

Quand une cellule eucaryote adopte un chromosome bactérien

Léa Meneu^{1,2} , Jacques Serizay¹ , Romain Koszul¹

¹Institut Pasteur, Régulation spatiale des génomes, CNRS UMR 3525, Université Paris Cité, Paris, France.

²Sorbonne université, Collège doctoral, Paris, France.
romain.koszul@pasteur.fr

➤ L'information génétique qui constitue un génome est contenue dans un chromosome (chez les bactéries) ou répartie dans plusieurs chromosomes (chez les eucaryotes), dont la molécule d'ADN « en double hélice » comporte une combinaison séquentielle de bases nucléotidiques, adénine (A), cytosine (C), guanine (G), ou thymine (T), appariées en AT et GC. Chaque espèce présente dans son génome des proportions relatives de paires de base AT et GC variables. L'ordonnement de ces bases le long de la molécule d'ADN non seulement individualise les différents gènes, mais aussi est impliqué dans la régulation de leur expression, la réplication du génome, ou son repliement tridimensionnel. Cette dernière propriété est essentielle pour permettre à des molécules dont la taille cumulée peut atteindre plusieurs mètres (environ deux mètres pour le génome humain) de trouver leur place dans un noyau cellulaire dont le diamètre mesure quelques micromètres. Cette compaction n'est cependant pas aléatoire : tous les génomes, qu'ils proviennent d'une bactérie, et contiennent alors environ 3 millions de paires de base, ou d'un mammifère, et en contiennent alors environ 3 milliards, sont structurés selon plusieurs niveaux d'organisation spatiale, allant de petits domaines liés à la transcription des gènes [m/s 1], à un ou plusieurs grands compartiments. Dans les génomes de mammifères, ces compartiments sont souvent associés à un état transcriptionnel actif ou inactif, mais leur fonction reste souvent à déterminer [2].

L'histoire évolutive d'une espèce s'accompagne de mutations des bases nucléotidiques de l'ADN qui en modifient la séquence, ce qui affecte les fonctions biologiques du génome. Mais la séquence nucléotidique modifiée peut aussi présenter des propriétés physiques différentes de la séquence ancestrale. Par exemple, la proportion de paires de bases GC, qui sont plus « rigides » que les paires de bases AT car maintenues par trois liaisons hydrogènes au lieu de deux pour les paires AT, pourra varier, perturbant alors les propriétés mécaniques de la molécule et modifiant sa capacité à se replier. L'évolution vers un génome plus riche en GC devra donc s'accompagner d'une évolution des mécanismes de repliement ou d'organisation spatiale de la molécule d'ADN. Cette coévolution entre les caractéristiques physiques de cette grande molécule et ses fonctions liées au métabolisme de la cellule rend complexe la distinction des liens de causalité entre la séquence nucléotidique de l'ADN, son repliement 3D, et l'activité des chromosomes en rapport avec le métabolisme, incluant la régulation de leur activité transcriptionnelle. Pourtant, il arrive assez fréquemment que des molécules d'ADN circulent d'une espèce à une autre, et se trouvent ainsi nouvellement confrontées à des règles métaboliques sous la contrainte desquelles elles n'ont pas évolué précédemment, par exemple, dans le cas d'infections de nouvelles espèces par des virus à ADN, mais aussi lors de transferts de gènes, ou encore d'hybridations entre espèces. Comment

la molécule d'ADN « étrangère » va-t-elle alors être incorporée par la cellule hôte, et quels facteurs favorisent son intégration fonctionnelle par cette cellule ?

Nous avons tenté de répondre à ces questions en utilisant des cellules de la levure de boulangerie *Saccharomyces cerevisiae*, un eucaryote unicellulaire modèle, dans le noyau desquelles nous avons transplanté des molécules d'ADN d'origine bactérienne [3].

La levure *S. cerevisiae*, dont le génome est réparti en 16 chromosomes, possède une remarquable tolérance à l'ADN étranger, ce qui en a fait un outil idéal pour l'intégration de longs fragments d'ADN exogène, comme ceux utilisés pour établir la cartographie physique du génome humain au cours des années 1990 [4], ou encore pour des approches de biologie synthétique [5]. La possibilité d'intégrer et de modifier des chromosomes complets de bactéries dans *S. cerevisiae* est connue depuis plus de 10 ans [6, 7]. Les séquences d'ADN bactériennes ont commencé à diverger de celles des eucaryotes il y a plus de 1,5 milliard d'années. Lorsque l'ADN bactérien est intégré dans le noyau de la levure sous la forme d'un chromosome surnuméraire, il représente alors près de 7 % du contenu en ADN de ces cellules.

Nous avons étudié deux souches modifiées de *S. cerevisiae*, chacune contenant un dix-septième chromosome préalablement linéarisé, d'origine bactérienne différente (Figure 1A) : *Mycoplasma pneumoniae*, dont le génome a un contenu en bases G

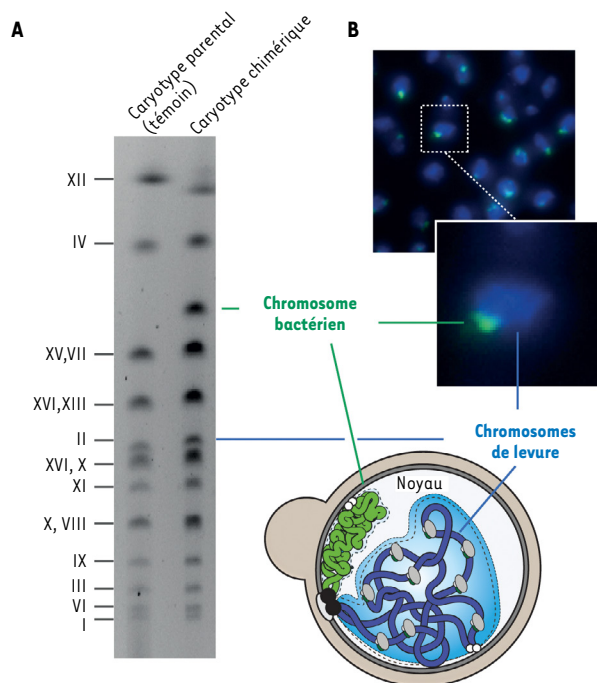


Figure 1. Intégration d'un chromosome bactérien dans le génome de la levure de boulangerie. A.

L'électrophorèse en champ électrique pulsé est une technique permettant de séparer de grandes molécules d'ADN en fonction de leur taille. Appliquée à la levure *S. cerevisiae*, cette technique permet ainsi de séparer les 16 chromosomes (même si certains migrent à des positions identiques dans le gel d'électrophorèse car ils ont presque la même taille). Dans le génome chimérique, on détecte une bande supplémentaire qui correspond à la taille du chromosome bactérien intégré.

B. Utilisation de la technique d'hybridation par fluorescence *in situ* (fluorescent *in situ* hybridization, FISH), qui permet de détecter des séquences spécifiques d'ADN par microscopie en ciblant par des sondes fluorescentes. L'ensemble de l'ADN est marqué en bleu par le fluorochrome DAPI. Le chromosome bactérien, quant à lui, est marqué non seulement en bleu car il est constitué d'ADN, mais également en vert par une sonde spécifique. Ce marquage différentiel permet de visualiser directement la position du chromosome bactérien dans l'espace nucléaire de la levure.

et C proche de celui de la levure (38 %), et *Mycoides mycoides*, dont le génome a un contenu en bases G et C inférieur (22 %). Une cartographie de l'organisation de ces chromosomes surnuméraires a montré que l'ADN bactérien de *M. pneumoniae* se structure de manière similaire aux chromosomes de levure. D'une part, il est enroulé et compacté autour d'histones en formant des nucléosomes, dont l'organisation est semblable à celle des nucléosomes des chromosomes de levure. D'autre part, le séquençage des ARN de ces levures a montré que l'ADN bactérien intégré dans la levure était spontanément transcrit en ARN dans le noyau, en suivant généralement l'orientation des gènes

bactériens sur le chromosome¹, suggérant ainsi l'existence d'un mécanisme d'orientation de la transcription précédant l'évènement de divergence entre eucaryotes et procaryotes. Au contraire, l'ADN bactérien de *M. mycoides*, dont la composition est pauvre en bases G et C, est transcriptionnellement inactif quand il est intégré dans un noyau de levure, bien qu'il soit lui aussi capable d'assembler des nucléosomes.

¹ La double hélice d'ADN peut être transcrite en ARN dans deux sens différents, selon le brin qui est « lu » par la machinerie de transcription. Dans un génome eucaryote ou procaryote, les gènes sont définis notamment par cette orientation de la transcription. Nous avons constaté que la transcription de l'ADN de *M. pneumoniae* intégré dans la levure se faisait généralement en respectant le sens de transcription des gènes bactériens disposés le long de la molécule.

En collaboration avec une autre équipe de recherche, nous avons pu localiser le chromosome de *M. mycoides* dans un compartiment distinct formant un domaine globulaire situé en périphérie du noyau (Figure 1B). Nous avons aussi utilisé une technique de capture de la conformation chromosomique (Hi-C), qui permet d'analyser la structure tridimensionnelle du génome, pour montrer que ce chromosome interagissait peu avec les autres chromosomes présents dans le noyau. Ainsi, ces chromosomes bactériens ont tous deux une organisation chromatiniennne, mais ont des propriétés transcriptionnelles et organisationnelles très différentes en fonction de leur contenu en bases G et C.

Nous nous sommes alors demandé comment des segments d'ADN bactérien et d'ADN de levure pourraient coexister sur un même chromosome. En utilisant la technique CRISPR-Cas9, nous avons fusionné des fragments d'ADN de levure et des fragments d'ADN bactérien de plusieurs centaines de kilobases afin d'obtenir un chromosome chimérique, composé en alternance d'ADN de levure et d'ADN bactérien. En utilisant la technique Hi-C, nous avons pu détecter une compartimentation séparant les segments d'ADN de levure (région transcriptionnellement active) et ceux d'ADN bactérien (région transcriptionnellement inactive), alors même que ces segments sont portés par le même chromosome. Cette compartimentation est d'autant plus étonnante que le génome de levure ne présente pas, dans des cellules normales, une telle partition entre des régions transcriptionnellement actives et inactives : tous les chromosomes sont transcrits, et sont mélangés dans l'espace nucléaire. Par contraste, cette compartimentation du chromosome chimérique est à rapprocher de celle des génomes de nombreux eucaryotes multicellulaires, qui sont segmentés en régions euchromatiniennes, transcriptionnellement actives, et hétérochromatiniennes, inactives (et dont la séquence nucléotidique est souvent constituée d'éléments répétés), régions qui se répartissent dans des compartiments différents du noyau lors de l'inter-



phase du cycle cellulaire [8]. Dans les cellules de levure contenant un chromosome chimérique, comme dans les cellules eucaryotes présentant des compartiments euchromatiniens et hétérochromatiniens, cette compartimentation est transitoirement perdue lors de la division cellulaire, quand les chromosomes sont compactés pour préparer leur ségrégation dans les cellules filles [9]. L'intégration, dans la levure, d'ADN exogène dont la séquence nucléotidique diffère fortement par son contenu en paires GC, produit donc spontanément des domaines chromatiniens dont l'organisation rappelle celle des génomes d'eucaryotes multicellulaires. Quels sont les mécanismes régissant le compartimentage de l'ADN de *M. mycoides* et de l'ADN de levure dans le chromosome chimérique ? En inhibant la transcription par l'ajout de thiolutine dans le milieu de culture, nous avons constaté la perte de la partition du chromosome chimérique entre les segments d'ADN qui étaient auparavant activement transcrits et les autres. La transcription joue donc un rôle essentiel dans l'organisation globale de ce chromosome. Cependant, chez les mammifères, la ségrégation entre euchromatine et hétérochromatine est principalement contrôlée par la formation d'hétérochromatine dite canonique via des modifications spécifiques des histones. Or, ce processus n'existe pas chez *S. cerevisiae*. Ainsi, la partition du chromosome chimérique, uniquement dépendante de l'activité transcriptionnelle, pourrait témoigner d'un processus ancestral de traitement de séquences exogènes très divergentes, avant l'apparition de mécanismes moléculaires plus complexes, comme celui de la formation de l'hétérochromatine canonique, dans l'évolution des génomes des eucaryotes. Enfin, en collaboration avec une autre équipe de recherche, nous avons montré que des modèles « d'apprentissage profond » (*deep learning*) entraînés sur des données expérimentales de levure étaient capables de prédire la composition en nucléosomes et l'accessibilité transcriptionnelle de l'ADN bactérien intégré dans cette levure. Ainsi, nous montrons que dans

un contexte biologique précis (étape du cycle cellulaire, état métabolique, etc.), le devenir fonctionnel d'une molécule d'ADN exogène peut être prédit simplement à partir de sa séquence nucléotidique. D'autres équipes de chercheurs ont tiré parti de la facilité avec laquelle on peut manipuler la levure *S. cerevisiae* pour y intégrer des fragments d'ADN synthétiques dont la séquence nucléotidique est aléatoire ou de l'ADN provenant d'autres organismes, et avaient ainsi pu montrer la transcription d'ADN exogènes avec des compositions en G et C relativement proches de celle du génome de *S. cerevisiae* [10-12]. L'originalité de notre stratégie de recherche est d'avoir exploré les conséquences de l'introduction d'un ADN exogène bactérien qui diffèrait fortement du génome de la levure par une proportion de paires de bases GC nettement moindre. La distinction que nous avons alors mise en évidence entre séquences transcriptionnellement actives (celles de l'ADN de levure, dont la proportion de GC est « normale ») et inactives (d'origine bactérienne, plus pauvres en GC) en fonction de leur composition en bases pourrait constituer un mécanisme primordial de sélection lors de l'acquisition de matériel génétique étranger, facilitant l'adaptation évolutive. Pour autant, l'ADN exogène intégré à la levure n'est pas immédiatement éliminé et s'avère relativement stable au cours des générations : il pourrait donc fournir un réservoir de séquences divergentes facilitant graduellement l'émergence de nouvelles fonctions au cours du temps. Par ailleurs, après l'intégration d'ADN dont la séquence nucléotidique est plus similaire au génome de la levure et donc plus à même de contenir des régions codantes d'intérêt, le fait que ces régions puissent alors être immédiatement transcrites pourrait là encore refléter un avantage évolutif. La différence des mécanismes de « sélection » des séquences exogènes à conserver ou à éradiquer chez les organismes eucaryotes plus complexes (par exemple les mammifères) pourrait expliquer l'existence de compartiments transcriptionnellement actifs ou inactifs dans ces organismes.

L'utilisation de modèles d'apprentissage profond pour prédire ces phénomènes constitue une avancée vers des applications concrètes. En comprenant comment un chromosome bactérien peut s'adapter à un environnement eucaryote, il devient envisageable de concevoir des séquences d'ADN optimisées ou même des génomes artificiels qui exploitent au mieux la machinerie cellulaire de l'hôte. Par exemple, en ajustant la composition en bases de certains gènes codant des protéines d'intérêt thérapeutique, il serait possible d'augmenter leur expression dans des cellules hôtes destinées à la production industrielle de biomolécules. ♦

When a eukaryotic cell adopts a bacterial chromosome

LIENS D'INTÉRÊT

Les auteurs déclarent ne pas avoir de lien d'intérêt.

RÉFÉRENCES

1. Bignaud A, Koszul R. La transcription, clé de voûte de l'architecture du chromosome bactérien. *Med Sci (Paris)* 2024 ; 40 : 412-3.
2. Mirny L, Dekker J. Mechanisms of chromosome folding and nuclear organization: Their interplay and open questions. *Cold Spring Harb Perspect Biol* 2022 ; 14 : a040147.
3. Meneu L, Chapard C, Serizay J, et al. Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus. *Science* 2025 ; 387 : eadm9466.
4. Schuler GD, Boguski MS, Stewart EA, et al. A gene map of the human genome. *Science* 1996 ; 274 : 540-6.
5. Koszul R. Beyond the bounds of evolution: Synthetic chromosomes... How and what for? *C R Biol* 2016 ; 339 : 324-8.
6. Lartigue C, Vashee S, Algire MA, et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* 2009 ; 325 : 1693-6.
7. Baby V, Labrousseau F, Brodeur J, et al. Cloning and transplantation of the *Mesoplasma florum* genome. *ACS Synth Biol* 2018 ; 7 : 209-17.
8. Lieberman-Aiden E, Van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009 ; 326 : 289-93.
9. Naumova N, Imakaev M, Fudenberg G, et al. Organization of the mitotic chromosome. *Science* 2013 ; 342 : 948-53.
10. Gvozdenov Z, Barcuatean Z, Struhl K. Functional analysis of a random-sequence chromosome reveals a high level and the molecular nature of transcriptional noise in yeast cells. *Mol Cell* 2023 ; 83 : 1786-97.e5.
11. Camellato BR, Brosh R, Ashe HJ, et al. Synthetic reversed sequences reveal default genomic states. *Nature* 2024 ; 628 : 373-80.
12. Luthra I, Jensen C, Chen XE, et al. Regulatory activity is the default DNA state in eukaryotes. *Nat Struct Mol Biol* 2024 ; 31 : 559-67.