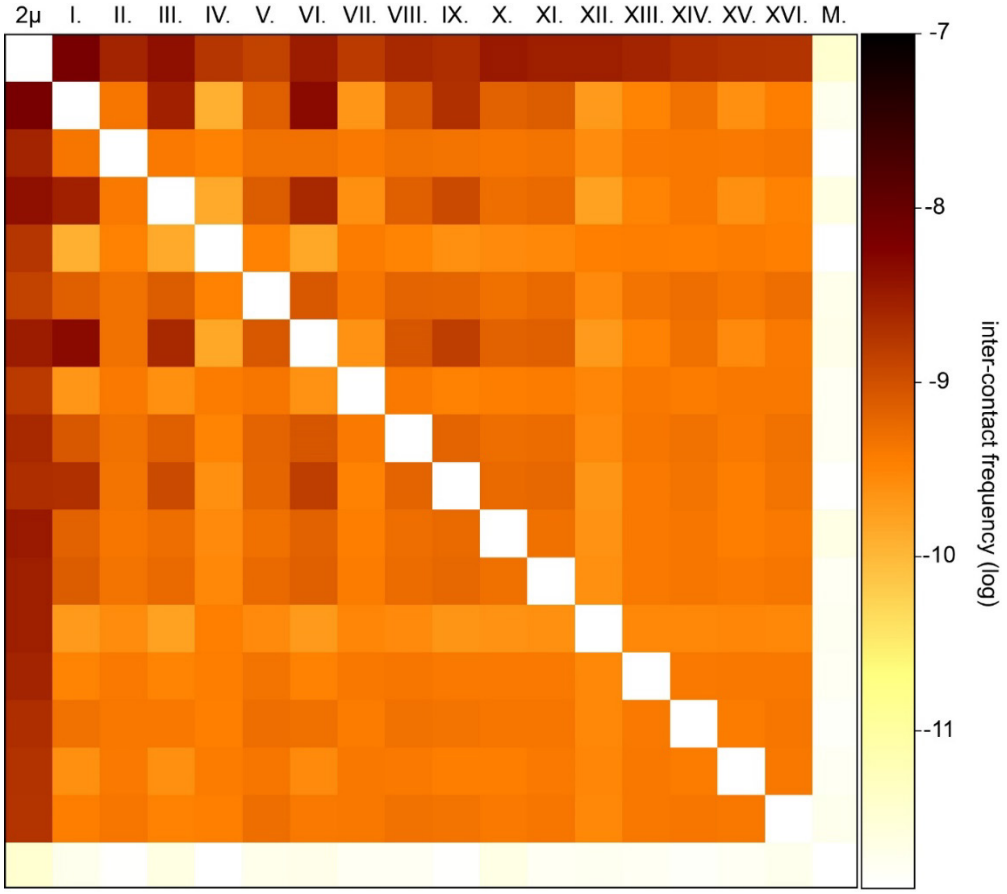


Figure 22.Contact map showing inter-contacts between S. cerevisiae chromosomes (roman capita number), the 2μ plasmid(2μ) and the mitochondrial (M.) genome. The log of contact frequency is represented here according to the colour scale.

without remit, there is more probability to sequence abundant DNA compared to rare DNA. For instance, EBV genome is 172 Kb and human genome is 3 Gb. An Hi-C derivatives exist and focuses on the physical contacts of the ecDNA vs the rest. We can cite the 4C-seq (one vs all)[159] or the Capture Hi-C (many vs all)[47] technique that uses different experimental procedure compared to Hi-C but output contact pairs focused on the genomic loci of interest. The contact between extra chromosomal DNA and host genomes is often non randomly distributed and

hotspots can be identified[16,155]. Upon the identification of hotspots, one could wonder what defines a hotspot, is it its sequence? a host protein occupancy or any other determinant?

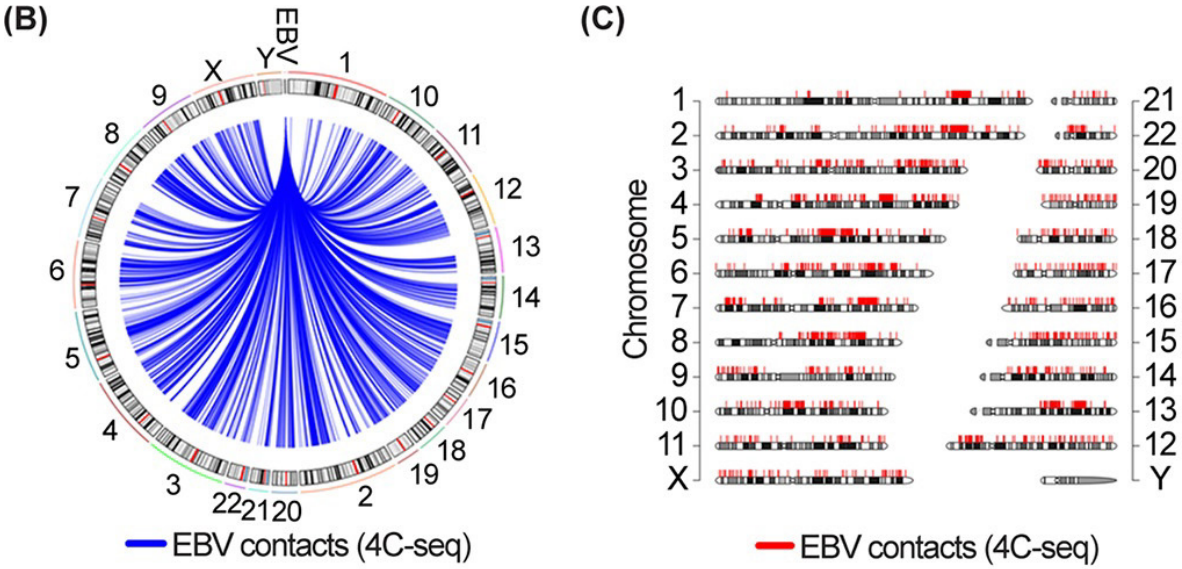


Figure 23. Results adapted from Kim et al. 2020 showing contacts between EBV episome and its host genome.

1.3.3 Intersecting evidence from different approach to access mechanism of interaction.

Reporting contact hotspot is a first step toward deciphering a mechanism of interaction between extra chromosomal DNA and host genome. This report highlights positions of interest on the host genome and looking for the common denominator between those regions can unravel key component of the strategy of extra chromosomal DNA to be maintained in host cell. The possibilities for genomic information to analyse is very vast since we can gather information on transcription, chromatin accessibility, nucleosomes organisation and protein

occupancy. If we are lucky enough, some tracks will be informative[16]. A way to deal with multiple input information and sort between uninformative and informative genomic characteristics is to use artificial intelligence. Recently [160] , Zhang et al. were able to predict Hi-C matrices in silico, using only genomic signals. If this kind of approach were to be used on

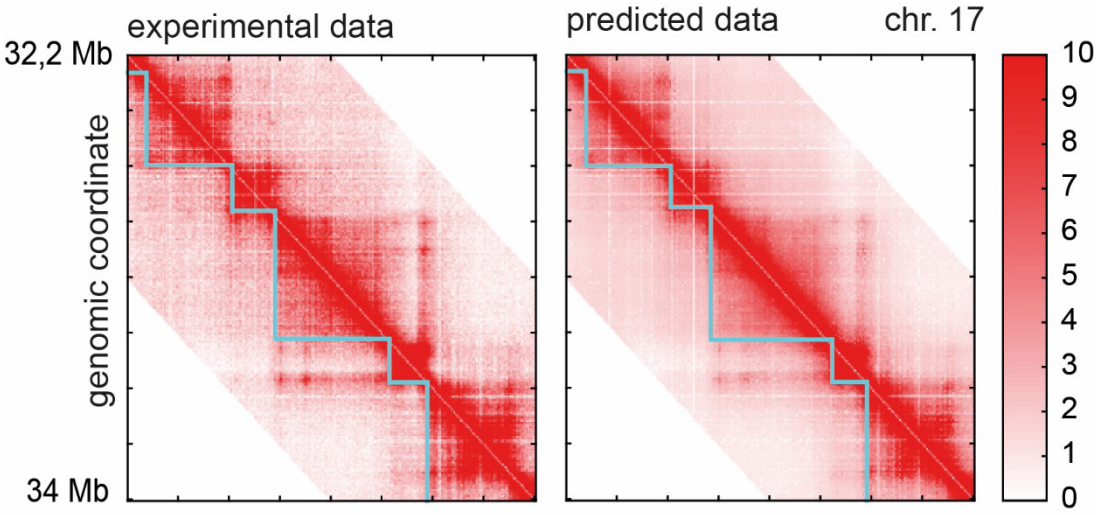


Figure 24. Comparison between experimental (left) and predicted (right) contact map adapted from Zhang et al.

extra chromosomal DNA and their physical contact frequency with host chromosome could be predicted by artificial intelligence, informative genomic characteristics would be identified without the bias of an experimenter choosing which genomic data to use.

1.4 PhD objectives

The 2µ plasmid study peaked around 1980's. The interest of research groups was mainly on its replication and its stability in its host cell. Along the way, the 2µ plasmid recombination machinery was used in the FRT/Flp technology. The stability of the 2µ plasmid was used to develop the first stable bio-engineered plasmid in *S. cerevisiae*. The interest for the 2µ plasmid slowly decayed. To our knowledge only 2 research groups in the world keep publishing on the 2 µ plasmid. Many questions remain unanswered. The 2µ plasmid is most certainly a very optimised systems with intricate aspect of its stability. We can distinguish: the plasmid copies number homeostasis, the repartition of plasmid copies between mother and daughter and finally its strategy to overcome the mother-biased segregation. It is still not clear how the 2µ plasmid segregates to both daughter and mother cells. And more precisely how the 2µ plasmid overcome the mother-biased segregation and how the 2µ plasmid copies are reparteed between mother and daughter cells. Until now, no 3C studies have been conducted on the 2 µ plasmid precisely. The 2µ plasmid have been overlooked in all previous 3C studies on *S. cerevisiae*, mainly because its sequence is excluded from the *S. cerevisiae* reference genome, and it is a decaying research study. I focused on the relationship between the 2µ plasmid DNA and its host chromatin. Here we used published dataset to scan many biological conditions and rule out some hypotheses. We produced a model explaining the tethering of the 2µ plasmid on its host chromosome. The tethering could be mediated by the histone H4 tail and occurs on long inactive host genomic regions. We decided to test our model with artificial biology approach. The 2µ plasmid was titrated from the host chromosome by the presence of a 1Mb long inactive artificial chromosome which is linearized *Mycoplasma mycoides* with telomeres, an origin of replication and a selection marker.

2 Results I: 2µ plasmid pre-print.

The following pages are from the pre-print article we recently published on biorxiv[161]. This works is the result from contributions of several scientists. My PhD directors and myself conceptualised the project, wrote the manuscript and designed the panels. All the analysis of published dataset were done by my co-director Axel Cournac. The FISH experiments were conducted by Myriam Ruault and Antoine Even in Angela Taddei's laboratory at Institut Curie, while the mycoplasma strain I used was constructed by Léa Meneu. The amoeba cells were cultured and chemically fixed by Sandrine Adiba at ENS Ulm. Finally, all new Hi-C experiments and their analysis were done by me, with experimental and computational training by Agnès Thierry and Axel Cournac, respectively.

The results also have been presented in several scientific gatherings:

- Oral presentation at the EMBO workshop: "the yin and yang of chromosomal and extra-chromosomal DNA" (2022, Ascona)
- Oral presentation at the third symposium on fungal genetics in IDF (2022, Paris)
- Oral presentation at the doctoral school day (2023, Paris)
- Oral presentation at a "Work in Progress" (WIP) from the genomes & genetics department of Institut Pasteur. (2022, Paris)
- Oral presentation at the "Yeast Club" of Paris (2022, Paris)

Anchoring of parasitic plasmids to inactive regions of eukaryotic chromosomes through nucleosome signal

Fabien Girard[1,2,3], Antoine Even[4], Agnès Thierry[1], Myriam Ruault[4], Léa Meneu[1,2], Sandrine Adiba[5], Angela Taddei[4], Romain Koszul[1,#], Axel Cournac[1,#]

[1] Institut Pasteur, CNRS UMR 3525, Université Paris Cité, Unité Régulation Spatiale des Génomes, 75015 Paris, France

[2] Sorbonne Université, Collège Doctoral, F-75005, Paris, France

[3] Departement de Biologie, Ecole Normale Supérieure Paris-Saclay, Université Paris-Saclay

[4] Institut Curie, Université PSL, Sorbonne University, CNRS, Nuclear Dynamics, Paris, France.

[5] Institut de Biologie de l'ENS, Département de biologie, Ecole Normale Supérieure, CNRS

[#] corresponding authors: romain.koszul@pasteur.fr, axel.cournac@pasteur.fr

2.1 Abstract

Natural plasmids are common in prokaryotes, but few have been documented in eukaryotes. The natural 2µ plasmid present in budding yeast *Saccharomyces cerevisiae* is one of the most well characterized. This highly stable genetic element coexists with its host for millions of years, efficiently segregating at each cell division through a mechanism that remains poorly understood. Using proximity ligation (Hi-C, Micro-C) to map the contacts between the 2µ and yeast chromosomes under dozens of different biological conditions, we found that the plasmid tether preferentially on regions with low transcriptional activity, often corresponding to long inactive genes. Common players in chromosome structure such as members of the structural maintenance of chromosome complexes (SMC) are not involved in these contacts which depend instead on a nucleosomal signal associated with a depletion of RNA Pol II. These contacts are stable throughout the cell cycle and can be established within minutes. This strategy may involve other types of DNA molecules and species other than *S. cerevisiae*, as suggested by the binding pattern of the natural plasmid along the silent regions of the chromosomes of *Dictyostelium discoideum*.

Keywords: eukaryotic plasmid; Hi-C; hitchhiking; *S. cerevisiae*; budding yeast; *Dictyostelium*; chromosome organization; cell cycle; SMC; transcription; chromatin; epigenetics.

2.2 Introduction

The way in which mobile DNA elements, such as plasmids, viruses, and transposons, are maintained within their hosts is a key question for understanding their phenotypes and their impact. Many eukaryotic DNA viruses are pathogens that can represent public health issues Some can even be retained over an extended period of time in the nucleus through several mechanisms. For instance, the hepatitis B virus (HBV) can integrate into the genome and remain in a latent state for very long periods (Dias et al., 2022). The Epstein Barr virus (EBV) of the herpesvirus family remains as an episome in the nucleus able to replicate and hitchhike on host chromosomes thus vertically transferred throughout cell division (Coursey and McBride, 2019; Kim et al., 2020). In contrast, few examples of plasmids naturally present in eukaryotic nuclei have been documented (Esser et al., 2012). To our knowledge, only two families of natural plasmid present in eukaryotic nuclei have been clearly identified and characterized: the *Saccharomyces cerevisiae* 2µ plasmid (Sau et al., 2019) and derivative in other ascomycota species, and the Ddp plasmid family in the social amoeba *Dictyostelium discoideum* (Rieben et al., 1998; Shammat et al., 1998). Why such a small number of plasmids remains unclear: either these objects have been less studied and are overlooked in sequencing data, or eukaryotes have developed sufficiently effective protection and/or defence mechanisms. In either case, it is not yet clear how these molecules are maintained across generations.

The 2µ plasmid is one of the most studied examples of a selfish DNA element, i.e. a molecule that does not appear to confer any fitness advantage on its host, without imposing a significant cost (Mead et al., 1986). The 6.3 kb sequence, named after its contour length when observed with electron microscopy (Sinclair et al. 1967), is replicated by the host replication machinery (Zakian et al., 1979). It is present in most *S. cerevisiae* natural isolates and laboratory strains (Peter et al., 2018) suggesting a very efficient and successful persistence mechanism. It encodes a partitioning system that includes Rep1 and Rep2, two proteins that associate with the plasmid STB repeat sequence that is essential for plasmid stability (McQuaid et al., 2019) and a specific recombinase Flp1. The origin of this plasmid is not well known. Various clues could point to the origin of bacteriophages. The specific Flp1 recombinase is a tyrosine recombinase like the Cre recombinase of bacteriophage P1. It has also been shown that bacteria can naturally transform *S. cerevisiae* yeast by conjugation (Heinemann and Sprague, 1989) making possible a transfer between different species of genetic material that evolved into the 2µ plasmid we know today. Several works point at a ‘chromosome hitchhiking’ mechanism by which the

plasmid binds to the chromosomes of the host, taking advantage of its segregation machinery during cell divisions (Sau et al., 2019). Microscopy investigations show that the 2µ plasmid colocalizes with the host chromatin, though its precise nuclear localisation remains unclear, with studies suggesting either a preferential position at the centre of the nucleus (Heun et al., 2001), or in nuclear periphery close to telomeres (Kumar et al., 2021).

De novo calling of DNA contacts between molecules is difficult to achieve with microscopy but can be done using capture of chromosome conformation approaches such as Hi-C (Lieberman-Aiden et al., 2009, Hsieh et al., 2016). We therefore used Hi-C and Micro-C data to map and quantify the plasmid physical contacts with the host chromosomes. We show that the plasmid contacts a set of discrete loci that remain remarkably stable in multiple growth and mutant conditions. The set of contact hotspots corresponds to relatively long, inactive regions. These contacts can evolve within a few minutes following environmental stress and depend on the nucleosomal H4 basic patch. Strikingly, a Mb long inactive artificial bacterial chromosome was able to bind plasmid molecules from their native binding positions, illustrating how the plasmid is spontaneously attracted by inactive chromatin, whatever its origin. Overall, these results point to the existence of a segregation mechanism whereby the plasmid may recognize a signal associated with chromatin structure to bind to the inactive chromatin regions of its hosts in a reversible way.

Very few eukaryotic plasmids have been identified. Other Saccharomycetaceae, such as the Lachancea species *Lachancea fermentati* and *waltii* that have diverged from Saccharomyces for over 100 My ago also display episomes homologous to the 2µ. Further in the phylogenetic tree, the amoeba *Dictyostelium discoideum* contains the Ddp5 plasmid. We show that these episomes also preferentially tether to inactive regions, suggesting that this may be a widespread strategy for eukaryotic plasmids to ensure their correct segregation during cell division in a way that does not disrupt host regulation.

2.3 Results

2.3.1 Chromatinization and 3D folding of the 2µ plasmid.

The 2µ plasmid sequence is typically filtered (or overlooked) from Hi-C analysis and more generally from any high-throughput sequencing data. We therefore revisited ten years of datasets to explore this overlooked sequence. Using Micro-C, a high-resolution Hi-C derivative that quantifies DNA contacts at the nucleosomal level (Hsieh et al., 2016; Swygert et al., 2019), we generated a plasmid contact map at 200 bp resolution from cells in exponential growth (**Fig. 1a**; **Methods**). The cis-contacts map revealed four small self-interacting regions corresponding to the four plasmid genes, reminiscent of the pattern observed at the level of active chromosomal genes (Hsieh et al., 2016). The ~1 kb STB-ORI region, positioned in between the RAF1 and REP2 genes, appeared as a constrained region with a stronger local enrichment in short range contacts. H3 chemical cleavage data (Chereji et al., 2018) shows the regular distribution of nucleosomes along the genes, with nucleosome free regions at transcription start sites (TSS) and an even distribution along the open reading frames (**Fig. 1b**), showing that the chromatin of the plasmid is similar to that of its host. RNA-seq data highlight a relatively low transcriptional activity of the 4 genes (taking into account its size and the number of copies per cell, the level of transcription of 2µ plasmid is 10 times less than the average for its host), and the presence of non-coding RNA as previously identified (Broach et al., 1979) (**Fig. 1b**). Finally, chromatin accessibility assessed by ATAC-seq confirms that the intergenic regions are the most open regions of 2µ plasmid (**Fig. 1b**). The cis-acting plasmid partitioning locus STB appears to be the most accessible region, which supports the notion that it acts as a gateway for known recruited host proteins (Chan et al., 2013). Note that no enrichment of the centromeric histone H3-like protein Cse4 was found on the plasmid sequence **(Supplementary Fig. 1**). Overall, these signals confirm that the chromatin composition and organization of the 2µ plasmid is very similar to that of its host's chromosomes (Nelson and Fangman, 1979). Discrete contacts between the 2µ plasmid and host chromosomes. To directly monitor the contacts made by the 2µ with the yeast genome, we plotted the relative contact frequencies between the plasmid and 16 yeast chromosomes from exponentially growing cells (Swygert et al., 2019) (**Fig. 1c, blue curve; Supplementary Fig. 2**). These curves can also be represented using a chromosomal heatmap diagram coloured along its linear axis according to a scale that reflects contact frequencies (Methods; **Fig. 1c, d,** chromosomal shapes**; Methods**).

The 2µ plasmid contacts with the hosts chromosomes were not evenly distributed, as reflected by the curve peaks (dotted black boxes) and darker stripes along the chromosomal

diagrams that represent hotspots of contacts (**Fig. 1c, black triangles**). A peak-calling algorithm yielded 73 such hotspots along the 16 host chromosomes, with most (66/73) located within chromosome arms, and 7 (more than randomly expected: 9.5% versus 3.9%) in sub telomeric regions (**Methods**). In contrast with past findings drawn from imaging approaches, centromeres were on average depleted in contact with the 2µ (**Fig. 1e**). The contact pattern measured by Hi-C correlates very well with the Rep1 occupancy signal measured by ChIP-seq (**Supplementary Fig. 3**), supporting that the plasmid is indeed positioned in the close vicinity of discrete loci within the yeast genome. In contrast, a plasmid carrying a yeast centromeric sequence colocalizes near the SPB with the 16 yeast centromeres and displays contacts only with these regions (**Supplementary Fig. 4a**). A replicative plasmid devoid of centromere (pARS) displays relatively even contacts throughout the genome and does not show contact enrichment around the regions identified with the 2µ plasmid (**Fig. 1d, Supplementary Fig. 4b**). Altogether, these data show that the 2µ makes specific, discrete contacts with dozens of loci interspersed over the entire genome, excluding pericentromeric regions.

2.3.2 REP1 and STB sequence are necessary for the specific attachment and stability.

The dimer Rep1/Rep2 has been shown to associate with the plasmid STB sequence to promote partitioning during cell division and both STB and REP1/REP2 are required for the plasmid stability in host cell (McQuaid et al., 2019; Mereshchuk et al., 2022). We therefore tested whether the distribution of contacts along the genome was dependent on these partners by generating Hi-C data of cells with a 2µ plasmid mutants lacking either REP1 or the STB loci (Kikuchi, 1983; McQuaid et al., 2019) (**Methods**). These mutant plasmids are unstable (McQuaid et al., 2019) and carry a marker gene to be retained in the cell. This instability is reflected by a relatively low copy number per cell, which can be assessed from the proportion of reads from 2µ in each library (**Supplementary Fig. 5**). In both mutants, there is no contact enrichment around the previously identified hotspots (**Supplementary Fig. 6 a, b**). Moreover, a plasmid mutant lacking only FLP1 shows contact enrichment around the identified hotspots (**Supplementary Fig. 6 c**) which indicates that this recombinase is not involved in establishing the specific contacts**.** These experiments show that the Rep1 protein and the STB sequence are essential for the establishment and/or maintenance of contacts between 2µ and specific regions of its host chromosomes.

2.3.3 Plasmid/chromosomes contact regions are stable under a range of conditions.

The diversity of published Hi-C and Micro-C experiments and the ubiquitous nature of the 2µ plasmid in yeast strains allowed sifting through existing data to detect variations in contact patterns between different genetic backgrounds, cell cycle stages, growth, and metabolic states. Laboratory strains (W303 and S288C) and natural strains (BHB and Y9, (Peter et al., 2018)) displaying high or low copy number of plasmids respectively presented nearly identical contact hotspots profiles (**Supplementary Fig. 7a,b**) (**Methods**). Furthermore, most hotspots were conserved throughout the mitotic cell cycle (**Fig. 1f**) (Costantino et al., 2020; Lazar-Stefanita et al., 2017) as well as during meiosis prophase (**Supplementary Fig. 8**) (Muller et al., 2017; Schalbetter et al., 2019). A small general increase in contact variability was observed at the later stages of meiosis prophase, which could reflect increased compaction and segregation of chromosomes into spores. The hotspot pattern was also conserved upon the degradation of chromatin associated protein complexes including members of the structural maintenance of chromosome (SMC) family cohesin and condensin (**Fig. 1g**), known to organize host genome (Bastié et al., 2022; Dauban et al., 2020) and proposed to be involved in 2µ plasmid stability (Kumar et al., 2021); in cells lacking the silencing complex member Sir3 (**Supplementary Fig. 7c**) that is able to act as a bridging complex (Ruault et al., 2021); or in cells depleted for the DNA replication initiation factor Cdc45 that reach mitosis without replicating (Dauban et al., 2020) (**Supplementary Fig. 7c**). The pattern of interaction was also conserved in cells treated with nocodazole (**Supplementary Fig. 7c**). Previous reports pointed that nocodazole treatment impaired the recruitment of Cse4 (Hajra et al., 2006), Kip1 or microtubule associated proteins Bim1 and Bik1 (Prajapati et al., 2017) meaning that those host factors are not involved in the anchoring of the 2µ plasmid on host chromosomes. The profile of contact of 2µ plasmid and host chromosomes appear very similar in other biological conditions like with double strand break (DSB) (Piazza et al., 2021), with hydroxyurea (HU) treatment (Jeppsson et al., 2022) and in different genetic mutants (**Supplementary Fig. 7c**).In sharp contrast, most hotspots were strongly attenuated in quiescent cells (Guidi et al., 2015; Swygert et al., 2021) (**Fig. 1g**), when the cells dramatically alter their transcription program (McKnight et al., 2015) and genome organization (Guidi et al., 2015; Swygert et al., 2021). The later observation prompted us to explore more closely the links between transcription and the plasmid-chromosomal contacts.

2.3.4 The 2µ preferentially tethers to inactive regions along the host's chromosomes.

To explore the links between transcription and plasmid contacts (**Fig. 2a)**, we first plotted the individual hotspots windows ordered by peak strength along with the corresponding transcription level and gene size annotation. This analysis reveals a contact enrichment with poorly transcribed regions often extending over several kbs (**Fig. 2b**), corresponding to relatively long genes (e.g., >4kb, to compare to a genome wide gene size median of 1kb) (**Fig. 2b**). They include 15 of the 19 genes longer than 7 kb that are non-expressed in these growth conditions, with the remaining four being transcribed. A statistical analysis shows that 2µ plasmid binding depends on the level of transcription and the size of the gene (**Supplementary Fig. 9**). We piled-up the contacts made by the 2µ plasmid, and 80 kb windows centered on the 73 peaks called on the contact profile, along with the pile-up transcription pattern of the 73 regions, revealing a strong depletion centred on the contact hotspot (**Fig. 2c; Methods**). The average GC% of the hotspots sequences is slightly lower than the genome average (36.8% versus 38.2%) (**Fig. 2d)**, and no consensus was identified when processing hotspots sequences using MEME algorithm (Bailey et al., 2015) (**Methods**). A magnification of contact distribution over the long inactive genes revealed a maximum enrichment in the middle of the sequence (e.g., DYN1 on **Fig. 2a, Fig. 2e**). In addition, the regions contacted by 2µ are not enriched in *cis* or *trans* contacts with each other, suggesting they do not colocalize in the nuclear space (**Fig. 2f, Supplementary Fig. 10a**). Finally, the contact signal measured by Micro-C reveals an enrichment in short-range contacts at the hotspots (**Supplementary Fig. 10b**).

2.3.5 Plasmid tethering is quickly reversible.

We then explored the dynamics of plasmid chromosome anchoring. To do this, we induced heat shock stress, known to modify transcriptome, chromatin state and protein-genome interactions (Kim et al., 2010; Vinayachandran et al., 2018), by transferring exponentially growing cells at 25°C to 37°C medium (**Methods**).Five minutes after heat shock, changes in the contact signal of 2µ plasmid were already observed. For instance, the contacts between the 2µ plasmid and the UTP20 gene (with a size of 7.4 kb) were strongly increased (**Fig. 2g**), while contact enrichment at the locus of FIR1 and RZG8 genes (with sizes of 2.6 kb and 3.2 kb) disappears (**Fig. 2g**). Precise kinetics with 4 time points show how contact points can appear (**Supplementary Fig. 11a**) or disappear (**Supplementary Fig. 11b**) in a matter of minutes. These results show that the plasmid can relocalize quickly to discrete regions.

2.3.6 2µ plasmid tether to exogenous artificial inactive chromatin

To further support the relationship between chromatin inactivity and 2µ contacts, we explored the plasmid behaviour in presence of a Mb-long exogenous sequence. The plasmid positioning in strains carrying the linearized sequence of the *Mycoplasma mycoides* (Mmyco) chromosome as supernumerary, artificial chromosome, was investigated using Hi-C (Chapard et al., 2023) (Lartigue et al., 2009) (**Fig. 2h**). Mmyco presents highly divergent sequence composition, as reflected by its GC% (24% to compare with yeast 38%). It is chromatinized by well-formed nucleosomes, imposes little fitness cost to the yeast, and segregates properly. Mmyco AT-rich sequence is devoid of transcription and shows no enrichment in RNA Pol II (Chapard et al., 2023). Strikingly, contacts between the 2µ sequence and the entire length of Mmyco's inactive 1.2Mb sequence were 6-fold higher than the average value on wild-type chromosomes. Our past work demonstrated that the Mmyco chromosome adopts a globular shape at the nuclear periphery (Chapard et al., 2023). In absence of Mmyco, the plasmid appears as several foci distributed in the nucleoplasm as previously reported (Heun et al., 2001; Velmurugan et al., 1998). In contrast, in the presence of the Mmyco, most of the 2µm signal concentrates and colocalizes with the bacterial chromosome., (**Fig. 2i**). This result demonstrates that the Hi-C data reflects titration of the 2µ plasmid from their hotspots by the long inactive sequence and shows that inactivity is one of the primary conditions of plasmid relocalization to a sequence. In addition, this result also demonstrates that the contacts quantified using Hi-C between the 2µ plasmid and genomic DNA do indeed correspond to physical relocalization of the molecules. We also analysed contacts between the 2µ and an artificial 9 kb array consisting of 200 lacO binding sites derived from *Escherichia coli* and introduced in chromosome VII (Guérin et al., 2019). The LacO array, which has a GC content of 41%, is not transcribed and is recognized by the DNA-binding repressor LacI put under the control of an inducible promoter (Guérin et al., 2019) (**Methods**). When LacI is not present, we observe a peak of contact but upon LacI binding, the array is not a contact hotspot anymore (**Fig. 2j**). These observations further support that a long (>9 kb) inactive region from a different organism but with a similar GC content than *S. cerevisiae* can be a contact hotspot for the 2µ plasmid. It has been shown that a high level of LacI binding results in nucleosome eviction (Loïodice et al., 2021). The observation that specific contact is lost when the LacI protein is attached to the region suggests that the resulting large nucleosome-free region could be responsible for the detachment of the 2µ plasmid.

2.3.7 Characterisation of hot spots of contact of the 2µ plasmid

To further characterize the composition of the regions contacted by the 2µ plasmid, we computed the average enrichment of various genomic signals (ChIP-seq, MNAse-seq, ATAC-seq) over the 73 contact hotspots. For example, we computed the histone H3 ChIP-seq average signal and observed an enrichment at the hotspots (**Fig. 3a**). In agreement with the transcriptional inactivity of the hotspots, chromatin accessibility (ATAC-seq) (Lee et al., 2018) shows that these regions are less accessible compared to the rest of the host genome (**Fig. 3b**). To determine the chromatin composition of the chromosomal regions contacted by the 2µ plasmid, we also took advantage of a recently generated ChIP-exo dataset to screen for the deposition of ~800 different proteins or histone marks along the *S. cerevisiae* genome (Rossi et al., 2021). For each genomic signal, the deposition profiles over the 73 contact hotspots were aggregated and tested for enrichment or depletion (**Methods**). In agreement with the low activity of the tested regions, most of the proteins associated with active transcription (e.g., general transcription factors or proteins of the SAGA complex) were depleted **(Fig. 3c**). On the other hand, histones H3, H2B, and histone marks like H3K79me3 and H3K36me3, were the only signals that were enriched over the contact hotspots (**Fig. 3c**). We tested for the influence of both marks by characterizing the plasmid contacts in absence of either Set2 or Dot1 methylase. Absence of Dot1, responsible for H3K79 methylation, did not affect plasmid hotspots of contact (**Fig. 3d**). In absence of Set2, which methylates H3K36, long genes are known to be derepressed (Li et al., 2007). Although contact specificity for the 2µ plasmid is still detectable in this mutant (**Fig. 3d, Supplementary Fig. 12b**), long genes have their contact signal with the 2µ plasmid greatly reduced while the level of transcription is slightly increased (**Supplementary Fig. 12c**). Therefore, these experiments further confirm that transcription activity is linked to plasmid contact profile. In absence of silent regions, the plasmid appears to relocalize to the telomeric regions (**Supplementary Fig. 12b**). The 2µ plasmid contact profile with the genome is also independent of remodeler complexes RSC2 or RSC1, and the centromere labelling protein HST2 (**Supplementary Figure 12d**). RSC2 was shown to be essential for the 2µm maintenance (Wong et al., 2002). We observed indeed a significant drop in the number of plasmids per cell in this mutant (**Supplementary Fig. 5**). However, its attachment to host chromosomes remain unchanged (**Supplementary Figure 12d**) suggesting that the mode of action of RSC2 is not directly linked to plasmid attachment. It was also shown for non-centromeric DNA circles that HST2 deacetylase was important for their condensation and propagation to daughter cells (Kruitwagen et al., 2018). However, we did not detect any changes in the contact profile of the HST2 deletion mutant (**Supplementary Figure 12d**).

Finally, the 2µ contact profile with the genome is also independent of the main chromatin remodelers, as shown by the lack of variation that follows degradation using auxin-inducible degron (AID) of Spt6, Isw1, Swr1, Fun30, Ino80, Chd1, and Isw2 (Jo et al., 2021) (**Supplementary Figure 12d**).

2.3.8 Plasmid anchoring depends on the H4 basic tail.

A recent study suggested that chromatin folding at the nucleosomal level is altered when five basic amino acids of histone H4 tail (basic patches aa 15 to 20) are converted into alanine in a mutant called H4 5toA (Swygert et al., 2021). This change takes place without affecting transcription of the contact hotspots identified in WT condition (**Supplementary Fig. 13 a**). The 2µ plasmid contacts with chromosomes in the H4 5toA mutant were quantified using Hi-C (**Fig. 3e)**. No more contact enrichment on hotspots was detected (**Fig. 3f)**, suggesting that the 2µ plasmid contact hotspots depend on the presence or composition of the H4 tail basic patch. Importantly, in the H4 5toA mutant, transcription is not modified on the contact hotspots compared to WT condition (**Supplementary Fig. 13 a**). This result shows that the plasmid - chromosome contacts can be suppressed not only by transcription activation, but also only through an alteration of the nucleosome H4 tail basic patch. The same analysis was replicated in quiescent cells: in that condition, the remaining contact specificity between the 2µ plasmid and the 73 hotspots is also lost (**Supplementary Fig. 13 d**), suggesting the tail patch or more generally chromatin folding plays an important role in their maintenance. A careful analysis showed that the Rep1 ChIP-seq signal is 90° phase-shifted with nucleosome position, (**Fig. 3g**) suggesting that Rep1 is not randomly contacting the host chromatin but is positioned in relation to the distribution of nucleosomes along the chromatin fibre. Taken together, these results point to a model in which Rep1/Rep2 proteins localise to large regions that both are transcriptionally inactive and display a nucleosome signal carried by the H4 tail (**Fig. 4a**). Note that these two features could be two faces of the same coin since transcription disturbs nucleosome distributions and could affect the histone signal specificity.

2.3.9 Other eukaryotic plasmids tethers to inactive chromatin

To assess whether this mechanism of binding to inactive sequences might concern other eukaryotic plasmids with "selfish" appearance, we analysed contact profiles of *Lachancea fermentati* and *Lachancea waltii* which also host natural plasmids related to the 2µ plasmid. In

these 2 organisms, we can also detect foci of contacts distributed across all the host chromosomes, far from the centromeres, and with also a bias towards long genes (**Fig. 4b, Supplementary Fig. 14a**). We also quantified the contacts made by the Ddp5 plasmid with *Dictyostelium discoideum* chromosomes (Rieben et al., 1998). We performed Hi-C experiment on *D. discoideum* cells in vegetative state and measured the trans contacts between Ddp5 plasmid and its host chromosomes similarly to the 2µ plasmid experiment (**Fig. 4c, Supplementary Fig. 14c, d**). Around a hundred hotspots were detected on the contact profile (**Methods**). We computed the averaged transcription profile over windows containing these hotspots (Wang et al., 2021), the regions display reduced transcription compared to the rest of the genome (**Fig. 4c**), reminiscent of the 2µ plasmid hotspots.

2.4 Discussion

In this work, we exploit overlooked genomic datasets to confirm that the 2µ plasmid chromatization is highly very similar to its host, potentially contributing to its long co-existence within its host's nucleus (Lieberman, 2006). We also investigated using chromosome conformation capture data the physical contacts between the 2µ plasmid and its host's chromosomes, revealing that the 2µ plasmid interacts to discrete positions along their arms. Most contact hotspots consist in long, poorly transcribed regions depleted in proteins of the transcription machinery or known DNA binding complexes and might be associated with nucleosome signals as it depends on the tail of histone H4.

Previous reports have pointed out that the STB region on the 2µ plasmid recruits several yeast factors necessary for its segregation into the two daughter cells, including Cse4 (Hajra et al., 2006), the microtubule associated proteins Bim1 and Bik1, and the motor protein Kip1 (Prajapati et al., 2017). The disruption of microtubules using nocodazole led to the depletion of these proteins at STB region and generated plasmid missegregation (Prajapati et al., 2017), suggesting that those factors are important for the proper repartition between mother and daughter cells. However, HiC contact maps in presence of nocodazole treatment show that the plasmid remains bound to host chromosomes, indicating that these factors are not needed to maintain attachment. Moreover, condensin (Kumar et al., 2021) and cohesin (Mehta et al., 2002) were also proposed to be necessary for the plasmid propagation, but our data show that chromosome binding is independent of any both complexes. It cannot be ruled out that these proteins are useful to the 2µ plasmid but for other processes.

Observations obtained during heat shock experiments show that the kinetics of this system are of the order of a few minutes. These rapid kinetics are hardly compatible with processes for writing or erasing epigenetic marks. For example, it has been shown using optogenetic control of Set2 that the histone mark H3K36me3 has a writing and erasing time of around 30 minutes (Lerner et al., 2020). More generally, these experiments show that the attachment of the 2µ plasmid is dynamic and non-specific to a DNA sequence, suggesting a high adaptability that can quickly adjust to the host metabolism.

Our experiments point to a model where the 2µ plasmid recognises through the REP1/REP2 proteins complex a structural signal involving several nucleosomes in the least active regions of its host chromosomes. The nature of this signal involves the basic tail of histone H4 but remains to be characterized. Since the 2µ plasmid binds to relatively long

inactive regions, a possibility is that chromatin geometry plays a role in the attachments. For instance, the 2µ plasmid system may recognize specific chromatin fibre patterns, such as the alpha or beta nucleosome motifs recently characterized in Hi-CO (Ohno et al., 2019). The positioning along silent regions would allow it not to interfere with the biological processes of its host and thus to preserve a certain neutrality that may account for its low fitness costs and its long cohabitation with *S. cerevisiae*.

The present results presented here are also molecular experimental evidence in favour of the hitchhiking model (Sau et al., 2019) which proposes that the 2µ plasmid physically attaches to the chromosomes of its host in order to take advantage of all the machinery and chromosome movements during segregation. The probability of contact between two different molecules (inter chromosomal contacts) is very low from a thermodynamic point of view. The fact that we observe robust contact enrichment between the 2µ plasmid and the identified host positions is a strong indication that the plasmid must be physically attached to these positions.

Interestingly, we observed very similar behaviour in other natural plasmids present in the nuclei of other eukaryotes: notably in the yeasts *L. waltii* and *L. fermentati*, but also in the amoeba *D. discoideum*, which is equidistant on the phylogenetic tree from the yeast *S. cerevisiae* and the human. The positioning characteristics of plasmid Ddp5 in *D. discoideum* are very similar to those we have demonstrated in *S. cerevisiae* (long non-transcribed regions). This suggests that the mechanism used may be the same between these different organisms. Whereas the CRISPR-cas9 system is based on a recognition mechanism that relies on a precise DNA sequence, the host-parasite system studied here seems to reveal a specificity mechanism based on a structural signal involving several nucleosomes. We envision that other biological processes depend on information encoded in nucleosomal availability and/or chromatin folding notably chromosome attachment of certain DNA viral episomes like Epstein Barr virus (EBV) (Kim et al., 2020) or Kaposi's sarcoma-associated herpesvirus (KSHV) or other extrachromosomal circular DNA (eccDNA) (Møller et al., 2015).

2.5 Acknowledgements

The authors thank M. Dobson for sharing 2µ plasmid mutant strains and Rep1 antibody. We thank G. Liti and J. Schacherer for sharing strains and suggestions. We thank S G Swygert and T Tsukiyama for sharing strains, Y. Barral, L. Baudry, G. Millot, P. Moreau, A. Piazza for useful discussions; C. Chapard, C. Matthey-Doret, J. Sérizay for technical advice. This work

used the computational and storage services (maestro cluster) provided by the IT department at Institut Pasteur, Paris. F.G is supported by an ENS Paris Saclay fellowship. This work has received support under the program Investissements d'Avenir launched by the French Government and implemented by ANR with the references ANR–10–LABX–54 MEMOLIFE and ANR–10–IDEX–0001–02 PSL Université Paris, Q-life ANR-17-CONV-6150005 (SA). The authors greatly acknowledge the PICT@Pasteur imaging facility of the Institut Curie, member of the France Bioimaging National Infrastructure (ANR-10-INBS-04). This research was supported by funding from the European Research Council under the Horizon 2020 Program (ERC grant agreement 771813) to R.K, from Agence Nationale pour la Recherche ANR-22-CE12-0013-01 to R.K. and A.T., and from Agence Nationale pour la Recherche JCJC 2019 (Apollo) to A.C.

2.6 Material and Methods

2.6.1 Strains and medium culture conditions

The genotype and background of strains used in this study are listed in the strain table (**Supplementary Table 3**). The jhd2::KANMX4, set2::KANMX4, dot1::KANMX4, rsc1::KANMX4, rsc2::KANMX4, hst2::KANMX4 strains were made using PCR amplified regions of strain from EUROSCARF collection. The diploid strain containing Mycoplasma chromosomes were made by crossing a strain containing a CRISPR linearized version of Mycoplasma chromosomes with BY4742. Culture media Liquid YPD media (1% Yeast extract, 2% peptone, 2% Glucose), containing 200 µg.mL^{-1} of Geneticin (Thermofisher CAT11811031) or not, and SD-HIS (0.17% Yeast Nitrogen Base, 0.5% Am-monium Sulfate,0.2% artificial dropout lacking histidine and 2% Glucose) were prepared according to standard protocols. *Dictyostelium discoideum* cells were cultured in 20 ml autoclaved SM medium (per L: 10g glucose, 10g proteose peptone, 1g yeast extract, 1g $MgSO_4*7H_2O$, 1.9g KH_2PO_4, 0.6g K_2HPO_4) with dead *Klebsellia pneumoniae* at 20°C and 130 rpm. After 4 days of growth, cells were centrifuged at 300 rpm during 10 min before performing Hi-C procedure.

2.6.2 Heat-shock experiment.

Fresh YPD media was inoculated with overnight culture of C+ BY4741 cells (YPD, 25°C) and grown at 25°C. When the culture reached 10^7 cells.mL^{-1}, heat shock was applied by adding warm (65°C) fresh YPD media to shift media temperature from 25°C to 37°C. Cells were grown at 37°C and cells were extracted at different timepoints for Hi-C and ChIP-seq.

2.6.3 ChIP-seq procedure.

ChIP was performed as described previously (Hu et al., 2015) without calibration strain. 15 OD_{600} unit of *S. cerevisiae* (approximately 1,5.10^8 cells) were harvested from an exponentially growing culture. Cell lysis was performed using Precellys in 2mL VK05 tubes and the sonication was performed on a Covaris S220 system as described previously (Piazza et al., 2021). To pull down Rep1 protein we used a polyclonal antibody, production has been previously described (Sengupta et al., 2001). Pull down chromatin was purified and prepared for paired end sequencing as described previously (Bastié et al., 2022).

2.6.4 FISH

FISH (Fluorescence In Situ Hybridization) experiments were performed as in (Gotta et al., 1996) with some modifications. The *M. mycoides* probe was obtained by direct labelling of the bacterial DNA (1.5 µg) using the Nick Translation kit from Jena Bioscience (Atto488 NT Labelling Kit). For the 2µ plasmid labelling, a 5-kb PCR fragment was amplified from the 2µ plasmid DNA using primer pair FG92 (TTTCTCGGGCAATCTTCCTA) / FG24 (GTATGCGCAATCCACATCGG). This PCR product (1.5 µg) was then labelled using the Nick Translation kit from Jena Bioscience (AF555 NT Labelling Kit). For the *M. mycoides* probe and the 2µ plasmid probe, the labelling reaction was performed at 15°C for 90 min and 30 min, respectively. The labelled DNA was purified using the Qiaquick PCR purification kit from Qiagen, eluted in 30 µl of water. The purified probe was then diluted in the probe mix buffer (50% formamide, 10% dextran sulphate, 2× SSC final). 20 OD of cells (1 OD corresponding to 10^7 cells) were grown to mid–logarithmic phase (1–2 × 10^7 cells/ml) and harvested at 1,200 *g* for 5 min at RT. Cells were fixed in 20 ml of 4% paraformaldehyde for 20 min at RT, washed twice in water, and resuspended in 2 ml of 0.1 M EDTA-KOH pH 8.0, 10 mM DTT for 10 min at 30°C with gentle agitation. Cells were then collected at 800 *g*, and the pellet was carefully resuspended in 2 ml YPD - 1.2 M sorbitol. Next, cells were spheroplasted at 30°C for 10 minutes with Zymolyase (60 µg/ml Zymolyase-100T to 1 ml YPD-sorbitol cell suspension). Spheroplasting was stopped by the addition of 40 ml YPD -1.2 M sorbitol. Cells were washed twice in YPD - 1.2 M sorbitol, and the pellet was resuspended in 1 ml YPD. Cells were put on diagnostic microscope slides and superficially air dried for 2 min. The slides were plunged in methanol at -20°C for 6 min, transferred to acetone at -20°C for 30 s, and air dried for 3 min. After an overnight incubation at RT in 4× SSC, 0.1% Tween, and 20 µg/ml RNase, the slides were washed in H_20 and dehydrated in ethanol 70%, 80%, 90%, and 100% consecutively at -20°C for 1 min in each bath. Slides were air dried, and a solution of 2× SSC and 70% formamide was added for 5 min at 72°C. After a second step of dehydration, the denatured probes were added to the slides for 10 min at 72°C followed by a 37°C incubation for 24h in a humid chamber. The slides were then washed twice in 0.05× SSC at 40°C for 5 min and incubated twice in BT buffer (0.15 M $NaHCO_3$, 0.1% Tween, 0.05% BSA) for 30 min at 37°C. For the DAPI staining, the slides were incubated in a DAPI solution (1µg/ml in 1× PBS) for 5 minutes and then washed twice in 1× PBS without DAPI.

2.6.5 Microscopy and image analysis.

For all fluorescent images, the axial (z) step is 200 nm and images shown are a maximum intensity projection of z-stack images. Images were acquired on a wide-field microscopy system based on an inverted microscope (TE2000; Nikon) equipped with a 100/1.4 NA immersion objective, a C-mos camera and a Spectra X light engine lamp (Lumencor, Inc) for illumination. The microscope is driven by the MetaMorph software (Molecular Devices). Images were not processed after acquisition. Images shown are maximum intensity projection of Z-stack acquisition.

2.6.6 Hi-C procedure and sequencing.

Cell fixation was performed with 3% Formaldehyde (Sigma-Aldrich cat no F8775) and performed as described previously (Dauban et al., 2020). Quenching of formaldehyde was done by adding 300 mM of glycine at room temperature for 20 mins. All the Hi-C were done using Arima Hi-C kit (Arima Genomics; restriction enzymes: DpnII, HinfI). Sequencing preparation was done using a Colibri ES DNA Library Prep kit for Illumina Systems (A38606024) and then sequenced on Illumina NextSeq500. The 2 Hi-C libraries concerning *Lachancea waltii* and *Lachancea fermentati* were generated using a different Hi-C protocol described previously (Lazar-Stefanita et al., 2017).

2.6.7 Data analysis

2.6.7.1 Contact data processing (Hi-C, Micro-C)

Hi-C and MicroC processing was performed using hicstuff package (Matthey-Doret et al., 2022). Briefly, the paired-end reads were aligned to the S288C reference genome (GCA-000146045.2, R64-1-1) and the 2µ plasmid sequence (GenBank accession number: CM007980) as well as with *M. mycoides* (GCA-006265075). For the two experiments containing the lacO site array, reads were aligned to genomes based on strain W303 containing the lacO site sequences (for details on the constructions, see (Guérin et al., 2019)). A threshold of 1 for mapping quality was used for these 2 experiments. Genomes for *Lachancea waltii* was CBS6430 and X56553.1 sequence was used for pKW1 plasmid. Genome for *Lachancea fermentati* was CBS 6772 and M18275.1 sequence was used for the plasmid pSM1. For *Dictyostelium discoideum*, AX4 reference genome sequence GCA_000004695.1 dicty_2.7 was used and NC_001889.1 sequence was used for Ddp5 plasmid. We used bowtie2 in its very

sensitive local mode (Langmead and Salzberg, 2012). Unique mapped paired reads are then assigned to restriction fragments and non-informative contacts are filtered (Cournac et al., 2012; Matthey-Doret et al., 2022). PCR/optical duplicates are discarded (i.e paired reads mapping at the same genomic positions). Contact signals were binned at 2 kb resolution except where noted (e.g., for the 2µ plasmid contact map or the averaged contact signal at long genes, a resolution of 200 bp was used).

2.6.7.2 Computation of contact signal of 2µ plasmid

To compute the contact signal of the 2µ plasmid with each bin of the host genome, we used the normalized following score:

$$S_i = \frac{cp_{ij}}{\sum_{j=1}^{N_{genome}} c_{ij}} / \frac{n_{plasmid}}{N_{total}}$$

$\mathbf{S_i}$ represents the contact score between 2µ plasmid and a bin $\mathbf{i}$ in the host genome. It corresponds to the proportion of contacts made with the 2µ plasmid for a bin $\mathbf{i}$ of the host genome normalized by the percentage of presence of 2µ plasmid in the library. $\mathbf{cp_{ij}}$ is the number of contacts detected between 2µ plasmid and host bin $\mathbf{i}$.$\sum_{j=1}^{N_{Genome}} c_{ij}$ is the total number of contacts detected for the host bin i. N_{genome} is the total number of genomic bins (in general 2 kb bins). $n_{plasmid}$ is the number of reads involving the plasmid. N_{total} is the total number of reads of the library. For ideogram visualization, the contact signal S_i is represented using R-ideogram package(Hao et al., 2020).

2.6.7.3 ChIP-Seq, Mnase-seq, ATAC-seq and RNA-seq processing.

ChIP-Seq, Mnase-seq, ATAC-seq and RNA-seq processing was performed using TinyMapper (https://github.com/js2264/tinyMapper). Paired end reads were aligned using the very sensitive and local mode of bowtie2 (Langmead and Salzberg, 2012), against S288C reference genome and the 2µ plasmid (CM007980) sequences . Only concordant pairs were retained and reads with a mapping quality larger than 0 were kept. PCR duplicates were removed using samtools. When available, the input was similarly processed. Coverage for each genomic position was computed using deeptools bamcoverage function then normalized by Count Per Million (CPM) method. ChIP signal was then computed by dividing immuno-

precipited normalized coverage by input coverage. Signals were then visualized with homemade python using pyBigWig package. RNA-seq signal was log represented.

2.6.7.4 Automatic detection of peaks of contact with 2µ plasmid

To detect contact peaks with the 2µ plasmid, we used the *find_peaks* function of the scipy package (with the following parameters: height = 0.8, distance=2). Before detection, the contact signal was interpolated using the *interp1d* function of the scipy package. In most of the average 2µ plasmid contact profiles shown, the set of contact peaks used is the one detected under normal log-phase culture conditions from Micro-C data (Swygert et al., 2019) corresponding to the 73 genomic positions given in **Extended Table 1**.

2.6.7.5 Averaged contact profile of 2µ plasmid around loci

To plot the average contact profile around loci of interest, we extracted the contact signals at windows +/- 40 kb centred at positions of interest using a 2 kb binned signal. In cases where the limits of the window exceed a chromosome, Nan values were used. The standard deviation is represented around the mean value.

2.6.7.6 Genomic features of regions contacted by 2µ plasmid.

To plot the aggregated profile of genomic signals, we convert contact data (cool file at 200 bp resolution) into a bw file using homem ade python code. We then used the functions *computeMatrix* and *plotProfile* from deepTools suite (Ramírez et al., 2016) to compute and plot the heatmaps. Same approach was used to plot the averaged contact signal inside long genes. Gene boxes at regions contacted by 2µ plasmid were generated using homemade python code and using the coordinates of genes of SGDdatabase (http://sgd-archive.yeastgenome.org/sequence/S288C_reference/orf_dna/). To plot the pileup plots around the pairs of genomic positions of peaks of contact with 2µ plasmid, we used the *quantify* mode of Chromosight (Matthey-Doret et al., 2020) with the following parameters --perc-undetected=100 --perc-zero=100. All possible pairs of peaks of contact in intra or inter configurations were generated using homemade python code and the function *combinations* from itertools package.

The computational screen uses a dataset of Chip-exo libraries for about 800 different proteins and genomic signals (Rossi et al., 2021). For each aligned library (bowtie2 alignment with very sensitive mode and mapping quality >0), Chip-exo signal was computed with homemade python code with a binning of 2 kb. An enrichment score was computed by taking the average of the ChIP-exo signal +/- 2kb signal around the bins of the positions of contact peaks detected in log phase (**Extended Table 1**). The different libraries were sorted according to the different categories identified in the UMAP analysis of (Rossi et al., 2021).

2.6.7.7 Data availability

Some of the data associated with this study are publicly available and their reference numbers are listed in Supplementary Tables 1 and 2.

2.6.7.8 Code availability.

All scripts required to reproduce figures and analyses are available at https://github.com/acournac/2micron-project.

2.7 Figures

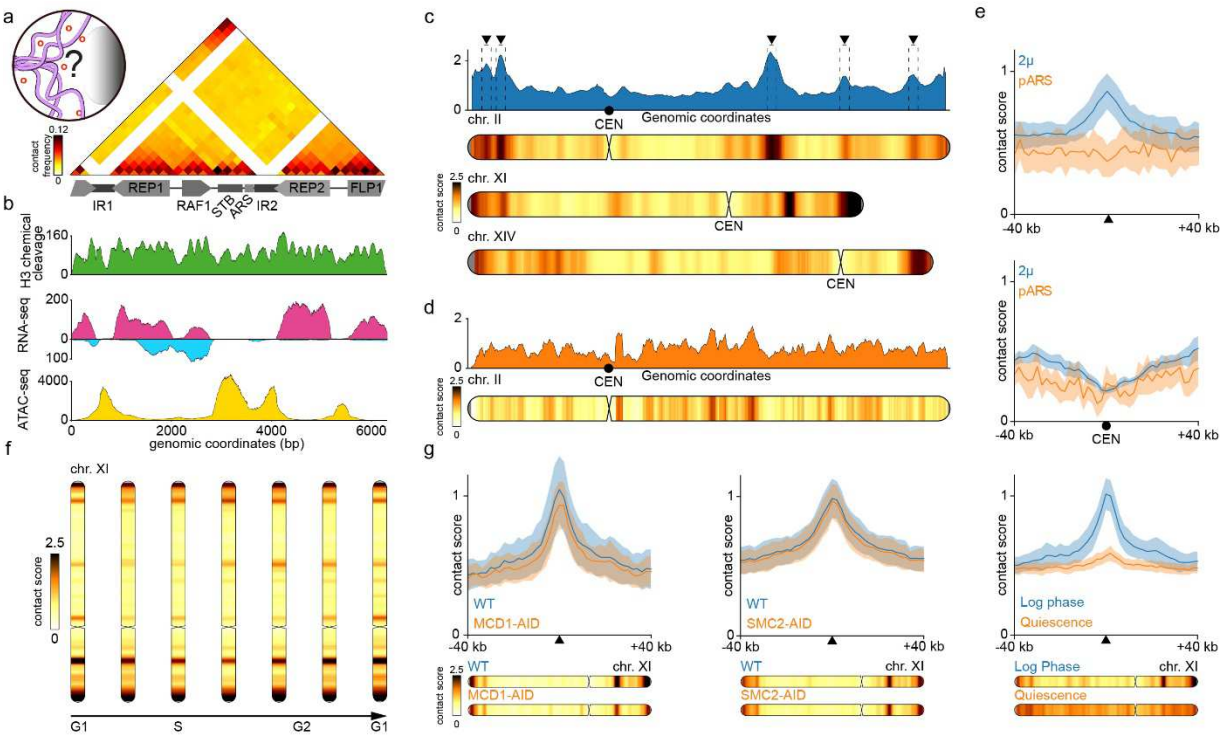


Figure 1. Specific positioning of the natural 2µ plasmid on several genomic regions of *S. cerevisiae* chromosomes.

a, contact map of the 2µ plasmid with bins of 200 bp based on Micro-C XL data in asynchronous cells (log phase). The 4 genes and cis-acting sequences of the 2µ plasmid genome are annotated below the map **b,** MNase-seq giving the nucleosome density, RNA-seq giving transcription level on forward and reverse strands and ATAC-seq giving chromatin accessibility along the 2µ plasmid. **c,** Contact profile of 2µ plasmid with several chromosomes of *S. cerevisiae* (binned at 2 kb). **d,** Contact profile of pARS plasmid (containing no 2µ or centromere systems) for chromosome II. **e,** Averaged contact signal of the 2µ and pARS plasmids at the positions automatically detected in WT, log phase for 2µ plasmid (top) and at centromeres positions (below). **f,** Contact profile of the 2µ plasmid with chr XI during the mitotic cell cycle. **g,** Averaged contact signal of the 2µ plasmid in mutants depleted in Mcd1 (cohesin subunit), mutant depleted in Smc2 (condensin subunit) and in quiescence state.

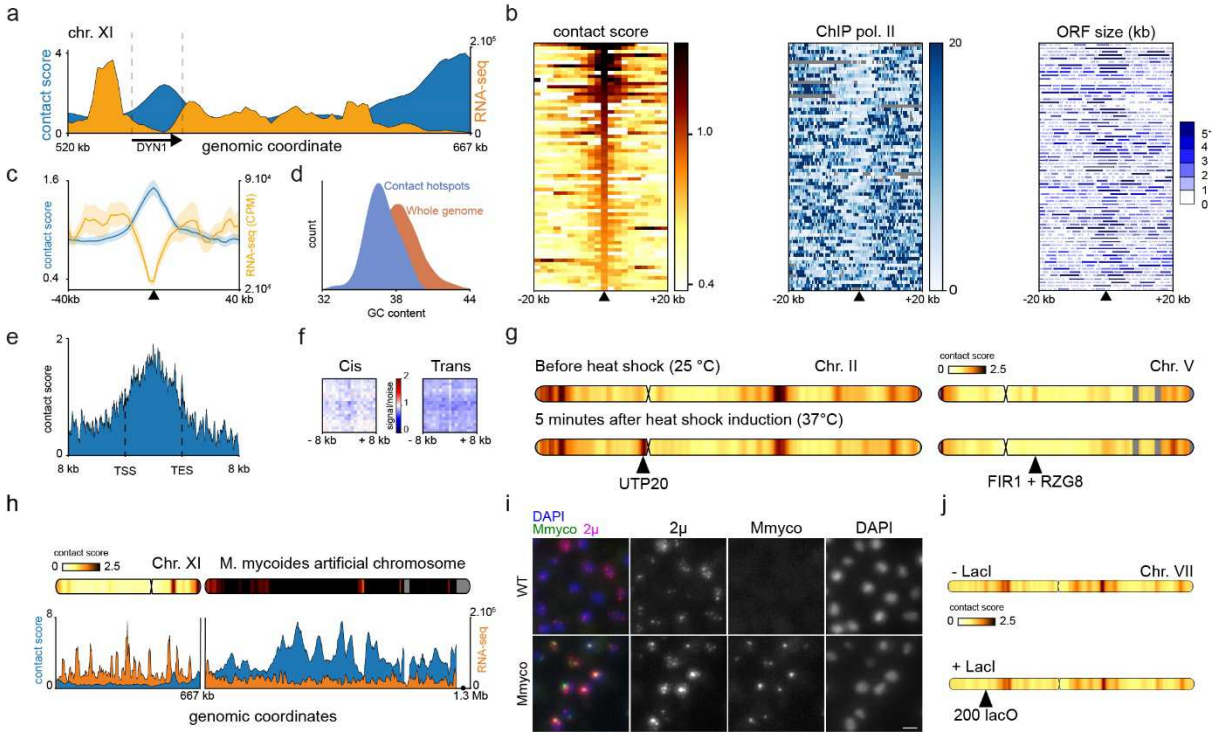


Figure 2. The contacted regions are depleted in transcription and more frequent at genes with long sizes.

a, Transcription signal and contact profile of the 2µ plasmid in a region of the Chr. XI **b**, Heat maps of 2µ plasmid contact signals, transcription level (ChIP-seq of Pol II) and gene structure sorted in descending order according to contact scores over the region − 20 to + 20 kb around the peaks of contact of 2µ plasmid. **c**, Averaged contact signal of the 2µ plasmid and transcription at the positions of contact automatically detected in WT, log phase for 2µ. **d**, Distribution of GC content for the group of sequences contacted by the 2µ plasmid and for the whole genome of S. cerevisiae. **e**, Contact profile of 2µ plasmid along long genes (> 7 kb), binned at 200 bp. **f**, Mean profile heatmap between hotspots of contact with 2µ plasmid belonging to the same (left) or different chromosome (right). **g**, Contact profile of the 2µ plasmid before and 5 min after a heat shock for Chr. II and Chr. V. Examples of regions where contact intensity varies significantly are marked with a vertical arrow. **h**, Contact profile of the 2µ plasmid in strains containing an additional bacterial chromosome M. mycoides as well as the transcription profile. **i**, Representative fluorescent images (Z-stack projection) of FISH (Fluorescence In Situ Hybridization) experiments using a probe specific for the M. mycoides genome and a probe specific for the 2µ sequence. The probes were hybridized on a wild-type strain or a strain carrying the M. mycoides genome fused to chromosome XVI. The scale bar is 2µm. **j**, Contact profile of the 2µ plasmid at an array of 200 lacO binding sites without and with LacI protein. The region is marked with a vertical arrowhead.

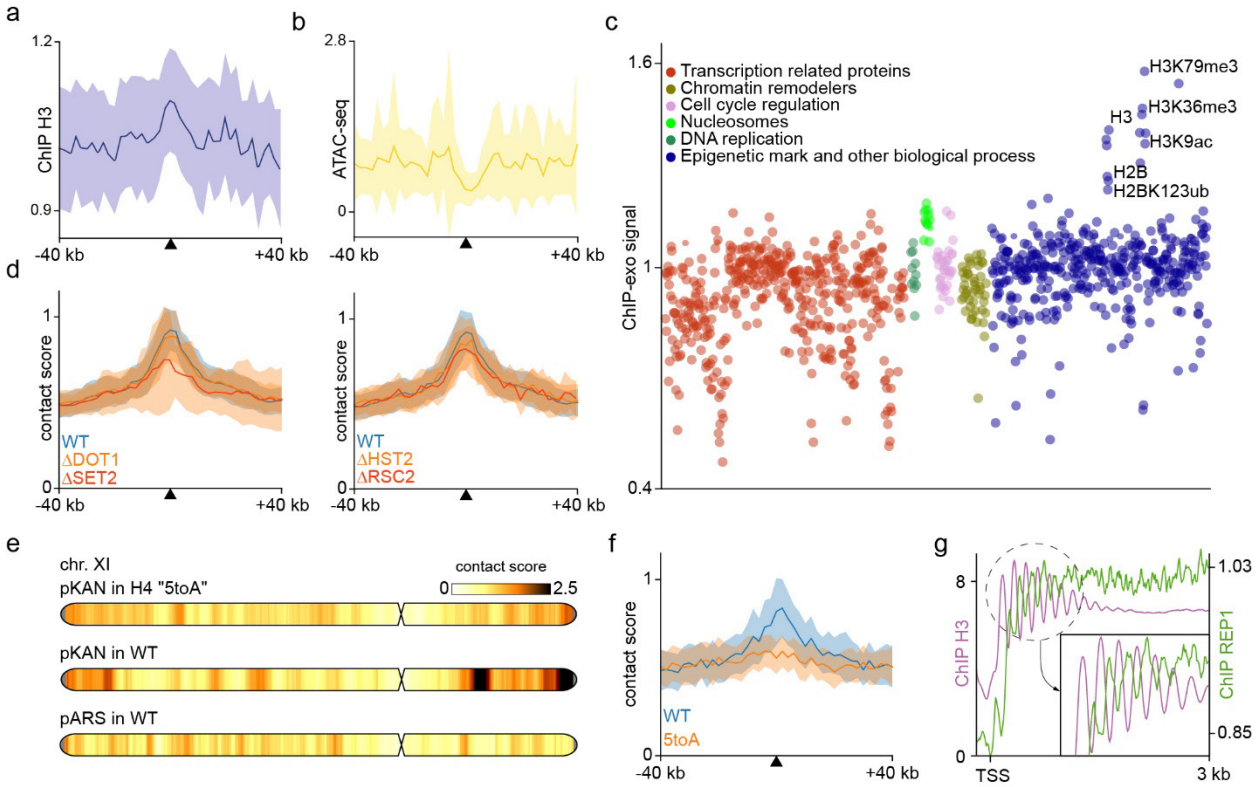


Figure 3. Specific positioning may be associated with nucleosome signal.

a, Averaged signal at the hotspots of contact with 2µ plasmid for nucleosome occupancy (ChIP-seq of H3 histone) and **b**, chromatin accessibility (ATAC-seq). **c,** Average value of signal at the hotspots of contact for 1251 ChIP-exo libraries sorted by general categories (Rossi et al., 2021). **d,** Averaged contact signal of the 2µ plasmid in mutants of epigenetic marks ΔDot1, ΔSet2 and mutant in chromatin remodeler ΔRSC2 and in deacetylase mutant ΔHST2. **e,** Contact profile of the 2µ plasmid (pKAN version) with chromosomes of *S. cerevisiae* in H4 5toA mutant, WT and of the pARS plasmid in WT for chromosome XI. **f,** Averaged contact signal of the 2µ plasmid in H4 5toA mutant and control. **g,** Averaged signal around TSS for nucleosomes and REP1 occupancy signals.

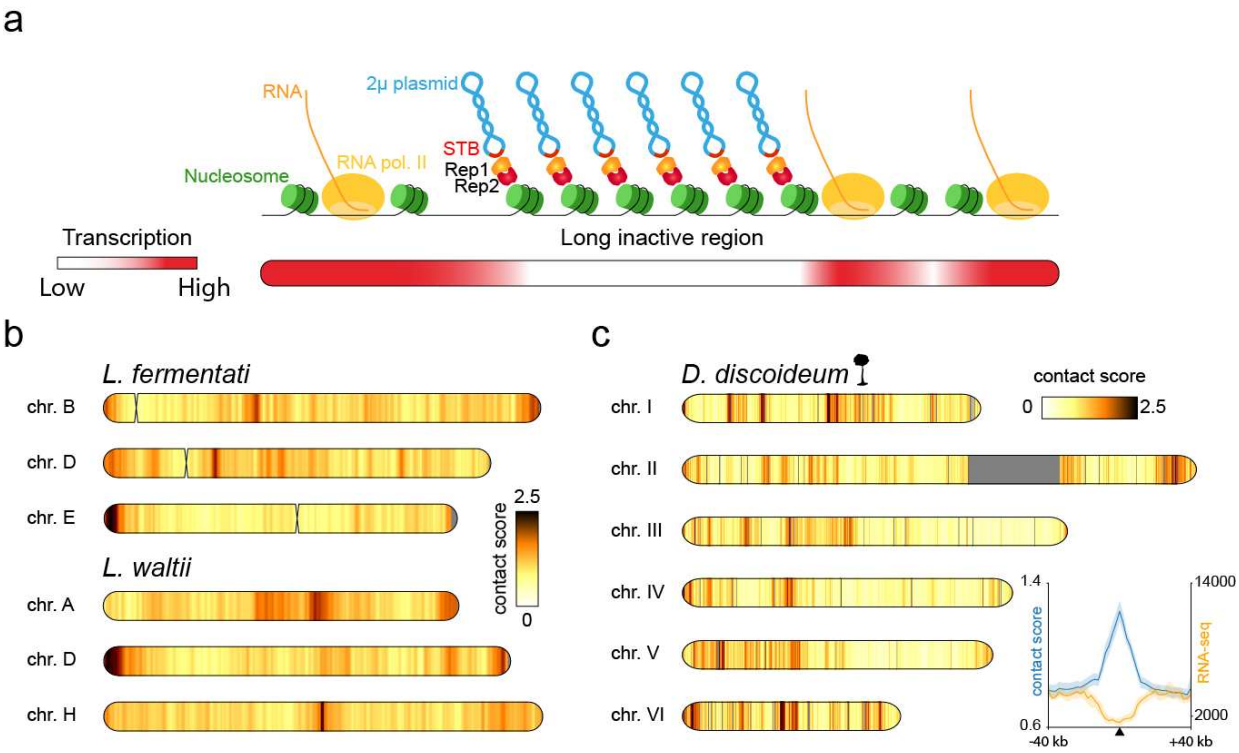


Figure 4. Contact specificity is also present in parasitic plasmids of other eukaryotes.

a, Model proposed involving the plasmid proteins REP1/REP2 and nucleosome signal to make attachment between 2µ plasmid and specific loci on host chromosomes. **b** Contact profiles of natural plasmids with several chromosomes of the yeasts *Lachancea fermentati* and *Lachancea waltii.* **c**, Contact profile of the natural plasmid Ddp5 with chromosomes of the social amoeba *Dictyostelium discoideum.* Averaged contact and transcription signals around the loci detected as peaks of contact with Ddp5 plasmid.

2.8 References

Bailey, T.L., Johnson, J., Grant, C.E., Noble, W.S., 2015. The MEME Suite. Nucleic Acids Res. 43, W39-49. https://doi.org/10.1093/nar/gkv416

Bastié, N., Chapard, C., Dauban, L., Gadal, O., Beckouët, F., Koszul, R., 2022. Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. Nat. Struct. Mol. Biol. 29, 575–585. https://doi.org/10.1038/s41594-022-00780-0

Broach, J.R., Atkins, J.F., McGill, C., Chow, L., 1979. Identification and mapping of the transcriptional and translational products of the yeast plasmid, 2mu circle. Cell 16, 827–839. https://doi.org/10.1016/0092-8674(79)90098-9

Chan, K.-M., Liu, Y.-T., Ma, C.-H., Jayaram, M., Sau, S., 2013. The 2 micron plasmid of Saccharomyces cerevisiae: A miniaturized selfish genome with optimized functional competence. Plasmid, Special Issue based on the International Society for Plasmid Biology Meeting: Santander 2012 70, 2–17. https://doi.org/10.1016/j.plasmid.2013.03.001

Chapard, C., Meneu, L., Serizay, J., Routhier, E., Ruault, M., Bignaud, A., Gourgues, G., Lartigue, C., Piazza, A., Taddei, A., Beckouët, F., Mozziconacci, J., Koszul, R., 2023. Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization. https://doi.org/10.1101/2022.12.21.520625

Costantino, L., Hsieh, T.-H.S., Lamothe, R., Darzacq, X., Koshland, D., 2020. Cohesin residency determines chromatin loop patterns. eLife 9, e59889. https://doi.org/10.7554/eLife.59889

Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R., Mozziconacci, J., 2012. Normalization of a chromosomal contact map. BMC Genomics 13, 436. https://doi.org/10.1186/1471-2164-13-436

Coursey, T.L., McBride, A.A., 2019. Hitchhiking of Viral Genomes on Cellular Chromosomes. Annu. Rev. Virol. 6, 275–296. https://doi.org/10.1146/annurev-virology-092818-015716

Dauban, L., Montagne, R., Thierry, A., Lazar-Stefanita, L., Bastié, N., Gadal, O., Cournac, A., Koszul, R., Beckouët, F., 2020. Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. Mol Cell. https://doi.org/10.1016/j.molcel.2020.01.019

Dias, J.D., Sarica, N., Cournac, A., Koszul, R., Neuveut, C., 2022. Crosstalk between Hepatitis B Virus and the 3D Genome Structure. Viruses 14, 445. https://doi.org/10.3390/v14020445

Esser, K., Kück, U., Lang-Hinrichs, C., Lemke, P., Osiewacz, H.D., Stahl, U., Tudzynski, P., 2012. Plasmids of Eukaryotes: Fundamentals and Applications. Springer Science & Business Media.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., Gasser, S.M., 1996. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. J. Cell Biol. 134, 1349–1363. https://doi.org/10.1083/jcb.134.6.1349

Guérin, T.M., Béneut, C., Barinova, N., López, V., Lazar-Stefanita, L., Deshayes, A., Thierry, A., Koszul, R., Dubrana, K., Marcand, S., 2019. Condensin-Mediated Chromosome Folding and Internal Telomeres Drive Dicentric Severing by Cytokinesis. Mol. Cell 75,131-144.e3. https://doi.org/10.1016/j.molcel.2019.05.021

Guidi, M., Ruault, M., Marbouty, M., Loïodice, I., Cournac, A., Billaudeau, C., Hocher, A., Mozziconacci, J., Koszul, R., Taddei, A., 2015. Spatial reorganization of telomeres in long-lived quiescent cells. Genome Biol. 16, 206.

Hajra, S., Ghosh, S.K., Jayaram, M., 2006. The centromere-specific histone variant Cse4p (CENP-A) is essential for functional chromatin architecture at the yeast 2-µm circle and partitioning locus and promotes equal plasmid segregation. J. Cell Biol. 174, 779–790. https://doi.org/10.1083/jcb.200603042

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., Chen, J., 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. 6, e251. https://doi.org/10.7717/peerj-cs.251

Heinemann, J.A., Sprague, G.F., 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 340, 205–209. https://doi.org/10.1038/340205a0

Heun, P., Laroche, T., Raghuraman, M.K., Gasser, S.M., 2001. The Positioning and Dynamics of Origins of Replication in the Budding Yeast Nucleus. J. Cell Biol. 152, 385. https://doi.org/10.1083/jcb.152.2.385

Hsieh, T.-H.S., Fudenberg, G., Goloborodko, A., Rando, O.J., 2016. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nat Methods 13, 1009–1011. https://doi.org/10.1038/nmeth.4025

Hu, B., Petela, N., Kurze, A., Chan, K.-L., Chapard, C., Nasmyth, K., 2015. Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. Nucleic Acids Res. 43, e132. https://doi.org/10.1093/nar/gkv670

Jeppsson, K., Sakata, T., Nakato, R., Milanova, S., Shirahige, K., Björkegren, C., 2022.

Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks. Sci. Adv. 8, eabn7063. https://doi.org/10.1126/sciadv.abn7063

Kikuchi, Y., 1983. Yeast plasmid requires a cis-acting locus and two plasmid proteins for its stable maintenance. Cell 35, 487–493. https://doi.org/10.1016/0092-8674(83)90182-4

Kim, K.-D., Tanizawa, H., De Leo, A., Vladimirova, O., Kossenkov, A., Lu, F., Showe, L.C., Noma, K., Lieberman, P.M., 2020. Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. Nat. Commun. 11, 877. https://doi.org/10.1038/s41467-019-14152-8

Kruitwagen, T., Chymkowitch, P., Denoth-Lippuner, A., Enserink, J., Barral, Y., 2018. Centromeres License the Mitotic Condensation of Yeast Chromosome Arms. Cell 175, 780-795.e15. https://doi.org/10.1016/j.cell.2018.09.012

Kumar, D., Prajapati, H.K., Mahilkar, A., Ma, C.-H., Mittal, P., Jayaram, M., Ghosh, S.K., 2021. The selfish yeast plasmid utilizes the condensin complex and condensed chromatin for faithful partitioning. PLOS Genet. 17, e1009660. https://doi.org/10.1371/journal.pgen.1009660

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lartigue, C., Vashee, S., Algire, M.A., Chuang, R.-Y., Benders, G.A., Ma, L., Noskov, V.N., Denisova, E.A., Gibson, D.G., Assad-Garcia, N., Alperovich, N., Thomas, D.W., Merryman, C., Hutchison, C.A., Smith, H.O., Venter, J.C., Glass, J.I., 2009. Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast. Science 325, 1693–1696. https://doi.org/10.1126/science.1173759

Lazar-Stefanita, L., Scolari, V.F., Mercy, G., Muller, H., Guérin, T.M., Thierry, A., Mozziconacci, J., Koszul, R., 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. EMBO J. https://doi.org/10.15252/embj.201797342

Lee, S., Oh, S., Jeong, K., Jo, H., Choi, Y., Seo, H.D., Kim, M., Choe, J., Kwon, C.S., Lee, D., 2018. Dot1 regulates nucleosome dynamics by its inherent histone chaperone activity in yeast. Nat. Commun. 9, 240. https://doi.org/10.1038/s41467-017-02759-8

Lerner, A.M., Hepperla, A.J., Keele, G.R., Meriesh, H.A., Yumerefendi, H., Restrepo, D., Zimmerman, S., Bear, J.E., Kuhlman, B., Davis, I.J., Strahl, B.D., 2020. An optogenetic switch for the Set2 methyltransferase provides evidence for transcription-dependent and-independent dynamics of H3K36 methylation. Genome Res. 30, 1605–1617. https://doi.org/10.1101/gr.264283.120

Li, B., Gogol, M., Carey, M., Pattenden, S.G., Seidel, C., Workman, J.L., 2007. Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. Genes Dev. 21, 1422–1430. https://doi.org/10.1101/gad.1539307

Lieberman, P.M., 2006. Chromatin regulation of virus infection. Trends Microbiol. 14, 132–140. https://doi.org/10.1016/j.tim.2006.01.001

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., others, 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science 326, 289–293.

Loïodice, I., Garnier, M., Nikolov, I., Taddei, A., 2021. An Inducible System for Silencing Establishment Reveals a Stepwise Mechanism in Which Anchoring at the Nuclear Periphery Precedes Heterochromatin Formation. Cells 10, 2810. https://doi.org/10.3390/cells10112810

Matthey-Doret, C., Baudry, L., Breuer, A., Montagne, R., Guiglielmoni, N., Scolari, V., Jean, E., Campeas, A., Chanut, P.H., Oriol, E., Méot, A., Politis, L., Vigouroux, A., Moreau, P., Koszul, R., Cournac, A., 2020. Computer vision for pattern detection in chromosome contact maps. Nat. Commun. 11, 5795. https://doi.org/10.1038/s41467-020-19562-7

Matthey-Doret, C., Baudry, L., Mortaza, S., Moreau, P., Koszul, R., Cournac, A., 2022. Normalization of Chromosome Contact Maps: Matrix Balancing and Visualization. Methods Mol. Biol. Clifton NJ 2301, 1–15. https://doi.org/10.1007/978-1-0716-1390-0_1

McKnight, J.N., Boerma, J.W., Breeden, L.L., Tsukiyama, T., 2015. Global Promoter Targeting of a Conserved Lysine Deacetylase for Transcriptional Shutoff during Quiescence Entry. Mol. Cell 59, 732–743. https://doi.org/10.1016/j.molcel.2015.07.014

McQuaid, M.E., Polvi, E.J., Dobson, M.J., 2019. DNA sequence elements required for partitioning competence of the Saccharomyces cerevisiae 2-micron plasmid STB locus. Nucleic Acids Res. 47, 716–728. https://doi.org/10.1093/nar/gky1150

Mead, D.J., Gardner, D.C.J., Oliver, S.G., 1986. The yeast 2 µ plasmid: strategies for the survival of a selfish DNA. Mol. Gen. Genet. MGG 205, 417–421. https://doi.org/10.1007/BF00338076

Mehta, S., Yang, X.M., Chan, C.S., Dobson, M.J., Jayaram, M., Velmurugan, S., 2002. The 2 micron plasmid purloins the yeast cohesin complex. J. Cell Biol. 158, 625–637. https://doi.org/10.1083/jcb.200204136

Mereshchuk, A., Johnstone, P.S., Chew, J.S.K., Dobson, M.J., 2022. The yeast 2-micron plasmid Rep2 protein has Rep1-independent partitioning function. Nucleic Acids Res. 50, 10571–10585. https://doi.org/10.1093/nar/gkac810

Møller, H.D., Parsons, L., Jørgensen, T.S., Botstein, D., Regenberg, B., 2015.Extrachromosomal circular DNA is common in yeast. Proc. Natl. Acad. Sci. 112, E3114–E3122. https://doi.org/10.1073/pnas.1508825112

Muller, H., Scolari, V., Mercy, G., Agier, N., Descorps-Declere, S., Fischer, G., Mozziconacci, J., Koszul, R., 2017. Redesigning chromosomes to optimize conformation capture (Hi-C) assays. bioRxiv 169847.

Ohno, M., Ando, T., Priest, D.G., Kumar, V., Yoshida, Y., Taniguchi, Y., 2019. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. Cell 176, 520-534.e25. https://doi.org/10.1016/j.cell.2018.12.014

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Piazza, A., Bordelet, H., Dumont, A., Thierry, A., Savocco, J., Girard, F., Koszul, R., 2021. Cohesin regulates homology search during recombinational DNA repair. Nat. Cell Biol. 23, 1176–1186. https://doi.org/10.1038/s41556-021-00783-x

Prajapati, H.K., Rizvi, S.M.A., Rathore, I., Ghosh, S.K., 2017. Microtubule-associated proteins, Bik1 and Bim1, are required for faithful partitioning of the endogenous 2 micron plasmids in budding yeast. Mol. Microbiol. 103, 1046–1064. https://doi.org/10.1111/mmi.13608

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160-165. https://doi.org/10.1093/nar/gkw257

Rieben, W.K., Gonzales, C.M., Gonzales, S.T., Pilkington, K.J., Kiyosawa, H., Hughes, J.E., Welker, D.L., 1998. Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1 and Ddp2 plasmid families. Genetics 148, 1117–1125. https://doi.org/10.1093/genetics/148.3.1117

Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R., Mairose, J.D., Basting, A.V., Mistretta, K.S., Rocco, D.J., Perkinson, E.S., Kellogg, G.D., Mahony, S., Pugh, B.F., 2021. A high-resolution protein architecture of the budding yeast genome. Nature 592, 309–314. https://doi.org/10.1038/s41586-021-03314-8

Ruault, M., Scolari, V.F., Lazar-Stefanita, L., Hocher, A., Loïodice, I., Koszul, R., Taddei, A., 2021. Sir3 mediates long-range chromosome interactions in budding yeast. Genome Res. https://doi.org/10.1101/gr.267872.120

Sau, S., Ghosh, S.K., Liu, Y.-T., Ma, C.-H., Jayaram, M., 2019. Hitchhiking on chromosomes: A persistence strategy shared by diverse selfish DNA elements. Plasmid 102, 19–28. https://doi.org/10.1016/j.plasmid.2019.01.004

Schalbetter, S.A., Fudenberg, G., Baxter, J., Pollard, K.S., Neale, M.J., 2019. Principles of meiotic chromosome assembly revealed in S. cerevisiae. Nat Commun 10, 4795. https://doi.org/10.1038/s41467-019-12629-0

Sengupta, A., Blomqvist, K., Pickett, A.J., Zhang, Y., Chew, J.S.K., Dobson, M.J., 2001. Functional Domains of Yeast Plasmid-Encoded Rep Proteins. J. Bacteriol. 183, 2306–2315. https://doi.org/10.1128/JB.183.7.2306-2315.2001

Shammat, I.M., Gonzales, C.M., Welker, D.L., 1998. Dictyostelium discoideum nuclear plasmid Ddp6 is a new member of the Ddp2 plasmid family. Curr. Genet. 33, 77–82. https://doi.org/10.1007/s002940050311

Stevens, B.J., Moustacchi, E., 1971. ADN satellite γ et molécules circulaires torsadées de petite taille chez la levure Saccharomyces cerevisiae. Exp. Cell Res. 64, 259–266. https://doi.org/10.1016/0014-4827(71)90075-9

Swygert, S.G., Kim, S., Wu, X., Fu, T., Hsieh, T.-H., Rando, O.J., Eisenman, R.N., Shendure, J., McKnight, J.N., Tsukiyama, T., 2019. Condensin-Dependent Chromatin Compaction Represses Transcription Globally during Quiescence. Mol. Cell 73, 533-546.e4. https://doi.org/10.1016/j.molcel.2018.11.020

Swygert, S.G., Lin, D., Portillo-Ledesma, S., Lin, P.-Y., Hunt, D.R., Kao, C.-F., Schlick, T., Noble, W.S., Tsukiyama, T., 2021. Local chromatin fiber folding represses transcription and loop extrusion in quiescent cells. eLife 10, e72062. https://doi.org/10.7554/eLife.72062

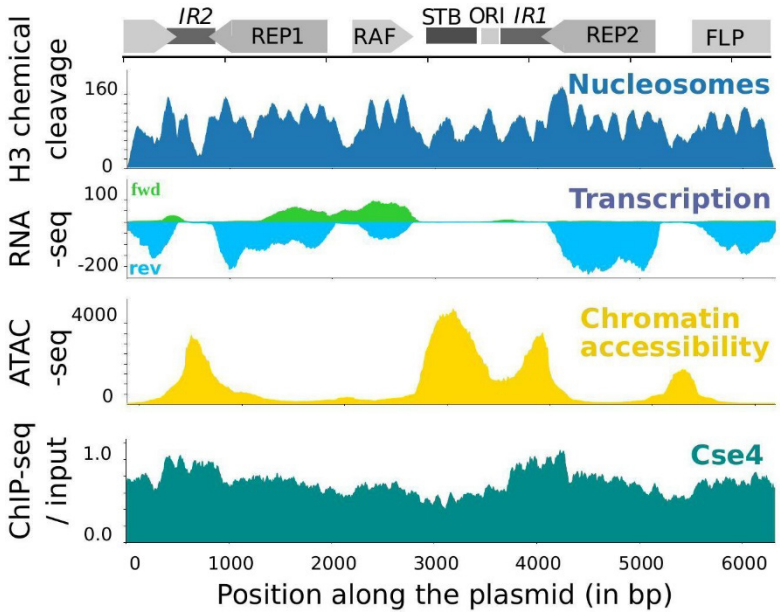
Velmurugan, S., Ahn, Y.-T., Yang, X.-M., Wu, X.-L., Jayaram, M., 1998. The 2µm Plasmid Stability System: Analyses of the Interactions among Plasmid- and Host-Encoded Components. Mol. Cell. Biol. 18, 7466–7477. https://doi.org/10.1128/MCB.18.12.7466

Wang, S.Y., Pollina, E.A., Wang, I.-H., Pino, L.K., Bushnell, H.L., Takashima, K., Fritsche, C., Sabin, G., Garcia, B.A., Greer, P.L., Greer, E.L., 2021. Role of epigenetics in unicellular to multicellular transition in Dictyostelium. Genome Biol. 22, 134. https://doi.org/10.1186/s13059-021-02360-9

Wong, M.C.V.L., Scott-Drew, S.R.S., Hayes, M.J., Howard, P.J., Murray, J.A.H., 2002. RSC2, Encoding a Component of the RSC Nucleosome Remodeling Complex, Is Essential for 2µm Plasmid Maintenance in Saccharomyces cerevisiae. Mol. Cell. Biol. 22, 4218–4229. https://doi.org/10.1128/MCB.22.12.4218-4229.2002

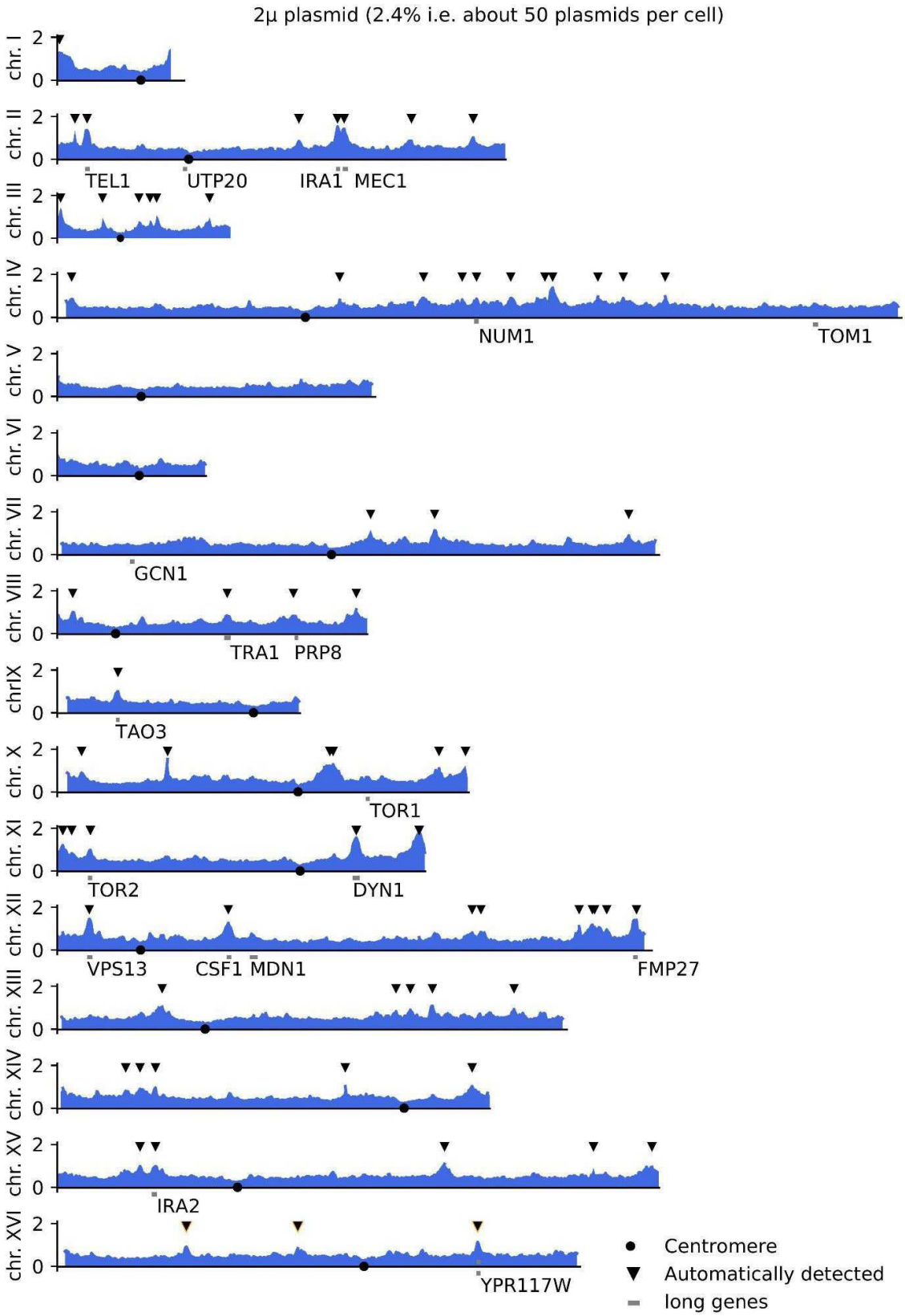
Zakian, V.A., Brewer, B.J., Fangman, W.L., 1979. Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase. Cell 17, 923–934. https://doi.org/10.1016/0092-8674(79)90332-5

2.9 Supplemental information

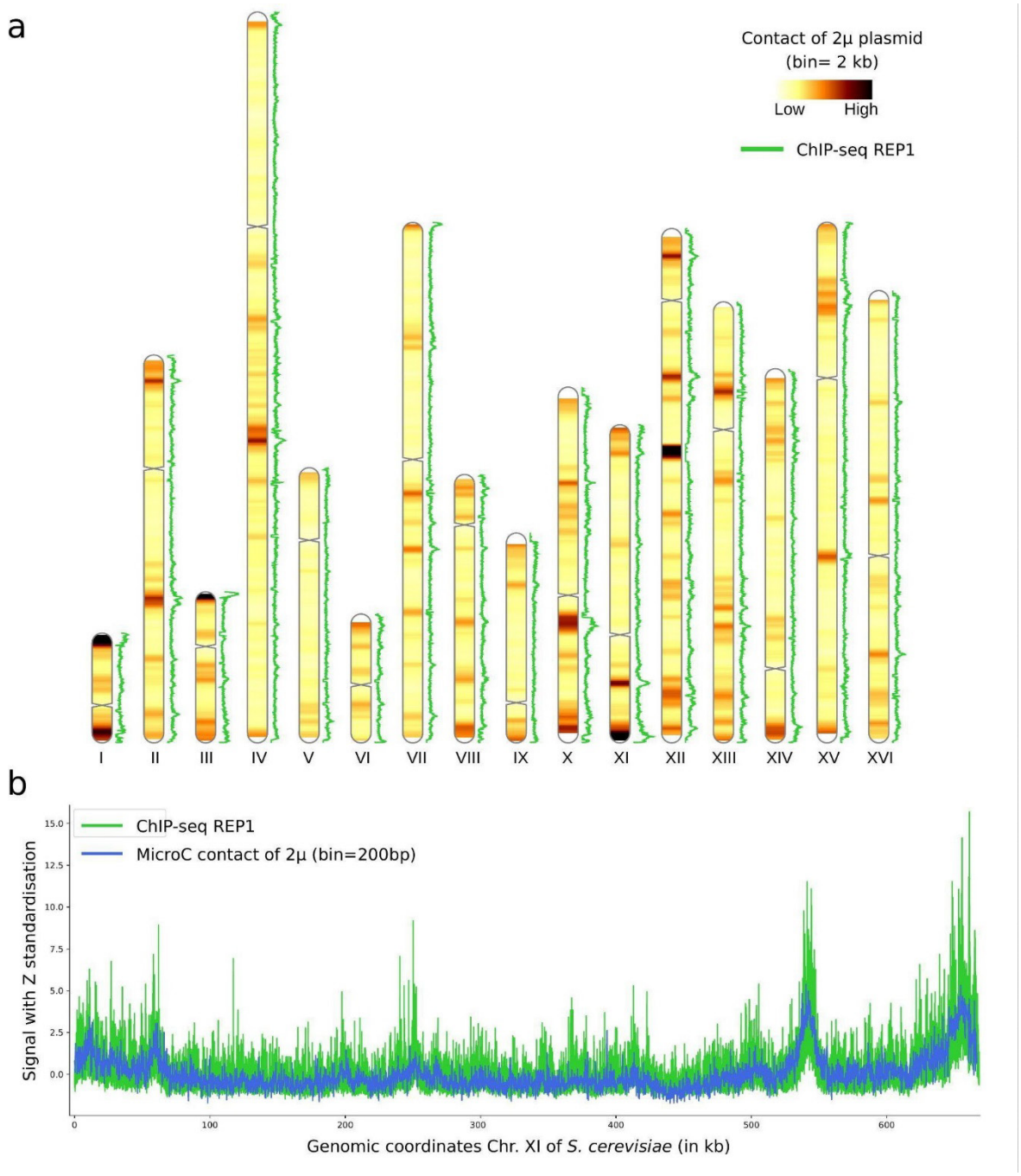


Supplementary Figure 1. Genomic signals along the 2*µ* plasmid of *Saccharomyces cerevisiae.*

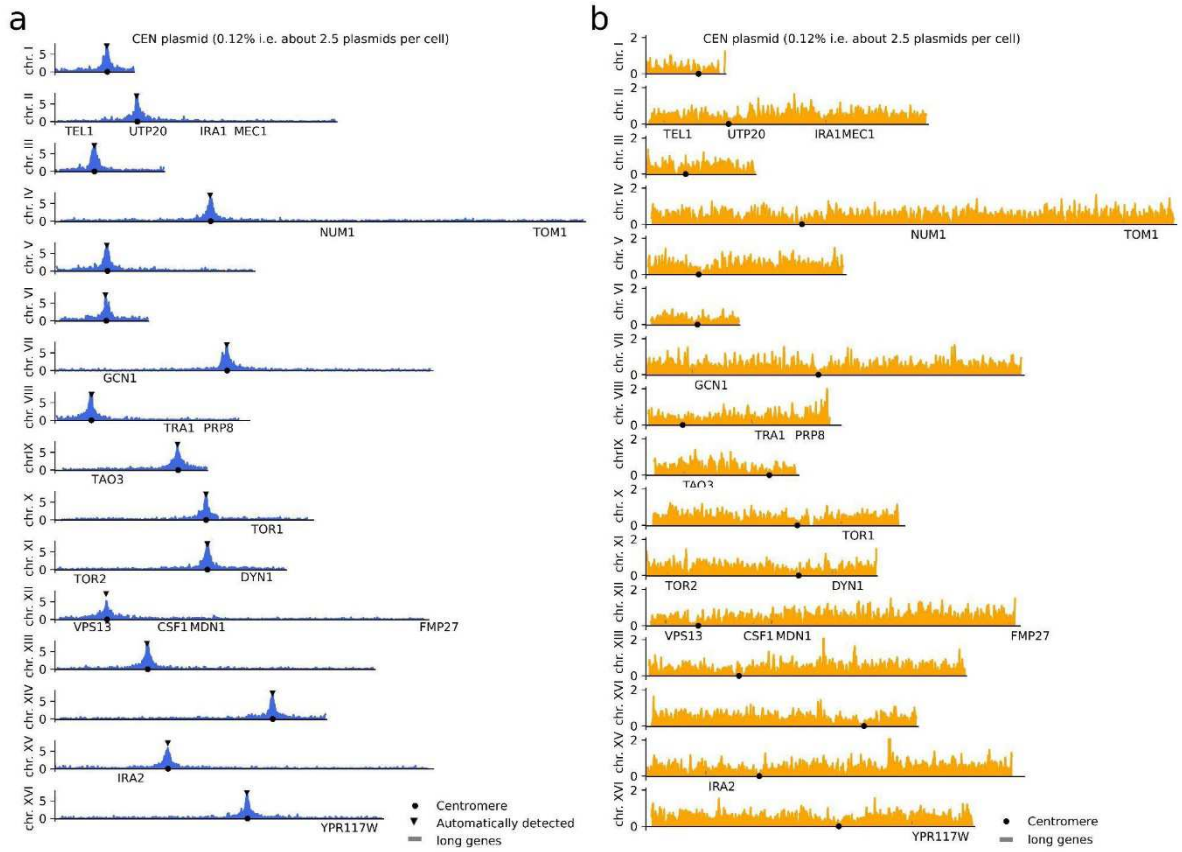
Nucleosomes signal along the 2*µ* plasmid from H3 chemical cleavage data [1]. Transcription signal along the 2*µ* plasmid from RNA-seq data of [2]. Chromatin accessibility signal along the 2*µ* plasmid from ATAC-seq data of [3]. Protein occupancy of Cse4 along the 2*µ* plasmid, from ChIP-seq data of [4].



Supplementary Figure 2. Contact signal of the 2μ plasmid along the 16 chromosomes of S. cerevisiae. **a**, The contact signal is binned at 2 kb, genes with size > 7 kb are annotated with grey rectangles and their names, (MicroC data from [5]). Automatically detected peaks of contact were annotated with black triangles.

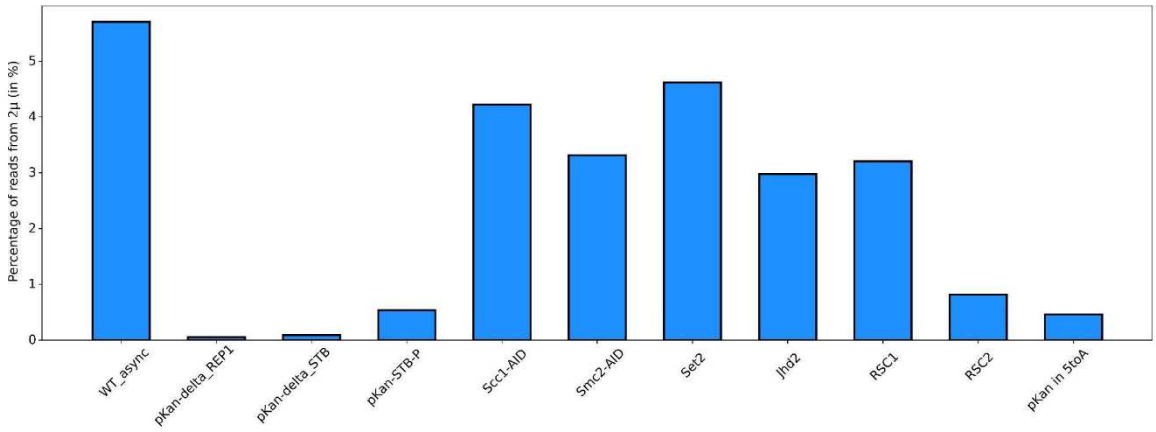


Supplementary Figure 3. ChIP-seq of REP1 protein from the 2μ plasmid. a, ChIP-seq of Rep1 protein along with the contact profile of 2μ plasmid with the chromosomes of *S. cerevisiae* (chromosomal heatmap diagram). **b**, ChIP-seq of Rep1 protein and contact profile of 2μ plasmid binned at 200 bp, (MicroC data from [5]) for the chromosome XI of *S. cerevisiae*.

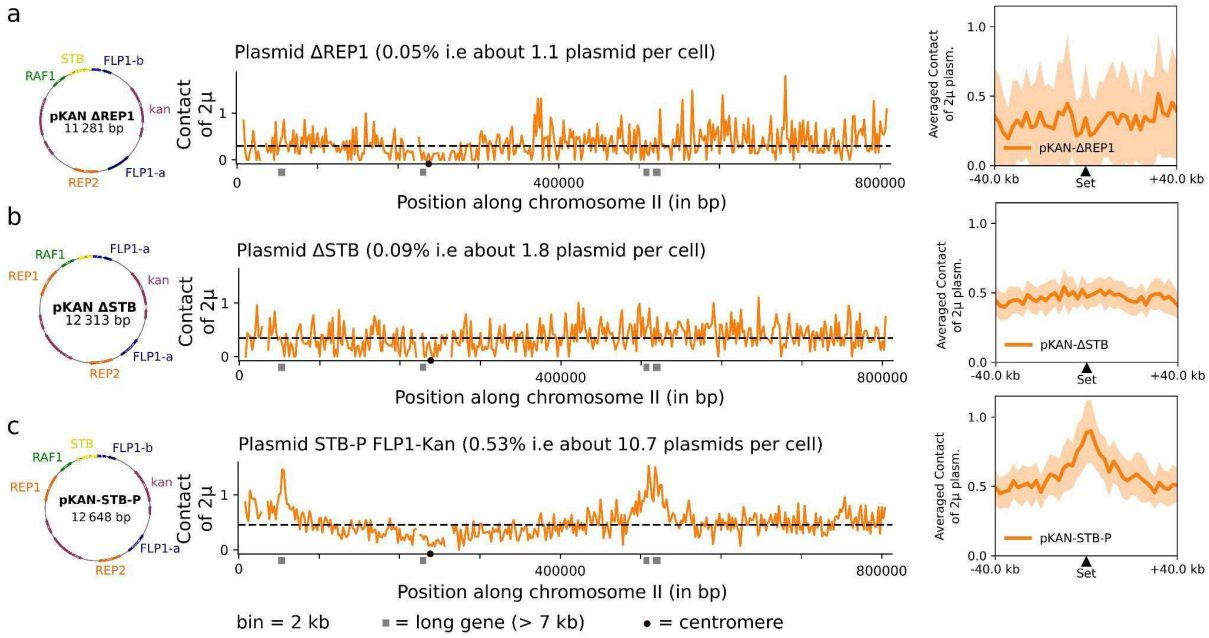


Supplementary Figure 4. Contact signal of the control plasmids along the 16 chromosomes of S. cerevisiae.

a, Contact signal of the yeast centromeric Plasmid (YCp) pRS416 along the 16 chromosomes of S. cerevisiae. The Hi-C contact signal is binned at 2 kb, names of genes with size >7 kb are annotated. Automatically detected peaks of contact were annotated with black triangles. **b**, Contact signal of a replicative plasmid devoid of centromere (pARS) and 2µ system along the 16 chromosomes of S. cerevisiae. The contact signal is binned at 2 kb, names of genes with size >7 kb are annotated.

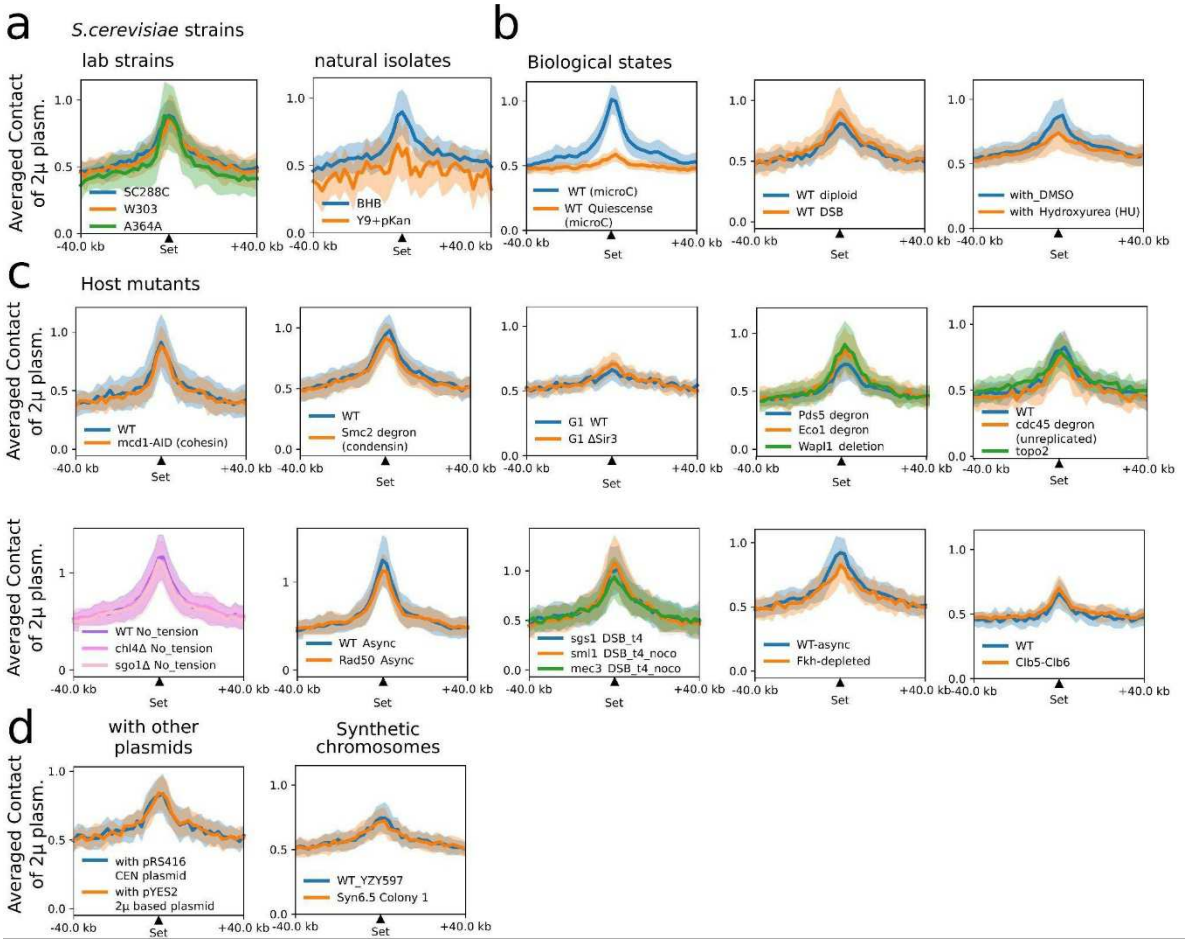


Supplementary Figure 5. Percentage of reads coming from 2µ plasmid sequence in WT and various mutants.



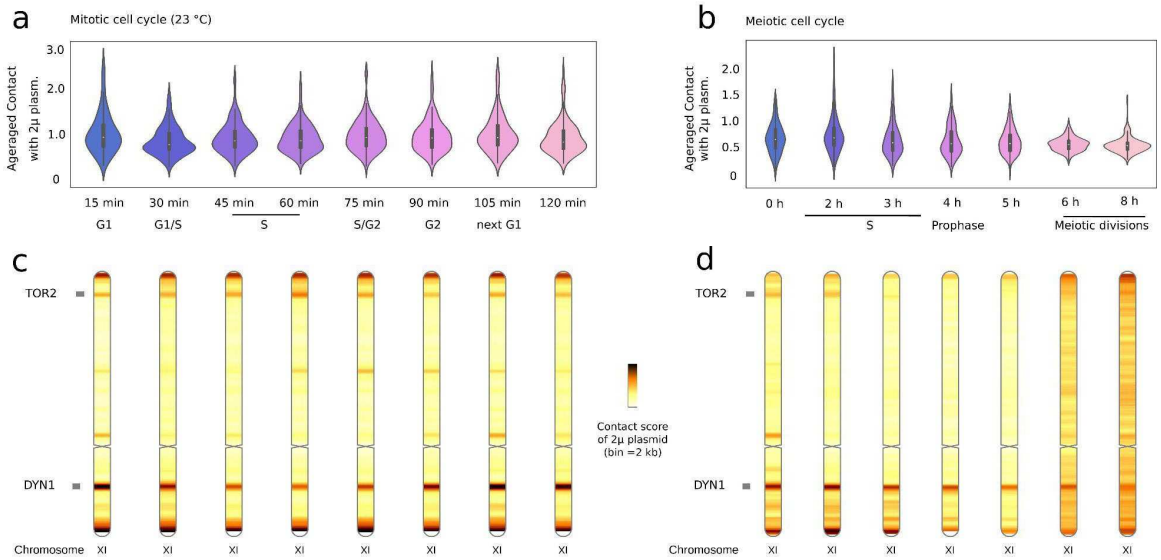
Supplementary Figure 6. Contact signal of 2*µ* plasmid mutants.

a, contact signal of the ΔREP1 mutant 2*µ* plasmid along the chromosome II of *S. cerevisiae* and the averaged contact signal on the hot spots of contact detected in WT, log phase condition. **b**, Same for the ΔSTB mutant 2*µ* plasmid. **c**, Same for the STB-P 2*µ* mutant plasmid.



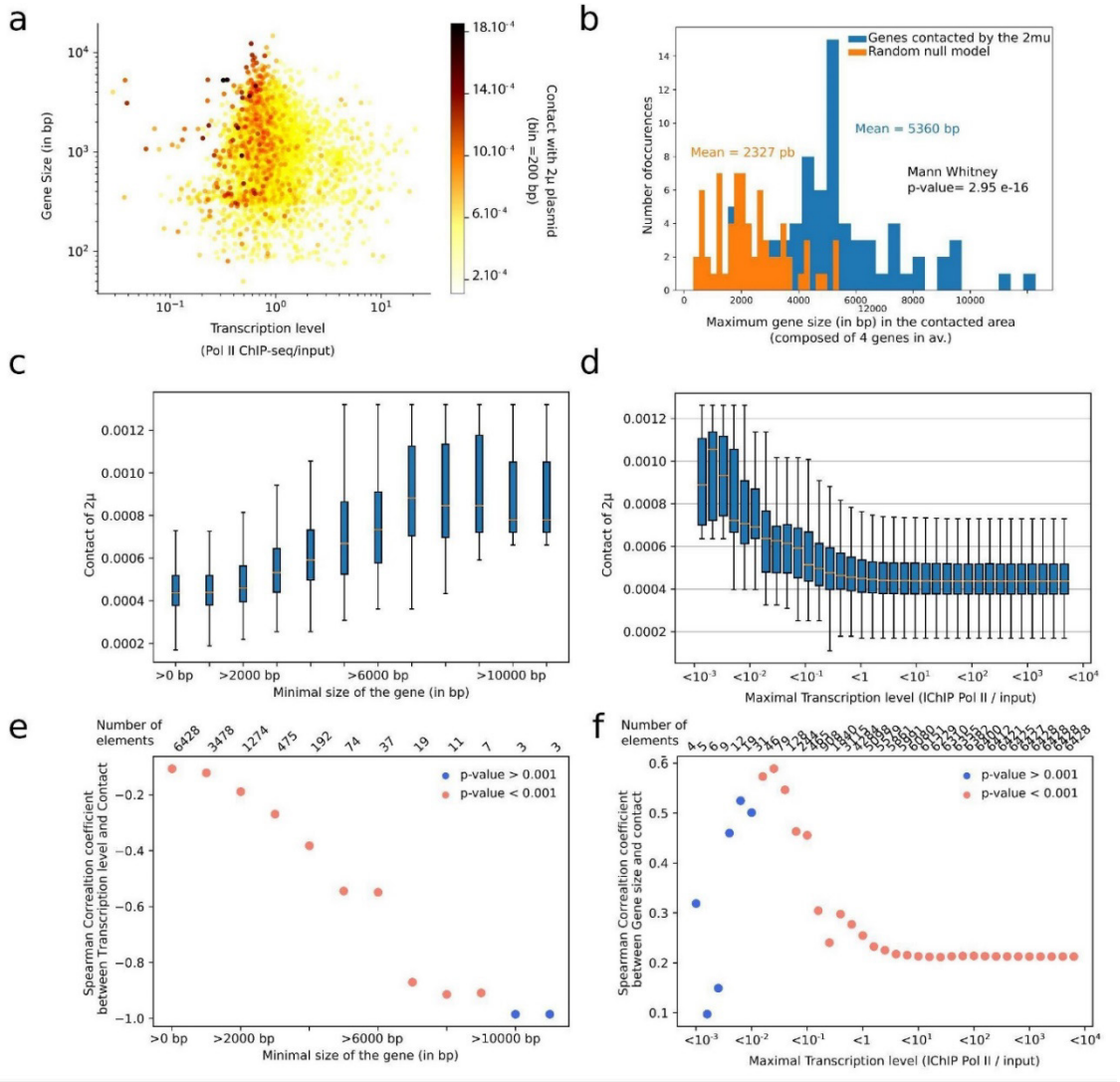
Supplementary Figure 7. The specific positioning of the 2μ plasmid is conserved under a wide variety of biological conditions and mutants.

a, Averaged 2μ plasmid contact signal over the hotspots of contact identified in WT, log phase in different lab strains of *S. cerevisiae*: SC288C, W303 [6], A364A [7] and strains from natural isolates [8]. **b**, Averaged 2μ plasmid contact signal over the hotspots of contact identified in WT, log phase in different biological states: in quiescence [9], in diploid stage, with double strand break of DNA [10], with DMSO or HU treatment [11]. **c**, Averaged 2μ plasmid contact signal over the hotspots of contact identified in WT, log phase in different mutants of *S. cerevisiae*: Mcd1 depleted (sub-unit of cohesin) [7], Smc2 depleted (sub-unit of condensin) [12], ΔSir3 [9], Pds5, EcoI, Wapl [6], in cdc-45 degron mutant (stopped replication) [6], ΔTOP2 [13], in condition with no tension of microtubules (nocodazole treatment), ΔChl4, ΔSgo1 [14], ΔRad50 [15], ΔSgs1, ΔSml1, ΔMec3 [10], in Fkh-depleted mutant [16], in Clb5-Clb6 mutant [17]. **d**, Averaged 2μ plasmid contact signal over the hotspots of contact identified in WT, log phase in presence of other plasmids (centromeric and 2µ based) and with artificial chromosomes [18].



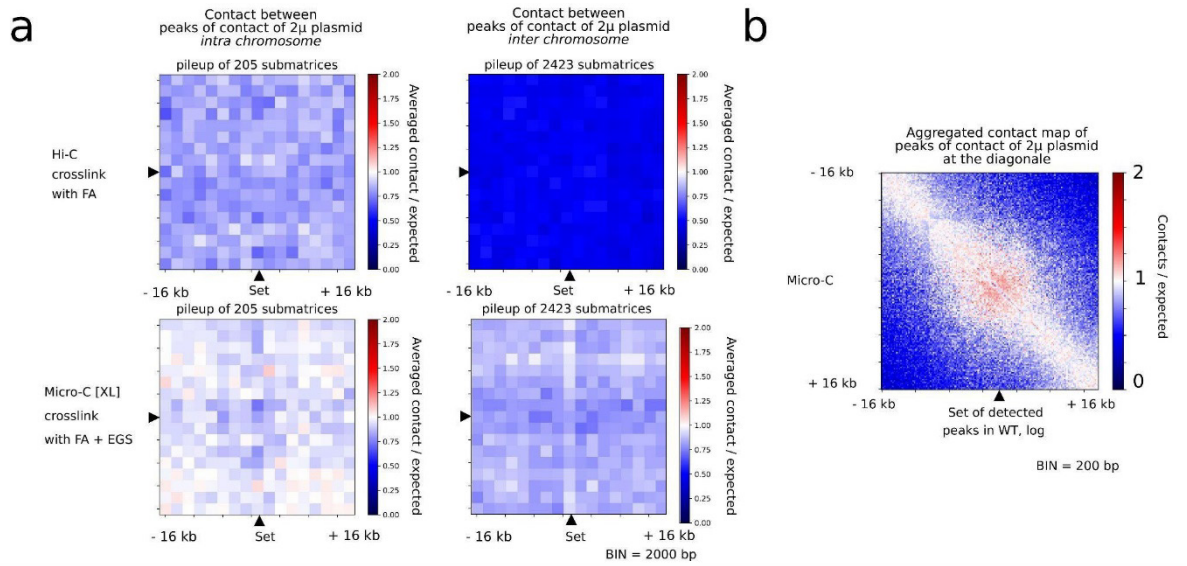
Supplementary Figure 8. Contact of the 2μ plasmid during mitotic and meiotic cell cycles.

a, Distribution of contact values of identified hotspots contacted by the 2μ plasmid identified in WT, log phase during the mitotic cell cycle (contact data reanalysed from [7]). **b**, Distribution of contact values of the identified hotspots contacted by the 2μ plasmid during the meiotic cell cycle (contact data reanalysed from [19]). **c**, Example of contact profile of 2μ plasmid with chromosome XI during mitotic cell cycle. **d**, Example of contact profile of 2μ plasmid with chromosome XI during meiotic cell cycle. Long genes (size > 7 kb) are annotated.



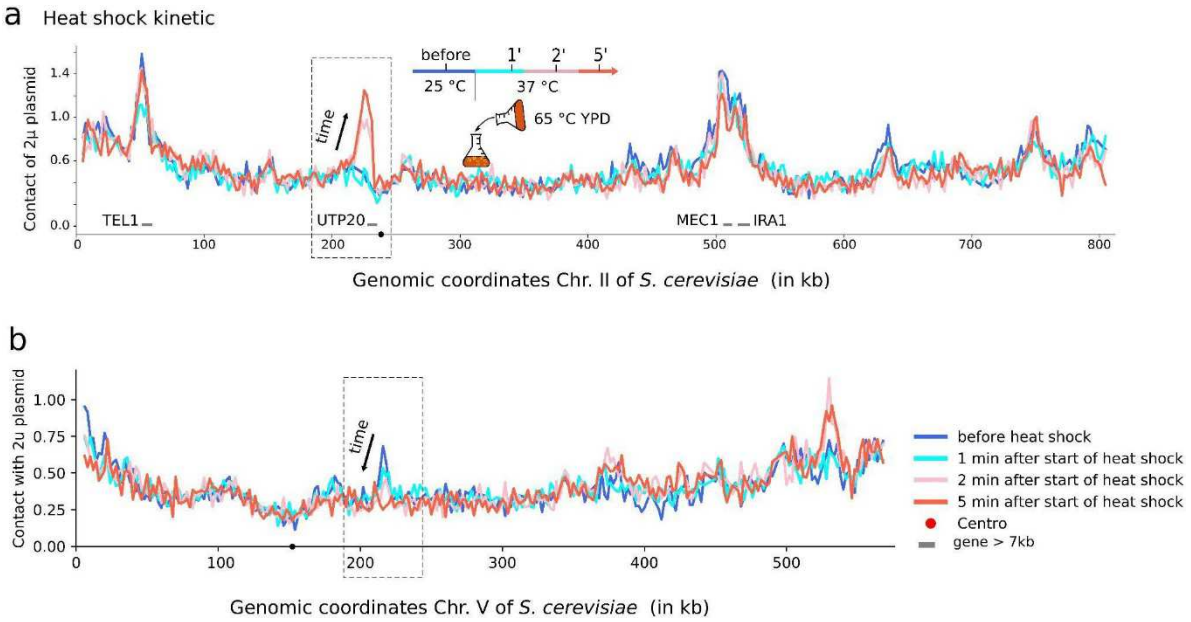
Supplementary Figure 9. Statistical analyses on the size and transcription level of genes contacted by 2μ plasmid.

a Scatter plot for all genes of *S. cerevisiae* represented in function of their transcription level (x-axis), their size in bp (y-axis) and their level of contact with 2μ plasmid represented by their colour (colour bar on the left). MicroC data were reanalysed from [5]. **b**, Distribution of maximum sizes of genes from loci contacted by 2μ plasmid and from a random group of loci with the associated statistical test. **c**, Contact level of 2μ plasmid in function of the minimal size of gene (in bp). **d**, Contact level of 2μ plasmid in function of the maximal transcription level (ChIP-seq data of Rpb3, sub-unit of PolII, [5]). **e**, Spearman correlation coefficient between transcription level and contact with 2μ plasmid in function of the minimal size of gene (in bp). **f**, Spearman correlation coefficient between gene size and contact with 2μ plasmid in function of the maximal transcription level (ChIP-seq data of Rpb3, sub-unit of PolII from [5]).



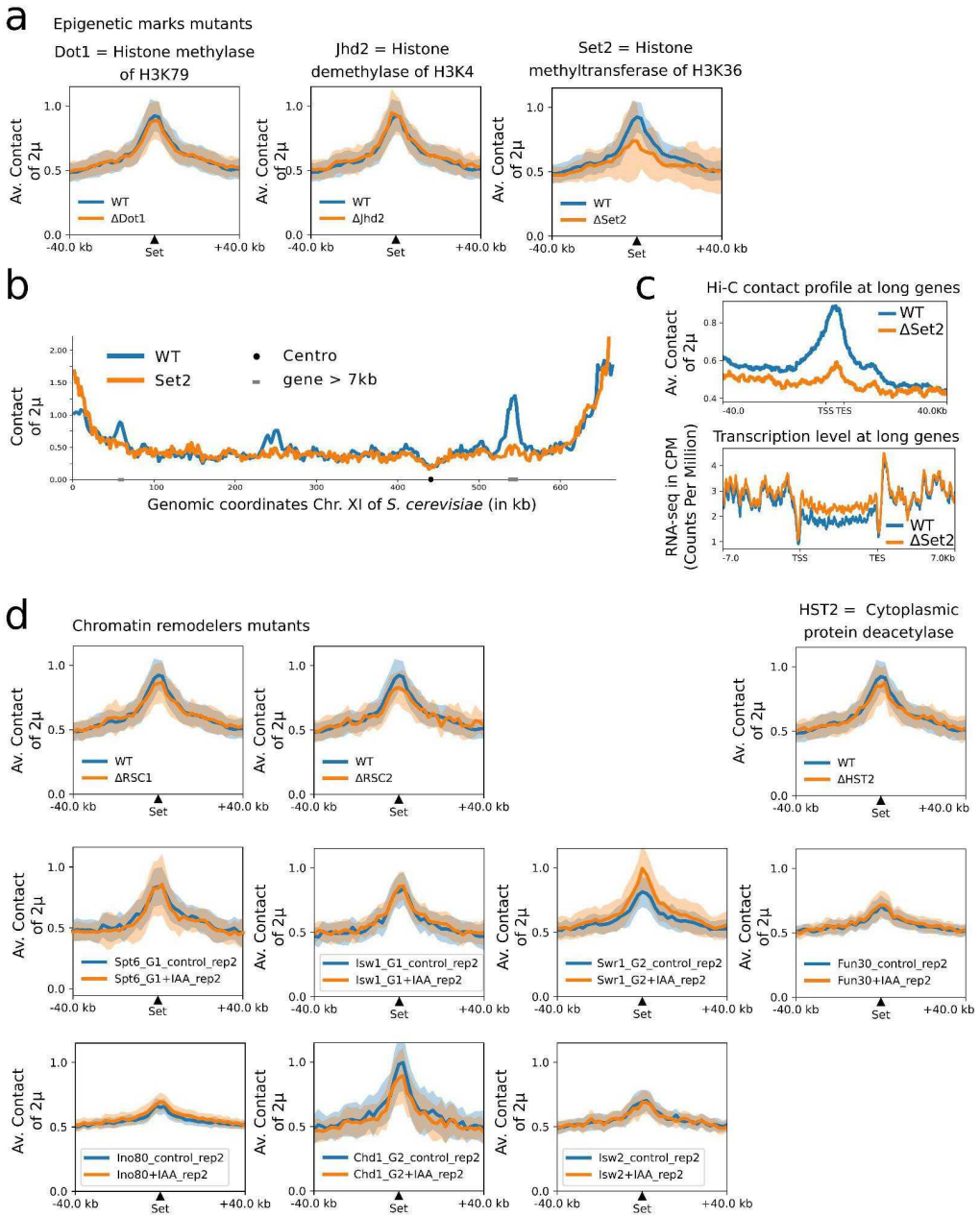
Supplementary Figure 10. Contact behaviour for the identified loci contacted by 2μ plasmid.

a, Agglomerated plot between pairs of loci contacted by 2μ plasmid belonging to the same chromosome (left) or belonging to different chromosomes (right) for two different contact technologies: Hi-C (top) and MicroC with dual crosslink (bottom) [5]. The signal represents the ratio between the contact measured between loci contacted by 2μ plasmid over random pairs separated by same genomic distances [20]. **b**, Agglomerated plot at the diagonal for the identified loci contacted by 2μ plasmid with bins of 200 bp (MicroC data reanalysed from [5]).



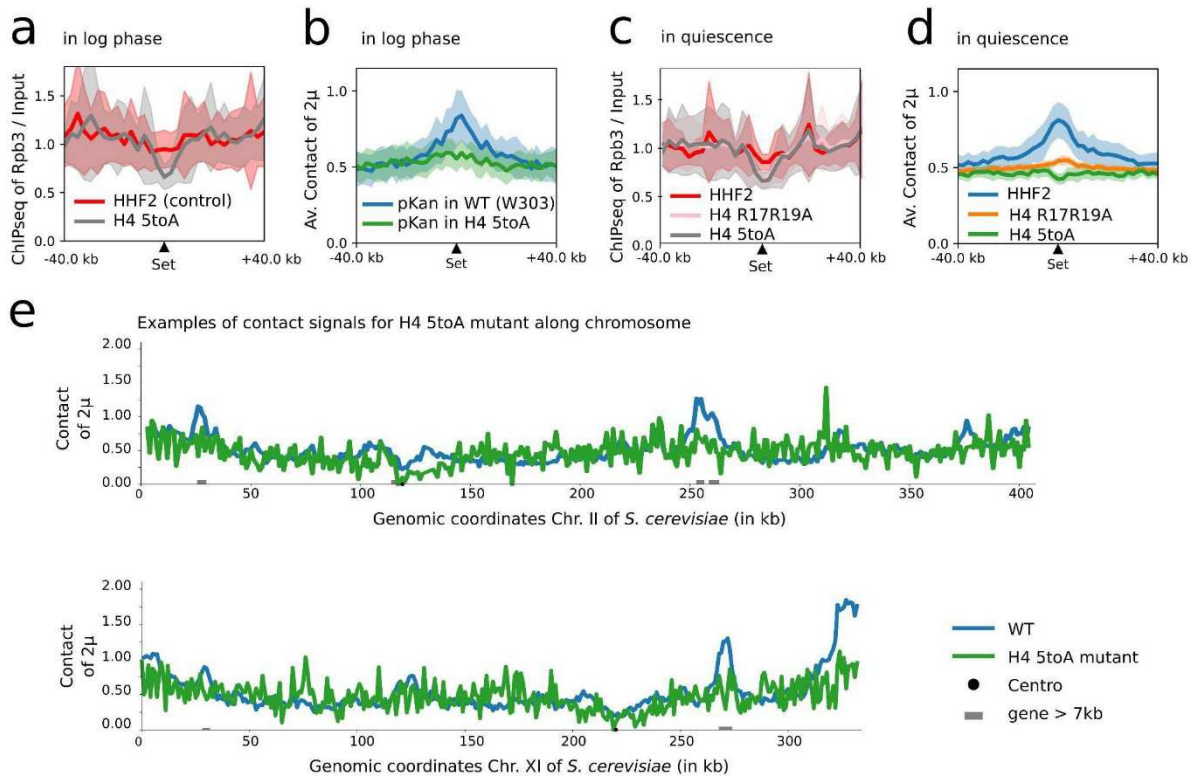
Supplementary Figure 11. Contact signal of 2*µ* plasmid during a heat shock.

a, Contact signal of 2*µ* plasmid along the chromosome II of *S. cerevisiae* for 4 time points: before heat shock, 1 min, 2 min and 5 min after heat shock. **b**, Contact signal of 2*µ* plasmid along the chromosome V of *S. cerevisiae* for 4 time points: before heat shock, 1 min, 2 min and 5 min after heat shock. Binning for the contact signals is 2 kb.



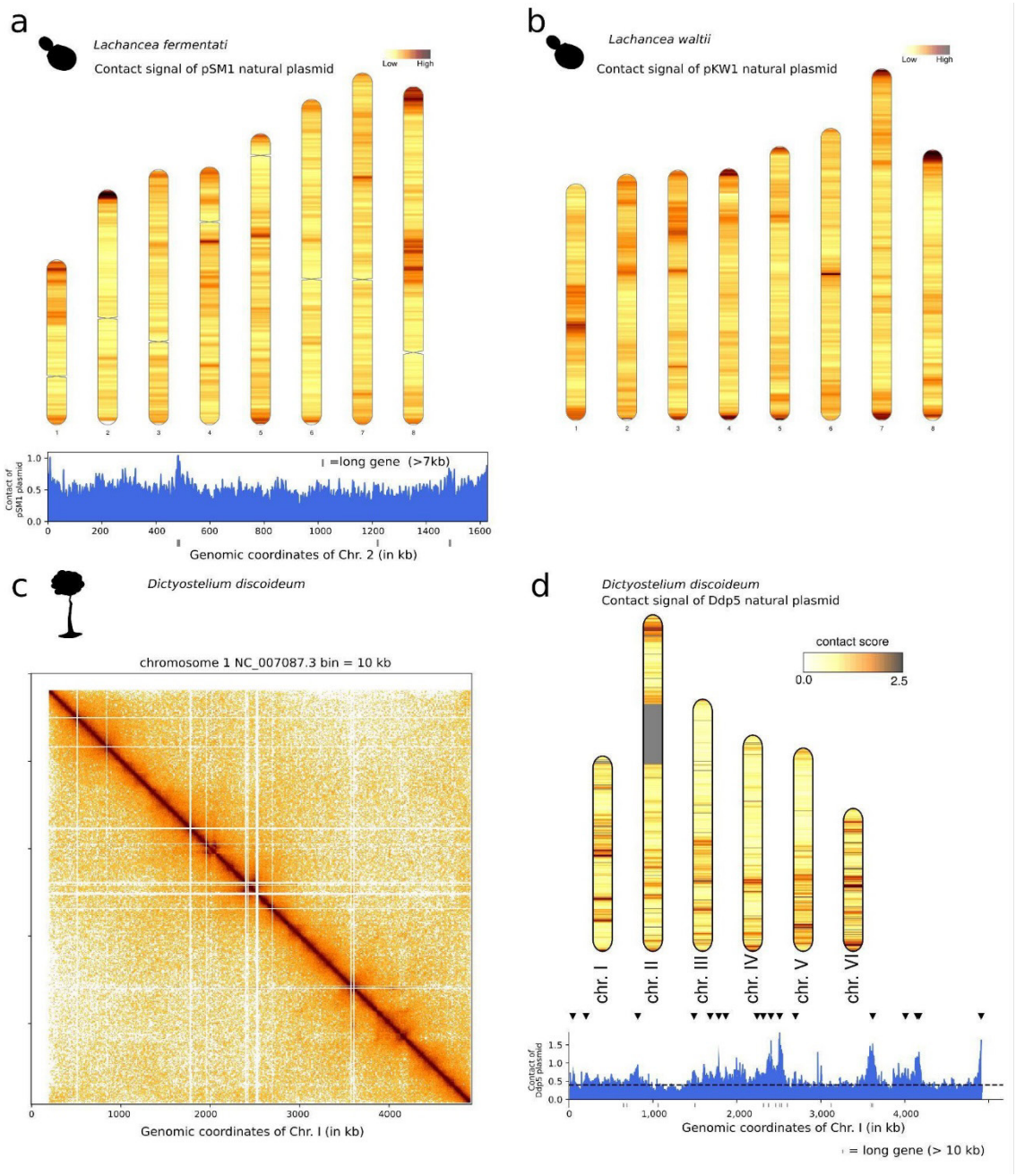
Supplementary Figure 12. Contact signal of 2*µ* plasmid in epigenetic marks and chromatin remodeler mutants

a, Averaged 2*µ plasmid* contact signal over the set of identified loci contacted by 2*µ* plasmid in WT, log phase condition for ΔDot1, ΔJhd2 and ΔSet2 mutants as well as for the ΔHST2 mutant. **b**, Contact signals of 2*µ* plasmid along chromosomes VII and XII for WT and ΔSet2 mutant. **c**, Average contact signal (top) and average transcription level (below) at long genes for WT and ΔSet2 mutant. **d**, Averaged 2*µ* plasmid contact signal over the set of genomic positions identified in WT, log phase condition for ΔRSC1, ΔRSC2 as well as for 7 chromatin remodelers degradation mutants (AID system) and their corresponding control: Spt6, Isw1, Swr1, Fun30, Ino80, Chd1, Isw2 (data from [21]).



Supplementary Figure 13. Contact signal of 2*µ* plasmid in the H4 5toA mutant.

a, Averaged transcription signal measured by Rpb3 (PolII sub-unit) ChIP-seq [22] over the set of identified loci contacted by 2*µ* plasmid in WT, log phase condition for the control and H4 5toA mutant in log phase. **b**, Averaged 2*µ* plasmid contact signal over the set of identified loci contacted by 2*µ* plasmid in WT, log phase condition for the control and H4 5toA mutant in log phase. pKAN version of the 2*µ* plasmid was used to ensure plasmid stability. **c**, Averaged transcription signal measured by Rpb3 (PolII sub-unit) ChIP-seq [22] over the set of identified loci contacted by 2*µ* plasmid in WT, log phase condition for the control, H4 R17R19A and H4 5toA mutants in quiescence phase [22]. **d**, Averaged 2*µ* plasmid contact signal over the set of identified loci contacted by 2*µ* plasmid in WT, log phase condition for the control, H4 R17R19A and H4 5toA mutant in quiescence phase [22].**e,** Examples of contact signals of 2*µ* plasmid (pKAN version) in WT and H4 5toA mutant.



Supplementary Figure 14. Contact signals of natural plasmids in other eukaryotes.

a, Contact map of chromosome 1 of *Dictyostelium discoideum* amoeba binned at 10 kb resolution. **b.** Contact signal of the Ddp5 natural plasmid with the chromosomes of *Dictyostelium discoideum* and contact profile for the chromosome 1. Orange triangles indicate automatically detected peaks and long genes are annotated as grey boxes. **c**, Contact signal of the pSM1 natural plasmid with the chromosomes of *Lachancea fermentati* yeast and contact profile for chromosome 2. Long genes are annotated as grey boxes. **d**, Contact signal of the pKW1 natural plasmid with the chromosomes of *Lachancea waltii* yeast.

Experiment type and condition	Figure	Reference	SRA accession number
Micro-C, WT log and quiescence phases	**Sup Fig 1,2,3**	[5]	SRR7939017 SRR7939018
Hi-C, WT W303, G1 arrest	**Sup Fig 7**	[10]	SRR12284705
MicroC, WT A364, asynchronous	**Sup Fig 7**	[7]	SRR11893084
Hi-C, WT diploid heterozygous	**Sup Fig 7**	[10]	SRR12284704
Hi-C, WT DSB t2 replicate2	**Sup Fig 7**	[10]	SRR12284704
Hi-C, WT, with DMSO	**Sup Fig 7**	[11]	SRR13147965
Hi-C, WT, with HU	**Sup Fig 7**	[11]	SRR13147975
Micro-C, Mcd1-AID (rep1) and control	**Fig1, Sup Fig 7**	[7]	SRR11893086 SRR11893085
Hi-C, Smc2-AID and control	**Fig1, Sup Fig 7**	[12]	SRR9040342 SRR9040345
Hi-C, Sir3-AID and control	**Sup Fig 7**	[9]	SRR12108219 SRR12108218
Hi-C, Pds5-degron, Eco1-degron, Wapl-degron	**Sup Fig 7**	[6]	SRR10687277 SRR10687278
Hi-C, cdc45-AID and control (alpha-factor G1)	**Sup Fig 7**	[6]	SRR10687274 SRR8769554
Hi-C, Top2	**Sup Fig 7**	[13]	
Hi-C, mutants Chl4Δ, sgo1Δ, nocodazole treatment	**Sup Fig 7**	[14]	SRR8718857, SRR8718853

MicroC, Rad50 mutant and control	**Sup Fig 7**	[15]	. SRR11489731 SRR11489732
Hi-C, ΔSgs1 DSB t4, ΔSml1 DSB t4 noco, ΔMec3 DSB t4 noco	**Sup Fig 7**	[10]	SRR12284714 SRR12284727 SRR13736493
Hi-C, Fkh mutant and control	**Sup Fig 7**	[16]	SRR5337954 SRR5337951
Hi-C, Clb5-Clb6 mutant and control	**Sup Fig 7**	[17]	SRR17873477 SRR17873476
Hi-C, Syn6.5 strain and control	**Sup Fig 7**	[23]	SRR22910279 SRR22910280
Micro-C, mitotic cell cycle	**Sup Fig 8**	[7]	SRR11893095 to SRR11893114
Hi-C, meiosis (t=0h,3h,4h, 6h)	**Sup Fig 8**	[24]	SRR7126297 SRR7126293 SRR7126301 SRR7340033
Hi-C, meiosis cell cycle, WT	**Sup Fig 8**	[19]	SRR8689946 to SRR8689952
Hi-C, chromatin remodelers mutants (Spt6, Isw1, Swr1, Fun30, Ino80, Chd1, Isw2 and controls)	**Sup Fig 12**	[21]	GSE158336

Supplementary Table 1: Contact datasets analysed in the present study.

The last column indicates either the identifier for the raw reads available on the Short Read Archive server (SRA) (https://www.ncbi.nlm.nih.gov/sra) or on Gene Expression Omnibus server (GEO) https://www.ncbi.nlm.nih.gov/geo.

Experiment type and condition	Figure	Ref	Identifier
Pol II ChIP-seq	Fig 3	[25]	SRR1916157 SRR1916162
H3 ChIP-seq, in Log Rep1	Fig 3	[22]	SRR13736587
RNA-seq	Fig 1,2 & Sup Fig 1	[2] [26]	SRR14693235 SRR7692240
ATAC-seq	Fig 1	[27]	SRR6246290
~1300 ChIP-exo covering ~800 different proteins and genomic signals	Fig 3	[28]	GSE147927
ATAC-seq	Fig1, Sup Fig 1	[3]	SRR11235539
Cse4 ChIP-seq	Sup Fig 1	[4]	SRR10765000 SRR10764999
Scc1 ChIP-seq	Sup Fig 1	[29]	SRR1103930 SRR1103928
Brn1 ChIP-seq	Sup Fig 1	[5]	SRR7175367 SRR7175368
RNA-seq in *Dictyostelium discoideum* (vegetative stage, rep1)	Fig 4	[30]	SRR10133961?

Supplementary Table 2: Genomic datasets (other than contact data) analysed in the present study.

The last column indicates either the identifier for the raw reads available on the Short Read Archive server (SRA) (https://www.ncbi.nlm.nih.gov/sra), the identifier of the cool files accessible on the Gene Expression Omnibus server (GEO)https://www.ncbi.nlm.nih.gov/geo

Name	genotype / specie if not S. cerevisiae	Origin
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cir+	Brachmann et al. 1998
BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	Brachmann et al. 1998
BY4743	MATa/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; met15Δ0/MET15; LYS2/lys2Δ0; ura3Δ0/ura3Δ0	Brachmann et al. 1998
W303	*MAT*a {leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15}	R. Rhotstein
Y9-4	strain from Indonesia (used for Ragi fermentation, finger millet)	Peter et al.2018
BHB	strain from bakeries in Australia	Peter et al. 2018
H4 5toA	MATa RAD5+ ura3-1 hht1-hhf1::Nat hht2- hhf2::Hyg trp1-1::pRS 404-HHT2-hhf2-K16A, R17A,H18A,R19A,K20A	Swygert et al 2021
RSGY 712	W303 *M. mycoides* linear	Chapard et al. 2023
RSGY 960	W303 [cir-] pKAN	McQuaid et al.2019
RSGY 960	BY4741 *M. pneumoniae* linear	Chapard et al. 2023
RSGT 1056	BY4741 jhd2::KANMX4 cir +	This study
RSGY 1055	BY4741 set2::KANMX4 cir+	This study
RSGY 1057	BY4742 X RSGY 960 cir+	This study
RSGY 1058	BY4742 X RSGY 712 cir +	This study
RSGY 1059	BY4741 cir-	This study
RSGY 1065	Y9-4 pKAN	This study
RSGY 1068	BY4741cir- pKAN-ΔREP1	This study
RSGY 1069	BY4741cir- pKAN-ΔSTB-P	This study
RSGY 1070	BY4741 cir-pKAN	This study
RSGY 1116	BY4741 dot1 ::KANMX4	This study
RSGY 1217	BY4741 hst2::KANMX4	This study
RSGY 1218	H4 5toA pKAN	This study
RSGY 1226	BY4741 rsc1 ::KANMX4	This study
RSGY 1252	BY4741 rsc2::KANMX4	This study
RSGY 1255	BY4741 cir- pARS	This study
	Lachancea waltii	Gilles Fisher
	Lachancea fermentati	Gilles Fisher
AX4	*Dictyostelium discoideum*	DBS0237907
yAT4593	*MATalpha rap1::RAP1-GFP(ADE2) (W303 background)*	Angela Taddei
yAT4595	*MATa rap1::RAP1-GFP(ADE2) XVI-mycoides-fusion (W303 background)*	Angela Taddei

Supplementary Table 3: List of strains used in the present study.

Name	Relevant genetic features	Origin
pKAN ΔREP1	pKAN ::ΔREP1	McQuaid et al.2019
pKAN ΔSTB	pKAN ::ΔSTB-P	McQuaid et al.2019
pKAN-STB-P	REP1 REP2 STB-P FLP1:KANMX4:FLP1 IR1 IR2 RAF1	McQuaid et al.2019
pARS	pRS413 ::CEN4	This study
pRSS413	CEN4 ARS HIS3	Sikorksi et al. 1989

Supplementary Table 4: List of plasmids used in this study

References

[1] R. V. Chereji, S. Ramachandran, T. D. Bryson, et S. Henikoff, « Precise genome-wide mapping of single nucleosomes and linkers in vivo », *Genome Biol.*, vol. 19, nº 1, p. 19, févr. 2018, doi: 10.1186/s13059-018-1398-0.

[2] J. Garcia-Luis *et al.*, « FACT mediates cohesin function on chromatin », *Nat. Struct. Mol. Biol.*, vol. 26, nº 10, p. 970-979, oct. 2019, doi: 10.1038/s41594-019-0307-x.

[3] V. Sánchez-Gaya *et al.*, « Elucidating the Role of Chromatin State and Transcription Factors on the Regulation of the Yeast Metabolic Cycle: A Multi-Omic Integrative Approach », *Front. Genet.*, vol. 9, p. 578, nov. 2018, doi: 10.3389/fgene.2018.00578.

[4] W.-C. Au *et al.*, « Skp, Cullin, F-box (SCF)-Met30 and SCF-Cdc4-Mediated Proteolysis of CENP-A Prevents Mislocalization of CENP-A for Chromosomal Stability in Budding Yeast », *PLoS Genet.*, vol. 16, nº 2, p. e1008597, févr. 2020, doi: 10.1371/journal.pgen.1008597.

[5] S. G. Swygert *et al.*, « Condensin-Dependent Chromatin Compaction Represses Transcription Globally during Quiescence », *Mol. Cell*, vol. 73, nº 3, p. 533-546.e4, 07 2019, doi: 10.1016/j.molcel.2018.11.020.

[6] L. Dauban *et al.*, « Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. », *Mol Cell*, janv. 2020, doi: 10.1016/j.molcel.2020.01.019.

[7] L. Costantino, T.-H. S. Hsieh, R. Lamothe, X. Darzacq, et D. Koshland, « Cohesin residency determines chromatin loop patterns », *eLife*, vol. 9, p. e59889, nov. 2020, doi: 10.7554/eLife.59889.

[8] J. Peter *et al.*, « Genome evolution across 1,011 Saccharomyces cerevisiae isolates », *Nature*, vol. 556, nº 7701, p. 339-344, avr. 2018, doi: 10.1038/s41586-018-0030-5.

[9] M. Ruault *et al.*, « Sir3 mediates long-range chromosome interactions in budding yeast », *Genome Res.*, févr. 2021, doi: 10.1101/gr.267872.120.

[10] A. Piazza *et al.*, « Cohesin regulates homology search during recombinational DNA repair », *Nat. Cell Biol.*, vol. 23, nº 11, p. 1176-1186, nov. 2021, doi: 10.1038/s41556-021-00783-x.

[11] K. Jeppsson, T. Sakata, R. Nakato, S. Milanova, K. Shirahige, et C. Björkegren, « Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks », *Sci. Adv.*, vol. 8, nº 23, p. eabn7063, juin 2022, doi: 10.1126/sciadv.abn7063.

[12] T. M. Guérin *et al.*, « Condensin-Mediated Chromosome Folding and Internal Telomeres Drive Dicentric Severing by Cytokinesis », *Mol. Cell*, vol. 75, nº 1, p. 131-144.e3, juill. 2019, doi: 10.1016/j.molcel.2019.05.021.

[13] L. Lazar-Stefanita *et al.*, « Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. », *EMBO J.*, juill. 2017, doi: 10.15252/embj.201797342.

[14] F. Paldi *et al.*, « Convergent genes shape budding yeast pericentromeres », *Nature*, vol. 582, nº 7810, p. 119-123, juin 2020, doi: 10.1038/s41586-020-2244-6.

[15] R. Forey *et al.*, « A Role for the Mre11-Rad50-Xrs2 Complex in Gene Expression and Chromosome Organization », *Mol. Cell*, vol. 81, nº 1, p. 183-197.e6, janv. 2021, doi: 10.1016/j.molcel.2020.11.010.

[16] U. Eser *et al.*, « Form and function of topologically associating genomic domains in budding yeast », *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, nº 15, p. E3061-E3070, avr. 2017, doi: 10.1073/pnas.1612256114.

[17] R. E. Barton, L. F. Massari, D. Robertson, et A. L. Marston, « Eco1-dependent cohesin acetylation anchors chromatin loops and cohesion to define functional meiotic chromosome domains », *eLife*, vol. 11, p. e74447, févr. 2022, doi: 10.7554/eLife.74447.

[18] Z. Zhao, A. C. Anselmo, et S. Mitragotri, « Viral vector-based gene therapies in the. doi: 10.1002/btm2.10258.

[19] S. A. Schalbetter, G. Fudenberg, J. Baxter, K. S. Pollard, et M. J. Neale, « Principles of meiotic chromosome assembly revealed in S. cerevisiae. », *Nat Commun*, vol. 10, nº 1, p. 4795, 2019, doi: 10.1038/s41467-019-12629-0.

[20] C. Matthey-Doret *et al.*, « Computer vision for pattern detection in chromosome contact maps », *Nat. Commun.*, vol. 11, nº 1, Art. nº 1, nov. 2020, doi: 10.1038/s41467-020-19562-7.

[21] H. Jo, T. Kim, Y. Chun, I. Jung, et D. Lee, « A compendium of chromatin contact maps reflecting regulation by chromatin remodelers in budding yeast », *Nat. Commun.*, vol. 12, nº 1, p. 6380, nov. 2021, doi: 10.1038/s41467-021-26629-6.

[22]S. G. Swygert *et al.*, « Local chromatin fiber folding represses transcription and loop extrusion in quiescent cells », *eLife*, vol. 10, p. e72062, nov. 2021, doi:10.7554/eLife.72062.

[23] Y. Zhao *et al.*, « Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions ». bioRxiv, p. 2022.04.11.486913, 11 avril 2022. doi: 10.1101/2022.04.11.486913.

[24]H. Muller *et al.*, « Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C », *Mol. Syst. Biol.*, vol. 14, nº 7, p. e8293, 16 2018, doi: 10.15252/msb.20188293.

[25]M. Morselli *et al.*, « In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse », *eLife*, vol. 4, p. e06205, avr. 2015, doi: 10.7554/eLife.06205.

[26]N. Grosjean *et al.*, « Combined omics approaches reveal distinct responses between light and heavy rare earth elements in Saccharomyces cerevisiae », *J. Hazard. Mater.*, vol. 425, p. 127830, mars 2022, doi: 10.1016/j.jhazmat.2021.127830.

[27]S. Lee *et al.*, « Dot1 regulates nucleosome dynamics by its inherent histone chaperone activity in yeast », *Nat. Commun.*, vol. 9, nº 1, p. 240, janv. 2018, doi: 10.1038/s41467-017-02759-8.

[28]M. J. Rossi *et al.*, « A high-resolution protein architecture of the budding yeast genome », *Nature*, vol. 592, nº 7853, p. 309-314, avr. 2021, doi: 10.1038/s41586-021-03314-8.

[29]K. F. Verzijlbergen *et al.*, « Shugoshin biases chromosomes for biorientation through condensin recruitment to the pericentromere », *eLife*, vol. 3, p. e01374, févr. 2014, doi: 10.7554/eLife.01374.

[30]S. Y. Wang *et al.*, « Role of epigenetics in unicellular to multicellular transition in Dictyostelium », *Genome Biol.*, vol. 22, nº 1, p. 134, mai 2021, doi: 10.1186/s13059-021-02360-9.

3 Results II: Multi-C protocol

3.1 Introduction

3.1.1.1 Using multi-contact: The case study of DNA loops

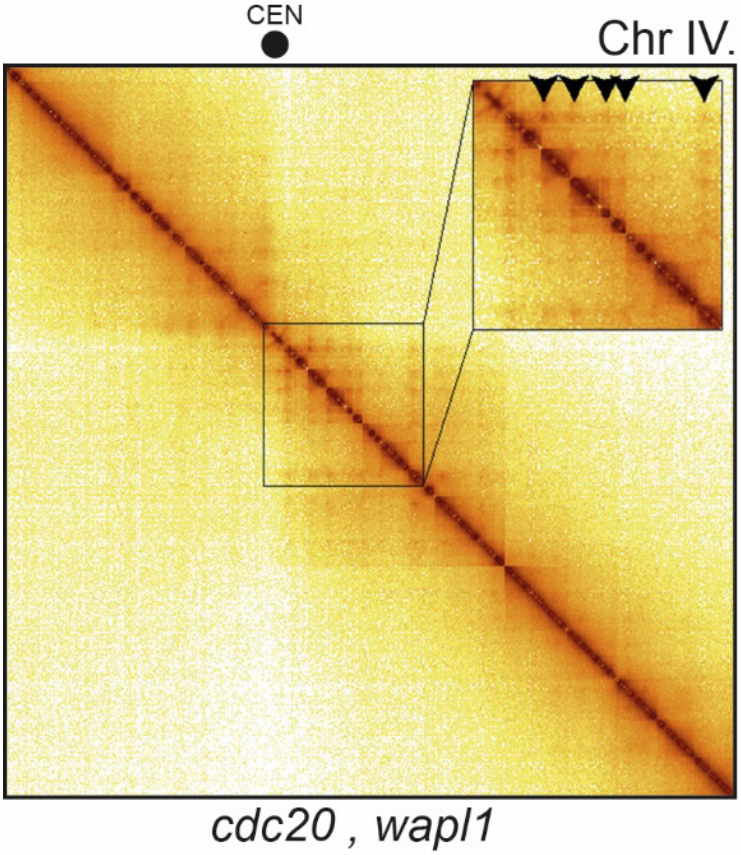


Figure 25. contact map adapted from Dauban et al. 2021 showing possible clustered loop anchors.

The canonical way to analyse 3C data is looking for contacts between two genomic loci. This is sufficient to answer many biological questions such as the genome organisation by the Structural Maintenance of Chromosome (SMC) protein family. More precisely, in G2 phase of the cell cycle in *S. cerevisiae* cohesins make loops in the genome[162]. This structure is not observed in G1 synchronised cells. On a contact map, the loops correspond to a strong enrichment of contact between two distant (10-20 kb) genomic loci, corresponding to a dot. The genomic loci that strongly interact together are name loop anchor. This pattern is so peculiar that it can be automatically detected by pattern recognition algorithm such as chromosight[163]. Sometimes, we observe pattern indicating that loop anchors from different loop interacts together (Figure 25). Several hypotheses could explain this patter. If we consider three genomic loci A, B and C. Those three loci forms loops with each other (A&B, B&C, A&C).

First, the 3C is the average of millions of nucleus organisation and the loops we observe is in fact a unique loop that involves two loci among A, B and C. Since there are several possibilities, the contact maps render all the possibility. Second, the loop anchors interact together and form a cluster together (Figure 26).

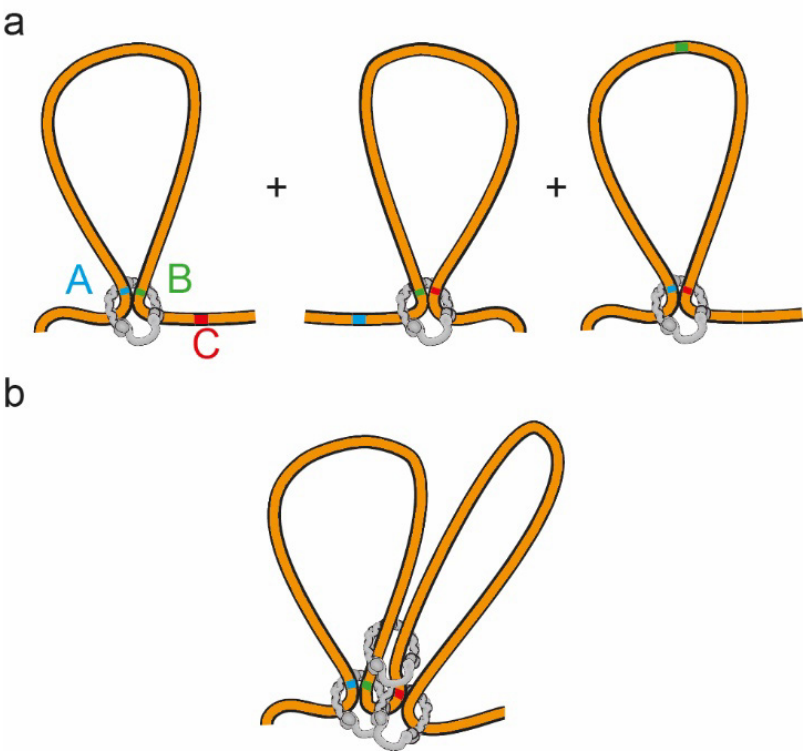


Figure 26.Drawing depicting the two possible situations that could explain the pyramidal pattern observed on the contact map of Figure 25. Graphical is inspired by Chapard, Meneu, Serizay et al.

To choose between those hypotheses, the contact between two loci is not sufficient. In the first hypothesis we would rarely find A, B and C in the same chimeric molecule while in the second hypothesis A, B and C would be in the same chimeric DNA molecule. Keeping in mind that if we observe A&B, B&C and A&C does not mean that A&B&C are together.

Several approaches exist to look at multi-contact but often are focused on the interaction between genomic loci of several Mb like Topologically Associated Domains (TADs)[164,165].

3.1.1.2 Available multi contact approach.

3.1.1.2.1 Pore-c

The Pore-C[165] approach is based on the nanopore sequencing of chimeric DNA molecules obtained after the 3C experimental procedure. There is no biotin pull down in this approach. In Desphande et al.[165] they were able to identify long range interaction between enhancers and promoters in human cell line. However, it is worth mentioning that for an experiment, the authors spend 2 PromethION sequencing flow-cell (about 1700 €) per experiment. Thus, the pore-C approach is very expensive and since there is no biotin pull-down, most of the sequenced reads would have to be discarded. Moreover, I discuss below about my experiment with the pore-C approach and the technical issue I faced.

3.1.1.2.2 Nano-C

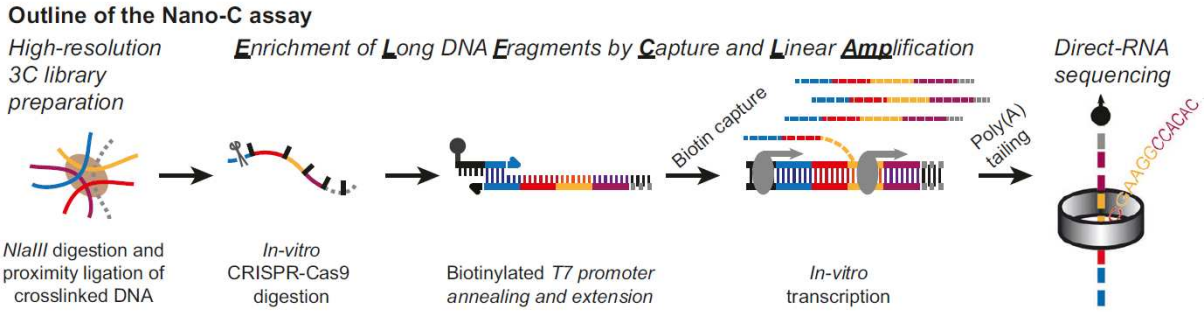


Figure 27. Experimental of the Nano-C adapted from Chang et al. 2023.

The Nano-C[164] approach is a one vs all multi contact approach (Figure 27). After de proximity ligation, chimeric DNA molecules are digested with Cas9 nuclease. The Cas9 nuclease is guided to digest a specific locus named viewpoint. Then, a biotinylated oligonucleotide hybridizes on the viewpoint. This oligonucleotide bears the T7 promoter sequence. After a streptavidin pull-down, the purified DNA molecules are transcribed in vitro. The transcripts are then sequenced with direct RNA sequencing from Oxford nanopore technology. Chang et al.[164] showed with this approach that CTCF clusters in more than two proteins.

3.2 Results.

When this project started, I needed to test experimental approach to sequence long chimeric molecules and build a bio-informatic pipeline of analysis. I did both in parallel. To develop my bio-informatical pipeline of analysis I used Hi-C library generated on G1 synchronised cells by myself and an Hi-C library generated on G2 synchronised cells by Christophe Chapard (CH84).

3.2.1.1 Multi-contacts can be extracted from Illumina paired end 150 base pair long reads.

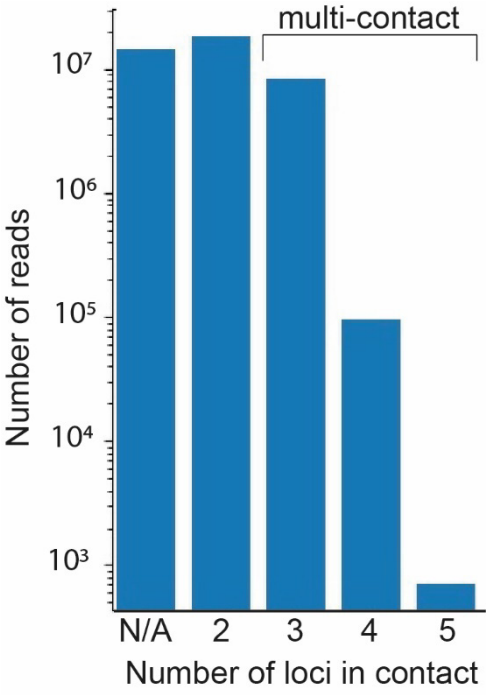


Figure 28. Histograms showing the abundance of multi-contact-event in the CH84 library.

Looking at multi-contact from reads is the same for every technology. The pipeline I developed could be applied to reads produced with Illumina, Nanopore… This strategy is inspired by hicstuff[152], the classical pipeline of the laboratory to analyse Hi-C data. Using my pipeline, I was able to recover multi-contact events. The multi-contact events correspond to reads where more than 1 relevant contacts were found. For instance, in the CH84 Hi-C library which generated with G2 synchronised *S. cerevisiae.* I obtained: 30% of uninformative reads; reads that does not contain any relevant ligation events, 48% of reads with one valid contact, 21% of reads with 2 valid contacts, 0,2 % of reads with 3 valid contacts and 0,002 % of reads with 4 valid contacts (Figure 28).

3.2.1.2 Cohesin loop anchors seems to cluster together.

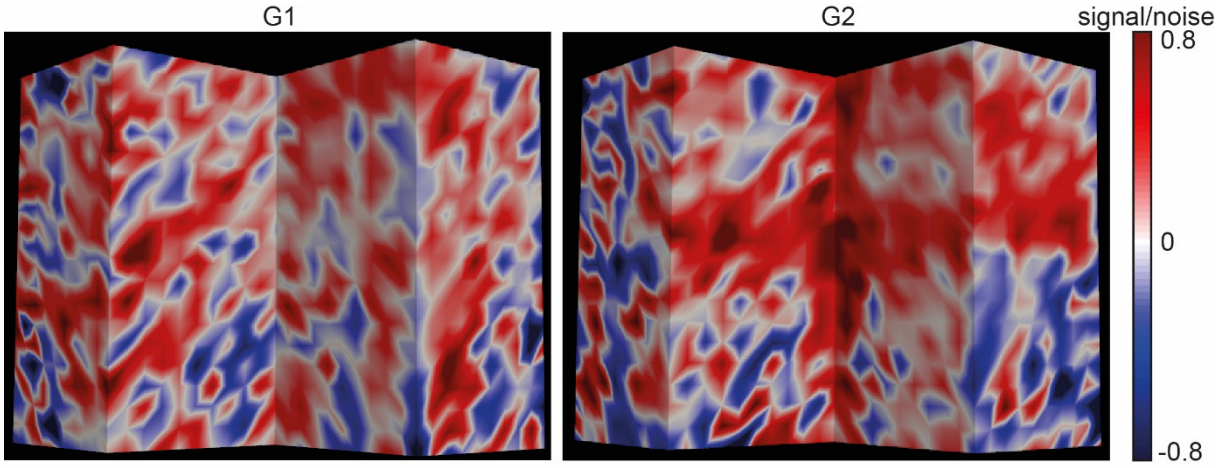


Figure 29. 3D representation of agglomerated cubes on cohesin loop anchor genomic positions in G1 or G2 synchronized S. cerevisiae cells.

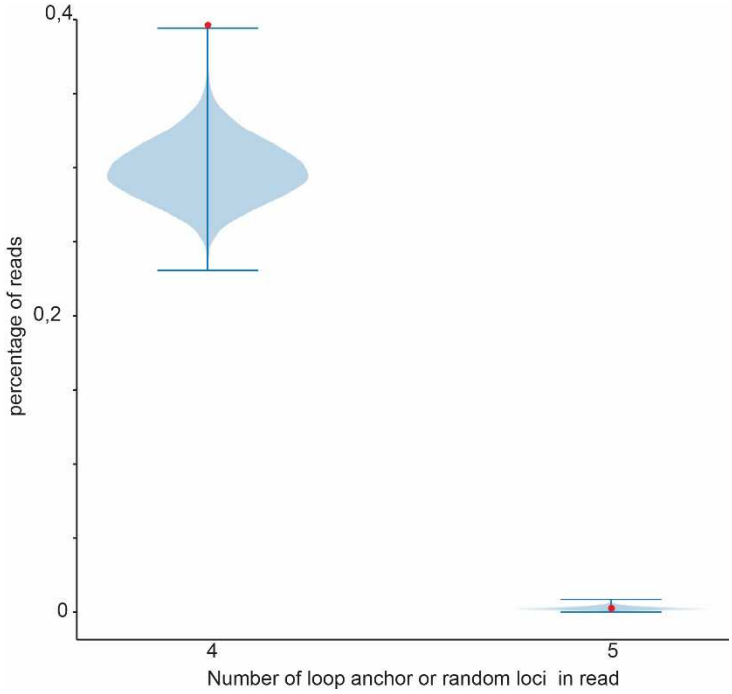


Figure 30. Representation of the bootstrap result. The Blue violin plot represent the distribution of reads with either four or five loci found in the random files. The red dots correspond to the result obtained for the cohesin loop anchor positions.

To test if the multi-contact found in CH84 had a biological significance, I asked whether the loop anchors colocalize together in clusters of 3,4 or 5. First, I represented agglomerated cubes of contact using reads where 3 loci were found together. I compared the agglomerated cubes between Hi-C library obtained from G2 and G1yeast synchronised cells. Both on 2D projection and 3D representation we can observe a strong enrichment of contact in the centre of each plot for the agglomerated cube obtained from G2 synchronised cell (Figure 29). This result is in accordance with the bootstrap test. The bootstrap test suggests that loops anchors clusters together by group of 4 (Figure 30). This result remains very preliminary since it is done on a small sample size (10000 reads had 4 validated loci in contact and 1000 had 5 validated loci in contact).

3.2.1.3 Nanopore sequencing yield too noisy reads both from 3C and Hi-C PCR amplified library

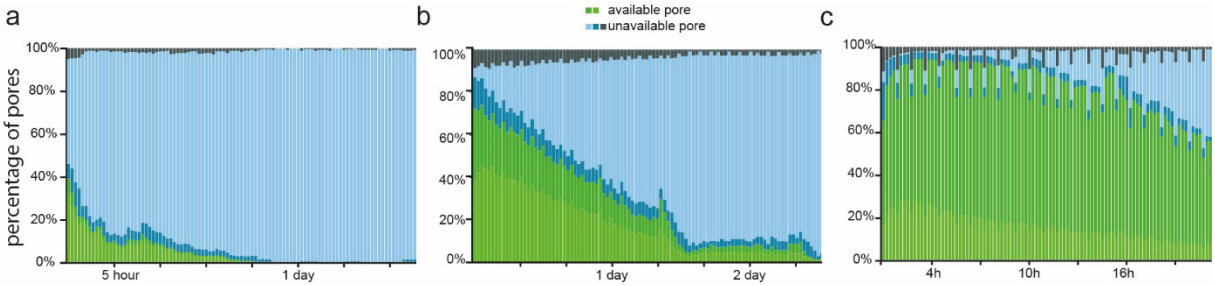


Figure 31. Histograms showing the proportion of available pores(green) and unavailable pores (blue and black) in one of my first sequencing run (a), a sequencing runs of PCR amplified Hi-C library (b) and a classical sequencing run on linear genomic DNA.

To go further and try to get long chimeric reads we wanted to use long read technology. We first use the Oxford Nanopore Technology. I tried to sequence DNA molecules obtained with the 3C procedure (without biotinylation step) and test the pore-C experimental approach.

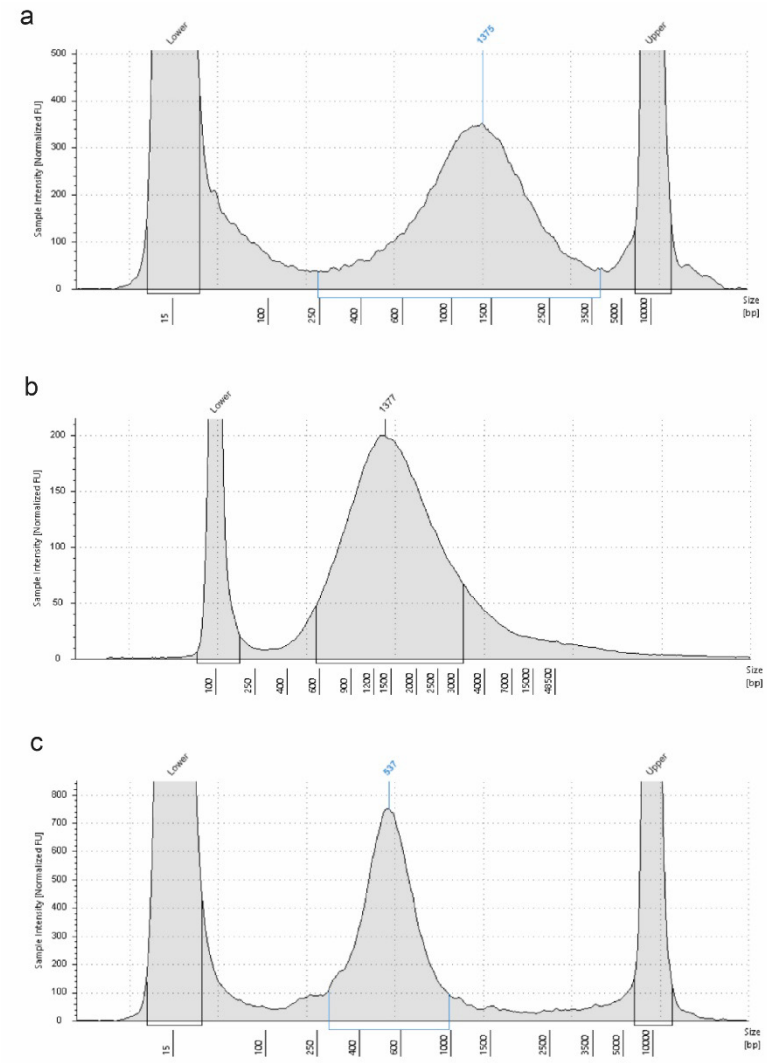


*Figure 32. Size distribution of amplicons obtained with Hi-C library generated on human Jurkat cells(**a**), S. cerevisiae cells (**b**) or P. aeruginosa (**c**). Each profile is obtained with the Agilent Tapestation system.*

The first sequencing runs yield were much below what could be expected (Figure 31). It appeared that un-amplified 3C library were harmful for the sequencing flow cell. I assumed it was because of the peculiar, branched topology of 3C molecules as reported by Noordermeer's team during the GDR ADN 2020. With this information, I decided to sequence PCR amplified libraries since PCR product topology is compatible with long read technologies. Moreover, it allowed me to use streptavidin pull-down and subsequently increase the quality of my libraries. The run duration and yield were greatly improved. However, the quality of the reads was not good enough to precisely map the genome. However, the sequenced reads were too short. I then moved to increase the size of my amplicons and switched to Pacific Bioscience sequencing technology. Unfortunately, the reads from all my tests were too short (median length was 1.5 kb) (Figure 32). Moreover, I sequenced 20 of them with Sanger sequencing. This analysis revealed that from 20 reads, around 20% were PCR duplicates and none of them contained a multi-contact read.

3.3 Discussion

The technological development of precise multi-contact Hi-C could help answers many biological questions. The results reported here are promising and still need development. I solved the sequencing problem with PCR amplification. However, if this step could be skipped, we could be able to directly sequence Hi-C reads with nanopore sequencing. Since nanopore sequencing can distinguish BrdU from the other canonical nucleotide. This could be of great interest when it comes to the interaction between the two sister chromatids in G2 phase. We could try to simplify the topology of 3C molecules by using a helicase coupled with a nuclease as for instance Mus81-Mms4 heterocomplex[166].

The principal technical problem I have for the multi-contact development is the size of the DNA molecules obtained after PCR. Interestingly, the size of the amplified molecules is rather the same for Human, *S. cerevisiae* and *P. aeruginosa* Hi-C library. To solve this issue, we could tweak the ligation step in the Hi-C protocol to produce bigger molecules or purify the largest DNA molecules obtained after the ligation step and only work with those very large DNA molecules.

3.4 Methods

3.4.1 Cell cultures and fixation

S. cerevisiae cells were grown overnight at 30C° in YPD liquid media at 180 RPM. The next day, cells were diluted in fresh media to 10^4 cells per millilitre and grown until it reaches 10^7 cells per millilitre. Then Nocodazole was added to a final concentration of 15 µg per millilitre and left at RT at 180RPM for 4 hours. Synchronisation was later checked using SytoxGreen and Myltenyl MACSquant cytometer. Fixation was done by adding 3% formaldehyde to the culture media for 30 min at RT, 100RPM. The Formaldehyde was quenched by adding Glycine at the 125 mM final concentration for 20 min at RT,100 RPM. Fixed cells were harvested by centrifugation at 5000g,5 min, RT and washed with PBS 1X.

Human cells were cultured and fixed by Thomas Verin from Marc Lavigne lab.

Pseudomas aeruginosa cells were grown in LB 37C° 180 RPM overnight and diluted in fresh media to an optical density at 600 nm (OD_{600}) of 0,1. Cells were grown until it reaches an OD_{600} of 0.5. Fixation was performed as *S. cerevisiae* cells.

3.4.2 3C and Hi-C procedure.

3C library were generated as described before[167]. Hi-C library we done using the ARIMA genomics HI-C kit. For Hi-C library, streptavidin pull-down was done as mentioned above in the pre-print article. However, no DNA fragmentation was performed prior to streptavidin pull-down. The sequencing library were prepared with sequencing adapters coming from SQK-LSK109 or SQK-RPB114.24 nanopore sequencing kit or SMRT bell PacBio sequencing kit. Because C1 streptavidin beads inhibits many enzymes, the end-preparation and adapters ligations were done using reagent from the Colibri kit (Invitrogen). The PCR was conducted using Takara Long Amp HF polymerase with compatible primers for each sequencing technologies. When applicable, the libraries were analysed prior to sequencing using the Agilent Tapestation system using a D5000 or D5000 high sensitivity DNA screentape.

3.4.3 Nanopore sequencing

The Nanopore sequencing was done on a MinION sequencer using MinION flow cell. The base calling was done on the GPUlab from Institut Pasteur (financed by the INCEPTION grant) using guppy 3.4.0 and aptainer.

3.4.4 Multi-contact bio-informatic pipeline and data visualisation

The analysis of reads either from PE 150 Illumina sequencing or nanopore sequencing were virtually digested. Thanks to the biotinylation step, I looked for pseudo duplication of restriction sites. If we take DpnII (one classical restriction enzyme used in Hi-C protocol for *S. cerevisiae)*. Its restriction site is 5'-*GATC-3', the star indicates the cut localisation. If DNA is digested, it will produce a 5' overhang which will filled during the biotinylation process. This, after DNA ligation will produce 5'-GATCGATC-3' pattern. Thus, I decided to splice reads when I found the exact suite of nucleotides corresponding to pseudo duplication of restriction sites from the enzymes used to generate the Hi-C library. This approach produced superior results for high quality reads from Illumina but was too stringent for low quality reads from Nanopore sequencing technology. For nanopore reads, I used a constant splicing size corresponding to the median restriction fragment. Next, I aligned the reads chunks using bowtie2 (Illumina reads) or Minimap2(for nanopore reads) against *S. cerevisiae* S288C-R64-1-1 reference genome. The raw multi-contact event were then filtered using thresholds for uninformative reads as described before[152]. Once the threshold for loops and uncuts were determined, I filtered events within

reads. I also removed the PCR/ optical duplicates. Once the multi-contact events were filtered, I generate a. multi file based on the. pairs file format. This file is then used to generates agglomerated cubes. The method to agglomerates cubes is the same compared to 2D agglomeration described previously[163] but adaptation had to be made for 3D agglomeration. The cubes were visualised using Mayavi2 python library. Code is available on my github account.

3.4.5 Bootstrap test

From the position of loops that were identified with chromosight[163] , I generated 10,000 random files. I kept the original genomic distances between positions. From this file, I asked for every multi-contact events how many positions matched between the read and the list of positions (randomly generated or corresponding to actual loop anchor positions). This allows to estimate the random probability of having 1 ,2,3,4 or 5 matching events. Then I compared the proportion obtained using the actual loop anchor position file to the distribution obtained with the 10,000 random files.

3.4.6 Sanger sequencing and analysis of amplified long Hi-C molecules.

Molecules from amplified Long Hi-C library were cloned in vectors using the Invitrogen zero blunt PCR cloning kit. 20 transformants were isolated and plasmids were extracted using nucleospin miniprep kit from Machenery-Nagel. The sanger sequencing was done using NEB Tubeseq services. The sequencing primer was the T7 promoter sequencing primer provided by NEB. Once the ligation event was located using the pseudo-duplication of restriction sites, each part of the read was aligned using BLAST.

4 Discussion

4.1 Molecular parasites tether to inactive genomic regions

Our pre-publication shows that known molecular parasites linked to plasmid 2µ and plasmid Ddp5 attach to inactive regions of their respective host genomes. The tethering of molecular parasites to inactive region of the genome could be part of the of the overall stealth strategy used by those parasites to avoid their detection and elimination from their host. We could hypotheses that the presence of the foreign DNA molecule docked onto inactive chromatin is less deleterious for the host cell compared to its docking on active genomic regions. We have further characterised the mechanism by which the 2µ plasmid attaches to its host chromosome. Anchoring could be directly mediated by the histone H4 tail. Our results show that when the H4 tail is replaced by alanine, contacts between the 2µ plasmid and host chromosomes disappear. However, we do not show any direct interaction between the Rep1/Rep2 complex and the histone H4 tail. We could, for example, look for the H4 tail in proteins co-immuno-precipitated with Rep1 or Rep2. In addition, an in vitro experiment showing the direct attachment of plasmid 2µ to reconstituted chromatin would provide further concrete evidence that plasmid 2µ attaches directly to host chromosomes by docking between nucleosomes as suggested by the Rep1 ChIP-seq signal on host chromatin.

The Rep1 ChIP signal on the 2µ plasmid is troubling (Figure 33). The previous reports pointed that Rep1 and Rep2 are enriched on the STB sequence but here we observe that Rep1

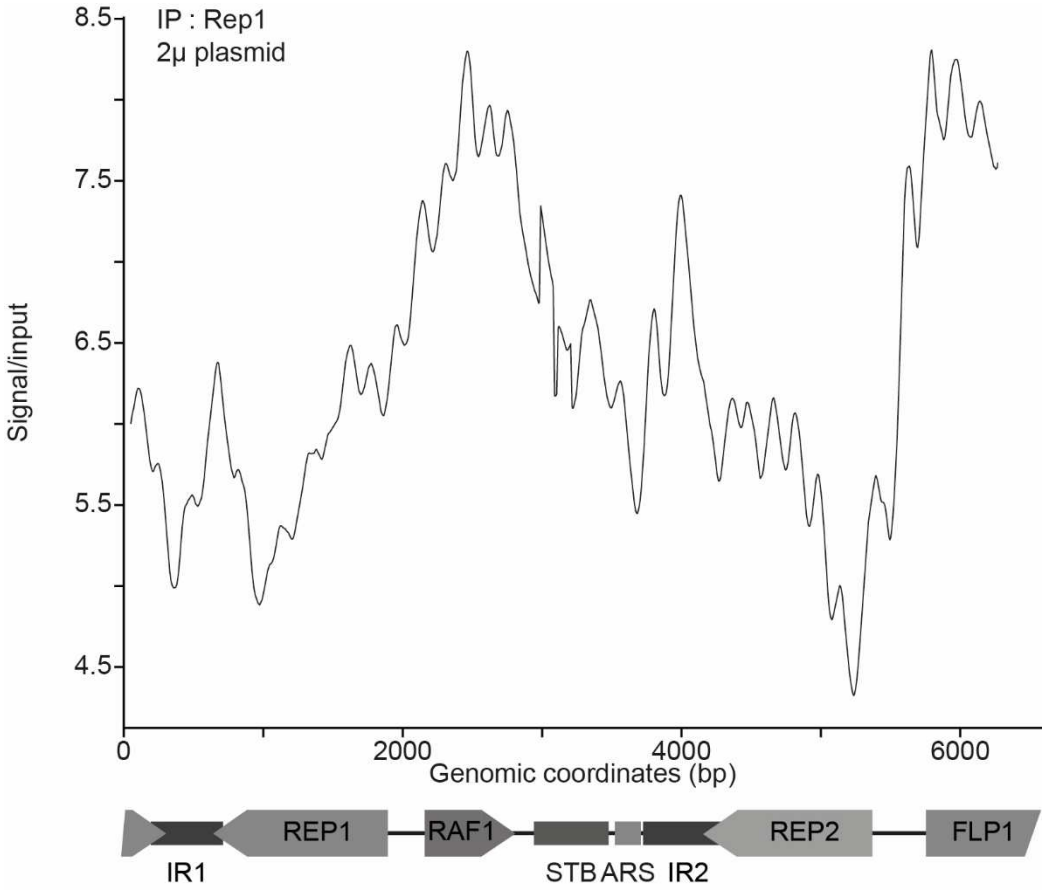


Figure 33. Plot showing the ChIP signal of Rep1 on the 2µ plasmid

is actually enriched on a large portion of the 2µ plasmid. One of the hypotheses we had about the tethering of the 2µ plasmid on host chromosomes was its winding around the host chromosomes that would be mediated by the Rep1/Rep2 complex. We could explore this assumption by inducing DNA breaks in the 2µ plasmid. The idea would be to evaluate upon break induction if the 2µ plasmid still tethers to host chromosome. This(ose) break(s) could be done using CRISPR/Cas system under an inducible promoter on a plasmid. We could induce multiple (3 breaks at the same time has been achieved in the laboratory using CRISPR/cas system) breaks at the different loci on the 2µ plasmid. We could target the ORFs, the IRs and even the STB locus.

Although *S. cerevisiae* is extensively studied, no essential functions associated with the 2µ plasmid have been reported. In my opinion, the 2µ plasmid does not bring any essential nor accessory functions to its host. Maybe we could define the 2µ plasmid as a commensal for *S. cerevisiae* and we could imagine that along the way, the 2µ plasmid could become a benefit for *S. cerevisiae* as for MGEs in prokaryotes that in certain cases improves host adaptability and genome plasticity.

4.2 *S. cerevisiae* defence mechanism

The field of prokaryotic defence mechanisms against infection is very dynamic. More and more defence mechanisms are being reported as research progresses. However, to my knowledge, the only defence mechanism reported in S. cerevisiae is the NPC. The NPC seems to filter the entry of genetic material into the nucleus and act on its expression in the nucleus. Yet S. cerevisiae is not massively invaded or threatened. The only known threats/parasites are the 2µ plasmid, retro-transposable elements, and cytoplasmic dsRNA viruses. Interestingly, dsRNA viruses and, in rare cases, Ty can be found as pseudocapsids in the cytoplasm. It seems they are trapped in the cell and can't escape it. Thus, we might ask whether new defence mechanisms have yet to be discovered in yeast, which might explain why so little threats to the *S. cerevisiae* genome are reported.

4.3 A common feature of eukaryotic parasitic plasmid/episome

During my PhD, we studied other plasmids in other organisms. The fact that they all seem to attach preferentially to inactive genomic regions could be used to search for other molecular parasites in eukaryotes. As mentioned in the introduction, molecular parasites of eukaryotes tend to be labelled as bacterial contaminants in shotgun-based assemblies. Firstly, we could use Hi-C information to locate plasmids in the nucleus and search for them. Secondly, we could look for plasmids that preferentially attach to weakly transcribed genomic regions. This can be done by working on raw Hi-C datasets generated on any organism, for example on the DNA zoo dataset.[168].

Moreover, the 2µ plasmid and derivatives, the Ddp plasmid family and EBV share some common peculiar genetic features such as the presence of repeated motif as the STB sequence of the 2µ plasmid. The EBV episomes also tethers to host chromosome with the help of viral encoded persistence protein EBNA1. This homology in the genetic structure and stability strategy could be the result of convergent evolution; the persistence issue in eukaryotes nucleus is the same for every persistent eukaryotic circular ecDNA and their respective strategies all converged to the hitchhiking strategy on inactive genomic regions. Or which could be very intriguing, all those selfish parasites share a common evolutive history. This could be studied by comparing the docking proteic complexes of each parasite and look for homology either in their quaternary structures, primary structure, and DNA sequence.

4.4 New project and biotechnological application of the 2µ plasmid system

The first biotechnology based on the 2µ plasmid were the FRT/Flp1 genomic engineering system and the 2µ-based plasmid in *S. cerevisiae*. Here we report that the 2µ plasmid tethers to inactive regions of its host regardless of its sequence. This result is the basis of the "chromoglue" project which will be conducted by my PhD co-director Axel Cournac and our collaborator Nolwenn Jouvenet. This project has received fundings from the "direction des Applications de la recherche et des relations industrielles" (DARRI) from the Institut Pasteur. Most available strategies to design stable vectors in human cells often involve viral genome that represent bio-safety risks. The idea is to design stable vector in human cells using the Rep1/Rep2 tethering system. This persistence strategy could allow the stable expression of genes in human cells and open the way to new gene therapy strategies that are independent of viral genomic elements and as unobtrusive as the 2µ plasmid is for its host.

Moreover, I started an experiment where I want to integrate STB sites in the *S. cerevisiae* genome and observe whether the integrated sites contact each other and if the STB site (and recruited proteins) blocked the cohesin complex. This experiment was based on a peculiar strain, available in the laboratory collection in which a 100kb synthetic and random sequence is integrated into the chr IV. In this playground, I managed to integrate one STB with the CRISPR technology.

5 Bibliography

1. Liu, Y. & Li, X. Darwin's Pangenesis, and molecular medicine. *Trends Mol. Med.* **18**, 506–508 (2012).
2. Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* **2**, 72–75 (2001).
3. Avery, O. T. & McCARTY, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES.
4. Haldane, J. The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* (1919).
5. Sturtevant, A. H. A Third Group of Linked Genes in Drosophila ampelophila. *Science* **37**, 990–992 (1913).
6. Winkler, H. & Winkler, H. *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche , p165*. 1–248 (G. FIscher, 1920). doi:10.5962/bhl.title.1460.
7. Goffeau, A. *et al.* Life with 6000 Genes. *Science* **274**, 546–567 (1996).
8. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
9. Completing human genomes. *Nat. Methods* **19**, 629–629 (2022).
10. Hall, J. P. J., Botelho, J., Cazares, A. & Baltrus, D. A. What makes a megaplasmid? *Philos. Trans. R. Soc. B Biol. Sci.* **377**, 20200472 (2022).
11. Lederberg, J. Cell Genetics and Hereditary Symbiosis. *Physiol. Rev.* (1952) doi:10.1152/physrev.1952.32.4.403.
12. Casali, N. & Preston, A. *E. coli Plasmid Vectors*. vol. 235 (Humana Press, 2003).
13. Lederberg, J. Plasmid (1952–1997). *Plasmid* **39**, 1–9 (1998).
14. Campbell, A. Episomes. in *Bacteria, Bacteriophages, and Fungi: Volume 1* (ed. King, R. C.) 295–307 (Springer US, 1974). doi:10.1007/978-1-4899-1710-2_18.
15. Broussard, G. & Damania, B. Regulation of KSHV Latency and Lytic Reactivation. *Viruses* **12**, 1034 (2020).
16. Kim, K.-D. *et al.* Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. *Nat. Commun.* **11**, 877 (2020).
17. Bastien, N. & McBride, A. A. Interaction of the Papillomavirus E2 Protein with Mitotic Chromosomes. *Virology* **270**, 124–134 (2000).
18. Barnett, J. A. A history of research on yeasts 10: foundations of yeast genetics1. *Yeast* **24**, 799–845 (2007).
19. Pasteur, L. *Mémoire sur la fermentation appelée lactique*. (1858).

20. Lindegren, C. C. & Lindegren, G. Segregation, Mutation, and Copulation in Saccharomyces cerevisiae. *Ann. Mo. Bot. Gard.* **30**, 453–468 (1943).
21. CHAUCER, G. The Canterbury Tales. *The British Library* (1400).
22. Lindegren, C. C. The yeast cell, its genetics and cytology. (1949).
23. Dujon, B. Mitochondrial genetics revisited. *Yeast* **37**, 191–205 (2020).
24. Hartwell, L. H., Culotti, J., Pringle, J. R. & Reid, B. J. Genetic Control of the Cell Division Cycle in Yeast. *Science* **183**, 46–51 (1974).
25. Sinclair, J. H., Stevens, B. J., Sanghavi, P. & Rabinowitz, M. Mitochondrial-satellite, and circular DNA filaments in yeast. *Science* **156**, 1234–1237 (1967).
26. Stevens, B. J. & Moustacchi, E. ADN SATELLITE y ET MOLBCULES CIRCULAIRES TORSADl?ES DE PETITE TAILLE CHEZ LA LEVURE SACCHAROMYCES CEREVISIAE. (1971).
27. Petes, T. D. & Williamson, D. H. Replicating circular DNA molecules in yeast. *Cell* **4**, 249–253 (1975).
28. Clark-Walker, G. D. & Miklos, G. L. G. Localization and Quantification of Circular DNA in Yeast. *Eur. J. Biochem.* **41**, 359–365 (1974).
29. Taketo, M., Jazwinski, S. M. & Edelman, G. M. Association of the 2-micron DNA plasmid with yeast folded chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3144–3148 (1980).
30. Hartley, J. L. & Donelson, J. E. Nucleotide sequence of the yeast plasmid. *Nature* **286**, 860–864 (1980).
31. Jayaram, M., Li, Y.-Y. & Broach, J. R. The yeast plasmid 2μ circle encodes components required for its high copy propagation. *Cell* **34**, 95–104 (1983).
32. Jayaram, M., Sutton, A. & Broach, J. R. Properties of REP3: a cis-acting locus required for stable propagation of the Saccharomyces cerevisiae plasmid 2 microns circle. *Mol. Cell. Biol.* **5**, 2466–2475 (1985).
33. Murray, J. A., Scarpa, M., Rossi, N. & Cesareni, G. Antagonistic controls regulate copy number of the yeast 2 mu plasmid. *EMBO J.* **6**, 4205–4212 (1987).
34. Futcher, A. B. & Cox, B. S. Maintenance of the 2 microns circle plasmid in populations of Saccharomyces cerevisiae. *J. Bacteriol.* **154**, 612–622 (1983).
35. Mead, D. J., Gardner, D. C. J. & Oliver, S. G. The yeast 2 μ plasmid: strategies for the survival of a selfish DNA. *Mol. Gen. Genet. MGG* **205**, 417–421 (1986).
36. Dawkins, R. *the selfish gene*. (1976).
37. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm, and genome evolution. *Nature* **284**, 601–603 (1980).

38. Beggs, J. D. Transformation of yeast by a replicating hybrid plasmid. *Nature* **275**, 104–109 (1978).
39. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).
40. Escudero, J. A., Loot, C., Nivina, A. & Mazel, D. The Integron: Adaptation On Demand. *Microbiol. Spectr.* **3**, 10.1128/microbiolspec.mdna3-0019–2014 (2015).
41. Grognet, P. *et al.* A RID-like putative cytosine methyltransferase homologue controls sexual development in the fungus Podospora anserina. *PLoS Genet.* **15**, e1008086 (2019).
42. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251–259 (2010).
43. Brouwer, M. S. M. *et al.* Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers. *Nat. Commun.* **4**, 2601 (2013).
44. Blaisonneau, J., Sor, F., Cheret, G., Yarrow, D. & Fukuhara, H. A Circular Plasmid from the YeastTorulaspora delbrueckii. *Plasmid* **38**, 202–209 (1997).
45. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
46. Metz, B. A., Ward, T. E., Welker, D. L. & Williams, K. L. Identification of an endogenous plasmid in Dictyostelium discoideum. *EMBO J.* **2**, 515–519 (1983).
47. Moreau, P. *et al.* Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. *Nat. Commun.* **9**, 4268 (2018).
48. Rush, M. G. & Misra, R. Extrachromosomal DNA in eucaryotes. *Plasmid* **14**, 177–191 (1985).
49. Zuo, S. *et al.* Extrachromosomal Circular DNA (eccDNA): From Chaos to Function. *Front. Cell Dev. Biol.* **9**, 792555 (2022).
50. Meinema, A. C. *et al.* DNA circles promote yeast ageing in part through stimulating the reorganization of nuclear pore complexes. *eLife* **11**, e71196 (2022).
51. Fernández-Domínguez, I. J. *et al.* The role of extracellular DNA (exDNA) in cellular processes. *Cancer Biol. Ther.* **22**, 267–278.
52. Gehlen, L. R. *et al.* Nuclear Geometry and Rapid Mitosis Ensure Asymmetric Episome Segregation in Yeast. *Curr. Biol.* **21**, 25–33 (2011).
53. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* **19**, 347–359 (2021).
54. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**, 563–602 (1994).

55. Harris, G. & Thompson, C. C. Alleged Transformation of Yeast. *Nature* **188**, 1212–1213 (1960).
56. Mitrikeski, P. T. Yeast competence for exogenous DNA uptake: towards understanding its genetic component. *Antonie Van Leeuwenhoek* **103**, 1181–1207 (2013).
57. Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F. & Arechaga, I. Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* **39**, 81–95 (2015).
58. Virolle, C., Goldlust, K., Djermoun, S., Bigot, S. & Lesterlin, C. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes* **11**, 1239 (2020).
59. San Millan, A. & MacLean, R. C. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiol. Spectr.* **5**, 10.1128/microbiolspec.mtbp-0016–2017 (2017).
60. Heinemann, J. A. & Sprague, G. F. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).
61. Heinemann, J. A. Genetics of gene transfer between species. *Trends Genet.* **7**, 181–185 (1991).
62. INFECTION : Définition de INFECTION. https://www.cnrtl.fr/definition/infection.
63. Deatherage, B. L. & Cookson, B. T. Membrane Vesicle Release in Bacteria, Eukaryotes, and Archaea: a Conserved yet Underappreciated Aspect of Microbial Life. *Infect. Immun.* **80**, 1948–1957 (2012).
64. Elzanowska, J., Semira, C. & Costa-Silva, B. DNA in extracellular vesicles: biological and clinical aspects. *Mol. Oncol.* **15**, 1701–1714 (2021).
65. Bhargava, A., Lahaye, X. & Manel, N. Let me in: Control of HIV nuclear entry at the nuclear envelope. *Cytokine Growth Factor Rev.* **40**, 59–67 (2018).
66. Ni, R., Feng, R. & Chau, Y. Synthetic Approaches for Nucleic Acid Delivery: Choosing the Right Carriers. *Life* **9**, 59 (2019).
67. Schenkel, L., Wang, X., Le, N., Burger, M. & Kroschewski, R. A special envelope separates extra-chromosomal from mammalian chromosomal DNA in the cytoplasm. 2023.03.02.530628 Preprint at https://doi.org/10.1101/2023.03.02.530628 (2023).
68. Hotta, Y. & Bassel, A. Molecular size and circularity of dna in cells of mammals and higher plants*. *Proc. Natl. Acad. Sci.* **53**, 356–362 (1965).
69. Zhao, Y., Yu, L., Zhang, S., Su, X. & Zhou, X. Extrachromosomal circular DNA: Current status and future prospects. *eLife* **11**, e81412 (2022).
70. Tao, Y. *et al.* The formation mechanism and homeostasis of extrachromosomal DNA. *Carcinogenesis* **43**, 815–825 (2022).

71. Motwani, M., Pesiridis, S. & Fitzgerald, K. A. DNA sensing by the cGAS–STING pathway in health and disease. *Nat. Rev. Genet.* **20**, 657–674 (2019).
72. Baños, R. C. *et al.* Differential Regulation of Horizontally Acquired and Core Genome Genes by the Bacterial Modulator H-NS. *PLOS Genet.* **5**, e1000513 (2009).
73. Jones, C. & Holland, I. B. Role of the SulB (FtsZ) protein in division inhibition during the SOS response in Escherichia coli: FtsZ stabilizes the inhibitor SulA in maxicells. *Proc. Natl. Acad. Sci.* **82**, 6045–6049 (1985).
74. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).
75. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
76. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258 (2011).
77. Cheng, K., Wilkinson, M., Chaban, Y. & Wigley, D. B. A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. *Nat. Struct. Mol. Biol.* **27**, 71–77 (2020).
78. Jin, S., Zhan, J. & Zhou, Y. Argonaute proteins: structures and their endonuclease activity. *Mol. Biol. Rep.* **48**, 4837–4849 (2021).
79. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
80. Girard, A. & Hannon, G. J. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* **18**, 136–148 (2008).
81. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020).
82. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).
83. Doyle, M. *et al.* An H-NS-like Stealth Protein Aids Horizontal DNA Transmission in Bacteria. *Science* **315**, 251–252 (2007).
84. Lang, K. S. & Johnson, T. J. Characterization of Acr2, an H-NS-like protein encoded on A/C2-type plasmids. *Plasmid* **87–88**, 17–27 (2016).
85. Ma, Z. & Damania, B. The cGAS-STING Defense Pathway and Its Counteraction by Viruses. *Cell Host Microbe* **19**, 150–158 (2016).

86. Wilkinson, M. *et al.* Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. *eLife* **5**, e22963 (2016).
87. Cullen, B. R. RNA Interference in Mammals: The Virus Strikes Back. *Immunity* **46**, 970–972 (2017).
88. Krüger, D. H. & Bickle, T. A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.* **47**, 345–360 (1983).
89. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Manipulating or Superseding Host Recombination Functions: A Dilemma That Shapes Phage Evolvability. *PLoS Genet.* **9**, e1003825 (2013).
90. Rocha, E. P. C. & Bikard, D. Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? *PLoS Biol.* **20**, e3001514 (2022).
91. Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561.e12 (2020).
92. Baharoglu, Z., Bikard, D. & Mazel, D. Conjugative DNA Transfer Induces the Bacterial SOS Response and Promotes Antibiotic Resistance Development through Integron Activation. *PLOS Genet.* **6**, e1001165 (2010).
93. Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740-753.e5 (2022).
94. Ghosh, S. K., Hajra, S., Paek, A. & Jayaram, M. Mechanisms for Chromosome and Plasmid Segregation. *Annu. Rev. Biochem.* **75**, 211–241 (2006).
95. Salje, J. Plasmid segregation: how to survive as an extra piece of DNA. *Crit. Rev. Biochem. Mol. Biol.* **45**, 296–317 (2010).
96. Lukacs, G. L. *et al.* Size-dependent DNA Mobility in Cytoplasm and Nucleus*. *J. Biol. Chem.* **275**, 1625–1629 (2000).
97. Sau, S., Ghosh, S. K., Liu, Y.-T., Ma, C.-H. & Jayaram, M. Hitchhiking on chromosomes: A persistence strategy shared by diverse selfish DNA elements. *Plasmid* **102**, 19–28 (2019).
98. Kruitwagen, T., Chymkowitch, P., Denoth-Lippuner, A., Enserink, J. & Barral, Y. Centromeres License the Mitotic Condensation of Yeast Chromosome Arms. *Cell* **175**, 780-795.e15 (2018).
99. Bouet, J.-Y. & Funnell, B. E. Plasmid Localization and Partition in Enterobacteriaceae. *EcoSal Plus* **8**, 10.1128/ecosalplus.ESP-0003–2019 (2019).
100. Salje, J., Gayathri, P. & Löwe, J. The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* **8**, 683–692 (2010).

101. Guynet, C., Cuevas, A., Moncalián, G. & Cruz, F. de la. The stb Operon Balances the Requirements for Vegetative Stability and Conjugative Transfer of Plasmid R388. *PLOS Genet.* **7**, e1002073 (2011).

102. De Leo, A., Calderon, A. & Lieberman, P. M. Control of Viral Latency by Episome Maintenance Proteins. *Trends Microbiol.* **28**, 150–162 (2020).

103. Daya, S. & Berns, K. I. Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* **21**, 583–593 (2008).

104. Thisted, T. & Gerdes, K. Mechanism of post-segregational killing by the hok/sok system of plasmid R1: Sok antisense RNA regulates hok gene expression indirectly through the overlapping mok gene. *J. Mol. Biol.* **223**, 41–54 (1992).

105. Strak, M. J. R., Boyd, A., Mileham, A. J. & Ramonos, M. A. The plasmid-encoded killer system of Kluyveromyces lactis: A review. *Yeast* **6**, 1–29 (1990).

106. Boynton, P. J. The ecology of killer yeasts: Interference competition in natural habitats. *Yeast* **36**, 473–485 (2019).

107. Mannazzu, I. *et al.* Yeast killer toxins: from ecological significance to application. *Crit. Rev. Biotechnol.* **39**, 603–617 (2019).

108. San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends Microbiol.* **26**, 978–985 (2018).

109. Bennett, P. M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.* **153**, S347–S357 (2008).

110. Coluzzi, C. *et al.* Chance favors the prepared genomes: horizontal transfer shapes the emergence of antibiotic resistance mutations in core genes. 2023.06.20.545734 Preprint at https://doi.org/10.1101/2023.06.20.545734 (2023).

111. Tandon, I., Pal, R., Pal, J. K. & Sharma, N. K. Extrachromosomal circular DNAs: an extra piece of evidence to depict tumor heterogeneity. *Future Sci. OA* **5**, FSO390 (2019).

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

113. Lin, C. *et al.* Encoding gene RAB3B exists in linear chromosomal and circular extrachromosomal DNA and contributes to cisplatin resistance of hypopharyngeal squamous cell carcinoma via inducing autophagy. *Cell Death Dis.* **13**, 1–10 (2022).

114. Koo, D.-H. *et al.* Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. *Proc. Natl. Acad. Sci.* **115**, 3332–3337 (2018).

115. San Millan, A. The journey of bacterial genes. *Nat. Ecol. Evol.* **6**, 498–499 (2022).

116. Morillon, A., Springer, M. & Lesage, P. Activation of the Kss1 Invasive-Filamentous Growth Pathway Induces Ty1 Transcription and Retrotransposition in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **20**, 5766–5776 (2000).

117. Gresham, D. *et al.* The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. *PLoS Genet.* **4**, e1000303 (2008).

118. Curcio, M. J., Lutz, S. & Lesage, P. The Ty1 LTR-Retrotransposon of Budding Yeast, Saccharomyces cerevisiae. *Microbiol. Spectr.* (2015) doi:10.1128/microbiolspec.mdna3-0053-2014.

119. Vial, L. & Hommais, F. Plasmid-chromosome cross-talks. *Environ. Microbiol.* **22**, 540–556 (2020).

120. Kirchberger, P. C., Schmidt, M. L. & Ochman, H. The Ingenuity of Bacterial Genomes. *Annu. Rev. Microbiol.* **74**, 815–834 (2020).

121. Heidelberg, J. F. *et al.* DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature* **406**, 477–483 (2000).

122. Val, M.-E., Soler-Bistué, A., Bland, M. J. & Mazel, D. Management of multipartite genomes: the Vibrio cholerae model. *Curr. Opin. Microbiol.* **22**, 120 (2014).

123. Trucksis, M., Michalski, J., Deng, Y. K. & Kaper, J. B. The Vibrio cholerae genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14464–14469 (1998).

124. Kirkup, B. C., Chang, L., Chang, S., Gevers, D. & Polz, M. F. Vibrio chromosomes share common history. *BMC Microbiol.* **10**, 137 (2010).

125. Harrison, P. W., Lower, R. P. J., Kim, N. K. D. & Young, J. P. W. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* **18**, 141–148 (2010).

126. Grewal, S. I. & Rice, J. C. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* **16**, 230–238 (2004).

127. Curcio, M. J., Hedge, A.-M., Boeke, J. D. & Garfinkel, D. J. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in Saccharomyces cerevisiae. *Mol. Gen. Genet. MGG* **220**, 213–221 (1990).

128. Bonnet, A., Chaput, C., Palmic, N., Palancade, B. & Lesage, P. A nuclear pore sub-complex restricts the propagation of Ty retrotransposons by limiting their transcription. *PLoS Genet.* **17**, e1009889 (2021).

129. Chapard, C. *et al.* Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization. 2022.12.21.520625 Preprint at https://doi.org/10.1101/2022.12.21.520625 (2023).

130. Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).

131. Argos, P. *et al.* The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**, 433–440 (1986).

132. Bates, S., Cashmore, A. M. & Wilkins, B. M. IncP Plasmids Are Unusually Effective in Mediating Conjugation of Escherichia coli and Saccharomyces cerevisiae: Involvement of the Tra2 Mating System. *J. Bacteriol.* **180**, 6538–6543 (1998).

133. Hayman, G. T. & Bolen, P. L. Movement of Shuttle Plasmids from Escherichia coli into Yeasts Other Than Saccharomyces cerevisiae Using Trans-kingdom Conjugation. *Plasmid* **30**, 251–257 (1993).

134. Dujon, B. Yeast Evolutionary Genomics. in *Yeast* 407–419 (John Wiley & Sons, Ltd, 2012). doi:10.1002/9783527659180.ch16.

135. Phung, H. T. T., Nguyen, H. L. H. & Nguyen, D. H. The possible function of Flp1 in homologous recombination repair in Saccharomyces cerevisiae. *AIMS Genet.* **5**, 161–176 (2018).

136. Mehta, S., Yang, X.-M., Jayaram, M. & Velmurugan, S. A novel role for the mitotic spindle during DNA segregation in yeast: promoting 2 microm plasmid-cohesin association. *Mol. Cell. Biol.* **25**, 4283–4298 (2005).

137. De Chiara, M. *et al.* Domestication reprogrammed the budding yeast life cycle. *Nat. Ecol. Evol.* **6**, 448–460 (2022).

138. Buser, C. C., Jokela, J. & Martin, O. Y. Scent of a killer: How could killer yeast boost its dispersal? *Ecol. Evol.* **11**, 5809–5814 (2021).

139. Rizvi, S. M. A., Prajapati, H. K. & Ghosh, S. K. The 2 micron plasmid: a selfish genetic element with an optimized survival strategy within Saccharomyces cerevisiae. *Curr. Genet.* **64**, 25–42 (2018).

140. Heun, P., Laroche, T., Raghuraman, M. K. & Gasser, S. M. The Positioning and Dynamics of Origins of Replication in the Budding Yeast Nucleus. *J. Cell Biol.* **152**, 385–400 (2001).

141. Hajra, S., Ghosh, S. K. & Jayaram, M. The centromere-specific histone variant Cse4p (CENP-A) is essential for functional chromatin architecture at the yeast 2-μm circle partitioning locus and promotes equal plasmid segregation. *J. Cell Biol.* **174**, 779–790 (2006).

142. Cui, H., Ghosh, S. K. & Jayaram, M. The selfish yeast plasmid uses the nuclear motor Kip1p but not Cin8p for its localization and equal segregation. *J. Cell Biol.* **185**, 251–264 (2009).

143. Kumar, D. *et al.* The selfish yeast plasmid utilizes the condensin complex and condensed chromatin for faithful partitioning. *PLoS Genet.* **17**, e1009660 (2021).

144. Liu, Y.-T., Ma, C.-H. & Jayaram, M. Co-segregation of yeast plasmid sisters under monopolin-directed mitosis suggests association of plasmid sisters with sister chromatids. *Nucleic Acids Res.* **41**, 4144–4158 (2013).

145. Ghosh, S. K., Hajra, S. & Jayaram, M. Faithful segregation of the multicopy yeast plasmid through cohesin-mediated recognition of sisters. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13034–13039 (2007).

146. Prajapati, H. K., Rizvi, S. M. A., Rathore, I. & Ghosh, S. K. Microtubule-associated proteins, Bik1 and Bim1, are required for faithful partitioning of the endogenous 2 micron plasmids in budding yeast. *Mol. Microbiol.* **103**, 1046–1064 (2017).

147. Pinder, J. B. Investigation of Post-Translational Modification and Function of the Yeast Plasmid Partitioning Proteins Rep1 and Rep2. (2012).

148. Hays, M., Young, J. M., Levan, P. F. & Malik, H. S. A natural variant of the essential host gene MMS21 restricts the parasitic 2-micron plasmid in Saccharomyces cerevisiae. *eLife* **9**, e62337 (2020).

149. Ma, C.-H. *et al.* A Flp-SUMO hybrid recombinase reveals multi-layered copy number control of a selfish DNA element through post-translational modification. *PLoS Genet.* **15**, e1008193 (2019).

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

151. Krietenstein, N. & Rando, O. J. Mammalian Micro-C-XLMicro-C-XL. in *Chromatin: Methods and Protocols* (eds. Horsfield, J. & Marsman, J.) 321–332 (Springer US, 2022). doi:10.1007/978-1-0716-2140-0_17.

152. Matthey-Doret, C. *et al.* Normalization of Chromosome Contact Maps: Matrix Balancing and Visualization. in *Hi-C Data Analysis: Methods and Protocols* (eds. Bicciato, S. & Ferrari, F.) 1–15 (Springer US, 2022). doi:10.1007/978-1-0716-1390-0_1.

153. Lazar-Stefanita, L. *et al.* Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.* **36**, 2684–2697 (2017).

154. Piazza, A. *et al.* Cohesin regulates homology search during recombinational DNA repair. *Nat. Cell Biol.* **23**, 1176–1186 (2021).

155. Wang, L. *et al.* Epstein-Barr Virus Episome Physically Interacts with Active Regions of the Host Genome in Lymphoblastoid Cells. *J. Virol.* **94**, e01390-20 (2020).

156. Baudry, L. *et al.* instaGRAAL: chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder. *Genome Biol.* **21**, 148 (2020).

157. Marbouty, M. & Koszul, R. Metagenomes Binning Using Proximity-Ligation Data. in *Hi-C Data Analysis: Methods and Protocols* (eds. Bicciato, S. & Ferrari, F.) 163–181 (Springer US, 2022). doi:10.1007/978-1-0716-1390-0_8.

158. Matthey-Doret, C. *et al.* Chromosome-scale assemblies of Acanthamoeba castellanii genomes provide insights into Legionella pneumophila infection–related chromatin reorganization. *Genome Res.* **32**, 1698–1710 (2022).

159. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).

160. Zhang, S., Chasman, D., Knaack, S. & Roy, S. In silico prediction of high-resolution Hi-C interaction matrices. *Nat. Commun.* **10**, 5449 (2019).

161. Girard, F. *et al.* Anchoring of parasitic plasmids to inactive regions of eukaryotic chromosomes through nucleosome signal. 2023.10.04.558402 Preprint at https://doi.org/10.1101/2023.10.04.558402 (2023).

162. Dauban, L. *et al.* Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol. Cell* **77**, 1279-1293.e4 (2020).

163. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020).

164. Chang, L.-H. *et al.* Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. *Nat. Commun.* **14**, 5615 (2023).

165. Deshpande, A. S. *et al.* Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).

166. Schwartz, E. K. *et al.* Mus81-Mms4 Functions as a Single Heterodimer To Cleave Nicked Intermediates in Recombinational DNA Repair. *Mol. Cell. Biol.* **32**, 3065–3080 (2012).

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

168. DNA Zoo. *DNA Zoo* https://www.dnazoo.org.